Imperial College London

MSI-BASED MAPPING STRATEGIES IN TUMOUR-HETEROGENEITY

Thesis submitted by

Emine Kazanc

For the degree of Doctor of Philosophy (PhD)

Imperial College London

Department of Metabolism, Digestion and Reproduction

Faculty of Medicine

May 2022

Declaration for Originality

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

Emine Kazanc

Copyright Declaration

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

Acknowledgement

I would like to thank my supervisors, Prof Zoltan Takats, Prof Sadaf Ghaem-Maghami and Dr George Poulogiannis, for their guidance and opportunity being a member of their research team. I would like to thank Professor Takats and Dr Poulogiannis lab members, past and present, for their support.

I am grateful to Dr Dan McGill, Dr Paolo Inglese, Dr Harry Whitwell, Dr Ambreen Ashraf, Dr Syabira Yusoff, and Dr Sara Heuss for their thorough thesis evaluation. I am particularly grateful to Dr Evi Karali for her careful review of the thesis and for taking the time to discuss unclear points, and for her patience and generous help all the way through.

Dr Paolo Inglese deserves a special appreciation for all of his bioinformatics assistance, for establishing a pipeline for data processing, and for putting up with my never-ending queries until my PhD ended.

I would also like to thank Dr Chelsea Nikula, Dr Weiwei Zhou, Dr Teresa Murta, Dr Alan Race from the National Physical Laboratory, and Dr Najah Islam from the Barts Cancer Institute for providing instrumentation and expertise. I must thank Nona and Cristina from the Imperial College Tissue Bank. I would also thank Dr Caroline Sands and Dr Maria Gomez Romero from the Imperial College MRC-NIHR Phenome Centre for LC-MS expertise.

I also want to thank my major supporters in Turkey, and my family, for always believing in me and being there for me when I needed them. I am incredibly thankful to the Turkish Republic's Ministry of National Education for assisting me in accomplishing my dream of pursuing a PhD by sponsoring my education.

Finally, I would like to thank my chosen family members, my friends: Zeynep, Nuray, Sophie, Syabira, Katia, Sara, Semra, Yasin, Gulsah, and Amina, for their kind and cheerful companionship during my stay in England. A special thanks to Seyma Turkseven for being "my partner in crime" during this adventure.

Abstract

Since the early 2000s, considerable innovations in MS technology and associated gene sequencing systems have enabled the "-omics" revolution. The data collected from multiple omics research can be combined to gain a better understanding of cancer's biological activity. Breast and ovarian cancer are among the most common cancers worldwide in women. Despite significant advances in diagnosis, treatment, and subtype identification, breast cancer remains the world's second leading cause of cancer-related deaths in women, with ovarian cancer ranking fifth. Tumour heterogeneity is a significant hurdle in cancer patient prognosis, response to therapy, and metastasis. As such, heterogeneity is one of the most significant and clinically relevant areas of cancer research nowadays. Metabolic reprogramming is a hallmark of malignancy that has been widely acknowledged in recent literature. Metabolic heterogeneity in tumours poses a challenge in developing therapies that exploit metabolic vulnerabilities.

Consequently, it is crucial to approach tumour heterogeneity with an unlabeled yet spatially specific read-out of metabolic and genetic information. The advantage of DESI-MSI technology originates from its untargeted nature, which allows for the investigation of thousands of component distributions, at a micrometre scale, in a single experiment. Most notably, using a DESI-MSI clustering approach could potentially offer novel insights into metabolism, providing a method to characterise metabolically distinct sub-regions and subsequently delineate the underlying genetic drivers through genomic analyses.

Hence, in this study, we aim to map the inter-and intra-tumour metabolic heterogeneity in breast and ovarian cancer by integrating multimodal MSI-based mapping strategies, comprising DESI and MALDI, with IMC (Imaging Mass Cytometry) analysis of the tumour section, using CyTOF, and high-throughput genetic characterisation of metabolically-distinct regions by transcriptomics. The multimodal analysis workflow was initially performed using sequential breast cancer Patient-Derived Xenografts (PDX) models and was expanded on primary tumour sections. Moreover, a newly developed DESI-MSI friendly, hydroxypropyl-methylcellulose and polyvinylpyrrolidone

(HPMC/PVP) hydrogel-based embedding was successfully established to allow simultaneous preparation and analysis of numerous fresh frozen core-size biopsies in the same Tissue Microarray (TMA) block for the investigation of tumour heterogeneity. Additionally, a single section strategy was combined with DESI-MSI coupled to Laser Capture Microdissection (LCM) application to integrate gene expression analysis and Liquid Chromatography-Mass Spectrometry (LC-MS) on the same tissue segment. The developed single section methodology was then tested with multi-region collected ovarian tumours. DESI-MSI-guided spatial transcriptomics was performed for co-registration of different omics datasets on the same regions of interest (ROIs). This co-registration of various omics could unravel possible interactions between distinct metabolic profiles and specific genetic drivers that can lead to intra-tumour heterogeneity.

Linking all these findings from MSI-based or guided various strategies allows for a transition from a qualitative approach to a conceptual understanding of the architecture of multiple molecular networks responsible for cellular metabolism in tumour heterogeneity.

Table of Contents

1 Introduction 16

1.1 Breast Cancer 16

1.2 Ovarian Cancer 17

1.3 Cancer Hallmark: Altered Oncogenic Signalling Pathways 18

1.3.1 PI3K signalling pathway 19

1.4 Lipid Hallmark: Altered Lipid Metabolism 21

1.5 Intra and Inter Tumour Heterogeneity (ITH) 26

1.6 Pre-clinical PDX models 28

1.7 Mass Spectrometry in Cancer Research 29

1.7.1 Mass Spectrometry 30

1.7.1.1 Ionisation Techniques 30

1.7.1.2 Mass Analyser 34

1.7.1.3 Detector 36

1.7.2 MS-based Platforms for Metabolite Profiling 37

1.7.2.1 MALDI-MSI 39

1.7.2.2 DESI-MSI 41

1.7.2.3 Imaging Mass Cytometry 42

1.7.2.4 LC-MS 43

1.8 Mapping Strategies of Tumour Heterogeneity in a Multi-modal Analysis 45

1.8.1 MSI-based mapping and spatial clustering analysis for Lipidomics 46

1.8.2 Complementary Spatial Transcriptomics to Map Alterations with MSI 47

1.8.3 Laser Capture microdissection system for MSI- guided sample preparation 49

1.8.4 Histological techniques 49

1.8.5 Embedding strategies for MSI 51

1.8.6 Data Analysis 52

1.8.6.1 Pre-processing data 52

1.8.6.2 Multivariate statistical analysis 53

1.8.6.2.1 Unsupervised classification 53

1.8.6.2.2 Supervised classification 54

1.8.6.3 Univariate analysis 56

1.8.6.4 Segmentation Analysis 57

2 Experimental Plan and Methods 60

2.1 Clinical Data 60

2.1.1 Ovarian Biopsies 60

2.1.2 Breast Biopsies 60

2.1.3 PDX models 61

2.2 A Modified Fast Snap-Frozen Standard Operating Procedure (SOP) for tissue sample collection 62

2.3 Sample Preparation 63

2.4 Imaging Mass Cytometry (IMC) Analysis 64

2.5 Mass Spectroscopy Imaging (MALDI-MSI) settings and Clustering Analysis 66

2.6 Desorption Electrospray Ionization Mass Spectroscopy Imaging (DESI-MSI) 66

2.6.1 DESI instrument setup for a fast sample run 66

2.6.2 DESI-MSI settings for standard samples run 67

2.6.3 Cold stage set up 68

2.7 Post-DESI-MSI histological staining ... 69
2.7.1 Modified RNase-free staining Protocols ... 70

2.8 LC-MS Sample Preparation ... 72
2.8.1 Lipid Extraction of the microdissected ROIs ... 72
2.8.2 The chromatographic gradient ... 72
2.8.2.1 Data analysis ... 73
2.8.2.2 Identification and annotation of metabolites ... 74

2.9 Gene Expression Analysis of Single Section Methodology ... 74
2.9.1 RNA isolation and Quality Assessment ... 74
2.9.2 qPCR ... 75

2.10 Laser Capture Microdissection ... 76

2.11 RNA sequencing sample preparation and Analysis ... 77
2.11.1 Smart Kit Amplification ... 77
2.11.1.1 Library preparation for transcriptome sequencing ... 78
2.11.1.1.1 Clustering and sequencing ... 79
2.11.1.2 Data Analysis Quality control ... 79
2.11.1.2.1 Error Rate ... 79
2.11.1.2.2 A/T/G/C Content Distribution ... 81
2.11.1.2.3 Data Filtering ... 82
2.11.1.3 Mapping to reference genome ... 83
2.11.1.3.1 Mapped Regions in Reference Genome ... 86
2.11.1.3.2 Visualisation of Mapping Status of Reads ... 87
2.11.1.3.3 Novel gene prediction ... 88
2.11.1.4 Quantification ... 88
2.11.1.5 Comparison between Gene Expression Levels ... 89
2.11.1.6 RNA-Seq Correlation ... 89
2.11.1.7 Differential expression analysis ... 89
2.11.1.7.1 List of Differentially Expressed genes ... 90
2.11.1.7.2 Screening of differentially expressed genes ... 91
2.11.1.8 Cluster Analysis of Gene Expression Differences ... 91
2.11.1.9 Enrichment analysis ... 92
2.11.1.9.1 GO enrichment analysis ... 92
2.11.1.9.2 KEGG Pathway enrichment analysis ... 92
2.11.1.10 Mutation analysis ... 93
2.11.1.11 Summary Workflow of RNA sequencing ... 93

2.12 The tissue embedding protocol ... 95
2.12.1.1 Chemical Materials ... 95
2.12.1.2 Animal Models ... 96
2.12.1.3 Embedding of biopsies ... 96
2.12.1.4 Tissue Sectioning ... 97
2.12.1.5 RNA extraction and analysis ... 98
2.12.1.6 cDNA preparation and RT-qPCR ... 99

2.13 Data Analysis ... 100
2.13.1 DESI-MSI ... 100
2.13.1.1 DESI-MSI Pre-processing of the datasets ... 100
2.13.1.2 Normalisation ... 101
2.13.1.3 Multivariate Statistical Analysis ... 101
2.13.1.3.1 Principal component Analysis ... 101
2.13.1.4 Univariate statistical methods ... 102
2.13.1.5 Clustering Analysis ... 103
2.13.1.6 MetaboAnalyst for statistical analysis of DESI-MSI ... 103
2.13.1.7 DESI-MSI Metabolite Identification ... 105
2.13.1.7.1 Data Input ... 106

2.13.1.7.2 Compound Matching Table 106
2.13.1.7.3 Mummichog Output 106
2.13.1.7.4 Mummichog Module analysis..... 107
2.13.1.7.5 Mummichog Pathway Analysis 109

3 Multi-modal mapping Strategy to reveal Tumour heterogeneity with Patient-Derived Xenografts (PDXs) models 110

3.1 Background 110

3.2 Aim and objectives 115

3.3 Summary Workflow of Experimental Plan 116

3.4 Results and Discussion 117

3.4.1 A pilot study of investigation metabolic heterogeneity in wild-type and mutant *PIK3CA*, PDX models by DESI-MSI analysis 117
3.4.2 Multi-modal spatial clustering strategy to reveal metabolic heterogeneity in a pair of *PIK3CA* mutant and wild-type PDX models by DESI-MSI..... 120
3.4.3 Comparison of DESI-MSI with MALDI-MSI for clustering analysis to reveal metabolic heterogeneity..... 125
3.4.4 DESI-MSI guided transcriptomics from microdissected Metabolic Clusters..... 128
3.4.4.1 Overview of data production quality and screening of differentially expressed genes 128
3.4.5 Correlation of Transcriptomics and metabolomics dataset to reveal driver pathways behind metabolic clusters defined by DESI-MSI 135
3.4.5.1 Comparison of Mutant *PIK3CA* PDX with WT PDX model and metabolic segments in Network Analysis 135
3.4.5.2 Comparison of microdissected MSI-guided ROIs in clusters of Mutant *PIK3CA* and Wild-type PDX model in Network Analysis..... 146

3.5 Discussion and Conclusions 157

3.6 Appendix 159

4 A novel embedding strategy for investigating tumour heterogeneity in breast tumour core biopsies 165

4.1 Background 165

4.2 Aim and Objectives..... 167

4.3 Results and Discussion 168

4.3.1 DESI-MSI Clustering Analysis on embedded fragile patient biopsies 168
4.3.2 The Impact of the Embedding Hydrogel on DNA and RNA Quality 170
4.3.3 Transcriptomics Analysis of the HPMC/PVP hydrogel embedded core biopsy..... 174
4.3.4 Correlation of Transcriptomics and metabolomics behind the DESI- MSI cluster of embedded core biopsy 176
4.3.4.1 Metabolite set enrichment analysis..... 176
4.3.4.2 Pathway Analysis..... 177
4.3.4.3 Network Explorer analysis..... 178

4.4 Discussion and Conclusion 183

5 Optimisation of a Multi-modal Analysis Strategy using a Single Fresh Frozen Tissue Section on a Technical Level 184

5.1 Background 184

5.2 Aim and Objectives..... 187

5.3 Summary Workflow of Experimental Plan 188

5.4 Results and Discussion 189
5.4.1 DESI-MSI set-up optimisation steps using a single tissue section for follow-up gene expression analysis 189
5.4.1.1 Evaluation of PEN-coated versus standard glass slides analysis 189
5.4.1.2 Evaluation of fast scan DESI-MSI setup using PEN-coated glass slides 191
5.4.1.3 Investigation of Cold Temperature DESI-MSI Stage's effect on the Imaging analyses 194
5.4.1.3.1 Comparison of the PEN-coated glass slides versus Standard Glass Slides for both scan rates in Cold Stage DESI- MSI Analysis 198
5.4.1.3.2 Comparison of PEN-coated glass slides with both scan rates on the Cold versus Room Temperature Stage of the DESI-MSI 199
5.4.2 Gene Expression Analyses for the post-DESI slides 201
5.4.2.1 Optimisations of the RNase Free Staining Protocol for the post-DESI slides ... 202
5.4.2.2 Extracted RNA Quality Assessment and RT-PCR from post-DESI microdissected samples 205
5.4.2.3 Extracted RNA Quality Assessment and RT-PCR from post-DESI microdissected samples under cold stage acquisition 207
5.4.3 LC-MS Analysis for the post-DESI slides 211
5.4.3.1 DESI-MSI clustering analysis to reveal metabolic segments for microdissection 212
5.4.3.2 Comparison of the IPA and MTBE-based sample preparation protocol for microdissected ROIs 216

5.5 Discussion and Conclusion 222

5.6 Appendix 225

6 The mapping of tumour heterogeneity in multi-site collected biopsies of ovarian tumours using single tissue section methodology 230

6.1 Background 230

6.2 Aim and Objectives 232

6.3 Summary workflow of Experimental Plan 233

6.4 Results 233
6.4.1 A pilot study for RNA sequencing optimisation steps 233
6.4.1.1 DESI-MSI clustering analysis to reveal heterogeneity in multi-site collected biopsies of multiple ovarian tumours 233
6.4.1.2 A complete Transcriptomics from microdissected ROIs of post-DESI tissue.... 236
6.4.2 Investigation of intra-tumour heterogeneity in multi-site collected biopsies of the same ovary 238
6.4.2.1 DESI-MSI clustering analysis to reveal metabolic intra-tumour heterogeneity in multi-site collected biopsies 238
6.4.2.2 Transcriptomics analysis to reveal intra-tumour heterogeneity in multi-site collected biopsies 241
6.4.2.3 Comparison of Cluster 2 (blue) metabolic segments in Network Analysis..... 241
6.4.2.4 Comparison of Cluster 3 (green) metabolic segments in Network Analysis ... 248
6.4.2.5 Comparison of Cluster 4 (purple) metabolic segments in Network Analysis .. 254

6.5 Discussion and Conclusion 260

6.6 Appendix 262

7 Final Remarks 266

7.1 The Dilemma of Tumor Heterogeneity in Breast and Ovarian Cancer 266

7.2 DESI-MS Imaging Technology in Multi-modal Analysis for Tumor-Heterogeneity Mapping 266

7.3 Research Impact **271**
7.4 Challenges and Future Work **271**
8 References ***273***
9 List of conference presentations and publications ***290***
10 Copyright and Permissions ***292***

Abbreviations

(9-AA)	9-Aminoacridine
(ACN)	Acetonitrile
(ADH1B)	Alcohol Dehydrogenase 1B (Class I), Beta Polypeptide
(AP)	Alkaline Phosphatase
(DART)	Direct Analysis in Real Time
(AA)	Arachidonic Acid
(BSA)	Bovine Serum Albumin
(BC)	Breast Cancer
(CA-125)	Carbohydrate Antigen-125
(CEL)	Carboxyl Ester Lipase
(Cams)	Cell Adhesion Molecules
(CI)	Chemical Ionisation
(cDNA)	Complementary DNA
(CN)	Copy Number
(CHCA)	Cyano-4- Hydroxycinnamic Acid
(CyTOF)	Cytometry by Time of Flight
(CPLA2)	Cytosolic Phospholipase A2
(DHB)	Dihydroxybenzoic Acid
(DNA)	Deoxyribonucleic Acid
(DESI)	Desorption Electrospray Ionisation
DAG)	Diacylglycerol
(DEGs)	Differentially Expressed Genes
(DGE)	Digital Gene Expression
(DC)	Direct Current
(DA)	Discriminant Analysis
(EASI)	Easy Ambient Sonic Spray Ionisation
(EI)	Electron Impact
(ETC)	Electron Transport Chain
(ESI)	Electrospray Ionisation
(ENCODE)	Encyclopedia of DNA Elements
(EOC)	Endometrioid Ovarian Cancer
(FDR)	False Discovery Rate
(FAB)	Fast Atom Bombardment
(FA)	Fatty Acid
(FD)	Field Desorption
(PC1)	First Principal Component
(FISH)	Fluorescence In-Situ Hybridization

(FC)	Fold Change
(FFPE)	Formalin-Fixed Paraffin-Embedded
(FT-ICR)	Fourier Transform Ion Cyclotron Resonance
(FT-Orbitrap)	Fourier Transform Orbitrap
(GC)	Gas Chromatography
(GO)	Gene Ontology
(GATK)	Genome Analysis Toolkit
(GAPDH)	Glyceraldehyde-3-Phosphate Dehydrogenase
(GPI)	Glycosylphosphatidylinositol
(GSEA)	Gene Set Enrichment Analysis
(H&E)	Haematoxylin-Eosin
(HexCer)	Hexosylceramides
(HCA)	Hierarchical Clustering Analysis
(HPLC)	High-Performance Liquid Chromatography
(HGSOC)	High-Grade Serous Ovarian Cancer
(HDI)	High-Definition Imaging
(HRP)	Horseradish Peroxidase
(HER2)	Human Epidermal Growth Factor Receptor 2
(Hodes)	Hydroperoxy Derivatives
(HpODEs)	Hydroperoxy-Octadecadienoic Acid
(HPMC)	Hydroxypropyl-Methylcellulose
(HIF-1)	Hypoxia-Inducible Factor-1
(IMC)	Imaging Mass Cytometry
(IF)	Immunofluorescence
(IHC)	Immunohistochemical
(ISH)	In Situ Hybridization
(IGV)	Integrative Genomics Viewer
(ITH)	Inter -/Intra-Tumour Heterogeneity
(IPA)	Isopropanol
(KO)	KEGG Orthologs
(KEGG)	Kyoto Encyclopedia of Genes and Genomes
(LacCer)	Lactosylceramide
(LAESI)	Laser Ablation Electrospray Ionisation
(LCM)	Laser Capture Microdissection
(LDA)	Linear Discriminant Analysis
(LC)	Liquid Chromatography
(LC-MS)	Liquid Chromatography-Mass Spectrometry
(LPIN 1)	Phosphatidate phosphatase-1 (lipin-1)

(LPA)	Lysophosphatidic Acid
(LPA)	Lysophosphatidic Acid
(LPC)	Lysophosphatidylcholine
(*m/z*)	Mass Over Charge
(MSI)	Mass Spectrometry Imaging
(MS)	Mass Spectrometry
(MALDI)	Matrix-Assisted Laser Desorption Ionisation
(MMC-DA)	Maximum Margin Criterion Discriminant Analysis
(MSEA)	Metabolite Set Enrichment Analysis
(MEOH)	Methanol
(MTBE)	Methyl Tert-Butyl Ether
(MAPK)	Mitogen-Activated Protein Kinase
(MG)	Monoacylglycerol
(GM3)	Monosialodihexosylganglioside
(MT)	Mutant
(DHB)	2,5-Dihydroxybenzoic Acid
(NGS)	Next-Generation Sequencing
(NF-Kappa B)	Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B Cells
(NMR)	Nuclear Magnetic Resonance
(ER)	Oestrogen Receptor
(OCT)	Optimal Cutting Temperature Medium
(OC)	Ovarian Cancer
(ORA)	Over Representation Analysis
(OXPHOS)	Oxidative Phosphorylation
(PDTCs)	Patient-Derived Tumour Cells
(PDX)	Patient-Derived Tumour Xenograft
(PFP)	Pentafluorophenyl
(PPIB)	Peptidyl-Prolyl Cis/Trans Isomerase B
(PTEN)	Phosphatase and Tensin Homolog
(PA)	Phosphatidic Acid
(Pcs)	Phosphatidylcholines
(Pes)	Phosphatidylethanolamines
(PI)	Phosphatidylinositol
(PIP2)	Phosphatidylinositol Biphosphate
(PIP3),	Phosphatidylinositol Trisphosphate
(PI3K)	Phosphoinositide 3-Kinase
(PD)	Plasma Desorption
(PVP)	Polyvinylpyrrolidone

(HPMA)	Poly(N-(2-Hydroxypropyl)-methacrylamide)
(PEN)	Polyethylene Naphthalate
(PCR)	Polymerase Chain Reaction
(PCA)	Principal Component Analysis
(PQN)	Probabilistic Quotient Normalisation
(PR)	Progesterone Receptor
(QC)	Quality Control
(RF)	Radio Frequency
(RPKM)	Reads Per Kilobase Of Exon Model Per Million Mapped Reads
(RT-qPCR)	Real-Time Quantitative-Polymerase Chain Reaction
(RMMC-DA)	Recursive Maximum Margin Criterion
(RABT)	Reference Annotation Based Transcript
(ROI)	Region-Of-Interest
(RT-PCR)	Reverse Transcription PCR
(RP)	Reversed-Phase
(RNA)	Ribonucleic Acid
(RIN)	RNA Integrity Number
(RNA- ISH)	RNA-In-Situ Hybridization
(SIMS)	Secondary Ion Mass Spectrometry
(SOM)	Self-Organizing Map
(Na-CMC)	Sodium-Carboxy-Methyl Cellulose
(CMC)	Carboxymethylcellulose
(S1P)	Sphingosine-1-Phosphate
(SOP)	Standard Operating Procedure
(HMDB)	The Human Metabolome Database
(ANOVA)	The One-Way Analysis of Variance
(TOF)	Time of Flight
(TMA)	Tissue Microarray
(TIC)	Total Ion Count
(TP53)	Tumor Protein P53
(TAG)	Triacylglycerol
(TFA)	Trifluoroacetic Acid
(TNBC)	Triple Negative Breast Cancer
(TBS)	Tris-Buffered Saline
(CT)	Ultrasound/Computed Tomography
(WT)	Wild Type

1 Introduction

1.1 Breast Cancer

More than 300,000 new cancer cases in the United Kingdom are diagnosed yearly, resulting in approximately 440 mortalities per day[1].

Breast cancer (BC) is the most common type of cancer among the women population in the world. Patients with breast cancer have a higher survival rate than those with other types of cancer[2]. Treatment outcomes differ considerably from one patient to another, as they are affected by various changes such as hormonal, genetic, and environmental factors[3].

Diagnosis methods for breast cancer can be made through self-examination, and tests and procedures are undertaken clinically, followed by mammography. It is also immunohistochemical (IHC) screening, which is based on oestrogen receptor (ER), and progesterone receptor (PR) absent/presence in breast tissue. These receptors identify a specific subtype of breast cancer in the clinic[4]. Breast cancer can be classified into different subtypes: the expression of the ER, PR and the human epidermal growth factor receptor 2 (HER2/ErbB2). Breast cancer subtype luminal (luminal A or B) refers to breast cancers which are positive for ER and PR. HER2 subtype includes breast cancer overexpressing HER2. The basal-like subtype has breast cancers that do not express ER, PR and HER2; this type is called triple-negative breast cancer (TNBC)[5].

For treatment, hormone receptors (ER+/PR+) can be treated with drugs such as tamoxifen, blocking the effect of oestrogen on tissue receptors. Patients generally have a better prognosis than breast cancer patients positive for HER2 amplification[6,7]. Breast cancer HER2- positive patients are treated with agents that block HER2, such as trastuzumab[8]. TNBC has a poor prognosis, and patients are treated mainly with chemotherapy because no approved targeted therapies are available[6]. However, the gene expression profiling techniques such as the Oncotype DX™, the MammaPrint, Mammostrat and PAM50 improve the prediction of the clinical outcome of the targeted therapy for breast cancer[8].

1.2 Ovarian Cancer

Ovarian cancer (OC) is the 5th most common and the most lethal gynecologic cancer among women in the world[9,10]. Approximately 90% of women with ovarian cancer survive for five or more years after being diagnosed at an early stage, while less than 30% of women are diagnosed at an advanced stage[11]. Diagnosis involves medical imaging such as ultrasound/computed tomography (CT) scan and blood tests examining levels of several biomarkers, including carbohydrate antigen-125 (CA-125). Once the preliminary investigation shows the presence of the tumour, histological diagnosis is required for the best diagnosis and treatment options based on the tumour. Although this traditional gold standard diagnosis has proven to be a simple and powerful practice, there is often inter-observer variability in grading the tumour resulting in poor prognosis[12]. In some cases, a further diagnosis such as immunohistochemistry and genetic screening may be required[13].

Approximately 90% of ovarian cancers are epithelial; the rare types are called germ cell tumours, stromal tumours and sarcomas. In addition, there are various types of epithelial cancers, including serous, endometrioid, clear cell, mucinous and unclassifiable based on their histopathology. Based on biomarker expression, the serous subtype has been divided into five subtypes, including high-grade serous and low-grade serous. High-grade serous ovarian cancer (HGSOC) is the most commonly diagnosed subtype of serous epithelial ovarian cancer (70%), followed by endometrioid (10%) and clear cell (10%), while mucinous (3%) and low-grade serous carcinomas (<5%) count for less than 10% of the cases together[14]. HGSOC have been identified as four molecular subtypes C1 (High stromal response), C2 (immunoreactive), C4 (low stromal response), and C5 (mesenchymal) using gene expression arrays, and the C1 subtype has the worst survival prognosis[15–17]. HGSOC is usually diagnosed in elderly patients at a late stage. Morphologically are composed of ciliated, columnar cells that form papillae, solid masses or slit-lake spaces with high-grade atypia. Immunohistochemistry is positive for CK7, PAX8, and WT1 but negative for CK20[18]. In current clinics, the main treatment for ovarian cancer is surgery, followed by several platinum-based chemotropic agents[11].

Despite significant advances in identifying subtypes of ovarian and breast cancer, several scientific and clinical problems remain, including early diagnosis, accurate characterisation and personalised management of the diseases. Because cancer is not a single disease but a highly heterogeneous and complex collection of different pathologies[18,19], the tumour-heterogeneity can explain the reasons for having different prognoses in similar tumours[3].
Further research and new multi-modal approaches involving better diagnosis are required to deal with this clinical tumour-heterogeneity challenge to decrease the burden on the traditional pathological diagnosis[20].

1.3 Cancer Hallmark: Altered Oncogenic Signalling Pathways

The development of the 'cancer niche' or 'tumour microenvironment has been reported as an important factor contributing to tumourigenesis[21,22]. Cancer cells have uncontrolled cell proliferation with the ability to escape apoptosis and metastatic behaviour. One of the critical alterations in cancer cells is the increased glycolysis rate under aerobic conditions, also named the 'Warburg effect.
Several signalling pathways have been associated with ovarian and breast cancers, including the Notch signalling pathway, the mitogen-activated protein kinase (MAPK) pathway, the phosphoinositide 3-kinase (PI3K)/AKT pathway, the lysophosphatidic acid (LPA) pathway, the Mullerian inhibitory substance receptor pathway, the ErbB activation pathway, the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kappa B) pathway, the proto-oncogene tyrosine-protein kinase Src pathway, the JAK/STAT3 pathway and the oestrogen receptors pathway[9]. Additionally, there is a direct relationship between the Akt/PI3K/mTOR pathway in the phospholipid metabolism and phosphatidic acid pathways[23,24]. There are associations with STAT-3-inducing cytokines, GFβ/BMP-triggered pathways, secretory macrophage-derived phospholipase PLA2 G7, its product arachidonic acid (AA) and signalling pathways controlled by the AA metabolites PGE2, PGI2, and LTB4[8,25,26].

1.3.1 PI3K signalling pathway

Phosphoinositide 3-kinases (PI3Ks) are a lipid kinase family that phosphorylates the 3′-hydroxyl group of phosphoinositides. PIP3, a critical second messenger in AKT for activation of growth, proliferation, and survival signalling is the most well-characterized outcome of this reaction[27,28]. The tumour suppressor PTEN dephosphorylation adversely controls PIP3. PIP3 production is a specific function of the class I subclass of PI3Ks, subdivided into classes IA and IB. Only the class IA PI3Ks have been linked to human cancer thus far. Still, frequent genetic mutations in these enzymes and their pathway effectors have made the PI3K pathway one of the most commonly dysregulated pathways in cancer. A subject to considerable has now been explored to identify pan-specific and isoform-specific PI3K inhibitors for cancer therapy.

Class IA PI3Ks are heterodimers with a regulatory subunit (p85, p55, p50, p85, p55) and a catalytic subunit (p110, p110, p110) that are induced downstream of receptor tyrosine kinases (RTKs)[29]. Activating mutations in PIK3CA, the gene encoding the PI3K p110 catalytic subunit, have previously been reported as advancements for triggering oncogenic PI3K activity[30]. These mutations comprise PIK3CA amplification, PTEN loss, AKT mutations, and RTK amplification in a group of common genomic abnormalities that promote carcinogenesis by activating the PI3K/AKT signalling pathway[27], **Figure 1.3.1.1.**

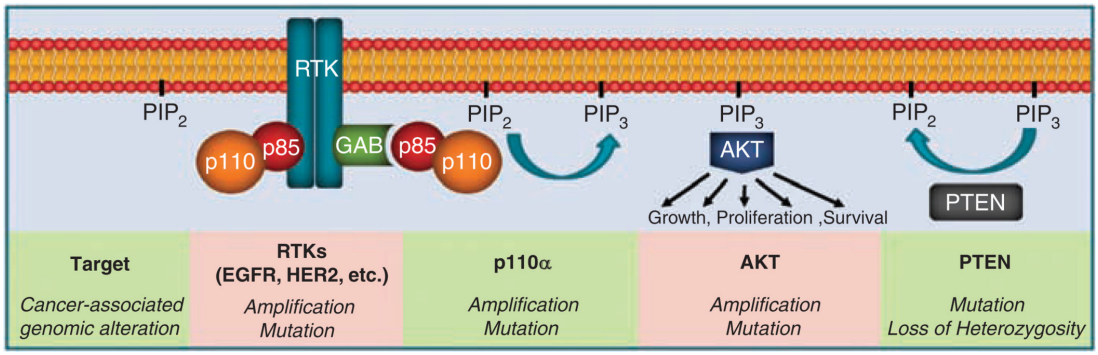


Figure 1.3.1.1. Common genomic abnormalities in the PI3K signalling pathway. RTK activation recruits PI3K directly or through adaptor proteins such as the GAB proteins. PI3K phosphorylates PIP2 to produce PIP3, which activates AKT, including several oncogenes that regulate crucial cellular activities in cancer cells. PTEN

controls this process adversely by dephosphorylating PIP3. In cancer, all critical components of this signalling pathway are typically altered. PI3K stands for phosphoinositide 3-kinase; RTKs stand for receptor tyrosine kinases. Reproduced with permission [27].

In human cancers, several different PIK3CA activating mutations have been described. H1047R and E542K/E545K are the two most prominent mutation "hotspots", **Figure 1.3.1.2.**[28] The H1047R mutation increases the kinase domain's interaction with membranes and eliminates the need for interaction with Ras.[31] However, the inhibitory interaction with the N-SH2 domains of the regulatory subunits is impaired by the mutations E542K and E545K[32]. Other, less frequent PIK3CA mutations (e.g., N345K, C420R) interrupt the C2 domain's interaction with iSH2. Other class I PI3K genes with tumour-associated modifications are incredibly uncommon. On the other hand, C-terminal truncation and deletion mutants that disrupt part of the iSH2 domains of regulatory subunits (encoded by PIK3R1 and PIK3R2) are C-terminal truncations and deletion mutants that disrupt carcinogenically and typically exist in the brain and endometrial malignancies. The activation of the p110 catalytic isoform is required for transformation by this regulatory subunit variants7. Activating mutations in PIK3CD encoding p110 (E1021K, E525K, N334) have been reported in roughly 100 individuals globally[33], with unique variants affecting additional domains identified[34].

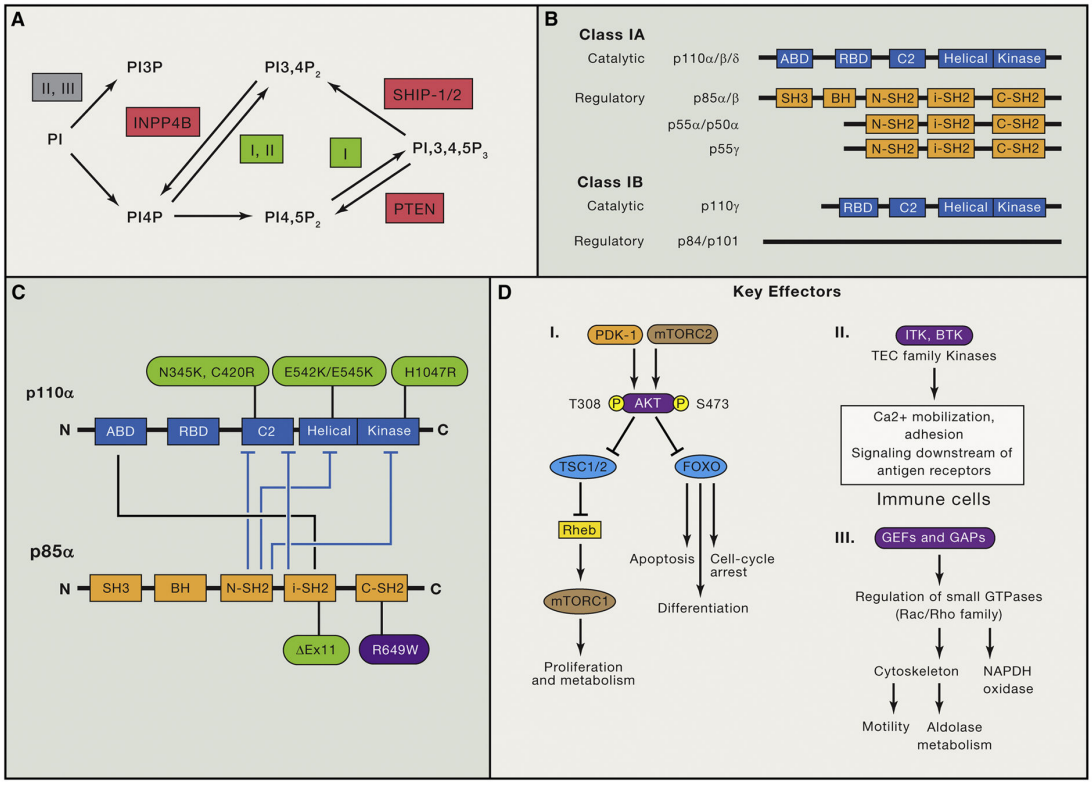


Figure 1.3.1.1. General mechanism of the PI3K Pathway in Cancer, a) Diagram of the essential synthesis and degradation mechanisms for PtdIns-3-P (PI3P), PtdIns-3,4-P2, and PtdIns-3,4,5-P3. The PI3K classes that induce activities (I, II, or III) are specified. The lipid phosphatases are highlighted in red. INPP4B is a type II inositol polyphosphate-4-phosphatase. b) Domain structure of the catalytic and regulatory subunits of class I PI3K. ABD stands for the adaptor-binding domain, RBD stands for the Ras-binding domain, and BH stands for breakpoint cluster region homology. c) Illustration of intramolecular connections between class IA catalytic and regulatory subunits (p110 and p85 are widely characterized examples). The ABD's tight binding to iSH2 imparts stability to p110. Blue-block arrows show other interactions that reduce baseline activity and are eased when regulatory subunits bind to pTyr. Green represents cancer-related activating mutations. The SHORT syndrome mutation in p85 (R649W) is highlighted in purple. d) A synopsis of critical PI3K effectors: PDK-1, AKT, TEC family kinases, and small GTPase GEFs/GAPs. It is unknown which GEFs cause PI3K-dependent Rac activation to increase motility and aldolase release. Reproduced with permission[28].

1.4 Lipid Hallmark: Altered Lipid Metabolism

One of the most important hallmarks of cancer is metabolic reprogramming, including uncontrolled cell growth and the production of high-energy and cell biomolecules. Most diseases related to the metabolic, immune and neurological systems, particularly cancer, are a result of changes in lipid metabolism. The essential role of lipids is coordinating membrane

homeostasis, regulatory and acting as first or second messengers in signalling transduction and energy storage in normal healthy cells, **Figure 1.4.1**.[19,35,36].

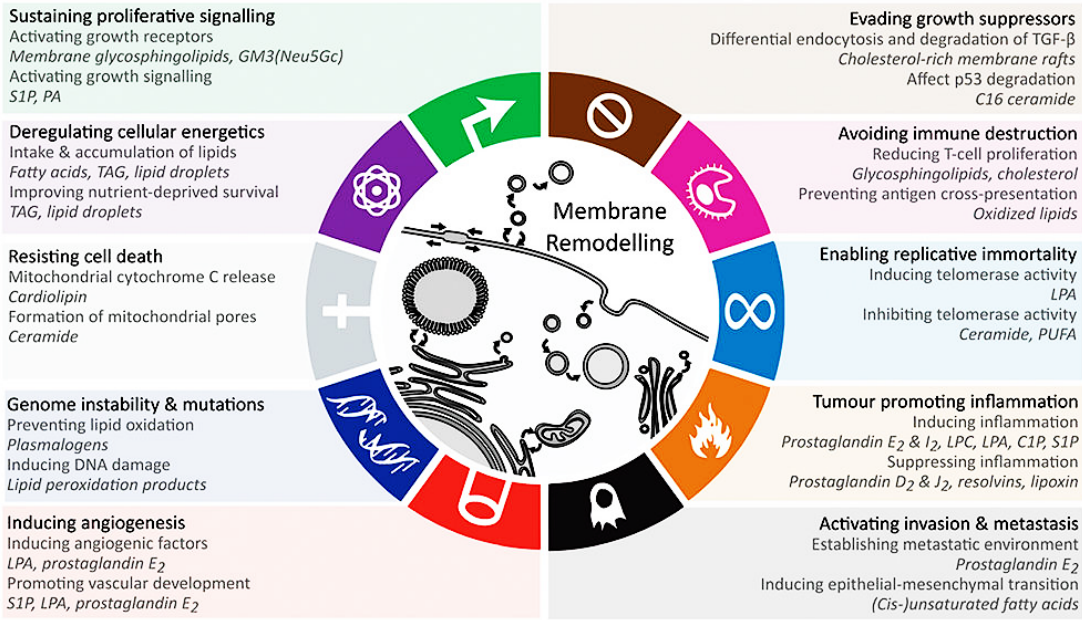


Figure 1.4.1. Lipid-mediated pathways enhance cancer's hallmarks. The lipid-related pathways are displayed alongside the relevant lipids in italics for each characteristic. At the centre, membrane remodelling for subcellular compartments is represented, where altered membrane composition via metabolism and trafficking permits many hallmarks. Abbreviations: GM3 – monosialodihexosylganglioside, LPA – lysophosphatidic acid, PA – phosphatidic acid, PI – phosphatidylinositol, PIP2 – phosphatidylinositol biphosphate, PIP3 – phosphatidylinositol trisphosphate, S1P – sphingosine-1-phosphate, TAG – triacylglycerol, Reproduced with permission[37].

Lipids are a large class of the total human metabolites occupying 70% of the Human Metabolome Database and already described over 37,500 lipids in the LIPID MAPS Structure Database[38,39]. A comprehensive classification of lipids, divided into eight subcategories, was contributed by Fahy et al., 2005[38] and updated in 2009[39] including lipids from non-mammalian sources such as plants, bacteria, and fungi. In eukaryotic and prokaryotic organisms, lipid subcategories have known as fatty acyls, glycerolipids, glycerophospholipids or phospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids, and polyketides, **Figure1.4.2**. [38,39]

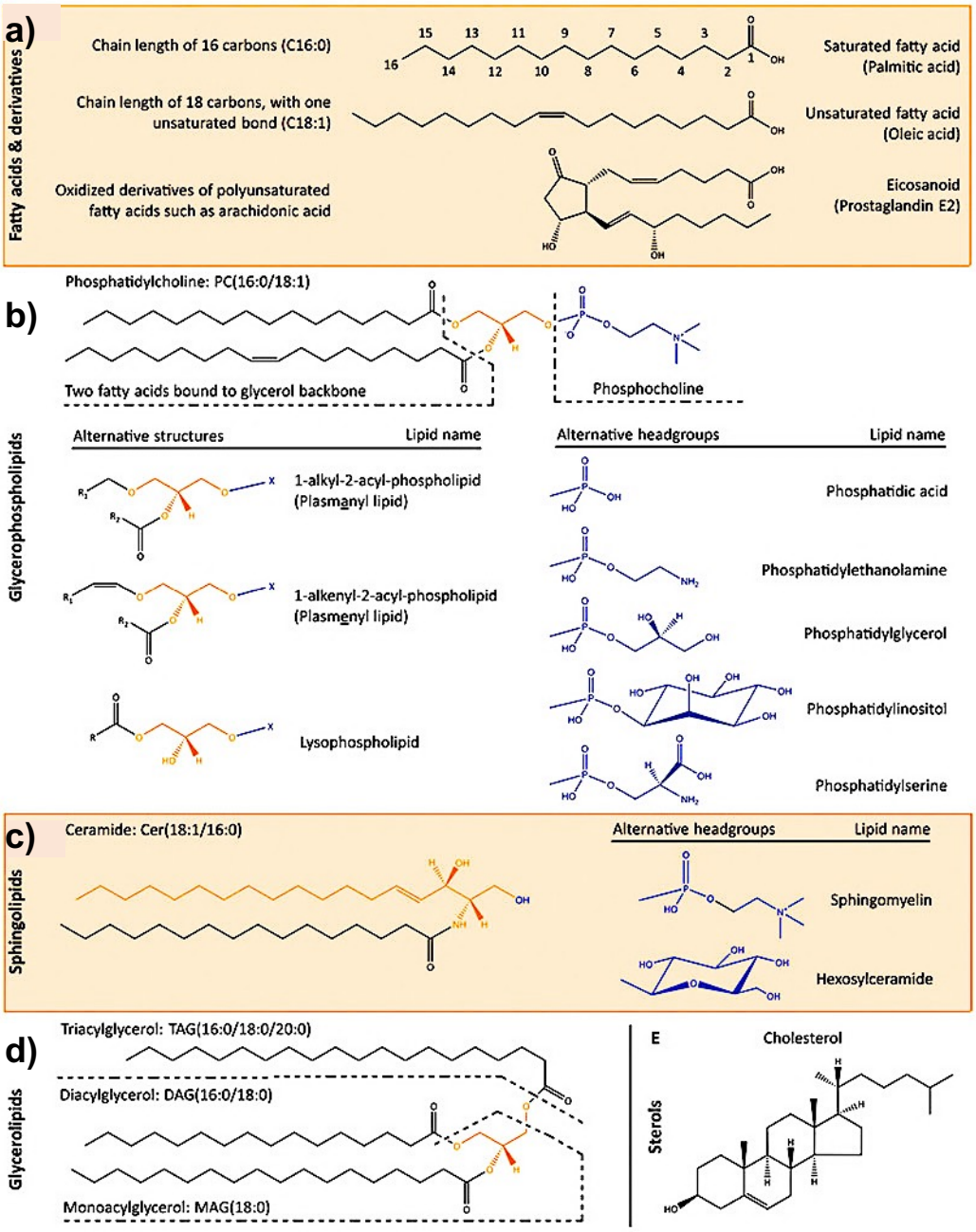


Figure 1.4.2. The most common types of cellular lipids include fatty acids, glycerophospholipids, sphingolipids, glycerolipids, and sterols. a) Fatty acids and derivatives are made up of a carboxyl group and a hydrocarbon chain with various lengths and unsaturated bonds, which affects the three-dimensional structure of lipids and, as a result, their functions and interactions. Lipid nomenclature represents the number of carbons and unsaturated bonds in the molecule's fatty acid chains. b) Glycerophospholipids have an orange glycerol backbone, a blue lipid headgroup, and at least one fatty acid or fatty alcohol chain. In the alternate structures, R denotes a fatty acid or fatty alcohol group, whereas X denotes the headgroup. c) Glycerophospholipids are fundamental components of cell membranes. Sphingolipids, as linked to glycerol, have a sphingoid backbone (orange), such as sphingosine or sphinganine. Sphingolipids are essential for membrane fluidity and cell signalling. Ceramides are diffusible signalling lipids with a

sphingoid backbone and a fatty acid chain but no polar headgroup. Sphingomyelins are similar to ceramides in structure but also feature a phosphocholine headgroup. The glycosphingolipids are structurally unique sphingolipids consisting of a ceramide molecule linked to a sugar residue. A glycosphingolipid containing a single hexose, such as glucose or galactose, is known as glucosylceramide or galactosylceramide. Gangliosides are glycosphingolipids with more complicated sugar structures and a sialic acid residue, of which GM2, GM3, GD2 and GD3 are well-known examples. d) Glycerolipids have a glycerol backbone with at least one fatty acid chain but no headgroup. They often serve as essential intermediates in glycerophospholipid production or lipid storage molecules. e) Sterols are a kind of steroid molecule that is attached to a hydroxyl group. Rather than relying on the head group or fatty acid chains for activity, cholesterol's distinctive rigid steroid structure disturbs the connection between fatty acid chains to allow for different biophysical activities. Furthermore, sterols' hydroxyl groups can be esterified, resulting in sterol esters like cholesteryl esters, often stored in lipid droplets. Reproduced with permission[37].

Lipid composition varies between organelle membranes, cell types, tissue types and between different disease grades. The lipid metabolism undergoes drastic alterations in tumorigenesis, essential in building new cell membranes, energy production, and regulating signalling cascades for cancer cell migration, invasion, metastasis and angiogenesis[40]. Metabolic alterations in lipids (i.e., exceptionally high abundance of phosphatidylcholines (PCs) and phosphatidylethanolamines (PEs) in the tumour tissues) have been found diversely in many cancer types[41]. Therefore, a deeper understanding of the identification of lipid compositions in normal and cancer cells might enlighten lipids functions significantly in tumourigenesis[42].

One of the critical alterations in cancer cells is the increased glycolysis rate under aerobic conditions, also named as ‘Warburg effect’, which plays a role in fatty acid (FA) synthesis for the cancer cells' growth. Under aerobic conditions, glycolysis of normal mammalian cells is the conversion of glucose to pyruvate and then to lactic acid. In cancer cells, some of the pyruvates are converted to Acetyl-coenzyme A, which takes a role in the prerequisite for *de novo* FA synthesis. The abnormal FA synthesis incorporates phospholipids in particular membrane lipids to promote cancer cell proliferation[43–46].

The alterations in lipid metabolism increased fatty acid β-oxidation, resulting in high levels of carnitine proteins, glycerolipids (i.e. monoacylglycerol (MG) and DAG), phospholipids (i.e. PCs, Pes, PGs, and lysophosphatidylcholine (LPC)), sphingolipids (i.e. SM, Cer, Glc series and Globo series), cholesterols

and alterations of long chain fatty acids such as a high abundance of palmitate, oleate and myristate in serum and tissue samples of the patients, **Figure 1.4.3**. [25,26,47–50].

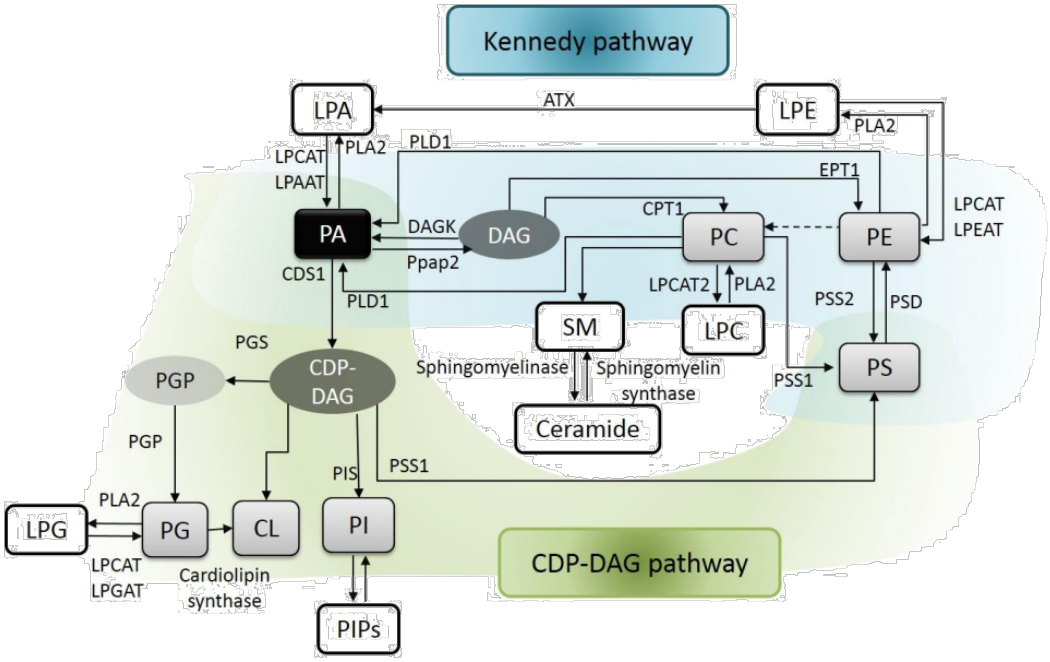


Figure 1.4.3. Several lipid signalling pathways, such as; Kennedy and CDP-DAG pathways. DAGK- diacylglycerol kinase, PLD- phospholipase D, PLA2- phospholipase A2, LPCAT- lysophosphatidylcholine acyltransferase, ATX- autotoxin, LPAAT- lysophosphatidic acid acyltransferase, LPAGT- lysophosphatidylglycerol acyltransferase, LPEAT- lysophosphatidylethanolamine acyltransferase, PGP- phosphoglycolate phosphatase, PIS- Phosphatidylinositol synthase, PSS- phosphatidylserine synthase, PSD- phosphatidylserine decarboxylases, CPT- Carnitine palmitoyl transferase, EPT- diacylglycerol ethanolamine phosphotransferase, LPA- lysophosphatidic acid, PA- phosphatidic acid, LPE- lysophosphatidylethanolamine, PE- phosphatidylethanolamine, DAG- diacylglycerol, PC- phosphatidylcholine, LPC- lysophosphatidylcholine, SM- Sphingomyelin, CDP-DAG- cytidine diphosphate diacylglycerol, PGP- phosphatidylglycerol phosphate, PI- phosphatidylinositol, PG- phosphatidylglycerol, LPG- lysophosphatidylglycerol, PIPs- phosphatidylinositol phosphate, note: broken arrows indicate more than one step.[35]

In ovarian cancer, high levels of lysophosphatidic acid (LPA) have been found in plasma samples, which has a vital role in metastases behaviour. Mainly, HGSOC were linked to abnormal lipid metabolism and increased amino acid with energy disorders [25,26,49,50].

In breast cancer, a high level of unsaturated fatty acids in the phosphatidylethanolamine (PE) and phosphatidylcholine (PC) has been found

in tissues compared to that of the healthy tissues correlating to tumourigenesis[47]. Phosphoinositides (PIs), which inactivate by phosphatase and tensin homolog (PTEN), have been associated with breast cancer progression mediating survival and migratory signals[48].

Studies have targeted lipids as a cancer therapy such as; inhibitor agents against lipogenic enzymes such as Fatty acid synthase (FASN) and ATP Citrate Lyase (ACLY), disrupting lipid-mediated tumour-stroma crosstalk and membrane cholesterol homeostasis inducing cell death signalling with modulated lipid raft components[36,51].

All these studies demonstrated the usefulness of metabolome/lipidome profiling for identifying ovarian and breast cancer as benign or malignant[13,52].

1.5 Intra and Inter Tumour Heterogeneity (ITH)

Tumour heterogeneity is one of the major clinical challenges in cancer therapy and even can be the leading cause of failed treatments and chemo-resistance. This problem stems from the fact that tumour tissue is not homogenous. These tumours have distinct morphological and phenotypical profiles, including histopathology, gene expression, proliferation and metastatic behaviour and, in particular different metabolic activities. This phenomenon occurs between different patient tumours named inter tumour- heterogeneity and within the same patient tumour named intra tumour-heterogeneity, **Figure 1.5. *1***.[18,53–57].

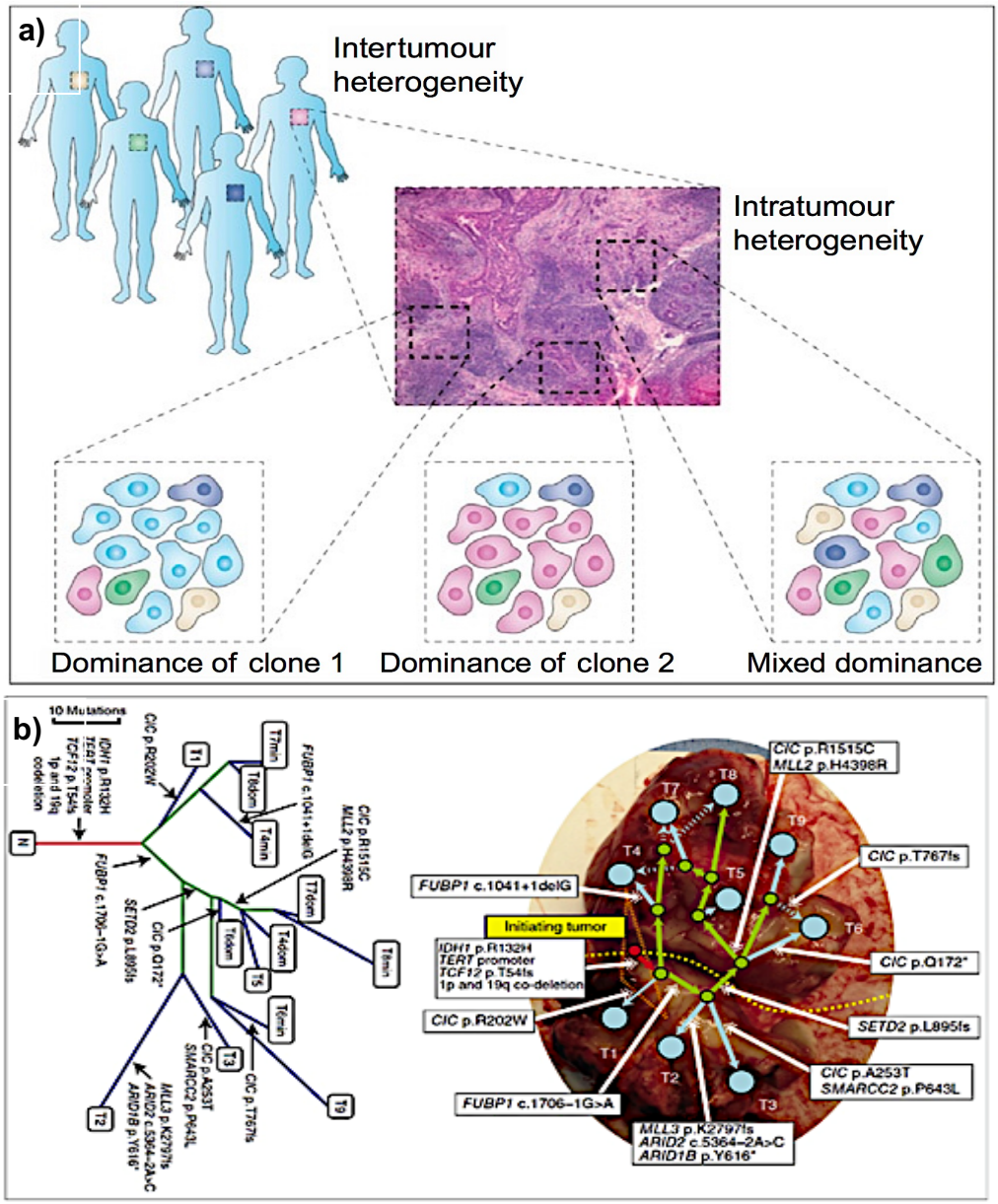


Figure 1.5. 1. Inter and intra tumour-heterogeneity. a) Clonal evaluation in inter and intra-tumour heterogeneity[55].b) The impact of a mutation by the spatial-temporal dynamics surrounding the process of clonal evaluation and mapping the phylogenetic evaluation of driver and parallel mutations into a spatial context, Reproduced with permission[54].

There is increasing evidence that not every metabolic alteration can be found in every cancer cell, even within the same tumour sample and across patients. The metabolic alterations caused by the expression of spatially mutated oncogenic genes, specific enzyme isoforms, transcription factors, lipids, and other metabolites, play an essential role in intra tumour heterogeneity within patients and inter tumour heterogeneity between patients[18].

Ovarian and breast cancers are a group of neoplasms with distinct clinicopathological, molecular, and prognostic features. Even though there are numerous cancer subtypes, they are regarded as a single disease. Some cells exhibit metastatic potential within a tumour by acquiring additional mutations and metabolic alterations that vary spatially from one site to another. They develop independently of the primary clone and undergo clonal evaluation during tumorigenesis. Tumours are thought to originate from a single mutated cell that accumulates additional mutations as it grows and divides, resulting in subpopulations with different mutations. According to the literature[58] some, sub-clones have an evolutionary advantage for being resistant to chemotherapy and dominating the tumour, resulting in a specific clonal expansion. This sub-clone phenomenon demonstrates that a single biopsy performed in a standard histopathology laboratory routine does not represent the entire tumour [58].

Several studies have shown that OC and BC have a high degree of ITH, which present a major challenge in developing effective therapy strategies[18,59,60]. A comprehensive ITH study could help clinicians predict which OC and BC patients could show the best prognosis after chemotherapy[3,16,61]. Thus, mapping the OC and BC to see whether tumour heterogeneity can explain the development of chemotherapy resistance, be useful in obtaining patient-specific prognostic markers, and even improve personalised, targeted therapy. High throughput 'omics' technologies have aided in investigating the complexity of the OC and BC, providing comprehensive knowledge of oncogenic pathways and potentially targetable metabolites. The classical histology classification of subtypes should not be neglected in terms of diagnosis but should be used to improve treatment strategies[18].

1.6 Pre-clinical PDX models

Patient-derived tumour xenograft (PDX) models have recently been a surge in interest in the development and characterisation of cancer research. The donor tumour's main histological and genetic characteristics are mostly retained in PDX models, which remain stable across passages[62,63]. PDX models, made up of immunodeficient mice engrafted with cancerous patient

material, provide more faithful representations than cancer cell lines, which tend to diverge over time[64]. These models have gained important platforms for cancer research regarding the predictability of clinical outcomes. They are used for preclinical drug testing, biomarker characterisation, biological studies, and personalised medicine approaches[65]. Their use in drug discovery allows for the pre-clinical evaluation of therapeutic agents and the exploration of novel drug combinations. Bruna et al.[66,67] in 2016 showed that molecularly characterised PDTXs and their matched original PDTX-derived tumour cells (PDTCs) retain this heterogeneity in short-term culture. Therefore, these pre-clinical models can be used as a platform for cancer drug screening with the potential to uncover molecular mechanisms of therapy response[66].

Investigation of molecular mechanisms such as the phosphoinositide 3-kinase (PI3K) pathway has been shown to play an important role in BC. Tumours with high *PIK3CA* copy number (CN) have more aggressive prognostic features, such as large tumour size, high tumour grade, and negative Hormone Receptor (HR) status. They are more likely to occur in patients with HR and HER2 negative disease[64]. Somatic mutations in the *PIK3CA* gene, which codes for the class IA PI3K p110 subunit, are the most common activating mutations, occurring in 30–50% of ER+/HER2 early BC and 28% of metastatic diseases. Many studies have been conducted to determine the prognostic significance of *PIK3CA* mutations in primary BC. In roughly half of the tumours, *PIK3CA* CN gain co-occurs with *PIK3CA* mutations[64,68]. Insights into changes in the PI3K-AKT-mTOR pathways would provide potentially promising targets and new metabolic vulnerabilities for BC management.

1.7 Mass Spectrometry in Cancer Research

Mass spectrometry (MS) is widely used to perform *in-situ* molecular analyses of cancerous cells. This analysis could identify tissues, classify primary tumour tissues based on chemo response and metastatic status, and identify diagnostic/prognostic markers with drug response rates and resistance. Understanding the abundance of peptides, proteins, lipids, glycans, and drug metabolites in cancerous tissues and cells can significantly improve disease diagnosis, staging, and treatment[69–73].

1.7.1 Mass Spectrometry

Mass spectrometry is an analytical technique that identifies the molecular components of a sample by using electromagnetic fields to separate vapour ions based on their mass over charge (*m/z*) ratio. MS has proven to be an effective scientific tool, capable of detecting thousands of analytes from complex samples in a single analysis. Similarly to the genomics field, the application of MS to complex biological matrices has transformed the disciplines of proteomics, metabolomics, and lipidomics over the last two or three decades.

Lipidomics is still a developing field compared to more established areas such as genomics and proteomics[74]. MS-based lipidomics presents a tough challenge because measuring the entire lipidome of a tissue or cell in a single experiment is difficult. This challenge is due to the lipid molecules' strong chemical and structural similarity and the approximately seven orders of the magnitude concentration range of physiologically significant lipids. MS-related sensitivity and specificity have been critical driving reasons behind the success of MS-based lipidomics research. However, some analytical issues remain that MS alone cannot handle satisfactorily. These include the structural characterisation of complex lipids due to difficult-to-determine double bond topologies that cannot be readily allocated based on MS analysis alone.

Historically, lipids have been analysed by GC-MS after selective extraction, hydrolysis, and chemical derivatisation[75,76]. The first reports on LC-MS-based lipid mixture analysis were published in the early 1990s[77] and were based on soft ionisation techniques, such as electrospray ionisation (ESI). MS is made up of three basic components: an ionisation source, a mass analyser, and a detector.

1.7.1.1 Ionisation Techniques

The use of MS involves the utilisation of gas-phase ions. This gas-phase is a barrier for the majority of biologically relevant molecular entities that are non-volatile in nature. Furthermore, the creation of gas phase ions is often desirable without requiring any changes to the molecular structure.

Electron impact (EI) ionisation was one of the first techniques established for the ionisation of organic compounds. EI entails the collision of molecules with a regulated kinetic energy electron beam. Because EI is a gas phase ionisation process, it is only appropriate for ionising volatile substances[78.] The procedure is referred to as "hard-ionisation" because an electron is stripped from the molecules. This ionisation results in positive radical ions with extremely high internal energy, leaving the molecule's electron-deficient and extremely fragmented. To circumvent the difficulty of detecting intact molecule ions, "softer ionisation" methods were devised, resulting in chemical ionisation (CI) creation. CI employs a high-pressure EI source to generate protonated or deprotonated ions of a reagent gas (Methane (CH4), or Ammonia [NH3]) that transfer charge to analyte molecules by protonation, deprotonation, or electron transfer. Although it creates less fragmentation than EI, it is still best suited to volatile and thermally stable chemicals. New desorption ionisation methods were developed to circumvent these restrictions and allow the study of non-volatile, thermally labile biological molecules (such as DNA, peptides, lipids, and proteins). Desorption ionisation has been used in secondary ion mass spectrometry (SIMS) since the 1940s[79]. This approach allows for analysing secondary ions produced by irradiating a sample surface with an intense primary ion beam. One of the key problems of early (dynamic) SIMS techniques was that "sputtering" destroyed the surface under inquiry. Modern SIMS devices, on the other hand, can ionise rather big physiologically important compounds with molecular weights of up to 2 kDa employing cluster ion beans69. The first attempts to ionise non-volatile and labile compounds using surface desorption were field desorption (FD) and plasma desorption (PD). However, the narrow mass range (2000-3000 Da) of FD and the radioactive nature of PD prohibited these techniques from being widely applied and commercialised. Liquid SIMS, also known as fast atom bombardment (FAB), was a significant breakthrough in investigating large thermally labile biomolecules, allowing compounds with masses greater than 10000 Da to be studied.

At the end of the 1980s, two new important soft ionisation techniques were introduced, quickly becoming the most extensively utilised for MS applications: electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation (MALDI). In ESI, an analyte solution, such as proteins and peptides, is pumped through a capillary, an electrical potential difference between the capillary tip and the surrounding atmosphere is applied, and the high voltage induces the breakdown of gas molecules into electrons and positive ions[80] . As seen in **Figure 1.7.1.1.1**., the high voltage causes spraying of the solution from the so-called "Taylor cone"

as it departs the capillary[81]. Highly charged small droplets with an excess of positive or negative charges form. The droplet shrinks, increasing surface charge density and causing the droplet to blow apart (Coulombic fission/explosion). These can further disintegrate until only individual analyte ions remain ("charged residue model") or discharge charged analyte molecules from their surface ("ion evaporation model"). Depending on the chemical composition and size of the analyte molecules, single or doubly charged ions can be formed[82]. The crucial step was to dissolve the organic chemical in a liquid matrix that was generally inert/involatile (glycerol). This liquid matrix meant that the analytes of interest were protected from the harmful effects of the primary high-energy Ar+ or Cs+ beam. The sample surface was constantly refilled, resulting in ion beams of intact molecules. One of the most significant shortcomings of FAB was that the matrix produced its signal across the whole mass range even in the absence of sample-matrix interactions.

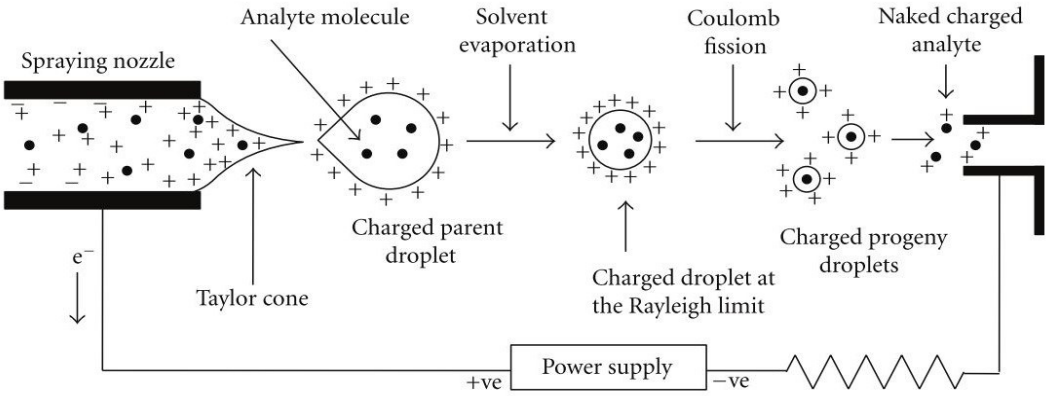


Figure 1.7.1.1.1. The electrospray ionisation process[81].

In the case of MALDI, the analyte is combined with a low molecular weight organic chemical known as the matrix, which is typically an aromatic carboxylic acid such as -cyano-4- hydroxycinnamic acid (CHCA) or 2,5-dihydroxybenzoic acid (DHB), and the mixture is deposited onto a target plate. A short pulsed laser beam irradiates the target plate, which is situated in a high vacuum source zone (most commonly a UV laser, i.e. N2 laser operating at 337 nm)[83]. As demonstrated in **Figure 1.7.1.1.2**, the analyte and matrix are combined and then placed on a target plate. A short-pulsed laser beam is employed in a high vacuum source region to irradiate the mixture. The matrix

absorbs a substantial amount of energy from the beam, resulting in its volatilisation and thermal degradation to gaseous products. It then leads to the desorption of analyte ions embedded in its crystal lattice. The high-intensity laser pulse excites the matrix, causing simultaneous desorption and ionisation of the sample, resulting in primarily singly charged ions[84]. Fragmentation is minimal to non-existent because the internal energy of the analyte molecules does not rise.

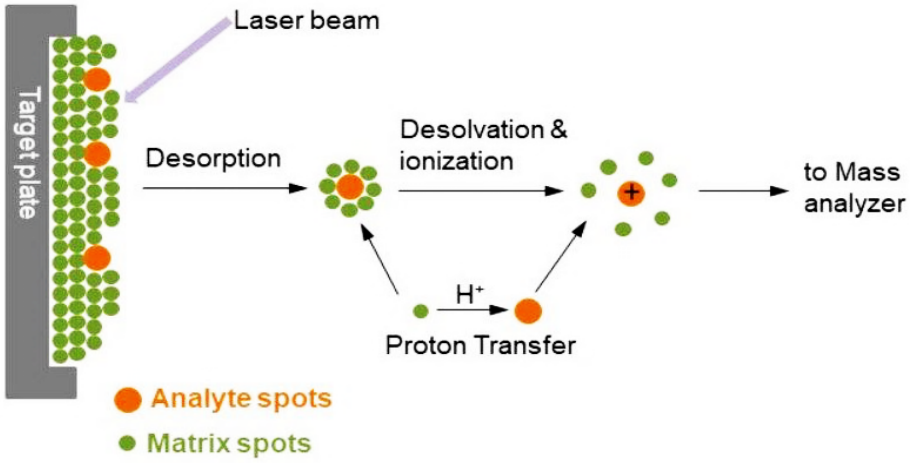


Figure 1.7.1.1.2. The ionisation process in MALDI[84].

Ambient ionisation technologies, which allow ionisation under air pressure with minimal or no sample preparation, have been developed over the last 15 years. While MALDI and SIMS ionisation is generally performed under vacuum, electrospray ionisation is carried out at ambient, with just the mass analysis conducted under vacuum. Zoltan Takats and his colleagues 2004 released a paper that DESI was offered one of the first ambient ionisation procedures[85]. As shown in **Figure 1.7.1.1.3**, the approach involves spraying an aqueous solvent onto a solid support housing the sample with a pneumatically assisted electrospray, resulting in charged secondary droplets. As a result, gaseous ions are created in a manner similar to ESI and transmitted to the mass spectrometer via the vacuum ion optics interface at atmospheric pressure. In contrast to ESI, a beam of mainly charged solvent droplets is directed onto the sample surface, extracting analyte molecules that are desorbed in secondary droplets and sucked into the mass spectrometer,

where they undergo desolvation and the production of gas phase molecular ion species.

More than 30 different ambient ionisation methods had been described in the literature by 2009[86]. DESI, on the other hand, is one of the most widely utilised ambient ionisation techniques due to its adaptability[87] and electrospray-like ionisation properties. Plasma-based ambient direct analysis in real-time (DART)[88], easy ambient sonic spray ionisation (EASI)[89], and laser ablation electrospray ionisation (LAESI)[90] are some of the other approaches that are often utilised.

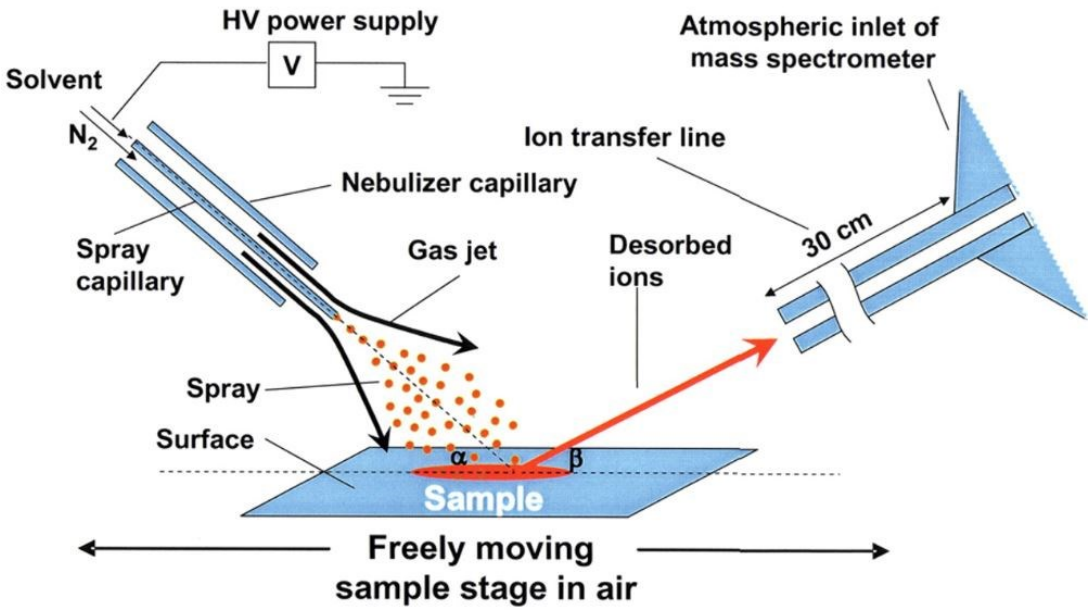


Figure 1.7.1.1.3.The ionisation process in desorption electrospray ionisation (DESI). An electrospray with pneumatic assistance is directed towards the sample surface. Desorbed molecules from the sample surface are protonated or deprotonated before being transported to the mass spectrometer via a capillary[85].

1.7.1.2 Mass Analyser

Ions must be separated based on their *m/z* after they have been created in the gas phase. For ion separation, mass spectrometers are engineered to utilise static or dynamic electric or magnetic fields, either alone or in combination. Quadrupoles, ion traps, time of flight (TOF) instruments, Fourier transform ion cyclotron resonance (FT-ICR), and Fourier transforms Orbitrap equipment are the most often used mass analysers for the measurement of metabolites and

lipids. Each type of mass analyser has distinct characteristics, summarised in **Table 1.7.1.2.1**[91].

One of the most prevalent types of mass analysers is the quadrupole analyser. It comprises four parallel rods subjected to a constant radio frequency (RF) and direct current (DC) voltage. The voltage amplitude dictates which range of *m/z* values have stable trajectories in the quadrupole, transferred from the ion source to the detector without colliding with the quadrupole filter's surface. The others follow an unsteady path and collide with the rods. Quadrupoles are generally inexpensive, simple to run, and simple to maintain. They, however, have a low resolution and a limited *m/z* range of roughly 2000-3000[92].

ToF analyser monitors how long it takes for accelerated ions to reach the detector[93]. The ions pass through a flight tube after being accelerated with a pre-set voltage. They are segregated based on their velocities once they enter this field-free region. In an ideal situation, all ions would have the same kinetic energy, but lighter ions would travel faster towards the detector than heavier ions due to their varied masses. A reflectron is an energy-correcting device found in more advanced ToF sensors. Following one flight distance, the ions enter an electrostatic mirror (i.e., the reflectron), which reverses their velocity and reaccelerates the ions back into the flight tube and down another field-free section. The extra distance traversed compensates for minor changes in velocities between ions with the same *m/z*.

Fourier transform ion cyclotron resonance (FT-ICR), and Fourier transforms Orbitrap (FT-Orbitrap) are the most remarkable mass resolution analysers. Ions are trapped in a chamber by the Lorentzian force generated by a strong magnetic field in FT-ICR. Ions orbit at an angular frequency called the ion cyclotron frequency, which is inversely proportional to the *m/z* ratio. The principle of FT-Orbitrap is identical to that of FT-ICR, except that an electric field substitutes the magnetic field. It comprises two electrodes, one on the outside with a barrel shape and one on the inside with a spindle shape. A constant electric potential is supplied between the two electrodes for ion trapping. Ion packets with different *m/z* ratios carry out axial oscillations at different frequencies[91].

Table 1.7.1.2.1. The most often used mass analysers in lipidomics and their properties[91].

Analyser type	Principle of separation	Mass range (*m/z*) limit	Mass resolution	Mass accuracy (ppm)
Quadrupole (Q)	*m/z* (trajectory stability)	2000	2000	100
Reflectron Time-of-Flight (ToF)	Velocity (flight time)	10000	20000	10
Fourier transform ion cyclotron resonance (FT-ICR)	*m/z* (resonance frequency)	30000	1000000	1
Fourier transform Orbitrap (FT-Orbitrap)	*m/z* (resonance frequency)	50000	150000	<5

1.7.1.3 Detector

There are many different types of detectors, but most works by providing an electronic signal based on the charge or current created when an ion strikes a surface. This ion strike is transformed into a graph that shows the relative current created by ions with varying *m/z* ratios. The detector records the abundance of each m/z value after the ions have been filtered through the mass analyser. The Faraday cup, for example, is a standard electrical detector in which the measured ion current is precisely proportional to the quantity of ions and charges per ion. It is a basic, low-cost, tough, and sturdy detector. Another type of detector is an electrical multiplier detector, which uses the

principle of secondary electron emission to increase the strength of the generated signal[91].

1.7.2 MS-based Platforms for Metabolite Profiling

In the mid-1990s, MS entered the cell biology field as its potential in identifying and quantifying targets[94]. Mass spectrometry imaging (MSI) has acquired a lot of traction in recent years. It has emerged as an indispensable method for seeing the spatial distribution of medicines, metabolites, peptides, and even proteins across various sample surfaces such as tissue sections without any prior labelling[95,96]. This function enables MSI to extract detailed biochemical information from the sample's numerous areas. A typical MSI experiment probes the sample surface (i.e., tissue slices) with an ionisation beam, which is then utilised to gather mass spectra from specific spots at defined x, and y coordinates. Each sample point has its mass spectrum. Software is used to recreate an image from the collective dataset, with each pixel including a mass spectrum, allowing mapping of individual compound distribution. Traditional immunohistochemistry (IHC) has been widely used to explore the localisation and distribution of biomarkers in malignant tumour cell diagnosis. MSI, like IHC, can be utilised for diagnosis to examine the identification and allocation of biochemical markers in cancer cells.

Ionisation techniques have a variety of applications, spatial resolution requirements, and sample preparation needs, all of which have been the topic of numerous reviews. MALDI, SIMS, and DESI are the three most often utilised ionisation techniques for MSI. **Figure 1.7.2. 1** depicts the strengths and weaknesses of each ionisation process[97].

SIMS is the most sensitive surface analysis technique, with detection limits ranging from parts per million to parts per billion, allowing research of single cells, subcellular organelles, and cell communities. On the other hand, SIMS experiments must be carried out in a high vacuum environment[98]. MALDI is the most often used imaging technology because it allows for studying various chemicals over a wide mass range. However, it depends on excellent preparation (coating sample with matrix), and sample preparation steps might

affect the spatial distribution and degradation of analytes during an analytical run.

DESI imaging has grown in popularity as an imaging technique over the last decade since it is a soft ionisation technique that can be done in ambient circumstances with minimal sample preparation steps. Although the spatial resolution of DESI (usually around 100 µm) is lower than that of SIMS and MALDI, it eliminates the constraints of high vacuum in SIMS imaging and sample preparation in MALDI imaging. Furthermore, because it is less invasive, it is appropriate for direct tissue analysis.

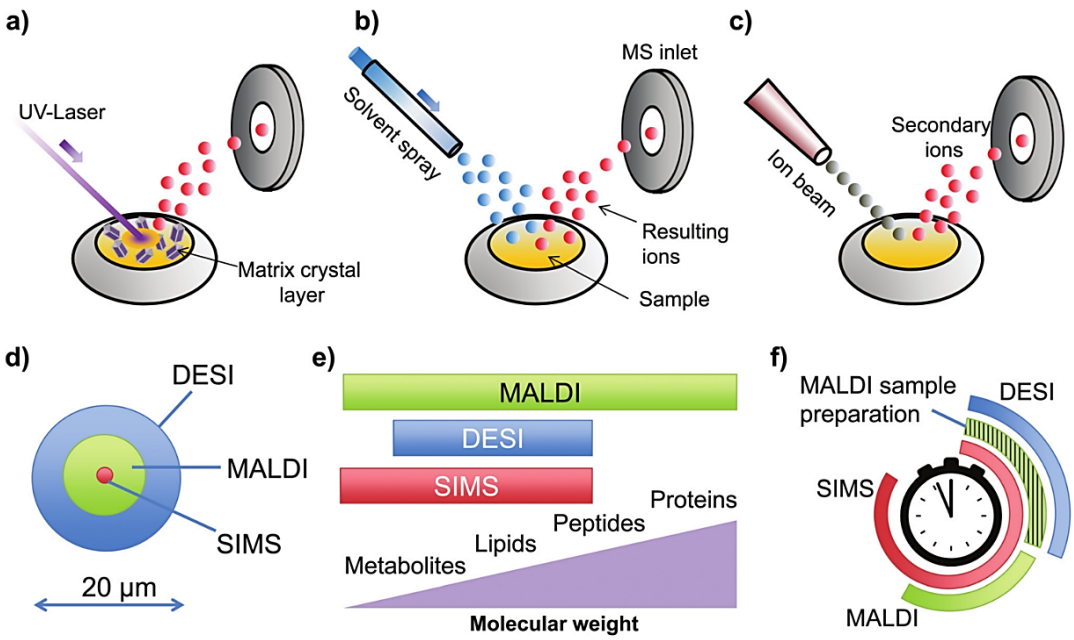


Figure 1.7.2. 1. Comparison of the most widely utilised MSI technologies. a) MALDI-MSI principle, b) DESI-MSI principle, c) SIMS-MSI concept, d) SIMS-MSI has the highest spatial resolution, followed by MALDI-MSI and DESI-MSI, e) MALDI-MSI has the most incredible flexibility in the mass range; and f) DESI-MSI is the fastest MSI method with the least amount of sample preparation[97].

Spatial-MS methods provide a platform for discovering and detecting novel biomarkers in the femtomole range of sensitivities relative to the spatial resolution range, **Figure 1.7.2. 2**. Furthermore, it allows for examining lipidomic alterations in diverse tissue locations, leading to a better understanding of the efficacy of various anticancer medication therapies. The spike in lipidomic discoveries highlights the usefulness of DESI-MSI and

related MSI techniques in obtaining valuable information on lipid composition[98,99].

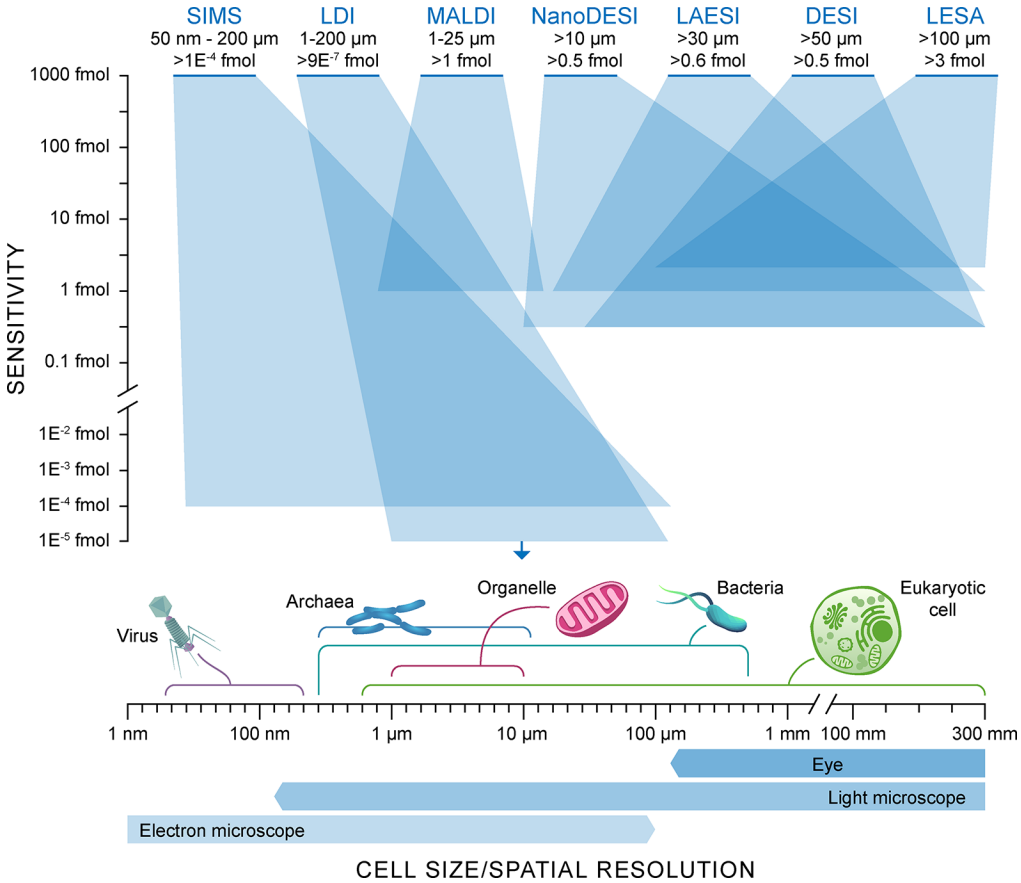


Figure 1.7.2. 2. Representation of the methodologies available for spatial metabolomics and proteomics using MS. For these spatial-MS techniques, the range of sensitivities in femtomoles (y-axis) is compared to the spatial resolution (x-axis). The transparent blue boxes depict the spatial dynamic range. Other structural imaging techniques' cell size dimensions and lateral resolution are given along the x-axis for comparison[100] .

1.7.2.1 MALDI-MSI

MALDI, first described[83] in 1985, is one of the most widely used ionisation techniques for analysing big molecules such as intact peptides, proteins, and synthetic polymers. Soft ionisation is aided by an organic matrix, which is either combined with the analyte solution or applied to the sample surface and re-crystallized via solvent evaporation. The matrix is irradiated with ultra-violet light generated by a pulsed laser, resulting in spontaneous ablation and the release of matrix and analyte molecules into the vacuum chamber. In the ablation plume, analyte molecules undergo protonation or deprotonation,

resulting in molecular ion species that can be accelerated and delivered into a mass analyser. The organic matrix substance and laser wavelength allow for selective ionisation of specific analytes[101]. MALDI-MSI, which was initially described[102] in 1997, is well suited for MSI because it allows for the capture of mass spectrum information while recording the coordinates for the spectra's origin, and it has subsequently emerged as the most often utilised ionisation technology for MSI. The MALDI sample preparation steps schematics are shown in **Figure 1.7.2.1.**

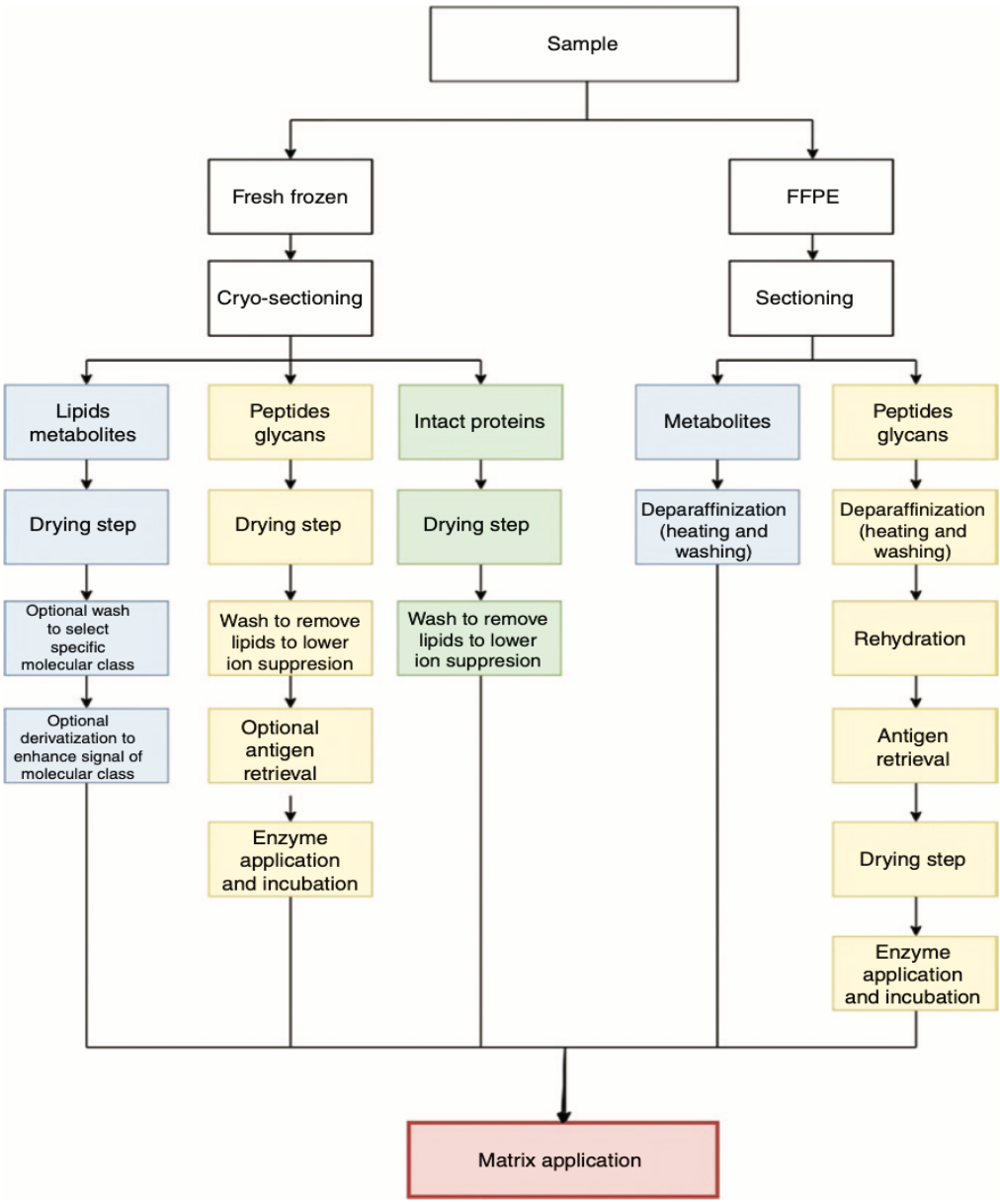


Figure 1.7.2.1.1. Typical MALDI-MSI sample preparation steps are depicted in a flowchart[103]. Different processes are required depending on the sample preservation and molecular class of interest. FFPE refers to formalin-fixed paraffin-embedded.

1.7.2.2 DESI-MSI

DESI is primarily a surface sampling technique; a methodology for creating surface images may be derived by operating it in a continuous scanning mode with set scan periods and lateral sample stage motions. It was first reported[85] in 2004, and the imaging experiment is carried out in continuous raster sampling mode, with the target moving in a 'typewriter-like' motion under the DESI spray. Spatial resolution is estimated using the target stage velocity, mass spectrometer sampling rate, a number of spectra averaged for a single

pixel, and distance between adjacent line scans. Although DESI has a low spatial resolution (75-150 μm), it requires little to no sample preparation. Tissue sections are put on glass slides, and the samples are taken at room temperature. Because DESI is an ambient ionisation technique, the source is positioned on the front plate of a mass spectrometer and is used in ambient settings. This operation allows for the reconstruction of analyte distribution maps for MSI, as the first reported[104].

1.7.2.3 Imaging Mass Cytometry

Imaging Mass Cytometry (IMC) employs antibodies tagged with metal isotopes from the lanthanide series, which have favourable chemistry and do not occur in biological systems; abundances per cell are measured using a time-of-flight mass spectrometer. In mass cytometry, ion counts are assumed to be proportional to the level of expression of the antibody-targeted antigens of interest. Mass cytometry engages stable isotopes of nonbiologic metals conjugated to conventional antibodies, which eliminates the issue of spectral overlap as the panel size builds up, **Figure 1.7.2.3.1**[105,106].

Bendall and colleagues[107] published a ground-breaking study in 2011 that used mass cytometry to map the human immune system, demonstrating the technique's impressive utility. IMC can simultaneously measure up to 35 proteins on tumour tissues in a spatially resolved manner (1 μm), including surface protein, an intracellular protein, cell cycle, viability, and other parameters. A single platform at a single cell level offers a comprehensive view of haematopoiesis, including immunophenotyping via surface antigen expression, intracellular signalling changes, and cell cycle progression[108–110]. Based on this research, mass cytometry can provide information on disease pathogenesis and assess and guide therapeutic regimens that would not be apparent without the analysis of subclones or small cell populations.

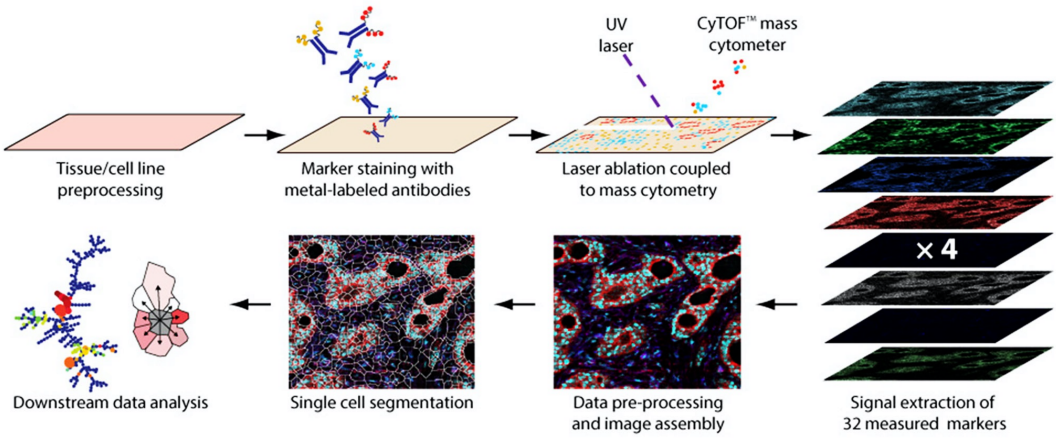


Figure 1.7.2.3.1. Imaging mass cytometry (IMC) workflow[105]

1.7.2.4 LC-MS

Liquid chromatography Mass Spectroscopy (LC-MS) separates analytes in complex mixtures using the principle of retention of analytes on a stationary phase to eliminate ion suppression effects caused by the sample matrix. The stationary phase is often made up of porous silica spheres chemically modified on the surface. Liquid samples are fed into a column filled with a stationary phase and eluted with a mobile phase to separate the analytes. In traditional column chromatography, gravity draws the mobile phase through the column. In (ultra) high-performance liquid chromatography (HPLC and UHPLC) setups, the mobile phase is typically given with increased pressures to maximise separation efficiency and shorten the time required for complete chromatographic separation. The chromatographic separation is based on the analytes' differential affinity for the stationary phase. The analytes should elute separated from one another under optimal chromatographic conditions.

In what is known as normal-phase chromatography, silica particles with no surface modifications are typically employed in conjunction with aprotic hydrocarbon-based mobile phases such as chloroform or hexane to separate analytes soluble in non-aqueous solvents. In normal-phase chromatography, the separation mechanism is mainly based on analyte adsorption effects on the stationary phase. A normal-phase column typically elutes analytes in the order of polarity, with lipophilic chemicals eluting first and the most polar compounds eluting last. Reversed-phase (RP) chromatography, which employs silica spheres with modified surface chemistry, is more widely utilised

for experimental methods in life science and pharmaceutical research. Lipophilic groups such as octyl (C-8), octadecyl (C-18), phenyl-hexyl, or pentafluorophenyl (PFP) are among the RP modifications, **Figure 1.7.2.4.1**. In RP chromatography, mobile phases are often aqueous solutions of polar-protic solvents such as methanol, 2-propanol, and acetonitrile which may contain modifiers such as formic, acetic, and trifluoroacetic acid, or ammonium formate and acetate. Surface interactions between non-polar residues of the analytes and the stationary phase are the primary route of separation in reversed-phased chromatography, with larger non-polar residues retaining analytes longer than those with large polarised surface areas. This technique permits analytes to be separated across a broad chemical spectrum, including biologically relevant analyte classes such as organic acids, lipids, proteins, and peptides.

Lipid extraction is required to differentiate cellular or fluidic lipids from proteins, carbohydrates, and pollutants while limiting lipid breakdown. They are utilised for lipid extraction due to the solubility of the hydrophobic chain of lipids in organic solvents[112]

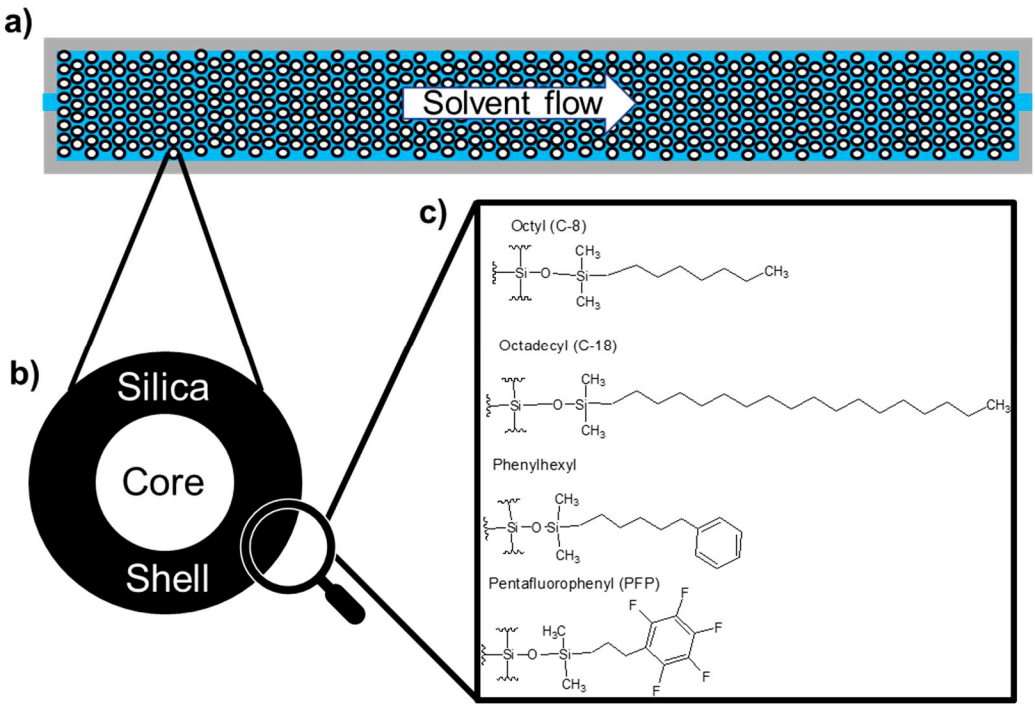


Figure 1.7.2.4.1 Diagram of a packed LC column. a) A column of stationary phase spheres encircled by a flowing mobile phase. b) Magnify a stationary phase sphere of a polymer core surrounded by a silica shell. b) A list of commonly utilised surface modifications for reversed-phase liquid chromatography.

1.8 Mapping Strategies of Tumour Heterogeneity in a Multi-modal Analysis

Combining two or more imaging modalities (i.e., 'multimodal imaging') provides benefits such as improved sensitivity and specificity of chemical classes that are difficult to identify or analyse using a single imaging modality alone and enhanced data mining capabilities. Furthermore, multiple spatially resolved MS tools can provide a more effective probing of the biomolecular complexity of biological systems by increasing confidence in molecular annotations and localizations.

All the techniques described in this section can provide a wealth of information. Still, when combined with other analytical approaches, they can provide a complete picture of complex biochemical processes. Furthermore, because each technique has its distinct advantage(s) for cancerous tissue analyses, their results can be improved through data combination as more biologically relevant signals are produced when used in a multi-modal fashion. MSI

experiments in multi-modal analysis can provide a wealth of chemical meaning, whereas transcriptomics can provide meaningful biochemical and phenotypic information, **Figure 1.8.1**.

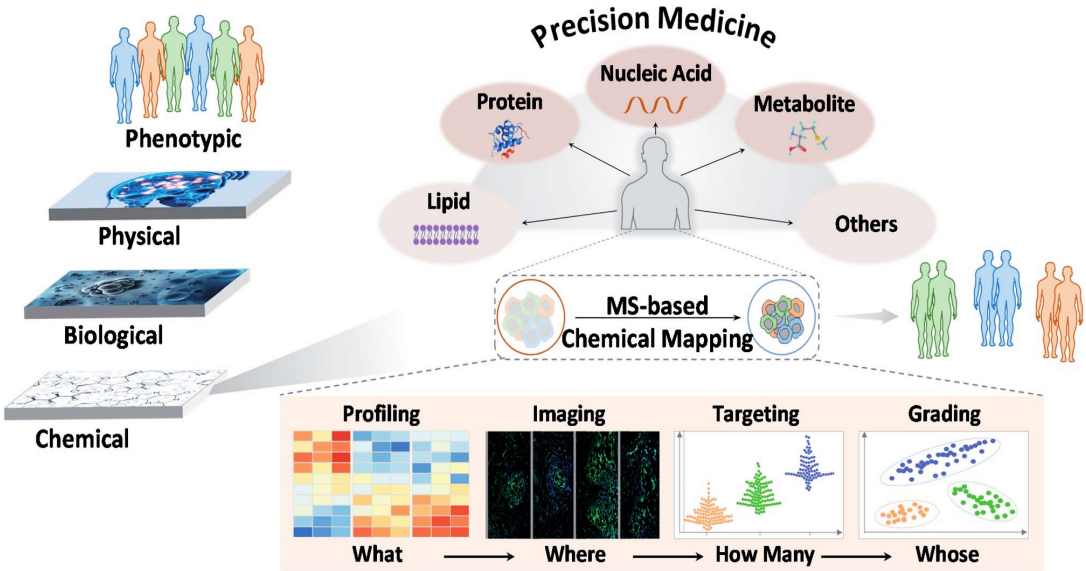


Figure 1.8.1. MS-based chemical mapping and profiling in Multi-modal analysis for precision medicine [113].

1.8.1 MSI-based mapping and spatial clustering analysis for Lipidomics

Cancer cells undergo metabolic reprogramming and heterogeneity within or between individuals, according to growing evidence[3,114]. In the case of solid tumours, the specific tissue site is thought to be the site of induced metabolic changes. The spatial and temporal variations associated with metabolic heterogeneity typically reflect tumour progression, metastasis, and prognosis and are extremely difficult to identify[3,61,115]. Spatially resolved metabolomics analysis based on MSI has been developed as a powerful analytical tool for investigating tumour metabolite chemo-spatial relationships[116–118]. MSI can preserve spatial information for large numbers of molecules that have been desorbed and ionised on-site simultaneously. Direct analysis of tissue samples without complex preparation is possible using ambient DESI-MSI[119]. MSI is a straightforward and promising technique for investigating tumour metabolism and interactions within the tumour microenvironment. Grid sampling and high-throughput mass analysis underpin MSI's ability to map

metabolites across entire tissue sections. Sequential scanning can be performed on a tissue surface that has been abstractly divided into individual pixels on a grid. Each pixel produces a mass spectrum, and each ion corresponds to a particular distribution. Thousands of pixel datasets must be processed through peak recognition and alignment, dimension reduction, and image reconstruction to produce a final ion mapping image118. Histology is also important in MSI analysis for identifying chemo spatial relationships. The spatial distribution of the target ion can be interpreted using the morphological features in the haematoxylin-eosin (H&E) stain image. Histology information is also used to guide the selection of the region of interest (ROI), which is especially important for tumour tissues with obvious tumour heterogeneity[121–124].

One study[125] proposed and developed a graphical data processing pipeline for MSI-based spatially resolved metabolomics, which included basic statistical algorithms, data pre-treatment, histology-driven and characterised ion image overlay, and metabolic feature-based spatial segmentation. The metabolic feature-based spatial segmentation pipeline proved to be a powerful and easy-to-use tool for better understanding tumour heterogeneity and identifying low-abundance microregion-specific biomarkers.

1.8.2 Complementary Spatial Transcriptomics to Map Alterations with MSI

Many tumour-related changes in lipid profiles are caused by changes in gene regulatory cascades, which are in turn influenced by oncogenes and tumour suppressors, epigenetics, and other adaptive mechanisms. As a result, complementary transcriptomic analysis can potentially enrich lipidomic profiles with gene expression data significantly. This complementary analysis assists in the biological interpretation of lipid metabolic changes and delineates pathogenic pathways or potential therapeutic targets. This omics field has evolved dramatically since the introduction of DNA microarray technologies, and it is now dominated by next-generation RNA sequencing. Its traditional application on tissue extracts has the same limitations as bulk lipidomics. It

provides a mean gene expression profile across cell populations, even when chosen for high tumour content.

Traditional molecular biology techniques, such as *in situ* hybridization (ISH) and immunohistochemistry (IHC), were the first to map DNA, RNA, and proteins to preserve spatial information within tissues. On the other hand, targeted approaches necessitate pre-selected markers, limiting the spatial analysis to a subset of genes and proteins at a time[126,127]. Techniques such as next-generation sequencing (NGS) are unbiased spatially resolved methods for querying the entire transcriptome from tissue sections. Therefore, it enables a high-resolution assessment of spatial gene expression across tissue sections, overcoming the limitations of pre-selected markers. This untargeted technology can query the entire transcriptome within a single tissue section for molecular profiling and system exploration. Transcriptomics combined with other unbiased omics technologies, such as spatial lipidomics, could provide a robust set of tools for characterising tumour heterogeneity and disease processes in intact tissue sections.

Transcriptomics technologies analyse gene expression by extracting and amplifying nucleic acids, most commonly ribonucleic acids[60,126]. Because of the potential for exponential amplification, this group of technologies is extremely sensitive and has been the driving force behind cell typing and classification[128.] The MSI and transcriptomics analysis combination is of great interest because it would allow for the simultaneous correlation of gene expression to gene products and bi-products. Most studies combining mass spectrometry and transcriptomics measurements have used liquid chromatography for bulk proteomics and/or metabolomics[129–131].This methodology incorporates spatial information on the microdissected structures, coupling MSI and transcriptomics

Furthermore, some exciting work has been done in which transcriptomics and proteomics information correlate on different bulk samples. Work combining two approaches on the same sample revealed information about the relationship between fatty acids and immunity in breast cancer[128] 131While there are only a few examples, the combination of transcriptomics and MSI

will expand in the coming years, with the potential to uncover new connections in molecular biology.

1.8.3 Laser Capture microdissection system for MSI- guided sample preparation

MSI (mass spectrometry imaging) is a technique for imaging molecules in biological tissue sections that are unlabelled and multiplexed. As a result, it enables spatial and molecular annotations of tissues in addition to histology. MSI has already been shown to guide subsequent material isolation technologies, such as laser capture microdissection (LCM), allowing for more in-depth molecular characterization of MSI-highlighted tissue regions. However, with MSI reaching spatial resolutions at the single-cell scale, precise co-registration between MSI and the LCM is required. On the other hand, the adaptive multi-modal analysis uses a single tissue section. MALDI-MSI paired to LCM has previously been reported[133] to analyse proteins, lipids, and metabolites from the same tissue sections utilizing liquid chromatography-mass spectrometry (LC-MS). This methodology[134] demonstrates how the strengths of MSI, histology, and extract-based omics can be combined to allow for a more comprehensive molecular characterization of in situ biological processes.

1.8.4 Histological techniques

Histological study of frozen or fixed tissue stained slices is routinely used to evaluate tissues. Tissue sections are often stained with dyes to analyse tissue morphology. Other information-rich techniques are based on affinity-based reactions to dye for specific proteins or ribonucleic acids (RNA).

Dye-based stains are routinely used to analyse the morphology of a specific tissue. Differences in dye affinity indicate the tissue structure. Counterstained tissues are frequently used to emphasise nuclei and identify individual cells. Hematoxylin and eosin are the most often used dye-based stain (H&E). Eosin is a lipophilic red dye that accumulates in lipid membranes, whereas

Hematoxylin stains chromatin and the rough endoplasmic reticulum, ribosomes, collagen, myelin, elastic fibres, and acid mucins. When both colours are used together, they allow for the analysis of cellular tissue composition and general morphology. A brightfield microscope is used to examine the dyed tissue sections.

Immunohistochemistry (IHC) uses antibodies to bind to peptides or proteins in tissue sections, allowing them to be seen. IHC is typically performed with a combination of a primary and secondary antibody; the primary antibody binds to the target structure within the tissue, while the secondary antibody is conjugated with alkaline phosphatase (AP) or horseradish peroxidase (HRP), detects species-specific epitopes on the primary antibody. The reporter enzymes AP and HRP catalyse reactions that result in the formation of a water-insoluble dye that, depending on the reporter molecule utilised, often precipitates as brown or red-pink spots where the target and hence the enzyme are present. A brightfield microscope is used to examine the dyed tissue sections.

A fluorescence microscope is used to examine the stained tissue sections. The fundamentals of Immunofluorescence (IF) are similar to those of IHC, except that the reporter molecule is a fluorescent dye conjugated to either the secondary antibody or, less typically, the primary antibody. There is no need for a secondary antibody if the label is bound to the primary antibody.

In-situ hybridization (ISH), as opposed to IHC, is used to visualise the distribution of deoxyribonucleic acid (DNA) and RNA. Typically, the probes are built on oligonucleotides that complement and hybridise with the target DNA or RNA at high temperatures. Washing solutions can be used to remove excess probes. The detergent addition, temperature, and osmolarity of the solutions can all be modified to leave only probes with full sequence binding to the target pH. The probes can be directly labelled, or additional reporter probes can be utilised to boost specificity. When numerous reporter probes may bind to the primary probe, the sensitivity of this dual-probe method increases. Alkaline phosphatase or, less typically, radionucleotides can be coupled to the labelled probe to detect the target fluorescent dyes.

Fluorescence microscopy, brightfield microscopy, or autoradiography are used to determine the distribution of the target DNA or RNA.

1.8.5 Embedding strategies for MSI

Various embedding strategies have been documented to optimise sample preparation for various imaging modalities, particularly for small or fragile samples and whole animals[135]. These approaches are frequently based on embedding in Optimal Cutting Temperature medium (OCT)[118,136], sodium-carboxy-methyl cellulose (Na-CMC)[137], gelatin, agarose or poly(N-(2-hydroxypropyl)-methacrylamide) (HPMA). HPMA provides tissue support and allows for cold embedding (-8 C), which reduces the temperature difference between the frozen tissue specimens and the embedding media, preventing the melting of tissues during preparation. Any embedding process must not change or thaw the samples to retain the tissue morphology for histological assessment. Furthermore, the embedding should not change the distribution of analytes within the tissue to obtain correct MSI data. In addition, the embedding media should not leak polymers that could contaminate the instrument or alter the mass spectrometric results. Finally, the embedding should enable seamless integration of the sample preparation workflow with orthogonal tissue-based imaging techniques such as IHC, imaging mass cytometry (IMC), fluorescence in-situ hybridization (FISH), or RNA-in-situ hybridization (RNA-in-situ hybridization) (RNA- ISH). These affinity-based techniques rely on probe hybridization with the target antigen or ribonucleic acid (RNA). The ideal embedding medium prevents unspecific probe binding or interferes with probe binding to their targets.

Regarding mechanical criteria, the embedding media must have sufficient viscosity and adherence to specimens to allow for precise specimen positioning and easy sectioning without moving or curling sections. Based on the criteria outlined above, a systematic review of previously published media was conducted; however, none fulfilled the standards. As a result, as part of the study discussed here, a new embedding medium was designed and evaluated[138]. Chapter 4 contains the experimental methodology and outcomes.

1.8.6 Data Analysis

Imaging research aims to identify *m/z* traits of biological relevance specific to a spatial region, such as those that distinguish a tumour region from a normal one. MSI investigations can produce profiles containing signals corresponding to hundreds of biomolecules derived from tissue samples. The data analysis technique is critical in MSI research because it connects biochemical and histological features.

In this study, various chemo-informatics methodologies based on four significant phases have been developed to meet the ultimate goal: 1) raw imaging data pre-processing, 2) unsupervised analysis as a preliminary stage to data investigation, 3) supervised classification, and 4) segmentation analysis for ITH.

1.8.6.1 Pre-processing data

Typically, the pre-processing of MSI data sets involves 1) normalisation, 2) peak picking, and 3) peak matching/spectra recalibration. Incorporating these pre-processing steps makes the MSI data comparable and manageable while accounting for bioanalytical complexities common to MSI studies.

Spectral normalisation, i.e. scaling each spectrum to a factor (usually dataset-specific), is one of the key steps in data pre-processing. This type of normalisation permits better inter-comparison of intensities between different mass spectra. Total ion count (TIC) normalisation is the most popular and common method recommended in general[139]. In TIC normalisation, the sum of all intensities (i.e. the TIC value) is calculated, and then all *m/z* intensities are divided by the TIC value. One major disadvantage of TIC normalisation is that the method can hugely influence the intensity of a single outlier *m/z* peak. Median normalisation has been suggested as an alternative to TIC and demonstrated to be the least compromised by biological variability.

Another important pre-processing step is peak picking, i.e. selection of *m/z* -values which correspond to high and relevant peaks. The peak picking step aims to reduce the number of *m/z* values by filtering out values corresponding

to noise-related signals. When comparing the lists of relevant peaks from different datasets, *m/z* values selected in different spectra for the same peak can slightly differ. Therefore, peak alignment is fundamental to correct these small *m/z* shifts, as this effect cannot be counterbalanced by instrument calibration[140]. Therefore, peak alignment is a key step in pre-processing. Several peak alignment algorithms exist, based on hierarchical clustering[141] and the use of an average spectrum to find common peaks[142].

1.8.6.2 Multivariate statistical analysis

Multivariate statistical analysis can be used to establish correlations in high-dimensional datasets. In MSI data analysis, by treating each pixel as an independent sample, multivariate techniques can be used where their output is shown as score plots in the spatial domain. Multivariate techniques commonly used for imaging datasets can be divided into two groups: unsupervised and supervised analysis.

1.8.6.2.1 Unsupervised classification

Principal component analysis (PCA) is the most common unsupervised technique in analysing imaging data[143]. In imaging, PCA can be applied to spectra from specific regions of interest and the entire tissue section without prior knowledge. PCA is a dimensionality reduction technique that involves reducing the number of variables with minimal information loss to reveal the data set's general distribution. It is achieved by finding relationships between variables (*m/z* features) and between samples (pixels) by finding the linear combination of variables with the highest variance. Linear orthogonal transformation helps to maximise the variance resulting in an ordered set of principal components. The first principal component (PC1) contains the maximum variance in the data, followed by PC2, which is orthogonal to PC1 and so on. Most of the variation is usually explained in the first three principal components. The data from PCA is generally reported in the form of score plots (coordinates of pixels in the transformed axes), loadings (represents the contribution of *m/z* peaks to the principal components) and variance (from

each principal component). It is often used as a way to briefly inspect the data prior to the implementation of advanced data analysis algorithms[144].

Hierarchical clustering analysis (HCA) is another unsupervised approach where the pixels that share similar MS profiles (similar *m/z* peaks with similar intensity) are grouped, and pixels with different MS profiles are separated. The data is presented as a dendrogram/ hierarchical tree using a similarity metric (often Euclidian distance in imaging) and a 'linkage' calculation. Hierarchical clustering is extensively used for gene expression analysis[145] and has also been used for MSI-based cancer studies[136,146]. The advantage of hierarchical clustering is the ability to investigate subgroups within the data interactively. However, it is best suited for comparatively small numbers of individual data points, as visualising large datasets gets challenging and removes the interactive nature by overcrowding the dendrogram.

K-means clustering is also an unsupervised analysis technique that incorporates the spatial component of the data into clustering analysis. It involves the segmentation of the imaging dataset into a pre-defined number, (k), of clusters, generally based on the Euclidian distances between the mass spectra. The clustering algorithm is initialised by selecting k-locations in the data space (k-means). Each pixel's mass spectrum is assigned to the cluster with the closest mean, then the centroid of each cluster is defined as the new mean, and the process is repeated until convergence is achieved. It is a powerful method for quickly assessing the spatial-chemical organisation of MSI datasets[147]. However, the results are highly dependent on the radius of pixel groups clustered together and the number of clusters.

Other unsupervised strategies, such as probabilistic latent semantic analysis, independent component analysis and non-negative matrix factorisation, have also been used for analysing imaging datasets[144].

1.8.6.2.2 Supervised classification

Discriminant analysis (DA) is the most widely used supervised strategy considering the data's group membership. It is done by maximising inter-class

variance (i.e. cancer vs normal) while simultaneously reducing intra-class variability. However, employing PCA before DA is based on the assumption that variance between different groups is a major source of the variance in the data. Linear discriminant analysis (LDA) is usually employed first to reduce the dimensionality of the data. Following this, LDA is applied, which finds linear combinations in m/z features that best explain the group membership. The resulting discriminant function can be used for classifying samples/patients by building a test set and a training set. The number of correct classifications then determines the quality of the model.

Additionally, LDA cannot be directly applied in circumstances where the number of variables exceeds the number of samples, widespread in MSI datasets[148,149]. On the other hand, PLS is a regression technique that combines features of PCA and multiple regression. PLS extracts latent variables as linear combinations of the original explanatory variables such that most of their association with the response variable is explained.

Maximum margin criterion discriminant analysis (MMC-DA) is another supervised multivariate technique adapted for lipidomics-based MSI studies as recursive maximum margin criterion (RMMC-DA)[150]. After dimensionality reduction, the RMMC feature extractor can extract the most discriminatory features by maximising the variance between and within classes. Additionally, unlike LDA, it does not suffer from the small sample size problem, and its feature-extracting capacity does not depend on the non-singularity of the within-class scatter matrix. Class-specific *m/z* signatures are derived based on weighted combinations of *m/z* feature patterns. The average margin between classes increases by maximising the distance between their most outer data points, thus improving class separation. The RMMC-DA approach also avoids selecting an optimal number of principal components before applying LDA and avoids the model over- or under-fitting. The advantages of RMMC-DA concerning other widely used discriminant analyses are summarised in **Table 1.8.6.2.2.1**[150].

Several other multivariate algorithms such as ones incorporating neuronal networks, random forest and support vector machines, to name a few, have

been used for MSI data analysis. The main feature of these strategies is that they build a statistical model to separate the training set [151], which can be used to classify unknowns.

Table 1.8.6.2.2.1. Main features of dimensionality reduction techniques used in MSI data analysis[150].

Dimensionality reduction technique	Components derivation
Principal component analysis (PCA)	Maximises overall dataset variation, without specific regard to between-class differences.
Partial least square (PLS-DA)	Maximises between-class variation (difference in class means), disregarding within-class variability.
Linear discriminant analysis (LDA)	Maximises ratio of between- versus within- class variation, with the condition that within-class variance is non-singular (*i.e.* number of samples >> number of variables).
Recursive maximum margin criterion (RMMC-DA)	Maximises the difference of between-class within-class variation.

1.8.6.3 Univariate analysis

In MSI, univariate tests are performed to compare spectral profiles from different cohorts (i.e. tumour vs non-cancerous) peak by the peak to identify potential biomarkers. In most MSI-based cancer studies, mean spectra of histologically defined regions are extracted for each patient and used for univariate analysis[152].

Statistical tests, such as Student's t-test and analysis of variance (ANOVA), are used to assess the likelihood that the sample groups under investigation are different. The t-test is probably the most commonly known statistical test, but it is often not applicable to clinical data[144]. For multi-class systems, analysis of variance (ANOVA) is better suited for clinical data analysis. Univariate statistical tests are individually applied to each MS signal/feature. The result of such tests is a p-value, which measures the probability that the observed effect is caused by chance. A p-value below 0.05 is often considered statistically significant (corresponding to a false positive rate of 5%).

1.8.6.4 Segmentation Analysis

The first paper in 2008 demonstrated MSI's ability to detect ITH and introduced hierarchical clustering of MSI data as a new tool for interpreting complex human cancers[153]. Hierarchical clustering is a procedure that iteratively merges single objects and groups of objects (clusters) based on their similarity into a tree-like structure known as a dendrogram. When applied to MSI, each tree branch can be considered a class, and all spectra (pixels) belonging to that class can be visualised with a single colour. These issues have recently been addressed by developing novel or modified multivariate segmentation techniques[154]. Alexandrov and his colleagues developed clustering algorithms that account for the spatial nature of MSI data by clustering based on a spectrum's spatial neighbourhood, **Figure 1.8.6.4.1**[155].

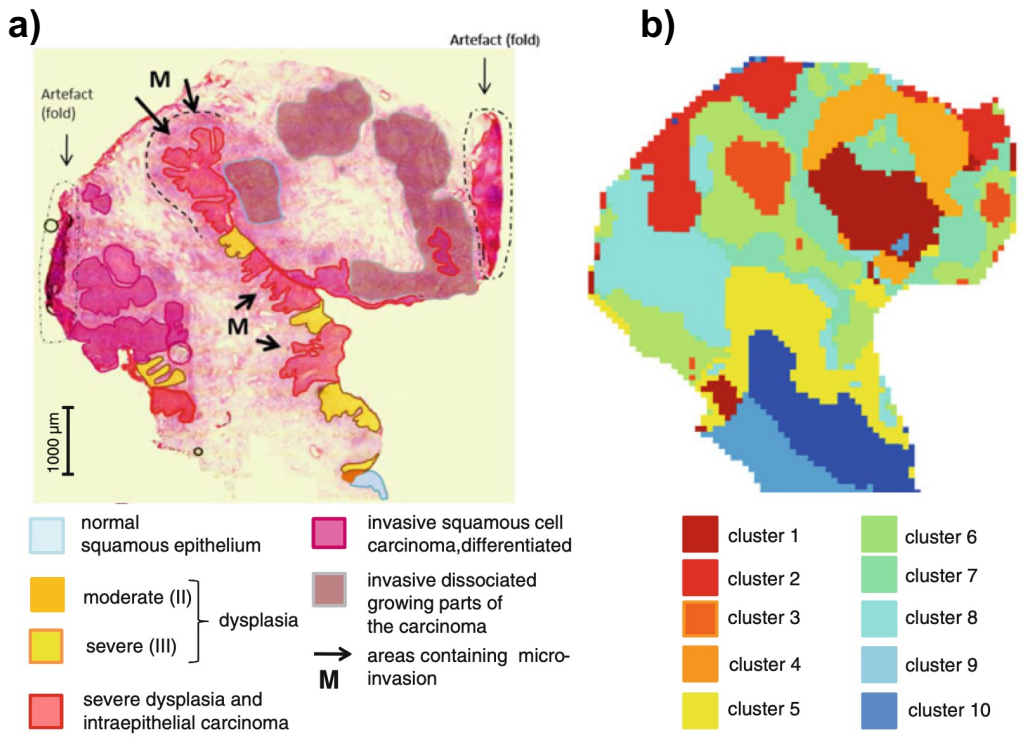


Figure 1.8.6.4.1. A segmentation map of 10 clusters was created using MALDI-imaging data from the section. a) After MALDI capture, a microscopic image of a larynx cancer segment stained with H&E is obtained. The annotation of regions of interest is done by hand. b) Colours in the segmentation map of 10 clusters derived from MALDI-imaging data, reproduced with permission[155].

The segmentation concept was used for several different cancer types and clinical endpoints by first obtaining spatially resolved proteomic data from primary tumour specimens using MALDI-MSI. Unlike genetic studies, MSI exploits the fact that cellular selection operates on phenotypes by detecting phenotypic information through mass spectral proteomic or metabolic profiles[66]. The number of clusters must be specified based on tissue complexity, an inherent problem in all clustering methods. A standard solution uses statistical measures to determine the optimal number of clusters[121].

The application of this data analysis strategy to other cancer types and clinical endpoints confirmed its general applicability. For example, in a cohort of 63 patients with intestinal-type gastric cancer, segmentation revealed significant ITH and a statistically significant association between the presence of one specific cluster and overall survival[156]. A study on sarcomas found that the

presence of a molecularly distinct tumour subpopulation is associated with a poor prognosis for the patients. The study's unique feature was that sarcomas have a high intra-tumour histological heterogeneity in terms of the grade of differentiation of the tumour cells. The various grades necessitated a more differentiated virtual microdissection of the tumour regions. Only in this way could a tumour region within moderately differentiated osteosarcoma cells be assigned a prognostic value.

It demonstrates that clustering in conjunction with histological annotation can identify microscopically indistinguishable tumour subpopulations that can negatively impact clinical outcomes. Although spatially resolved metabolomics is an excellent tool for elucidating *in situ* molecular events, its application remains challenging due to the complexity of endogenous metabolites in tissue heterogeneity.

2 Experimental Plan and Methods

2.1 Clinical Data

2.1.1 Ovarian Biopsies

The Imperial College Healthcare National Health Service Trust's institutional review board approved collecting all samples for this investigation (Tissue Bank sub-collection-number iKnife/R14090-3A). All the patients permitted their samples to be used in this investigation. All methods were carried out following institutional and ethical standards. Patients undergoing gynaecological surgery were enlisted, and 22 tissue samples from various HGSOC and Endometrioid Ovarian Cancer (EOC) tumours were collected, **Table 2.1.1. 1**. Immediately after collection, all tissue samples were kept at -80°C.

Table 2.1.1. 1. Patient samples were used during this study.

Patient ID	Diagnosis	Biopsy numbers
T19-221	HGS ovarian carcinoma 4	5
T19-121	HGS ovarian carcinoma 1c	6
T18-124	G2 endometrioid carcinoma of ovary 2b	6
T18-161	G1 endometrioid carcinoma of ovary 1a	5

2.1.2 Breast Biopsies

The institutional review board approved collecting all samples for this study at Imperial College Healthcare National Health Service Trust (Tissue Bank sub-collection number SUR-ZT-14-043). All patients provided their consent to use their samples in this study. All methods were performed according to institutional and ethical guidelines. Patients undergoing surgery were recruited, and 23 tissue samples were taken from a range of Luminal A, Luminal B, triple -ve/basal-like, HER-2 enriched, normal-like tumours identified in the histopathology assessment, **Table 2.1.1. 2**. Upon collection, all tissue samples were stored at -80°C.

Table 2.1.1. 2. Patient samples were used during this study.

DESI ID	Histology	ER	PR	HER2	Tumour Subtype
B1	IDC	-	-	-	TNBC
B2	IDC	8/8	8/8	-	LUMINAL A
B3	ILC	2/8	0/8	neg	LUMINAL A
B4	IDC	8/8	0/8	BORDERLINE	LUMINAL A
B5	IDC	8/8	0/8	-	LUMINAL A
B6	IDC	8/8	8/8	-	LUMINAL A
B7	IDC	8/8	6/8	neg	LUMINAL A
B8	IDC	8/8	7/8	neg	LUMINAL A
B9	IDC	-	-	-	TNBC
B10	IDC	-	-	-	TNBC
B11	ILC	8/8	0/8	-	LUMINAL A
B12	IDC	-	-	-	TNBC
B13	IDC	7/8	0/8	+	LUMINAL B
B14	IDC	8/8	8/8	pos	LUMINAL B
B15	IDC	8/8	5/8	neg	LUMINAL A
B16	IDC	8/8	8/8	pos	LUMINAL B
B17	ILC	8/8	7/8	neg	LUMINAL A
B18	IDC	8/8	2/8	neg	LUMINAL A
B19	IDC	8/8	5/8	neg	LUMINAL A
B20	IDC	8/8	6/8	neg	LUMINAL A
B21	IDC	8/8	8/8	neg	LUMINAL A
B22	IDC	8/8	7/8	neg	LUMINAL A
B23	IDC	8/8	8/8	-	LUMINAL A

2.1.3 PDX models

Fresh frozen breast patient-derived xenograft (PDX) tumours were obtained from Crown Bioscience and Champions Oncology, two commercial companies. The *PIK3CA* mutational status for all 10 PDX tumour samples is summarised in **Table 2.1.1. 3**.

Table 2.1.1. 3. Summary of *PIK3CA* mutation status for all 10 PDX tumour models profiled with DESI-MSI. MT: Mutant, WT: wild-type

Sample	PIK3CA Mutation	Model ID
TM00112	MUT,I391M	1
BR1458	WT	2
J000104720	WT	3
TM0098	WT	4
TM0091	WT	5
J000100674	WT	6
BR1282	MUT,C420R	7
TM01273	MUT,I391M	8
TM01117	MUT,D915Y	9
TM0099	WT	10

2.2 A Modified Fast Snap-Frozen Standard Operating Procedure (SOP) for tissue sample collection

The biopsies were submitted to the pathology lab at the same hospital as soon as they were removed from the operating room. A histopathologist took five or six biopsies to use in the research study after examining tumour size and content of tumour vs normal tissue for best availability. The ovarian biopsies from multiple regions were snap-frozen with a cool rack on a -80°C freezer, and the entire procedure took around 60 minutes. This modified quick snap-frozen SOP provides exceptional specimen integrity and a wide range of tissue analysis choices, including protein, DNA, and RNA extraction. The specimens are then stored at -80°C until DESI-MSI analysis is performed. **Figure 2.2. 1**. depicts the expected collection areas from tumour tissue.

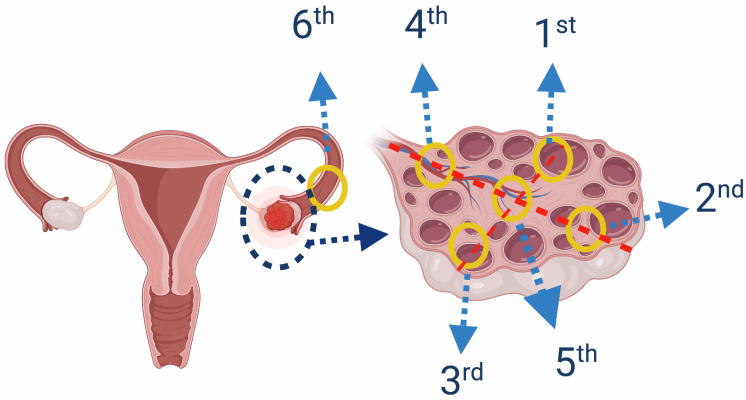


Figure 2.2. 1. Multi-site biopsy collection from Ovarian tumour. The collection strategy for the multi-site biopsies (six tumour tissue samples) from the same ovarian cancer patient.

2.3 Sample Preparation

Each sample was cryosectioned into 10 mm thick parallel sections using a Cryostat HM550 cryomicrotome (Thermo Scientific Inc, Waltham, MA, USA) set to -20°C in the chamber and -17°C in the sample holder, **Figure 2.3. 1**. Tissue sections were put onto Super Frost plus Glass slides (Thermo Fisher Scientific Inc, USA) and PEN membrane slides after sectioning. Before use in sectioning, PEN slides are routinely decontaminated with RNAzap and U.V. irradiation for 30 minutes. This step is taken to protect the RNA integrity of the tissue. The slides were vacuum-packed and stored at -80°C after the sectioning phase was completed.

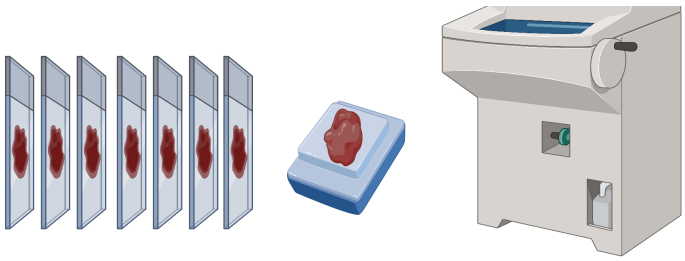


Figure 2.3. 1. Consecutive sectioning of the Tumour samples

2.4 Imaging Mass Cytometry (IMC) Analysis

An antibody panel of 25 protein markers, comprising immunological, phenotypic, signalling, and architectural proteins, was created to differentiate cell types and states. Fluidigm (https://www.fluidigm.com) supplied twenty-five labelled antibodies, **Table 2.4. 1**. Following the manufacturer's instructions, antibodies were conjugated with metals using the Max-par X8 Multimetal Labeling Kit (Fluidigm, 201300). The titration and specificity of all conjugated antibodies were determined by visually comparing IMC pictures of various spare tissue slides from breast cancer patients.

On SuperFrost glass slides, fresh-frozen PDX breast tumours were sectioned at a thickness of 10 μm. The slides were then transferred to a pre-chilled slide box at -20°C and allowed to equilibrate for 1 hour before being kept at 4 °C for 10-15 minutes. Following that, the slides were air-dried for 30 minutes at room temperature. Fixation was performed with 4% paraformaldehyde (PFA), then washed twice with 1x Tris-buffered saline (TBS), and a final wash with 1x TBS+ 0.05% Tween for 5 minutes on an orbital shaker. The slides were then incubated in a humid room for 1 hour in blocking buffer (5 per cent bovine serum albumin in TBS buffer).

Meanwhile, antibody cocktails were made with 0.5% Bovine Serum Albumin (BSA) buffer. The antibody mixture was applied to the sections and incubated overnight in a humid room at 4 degrees. Each slide was washed twice with 1x TBS on an orbital shaker with 0.05 per cent Tween for 5 minutes. Additionally, a 1:500 dilution of Intercalator (DNA staining) was made in 1x TBS and incubated slides in a humid chamber for 30 minutes at room temperature. Finally, slides were washed once with 1x TBS with 0.05% Tween for 5 minutes on the orbital shaker, then three times with distilled water for 5 minutes each. The slides were then air-dried for 30-40 minutes before being kept in a dust-free environment.

A Hyperion Imaging System was used to capture the images. All procedures were carried out following the manufacturer's instructions. In brief, ROIs in the CT or IM areas were chosen randomly from an H&E-stained serial tissue slice

obtained from a professional pathologist. Images were laser ablated at a rate of 200 Hz, and raw data were collected using commercial capture software (Hyperion Imaging System, Fluidigm). The Hyperion Imaging System's status was monitored through the interspersed capture of data from the tuning slide (Fluidigm). The MCD Viewer was used to examine all marker staining patterns and generated images (Fluidigm, v1.0.560.2).

Table 2.4. 1. Fluidigm antibody panel of 25 proteins for immunological, phenotypic, signalling, and architectural markers.

Identify	Annotation	Localisation	Marker
Phenotype	Membrane/ Stroma- Fibroblast and Endothelial	cytoplasm, cytoskeletal (stroma)	αSMA
Immune	dendritic cells> monocytes and granulocytes	Membrane	CD11c
Phenotype	Mesenchymal, some expression in T- and B-lymphocytes	Cytoplasm	Vimentin
Signalling	cellular signaling, unregulated tyrosine kinase activity can drive malignancy and tumor	Cytoplasm= Membrane	p-tyrosine (Tyr 100)
Immune	CD4 T cells, lymphocytes	Cell membrane	CD4
	DNA damage signaling and repair, phosphorylated at Ser139 by DNA-PK in response to cell death receptor activation	Nuclear	p-Histone H2A.X [Ser139]
Architecture	epithelial cells	Membrane	Pan-Keratin
	Stress-responsive protein, Acts as a tumor suppressor in many cell types	Mainly cytoplasmic	p-NDRG1 (Thr346)
Immune	predominantly on neutrophils	Membrane	Ly-6G
Signalling	modulator of the AKT-mTOR signaling pathway, pSer-473 leading to activation and translocates to the nucleus	Cytoplasm. Nucleus. Cell membrane	pAkt [S473]
Architecture	High expression in BC	Membrane	Alpha vbeta-6
Architecture	hydrolysis of ATP coupled with the exchange of sodium and potassium ions across the plasma membrane	Membrane	Na-K-ATPase
Architecture	epithelial cells	Membrane	E-Cadherin/ CD324
Architecture		Membrane	Alpha vbeta-3
Signalling	phosphorylates and activates PKC zeta in the PI 3-kinase signaling pathway	Cytoplasm. Endosome	Anti-PKC zeta (phospho T560)
Phenotype	Cell proliferation	Nuclear	Ki-67
Signalling	Tumor suppressor, antagonizes the PI3K-AKT/PKB signaling pathway	Nuclear> Cytoplasm, Secreated	PTEN
Architecture	endothelial cells, on platelets and leukocytes	Membrane. Cell junction	CD31/PECAM-1
Architecture	epithelial cells	Membrrane	CD326/EpCAM
Signalling	Akt substrate,	Cytoplasm > cytosol	p-PRAS40 (Thr246)
Immune	CD8 ctotoxic T cells, lymphocytes	Cell membrane	CD8a
Architecture	Stroma	Secreted > extracellular space > extracellular matrix	Collagen Type I
Phenotype	Activation cascade of caspases responsible for apoptosis execution	Cytoplasm	Cleaved Caspase 3
Phenotype		Cytoplasm, nuclear	pS6 [S235/236]
Phenotype		Nuclear	pHistone H3 [Ser28]
DNA intercalater like Dapi or Hematoxilyn			Intercalator

2.5 Mass Spectroscopy Imaging (MALDI-MSI) settings and Clustering Analysis

Fresh-frozen PDX breast tumours were sectioned to 10 μm on SuperFrost glass slides. For MALDI-MSI analysis, a uniform coating of 9-Aminoacridine (9-AA) matrix solution was applied to designated tissue sections using an automated spray system (TM Sprayer, HTX Technologies, NC, USA). MALDI-MSI analysis was carried out using an SYNAPT G2 high definition mass spectrometer (Waters Corporation, Manchester, UK) equipped with an Nd: YAG (neodymium-doped yttrium aluminium garnet) laser for ionisation on the sample. Spectra were collected in negative ion mode at 20 pixels/sec with the mass range *m/z* 50 to 1200. The pixel size was determined to be 45x45 μm. Additional information on the analytical parameters and pre-processing data methods were done in-house in the R and Python environments. K-means clustering and probabilistic quotient normalisation (PQN) were used to analyse the MALDI-MSI dataset for clustering.

2.6 Desorption Electrospray Ionization Mass Spectroscopy Imaging (DESI-MSI)

2.6.1 DESI instrument setup for a fast sample run

The DESI-MSI analysis was carried out on a Xevo G2-XS QToF (Waters, Manchester, UK) utilising Mass Lynx software V4.1 SCN958, 2015 (Waters, Milford, MA, USA). The negative and positive ion modes were outfitted with a 2D sampling stage (Prosolia Inc., Indianapolis, IN, USA) and a 500°C-heated intake capillary.

A novel re-designed DESI sprayer (Waters, Milford, MA, USA) was modified to provide a high-speed scan rate of 10 scans per second completed in under 30 mins[157], used to analyse samples. It consists of a 400 μm aperture, stainless steel gas nozzle and a 360 μm O.D. and 20 μm ID Taper Tip TM emitter. A stainless-steel emitter guide secured the emitter, which slides into

a stainless-steel positioning disc with radially arranged holes for the nebulising gas flow, as shown in **Figure 2.6.1. 1**.

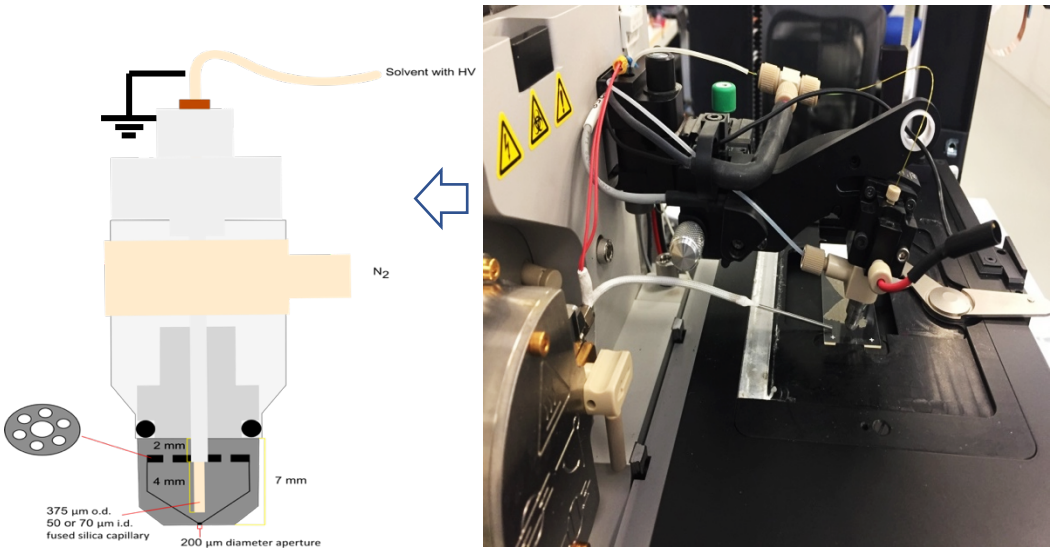


Figure 2.6.1. 1. Detailed schematic details about the new DESI sprayer, design and the DESI-MSI stage condition.

2.6.2 DESI-MSI settings for standard samples run

Using the custom-built DEFFI sprayer, the acquisition speed was set to 10 and 2.5 scans per second (scan/s) (Waters, Manchester, U.K.)[157]. The following instrumental parameters were used: the mass resolution of 100,000 (mass accuracy of <3ppm); injection time of 1500 ms; mass to charge (*m/z*) range of 50–1.500; the capillary temperature of 250°C; capillary voltage of 30V; tube lens voltage of -150V; and skimmer voltage of -40V. The Q-ToF setup was operated the solvent used was 95% of methanol with 5% water (Sigma-Aldrich, St. Lewis, Missouri, USA). The instrument flow rate was 1.5µL/min, and the nebuliser gas pressure was set to 5 bar. The height distance between the DESI sprayer and the sample surface was 2 mm. The distance between the sprayer and the heated capillary was 10 mm. The distance between the sample surface and the inlet capillary was less than 1 mm. The sprayer tip's angle concerning the sample surface was set to 75°C. 10°C was selected as the angle between the inflow capillary and the sample surface. The nominal pixel size was determined to be 100x100 m. Following the acquisition of the DESI-MSI, the High-Definition Imaging (HDI) software V1.4 (Waters, Milford,

MA, USA) was used to choose the region for the imaging experiments and to combine the series of raw files for each imaging dataset. The following files were read in the Python and R programming environments using an in-house pipeline.

2.6.3 Cold stage set up

The cold stage set-up was one of the needs of this study to test. So, a small Peltier device PE120-XY (Linkam Scientific, Surrey, UK), a T96 Peltier LinkPad controller (Linkam Scientific, Surrey, UK), and a water pump chiller K39 (Applied Thermal Control, Loughborough, U.K.), were customised and assembled by Agar Scientific (Stansted, U.K.). **Figure 2.6.3. 1**. depicts the construction of a closed chamber by putting sheets of plastic bags on the existing DESI chamber and attaching two long sleeve gloves on the side of the chamber.

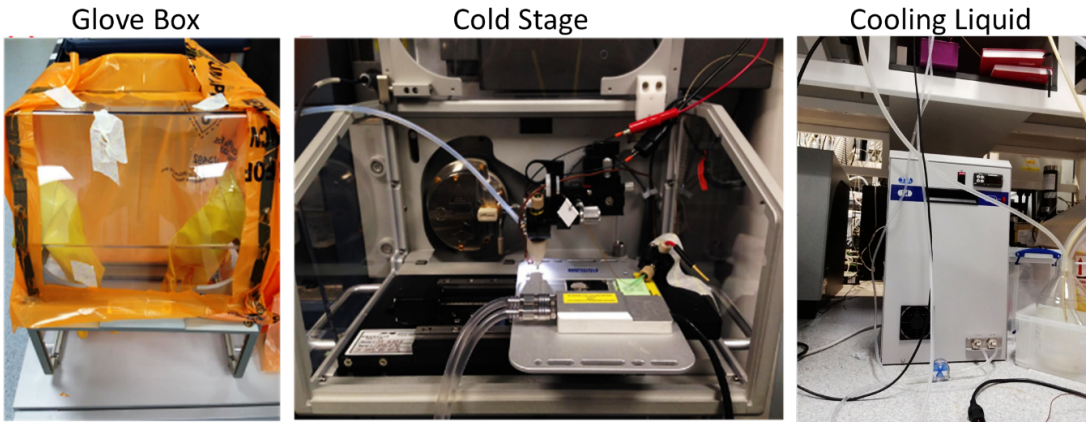


Figure 2.6.3. 1. Detailed schematic details about the cold-stage set-up. Photographs of the glovebox (left), the cold stage Peltier device mounted on the Prosolia 2D moving stage, and the liquid coolant device used to cool down the Peltier device (right).

The nitrogen supply for the DESI-based sprayer was cooled by a looping PTFE tube stacked within a Styrofoam box filled with dry ice, where the temperature dropped to 4°C, **Figure 2.6.3. 2**.

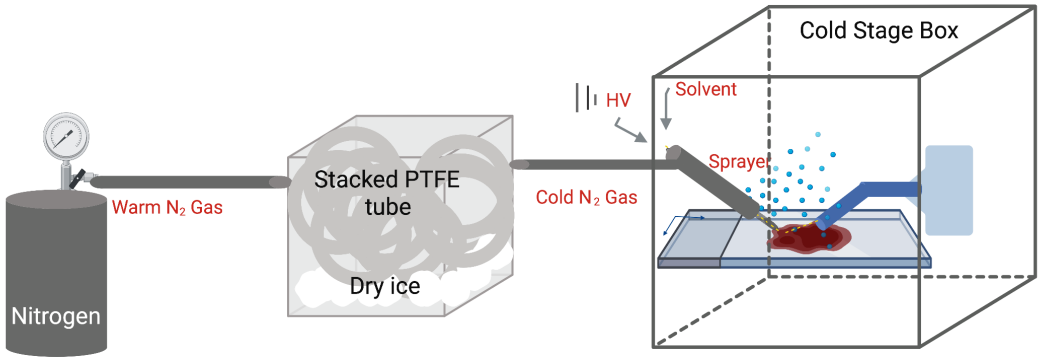


Figure 2.6.3. 2. Schematics representation of Cold stage cooling system. N_2 gas transfer by looping it within Styrofoam with a stacked PTFE tube filled with dry ice.

2.7 Post-DESI-MSI histological staining

The tissue sections were stained with H&E for histological tissue classification by a pathologist after DESI-MSI analysis. **Table 2.7. 1**. outlines the H&E staining procedure. A histopathologist can distinguish tissue types based on the contrast and colours of staining, as well as the morphological/architectural detail of the cells. Haematoxylin is a dark blue/violet basic/positive stain that binds to basophilic DNA/RNA in the nucleus. Eosin is a red/pink stain that attaches to acidophilic molecules in the cytoplasm, such as proteins.

Table 2.7. 1. The H&E staining protocol.

Step	Time
Stain nuclei in Harris Hematoxylin	1 minute
Rinse (immerse in a dish with running water)	10 seconds
Differentiate in 1% acid alcohol	2-3 dips
Rinse (immerse in a dish with running water)	10 seconds
"blue" in Scott's water	20 seconds
Rinse (immerse in a dish with running water)	10 seconds
Stain in 1%eosin	45 seconds
Rinse (immerse in a dish with running water)	10 seconds
Dehydrate in 99% alcohol x3	20-30 dips

2.7.1 Modified RNase-free staining Protocols

The quality and quantity assessment of isolated RNA have been measured using the Agilent Bioanalyser system from the 10 mm^2 pork liver microdissected sections. Samples previously analysed by the DESI-MSI were stained. It showed that H&E staining could affect RNA quality and quantity; therefore, different staining dyes were compared. The main disadvantage of the H&E staining was that the water-based solutions (dyes) activate RNases; this can lead to RNA degradation, which significantly impacts the following gene expression profiling results.

Three different stainings were compared for Harris Modified (Hematoxylin Solution, Harris modified, Sigma Diagnostics, Order no. HHS-16) and Gill No:3 & Eosin Y, (Eosin Y solution, alcoholic with phloxine B, Sigma Diagnostics, Order no. HT110-3-32), **Table 2.7. 2**. and Cresyl Violet, **Table 2.7. 3**.

Table 2.7. 2. Modified Staining Protocols for RNA-Research with H&E.

Step	Time
Dip in Rnase Free water	30 seconds
Stain in Hematoxylin (Harris Modified or Gill No:3)	30 seconds
Rinse in Rnase Free water	30 seconds
Stain in Eosin Y solution	30 seconds
Dip in 70% ethanol	30 seconds
Dip in 95% ethanol	30 seconds
Dip in 100% ethanol	30 seconds
Air-dry at room temperature	30 seconds

Table 2.7. 3. Cresyl Violet Staining with Modifications for RNA Research.

Step	**Time**
Mount section and fix in cold (-20°C) 75% ethanol	2 minutes
Apply Cresyl violet staining solution directly with 0.2 mm syringe and sterile filter to the section and incubate	1 minute
Swivel gently and dip in 75% ethanol	5 seconds
Dip in 90% ethanol	5 seconds
Dip in 100% ethanol	5 seconds
Final fixation is done in fresh 100% ethanol	1 minute
Dry sample in drying chamber with desiccant	30 seconds

2.8 LC-MS Sample Preparation

2.8.1 Lipid Extraction of the microdissected ROIs

For the microdissected tissue sample preparation, two different sampling procedures were used. The tubes were homogenised using a tissue homogenizer after adding zirconia beads (Sigma-Aldrich Company Ltd, Dorset, UK) and 0.5 mL ice-cold a mix of water: isopropanol (IPA) (1:5;v/v) (per 5 mm^2 regions of microdissected tissue) (Precelllys 24, Bertin Technologies SAS, Montigny-le-Bretonneux, France). The homogenization was done in two cycles, each consisting of 45 seconds of shaking at 6500 rpm, 15 seconds of rest, and 45 seconds of shaking. In between cycles, tubes were placed on ice for 5 minutes to allow samples to cool down. The tubes were vortexed for 1 minute and then placed at -20°C for 10 minutes before being vortexed again; the tubes were then placed at 4°C for 2 hours for protein precipitation. The tubes were then centrifuged at maximum rpm for 10 minutes at 4°C, after which the supernatant was transferred to a new Eppendorf tube and dried in nitrogen flow before being kept at -80°C.Before LC-MS analysis, the samples were resuspended in 100 µL per 5 mm^2 microdissected tissue of a previously prepared solution of 1:5 (v/v) H2O: IPA.
In the second sample preparation, 0.5 mL of Methyl tert-butyl ether (MTBE) was added, and homogenization was carried out in the same manner as the previous preparation. The samples were then speed vacuumed for 2-6 hours at 45°C until dry before being stored at -80°C. The samples were resuspended in a combination of water/isopropanol/acetonitrile (ACN) (1:1:1;v/v/v) in 100 µL per 5 mm^2 of microdissected tissue before LC-MS analysis. At the same time as the sample preparation, blank control tubes with the exact quantities (without tissue and polyethene naphthalate (PEN) membrane) were prepared.

2.8.2 The chromatographic gradient

For lipid profiling, the National Phenome Centre(NPC), Imperial College of London protocols are followed; the facility uses a reverse-phase LC-MS approach that takes less than fifteen minutes. At a temperature of 55°C, the

procedure was run on an ACQUITY UPLC BEH C8 column (1.7 m 2.1 x 100mm). **Table 2.8.2. 1**. shows the linear gradient that was utilised to achieve chromatographic separation. The organic phase (B) was composed of isopropanol/acetonitrile (50:50; v/v) with 5 mM ammonium acetate and 0.05 per cent acetic acid, whereas the aqueous mobile phase (A) was composed of water/isopropanol/acetonitrile (50:25:25; v/v/v) with 5 mM ammonium acetate and 0.05 per cent acetic acid. To improve the peak shape of anionic, notably phosphate-containing metabolites, 20 M phosphoric acid was added to mobile phase A. The injection volume was set at 10 µL. During the analysis, the LC-MS sample injecting chamber temperature was held at 4°C.

Table 2.8.2. 1. LC gradient from a C8 column using a lipid reversed-phase UPLC-MS technique.

Time (min)	Flow (mL/min)	% A	% B
Initial	0.6	99.0	1.0
0.10	0.6	99.0	1.0
2.00	0.6	70.0	30.0
11.50	0.6	10.0	90.0
12.00	1.0	0.1	99.9
12.50	1.0	0.1	99.9
12.55	0.9	35.0	65.0
12.65	0.8	70.0	30.0
12.75	0.7	99.0	1.0
12.95	0.6	99.0	1.0
13.25	0.6	99.0	1.0

2.8.2.1 Data analysis

NPC facility conducted pre-processing data analysis to extract the relative abundance of observed peaks in each sample run. Novel non-linear retention time alignment, matched filtration, peak detection, and peak matching were all implemented using the XCMS software. Following that, relative metabolite ion

intensities are directly compared to detect changes in a given comparison set, such as comparing different lipid extraction procedures.

2.8.2.2 Identification and annotation of metabolites

The exact measured mass was attempted to be converted into metabolite annotations for features with significant changes in expression based on accurate mass matching against established online databases Human Metabolome Database (HMDB, https://hmdb.ca/) and LIPID MAPS® (https://www.lipidmaps.org/). This method is known as a tentative annotation. The structural elucidation of lipids was then carried out using tandem MS spectra.

2.9 Gene Expression Analysis of Single Section Methodology

2.9.1 RNA isolation and Quality Assessment

The RNeasy Micro Kit was used for the RNA isolation (Qiagen, Valencia, CA, USA). Total RNA was purified using a procedure designed specifically for Microdissected Cryosections. The tissue section tubes were filled with RLT buffer and then passed to the purification column. To remove genomic DNA, an on-column DNase digestion was performed, in which the RNA attached to the column was treated for 20 minutes with 10 µL of DNase I diluted in RDD buffer (RNase-Free DNase Set, Qiagen, Valencia, CA, USA). RNase Nuclease-free water was used to elute the RNA from the purification column into RNase Nuclease-free tubes. The RNA extracts were kept at –80°C until they were analysed.

The RNA quality was determined using the Agilent RNA 6000 Pico Lab Chip Kit and microfluidic capillary electrophoresis performed on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). The RNA Integrity Number (RIN) was used to assess each sample's RNA degradation. The RIN was calculated by the 2100 Bioanalyzer equipment using an Expert software algorithm. The fragment size distribution, which includes two peaks relating to 18S and 28S

ribosomal RNAs and a signal in the small molecular weight area related to short RNAs, is used to determine RNA quality in each sample. After that, the distribution is compared to an extensive database of RNA samples, and a RIN is calculated **Figure 2.9.1. 1**.

In summary, the state of RNA integrity is assessed using a RIN supplied by an electropherogram to evaluate RNA degradation[158]. A RIN quality of 5 or higher is appropriate for undertaking downstream analysis. To avoid RNA breakdown, RNase inhibitor has been added to dye solutions in all water-based stages (SUPERase, Invitrogen)[159]. RNA concentration was determined using a NanoDrop® ND-1000 UV-Vis Spectrophotometer and fluorometric quantification on a Qubit® 2.0 Fluorometer using the Qubit® RNA HS Assay kit (Thermo Fisher Scientific, Waltham, MA, USA). For each condition, all preparations were done in technical triplicate.

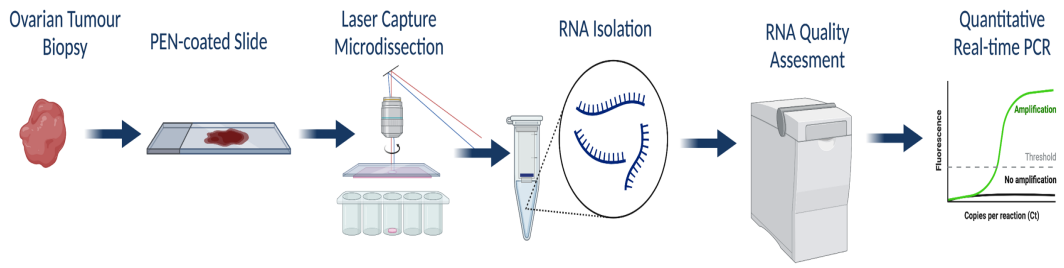


Figure 2.9.1. 1. Procedural overview of the RNA isolation from fresh frozen microdissected tissue.

2.9.2 qPCR

Due to the small amount of starting material of isolated RNA from microdissected segments, the TaqMan™ probe and commercial primers were employed for the real-time quantitative-polymerase chain reaction (rt-qPCR). This experiment provided a more sensitive and optimised method for obtaining data. The SuperScript™ VILO™ cDNA Synthesis Kit was used to reverse transcription to employ extracted RNAs in qPCR. In a 25 µL complementary DNA (cDNA) kit, the concentration of extracted RNAs was normalised to 7.5ng. The reactions were placed on ice after being incubated in a thermocycler. In a total volume of 10 µL, the RT-qPCR analysis was carried

out with 2 µL (0.5 ng) of template DNA, 0.15 µL of each primer, 5 µL of TaqMan Master Mix, and 2.7 µL sterile distilled water. Quantitative rt-qPCR reactions were performed under the following conditions: 95°C for 3 min followed by 40 cycles for Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) and 60 cycles for genes listed in **Table 2.9.2.1,** of 95°C for 5 sec, 60°C for 1 min and 72°C for 15 sec. A no-template control was used as a negative control. For data analysis, the $2^{-\Delta\Delta}$Ct method and mean fold changes in gene expression relative to mock (healthy) control were used. All preparations were carried out as three analytical duplicates for each extracted RNA sample. GraphPad Prism was used to represent the findings of the final statistical test.

Table 2.9.2.1. Commercial primers and catalogue numbers were used to quantify gene expression using rt-qPCR.

Gene	Catalogue Number
PIK3CA	4448892
TP53	4331182
CEL	4448892
LPIN1	4453320
GAPDH	4453320

2.10 Laser Capture Microdissection

The Leica LCM 7000 system uses the non-contact method. The laser was used to cut the area of interest, and the microdissected areas were collected directly into the cap of a PCR tube with the appropriate buffer (RLT) for further analysis. After that, sections were placed directly on Polyethylene naphthalate (poly(ethylene 2,6-naphthalate or PEN) type slides. Laser and microscope control, as well as power: 28, speed: 7, specimen balance: 15, and pulse frequency: 120, have all been calibrated in the LCM. **Figure 2.10. 1**. shows a general information layout of the LCM software setup, including ROI selection, microdissection, and ROI collection.

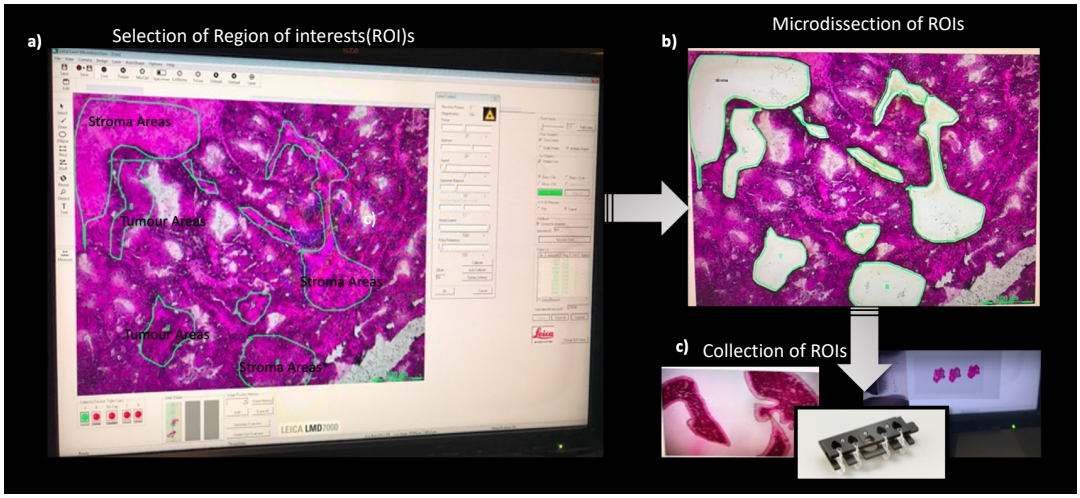


Figure 2.10. 1. General schematic of the live-view LCM set up in the software. a) Selection of the H&E stained ROIs, b) microdissection of ROIs, c) collection of ROIs.

2.11 RNA sequencing sample preparation and Analysis

RNA integrity and quantitation were assessed using the RNA Nano 6000 Assay of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). RNA purity was tested using the NanoPhotometer® and spectrophotometer (IMPLEN, CA, USA). RNA degradation and contamination were monitored on 1% agarose gels.

2.11.1 Smart Kit Amplification

SMART technology (www.takarabio.com) adaptations for next-generation sequencing (NGS) were incorporated into the first generation of our ultra-low input mRNA-seq kit (the SMARTer® Ultra Low RNA Kit for Illumina Sequencing) and published as the SMART-Seq method [160]. This protocol has been optimised for cDNA synthesis, starting with 10 pg of total RNA. Total RNA should be purified in nuclease-free water. **Figure 2.11.1. 1**. depicts a schematic outline of the technology and workflow.

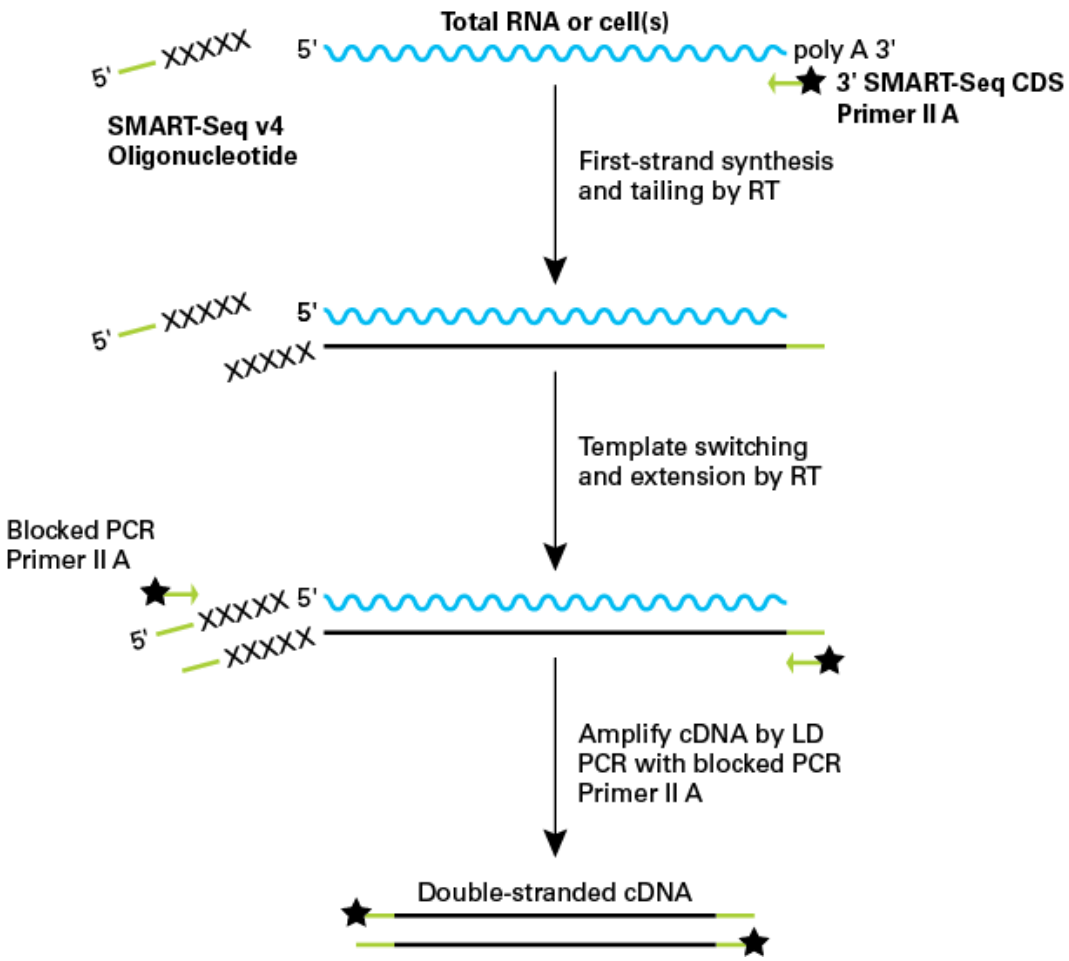


Figure 2.11.1. 1. Flowchart of SMART cDNA synthesis. The SMART-Seq v4 Oligonucleotide, 3' SMART-Seq CDS Primer II A, and PCR Primer II A all contain a stretch of identical sequence. The black star indicates a chemical block on the 5' end of the oligonucleotide.

2.11.1.1 Library preparation for transcriptome sequencing

Following the manufacturer's recommendations, sequencing libraries were generated using NEBNext® Ultra™RNA Library Prep Kit for Illumina™(NEB, USA). A total amount of 1 µg RNA per sample was used as input material for the RNA sample preparations. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was performed in NEBNext First Strand Synthesis Reaction Buffer (5X) using divalent cations under elevated temperature. First-strand cDNA was synthesised using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H). Second strand cDNA synthesis was performed using DNA Polymerase I and

RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridisation. To select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 µl USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37 °C for 15 min, followed by 5 min at 95 °C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system), and library quality was assessed on the Agilent Bioanalyzer 2100 system.

2.11.1.1.1 Clustering and sequencing

The index-coded samples were clustering on a cBot Cluster Generation System using P.E. Cluster Kit cBot-HS (Illumina). After cluster generation, the library preparations were sequenced on an Illumina platform, and paired-end reads were generated.

2.11.1.2 Data Analysis Quality control

Raw data (raw reads) of FASTQ format were firstly processed through in-house scripts[161]. This step obtained clean data (clean reads) by removing reads containing adapter and poly-N sequences and low-quality reads from raw data. At the same time, in Q20, the pure data's Q30 and G.C. content were calculated[126,162].

2.11.1.2.1 Error Rate

The Phred score could transform the error rate for each base as in equation 1 (equation 1: $Q_{phred} = -10\log_{10}{}^{e}$). The relationship between Phred quality scores Q and base-calling error "e" were given in **Table 2.11.1.2.1. 1**:

Table 2.11.1.2.1. 1. Base Quality and Phred score relationship with the Illumina CASAVA software.

Phred score	The base Calling error rate	Base Calling correct rate	Q-score
10	1/10	90%	Q10
20	1/100	99%	Q20
30	1/1000	99.9%	Q30
40	1/10000	99.99%	Q40

Sequencing error rate and base quality were determined by sequencer, reagent availability, and samples.

(1) Error rate increases as reads extend and sequencing reagents run out. (2) A relatively high error rate was standard at the first six bases due to using random hexamer primers in cDNA synthesis.

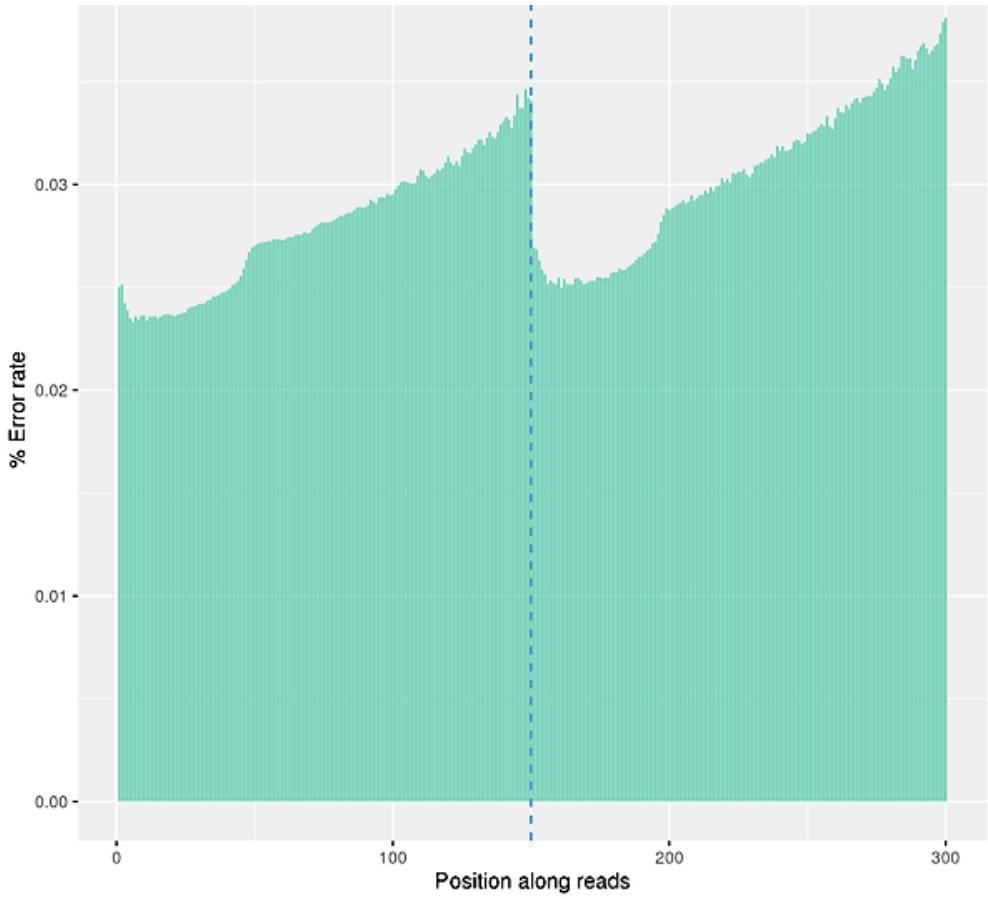


Figure 2.11.1.2.1. 1. An example of Error Rate Distribution. The x-axis shows the base position and each sequencing read, and the y-axis shows the base error rate.

2.11.1.2.2 A/T/G/C Content Distribution

G.C. content distribution was evaluated to detect potential AT/GC separation, affecting gene expression quantification. Theoretically, the number of G equals the number of C, and the number of A equals the number of T throughout the sequencing process for non-stranded libraries. However, AT/GC separation is generally observed in stranded libraries. For Digital Gene Expression (DGE) libraries, more errors in the first 6-7 bases are allowed due to using random primers in library construction.

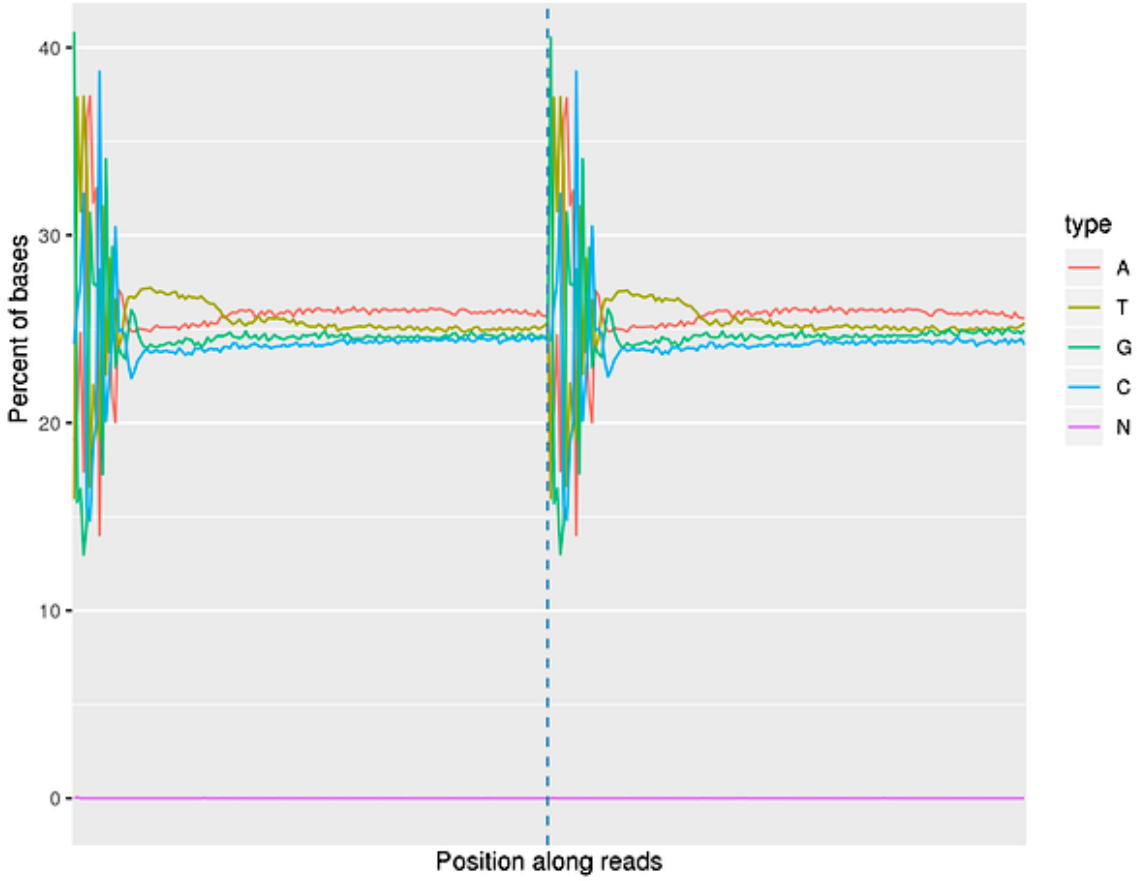


Figure 2.11.1.2.2. 1. An example of GC content distribution. The x-axis shows each base position within a read, and the y-axis shows the percentage of each base, with each base represented by a different colour.

2.11.1.2.3 Data Filtering

Raw reads were filtered to remove reads with adapter contamination or low-quality ones. Only clean reads were used in the downstream analyses, **Figure 2.11.1.2.3. 1** and **Table 2.11.1.2.3. 1**.

The filtering process was as follows:
(1) Discard reads with adaptor contamination.
(2) Discard reads when uncertain nucleotides constitute more than 10% of either read (N > 10%).
(3) Discard reads when low-quality (base quality less than 20) constitutes more than 50% of the read.
RNA-seq Adapter sequences (Oligonucleotide sequences of adapters from TruSeqTM RNA and DNA Sample Prep Kits):

RNA 5' Adapter (RA5):
5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
RNA 3' Adapter (RA3):
5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC (6-nucleotide index) ATCTCGTATGCCGTCTTCTGCTTG-3'

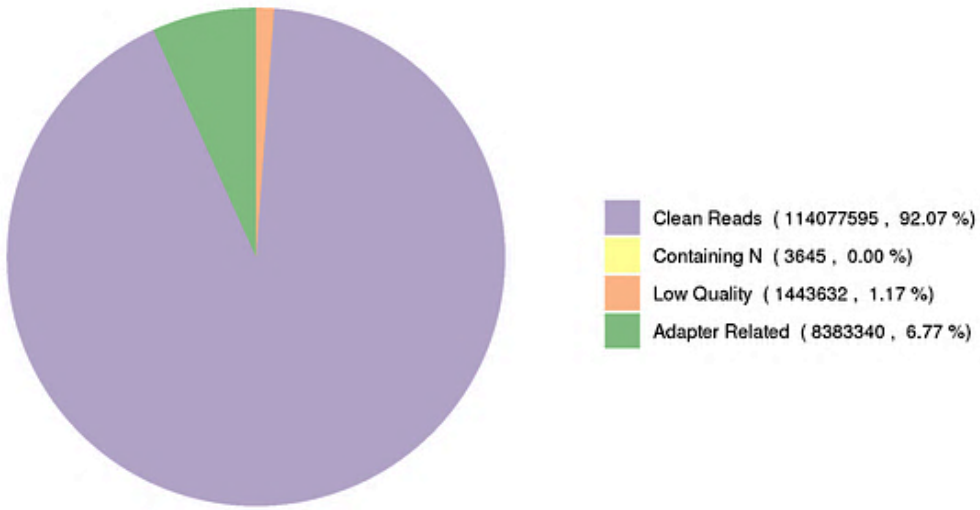


Figure 2.11.1.2.3. 1. An example of Raw Reads Components. Results are shown as a percentage of total raw reads. ((1) Clean Reds: reads that passed quality control (2) Containing N: reads in which uncertain nucleotides constituted more than 10 per

cent of the read. (3) Low Quality: reads in which low-quality nucleotides constituted more than 50 per cent of the read. (4) Adapter related: reads that had adapter contamination.

Table 2.11.1.2.3. 1. Overview of data production quality.

Sample	raw_reads	clean_reads	raw_data(G)	clean_data(G)	error_rate(%)	Q20(%)	Q30(%)	GC_content(%)
OC	123908212	114077595	37.2	34.2	0.03	96.07	90.99	48.75

Detail statistics of sequencing data:

(1) Sample: The names of samples.

(2) raw_Reads: The original sequencing reads counts.

(3) clean_Reads: Number of reads after filtering

(4) raw_data: Raw reads number multiply read length, saved in G unit.

(5) clean_data: Clean reads number multiply read length, kept in G unit.

(6) error_rate: Average sequencing error rate calculated by Qphred=-10log10(e).

(7) Q20: Percentages of bases whose correct base recognition rates are greater than 99% in total bases.

(8) Q30: Percentages of bases whose correct base recognition rates are greater than 99.9% in total bases.

(9) GC_content: Percentages of G and C on total bases.

2.11.1.3 Mapping to reference genome

Reference genome and gene model annotation files were downloaded directly from the genome website browser,2019 version (NCBI/UCSC/Ensembl). Paired-end clean reads were mapped to the reference genome using HISAT2 software[163]. HISAT2 uses a large set of small GFM indexes that collectively cover the whole genome. These small indexes (called local indexes), combined with several alignment strategies, enable rapid and accurate alignment of sequencing reads. HISAT2 is a fast and sensitive alignment program mapping next-generation sequencing reads, **Figure 2.11.1.3. 1**. HISAT2 uses a graph-based alignment and has succeeded HISAT and

TOPHAT2. Besides one global GFM index, HISAT2 also includes a large set of small GFM indexes that collectively cover the whole genome. Combined with multiple alignment strategies, these small indexes (local indexes) enabled effective alignment of RNA-seq reads, mainly spanning multiple exons.

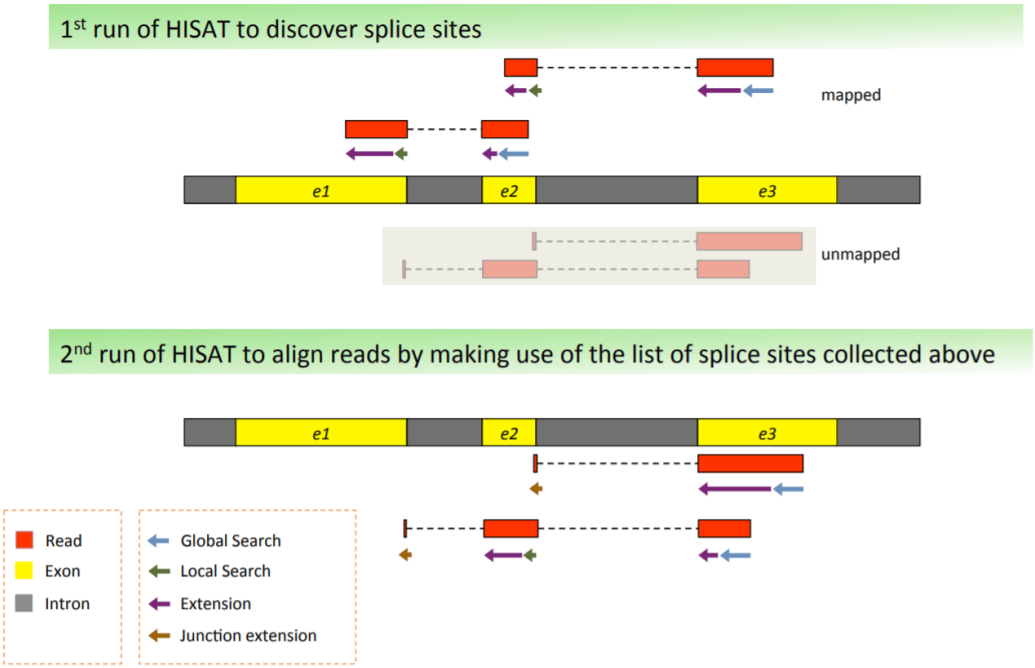


Figure 2.11.1.3. 1. Two-step approach version of HISAT, allowing alignment of junction reads with small anchors.

Table 2.11.1.3. 1. Overview of Mapping Status.

Sample name	O.C.
Total reads	228155190
Total mapped	212421204 (93.10%)
Multiple mapped	7907480 (3.47%)
Uniquely mapped	204513724 (89.64%)
Read-1	102736654 (45.03%)
Read-2	101777070 (44.61%)
Reads map to '+'	102467164 (44.91%)
Reads map to '-'	102046560 (44.73%)
Non-splice reads	114727626 (50.28%)
Splice reads	89786098 (39.35%)

Mapping Results Details:

(1) Total reads: Total number of filtered reads (Clean data).

(2) Total mapped: Total number of reads that can be mapped to the reference genome.

(3) Multiple mapped: Number of reads that can be mapped to multiple sites in the reference genome.

(4) Uniquely mapped: Number of reads that can be uniquely mapped to the reference genome.

(5) Read-1: Number of left reads that can be mapped to the reference genome.
(6) Read-2: Number of right reads that can be mapped to the reference genome.
(7) Reads map to '+': Number of reads that can be mapped to the positive strand (+).
(8) Reads map to '-': Number of reads that can be mapped to the minus strand (-).
(9) Non-splice reads: Number of reads that can be mapped entirely to a single exon.
(10) Splice reads: Number of reads that can be segmented and mapped to two exons (also named junction reads).

2.11.1.3.1 Mapped Regions in Reference Genome

Mapped regions can be classified as exons, introns, or intergenic regions. When the reference genome is well-annotated, exons should be the most abundant type of mapped reads. Introns may be derived from pre-mRNA contamination or intron-retention events from alternative splicing. Reads mapped to intergenic regions are mainly because of weak annotation of the reference genome.

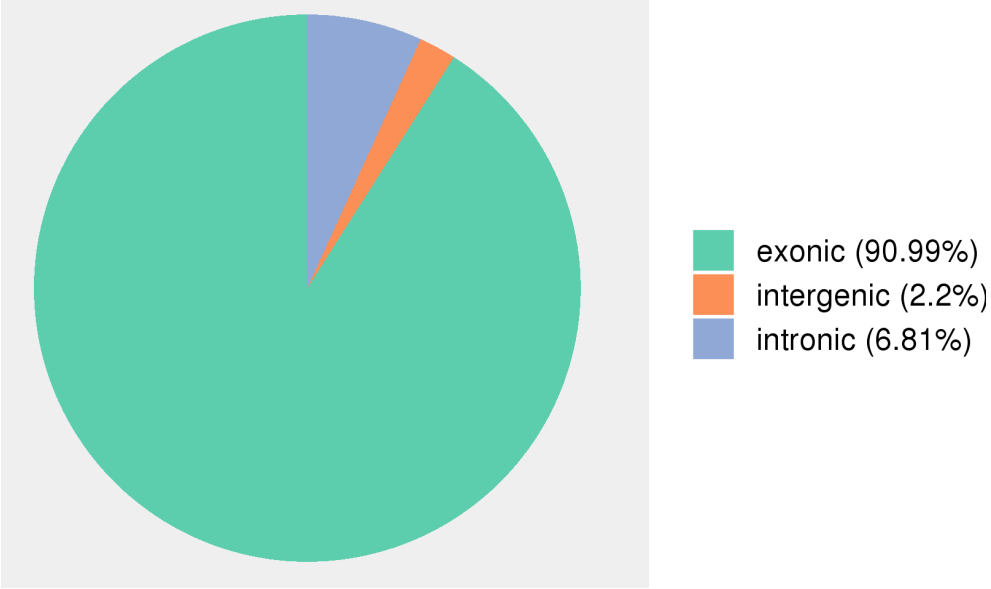


Figure 2.11.1.3.1. 1. An example of Classification of Reads According to Mapped Region.

2.11.1.3.2 Visualisation of Mapping Status of Reads

Files were provided in BAM format, a standard file format that contains mapping results. The Integrative Genomics Viewer (IGV)[164] is recommended to visualise data from BAM files.

The IGV has several features:

(1) It displays the positions of single or multiple reads in the reference genome, as well as read distribution between annotated exons, introns or intergenic regions, both on the adjustable scale;

(2) It displays the read abundance of different regions to demonstrate their expression levels on the adjustable scale;

(3) It provides annotation information for both genes and splicing isoforms;

(4) It provides other related annotation information;

(5) It displays annotations downloaded from remote servers and/or imported from local machines.

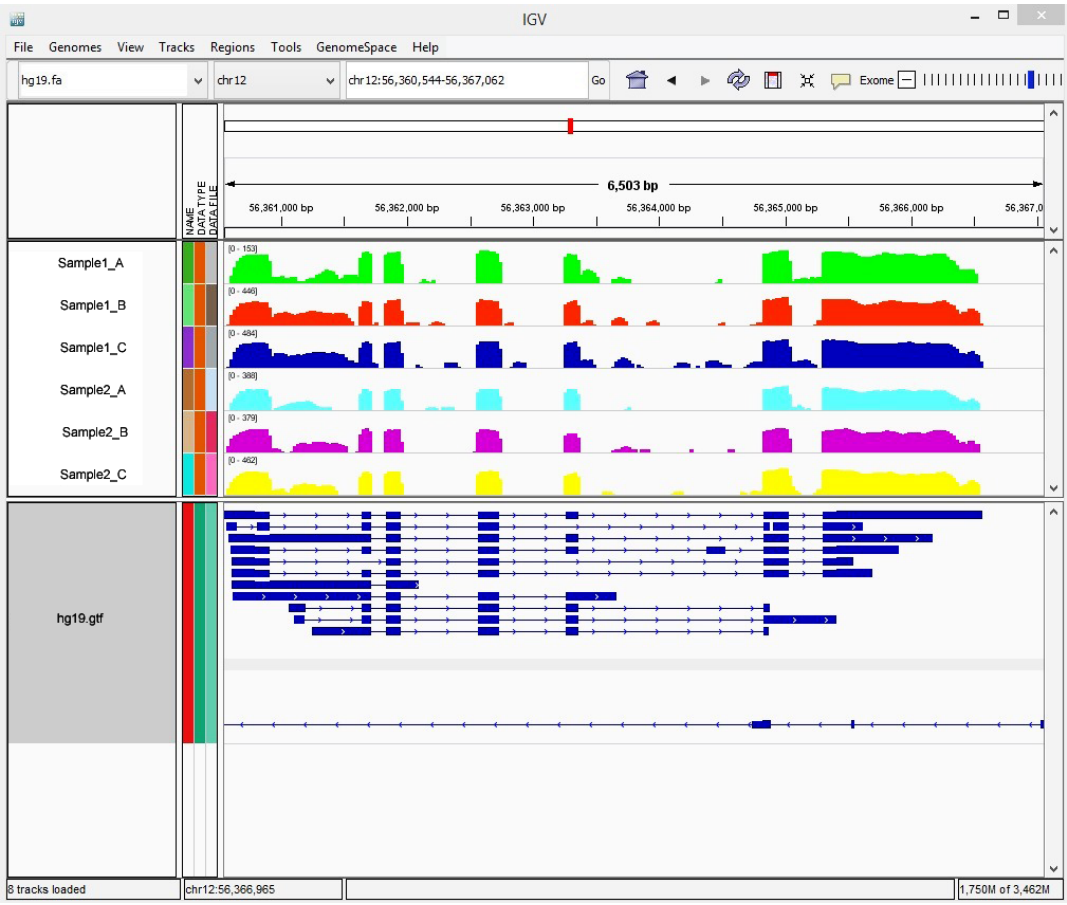


Figure 2.11.1.3.2. 1. Interface of Integrative Genomics Viewer (IGV)[164]

2.11.1.3.3 Novel gene prediction

Because transcriptome annotations are still incomplete, most RNA-seq studies will reveal novel genes and transcripts. The Cufflinks Reference Annotation Based Transcript (RABT)[165] assembly method was used to assemble the set of transcript isoforms of each bam file obtained in the mapping step. Cufflinks include a utility program called 'Cuffcompare' that can compare Cufflinks assemblies to reference annotation files and help sort out new genes from known ones. The Cuffcompare class codes underlying transcript classification were: "u" for "Potentially novel gene"; the set of "j", "u", "i", "o" and "x" for "Potentially novel isoform".

2.11.1.4 Quantification

HTSeq was used to count the read numbers mapped of each gene, including known and novel genes[166]. And then, the Reads Per Kilobase of exon model per Million mapped reads (RPKM) of each gene was calculated based on the length of the gene and the reads count mapped to this gene. RPKM considers the effect of sequencing depth and gene length for the read count simultaneously and is currently the most commonly used method for estimating gene expression levels.

Gene expression level was measured by transcript abundance. The greater the abundance, the higher the gene expression level. Our RNA-seq analysis estimated gene expression levels by counting the reads that map to genes or exons. The read count was proportional to the actual gene expression level, gene length, and sequencing depth. In RNA-seq, FPKM, standing for the expected number of Fragments Per Kilobase of transcript sequence per Million of base pairs sequenced, is the most common method of estimating gene expressions, which considers both effects of sequencing depth and gene length on the counting of fragments. For gene expression levels estimated from different genes and experiments to be comparable, the FPKM is used[165]. HTSeq software was used to analyse gene expression levels in this experiment, using the union mode. The result files present the number of

genes with different expression levels and the expression level of single genes. In general, an FPKM value of 1 is set as the threshold for determining whether the gene was expressed or not.

Table 2.11.1.4. 1. The number of genes with different expression levels.

FPKM_Interval	OC
0~1	44430(73.5%)
1~3	3748(6.2%)
3~15	5788(9.58%)
15~60	4088(6.76%)
>60	2394(3.96%)

2.11.1.5 Comparison between Gene Expression Levels

An FPKM distribution diagram and Box Plot were used to compare gene expression levels under different conditions. For biological replicates, the final FPKM would be the mean value.

2.11.1.6 RNA-Seq Correlation

The correlation between samples was an important indicator for testing the experiment's reliability. The closer the correlation coefficient was to 1, the greater the similarity of the samples. The Encyclopedia of DNA Elements (ENCODE) suggested that the square of the Pearson correlation coefficient should be more significant than 0.92 under ideal experimental conditions. In this study, the R^2 should be more powerful than 0.8.

2.11.1.7 Differential expression analysis

Differential expression analysis between two conditions/groups (three biological replicates per condition) was performed using the DESeq2 R package. DESeq2 provides statistical routines for determining differential

expression in digital gene expression data using a model based on the negative binomial distribution[167,168]. The resulting P values were adjusted using Benjamini and Hochberg's approach for controlling the False Discovery Rate (FDR). Genes with an adjusted P-value < 0.05 found by DESeq2 were assigned as differentially expressed. Before differential gene expression analysis, for each sequenced library, the read counts were adjusted by edgeR program[169] through one scaling normalised factor. Differential expression analysis of two conditions was performed using the DEGseq R package. The P values were adjusted using the Benjamini and Hochberg methods. Corrected p-value of 0.05 and |log2(Fold Change)| of 1 were set as the threshold for significantly differential expression.

2.11.1.7.1 List of Differentially Expressed genes

The input data for differential gene expression analysis were read counts from gene expression level analysis. The differential gene expression analysis contained three steps:

1) Read counts Normalisation
2) Model-dependent p-value estimation
3) FDR value estimation based on multiple hypothesis testing

Different software and parameter sets were applied in different situations. The analysis methods are listed below:

The read count value of the i-th gene in the j-th sample is K_{ij}, then Negative binomial distribution: $K_{ij} \sim N.B. (\mu_{ij},\sigma_{ij}^2)$

Poisson distribution: $K_{ij} \sim P(\mu_{ij})$

Type	Software	Normalisation method	p-value estimation model	FDR estimation method	Differentially expressed gene screening standard
With biological duplicates	DESeq2	DESeq	negative binomial distribution	BH	padj < 0.05
Without biological duplicate	DEGseq	TMM	Poisson distribution	B.H.	$\lvert\log_2(\text{FoldChange})\rvert > 1$&padj < 0.005

Table 2.11.1.7.1. 1. Example of top 4 Differentially Expressed Genes.

Gene Id	PIK3CA.value	WT.value	log_2FoldChange
ENSG00000205076	1980.38921839784	0.562303106839183	11.7821449665148
ENSG00000189334	1038.0063429481	0.383511146375281	11.4022560309317
ENSG0000017 8363	151.248920301591	0.0405710623126586	11.8641821955245
ENSG00000186081	3522.18727857515	1.85309255748777	10.892318560581

Differentially Expressed Genes List includes:

(1) Gene Id: Gene Id.

(2) *PIK3CA*. Value: The mean of normalised FPKMs in group *PIK3CA*.

(3) W.T.value: The mean of normalised FPKMs in group WT.

(4) log_2FoldChange: log_2(*PIK3CA*/WT).

2.11.1.7.2 Screening of differentially expressed genes

Volcano plots were used to infer the overall distribution of differentially expressed genes. For experiments without biological replicates, the threshold was normally set as: $|log_2(\text{Fold Change})| > 1$ and padj < 0.005. Our threshold for experiments with biological replicates was normally set as: padj < 0.05.

2.11.1.8 Cluster Analysis of Gene Expression Differences

Cluster analysis was used to find genes with similar expression patterns under various experimental conditions. In hierarchical clustering, areas of different colours denote different groups (clusters) of genes, and genes within each cluster may have similar functions or participate in the same biological process. Clustering genes with similar expression patterns made it possible to discern unknown functions of previously characterised genes or functions of unknown genes.

In addition to the FPKM cluster, the H-cluster, K-means and self-organizing map (SOM) were also used to cluster the log_2 (ratios). Genes within the cluster exhibited the same trend in expression levels under different conditions.

2.11.1.9 Enrichment analysis

A common way to search for shared functions among genes is to incorporate the biological knowledge provided by biological ontologies. The Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg/) annotates genes to pathways. Gene Ontology (GO, http://geneontology.org/docs/go-enrichment-analysis/) annotates genes to biological processes, molecular functions, and cellular components in a directed acyclic graph structure.

2.11.1.9.1 GO enrichment analysis

Gene Ontology is a major bioinformatics initiative to unify the presentation of gene and gene product attributes across all species. GO enrichment analysis of differentially expressed genes was implemented by the GOseq R package[170], correcting gene length bias. Differentially expressed genes significantly enriched GO terms with a corrected P value of less than 0.05. DEGs refer to differentially expressed genes. GO GOseq used enrichment analysis based on Wallenius' non-central hypergeometric distribution. Its characteristics are: the probability of drawing an individual from a particular category differs from pulling it from outside the class. This difference is obtained from estimating the preference of gene length.

2.11.1.9.2 KEGG Pathway enrichment analysis

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (http://www.genome.jp/kegg/)11. KOBAS software[172] was used to test the statistical enrichment of differentially expressed genes in KEGG pathways.
The interactions of multiple genes may be involved in specific biological functions. KEGG is a collection of manually curated databases dealing with

genomes, biological pathways, diseases, drugs, and chemical substances. KEGG is utilised for bioinformatics research and education, including data analysis in genomics, metagenomics, metabolomics and other omics studies. Pathway enrichment analysis identifies significantly enriched metabolic pathways or signal transduction pathways associated with differentially expressed genes than the whole genome background.

2.11.1.10 Mutation analysis

Firstly, Picard tools and Samtools were used to sort and mark duplicated reads and reorder the bam alignment results of each sample. Then the tool HaplotypeCaller in The Genome Analysis Toolkit (GATK) software[173] was used to perform variant discovery. Raw VCF files were filtered with GATK standard filter method and using the following parameters: cluster, 3; window size, 35; QD < 2.0 or FS > 30.0. Finally, ANNOVAR was used to functionally annotate genetic variants detected from diverse genomes with user-specified versions of genome builds.

2.11.1.11 Summary Workflow of RNA sequencing

From the RNA sample to the final data, data quality was tested at every step, including sample testing, library preparation, and sequencing. The quality of the data had a direct impact on the analysis results. Quality control (QC) was undertaken at each step of the method to ensure the data's reliability. The following is the procedure represented in **Figure 2.11.1.11. 1**.

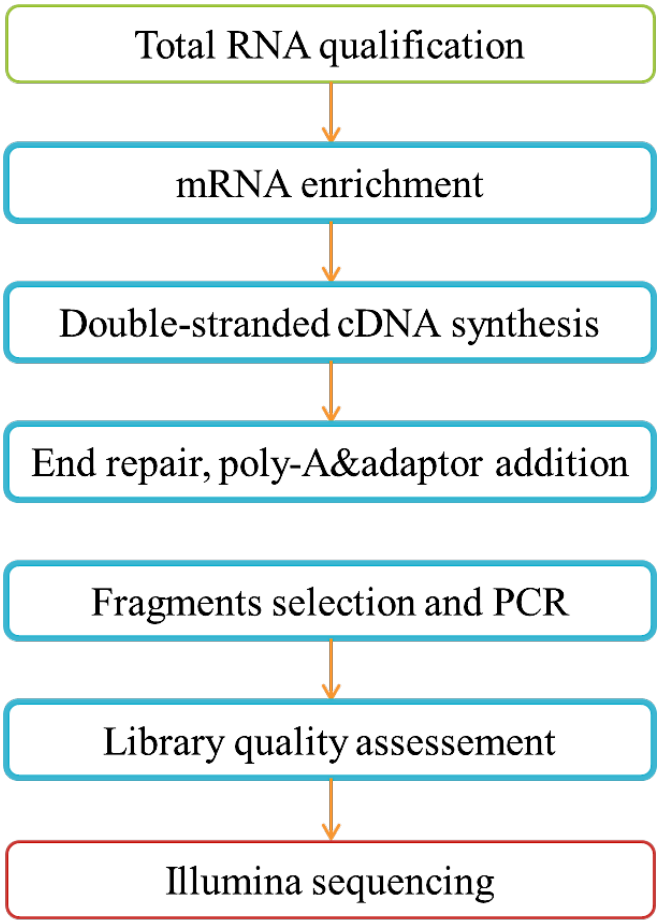


Figure 2.11.1.11. 1. Summary Workflow of Experimental Procedure.

The analysis workflow is shown in **Figure 2.11.1.11. 2**. data with a reference genome was as follows, and the software used in the analysis is listed in **Table 2.11.1.11. 1**.

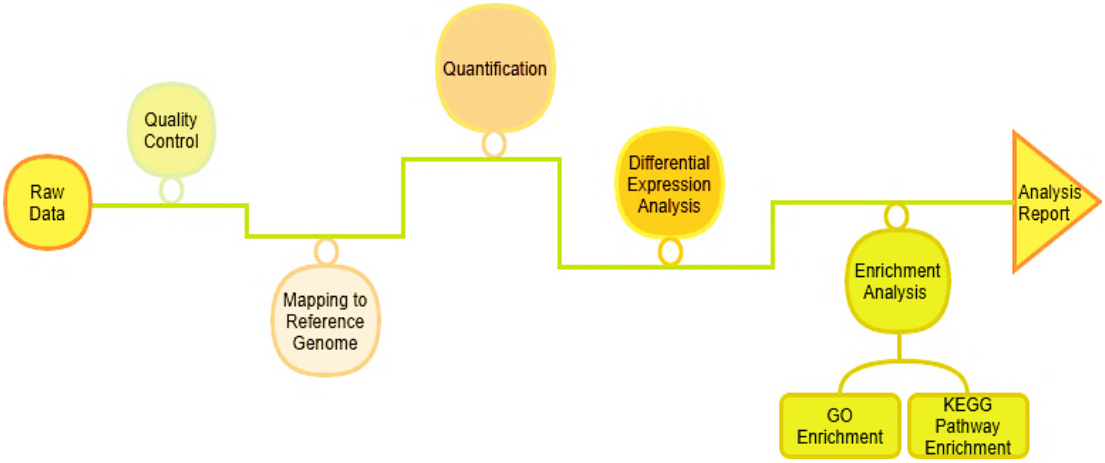


Figure 2.11.1.11. 2. Summary Workflow of the data analysis procedure.

Table 2.11.1.11. 1. The software used in the Data analysis.

Analysis	Software	Version	Parameters	Remarks
Mapping to reference genome	HISAT2	V2.0.5	default parameter	
Quantification	HTSeq	v0.6.1	-m union	
Differential Expression Analysis	DEGseq	v1.34.1	|$\log_2$Fold change| > 1 && Padj < 0.005	Without biological replicates
	DESeq2	v1.20.0	Padj < 0.05	With biological replicates
GO Enrichment	GOSeq, topGO,hmmscan	v1.34.1	Padj < 0.05	padj < 0.05 were considered significantly enriched
KEGG Pathway Enrichment	KOBAS	v3.0	Padj < 0.05	padj < 0.05 were considered significantly enriched

2.12 The tissue embedding protocol

2.12.1.1 Chemical Materials

Gelatin, poly N-(hydroxypropyl)-methacrylate, polyvinylpyrrolidone (PVP,MW 360 kDa), (Hydroxypropyl)-methylcellulose (HPMC) (viscosity 40−60 cP, 2% in H2O (20 C)), low and medium viscosity Na-carboxymethyl cellulose, 2,5-dihydroxybenzoic acid (DHB), trifluoroacetic acid (TFA), Triton-X 100, and 9-aminoacridine (9-AA) were purchased from Merck (Darmstadt, Germany). Optimal cutting temperature medium, methanol, water, iso-pentane, isopropyl

alcohol, and acetonitrile were obtained from Fisher Scientific (Waltham, MA). All solvents used were of analytical grade or higher.

2.12.1.2 Animal Models

Adult male Hans Wistar rats were sourced from Charles River Laboratories (Margate, Kent, UK) and reacclimated on-site for a minimum of 3 days. All tissue dissection was carried out by trained AstraZeneca specialists agreeing with institutional review panel standards (AstraZeneca Project licence 40/3484, procedure number 10).

2.12.1.3 Embedding of biopsies

Gelatine, N-(2-Hydroxypropyl)-methacrylamide (HPMA), Sodium-Carboxymethylcellulose (CMC), (Hydroxypropyl-) methylcellulose (HPMC), and Polyvinylpyrrolidone were used to embed tissue (PVP). All cellulose derivatives and PVP were thoroughly mixed with Milli-Q water (Merck, Darmstadt, Germany) and stored at 4°C overnight to completely dissolve. A gelatine solution was prepared by dispersing 10% (w/w) gelatine in water at 70°C with continuous stirring until the gelatine was fully dissolved. All solutions were cooled on ice for half an hour before tissue insertion. All tissue embedding was completed on ice except gelatine embedding procedures to avoid warming the embedding fluid. Before adding rat tissue specimens, embedding moulds were filled with each embedding protocol media for RNA and DNA quality evaluation. For the core biopsies, **Figure 2.12.1.3. 1a,** ready-made HPMC/PVP hydrogel moulds with Simport® T-Sue™ microarray (TMA) moulds, **Figure 2.12.1.3. 1b-c** were used to embed small size (3-6 cm length) core biopsies, **Figure 2.12.1.3. 1d,** and sectioned on standard glass slides for DESI-MSI, **Figure 2.12.1.3. 1e.** The moulds were snap-frozen in ice-chilled isopentane and isopropanol and stored in the -80 °C freezer until entirely frozen[174].

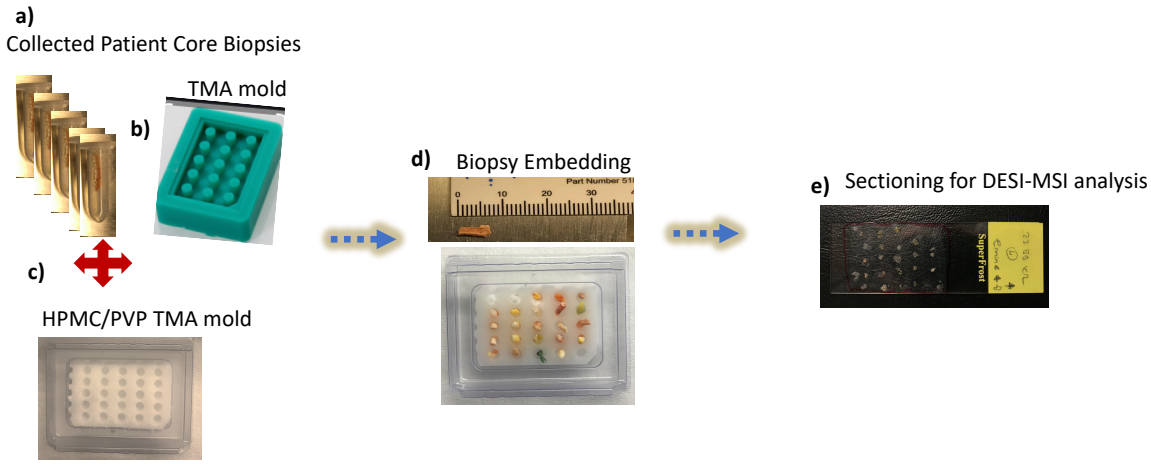


Figure 2.12.1.3. 1. A schematic workflow of Core Biopsies Embedding. a) Patient core biopsies b) TMA The Simport® T-Sue™ microarray (TMA) moulds c) pre-casted HPMC/PVP hydrogel d) placed and embedded fragile breast cancerous core biopsies e) sectioned block on a glass slide for DESI-MSI.

2.12.1.4 Tissue Sectioning

Sectioning was performed on a Leica Biosystems CM3050 S cryostat (Nussloch, Germany) with a section thickness of 10 μm. Milli-Q water (Merck, Darmstadt, Germany) was used to mount the specimens onto the sample holder, **Figure 2.12.1.4. 1a.** The chamber temperature was kept at -20 °C, while the sample temperature was retained at -18 °C. The cryostat chamber and all the tools were sterilised using RNAzap to eliminate RNases to reduce RNA degradation during sample preparation. Then, the cleaning was followed by ethanol to remove commercially available cleaning agents before sectioning (RNase away, Molecular BioProducts, San Diego, CA). An example of DESI-MSI tissue sections was ready for thawed and positioned on SuperFrost glass TM microscope slides (Thermo Scientific, Waltham, MA), **Figure 2.12.1.4. 1b**. Tissue sections were inserted into RNase-free tubes for DNA and RNA extraction (Eppendorf, Hamburg, Germany).

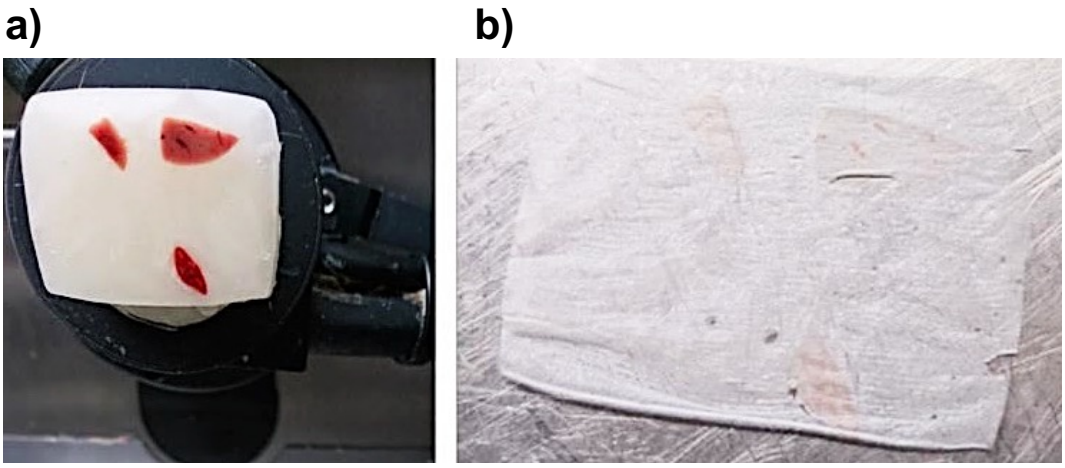


Figure 2.12.1.4. 1. An example section of HPMC/PVP hydrogel embedded samples. a) sample mounted on the specimen holder. b) Ready section of HPMC/PVP hydrogel embedded tissue to place on glass slides for DESI-MSI.

2.12.1.5 RNA extraction and analysis

The RNeasy Micro Kit (Qiagen, Valencia, CA, USA) was used for total RNA extraction. Total RNA was isolated using the used according to the manufacturer's recommendations for microdissected cryosections. Briefly, RLT buffer (a guanidine-thiocyanate containing lysis buffer including 2-Mercaptoethanol (2-M.E.) as a reducing agent) was added into the Eppendorf tubes with tissue sections, and the solution was transferred onto a filter column. To remove genomic DNA, the RNA bound to the column was incubated for 20 minutes with 10 μL of DNase diluted in RDD buffer (RNase-Free DNase Set, Qiagen, Valencia, CA, USA). Nuclease-free water eluted the RNA from the purification column into RNase-free tubes. The RNA extracts were stored at –80°C until they were examined.

Microfluidic capillary electrophoresis was used to measure RNA quality using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) with Agilent RNA 6000 Pico Lab Chip Kit (Agilent Technologies, Santa Clara, CA, USA). The RNA Integrity Number (RIN) was used to evaluate the RNA degradation of each sample. The RIN number was calculated using the 2100 Bioanalyzer Expert software algorithm. In detail, RNA quality in each sample was assessed on the fragment size distribution, including two peaks corresponding to 18S and 28S ribosomal RNAs and a small molecular weight

area signal related to short RNAs. This distribution was then compared to an array of RNA samples to generate a RIN. RNA purity and concentration were ascertained using a NanoDrop ND-1000 UV-Vis Spectrophotometer, and fluorometric quantification was measured using the RNA HS Assay kit (Thermo Fisher Scientific, Waltham, MA, USA) with a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). To adjust the comparison for the average amounts per embedded tissue, the per tissue areas of consecutive sections were H&E stained. The area used in extraction was normalized based on this. For each condition, all preparations were performed in technical triplicate.

2.12.1.6 cDNA preparation and RT-qPCR

The Quantitect Reverse Transcription Kit was used to reverse-transcribe isolated RNAs (Qiagen, Germany). The extracted RNA content in RNase-free water was normalized to 100ng/12 μL. Genomic DNA was eliminated by the reaction buffer consisting of 2 μL gDNA Wipeout 7x Buffer and 12 μL template RNA. The reactions were incubated for 2 min at 42°C and immediately placed on ice. The reverse-transcription master mix consisted of 4 μL of Quantiscript 5xRT buffer (includes Mg^{2+} and dNTPs), 1μl of R.T. primer mix, 1 μL of Quantiscript Reverse Transcriptase (containing Rnase inhibitor), and 14 μl Template RNA. The samples were then incubated for 30 min at 42°C, followed by an incubation of 3 min at 95°C. The reaction tubes were subsequently transferred onto the ice before RT-qPCR.

RT-qPCR analysis was performed in a total volume of 10 μL, containing 1 μL (5 ng) of template DNA, 0.5 μM of each primer, **Table 2.12.1.6. 1**, 5 μL of SYBR Select Master Mix (Applied Biosystems, USA), and 3 μL sterile distilled water. Quantitative RT-qPCR reactions were performed using the professional Thermocycler (Biometra, Germany) under the following conditions: 95°C for 3 min followed by 40 cycles for GAPDH and 60 cycles for PPIB of 95°C for 5 s, 60°C for 1 min and 72°C for 15s. A no-template sample was used as a negative control. For data analysis, the $2^{-\Delta\Delta}$Ct technique was utilized. The mean fold changes of the housekeeping genes glyceraldehyde 3-phosphate

dehydrogenase (GAPDH) and peptidyl-prolyl cis/trans isomerase B (PPIB) to the mock control were calculated. Each extracted RNA sample was performed in an analytical duplicate. For multiple comparisons, two-tailed t-tests were used to determine the significance and the absolute distance of normalized expression value from the control mean.

Table 2.12.1.6. 1. Primers are used to quantify gene expression using RT-qPCR.

Gene	Forward Primer (3'-5')	Reverse Primer (5'-3')
GAPDH	GACATGCCGCCTGGAGAAAC	AGCCCAGGATGCCCTTTAGT
PPIB	CAAGACCTCCTGGCTAGACG	GCTTAGGTGCCTTTGGGGT

2.13 Data Analysis

2.13.1 DESI-MSI

2.13.1.1 DESI-MSI Pre-processing of the datasets

DESI-MSI detects *m/z* species in the order of thousands; solvent and noise-related peak subtraction features are essential to ensure that the extracted profiles are of genuine biological relevance. Data for negative ion modes were acquired in the 50-1000 *m/z* range. GPLs in the 600-1000 *m/z* range and Fatty acids were analysed in the 200-400 *m/z* range. Although the Xevo QToF instruments are calibrated before each DESI-MSI run, non-linear peak shifts are quite common. Before performing multiple sample comparisons, a peak alignment and noise/baseline subtraction algorithm developed in-house was used (dynamic programming, 10 ppm matched). Subsequent data pre-processing of raw data was then performed on the selected *m/z* ranges of interest. From the selected regions/pixels of interest, mean average spectra were calculated and treated as single data units. MS data pre-processing was undertaken using an in-house developed algorithm[175] in Python. The lists consisting of *m/z* values and relative signal intensities for each feature detected were generated using an in-house algorithm. Further multivariate analysis was applied after pre-processing.

2.13.1.2 Normalisation

The aligned data were TIC normalised to account for differences in signal intensities of molecular species. These variations in intensity can arise due to instrument variations and differences in cellular density in these heterogeneous tissues. By imposing TIC normalisation, the same overall sum intensity of peaks is allocated for a given mass range, the same analyses sample before. Therefore, the only difference will be the relative abundance of the spectral features within that mass range.

2.13.1.3 Multivariate Statistical Analysis

The processed data were subjected to supervised and unsupervised analysis strategies to reduce their complexity and identify spectral features of discriminatory power. Multivariate statistical methods allow finding similarities and differences between samples in a dataset. In an unsupervised strategy, such as Principal Component Analysis (PCA), the classification is based on data analysis without the user providing sample classes. The method provides a relative measure of the relationship between samples and is therefore used to identify the most significant variance without external factors. PCA is an unsupervised method aiming to find the Orthogonal factors and directions that best explain the variance in a dataset. The data were summarised into fewer variables called scores, a weighted average of the original variables (loadings). The weighting profiles are called loadings.

2.13.1.3.1 Principal component Analysis

PCA can be applied to the extracted mass spectra from selected regions or the entire tissue section in MSI. In this thesis, the average DESI-MSI profiles extracted from different patients were analysed by PCA. PCA was used to identify trends between various tissue classes based on their lipidomic profiles without any prior histological information. The results from the PCA were presented in the form of score plots and the percentage of variance for a given principal component. The first three components explained most variation in

most cases. The closer the points were on the scores plot, the closer the similarity in the data.

2.13.1.4 Univariate statistical methods

The one-way Analysis of Variance (ANOVA) was utilised in the univariate analysis of multiple groups, determining whether the overall comparison was significant. To determine which two comparisons are different in subjected analysis, a Tukey's Honestly Significant Difference (HSD) posthoc, Satterthwaite approximation in a two-sample t-test for degrees of freedom and Wilcoxon signed-rank test analysis were utilised. Under the study conditions, the univariate analyses offered a preliminary overview of possibly significant aspects of interest.

The mass spectra of distinct tissue types were removed to establish a mean spectrum for each patient or sample. Then, using an ANOVA test, univariate analysis was performed to compare fatty acid and lipid intensities amongst different tissue subclasses. For each *m/z* characteristic, the ANOVA results were expressed as p-values. The p-values indicate the likelihood that the observed effect is significantly different from random changes. For example, if the p-value was less than 0.05, the variation in the intensity of the *m/z* feature was regarded as potentially significant, implying that there is a less than 5% chance that the identical results were seen by chance. Throughout this thesis, boxplots were also provided with ANOVA findings to determine group overlap between individual data points.

Because many variables (i.e., *m/z* characteristics) were compared between various tissue classes, the significant p-value was most likely due to chance. To account for this, the calculated p-values were converted to q-values with a false discovery rate (FDR) of 0.001. The FDR is the probability of getting a significant p-value without a significant outcome. An FDR of 0.05, for example, suggests that 5% of the accurate results will be false positives.

2.13.1.5 Clustering Analysis

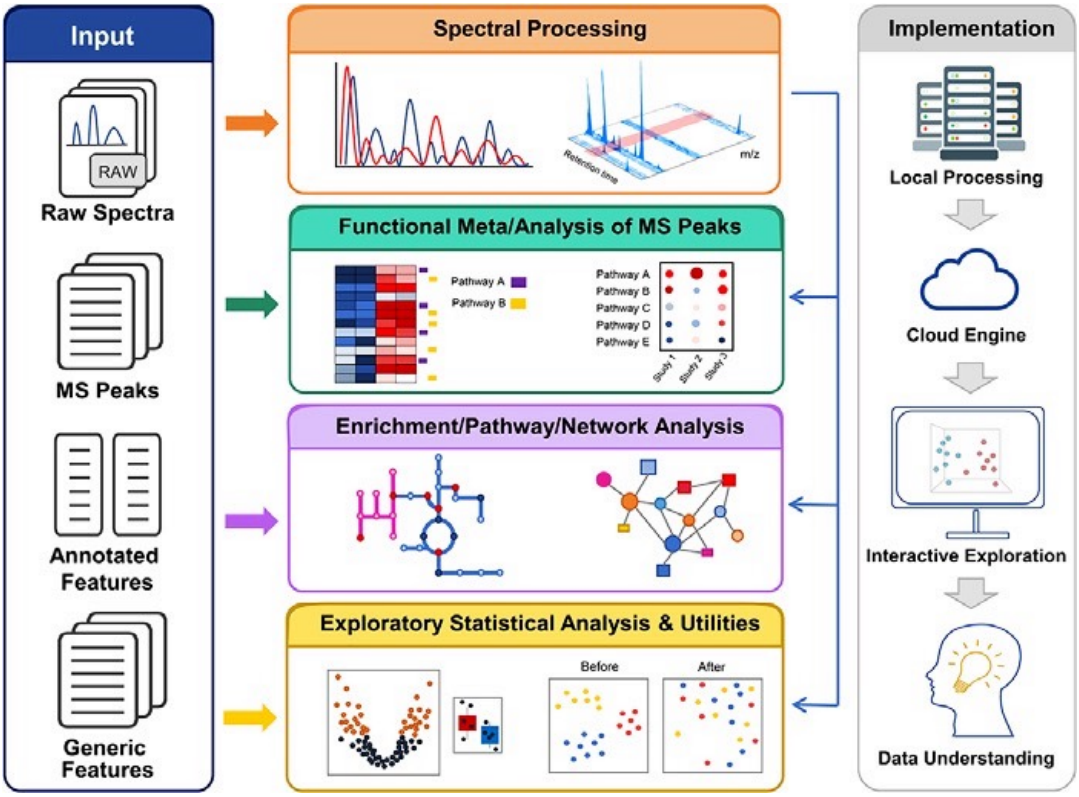
In the literature, one clustering analysis study[122] proved that DESI mass spectrometry imaging could quantify the abundance of the identified metabolites in a co-expression network. The ion vectors of tumour mass spectra with disconnected sub-networks provide a powerful 3D visualisation tool for molecular tumour heterogeneity analysis. The 3D-MSI visualisation, underlying biological composition, interactions, and the likely reason for its heterogeneous structure allow sub-regions in the tumour tissue to be identified.

This study specifically identified highly colocalised ions using Spearman's ion pair correlation. A matrix representing the pairwise correlations across all the tissue-related pixels was estimated and used to define a network. In this network, the nodes corresponded to the ions and the edges to their mutual correlations. A bisecting k-means, hierarchical clustering (average linkage) was employed to the ions in highly correlated. These ion lists were assumed to capture local metabolic interactions and, as a result, could be connected to the phenotypic of malignant cells in the area. The representative spatial distribution of the groups of strongly colocalised ions was derived using the first principal component scores, which were calculated using the ions belonging to each cluster.

2.13.1.6 MetaboAnalyst for statistical analysis of DESI-MSI

A data matrix was generated and pre-processed for normalisation in csv files and uploaded to MetaboAnalyst 5.0 (http://www.metaboanalyst.ca/) using required online analytic workflows, **Figure 2.13.1.3.1.**

a)



b)

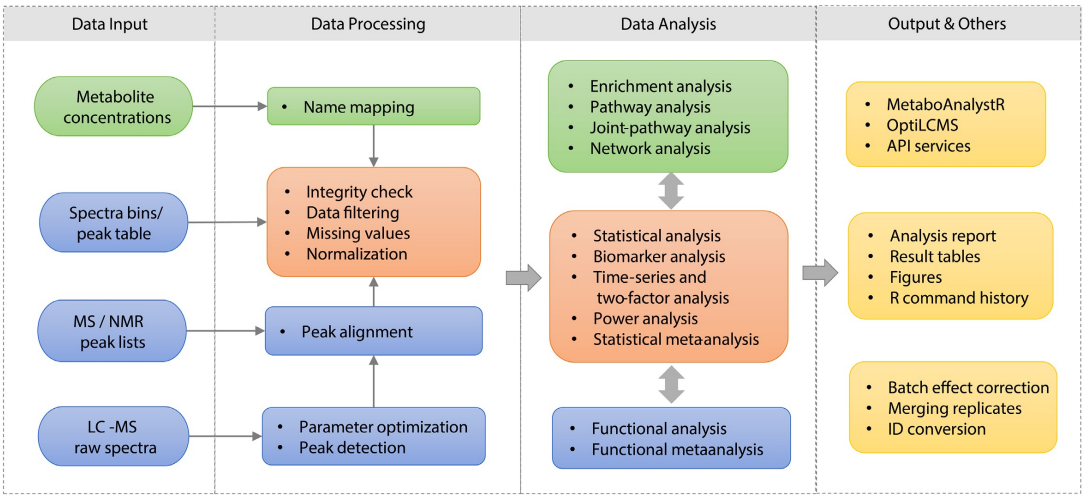


Figure 2.13.1.3. 1. Overview of MetaboAnalyst v5.0 workflows. a) Graphical Abstract of v5.0 workflow, b) Green boxes denote steps for targeted metabolomics, blue boxes denote untargeted metabolomics, and steps in orange can be utilised for both. To analyse outside of the pre-defined regular workflows, experienced users can use numerous utility functions or install the necessary R packages (yellow boxes)[176].

2.13.1.7 DESI-MSI Metabolite Identification

The features shown that it is particularly selective across tissue types and clusters were later spectrally annotated. Mummichog software[177] in MetaboAnalyst v5.0[176] was used to perform this annotation, which included querying databases for specific m/z values and isotope distributions. This programme allows for the functional analysis of untargeted metabolomics data from high-resolution mass spectrometry (HRMS). The critical concept is that putative annotation at the chemical level can anticipate changes at functional levels defined by metabolite sets or pathways. This is because, as established by Li et al.[177], changes at the group level rely on "collective behaviour," which is more tolerant of random errors in the compound annotation. The feature numeric mass (m/z) values for putative annotation supplied in the Functional analysis module must be included in the input peak list or peak table to apply this method. The mummichog programme is the first to use this approach, implying pathway activity from a sorted list of MS peaks detected by untargeted metabolomics. The original methodology used an over-representation analysis (ORA) technique to evaluate pathway-level enrichment based on relevant characteristics. This module must specify a pre-defined cut-off based on p-values. The Gene Set Enrichment Analysis (GSEA) method, frequently used to derive biological meaning from a ranked gene list, is a complementary approach. Unlike ORA, this approach considers the total rankings of MS peaks rather than a significant cut-off. It can detect subtle and consistent changes that ORA approaches may miss. In the MS Peaks to Paths module, the mummichog algorithm (Version 1.0.10) is adequately translated from Python programming to R, and the customised GSEA approach has been implemented. In addition, the module has an extended knowledge base of 21 species for pathway analysis. The MS Peaks to Pathways module is divided into three steps: uploading the user's data, choosing a pathway library, and analysing the pathway.

2.13.1.7.1 Data Input

The MS Peaks in this module are supplied in a three-column table in.txt format, with the m/z values, p-values, and statistical scores, t-scores. The mummichog and GSEA algorithms can be used if the input is a three-column table. The mass accuracy/tolerance of the module was set at 5 ppm. The analytical mode of the instrument was set to negative, and the m/z peak range was adjusted to 50-2000. The p-value threshold was used to distinguish between considerably and non-significantly enriched m/z features (for mummichog only). The chosen p-value cut-off was 0.05. M-H [1-], M-H2O-H [1-], and M+Cl [1-] adducts were chosen based on DESI-MSI feasible adduct forms for negative ionisation.

This module's knowledge base consists of five genome-scale metabolic models collected from the original Python version, either personally curated or downloaded from BioCyc, an enhanced library of 21 species derived from KEGG metabolic pathways, and 10 unique metabolites set libraries. Users must choose between 21 KEGG pathway libraries and five metabolic models. Because of the availability of the DESI-MSI dataset, Homo sapiens (human) [MFN] and sub-lipids class libraries were used.

2.13.1.7.2 Compound Matching Table

The MS Peaks output in this module also includes a thorough table comprising compound matching information for all user-uploaded m/z features. The table consists of four columns: the Query Mass for each feature, the predicted Matched Compound for each feature, the Matched Form, and the Mass Diff.

2.13.1.7.3 Mummichog Output

This module aims to use the power of existing metabolic models/pathways to get functional insight precisely from m/z features. The mummichog algorithm has three steps: 1) Permutations: From the user's uploaded a collection of m/z features, a list of metabolites (the same length as the number of significant m/z features) is inferred, taking into account all possible matches (isotopes/adducts). These potential chemicals are then mapped onto known

metabolic pathways for the organism of interest. A hypergeometric p-value is generated for each pathway. 2) Step 1 is performed several times to determine the null distribution of p-values for all pathways, represented by a Gamma distribution. 3) The important m/z characteristics are then utilised to generate the p-values for each route (Step 1). The p-values are then adjusted to reflect the permutations.

2.13.1.7.4 Mummichog Module analysis

Mummichog analyses are the pre-defined reference metabolic network model M for all modules that may be created using user input data and computing their activity scores. This technique is performed several times to estimate the background null distribution for the permutation data. Finally, depending on user data, the statistical significance of modules is assessed using the null distribution. The following are the particular steps:

1. The significant list of m/z features, Lsig, is used to generate a list of "input metabolites." All common isotopic derivatives and adducts are considered in the m/z - metabolite matching.
2. M contains all subnetworks that connect the input metabolites from Step 1 inside a single reaction. Within two, three, and four responses, this is repeated for subnetworks. These subnetworks include "input metabolites" and additional metabolites from M.
3. Newman's spectral split technique [178] is used to identify structural modules within these subnetworks. This results in a large number of modules of varying sizes.
4. Step 2 subnetworks and Step 3 modules are combined to form a collection of candidate modules (graphs), G1, G2,:::, Gn. Protruding edges are removed from each graph if they do not connect input metabolites and so do not contribute to the activity score.
5. Both the enrichment of "input metabolites" and the modularity should be reflected in the activity score. A candidate module's activity score is calculated as follows: The activity score for a subgraph G is:

$$\hat{A} = Q \cdot \frac{N_{I,G}}{N_G},$$

Where NG represents the number of metabolites in G, NI, G represents the number of "input metabolites" in G, and Q represents the adjusted Newman-Girvan modularity:

$$Q = \sqrt{\frac{N_I}{N_G}} \cdot \left(\frac{E_G}{m} - \sum_{i,j} \frac{k_i}{2m} \frac{k_j}{2m} \right) \quad , \quad i,j \in G$$

Where ki represents the network degree of metabolite i, m represents the total number of edges in the metabolic network M, EG represents the total number of edges in G, and NI represents the number of "input metabolites."

The original Newman-Girvan modularity favours extensive modules. It $\sqrt{\frac{N_I}{N_G}}$ is added to reduce this bias.

6. By randomly sampling Lref, a list of permutation features (equivalent to Lsig) is formed. The activity scores of modules from this permutation list are calculated similarly to Steps 1-5.

7. Step 6 should be repeated several times to get a list of scores from random modules. These scores are modelled as a Gamma distribution (null distribution) using maximum likelihood estimation, and a cumulative distribution function (CDF) is created.

8. On the CDF of the null distribution, the p-value for each module from input metabolites is determined.

2.13.1.7.5 Mummichog Pathway Analysis

The mummichog analysis produces a table with ranked pathways enriched by user-supplied data. The table shows the overall number of hits per pathway (all, significant, and anticipated), raw p-values (Hypergeometric), and p-values based on user data using a Gamma distribution.

All matching paths are shown as circles in the pathway summary. Each circle's colour and size correlate to its p-value and enrichment factor. A route's enrichment factor is computed as the ratio of significant pathway hits to the predicted number of compound hits inside the pathway.

Mummichog additional tests for pathway analysis were chosen, including the enrichment technique (Hypergeometric Test) and topological analysis (Relative-betweenness Centrality) with the Homo sapiens (KEGG) pathway library. Pathway Database (Metabolic pathways (integrated), Algorithm Selection, Enrichment method (Hypergeometric Test), and topology analysis (Degree Centrality) with Integration technique (Combine queries) were used for joint-pathway analysis. Gene-Metabolite Interaction networks were chosen for network analysis to observe subnetworks.

Overall, this mummichog method's efficacy in identifying metabolites is potentially significant and can implement multiple analyses[131,179,180]. Because numerous metabolite annotations belong to the same group, comparing various groups could still offer valuable information on overall effect levels (i.e. Oxylipids). To have a more robust identification, it only relies on precise mass measurements. Additional validation using tandem-MS, LC-MS, or Nuclear magnetic resonance (NMR) experiments is usually required for thorough structure elucidation.

3 Multi-modal mapping Strategy to reveal Tumour heterogeneity with Patient-Derived Xenografts (PDXs) models

3.1 Background

Tumour heterogeneity is a crucial clinical obstacle in cancer treatment, and it can potentially be the primary cause of treatment failure and chemoresistance. These tumours have distinct morphological and phenotypic profiles, including histopathology, gene expression, proliferation, metastatic behaviour, and metabolism. There is growing evidence to suggest that not every metabolic alteration in a cancer cell will be found across every cancerous cell in the same tumour sample and among patients. This phenomenon is referred to as "inter-tumour heterogeneity" when it arises between different patients' tumours, whereas "intra-tumour heterogeneity" occurs within the same patient's tumour[18,55,114,115,181,182].

Human breast tumours are formed of clones with various levels of mutation and metabolic abnormalities, resulting in intra-tumour genetic heterogeneity[668,61]. Although intra-tumour heterogeneity varies amongst cancers, it can be already evident at diagnosis and dynamically evolves spatially and through the disease timeline[183]. Tumour heterogeneity frequently influences prognosis and treatment responsiveness[183,184]. More recently, it has been acknowledged that the metabolic variability of human tumours can also complicate designing therapeutics that target metabolic vulnerabilities. Although the data from cancer cells cultivated in vitro are significant, they lack information on the tumour microenvironment and metabolism. This restricts the applicability of the findings based on this type of data due to biochemical differences associated with in-vitro adaptations. It has been shown that the high degree of molecular heterogeneity causes treatment failure in breast cancer.

Similarly, recent findings have shown that tumours exhibit reprogrammed metabolic activities that promote cancer progression. Thus, pre-clinical

translational studies could be important to focus on identifying potentially novel molecular vulnerabilities. The patient-derived tumour xenografts (PDTXs) model offers the advantage of capturing the inter-and intra-tumour heterogeneity in pre-clinical models. The morphological and genetic properties of the original tumour are retained via animal passaging[66,67]. To some extent, PDX models are able to mimic the tumour and features of the microenvironment. These models offer an optimal environment to investigate the specific pathways' roles in carcinogenesis and tumour progression.

Hyperactivation of specific genes can initiate metabolic alterations, and one of the most critical pathways, the PI3K/AKT pathway, plays a crucial role in cancer oncogenesis. The *PIK3CA* gene, which codes for the p110 subunit of PI3K, is altered in 40–50% of ER+ tumours, indicating that ER+ breast cancer cells rely on this pathway[185]. Given PI3K involvement in promoting proliferation, survival, and hormone receptor pathway activity, activating the PI3K/AKT/mTOR pathway increases disease progression and resistance to endocrine treatment[24,66,186]. The PI3K/AKT pathway is frequently overactivated in cancer cells, resulting in increased cell growth and proliferation[24,186]. Several studies have demonstrated that the PI3K/AKT pathway may alter cell lipid metabolism and reprogram cancer cell features [23,24].

Imaging Mass Cytometry (IMC) is a multiplexed approach for assessing tissue phenotyping, signalling pathways, and cell markers at sub-cellular spatial resolution (1 μ)[187]. This pipeline employs MSI techniques to simultaneously detect and measure up to 30 distinct antibodies within a tissue segment. Traditionally, sections from conventional histology (typically haematoxylin and eosin-stained) are overlaid with MSI-derived molecular profiles. As a result, MSI-based molecular histology gives a snapshot of the tissue milieu. It allows associating medicines, metabolites, lipids, peptides, or proteins with histological/pathological characteristics or tissue substructures. With other molecular imaging methods, mass spectrometry (MSI) can also provide beneficial supplemental information for exploring new scientific subjects [188].

Matrix-assisted laser desorption ionisation imaging mass spectrometry (MALDI-MSI) is the gold standard for intact tissue analysis. It allows for the simultaneous localization and quantification of biomolecules in different histological areas of interest. This method can provide novel insights into tumour microenvironmental (TME) biology and the study of intra-tumour heterogeneity (ITH)[189–191].

Desorption electrospray ionisation (DESI) is another MSI technique that is widely employed[188] and is currently being used to investigate lipids and metabolites on tissue surfaces[188]. DESI-MSI is extensively used among MSI methods because of several benefits, including working at atmospheric pressure and the requirement for minimal sample preparation, allowing for immediate analysis followed by tissue sectioning on slides[85]. Compared to MALDI-MSI, the lack of a matrix makes DESI imaging data less complicated and makes lipid assignments more straightforward in the absence of matrix ions. Conversely, the lateral resolution with DESI is generally 75–200 μm, which compares unfavourably to high-definition MALDI imaging resolution, which can be below 75 μm. In two and three dimensions, DESI-MSI has effectively connected lipid and metabolic content, distributions, and patterns with tissue shape [192]. Using DESI-MSI in the lipidomics approach for cancer diagnostics has proven its significance by distinguishing the metabolic profiles of cancerous tissues from healthy tissues[117,193,194], particularly in breast[116,195,196], and other types of cancer[35,197]. By focusing on the patient's tumour profiles, this detailed metabolic profiling enables the investigation of Inter-/Intra-Tumor Heterogeneity (ITH)[198]. Furthermore, it allows the examination of metabolic changes in diverse tissue locations and a better spatial comprehension of molecular dynamics.

Advancements in bioinformatics have allowed MSI technologies to overcome some limitations in identifying molecular compositions, which are essential in studying the ITH involved in cancer development and progression. One of the most notable accomplishments of bioinformatics in MSI is measuring the intensity level of the molecules directly correlates with anatomical and

histological characteristics in tumours and healthy tissues at the cellular level, which provided an in-depth investigation of molecular heterogeneity [122,190,199].

The capacity to selectively dissect the various areas of tumour tissue for follow-up experiments, such as RNA analysis, is crucial in the multi-modal analysis workflow for a better understanding of tumour biology. When RNA is extracted from bulk tissue, it is difficult to pinpoint the source of a given signal or the exact cell type in which a specific biochemical pathway is activated. Laser Capture Microdissection (LCM) is a well-known technology for overcoming the heterogeneity obstacle associated with different subtypes of tumour tissue. Under microscopic observation, LCM allows the separation of homogeneous cellular populations from heterogeneous cellular mixes of tumour tissue. LCM is a non-contact technique; tissue slices are positioned directly onto polymer-based foil that is adhered to a glass slide. These slides permit the microdissection of regions of interest (ROIs) from tissue using an ultraviolet laser. The ROIs are collected straight into the cap of a PCR tube for further processing 193, such as LC-MS and gene expression analysis[200]. Since tissue heterogeneity is a challenge to molecular research. This novel combination of different methodologies, such as metabolomics and transcriptomics, could offer an exceptional opportunity to investigate in-depth ITH. Since the heterogeneous nature of tissue is a hurdle for molecular research, this novel combination of different methodologies provides an excellent opportunity for investigating ITH. It is difficult to pinpoint the origin of a given signal or the exact cell type in which a specific biochemical pathway is activated when RNA is extracted from bulk tissue.

Co-registration of several types of information with a multi-modal analysis will likely yield a comprehensive measurement of proteins, lipids, metabolites, and RNA/DNA in the ROIs. This approach enables spatial omics technologies on multi-dimensional ROIs to acquire precise metabolic/genetic profiling of diseases. This method could be advantageous in mapping tumour heterogeneity in cancer [105,121,197]. The colocalization strategy was used to correlate diverse omics datasets in the final result. The colocalization patterns of the molecules involved in the mechanisms express local metabolic

pathways. One early study from the lab uncovered this phenomenon by looking for two types of biological correlations: (i) when the two correlated species are the substrate and product of the same enzyme, or (ii) when both are products or substrates of the same enzyme/transporter. In this study, a similar methodology was employed, with the most enriched colocalised metabolites and genes considered to be involved in the same signalling pathways.

The goal of this study is to map the inter- and intra-tumour metabolic heterogeneity of breast cancer by combining a multi-modal Mass Spectrometry Imaging (MSI) approach that includes DESI and MALDI with Imaging Mass Cytometry (IMC) analysis of the tumour section using Cytometry by the time of flight CyTOF and high-throughput genetic characterization of metabolically distinct regions using transcriptomics. A multi-modal analytic approach was then correlated in network analysis to find driver signalling pathways on unique metabolic clusters in the sequential sections of PDX models, targeting hyperactivation of the PI3K/AKT signalling pathway by gaining or activating mutations in *PIK3CA*[23].

3.2 Aim and objectives

The aim of this chapter was to incorporate a multi-modal analysis workflow into a segmentation strategy to reveal spatially co-localized ions clusters following DESI-MSI, using breast cancer patient-derived xenograft models with hyperactivation of the PI3K/AKT signalling pathway by gain or activation of mutations in *PIK3CA*. Furthermore:

- To map the lipid compositions, patterns, and spatial distributions in a segmentation strategy from breast cancer PDX models using DESI-MSI technology and to explore the function of hyperactivation of the PI3K/AKT signalling pathway in inter/intra-tumour heterogeneity.
- To perform network and gene enrichment analysis on ROIs derived from DESI-MSI segmentation for spatial transcriptomics to determine whether there is a relationship between metabolic signatures and genetic traits in *PIK3CA* mutant and wild-type PDX models.
- To supplement the DESI finding using a protein profiling approach by performing a CyTOF analysis with a customised panel.

3.3 Summary Workflow of Experimental Plan

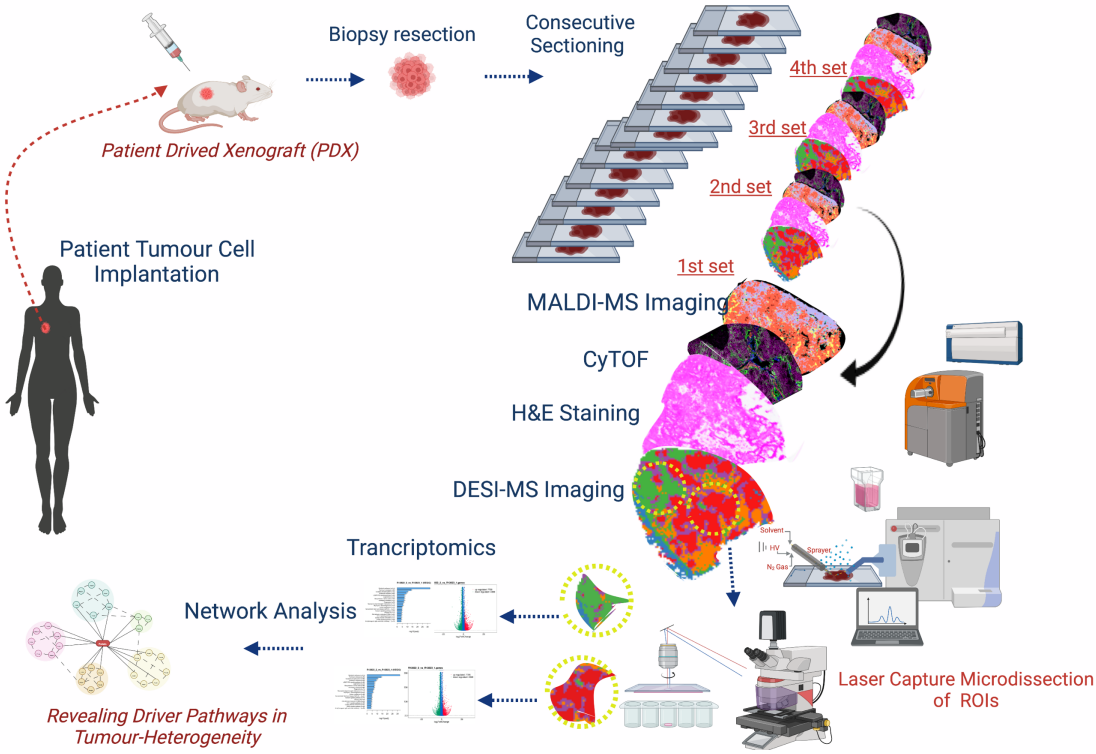


Figure 3.3. 1. Multi-modal spatial analysis revealed tumour heterogeneity within PDX models. Four sets of *PIK3CA* mutant and wild-type PDX models were sectioned at 10 µm and placed on SuperFrost and PEN membrane glass slides. The specimen was imaged using mass spectrometry, DESI-MSI, MALDI-MSI, with IMC (Imaging Mass Cytometry) with a CyTOF panel including 25 protein markers. Hierarchical clustering of the MSI spectra was used for data segmentation analysis. Several spatial segments were generated (depicted here with distinctive colours such as orange, red, green, etc.). The outlines of each cluster of interest were precisely assigned to the optical image of the following section utilising MSI data and the LCM live-view image. This co-registration enabled the segments to be microdissected. The extracted RNA material from the dissected segments was subsequently processed for high-throughput genetic characterization of metabolically distinct regions utilising transcriptomics.

3.4 Results and Discussion

3.4.1 A pilot study of investigation metabolic heterogeneity in wild-type and mutant *PIK3CA*, PDX models by DESI-MSI analysis

In this part of the study, a DESI-MSI clustering analysis on multiple breast cancer PDX models with hyperactivation of the PI3K/AKT signalling pathway by gaining or activating mutations in *PIK3CA* was used to address the metabolic heterogeneity. DESI imaging in negative ion mode was performed at 1 scan/s (one scan per second) and 100 µm pixel size, followed by clustering analysis utilising bisecting k-means to produce the particular ion-related segments described in section 2.6.2.

Following DESI runs, segmentation analysis was performed on the 10 different PDX models (4 mutant (MT), 6 wild types (WT)), revealing the presence of metabolically distinct regions. In more detail, the bisect k-means clustering analysis has produced five different metabolic clusters, **Figure 3.4.1. 1a.** The score from the first two PCA components of all PDX models (batch 1-10) is shown in **Figure 3.4.1. 1b**. Each cluster is characterised by a list of *m/z* values representing the spatial distribution of ions with a significantly different average intensity compared to the remaining regions (Wilcoxon test, p-value < 0.05 after Benjamini Hochberg correction). The fold change (FC) of these metabolites is displayed in the volcano plots of each cluster vs all remaining clusters, **Figure 3.4.1. 1c-g.**The heat map compares and contrasts the average intensity of the detected features identified by DESI-MSI in each cluster per sample set by emphasising similarities and differences using hierarchical (agglomerative) cluster analysis **Figure 3.4.1. 1h. Table 3.6.1** in the Appendix depicted a list of m/z peaks from the heatmap.

The total distribution of differentially detected metabolites is reported using volcano plots. When compared to the other clusters, cluster 1 (red, **Figure 3.4.1. 1c**) had the most significant increased *m/z* variables (p-value < 0.05, logFC > 1.0). Cluster 3 (green, **Figure 3.4.1. 1e**) incorporated the lower *m/z* numbers. Whereas Clusters 1 (red), 2 (blue), 3 (green), and 4 (purple) were

shared between all PDX models, interestingly, Cluster 5 (orange) was unique and mainly present in the mutant *PIK3CA* models and partly in the wild type PDXs. Thus cluster 5 seemed to be a promising target for further investigation using multi-modal spatial clustering. Next, one wild type (PDX model 2) and one mutant (PDX model 7) were chosen to proceed with consecutive sectioning and multi-modal pipeline use. These sections were employed in the respective omics approaches to map localised metabolic hotspots to identify the potential presence of intra- and inter-tumour heterogeneity in the PDX models.

Furthermore, the post-DESI tissue section was used for H&E staining for histological examination, **Figure 3.4.1. 1a.** In the H&E staining, the tissue subtypes were differentiated by staining aligned to the DESI-MSI cluster regions, as seen in the exemplary samples for PDX model 1 (cluster 1, red) and 10 (cluster 2, blue). To demonstrate matched areas, careful co-registration of the scanned H&E images on the envisioned DESI-MSI cluster images, as seen in the yellow and white draw areas, **Figure 3.4.1. 1a, (1)-(2),** was used. These results suggested that clustering analysis could be used to assist in the histological annotation of tumour samples.

To summarise the clustering analysis in this work, hierarchical segmentation analysis was utilised to find unique clusters. To discover highly colocalized ions, ion pairwise Spearman's correlation was employed. The initial principal component scores, calculated using ions from each cluster, were defined using a matrix encapsulating the pairwise correlations across all tissue-related pixels. The PC1 illustrates the spatial distribution of groupings of strongly colocalised ions. Each tissue section entails many clusters with distinct colour modules, each with its *m/z* value list (csv file) of detected ions with a sub-network in the tissue. These clusters can capture local metabolic interactions, possibly related to the behaviour of regional subgroups of malignant cells. These local subgroups of cell clones indicate that metabolic heterogeneity may exist within the same tumour.

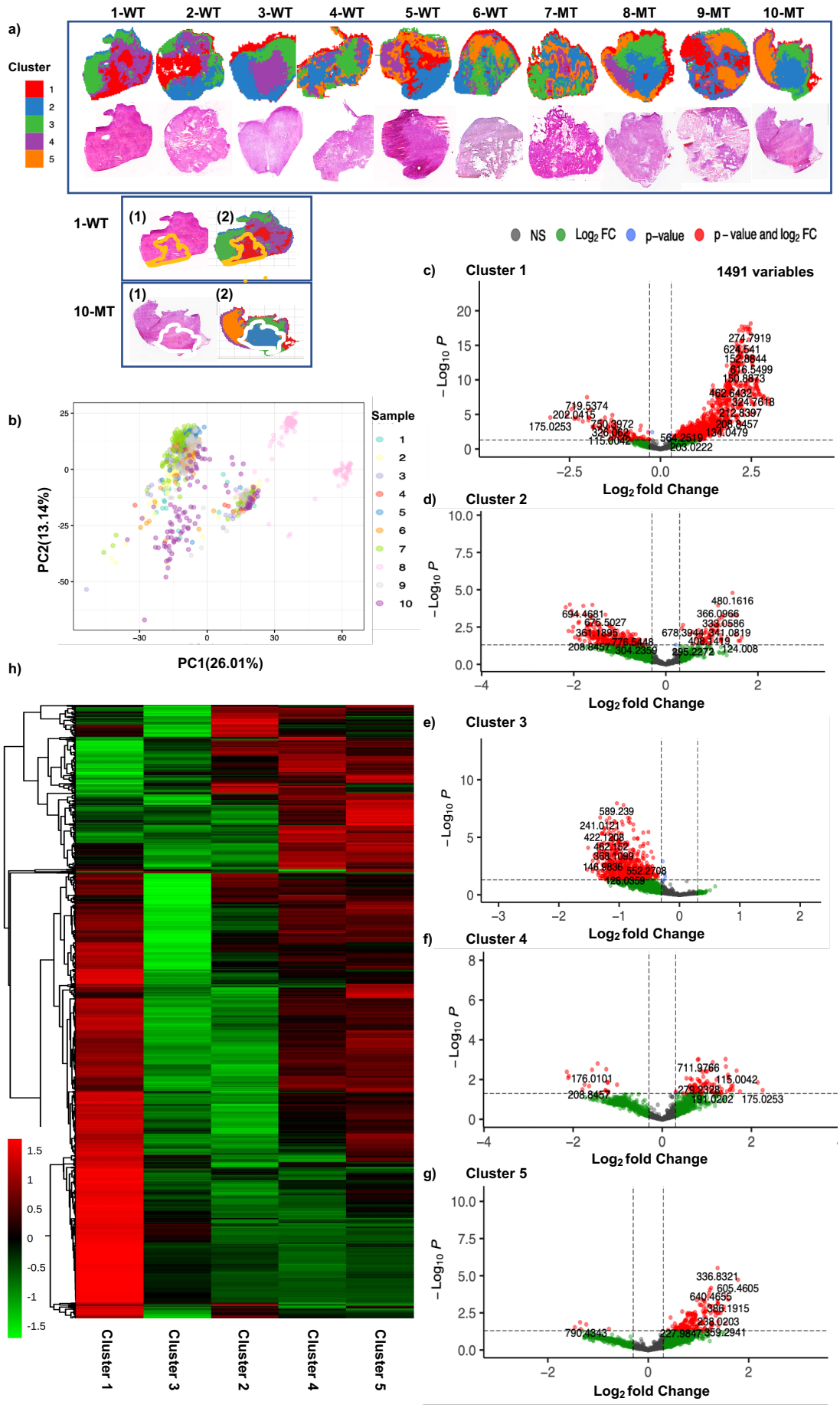

a)
1-WT
2-WT
3-WT
4-WT
5-WT
6-WT
7-MT
8-MT
9-MT
10-MT
Cluster
1
2
3
4
5
1-WT
10-MT
(1)
(2)
b)
Sample
PC2(13.14%)
PC1(26.01%)
NS
Log2 FC
p-value
p-value and log2 FC
c)
Cluster 1
1491 variables
−Log10 P
Log2 fold Change
d)
Cluster 2
e)
Cluster 3
f)
Cluster 4
g)
Cluster 5
h)
Cluster 1
Cluster 3
Cluster 2
Cluster 4
Cluster 5

Figure 3.4.1. 1. A pilot study to reveal metabolic heterogeneity in *PIK3CA* mutant and wild-type PDX models by DESI-MSI. a) 5 different metabolic clusters within 10 PDX models (6 wt, 4 *PIK3CA* mt), highly colocalised ions were identified using the ion pair Spearman's correlation and hierarchical clustering (average linkage). The first PC scores, calculated using the ions belonging to each cluster, were used as the representative spatial distribution of the groups of highly colocalised ions. H&E staining of post-DESI slides of all the 10 PDX models and co-registration of two representative slides to MSI cluster regions (lower panel).b) Loadings plot from first PCA components. c-g) Volcano plots of 5 different clusters from different patient samples with the $|\log_2(FC)| > 1$ and p-value < 0.05. h) The Heatmap depicted the grouping results of all clusters between *m/z* values (distance measure using the Euclidean, and clustering algorithm using average).

3.4.2 Multi-modal spatial clustering strategy to reveal metabolic heterogeneity in a pair of *PIK3CA* mutant and wild-type PDX models by DESI-MSI

Following the success of the pilot study, a representative model of both a wild type and a *PIK3CA* mutant PDX model was chosen for analysis with serial sections. Although a single section may only provide a snapshot of the underlying heterogeneity complexity, serial section analysis may provide more robust information regarding the molecular composition dynamic of tumour sub-regions. The next step was to proceed with DESI imaging in four consecutive sections set 100 µm apart.

Cluster analysis was utilised to distinguish tissue regions representing metabolically distinct areas in the mutant (MT) *PIK3CA*, BR1282 PDX model, **Figure 3.4.2. 1** and wild-type (WT), BR1458 PDX model, **Figure 3.4.2. 2**. Each model was analysed, and the bioinformatic analysis revealed five distinct metabolic clusters, **Figure 3.4.2. 1a** and **Figure 3.4.2. 2a** with the first two PCA components' scores were shown for each model, **Figure 3.4.2. 1b** and **Figure 3.4.2. 2b.** Each cluster had an *m/z* value list of the metabolites that represented the significantly more abundant compared to the remaining tissue parts (Wilcoxon test, p-value < 0.05, Benjamini Hochberg correction). The fold change (log2FC) of these metabolites in each cluster vs all clusters was displayed with the volcano plots, **Figure 3.4.2. 1c-g** and **Figure 3.4.2. 2c-g.** Volcano plots were used to determine the distribution of differently identified metabolites in all datasets.

In the *PIK3CA* MT PDX model, there were 1890 *m/z* values detected, within cluster 1 (red, **Figure 3.4.2. 1c**) having the most significant *m/z* variables (p-value 0.05, logFC > 1.0) when compared to the other clusters. Cluster 4 (purple, **Figure 3.4.2. 1f**) had the lowest *m/z* ratio. Intriguingly, Clusters 3 (green) and 5 (orange) demonstrated a dominating profile over tissue sections in the mutant *PIK3CA* PDX model, and both clusters were enlarged across whole tissue morphology Cluster 1 (red), on the other hand, was nearly identical in size-wise throughout the tissue slices. These three clusters (red, green, and orange) may be important in intra-tumour heterogeneity based on their dynamic pattern in the tissue. In addition, when comparing the other clusters' significant DESI peaks, **Figure 3.4.2. 1c-g** and heat map, **Figure 3.4.2. 1h,** they had the highest abundance peaks **Table 3.6.2** in the Appendix depicted a list of m/z peaks from the heatmap**.** Clusters 2 (blue) and purple (4) appear to have a few high abundance peaks, while cluster 2 (blue) appears to have less than the other clusters. A histopathologist annotated two separate histological areas in the H&E image: tumour and stroma. The solid tumour seemed histologically homogenous. Interestingly, the stroma was matched with cluster 1 (red), and three separate clusters shared tumour areas in the DESI-MSI clustering (purple, green and orange). A detailed annotation was depicted in **Figure 3.6.1**. the Appendix.

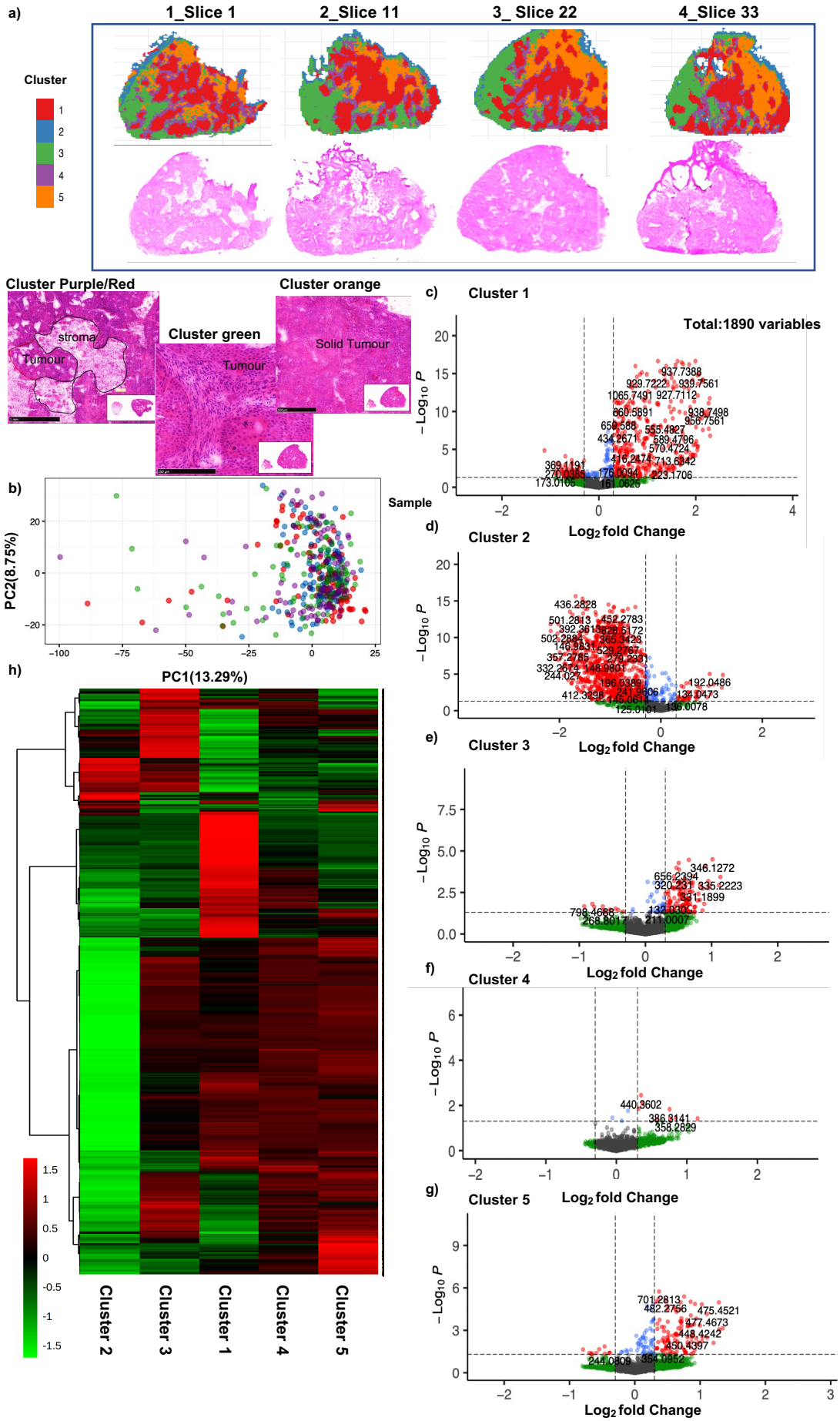

a)
1_Slice 1
2_Slice 11
3_ Slice 22
4_Slice 33
Cluster
1
2
3
4
5
Cluster Purple/Red
stroma
Tumour
Cluster green
Tumour
Cluster orange
Solid Tumour
b)
PC2(8.75%)
PC1(13.29%)
c)
Cluster 1
Total:1890 variables
−Log10 P
Log2 fold Change
Sample
d)
Cluster 2
e)
Cluster 3
f)
Cluster 4
g)
Cluster 5
h)
Cluster 2
Cluster 3
Cluster 1
Cluster 4
Cluster 5

Figure 3.4.2. 1. Multi-modal clustering strategy revealed metabolic heterogeneity in mutant *PIK3CA* PDX models by DESI-MSI. a) 5 different metabolic clusters between four sequential sections 100 μm apart, highly colocalised ions were identified using the ion pair Spearman's correlation d a hierarchical clustering (average linkage). The first PC scores, calculated using the ions belonging to each cluster, were used as the representative spatial distribution of the groups of highly colocalised ions. b) Loadings plot from first PCA components c-g) Volcano plots of 5 different clusters from different patient samples with the $|\log_2(FC)| > 1$ and p-value < 0.05. h) The Heatmap depicted the grouping results of all clusters between *m/z* values (distance measure using the Euclidean, and clustering algorithm using average).

In the wt PDX model, there were 1654 detected *m/z* values, with cluster 1 (red, **Figure 3.4.2. 2c**) having the most significant *m/z* variables detected at a higher intensity (p-value 0.05, logFC > 1.0) when compared to the other clusters. Cluster 2 (blue, **Figure 3.4.2. 2d**) had the fewest significant *m/z* values detected at a lower intensity. In particular, clusters 1 (red) and 5 (orange) demonstrated a dynamic profile over tissue sections in the WT PDX model**, Figure 3.4.2. 2a** and both clusters are large, whereas other clusters occupied a smaller tissue area; Clusters 2 (orange), 3 (green), and 4(purple). These two red and orange clusters may play an essential role in intra-tumour heterogeneity since they have the most significant high abundance identified peaks displayed with volcano plots, **Figure 3.4.2. 2c-g**, and heatmap, **Figure 3.4.2. 2h**. **Table 3.6.3** in the Appendix depicted a list of m/z peaks from the heatmap. In the H&E image, there were histological differences in the areas between necrotic and tumour/stroma (red and orange clusters). A detailed annotation was depicted in **Figure 3.6.2**. the Appendix. Although the solid tumour seemed histologically homogenous in the region of orange, blue and purple clusters, that differs from the metabolomic alterations. The MSI clustering revealed that necrotic tissue was matched with cluster 1 (red), tumour/stroma/necrotic areas with cluster 2 (blue), and tumour portions were shared by three separate clusters (purple, green and orange). This heterogeneity might imply that the tumour subclones were molecularly diverse and detectable by MSI but histologically indistinguishable. This example demonstrated how DESI-MSI clustering analysis might be used to examine tumours and associated tumour heterogeneity interactively.

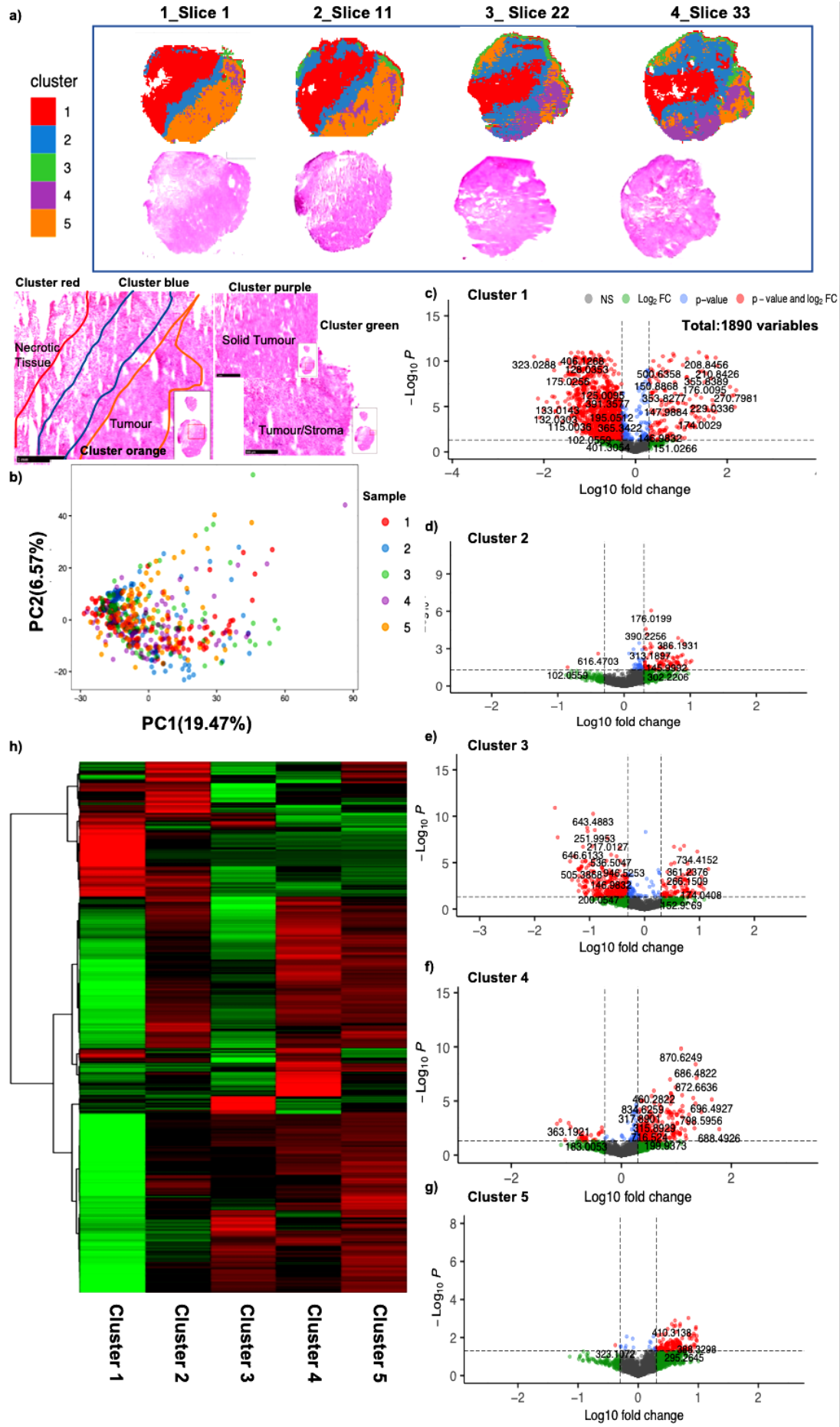

a)
1_Slice 1
2_Slice 11
3_ Slice 22
4_Slice 33
cluster
1
2
3
4
5
Cluster red
Cluster blue
Cluster purple
Cluster green
Necrotic Tissue
Solid Tumour
Tumour
Tumour/Stroma
Cluster orange
b)
Sample
PC2(6.57%)
PC1(19.47%)
c)
Cluster 1
NS
Log2 FC
p-value
p - value and log2 FC
Total:1890 variables
−Log10 P
Log10 fold change
d)
Cluster 2
e)
Cluster 3
f)
Cluster 4
g)
Cluster 5
h)
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5

Figure 3.4.2. 2. Multi-modal clustering strategy to reveal metabolic heterogeneity in wild-type PDX model by DESI-MSI. a) 5 different metabolic clusters between four sequential sections 100 μm apart, highly colocalised ions were identified using the ion pair Spearman's correlation and hierarchical clustering (average linkage). The first PC scores, calculated using the ions belonging to each cluster, were used as the representative spatial distribution of the groups of highly colocalised ions. b) Loadings plot from first PCA components c-h) Volcano plots of 5 different clusters from different patient samples with the $|\log_2(FC)| > 1$ and p-value < 0.05. h) The Heatmap depicted the grouping results of all clusters between *m/z* values (distance measure using the Euclidean, and clustering algorithm using average).

3.4.3 Comparison of DESI-MSI with MALDI-MSI for clustering analysis to reveal metabolic heterogeneity

In an attempt to uncover metabolic heterogeneity, DESI and MALDI-MSI modalities were utilised. To summarise the methodology used, four sequential sections of each PDX model were imaged by DESI, and the data was acquired at 10 scans/s with the mass range *m/z* 50 to 1000 and 100x100 μm pixel size in the negative ionisation mode (Xevo G2-XS QToF, Waters). Similarly, MALDI-MSI analysis was performed at 20 pixels/sec with the mass range *m/z* 50 to 1200 and 45x45 μm pixel size in the negative modes (SYNAPT G2, Waters).

Figure 3.4.3. 1a shows that the DESI-MSI clustering for the *PIK3CA* mutant PDX model revealed 5 clusters, while the MALDI-MSI clustering revealed 3 clusters. Despite having a lesser lateral resolution of 45 μm compared to DESI lateral resolution of 100 μm, MALDI-MSI was unable to identify smaller regions such as cluster 3 (green) in the WT PDX model DESI-MSI recognised, but MALDI-MSI did not **Figure 3.4.3. 1a**. The reason could be the reduced number of identified peaks in MALDI-MSI. DESI-MSI found 9338 distinct peaks compared to 678 by MALDI-MSI, implying that DESI-MSI recognised approximately ten times as many features, **Figure 3.4.3. 1c**. Furthermore, the DESI-MSI samples show greater intensity in spectral comparison (**Figure 3.4.3. 1b.**

Surprisingly, when comparing the data acquired utilising the breast PDX tissues in both quantitative and qualitative terms, DESI-MSI delivered the best

results. This can be partly explained by the discovery of smaller regions in the DESI-MSI clustering analysis, which shows that the more characteristics found, the more likely it is that additional hidden metabolic sub-interactions would be revealed.

One combination study[201] of MALDI and DESI revealed that both modalities allow for identifying various lipid species and co-mapping on the tissue surface. Fewer peak identification with MALDI-MSI could be attributed to using several matrix types during the sample preparation stage before data acquisition. Another significant technical difficulty could be for MALDI-MSI, the necessity to image low-abundance metabolites and lipids without sacrificing sensitivity while preserving feasible spatial resolution. Lipid spectra, for example, are often dominated by signals of highly abundant or easily ionisable lipids such as PCs, while other PLs such as PE are barely detectable[140,202]. Based on this literature information and our findings, the DESI-MSI appears to be the most promising modality for follow-up gene expression analysis. Along with the fact that with no prior sample preparation of DESI-MSI seems to be the optimal method for a shorter turnaround time.

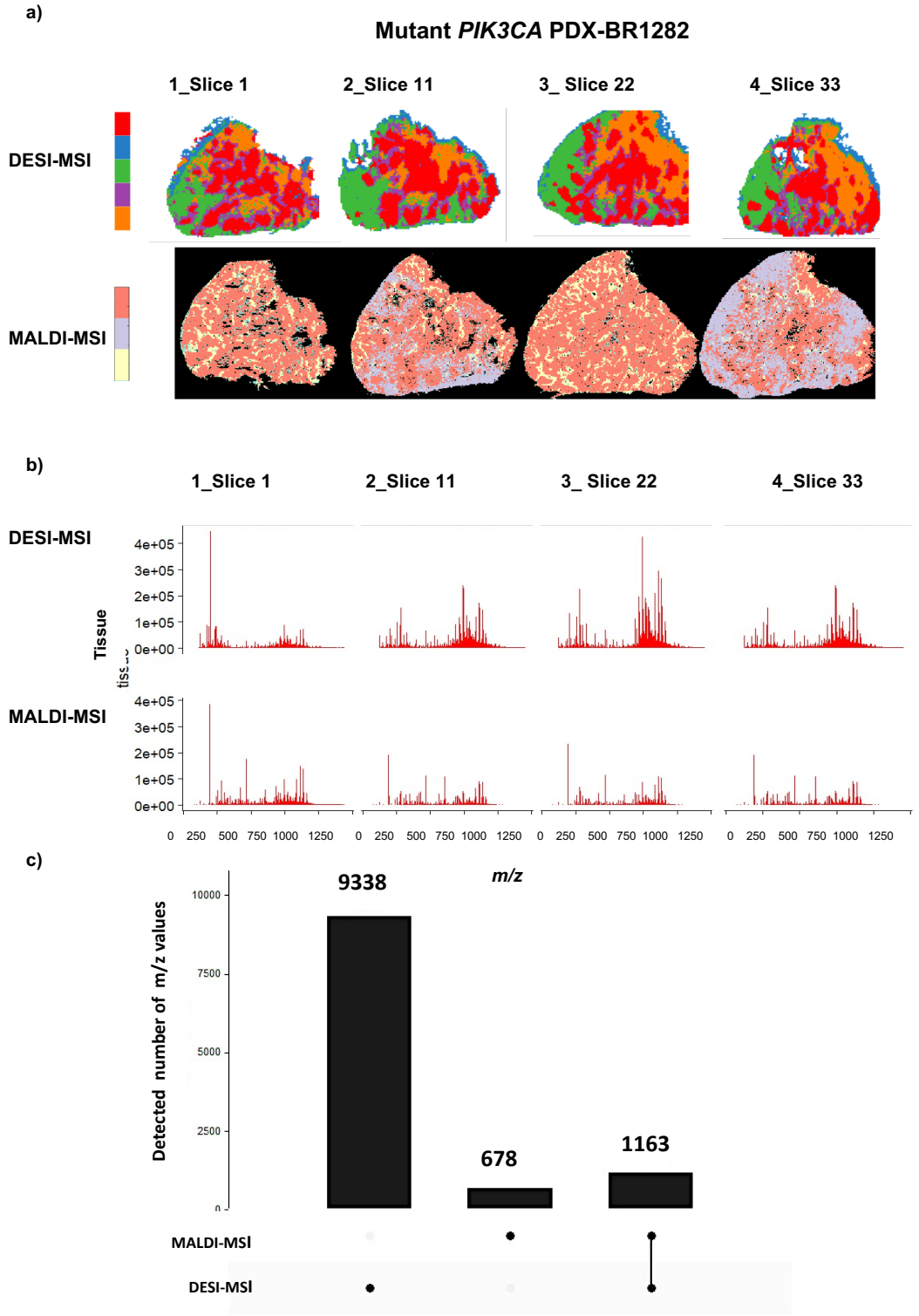


Figure 3.4.3. 1. Comparison of DESI-MSI with MALDI-MSI for clustering analysis to reveal metabolic heterogeneity. a) *PIK3CA* mutant PDX model clustering, DESI-MSI in four consecutive sections set 100 µm apart: five separate metabolic clusters were discovered using bisecting k-means. MALDI-MSI was performed in four consecutive sections 100 µm apart to identify four separate metabolic groups using k-means clustering and PQN normalisation. c) a comparison of peak intensities in four consecutive sections distanced by 100 µm, d) The detected

peaks, unique and in common with DESI and MALDI-MSI. A homogeneous coating of a 9-Aminoacridine (9-AA) matrix solution was used for MALDI-MSI analysis.

3.4.4 DESI-MSI guided transcriptomics from microdissected Metabolic Clusters

As described before, visualising the distinct metabolic regions following DESI-MSI acquisition has demonstrated the ability to identify tumour tissue sub-regions as metabolically differentiated segments, perhaps representing the underlying biological composition, interactions, and the likely cause of its heterogeneous structure. This DESI-MSI guided LCM microdissected tissue region for RNA sequencing. Preliminary assessment in the preceding section determines whether clusters may be an ideal possibility for identifying gene profiles underlying this DESI-MSI detected metabolic cluster. The adjacent slice of mutant *PIK3CA* model clusters 1 (red) and 3 (green), **Figure 3.4.2. 1a**, as well as the wild-type model clusters 1 (red) and 5 (orange), **Figure 3.4.2. 2a** was chosen for sample preparation.

3.4.4.1 Overview of data production quality and screening of differentially expressed genes

From serial slides, tissue slice 22 of both PDX models were chosen to continue cluster microdissection. The extracted RNA materials were then provided to a commercial sequencing company (Novogene, Cambridge, UK) to conduct the transcriptomic analysis. Section 2.11 provides the initial quality assessment of the extracted RNA's sequenced data. Raw readings were screened to eliminate reads that have adaptor contamination or poor quality. In the subsequent studies, only clean reads were employed. Numerous factors had to be examined to get a complete statistical analysis of sequencing data, which are presented in **Table 3.4.4.1. 1** and the percentage of clean reads per sample in **Figure 3.4.4.1. 1**. Most crucially, Q30 refers to the percentage of bases with proper base recognition - rates ranged from 90 to 92 per cent in total bases, with a

minimum requirement of 80 per cent. According to these findings, the mapping percentage was high enough to obtain complete transcriptome information from the microdissected tissue slice. Appendix 1 shows the quality assessment of isolated RNA in terms of RNA integrity number (RIN) and quantities.

Table 3.4.4.1. 1. Overview of RNA sequencing data production quality.(1) Sample: The names of samples. (2) raw_Reads: The original sequencing reads counts.(3) clean_Reads: Number of reads after filtering. (4) raw_data: Raw reads number multiply read length, saved in G unit. (5) clean_data: Clean reads number multiply read length, saved in G unit. (6) error_rate: Average sequencing error rate, which is calculated by Qphred=-10log10(e). (7) Q20: Percentages of bases whose correct base recognition rates are 95-97%. (8) Q30: Percentage of bases whose correct base recognition rates are between 90-92% in total bases.

PDX models	Cluster ID	raw_reads	clean_reads	raw_data(G)	clean_data(G)	Q20(%)	Q30(%)
WT	orange	148853258	140658911	44.7	42.2	97.01	92.58
WT	red	139706605	132339994	41.9	39.7	97.1	92.72
PIK3CA MT	green	123295509	115067174	37	34.5	96.94	92.69
PIK3CA MT	red	114449821	108034356	34.3	32.4	95.7	90.94

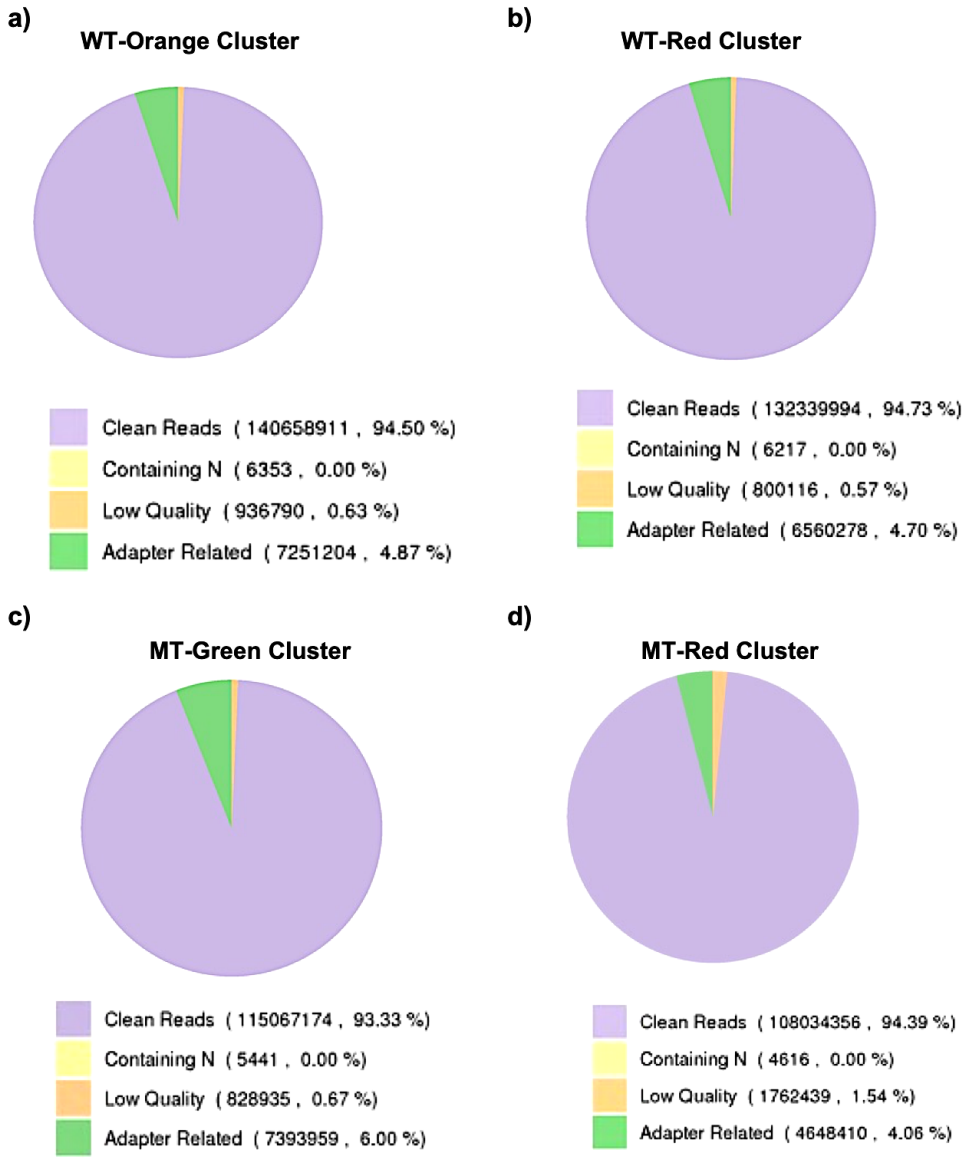


Figure 3.4.4.1. 1. Percentage of total Raw Reads Components of Sequenced Samples. a) wild-type cluster 5 (orange), b) wild-type cluster 1 (red), c) *PIK3CA* mutant cluster 3 (green), d) *PIK3CA* mutant cluster 1 (red), Clean Reads: reads that passed quality control. Containing N: reads in which uncertain nucleotides constituted more than 10 per cent of the read. Low Quality: reads in which standard quality nucleotides constituted more than 50 per cent of the read. Adapter related: reads that had adapter contamination.

The three categories of mapped areas were exons, introns, and intergenic regions. Exons (Q30) are the most prevalent kind of mapped rate when the reference genome is well-annotated. The exonic quality level (Q30) should be

more than 80%. The fraction of exonic readings mapped to the reference genome ranged from 91 to 97 per cent in **Figure 3.4.4.1. 2**.

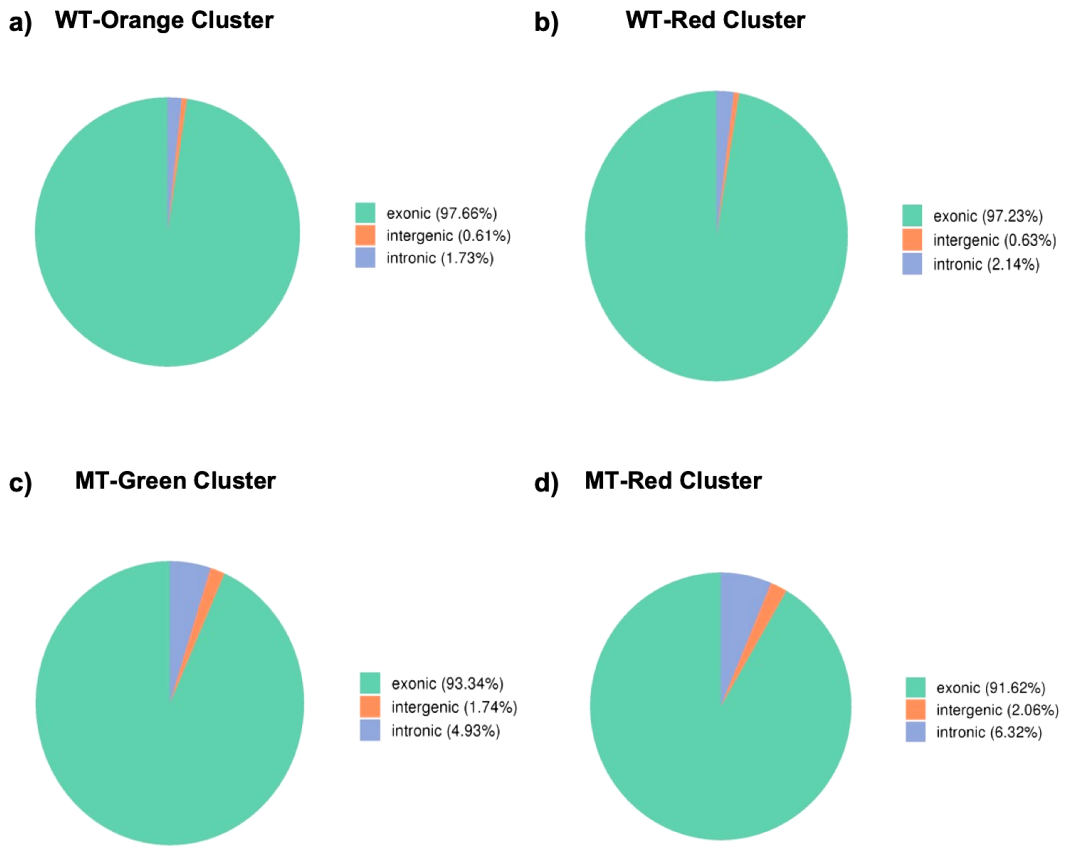


Figure 3.4.4.1. 2. Classification of Reads According to Mapped Region. a) wild-type cluster 5 (orange), b) wild-type cluster 1(red), c) *PIK3CA* mutant cluster 3 (green), d) *PIK3CA* mutant cluster 1 (red), Exons, introns, and intergenic regions are the three types of mapped areas. When the reference genome is well-annotated, exons should be the most abundant type of mapped reads. Introns may be derived from pre-mRNA contamination or intron-retention events from alternative splicing. Reads mapped to intergenic regions are mainly because of weak annotation of the reference genome. The exonic quality level (Q30) should be above 80%. All reads mapped be above 80%, and that passed quality control.

Numerous factors had to be examined to provide a complete statistical analysis of sequencing data, described in **Table 3.4.4.1. 2** and the percentage of reads to mapped region per sample in **Figure 3.4.4.1. 2**.

Table 3.4.4.1. 2. Overview of Mapping Status. Total reads Total number of filtered reads (Clean data). Total mapped: Total number of reads that can be mapped to the reference genome. Multiple mapped: Number of reads that can be mapped to multiple sites in the reference genome. Uniquely mapped: Number of reads that can be uniquely mapped to the reference genome. Read-1: Number of left read that can be mapped to the reference genome. Read-2: Number of right read that can be mapped to the reference genome. Reads map to '+': Number of reads that can be mapped to the positive strand (+). Reads map to '-': Number of reads that can be mapped to the minus strand (-). Non-splice reads: Number of reads that can be mapped entirely to a single exon. Splice reads: Number of reads that can be segmented and mapped to two exons (also named junction reads).

Sample name	PIK3CA MT	R12823_1	S14583B2	S14585C1
Cluster ID	red	green	orange	red
Total reads	216068712	230134348	264679988	281317822
Total mapped	182201484 (84.33%)	197105720 (85.65%)	163757967 (61.87%)	257965480 (91.70%)
Multiple mapped	7994386 (3.70%)	6982296 (3.03%)	5991948 (2.26%)	10375086 (3.69%)
Uniquely mapped	174207098 (80.63%)	190123424 (82.61%)	157766019 (59.61%)	247590394 (88.01%)
Read-1	87251661 (40.38%)	95215245 (41.37%)	79056721 (29.87%)	123881055 (44.04%)
Read-2	86955437 (40.24%)	94908179 (41.24%)	78709298 (29.74%)	123709339 (43.97%)
Reads map to '+'	87672503 (40.58%)	95169227 (41.35%)	79027498 (29.86%)	124007615 (44.08%)
Reads map to '-'	86534595 (40.05%)	94954197 (41.26%)	78738521 (29.75%)	123582779 (43.93%)
Non-splice reads	99596395 (46.09%)	104528876 (45.42%)	76673137 (28.97%)	120200725 (42.73%)
Splice reads	74610703 (34.53%)	85594548 (37.19%)	81092882 (30.64%)	127389669 (45.28%)

The Fragments per kilobase of transcript per million fragments (FPKM) mapped were used to compare gene expression levels estimated from different genes and studies. This was the most often used technique of calculating gene expression levels, considering the impacts of sequencing depth and gene length on fragment counts. The FPKM distribution diagram and boxplot were used to compare gene expression levels in each sample **Figure 3.4.4.1. 3a.**

Cluster analysis was performed to identify the genes from distinct metabolic regions with similar expression patterns. It is feasible to detect unknown functions of previously defined genes or roles of new genes by grouping genes with comparable expression patterns. Areas of various colours signify different groups (clusters) of genes in hierarchical clustering, and genes within each set may have comparable functions or participate in the same biological process. After that, the whole transcriptome of the samples was heat mapped using FPKM cluster analysis and the log10 (FPKM+1) value, **Figure 3.4.4.1. 3b.** This cluster analysis revealed the expression patterns of up-and

down-regulated genes. Genes with high levels of expression were emphasised in red, whereas genes with low levels of expression were highlighted in blue.

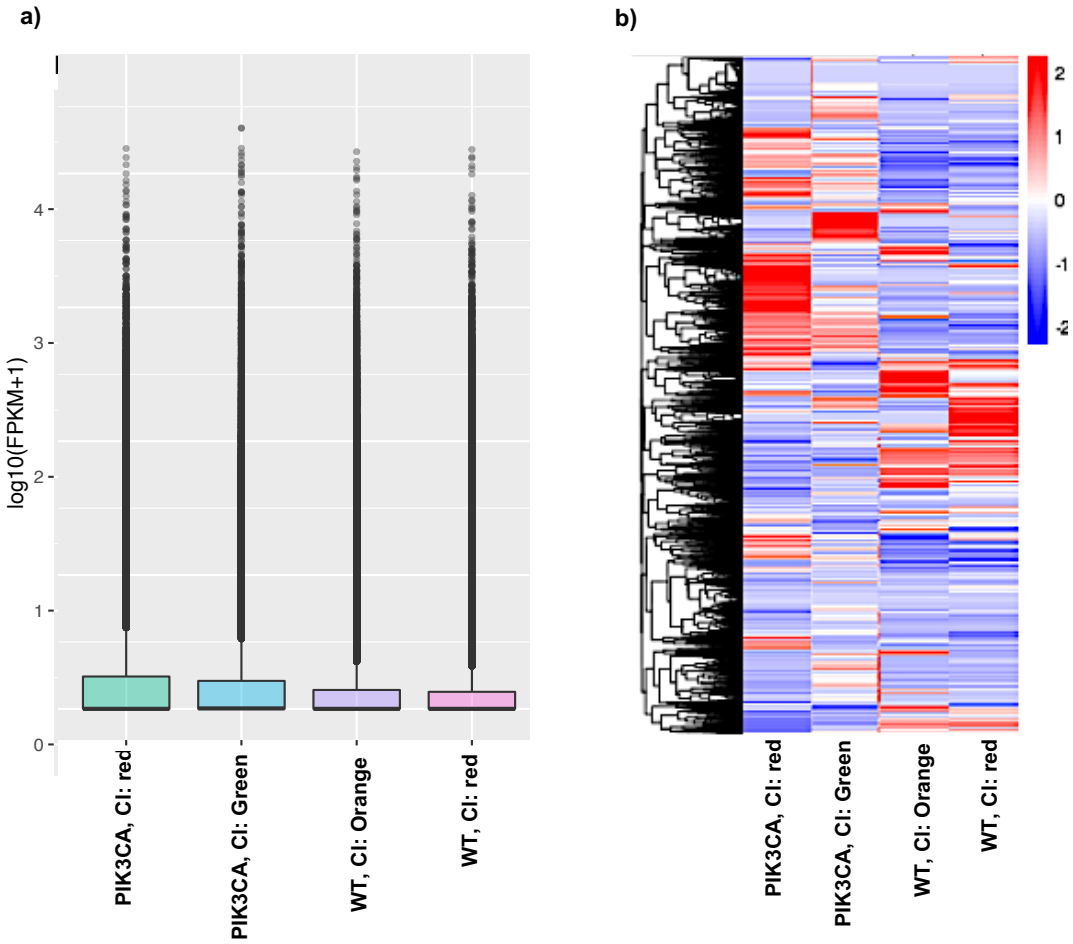


Figure 3.4.4.1. 3. FPKM distribution of different gene expression levels and Cluster analysis per sample. a) FPKM box Plot, the x-axis shows the sample names, and the y-axis shows the $\log_{10}$(FPKM+1). Each box had five statistical magnitudes (max value, upper quartile, median, lower quartile and min value). The box width shows the gene density. b) The overall results of FPKM cluster analysis, clustered using the $\log_{10}$(FPKM+1) value. The colour range from red to blue represents the high to low $\log_{10}$(FPKM+1) value. Red denotes genes with high expression levels, and blue indicates low expression levels.

DEGs were identified in these three different comparison datasets using *PIK3CA* MT against wild-type genes from the PDX models. The database analysis outlined in section 2.11.1.7 yielded 6606 DEGs, which included 3670 upregulated genes and 2396 downregulated genes, **Figure 3.4.4.1. 4a and d**.

12104 DEGs were found for all metabolic segments in the mutant PDX model (cluster red versus green, **Figure 3.4.4.1. 4b and e**), comprising 7706 upregulated genes and 4398 downregulated genes. The wild-type PDX model (red versus orange cluster, **Figure 3.4.4.1. 4c and f**) found 9152 DEGs, comprising 3709 upregulated genes and 5443 downregulated genes. For screening differentially expressed genes, volcano plots were used to determine the general distribution. For experiments, the standard thresholds are |log2(Fold Change)| > 1 and padj < 0.05 (figure a-c), as well as the number of differentially expressed genes per sample in a Venn diagram **Figure 3.4.4.1. 4d-f**.

The complete transcriptome from each sample set showed unique Differentially expressed genes (DEG) patterns and predominantly upregulated genes in the different regions of the same model, suggesting that the active driving pathways were not shared universally across the tissue.

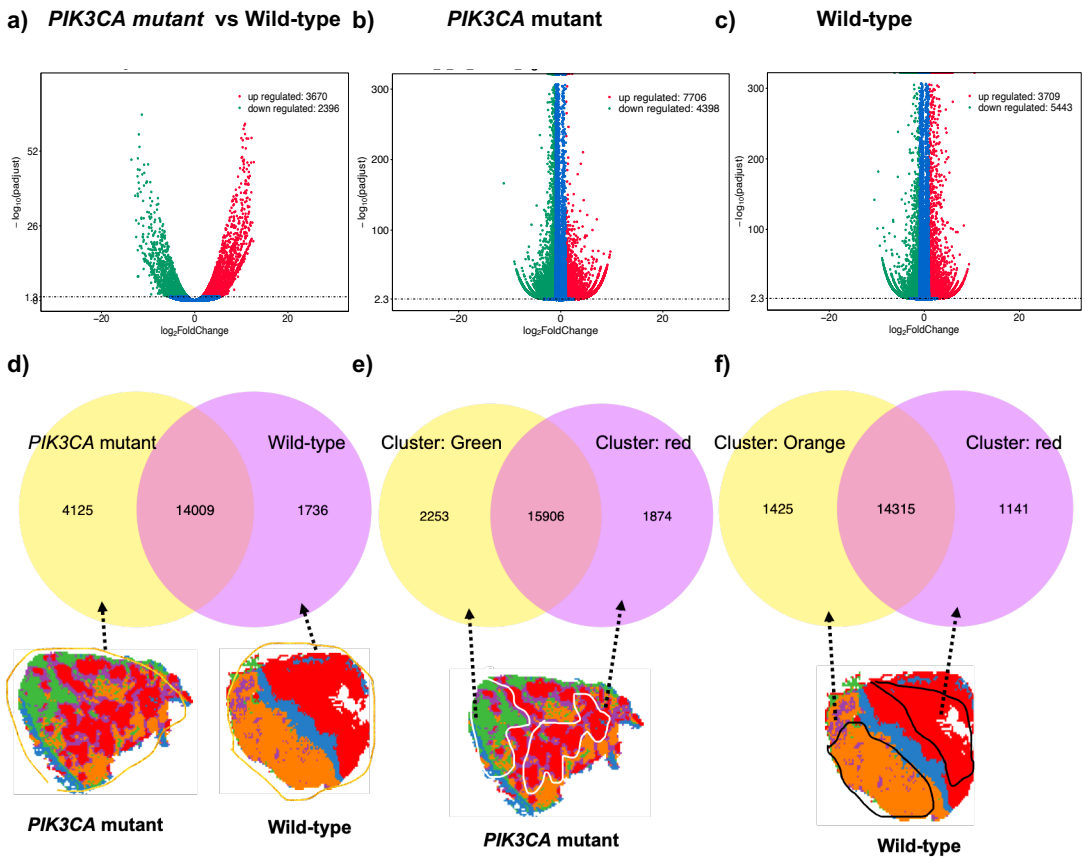


Figure 3.4.4.1. 4. Volcano plot for differentially expressed genes. The x-axis shows the fold change in gene expression between different samples, and the y-axis shows the statistical significance of the differences. Significantly up and down-regulated genes are highlighted in red and green, respectively. Genes that did not express differently between the treatment and control groups are blue. a) mutant *PIK3CA* versus wild-type PDX, b) mutant *PIK3CA* PDX: cluster green versus red, c) wild-type PDX: cluster orange versus red, d) the numbers of shared and unique differentially expressed genes comparison PDX versus wild-type sample set in the Venn diagram, e) mutant *PIK3CA* PDX: cluster green versus red, and f) wild-type PDX: cluster orange versus red $|\log_2(\text{Fold Change})| > 1$ and padj < 0.005.

3.4.5 Correlation of Transcriptomics and metabolomics dataset to reveal driver pathways behind metabolic clusters defined by DESI-MSI

The purpose of this study was to use the transcriptomics and metabolomics datasets in the MetaboAnalyst software to find possible correlations in the region of tissues defined by DESI-MSI clustering analysis. First, the DESI-MSI features were identified and classified utilising enrichment and pathway analysis by Metabolite Set Enrichment Analysis (MSEA). Next, the information from the microdissected metabolic clusters' expressed genes enrichment analysis in KEGG was connected to network analysis to find potential driving pathways.

3.4.5.1 Comparison of Mutant *PIK3CA* PDX with WT PDX model and metabolic segments in Network Analysis

MSEA revealed physiologically important patterns in metabolomic data by providing a global picture of changed biochemical processes and identifying substantially altered compounds. This study assessed individual metabolites for their significance in each sample set in the software analysis. The compounds that had attained a high enough relevance based on the preselected parameters were combined to determine if any relevant patterns emerged. Lactosylceramides (LacCer) were the most significant lipid class (p-value < 0.05, [FDR] < 0.05) in all PDX models based on this enriched metabolomic dataset from DESI-MSI of sample sets. Other lipid classes included O-PA, PI-Cer. and PA lipid classes, etc., **Table 3.4.5.1. 1**.

Table 3.4.5.1. 1. The most enriched lipid class of mutant *PIK3CA* PDX versus wild-type PDX model. The 'total' was the total number of compounds in the lipid classes; the 'hits' was the number of matched combinations from the uploaded data; the 'raw p' is the original p-value calculated from the enrichment analysis; the' holm p' is the p-value adjusted using the Holm-Bonferroni method, and the 'FDR' is the p-value adjusted using False Discovery Rate.

Lipid Classes	total	expected	hits	Raw p	Holm p	FDR
LacCer	4580	129	512	5.86E-156	4.29E-153	4.29E-153
O-PA	965	27.2	230	2.10E-139	1.53E-136	7.67E-137
PI-Cer	2290	64.5	303	2.48E-110	1.80E-107	6.03E-108
PA	1950	55	244	9.51E-84	6.93E-81	1.74E-81
PS	1970	55.5	240	4.48E-80	3.26E-77	6.55E-78
PE	1970	55.6	235	1.90E-76	1.38E-73	2.32E-74
PG	1950	54.8	226	1.74E-71	1.26E-68	1.81E-69
HETE	60	1.69	51	7.96E-70	5.76E-67	7.27E-68
HexCer	8690	245	543	1.06E-66	7.64E-64	8.59E-65
Unsaturated Fatty Acids	267	7.52	80	1.01E-57	7.26E-55	7.35E-56
SM	2320	65.2	196	2.38E-41	1.72E-38	1.58E-39
PC	1980	55.8	168	9.81E-36	7.06E-33	5.98E-34
PE-Cer	2290	64.5	178	5.19E-33	3.73E-30	2.92E-31
PI	1960	55.3	161	1.45E-32	1.04E-29	7.59E-31
Cer-1-P	2290	64.6	174	4.35E-31	3.12E-28	2.12E-29
O-PS	984	27.7	103	1.46E-29	1.05E-26	6.68E-28
HODEs	18	0.507	18	1.21E-28	8.67E-26	5.21E-27
Hydroxy Fatty Acids	274	7.72	53	1.99E-28	1.42E-25	8.07E-27
DGDG	1730	48.7	136	5.00E-26	3.56E-23	1.92E-24
O-PG	964	27.2	95	1.91E-25	1.36E-22	6.99E-24
O-PE	1100	30.8	101	9.43E-25	6.70E-22	3.28E-23
O-PI	964	27.2	90	1.40E-22	9.94E-20	4.65E-21
O-PC	1040	29.4	94	2.06E-22	1.46E-19	6.55E-21
EpETrE	12	0.338	12	2.46E-19	1.74E-16	7.51E-18
Cer	1610	45.4	115	5.28E-19	3.73E-16	1.54E-17

The Pathway Analysis module combines robust enrichment analysis results with pathway topology to identify the most relevant pathways involved in this study. The findings of the pathway analysis were depicted graphically in **Figure 3.4.5. 1c** and in-depth in **Table 3.4.5.1. 2**. To enable data exploration, an interactive visualisation system was built by the module. The top three most impacted pathways were arachidonic acid metabolism, biosynthesis of unsaturated fatty acids, and linoleic acid metabolism. The most impacted pathway, arachidonic acid metabolism, was chosen as a representative to provide graphical output from the metabolome view, **Figure 3.4.5. 1c**. The selected pathway library code was hsa, and pathway views and compound views were generated dynamically based on interactions with the visualisation system (KEGG organisms abbreviation). The red highlighted matched KEGG

IDs (C00157, C00350, C02737, etc.) in the Arachidonic acid belonged to the PC, PE, and PS Lipid classes, **Figure 3.4.5. 1c**.

Table 3.4.5.1. 2. The most relevant enriched pathways of mutant *PIK3CA* PDX versus wild-type PDX model in the metabolomics dataset.

Pathways	Total	Expected	Hits	Raw p	Holm p	FDR
Arachidonic acid metabolism	36	2.9265	23	1.14E-17	16.942	9.61E-16
Biosynthesis of unsaturated fatty acids	36	2.9265	12	1.15E-05	4.9381	0.00048437
Linoleic acid metabolism	5	0.40645	3	0.0046478	2.3328	0.13014
alpha-Linolenic acid metabolism	13	1.0568	2	0.28577	0.54398	1
Lipoic acid metabolism	8	0.65032	1	0.49332	0.30687	1
Sphingolipid metabolism	21	1.7071	2	0.51963	0.28431	1
Glycerophospholipid metabolism	36	2.9265	3	0.57155	0.24295	1
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	14	1.1381	1	0.69645	0.15711	1
Glycerolipid metabolism	16	1.3006	1	0.74423	0.1283	1
Fatty acid elongation	39	3.1703	2	0.84047	0.075477	1
Fatty acid degradation	39	3.1703	2	0.84047	0.075477	1
Phosphatidylinositol signaling system	28	2.2761	1	0.9089	0.041482	1
Steroid biosynthesis	42	3.4142	1	0.97298	0.011897	1
Primary bile acid biosynthesis	46	3.7394	1	0.98095	0.0083553	1
Fatty acid biosynthesis	47	3.8206	1	0.98254	0.0076491	1
Steroid hormone biosynthesis	85	6.9097	1	0.9994	0.00026065	1

The enrichment analysis could determine biological functions or pathways significantly associated with DEGs. The interactions of multiple genes may be involved in specific biological functions. The analysis identified the pathways which had the most significant involvement with the upregulated and downregulated. DEGs were submitted for the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. KEGG pathway analysis demonstrated that upregulated DEGs were significantly enriched for the *PIK3CA* MT versus WT PDX model in Metabolic Pathways, Ribosome, Cell Adhesion Molecules (CAMs), and phosphatidylinositol 3 kinase-protein kinase B (PI3K-AKT) signalling pathway respectively, **Figure 3.4.5. 2b.**

In the final step, the Network explorer analysis module augmented the MetaboAnalyst joint-pathway analysis module by detecting links in a gene-metabolite interaction network that correlates to features on specific pathways. This interaction allowed for a more comprehensive perspective of the pathways, which would not have been possible if they had been evaluated individually, **Figure 3.4.5. 2a**. The most interest in the gene-metabolite pathway interaction as a representative was arachidonic acid metabolism, **Figure 3.4.5. 2c** and others in **Figure 3.4.5. 2c.**

Table 3.4.5.1. 3. This network was built using HMDB (https://hmdb.ca) and STITCH (https://www.stitchdata.com) databases and only applies to human studies. The graphical output contained the metabolome view with the corresponding genes/compounds presented in

Table 3.4.5.1. ***4***. As seen in the joint-pathway analysis, KEGG orthologs (KO) as gene IDs in the arachidonic acid metabolic pathway were coloured, while genes up-regulated in red, and down-regulated in green. Yellow circles represented matched compounds to metabolites, **Figure 3.4.5. 2a**. Additionally, a model of the network of metabolites and genes assembled in

arachidonic acid metabolism are represented by blue squares and are associated with multiple genes, defined as blue circles, **Figure 3.4.5. 2c**.

Table 3.4.5.1. 3. The crossed pathways by detecting in gene-metabolite interaction network module of *mutant* PIK3CA PDX *versus wild-type* PDX model.

Network Pathways	Total	Expected	Hits	Raw p	Holm p	FDR
Arachidonic acid metabolism	81	11.884	37	9.16E-12	11.038	7.70E-10
Glycerolipid metabolism	35	5.1352	17	1.70E-06	5.7697	7.14E-05
Linoleic acid metabolism	17	2.4943	11	3.17E-06	5.4986	8.88E-05
Mucin type O-glycan biosynthesis	22	3.2279	12	1.34E-05	4.8714	0.00026623
Glycerophospholipid metabolism	86	12.618	28	1.58E-05	4.8	0.00026623
Ether lipid metabolism	39	5.7221	16	4.95E-05	4.3055	0.00069289
Nitrogen metabolism	10	1.4672	7	0.00011151	3.9527	0.0013381
Retinol metabolism	47	6.8959	17	0.00018696	3.7282	0.0019631
alpha-Linolenic acid metabolism	22	3.2279	10	0.00050793	3.2942	0.0047407
Glutathione metabolism	56	8.2164	17	0.0018735	2.7274	0.015737
Phosphatidylinositol signaling system	74	10.857	19	0.0082527	2.0834	0.057072
Lysine degradation	49	7.1893	14	0.0084198	2.0747	0.057072
Synthesis and degradation of ketone bodies	10	1.4672	5	0.0088326	2.0539	0.057072
Drug metabolism - other enzymes	70	10.27	18	0.009865	2.0059	0.05919

Table 3.4.5.1. 4. The matched genes/compounds in the gene-metabolite interaction network module of mutant *PIK3CA* PDX versus wild-type PDX model.

Network Pathways	matched_features (genes/compounds)
Arachidonic acid metabolism	cpd:C05966; cpd:C05356; cpd:C00157; cpd:C00909; cpd:C02165; cpd:C02166; cpd:C14823; cpd:C14812; cpd:C05965; cpd:C04853; hsa:1573; hsa:247; hsa:242; hsa:4051; hsa:8529; hsa:2877; hsa:2882; hsa:493869; hsa:11283; hsa:246; hsa:5742; hsa:5743; hsa:123745; hsa:151056; hsa:255189; hsa:283748; hsa:5321; hsa:5322; hsa:64600; hsa:8605; hsa:240; hsa:2678; hsa:2687; hsa:5740; hsa:5730; hsa:873; hsa:874
Glycerolipid metabolism	cpd:C00416; hsa:119548; hsa:8513; hsa:9388; hsa:64900; hsa:80168; hsa:150763; hsa:57016; hsa:217; hsa:224; hsa:4023; hsa:2717; hsa:1606; hsa:1607; hsa:160851; hsa:9162; hsa:132158
Linoleic acid metabolism	cpd:C00157; hsa:1573; hsa:246; hsa:123745; hsa:151056; hsa:255189; hsa:283748; hsa:5321; hsa:5322; hsa:64600; hsa:8605
Glycerophospholipid metabolism	cpd:C00350; cpd:C00157; cpd:C00416; hsa:10390; hsa:56994; hsa:1119; hsa:43; hsa:1103; hsa:123745; hsa:151056; hsa:255189; hsa:283748; hsa:5321; hsa:5322; hsa:64600; hsa:8605; hsa:1606; hsa:1607; hsa:160851; hsa:9162; hsa:64900; hsa:122618; hsa:5337; hsa:162466; hsa:150763; hsa:375775; hsa:1040; hsa:23659
alpha-Linolenic acid metabolism	cpd:C00157; hsa:123745; hsa:151056; hsa:255189; hsa:283748; hsa:5321; hsa:5322; hsa:64600; hsa:8605; hsa:9415
Phosphatidylinositol signaling system	cpd:C00416; hsa:4534; hsa:55613; hsa:9108; hsa:51196; hsa:5332; hsa:84812; hsa:89869; hsa:5288; hsa:1040; hsa:1606; hsa:1607; hsa:160851; hsa:9162; hsa:8821; hsa:3613; hsa:3705; hsa:3635; hsa:117283
Sphingolipid metabolism	cpd:C00195; cpd:C01190; hsa:130367; hsa:10715; hsa:204219; hsa:123099; hsa:55304; hsa:2531; hsa:2717; hsa:4758; hsa:7368; hsa:125981; hsa:8877; hsa:166929

Using these numerous analyses to reveal enriched pathways, compounds and genes behind these metabolic clusters identified by DESI-MSI could aid in understanding the complexity of heterogeneous tumour tissues. As an encouraging outcome of our exploratory research, the most critical pathway findings from each type of analysis were matched to previously reported results in the literature.

The top lipid class found by enrichment analysis in this dataset were ceramides, especially Lactosylceramides (LacCer), a class of glycosphingolipids composed of a variable hydrophobic ceramide lipid, a

hydrophilic sugar moiety. LacCer activates an enzyme, cytosolic phospholipase A2 (cPLA2), to generate arachidonic acid (AA), a precursor to prostaglandins[203]. In turn, arachidonic acid serves as a precursor to eicosanoids and prostaglandin, which transduce a cascade of reactions leading to inflammation, cell proliferation, migration/infiltration, adhesion, angiogenesis, apoptosis, autophagy, and mitochondrial dysfunction. A significant phenotype underscored the initiation and progression of several debilitating diseases, such as breast cancer[203].

These findings of enriched lipids, as seen in **Figure 3.4.5. 1b**, the lipid class were consistent with the fact that the most enriched pathways were arachidonic acid metabolism in **Figure 3.4.5. 1c**, also from the analysis of the gene expressions in **Figure 3.4.5. 2b**. The role of arachidonic acid metabolism confirms that the cascade of reactions activated multiple metabolic pathways, resulting in the top enriched pathway in the KEGG analysis in **Figure 3.4.5. 2b.** The second top enriched pathway was the Ribosome pathway. The role of dysregulated ribosome biogenesis in the development and progression of most cancers has already been established[204,205]. However, the surprising finding was a potential link of ribosome activity with arachidonic acid, which is currently not well characterised. Only in a research study[206], AA supplementation on a group of men, following 48 h of recovery, markers of ribosome biogenesis increased significantly compared with placebo[206]. The explanation of a potential link between AA supplementation and ribosome biogenesis could be via the inflammatory cytokine IL-6, which has been shown to promote pre-rRNA synthesis in cancer cells[207]. The other enriched pathway was the PI3K-AKT signalling, which was expected as the PDX model having *PIK3CA* mutant[24].

Finally, network analysis based on the top hit genes and compounds in the whole metabolome and transcriptome information matched the arachidonic acid metabolism pathway in **Figure 3.4.5. 2b-c**. This network confirmed the previous findings of the separate KEGG gene and MSEA metabolite enrichment analysis, where arachidonic acid metabolism was the top enriched in the *PIK3CA* mutant PDX model.

Furthermore, IMC images of particular ROIs from mutant *PIK3CA* and WT PDX models revealed protein profiling with selected signalling and architecture markers, focusing on PTEN and Avb3 protein expression, as shown in **Figure 3.4.5.1d.** Even though PTEN loss of expression (20%) but not mutations (10%) is the most common alteration in *PIK3CA* (30%) mutated tumours[208,209] , this suggests that there are different genetic abnormalities in the clusters defined by DESI-MSI in mutant PDX model. The green cluster has no PTEN loss and considerable expression (yellow marker). Indeed, the differential expression analysis of the transcriptomics dataset showed a significant fold change in PTEN protein when comparing cluster red against green. In summary, there is no concordance between mutant PTEN and *PIK3CA* in the green cluster from the *PIK3CA* mutated PDX model and the red cluster from the wild-type PDX model.

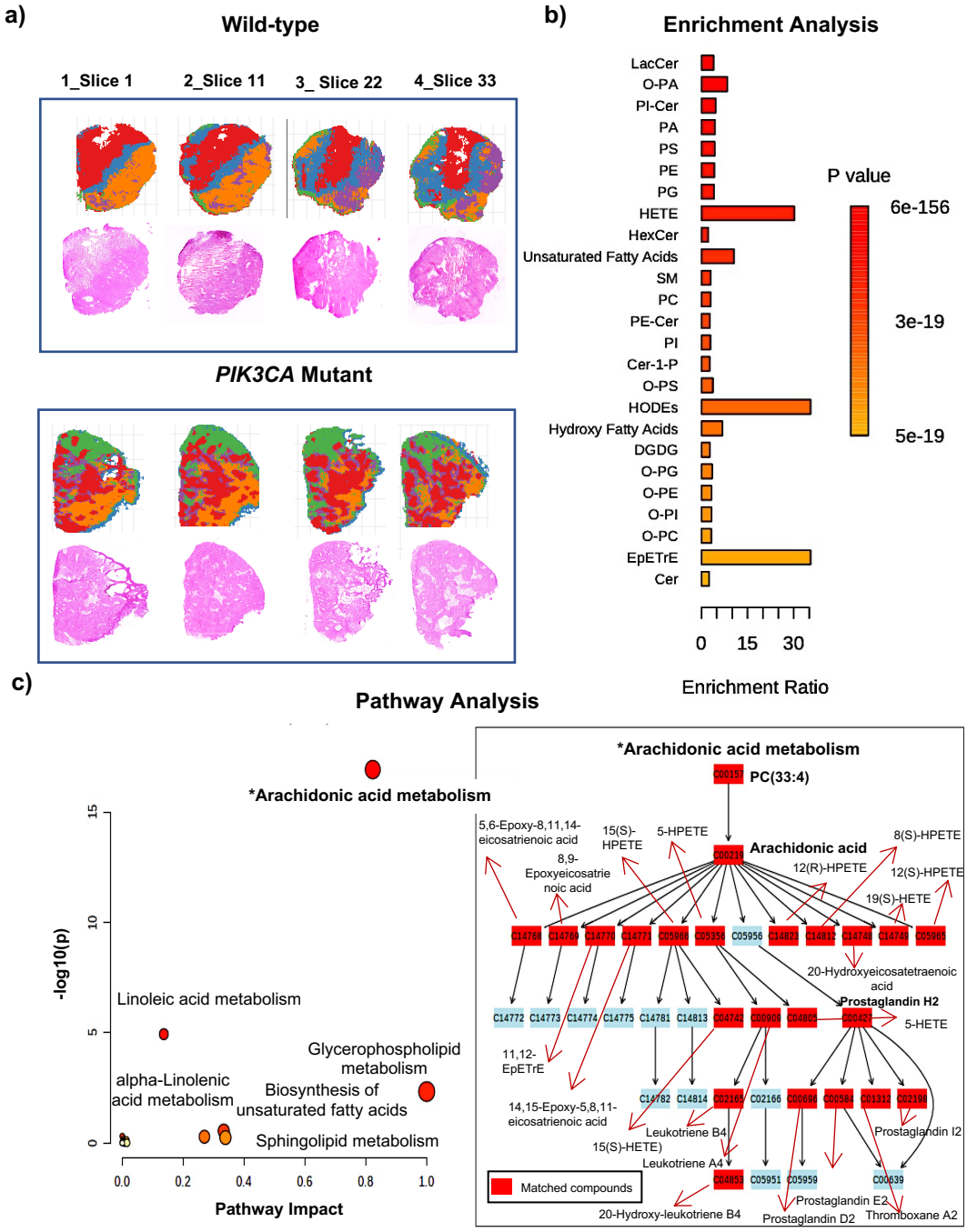


Figure 3.4.5. 1. Metabolic Enrichment Analysis of the DESI- MSI cluster of Mutant *PIK3CA* PDX with WT PDX model. a) *PIK3CA* mutant and wild-type PDX model sample set b) Overview of enriched metabolomic content, the significant lipid classes in colour grading red (high) to white (less) for the most significant (p-value< 0.05). c) The global view of the pathways in the interaction network to show graphical output contained the metabolome view with the related compounds -the most impacted on the metabolite pathway interaction was arachidonic acid metabolism, red highlighted KEGG labelled compounds matched tentatively with PC., Cer. and Eicosatrioneic acid and Prostaglandin classes. All statistical analysis threshold, p-value < 0.05, [FDR] < 0.05.

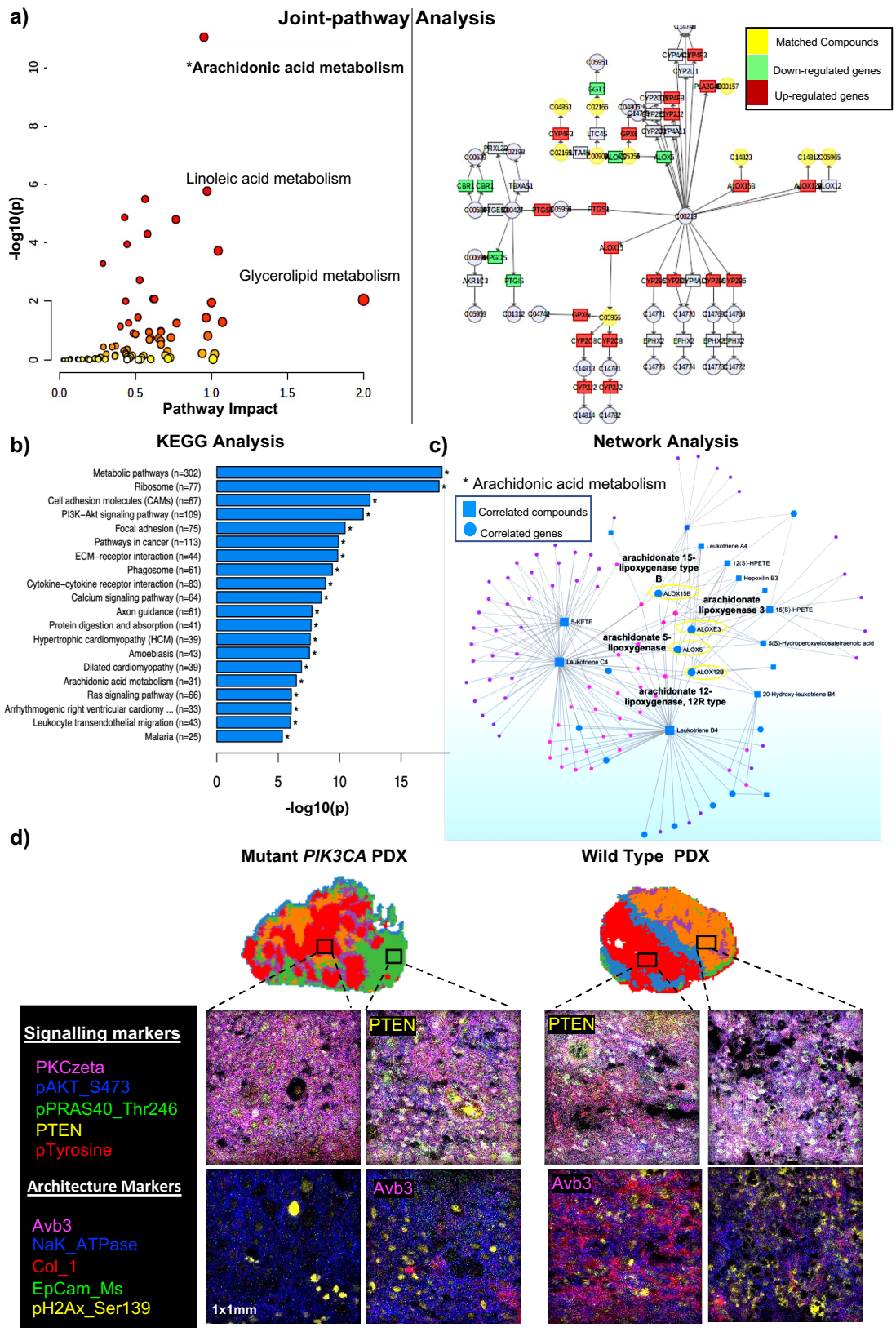


Figure 3.4.5. 2. Correlation of Transcriptomics and metabolomics behind the DESI- MSI cluster of Mutant *PIK3CA* PDX with WT PDX model. a) Summary of MetaboAnalyst joint-Pathway Analysis in Network Explorer module, all the matched pathways are displayed as circles. The colour and size of each circle are based on the p-value and pathway impact value, respectively. Arachidonic Acid metabolism

had the most impact on the gene-metabolite pathway connection. The circle in yellow differential metabolites and the nodes in red indicate up-regulated and green down-regulated expression of genes, b) KEGG enrichment analysis to identify the pathways which had the most significant involvement with the upregulated and downregulated genes. c) The plot of the selected interaction sub-network, arachidonic acid, in the Network Explorer. d) The IMC images show the protein level differences of PTEN and Avb3 protein markers, All statistical analysis thresholds, p-value < 0.05, [FDR] < 0.05.

3.4.5.2 Comparison of microdissected MSI-guided ROIs in clusters of Mutant *PIK3CA* and Wild-type PDX model in Network Analysis

The global gene expression of mutant versus wild-type was discussed in the previous section. This section investigates the selected ROIs on the MSI clusters for intra-tumour heterogeneity in the individual PDX model.

In mutant *PIK3CA* PDX models, before correlating the transcriptome dataset, first, the features detected by DESI-MSI were studied with MSEA. Based on this enriched metabolomic dataset from DESI-MSI of sample sets, the most significant lipid class (p-value < 0.05, [FDR] < 0.05) was Hexosylceramides (HexCer). The 24 distinct lipid classes followed were Cer, PI-Cer, DHCer, MGDG, etc., **Figure 3.4.5.2. 1b** and **Table 3.4.5.2.1.** In wild-type PDX models, the most significant lipid class (p-value < 0.05, [FDR] < 0.05) was Hydroxy Fatty Acids and followed by hydroperoxy-octadecadienoic acid (HpODEs), DiHODEs, EpODEs, and PE etc., **Figure 3.4.5.2.2b** and **Table 3.4.5.2.2.**

The Pathway Analysis module combines the most significant lipid classes to identify the most relevant pathways involved in these MSI clusters. The findings of the pathway analysis were depicted graphically in **Figure 3.4.5.2.1c and Figure 3.4.5.2c.** In mutant *PIK3CA* PDX models, the top three were sphingolipid metabolism, glycosylphosphatidylinositol (GPI)- anchor biosynthesis, and glycerophospholipid metabolism most enriched pathways. The most enriched pathway, sphingolipid metabolism, was chosen as a representative to provide graphical output from the metabolome view in **Figure 3.4.5.2. 1c** and **Table 3.4.5.2.1**. In wild-type PDX models, linoleic acid metabolism, glycerophospholipid metabolism and alpha-linolenic acid metabolism were the top three most enriched pathways. The most enriched

pathway, linoleic acid metabolism, was chosen as a representative to provide graphical output from the metabolome view in **Figure 3.4.5.2.2c** and **Table 3.4.5.2.2.**

Table 3.4.5.2.1. The most relevant enriched pathways in the metabolomics dataset of the PIK3CA mutant PDX model (cluster red with green).

Lipid Classes	Total	Expected	Hits	Raw p	Holm p	FDR
Sphingolipid metabolism	21	0.40645	4	0.00054386	3.2645	0.045685
Glycerophospholipid metabolism	36	0.69677	4	0.0043699	2.3595	0.18354
Glycerolipid metabolism	16	0.30968	2	0.036734	1.4349	1
Linoleic acid metabolism	5	0.096774	1	0.093216	1.0305	1
alpha-Linolenic acid metabolism	13	0.25161	1	0.22514	0.64754	1
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	14	0.27097	1	0.24027	0.6193	1
Phosphatidylinositol signaling system	28	0.54194	1	0.42427	0.37236	1
Arachidonic acid metabolism	36	0.69677	1	0.50922	0.2931	1
Fatty acid degradation	39	0.75484	1	0.53784	0.26935	1
Steroid hormone biosynthesis	85	1.6452	1	0.81886	0.08679	1

Table 3.4.5.2.2. The most relevant enriched pathways in the metabolomics dataset of Wild type PDX model(cluster red with orange).

Enriched Pathways	Total	Expected	Hits	Raw p	Holm p
Linoleic acid metabolism	5	0.20968	2	0.015958	1.797
Glycerophospholipid metabolism	36	1.5097	4	0.060575	1.2177
alpha-Linolenic acid metabolism	13	0.54516	2	0.10034	0.99855
Biosynthesis of unsaturated fatty acids	36	1.5097	3	0.18925	0.72297
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	14	0.5871	1	0.45248	0.34441
Glycerolipid metabolism	16	0.67097	1	0.49785	0.3029
Phosphatidylinositol signaling system	28	1.1742	1	0.7019	0.15372
Glycine, serine and threonine metabolism	33	1.3839	1	0.76042	0.11895
Arachidonic acid metabolism	36	1.5097	1	0.78993	0.10241
Fatty acid degradation	39	1.6355	1	0.81586	0.088385
Steroid biosynthesis	42	1.7613	1	0.83863	0.07643
Steroid hormone biosynthesis	85	3.5645	1	0.97639	0.010376

KEGG enrichment could determine which biological functions or pathways are significantly associated in the transcriptome dataset from the microdissected

tissue ROIs. KEGG analysis demonstrated that the top three pathways that were significantly enriched for mutant *PIK3CA* PDX model in the clusters 1 (red) and 3 (green) ROIs were Metabolic Pathways, Parkinson's disease and Oxidative phosphorylation in **Figure 3.4.5.2. 1d.** For wild type PDX model, in the clusters 1 (red) and 5 (orange) ROIs, the top three enriched pathways were Metabolic Pathways, HIF-1 signalling pathway and Neuroactive ligand-receptor interaction in **Figure 3.4.5.2.2d.**

In the final step, the Network explorer analysis module augmented the MetaboAnalyst joint-pathway analysis module by detecting links in a gene-metabolite interaction network that correlates to features on specific pathways. In the mutant *PIK3CA* PDX model, purine metabolism was most interested in the gene-metabolite pathway interaction as a representative in **Figure 3.4.5.2.1e** and others in **Table 3.4.5.2.3** with matched genes and compounds**.** In the wild-type PDX model, the most interest in the gene-metabolite pathway interaction as a representative was glycerophospholipid metabolism in **Figure 3.4.5.2.2e** and others in **Table 3.4.5.2.4** with matched genes and compounds**.**

Table 3.4.5.3. The matched genes/compounds in the gene-metabolite interaction network module of the PIK3CA mutant PDX model (cluster red with green).

Network Pathways	matched_features (genes/compounds)
Purine metabolism	cpd:C00117; cpd:C00620; hsa:50484; hsa:6240; hsa:6241; hsa:7498; hsa:221823; hsa:5631; hsa:5634; hsa:5471; hsa:2618; hsa:471; hsa:158; hsa:10606; hsa:10846; hsa:50940; hsa:5136; hsa:5137; hsa:5138; hsa:5139; hsa:5140; hsa:5153; hsa:27115; hsa:5141; hsa:5142; hsa:5143; hsa:5144; hsa:5150;
Drug metabolism - other enzymes	hsa:7172; hsa:3251; hsa:8833; hsa:3704; hsa:7498; hsa:10; hsa:9; hsa:151531; hsa:7378; hsa:1890; hsa:1806; hsa:1807; hsa:51733; hsa:7372; hsa:54963; hsa:7371; hsa:83549; hsa:7083; hsa:7084; hsa:3614; hsa:3615; hsa:1548; hsa:978; hsa:1066; hsa:8824; hsa:10720; hsa:10941;
Glycerophospholipid metabolism	cpd:C00346; cpd:C00093; cpd:C01233; hsa:81490; hsa:9791; hsa:8443; hsa:10162; hsa:254531; hsa:54947; hsa:79888; hsa:10390; hsa:56994; hsa:5130; hsa:9468; hsa:1119; hsa:1120; hsa:43; hsa:1103; hsa:5833; hsa:55224; hsa:55500; hsa:100137049; hsa:123745; hsa:151056; hsa:255189; hsa:26279; hsa:283748;
Steroid hormone biosynthesis	cpd:C05498; hsa:412; hsa:6820; hsa:1583; hsa:1586; hsa:1589; hsa:3283; hsa:3284; hsa:9420; hsa:1584; hsa:1585; hsa:100861540; hsa:1543; hsa:1544; hsa:1551; hsa:1577; hsa:1571; hsa:1576; hsa:8644; hsa:6715; hsa:6716; hsa:79644; hsa:1646; hsa:6718; hsa:1109; hsa:1645; hsa:3290;
Arachidonic acid metabolism	cpd:C14781; cpd:C14813; cpd:C05966; cpd:C05356; cpd:C05956; cpd:C02165; cpd:C00427; cpd:C00696; cpd:C00584; cpd:C14782; cpd:C14814; cpd:C14823; cpd:C14812; cpd:C05965; cpd:C04853; cpd:C01312; cpd:C02198; cpd:C05959; cpd:C00639; hsa:2053; hsa:1555; hsa:1573; hsa:1559; hsa:1558; hsa:1579;
Glutathione metabolism	cpd:C00051; hsa:290; hsa:51056; hsa:124975; hsa:2678; hsa:2686; hsa:2687; hsa:257202; hsa:2876; hsa:2877; hsa:2878; hsa:2880; hsa:2882; hsa:493869; hsa:2879; hsa:51060; hsa:2539; hsa:3417; hsa:3418; hsa:2936; hsa:2937; hsa:2729; hsa:2730; hsa:26873; hsa:79017; hsa:51471; hsa:9027;
Amino sugar and nucleotide sugar metabolism	cpd:C01674; cpd:C00329; hsa:55577; hsa:5238; hsa:6675; hsa:91373; hsa:10020; hsa:54187; hsa:140838; hsa:55907; hsa:2592; hsa:2582; hsa:7360; hsa:7358; hsa:80146; hsa:10007; hsa:132789; hsa:51005; hsa:4351; hsa:5372; hsa:5373; hsa:2762; hsa:5236; hsa:55276;
Glycerolipid metabolism	cpd:C00093; hsa:1056; hsa:119548; hsa:3990; hsa:5406; hsa:5407; hsa:5408; hsa:57104; hsa:80339; hsa:8513; hsa:9388; hsa:346606; hsa:84649; hsa:8694; hsa:23175; hsa:64900; hsa:9663; hsa:116255; hsa:80168; hsa:10554; hsa:10555; hsa:129642; hsa:154141; hsa:253558;

Table 3.4.5.4. The matched genes/compounds in the gene-metabolite interaction network module of Wild type PDX model(cluster red with orange).

Network Pathways	matched_features (genes/compounds)
Glycerophospholipid metabolism	cpd:C00346; cpd:C00093; cpd:C01233; hsa:81490; hsa:9791; hsa:8443; hsa:10162; hsa:254531; hsa:54947; hsa:79888; hsa:10390; hsa:56994; hsa:5130; hsa:9468; hsa:1119; hsa:1120; hsa:43; hsa:1103; hsa:5833; hsa:55224; hsa:55500; hsa:100137049; hsa:123745; hsa:151056; hsa:255189; hsa:26279;
Purine metabolism	cpd:C00262; cpd:C00387; hsa:50484; hsa:6240; hsa:6241; hsa:7498; hsa:221823; hsa:5631; hsa:5634; hsa:5471; hsa:2618; hsa:471; hsa:158; hsa:10606; hsa:10846; hsa:50940; hsa:5136; hsa:5137; hsa:5138; hsa:5139; hsa:5140; hsa:5153; hsa:27115; hsa:5141; hsa:5142; hsa:5143; hsa:5144; hsa:5150; hsa:5151;
Phosphatidylinositol signaling system	cpd:C03546; cpd:C01177; cpd:C04006; hsa:5728; hsa:4534; hsa:55613; hsa:64419; hsa:8776; hsa:8897; hsa:8898; hsa:9107; hsa:9108; hsa:9110; hsa:64768; hsa:113026; hsa:23236; hsa:51196; hsa:5330; hsa:5331; hsa:5332; hsa:5333; hsa:5335; hsa:5336; hsa:84812; hsa:89869; hsa:10423;
Inositol phosphate metabolism	cpd:C00092; cpd:C03546; cpd:C04006; cpd:C01177; hsa:5305; hsa:79837; hsa:8396; hsa:3633; hsa:4952; hsa:56623; hsa:8867; hsa:8871; hsa:253430; hsa:5728; hsa:9562; hsa:27124; hsa:3632; hsa:51763; hsa:3706; hsa:3707; hsa:80271; hsa:3705; hsa:3628; hsa:3631; hsa:8821; hsa:4534; hsa:51477; hsa:3612;
Glycolysis or Gluconeogenesis	cpd:C05345; cpd:C00668; cpd:C00103; cpd:C01172; hsa:217; hsa:219; hsa:223; hsa:224; hsa:501; hsa:218; hsa:220; hsa:221; hsa:222; hsa:10327; hsa:124; hsa:125; hsa:126; hsa:127; hsa:128; hsa:130; hsa:131; hsa:5160; hsa:5161; hsa:5162; hsa:1737; hsa:160287; hsa:3939; hsa:3945; hsa:3948; hsa:92483
Metabolism of xenobiotics by cytochrome P450	cpd:C19562; hsa:1571; hsa:1543; hsa:1559; hsa:1576; hsa:1545; hsa:1544; hsa:119391; hsa:221357; hsa:27306; hsa:2938; hsa:2939; hsa:2940; hsa:2941; hsa:2944; hsa:2946; hsa:2947; hsa:2948; hsa:2949; hsa:2953; hsa:373156; hsa:4257; hsa:4258; hsa:4259; hsa:653689; hsa:9446; hsa:10720; hsa:10941;
Pyrimidine metabolism	cpd:C00105; hsa:50484; hsa:6240; hsa:6241; hsa:790; hsa:7372; hsa:10201; hsa:221264; hsa:29922; hsa:4830; hsa:4831; hsa:4832; hsa:4833; hsa:654364; hsa:129607; hsa:51727; hsa:1503; hsa:56474; hsa:377841; hsa:953; hsa:956; hsa:124583; hsa:955; hsa:957; hsa:9583; hsa:139596; hsa:54963; hsa:7371;
Glutathione metabolism	cpd:C00051; cpd:C00025; cpd:C01879; hsa:290; hsa:51056; hsa:124975; hsa:2678; hsa:2686; hsa:2687; hsa:257202; hsa:2876; hsa:2877; hsa:2878; hsa:2880; hsa:2882; hsa:493869; hsa:2879; hsa:51060; hsa:2539; hsa:3417; hsa:3418; hsa:2936; hsa:2937; hsa:2729; hsa:2730; hsa:26873; hsa:79017; hsa:51471;

Findings on biochemical activities in the mutant *PIK3CA* PDX model were confirmed with reference to the literature. First, it was started with enrichment

analysis of DESI-MSI detected metabolites; the top ones were Ceramides, especially HexCer (monohexosylceramides). The HexCer were the most abundant form with SM, serving as a substrate for synthesising more complex sphingolipids[210]. These findings of the enriched lipid class, HexCer are consistent with the pathway analysis results; as seen in **Figure 3.4.5.2. 1c**, the most enriched pathways were sphingolipid metabolism. One published study has reported that sphingolipid metabolism is a critical regulatory molecule that controls various aspects of cell growth and proliferation in cancer and is especially important in breast cancer progression[210]. The correlation of the sphingolipid metabolism function in the KEGG analysis also reveals that high activity of growth and proliferation outcomes as metabolic pathways are the most significantly enriched compared to other pathways.

Interestingly, the second most enriched pathway was Parkinson's disease. The role of sphingolipids in Parkinson's and most neurodegenerative diseases such as Alzheimer's (4th most enriched pathway in the KEGG, **Figure 3.4.5.2. 1d**) has been reported over the past decade. It was found that relatively simple sphingolipids, such as ceramide, play essential roles in neuronal functions by regulating neuronal growth and differentiation rates. Numerous experiments have highlighted the involvement of ceramide, especially HexCer metabolism, in the pathophysiology of the sporadic forms of Parkinson's disease [211]. As a result, this emergence could be linked to Parkinson's disease rather than cancer in the first instance. However, the promising outcome is that the transcriptome-metabolome data correlation could reflect the disease phenotype.

The next enriched pathway was oxidative phosphorylation. As reported previously in this study[212], the increased intracellular concentration of sphingolipids, especially ceramides, impedes mitochondrial function by interfering with multiple aspects of mitochondrial electron transport chain (ETC) and oxidative phosphorylation (OXPHOS), ATP production and mitochondrial biogenesis[212]. These findings also were correlated with the network analysis results in **Figure 3.4.5.2. 1e**. The most enriched pathway was purine metabolism, which could be enhanced by mitochondria

dysregulation (i.e., defect in electron transport or oxidative phosphorylation)[213]. Additionally, purine metabolism is enhanced in tumour cells because purine nucleotides are fundamental and necessary for tumour cell proliferation, affecting the growth, invasiveness, and metastasis of the tumour[213]. The IMC images show the protein level differences of pHistone3 and ki67 correlated with mitotic activity/tumour cell proliferation might reflect on the high purine metabolism activity. This protein level alteration was more noticeable in the *PIK3CA* mutant model's cluster red in **Figure 3.4.5.2. 1f.**

Overall, these network analysis findings based on the top hit genes and compounds in the whole metabolome and transcriptome information are consistent with the activation of purine metabolism in **Figure 3.4.5.2. 1e** and are encouraging. This joint network is consistent with the previous findings of the KEGG gene and MSEA metabolite enrichment analysis that sphingolipid metabolism and oxidative phosphorylation were the top enriched driver pathways in the cluster red of the *PIK3CA* mutant PDX model.

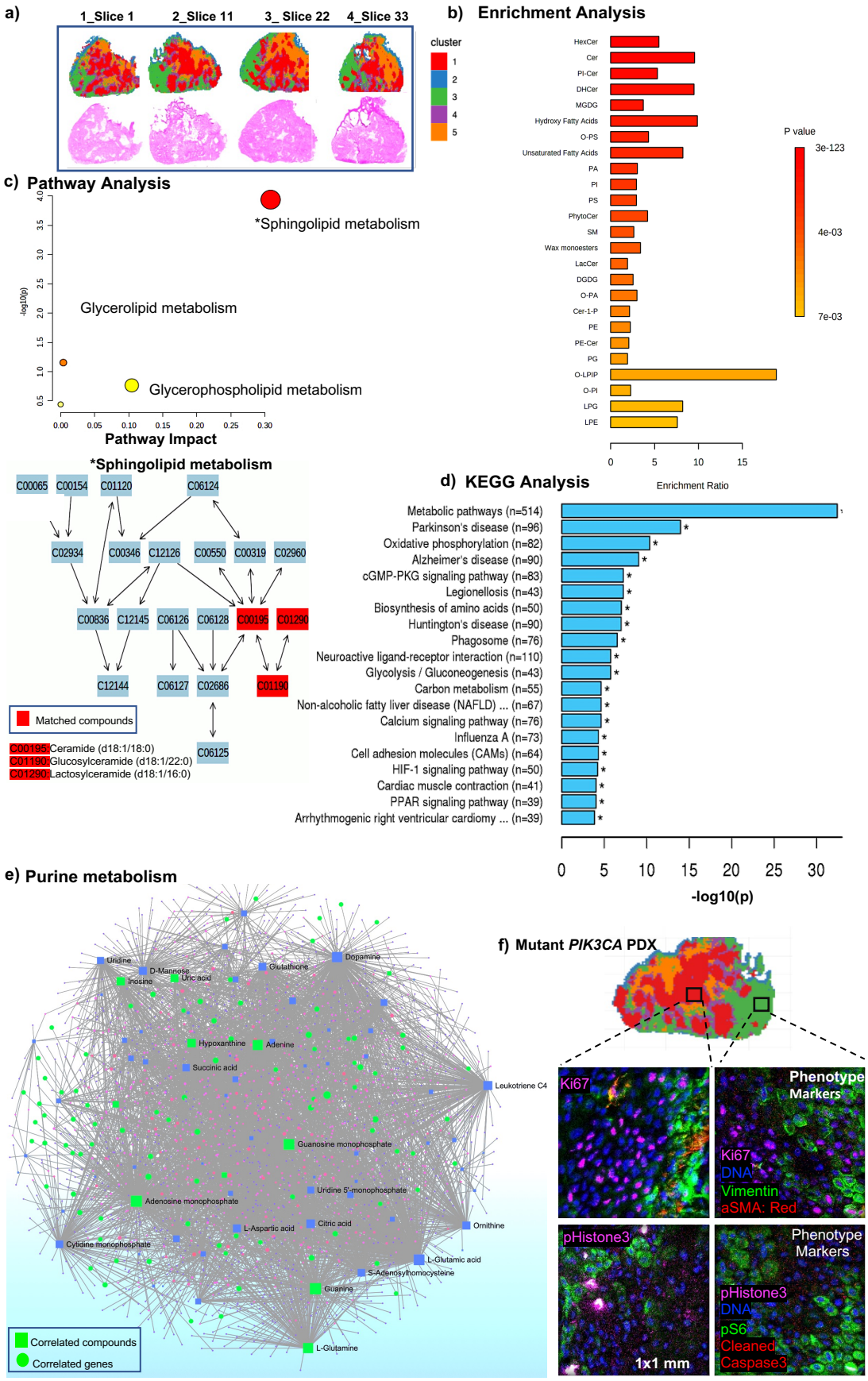


Figure 3.4.5.2. 1. Correlation of Transcriptomics and metabolomics behind the DESI- MSI cluster of *PIK3CA* mutant PDX model (cluster red with green) .a)

PIK3CA mutant PDX model sample set. b) Metabolite Enrichment Overview, Summary Plot for Over Representation Analysis (ORA), The enriched metabolomic content, the significant lipid classes in colour grading red (high) to white (less) for the most significant (p-value< 0.05). c) The global view of the pathways in the interaction network to show graphical output contained the metabolome view with the related compounds - the most impacted on the metabolite pathway interaction was Sphingolipid metabolism, red highlighted KEGG labelled compounds matched tentatively with Cer, Lactosylceramide and Glucosylceramide classes d) KEGG enrichment analysis to identify the pathways which had the most significant involvement with the upregulated and downregulated genes e) The plot of the selected interaction sub-network, Purine metabolism in the Network Explorer. f) The IMC images show the protein level differences of Ki67 and pHistone3 protein markers, All statistical analysis thresholds, p-value < 0.05, [FDR] < 0.05.

The biochemical activities discovered in the wild-type PDX model, starting with metabolite enrichment analysis, the most enriched metabolites were Hydroxy Fatty Acids, the most abundant fatty acid in Linoleic acid metabolism. Oxidation of linoleic acid to hydroperoxy derivatives (HODEs) with different forms was also the most enriched lipid class in this analysis, such as HpODEs, DiHODEs, EpODEs etc. [214,215]. These findings of enriched lipid class, Hydroxy Fatty Acids, could confirm the pathway analysis results; as seen in **Figure 3.4.5.2.2b,** the most enriched pathways were linoleic acid metabolism due to a possible high abundance of hydroxy fatty acids. Published studies have reported that linoleic acid has one of the most common cancer-promoting effects with arachidonic acid in breast cancer[216,217]. The linoleic acid metabolism role in the transcriptome dataset also confirms that high growth and proliferation activity results in significantly enriched metabolic pathways while comparing other pathways in the KEGG analysis in **Figure 3.4.5.2.2d.** The second top enriched pathway was the HIF-1 signalling pathway, which promotes glycolysis and fatty acid oxidation, including linoleic acid metabolism, results in malignant cells survival and has crosstalk between PTEN signalling [218,219].

The IMC images show the protein level differences of PTEN and Avb3 protein markers correlated with Hypoxia-inducible factor-1 (HIF-1), and this protein level alteration was more noticeable in the wild-type model's cluster 1 (red) compared to cluster 3 (green) in **Figure 3.4.5.2.2f.** Alteration in the entire

antibody panel of 25 proteins, including immunological, phenotypic, signalling, and architectural proteins, differentiate cell types and states in the Appendix, **Figure 3.6.3.** Additionally, in the H&E staining, this histological region of wild-type cluster 1 had necrotic tissue features; however, there were tumour cells present, but they were not distinguishable with histological staining in **Figure 3.4.5.2a.** Thus, these results suggest that MSI with the transcriptomics dataset is a promising solution to investigate metabolic profiling of the tissues where the histological annotation is limited. These findings also match with network analysis in **Figure 3.4.5.2.2c.** The top enriched pathway was glycerophospholipid metabolism, which could be enhanced by the direct and indirect roles of HIFs in regulating genes involved in this lipid metabolism[220]. In summary, the findings of network analysis based on the top hit genes and compounds in the whole metabolome and transcriptome information matched the glycerolipid metabolism in **Figure 3.4.5.2.2e.** This network confirmed the previous findings of the KEGG gene and MSEA that linoleic acid metabolism and HIF-1 signalling were the top enriched driver pathways in the wild-type PDX model cluster red.

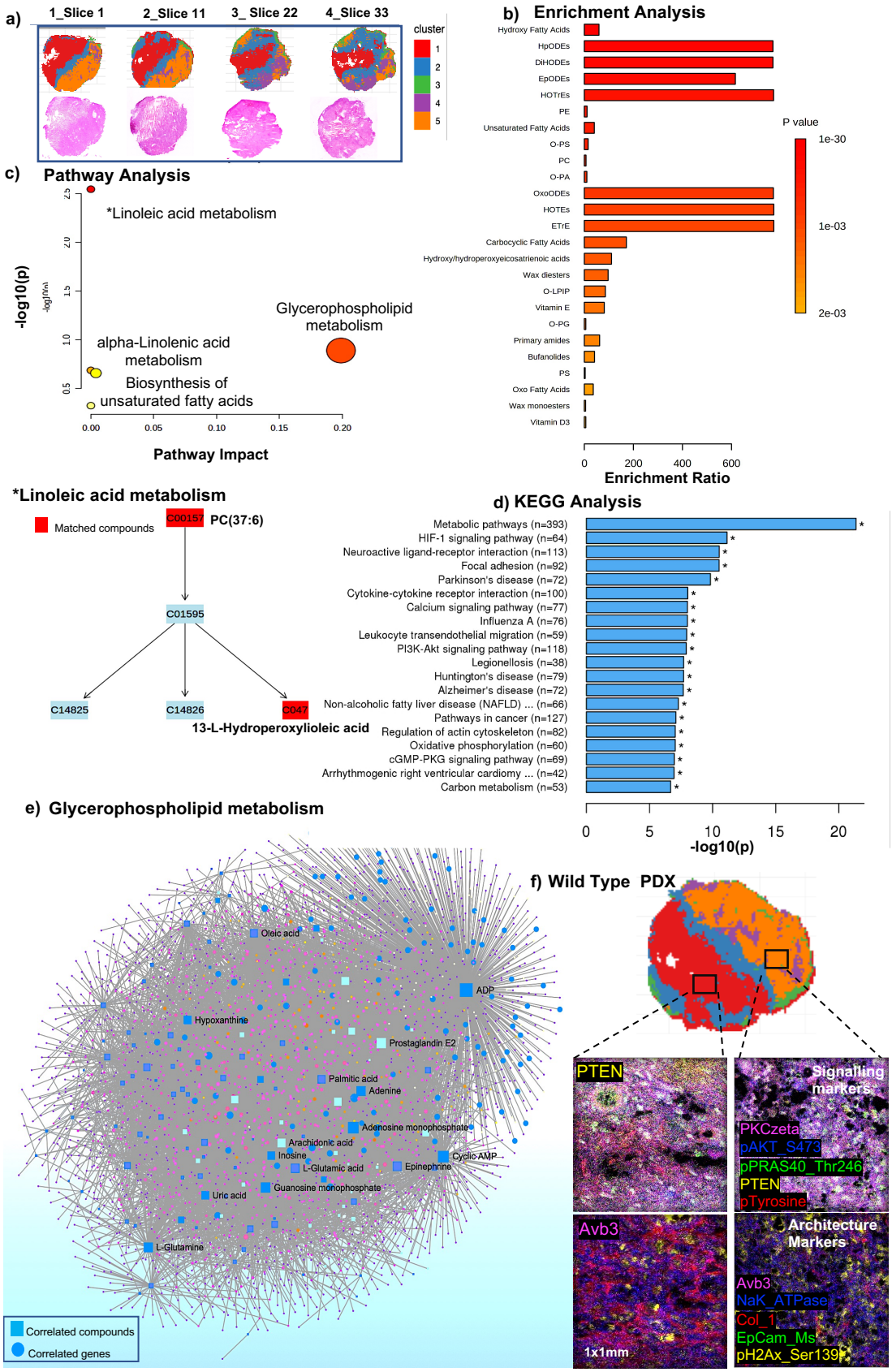


Figure 3.4.5.2. 2. Correlation of Transcriptomics and metabolomics behind the DESI- MSI cluster of Wild type PDX model(cluster red with orange). a)Wild type PDX model sample set b) Metabolite Enrichment Overview, Summary Plot for Over

Representation Analysis (ORA), The enriched metabolomic content, the significant lipid classes in colour grading red (high) to white (less) for the most significant (p-value< 0.05). c) The global view of the pathways in the interaction network to show graphical output contained the metabolome view with the related compounds, red highlighted KEGG labelled compounds matched tentatively with PC. Hyroperoxylioleic acid classes d) KEGG enrichment analysis to identify the pathways which had the most significant involvement with the upregulated and downregulated genes e) The plot of the selected interaction sub-network, Glycerophospholipid metabolism in the Network Explorer. f) The IMC images show the protein level differences of PTEN and Avb3 protein markers. All statistical analysis threshold, p-value < 0.05, [FDR] <0.05.

3.5 Discussion and Conclusions

The findings in this chapter expand the application area of DESI-MSI technology and its feasibility for a multi-modal data integration of breast tumours to reveal inter-and intra-tumour biochemical heterogeneity. Clustering with a serial segmentation analysis of breast cancer PDX models (WILDTYPE-WT and MUTANT-MT) containing the *PIK3CA* mutation reveals a degree of heterogeneity across samples. The presence of metabolically different clusters in the PDX models suggests that unique genetic changes may be driving other metabolic activity. The DESI-MSI clusters ROIs were then employed for transcriptomics investigation with laser capture microdissection. Transcriptomic findings demonstrated gene expression differences between wild-type and *PIK3CA* mutant and across metabolically separate sub-regions identified by DESI analysis within the same tumour. MSI and RNA-seq profiling were augmented with a CyTOF panel that included structural, immunological, and PI3K pathway markers to capture protein expression patterns. The key pathways underlying metabolically diverse clusters were identified via network analysis, and correlations between metabolic phenotypes and genotypes suggested that the integration methodology can better understand the underlying biochemical mechanism.

Emerging information on metabolic and lipid rewiring was fingerprinted using multi-modal spatial omics techniques to characterise molecular tumour patterns. A thorough assessment of breast cancer metabolite and genetic changes in wild-type and PIKC3CA mutant PDX model tissues revealed non-obvious regional heterogeneity. This multi-modal investigation proved the

complementarity of many analytical methods in the non-target approach, allowing us to achieve more robust results through the consistency of orthogonal measurements. Ceramides, hydroxy fatty acids, sphingolipid metabolism, linoleic acid metabolism, oxidative phosphorylation, HIF-1 signalling, and, most crucially, arachidonic acid metabolism were detected in both metabolomic and genomic data. Most of the observed changes aligned with the metabolic markers previously reported for breast cancer. The ensuing genetic mapping of the tumour ecosystem can provide unprecedented insight into altered cellular activities and interactions *in situ*, with the potential to find novel disease-specific targets and biomarkers.

The combination of DESI-MSI, spatial transcriptomics and IMC can identify tumour subpopulations and enrich the information gained by each technology separately. In this chapter, metabolite changes in specific clusters were discovered using DESI-MSI technology and clustering analysis, which were then linked to the genetic expression underlying these metabolic changes. In the future, a comprehensive molecular characterisation and deeper insights into cancer biological processes may be useful for a more precise diagnosis in clinics. DESI-MSI, on the other hand, may be used for the *de novo* discovery and description of ITH based on known clinically essential traits.

Based on DESI-MSI-guided omics data, this spatial multi-modal omics strategy should be helpful in patient stratification. The cost-effectiveness of lipid-based DESI-MSI analysis could allow it to take an active role in clinical-pathological settings than other spatial omics approaches. The rapid adoption of DESI-MSI spatial omics could improve routine clinical decision-making, resulting in better and more personalised patient care.

3.6 Appendix

Table 3.6.1. The heatmap of identified top 50 m/z values was compared in Hierarchical clustering in *PIK3CA* mutant and wild-type PDX models by DESI-MSI.

mz	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change
560.5887	2.99680587	-1.7980837	0.4223441	-1.7268449	-1.0941692
562.5857	2.98139674	-1.7290852	0.36446598	-1.6204302	-1.1624685
558.5917	2.95277596	-1.6512053	0.38314912	-1.5487819	-1.2778766
564.5828	2.84942596	-1.5410007	0.32406723	-1.4441976	-1.2710793
502.6304	2.84642178	-1.6810393	0.37784015	-1.5663607	-1.1003784
504.6274	2.82421401	-1.4614818	0.28837	-1.4815099	-1.2367508
500.6334	2.81210254	-1.5865174	0.36703485	-1.5226113	-1.1660132
444.672	2.74367917	-1.4182638	0.26984772	-1.353005	-1.266043
442.6751	2.71498048	-1.3791076	0.23766445	-1.3163914	-1.2604611
506.6245	2.69650108	-1.2829561	0.10913789	-1.2324155	-1.2499537
556.5947	2.69317195	-1.2888018	0.12226572	-1.2314699	-1.2563264
1030.2443	2.67032644	-1.1307495	-0.0883966	-1.1627441	-1.1889018
446.669	2.66967476	-1.2479037	0.04589041	-1.1925969	-1.2110977
768.4811	2.65172481	-0.4562216	-0.8213418	-1.0719074	-1.0085011
566.5804	2.65004727	-1.2224776	0.0194705	-1.1771254	-1.1928757
1320.0162	2.64775574	-1.2202714	-0.0881442	-1.1020037	-1.1365538
1026.2492	2.64636417	-1.2150857	0.00216398	-1.1681356	-1.1831719
498.6364	2.63298202	-1.1982232	-0.0228356	-1.1503328	-1.1688002
1028.247	2.62933043	-1.2431917	0.16456767	-1.2739032	-1.2281456
618.547	2.61317115	-1.2506464	0.06562892	-1.1828709	-1.169946
1322.0143	2.59884895	-1.1547429	0.00082508	-1.1796805	-1.1656785
794.4982	2.59790992	-0.4368388	-0.7699191	-1.3229631	-0.7992985
888.6655	2.59082097	-1.4846142	-0.6492404	-1.0768395	-0.2309979
688.4899	2.58947634	-0.8508703	-1.1444449	-0.5306545	-0.6973245
968.2937	2.57476751	-0.9738683	-0.2743427	-1.0920236	-1.0606781
324.7618	2.56015535	-1.3546428	0.22060263	-1.2977909	-1.0887564
861.6381	2.55797747	-2.1147681	-0.3720031	-0.9431381	-0.0709568
1032.2422	2.55138156	-1.1574956	-0.0224953	-1.1295477	-1.1223957
678.5018	2.54347709	-1.094131	-0.2891745	-1.0290287	-0.9560513
972.2904	2.54315908	-1.1206871	-0.158795	-1.0629666	-1.0492582
889.6684	2.53533308	-1.1135506	-0.7568734	-0.9995017	-0.4309984
620.5419	2.53053893	-1.2193502	0.13821985	-1.2138523	-1.1501585
576.5626	2.52610734	-1.5367114	0.35911597	-1.4995826	-0.8694468
970.2928	2.52530858	-0.9591287	-0.1615026	-1.1235807	-1.1118142
440.678	2.52421018	-1.0531561	-0.2409706	-1.0351352	-1.0166249
616.5499	2.51009913	-1.0240624	-0.2488448	-1.069412	-0.9860044
1335.9898	2.49477366	-0.9707809	-0.3254688	-1.0444519	-0.9491954
578.5597	2.49376375	-1.5100984	0.3344272	-1.4381853	-0.8783961
1044.22	2.4902316	-0.9734381	-0.3252713	-1.0192211	-0.9634089
448.6661	2.48139853	-0.8808226	-0.5679018	-0.8981013	-0.8660648
1333.9946	2.47927986	-0.9705641	-0.4506011	-0.9232753	-0.8952493
1024.2532	2.47423942	-1.0919236	-0.1549332	-1.0149574	-1.0355784
330.7533	2.47068177	-1.2502181	0.08371068	-1.1616403	-1.0341816
574.5656	2.47005399	-1.4027447	0.24204838	-1.3546943	-0.9112466
272.7949	2.46248397	-1.8669474	0.28442298	-1.7903594	-0.1695348
274.7919	2.46115249	-0.9464303	-0.4213178	-0.9492592	-0.9042069
1046.2178	2.46114276	-0.948174	-0.388951	-0.968112	-0.9225954
580.5565	2.44142674	-1.2828151	0.14299468	-1.2769017	-0.9365496
662.4713	2.44035109	-0.8960597	-0.5893823	-1.0386971	-0.6566

Table 3.6.2. The heatmap of identified top 50 m/z values was compared in Hierarchical clustering in *PIK3CA* mutant

mz	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change
964.7687	2.29659468	-0.9417731	-0.8279769	0.25752553	-0.7843702
965.7739	2.28788474	-0.6248968	-0.6657083	-0.2497029	-0.7475767
966.7779	2.26732067	-0.8367571	-0.8003432	0.12924813	-0.7594686
938.7498	2.26382934	-0.7018965	-0.696665	-0.1792956	-0.6859723
990.7809	2.26139009	-0.720001	-0.7091441	-0.0970989	-0.7351461
962.7524	2.2594403	-0.7896745	-0.6644764	-0.0614843	-0.7438051
989.7739	2.2122517	-0.7623499	-0.686764	-0.0448664	-0.7182714
956.7561	2.18952978	-0.6579493	-0.6594804	-0.0680517	-0.8040483
1016.7966	2.16545783	-0.7261686	-0.6838545	-0.071741	-0.6836937
992.7914	2.1492896	-0.5773605	-0.5823094	-0.3429547	-0.6466651
1032.7919	2.14527025	-0.8632558	-0.770008	0.13924237	-0.6512488
967.7835	2.1375228	-0.5358891	-0.5726019	-0.3644329	-0.6645989
987.7576	2.12610821	-0.7647466	-0.6268815	-0.0255373	-0.7089428
1034.8059	2.11807236	-0.5678951	-0.5931565	-0.3192475	-0.6377733
1015.7891	2.081151	-0.7331381	-0.648386	-0.0141154	-0.6855115
939.7561	2.06742728	-0.532224	-0.5396464	-0.3737865	-0.6217703
953.7388	2.0459951	-1.0758163	-0.6639106	0.45403129	-0.7602995
977.7388	2.04238787	-0.6223941	-0.5505996	-0.2067158	-0.6626784
961.7424	2.04102971	-0.5782806	-0.5413298	-0.3020326	-0.6193867
1033.7976	2.03454521	-0.5616011	-0.5439336	-0.3401768	-0.5888338
1011.7894	2.03340758	-1.0337334	-0.8215497	0.40843923	-0.5865637
1010.8013	2.03199855	-1.0593996	-0.9391287	0.52571356	-0.5591838
976.7343	2.01129505	-0.9820793	-0.6898078	0.33730218	-0.6767102
985.7496	2.00475271	-0.8098492	-0.6104549	0.09160875	-0.6760574
1009.7992	1.99567041	-1.2200508	-0.9542021	0.61247067	-0.4338882
960.7372	1.99516181	-1.0534017	-0.758046	0.29835839	-0.4820725
963.7587	1.99221058	-0.4914626	-0.4897652	-0.4358421	-0.5751407
1003.7532	1.96720798	-0.6223587	-0.5461824	-0.1635066	-0.6351603
1030.7775	1.95263179	-1.0222751	-0.685513	0.34905883	-0.5939025
979.7551	1.9474233	-1.252731	-0.7322464	0.5788578	-0.5413037
1007.7849	1.93779108	-1.2552887	-0.8879046	0.61071405	-0.4053118
955.7542	1.93693389	-1.1834557	-0.7950404	0.61968934	-0.5781272
1031.782	1.93336113	-0.5200688	-0.5165528	-0.3015697	-0.5951699
1029.7672	1.91656313	-0.7049563	-0.5407519	-0.0521311	-0.6187239
1041.8042	1.88395178	-0.6196128	-0.5821376	-0.1210153	-0.5611861
988.7687	1.88389146	-1.2632916	-0.8088477	0.54961242	-0.3613646
1039.7888	1.88102109	-0.7375943	-0.6369873	0.00830847	-0.514748
1013.7751	1.87782188	-1.1523095	-0.6814071	0.45687006	-0.5009754
991.7875	1.86416671	-0.4115613	-0.4627907	-0.4458479	-0.5439669
1042.809	1.85172561	-0.5262511	-0.5239722	-0.2342555	-0.5672468
983.7844	1.84841598	-1.2321194	-0.9138168	0.66651035	-0.3689901
968.751	1.82996983	-0.7509355	-0.3697956	-0.0338563	-0.6753824
1005.7696	1.81721936	-1.3554167	-0.8097171	0.53813814	-0.1902237
950.7158	1.8153096	-0.5275144	-0.4382818	-0.2686488	-0.5808646
1002.749	1.79763499	-1.1309762	-0.6076664	0.46447335	-0.5234657
926.6987	1.78184946	-0.6857813	-0.5303372	-0.1162877	-0.4494432
1056.789	1.76176216	-0.5186513	-0.4596971	-0.24093	-0.5424838
1000.7388	1.74914221	-0.9259533	-0.4845604	0.18589426	-0.5245228
1017.8031	1.73867238	-0.3953649	-0.4391716	-0.3973875	-0.5067483

Table 3.6.3. The heatmap of identified top 50 m/z values was compared in Hierarchical clustering in wild-type PDX models by DESI-MSI

mz	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change
270.7981	2.19784402	0.37081994	-0.7572624	-0.8885652	-0.9228364
268.8014	2.16879016	0.82207855	-0.7552765	-1.0973283	-1.138264
266.8044	2.15361284	0.48376949	-0.7118365	-0.9390968	-0.9864491
322.856	2.10426128	0.47378159	-0.7833015	-0.8794246	-0.9153167
320.859	2.06259217	0.27310314	-0.7159351	-0.8051242	-0.814636
257.1081	1.89483404	0.24406893	-0.2387248	-0.9372086	-0.9629696
210.8426	1.78741539	-0.019133	-0.5754729	-0.590145	-0.6026646
324.853	1.74977779	-0.1318465	-0.5338458	-0.5403779	-0.5437075
244.0269	1.74482641	0.72078525	-0.9322818	-0.697586	-0.8357438
284.7751	1.70343707	0.97215831	-0.9297102	-0.7750181	-0.9708671
276.0842	1.69144568	0.55252179	-0.7249464	-0.8159184	-0.7031026
286.7722	1.68329527	0.67312429	-0.7897206	-0.7178489	-0.84885
229.0336	1.67667669	0.36126952	-0.5960494	-0.7292589	-0.7126379
357.8387	1.65545786	0.06861648	-0.4692211	-0.6257099	-0.6291433
588.4752	1.60628109	0.64979276	-0.7113991	-0.8613608	-0.6833139
266.0757	1.60308512	0.7693583	-0.9650811	-0.5912063	-0.8161561
594.5287	1.56236771	-0.1618882	-0.4515265	-0.4801793	-0.4687738
282.7781	1.55977854	0.56985015	-0.7176616	-0.6317167	-0.7802504
338.8299	1.55275828	0.53660741	-0.7430039	-0.6086	-0.7377618
208.8456	1.5502446	-0.1314897	-0.4632202	-0.4748119	-0.4807228
355.8389	1.54636688	0.00822661	-0.3827059	-0.5826233	-0.5892643
176.0095	1.50863222	0.16847617	-0.5437211	-0.5860314	-0.5473559
203.0426	1.49907101	0.57445146	-1.2008542	-0.2061679	-0.6665004
644.5979	1.46574366	0.06672998	-0.5594185	-0.4324277	-0.5406274
564.5357	1.43195952	0.43357742	-0.67038	-0.6205832	-0.5745737
174.0029	1.410795	0.67559037	-0.4295864	-0.7443882	-0.9124108
590.4799	1.39592662	0.10536573	-0.4808998	-0.5202407	-0.5001519
272.7952	1.38453872	-0.2108376	-0.3880803	-0.3928104	-0.3928104
226.0701	1.37609908	0.78730731	-1.0962378	-0.3169825	-0.7501861
602.5143	1.36579814	0.60631153	-0.7710944	-0.6561561	-0.5448591
601.5155	1.33518837	0.52377514	-0.7087558	-0.6075712	-0.5426366
231.0622	1.3335941	0.66506633	-0.3568435	-0.8084135	-0.8334034
336.8327	1.32087321	0.2184356	-0.5285795	-0.4750835	-0.5356458
656.5727	1.30529103	0.62815506	-0.8844911	-0.5000106	-0.5489444
237.024	1.30315423	0.93662077	-1.0629994	-0.3371098	-0.8396658
238.0199	1.29749839	0.89711553	-0.881121	-0.4560698	-0.8574232
510.4161	1.25875918	0.29490521	-0.6498818	-0.5246798	-0.3791027
647.6194	1.25392062	0.35107113	-0.7981458	-0.2677692	-0.5390767
538.448	1.25211292	0.50411201	-0.9124358	-0.4874902	-0.3562989
570.467	1.22597533	0.40925245	-0.7216048	-0.4008104	-0.5128127
301.1281	1.22457317	0.53970597	-0.6092724	-0.6151356	-0.5398711
681.5791	1.20633848	0.81224375	-1.0366583	-0.3712564	-0.6106675
553.4696	1.1882075	0.28522556	-0.5294044	-0.6670001	-0.2770285
260.0564	1.18681448	0.71271887	-1.1053642	-0.309267	-0.4849022
649.6338	1.18640788	0.06867372	-0.544812	-0.3184843	-0.3917853
554.4814	1.182851	0.18442104	-0.3923124	-0.5324862	-0.4424735
575.4868	1.1595034	0.69130194	-1.0148182	-0.4714763	-0.3645108
589.478	1.12372058	-0.0044491	-0.3724044	-0.3767963	-0.3700708
658.5776	1.12354596	0.11894849	-0.4375176	-0.4005928	-0.404384

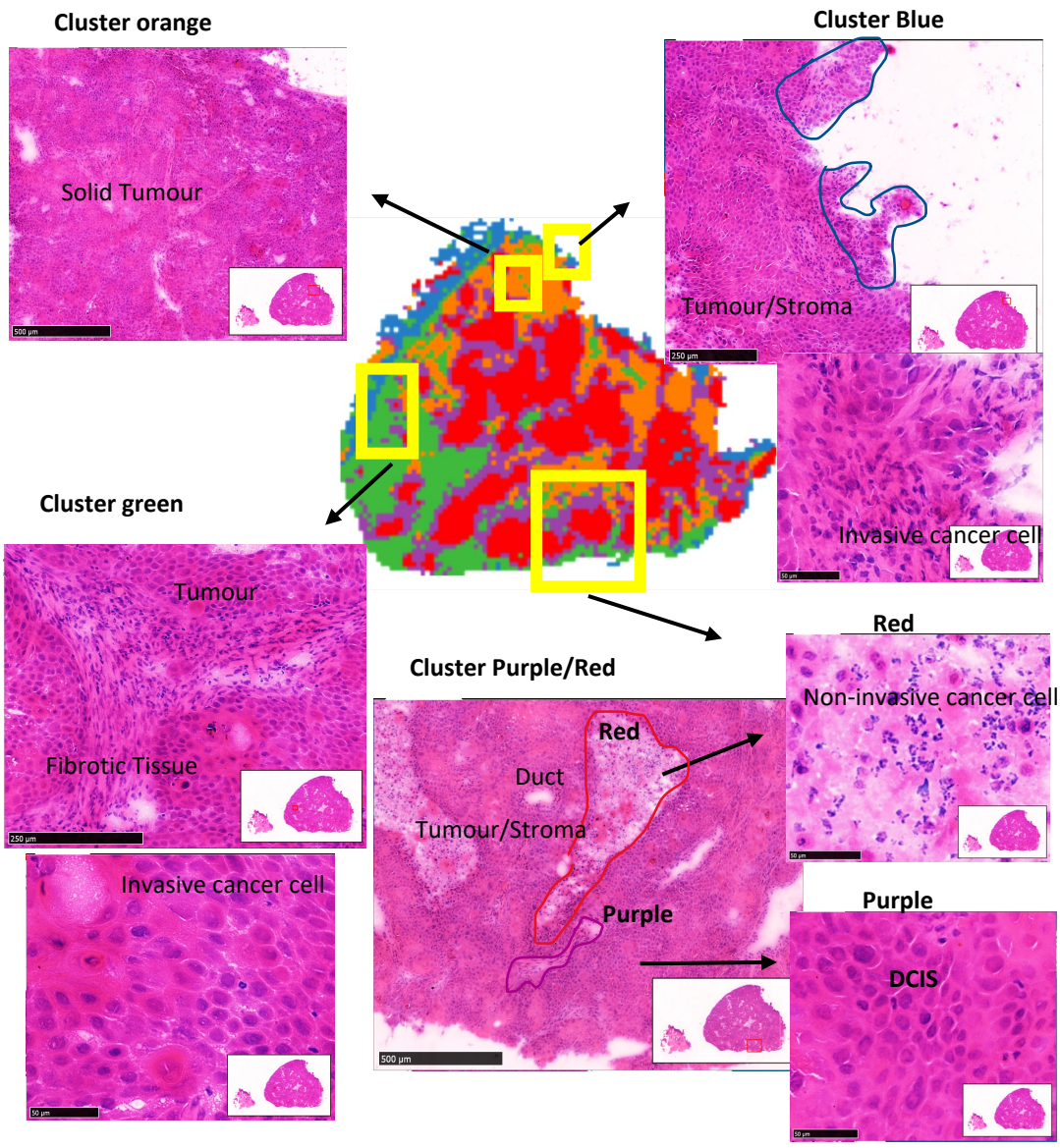


Figure 3.6.1. Detailed H&E annotation of the clusters in. the PIK3CA mutation PDX model.

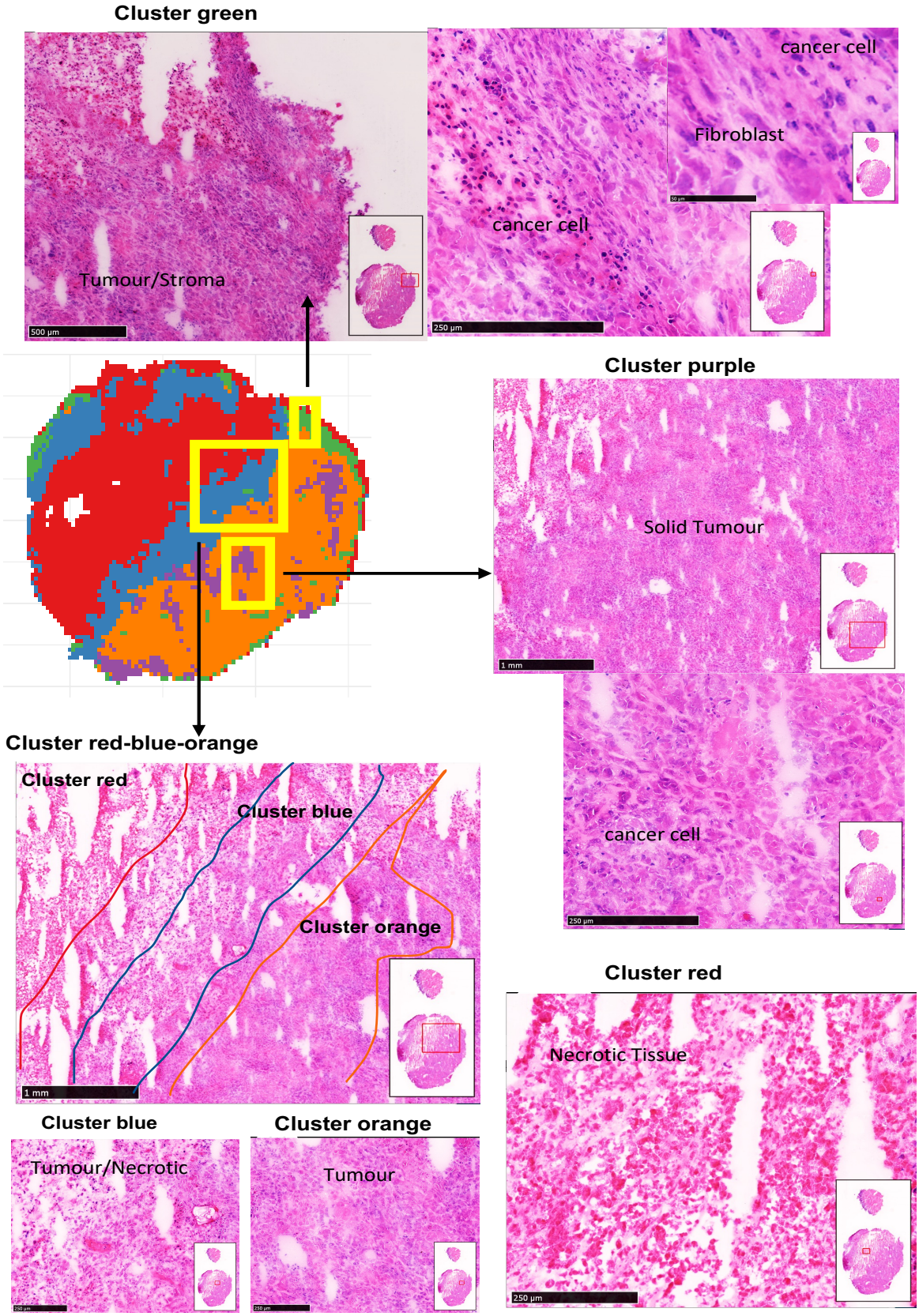


Figure 3.6.2. Detailed H&E annotation of the clusters in. the wild-type PDX model.

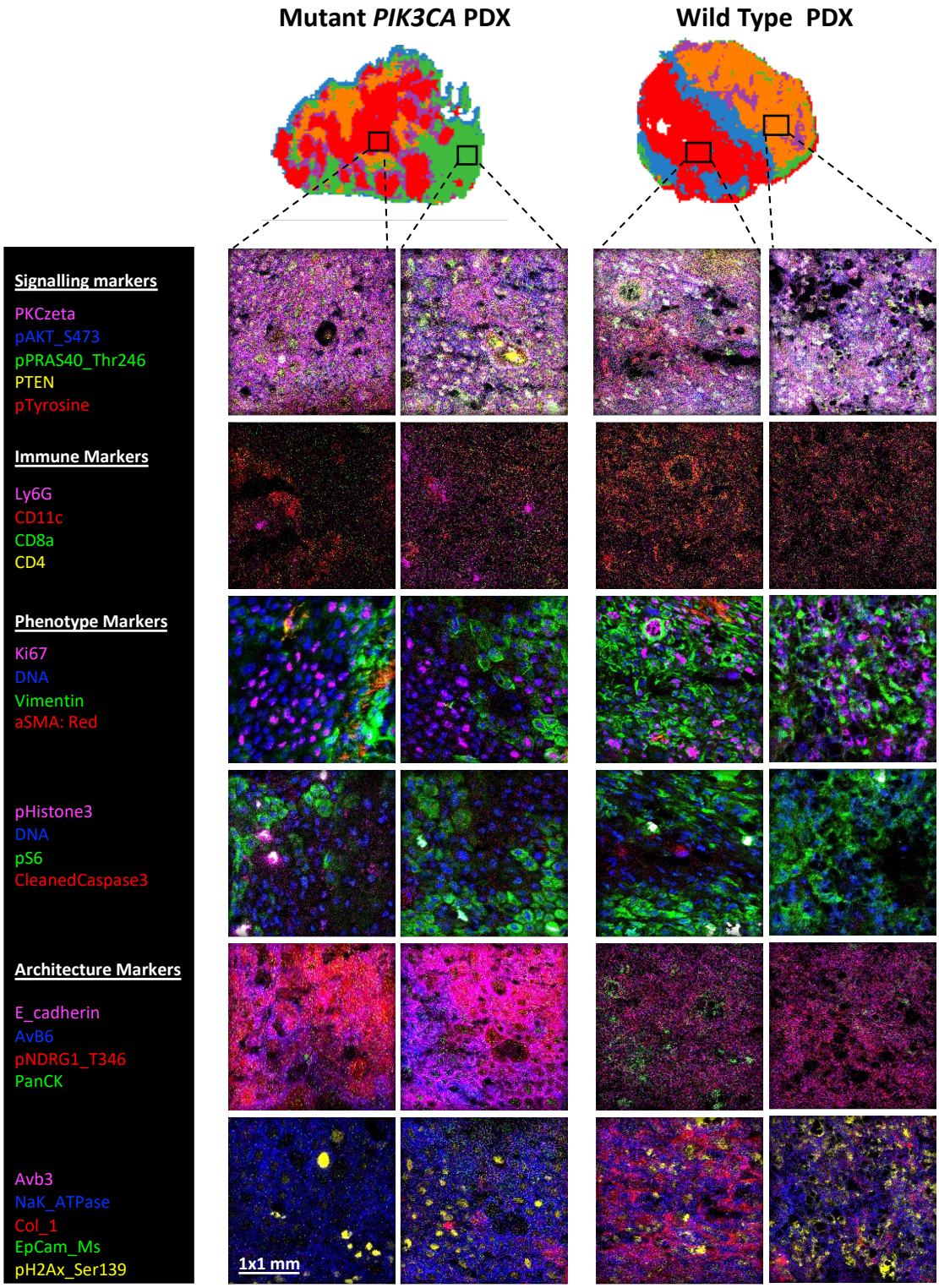


Figure 3.4.5.2. 3. IMC of the all-antibody panel of 25 proteins. Comprising immunological, phenotypic, signalling, and architectural proteins, was created to differentiate cell types and states for selected ROIs from *PIK3CA* mutant and wild-type PDX models (Fluidigm Hyperion Imaging System v1.0.560.2 and MCD Viewer).

4 A novel embedding strategy for investigating tumour heterogeneity in breast tumour core biopsies

4.1 Background

DESI-MSI technology has already proved to be a powerful technique for evaluating tumour heterogeneity because it can disclose detailed information on the spatial distribution of metabolites, therapeutic molecules, and lipids in large tissue slices[35,197,221]. Although the investigation of tumour heterogeneity on small-size core biopsies can be laborious, a tailored embedding media, in which samples could be placed in a block as a Tissue Micro Array (TMA), allows for the simultaneous sectioning and DESI-MSI analysis of numerous clinical samples[135]. Various embedding strategies have been reported in the literature to improve the sample preparation for diverse imaging techniques, especially for preparing small or fragile samples and whole animals[135]. These strategies are commonly based on embedding in Optimal Cutting Temperature medium (OCT)[118,222], sodium-carboxy-methyl cellulose (Na-CMC)[137], gelatin[223], agarose[224] or poly(N-(2-hydroxypropyl)-methacrylamide) (HPMA)[225]. The embedding media should not alter or thaw the samples to preserve the tissue morphology for histological evaluation.

Additionally, the embedding should not alter the distribution of analytes within the tissue to allow accurate MSI results. Most importantly, embedding media should neither affect DNA or RNA extraction and purity nor the efficient amplification of molecules of interest where MSI is paired with molecular biology applications, such as polymerase chain reaction (PCR) or reverse transcription PCR (RT-PCR). Consequently, the extracted RNA and DNA from the implanted biopsies must meet optimal quality and quantity standards as discussed in the previous chapter 3, section 3.4.4. Comparison of different polymers and the freezing procedures for tissue embedding is crucial in achieving the most reliable possible MSI results. Since none of the embedding media assessed in the current literature[226–231] congregates the above characteristics. To accomplice that, a novel universal embedding hydrogel, a matrix of hydroxypropyl methylcellulose (HPMC) and polyvinylpyrrolidone

(PVP), was assessed to construct a robust tissue embedding protocol for MSI analysis and follow-up analyses such as RT-PCR and was published in 2020, in Analytical Chemistry[174].

This newly developed hydrogel media was used in large biopsies but not in core biopsies, specifically in human breast tissues with a high-fat concentration, making them more delicate. So, it was needed to assess whether this hydrogel could be used with such fragile fresh frozen core biopsies for DESI-MSI and follow-up high enough standard RNA material extraction for transcriptomics analysis. After this first step evaluation, the next step was to examine tumour heterogeneity with DESI-MS clustering analysis and then continue a correlation analysis with a representative biopsy. The Network Explorer module in the MetaboAnalyst software was used for the correlation analysis to reveal any associations between DESI-MSI metabolomics and transcriptomics datasets. This integration, particularly in the specific sub-heterogeneous region of the tumour, could lead to the identification of key signalling pathways characterised by the selected DESI-MSI cluster.

4.2 Aim and Objectives

The aim of this chapter is to investigate the applicability of the hydrogel embedding media using small-sized (i.e. 3 mm length) core biopsies collected from breast cancer patients.

- To determine the degree of intra- and inter-tumour heterogeneity by performing DESI-MSI clustering analysis embedded into the HPMC/PVP hydrogel core biopsies.
- To evaluate hydrogel media compliance for downstream gene expression analysis such as rt-PCR and RNA sequencing by checking the quality and quantity requirements of isolated DNA and RNA material from the embedded biopsies.
- To investigate the potential correlation with the DESI-MSI metabolomics and transcriptomics datasets by applying network analysis to discover any key signalling pathways in sub-heterogeneous regions defined by DESI-MSI clustering.

4.3 Results and Discussion

4.3.1 DESI-MSI Clustering Analysis on embedded fragile patient biopsies

The primary work of this chapter was focused on the embedded core biopsies evaluation in the new hydrogel media. 23 fresh-frozen patient core biopsy samples diagnosed with breast cancer, provided through the ICL Tissue Bank, were collected. After their embedding, the samples were cryosectioned at 10 um thickness and mounted on standard glass and PEN membrane glass slides. As a first step, DESI imaging was performed in negative ion mode at 1s/s scan rate and 100 µm pixel size**, Figure 4.3.1. 1a.**The MSI data were then processed by segmentation analysis utilizing the ion pair Spearman's correlation and hierarchical clustering (average linkage) described in section 2.13.1. Clustering analysis was performed to distinguish segments in the tissue section that represented metabolically distinct areas. This clustering has produced 6 different metabolic groupings in **Figure 4.3.1. 1c-d.** The similarity distance of this metabolic grouping in the cluster analysis between samples was displayed in a dendrogram (distance measured using Pearson and clustering technique using average). A PC1 analysis was used to depict the spatial distribution of strongly co-localised ion groups across the samples in **Figure 4.3.1. 1e.** These clusters were assumed to reflect local metabolic interactions and thus be conceivably related to the behaviour of regional subgroups of malignant cells. The DESI-MSI data have indicated that the metabolic variability across patients occurs between patient samples inter-tumour and within the same biopsy intra tumour. The post-DESI tissue segment was further used for the histological assessment by H&E staining in **Figure 4.3.1. 1b.**

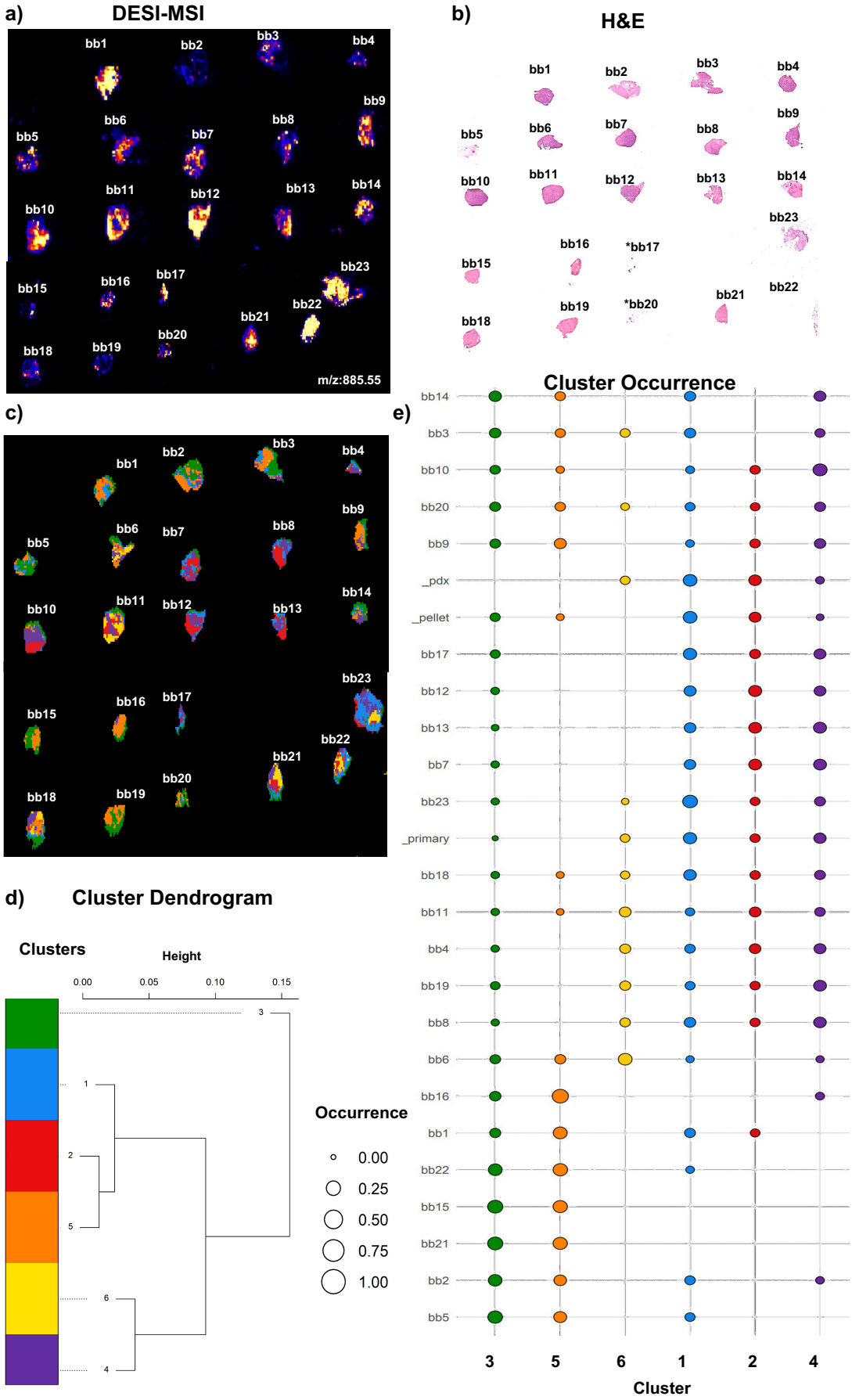

a)
DESI-MSI
bb1
bb2
bb3
bb4
bb5
bb6
bb7
bb8
bb9
bb10
bb11
bb12
bb13
bb14
bb15
bb16
bb17
bb18
bb19
bb20
bb21
bb22
bb23
m/z:885.55
b)
H&E
bb1
bb2
bb3
bb4
bb5
bb6
bb7
bb8
bb9
bb10
bb11
bb12
bb13
bb14
bb15
bb16
*bb17
bb18
bb19
*bb20
bb21
bb22
bb23
c)
d)
Cluster Dendrogram
Clusters
Height
0.00
0.05
0.10
0.15
3
1
2
5
6
4
Occurrence
0.00
0.25
0.50
0.75
1.00
e)
Cluster Occurrence
bb14
bb3
bb10
bb20
bb9
_pdx
_pellet
bb17
bb12
bb13
bb7
bb23
_primary
bb18
bb11
bb4
bb19
bb8
bb6
bb16
bb1
bb22
bb15
bb21
bb2
bb5
3
5
6
1
2
4
Cluster

Figure 4.3.1. 1. DESI-MSI Clustering Analysis on embedded breast tumour core biopsies. a) Viewing ion image (*m/z*: 855.55) of the core biopsies embedded in a Hydrogel TMA in HDI software (Waters) for DESI-MSI analysis b) H&E staining of the embedded biopsies tissue section after DESI-MSI (*staining artefacts). c) The PC1 scores, calculated using the ions belonging to each cluster, were used as the representative spatial distribution of the groups of highly colocalized ions. d) clusters dendrogram showed the similarity of clusters e) 6 different metabolic clusters between 23 biopsies, highly colocalized ions were identified using the ion pair Pearson correlation and hierarchical clustering (average linkage) and bubble plot displayed the occurrence rate of clusters in each sample.

Concludingly, the findings presented in this study have indicated that the hydrogel embedding media enabled the accessibility of high-quality DESI-MSI clustering analysis of the multiple breast tissue core biopsies in a single stroke. While it allowed the quantification of the abundance of identified metabolites in clustering analysis, the ion vectors of tumour mass spectra with unconnected sub-networks provided a robust visualization tool for molecular analysis of tumour heterogeneity even when exploiting a small, 1-2 mm^2 size biopsy.

4.3.2 The Impact of the Embedding Hydrogel on DNA and RNA Quality

Linking information of identified metabolic features with gene expression could provide valuable information to underly altered signalling pathways. For the sample preparation of gene expression analysis steps, the RNA in **Figure 4.3.1.2a-b** and DNA in **Figure 4.3.1.2d-e**, isolated from embedded mouse kidney sections, were compared. For comparison, those embedded sections in OCT, gelatin, and non-embedded as controls were assessed the compatibility of the newly developed embedding media with molecular genetics experiments. The purity of the extracted DNA was determined using the absorbance quotient A260/280 in **Figure 4.3.1.2d,** and the extracted RNA was determined by the RIN number in **Figure 4.3.1.2a,** calculated using the 2100 Agilent Bioanalyzer. The RNA extracted from the sample embedded in the HPMC/PVP hydrogel media was similar in purity and quality to the RNA obtained from the unembedded controls. Unlike the gelatin-embedded sample, the RNA from the OCT-embedded specimen did not differ significantly

from the control in **Figure 4.3.1.2a.** However, one OCT and two gelatin samples fell below the in-house quality criteria of a RIN of 5, indicating RNA degradation[232,233].

RT-PCR experiments utilized the housekeeping genes glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) and peptidyl-prolyl cis/trans isomerase B (*PPIB*) as targets' ability to transcribe RNA into cDNA and amplify the transcribed DNA in

Figure 4.3.1. 2b. These housekeeping genes were chosen since they have roughly comparable abundance and distribution across the kidney, minimizing the effects of biological variability in total expressed RNA within the kidney specimen in **Figure 4.3.1.2c.** The extracted genomic DNA was found to meet the requirements from all embedding media as measured by the absorption ratio at A260/280 in Figure 4.3.1.2d. Compared to the unembedded control, samples embedded in the novel HPMC+PVP medium and gelatin showed no interference with amplification, while OCT revealed considerably reduced relative quantification values in **Figure 4.3.1.2b.** All embedding protocols yielded equivalent amounts of DNA on average, while the unembedded control and gelatin-embedded samples yielded lower amounts in one replicate in **Figure 4.3.1.2e.** This variability could be related to the reduced repeatability of sectioning for these embedding settings, similar to the tissue loss seen in the histological evaluation with RNA-ISH in **Figure 4.3.1.2c.**

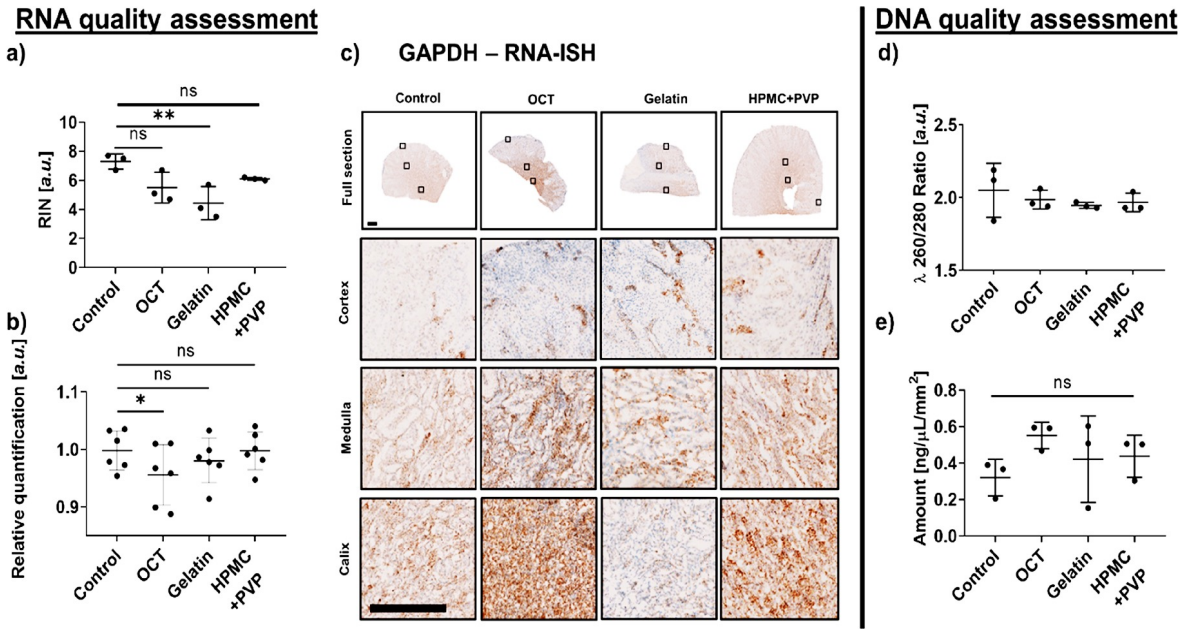


Figure 4.3.1. 2. Effects of the embedding media on DNA and RNA extraction. a) RNA integrity for unembedded controls and tissues embedded in OCT, gelatin, or HPMC+PVP, respectively (n = 3 technical replicates/condition). b) Relative quantification of the housekeeping gene GAPDH for the different embedding conditions via rt-PCR using *PPIB* expression as a reference gene. (n = 6 with two analytical replicates/technical replicate). c) RNA-ISH for *GAPDH* showing the distribution of expressed mRNA (brown) against an unembedded control and embedded tissue sections. The nuclei were counter-stained with Hematoxylin (blue). The squares in the full section indicate the regions chosen for the zoomed panels. The scale bar for the full section view is 1 mm, and the scale bar in the zoomed panels marks 250 µm. d) Purity of the extracted DNA for all embedding conditions (n=3 per condition). e) Amount of DNA determined by Qubit normalized to the mean area of the tissue section (n=3 per condition). The bars indicate the mean ± S.D. for all replicates. p-values: * = $p < 0.05$; ** = $p < 0.01$, *** = $p < 0.001$.

Figure 4.3.1. 3 illustrated for all replicates the absorbance quotient A260/280 in **Figure 4.3.1.3a,** the extracted amount in **Figure 4.3.1.3b,** and the electropherograms used to calculate the RIN in **Figure 4.3.1.3c.** The total isolated RNA have a similar average amount except for HPMC/PVP in **Figure 4.3.1.3b.** This variability could be related to the reduced repeatability of sectioning for these embedding settings, similar to the tissue loss seen in the histological evaluation with RNA-ISH in **Figure 4.3.1.2c.**

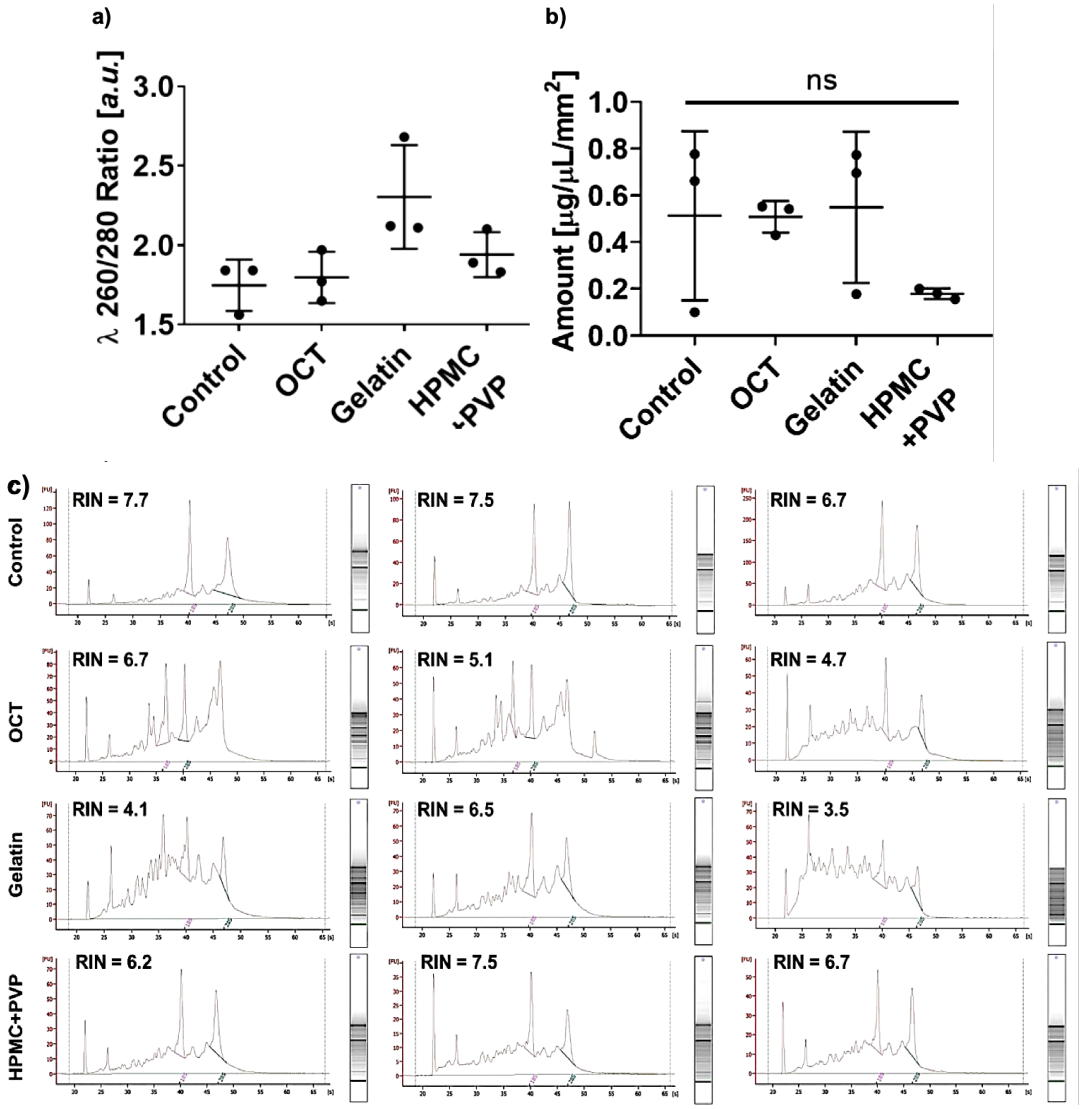


Figure 4.3.1. 3. Analysis of RNA integrity. a) The purity of extracted RNA from all embedding conditions (n = 3).b) Extracted RNA amount normalised to tissue section mean area (n = 3 per condition). c) Purity of the extracted RNA for all embedding conditions (n=3 per condition). d) shows the electropherograms for each replicate analyzed on the 2100 Bioanalyzer and the computed RIN based on the integrity of the extracted RNA. The bars in b and c indicate the mean ± S.D. for all replicates. p-values: * = $p < 0.05$; ** = $p < 0.01$, *** = $p < 0.001$.

The extracted DNA has shown high quality, with no evidence of DNA degradation (leading to substantial fragmentation) in all of the conditions studied in **Figure 4.3.1. 4**.

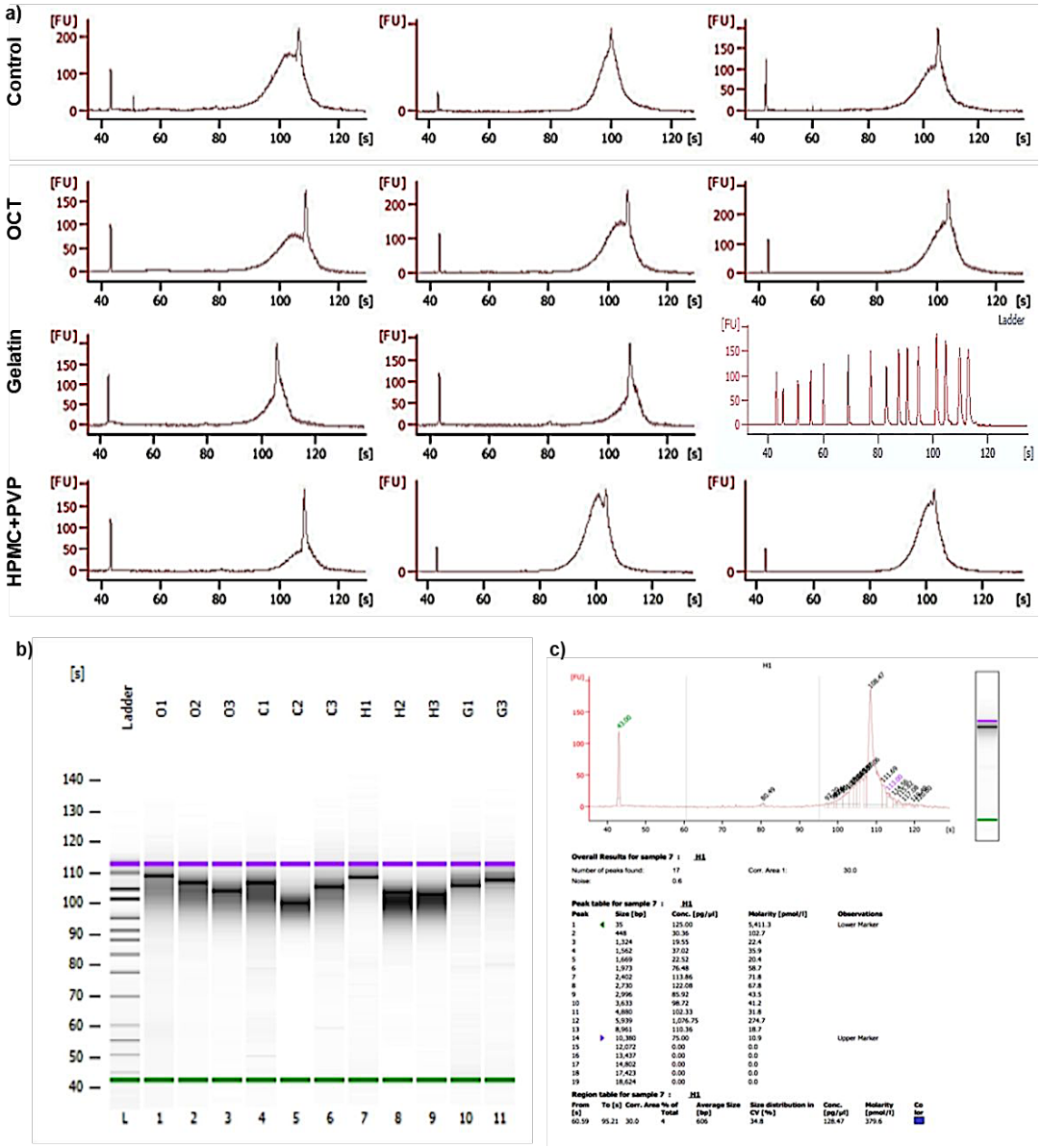


Figure 4.3.1. 4. Analysis of DNA integrity. a) Electropherograms for DNA extracts of all replicates of OCT, gelatin, and HPMC+PVP embedded kidney specimens compared to unembedded controls. The data was quantitatively and qualitatively assessed for DNA concentration, degradation/fragmentation, and fragment sizing, as seen in the pseudo-gel image b). The isolated DNA size (bp) ranges between the standard upper and lower marker. The size of each isolated DNA was calculated based on the 16 fragmented peaks of the ladder. i.e. c) An electropherogram of sample H1 has 23 DNA peaks, and the fragment sizes range from 35bp to 10.380bp.

4.3.3 Transcriptomics Analysis of the HPMC/PVP hydrogel embedded core biopsy

DESI-MSI acquisition has made it possible to identify sub-regions in core tumour tissue with metabolic segments and their underlying biological makeup, interactions, and the likely rationale for their heterogeneous

structure. First, one of the representative core biopsy utilized for RNA isolation and then transcriptomics was performed by a commercial entity using the Illumina platform RNA sequencing technology. One sample, bb7, was randomly chosen to be used in the optimization phase for the RNA sequencing procedure to see if the extracted RNA matches the ideal requirements in **Figure 4.3.3. 1a.** In the previous section,

Figure 4.3.1. 2b, animal kidney tissues were utilized to test the RT-PCR. Additionally, the results published in a part of the HMPC/PVP hydrogel embedding methodology[174,234], the data published before have never investigated the embedded core biopsies. The present evaluation takes under consideration the ideal RNA quality (RIN), the adequate isolated amount (> 10 ng RNA in total eluted 14 µL volume), and the high mapping percentage (> 80%) of the human genome's transcriptome data. The total amount of retrieved RNA was 20 ng, extracted in a final volume of 10 µL from a 20 μm^2 dissected area, while the RIN quality was 7.5.

The RNA sequencing data robustness is ensured when the exonic quality level (Q30) is 80%. Excitingly, the percentage of exonic readings mapped to the reference genome, in this case, was 96.06% in **Figure 4.3.3. 1b.** The complete transcriptome of sample bb7 was depicted with a heat map using FPKM cluster analysis and the log10 (FPKM+1) value in the dark blue colour rectangle in **Figure 4.3.3. 1c**. This clustering analysis revealed the expression patterns of up and down-regulated genes. Genes with high expression levels were highlighted in red, whereas genes with low expression levels were highlighted in blue. The log10 (FPKM+1) value was shown on a red to blue scale from high to low.

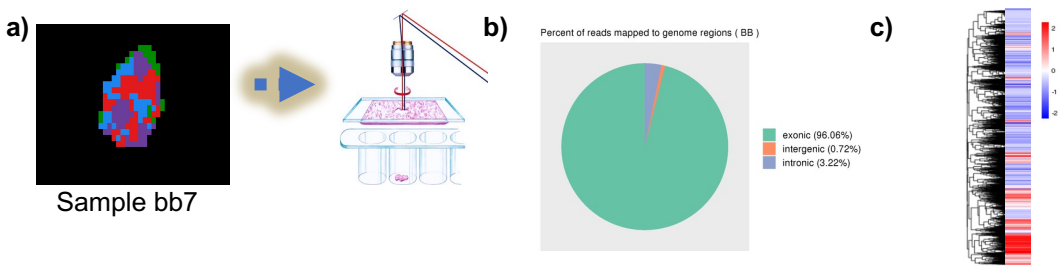


Figure 4.3.3. 1. Transcriptome analysis of microdissected Metabolic Segment.

a) microdissection of the clustered blue from sample bb7 b) Read Classification Based on Mapped Region c) Heatmap visualization of expressed genes in the overall FPKM cluster analysis, using the log10 (FPKM+1) value.

Interestingly, this pilot study found a complete transcriptome data set from the microdissected area using the HPMC/PVP embedded patient breast core biopsy possible. These findings would help to acquire the RNA expression profile behind the metabolic segments found by DESI-MSI clustering analysis.

4.3.4 Correlation of Transcriptomics and metabolomics behind the DESI- MSI cluster of embedded core biopsy

The purpose of this study was to investigate any possible correlation between transcriptomics and Metabolite Set Enrichment Analysis (MSEA) on a small-size core biopsy. For MSEA, Pathway and Network analysis modules in MetaboAnalyst software were performed to interact with various omics datasets on the MSI cluster region of breast cancer core biopsy. First, the DESI-MSI features were identified and classified using enrichment and pathway analysis. The expressed genes in transcriptomics from the sample bb7 were then connected in network analysis to find potential driving pathways. Therefore, it was reasonable to conclude that these enriched pathways contribute to tissue heterogeneity's transcriptome and metabolome activity.

4.3.4.1 Metabolite set enrichment analysis

MSEA was already mentioned in the previous chapter that discovered biologically important patterns in metabolomic data to provide a global picture of changed biochemical processes and identify substantially altering compounds. In the software analysis, this study assessed individual metabolites for their significance in understudy scenarios. Based on the programme parameters, the compounds that had attained a high enough degree of relevance were combined to determine if any relevant patterns emerged. PI is the most significant lipid class (p-value < 0.05, [FDR] < 0.05)

in the sample bb7 based on this enriched metabolomic dataset from DESI-MSI. Other lipid classes included HexCer, PI-Cer, O- PS, PE, and others (

Figure 4.3.4.3. 1a and

Table 4.3.4.1. *1*).

Table 4.3.4.1. 1. The most enriched lipid of hydrogel embedded sample bb7. The 'total' refers to the total number of compounds in the lipid classes; the 'hits' refer to the number of matched combinations from the uploaded data; the 'raw p' represents the original p-value calculated from the enrichment analysis; the 'holm p' was the p-value adjusted using the Holm-Bonferroni method, and the 'FDR p' was the p-value adjusted using False Discovery Rate.

Lipid Classes	total	expected	hits	Raw p	Holm p	FDR
PI	1960	17.4	145	2.58E-84	1.89E-81	1.89E-81
HexCer	8690	76.8	275	8.29E-75	6.05E-72	3.03E-72
PI-Cer	2290	20.2	114	6.53E-49	4.76E-46	1.59E-46
O-PS	984	8.69	77	6.28E-47	4.57E-44	9.81E-45
PE	1970	17.4	104	6.71E-47	4.88E-44	9.81E-45
LacCer	4580	40.5	146	1.49E-39	1.08E-36	1.81E-37
HETE	60	0.53	27	1.93E-39	1.40E-36	2.02E-37
O-PA	965	8.53	64	5.46E-35	3.95E-32	4.99E-33
HEPE	41	0.362	18	1.65E-26	1.19E-23	1.34E-24
PS	1970	17.4	76	5.69E-26	4.11E-23	4.16E-24
O-PI	964	8.52	52	1.52E-24	1.10E-21	1.01E-22
PC	1980	17.5	73	7.45E-24	5.37E-21	4.54E-22
PG	1950	17.2	65	3.27E-19	2.35E-16	1.84E-17
DGDG	1730	15.3	58	2.24E-17	1.61E-14	1.17E-15
O-PC	1040	9.22	40	3.00E-14	2.15E-11	1.46E-12
Other Eicosanoids	16	0.141	8	4.42E-13	3.16E-10	2.02E-11
LPS	63	0.557	10	2.36E-10	1.69E-07	1.02E-08
LPE	67	0.592	10	4.44E-10	3.17E-07	1.80E-08
Leukotrienes	69	0.61	10	5.99E-10	4.27E-07	2.31E-08
O-PE	1100	9.68	31	2.80E-08	2.00E-05	1.02E-06
LPI	61	0.539	8	7.11E-08	5.06E-05	2.47E-06
SM	2320	20.5	48	1.01E-07	7.18E-05	3.36E-06
PA	1950	17.2	42	2.36E-07	0.000167	7.49E-06
Cer-1-P	2290	20.3	46	4.57E-07	0.000324	1.39E-05
Unsaturated Fatty Acids	267	2.36	12	5.92E-06	0.00419	0.000173

4.3.4.2 Pathway Analysis

The Pathway Analysis module combines robust pathway enrichment analysis results with pathway topology analysis to identify the most relevant pathways

involved in this study. The findings of the pathway analysis were depicted graphically in **Figure 4.3.4.3. 1b** and **Table 4.3.4.1. 2**. To enable data exploration, an interactive visualisation system was built. Arachidonic acid metabolism, glycerophospholipid, and sphingolipid metabolism were the top three most significantly (p-value < 0.05, [FDR] < 0.05) impacted pathways. The most impacted pathway, Arachidonic acid metabolism, was chosen as a representative to provide graphical output from the metabolome view in **Figure 4.3.4.3. 1b**. The selected pathway library code was hsa, and pathway views and compound views were generated dynamically based on interactions with the visualisation system (KEGG organisms abbreviation). The red highlighted matched KEGG IDs (C00157, C05966, C05356, C14823 etc.) in the Arachidonic acid belonged to the PC, PE, and PS Lipid classes.

Table 4.3.4.1. 2. The most relevant enriched pathways in the hydrogel-embedded metabolomics dataset bb7.

Enriched Pathways	Total	Expected	Hits	Raw p	Holm P	FDR
Arachidonic acid metabolism	36	0.85935	10	2.71E-09	8.5662	2.28 E-07
Glycerophospholipid metabolism	36	0.85935	3	0.052058	1.2835	1
Sphingolipid metabolism	21	0.50129	2	0.087641	1.0573	1
Linoleic acid metabolism	5	0.11935	1	0.11393	0.94336	1
Lipoic acid metabolism	8	0.19097	1	0.17612	0.7542	1
alpha-Linolenic acid metabolism	13	0.31032	1	0.27045	0.56792	1
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	14	0.33419	1	0.28801	0.54059	1
Glycerolipid metabolism	16	0.38194	1	0.32191	0.49227	1
Phosphatidylinositol signaling system	28	0.66839	1	0.49466	0.30569	1
Biosynthesis of unsaturated fatty acids	36	0.85935	1	0.58516	0.23272	1

4.3.4.3 Network Explorer analysis

In the final step, the Network Explorer analysis module supplemented the MetaboAnalyst joint-pathway Analysis module by detecting links in a gene-metabolite interaction network that crosses pathway borders. This interaction allowed for a more comprehensive perspective of the routes, which would not have been possible if they had been evaluated individually in **Figure 4.3.4.3.**

1c-d. The most interest in the gene-metabolite pathway interaction as a representative was Arachidonic acid metabolism in the red cluster versus all other clusters in this bb7 biopsy with other biopsies, **Figure 4.3.4.3. 1d**. Furthermore, the bb7 biopsy indicated that it mutated for *PIK3CA* in the gene enrichment analysis. This network was built using HMDB and STITCH databases and is only applicable to human studies. The graphical output contained the metabolome view with the corresponding genes/compounds presented in the red highlighted KEGG orthologs (KO) as gene IDs in the Arachidonic acid metabolic pathway. Yellow circles represent metabolites.

MetaboAnalyst and a model of the network of metabolites and genes were assembled. As seen in the network image, **Figure 4.3.4.3. 1d**, metabolites/compounds are involved in arachidonic acid metabolism represented by blue squares and associated genes defined as blue circles.

In addition, the following post-DESI tissue segment was used for H & E staining for histological examination in **Figure 4.3.4.3. 1f**. In the H&E staining, the tissue subtypes were differentiated by staining that aligned to the DESI-MSI cluster regions, as seen in the exemplary samples for bb7 out of 23 core biopsies (cluster 2, red) **Figure 4.3.4.3. 1e.** Following careful co-registration of the scanned H&E images with the core biopsy DESI-MSI cluster image, as seen in the white draw areas in **Figure 4.3.4.3. 1g**. Surprisingly, results suggested that clustering pattern are matching with the histological annotations of tumour samples.

Table 4.3.4. 1. The matched genes/compounds in the gene-metabolite interaction network module of hydrogel embedded sample bb7.

Network Pathways	matched_features (genes/compounds)
Arachidonic acid metabolism	cpd:C05966; cpd:C05356; cpd:C00157; cpd:C00909; cpd:C02165; cpd:C02166; cpd:C14823; cpd:C14812; cpd:C05965; cpd:C04853; hsa:26279; hsa:50487; hsa:240; hsa:4056; hsa:127281
Glycerophospholipid metabolism	cpd:C00350; cpd:C00157; cpd:C00416; hsa:79888; hsa:26279; hsa:50487; hsa:55326; hsa:79143
Linoleic acid metabolism	cpd:C00157; hsa:26279; hsa:50487
Sphingolipid metabolism	cpd:C00195; cpd:C01190; hsa:81537; hsa:4758; hsa:2629
alpha-Linolenic acid metabolism	cpd:C00157; hsa:26279; hsa:50487
Glycerolipid metabolism	cpd:C00416; hsa:8694; hsa:55326
Phosphatidylinositol signaling system	cpd:C00416; hsa:5728; hsa:8396; hsa:3635
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	cpd:C00350; hsa:284098

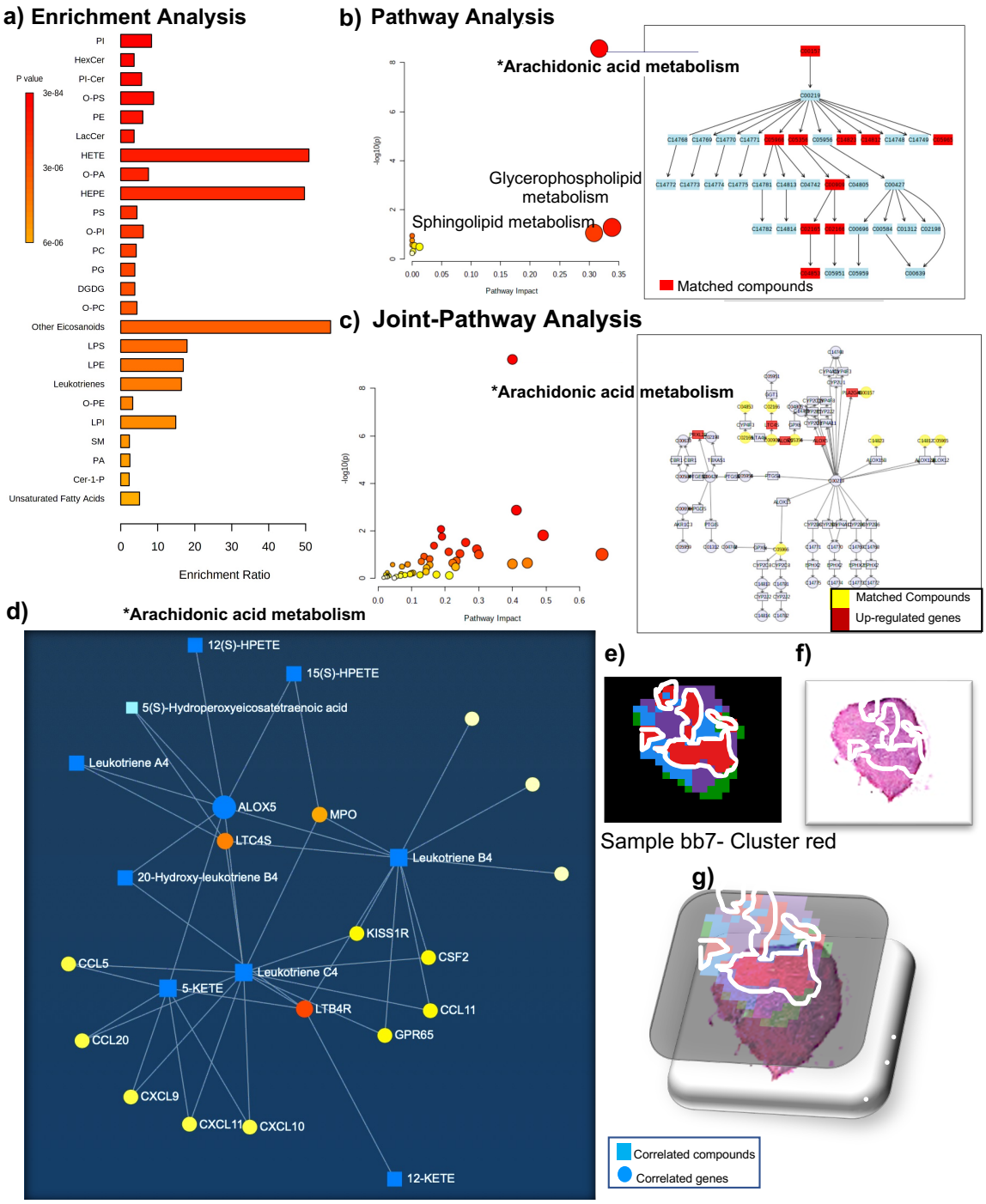


Figure 4.3.4.3. 1. Correlation of transcriptomics and metabolomics behind the DESI- MSI cluster of embedded core biopsy. a) Metabolite Enrichment Overview, Summary Plot for Over Representation Analysis (ORA), The enriched metabolomic content of sample bb7, the significant lipid classes in colour grading red (high) to white (less) for the most significant (p-value < 0.05, [FDR] < 0.05). b)Summary of Pathway Analysis, the global view of the pathways in the interaction network, the most impacted on the metabolite pathway interaction was Arachidonic Acid metabolism to show graphical output contained the metabolome view with the related compounds, red highlighted KEGG labelled compounds matched tentatively with lipid classes c) Summary of MetaboAnalyst joint-Pathway Analysis in Network Explorer module, all the matched pathways are displayed as circles. The colour and size of each circle

are based on the p-value, and pathway impact value, respectively. The most impacted gene-metabolite pathway interaction was Arachidonic Acid metabolism, which showed that graphical output contained the metabolome view with the related genes. The circle in yellow differential metabolites and the red nodes indicate increased gene expression. d) The plot of the selected interaction sub-network, arachidonic acid, in the Network Explorer. e) The selected MSI cluster 2, red of the sample bb7. f) The regions belonging to the cluster red were accurately co-registered to the H&E stained image. g) Co-registration image of MSI cluster and H&E staining.

Numerous analyses based on this could aid in understanding the complexity of heterogeneous core biopsy by revealing enriched pathways, compounds, and genes behind the metabolic cluster red identified by DESI-MSI. In this case, the information was mainly matched to the literature to explain biochemical activities due to the findings from each different type of analysis. To begin with a metabolite enrichment analysis, the top enriched metabolites in this dataset were phosphatidylinositol (PI). The most common fatty acids in almost all animal tissues are PI, which has a high content of stearic acid (18:0) and arachidonic acid (AA). PIP and PIP2 can be formed by phosphorylating the inositol head group[235,236]. Arachidonic acid, in turn, is a precursor to eicosanoids and prostaglandins, which cause inflammation, cell proliferation, migration/infiltration, adhesion, angiogenesis, apoptosis, autophagy, and mitochondrial dysfunction[237]. An emerging metabolism of evidence suggests that cancer cells exhibit specific alterations in phospholipid metabolism that contribute to tumour progression in various cancers, including breast cancer[238,239]. These findings of an enriched lipid class, PI, could confirm the pathway analysis results; as shown in **Figure 4.3.4.3. 1b**, the most enriched pathways were arachidonic acid metabolism due to PI activity. A network analysis based on the top hit genes and compounds in the entire metabolome and transcriptome information matched the arachidonic acid metabolism in **Figure 4.3.4.3. 1c-d**. This network confirmed the previous findings of the KEGG gene and MSEA metabolite enrichment analyses, which showed that arachidonic acid metabolism was the most enriched in the bb7 sample.

4.4 Discussion and Conclusion

The sample preparation processes of small-sized (3 mm) frozen tissue specimens for mass spectroscopy imaging, histological analysis, and transcriptome analysis can be simplified using the presented embedding protocol[174]. This study has confirmed the practicality of gathering these datasets in a co-registered methodology for fragile cancerous core biopsies. Overcoming the batch effect concern was another critical advancement. Individual imaging datasets from tiny or large biopsies frequently exhibit substantial batch effects, which may obscure the researched biological processes.

The data acquisition of a single tissue using MSI approaches can take several hours, depending on the instrument employed. Data collection on multiple core size samples can take several weeks, leading to the imposed batch effects. By contrast, the previously described strategy allows the development of fresh frozen tissue microarrays for core biopsies, enabling uniform description and equivalent analytical results across overall sample sets.

Changes in more key network sites were expected to significantly influence the pathway than changes in marginal or relatively isolated positions in the cancerous tissue. Integrating network topological analysis, interactive network exploration, and functional enrichment analysis yielded the most significantly (p-value < 0.05, [FDR] < 0.05) enriched pathway was Arachidonic acid. In this scenario, interpreting metabolomics and gene expression data was incredibly rewarding and would lead to developing testable experimental hypotheses.

5 Optimisation of a Multi-modal Analysis Strategy using a Single Fresh Frozen Tissue Section on a Technical Level

5.1 Background

A definitive clinical cancer diagnosis is usually based on histological information and medical tests[9,11,17,240–242]. One of the major limitations in diagnosis is the manual analysis of morphological information in tissue. This manual analysis depends on the researcher's prior experience, which might lead to misinterpretation or question the results' certainty. Using histology information in conjunction with mass spectrometric imaging (MSI) can provide a better knowledge of the spatially localised features that can be used to identify tissue morphologies [69,190,243,244].

Using DESI-MSI in the lipidomics approach for cancer diagnostics[193], particularly in ovarian tumour cases, this study has proven its significance by distinguishing the metabolic profiles of cancerous tissues from healthy tissues[35]. Furthermore, it enables the examination of metabolic changes in diverse tissue locations and a better spatial comprehension of molecular dynamics. By focusing on the patient's tumour profiles, this detailed metabolic profiling enables the investigation of Intra Tumor Heterogeneity (ITH)[198.]

Co-registration of several types of information in the consecutive sections had already been discussed in chapter 1. Still, to the extent, this methodology's feasibility using the same tissue section had not been investigated yet. Using several techniques in a multi-modal workflow to analyse the same single tissue section will likely yield more precise information on protein, lipids, metabolites, and RNA/DNA in the region of interest (ROIs). While using adjacent sections for each approach may contribute to inevitable variability, the single section methodology minimises possible information loss[133,245–249]. This approach enables using spatial omics technologies on multi-dimensional ROIs to acquire precise metabolic profiling of diseases. This single tissue section strategy could be beneficial in mapping tumour heterogeneity in cancer [105,121,197].

The Laser Capture Microdissection (LCM) System is primarily utilised to capture ROIs from the same tissue section for follow-up procedures for adaptive multi-modal analysis using a single tissue section. MALDI-MSI paired to LCM has previously been reported [133] to analyse proteins, lipids, and metabolites from the same tissue sections utilizing liquid chromatography-mass spectrometry (LC-MS).

Extracting RNA from fresh frozen tissue slices that have already been scann ed with MALDI-MSI has some constraints regarding the acquisition. Because RNA degradation has a short lifetime[191]. Therefore, to preserve the RNA integrity, fresh frozen samples frequently undergo a fixation process[128,250]. However, fixed samples such as Formalin-Fixed Paraffin-Embedded (FFPE) are characterised by exceptionally well-preserved tissue morphology but poor RNA integrity compared to fresh frozen tissues[250]. Furthermore, one common issue is RNA impurity caused by crosslinking of formalin fixation with nucleic acids (DNA, RNA) and proteins. MALDI-MSI is superior for formalin-fixed tissues; however, using the fixed same ROIs after imaging for RNA extraction is not the best strategy because it may result in RNA contamination[250].

Furthermore, fresh frozen samples utilised for MALDI-MSI are not optimal for follow-up gene expression investigation from the same ROI. Under the experimental conditions, a large proportion of important RNA could be destroyed during matrix coating and acquisition[30,32] . Therefore, an alternative method of a multi-modal analysis on the fresh frozen tissue section is required to preserve the RNA integrity while acquiring MSI data.

This considerable alternative could be DESI-MSI, which allows data acquisition without matrix coating and has a fast imaging set-up to reduce acquisition time[157]. As a consequence, this could aid in reducing RNA degradation and maintaining maximal RNA integrity under ambient circumstances. The RNA material recovered from the ROIs of fresh frozen tissue sections imaged by DESI-MSI could then be preserved for transcriptomics. Follow-up transcriptomics with post-DESI tissue sections, since the heterogeneous nature of the tissue, is a hurdle for molecular research. This novel combination of different methodologies on the same

selected ROIs using the same tissue material provides an excellent opportunity for investigating ITH.

The study discussed in this chapter aims to develop an adapted strategy that incorporates numerous techniques using a single fresh frozen tissue section for DESI-MSI followed by LC-MS or gene expression analysis. In the suggested combination analysis, several parameters must be evaluated. These include selecting the optimal DESI-MSI run time to avoid RNA degradation at ambient temperature and to enhance LC-MS sample preparation for high-throughput analysis of microdissected ROIs. The isolated RNA from that post-DESI segment, in particular, must be of adequate quality and quantity to proceed with the downstream gene expression study, which will involve the use of technologies such as real-time PCR (rt-PCR) and RNA sequencing for transcriptomics [251].

5.2 Aim and Objectives

The research aims to optimise a multi-modal mapping strategy on a single fresh frozen tissue section from ovarian tumour specimens. The methodology makes advantage of first-time literature to do follow-up gene expression analysis or LC-MS on the same area using DESI-MSI technology coupled LCM system. For that purpose, the study's precise objectives are as follows:

- To determine whether the DESI-MSI fast scan setup is more limiting for examining lipid compositions than the conventional scan rate.
- To evaluate whether a DESI-MSI cold stage setup preserves the highest possible RNA quality and quantity level for subsequent studies such as rt-PCR.
- To perform a quality check on all ROIs extracted from post-DESI-MSI sections, rt-PCR was used to examine gene expression.
- To investigate the ideal sample preparation protocol for microdissected ROIs for LC-MS analysis by comparing the intensity of the same features.

5.3 Summary Workflow of Experimental Plan

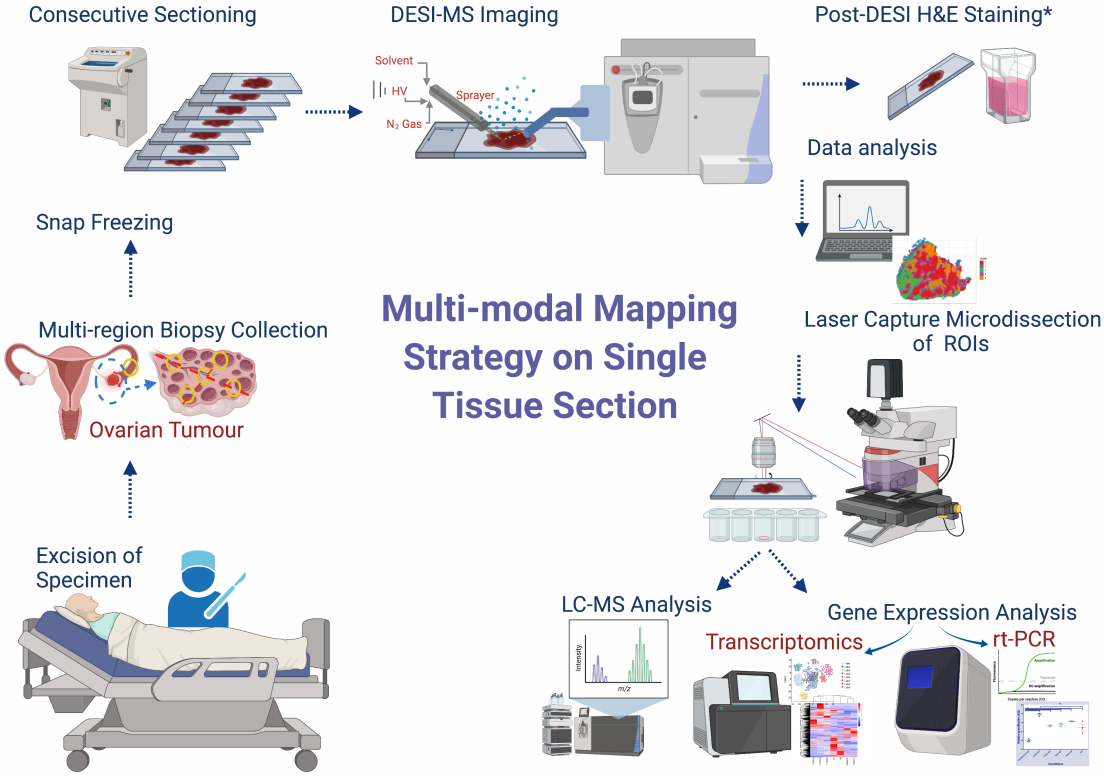


Figure 5.3. 1. The summary workflow of a multi-modal Mapping Strategy on a single tissue section.

A tissue segment was placed on a PEN membrane slide. Desorption/ionisation mass spectrometry imaging (DESI-MSI) was used to generate the tissue spectra. K-mean clustering of the MSI spectra yielded multiple spatial segments (here depicted as blue, red, green, etc.), the outlines of which could be precisely transferred to the optical image of the sample (with the MSI data) to the live-view image in the LCM. This co-registration allowed for the microdissection of the segments, the material of which was then processed for analysis with liquid chromatography-mass spectrometry (LC-MS), resulting in the identification of lipids by the same tissue section.

5.4 Results and Discussion

In this chapter, several technical optimisation studies for multi-modal analysis of a single tissue section were carried out step by step.

5.4.1 DESI-MSI set-up optimisation steps using a single tissue section for follow-up gene expression analysis

5.4.1.1 Evaluation of PEN-coated versus standard glass slides analysis

The same tissue sections for multi-modal analysis necessitated using particular glass slides that allowed for ROI isolation using LCM. PEN-coated glass slides were commonly used for LCM[252] . PEN-coated glass slides differ from typical super frost glass slides used in DESI-MSI because they have a polymer-based membrane on the glass surface. The PEN slides have never been reported to be used in MSI technology before, thus, their appropriateness for DESI-MSI was assessed.

Consecutive ovarian tumour sections on both glass slide types were analysed in negative ion mode and set up with a fast scan rate to test DESI-MSI applicability with PEN slides, as indicated in section 2.6.1. The collected mass spectra on PEN-coated glass slides were compared to the conventional super frost glass slides to see if there were any differences in spectral quality. To further investigate, the quality of the spectral data was evaluated using statistical analysis, as detailed in section 2.13.1, to identify any potential drawbacks, such as fewer detected peaks and overall lower intensity levels than conventional super frost glass slides. In MassLynx, each spectrum was normalised to the strength of the greatest ion peak (base-peak chromatogram). The mass spectrum from tissue and background (slide-only, no tissue) indicated that the PEN-coated slide did not give any chemical background, despite the presence of the PEN membrane on the glass surface, which might adversely impact spectral data quality in **Figure 5.4.1.1. 1a.** The treated data yielded over 600 *m/z* peaks, including specifics of the processing pipeline in section 2.13.1.1, such as noise/baseline subtraction[122,175]. In the

spectra from both slides, the number and intensity of detected peaks were essentially the same in **Figure 5.4.1.1. 1b-c**. When compared to glass slides, 534 with 41 per cent and common 481 peaks, PEN slides provide a slightly less unique number of detected peaks, 285 with 21.9 per cent in **Figure 5.4.1.1. 1d**. As a consequence, compared to the super frost glass slides, the spectral data quality of the PEN-coated glass slides implies that the PEN-coated glass slides can be used with confidence for DESI-MSI research.

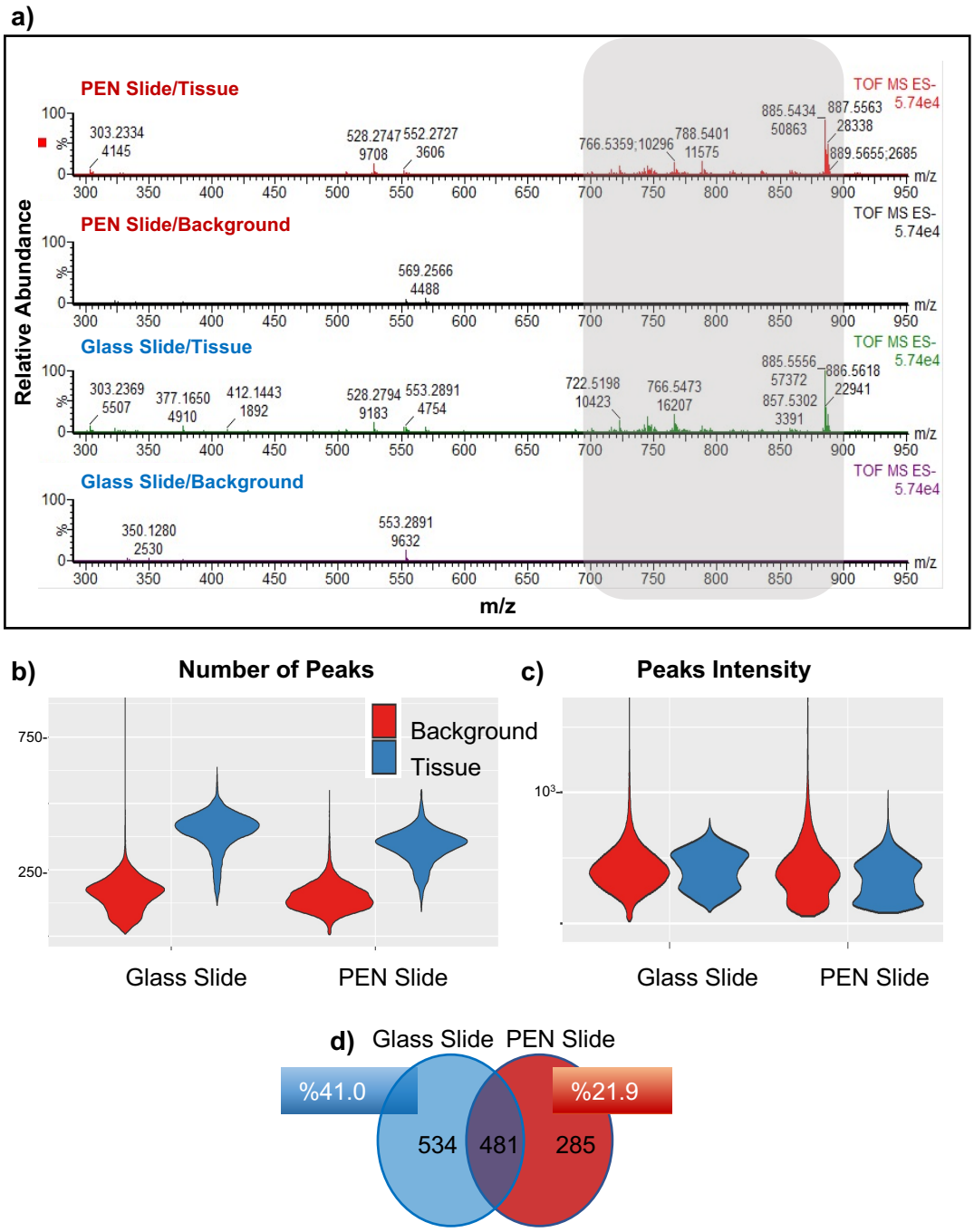


Figure 5.4.1.1. 1. Comparison of spectra from tissue areas on PEN-coated and Super Frost glass slides. a) Each spectrum displayed tissue areas and slide's

background from PEN-coated and Super Frost glass slides. Each spectrum was normalised to the highest ion peak intensity in the grey shaded area, 5.74 e^4 in Mass Lynx software (Waters). b) The total number of peaks shown in the violin plot by detected *m/z* values in each spectrum. The Tukey method was employed as a pairwise comparison and statistical test for the number of peaks c) The peak intensities in each spectrum of the tissue were calculated with the Satterthwaite approximation, a two-sample t-test for degrees of freedom. d) A Venn diagram represents the percentage of peaks discovered from each glass slide.

5.4.1.2 Evaluation of fast scan DESI-MSI setup using PEN-coated glass slides

The earlier study discussed the fast scan setup for DESI-MSI technology when analysing the compatibility of PEN slides. Due to many advantages over the standard scan setup, the DESI-MSI fast scan setup was used throughout this chapter's research. This quick scan setup was carried out with a custom-made sprayer[157] designed to acquire DESI-MSI at a high rate and in a short time.

One of the chapter's conceptual goals was to isolate high-quality RNA from the same post-DESI tissue segment; however, RNA is a delicate biological material. It has been reported that when sectioned samples are maintained at room temperature for an extended time (more than 30 minutes) after defrosting, Rnase enzyme activity can occur during the DESI-MSI acquisition, resulting in RNA degradation[250]. Because this degradation can considerably negatively impact gene expression analyses [159,253,254]; a shorter DESI-MSI period may give a decisive alternative for follow-up RNA studies employing the same post-DESI-MSI sections. The goal of this investigation was to examine if the fast scan setup, which included a custom-made sprayer and a 30-minute DESI-MSI acquisition time, resulted in any spectrum drawbacks, such as fewer identified ion counts or less intense ions. This part's RNA assessment will be discussed in section 5.4.2.

To acquire DESI-MSI at high speed without decreasing sensitivity, scan rates of 10 scans per second (10 scan/s) and 2.5 scans per second (2.5 scan/s) were compared in **Figure 5.4.1.2. 1**. By retaining sensitivity, the faster and shorter sample run analysis could still successfully discriminate chemically distinct tissue types in cancer tissues when compared to the routine analysis. DESI-MS images are traditionally acquired at one or two scans (pixels) per second, leading to several hours consumed for large tissue sections. 2.5 scan

setup was chosen because it allows for the shortest possible capture time and the shortest possible exposure time of RNA in the tissue segment to ambient temperature. As previously stated, reducing acquisition time at room temperature considerably benefits the preservation of high-quality RNA.

To compare ion intensity, the average triplicate spectra of each scan rate were normalised to the highest ion peak intensity, 7.43 e^6 in MassLynx. For this experimental comparison using raw data, MassLynx was used to pick the peaks in the chromatogram between 40-45 minutes from the 2.5 scans per second rate runs (2.5 scans/s, standard rate) and 20-25 minutes from the 10 scans per second rate (10 scans/s, fast scan rate). Both sample sets' time points were randomly selected. Taking a closer look at the fast scan rate, the spectra revealed a drop in the intensity of the targeted *m/z* 600-900 phospholipid range area, particularly *m/z* 885.5, in **Figure 5.4.1.2. 1a** .The *m/z:*885.55 display was chosen because it is used as an internal standard in the pre-processing data step and thus provides a valid signature. PCA transformed the intensity of total detected *m/z* ions to a colour range, and the 2.5scan/s sample had a little greater intensity in **Figure 5.4.1.2. 1a.**

As illustrated by the boxplots, one condition produces more peaks and a more extensive spread of numerous peaks in **Figure 5.4.1.2. 1b**. Tissues examined with a slower run (2.5 scan/s) had more identified peaks than those examined with a faster run (10 scan/s). Following that, a comparison of the total number of identified peaks represents the number of *m/z* values over the noise threshold for each scan rate spectrum in triplicate, as represented by the boxplots. Boxplots with duplicates (n=3) demonstrated that the number of observed peaks was statistically reliable for evaluating the signal. The constancy of the numbers, as revealed by the boxplots, confirmed that this was a compelling statistic in **Figure 5.4.1.2. 1b**. To understand the patterns in intensities, the 90th percentile was determined for the peak intensities in each tissue spectrum, and these values were compared for the various situations in **Figure 5.4.1.2. 1c**. The repeatability of the numbers, as revealed by the boxplots, confirmed that this was a credible statistic. Tissues examined with a slower run (2.5 scan/s) had slightly higher intensity than those investigated with a faster run (10 scan/s) in **Figure 5.4.1.2. 1c.** As a percentage, the fast

runs (10 scan/s) only did not capture 10% of the total peaks from all six datasets in **Figure 5.4.1.2. 1d**.

According to our findings, a short run time in DESI fast scan procedure on consecutive sections from the same tissue sample resulted in slightly less intensity of the identified peaks. Compared to the long and slow run (2.5 scan/s, standard rate), just 10% of the detected peaks could not be captured at fast speed DESI-MSI analysis.

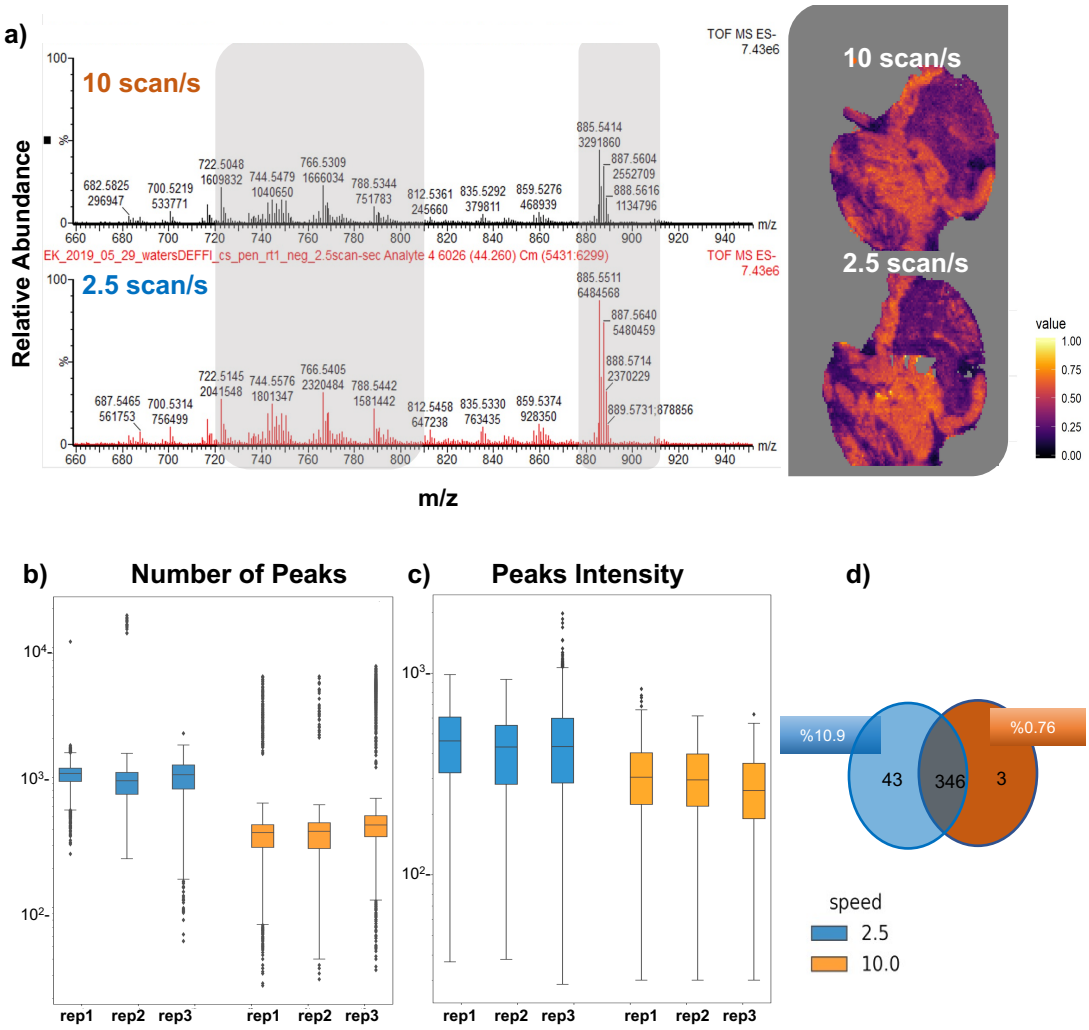


Figure 5.4.1.2. 1. A comparison of the fast Scan DESI-MSI setup using PEN-coated glass slides. a) Each spectrum displayed 2.5 and 10 s/s scan rates from PEN-coated glass slides. Each spectrum was normalised to the highest ion peak intensity in the grey shaded area, 7.43 e^6 in Mass Lynx software (Waters). Comparison of normalised intensities between different scan rates on a cancerous ovarian sample, the intensity of total detected *m/z* ions was transformed to a colour range by PCA, b) The total number of peaks shown in the boxplot by detected *m/z* values in each spectrum. The Tukey method was employed as a pairwise comparison and statistical test for the number of peaks (p.value 0.0001, z.ratio: 4.933, df: Inf, SE:

00.134, estimate: 0.663). c) The peak intensities in each spectrum of the tissue were calculated to the 90th percentile. The Satterthwaite approximation was utilised in a two-sample t-test for degrees of freedom. The level of confidence was 0.95. (p-value: 0.0007, t-ratio: 6.206, df:9, SE: 21.9, estimate: 136.1.) d) A Venn diagram represents the percentage of peaks discovered from each scan rate. n=3 per condition.

5.4.1.3 Investigation of Cold Temperature DESI-MSI Stage's effect on the Imaging analyses

After evaluating the fast scan setup and sample collecting protocol, the next goal was to examine the DESI-MSI cold stage setup. Subzero temperatures are required for RNA research. As a result, the analysis predicted that a cold stage (-20°C) could be a useful strategy for generating DESI-MSI datasets because the cold environment might assist in minimising RNA degradation.

In addition to evaluating RNA quality, quantity, and gene expression (which will be discussed in section 5.4.2), statistical analysis was utilised to determine whether the cold stage DESI-MSI method results in more detected peaks and higher intensity than the room temperature stage.

The data were collected in negative ion mode utilising a fast scan setup with DESI-MSI in triplicates on an ovarian cancer sample set. Six distinct conditions were compared one by one. Several DESI-MSI scan speeds (2.5 and 10 scan/s) were tested on super frost and PEN-coated glass slides on the cold stage. The identical experimental design was carried out at room temperature with the DESI-MSI stage. The observed peaks for each condition are depicted on the Venn diagram. Compared to other conditions, the fast scan setup and cold stage DESI-MSI setup using a super frost glass slide produced the most identified peaks in **Figure 5.4.1.3. 1**. The highest and lowest identified peaks were compared as a first assessment to distinguish similarities and differences between multiple conditions. The maximum identified peaks were achieved using 2.5 scans/s setups on the pen glass slide at room temperature, 3238 (pink). This slower scan rate suggested the best situation for more observed peaks, but a shorter DESI-MSI analysis time in our follow-up RNA investigation was a vital milestone. The total detected peaks obtained with 10 scans/s sets on the PEN glass slide under the cold stage,460 (yellow), were

then compared, followed by the room temperature stage 699 (grey). Based on this preliminary assessment, the room temperature configuration was optimal for obtaining more observed peaks. In addition, the Venn diagram provided a unique detected number for each condition. As unique detected peaks obtained with 10 scan/s setups of glass slide under the cold stage, 521 (purple) while comparing the condition using PEN slide, 0 (yellow). This unique detected peaks comparison concluded that using a glass slide was ideal for the cold stage DESI-MSI analysis rather than the PEN glass slide. To investigate those first assessments based on the Venn diagram of the discovered peak, which will be explored in the following sections of this chapter.

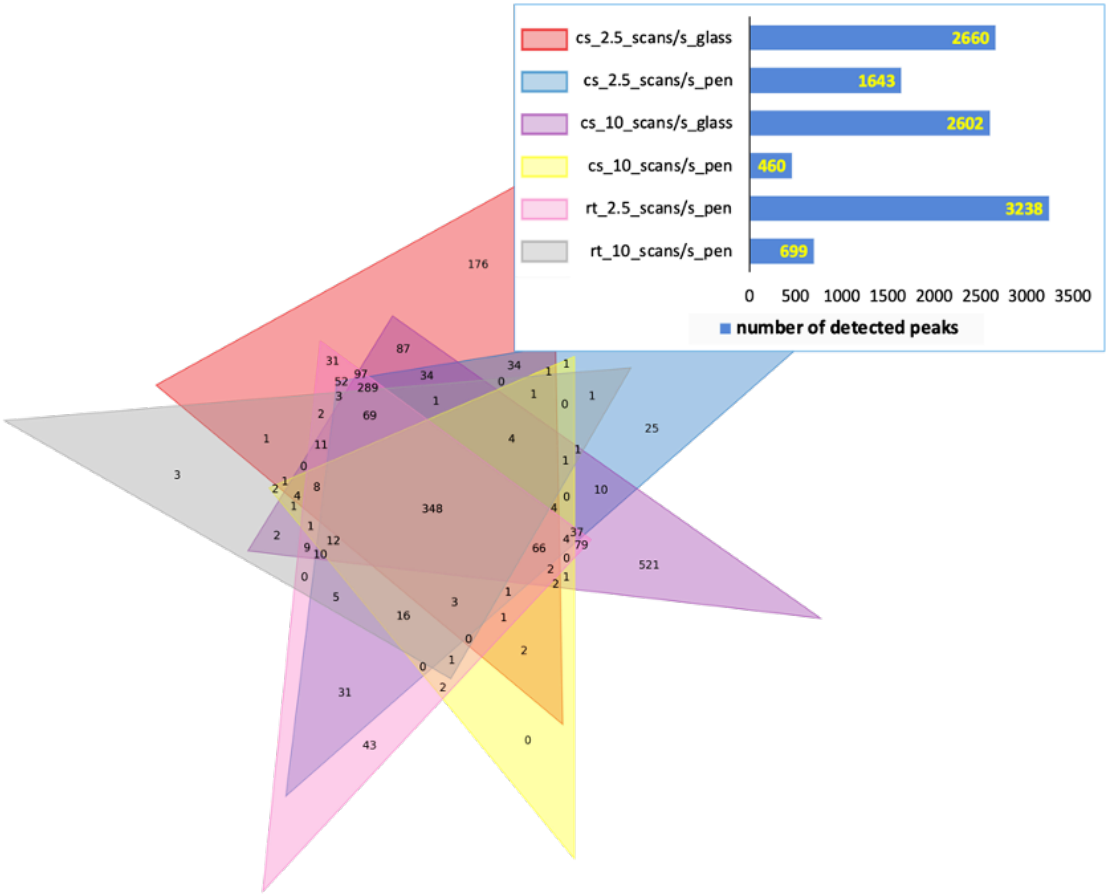


Figure 5.4.1.3. 1. A Venn diagram depicted the number of detected peaks for each condition. The Tukey method was used as a pairwise comparison and statistical test for the number of peaks. rt: room temperature, cs: cold stage.

Hierarchical clustering of the observed *m/z* values was presented as a heatmap in **Figure 5.4.1.3.2.** This heat map could compare and contrast conditions and replicates by highlighting similarities and contrasts. The heatmap demonstrated that replicates clustered together consistently when comparing detected *m/z* values intensity for each condition. Interestingly, even if the replicates are from the same patient, the various circumstances may vary the measured ion intensities. The Appendix contains the list of the top 35 discovered m/z peaks, **Table.5.6.1.**The samples from cs and rt 10s/s on PEN slide (*) displayed a similar intensity pattern.

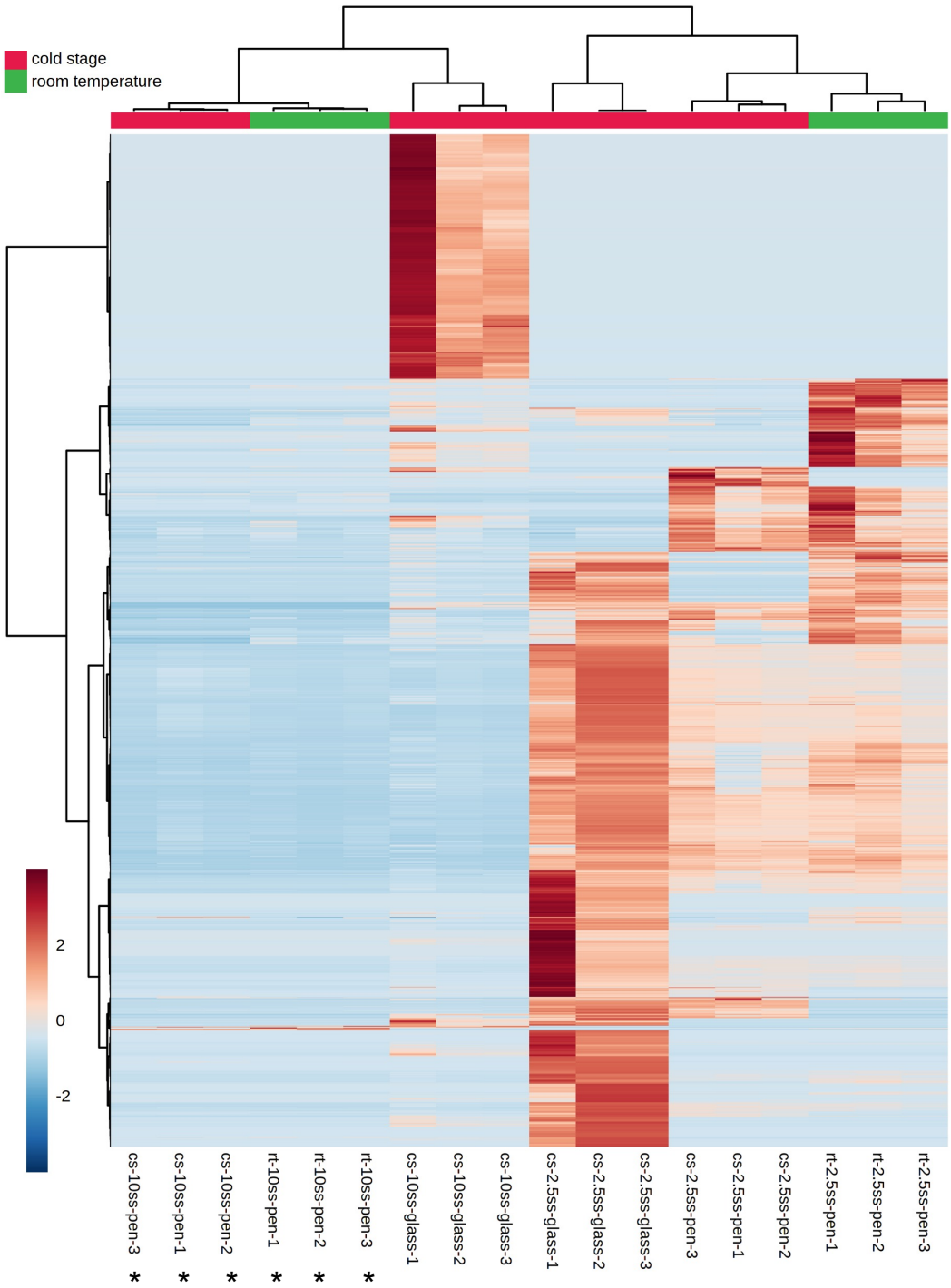


Figure 5.4.1.3. 2. Hierarchical clustering for comparing different conditions in Cold versus Room Temperature Stage of the DESI-MSI. The Heatmap depicted the grouping results between peaks (distance measure using the Euclidean and clustering algorithm using average). Across all conditions and replicates, all identified *m/z* values were compared. cs: cold stage, rt: room temperature; 2.5ss/10ss: scan numbers per second, pen: PEN glass slide, glass: Standard glass slide.

5.4.1.3.1 Comparison of the PEN-coated glass slides versus Standard Glass Slides for both scan rates in Cold Stage DESI- MSI Analysis

Following the initial assessment with the Venn diagram, it appeared that employing glass slides under the cold stage for DESI-MSI had advantages over the PEN slide. Nevertheless, the appropriateness of the PEN-coated glass slides was required to check on the cold stage DESI-MSI due to coated material. This material is a polymer-based membrane known as PEN, which may interfere with tissue analysis in a cold environment. The spectral quality of DESI-MSI data from PEN-coated (blue) and standard Super Frost (orange) glass slides was compared to identify undesired effects such as background peaks or solvent interference from the DESI-MSI stage linked to a cold environment. For both 2.5 and 10 scans/s rates, the number of identified ions was slightly higher in the glass slide in **Figure 5.4.1.3.1. 1a and d.** In terms of intensity, 2.5 scan/s with glass slide was higher than PEN glass slide in **Figure 5.4.1.3.1. 1b**, whereas 10 scans/s with PEN glass slide was higher than glass slide in **Figure 5.4.1.3.1. 1e**.

On the cold stage DESI-MSI, a comparison of PEN-coated glass slides and standard glass slides revealed that 8.2% of the total identified ions were not found when using the PEN in **Figure 5.4.1.3.1. 1f**. However, employing a PEN-coated glass slide could still be beneficial because the strength of the observed peaks at cold stage DESI-MSI rather than the ambient temperature was higher with the fast run (10 scan/s) than with the standard glass slide.

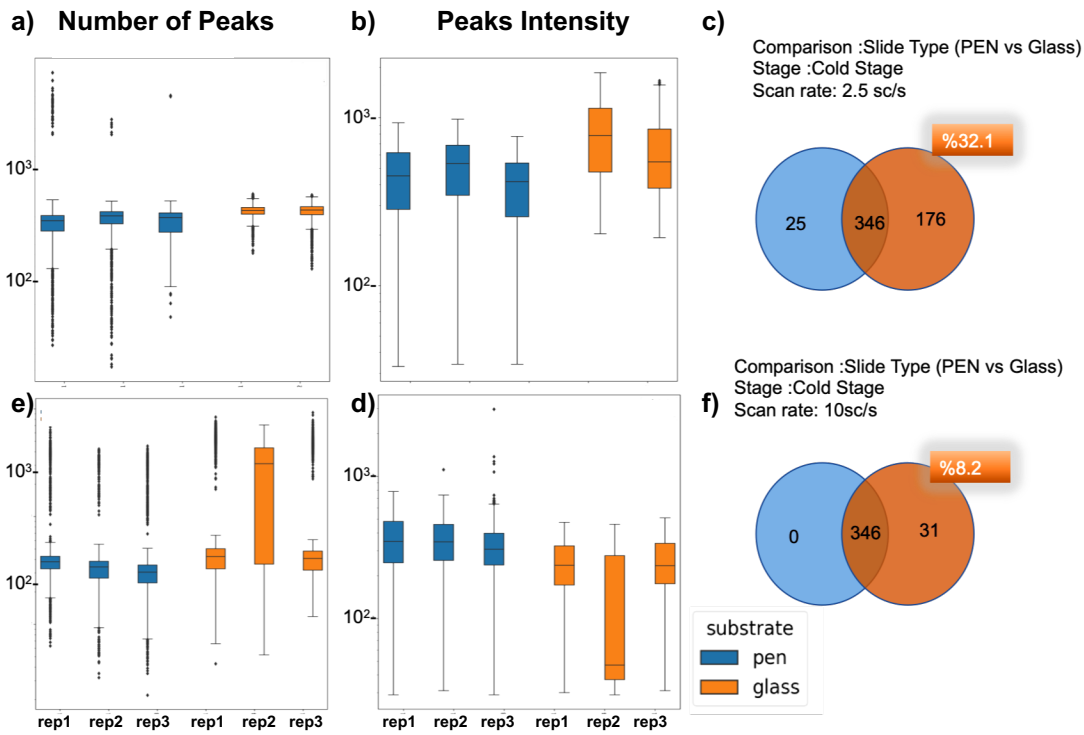


Figure 5.4.1.3.1. 1. Comparison of the PEN-coated versus Standard Glass Slides in Cold Stage of DESI-MSI Analysis. a) The number of peaks above the noise threshold in each spectrum with a scan rate of 2.5 scan/s. b) The 90th percentile calculated the peak intensities in each spectrum of the tissue with a scan rate of 2.5 scan/s. c) The detected peaks are unique and common, with percentages for a scan rate of 2.5 scan/s. d) Comparison of the number of peaks detected above the noise threshold in each spectrum with a scan rate of 10 scans/s e) The 90th percentile showed the calculation of the peaks' intensities in each spectrum of the tissue with a scan rate of 10 scan/s f) The detected peaks, unique and in common with percentages for a scan rate of 10 scan/s. The Tukey method was used as a pairwise comparison and statistical test for the number of peaks. For the 90th percentile of the peak intensity, the Satterthwaite approximation was used in a two-sample t-test for degrees of freedom. The confidence level used was 0.95, n=3 per condition. 2.5sc/s, 10sc/s: scan numbers per second.

5.4.1.3.2 Comparison of PEN-coated glass slides with both scan rates on the Cold versus Room Temperature Stage of the DESI-MSI

Following a cold stage evaluation of both glass slide types with varying scan rates, the following assessment examines the temperature effect on the DESI-MSI analysis. The samples on the PEN-coated glass slides were analysed using DESI-MSI at cold and room temperatures, with longer (2.5 scan/s) and shorter (10 scan/s) acquisition times in **Figure 5.4.1.3.2. 1**. Furthermore, these

datasets were utilised to test whether the spectra from the DESI-MSI stage have undesired impacts such as fewer detected peaks and lower peak intensities in a cold environment. For these potential effects, the spectral quality of DESI-MSI data from PEN-coated slides was examined across conditions.

As revealed by the boxplots, the detected number of peaks for 2.5 scan/s samples and the intensity of the identified peaks were higher from both DESI-MSI stages compared to 10 scan/s in **Figure 5.4.1.3.2. 1a and d.** The slower scan rate yielded 6.7% more *m/z* values in the longer DESI-MSI run time under cold stage conditions in **Figure 5.4.1.3.2. 1c**, corroborating the earlier results that a slower scan rate provided more detected peaks. A comparison of the room temperature stage for a scan rate of 2.5 scan/s against 10 scan/s revealed that the lengthier run discovered 10.9 % more peaks in **Figure 5.4.1.3.2. 1f**. Overall, the DESI-MSI analysis at room temperature for both scan rates produced slightly more unique identified peaks than the cold stage. Interestingly, it is possible to conclude that the cold stage with a fast scan setup does not significantly modify the identified unique peak numbers compared to the room temperature stage. It also did not substantially improve the detection of more peaks than at the ambient temperature stage. In addition, it appeared that the glass slide had an advantage over the PEN glass slide during the cold stage in terms of getting more unique peaks with glass slide conditions. Based on these findings, utilising the PEN slide at the cold stage is not adequate for experimenting. It is preferable to complete the remaining studies at room temperature.

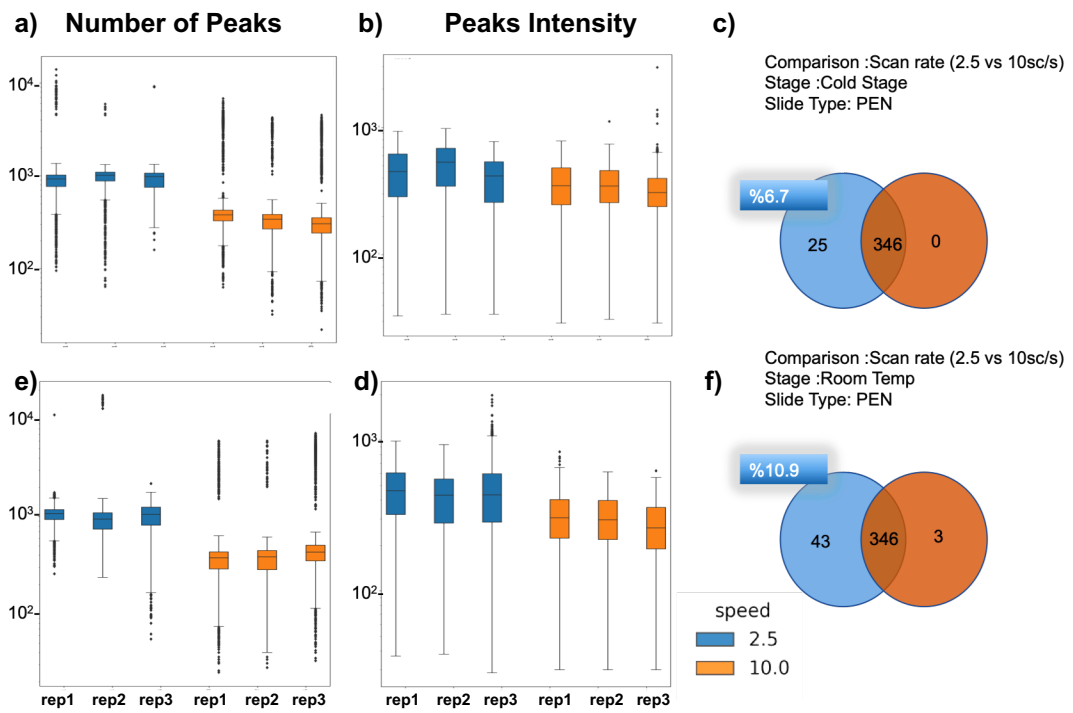


Figure 5.4.1.3.2. 1. Comparison of PEN-coated glass slides in the DESI MSI's Cold versus Room Temperature Stages. a) Comparison of the number of peaks represented by the number of *m/z* values in each spectrum at scan rates of 2.5 scan/s and 10 scan/s from the cold stage. b) The 90th percentile calculated the intensities of the peaks in each spectrum of the tissue with scan rates of 2.5 scan/s against 10 scan/s during the Cold Stage. c) The observed peaks separately and in common with percentages of 2.5 scan/s versus 10 scan/s during the Cold Stage. d) A comparison of the number of peaks represented by the number of *m/z* values detected over the noise threshold in each spectrum at 2.5scan/s and 10scan/s from the room temperature stage. e) The 90th percentile calculated the peak intensities in each spectrum of the tissue at scan rates of 2.5scan/s and 10scan/s from the room temperature stage. f) The identified peaks were detected separately and in common with percentages from the room temperature stage at scan rates of 2.5scan/s and 10scan/s. The Tukey method was utilised as a pairwise comparison and statistical test for the number of peaks. In a two-sample t-test for degrees of freedom, the Satterthwaite approximation was employed for the 90th percentile of peak intensity. The confidence level utilised was 0.95, with n=3 per condition. 2.5sc/s, 10sc/s: scan numbers per second.

5.4.2 Gene Expression Analyses for the post-DESI slides

The mass spectrometric parameters for fast scan setup and sample collection were effectively obtained for the multi-modal analytic procedure, leading to further evaluation at the gene level. Following DESI-MS imaging, the sections were subjected to laser capture microdissection (LCM). Excision of the

predefined ROIs, determined using DESI-MSI, were microdissected for further gene expression investigation, such as transcriptomic.

The importance of RNA integrity in transcriptomics cannot be overstated. RNA is a delicate biological molecule quickly degraded at ambient temperature, particularly by air ozone, which reacts quickly with RNA. These reactions result in the spontaneous breaking of phosphodiester linkages and the degradation[233,255–257]. It was necessary to keep it under control during DESI imaging; hence many optimization techniques were taken to optimise the RNA quantity and quality of the microdissected areas. This potential co-registered gene expression study on post-DESI slides could provide comprehensive information for identifying tissue phenotype and specificity.

5.4.2.1 Optimisations of the RNase Free Staining Protocol for the post-DESI slides

In metabolomics research, mass spectrometric analysis of tissue sections is usually followed by H&E staining to localise the particular patterns identified in stroma and tumour tissues. However, as described in the literature, applying H&E staining on tissue has a deleterious impact on RNA quality and quantity[200]. According to this study, H&E staining can cause roughly 33% degradation in the total RNA of sections. This inverse association is based on the premise that water-based solutions (dyes) activate RNAse, which results in rapid RNA breakdown on tissue [233]. This degradation percentage of isolated RNA may substantially impact gene expression profiling results; the staining process must be optimised using several staining techniques to reveal optimal RNAse-free staining. This section examined the contribution of several staining processes to the transcriptome data produced to address this issue. Due to the limitation of ovarian tissue, staining techniques were optimised using pork liver tissue sections.

The extracted RNA from the pork liver tissue section was evaluated from the complete microdissected sections, which had previously been analysed by DESI-MSI and stained with H&E. Moreover, No LCM/DESI/H&E control

sections were used since they had no DESI and staining. It resulted in 37.5 % degradation of total extracted RNA in **Figure 5.4.2.1. 1a**, validating the literature-based knowledge of a 33% reduction.

The extracted RNA from the pork liver tissue section was examined from the 10 mm^2 areas, microdissected sections that had previously been analysed by DESI-MSI and stained by three different staining techniques, including a control sample without staining in **Figure 5.4.2.1. 1b**. The quantities in **Figure 5.4.2.1. 1c** and RNA Integrity Number (RIN) in **Figure 5.4.2.1. 1d**, of scanned images from different stained samples and isolated RNAs, were assessed. It was a conventional assessment to check the degradation level of the isolated material. In terms of yield, Cresyl Violet offered a higher yield than other stainings; however, the RIN quality did not match the ideal criteria, which should be greater than 5[258] . Following statistical analysis for the integrity comparison, the Harris Modified & Eosin Y technique showed the best quality extracted RNA, with the lowest standard deviation (SD). It displayed reduced variability compared to Cresyl Violet and Gill No:3 & EosinY. After comparing various stainings, Harris Modified & Eosin Y was chosen to have a more isolated yield with an optimal RIN value.

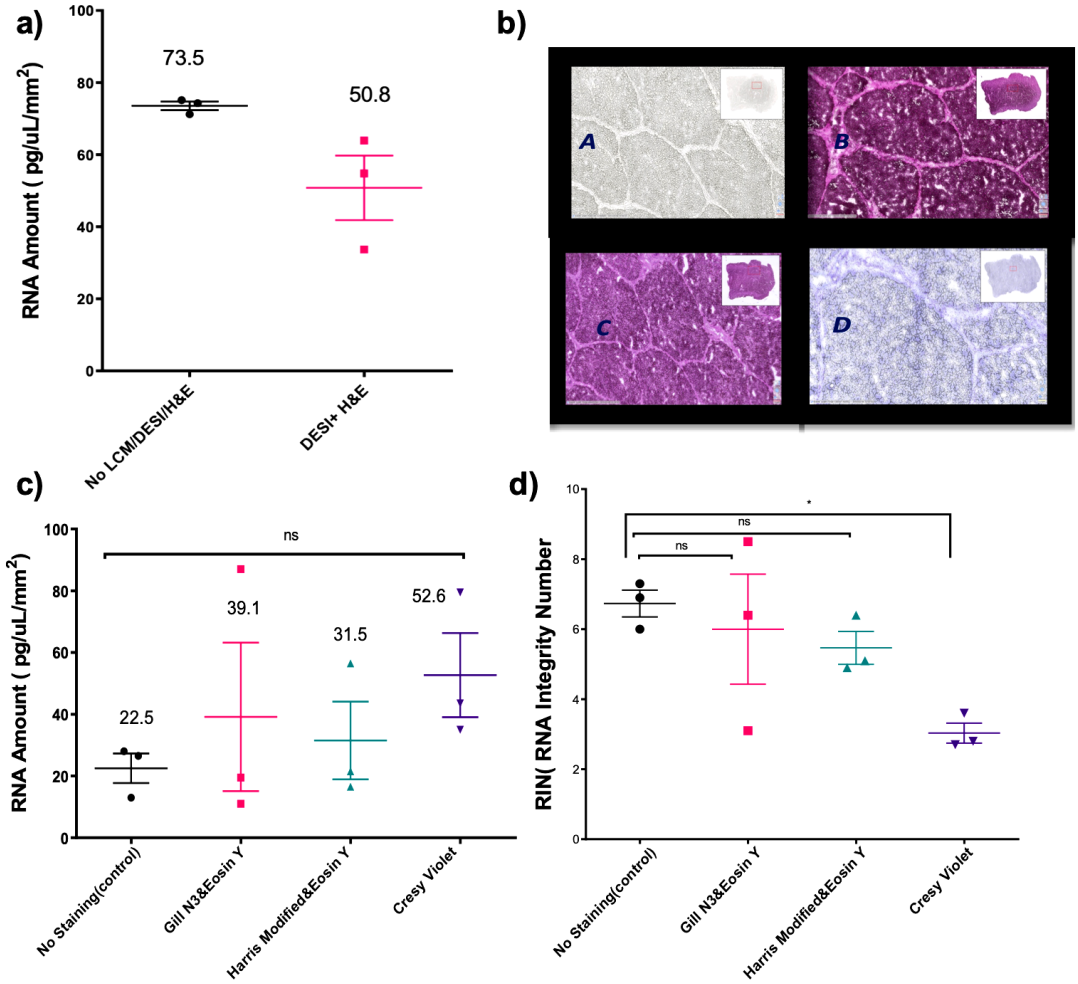


Figure 5.4.2.1. 1. Comparison of the amount of isolated RNA RNA's quantity and quality after DESI-MSI and different staining dyes. a) The average quantity of the isolated RNA from the pork liver tissue stained by H&E after DESI-MSI and the control sample without H&E staining. b) Different staining dyes with pork liver tissue sections A. No Staining, B. Harris Modified & Eosin Y, C. Gill No. 3 & Eosin Y, D. 2% Cresyl Violet perchlorate, c) The average amounts of the isolated RNA, d) The average RIN numbers of the isolated RNA (n = 3 per condition). The bars indicate the mean ± S.D. for all replicates. p-values: * = $p < 0.05$; ** = $p < 0.01$, *** = $p < 0.001$; ns=no significance.

5.4.2.2 Extracted RNA Quality Assessment and RT-PCR from post-DESI microdissected samples

After evaluation, the Harris Modified & Eosin Y staining methodology was optimal for isolated RNA yield and quality (RIN). The staining and DESI-MSI studies were conducted in various settings to assess the effects of staining and DESI-MSI on the extracted RNA from the post-DESI tissue section. Multiple conditions were used, including sections with DESI-MSI analysis and H&E staining, sections with only H&E and DESI-MSI, and sections with neither in **Figure 5.4.2.2. 1**. The results were obtained by micro dissecting 10 mm^2 regions from post-DESI ovarian tumour biopsy sections acquired utilising both SOPs (modified fast snap-frozen and tissue bank). The RNA quantities and RIN counts of extracted RNA from microdissected regions were compared **Figure 5.4.2.2. 1a-b.**

The RNA concentration in the sample was 52.8 pg/µL in **Figure 5.4.2.2. 1a**. The total RNA amount in DESI-MSI-analyzed slides was slightly lower, 50.8 pg/µL. This reduction can be due to extended room-temperature exposure of tissue sections without any precautions, which leads to RNA breakdown[254]. When comparing sections that did not contain DESI-MSI and staining, there was 3.8% RNA degradation compared to prior reported research's RNA degradation (33%). These degradations were less severe[200]. When the findings were compared, the RIN numbers in **Figure 5.4.2.2. 1b** were greater than 5, indicating that the minimum requirement for an ideal quality evaluation was met[258].

The extracted RNA study required gene expression analysis utilising rt-PCR confirmation once the RNA assessment was completed in **Figure 5.4.2.2. 1c**. Following a review of the literature, the genes *TP53* (tumour protein p53), *LPIN1* (Lipin1), *CEL* (Carboxyl Ester Lipase), and *ADH1B* (Alcohol Dehydrogenase 1B(Class I), Beta Polypeptide) were chosen for validation[259,260]. These genes have important roles in carcinogenesis, especially in lipid metabolism.

The TaqMan probe and commercial primers were used for the rt-PCR experiment since they had previously been optimised for a more sensitive assay in many published investigations. This sensitive method aided in

obtaining results from a modest amount of starting material extracted RNA from microdissected post-DESI slide segments. GAPDH was chosen as housekeeping genes because they showed comparable distribution patterns across tumour tissue, reducing the effects of biological variability in total expressed RNA within the material. Following that, we used RNA samples to compare the impact of post-DESI-MSI on gene expression analysis. These results showed that the extracted RNA samples did not affect the RNA expression of the selected genes in **Figure 5.4.2.2. 1c**. The ct values details were in the Appendix, **Table 5.6.4.**

In conclusion, even though the tissue was processed utilising distinct techniques and analyses, such as DESI-MSI and H&E staining, the quality and amount of RNA were adequate for downstream analysis, such as gene expression study using rt-qPCR.

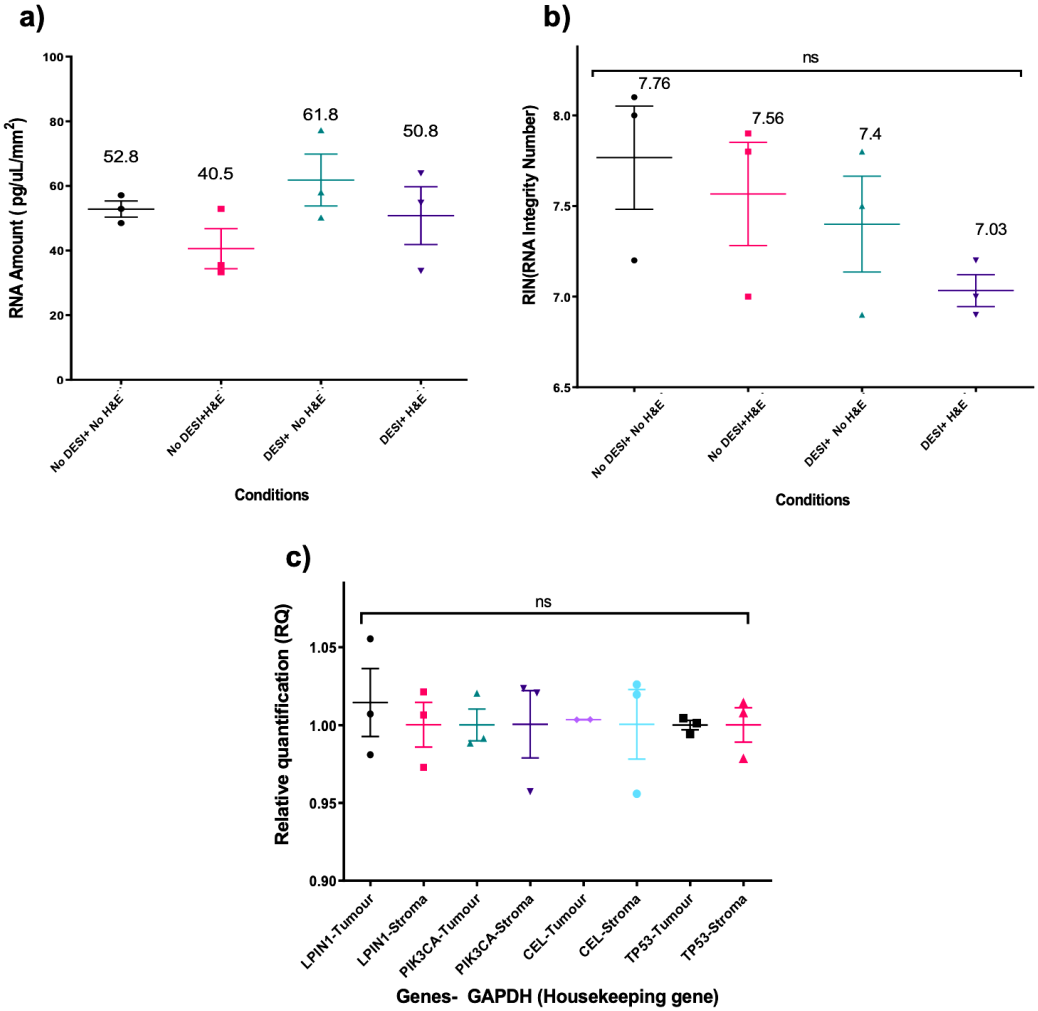


Figure 5.4.2.2. 1. Comparison of the gene expression and quantity/quality of the isolated RNA samples in different conditions. a) RNA yield from different conditions, b) RIN numbers from different conditions. RIN numbers were calculated using the Agilent Bioanalyzer system for Quality Assessment. c) Relative quantification of the housekeeping gene *GAPDH* with samples via rt-PCR using gene of interest *TP53, LPIN1, CEL*, and *PIK3CA* expression with ovarian tumour tissue. Extrapolating the relative expression values, delta ct values were normalised to the average of replicates ct values as control due to lack of wild-type control (n=3 per condition). The bars indicate the mean ± S.D. for all replicates. p-values: * = $p < 0.05$; ** = $p < 0.01$, *** = $p < 0.001$; ns=no significance.

5.4.2.3 Extracted RNA Quality Assessment and RT-PCR from post-DESI microdissected samples under cold stage acquisition

As previously stated, samples for RNA extraction after DESI-MSI could only be kept at room temperature for 30 minutes. As a result, the next step was to

examine RNA yield and RNA integrity. These tests were carried out to see if cold stage extended runs could produce a minus zero atmosphere, preventing RNA degradation if the tissue required more than 30 minutes for DESI-MSI acquisition.

Previously DESI-MS imaged sections were compared for RNA quantities and RIN counts, and it was determined to continue using the same section for follow-up studies. The ovarian tumour tissues were microdissected into 10 mm^2 regions. The measured RNA yield from 10 scan/s cold stage samples was nearly three times higher **Figure 5.4.2.3. 1b.** For example, the average RNA amount from the 10 scans/s cold stage sample was 28.3 ng/μL, while the RNA amount from the 10 s/s room temperature stage sample was 9.2 ng/μL. The reduced amount of RNA for 10s/s could be attributed to DESI-MSI analysis at ambient temperature. At both stages, the extracted RNA in **Figure 5.4.2.3. 1b** and matching RIN scores in **Figure 5.4.2.3. 1a** was shown for varied scan rate datasets ranging from 10 scans/s to 2.5 scan/s. The RNA yield for the samples at 2.5 scan/s on the cold stage was higher than at room temperature.

In contrast, the RIN score for the 10 scan/s run was much higher at room temperature, while the RIN score for the 10 scan/s run at the cold stage was slightly higher than both of those from 2.5 scan/s runs. Exposure to room temperature settings via the LCM system, which could take up to 30 minutes, could be a potential explanation for the slight differences.

Nonetheless, there was some indication that cold samples could promote RNA yield and that the shorter 30-minute run could greatly aid a better RIN score. A RIN score of 5.0 shows minimal contamination of small degraded RNA products and is widely considered the lowest value required to achieve acceptable results using RNA-Seq procedures. RIN quality greater than 5 is acceptable [258,261] for additional downstream analysis.

Figure 5.4.2.3. 1b shows that all the sample RIN numbers were greater than five, indicating that all extracted RNA integrity was in good condition. All samples were appropriate for transcriptome analysis when run at 10 scan/s for 30 minutes and 2.5 scan/s at -20°C and room temperature.

The gene expression analysis of the extracted RNA was performed under the cold stage and room temperature acquisition in **Figure 5.4.2.3. 1c**. The rt-PCR experiment used *GAPDH* as a housekeeping gene and *TP53* as a target gene. Based on a thorough literature study, *TP53*[260,262] was chosen for this validation section. Some lipid metabolism genes are controlled by *TP53*[259] and directly relate to High-grade Serious Ovarian Carcinoma (HGSOC) survival. The effect of DESI-MSI analysis with cold and room temperature stages on RNA expression under various settings was compared to healthy control. Compared to the cold stage or room temperature, isolated RNAs from diverse circumstances had no significant impact on the RNA expression of the targeted gene. Only the healthy control exhibited significantly lower values in the relative quantification, as expected in **Figure 5.4.2.3. 1c**.

In conclusion, even though the tissue was processed at a cold stage, the yield was higher, and the RIN quality was nearly similar. Because of the additional technical preparation required for the cold stage, running DESI-MSI with a cold stage set-up is not optimal for future studies. As a result, the room temperature setup is optimal for subsequent experiments.

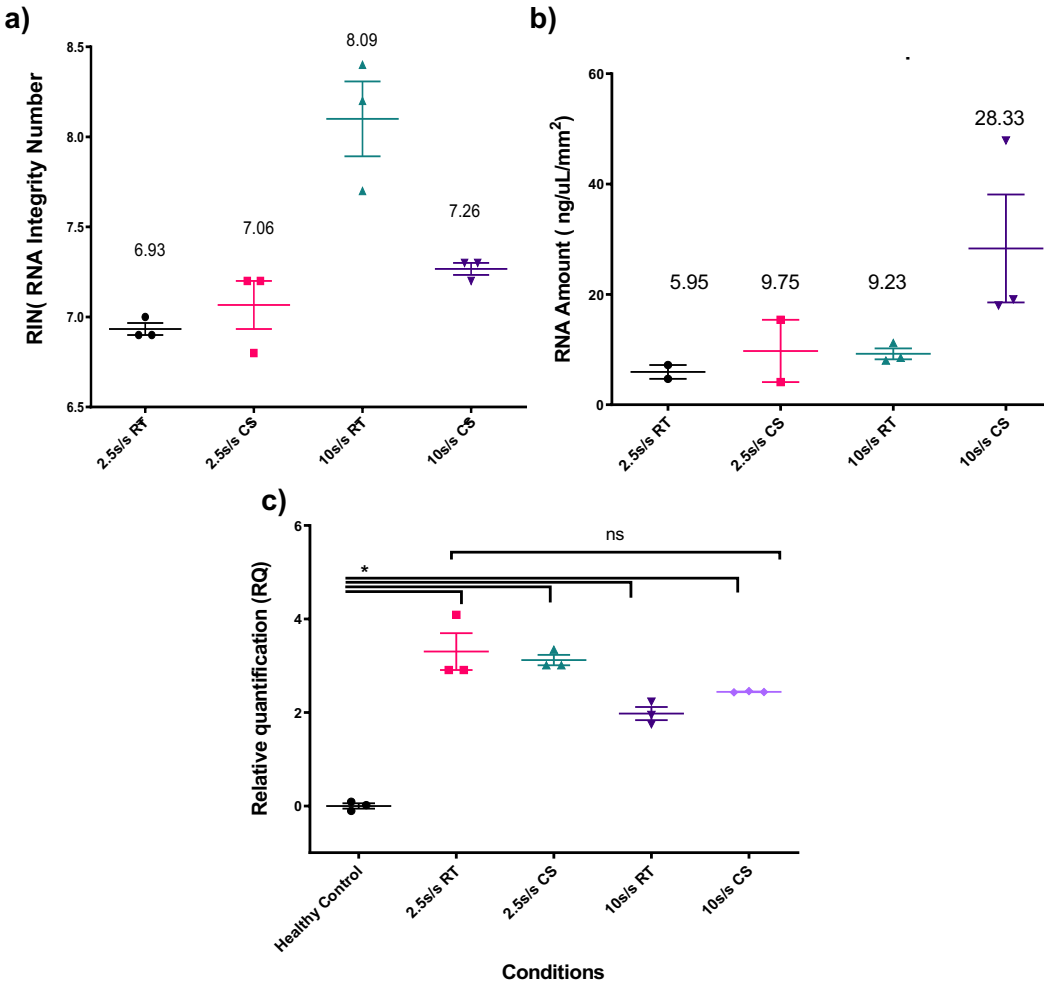


Figure 5.4.2.3. 1. The gene expression and quantity/quality of extracted RNA samples from post-DESI sections were compared in room temperature and cold stage conditions with different run times (2.5s/s and 10s/s). a) RNA yield from different conditions, **b)** RIN numbers from different conditions**.** RIN numbers were calculated using the Agilent Bioanalyzer system for Quality Assessment. **c)** Relative quantification of the housekeeping gene *GAPDH* with samples via rt-PCR using gene of interest *TP53* expression with tumour and healthy ovarian tissue. Extrapolating the relative expression values, delta ct values were normalised to the average of replicates ct values as control due to lack of wild-type control (n=3 per condition). The bars indicate the mean ± S.D. for all replicates. p-values: * = $p < 0.05$; ** = $p < 0.01$, *** = $p < 0.001$; ns=no significance. 2.5s/s, 10s/s: scan numbers per second,

5.4.3 LC-MS Analysis for the post-DESI slides

To extend this multi-modal workflow to another strategy, region-specific DESI-MSI data was used to guide the laser microdissection (LCM) system for liquid chromatography-mass spectrometry sample preparation (LC-MS) analyses on a single fresh frozen tissue section. It is already known that the spatial distribution of tissue features can be investigated using desorption electrospray ionisation mass spectrometric imaging (DESI-MSI). Considering DESI-MSI with more sensitive local extraction-based LC-MS may enable thorough molecular characterisation from the same tissue section[263–267]. Sample preparation methodologies were evaluated to find the optimal protocol for microdissected tissues. Two common sample preparation protocols for lipidomics were modified, as described in section 2.8.1, to address using micro-scale tissue regions (4.5mm^2) in LC-MS analysis. The first was a water-soluble organic solvent such as isopropanol (IPA)[268,269] . The other was the MTBE-MeOH (Methanol) (mMM) method of the widespread liquid-liquid extraction for shotgun lipidomics[270270,271].

The advantage of this study is the prospect of combining the molecular spatial information supplied by DESI-MSI data clustering analysis with micro-lipidomics characterization by LC-MS on the same tissue section using a Leica LCM 6500 system in **Figure 5.4.3. 1**.

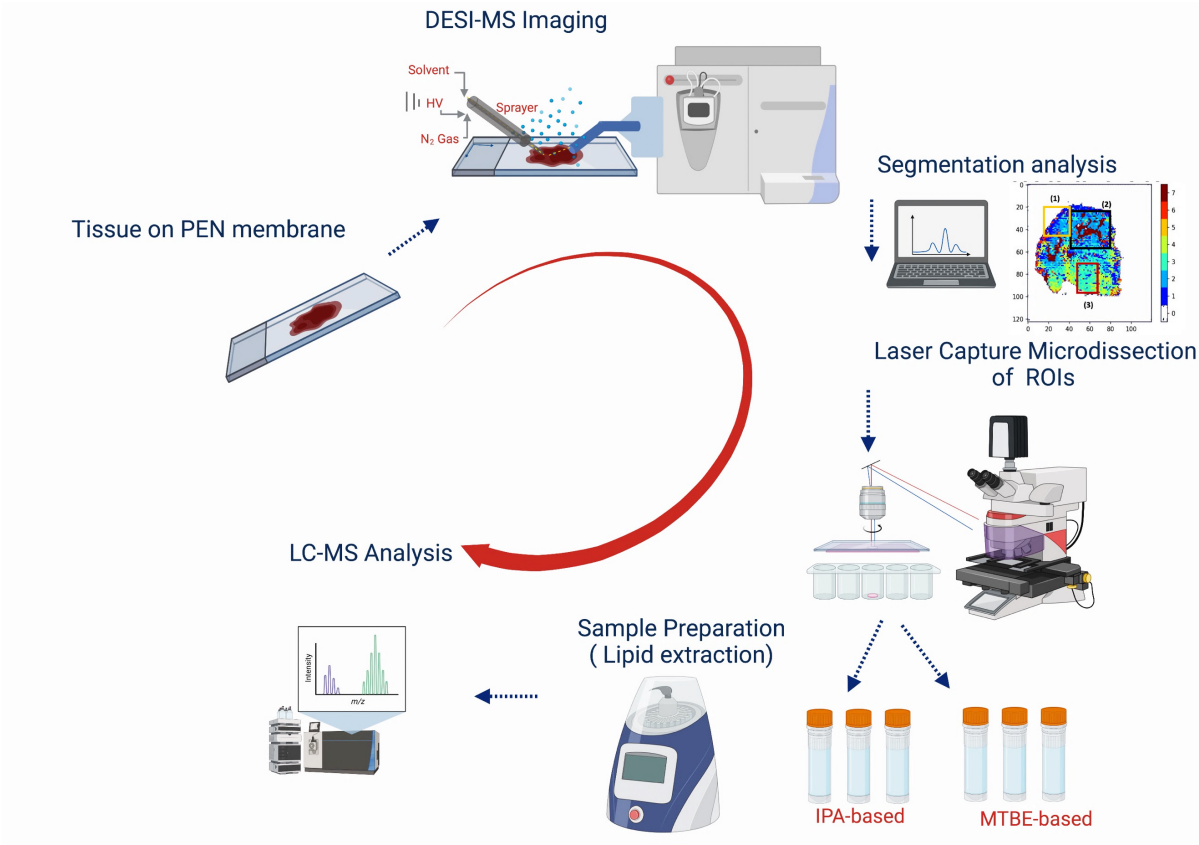


Figure 5.4.3. 1. DESI-MSI-guided LC-MS analysis summary workflow on a single tissue section. The tissue is sectioned and placed on a PEN membrane slide. The specimen is imaged using the imaging technique of desorption/ionisation mass spectrometry (DESI-MSI). Data segmentation analysis is performed using K-mean clustering of the MSI spectra. Multiple spatial segments were generated (depicted here with distinctive colours such as blue, red, green, etc.). The outlines of each cluster of interest can be precisely reassigned to the optical image of the sample with the MSI data and the live-view image in the LCM. This co-registration allows for the microdissection of the segments. The material is then processed for analysis with liquid chromatography-mass spectrometry (LC-MS), resulting in the identification of lipids using a single tissue section.

5.4.3.1 DESI-MSI clustering analysis to reveal metabolic segments for microdissection

Region-specific DESI-MSI data was used to guide a laser microdissection system for sample preparation of lipidomics analysis. DESI-MSI was applied to a 10 µm thick section of fresh-frozen ovarian cancer tissue collected by Imperial College London's Tissue Biobank. The section was mounted on

polyethene naphthalene (PEN) membrane slide, which is compatible with the MSI method required for the subsequent LCM. Six consecutive sections of DESI imaging were performed in negative ion mode (Xevo G2-XS QToF, Waters) at 10 scans/s and 100µm pixel size, followed by hierarchical segmentation analysis (average linkage) clustering to achieve specific ion-associated segments. Seven different metabolic clusters were identified in **Figure 5.4.3.1. 1a.** Each cluster was associated with a set of *m/z* value lists representing the spatial distribution of significantly colocalised ions. In the volcano plots of each cluster against all clusters, the fold change (FC) of the high and low abundance of detected features was shown in **Figure 5.4.3.1. 1b-h**. Volcano plots were used to obtain the maximum distribution of differentially expressed metabolites. The heat map depicts the comparison and contrast of the average intensity of the detected features identified by DESI-MSI in each cluster per sample set by emphasising similarities and differences using hierarchical (agglomerative) cluster analysis in **Figure 5.4.3.1. 1i.** The Appendix contains the complete list of the top 35 discovered m/z peaks, **Table.5.6.2.**The standard criterion for this experiment was |log2(FC)| > 1, and the p-value was below 0.05. The total number of detected *m/z* values in the clusters was 752 – clusters two (blue), three (green), and seven (red) were chosen randomly as representative examples from the seven clusters for microdissection by LCM.

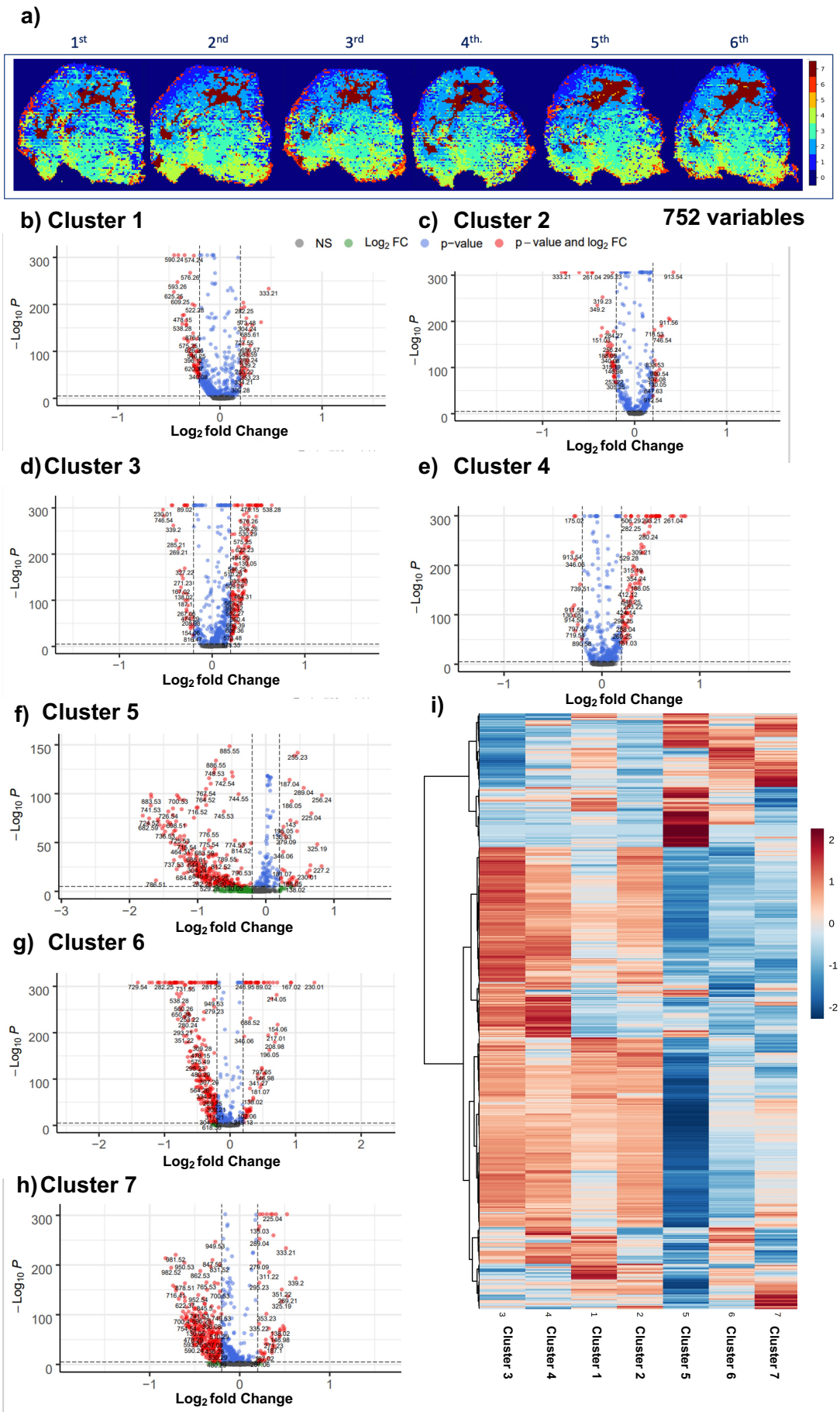

a)
1st
2nd
3rd
4th.
5th
6th
b) Cluster 1
c) Cluster 2
752 variables
NS
Log2 FC
p-value
p-value and log2 FC
−Log10 P
Log2 fold Change
d) Cluster 3
e) Cluster 4
f) Cluster 5
g) Cluster 6
h) Cluster 7
i)
Cluster 3
Cluster 4
Cluster 1
Cluster 2
Cluster 5
Cluster 6
Cluster 7

Figure 5.4.3.1. 1. DESI-MSI clustering analysis reveals metabolic segments for ROIs microdissection. a) Seven different metabolic clusters were identified between six consecutive sections, with highly colocalised ions identified using ion-pair Spearman's correlation and hierarchical clustering (average linkage). The first PC scores, obtained using the ions in each cluster, were used as the representative spatial distribution of the groups of highly colocalised ions. b-h) Volcano plots of 7 different clusters from the sample with |log2(F.C.)| > 1 and p-value < 0.05 i) The Heatmap depicted the grouping results of all clusters between *m/z* values (distance measure using the Euclidean and clustering algorithm using average).

After manually co-registering selected segments in the rectangular areas by the DESI-MSI in **Figure 5.4.3.1. 2a**, these segments were scanned by the LCM software system (Leica LCM 7000, Leica Microsystems, Germany) in **Figure 5.4.3.1. 2b**. Three segments (blue (1), red (2), and green (3)) were selected and microdissected (total area 4.5 mm^2 per segment, **Figure 5.4.3.1. 2c**) from every six consecutive slides for two further lipid extraction procedures (IPA and MTBE-based), resulting in a total of 18 samples and controls (only PEN membrane, protocols solvents, blank solvent etc.). To see a live-view video of the cluster red microdissection procedure, please scan the QR code in **Figure 5.4.3.1. 2d.** These samples were then run for LC-MS analysis (Xevo G2 XS QToF, Waters).

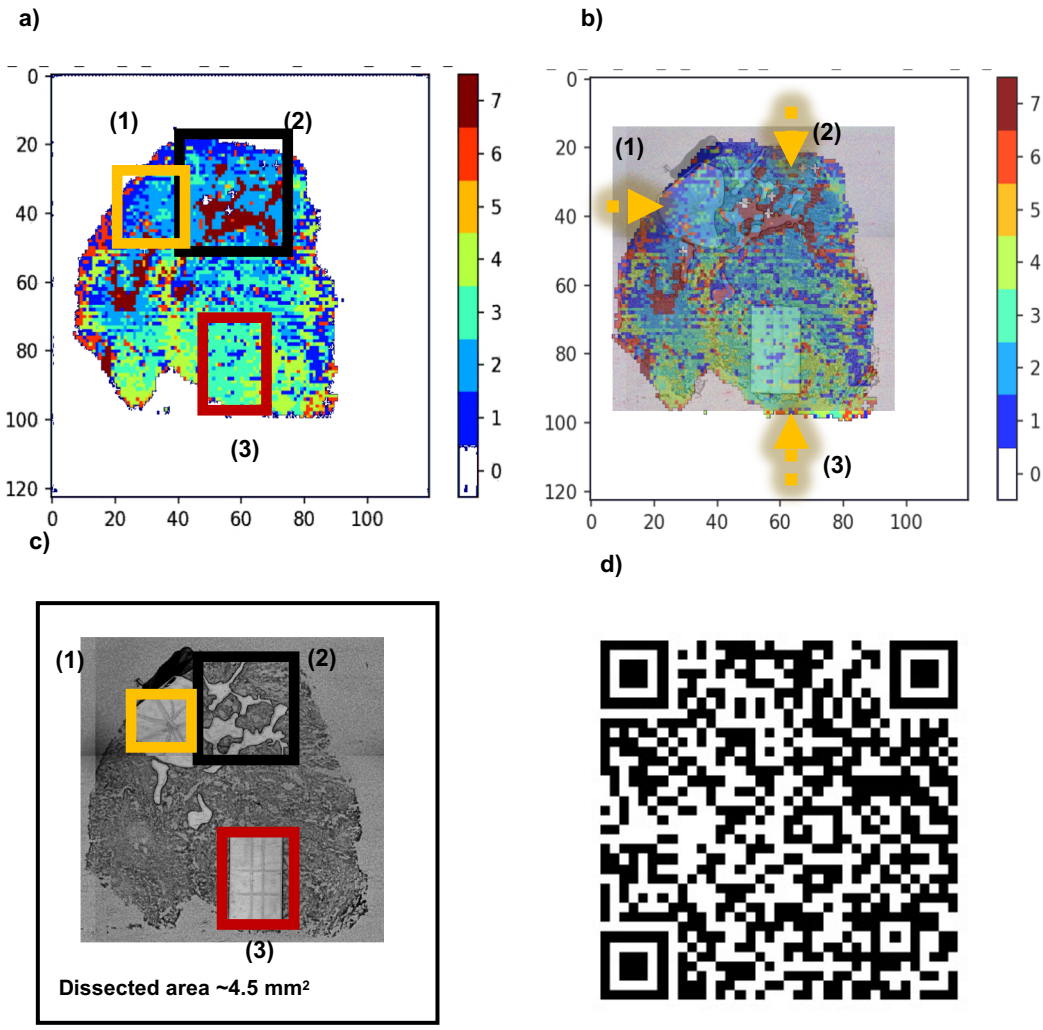


Figure 5.4.3.1. 2. Co-registration of DESI-MSI Segments to optical image for laser microdissection (LCM).a) The selected tumour regions of an ovarian cancer sample after clustering using the k-means algorithm. b) The regions belonging to the three DESI-MSI segments were accurately co-registered to the optical image. c) The microdissection of regions in the live-view image in the LCM system d) QR code of live-view video for microdissection procedure of red cluster segment.

5.4.3.2 Comparison of the IPA and MTBE-based sample preparation protocol for microdissected ROIs

To compare the sample runs, hierarchical (agglomerative) cluster analysis was used to begin each sample aggregation until they all belonged to the same group. A heatmap was created from the sample runs in the hierarchical

clustering of all samples in **Figure 5.4.3.2. 1a.** This heat map examines the similarities and differences between the two distinct extraction techniques with dissected three regions' technical replicates (IPA-C2/C3/C7 and MTBE-C2/3/7) and benchmark controls (IPA-T and MTBE-T), which are bulk tissue sections and PEN membrane controls (no tissue), respectively (IPA-P and MTBE-P). The Appendix contains the complete list of the top 35 discovered m/z peaks, **Table.5.6.3.** After running these statistical tests and comparisons on the identified peaks above the noise threshold and the intensity of the detected peaks, it was determined that the extracted samples using the IPA-based process in **Figure 5.4.3.2. 1a** and **c** had more observed peaks than the samples using the MTBE-based approach in **Figure 5.4.3.2. 1a** and **d**. The IPA-based technique samples covered 85.6 % of the total peaks found, while the MTBE-based protocol samples supplied 14.2 %. Using the IPA-based technique, only a small number of features (25) were not recognised at high-intensity levels **Figure 5.4.3.2. 1a**.

Even though the duplicates are from the same patient, the measured ion intensities and recognised peaks differ depending on the lipid extraction process. For lipidomics analysis from quantity-limited microdissected materials, the IPA-based methodology approach appears superior to the MTBE-based methodology.

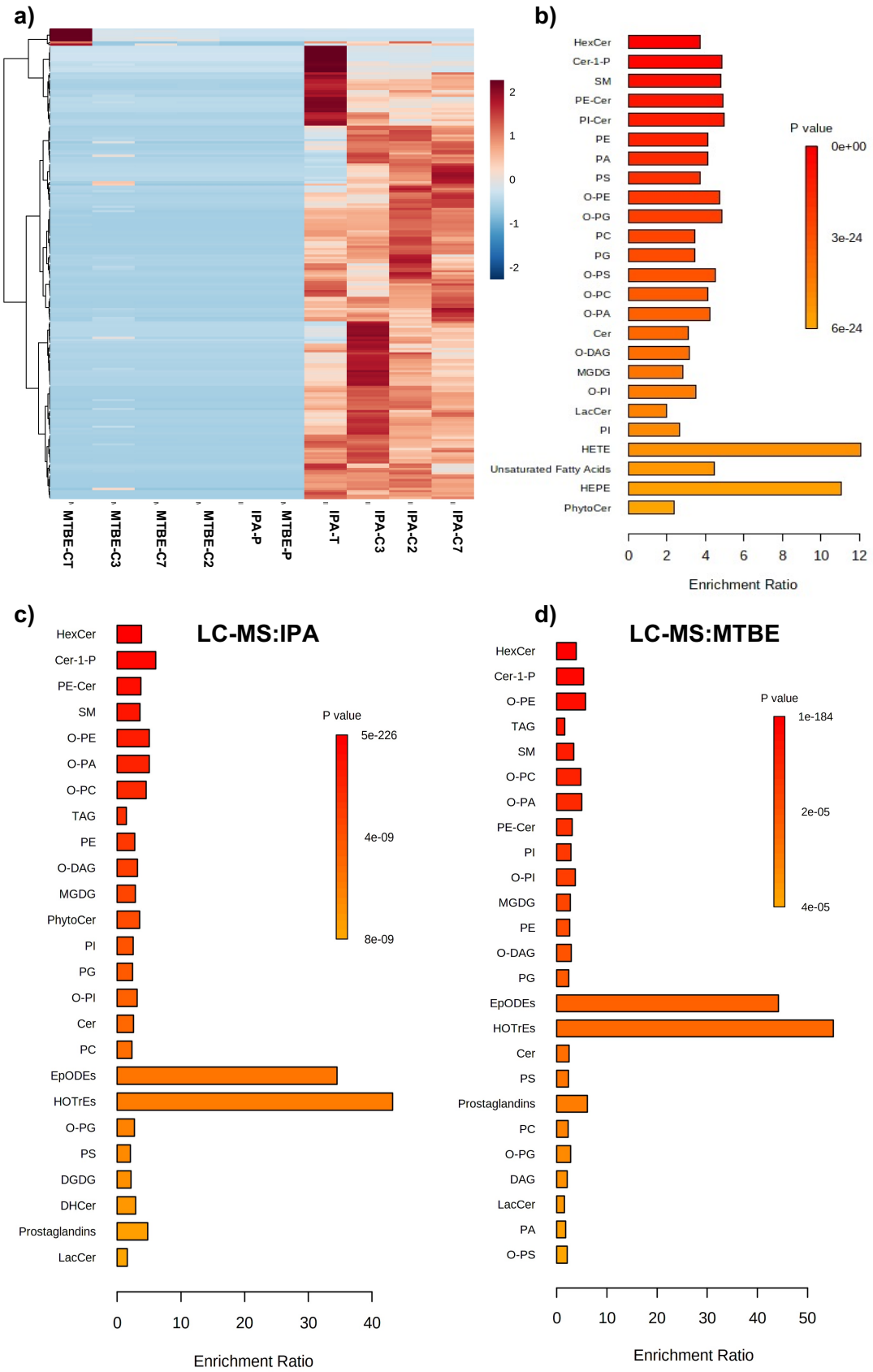


Figure 5.4.3.2. 1. Comparison of two lipid extraction methods, IPA and MTBE-based, for LC-MS micro-lipidomics characterisation. a) Heatmap displaying the grouping results for all peaks (distance measure using Euclidean, and clustering

process using average) dissected three regions(IPA-C2/C3/C7 and MTBE-C2/3/7) and benchmark controls (IPA-T and MTBE-T), which are bulk tissue sections and PEN membrane controls, respectively (IPA-P and MTBE-P). b) DESI-MSI Metabolite Enrichment Content of the Ovarian Sample c) LC-MS analysis of IPA-based lipid extraction d) LC-MS analysis of MTBE-based lipid extraction. All samples' enhanced metabolomic content, with hexosylceramides (HexCer) being the most significant (p-value 0.05, FDR 1) lipid class, followed by 25 distinct lipid classes such as Cer-1-P, PE-Cer, SM, PI-Cer, O-PE, etc.

Following the heatmap analysis, Metabolite set enrichment analysis (MSEA) is performed to analyse metabolomics data to uncover biologically significant patterns of these features identified in high abundance in the MTBE-based dataset. When comparing three datasets: DESI-MSI in **Figure 5.4.3.2. 1b** and **Table 5.4.3.2. 1**, LC-MS with IPA in **Figure 5.4.3.2. 1c** and **Table 5.4.3.2. 2** ; MTBE in **Figure 5.4.3.2. 1d** and **Table 5.4.3.2. 3**, the top significant enriched lipid classes, hexosylceramides (HexCer) (p-value 0.05), of the ovarian sample were similar. Compared to the DESI-MSI dataset, the IPA-based LC-MS has more matching lipid content. It could imply that IPA-based extraction has already demonstrated higher intensity for identified peaks and more overlap with DESI-MSI content. **Table 5.4.3.2. 1**, **Table 5.4.3.2. 2** and **Table 5.4.3.2. 3** summarise the top enriched 25 MSEA findings.

Table 5.4.3.2. 1. The complete enhanced metabolomic content of the Ovarian sample by DESI-MSI. The 'raw p' is the original p-value determined from the enrichment analysis, the 'holm p' is the p-value corrected using the Holm-Bonferroni method, and the 'FDR p' is the p-value adjusted using the False Discovery Rat

Lipid Classes	total	expected	hits	Raw p	Holm p	FDR
HexCer	8690	201	765	4.52E-226	3.31E-223	3.31E-223
Cer-1-P	2290	53.1	321	1.77E-147	1.29E-144	6.48E-145
PE-Cer	2290	53.1	197	7.18E-55	5.23E-52	1.75E-52
SM	2320	53.7	192	3.70E-51	2.69E-48	6.76E-49
O-PE	1100	25.4	128	3.61E-50	2.62E-47	5.27E-48
O-PA	965	22.4	113	1.59E-44	1.15E-41	1.93E-42
O-PC	1040	24.2	110	4.53E-39	3.29E-36	4.73E-37
TAG	27100	627	888	2.31E-27	1.67E-24	2.11E-25
PE	1970	45.8	126	1.31E-23	9.44E-21	1.06E-21
O-DAG	1270	29.5	94	3.53E-22	2.55E-19	2.58E-20
MGDG	1730	40.1	113	4.67E-22	3.37E-19	3.10E-20
PhytoCer	805	18.7	66	2.81E-18	2.02E-15	1.71E-16
PI	1960	45.5	113	5.69E-18	4.09E-15	3.20E-16
PG	1950	45.1	109	1.37E-16	9.83E-14	7.15E-15
O-PI	964	22.3	70	1.54E-16	1.10E-13	7.48E-15
Cer	1610	37.4	96	1.98E-16	1.42E-13	9.06E-15
PC	1980	45.9	106	6.48E-15	4.64E-12	2.79E-13
EpODEs	10	0.232	8	3.57E-12	2.55E-09	1.38E-10
HOTrEs	7	0.162	7	3.58E-12	2.55E-09	1.38E-10
O-PG	964	22.3	60	1.34E-11	9.52E-09	4.88E-10
PS	1970	45.7	95	4.95E-11	3.52E-08	1.72E-09
DGDG	1730	40.1	86	9.13E-11	6.48E-08	3.03E-09
DHCer	697	16.2	47	1.55E-10	1.10E-07	4.92E-09
Prostaglandins	199	4.61	22	2.01E-09	1.42E-06	6.11E-08
LacCer	4580	106	168	8.11E-09	5.73E-06	2.37E-07

Table 5.4.3.2. 2. The complete enhanced metabolomic content of the Ovarian sample by IPA-based LC-MS.

Lipid classes	total	expected	hits	Raw p	Holm p	FDR
HexCer	8690	201	765	4.52E-226	3.31E-223	3.31E-223
Cer-1-P	2290	53.1	321	1.77E-147	1.29E-144	6.48E-145
PE-Cer	2290	53.1	197	7.18E-55	5.23E-52	1.75E-52
SM	2320	53.7	192	3.70E-51	2.69E-48	6.76E-49
O-PE	1100	25.4	128	3.61E-50	2.62E-47	5.27E-48
O-PA	965	22.4	113	1.59E-44	1.15E-41	1.93E-42
O-PC	1040	24.2	110	4.53E-39	3.29E-36	4.73E-37
TAG	27100	627	888	2.31E-27	1.67E-24	2.11E-25
PE	1970	45.8	126	1.31E-23	9.44E-21	1.06E-21
O-DAG	1270	29.5	94	3.53E-22	2.55E-19	2.58E-20
MGDG	1730	40.1	113	4.67E-22	3.37E-19	3.10E-20
PhytoCer	805	18.7	66	2.81E-18	2.02E-15	1.71E-16
PI	1960	45.5	113	5.69E-18	4.09E-15	3.20E-16
PG	1950	45.1	109	1.37E-16	9.83E-14	7.15E-15
O-PI	964	22.3	70	1.54E-16	1.10E-13	7.48E-15
Cer	1610	37.4	96	1.98E-16	1.42E-13	9.06E-15
PC	1980	45.9	106	6.48E-15	4.64E-12	2.79E-13
EpODEs	10	0.232	8	3.57E-12	2.55E-09	1.38E-10
HOTrEs	7	0.162	7	3.58E-12	2.55E-09	1.38E-10
O-PG	964	22.3	60	1.34E-11	9.52E-09	4.88E-10
PS	1970	45.7	95	4.95E-11	3.52E-08	1.72E-09
DGDG	1730	40.1	86	9.13E-11	6.48E-08	3.03E-09
DHCer	697	16.2	47	1.55E-10	1.10E-07	4.92E-09
Prostaglandins	199	4.61	22	2.01E-09	1.42E-06	6.11E-08
LacCer	4580	106	168	8.11E-09	5.73E-06	2.37E-07
DAG	1840	42.6	78	4.62E-07	0.000326	1.30E-05

Table 5.4.3.2. 3. The complete enhanced metabolomic content of the Ovarian sample by MTBE-based LC-MS.

Lipid Classes	total	expected	hits	Raw p	Holm p	FDR
HexCer	8690	157	611	1.17E-184	8.56E-182	8.56E-182
Cer-1-P	2290	41.5	222	8.59E-91	6.27E-88	3.14E-88
O-PE	1100	19.8	114	3.51E-50	2.56E-47	8.56E-48
TAG	27100	490	772	1.10E-38	8.02E-36	2.01E-36
SM	2320	41.9	142	1.88E-35	1.36E-32	2.74E-33
O-PC	1040	18.9	91	2.93E-34	2.13E-31	3.57E-32
O-PA	965	17.5	87	6.73E-34	4.88E-31	7.03E-32
PE-Cer	2290	41.5	127	8.94E-28	6.47E-25	8.17E-26
PI	1960	35.6	100	1.38E-19	9.98E-17	1.12E-17
O-PI	964	17.4	64	1.95E-18	1.40E-15	1.42E-16
MGDG	1730	31.3	85	5.71E-16	4.12E-13	3.79E-14
PE	1970	35.7	91	2.23E-15	1.61E-12	1.36E-13
O-DAG	1270	23	66	8.41E-14	6.05E-11	4.73E-12
PG	1950	35.2	85	3.38E-13	2.43E-10	1.77E-11
EpODEs	10	0.181	8	4.97E-13	3.56E-10	2.42E-11
HOTrEs	7	0.127	7	6.32E-13	4.52E-10	2.89E-11
Cer	1610	29.2	72	7.47E-12	5.34E-09	3.21E-10
PS	1970	35.6	82	9.41E-12	6.72E-09	3.82E-10
Prostaglandins	199	3.6	22	2.04E-11	1.45E-08	7.84E-10
PC	1980	35.8	81	2.89E-11	2.05E-08	1.05E-09
O-PG	964	17.4	48	7.04E-10	5.00E-07	2.45E-08
DAG	1840	33.3	69	2.43E-08	1.72E-05	8.06E-07
LacCer	4580	82.9	126	3.98E-06	0.00282	0.000127
PA	1950	35.3	62	2.27E-05	0.0161	0.000691
O-PS	984	17.8	37	3.64E-05	0.0257	0.00103

5.5 Discussion and Conclusion

The work presented in this chapter concluded that this was the first time in the literature that DESI-MSI technology had been combined with an LCM system to perform follow-up gene expression analysis or LC-MS using the same single section. The different technical optimisations for an operational workflow had been standardised. Three main findings had been identified: i) an optimised RNAse free H&E staining and fast run DESI-MSI setup allow conserving RNA quantities and quality so that the same tissue section can be used after DESI-MSI analysis, and ii) an established cold stage set up for DESI-MSI analysis to prevent RNA degradation less rather than room temperature analysis and

isolate more RNA amount from the post-DESI tissue section, and iii) an enhanced IPA-based lipid extraction protocol for microdissected tissue areas to identify more metabolites using LC-MS analysis.

RNA quantities may be conserved, and the same tissue slice can be utilised after DESI-MSI analysis with the use of optimised RNAse-free H&E staining and a fast-run DESI-MSI setup. When executing a fast run set-up (10 scan/s versus 2.5 scan/s), the signal intensity and number of peaks recorded from pixels were lowered by 10%. These findings revealed that when the faster scan rate was used, some spectral information was lost, resulting in quicker imaging time than a standard examination of each pixel area on the tissue. Despite this, some advantages to adopting a high-speed analysis include saving instrumental time.

More notably, this was the first time RNA was isolated from the same slide previously imaged by DESI. If the DESI-MSI analysis time is short enough, such as 30 minutes, the RNA material from the same slides can still be used for gene expression analysis such as rt-PCR. A viable methodology for multi-modal analysis using cold stage DESI-MS imaging has been devised for the Xevo G2-XS QToF. While there is no significant difference in RIN quality between samples tested at cold (-20°C) and room temperature, the total extracted RNA yield is higher. Total duration of 30 minutes or less appears to be desirable for higher RNA production and quality. RNA of sufficient quality and quantity has been acquired for the next phases, which include NGS analysis. As a future direction, transcriptomics analysis will be completed with microdissected ROIs from multi-site biopsies, which will aid in revealing the inter-tumour heterogeneity profile in the following chapter.

Conclusively for LC-MS micro-lipidomics characterisation, the sample extraction from microdissected areas using IPA demonstrated a suitable methodology to cover the molecular profile for lipidomics analysis, as previously published articles[268,269] indicated. Sample preparation protocols should be carefully optimised based on the attributes and needs of the target molecular species, such as structure, hydrophobicity/hydrophilicity etc. As a result, this methodology provides a valuable strategy for examining localised lipidomics and co-registering DESI-MSI and LC-MS datasets on the same fresh-frozen tissue areas.

This workflow allowed for a thorough metabolic, transcriptomic, and histological study of a single tissue section strategy, which was especially useful for restricted clinical samples. LCM, combined with DESI-MSI, appeared to have the potential to gain more in-depth knowledge of the regions of interest from the same tissue section by integrating several techniques. These multi-omics analyses could be utilised to comprehensively identify patient samples by providing molecular information from lipids, metabolites, and RNA. Using the same material for RNA research and DESI-MS imaging, especially in clinical studies, enables a more reliable and thorough diagnosis. It could open the path for tailored therapy by profiling tumours in several dimensions on the same ROIs.

5.6 Appendix

Table 5.6.1. Across all conditions and replicates, cs: cold stage, rt: room temperature; 2.5ss/10ss: scan numbers per second, pen: PEN glass slide, glass: Standard glass slide. The identified top 35 m/z values were compared in Hierarchical clustering for comparing different conditions in the Cold versus Room Temperature Stage of the DESI-MSI.

m.z	cs-2.5ss-pen-1 fold change	cs-2.5ss-pen-2 fold change	cs-2.5ss-pen-3 fold change	rt-10ss-pen-1 fold change	rt-10ss-pen-2 fold change	rt-10ss-pen-3 fold change	rt-2.5ss-pen-1 fold change	rt-2.5ss-pen-2 fold change	rt-2.5ss-pen-3 fold change
265.1473409	16601.82819	23369.75986	35287.83049	2695.163679	6707.744061	7446.157498	51265.26658	29880.52835	20411.88864
201.0374973	5136.438523	7883.950038	10821.66306	1520.662558	1137.432657	1494.233917	9368.991518	7518.252235	5248.22327
885.5484964	7070.335203	6288.415553	7398.353724	1296.702988	1411.347222	1794.028394	6489.944041	7189.864368	6129.757986
887.5612991	5340.908135	4727.708374	5563.513505	891.8654528	1036.454039	1286.518523	4963.951972	5805.505619	4914.985027
175.0242301	1581.811477	2888.459759	3740.565938	830.9797386	721.8125	914.3087844	3616.555519	3659.345977	2538.948758
886.5528996	3468.201496	3132.490171	3693.474876	648.5593838	695.8815789	884.0488021	3151.406808	3568.220562	3040.746894
766.5363007	2861.973469	2547.547518	3110.280536	648.0830065	604.5769371	761.6093611	2585.734743	2604.833716	2123.694876
722.5116979	2935.370968	2492.605192	3060.07567	636.105042	548.0469664	699.0787489	2183.740141	2063.209195	1731.630878
124.0069291	764.6811594	2053.12762	2555.877458	552.1518207	470.7044956	619.3014641	3320.754792	3390.863985	2555.022072
528.27108	348.657784	492.1857283	704.845283	436.3132586	326.7738487	421.2168367	1168.896673	1091.990805	798.5217391
750.5445715	1900.106942	1692.198219	2059.776961	435.8512605	377.3908991	473.7998004	1443.496585	1481.976501	1222.583075
748.5282552	1839.385577	1577.162376	1972.987587	415.073296	363.6582602	461.7399068	1482.986451	1405.183142	1165.833629
744.5516465	2099.979897	1842.42598	2122.735253	415.0187675	392.2222222	503.8830967	1970.299405	2457.687612	1808.048026
888.5642848	2353.71108	2124.59357	2465.373784	401.8515406	457.0861659	570.8399512	2170.424873	2580.298851	2200.874335
266.1500834	2088.581931	2851.334963	4258.815789	356.5771242	795.0611294	913.6011535	5799.696189	3748.581226	2510.928682
768.5519771	1607.466456	1460.699359	1759.047269	349.1933707	331.7823465	417.7904836	1492.027649	1694.608685	1305.946761
746.5140737	1606.698457	1396.080591	1731.291857	347.2968254	304.8866045	389.7607587	1252.181207	1177.525926	988.342835
742.5410729	1553.066737	1382.112523	1669.581231	346.8606909	323.2709247	409.0272848	1422.619079	1606.215709	1249.989574
788.5410725	1358.157083	1374.983274	1485.659484	339.3309991	356.2036732	435.8011313	1465.23937	1899.660536	1596.008651
187.0413877	54.83228144	426.6768763	610.7889772	317.4084967	260.0131579	399.9621783	2006.036021	1290.244828	1038.25976
716.5213916	1370.200795	1218.572065	1431.617676	309.6764706	295.6690424	374.8232032	1250.8026	1305.03129	1097.443988
767.5439017	1320.978962	1201.210709	1449.071897	307.2573296	284.9749635	361.9210293	1183.930381	1221.734227	1003.619676
723.5203899	1301.138733	1131.492886	1376.267329	291.1867414	253.0285088	322.7684117	987.3174708	955.444189	800.8317436
377.1604954	1557.164563	2223.369501	3984.874181	277.194211	502.735106	613.8778838	3611.87332	1923.022989	1264.473824
747.5161923	1057.377396	959.5904203	1191.586892	260.2001867	228.9489218	291.1363132	901.0601454	860.6057471	702.516748
195.0504079	523.4354839	743.7594222	872.1979146	259.3010271	213.4689327	269.3644632	936.1245869	1040.541635	729.3787711
572.4813828	400.2734923	474.9636146	615.9268123	252.1612512	181.9998173	221.8417258	680.3207755	419.9550447	398.8772183
700.5294202	1086.751519	926.7068535	1079.481927	211.8366947	195.4874817	250.998669	782.3886319	747.9970626	694.5571207
790.5416817	874.8980832	825.0266102	992.5370407	208.5827264	192.7306287	240.9444321	789.062679	902.8895275	729.2788376
764.5222069	837.3177887	731.5014663	927.4076465	206.6689076	185.9154788	232.2547693	777.0034148	742.3347382	608.6634871
751.5478421	859.4115241	777.2804388	935.748858	205.0696545	176.8893458	223.2975821	665.5949548	690.9624521	573.0552351
734.5363521	599.693899	593.1156729	759.34429	195.096732	160.2987025	202.5391526	654.1899097	522.0942529	461.4409938
859.5290658	940.9953249	887.9790377	1071.5857	190.402521	201.4608004	249.9932343	892.0117867	1039.894764	865.1629326
745.560703	920.4740533	822.9359183	940.3639523	189.7129785	178.1670322	229.5544587	862.6151135	1083.769732	808.3625776

Table 5.6.1.continuous, The identified top 35 m/z values were compared in Hierarchical clustering for comparing different conditions in the Cold versus Room Temperature Stage of the DESI-MSI.

m.z	cs-10ss-glass-1	cs-10ss-glass-2	cs-10ss-glass-3	cs-10ss-pen-1	cs-10ss-pen-2	cs-10ss-pen-3	cs-2.5ss-glass-1	cs-2.5ss-glass-2	cs-2.5ss-glass-3
	fold change	fold change	fold change	fold change	fold change	fold change	fold change	fold change	fold change
265.14734	5.771885522	9.201137467	1927.218406	4881.779	5240.53083	4865.48272	1.154377104	1.154377104	1.154377104
201.0375	1363.914036	2296.105008	1698.184512	1479.16104	1323.86374	1343.30531	9723.452556	10360.52673	10360.52673
885.5485	882.7092803	1218.963934	1062.250168	2386.60303	1867.78788	1418.22099	8761.684725	14578.2165	14578.2165
887.5613	633.9681187	831.43196	779.587991	1637.66679	1268.61353	945.388889	7059.142226	11777.10929	11777.10929
175.02423	635.6960227	920.6702733	737.2732884	1030.79879	787.147833	694.518025	3105.665369	5535.659029	5535.659029
886.5529	453.944234	621.524622	539.8771044	1162.83221	919.779461	712.972346	4435.749527	7227.831117	7227.831117
766.5363	494.4877946	724.0539603	588.8766554	1091.20382	930.677189	754.761235	4649.474332	6633.657237	6633.657237
722.5117	485.37742	736.4469413	542.8007856	1253.47956	1085.6956	880.715309	4565.718914	7057.31108	7057.31108
124.00693	453.3746843	747.4161465	580.8159371	559.047518	484.426347	477.718889	3995.74595	4428.173692	4428.173692
528.27108	297.1163721	469.5687335	367.4043771	138.334044	114.236427	119.237531	518.0989901	953.922678	953.922678
750.54457	335.266835	511.5020114	368.7022447	863.10096	740.271465	607.544691	3037.306964	4462.590005	4462.590005
748.52826	319.5427189	480.1070884	367.3656566	751.007228	632.666141	505.359877	2988.980644	4482.376426	4482.376426
744.55165	308.5631313	452.0077681	397.6216611	647.971916	557.1492	469.582963	4521.934673	5163.877322	5163.877322
888.56428	292.2148569	378.1856013	356.6063973	713.626022	558.381103	424.018025	3216.141279	5256.128147	5256.128147
266.15008	71.31542484	71.31542484	71.31542484	635.43216	669.369003	626.061728	71.31542484	71.31542484	71.31542484
768.55198	270.9707492	390.5961992	329.8227834	558.041711	474.138152	386.646173	2862.340732	3801.04479	3801.04479
746.51407	267.1959175	402.2528783	308.4295174	650.62543	566.125947	459.553827	2542.907532	3856.501744	3856.501744
742.54107	271.2648359	393.4079623	331.8910213	544.405854	458.908565	382.124198	2734.753734	3656.562376	3656.562376
788.54107	237.3559554	337.9955611	274.1161616	556.384524	420.59154	332.766296	1816.724595	2366.639321	2366.639321
187.04139	239.0909091	330.0772645	295.6979798	10.9664563	10.9664563	10.9664563	249.9763307	378.3439887	378.3439887
716.52139	243.7497896	339.5448745	303.7952862	485.452186	416.523569	345.435679	2316.45445	3273.02414	3273.02414
767.5439	238.8055556	345.8044112	282.8744108	503.079156	432.676136	358.557901	2164.844835	3058.448373	3058.448373
723.52039	227.3255471	338.8708559	254.1841751	562.265908	487.378998	398.941235	2088.806228	3173.884583	3173.884583
377.1605	55.4388422	55.4388422	55.4388422	401.534187	522.655619	563.228272	55.4388422	55.4388422	55.4388422
747.51619	198.8174453	301.6888611	234.6481481	438.621401	382.076915	309.418519	1701.72975	2683.365865	2683.365865
195.05041	214.3506944	320.8077403	264.3230079	261.3959	207.418035	205.980741	780.300021	1168.976143	1168.976143
572.48138	178.5651305	251.7257595	175.6868687	230.517597	212.920455	164.520247	114.8723964	174.1698256	174.1698256
700.52942	167.2805135	239.3841032	193.9023569	440.728641	379.978009	318.863827	1682.454871	2581.485149	2581.485149
790.54168	159.6419402	235.478291	187.4041526	332.796421	280.454861	233.238889	1424.792447	1988.785856	1988.785856
764.52221	163.5420875	232.3395755	193.0430976	313.345776	262.575442	214.761975	1321.064801	1918.916737	1918.916737
751.54784	161.3357534	241.91469	177.8869809	394.649129	342.462542	283.207161	1400.608879	2020.728901	2020.728901
734.53635	142.6742424	205.2532945	168.8518519	218.254651	202.763573	173.458148	545.3492531	964.5920792	964.5920792
859.52907	141.7443182	189.5971702	170.559596	313.067781	249.311448	195.627161	1322.000842	2123.501556	2123.501556
745.5607	144.3519571	208.3886808	184.4709315	290.015286	249.858375	213.837901	1984.449295	2253.614993	2253.614993

Table 5.6.2. The identified top 35 m/z values for the Heatmap of DESI-MSI clustering analysis to reveal metabolic segments for ROIs microdissection. Seven metabolic clusters were identified between six consecutive sections, with highly colocalised ions identified using ion-pair Spearman's correlation and hierarchical clustering (average linkage). |log2(F.C.)| > 1 and p-value < 0.05 (distance measure using the Euclidean, and clustering algorithm using average).

m.z	1	2	3	4	5	6	7
	fold Change	fold Change	fold Change	fold Change	fold Change	fold Change	fold Change
89.02419839	2.169568094	1.991515344	1.71096385	2.074883575	1.727305446	2.152372259	2.473799948
92.92811544	0.016206795	0.005385146	0.00088457	0.007127684	0.041565692	0.039326979	0.03745603
94.925216	0.00551904	0.000552914	-0.000514761	0.003748761	0.147052545	0.010949263	0.030433653
102.0561328	0.473255011	0.567830561	0.604682827	0.456855512	0.348966508	0.464920885	0.802397457
124.0074304	2.596402599	2.59937799	2.44290301	2.563201425	2.607523779	2.727090567	2.948142555
128.0353978	1.968217241	1.971103381	2.030643854	1.978378356	1.808551542	1.859642084	2.324024127
130.0511174	0.749771926	1.009486837	1.275588567	0.784108134	0.728782133	0.606855131	0.801348172
133.0145646	0.31169154	0.291010606	0.361162501	0.372183449	0.264874481	0.366107702	0.38012813
134.0472612	0.007747454	0.010132887	0.004870764	0.002499811	0.221972986	0.014455454	0.126921789
135.0312641	2.351021448	2.152583612	2.048111248	2.318137169	2.500774868	2.489186618	2.614042993
138.0200529	1.039161648	0.717538426	0.576588436	1.012230356	1.129244139	1.24097393	1.135718453
142.9986452	2.433262388	2.404394783	2.259221393	2.403731593	2.463134631	2.490745036	2.684930754
145.0617915	1.846850621	1.749167936	1.61986709	1.879229657	1.547670919	1.88470424	2.344763061
146.0459273	2.454989019	2.519580298	2.506184393	2.47105751	2.174894267	2.455690294	2.661940216
146.9833446	0.573051091	0.225454866	0.139710654	0.539222468	0.66351287	0.831002628	0.969188188
151.0265915	1.237120375	0.797784604	0.84759989	1.268852174	1.256395787	1.307046412	1.605949004
154.0622344	0.573375515	0.3827877	0.297666071	0.481682127	0.397314843	0.526905125	1.088291598
167.021446	0.716425095	0.556221461	0.513967125	0.60558711	0.786740206	0.961460474	1.307864563
171.1027561	0.651678431	0.496292424	0.353534443	0.585810091	0.296035599	0.865674996	0.62800543
174.0407982	2.65239422	2.745408528	2.790509153	2.690222248	2.379837686	2.510896429	2.750281873
175.0249176	2.67379923	2.810596152	2.499357306	2.229007121	2.229557559	2.343448264	2.950222114
181.0721662	0.234850144	0.149253701	0.088103567	0.27491984	0.297141947	0.362734729	0.619537646
186.0455341	2.405346807	2.365228616	2.397729441	2.520814507	2.719476886	2.515919942	2.601583243
187.0419619	3.061737014	2.987316733	3.00660712	3.206091212	3.481834967	3.244311851	3.341599102
187.0978876	0.706581063	0.598745692	0.387294153	0.652168323	0.654166001	0.7874992	0.883129309
188.0456208	0.929469239	0.842278347	0.909829433	1.269481833	1.520197536	1.136100307	1.292915845
195.0509943	2.673447393	2.639740138	2.540039479	2.687988732	3.030247396	2.870840568	3.095647058
196.054942	0.168008689	0.165833643	0.11371531	0.150780719	0.495542426	0.235672035	0.627864029
201.0381807	2.559035481	2.30890104	2.087343427	2.513770923	2.734248255	2.781360568	2.87245065
201.1135171	0.249819976	0.227635704	0.082808897	0.151872046	0.397165742	0.393306944	0.471623045
208.9842384	0.249787318	0.215348834	0.129851897	0.204982915	0.670088343	0.488465431	0.834931605
214.048648	0.699165303	0.742923273	0.693653815	0.919238127	0.489268776	0.697712255	1.233715584
215.1290449	0.280315231	0.195060656	0.067960187	0.14810117	0.233089766	0.361316436	0.364674767
217.0124353	0.580321402	0.384038132	0.216813972	0.469979133	0.627810298	0.623163052	1.284520597

Table. 5.6.3. Comparison of two lipid extraction methods, IPA and MTBE-based, for LC-MS micro-lipidomics characterisation. a) Heatmap displaying the grouping results for all peaks (distance measure using Euclidean, and clustering process using average) dissected three regions(IPA-C2/C3/C7 and MTBE-C2/3/7) and benchmark controls (IPA-T and MTBE-T), which are bulk tissue sections and PEN membrane controls, respectively (IPA-P and MTBE-P).

m.z	IPA-C2	IPA-C3	IPA-C7	IPA-T	IPA-P	MTBE-C7	MTBE-C3	MTBE-C2	MTBE-T	MTBE-P
	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change	Fold Change
696.817023	3827391.35	2989691.19	3038080.28	654486.671	31.8872154	31.8872154	159.436077	31.8872154	31.8872154	31.8872154
557.494342	2313694.88	1330262.29	986476.894	909441.278	69.3155856	346.577928	61341.9863	69.3155856	669.56659	69.3155856
857.557082	2196430.85	1121100.45	742734.447	2604802.35	8.87073142	8.87073142	8.87073142	44.3536571	8.87073142	8.87073142
654.660621	1599007.9	1875316.51	2144686.95	1186292.69	237258.538	237258.538	237258.538	237258.538	237258.538	237258.538
473.433749	1562131.25	1139108.7	1022876.64	873983.821	174796.7642	174796.7642	174796.7642	174796.7642	174796.7642	174796.7642
775.610502	1532117.98	949844.914	1823770.19	1985057.51	58.6664166	58.6664166	293.332083	58.6664166	58.6664166	58.6664166
774.540527	1408817.32	1753239.16	2400020.46	1720247.7	62.2849146	311.424573	233842.618	62.2849146	27671.6478	62.2849146
690.56185	1404450.46	2003398.41	698275.111	595515.81	119103.162	119103.162	119103.162	119103.162	119103.162	119103.162
283.551583	1317662.79	1844670.19	2759615.53	1432913.41	3532.51082	3532.51082	17662.5541	3532.51082	3532.51082	3532.51082
283.655674	1316333.03	2792272.56	2244541.64	2115106.16	263266.606	263266.606	263266.606	263266.606	263266.606	263266.606
694.639015	1204286.62	2890943.09	1255164.19	2016266.82	240857.324	240857.324	240857.324	240857.324	240857.324	240857.324
414.875333	1113252.99	1143539.1	1310921.53	535608.049	107121.6098	107121.6098	107121.6098	107121.6098	107121.6098	107121.6098
529.501729	1102377.81	808902.314	1011025.07	819841.31	167.528845	167.528845	167.528845	167.528845	837.644225	167.528845
696.477513	1094538.67	1632547.55	1249092.45	1176274.95	3685.5076	3685.5076	18427.538	3685.5076	3685.5076	3685.5076
476.443238	1025845.98	1042039.02	1156397.32	974237.825	69.5078	347.539	69.5078	69.5078	69.5078	69.5078
283.456557	872664.547	1089831.95	2418708.83	1185871.73	2180.33384	2180.33384	10901.6692	2180.33384	2180.33384	2180.33384
530.504889	719355.485	546542.746	692658.791	549372.796	109308.5492	109308.5492	109308.5492	109308.5492	109308.5492	109308.5492
677.465846	572575.98	438288.429	474857.295	297551.517	59510.3034	59510.3034	59510.3034	59510.3034	59510.3034	59510.3034
1242.64458	554397.773	233448.705	692212.832	280112.511	46689.741	46689.741	46689.741	46689.741	46689.741	46689.741
729.60354	540927.758	833036.073	588656.959	618182.78	108185.5516	108185.5516	108185.5516	108185.5516	108185.5516	108185.5516
771.633824	499822.473	1458394.8	572411.66	669402.403	104.0705492	104.0705492	104.0705492	520.352746	1031.7656	104.0705492
311.189494	485815.284	633371.42	447080.333	727891.449	89416.0666	89416.0666	89416.0666	89416.0666	89416.0666	89416.0666
678.46787	476020.971	367275.381	368504.342	131481.348	26296.2696	26296.2696	26296.2696	26296.2696	26296.2696	26296.2696
844.613584	454585.986	514845.906	194521.632	663381.642	38904.3264	38904.3264	38904.3264	38904.3264	38904.3264	38904.3264
858.560211	452108.377	257006.522	151193.84	577999.26	27.9042806	27.9042806	27.9042806	27.9042806	139.521403	27.9042806
700.788277	451769.515	436449.015	156853.692	423660.597	1492.787716	1492.787716	7463.93858	1492.787716	1492.787716	1492.787716
456.878191	379154.04	260907.405	337766.101	259286.833	10288.37936	10288.37936	51441.8968	10288.37936	10288.37936	10288.37936
767.546214	378433.614	208921.395	441176.845	478191.246	45.288338	45.288338	45.288338	45.288338	226.44169	45.288338
473.433936	374864.21	817336.558	368555.487	241068.54	1.993971316	1.993971316	9.96985658	1.993971316	1.993971316	1.993971316
820.759067	365264.144	419205.915	465379.156	1553242.83	73052.8288	73052.8288	73052.8288	73052.8288	73052.8288	73052.8288
540.823533	347510.333	536256.636	299632.577	199564.241	275.130422	275.130422	1375.65211	275.130422	275.130422	275.130422
710.670513	300275.966	519413.718	672249.722	356788.2	60055.1932	60055.1932	60055.1932	60055.1932	60055.1932	60055.1932
298.267042	289425.365	140769.504	423580.215	2863.0907	572.61814	572.61814	572.61814	572.61814	572.61814	572.61814
314.925996	271108.224	253056.006	228898.695	223719.153	44743.8306	44743.8306	44743.8306	44743.8306	44743.8306	44743.8306

Table 5.6.4. Details of Relative quantification of the housekeeping gene *GAPDH* with microdissected tumour and stroma samples via rt-PCR using gene of interest *TP53, LPIN1, CEL,* and *PIK3CA* expression.

Stroma				Tumour			
Replicates	**1**	**2**	**3**	**Replicates**	**1**	**2**	**3**
TP53	30.72	31.07	30.3	**TP53**	30.22	29.89	29.97
normalisation	1.00076012	1.01216201	0.987077859	**normalisation**	1.006438721	0.99544849	0.998112789
GAPDH	30.54	30.61	28.58	**GAPDH**	24.66	24.28	24.1
normalisation	1.02106319	1.023403544	0.955533266	**normalisation**	1.01286966	0.997261774	0.989868565
delta dt	-0.02030306	-0.011241529	0.031544592	**delta dt**	-0.006430939	-0.001813284	0.008244223
RQ	1.014172502	1.007822471	0.978372262	**RQ**	1.004467537	1.001257663	0.994301836

Replicates	**1**	**2**	**3**	**Replicates**	**1**	**2**	**3**
PIK3CA	30.87	31.07	31.84	**PIK3CA**	27.43	28.28	27.95
normalisation	0.987523992	0.993921945	1.018554063	**normalisation**	0.983624193	1.01410471	1.002271097
GAPDH	30.54	30.61	28.58	**GAPDH**	24.66	24.28	24.1
normalisation	1.02106319	1.023403544	0.955533266	**normalisation**	1.01286966	0.997261774	0.989868565
delta dt	-0.033539197	-0.029481599	0.063020796	**delta dt**	-0.029245467	0.016842935	0.012402532
RQ	1.023519932	1.020645313	0.957257659	**RQ**	1.020478274	0.988393251	0.991440066

Replicates	**1**	**2**	**3**	**Replicates**	**1**	**2**	**3**
CEL	30	30.35	31.12	**CEL**	30.84	30.37	0
normalisation	0.983929157	0.995408331	1.020662512	**normalisation**	1.007678484	0.992321516	0
GAPDH	30.54	30.61	28.58	**GAPDH**	24.66	24.28	24.1
normalisation	1.02106319	1.023403544	0.955533266	**normalisation**	1.01286966	0.997261774	0.989868565
delta dt	-0.037134032	-0.027995213	0.065129246	**delta dt**	-0.005191177	-0.004940258	0
RQ	1.026073467	1.0195943	0.955859682	**RQ**	1.003604731	1.003430196	

Replicates	**1**	**2**	**3**	**Replicates**	**1**	**2**	**3**
LPIN1	31.12	30.54	30.61	**LPIN1**	28.37	28.14	28.38
normalisation	1.011813157	0.992955457	0.995231386	**normalisation**	1.002591589	0.919457605	1.017690653
GAPDH	30.54	30.61	28.58	**GAPDH**	24.66	24.28	24.1
normalisation	1.02106319	1.023403544	0.955533266	**normalisation**	1.01286966	0.997261774	0.989868565
delta dt	-0.009250033	-0.030448087	0.03969812	**delta dt**	-0.010278071	-0.077804169	0.027822087
RQ	1.006432232	1.021329291	0.972858494	**RQ**	1.007149654	1.055410447	0.980899961

6 The mapping of tumour heterogeneity in multi-site collected biopsies of ovarian tumours using single tissue section methodology

6.1 Background

Ovarian cancer (OC) consists of a collection of neoplasms with distinctive clinicopathological molecular and prognostic characteristics[9]. Even though there are several forms of ovarian cancer, they are always treated as a single disease[18]. It has been shown that[25,241,272] cells within a tumour exhibit metastatic potential by accumulating additional mutations and metabolic changes that differ from one area to another. These cells can evolve independently of the leading clone during carcinogenesis and undergo clonal evolution. Tumours are thought to arise from a single mutated cell that accumulates additional mutations as it grows and divides, leading to production subpopulations with different mutations. According to studies[58–60,182,273] for the concept of cancer evolution, some sub-clones acquire an evolutionary advantage of being resistant to chemotherapy and dominating the tumour, resulting in a specific clonal expansion. Collecting a single biopsy for a routine histopathology assessment does not represent the whole tumour[58]. Several studies[59,60] have shown that OC has a high degree of intra-tumour heterogeneity (ITH), which causes a critical challenge in designing effective therapy strategies. Thus, mapping the OC with multi-region collected biopsies, whether for clonal evaluation or tumour heterogeneity, might explain the development of chemotherapy resistance. This process could determine patient-specific prognostic markers and improve personalised therapy.

High-throughput omics technologies such as transcriptomics and metabolomics have helped investigate the complexity of the OC by providing comprehensive knowledge of oncogenic pathways and metabolites that are potentially targetable. The classic histological classification of the subtypes is an important diagnostic tool and has improved the treatment strategies.

Collecting and evaluating multi-site biopsies can provide an excellent way to assess how the metabolic profiling of spatially-separate biopsies differs within the same ovarian tumour. The aim of this chapter was to optimise a multi-modal analysis workflow within a single fresh frozen tissue section from the multi-site collected biopsies of the same cancerous ovary. DESI-MSI were conducted to investigate lipid compositions, patterns and distributions of multiple spatially-separated biopsies from the primary ovary; then, selected ROIs defined by DESI-MSI clustering analysis were microdissected for gene expression analysis. Therefore the co-registration of DESI-MSI and transcriptomics information in the same MSI-guided tissue segment could assist in mapping intra tumour heterogeneity across the whole ovary. This co-registration will capture as much heterogeneity across the entire variety as possible. Hence, the multi-region molecular profiling could answer important questions such as; similarities or differences in the metabolites' and lipids' composition, patterns, and distribution among the whole cancerous ovary.

6.2 Aim and Objectives

The proposed aim of this chapter's research would be whether spatial metabolomics using DESI-MSI technology can be extended to guide spatial transcriptomics using a single tissue section strategy. To test this methodology, the comparison of multi-site collected biopsies within the same patient and among different patients was utilised to capture inter and intra-tumour heterogeneity. For that purpose, the following are the study's precise objectives:

- To map the lipid compositions, patterns, and distributions of multiple spatially separated biopsies from primary ovarian tumours using DESI-MSI technology, then employ this metabolic profile in DESI-MSI segmentation analysis on the tissue section to unravel inter/intra-tumour heterogeneity.
- To evaluate microdissection of metabolic segments defined by DESI-MSI from post-DESI tissue sections to see if extracted RNA material can be adequate for meeting the standards of DESI-MSI guided spatial transcriptomics.
- To perform network analysis to determine whether there is a link between metabolically enriched signatures and genetic phenotypes in each revealed DESI-MSI segments of spatially separated biopsies from the same ovarian tumour.

6.3 Summary workflow of Experimental Plan

6.4 Results

6.4.1 A pilot study for RNA sequencing optimisation steps

The work presented in this chapter was to examine multi-site gathered biopsies from the same ovary to test the existing multi-modal analytic workflow in the single fresh frozen tissue section. First, the DESI-MSI data was obtained with a fast scan rate; afterwards, cluster analysis was carried out to display various segments in the tissue section representing metabolically distinct areas. Then, the DESI-MSI-guided metabolic segments of the post-DESI slides were microdissected using the LCM equipment for RNA sequencing sample preparation.

6.4.1.1 DESI-MSI clustering analysis to reveal heterogeneity in multi-site collected biopsies of multiple ovarian tumours

Following the successful optimisation of using a single tissue section for multi-modal analysis in the previous chapter, the next step was to proceed with the sequencing procedure with a pilot study to determine whether the extracted RNA from the same post-DESI tissue generated a decent transcriptome dataset. For this pilot study, 12 multi-region harvested fresh-frozen human ovarian tumour tissue biopsies from three different patients, provided by ICL Tissue Bank. DESI imaging was performed in negative ion mode at 10s/s (scan per second) and 100 µm pixel size, followed by segmentation analysis using hierarchical (average linkage) clustering to achieve specific ion-associated segments. There were five distinct metabolic clusters in **Figure 6.4.1.1. 1a** and loadings plot from the first PCA components of each patient set in **Figure 6.4.1.1. 1b**. Each cluster had a set of *m/z* value lists that represented the spatial distribution of highly colocalised ions. The fold changes (FC) of these metabolites were shown in the volcano plots of each cluster versus all clusters in **Figure 6.4.1.1. 1b-g**. The total distribution of

differentially detected metabolites was concluded using volcano plots. The total number of identified *m/z* values in the clusters was 831, with cluster 1 (red) having the most significant *m/z* variables ($|\log_2(FC)| > 1$ and p-value < 0.05) compared to the other clusters. Cluster 2 (blue) had the fewest *m/z* numbers, only five *m/z*. This detailed mapping of localised metabolic regions revealed intra- and inter-tumour heterogeneity in malignant samples from the same and different patients. Furthermore, the next subsequent tissue section was employed for H & E staining for histological evaluation.

These data demonstrated that imaging with DESI mass spectrometry could quantify the abundance of identified metabolites in a co-expression network. Ion vectors of tumour mass spectra displaying disconnected sub-networks provide a sturdy visualisation tool for molecular investigation of tumour heterogeneity.

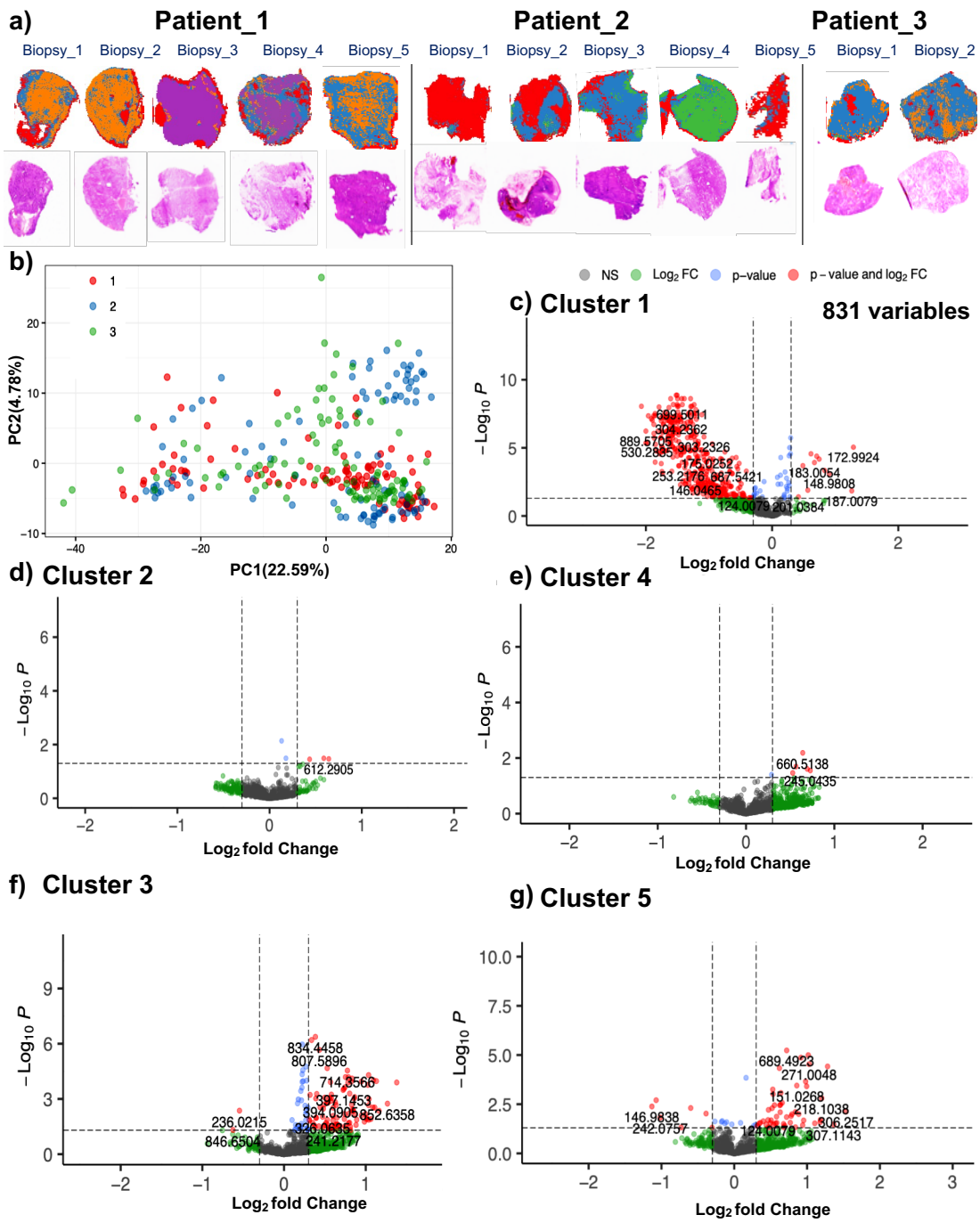


Figure 6.4.1.1. 1. Distinct metabolic segments revealed with DESI-MSI clustering Analysis. a) 5 different metabolic clusters between 12 samples-3 patients, highly colocalised ions were identified using the ion pair Spearman's correlation hierarchical clustering (average linkage). The first PC. scores, calculated using the ions belonging to each cluster, were used as the representative spatial distribution of the groups of highly colocalised ions. b) Loadings plot from first PCA components c-g) Volcano plots of 5 different clusters from different patient samples with the $|\log_2(FC)| > 1$ and p-value < 0.05.

6.4.1.2 A complete Transcriptomics from microdissected ROIs of post-DESI tissue

The DESI-MSI visualisation enabled the identification of tumour tissue sub-regions as metabolic segments, which underly biological composition, interactions, and the likely reason for its heterogeneous structure. The next step was to select the metabolic segments used for RNA isolation and sequencing. One of the samples was randomly chosen to optimise RNA sequencing procedures to determine whether the isolated RNA fulfils the ideal requirements. This perfect requirement included several assessments such as optimal quality (RIN), a sufficient isolated amount, and a high mapping percentage of the human genome's transcriptome data. The Illumina platform technology was used to process the isolated RNA samples via Novogene, a commercial sequencing company. Due to the high cost of sequencing technology, one cluster from one patient was decided for the evaluation of RNA sequencing. Post-DESI tissue section from Patient 2, biopsy 2 samples, was chosen to continue microdissection for RNA isolation; these samples exhibited two distinct clusters and a heterogeneous phenotype. In particular, clusters 1 (red) and 2 (blue) displayed a dominant profile over tissue sections and may play a key role in intra-tumour heterogeneity. The H&E image showed a histological difference between these two clusters. The blue from sample Patient 2- Biopsy 2 (OC) was microdissected in **Figure 6.4.1.2.1a,** and the quality assessment was completed in **Figure 6.4.1.2.1b**. The extracted RNA was five ng in total 10 µL volume from a 10 mm^2 dissected area, and the RIN quality was 7.6.

Exons, introns, and intergenic regions are the three types of mapped areas. When the reference genome is well-annotated, exons (Q30) should be the most common type of mapped rate. The exonic quality level (Q30) should be above 80%. The proportion of exonic readings mapped to the reference genome was 90.99 % **Figure 6.4.1.2.1b**. The complete transcriptome of sample OC was then heat mapped using FPKM cluster analysis and the log10 (FPKM+1) value in **Figure 6.4.1.2.1c**. This cluster analysis revealed the expression patterns of genes that were up- and down-regulated. Genes with high expression levels are shown in red, while genes with low levels are

highlighted in blue. The log10 (FPKM+1) value was red to blue from large to small.

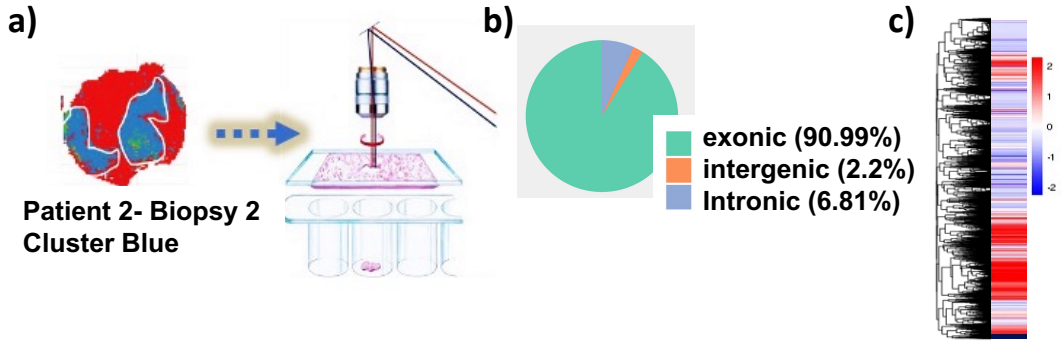


Figure 6.4.1.2.1. Transcriptome analysis of the microdissected Metabolic Segment. a) Microdissection of the clustered blue from sample Patient 2- Biopsy 2 OC.) b) Classification of Reads According to Mapped Region c) Heatmap visualisation of the overall results of FPKM cluster analysis, clustered using the $\log_{10}$(FPKM+1) value.

To gain a detailed statistical analysis of sequencing data, numerous parameters had to be checked, which are listed in **Table 6.4.1.2. 1**. Most importantly, Q30:Percentage of bases whose correct base recognition rates are more significant than 90.99% in total bases, the minimum requirement was 80%. This result indicated that the mapping percentage is high enough to get vital transcriptome information from the post-DESI tissue section.

Table 6.4.1.2. 1. Overview of data production quality of sample OC. Patient 2-Biopsy 2**,** (1) Sample: The names of samples.(2) raw_Reads: The original sequencing reads counts.(3) clean_Reads: Number of reads after filtering. (4) raw_data: Raw reads number multiply read length, saved in G unit. (5) clean_data: Clean reads number multiply read length, saved in G unit. (6) error_rate: Average sequencing error rate, which is calculated by Qphred=-10log10(e). (7) Q20: Percentages of bases whose correct base recognition rates are greater than 96%. (8) Q30: Percentage of bases whose correct base recognition rates are more significant than 90.99% in total bases

Sample	raw_reads	clean_reads	raw_data(G)	clean_data(G)	Q20(%)	Q30(%)
OC	123908212	114077595	37.2	34.2	96.07	90.99

This pilot investigation extracted a complete transcriptome data set from the microdissected area using the same fresh-frozen post-DESI section. The

same sections following DESI-MSI were employed for RNA sequencing for the first time. This optimisation proved helpful in acquiring RNA expression profiling behind the metabolic segments discovered using DESI-MSI for further samples.

6.4.2 Investigation of intra-tumour heterogeneity in multi-site collected biopsies of the same ovary

6.4.2.1 DESI-MSI clustering analysis to reveal metabolic intra-tumour heterogeneity in multi-site collected biopsies

Following the optimisation of RNA sequencing analysis with a test sample, the next step was to proceed with more patient samples. Six multi-region acquired fresh-frozen human ovarian tumour tissue samples (samples A, M, P, S, FT, L) from the same ovary were acquired then, and the DESI imaging was performed as triple replicates from sequential sections at ten scans/s and 100 µm pixel size in the negative ionisation modes (Xevo QToF, Waters). Following data acquisition, particular ion-related segments were created using hierarchical (average linkage) and k-mean clustering. There were four metabolic clusters for the – ion mode in **Figure 6.4.2.1. 1a,** as previously stated, each cluster included several lists and was used as the representative spatial distribution of highly colocalised ions. To determine the distribution of differentially detected metabolites, volcano plots were utilised to visualise those *m/z* value lists in **Figure 6.4.2.1. 1c-f**. The fold change (FC) of these up and down expressed metabolites was shown in the volcano plots of each cluster against all clusters. For this experiment, the standard threshold was |log2FC)| > 1, and the p-value was < 0.05. The total number of identified *m/z* values in the clusters was 498, with cluster 2 (blue) having the most significant *m/z* variables ($|\log_2(FC)| > 1$ and p-value < 0.05) compared to the other clusters. Cluster 4 (purple) had the fewest *m/z values*, less than 25. The heat map compares and contrasts the average intensity of the detected features identified by DESI-MSI in each cluster. Per sample set emphasise similarities and differences using hierarchical (average linkage) cluster analysis in **Figure**

6.4.2.1. 1g. Top detected *m/z* list in Appendix, **Table 6.6.1.** This thorough mapping of localised metabolic regions revealed intra- and inter-tumour heterogeneity in malignant tissues.

The intra-tumour heterogeneity in the same malignant samples was shown by localised metabolic segments utilising DESI-MSI technology. The following study revealed these metabolic segments' genetic profiling by transcriptome information.

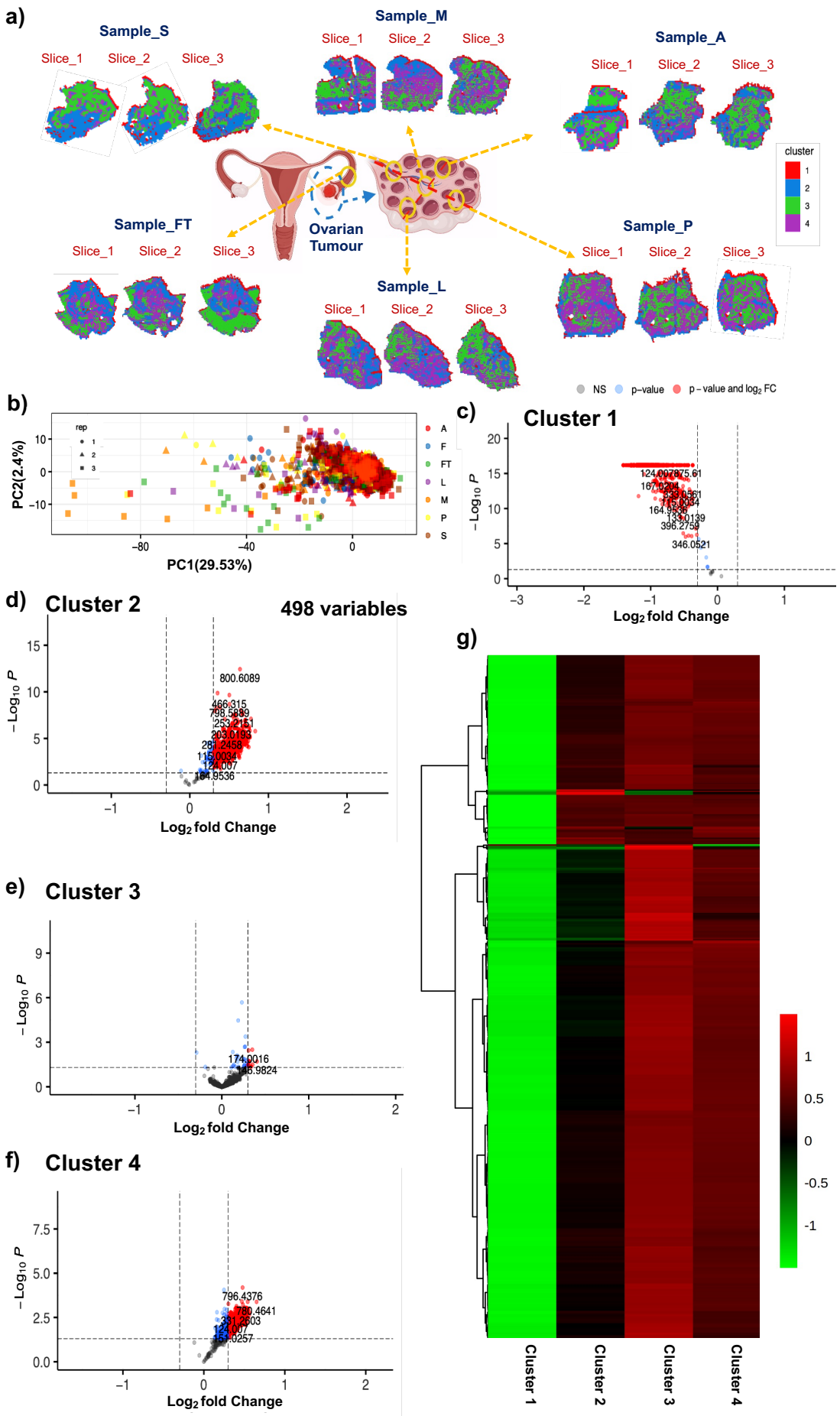

a)
Sample_S
Slice_1
Slice_2
Slice_3
Sample_M
Slice_1
Slice_2
Slice_3
Sample_A
Slice_1
Slice_2
Slice_3
cluster
1
2
3
4
Ovarian Tumour
Sample_FT
Slice_1
Slice_2
Slice_3
Sample_P
Slice_1
Slice_2
Slice_3
Sample_L
Slice_1
Slice_2
Slice_3
NS
p-value
p-value and log2 FC
b)
rep
1
2
3
A
F
FT
L
M
P
S
PC2(2.4%)
PC1(29.53%)
c)
Cluster 1
−Log10 P
Log2 fold Change
d)
Cluster 2
498 variables
800.6089
466.315
798.5889
253.2151
203.0193
281.2458
115.0034
124.007
164.9536
−Log10 P
Log2 fold Change
e)
Cluster 3
174.0016
146.9824
−Log10 P
Log2 fold Change
f)
Cluster 4
796.4376
780.4641
331.2603
124.007
153.0257
−Log10 P
Log2 fold Change
g)
1
0.5
0
-0.5
-1
Cluster 1
Cluster 2
Cluster 3
Cluster 4

Figure 6.4.2.1. 1. Revealing distinct metabolic segments with DESI-MSI clustering Analysis of multi-site collected biopsies from the same tumour. a) Metabolite distributions by DESI-MSI in three consecutive slides from ovarian cancer sample with a heterogeneous visualisation pattern within the tumour, a segmentation map of 4 clusters for – ion mode. Each cluster is represented by a different colour (red, green, blue, purple), corresponding to a group of highly colocalised ions (list at right). Other metabolic clusters between 18 samples of highly colocalised ions were identified using the ion pair Spearman's correlation and hierarchical clustering (average linkage) b) The first PC. scores, c-f) Volcano plots of 4 different clusters with the $|\log_2 FC.)| > 1$ and p-value < 0.05. (n=3) g) The Heatmap depicted the grouping results of all clusters between *m/z* values (distance measure using the Euclidean, and clustering algorithm using average).

6.4.2.2 Transcriptomics analysis to reveal intra-tumour heterogeneity in multi-site collected biopsies

In this study, the transcriptomics and metabolomics datasets were used in the MetaboAnalyst software to find possible correlations in the same MSI-defined region of tissues using a single fresh frozen tissue section methodology. First, the DESI-MSI features were identified and classified using Metabolite Set Enrichment Analysis (MSEA). Next, the information from the microdissected metabolic clusters' expressed genes enrichment analysis in KEGG was connected to a joint-pathway and network analysis to find potential vital pathways.

The quality of sequencing data had been checked and passed, and the details of RIN, total amount, and microdissected areas were displayed in **Figure 6.6.1** and **Table 6.6.2** in the Appendix. A summary of the quality results indicated that the mapping percentage is high enough to get complete transcriptome information from the post-DESI tissue section.

6.4.2.3 Comparison of Cluster 2 (blue) metabolic segments in Network Analysis

MSEA revealed physiologically important patterns in metabolomic data by providing a global picture of changed biochemical processes and identifying substantially altered compounds. This study assessed individual metabolites for their significance in each sample set in the software analysis. The compounds that had attained a high enough relevance based on the preselected parameters were combined to determine if any relevant patterns emerged. Hexosylceramides (HexCer) was the most significant lipid class (p-value < 0.05, [FDR] < 0.05) in all cluster 2 (blue, **Figure 6.4.2.3.1a**) segments based on this enriched metabolomic dataset from DESI-MSI of sample sets in **Figure 6.4.2.3.1b**. Other lipid classes included other forms of Ceramides such as LacCer, PE-Cer, PI-Cer and others in **Table 6.4.2.3. 1**.

Table 6.4.2.3. 1. The most enriched lipid class of Cluster 2 (blue). The 'total' was the total number of compounds in the lipid classes; the 'hits' was the number of matched combinations from the uploaded data; the 'raw p' is the original p-value calculated from the enrichment analysis; the' holm p' is the p-value adjusted using the Holm-Bonferroni method, and the 'FDR p' is the p-value adjusted using False Discovery Rate.

Lipid Classes	total	expected	hits	Raw p	Holm p	FDR
HexCer	8690	275	1450	0	0	0
LacCer	4580	145	576	5.65E-177	4.12E-174	2.06E-174
PE-Cer	2290	72.4	371	2.24E-148	1.63E-145	5.45E-146
PI-Cer	2290	72.3	367	2.87E-145	2.09E-142	5.24E-143
SM	2320	73.1	352	4.04E-132	2.94E-129	5.90E-130
O-PA	965	30.5	226	6.95E-125	5.05E-122	8.47E-123
O-PC	1040	33	193	1.68E-87	1.22E-84	1.75E-85
Cer-1-P	2290	72.4	267	3.60E-74	2.60E-71	3.29E-72
O-PE	1100	34.6	175	2.60E-69	1.88E-66	2.11E-67
PG	1950	61.5	219	2.23E-58	1.61E-55	1.63E-56
MGDG	1730	54.7	204	1.10E-57	7.96E-55	7.33E-56
PS	1970	62.2	204	8.15E-49	5.87E-46	4.96E-47
PA	1950	61.6	200	3.21E-47	2.31E-44	1.81E-45
O-PI	964	30.4	128	3.83E-42	2.75E-39	2.00E-40
PE	1970	62.3	191	1.33E-41	9.51E-39	6.47E-40
PC	1980	62.5	166	4.29E-29	3.07E-26	1.96E-27
HODEs	18	0.569	18	9.56E-28	6.84E-25	4.11E-26
O-PG	964	30.4	91	6.73E-20	4.81E-17	2.73E-18
PI	1960	62.1	118	5.93E-11	4.23E-08	2.28E-09
Unsaturated Fatty Acids	267	8.43	32	1.55E-10	1.10E-07	5.66E-09
Docosanoids	43	1.36	13	4.63E-10	3.29E-07	1.61E-08
MAG	71	2.24	13	3.22E-07	0.000228	1.07E-05
Prostaglandins	199	6.29	20	6.15E-06	0.00436	0.000195
DGDG	1730	54.6	81	0.000386	0.273	0.0118

The enrichment analysis could determine biological functions or pathways significantly associated with DEGs. The interactions of multiple genes may be involved in specific biological processes. To identify the pathways which had the most significant involvement with the upregulated and downregulated. DEGs were submitted for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. KEGG pathway analysis demonstrated that upregulated DEGs were significantly enriched for cluster 2 (blue) versus all clusters in the NF-kappaB (nuclear factor kappa-light-chain-enhancer of activated B cells) signalling pathway **Figure 6.4.2.3.1d.**

In the final step, the Network explorer analysis module augmented the MetaboAnalyst joint-pathway analysis module by detecting links in a gene-metabolite interaction network that correlates to features on specific pathways. This interaction allowed for a more comprehensive perspective of the pathways, which would not have been possible if they had been evaluated individually in **Figure 6.4.2.3.1c**. The most interested in the gene-metabolite pathway interaction as a representative was arachidonic acid metabolism **Figure 6.4.2.3.1c** and others **Table 6.4.2.3. 2**. This network was built using HMDB (https://hmdb.ca) and STITCH (https://www.stitchdata.com) databases and is only applicable to human studies. The graphical output contained the metabolome view with the corresponding genes/compounds presented in **Table 6.4.2.3. 3**. As seen in the joint-pathway analysis, KEGG orthologs as geneIDs in the arachidonic acid metabolic pathway were coloured, while genes upregulated in red and down-regulated in green. Yellow circles represented matched compounds to metabolites in **Figure 6.4.2.3.1c**. Additionally, a model of the network of metabolites and genes assembled in arachidonic acid metabolism is represented by blue squares and is associated with multiple genes, defined as blue circles in **Figure 6.4.2.3.1e**.

Table 6.4.2.3. 2. The crossed pathways detected in the gene-metabolite interaction network module of Cluster 2 (blue).

Pathways	Total	Expected	Hits	Raw p	Holm p	FDR
Arachidonic acid metabolism	81	2.2242	10	4.93E-05	4.3067	0.0041453
Biosynthesis of unsaturated fatty acids	47	1.2906	5	0.008435	2.0739	0.35427
Glycerolipid metabolism	35	0.96107	4	0.014331	1.8437	0.40127
Sphingolipid metabolism	58	1.5926	5	0.019982	1.6994	0.41961
Taurine and hypotaurine metabolism	16	0.43934	2	0.069597	1.1574	1
Linoleic acid metabolism	17	0.4668	2	0.077514	1.1106	1
Glycerophospholipid metabolism	86	2.3615	5	0.084671	1.0723	1
Neomycin, kanamycin and gentamicin biosynthesis	4	0.10984	1	0.10546	0.97693	1
Mucin type O-glycan biosynthesis	22	0.6041	2	0.12075	0.91812	1
alpha-Linolenic acid metabolism	22	0.6041	2	0.12075	0.91812	1
Phosphatidylinositol signaling system	74	2.032	4	0.14335	0.84361	1
Glutathione metabolism	56	1.5377	3	0.19753	0.70436	1
Nitrogen metabolism	10	0.27459	1	0.24342	0.61364	1
Inositol phosphate metabolism	69	1.8947	3	0.29388	0.53183	1
Citrate cycle (TCA cycle)	42	1.1533	2	0.32158	0.49271	1
Starch and sucrose metabolism	43	1.1807	2	0.33179	0.47913	1
Pyruvate metabolism	45	1.2357	2	0.35209	0.45335	1
Arginine and proline metabolism	78	2.1418	3	0.36265	0.44051	1
Lysine degradation	49	1.3455	2	0.39203	0.40668	1
Glycolysis or Gluconeogenesis	61	1.675	2	0.50451	0.29713	1
Arginine biosynthesis	27	0.74139	1	0.53039	0.2754	1
Butanoate metabolism	29	0.79631	1	0.55611	0.25484	1
Drug metabolism - other enzymes	70	1.9221	2	0.57983	0.2367	1
One carbon pool by folate	31	0.85123	1	0.58045	0.23624	1
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	31	0.85123	1	0.58045	0.23624	1

Table 6.4.2.3. 3. The matched genes/compounds in the gene-metabolite interaction network module of Cluster 2 (blue).

Correlated Pathways	matched_features (genes/compounds)
Arachidonic acid metabolism	cpd:C00157; cpd:C00427; cpd:C00696; cpd:C00584; cpd:C04853; cpd:C01312; cpd:C02198; hsa:2678; hsa:2687; hsa:6916
Biosynthesis of unsaturated fatty acids	cpd:C00249; cpd:C00712; cpd:C01595; cpd:C06426; cpd:C06427
Glycerolipid metabolism	cpd:C00422; cpd:C00416; hsa:23175; hsa:84803
Sphingolipid metabolism	cpd:C00550; cpd:C00195; cpd:C01290; cpd:C01190; hsa:55331
Glycerophospholipid metabolism	cpd:C00350; cpd:C00157; cpd:C00416; hsa:23175; hsa:84803
Phosphatidylinositol signaling system	cpd:C00416; hsa:5298; hsa:8867; hsa:27124
Fatty acid elongation	cpd:C00249; hsa:51144
Fatty acid biosynthesis	cpd:C00249; hsa:109703458

Findings on biochemical activities in cluster 2 (blue) segments were confirmed in the literature. First, we started with enrichment analysis of DESI-MSI detected metabolites. The top ones were Ceramides (Cer), especially HexCer (monohexosylceramides), **Figure 6.4.2.3.1b**. Cer is a lipid found in eukaryotic cells that acts as a lipid second messenger in cell signalling transductions involving various cellular activities such as cell proliferation, differentiation, and viability regulation. Therefore, a newly published study[274] reported the lipidomic profile associated with Ovarian cancer, specifically High-Grade Serious Ovarian Cancer (HGSOC) development and progression. The results revealed a dysregulated lipidome profile, with ceramides (Cer) and hexosylceramides (HexCer) showing subtle increases in abundance. These findings of the enriched lipid class, HexCer are consistent with the joint-pathway analysis results; as seen in **Figure 6.4.2.3.1c**, one of the most enriched joint-pathway was Arachidonic acid (AA) metabolism and sub-network details from gene-metabolite explorer analysis in **Figure 6.4.2.3.1e**. Ceramides, especially LacCer (2nd top enriched lipid class, **Figure 6.4.2.3.1b**), activate an enzyme, cytosolic phospholipase A2 (cPLA2), to generate arachidonic acid (AA), a precursor to prostaglandins[203]. In turn, arachidonic

acid serves as a precursor to eicosanoids and prostaglandin, which transduce a cascade of reactions leading to inflammation, cell proliferation, migration/infiltration, adhesion, angiogenesis, apoptosis, autophagy, and mitochondrial dysfunction. A significant phenotype underscored the initiation and progression of several debilitating diseases, such as cancer[203]. One study discovered that adipocytes and their secretory AA played a significant function in Ovarian Cancer, such as acting as a mediator for chemoresistance. This chemo-protective lipid mediator is arachidonic acid, which may be a possible therapeutic target for tackling OC chemoresistance[275].

The correlation of the Arachidonic acid metabolism function in the KEGG analysis also revealed the high activity of the F-kappa B signalling pathway, which was the most significantly enriched compared to other pathways in **Figure 6.4.2.3.1d**. In ovarian cancer (OC), the NF- kappa B signalling system plays a crucial role in chemoresistance, cancer stem cell maintenance, metastasis, and immune evasion. Many signalling pathways are dysregulated in OC and can stimulate NF- kappa B signalling via conventional or non-canonical pathways, all of which play overlapping roles in tumour growth[276]. A study[277] conducted in 1997 revealed that inhibitors of arachidonate oxidative metabolism were discovered to block NF-kappa B. In addition, one study[278] reported that AA, independent of its metabolites, can play a significant role in endothelial cell (EC) gene regulation in cancer. AA inhibits NF-kappa B translocation, given the importance of NF-kappa B in the upregulation of proinflammatory genes with EC activation.

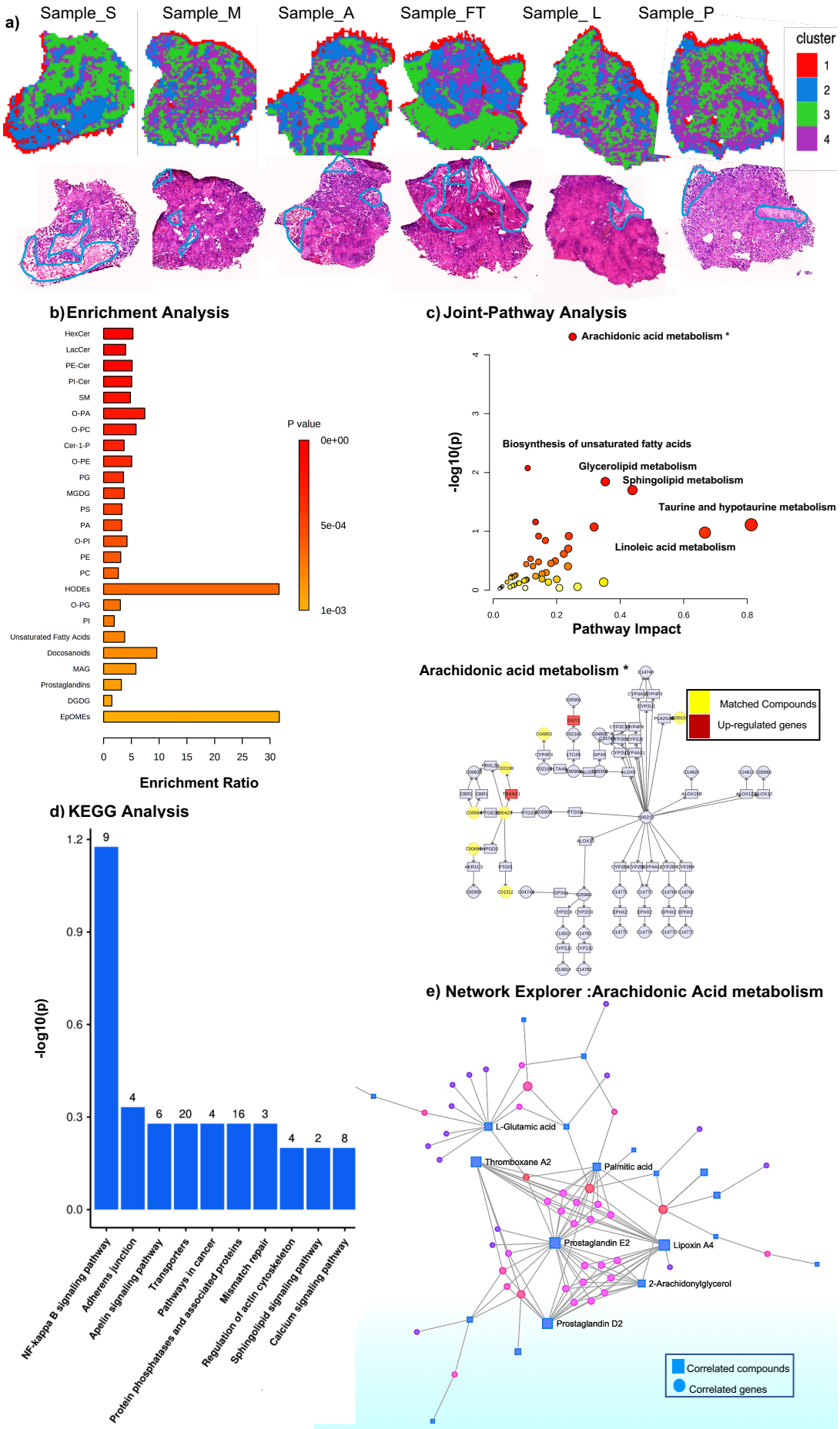

a)
Sample_S
Sample_M
Sample_A
Sample_FT
Sample_ L
Sample_P
cluster
1
2
3
4
b) Enrichment Analysis
HexCer
LacCer
PE-Cer
PI-Cer
SM
O-PA
O-PC
Cer-1-P
O-PE
PG
MGDG
PS
PA
O-PI
PE
PC
HODEs
O-PG
PI
Unsaturated Fatty Acids
Docosanoids
MAG
Prostaglandins
DGDG
EpOMEs
P value
0e+00
5e-04
1e-03
0
5
10
15
20
25
30
Enrichment Ratio
c) Joint-Pathway Analysis
Arachidonic acid metabolism *
Biosynthesis of unsaturated fatty acids
Glycerolipid metabolism
Sphingolipid metabolism
Taurine and hypotaurine metabolism
Linoleic acid metabolism
-log10(p)
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
Pathway Impact
Arachidonic acid metabolism *
Matched Compounds
Up-regulated genes
d) KEGG Analysis
-log10(p)
1.2
0.9
0.6
0.3
0.0
9
4
6
20
4
16
3
4
2
8
NF-kappa B signaling pathway
Adherens junction
Apelin signaling pathway
Transporters
Pathways in cancer
Protein phosphatases and associated proteins
Mismatch repair
Regulation of actin cytoskeleton
Sphingolipid signaling pathway
Calcium signaling pathway
e) Network Explorer :Arachidonic Acid metabolism
L-Glutamic acid
Thromboxane A2
Palmitic acid
Prostaglandin E2
Lipoxin A4
2-Arachidonylglycerol
Prostaglandin D2
Correlated compounds
Correlated genes

Figure 6.4.2.3.1. Correlation of Transcriptomics and metabolomics behind Cluster 2 by DESI- MSI of ovarian cancer (cluster blue versus all). a) MSI Clustering of Multi-site collected ovarian biopsies set b) Metabolite Enrichment Overview, Summary Plot for Over Representation Analysis (ORA), The enriched metabolomic content, the significant lipid classes in colour grading red (high) to white (less) for the most significant (p-value< 0.05). c) The global view of the pathways in the interaction network to show graphical output contained the view with the related genes and compounds - the most impacted on the metabolite pathway interaction as Arachidonic acid metabolism, KEGG orthologs as geneIDs in the arachidonic acid metabolic pathway were coloured, while genes upregulated in red and down-regulated in green. Yellow circles represented matched compounds to metabolites d) KEGG enrichment analysis to identify the pathways which had the most significant involvement with the upregulated and downregulated genes e) The plot of the selected interaction sub-network, Arachidonic acid metabolism in the Network Explorer, represented by blue squares and are associated with multiple genes, which are defined as blue circles. All statistical analysis threshold, p-value < 0.05, [FDR] < 0.05.

6.4.2.4 Comparison of Cluster 3 (green) metabolic segments in Network Analysis

Findings on biochemical activities in cluster 3 (green) segments were confirmed in the literature. First, we started with enrichment analysis of DESI-MSI detected metabolites. The top ones were Ceramides, especially HexCer (monohexosylceramides) in **Figure 6.4.2.4.1b.** The previous section mentioned Ceramides' role, which is needed for HGSOC cancer cell development and progression[274]. Other lipid classes included other forms of Ceramides such as LacCer, PI-Cer, O-PA, PE-Cer and others in **Table 6.4.2.1.**

Table 6.4.2.4.1. The most enriched lipid class of Cluster 3 (Green).

Lipid Classes	total	expected	hits	Holm p	FDR
HexCer	8690	175	825	3.8e-310	3.8e-310
PI-Cer	2290	46.1	293	3.65E-137	1.83E-137
O-PA	965	19.4	199	3.38E-132	1.13E-132
PE-Cer	2290	46.1	277	3.23E-123	8.11E-124
LacCer	4580	92.2	384	8.67E-120	1.74E-120
SM	2320	46.6	264	3.25E-111	5.45E-112
PA	1950	39.3	161	1.34E-47	1.94E-48
MGDG	1730	34.9	143	4.17E-42	5.26E-43
PS	1970	39.6	145	5.21E-37	5.86E-38
O-PE	1100	22	103	1.05E-34	1.06E-35
O-PI	964	19.4	84	9.86E-26	9.09E-27
O-PC	1040	21	87	2.50E-25	2.12E-26
Cer-1-P	2290	46.1	133	9.82E-24	7.68E-25
O-PG	964	19.4	76	1.70E-20	1.24E-21
DGDG	1730	34.8	98	2.00E-16	1.36E-17
PE	1970	39.7	95	1.63E-11	1.04E-12
PG	1950	39.2	90	6.98E-10	4.20E-11
MAG	71	1.43	11	0.00013	7.41E-06
PI	1960	39.5	75	0.000146	7.87E-06
Unsaturated Fatty Acids	267	5.37	17	0.0261	0.00134
O-LPG	14	0.282	4	0.0991	0.00485
HOMEs	2	0.0402	2	0.287	0.0135
LPE	67	1.35	6	1	0.0736
PC	1980	39.8	57	1	0.165
LPC	79	1.59	5	1	0.637

For the joint-pathway analysis results, as seen in **Figure 6.4.2.4.1c**, the most enriched joint-pathway was glycerolipid metabolism and others **Table 6.4.2.4.2.** Glycerolipids were biological molecules needed for membrane development, caloric storage, and intracellular signalling. There seems to be no direct relationship between ceramide abundance and glycerolipid metabolism. According to scientific literature, glycerolipids and sphingolipids can interact to alter signalling inside the cell, simulating Ceramides. Ceramides are thus essential for producing the bulk of complex sphingolipids, including sphingomyelins, glucosylceramides, and sphingosine[279]. Therefore, sphingolipid metabolism was also one of the enhanced joint-pathway metabolisms in **Figure 6.4.2.4.1c**. Interestingly, the Phosphatidylinositol signalling system was one of the top metabolic activities and the most enriched pathway in the KEGG investigation in **Figure 6.4.2.4.1d**. The cluster green metabolic segments appear to have significant *PIK3CA-related* gene expression activity, reflected in both MSEA and GSEA data.

Table 6.4.2.4.2. The crossed pathways by detected in the gene-metabolite interaction network module of Cluster 3 (green).

Top Enriched Joint-Pathways	Total	Expected	Hits	Raw p	Holm p	FDR
Glycerolipid metabolism	35	2.2234	9	0.00022573	3.6464	0.0077413
Glycerophospholipid metabolism	86	5.4631	15	0.00023661	3.626	0.0077413
Mucin type O-glycan biosynthesis	22	1.3975	7	0.00027648	3.5583	0.0077413
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	18	1.1434	6	0.00058467	3.2331	0.012278
Sphingolipid metabolism	58	3.6844	10	0.0029421	2.5313	0.049427
Phosphatidylinositol signaling system	74	4.7008	11	0.0060564	2.2178	0.084789
Inositol phosphate metabolism	69	4.3832	10	0.010467	1.9802	0.12561
Retinol metabolism	47	2.9857	7	0.026661	1.5741	0.25614
Arachidonic acid metabolism	81	5.1455	10	0.029991	1.523	0.25614
Lysine degradation	49	3.1127	7	0.032755	1.4847	0.25614
Ether lipid metabolism	39	2.4775	6	0.033961	1.469	0.25614
Glycolysis or Gluconeogenesis	61	3.875	8	0.036591	1.4366	0.25614
alpha-Linolenic acid metabolism	22	1.3975	4	0.046709	1.3306	0.28037
Nicotinate and nicotinamide metabolism	42	2.668	6	0.046729	1.3304	0.28037
Linoleic acid metabolism	17	1.0799	3	0.088747	1.0518	0.49698
Riboflavin metabolism	9	0.57172	2	0.10758	0.96829	0.56477
Amino sugar and nucleotide sugar metabolism	79	5.0184	8	0.12484	0.90364	0.60225
Phosphonate and phosphinate metabolism	10	0.63525	2	0.12905	0.88923	0.60225
Histidine metabolism	32	2.0328	4	0.1416	0.84893	0.60785
Phenylalanine metabolism	21	1.334	3	0.14473	0.83945	0.60785
Fatty acid degradation	102	6.4795	9	0.19588	0.70801	0.78351
Arginine and proline metabolism	78	4.9549	7	0.22347	0.65079	0.84374
Neomycin, kanamycin and gentamicin biosynthesis	4	0.2541	1	0.23102	0.63634	0.84374
Butanoate metabolism	29	1.8422	3	0.27882	0.55468	0.91396
Glutathione metabolism	56	3.5574	5	0.28138	0.5507	0.91396

Finally, network analysis using the top hit genes and compounds in the whole metabolome and transcriptome data matched the alpha-linolenic acid metabolic pathway in **Figure 6.4.2.4.1e** and others in **Table 6.4.2.4.3.** As a component of membrane phospholipids, Linoleic acid serves as a structural component in the maintenance of the epidermis' transdermal water barrier and membrane fluidity[280]. This network supported prior studies on why sphingolipid metabolism is significant in the joint-pathway result. Consequently, multiple joint pathways influence metabolic complexity in this green metabolic segment rather than only one joint-pathway activity.

Table 6.4.2.4.3. The matched genes/compounds in the gene-metabolite interaction network module of Cluster 3 (green).

Network Pathways	matched_features
Glycerolipid metabolism	cpd:C00422; cpd:C00416; hsa:196051; hsa:8611; hsa:8613; hsa:11343; hsa:84803; hsa:219; hsa:8526
Glycerophospholipid metabolism	cpd:C00350; cpd:C00157; cpd:C04230; cpd:C00416; hsa:5130; hsa:5833; hsa:5322;hsa:8399; hsa:8526; hsa:196051; hsa:8611; hsa:8613; hsa:23646; hsa:84803; hsa:2820
Sphingolipid metabolism	cpd:C00550; cpd:C00195; cpd:C01290; hsa:55331; hsa:8611; hsa:8613; hsa:6609; hsa:6610; hsa:2531; hsa:8877
Phosphatidylinositol signaling system	cpd:C00416; hsa:9107; hsa:5336; hsa:5293; hsa:23396; hsa:5287; hsa:8526; hsa:8821; hsa:8871; hsa:5305; hsa:3632
Arachidonic acid metabolism	cpd:C00157; hsa:2878; hsa:5743; hsa:5322; hsa:8399 hsa:2687; hsa:5740; hsa:9536; hsa:8644; hsa:5730
alpha-Linolenic acid metabolism	cpd:C00157; cpd:C06427; hsa:5322; hsa:8399
Linoleic acid metabolism	cpd:C00157; hsa:5322; hsa:8399

The KEGG analysis of the correlation of alpha-linolenic acid function indicated the high activity of the PI3K-Akt signalling pathway, which was the most substantially enriched compared to other pathways in **Figure 6.4.2.4.1d.** In ovarian cancer (OC) cases, The PI3K/AKT/mTOR pathway is activated in about 70% of ovarian cancer patients, promoting cellular growth, proliferation, and survival via a circuitous sequence of hyperactive signalling cascades. Its activation is associated with higher invasive and migratory capacities across heterogeneous cell subpopulations coexisting in human ovarian cancer. This association could suggest that the PI3K/AKT/mTOR pathway is a potential predictor of invasiveness for ovarian tumour cells[281]. One of the study[282] findings emphasises the relevance of PI3K signalling, as significantly enhanced PI3K pathway activity was identified in Fallopian tube epithelium (FTE) and HGSOC samples compared to healthy FTE samples. Another well-known fact, Despite mounting evidence that Fallopian tube epithelium (FTE)

is the cell of origin for most HGSC. However, the early molecular pathways driving the transition of FTE into serous tubal intraepithelial carcinoma (STIC) and high-grade serous ovarian carcinoma (HGSOC) remain unknown. **Figure 6.4.2.4.1a (Sample_FT)** demonstrates that cluster green exhibited distinctive metabolic regions with significant PI3K-Akt signalling activity. As a result, according to a newly published research in 2022[283], neoplastic transformation of FTE was expected to need abnormal signalling networks rather than a single mechanism. Aside from increased PI3K pathway activity. The study discovered decreased oestrogen receptor (ER) pathway activity in serous tubal intraepithelial carcinoma (STIC) and high-grade serous ovarian carcinoma (HGSOC) samples which was caused by tumour-driving effects of aberrant PI3K signalling. Nonetheless, loss of ER pathway functional activity could explain why people with HGSC have a poor response to anti-estrogen receptor therapeutic targets.

These findings, multiple analyses based on the top hit genes and compounds in the whole metabolome, and transcriptome data matched the PI3K-Akt signalling pathway with alpha-linolenic acid metabolism. This finding may suggest that cluster green in the Fallopian tube is the origin of HGSOC progression.

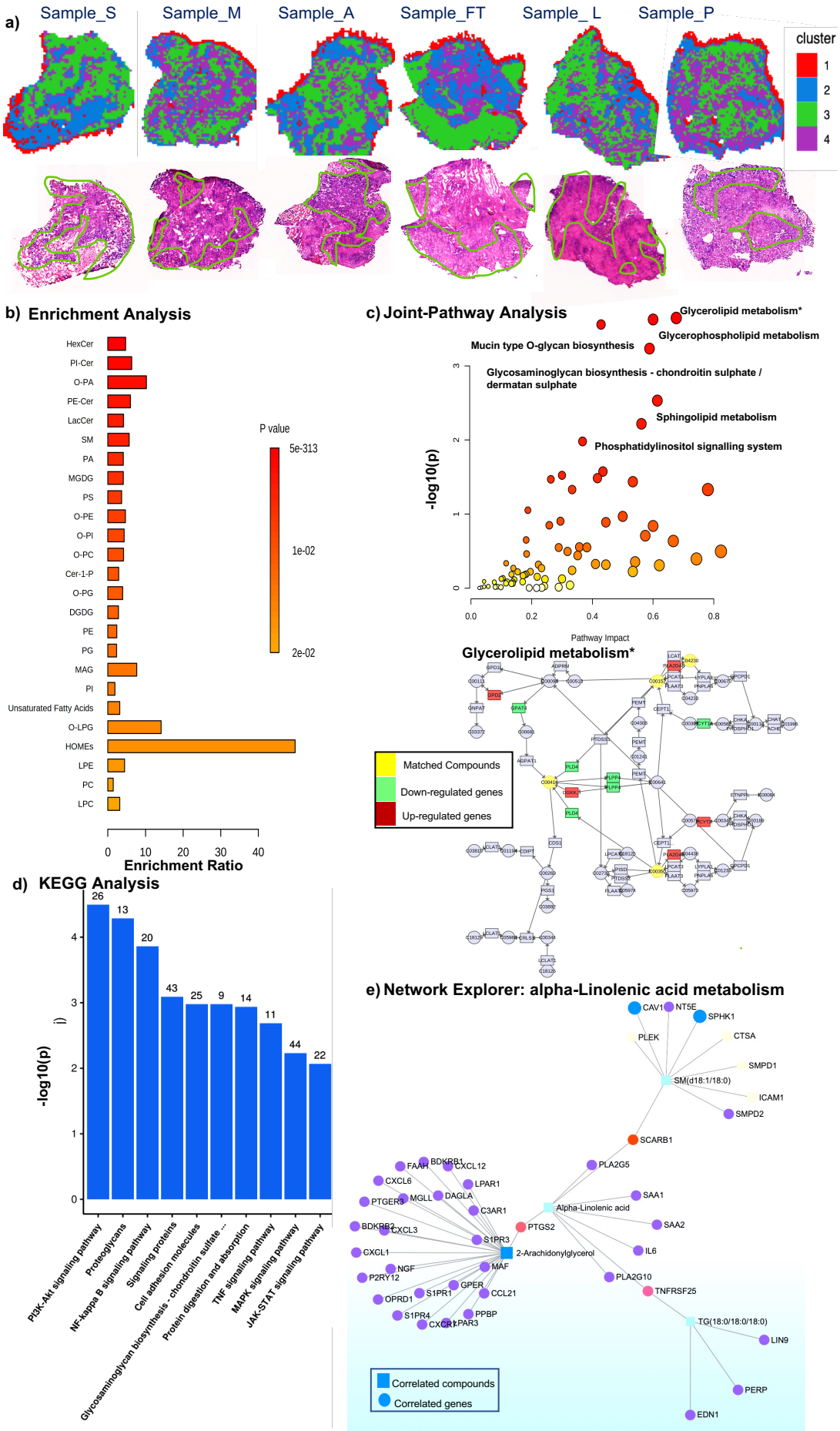

a)
Sample_S
Sample_M
Sample_A
Sample_FT
Sample_ L
Sample_P
cluster
1
2
3
4
b) Enrichment Analysis
HexCer
PI-Cer
O-PA
PE-Cer
LacCer
SM
PA
MGDG
PS
O-PE
O-PI
O-PC
Cer-1-P
O-PG
DGDG
PE
PG
MAG
PI
Unsaturated Fatty Acids
O-LPG
HOMEs
LPE
PC
LPC
P value
5e-313
1e-02
2e-02
0
10
20
30
40
Enrichment Ratio
c) Joint-Pathway Analysis
Glycerolipid metabolism*
Glycerophospholipid metabolism
Mucin type O-glycan biosynthesis
Glycosaminoglycan biosynthesis - chondroitin sulphate / dermatan sulphate
Sphingolipid metabolism
Phosphatidylinositol signalling system
-log10(p)
0
1
2
3
0.0
0.2
0.4
0.6
0.8
Pathway Impact
Glycerolipid metabolism*
Matched Compounds
Down-regulated genes
Up-regulated genes
d) KEGG Analysis
-log10(p)
0
1
2
3
4
26
13
20
43
25
9
14
11
44
22
PI3K-Akt signaling pathway
Proteoglycans
NF-kappa B signaling pathway
Signaling proteins
Cell adhesion molecules
Glycosaminoglycan biosynthesis - chondroitin sulfate ...
Protein digestion and absorption
TNF signaling pathway
MAPK signaling pathway
JAK-STAT signaling pathway
e) Network Explorer: alpha-Linolenic acid metabolism
CAV1
NT5E
SPHK1
PLEK
CTSA
SMPD1
SM(d18:1/18:0)
ICAM1
SMPD2
SCARB1
PLA2G5
FAAH
BDKRB1
CXCL12
CXCL6
LPAR1
PTGER3
MGLL
DAGLA
SAA1
C3AR1
Alpha-Linolenic acid
SAA2
BDKRB2
CXCL3
PTGS2
S1PR3
IL6
CXCL1
2-Arachidonylglycerol
NGF
MAF
PLA2G10
P2RY12
GPER
TNFRSF25
OPRD1
S1PR1
CCL21
S1PR4
PPBP
CXCR4
LPAR3
TG(18:0/18:0/18:0)
LIN9
PERP
EDN1
Correlated compounds
Correlated genes

Figure 6.4.2.4.1. Correlation of Transcriptomics and metabolomics behind Cluster 3 by DESI- MSI of ovarian cancer (cluster green versus all). *a)* MSI Clustering of Multi-site collected ovarian biopsies set b) Metabolite Enrichment Overview, Summary Plot for Over Representation Analysis (ORA), The enriched metabolomic content, the significant lipid classes in colour grading red (high) to white (less) for the most significant (p-value< 0.05). c) The global view of the pathways in the interaction network to show graphical output contained the view with the related *genes and* compounds - the most impacted on the metabolite pathway interaction as *Glycerolipid* metabolism*,* KEGG orthologs as geneIDs in the arachidonic acid metabolic pathway were coloured, while genes upregulated in red and down-regulated in green. Yellow circles represented matched compounds to metabolites d) KEGG enrichment analysis to identify the pathways which had the most significant involvement with the upregulated and downregulated genes e) The plot of the selected interaction sub-network, alpha-linolenic acid metabolic pathway in the Network Explorer*,* represented by blue squares and are associated with multiple genes, which are defined as blue circles. All statistical analysis threshold, p-value < 0.05, [FDR] < 0.05.

6.4.2.5 Comparison of Cluster 4 (purple) metabolic segments in Network Analysis

Findings on biochemical activities in cluster 4 (purple) segments were confirmed in the literature. First, we started with enrichment analysis of DESI-MSI detected metabolites. The top ones were Ceramides, especially HexCer (monohexosylceramides) in **Figure 6.4.2.5.1b** and others in **Table 6.4.2.5.1.** The previous section mentioned Ceramides' role, which is needed for HGSOC cancer cell development and progression[274].

Table 6.4.2.5.1. The most enriched lipid class of Cluster 4 (Purple).

Lipid Classes	total	expected	hits	Raw p	Holm p	FDR
HexCer	8690	175	825	5.20E-319	3.8e-310	3.8e-310
PI-Cer	2290	46.1	293	5.01E-140	3.65E-137	1.83E-137
O-PA	965	19.4	199	4.64E-135	3.38E-132	1.13E-132
PE-Cer	2290	46.1	277	4.44E-126	3.23E-123	8.11E-124
LacCer	4580	92.2	384	1.19E-122	8.67E-120	1.74E-120
SM	2320	46.6	264	4.47E-114	3.25E-111	5.45E-112
PA	1950	39.3	161	1.85E-50	1.34E-47	1.94E-48
MGDG	1730	34.9	143	5.75E-45	4.17E-42	5.26E-43
PS	1970	39.6	145	7.21E-40	5.21E-37	5.86E-38
O-PE	1100	22	103	1.45E-37	1.05E-34	1.06E-35
O-PI	964	19.4	84	1.37E-28	9.86E-26	9.09E-27
O-PC	1040	21	87	3.47E-28	2.50E-25	2.12E-26
Cer-1-P	2290	46.1	133	1.37E-26	9.82E-24	7.68E-25
O-PG	964	19.4	76	2.37E-23	1.70E-20	1.24E-21
DGDG	1730	34.8	98	2.79E-19	2.00E-16	1.36E-17
PE	1970	39.7	95	2.27E-14	1.63E-11	1.04E-12
PG	1950	39.2	90	9.76E-13	6.98E-10	4.20E-11
MAG	71	1.43	11	1.82E-07	0.00013	7.41E-06
PI	1960	39.5	75	2.05E-07	0.000146	7.87E-06
Unsaturated Fatty Acids	267	5.37	17	3.67E-05	0.0261	0.00134
O-LPG	14	0.282	4	0.000139	0.0991	0.00485
HOMEs	2	0.0402	2	0.000405	0.287	0.0135
LPE	67	1.35	6	0.00232	1	0.0736
PC	1980	39.8	57	0.00542	1	0.165

For the joint-pathway analysis results, as seen in **Figure 6.4.2.5.1c**, the most enriched joint-pathway was glycerophospholipid metabolism and others in **Table 6.4.2.5.2.** Glycerophospholipids and sphingolipids are the essential components of the cell membrane. High Ceramide abundance and glycerophospholipid metabolism appear to have no direct link. Similar to the previous section, glycerolipids and sphingolipids can interact in several ways to affect signalling inside the cell mediated by Ceramides, according to recent studies [279]. As a result, one of the enhanced joint-pathway metabolisms was sphingolipid metabolism in **Figure 6.4.2.5.1c** and network explorer analysis matched glycerolipids metabolism in **Figure 6.4.2.5.1e** and others in **Table 6.4.2.5.3.** Interestingly, both N-Glycan biosynthesis and Glycosaminoglycan biosynthesis - chondroitin sulphate / dermatan sulphate had high metabolic activity and fit the most enriched pathway, proteoglycans, in the KEGG analysis, **Figure 6.4.2.5.1d.**

Table 6.4.2.5.2. The crossed pathways by detected in the gene-metabolite interaction network module of Cluster 4 (Purple).

Top Enriched Joint-pathways	Total	Expected	Hits	Raw p	Holm p	FDR
Glycerophospholipid metabolism	86	2.432	10	0.00010792	3.9669	0.0090654
Glycerolipid metabolism	35	0.98975	6	0.00034894	3.4572	0.014656
Phosphatidylinositol signalling system	74	2.0926	7	0.0042401	2.3726	0.10946
Sphingolipid metabolism	58	1.6402	6	0.0052124	2.283	0.10946
N-Glycan biosynthesis	77	2.1775	5	0.064206	1.1924	1
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	18	0.50902	2	0.090154	1.045	1
Glycolysis or Gluconeogenesis	61	1.725	4	0.091648	1.0379	1
alpha-Linolenic acid metabolism	22	0.62213	2	0.12676	0.89701	1
Inositol phosphate metabolism	69	1.9512	4	0.12877	0.89019	1
Pyruvate metabolism	45	1.2725	3	0.1326	0.87747	1
Pentose phosphate pathway	47	1.3291	3	0.14574	0.83641	1
Glutathione metabolism	56	1.5836	3	0.20948	0.67885	1
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	31	0.87664	2	0.21794	0.66167	1
Riboflavin metabolism	9	0.25451	1	0.22787	0.64231	1
Various types of N-glycan biosynthesis	36	1.018	2	0.27091	0.56717	1
Purine metabolism	166	4.6943	6	0.3276	0.48465	1
Thiamine metabolism	14	0.3959	1	0.33149	0.47954	1
Citrate cycle (TCA cycle)	42	1.1877	2	0.33439	0.47575	1
Starch and sucrose metabolism	43	1.216	2	0.34485	0.46237	1
Arginine and proline metabolism	78	2.2057	3	0.38037	0.41979	1
Biosynthesis of unsaturated fatty acids	47	1.3291	2	0.38615	0.41325	1
Linoleic acid metabolism	17	0.48074	1	0.38694	0.41235	1
Porphyrin and chlorophyll metabolism	53	1.4988	2	0.44591	0.35075	1
Mucin type O-glycan biosynthesis	22	0.62213	1	0.46946	0.3284	1
Pyrimidine metabolism	99	2.7996	3	0.53705	0.26998	1

Table 6.4.2.5.3. The matched genes/compounds in the gene-metabolite interaction network module of Cluster 4 (Purple).

Network Pathways	matched_features (genes/compounds)
Glycerophospholipid metabolism	cpd:C00350; cpd:C00157; cpd:C04230; cpd:C00416; hsa:139189; hsa:1609; hsa:8526; hsa:2819; hsa:11313; hsa:23659
Glycerolipid metabolism	cpd:C00422; cpd:C00416; hsa:5407; hsa:139189; hsa:1609; hsa:8526
Phosphatidylinositol signalling system	cpd:C00416; hsa:51196; hsa:5330; hsa:139189; hsa:1609; hsa:8526; hsa:3628
Sphingolipid metabolism	cpd:C00550; cpd:C00195; cpd:C01290; hsa:6610; hsa:2629; hsa:2581
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	cpd:C00350; hsa:5283
Arachidonic acid metabolism	cpd:C00157; hsa:493869

Proteoglycans are glycocalyx components formed of a protein core and long glycosaminoglycan (GAG) carbohydrate chains covalently bonded[284]. Proteoglycans are glycoproteins located on cell surfaces and in the extracellular matrix (ECM). GAGs are non-branched linear chains of repeating disaccharide units produced at the plasma membrane, including heparan sulphate, chondroitin sulphate (CS), dermatan sulphate keratan, and hyaluronic acid/hyaluronan[284,285]. According to one study[286], a tumour cell surface neoantigen, chondroitin sulphate proteoglycan 4 (CSPG4), was an independent risk factor for poor survival of patients with EOC, as shown in **Figure 6.4.2.5.1d**. Their findings reveal that CSPG4 enhances cell invasion, cisplatin resistance, spheroid formation in vitro, and tumour development *in vivo*. Spheroid development and tumour cell invasion are caused by CSPG4-induced production of the mesenchymal transcription factor ZEB1. As a result, they propose that CSPG4 expression is responsible for phenotypic heterogeneity and malignant development in EOC tumours. These results show that CSPG4 expression levels are a possible diagnostic biomarker in EOC and that targeting cells that express this neoantigen may minimise recurrence and enhance outcomes in EOC patients[287]. Multiple studies based on the top hit genes and substances in the whole metabolome and transcriptome data matched the high proteoglycan activity, evident in both MSEA and GSEA data.

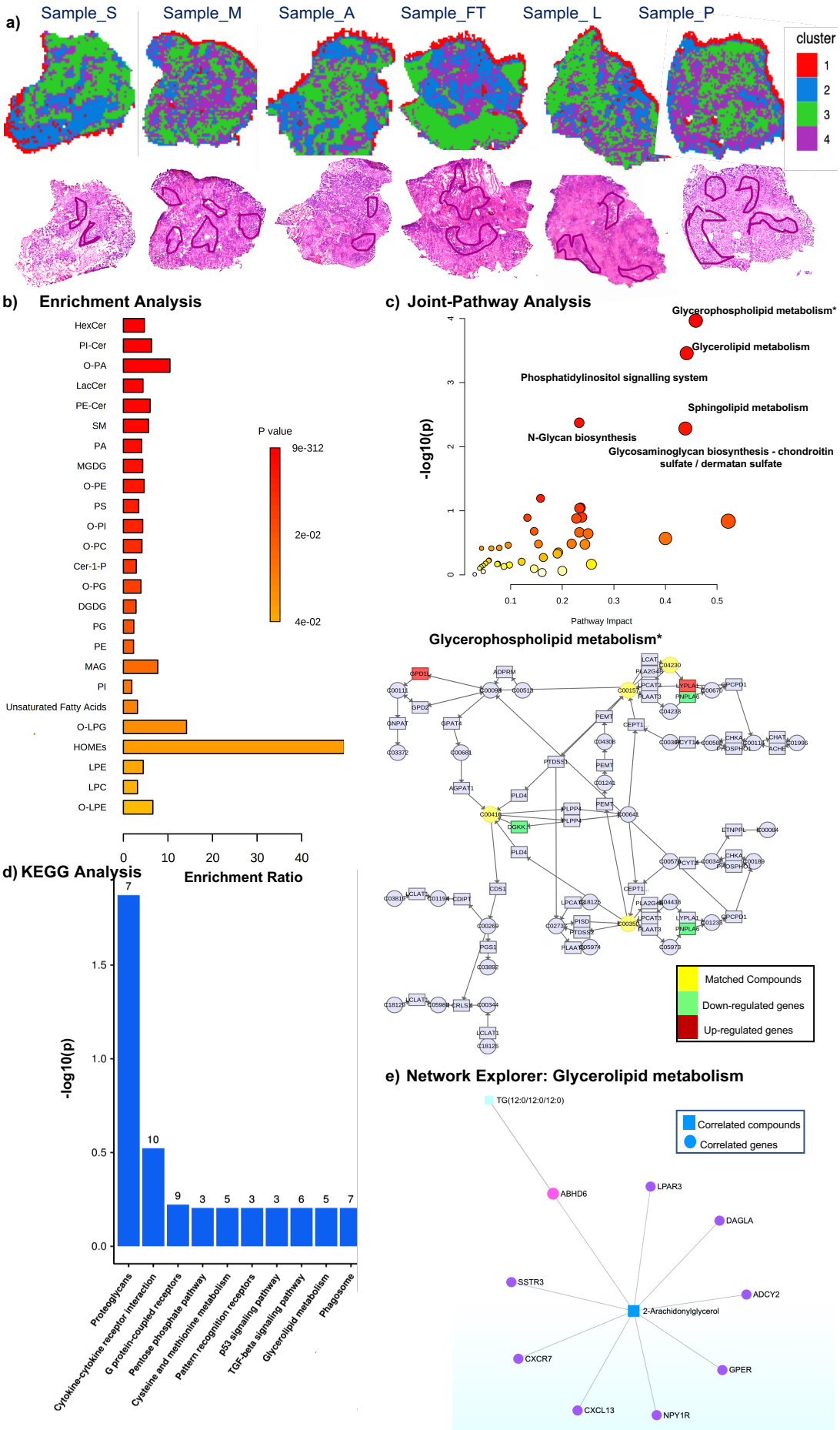

a)
Sample_S
Sample_M
Sample_A
Sample_FT
Sample_ L
Sample_P
cluster
1
2
3
4
b) Enrichment Analysis
HexCer
PI-Cer
O-PA
LacCer
PE-Cer
SM
PA
MGDG
O-PE
PS
O-PI
O-PC
Cer-1-P
O-PG
DGDG
PG
PE
MAG
PI
Unsaturated Fatty Acids
O-LPG
HOMEs
LPE
LPC
O-LPE
P value
9e-312
2e-02
4e-02
0
10
20
30
40
Enrichment Ratio
c) Joint-Pathway Analysis
Glycerophospholipid metabolism*
Glycerolipid metabolism
Phosphatidylinositol signalling system
Sphingolipid metabolism
N-Glycan biosynthesis
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate
-log10(p)
Pathway Impact
Glycerophospholipid metabolism*
Matched Compounds
Down-regulated genes
Up-regulated genes
d) KEGG Analysis
-log10(p)
7
10
9
3
5
3
3
6
5
7
Proteoglycans
Cytokine-cytokine receptor interaction
G protein-coupled receptors
Pentose phosphate pathway
Cysteine and methionine metabolism
Pattern recognition receptors
p53 signaling pathway
TGF-beta signaling pathway
Glycerolipid metabolism
Phagosome
e) Network Explorer: Glycerolipid metabolism
TG(12:0/12:0/12:0)
Correlated compounds
Correlated genes
ABHD6
LPAR3
DAGLA
SSTR3
ADCY2
2-Arachidonylglycerol
CXCR7
GPER
CXCL13
NPY1R

Figure 6.4.2.5.1. Correlation of Transcriptomics and metabolomics behind Cluster 4 by DESI-MSI of ovarian cancer (cluster purple versus all). a) MSI Clustering of Multi-site collected ovarian biopsies set b) Metabolite Enrichment Overview, Summary Plot for Over Representation Analysis (ORA), The enriched metabolomic content, the significant lipid classes in colour grading red (high) to white (less) for the most significant (p-value< 0.05). c) The global view of the pathways in the interaction network to show graphical output contained the view with the related genes and compounds - the most impacted on the metabolite pathway interaction as Glycerophospholipids metabolism, KEGG orthologs as geneIDs in the arachidonic acid metabolic pathway were coloured, while genes upregulated in red and down-regulated in green. Yellow circles represented matched compounds to metabolites d) KEGG enrichment analysis to identify the pathways which had the most significant involvement with the upregulated and downregulated genes e) The plot of the selected interaction sub-network, Glycerolipid metabolic pathway in the Network Explorer, represented by blue squares and are associated with multiple genes, which are defined as blue circles. All statistical analysis threshold, p-value < 0.05, [FDR] < 0.05.

6.5 Discussion and Conclusion

The findings in this chapter broaden the scope of DESI-MSI technology's application and demonstrate its practicality in multi-modal data integration utilising a single tissue section strategy. This strategy is the first time DESI-MSI has been used in conjunction with an LCM system to conduct a transcriptomics follow-up investigation using a post-DESI tissue slice. Comparison of multi-site collected ovarian biopsies in three consecutive sections and segmentation analysis demonstrated a degree of heterogeneity among samples. The DESI-MSI clusters 2, 3, and 4 ROIs were then microdissected by a laser capture system for transcriptomics study. Transcriptomic data revealed gene expression differences between samples and metabolically distinct sub-regions identified by DESI analysis inside the same biopsy. Metabolically different clusters in each biopsy show that specific genetic alterations may drive diverse metabolic activity.

Several technical optimisations for an operational workflow were standardised using the single tissue section for transcriptomics. Because tissue morphology was retained following DESI-MSI acquisitions, the quantity and quality of RNA collected from tissue slices demonstrated that the isolated starting material was valuable and sufficient for a transcriptome analysis such as RNA sequencing. Due to its speed and adaptability, DESI-MS imaging is suitable for follow-up gene expression analysis under ambient circumstances. This multi-modal investigation proved the complementarity of many analytical methods in the non-target approach, allowing us to achieve more robust results through the consistency of orthogonal measurements. Ceramide lipid classes (HexCer, LacCer, PE-Cer, and others), NF-kappa B signalling, PI3K-Akt signalling pathway, Proteoglycans, Glycerophospholipids, Sphingolipids and, most importantly, arachidonic acid metabolism were all observed in both metabolomic and genomic data. Most observed alterations were consistent with previously identified ovarian cancer metabolic variables. The associated genomic mapping of the tumour environment can give a unique insight into altered cellular activities and interactions *in situ*, potentially leading to the discovery of novel disease-specific targets and biomarkers.

The present multi-modal analysis with a single section strategy can distinguish metabolically distinct subclones within the same tumour. This transcriptomic-based subcloning method can detect tumour heterogeneity. RNA sequencing data might determine gene expressions of oncogenic pathways associated with the DESI-MSI-observed tumour phenotype. This strategy accurately maps biochemical and genetic information from two separate omics datasets of molecules to phenotypic traits. The DESI and transcriptomics technologies were combined with increasing the amount of in-depth molecular coverage obtained from a single tissue section. This study presents a novel MSI-based mapping strategy that might improve diagnostic skills, shed light on disease pathophysiology, and be used in routine MS applications.

6.6 Appendix

Table 6.6.1. Top Detected peaks details for the heatmap features

mz	Cluster 1	1_p.value	Cluster 2	2_p.value	Cluster 3	3_p.value	Cluster 4	4_p.value
800.6089	-0.477941908	1.06E-06	-0.290438382	0.00500455	0.638231987	3.72E-13	0.130148303	0.24350332
575.4793	-0.379816542	8.39E-13	-0.034419317	0.64349484	0.352707307	1.37E-10	0.061528552	0.37586715
466.315	-0.525732929	1.24E-11	-0.090566869	0.36207377	0.502332963	2.21E-10	0.113966834	0.235467
798.5889	-0.502234111	2.53E-09	-0.189464897	0.04630909	0.504088827	2.10E-09	0.187610182	0.04874877
511.2935	-0.420179797	2.59E-11	-0.098283189	0.20169365	0.383242152	4.71E-09	0.135220835	0.06764245
269.2462	-0.370795317	2.06E-11	-0.017301034	0.83040668	0.335655228	5.69E-09	0.052441123	0.47073839
712.501	-0.771169881	1.61E-13	-0.1357161	0.31877467	0.64765144	9.47E-09	0.259234541	0.0395946
600.5019	-0.36648109	4.32E-12	-0.011121452	0.88836754	0.319787107	1.11E-08	0.057815435	0.40208743
755.553	-0.840330105	7.27E-15	-0.127345437	0.38375243	0.672491029	1.89E-08	0.295184513	0.02557322
253.2151	-0.721114968	3.72E-15	-0.052292586	0.70118086	0.569171958	2.65E-08	0.204235595	0.07420257
505.2684	-0.822002388	9.87E-14	-0.137639729	0.3427552	0.669102281	2.92E-08	0.290539837	0.02908536
758.5553	-0.723241832	3.27E-14	-0.128270035	0.30525246	0.579737181	3.34E-08	0.271774686	0.01830799
757.5539	-0.705293042	1.28E-15	-0.08836982	0.4726567	0.543669604	3.91E-08	0.249993259	0.02126627
476.272	-0.965220756	3.35E-13	-0.134045071	0.45202544	0.777760109	7.92E-08	0.321505718	0.0432258
573.4761	-0.775388849	0	0.023849643	0.87173988	0.56262741	8.29E-08	0.188911796	0.1080146
339.2012	-0.579798167	3.72E-15	8.92E-05	0.99917559	0.4442238	8.68E-08	0.135485168	0.14940773
849.5317	-0.843800208	0	-0.093046587	0.51289384	0.604642415	9.82E-08	0.33220438	0.00662117
720.4906	-0.90692547	2.65E-14	-0.200787782	0.18519145	0.705293468	1.03E-07	0.402419785	0.00480257
479.2912	-0.90310575	5.93E-14	-0.127328032	0.43348327	0.706098548	1.17E-07	0.324335233	0.0254938
509.283	-0.961214773	0	-0.129035453	0.42673763	0.702632327	1.39E-07	0.387617899	0.00691794
841.56	-0.872974891	1.28E-15	-0.053751276	0.74484666	0.650139911	1.51E-07	0.276586257	0.04084205
847.526	-0.827200179	0	-0.115868474	0.4102794	0.607169796	1.93E-07	0.335898857	0.00725094
801.5503	-0.671998046	2.21E-12	-0.127580026	0.29564841	0.538739569	2.34E-07	0.260838504	0.0206257
905.5143	-0.73793569	1.38E-11	-0.139292985	0.31286972	0.605943811	2.49E-07	0.271284864	0.03333776
616.2763	-0.685585023	1.34E-12	-0.07092753	0.59746718	0.544668079	2.55E-07	0.211844474	0.06737702
805.4868	-0.787123117	1.28E-15	-0.096729639	0.48289314	0.576921802	2.75E-07	0.306930954	0.01084724

Table 6.6.1. continuous, Top Detected peaks details for the heatmap features

756.5493	-0.987472864	0	-0.049376768	0.78072904	0.673249001	5.04E-07	0.363600632	0.01107308
504.2669	-1.108052381	0	-0.070417018	0.72150868	0.757426886	5.14E-07	0.421042513	0.00882679
761.5182	-1.02036478	0	-0.105960042	0.55545323	0.716760363	5.23E-07	0.409564458	0.00703974
875.5401	-0.795119679	3.72E-15	-0.091525869	0.52356752	0.57517442	6.91E-07	0.311471128	0.01168023
842.5676	-0.763170476	0	0.004953326	0.97171199	0.49661558	7.39E-07	0.26160157	0.01454782
309.2763	-0.765866987	2.11E-13	-0.061797441	0.67814158	0.576085705	7.46E-07	0.251578723	0.04508564
866.5642	-0.743199476	0	-0.003784523	0.97933246	0.512161176	7.51E-07	0.234822824	0.03497704
483.2668	-0.548970563	9.16E-12	-0.073434713	0.49358285	0.433540201	7.54E-07	0.188865075	0.04573668
619.2858	-0.932725477	0	-0.078602572	0.64009714	0.654984363	8.17E-07	0.356343686	0.0117448
785.4989	-0.788571492	6.70E-15	-0.063801271	0.66327828	0.569906743	8.22E-07	0.28246602	0.02245988
462.2924	-1.072426916	0	-0.067021091	0.7393971	0.752899297	8.27E-07	0.38654871	0.01792168
742.5292	-0.838842836	0	-0.025537151	0.86619123	0.534995692	8.43E-07	0.329384295	0.00415743
728.5461	-0.975924523	0	0.009902174	0.95340824	0.612327291	9.59E-07	0.353695058	0.00753604
533.2803	-0.540533617	3.18E-12	-0.080310407	0.43151441	0.416961197	9.88E-07	0.203882827	0.02543935
910.547	-0.896359509	0	-0.109956291	0.47062717	0.612692401	1.12E-06	0.393623399	0.0030803
744.544	-0.784526889	0	-0.026285217	0.84986848	0.495344888	1.23E-06	0.315467218	0.00342352
450.2914	-0.558255698	4.04E-11	-0.054479119	0.63703469	0.442727709	1.32E-06	0.170007108	0.08796017
280.2337	-0.990141984	0	0.029566347	0.86955248	0.630297637	1.36E-06	0.330278001	0.01744826
732.5392	-0.741368256	0	-0.04274336	0.75771203	0.510978419	1.37E-06	0.273133198	0.01513411
506.2823	-1.090025416	0	-0.012501477	0.94875465	0.686051961	1.41E-06	0.416474932	0.00550084
618.2946	-1.098460486	1.28E-15	-0.047995714	0.8216562	0.765342773	1.55E-06	0.381113427	0.02488318
481.3046	-0.962850967	0	-0.025984811	0.88650998	0.637058369	1.56E-06	0.351777409	0.01234271
846.5725	-0.743789171	0	-0.042105492	0.75270374	0.492718517	1.56E-06	0.293176146	0.00674281
562.3417	-0.614779068	1.28E-15	-0.084852267	0.42541182	0.427891574	1.61E-06	0.271739761	0.00380902
531.2767	-1.02425219	0	-0.065805881	0.71699304	0.671467084	1.63E-06	0.418590987	0.00451933
478.2873	-1.298948302	0	-0.03326076	0.88836754	0.832136105	1.66E-06	0.500072957	0.00632552
916.5608	-0.690789824	4.31E-15	-0.05939145	0.64132002	0.482566564	2.00E-06	0.26761471	0.01264922
873.5351	-1.08689167	0	-0.077387187	0.68674771	0.715735142	2.03E-06	0.448543716	0.00460143
574.4781	-0.972875299	0	0.079321987	0.63530643	0.629260303	2.09E-06	0.264293009	0.06431188
831.4991	-0.710418008	1.91E-15	-0.037522872	0.78207995	0.490152975	2.26E-06	0.257787904	0.01913833

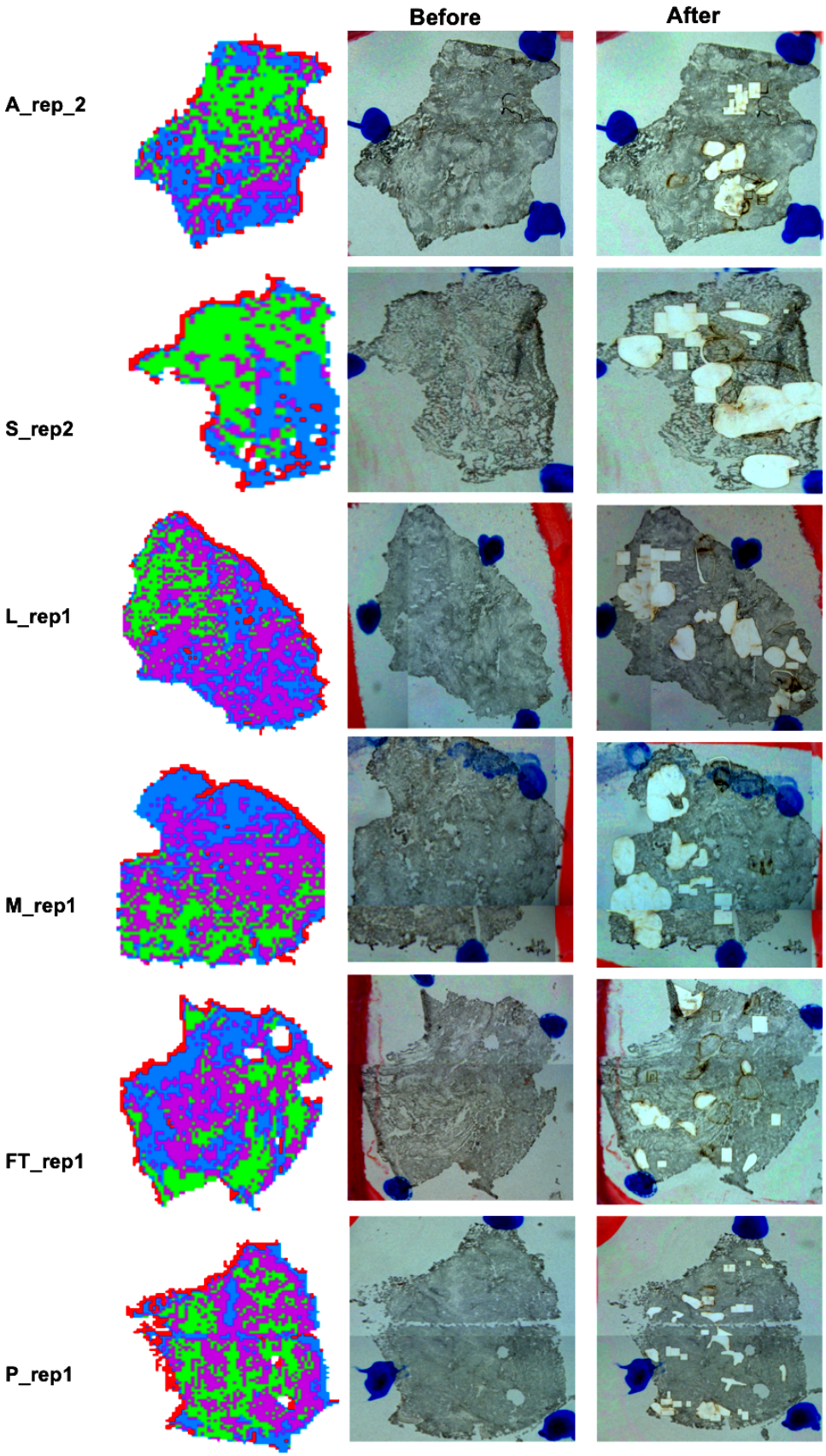


Figure 6.6.1. Before and after microdissected regions on all samples.

Table 6.6.2. RIN number and total isolated amount of microdissected samples.

Sample ID	Cluster ID	Total Amount (ng)	RIN
A-2	**blue**	10.428	8.9
A-3	**green**	6.444	7.6
A-4	**purple**	17.2	7.7
FT-2	**blue**	9.54	7
FT-3	**green**	17.28	8.5
FT-4	**purple**	11	6.5
L-2	**blue**	49.92	8.2
L-3	**green**	21.6	8.3
L-4	**purple**	19.5	8.6
M-2	**blue**	15.84	8.5
M-3	**green**	24.48	7.5
M-4	**purple**	17.28	7.9
P-2	**blue**	19.3	7.9
P-3	**green**	25.56	8.5
P-4	**purple**	8.352	7.8
S-2	**blue**	39.6	8
S-3	**green**	8.35	8.6
S-4	**purple**	16.44	7.9

7 Final Remarks

7.1 The Dilemma of Tumor Heterogeneity in Breast and Ovarian Cancer

Breast and ovarian cancer are among the top ten most frequent cancers in women globally[2,288]. Despite significant breakthroughs in diagnosis, treatment innovations, and subtype identification, breast cancer remains the second biggest cause of cancer-related deaths among women worldwide, with ovarian cancer ranking fifth[9,289]. Several scientific and clinical issues influence patient survival, including early diagnosis, precise characterization, the development of resistance to therapy, metastasis, personalized prognosis management, and tumour heterogeneity[289]. Tumour heterogeneity is a significant clinical problem in cancer therapy, and it can potentially be the primary cause of treatment failure and chemo-resistance. This difficulty originates from tumour tissue being a diverse and complicated assemblage of numerous illnesses[18,19]. These tumours exhibit different physical and behavioural characteristics, including histopathology, gene expression, proliferation and metastatic behaviour, and metabolism. Cancer cells' metabolism, such as glycolysis, oxidative phosphorylation, glutamine anaplerosis, and lipid metabolism, are now well highlighted to support their high proliferation demands and growth[290,291]. Importantly, there is considerable cross-talk between genetic and metabolic processes, with *PIK3CA* mutations/hyperactive PI3K signalling altering glucose and glutamine metabolism being prominent examples[292]. These examples highlight the necessity of knowing the relationship between specific genetic events and metabolic abnormalities to uncover novel treatment prospects.

7.2 DESI-MS Imaging Technology in Multi-modal Analysis for Tumor-Heterogeneity Mapping

A definite diagnosis is based on histology information provided by a pathologist and supplemented by medical investigations. Traditional immunohistochemistry (IHC) has been widely employed in the histological

diagnosis of malignant tumour cells. IHC may investigate the location and distribution of specific biomarkers in cancer cells. However, in terms of sensitivity and selectivity, it is a reliable and consistent approach dependent on specific antibodies' availability. The advancement of DESI-MS imaging technology holds considerable potential for addressing the classic IHC's sensitivity and selectivity limitations. The DESI-MSI technology has been gaining significant momentum due to its capability to provide metabolic profiling of thousands of compounds. It has been used to visualize the spatial distribution of medicines, metabolites, peptides, and proteins across various sample surfaces such as tissue slices[188]. This feature obtains biological information from mass spectra in numerous areas across the sample surface[52,293]. Correlating Mass spectroscopy imaging with histological characteristics directly can offer a regionally localized metabolic profile. And this association might be utilized to supplement histological categorization methods[294,295]. In ovarian[35], breast[116] , colorectal[194], prostate[296] and many other tumour types[117,188,295], the lipidomics approach employing DESI-MSI technology has shown to be a powerful tool for separating the metabolic patterns of malignant from morphologically normal tissue. Furthermore, it allows for examining and comprehending metabolic changes in distinct tissue locations. This upsurge in lipidomics discoveries demonstrates the significance of DESI-MSI technology in gaining information on biochemical composition to research tumour heterogeneity by concentrating on patients' metabolic profiles[121,122,124] .

Micro-heterogeneity investigation necessitated an unlabelled yet spatially precise read-out of the tumour's genetic information at a micrometre scale, comprising subclones ranging from a few thousand cancer[114,273]. A combined analysis using DESI-MSI technology and Next Generation Sequencing may be able to meet these requirements and contribute to the creation of a map of molecular ITH in cancerous tissues. To reduce the demand for traditional pathology diagnosis, further research and innovative multi-modal techniques incorporating precise diagnosis are necessary to cope with this clinical tumour-heterogeneity dilemma.

The primary purpose of this study was to evaluate the prospect of extending the molecular mapping capabilities of the DESI-MSI technology to unravel

metabolic heterogeneity in breast and ovarian cancer and integrate multi-omics technologies in unique multi-modal analysis. The specific aims of this project were as follows:

1) To integrate a multi-modal analysis workflow using sequential sections to potentially reveal spatially co-localized ion clusters following DESI-MSI analysis on breast cancer Patient-Derived Xenografts (PDX) models.

2) To evaluate the compatibility of a newly developed hydrogel-based embedding protocol[174] on multiple fresh frozen breast core biopsies in a Tissue Microarray (TMA) block. Thus, engaging the developed multi-modal analysis workflow in one stroke achieves optimal time compared to individual slide preparation and DESI-MSI analysis of the biopsies.

3) To optimize a new multi-modal analysis on a single fresh frozen tissue section from ovarian tumour specimens. Mainly, this methodology aimed to exploit and optimise the integration of gene expression analysis and Liquid Chromatography-Mass Spectrometry (LC-MS) in conjunction with DESI-MSI technology combined with Laser Capture Microdissection (LCM) application on the same single tissue segment.

4) To employ DESI-MSI-guided spatial Transcriptomics using the single tissue section methodology and co-registration of different omics datasets on the same regions of interest (ROIs) to unravel possible interactions between distinct metabolic profiles and specific genetic drivers that can lead to tumour heterogeneity.

By addressing these objectives, DESI-MSI technology can be exploited as a novel strategy for mapping tumour heterogeneity beyond a normal vs cancerous classification tool, as previously demonstrated in the literature[35,116,117,188,194,197,296]. The key findings presented in this thesis primarily

focus on metabolic and lipidomic alterations to map tumour heterogeneity in a primary patient and PDX tissues using a multi-modal analytic workflow.

The findings presented in Chapter 3 intended to expand metabolic heterogeneity mapping by incorporating more than one modality using *PIK3CA* mutant and wild-type PDX tumour models. Following DESI-MSI acquisition, this study has successfully revealed several spatially co-localized ions that have formed distinct clusters in all four consecutive sections. Subsequently, a k-means hierarchical clustering analysis divulges the ability to capture metabolic complexity and heterogeneity between and within tissues. It has been previously shown that hyperactivation of the PI3K/AKT pathway can play a significant role in cancer oncogenesis [24]. Hence, enrichment analysis of metabolites and genes revealed potential novel interactions between metabolic phenotypes and genotypes. This correlation could be a starting point for exploring metabolic vulnerabilities that can be therapeutically exploited in vivo. A topological network analysis provided deeper insights into metabolic pathways related to individual genes by integrating interactive network exploration and functional enrichment analysis from the dataset produced by DESI-MSI-guided transcriptomics. The selected interaction sub-networks displayed in this study were used to visualise the most significant interaction network between genes and metabolites. Interestingly, Ceramides, hydroxy fatty acids, sphingolipid metabolism, linoleic acid metabolism, oxidative phosphorylation, HIF-1 signalling, and, most crucially, arachidonic acid metabolism are beyond the main findings identified by the network enrichment analysis. In this context, linking metabolomics and gene expression data would be more advantageous and could be acknowledged as a promising tool for uncovering new pharmaceutical targets.

The feasibility of the HPMC/PVP embedding methodology for fresh frozen core biopsies was explored in Chapter 4. It was demonstrated that the newly developed HPMC/PVP hydrogel could simplify and optimise the sample preparation steps of small-sized (3 mm) primary breast specimens used for mass spectroscopic imaging, histological analysis, and transcriptome analysis. Some results of this part of the study were recently published[174] and

demonstrated the practicability of collecting multiple datasets in a co-registration mode for biopsies. Another key accomplishment of DESI-MSI analysis was overcoming the batch effect risk. Individual imaging datasets from small or large biopsies are typically characterised by a batch effect, which can significantly obscure the biological interpretation of the data produced. Depending on the instrument used, data collection of individual tissues utilizing MSI methods might take several hours.
Consequently, data acquisition for several core-size samples might take many weeks, possibly resulting in the batch effect. The stated technique enables the construction of fresh frozen tissue microarrays for core biopsies, allowing for consistent description and comparable analytical results across entire sample sets. This study validated that this novel embedding media enabled core biopsies to be subjected to DESI-MSI analysis in a single step, thus assessing and quantifying the abundance of identified metabolites in a co-expression network. Ion vectors' visual representation of tumour mass spectra connected with the gene-metabolite sub-networks could provide a reliable tool for molecular investigation of tumour heterogeneity in core-size biopsies.

The results presented in Chapter 5 indicated that coupling DESI-MSI with LCM microdissection could lead to a follow-up study such as gene expression and LC-MS analysis using the single section. Additionally, several different technical optimisations have been accessed in this chapter. Essentially, three main findings have been summarised: i) an optimisation protocol of RNAse-free H&E staining and fast DESI-MSI run setup, allowing the preservation of RNA quantity and quality so that the same tissue section can be used for follow-up analyses, and ii) an established cold stage set up for DESI-MSI analysis to prevent RNA degradation compared to room temperature sample run; besides achieving a higher RNA amount using a post-DESI tissue section, final iii) an IPA-based sample preparation methodology developed for LC-MS analysis of microdissected tissue regions.

Finally, in Chapter 6, DESI-MSI technology was used to map the lipid compositions, patterns, and distributions of multi-region biopsies from primary ovarian tumours. The metabolic profiling was identified by DESI-MSI

segmentation analysis on the tissue section unravelled a degree of inter/intra-tumour heterogeneity. Succeeding, DESI-MSI analysis produced metabolically distinct ROIs and these unique segments were used for spatial transcriptomics via microdissection on post-DESI tissue sections. The data have displayed that the metabolically identified ROIs offered sufficient quality and quantity material for the subsequent transcriptomics review of the isolated material. In summary, NF-kappa B signalling, PI3K-Akt signalling pathway, Proteoglycans, Glycerophospholipids, Sphingolipids, and, most critically, arachidonic acid metabolism were all seen in both metabolomic and genomic data. To conclude, the results shown in this chapter encourage the assignment of a single section to be used to produce maximum data through the co-registration of various omics on the same segments of the tissues.

7.3 Research Impact

Using the same sample for DESI-MSI technology, histology, and gene expression-related technologies in academia, industry, and medical diagnostics allows for unique information integration between divergent datasets. The importance of DESI-MSI stems from its untargeted nature, which enables the investigation of the distributions of thousands of components in a single experiment. The DESI-MSI clustering methodology provides valuable insights into metabolism and the capture of metabolic heterogeneity by segmentation maps. Furthermore, these untargeted findings usually need validation by complementary approaches such as transcriptome analysis, which places the metabolic data in context with the key signalling pathways. Linking results from various methodologies facilitates a shift from a qualitative approach to a conceptual understanding of the architecture of numerous molecular networks responsible for cellular metabolism in normal and cancer.

7.4 Challenges and Future Work

MS is already playing an important part in the research of '-omics,' as it fits all of the criteria for tackling a variety of complicated tasks, such as high sensitivity, selectivity, throughput, robustness, flexibility, and linear range of

quantification of complex biological samples. Human genomes, transcriptomics, proteomics, metabolomics, and lipidomics data are now abundant in the literature, but their integration has yet to be adequately addressed by computational biology tools. The various spatial resolutions of the different techniques and the necessity for integration with bulk omics data, such as bulk lipidomics and single nucleus RNA-sequencing, present challenges to combining spatial omics techniques. Further technological advancements and computational analyses will transform research by allowing for high-resolution mapping of expression patterns and biomolecules. As a result, researchers will be able to analyse in depth how tumour cells interact with neighbouring cells or the surrounding microenvironment. One major impediment when adopting spatial omics platforms in clinical settings is that these technologies are currently relatively low-throughput, expensive, technically complicated, and might benefit from improved spatial resolution. A persistent difficulty shared by all multi-omics approaches is the requirement for new methodologies to efficiently integrate data across diverse modalities to improve diagnoses and the development of tailored medicine for health monitoring and prevention.

8 References

1. Kamangar, F., Dores, G. M. & Anderson, W. F. Patterns of cancer incidence, mortality, and prevalence across five continents: Defining priorities to reduce cancer disparities in different geographic regions of the world. *Journal of Clinical Oncology* vol. 24 2137–2150 Preprint at https://doi.org/10.1200/JCO.2005.05.2308 (2006).
2. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* **136**, E359–E386 (2015).
3. Polyak, K. Heterogeneity in breast cancer. *Journal of Clinical Investigation* vol. 121 3786–3788 Preprint at https://doi.org/10.1172/JCI60534 (2011).
4. Hammond, M. E. H. *et al.* American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Journal of Clinical Oncology* vol. 28 2784–2795 Preprint at https://doi.org/10.1200/JCO.2009.25.6529 (2010).
5. Langlands, F. E., Horgan, K., Ddodwell, D. & Smith, L. Breast cancer subtypes: Response to radiotherapy and potential radiosensitisation. *British Journal of Radiology* vol. 86 Preprint at https://doi.org/10.1259/bjr.20120601 (2013).
6. Chavez, K. J., Garimella, S. v. & Lipkowitz, S. Triple negative breast cancer cell lines: One tool in the search for better treatment of triple negative breast cancer. *Breast Disease* **32**, 35–48 (2010).
7. Sauter, E. R. Reliable Biomarkers to Identify New and Recurrent Cancer. *European Journal of Breast Health* **13**, 162–167 (2017).
8. Arpino, G. *et al.* Gene expression profiling in breast cancer: A clinical perspective. *Breast* vol. 22 109–120 Preprint at https://doi.org/10.1016/j.breast.2013.01.016 (2013).
9. Reid, B. M., Permuth, J. B. & Sellers, T. A. Epidemiology of ovarian cancer: a review. *Cancer Biology and Medicine* vol. 14 9–32 Preprint at https://doi.org/10.20892/j.issn.2095-3941.2016.0084 (2017).
10. Ligament, B. Special Section : Ovarian Cancer. *American Cancer Society* **45**, 28–43 (2018).
11. Bast, R. C., Hennessy, B. & Mills, G. B. The biology of ovarian cancer: New opportunities for translation. *Nature Reviews Cancer* vol. 9 415–428 Preprint at https://doi.org/10.1038/nrc2644 (2009).
12. Kwak, J. T., Hewitt, S. M., Sinha, S. & Bhargava, R. Multimodal microscopy for automated histologic analysis of prostate cancer. *BMC Cancer* **11**, (2011).
13. Dória, M. L. *et al.* Epithelial ovarian carcinoma diagnosis by desorption electrospray ionization mass spectrometry imaging. *Scientific Reports* **6**, 1–11 (2016).
14. Prat, J. Ovarian carcinomas: Five distinct diseases with different origins, genetic alterations, and clinicopathological features. *Virchows Archiv* vol. 460 237–249 Preprint at https://doi.org/10.1007/s00428-012-1203-5 (2012).
15. Leong, H. S. *et al.* Efficient molecular subtype classification of high-grade serous ovarian cancer. *Journal of Pathology* **236**, 272–277 (2015).
16. Blagden, S. P. Harnessing pandemonium: The clinical implications of tumor heterogeneity in ovarian cancer. *Frontiers in Oncology* **5**, 1–6 (2015).

17. Palmirotta, R., Silvestris, E., D'Oronzo, S., Cardascia, A. & Silvestris, F. Ovarian cancer: Novel molecular aspects for clinical assessment. *Critical Reviews in Oncology/Hematology* **117**, 12–29 (2017).
18. Kossaï, M., Leary, A., Scoazec, J. Y. & Genestie, C. Ovarian Cancer: A Heterogeneous Disease. *Pathobiology* **85**, 41–49 (2018).
19. Braicu, E. I. *et al.* High-grade ovarian serous carcinoma patients exhibit profound alterations in lipid metabolism. *Oncotarget* **8**, 102912–102922 (2017).
20. Bartels, P. H. *et al.* Machine Vision-based Histometry of Premalignant and Malignant Prostatic Lesions. *Pathology Research and Practice* **191**, 935–944 (1995).
21. Barcellos-Hoff, M. H., Lyden, D. & Wang, T. C. The evolution of the cancer niche during multistage carcinogenesis. *Nature Reviews Cancer* vol. 13 511–518 Preprint at https://doi.org/10.1038/nrc3536 (2013).
22. Albini, A. & Sporn, M. B. The tumour microenvironment as a target for chemoprevention. *Nature Reviews Cancer* vol. 7 139–147 Preprint at https://doi.org/10.1038/nrc2067 (2007).
23. Isakoff, S. J. *et al.* Breast cancer-associated PIK3CA mutations are oncogenic in mammary epithelial cells. *Cancer Research* **65**, 10992–11000 (2005).
24. Koundouros, N. *et al.* Metabolic Fingerprinting Links Oncogenic PIK3CA with Enhanced Arachidonic Acid-Derived Eicosanoids. *Cell* **181**, 1596-1611.e27 (2020).
25. Ke, C. *et al.* Metabolic phenotyping for monitoring ovarian cancer patients. *Scientific Reports* **6**, 1–9 (2016).
26. Cuello, M. A., Kato, S. & Liberona, F. The impact on high-grade serous ovarian cancer of obesity and lipid metabolism-related gene expression patterns: the underestimated driving force affecting prognosis. *Journal of Cellular and Molecular Medicine* **22**, 1805–1815 (2018).
27. Yuan TL; Cantley LC. PI3K pathway alterations in cancer. *Oncogene* **27**, 5497–5510 (2008).
28. Fruman, D. A. *et al.* The PI3K Pathway in Human Disease. *Cell* **170**, 605–635 (2017).
29. Engelman, J. A., Luo, J. & Cantley, L. C. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nature Reviews Genetics* **7**, 606–619 (2006).
30. Samuels, Y. *et al.* High Frequency of Mutations of the PIK3CA Gene in Human Cancers. *Science (1979)* **304**, 554 (2004).
31. Burke, J. E. & Williams, R. L. Synergy in activating class I PI3Ks. *Trends in Biochemical Sciences* **40**, 88–100 (2015).
32. Miled, N. *et al.* Mechanism of two classes of cancer mutations in the phosphoinositide 3-kinase catalytic subunit. *Science (1979)* **317**, 239–242 (2007).
33. Lucas, C. L., Chandra, A., Nejentsev, S. & Condliffe, A. M. Europe PMC Funders Group PI3K δ and primary immunodeficiencies. **16**, 702–714 (2017).
34. Leontiadou, H., Galdadas, I., Athanasiou, C. & Cournia, Z. Insights into the mechanism of the PIK3CA E545K activating mutation using MD simulations. *Scientific Reports* **8**, 1–16 (2018).
35. Dória, M. L. *et al.* Epithelial ovarian carcinoma diagnosis by desorption electrospray ionization mass spectrometry imaging. *Scientific Reports* **6**, 1–11 (2016).

36. Beloribi-Djefaflia, S., Vasseur, S. & Guillaumond, F. Lipid metabolic reprogramming in cancer cells. *Oncogenesis* **5**, e189–e189 (2016).
37. Benjamin Chun-Kit Tong. 乳鼠心肌提取 HHS Public Access. *Physiol Behav* **176**, 139–148 (2017).
38. Fahy, E. *et al.* A comprehensive classification system for lipids. *Journal of Lipid Research* **46**, 839–861 (2005).
39. Fahy, E. *et al.* Update of the LIPID MAPS comprehensive classification system for lipids. *Journal of Lipid Research* vol. 50 S9 Preprint at https://doi.org/10.1194/jlr.R800095-JLR200 (2009).
40. Baenke, F., Peck, B., Miess, H. & Schulze, A. Hooked on fat: The role of lipid synthesis in cancer metabolism and tumour development. *DMM Disease Models and Mechanisms* **6**, 1353–1363 (2013).
41. Ackerstaff, E., Glunde, K. & Bhujwalla, Z. M. Choline phospholipid metabolism: A target in cancer cells? *Journal of Cellular Biochemistry* vol. 90 525–533 Preprint at https://doi.org/10.1002/jcb.10659 (2003).
42. van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: Where they are and how they behave. *Nature Reviews Molecular Cell Biology* vol. 9 112–124 Preprint at https://doi.org/10.1038/nrm2330 (2008).
43. Igney, F. H. & Krammer, P. H. Immune escape of tumors: apoptosis resistance and tumor counterattack. *J Leukoc Biol* **71**, 907–20 (2002).
44. Warburg, O. On the origin of cancer cells. *Science (1979)* **123**, 309–314 (1956).
45. Swinnen, J. v. *et al.* Fatty acid synthase drives the synthesis of phospholipids partitioning into detergent-resistant membrane microdomains. *Biochemical and Biophysical Research Communications* **302**, 898–903 (2003).
46. Menendez, J. A. & Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nature Reviews Cancer* vol. 7 763–777 Preprint at https://doi.org/10.1038/nrc2222 (2007).
47. Punnonen, K., Hietanen, E., Auvinen, O. & Punnonen, R. Phospholipids and fatty acids in breast cancer tissue. *Journal of Cancer Research and Clinical Oncology* **115**, 575–578 (1989).
48. Dillon, L. & Miller, T. Therapeutic Targeting of Cancers with Loss of PTEN Function. *Current Drug Targets* **15**, 65–79 (2014).
49. Reinartz, S. *et al.* A transcriptome-based global map of signaling pathways in the ovarian cancer microenvironment associated with clinical outcome. *Genome Biol* **17**, 108 (2016).
50. Tania, M., Khan, M. A. & Song, Y. Association of lipid metabolism with ovarian cancer. *Current Oncology* vol. 17 6–11 Preprint at https://doi.org/10.3747/co.v17i5.668 (2010).
51. Zhou, Y. *et al.* Inhibition of fatty acid synthase suppresses neovascularization via regulating the expression of VEGF-A in glioma. *Journal of Cancer Research and Clinical Oncology* **142**, 2447–2459 (2016).
52. Eberlin, L. S., Ferreira, C. R., Dill, A. L., Ifa, D. R. & Cooks, R. G. Desorption electrospray ionization mass spectrometry for lipid characterization and biological tissue imaging. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* vol. 1811 946–960 Preprint at https://doi.org/10.1016/j.bbalip.2011.05.006 (2011).
53. Lipinski, K. A. *et al.* Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine. *Trends in Cancer* vol. 2 49–63 Preprint at https://doi.org/10.1016/j.trecan.2015.11.003 (2016).

54. Suzuki, H. *et al.* Mutational landscape and clonal architecture in grade II and III gliomas. *Nature Genetics* **47**, 458–468 (2015).
55. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: A looking glass for cancer? *Nature Reviews Cancer* vol. 12 323–334 Preprint at https://doi.org/10.1038/nrc3261 (2012).
56. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* vol. 501 338–345 Preprint at https://doi.org/10.1038/nature12625 (2013).
57. Swanton, C. Intratumor heterogeneity: Evolution through space and time. *Cancer Research* vol. 72 4875–4882 Preprint at https://doi.org/10.1158/0008-5472.CAN-12-2217 (2012).
58. Bashashati, A. *et al.* Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *Journal of Pathology* **231**, 21–34 (2013).
59. Schwarz, R. F. *et al.* Spatial and Temporal Heterogeneity in High-Grade Serous Ovarian Cancer: A Phylogenetic Analysis. *PLoS Medicine* **12**, 1–20 (2015).
60. Lee, J. Y. *et al.* Tumor evolution and intratumor heterogeneity of an epithelial ovarian cancer investigated using next-generation sequencing. *BMC Cancer* **15**, 1–9 (2015).
61. Polyak, K. Review series introduction Heterogeneity in breast cancer. *J.Clin.Invest.* **121**, 2011–2013 (2011).
62. Sun, H. *et al.* Comprehensive characterization of 536 patient- derived xenograft models prioritizes candidates for targeted treatment. (2021).
63. Hidalgo, M. *et al.* Europe PMC Funders Group Patient Derived Xenograft Models : An Emerging Platform for Translational Cancer Research. **4**, 998–1013 (2015).
64. Migliaccio, I. *et al.* PIK3CA co-occurring mutations and copy-number gain in hormone receptor positive and HER2 negative breast cancer. *npj Breast Cancer* **8**, 1–10 (2022).
65. Yoshida, G. J. Applications of patient-derived tumor xenograft models and tumor organoids. *Journal of Hematology and Oncology* **13**, 1–16 (2020).
66. Bruna, A. *et al.* A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. *Cell* **167**, 260-274.e22 (2016).
67. Cassidy, J. W., Batra, A. S., Greenwood, W. & Bruna, A. Patient-derived tumour xenografts for breast cancer drug discovery. *Endocrine-Related Cancer* **23**, T259–T270 (2016).
68. Coussy, F. *et al.* Combination of PI3K and MEK inhibitors yields durable remission in PDX models of PIK3CA-mutated metaplastic breast cancers. *Journal of Hematology and Oncology* **13**, 1–10 (2020).
69. Arentz, G. *et al.* Applications of Mass Spectrometry Imaging to Cancer. *Advances in Cancer Research* **134**, 27–66 (2017).
70. Kim. Mass Spectrometry-based Metabolomics in Translational Research. *Advances in Experimental Medicine and Biology* (2021) doi:10.1007/978-981-33-6064-8_19.
71. McDonnell, L. A., Angel, P. M., Lou, S. & Drake, R. R. *Mass Spectrometry Imaging in Cancer Research: Future Perspectives*. *Advances in Cancer Research* vol. 134 (Elsevier Inc., 2017).

72. Girolamo, F., Lante, I., Muraca, M. & Putignani, L. The Role of Mass Spectrometry in the “Omics” Era. *Current Organic Chemistry* **17**, 2891–2905 (2013).
73. Chughtai, K. & Heeren, R. M. A. Mass Spectrometric Imaging for Biomedical Tissue Analysis - Chemical Reviews (ACS Publications). *Chem Rev* **110**, 3237–3277 (2011).
74. Wenk, M. R. Lipidomics: New tools and applications. *Cell* **143**, 888–895 (2010).
75. Xue. Recent advances in ambient mass spectrometry imaging. *TrAC Trends in Analytical Chemistry* (2019) doi:10.1016/j.trac.2019.115659.
76. Lv, W. & Yang, T. Identification of possible biomarkers for breast cancer from free fatty acid profiles determined by GC-MS and multivariate statistical analysis. *Clinical Biochemistry* **45**, 127–133 (2012).
77. Kim, H. Y., Ma, Y. C. & Wang, T. C. L. Liquid Chromatography/Mass Spectrometry of Phospholipids Using Electrospray Ionization. *Analytical Chemistry* **66**, 3977–3982 (1994).
78. Boxer, S. G., Kraft, M. L. & Weber, P. K. Advances in imaging secondary ion mass spectrometry for biological samples. *Annual Review of Biophysics* **38**, 53–74 (2009).
79. Vickerman, J. C. Molecular imaging and depth profiling by mass spectrometry - SIMS, MALDI or DESI? *Analyst* **136**, 2199–2217 (2011).
80. Fenn, J. B., Mann, M., Meng, C. K. A. I., Wong, S. F. & Whitehouse, C. M. Jllll ’ ’ ~ apilar.
81. Banerjee, S. & Mazumdar, S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry* **2012**, 1–40 (2012).
82. Barber, M., Bordoli, R. S., Sedgwick, R. D. & Tyler, A. N. Fast atom bombardment of solids (F.A.B.): A new ion source for mass spectrometry. *Journal of the Chemical Society, Chemical Communications* **554**, 325–327 (1981).
83. Karas, M. & Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Analytical Chemistry* **60**, 2299–2301 (1988).
84. Karas, M., Bachmann, D. & Hillenkamp, F. Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules. *Analytical Chemistry* **57**, 2935–2939 (1985).
85. Takáts, Z., Wiseman, J. M., Gologan, B. & Cooks, R. G. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science (1979)* **306**, 471–473 (2004).
86. Weston, D. J. Ambient ionization mass spectrometry: Current understanding of mechanistic theory; Analytical performance and application areas. *Analyst* **135**, 661–668 (2010).
87. Takáts. Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology. *J Mass Spectrom* (2005) doi:10.1002/jms.922.
88. Cody, R. B. Versatile New Ion Source for the Analysis of Materials in Open Air under Ambient Conditions. *Anal Chem* (2005) doi:10.1021/ac050162j.
89. Haddad, R., Sparrapan, R., Kotiaho, T. & Eberlin, M. N. Easy ambient sonic-spray ionization-membrane interface mass spectrometry for direct analysis of solution constituents. *Analytical Chemistry* **80**, 898–903 (2008).

90. Nemes, P. & Vertes, A. Laser ablation electrospray ionization for atmospheric pressure molecular imaging mass spectrometry. *Methods in Molecular Biology* **656**, 159–171 (2010).
91. Watson, J. T. & Sparkman, O. D. *Introduction to Mass Spectrometry INTRODUCTION TO Adjunct Professor of Chemistry*. (2007).
92. Yinon, J. & Klein, F. The quadrupole and its applications in vacuum technology and mass spectrometry. *Vacuum* **21**, 379–383 (1971).
93. Glish, G. L. & Vachet, R. W. The basics of mass spectrometry in the twenty-first century. *Nature Reviews Drug Discovery* **2**, 140–150 (2003).
94. Sidoli, S., Kulej, K. & Garcia, B. A. Why proteomics is not the new genomics and the future of mass spectrometry in cell biology. *Journal of Cell Biology* **216**, 21–24 (2017).
95. McDonnell, L. A. & Heeren, R. M. A. Imaging mass spectrometry. *Mass Spectrometry Reviews* vol. 26 606–643 Preprint at https://doi.org/10.1002/mas.20124 (2007).
96. Addie, R. D., Balluff, B., Bovée, J. V. M. G., Morreau, H. & McDonnell, L. A. Current State and Future Challenges of Mass Spectrometry Imaging for Clinical Research. *Analytical Chemistry* **87**, 6426–6433 (2015).
97. Vaysse, P. M., Heeren, R. M. A., Porta, T. & Balluff, B. Mass spectrometry imaging for clinical research-latest developments, applications, and current limitations. *Analyst* **142**, 2690–2712 (2017).
98. Kraft, M. L. & Klitzing, H. A. Imaging lipids with secondary ion mass spectrometry. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1841**, 1108–1119 (2014).
99. Gode, D. & Volmer, D. A. Lipid imaging by mass spectrometry-a review. *Analyst* **138**, 1289–1315 (2013).
100. Taylor, M. J., Lukowski, J. K. & Anderton, C. R. Spatially Resolved Mass Spectrometry at the Single Cell: Recent Innovations in Proteomics and Metabolomics. *J Am Soc Mass Spectrom* **32**, 872–894 (2021).
101. Koichi Tanaka, Hiroaki Waki, Yukata Ido, Satoshi Akita, Yoshikazu Yoshida, T. Y. 1988_Tanaka2.pdf. *Rapid Communications in Mass Spectrometry* vol. 2 151–153 Preprint at http://masspec.scripps.edu/mshistory/timeline/time_pdf/1988_Tanaka2.pdf (1988).
102. Caprioli, R. M., Farmer, T. B. & Gile, J. Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Analytical Chemistry* **69**, 4751–4760 (1997).
103. Ščupáková, K. *et al.* Cellular resolution in clinical MALDI mass spectrometry imaging: The latest advancements and current challenges. *Clinical Chemistry and Laboratory Medicine* **58**, 914–929 (2020).
104. Wiseman, J. M. Tissue Imaging at Atmospheric Pressure Using Desorption Electrospray Ionization (DESI) Mass Spectrometry. *Angew Chem Int Ed* (2006) doi:10.1002/anie.200602449.
105. Do, P. & Byrd, J. C. Mass cytometry: A high-throughput platform to visualize the heterogeneity of acute myeloid leukemia. *Cancer Discovery* **5**, 912–914 (2015).
106. Xiao, X. *et al.* Multiplexed imaging mass cytometry reveals distinct tumor-immune microenvironments linked to immunotherapy responses in melanoma. *bioRxiv* 2022.01.15.476480 (2022).
107. Bendall, S. C. *et al.* NIH Public Access. **332**, 687–696 (2012).

108. Schulz, D. *et al.* Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples by Mass Cytometry. *Cell Systems* **6**, 25-36.e5 (2018).
109. Simmons, A. J. *et al.* Impaired coordination between signaling pathways is revealed in human colorectal cancer using single-cell mass cytometry of archival tissue blocks. *Science Signaling* **9**, 1–13 (2016).
110. Winter, D. R., Ledergor, G. & Amit, I. From mass cytometry to cancer prognosis. *Nature Biotechnology* **33**, 931–932 (2015).
111. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Methods* **11**, 417–422 (2014).
112. Yang, K. & Han, X. Lipidomics: Techniques, Applications, and Outcomes Related to Biomedical Sciences. *Trends in Biochemical Sciences* **41**, 954–969 (2016).
113. Hu, Y. *et al.* Mass spectrometry-based chemical mapping and profiling toward molecular understanding of diseases in precision medicine. *Chemical Science* **12**, 7993–8009 (2021).
114. Swanton, C. Intratumor heterogeneity: Evolution through space and time. *Cancer Research* **72**, 4875–4882 (2012).
115. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345 (2013).
116. Santoro, A. L. *et al.* In situ Desi-MSI lipidomic profiles of breast cancer molecular subtypes and precursor lesions. *Cancer Research* **80**, 1246–1257 (2020).
117. Yang, X. *et al.* In situ DESI-MSI lipidomic profiles of mucosal margin of oral squamous cell carcinoma. *EBioMedicine* **70**, 103529 (2021).
118. Peukert, M. *et al.* Spatially resolved analysis of small molecules by matrix-assisted laser desorption/ionization mass spectrometric imaging (MALDI-MSI). *New Phytologist* **193**, 806–815 (2012).
119. Takáts, Z., Wiseman, J. M. & Cooks, R. G. Ambient mass spectrometry using desorption electrospray ionization (DESI): Instrumentation, mechanisms and applications in forensics, chemistry, and biology. *Journal of Mass Spectrometry* **40**, 1261–1275 (2005).
120. Qin. Recent advances in matrix-assisted laser desorption/ionisation mass spectrometry imaging (MALDI-MSI) forin situanalysis of endogenous molecules in plants. *Phytochemical Analysis* (2018) doi:10.1002/pca.2759.
121. Balluff, B., Hanselmann, M. & Heeren, R. M. A. Mass Spectrometry Imaging for the Investigation of Intratumor Heterogeneity. in *Advances in Cancer Research* vol. 134 201–230 (Academic Press Inc., 2017).
122. Inglese, P. *et al.* Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer. *Chemical Science* **8**, 3500–3511 (2017).
123. Zhang, Q. *et al.* Integrated multiomic analysis reveals comprehensive tumour heterogeneity and novel immunophenotypic classification in hepatocellular carcinomas. *Gut* **68**, 2019–2031 (2019).
124. Gawin, M. *et al.* Intra-tumor heterogeneity revealed by mass spectrometry imaging is associated with the prognosis of breast cancer. *Cancers (Basel)* **13**, 1–14 (2021).
125. Huang. A graphical data processing pipeline for mass spectrometry imaging-based spatially resolved metabolomics on tumor heterogeneity. *Anal Chim Acta* (2019) doi:10.1016/j.aca.2019.05.068.

126. Erlich, Y., Mitra, P. P., Mccombie, W. R. & Hannon, J. sequencing. **5**, 679–682 (2010).
127. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
128. Hedegaard, J. *et al.* Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS ONE* **9**, (2014).
129. Mutuku, S. M. *et al.* Unravelling Prostate Cancer Heterogeneity Using Spatial Approaches to Lipidomics and Transcriptomics. *Cancers (Basel)* **14**, 1702 (2022).
130. Ruiying, C. *et al.* A comprehensive analysis of metabolomics and transcriptomics in non-small cell lung cancer. *PLoS ONE* **15**, 1–16 (2020).
131. Li, S. *et al.* Metabolic Phenotypes of Response to Vaccination in Humans. *Cell* **169**, 862-877.e17 (2017).
132. Kawashima, M. *et al.* High-resolution imaging mass spectrometry combined with transcriptomic analysis identified a link between fatty acid composition of phosphatidylinositols and the immune checkpoint pathway at the primary tumour site of breast cancer. *British Journal of Cancer* **122**, 245–257 (2020).
133. Dilillo, M. *et al.* Mass Spectrometry Imaging, Laser Capture Microdissection, and LC-MS/MS of the Same Tissue Section. *Journal of Proteome Research* **16**, (2017).
134. Dewez, F. *et al.* Precise co-registration of mass spectrometry imaging, histology, and laser microdissection-based omics. *Analytical and Bioanalytical Chemistry* **411**, 5647–5653 (2019).
135. Goodwin, R. J. A. Sample preparation for mass spectrometry imaging: Small mistakes can lead to big consequences. *Journal of Proteomics* **75**, 4893–4911 (2012).
136. Schwartz. Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation. *Journal of Mass Spectrometry* (2003) doi:10.1002/jms.505.
137. Stoeckli, M., Staab, D., Schweitzer, A., Gardiner, J. & Seebach, D. Imaging of a β-Peptide Distribution in Whole-Body Mice Sections by MALDI Mass Spectrometry. *J Am Soc Mass Spectrom* **18**, 1921–1924 (2007).
138. Dannhorn, A. *et al.* Universal Sample Preparation Unlocking Multimodal Molecular Tissue Imaging. *Analytical Chemistry* **92**, 11080–11088 (2020).
139. Deininger, S. O. *et al.* Normalization in MALDI-TOF imaging datasets of proteins: Practical considerations. *Analytical and Bioanalytical Chemistry* **401**, 167–181 (2011).
140. Alexandrov, T. MALDI imaging mass spectrometry: statistical data analysis and current computational challenges. *BMC Bioinformatics* **13 Suppl 1**, (2012).
141. Tibshirani. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B (Methodological)* (1996) doi:null.
142. Morris, J. S., Coombes, K. R., Koomen, J., Baggerly, K. A. & Kobayashi, R. Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics* **21**, 1764–1775 (2005).
143. McCombie, G., Staab, D., Stoeckli, M. & Knochenmuss, R. Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. *Analytical Chemistry* **77**, 6118–6124 (2005).

144. Jones, E. A., Deininger, S. O., Hogendoorn, P. C. W., Deelder, A. M. & McDonnell, L. A. Imaging mass spectrometry statistical analysis. *Journal of Proteomics* **75**, 4962–4989 (2012).
145. Scherf, U. *et al.* A gene expression database for the molecular pharmacology of cancer. *Nature Genetics* **24**, 236–244 (2000).
146. Schwartz, S. A., Weil, R. J., Johnson, M. D., Toms, S. A. & Caprioli, R. M. Protein Profiling in Brain Tumors Using Mass Spectrometry: Feasibility of a New Technique for the Analysis of Protein Expression. *Clinical Cancer Research* **10**, 981–987 (2004).
147. Alexandrov, T. & Kobarg, J. H. Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics* **27**, 230–238 (2011).
148. Luo, D., Ding, C. & Huang, H. Linear discriminant analysis: New formulations and overfit analysis. *Proceedings of the National Conference on Artificial Intelligence* **1**, 417–422 (2011).
149. Yuan, H., Tang, Y. Y., Lu, Y., Yang, L. & Luo, H. Spectral-spatial classification of hyperspectral image based on discriminant analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **7**, 2035–2043 (2014).
150. Veselkov, K. A. *et al.* Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. *Proc Natl Acad Sci U S A* **111**, 1216–1221 (2014).
151. Hilario, M., Kalousis, A., Pellegrini, C. & Müller, M. Processing and classification of protein mass spectra. *Mass Spectrometry Reviews* **25**, 409–449 (2006).
152. Balluff, B. *et al.* Classification of HER2/neu status in gastric cancer using a breast-cancer derived proteome classifier. *Journal of Proteome Research* **9**, 6317–6322 (2010).
153. Deininger, S. O., Ebert, M. P., Fütterer, A., Gerhard, M. & Röcken, C. MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers. *Journal of Proteome Research* **7**, 5230–5236 (2008).
154. Huang, L. *et al.* A graphical data processing pipeline for mass spectrometry imaging-based spatially resolved metabolomics on tumor heterogeneity. *Analytica Chimica Acta* **1077**, 183–190 (2019).
155. Alexandrov, T. & Kobarg, J. H. Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics* **27**, 230–238 (2011).
156. Balluff, B. *et al.* De novo discovery of phenotypic intratumour heterogeneity using imaging mass spectrometry. *Journal of Pathology* **235**, 3–13 (2015).
157. Tillner, J. *et al.* Faster, More Reproducible DESI-MS for Biological Tissue Imaging. *J Am Soc Mass Spectrom* **28**, 2090–2098 (2017).
158. Lightfoot, S. Quantitation comparison of total RNA using the Agilent 2100 bioanalyzer , ribo- green analysis and UV spectrometry Application. 1–8 (2002).
159. Butler, A. E. *et al.* Recovery of high-quality RNA from laser capture microdissected human and rodent pancreas. *Journal of Histotechnology* **39**, (2016).

160. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology* **30**, 777–782 (2012).
161. Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L. & Rice, P. M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research* **38**, 1767–1771 (2009).
162. Jiang, L. *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome Research* **21**, 1543–1551 (2011).
163. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**, 907–915 (2019).
164. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178–192 (2013).
165. Trapnell, C. *et al.* Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nature Biotechnology* **28**, 511–515 (2011).
166. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
167. Anders, S. & Huber, W. Differential expression of RNA-Seq data at the gene level–the DESeq package. *DESq Manual* 1–28 (2012).
168. Wang, L., Feng, Z., Wang, X., Wang, X. & Zhang, X. DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138 (2009).
169. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
170. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**, (2010).
171. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Research* **36**, 480–484 (2008).
172. Mao, X., Cai, T., Olyarchuk, J. G. & Wei, L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**, 3787–3793 (2005).
173. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–1303 (2010).
174. Dannhorn, A. *et al.* Universal Sample Preparation Unlocking Multimodal Molecular Tissue Imaging. *Analytical Chemistry* **92**, 11080–11088 (2020).
175. Inglese, P., Correia, G., Takats, Z., Nicholson, J. K. & Glen, R. C. SPUTNIK: An R package for filtering of spatially related peaks in mass spectrometry imaging data. *Bioinformatics* **35**, 178–180 (2019).
176. Pang, Z. *et al.* MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research* **49**, W388–W396 (2021).
177. Li, S. *et al.* Predicting Network Activity from High Throughput Metabolomics. *PLoS Computational Biology* **9**, (2013).
178. Newman, M. E. J. Modularity and community structure in networks. *Proc Natl Acad Sci U S A* **103**, 8577–8582 (2006).
179. Li, S. Pathway / network analysis for metabolomics. (2014).

180. Huan, T. *et al.* Systems biology guided by XCMS Online metabolomics. *Nature Methods* **14**, 461–462 (2017).
181. Suzuki, H. *et al.* Mutational landscape and clonal architecture in grade II and III gliomas. *Nature Genetics* **47**, 458–468 (2015).
182. Lipinski, K. A. *et al.* Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine. *Trends in Cancer* **2**, 49–63 (2016).
183. Curtis, C., Shah, S. P., Chin, S. & Turashvili, G. Europe PMC Funders Group The genomic and transcriptomic architecture of 2 , 000 breast tumours reveals novel subgroups. **486**, 346–352 (2012).
184. Reina-Campos, M. *et al.* Breast cancer statistics, 2013. *CA: A Cancer Journal for Clinicians* **23**, T259–T270 (2016).
185. Miller, T. W. *et al.* ERα-dependent E2F transcription can mediate resistance to estrogen deprivation in human breast cancer. *Cancer Discovery* **1**, 338–351 (2011).
186. Jacquemetton, J. *et al.* Analysis of genomic and non-genomic signaling of estrogen receptor in PDX models of breast cancer treated with a combination of the PI3K inhibitor alpelisib (BYL719) and fulvestrant. *Breast Cancer Research* **23**, 1–15 (2021).
187. Nowicka, M. CyTOF workflow : differential discovery in Have you analyzed CyTOF or flow cytometry data before ? ! (2017).
188. Thoma, C. Diagnosis: Making DESI-MSI desirable. *Nature Reviews Urology* **14**, 325 (2017).
189. Oetjen, J. *et al.* Benchmark datasets for 3D MALDI- and DESI-imaging mass spectrometry. *Gigascience* **4**, (2015).
190. Mirnezami, R. *et al.* Chemical mapping of the colorectal cancer microenvironment via MALDI imaging mass spectrometry (MALDI-MSI) reveals novel cancer-associated field effects. *Molecular Oncology* **8**, 39–49 (2014).
191. Seeley, E. H. & Caprioli, R. M. MALDI imaging mass spectrometry of human tissue: Method challenges and clinical perspectives. *Trends in Biotechnology* vol. 29 Preprint at https://doi.org/10.1016/j.tibtech.2010.12.002 (2011).
192. St John, E. R., Rossi, M., Pruski, P., Darzi, A. & Takats, Z. Intraoperative tissue identification by mass spectrometric technologies. *TrAC - Trends in Analytical Chemistry* **85**, 2–9 (2016).
193. Abbassi-Ghadi, N. *et al.* Repeatability and reproducibility of desorption electrospray ionization-mass spectrometry (DESI-MS) for the imaging analysis of human cancer tissue: A gateway for clinical applications. *Analytical Methods* **7**, 71–80 (2015).
194. Mroz, A., Doria, L., Claude, E. & Takats, Z. Metabolic Phenotyping of Colorectal Tissue Samples by DESI-MSI for Clinical Research.
195. Calligaris. Application of desorption electrospray ionization mass spectrometry imaging in breast cancer margin analysis. *PNAS* (2014) doi:10.1073/pnas.1408129111.
196. Guenther. Spatially resolved metabolic phenotyping of breast cancer by desorption electrospray ionization mass spectrometry. *Cancer Res* (2015) doi:10.1158/0008-5472.CAN-14-2258.
197. Inglese, P. *et al.* Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer. *Chemical Science* **8**, 3500–3511 (2017).

198. Balluff, B., Hanselmann, M. & Heeren, R. M. A. *Mass Spectrometry Imaging for the Investigation of Intratumor Heterogeneity*. *Advances in Cancer Research* vol. 134 (Elsevier Inc., 2017).
199. Niemelä, P. S., Castillo, S., Sysi-Aho, M. & Orešič, M. Bioinformatics and computational methods for lipidomics. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* vol. 877 2855–2862 Preprint at https://doi.org/10.1016/j.jchromb.2009.01.025 (2009).
200. Santoro, S. *et al.* Laser capture microdissection for transcriptomic profiles in human skin biopsies. *BMC Molecular Biology* **19**, 7 (2018).
201. Eberlin, L. S. *et al.* Desorption electrospray ionization then MALDI mass spectrometry imaging of lipid and protein distributions in single tissue sections. *Analytical Chemistry* **83**, 8366–8371 (2011).
202. Ščupáková, K. *et al.* Cellular resolution in clinical MALDI mass spectrometry imaging: The latest advancements and current challenges. *Clinical Chemistry and Laboratory Medicine* **58**, 914–929 (2020).
203. Chatterjee, S., Balram, A. & Li, W. Convergence: lactosylceramide-centric signaling pathways induce inflammation, oxidative stress, and other phenotypic outcomes. *International Journal of Molecular Sciences* **22**, 1–25 (2021).
204. Pelletier, J., Thomas, G. & Volarevic, S. Players and Therapeutic Avenues. *Nature Publishing Group* **18**, 51–63 (2017).
205. Ebright, R. Y. *et al.* Translation Promotes Breast Cancer Metastasis. *Science (1979)* **1473**, 1468–1473 (2020).
206. Mitchell, C. J. *et al.* Effect of dietary arachidonic acid supplementation on acute muscle adaptive responses to resistance exercise in trained men: A randomized controlled trial. *Journal of Applied Physiology* **124**, 1080–1091 (2018).
207. Brighenti, E. *et al.* Interleukin 6 downregulates p53 expression and activity by stimulating ribosome biogenesis: a new pathway connecting inflammation to cancer. *Oncogene* **33**, 4396–4406 (2014).
208. Adam Moser, Kevin Range, and D. M. Y. *et al.* 基因的改变NIH Public Access. *Bone* **23**, 1–7 (2008).
209. Lin, F. R., Niparko, J. K. & Ferrucci, and L. 基因的改变NIH Public Access. *Bone* **23**, 1–7 (2014).
210. Snider, J. M., Snider, A. J., Obeid, L. M., Luberto, C. & Hannun, Y. A. Probing de novo sphingolipid metabolism in mammalian cells utilizing mass spectrometry. *Journal of Lipid Research* **59**, 1046–1057 (2018).
211. Alessenko, A. v. & Albi, E. Exploring Sphingolipid Implications in Neurodegeneration. *Frontiers in Neurology* **11**, 1–13 (2020).
212. Thomas, J. M. *et al.* The Role of Sphingolipid Signaling in Oxidative Lung Injury and Pathogenesis of Bronchopulmonary Dysplasia. *International Journal of Molecular Sciences* **23**, (2022).
213. Yin, J. *et al.* Potential Mechanisms Connecting Purine Metabolism and Cancer Therapy. *Frontiers in Immunology* **9**, 1–8 (2018).
214. Robichaud, P. P., Boulay, K., Munganyiki, J. É. & Surette, M. E. Fatty acid remodeling in cellular glycerophospholipids following the activation of human T cells. *Journal of Lipid Research* **54**, 2665–2677 (2013).

215. Jira, W., Spiteller, G., Carson, W. & Schramm, A. Strong increase in hydroxy fatty acids derived from linoleic acid in human low density lipoproteins of atherosclerotic patients. *Chemistry and Physics of Lipids* **91**, 1–11 (1997).
216. MacLennan, M. & Ma, D. W. L. Role of dietary fatty acids in mammary gland development and breast cancer. *Breast Cancer Research* **12**, 1–10 (2010).
217. Zock, P. L. & Katan, M. B. Linoleic acid intake and cancer risk: A review and meta-analysis. *American Journal of Clinical Nutrition* **68**, 142–153 (1998).
218. Meilhac, O., Zhou, M., Santanara, N. & Parthasarathy, S. Lipid peroxides induce expression of catalase in cultured vascular cells. *Journal of Lipid Research* **41**, 1205–1213 (2000).
219. Zhang, H. HIF-1 suppresses lipid catabolism to promote cancer progression. *Molecular and Cellular Oncology* **2**, 6–8 (2015).
220. Mylonis, I., Simos, G. & Paraskeva, E. Hypoxia-Inducible Factors and the Regulation of Lipid Metabolism. *Cells* **8**, 214 (2019).
221. Porta Siegel, T. *et al.* Mass Spectrometry Imaging and Integration with Other Imaging Modalities for Greater Molecular Understanding of Biological Tissues. *Molecular Imaging and Biology* **20**, 888–901 (2018).
222. Schwartz, S. A., Reyzer, M. L. & Caprioli, R. M. Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: Practical aspects of sample preparation. *Journal of Mass Spectrometry* **38**, 699–708 (2003).
223. Luis, F. & Moncayo, G. No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title.
224. Nelson, K. A., Daniels, G. J., Fournie, J. W. & Hemmer, M. J. Optimization of whole-body zebrafish sectioning methods for mass spectrometry imaging. *Journal of Biomolecular Techniques* **24**, 119–127 (2013).
225. Pertici, V. *et al.* The use of poly(N-[2-hydroxypropyl]- methacrylamide) hydrogel to repair a T10 spinal cord hemisection in rat: A behavioural, electrophysiological and anatomical examination. *ASN Neuro* **5**, 149–166 (2013).
226. Nguyen, S. N. *et al.* Towards High-Resolution Tissue Imaging Using Nanospray Desorption Electrospray Ionization Mass Spectrometry Coupled to Shear Force Microscopy. *J Am Soc Mass Spectrom* **29**, 316–322 (2018).
227. Dong, Y. *et al.* Sample preparation for mass spectrometry imaging of plant tissues: A review. *Frontiers in Plant Science* **7**, (2016).
228. Schwartz, S. A., Reyzer, M. L. & Caprioli, R. M. Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: Practical aspects of sample preparation. *Journal of Mass Spectrometry* **38**, 699–708 (2003).
229. Nelson, K. A., Daniels, G. J., Fournie, J. W. & Hemmer, M. J. Optimization of whole-body zebrafish sectioning methods for mass spectrometry imaging. *Journal of Biomolecular Techniques* **24**, 119–127 (2013).
230. Miranda ML, E. S. 基因的改変NIH Public Access. *Bone* **72**, 132–5 (2011).
231. Rose Khavari Nicholas Dias, Y. P. 乳鼠心肌提取 HHS Public Access. *Physiol Behav* **176**, 139–148 (2017).
232. Santoro, S. *et al.* Laser capture microdissection for transcriptomic profiles in human skin biopsies. *BMC Molecular Biology* **19**, 1–11 (2018).

233. Seyhan, A. A. & Burke, J. M. Mg2+-independent hairpin ribozyme catalysis in hydrated RNA films. *RNA* **6**, (2000).
234. Serra, M. P., Vorng, J., Inglese, P., Maglennon, G. & Hamm, G. Supplementary Information: ”Universal sample preparation unlocking multimodal molecular tissue imaging” Andreas Dannhorn,. 1–20.
235. Lee, H. C. *et al.* LPIAT1 regulates arachidonic acid content in phosphatidylinositol and is required for cortical lamination in mice. *Molecular Biology of the Cell* **23**, 4689–4700 (2012).
236. Yang, B. *et al.* Yin-yang mechanisms regulating lipid peroxidation of docosahexaenoic acid and arachidonic acid in the central nervous system. *Frontiers in Neurology* **10**, 1–14 (2019).
237. Sakai, M. *et al.* Arachidonic acid and cancer risk: A systematic review of observational studies. *BMC Cancer* **12**, (2012).
238. Borin, T. F., Angara, K., Rashid, M. H., Achyut, B. R. & Arbab, A. S. Arachidonic acid metabolite as a novel therapeutic target in breast cancer metastasis. *International Journal of Molecular Sciences* **18**, 1–23 (2017).
239. Hiraide, T. *et al.* Accumulation of arachidonic acid-containing phosphatidylinositol at the outer edge of colorectal cancer. *Scientific Reports* **6**, 1–11 (2016).
240. Sundar, S., Neal, R. D. & Kehoe, S. Diagnosis of ovarian cancer. *BMJ (Online)* **351**, 1–7 (2015).
241. National Collaborating Centre. Ovarian cancer: The recognition and initial management of ovarian cancer. *Health (San Francisco)* **NICE clini**, (2011).
242. Group, A. P. Diagnosing ovarian cancer sooner: what more can be done? *All Party Parliamentary Group on Ovarian Cancer* (2018).
243. Vickerman, J. C. Molecular imaging and depth profiling by mass spectrometry - SIMS, MALDI or DESI? *Analyst* vol. 136 2199–2217 Preprint at https://doi.org/10.1039/c1an00008j (2011).
244. Schwamborn, K. *The Importance of Histology and Pathology in Mass Spectrometry Imaging*. *Advances in Cancer Research* vol. 134 (Elsevier Inc., 2017).
245. Mustafa, D., Kros, J. M. & Luider, T. Combining laser capture microdissection and proteomics techniques. *Methods in Molecular Biology* **428**, (2007).
246. Xu, B. J. Combining laser capture microdissection and proteomics: Methodologies and clinical applications. *Proteomics - Clinical Applications* vol. 4 Preprint at https://doi.org/10.1002/prca.200900138 (2010).
247. Zimmerman, M., Blanc, L., Chen, P. Y., Dartois, V. & Prideaux, B. Spatial quantification of drugs in pulmonary tuberculosis lesions by laser capture microdissection liquid chromatography mass spectrometry (LCM-LC/MS). *Journal of Visualized Experiments* **2018**, (2018).
248. Xu, B. J., Caprioli, R. M., Sanders, M. E. & Jensen, R. A. Direct analysis of laser capture microdissected cells by MALDI mass spectrometry. *J Am Soc Mass Spectrom* **13**, (2002).
249. Ezzoukhry, Z. *et al.* Combining laser capture microdissection and proteomics reveals an active translation machinery controlling invadosome formation. *Nature Communications* **9**, (2018).
250. Esteve-Codina, A. *et al.* A comparison of RNA-Seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples. *PLoS ONE* **12**, (2017).

251. Wimmer, I. *et al.* Systematic evaluation of RNA quality, microarray data reliability and pathway analysis in fresh, fresh frozen and formalin-fixed paraffin-embedded tissue samples. *Scientific Reports* **8**, 1–17 (2018).
252. Fasulka, B., Free, T. & Support, A. Leica Foil Membrane Slides for Laser Microdissection and Use with the Leica AS LMD, LMD6000 and LMD6500/7000 systems. *Bob Fasulka Cell* (2012).
253. May-Zhang, A. A., Deal, K. K. & Southard-Smith, E. M. Optimization of laser-capture microdissection for the isolation of enteric ganglia from fresh-frozen human tissue. *Journal of Visualized Experiments* **2018**, (2018).
254. Clément-Ziza, M., Munnich, A., Lyonnet, S., Jaubert, F. & Besmond, C. Stabilization of RNA during laser capture microdissection by performing experiments under argon atmosphere or using ethanol as a solvent in staining solutions. *RNA* **14**, 2698–2704 (2008).
255. Perreault, D. M. & Anslyn, E. v. Unifying the Current Data on the Mechanism of Cleavage-Transesterification of RNA. *Angewandte Chemie (International Edition in English)* vol. 36 Preprint at https://doi.org/10.1002/anie.199704321 (1997).
256. Oivanen, M., Kuusela, S. & Lönnberg, H. Kinetics and mechanisms for the cleavage and isomerization of the phosphodiester bonds of RNA by bronsted acids and bases. *Chemical Reviews* **98**, (1998).
257. Fabre, A. L., Colotte, M., Luis, A., Tuffet, S. & Bonnet, J. An efficient method for long-term room temperature storage of RNA. *European Journal of Human Genetics* **22**, (2014).
258. Masotti, A. & Preckel, T. Analysis of small RNAs with the Agilent 2100 Bioanalyzer. *Nature Methods* **3**, (2006).
259. Parrales, A. & Iwakuma, T. p53 as a regulator of lipid metabolism in cancer. *International Journal of Molecular Sciences* vol. 17 Preprint at https://doi.org/10.3390/ijms17122074 (2016).
260. Goldstein, I. & Rotter, V. Regulation of lipid metabolism by p53 - fighting two villains with one sword. *Trends in Endocrinology and Metabolism* vol. 23 567–575 Preprint at https://doi.org/10.1016/j.tem.2012.06.007 (2012).
261. Kap, M., Oomen, M., Arshad, S., de Jong, B. & Riegman, P. Fit for purpose frozen tissue collections by RNA integrity number-based quality control assurance at the Erasmus MC tissue bank. *Biopreservation and Biobanking* **12**, (2014).
262. Goldstein, I. *et al.* P53, a novel regulator of lipid metabolism pathways. *Journal of Hepatology* **56**, 656–662 (2012).
263. Murray, G. I. *Laser Capture Microdissection Methods in Molecular Biology 1723*. (2018).
264. Dannhorn, A. *et al.* metabolites Correlating Mass Spectrometry Imaging and Liquid Chromatography-Tandem Mass Spectrometry for Tissue-Based Pharmacokinetic Studies. 1–14 (2022).
265. Knittelfelder, O. *et al.* Shotgun Lipidomics Combined with Laser Capture Microdissection: A Tool to Analyze Histological Zones in Cryosections of Tissues. *Analytical Chemistry* **90**, 9868–9878 (2018).
266. Zhang, Y., Yuan, Z., Dewald, H. D. & Chen, H. Coupling of liquid chromatography with mass spectrometry by desorption electrospray ionization (DESI). *Chemical Communications* **47**, 4171–4173 (2011).

267. Herrera, J. A. *et al.* Laser capture microdissection coupled mass spectrometry (LCM-MS) for spatially resolved analysis of formalin-fixed and stained human lung tissues. *Clinical Proteomics* **17**, 1–12 (2020).
268. Satomi, Y., Hirayama, M. & Kobayashi, H. One-step lipid extraction for plasma lipidomics analysis by liquid chromatography mass spectrometry. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **1063**, 93–100 (2017).
269. Sarafian, M. H. *et al.* Objective set of criteria for optimization of sample preparation procedures for ultra-high throughput untargeted blood plasma lipid profiling by ultra performance liquid chromatography-mass spectrometry. *Analytical Chemistry* **86**, 5766–5774 (2014).
270. Matyash, V., Liebisch, G., Kurzchalia, T. v., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl-terf-butyl ether for high-throughput lipidomics. *Journal of Lipid Research* **49**, 1137–1146 (2008).
271. Aldana, J., Romero-otero, A. & Cala, M. P. Exploring the lipidome: Current lipid extraction techniques for mass spectrometry analysis. *Metabolites* **10**, 1–32 (2020).
272. Meinhold-Heerlein, I. & Hauptmann, S. The heterogeneity of ovarian cancer. *Archives of Gynecology and Obstetrics* **289**, 237–239 (2014).
273. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* vol. 501 338–345 Preprint at https://doi.org/10.1038/nature12625 (2013).
274. Sah, S. *et al.* Serous Ovarian Cancer Mouse Model. 1–21 (2022).
275. Yang, J. *et al.* Adipocytes promote ovarian cancer chemoresistance. *Scientific Reports* **9**, 1–12 (2019).
276. Harrington, B. S. & Annunziata, C. M. Nf-κb signaling in ovarian cancer. *Cancers (Basel)* **11**, (2019).
277. Camandola, S. *et al.* Nuclear Factor kB Is Activated by Arachidonic Acid but Not by Eicosapentaenoic Acid abundant in cell membranes . Since mammalian tissues are not able to synthetize it , cellular AA rated fatty acid ratio in the membrane has recently been shown to decrease. *Biochemical and biophysical research communication* **229**, 643–647 (1996).
278. Stuhlmeier, K. M., Kao, J. J. & Bach, F. H. Arachidonic acid influences proinflammatory gene induction by stabilizing the inhibitor-κBα/nuclear factor-κB (NF-κB) complex, thus suppressing the nuclear translocation of NF-κB. *Journal of Biological Chemistry* **272**, 24679–24683 (1997).
279. Smith, P. G. *et al.* The Role of Eicosanoids in Gynecological Malignancies. *Frontiers in Pharmacology* **11**, 1–19 (2020).
280. Zhang, W. *et al.* Alpha-Linolenic Acid Exerts an Endothelial Protective Effect against High Glucose Injury via PI3K/Akt Pathway. *PLoS ONE* **8**, (2013).
281. Rinne, N. *et al.* Targeting the PI3K/AKT/mTOR pathway in epithelial ovarian cancer, therapeutic treatment options for platinum-resistant ovarian cancer. *Cancer Drug Resistance* **4**, 573–595 (2021).
282. van der Ploeg, P. *et al.* Signal transduction pathway activity in high-grade serous carcinoma, its precursors and Fallopian tube epithelium. *Gynecologic Oncology* **165**, 114–120 (2022).
283. Wang, Y. C. *et al.* Docosahexaenoic Acid Modulates Invasion and Metastasis of Human Ovarian Cancer via Multiple Molecular Pathways. *International Journal of Gynecological Cancer* **26**, 994–1003 (2016).

284. Ahrens, T. D. *et al.* The Role of Proteoglycans in Cancer Metastasis and Circulating Tumor Cell Analysis. *Frontiers in Cell and Developmental Biology* **8**, 1–27 (2020).
285. Vitale, D. *et al.* Proteoglycans and glycosaminoglycans as regulators of cancer stem cell function and therapeutic resistance. *FEBS Journal* **286**, 2870–2882 (2019).
286. Yang, J. *et al.* Chondroitin sulfate proteoglycan 4, a targetable oncoantigen that promotes ovarian cancer growth, invasion, cisplatin resistance and spheroid formation. *Translational Oncology* **16**, 101318 (2022).
287. Yang, J. *et al.* Chondroitin sulfate proteoglycan 4, a targetable oncoantigen that promotes ovarian cancer growth, invasion, cisplatin resistance and spheroid formation. *Translational Oncology* **16**, 101318 (2022).
288. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2021. *CA: A Cancer Journal for Clinicians* **71**, 7–33 (2021).
289. DeSantis, C., Ma, J., Bryan, L. & Jemal, A. Breast cancer statistics, 2013. *CA: A Cancer Journal for Clinicians* **64**, 52–62 (2014).
290. de Berardinis, R. J. & Chandel, N. S. Fundamentals of cancer metabolism. *Science Advances* **2**, (2016).
291. Pavlova, N. N. & Thompson, C. B. The Emerging Hallmarks of Cancer Metabolism. *Cell Metabolism* **23**, 27–47 (2016).
292. Makinoshima, H. *et al.* Signaling through the phosphatidylinositol 3-kinase (PI3K)/mammalian target of rapamycin (mTOR) axis is responsible for aerobic glycolysis mediated by glucose transporter in epidermal growth factor receptor (EGFR)-mutated lung adenocarcinoma. *Journal of Biological Chemistry* **290**, 17495–17504 (2015).
293. Wiseman, J. M. *et al.* Desorption electrospray ionization mass spectrometry: Imaging drugs and metabolites in tissues. *Proc Natl Acad Sci U S A* **105**, 18120–18125 (2008).
294. Mirnezami, R. *et al.* Chemical mapping of the colorectal cancer microenvironment via MALDI imaging mass spectrometry (MALDI-MSI) reveals novel cancer-associated field effects. *Molecular Oncology* **8**, 39–49 (2014).
295. Takats, Z., Strittmatter, N. & McKenzie, J. S. Ambient Mass Spectrometry in Cancer Research. in *Advances in Cancer Research* vol. 134 231–256 (Academic Press Inc., 2017).
296. Banerjee, S. *et al.* Diagnosis of prostate cancer by desorption electrospray ionization mass spectrometric imaging of small metabolites and lipids. *Proc Natl Acad Sci U S A* **114**, 3334–3339 (2017).

9 List of conference presentations and publications

Conferences Attended

• **Emine Kazanc;** Evdoxia Karali; Vincen Wu; Paolo Inglese, James McKenzie; Aurelian Trip; Nikos Koundouros; Thanasis Tsalikis; Hiromi Kudo; George Poulogiannis; Sadaf Ghaem-Maghami and Zoltan Takats," A new Multimodal Analysis in Breast Cancer: Revealing metabolic heterogeneity using DESI-MS imaging with Laser-microdissection coupled transcriptome approach", **Metabolomics Virtual Conference**, October 27-29, 2020 **(Poster)**

• **Emine Kazanc;** Evdoxia Karali; Vincen Wu; Paolo Inglese, James McKenzie; Aurelian Trip; Nikos Koundouros; Thanasis Tsalikis; Hiromi Kudo; George Poulogiannis; Zoltan Takats," A Multimodal Analysis in Breast Cancer: Revealing metabolic heterogeneity using DESI-MS imaging with Laser-microdissection coupled transcriptome approach", **AACR Virtual Special Conference: Tumor Heterogeneity**: From Single Cells to Clinical Impact, September 17-18,2020 (**Video Poster)**

• **Emine Kazanc**; Evdoxia Karali; Vincen Wu; James McKenzie; Olof Isberg, Andreas Dannhorn; Paolo Inglese, Sadaf Ghaem-Maghami, George Poulogiannis, Zoltan Takats, "A Single Tissue Section Analysis Using Desi-MSI and Laser Capture Microdissection for a Multimodal Cancer Research", **Metabolomics Conference**, The Hague, Netherlands, June 23-27,2019 **(Poster)**

• **Emine Kazanc**; Evdoxia Karali; Vincen Wu; James McKenzie; Olof Isberg, Andreas Dannhorn; Paolo Inglese, Sadaf Ghaem-Maghami, George Poulogiannis, Zoltan Takats, "A Multimodal Approach Using DESI-MSI and Laser Capture Microdissection for Single Tissue Section Analysis" 67th **American Society for Mass Spectrometry**, **ASMS**, Atlanta, Georgia, June 2-6, 2019**(Poster)**

• **Emine Kazanc,** Evi Karali, Vincen Wu, Andreas Dannhorn, Paolo Inglese, James McKenzie, George Poulogiannis and Zoltan Takats," A Combination Analysis of DESI-MSI with Laser Capture Microdissection using Same Tissue Section", **Mass Spectrometry: Applications to the Clinical Lab, MSACL** US, Palm Springs, California- March 31- April 4, 2019, ***Young Investigator Grantee*(Oral)**

• **Emine Kazanc,** Andreas Dannhorn, James McKenzie, Paolo Inglese, Zoltan Takats, "Evaluation of Different Embedding Methods for Fragile and Small Size Clinical Samples in MSI Analysis" Mass Spectrometry: Applications to the

Clinical Lab, MSACL EU, Salzburg, Austria- 9-13, September 2018, ***Young Investigator Grantee (Poster)**

• **Emine Kazanc,** Andreas Dannhorn, Paolo Inglese, James McKenzie, Zoltan Takats, "Evaluation of different Embedding methods for fragile and small size samples in MSI analysis"**, International Mass spectrometry Conference, IMSC,** Italy, Florence – August 26-31, 2018 **(Poster)**

Publications

• Andreas Dannhorn; **Emine Kazanc**;Gregory Hamm;John G. Swales;Nicole Strittmatter;Gareth Maglennon;Richard J. A. Goodwin;and Zoltan Takats "Correlating Mass Spectrometry Imaging and Liquid Chromatography-Tandem Mass Spectrometry for Tissue-Based Pharmacokinetic Studies" Metabolites 2022,12,261, Impact Factor: 4.097 https://doi.org/ 10.3390/metabo12030261

• Andreas Dannhorn; **Emine Kazanc**; Stephanie Ling; Chelsea Nikula; Evdoxia Karali; Maria Paola Serra; Jean-Luc Vorng; Paolo Inglese; Gareth Maglennon; Gregory Hamm; John Swales; Nicole Strittmatter; Simon T. Barry, Owen J. Sansom, George Poulogiannis; Josephine Bunch; Richard J.A. Goodwin; Zoltan Takats. "Universal sample preparation unlocking multimodal molecular tissue imaging" Analytical. Chemistry, ACS, 92, 16, (2020), 11080–11088, Impact Factor: 7.03 https://doi.org/10.1021/acs.analchem.0c00826

• **Emine Kazanc**; Evdoxia Karali; Paolo Inglese, Vincen Wu; George Poulogiannis; Sadaf Ghaem-Maghami; Zoltan Takats, "A Multimodal Analysis in Ovarian Cancer: Revealing metabolic heterogeneity using the same section for DESI-MS imaging with Laser-microdissection coupled transcriptome approach."(**Under Manuscript preparation)**

10 Copyright and Permissions

Figure 1.3.1.1.

Licensed Content

Licensed Content Publisher	Springer Nature
Licensed Content Publication	Oncogene
Licensed Content Title	PI3K pathway alterations in cancer: variations on a theme
Licensed Content Author	T L Yuan et al
Licensed Content Date	Sep 15, 2008

Order Details

Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no

About Your Work

Title	MSI based mapping strategies
Institution name	Imperial College of London
Expected presentation date	Aug 2022

Additional Data

Portions	FIGURE 1

Requestor Location

Emine Kazanc
Imperial College of London

Tax Details

Figure 1.3.1.2.

This Agreement between Emine Kazanc ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	5365470876600
License date	Aug 10, 2022

Printable Details

Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell
Licensed Content Title	The PI3K Pathway in Human Disease
Licensed Content Author	David A. Fruman,Honyin Chiu,Benjamin D. Hopkins,Shubha Bagrodia,Lewis C. Cantley,Robert T. Abraham
Licensed Content Date	Aug 10, 2017
Licensed Content Volume	170
Licensed Content Issue	4
Licensed Content Pages	31

Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

About Your Work

Title	MSI based mapping strategies
Institution name	Imperial College of London
Expected presentation date	Aug 2022

Additional Data

Portions	FIGURE 1

Figure 1.4.1 and Figure 1.4.2

CCC | Marketplace™

This is a License Agreement between Emine Kazanc ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the CCC Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the CCC Terms and Conditions below.

Order Date	25-Apr-2022	**Type of Use**	Republish in a thesis/dissertation
Order License ID	1214842-1	**Publisher**	Royal Society of Chemistry
ISSN	2515-4184	**Portion**	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Molecular Omics	**Start Page**	6
Article Title	Lipid mechanisms in hallmarks of cancer.	**End Page**	18
Date	01/01/2018	**Issue**	1
Rightsholder	Royal Society of Chemistry	**Volume**	16
Publication Type	Journal		

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	**Distribution**	U.K. and Commonwealth (excluding Canada)
Number of charts / graphs / tables / figures requested	2	**Translation**	Original language of publication
Format (select all that apply)	Print	**Copies for the disabled?**	No
Who will republish the content?	Not-for-profit entity	**Minor editing privileges?**	Yes
Duration of Use	Life of current edition	**Incidental promotional use?**	No
Lifetime Unit Quantity	Up to 499	**Currency**	GBP
Rights Requested	Main product		

NEW WORK DETAILS

Title	MSI Based mapping strategies	**Institution name**	Imperial College of London
Instructor name	EmineKazanc	**Expected presentation date**	2022-04-29

Figure 1.5.1

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Apr 25, 2022

This Agreement between Emine Kazanc ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5296180186016
License date	Apr 25, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Genetics
Licensed Content Title	Mutational landscape and clonal architecture in grade II and III gliomas
Licensed Content Author	Hiromichi Suzuki et al
Licensed Content Date	Apr 13, 2015
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute

Figure 1.8.6.4.1.

Thank you for your order.

This Agreement between Emine Kazanc ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

Printable Details

License Number	5296161032662
License date	Apr 25, 2022

Licensed Content

Licensed Content Publisher	Springer Nature
Licensed Content Publication	Journal of Cancer Research and Clinical Oncology
Licensed Content Title	MALDI-imaging segmentation is a powerful tool for spatial functional proteomic analysis of human larynx carcinoma
Licensed Content Author	Theodore Alexandrov et al
Licensed Content Date	Sep 6, 2012

Order Details

Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no

About Your Work

Title	MSI based mapping strategies
Institution name	Imperial College of London
Expected presentation date	May 2022

Additional Data

Portions	Figure 2

Requestor Location

Emine Kazanc
Imperial College of London

Tax Details