

# **VOLATILE HYPOXIA SIGNATURES IN OESOPHAGEAL ADENOCARCINOMA**

**Bibek Das MBBS BSc (Hons) MRCS (Eng)**

Department of Surgery and Cancer, Imperial College London

**v2023**

A thesis submitted to Imperial College London for the degree of Doctorate of Philosophy

**Supervisors**

**Professor George B Hanna PhD FRCS**

**Dr Stefan Antonowicz PhD FRCS**

Department of Surgery and Cancer, Imperial College London

# THESIS DECLARATION

I hereby declare that I am the sole author of this Thesis and that all the work within it is my own. Any individuals who carried out collaborative work are appropriately credited.

1<sup>st</sup> December 2023

Bibek Das

# COPYRIGHT DECLARATION

'The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND). Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.'



# ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to Professor George Hanna, my primary supervisor and mentor. His unwavering commitment to high quality standards in surgery and scientific research has profoundly influenced my philosophy over the past four years.

I extend my sincere appreciation to Mr Stefan Antonowicz whose supervision and guidance has been invaluable. His principles of objectivity, patience and clarity in communication have become the foundations of my approach to research.

I would like to extend sincere appreciation to Professor Robert Bristow for his time and guidance in all discussions related to tumour hypoxia. Special thanks to the OCCAMS collaborative for providing the OCCAMS data, and to Dr Richard Turkington for providing his gene expression data. I owe particular thanks to Mr Aaron Parker who provided vital guidance with mass spectrometry techniques.

My thanks also go to Hanna lab members: Philip Leung, Illaria Belluomo, Antonis Myridakis, Sara Jamel, Yan Mei Goh, Anuja Mitra, Munir Tarazi, Bhamini Vadhvana, Michael Hewitt, Qing Wen, Boyu Xie, Katja Christodolou, Valerio Converso, Jim Ellis, Michael Fadel, Caoimhe Walsh, Guillaume Lafaurie, Pallavi Arya, Sameera Sharma and Dougie Robb.

Most importantly, I owe the utmost gratitude to my family, especially my wife, Sarah, who has supported and encouraged me at every step of this journey.

# ABSTRACT

The UK has one of the highest rates of oesophageal adenocarcinoma (OAC) in the world. Tumour hypoxia is an attractive therapeutic target, but clinically feasible and accurate biomarkers are needed to advance precision medicine trials of hypoxia-targeting therapy (HTT). Mass spectrometric analysis of exhaled volatile organic compounds (VOCs) is a promising approach to diagnose OAC and may also offer a unique opportunity to assess tumour hypoxia non-invasively.

Using experimental and bioinformatic approaches, a novel 12-gene OAC-specific hypoxia-associated gene expression signature (OAC-Hyp) was derived which was prognostic for overall survival in independent cohorts (HR 1.12 (95% CI 1.03-1.22) on meta-analysis). This signature out-performed existing literature gene signatures and identified OAC subgroups with aggressive molecular characteristics.

A pilot clinical phenotyping study was then performed in OAC patients (n = 30) stratified by hypoxic gene expression (OAC-Hyp). This utilised a laser capture microdissection-RNA-seq workflow and a TD-GC-ToF-MS breath analysis pipeline with stringent quality control. A multi-dimensional sequential re-collection pipeline was developed to improve compound identification. This identified eight candidate hypoxia-associated VOCs with good discriminatory performance (AUC 0.888 95% CI 0.770 – 0.975) suitable for future validation studies.

To explore the effects of hypoxia *in vitro*, a multi-omic phenotyping study was performed on OAC cells. A bespoke programming package was developed for the semi-automated pre-processing of biofluid VOC data for this purpose. Hypoxia was found to induce transcriptional metabolic reprogramming which generated distinct lipidomic and volatilomic signatures associated with redox dysfunction.

Finally, an *in silico* screening approach identified PARP inhibitors as candidate HTT in OAC. These findings were validated *in vitro*, providing the rationale for evaluating breath analysis in an active clinical trial.

These data offer insights into the potential of breath analysis to detect OAC tumour hypoxia non-invasively. Future research will explore the translational utility of exhaled VOCs in clinical trials of HTT.

# CONTENTS PAGE

|                                                                                                               |           |
|---------------------------------------------------------------------------------------------------------------|-----------|
| <b>THESIS DECLARATION .....</b>                                                                               | <b>2</b>  |
| <b>COPYRIGHT DECLARATION .....</b>                                                                            | <b>2</b>  |
| <b>ACKNOWLEDGEMENTS .....</b>                                                                                 | <b>3</b>  |
| <b>ABSTRACT .....</b>                                                                                         | <b>4</b>  |
| <b>CONTENTS PAGE.....</b>                                                                                     | <b>5</b>  |
| <b>LIST OF FIGURES .....</b>                                                                                  | <b>10</b> |
| <b>LIST OF TABLES .....</b>                                                                                   | <b>12</b> |
| <b>LIST OF ABBREVIATIONS .....</b>                                                                            | <b>13</b> |
| <b>CHAPTER 1 – INTRODUCTION .....</b>                                                                         | <b>15</b> |
| <b>1.1 Oesophageal adenocarcinoma .....</b>                                                                   | <b>16</b> |
| 1.1.1 Background.....                                                                                         | 16        |
| 1.1.2 Treatment of advanced OAC .....                                                                         | 16        |
| <b>1.2 Tumour hypoxia – a cancer hallmark, mechanism of treatment resistance and target for therapy .....</b> | <b>17</b> |
| 1.2.1 Background.....                                                                                         | 17        |
| 1.2.2 Hypoxia-targeting therapeutics .....                                                                    | 18        |
| 1.2.3 Hypoxia biomarkers.....                                                                                 | 19        |
| 1.2.4 Literature review of hypoxia biomarkers in OAC.....                                                     | 23        |
| <b>1.3 Volatile organic compounds in OAC .....</b>                                                            | <b>28</b> |
| 1.3.1 Background.....                                                                                         | 28        |
| 1.3.2 Breath testing for OG cancer .....                                                                      | 28        |
| 1.3.3 Mass spectrometry techniques relevant for VOC analysis .....                                            | 29        |
| 1.3.4 Mechanisms of VOC production.....                                                                       | 31        |
| 1.3.5 Metabolic reprogramming and redox homeostasis under hypoxia .....                                       | 32        |
| 1.3.6 Lipid metabolism under hypoxia .....                                                                    | 34        |
| 1.3.7 Hypoxia and VOCs.....                                                                                   | 36        |
| <b>1.4 Gaps in the literature .....</b>                                                                       | <b>37</b> |
| <b>1.5 Thesis hypothesis and objectives .....</b>                                                             | <b>38</b> |
| <b>CHAPTER 2 – A HYPOXIA GENE EXPRESSION SIGNATURE FOR OESOPHAGEAL ADENOCARCINOMA .....</b>                   | <b>39</b> |
| <b>Summary.....</b>                                                                                           | <b>40</b> |
| <b>2.1 Methodological rationale.....</b>                                                                      | <b>42</b> |
| 2.1.1 Using a gene expression signature as a reference test for tumour hypoxia .....                          | 42        |
| 2.1.2 Signature derivation and validation .....                                                               | 43        |
| 2.1.3 Choice of oxygen concentration .....                                                                    | 44        |

|                                                                                          |           |
|------------------------------------------------------------------------------------------|-----------|
| <b>2.2 Hypothesis and aims .....</b>                                                     | <b>45</b> |
| <b>2.3 Methods .....</b>                                                                 | <b>46</b> |
| 2.3.1 Review of literature gene signatures.....                                          | 46        |
| 2.3.2 Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) .....       | 48        |
| 2.3.3 Retrospective gene-VOC analysis .....                                              | 49        |
| 2.3.4 Cell culture.....                                                                  | 49        |
| 2.3.5 Hypoxia treatment.....                                                             | 49        |
| 2.3.6 Immunoblot .....                                                                   | 50        |
| 2.3.7 RNA extraction.....                                                                | 50        |
| 2.3.8 RNA sequencing.....                                                                | 51        |
| 2.3.9 Clinical gene expression datasets .....                                            | 51        |
| 2.3.10 Development and validation of an OAC mRNA hypoxia-associated gene signature ..... | 52        |
| 2.3.11 Differential gene expression analysis .....                                       | 54        |
| 2.3.12 Differential immune infiltration analysis .....                                   | 54        |
| 2.3.13 Statistical analysis.....                                                         | 54        |
| <b>2.4 Results.....</b>                                                                  | <b>55</b> |
| 2.4.1 Review of literature hypoxia gene signatures .....                                 | 55        |
| 2.4.3 Development of a novel OAC hypoxia-associated gene signature .....                 | 59        |
| 2.4.4 Prognostic and predictive utility.....                                             | 62        |
| 2.4.5 Differential gene expression analysis .....                                        | 65        |
| 2.4.6 Immune response signature .....                                                    | 65        |
| <b>2.5 Discussion.....</b>                                                               | <b>67</b> |
| <b>CHAPTER 3 – HYPOXIA-ASSOCIATED EXHALED VOCs IN OESOPHAGEAL ADENOCARCINOMA .....</b>   | <b>69</b> |
| <b>Summary.....</b>                                                                      | <b>70</b> |
| <b>3.1 Methodological rationale.....</b>                                                 | <b>72</b> |
| 3.1.1 Pilot study design.....                                                            | 72        |
| 3.1.2 Laser capture micro-dissection.....                                                | 72        |
| 3.1.3 Choice of breath analysis technique.....                                           | 73        |
| 3.1.4 Sequential re-collection principle .....                                           | 73        |
| <b>3.2 Hypothesis and aims .....</b>                                                     | <b>75</b> |
| <b>3.3 Methods .....</b>                                                                 | <b>76</b> |
| 3.3.1 Protocol for the pilot prospective phenotyping study .....                         | 76        |
| 3.3.2 Laser capture microdissection (LCM) .....                                          | 76        |
| 3.3.3 RNA extraction.....                                                                | 78        |
| 3.3.4 RNA sequencing.....                                                                | 78        |
| 3.3.5 Principles of thermal desorption for breath VOC capture.....                       | 79        |
| 3.3.6 Quality assurance .....                                                            | 79        |
| 3.3.7 Preparation of TD tubes for breath sampling.....                                   | 80        |

|                                                                                                       |            |
|-------------------------------------------------------------------------------------------------------|------------|
| 3.3.8 GC-FID method .....                                                                             | 80         |
| 3.3.9 Surrogate addition .....                                                                        | 82         |
| 3.3.10 Breath sampling procedure .....                                                                | 82         |
| 3.3.11 Sample preparation for storage .....                                                           | 84         |
| 3.3.12 Preparation of calibration mixes .....                                                         | 85         |
| 3.3.13 Internal standard addition.....                                                                | 85         |
| 3.3.14 TD-GC-MS methods .....                                                                         | 86         |
| 3.3.15 GC-MS data analysis .....                                                                      | 88         |
| 3.3.16 Statistical analysis.....                                                                      | 90         |
| <b>3.4 Results.....</b>                                                                               | <b>92</b>  |
| 3.4.1 Sample QC and cohort selection .....                                                            | 92         |
| 3.4.4 Sequential re-collection pipeline method development .....                                      | 98         |
| 3.4.2 Targeted GC-MS analysis .....                                                                   | 101        |
| 3.4.3 Untargeted GC-MS analysis .....                                                                 | 105        |
| <b>3.5 Discussion.....</b>                                                                            | <b>109</b> |
| <b>CHAPTER 4 – METABOLIC EFFECTS OF HYPOXIC REPROGRAMMING IN OESOPHAGEAL<br/>ADENOCARCINOMA .....</b> | <b>113</b> |
| <b>Summary.....</b>                                                                                   | <b>114</b> |
| <b>4.1 Methodological rationale.....</b>                                                              | <b>115</b> |
| 4.1.1 In vitro VOC analysis .....                                                                     | 115        |
| 4.1.2 Modelling lipid to VOC dynamics.....                                                            | 115        |
| 4.1.3 Headspace VOC analysis.....                                                                     | 117        |
| 4.1.4 Designing an automated pipeline for GC-MS data analysis.....                                    | 118        |
| <b>4.2 Hypothesis and aims .....</b>                                                                  | <b>119</b> |
| <b>4.3 Methods .....</b>                                                                              | <b>120</b> |
| 4.3.1 Cell culture.....                                                                               | 120        |
| 4.3.2 Hypoxia treatment.....                                                                          | 121        |
| 4.3.3 Immunoblot.....                                                                                 | 121        |
| 4.3.4 RNA extraction and sequencing .....                                                             | 121        |
| 4.3.4 Differential gene expression analysis .....                                                     | 121        |
| 4.3.7 Redox assays (flow cytometry) .....                                                             | 122        |
| 4.3.8 Redox assays (fluorescence microscopy).....                                                     | 122        |
| 4.3.9 Redox assays (antioxidant co-factors) .....                                                     | 122        |
| 4.3.10 Sample preparation for paired lipidomics and volatilomics .....                                | 123        |
| 4.3.11 Ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) method ....                | 123        |
| 4.3.12 Liposome preparation for headspace analysis .....                                              | 124        |
| 4.3.13 Preparation of sorbent probes for headspace sampling .....                                     | 124        |
| 4.3.14 GC-FID method .....                                                                            | 125        |

|                                                                            |            |
|----------------------------------------------------------------------------|------------|
| 4.3.15 Headspace extraction .....                                          | 126        |
| 4.3.16 TD-GC-MS methods .....                                              | 126        |
| 4.3.17 GC-MS data analysis .....                                           | 127        |
| 4.3.18 Statistical analysis.....                                           | 128        |
| <b>4.4 Results.....</b>                                                    | <b>130</b> |
| 4.4.1 Transcriptomics .....                                                | 130        |
| 4.4.2 Lipidomics.....                                                      | 135        |
| 4.4.3 Volatilomics .....                                                   | 139        |
| 4.4.4 Redox analysis .....                                                 | 144        |
| <b>4.5 Discussion.....</b>                                                 | <b>146</b> |
| <b>CHAPTER 5 – TRANSLATIONAL RELEVANCE OF A HYPOXIC VOC SIGNATURE.....</b> | <b>149</b> |
| <b>Summary.....</b>                                                        | <b>150</b> |
| <b>5.1 Methodological rationale.....</b>                                   | <b>151</b> |
| 5.1.1 Hypoxia and response to therapy.....                                 | 151        |
| 5.1.2 PARP inhibitors and contextual synthetic lethality.....              | 151        |
| <b>5.2 Aims .....</b>                                                      | <b>153</b> |
| <b>5.3 Methods .....</b>                                                   | <b>154</b> |
| 5.3.1 In silico prediction of drug response .....                          | 154        |
| 5.3.2 Cell culture.....                                                    | 154        |
| 5.3.3 Therapeutics .....                                                   | 154        |
| 5.3.4 Hypoxia treatment.....                                               | 155        |
| 5.3.5 Immunoblot.....                                                      | 155        |
| 5.3.6 RNA extraction and sequencing .....                                  | 155        |
| 5.3.7 Differential gene expression analysis .....                          | 155        |
| 5.3.8 Cell cycle analysis (flow cytometry).....                            | 155        |
| 5.3.9 Immunofluorescence.....                                              | 156        |
| 5.3.10 Sulforhodamine B (SRB) assay .....                                  | 156        |
| 5.3.11 Clonogenic assay .....                                              | 157        |
| <b>5.4 Results.....</b>                                                    | <b>158</b> |
| 5.4.1 In silico analysis .....                                             | 158        |
| 5.4.2 Evaluation of PARP inhibitors in vitro.....                          | 160        |
| <b>5.5 Discussion.....</b>                                                 | <b>163</b> |
| <b>5.6 SOLAR study .....</b>                                               | <b>164</b> |
| <b>CHAPTER 6 – THESIS CONCLUSIONS AND FUTURE WORK.....</b>                 | <b>165</b> |
| <b>6.1 Conceptual gains from this thesis.....</b>                          | <b>166</b> |
| <b>6.2 Technical gains from this thesis.....</b>                           | <b>168</b> |
| <b>6.3 Future research.....</b>                                            | <b>169</b> |
| 6.3.1 Genetic framework for hypoxia.....                                   | 169        |

|                                                                                                             |            |
|-------------------------------------------------------------------------------------------------------------|------------|
| 6.3.2 Description of the hypoxic volatilome of OAC organoid models.....                                     | 169        |
| 6.3.3 Hypoxia-Driven Response in Oesophageal Adenocarcinoma (HYDRA) study.....                              | 169        |
| 6.3.4 Stakeholder analysis on the breath test-assisted treatment pathway.....                               | 170        |
| <b>References .....</b>                                                                                     | <b>171</b> |
| <b>Appendix 1: PRISMA Flow Chart .....</b>                                                                  | <b>186</b> |
| <b>Appendix 2: REMARK tool to assess reporting criteria of prognostic biomarkers.....</b>                   | <b>187</b> |
| <b>Appendix 3: Primer sequences for the Toustrup hypoxia gene signature .....</b>                           | <b>188</b> |
| <b>Appendix 4: Quality control for cell line RNA sequencing.....</b>                                        | <b>189</b> |
| <b>Appendix 5: Prognostic utility of the Toustrup hypoxia gene signature.....</b>                           | <b>190</b> |
| <b>Appendix 6: Baseline characteristics of patients in OCCAMS and TCGA-ESCA (OAC) datasets..</b>            | <b>191</b> |
| <b>Appendix 7: Ethical approval for biosample collection.....</b>                                           | <b>192</b> |
| <b>Appendix 8: Targeted quantification settings for the TD-GC-ToF-MS mid-polar method .....</b>             | <b>195</b> |
| <b>Appendix 9: Targeted quantification settings for the TD-GC-ToF-MS polar method.....</b>                  | <b>196</b> |
| <b>Appendix 10: compiler.R .....</b>                                                                        | <b>197</b> |
| <b>Appendix 11: composite_file_processing.R.....</b>                                                        | <b>198</b> |
| <b>Appendix 12: biofluids_normalization_filtering.R .....</b>                                               | <b>199</b> |
| <b>Appendix 13: Gene Set Enrichment Analysis of OAC cell lines – Common Hypoxia Regulated Pathways.....</b> | <b>205</b> |
| <b>Appendix 14: oncoPredict drug predictions common to all datasets .....</b>                               | <b>206</b> |
| <b>Appendix 15: Ethical approval for breath sampling during the SOLAR trial.....</b>                        | <b>208</b> |

# LIST OF FIGURES

|                                                                                                                                            |     |
|--------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 1. The need to stratify trials of hypoxia-targeting therapy .....                                                                   | 20  |
| Figure 2. QUIPS tool for risk of bias.....                                                                                                 | 24  |
| Figure 3. Volatile products can be traced to the lipidome.....                                                                             | 32  |
| Figure 4. Maintaining lipid homeostasis under hypoxia .....                                                                                | 36  |
| Figure 5. Summary of OAC hypoxia-associated gene signature derivation method .....                                                         | 41  |
| Figure 6. Western blot analysis of hypoxia pathways in FLO1 cells cultured at 1% O <sub>2</sub> and 0.2% O <sub>2</sub> for 48 hours ..... | 44  |
| Figure 7. Hypoxia scoring equation .....                                                                                                   | 53  |
| Figure 8. Performance of literature hypoxia signatures in vitro .....                                                                      | 56  |
| Figure 9. Performance of literature gene signatures in vivo .....                                                                          | 57  |
| Figure 10. Retrospective analysis of VOC data stratified by the Toustrup hypoxia signature .....                                           | 58  |
| Figure 11. Derivation of the OAC-Hyp novel hypoxia-associated gene signature .....                                                         | 60  |
| Figure 12. Quality evaluation of OAC-Hyp and Toustrup gene signatures .....                                                                | 61  |
| Figure 13. Prognostic utility of OAC-Hyp signature for overall survival in independent datasets .....                                      | 62  |
| Figure 14. Forest plot of univariate hazard ratios (OAC-Hyp) .....                                                                         | 63  |
| Figure 15. Forest plot of multivariate hazard ratios (OAC-Hyp) .....                                                                       | 63  |
| Figure 16. Molecular and immune phenotypes of hypoxic OAC tumours .....                                                                    | 66  |
| Figure 17. Summary of the pilot clinical phenotyping study .....                                                                           | 71  |
| Figure 18. Steps of laser capture microdissection .....                                                                                    | 77  |
| Figure 19. Laser capture microdissection can be used to isolate pure tumour cell populations.....                                          | 77  |
| Figure 20. Complete TD tube workflow .....                                                                                                 | 80  |
| Figure 21. GC-FID QC log file .....                                                                                                        | 81  |
| Figure 22. Breath collection bag.....                                                                                                      | 83  |
| Figure 23. Correct setup of the breath collection bag, TD tubes and gas sampling pump .....                                                | 83  |
| Figure 24. Sequential re-collection workflow .....                                                                                         | 87  |
| Figure 25. Manually setting retention times and integration windows for the targeted analysis.....                                         | 89  |
| Figure 26. Quant Result browser for manual peak checking and quantification .....                                                          | 89  |
| Figure 27. Mid-polar inter-sample ISTD/surrogate variance.....                                                                             | 93  |
| Figure 28. Polar inter-sample ISTD/surrogate variance .....                                                                                | 93  |
| Figure 29. Density plot of acetone abundance on breath sample and room air TD tubes .....                                                  | 94  |
| Figure 30. Quality control for RNA extraction and sequencing.....                                                                          | 95  |
| Figure 31. Summary of patient selection for the pilot clinical phenotyping study .....                                                     | 96  |
| Figure 32. Relative responses after single tube or sequential re-collection cycles (raw, regression model and average) .....               | 99  |
| Figure 33. Relative responses of analytical standards on TD-GCxGC-TOF-MS after re-collection on TD-GC-TOF-MS.....                          | 100 |
| Figure 34. Relative responses of breath VOCs from healthy individuals on GCxGC-TOF-MS after re-collection on GC-MS.....                    | 100 |
| Figure 35. Resolution of co-eluting VOCs in breath after single (1-1) (black) and sequential (2-1) (red) re-collection .....               | 101 |
| Figure 36. Calibration curves for the targeted mid-polar method .....                                                                      | 102 |
| Figure 37. Calibration curves for the targeted polar method .....                                                                          | 103 |
| Figure 38. Significantly different VOCs between hypoxic classes and room air with a targeted analysis .....                                | 104 |
| Figure 39. TD-GC-TOF-MS (mid-polar) untargeted analysis .....                                                                              | 106 |
| Figure 40. TD-GC-TOF-MS (mid-polar) untargeted analysis .....                                                                              | 107 |
| Figure 41. Representative GCxGC-TOF-MS chromatogram .....                                                                                  | 108 |



|                                                                                                                            |     |
|----------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 42. ROC analysis of the 8 VOC model using a Random Forests algorithm .....                                          | 108 |
| Figure 43. Choice of sample matrix for <i>in vitro</i> VOC analysis.....                                                   | 115 |
| Figure 44. Liposome preparation for volatilomic analysis .....                                                             | 116 |
| Figure 45. Volatile products can be traced to the lipidome (dynamic profiling).....                                        | 117 |
| Figure 46. High capacity sorptive extraction-thermal desorption method for static headspace<br>extraction .....            | 118 |
| Figure 47. Final experimental lipid-VOC workflow.....                                                                      | 120 |
| Figure 48. Final standardized GC-MS data pre-processing pipeline.....                                                      | 128 |
| Figure 49. Volcano plots demonstrating differential gene expression analysis of OAC cell lines<br>exposed to hypoxia ..... | 131 |
| Figure 50. Common hypoxia up-regulated genes and STRING protein association network.....                                   | 133 |
| Figure 51. GSVA analysis of regulators of unsaturated fatty acid metabolism .....                                          | 134 |
| Figure 52. Unsupervised and supervised analyses of cell line lipidomic data.....                                           | 136 |
| Figure 53. Core OAC hypoxia-associated lipid signature.....                                                                | 137 |
| Figure 54. Correlation analysis of core lipids associated with the hypoxic response in OAC cells.....                      | 138 |
| Figure 55. PCA score plots before and after GC-MS data pre-processing.....                                                 | 139 |
| Figure 56. PCA, OPLS-DA and permutation plots for VOC data.....                                                            | 140 |
| Figure 57. Representative GCxGC-TOF-MS chromatogram from a pooled QC sample.....                                           | 141 |
| Figure 58. Univariate and ROC curve analysis of VOCs separating hypoxic and normoxic cells .....                           | 142 |
| Figure 59. Lipid-VOC correlation analysis.....                                                                             | 143 |
| Figure 60. Redox state of OAC cells under hypoxia.....                                                                     | 145 |
| Figure 61. PARP inhibitors and synthetic lethality .....                                                                   | 152 |
| Figure 62. Volcano plots showing hypoxia score-imputed drug response correlations in OAC datasets<br>.....                 | 159 |
| Figure 63. DNA repair pathway activity in cell lines under hypoxia .....                                                   | 160 |
| Figure 64. DNA repair pathway in TCGA-ESCA (OAC) stratified by the OAC-Hyp gene signature .....                            | 160 |
| Figure 65. Western blot analysis of key DNA repair proteins under hypoxia in OAC cells.....                                | 160 |
| Figure 66. Cell cycle analysis using BrDU-PI of OAC cells under normoxic and hypoxic conditions...                         | 161 |
| Figure 67. Evaluation of PARP inhibitors in OAC cells under normoxic and hypoxic conditions.....                           | 162 |
| Figure 68. Summary of the SOLAR trial protocol.....                                                                        | 164 |

# LIST OF TABLES

|                                                                                                  |     |
|--------------------------------------------------------------------------------------------------|-----|
| Table 1. Hypoxia biomarkers and prognosis in OAC.....                                            | 24  |
| Table 2. Hypoxia gene signatures from RadiationSigDB .....                                       | 46  |
| Table 3. Microarray expression datasets from GEO Datasets .....                                  | 47  |
| Table 4. Authenticated OAC cell lines used in this work .....                                    | 49  |
| Table 5. Novel hypoxia-associated gene signature (OAC-Hyp) .....                                 | 60  |
| Table 6. Hazard ratios (OS) from Cox proportional hazards model (OCCAMS) .....                   | 64  |
| Table 7. Hazard ratios (OS) for Cox proportional hazards model (TCGA-ESCA (OAC)) .....           | 64  |
| Table 8. Stable isotope-labelled internal standards.....                                         | 86  |
| Table 9. Baseline characteristics of patients in the pilot phenotyping study.....                | 97  |
| Table 10. Candidate exhaled VOCs for OAC hypoxia classification .....                            | 105 |
| Table 11. Common hypoxia up-regulated genes and protein products.....                            | 132 |
| Table 12. Core OAC hypoxia-associated lipid signature .....                                      | 138 |
| Table 13. Significant VOCs separating hypoxic classes on multivariate analysis.....              | 141 |
| Table 14. Significant hypoxia score-imputed drug response correlations for PARP inhibitors ..... | 158 |

# LIST OF ABBREVIATIONS

|           |                                                                            |
|-----------|----------------------------------------------------------------------------|
| CAS       | Chemical Abstracts Service                                                 |
| CV        | coefficient of variance                                                    |
| EMT       | epithelial-mesenchymal transition                                          |
| ESI       | electrospray ionisation                                                    |
| FBS       | foetal bovine serum                                                        |
| GC        | gas chromatography                                                         |
| GEO       | Gene Expression Omnibus                                                    |
| GO        | Gene Ontology                                                              |
| GSEA      | gene set enrichment analysis                                               |
| GSVA      | gene set variation analysis                                                |
| HTT       | hypoxia targeting therapy                                                  |
| IHC       | immunohistochemistry                                                       |
| IF        | immunofluorescence                                                         |
| ISTD      | stable isotope labelled internal standard                                  |
| KEGG      | Kyoto Encyclopaedia of Genes and Genomes                                   |
| LD        | lipid droplet                                                              |
| mRNA      | messenger ribonucleic acid                                                 |
| MF        | NIST library match factor (forward and reverse)                            |
| MS        | mass spectrometry                                                          |
| MW        | molecular weight                                                           |
| m/z       | mass-to-charge ratio                                                       |
| NES       | normalised enrichment scores                                               |
| NIST      | National Institute of Standards and Technology                             |
| OAC       | oesophageal adenocarcinoma                                                 |
| OAC-Hyp   | novel OAC hypoxia-associated gene expression signature                     |
| OCCAMS    | Oesophageal Cancer Clinical And Molecular Stratification                   |
| OG        | oesophago-gastric                                                          |
| OS        | overall survival                                                           |
| OSCC      | oesophageal squamous cell carcinoma                                        |
| PARP      | poly (adenosine diphosphate-ribose) polymerase                             |
| PBS       | phosphate buffered saline                                                  |
| PCA       | principle component analysis                                               |
| PFI       | progression-free interval                                                  |
| (O)PLS-DA | (orthogonal) partial least squares regression                              |
| QC        | quality control                                                            |
| QUIPS     | Quality In Prognosis Studies                                               |
| REMARK    | REporting recommendations for tumour MARKer prognostic studies             |
| ROS       | Reactive oxygen species                                                    |
| RSD       | relative standard deviation                                                |
| RT-qPCR   | real-time quantitative polymerase chain reaction                           |
| SIFT      | selected ion flow tube                                                     |
| SOLAR     | single agent olaparib in the treatment of advanced oesophagogastric cancer |
| SPME      | solid phase microextraction                                                |
| SVM       | support vector machine                                                     |
| TCGA      | The Cancer Genome Atlas                                                    |
| TD        | thermal desorption                                                         |

|      |                                         |
|------|-----------------------------------------|
| ToF  | time of flight                          |
| UPLC | ultra performance liquid chromatography |
| VIP  | variable importance in projection       |
| VOC  | volatile organic compound               |

## **CHAPTER 1 – INTRODUCTION**

## **1.1 Oesophageal adenocarcinoma**

### **1.1.1 Background**

Oesophageal cancer is the eighth most common cancer and the sixth most common cause of cancer death worldwide<sup>1</sup>. The two main histological subtypes of oesophageal cancer are squamous cell carcinoma (OSCC) and adenocarcinoma (OAC). OAC is the dominant subtype in Western countries and is associated with chronic gastro-oesophageal reflux, obesity, and the development of intestinal metaplasia (Barrett's oesophagus) which is pre-malignant<sup>2</sup>. The incidence of OAC in UK has increased 700% since the 1970s and is now one of the highest in the world<sup>3</sup>. The survival of patients with OAC is stage-dependent, with 5-year survival rates ranging from 55% at stage 1 (local disease) to 15% at stage 3 (spread to regional lymph nodes)<sup>4</sup>. Symptoms are usually absent or non-specific in the early stages whereas symptoms such as dysphagia usually develop at later stages. Given the lack of a screening programme to detect early stage OAC, the majority of patients with OAC in the UK currently present with locally advanced or metastatic disease.

### **1.1.2 Treatment of advanced OAC**

Treatment for locally advanced OAC is performed with a curative or palliative intent. Curative treatment is recommended for non-metastatic OAC in fit patients with resectable OAC. This involves a multimodal strategy which integrates systemic treatment with radical surgery. In the UK, the current standard of care for systemic treatment is neoadjuvant chemotherapy (nCT) such as the FLOT regimen<sup>5</sup>. An alternative treatment choice is neoadjuvant chemo-radiotherapy (nCRT) such as the CROSS regimen<sup>6</sup>. The response rates to nCT and nCRT, however, are variable. In the CROSS study, 49% of OSCC patients achieved a pathological complete response compared with just 23% in OAC<sup>6</sup>. Since only responders appear to have a survival benefit from nCT/nCRT<sup>7</sup>, this suggests that the majority of OAC patients receive do not receive benefit from these regimens despite being exposed to the risk of serious adverse events.

Palliative treatment is recommended for inoperable or metastatic OAC. This encompasses the majority of patients diagnosed with OAC in the UK<sup>8</sup>. The current recommended treatment options are chemotherapy +/- radiotherapy<sup>9</sup>. First-line combination chemotherapy involves a platinum-based combination +/- trastuzumab for HER-2 positive tumours. Disease progression after first-line palliative chemotherapy, however, is common. Treatment options after progressive disease involves second-

line chemotherapy or entry into clinical trials of new therapies such as immunotherapy. The KEYNOTE-180 phase II trial of the immunotherapy drug pembrolizumab reported a disease control rate of 40% in OSCC compared with 21% in OAC, and no patients experienced complete response<sup>10</sup>. Survival of patients with metastatic OAC is bleak, with <5% surviving 5-years after diagnosis<sup>11</sup>.

In summary, despite improvements in treatment regimens, advanced OAC still carries a very poor prognosis due to a combination of exceptional resistance to conventional treatments and an aggressive phenotype. This highlights an unmet need to identify and treat the causes of treatment resistance and develop new therapies in OAC.

## **1.2 Tumour hypoxia – a cancer hallmark, mechanism of treatment resistance and target for therapy**

### **1.2.1 Background**

Regional hypoxia is common in solid tumours and is associated with treatment resistance and the development of metastasis<sup>12</sup>. Hypoxia results from the imbalance between the high oxygen demands of oncogene-driven rapid cellular proliferation and inadequate oxygen supply from an abnormal vasculature<sup>12–14</sup>. Two forms of hypoxia have been described in solid tumours – chronic (diffusion-limited) and acute/cycling (perfusion-limited) hypoxia. Oxygen is unable to diffuse through respiring cells farther than ~150µm from blood vessels which generates an oxygen gradient (diffusion-limited chronic hypoxia). Transient changes in blood flow within the disordered tumour vasculature leads to cycles of hypoxia and reperfusion (perfusion-limited cycling hypoxia). Both forms of hypoxia exist within tumours, leading to a wide range of oxygen tensions between 0.02-8%. Given that the normal physiological pO<sub>2</sub> of tissues can vary depending on their metabolic requirements, it is difficult to define an absolute pO<sub>2</sub> value that can be considered hypoxic. Most solid tumours have a lower median pO<sub>2</sub> than their corresponding normal tissue<sup>12</sup>. However, there appears to be a critical pO<sub>2</sub> level below 1% which can have adverse effects for cells and triggers oxygen-sensing adaptive pathways<sup>15</sup>.

Hypoxia mediates treatment resistance through a number of mechanisms. Long diffusion distances from the disordered tumour vasculature limits the delivery and effective dose of chemotherapeutics to hypoxic cells. Radiation resistance of hypoxic cells at pO<sub>2</sub> <0.2% is explained by the oxygen fixation hypothesis. Ionizing radiation reacts with water in cells to generate hydroxyl radicals which causes reversible DNA damage. When free radicals react with molecular oxygen, however, the resulting

peroxyl radicals can cause permanent and irreparable DNA damage. Molecular oxygen thus acts as a radio-sensitiser by 'fixing' radiation-induced damaged<sup>16</sup>. Hypoxia also generates a functional impairment of the cellular immune response, limiting the effectiveness of immunotherapy<sup>17</sup>. These factors collude to protect hypoxic cancer cells from therapeutic cytotoxicity. In addition to treatment resistance, hypoxia is also associated with a transcriptional program that favours an aggressive phenotype and genomic instability<sup>18,19</sup>. Consequently, several clinical studies have found that tumours with a greater hypoxic fraction are associated with poor prognosis, resistance to treatment and more likely to metastasise<sup>20–26</sup>.

### 1.2.2 Hypoxia-targeting therapeutics

Given its negative implications, several therapeutic strategies have been developed to reduce tumour hypoxia<sup>27</sup>. Hypoxia is an appealing target for therapeutic interventions as it is one of the major physiological differences between tumour and healthy tissue. Examples of hypoxia-targeting therapies (HTT) include:

***Oxygen therapy and oxygen mimetics.*** Hyperbaric oxygen therapy has been used with some success to improve response rates to radiotherapy, but it is challenging to implement clinically<sup>28</sup>. Carbogen is the combination of 2% carbon dioxide with 98% oxygen. In combination with nicotinamide, a vasodilator, it can be administered to patients immediately before and during radiotherapy to enhance oxygen delivery to tumours. The combination of carbogen and nicotinamide with radiotherapy in bladder cancer (BCON) has shown long-term benefits in survival and has been approved for clinical use in the UK<sup>29</sup>. Nimorazole is a radiosensitiser which mimics the ability of oxygen to fix radiation-induced DNA damage. Its use was associated with significant improvement in locoregional tumour control in head & neck cancer<sup>30</sup>.

***Hypoxia-activated prodrugs.*** Hypoxia upregulates one-electron reductase enzymes which can be exploited for selective drug delivery by hypoxia-activated prodrugs<sup>31</sup>. The majority of prodrugs under investigation use a nitroaromatic bioreductive group to trigger the release of the therapeutic molecule. An example is TH-302 (evofosfamide) which releases a nitrogen mustard under hypoxia that alkylates basic residues in DNA leading to irreversible damage<sup>32</sup>. In OAC murine xenografts, TH-302 led to a significant radio-sensitising effect with minimal normal tissue toxicity<sup>33</sup>.



**DNA damage response inhibitors.** Severe hypoxia can induce a DNA damage response (DDR) in response to replication arrest<sup>34</sup>. Selectively targeting the DDR in hypoxic regions can sensitize previously resistant hypoxic cells to the effects of DNA damage. For example, VX-970, an inhibitor of the DDR-associated protein ATR, significantly increased the lethality of CRT in hypoxic oesophageal cancer cells in a pre-clinical study<sup>35</sup> and a phase I clinical trial is on-going. The homologous recombination pathway which repairs DNA double strand breaks is also compromised by hypoxia<sup>36,37</sup>. Since poly-(ADP-ribose) polymerase (PARP) is required to repair double standard breaks in homologous recombination-deficient cells, the use of PARP inhibitors such as olaparib to target hypoxic cells is a promising approach<sup>38</sup>. The combination of drugs targeting the DNA damage response with conventional chemo-radiotherapy may be an effective strategy of overcoming hypoxia-induced treatment resistance<sup>39</sup>.

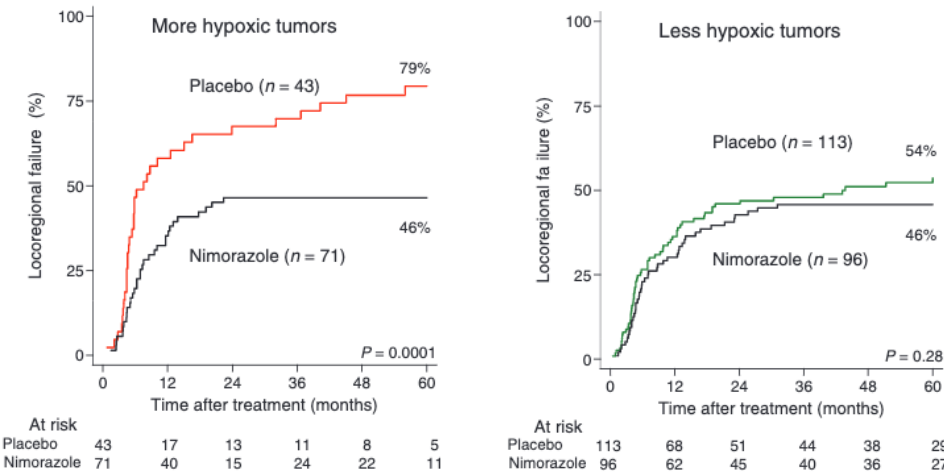
**OXPHOS inhibitors.** Reducing cellular oxygen consumption is a relatively novel method to reduce tumour hypoxia but may be much more effective than measures to increase oxygen delivery<sup>40,41</sup>. Investigational work has focused on mitochondrial complex inhibitors such as papaverine and metformin to induce 'metabolic radiosensitization' of hypoxic tumours<sup>42,43</sup>. Metformin in particular can substantially reduce the fraction of viable radio-resistant hypoxic cells in vitro<sup>42</sup> and has been associated with improved response to CRT in OAC<sup>44</sup>.

### 1.2.3 Hypoxia biomarkers

Although hypoxia is therapeutically actionable by a number of modalities, few hypoxia-targeting therapeutics have progressed beyond clinical trials<sup>45</sup>. A widely held criticism is inadequate stratification<sup>45,46</sup> since studies utilising intratumoural oxygen measurements have shown that the degree of hypoxia varies widely between tumours<sup>13</sup>. Toustrup et al, for example, showed that only a subset of HNSCC patients with very hypoxic tumours benefited from nimorazole in the DAHANCA trial<sup>47</sup> (Figure 1). Since phase III randomized-controlled trials of hypoxia-targeting therapy have not stratified patients for entry by hypoxic status, the power of these trials to detect benefit from hypoxic modification is significantly reduced<sup>40,46</sup>. Therefore, a clinically feasible and accurate method of detecting tumoral hypoxia is needed. The ideal biomarker would be non-invasive, specific for hypoxia and allow serial measurements for treatment monitoring.

Due to the need to stratify entry into clinical trials of targeted therapy, several predictive biomarkers have been developed. The main hypoxia biomarkers are reviewed, and their advantages and disadvantages summarised in Figure 1.

a



b

|                                | Oxygen electrode | Gene signature | Endogenous IHC marker | Exogenous IHC marker | Serum marker | Imaging |
|--------------------------------|------------------|----------------|-----------------------|----------------------|--------------|---------|
| Invasive                       | Yes              | Yes            | Yes                   | Yes                  | No           | No      |
| Cost                           | Medium           | Medium         | Medium                | Medium               | Low          | High    |
| Risk of sampling bias          | Yes              | Yes            | Yes                   | Yes                  | No           | No      |
| Transferable across cancers    | Yes              | No             | Yes                   | Yes                  | Yes          | Yes     |
| Spatial information            | No               | No             | No                    | No                   | No           | Yes     |
| Specificity for tumour hypoxia | High             | Medium         | Medium                | High                 | Low          | High    |
| Serial measurements            | No               | No             | No                    | No                   | Yes          | Yes     |
| In routine clinical use        | No               | No             | No                    | No                   | No           | No      |
| Clinical feasibility           | Low              | High           | High                  | Low                  | High         | Medium  |

Closest to clinical application

**Towards precision medicine trials of hypoxia-modifying therapy.** a, Locoregional failure of HNSCC to radiotherapy +/- nimorazole in the DAHANCA trial stratified by a 15-gene hypoxia gene expression signature (Figure from Toustrup et al. (2011)). b, Biomarkers of tumour hypoxia and attributes which may influence clinical translation (Figure adapted from Yang et al. (2018)). Abbreviations: IHC, immunohistochemistry.

**Figure 1. The need to stratify trials of hypoxia-targeting therapy**

**Oxygen electrodes.** Polarographic oxygen electrodes provide direct measurements of intratumoural oxygen tension. A needle electrode is introduced into the tumour and pO<sub>2</sub> measurements are directly taken with multiple passes. The readings are plotted on a histogram and the proportion of measurements below a certain threshold are reported as the hypoxic fraction. Although it has been considered a 'gold standard' method for measuring hypoxia, its use is limited to easily accessible tumours. Moreover, its application in large phase III trials of HTT is further limited by high cost, inter-operator variability, intratumoural heterogeneity and overestimation of hypoxia in regions of tissue necrosis<sup>12,48</sup>.

**Gene expression signatures.** Hypoxia gene signatures are sets of genes which are up- or down-regulated by hypoxia. Gene expression can be determined from the index biopsy to calculate a 'hypoxia score' which can be used to provide prognostic information and predict the response to HTT<sup>47</sup>. The transcriptional response to hypoxia seems to be tissue- and tumour-site specific<sup>49</sup> therefore over 30 hypoxia gene signatures have been curated from multiple tissue sources with various degrees of overlap<sup>50</sup>. The majority of hypoxia gene signatures have been curated by analysing the differential gene expression of cell lines exposed to normoxia and hypoxia. Alternatively, bioinformatics methods have been used to aggregate highly correlating and prognostic hypoxia-regulated genes from clinical transcriptomic data<sup>50</sup>. One hypoxia gene signature (Toustrup) has been validated against oxygen electrode measurements in human tumours<sup>47</sup>. Compared with other tissue biomarkers, gene signatures are less prone to intratumoral heterogeneity and display consistent expression across multiple biopsies from different tumour regions<sup>51,52</sup>. Large retrospective studies suggest that gene signatures can accurately identify patients who benefit from HTT<sup>47,49,53</sup>. Two gene signatures are undergoing prospective validation in large clinical trials of HTT in HNSCC suggesting that hypoxia gene signatures are closest to clinical application<sup>54–56</sup> although no hypoxia gene signatures have been evaluated in OAC. The clinical limitation of this approach is the requirement for tissue and risk of sampling bias.

**Immunohistochemistry.** Several endogenous markers of hypoxia detected by immunohistochemistry (IHC) have been evaluated. The most well-studied include HIF1 $\alpha$  and the downstream targets CAIX and GLUT-1<sup>57,58</sup>. Increased expression of these markers is associated with tumour hypoxia and worse overall prognosis in several cancers<sup>59–61</sup>. CAIX has a longer half-life than HIF1 $\alpha$  and may be a more stable and specific biomarker of hypoxia than HIF1 $\alpha$  which is regulated by a number of hypoxia-independent pathways<sup>62</sup>. Pimonidazole and pentafluorinated etanidazole (EF5) are exogenous hypoxia biomarkers. After administration, these compounds are selectively reduced in hypoxic cells

and form stable adducts with cellular thiol groups which can be detected on tissue sections<sup>63</sup>. Pimonidazole binding may be more accurate in predicting benefit from HTT in HNSCC than endogenous markers of hypoxia such as CAIX<sup>64</sup>. All IHC markers, however, suffer from the same limitations: a requirement for tissue, inter-observer variability and potential for sampling bias.

**Serum markers.** The theoretical advantage of a serum biomarker is the low cost and ability for serial measurements during therapy to monitor treatment response. Serum osteopontin (OPN) correlates with tumour pO<sub>2</sub> in HNSCC<sup>65</sup> and high levels were associated with treatment failure in the DAHANCA trial of nimorazole<sup>66</sup>. OPN is also highly overexpressed in OAC but its association with clinical outcomes and hypoxia is unknown<sup>67</sup>. Serum biomarkers like OPN, however, are regulated by factors other than hypoxia in tumours and can be influenced by systemic processes such as inflammation<sup>68,69</sup>.

**Imaging.** Functional imaging to detect tumour hypoxia is an attractive modality as it is non-invasive and may provide spatial information to guide targeted radiotherapy<sup>70</sup>. PET-based hypoxia imaging protocols employ radiolabelled compounds that selectively accumulate in hypoxic tissue. These hypoxia tracers have shown good correlation with both oxygen electrode measurements and pimonidazole staining<sup>71–73</sup>. Studies using <sup>18</sup>F-MISO have shown a high prevalence of hypoxia in OC, with a greater prevalence in OAC compared to OSCC<sup>74</sup>. The main limitations include the high cost of the tracers, lack of standardized protocols, limited resolution, and slow uptake after administration<sup>75</sup>. MRI-based methods include standard dynamic contrast-enhanced MRI (DCE-MRI) or blood oxygen level-dependent (BOLD) MRI. These modalities do show some correlation with direct pO<sub>2</sub> measurements in human tumours<sup>76,77</sup> but only provide indirect evidence of tumour hypoxia. Oxygen-enhanced MRI (OE-MRI) can provide dynamic spatial mapping of tumour oxygen delivery. The fraction of tumour tissue refractory to an oxygen challenge correlates strongly with pimonidazole, in contrast to hypoxia metrics derived from DCE-MRI and BOLD-MRI<sup>78</sup>. The clinical utility of these imaging modalities has yet to be clearly defined, however, and MRI has not been used to investigate tumour hypoxia in OC.

**Novel methods.** Several novel methods of assessing hypoxia are being developed. These include photoacoustic tomography<sup>79</sup>, phosphorescence quenching<sup>80</sup>, electron paramagnetic resonance<sup>81</sup>, <sup>19</sup>F-magnetic resonance spectroscopy<sup>82</sup> and Overhauser-enhanced MRI<sup>83</sup>. These remain experimental methods which have predominantly been used in animal models with limited experience in humans.

In summary, no biomarker of tumour hypoxia is validated for clinical use, and most are unsuited to monitoring treatment response. A major problem is the spatio-temporal heterogeneity of hypoxia and the lack of a clear gold-standard reference test<sup>84</sup>. Biomarkers of hypoxia such as CAIX expression, pimonidazole binding and oxygen electrode measurements often do not correlate, which makes validation challenging<sup>64,85</sup>. Consequently, a recent review noted that ‘hypoxia may be the best-validated therapeutic target in oncology that remains to be exploited...[but] the need for biomarkers to guide clinical development of hypoxia-targeted agents is paramount’<sup>27</sup>.

#### **1.2.4 Literature review of hypoxia biomarkers in OAC**

**Aim.** Given the strong need for a reliable hypoxic biomarker in OAC, a systematic review of the literature was performed to identify and compare OAC-specific candidate hypoxia biomarkers.

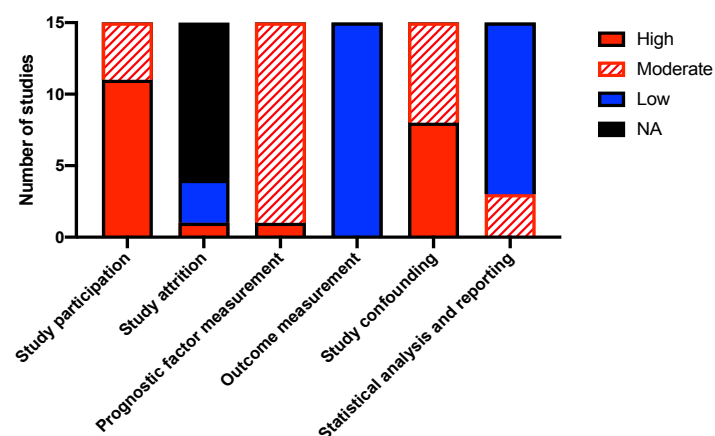
**Methods.** A systematic search of the literature was performed using Pubmed (1951-2023), Embase (1974-2023), Ovid MEDLINE (1948-2023), Web of Science (1970-2023), the Cochrane Database and a compilation of grey literature sources defined previously<sup>86</sup> up to and including 30<sup>th</sup> January 2023. This review has been conducted in line with PRISMA guidelines<sup>87</sup>. The search was conducted using the following MeSH terms separated by the Boolean operators (AND) or (OR): ‘hypox\*’, ‘tumor’, ‘cancer’, ‘neoplasm’, ‘esophag\*’, ‘stomach’, ‘gastric’ and ‘esophagogastric’. Full text articles were then retrieved and assessed to determine eligibility for inclusion. The reference lists of these articles and systematic reviews on hypoxia biomarkers in OG cancer were also searched to identify further potentially relevant articles. Original full-text articles in English were selected if the study evaluated clinical outcomes (survival and/or treatment response) associated with hypoxia biomarkers in adult patients with OAC (including Siewert type I-II junctional tumours). Papers were included if they used either established hypoxia biomarkers<sup>84</sup> or if the authors provided evidence that the biomarker was associated with hypoxia. Studies using only cell lines or animal models were excluded. The primary extraction items were survival and treatment response. Reporting quality in papers with survival data was assessed using the REMARK criteria for prognostic tumour biomarkers<sup>88</sup> and risk of bias with the QUIPS tool<sup>89</sup>. The QUIPS tool is recommended by the Cochrane group and uses six domains to assess risk of bias in prognostic studies: study participation, study attrition (prospective studies only), prognostic factor measurement, outcome measurement, study confounding and statistical analysis and reporting.

| Author                         | Specimen            | Type     | Biomarker           | Threshold                                | Tumour type | Number of samples | Meeting threshold (%) | Independent prognostic factor* |
|--------------------------------|---------------------|----------|---------------------|------------------------------------------|-------------|-------------------|-----------------------|--------------------------------|
| Blayney, 2018 <sup>115</sup>   | Surgical            | IHC      | GLUT-1              | Membranous or cytoplasmic staining       | OAC         | 127               | 69.5                  | Y                              |
| Drenckhan, 2018 <sup>116</sup> | Surgical            | IHC      | CAIX                | Moderate membranous expression           | OAC         | 69                | 37.7                  | -                              |
| Huhta, 2016 <sup>117</sup>     | Surgical            | IHC      | MCT1                | MCT1 histoscore < 150                    | OAC         | 96                | 55.2                  | N                              |
| Huhta, 2016 <sup>117</sup>     | Surgical            | IHC      | MCT4                | MCT4 histoscore > 110                    | OAC         | 98                | 49                    | Y                              |
| Huhta, 2016 <sup>117</sup>     | Surgical            | IHC      | MCT1 and MCT4       | As above                                 | OAC         | 96                | 31.3                  | N                              |
| Huber, 2015 <sup>118</sup>     | Surgical            | IHC      | CAIX                | >10% stromal staining                    | OAC         | 112               | 90                    | -                              |
| Jomrich, 2014 <sup>119</sup>   | Surgical            | IHC      | CAIX                | >10% stromal staining                    | OAC         | 206               | 11.2                  | N                              |
| Shreurs, 2014 <sup>120</sup>   | Patient             | PET scan | <sup>18</sup> F-FDG | SUVmax (FDG) > 3.67                      | OAC & OSCC  | 47 (OAC = 38)     | 23.3                  | -                              |
| Goscinski, 2013 <sup>121</sup> | Surgical            | IHC      | HIF-1α              | >25% nuclear staining                    | OAC         | 28                | 35.7                  | -                              |
| Mirza, 2012 <sup>122</sup>     | Biopsy and surgical | IHC      | Pimonidazole        | >1% cytoplasmic staining                 | OAC         | 34                | 77                    | -                              |
| Birner, 2011 <sup>123</sup>    | Surgical            | IHC      | CAIX                | >median staining intensity staining rate | OAC         | 182               | 46.7                  | Y                              |
| Takala, 2011 <sup>124</sup>    | Surgical            | IHC      | HIF-1α              | >5% nuclear staining                     | OAC & OSCC  | 80 (OAC = 20)     | 71.3                  | -                              |
| Takala, 2011 <sup>124</sup>    | Surgical            | IHC      | VEGF                | >25% cytoplasmic staining                | OAC & OSCC  | 84 (OAC = 20)     | 71.4                  | -                              |
| Ling, 2009 <sup>125</sup>      | Surgical            | IHC      | HIF-1α              | Nuclear staining                         | OAC         | 53                | 45.3                  | -                              |
| Griffiths, 2008 <sup>126</sup> | Surgical            | IHC      | HIF-2α              | Nuclear staining                         | OAC & GC    | 172 (OAC = 92)    | 61.6                  | -                              |
| Griffiths, 2007 <sup>127</sup> | Surgical            | IHC      | HIF-1α              | Nuclear staining                         | OAC & GC    | 176 (OAC = 96)    | 54.5                  | N                              |
| Griffiths, 2007 <sup>127</sup> | Surgical            | IHC      | HIF-1α              | Invasive edge staining                   | OAC & GC    | 173 (OAC = 95)    | 54.9                  | N                              |
| Driessen, 2006 <sup>128</sup>  | Surgical            | IHC      | CAIX                | >median membranous staining              | OAC         | 39                | 49                    | Y                              |
| Driessen, 2006 <sup>128</sup>  | Surgical            | IHC      | VEGF                | >median cytoplasmic staining             | OAC         | 39                | 69.2                  | N                              |
| Ling, 2006 <sup>129</sup>      | Biopsy and surgical | IHC      | HIF-1α protein      | Nuclear staining                         | OAC & OSCC  | 53 (OAC = 22)     | 33.3                  | -                              |
| Ling, 2006 <sup>129</sup>      | Biopsy and surgical | RNA      | HIF-1α mRNA         | Ratio tumour-to-normal epithelium        | OAC         | 22                | 15.4                  | -                              |

\*Defined as prognostic (overall survival) on multivariate analysis. Abbreviations: CAIX, carbonic anhydrase 9; <sup>18</sup>F-FDG, fluorodeoxyglucose; GC, gastric cancer; GLUT-1, glucose transporter 1; IHC, immunohistochemistry; MCT1/4, monocarboxylate transporter 1/4; OAC, oesophageal adenocarcinoma; OSCC, oesophageal squamous cell carcinoma; PET, positron emission tomography; RNA, ribonucleic acid.

**Table 1. Hypoxia biomarkers and prognosis in OAC**

**Results & Discussion.** A total of 3728 papers were screened and 16 relevant studies were identified. The search strategy is summarised in a PRISMA flow chart (Appendix 1) and Table 1. There were fewer OAC hypoxia biomarker studies, compared to OSCC (n = 37) and GC (n = 68). The majority (11/16) were retrospective IHC studies on archival surgical specimens correlated to clinical endpoints. No studies used direct tumour oxygen measurements as a reference standard. Included papers contained survival data only (n = 14), treatment response data only (n = 1) or both (n = 1). Four studies identified biomarkers which were multivariate independent prognostic factors. One study (Blayney, 2018) externally validated the prognostic utility of the biomarker. No biomarkers with prognostic or predictive potential were evaluated in further studies, trials or adopted clinically.



**Figure 2. QUIPS tool for risk of bias**

The combined risk of bias was assessed using the QUIPS tool (Figure 2). The majority of studies showed high risk of bias in three domains: study participation (domain 1), prognostic factor measurement (domain 3) and study confounding (domain 5). Tissue was often selected only from treatment-naïve surgical specimens, and/or no information was given on how samples were included or excluded; these studies were ranked as having a high risk of bias in domain 1. Despite long time periods between sample retrieval and analysis, no studies reported on assay validity, particularly the use of quality control procedures, stability of markers, antibody validity, reproducibility and reporting protocols. Therefore, these were deemed at moderate to high risk of bias in domain 3. Moreover, several studies did not consider the influence of confounding variables or account for them using regression models. For example, several studies analysed OAC in a pooled cohort with other cancers (OSCC or GC). These studies were ranked as having moderate to high risk of bias in domain 5. Reporting quality was reviewed according to the REMARK criteria (Appendix 2). The lowest scoring criteria were details of specimen characteristics and assay methods. The majority of papers did not perform multivariate analysis to estimate the association of a biomarker with clinical outcome after accounting for other important prognostic variables (e.g. age, gender or stage). Only one paper validated the prognostic ability of their biomarker in an independent dataset<sup>90</sup>.

Mirza and colleagues performed the only pimonidazole-based study in gastro-oesophageal adenocarcinoma, which currently remains unpublished<sup>91</sup>. Despite recruiting 57 patients, only 34 patients had samples which were available for analysis. The remainder were necrotic or normal tissue, which highlights problems with the sampling methodology used in this study. Despite these shortcomings, this study provides direct evidence of tumour hypoxia in OAC and suggests that this is prevalent (77% were rated as pimonidazole positive). Unfortunately, the final sample size was too small to draw conclusions on clinical outcomes.

The majority of studies used IHC biomarkers of HIF proteins and their downstream products (GLUT-1, CAIX, VEGF and MCT4). HIF1 $\alpha$  can be stabilised through numerous hypoxia-independent pathways and therefore its utility as a specific hypoxia biomarker may be limited. Most studies did not find an association between HIF1 $\alpha$  and prognosis, although Griffiths et al. found that a subgroup of tumours with invasive edge staining carried a negative prognosis compared to focal positivity which had a positive effect<sup>92</sup>. The authors speculated that the balance between pro- and antiapoptotic effects of HIF1 $\alpha$  may explain this difference. Griffiths et al. later reported that HIF2 $\alpha$ , which may activate different hypoxia pathways and be more specific for hypoxia, carries univariate but not multivariate

prognostic significance in OG cancer<sup>93</sup>. Blayney et al. evaluated gene expression in 60 pre-treatment OAC biopsies and found that the HIF signalling pathway, which included *SLC2A1*/GLUT-1, was the top ranked gene set associated with pathological non-response to neoadjuvant chemotherapy. GLUT-1 and its overexpression was highly prognostic in both training and validation datasets, although the study was limited by its use of pre-chemotherapy tissue for discovery and post-chemotherapy tissue for validation. MCT4 is a HIF-regulated protein involved in lactate transport and its high expression was found to be a significant multivariate predictor of survival in one study<sup>94</sup>, although this has not been externally validated. Finally, CAIX is a classical hypoxia biomarker but its prognostic significance in OAC was variable. In one study only moderate CAIX expressing tumours carried negative prognosis whereas strong or weak expression did not<sup>95</sup>. The reasons for the variability in outcomes are likely common to all included IHC studies - intratumoral heterogeneity, different grading criteria, immunostaining protocols and assay validity. A greater emphasis on standardisation and reproducibility will be needed to validate IHC hypoxia biomarkers in the future.

Shreurs et al. used <sup>18</sup>FDG-PET in a predominantly adenocarcinoma cohort of OC patients and found that a high SUV<sub>max</sub> was significantly associated with disease-free survival. There was no correlation, however, between HIF1α or GLUT-1 and SUV<sub>max</sub> in OAC, in contrast to OSCC, which may be due to differential FDG uptake in mucous producing cells<sup>96</sup>. The use of <sup>18</sup>FDG-PET to detect hypoxia in OAC is now being superseded by hypoxia-specific tracers<sup>97,98</sup>.

The association between hypoxia and treatment response was only evaluated in two studies. Ling et al. found no association between HIF1α protein or mRNA expression in OAC and tumour regression after radiotherapy although the sample size was small (n = 22)<sup>99</sup>. Smit et al. showed that CD44<sup>+</sup>/CD24<sup>-</sup> tumour cells localised to hypoxic zones detected by pimonidazole in a xenograft model of OAC<sup>100</sup>. In 27 pre-CRT biopsies, CD44<sup>+</sup>/CD24<sup>-</sup> cells were only identified in non-responders (50%) and were not found in responders (0%). There is evidence in the literature to suggest that this cell population is expanded by hypoxia in other tumour types<sup>101,102</sup>. Further work is required to establish the role of CD44<sup>+</sup>/CD24<sup>-</sup> tumour cells as biomarkers of hypoxia and radiation response.

**Conclusion.** The aim of this systematic review was to identify clinically validated biomarkers of tumour hypoxia in OAC. No studies in this review correlated direct oxygen measurements from OAC with clinical outcomes in humans (OAC is less accessible than head and neck tumours, in which the direct oxygen measurement approach has been implemented). Aside from pimonidazole, however, the association between most of the included biomarkers and actual intratumoural oxygen levels is



unknown. This is a recognised weakness of most hypoxia biomarker studies<sup>56,84</sup>. Furthermore, no studies assessed whether hypoxia biomarkers could predict benefit from HTT.

In summary, despite evidence that hypoxia is prevalent in OAC, this review found relatively few studies evaluating tumour hypoxia and clinical outcomes in OAC. The limited data from these studies suggests that hypoxia may be associated with worse prognosis and advanced tumour stage in OAC, but the risk of bias in these studies is high. Future studies must address the methodological limitations highlighted by this review such as standardized sampling protocols, validation against a reference test, external validation, and a clear demonstration of clinical utility e.g. prediction of benefit from HTT.

## 1.3 Volatile organic compounds in OAC

### 1.3.1 Background

Volatile organic compounds (VOCs) are carbon-containing compounds that have a vapour pressure ( $>0.01$  kPa) at room temperature (293.15 K). Over 3000 VOCs can be detected in human samples, including exhaled breath, exhaled breath condensate, skin secretions, urine, stool, and breast milk<sup>103</sup>. 'Volatilomics' is a subset of metabolomics based on the study of endogenously produced VOCs, also known as volatile organic metabolites. In recent decades, interest in measuring the changes in VOC concentration in pathological states has grown as this could permit the non-invasive detection and monitoring of various human diseases, including cancer.

### 1.3.2 Breath testing for OG cancer

Early cancer detection often leads to better survival, because pre-malignant and early-stage cancers can be more effectively treated<sup>104</sup>. Alerting symptoms of upper GI cancer, however, are often associated with advanced disease, whereas symptoms in early stages are typically non-specific or absent. Endoscopy is the gold-standard diagnostic test for upper GI malignancy but is an invasive procedure with relatively high cost. This precludes its use as an unselected screening modality. There is an evident need to develop a non-invasive and low-cost screening tool to detect early upper GI cancer in the primary care setting.

Exhaled breath analysis has high translational utility, as breath collection is safe, acceptable to patients, easily repeatable and cost-effective at scale. The MAGIC trial evaluated the feasibility and acceptability of breath collection in primary care among 1002 patients<sup>105</sup>. This showed that breath testing received almost universal acceptance from patients and primary care physicians and demonstrated the feasibility of conducting high-quality breath analysis of large numbers of samples from primary care.

Three studies have evaluated the diagnostic potential of breath VOCs in OAC<sup>106–109</sup>. Kumar et al. reported the first clinical data in a pilot study which demonstrated the potential of exhaled VOCs (fatty acids and phenols) for OG cancer detection<sup>108</sup>. A follow-up study was conducted involving 48 patients with oesophageal adenocarcinoma, 33 patients with gastric adenocarcinoma, 16 patients with Barrett's metaplasia, 62 patients with benign upper gastrointestinal diseases, and 51 patients with a

normal upper gastrointestinal tract. Twelve VOCs (a mixture of aldehydes, fatty acids, and phenols) were found to accurately discriminate cancer patients from healthy controls and benign disease (AUROC 0.87). Finally, a follow-up multi-centre clinical study by Markar et al. included an analysis of breath VOC data from 163 OG cancer patients and 172 control patients which demonstrated the diagnostic potential of a simplified 5-VOC diagnostic model (butanal, decanal, butanoic acid, hexanoic acid and pentanoic acid) (AUROC 0.85). These studies demonstrate the potential of breath analysis in the detection of OG cancer. Trials are now underway to evaluate the potential of breath testing to diagnose high-grade dysplasia and early cancer (NCT04001478).

### 1.3.3 Mass spectrometry techniques relevant for VOC analysis

Mass spectrometry (MS) is a powerful analytical technique that is used to identify and quantify analytes in a sample based on their mass and chemical structure. Different MS approaches with various sampling, separation and ionization techniques have been used in the field of metabolomics. Gas Chromatography (GC) coupled with MS is considered the gold-standard technique for the analysis for VOCs, whereas Liquid Chromatography (LC)-MS is the preferred approach for non-volatile metabolites<sup>110</sup>.

**Gas Chromatography-Mass Spectrometry.** GC-MS begins with vaporisation of a sample which is then injected into a GC column. Here, compounds are separated based on their interaction with the column's stationary phase and their volatility. As compounds exit the GC column, they are ionized, typically through electron ionization (EI) in the MS unit. The ionized molecules are fragmented, producing a unique spectrum of ion fragments that are detected and analysed. The mass analyser sorts these fragments, generating a mass spectrum for each compound. Compounds are identified by comparing their retention indices and mass spectra to reference libraries. GC-MS offers high sensitivity, specificity, and the ability to analyse complex mixtures. Challenges include matrix effects that can influence ionization efficiency and the need for careful calibration and maintenance to ensure accurate, reliable results<sup>111</sup>.

**Thermal desorption-GC-MS.** Thermal desorption (TD) technology enhances the capability of GC-MS for analysing VOCs by efficiently pre-concentrating and introducing trace-level VOCs into the GC-MS system<sup>112</sup>. In this process, a sample is heated in a TD unit, causing the compounds to vaporize from the sorbent. These vapours are then either trapped in a focusing trap or directly introduced to the GC column for subsequent analysis. This process improves analytical sensitivity and minimizes sample

preparation steps. Combining GC-MS with TD provides a robust platform for the comprehensive analysis of VOCs, enabling accurate identification and quantification of these compounds even in complex samples and at trace-level concentrations.

**Liquid Chromatography-Mass Spectrometry.** Unlike GC-MS, which is tailored for volatile and semi-volatile compounds, LC-MS is particularly suited for analysing large, non-volatile, and polar molecules, including metabolites, proteins, peptides, and other biomolecules. This makes LC-MS less traditionally associated with VOC analysis compared to GC-MS. However, LC-MS can be applied to certain volatile compounds when they are part of larger, non-volatile complexes or when they have been derivatized into less volatile derivatives<sup>113</sup>. In the LC component, compounds in a liquid sample are separated based on their interactions with a stationary phase and a mobile phase. The separation is influenced by factors such as polarity, charge, and molecular size. After separation, the compounds are ionized, typically using techniques like Electrospray Ionization (ESI) or Atmospheric Pressure Chemical Ionization (APCI), and then introduced into the MS. The MS analyses these ions based on their mass-to-charge ratio ( $m/z$ ), enabling precise identification and quantification.

**Quadrupole vs time-of-flight mass analysers.** Among mass analysers used in VOC analysis, quadrupole, and time-of-flight (TOF) MS are two important types, each with unique characteristics and applications. Quadrupole MS uses four parallel metal rods connected to DC (direct current) and RF (radiofrequency) voltage which create an oscillating electric field. The stability of ions through this field depends on the ion's mass-to-charge value and the applied electrical voltages. Varying the RF and DC voltage (scanning) enables selective ion transfer and detection. By sequentially allowing ions of varying  $m/z$  to be detected, this permits mass analysis. In contrast, in a TOF mass spectrometer, an electric field accelerates ions to the same kinetic energy, irrespective of mass. These ions then pass through a drift tube to the detector. Due to equal kinetic energy, heavier ions move slower than lighter ones. The detector measures the time taken by each ion to arrive, which is proportional to the square root of the ion's mass, enabling the determination of its  $m/z$  ratio. This allows TOF spectrometers to detect all mass-to-charge ratios simultaneously, offering efficient mass analysis<sup>114</sup>.

TOF mass spectrometers surpass quadrupole systems in several areas<sup>115</sup>. They offer a wider mass range and higher resolution, crucial for analysing large biomolecules and complex mixtures such as breath. Their capability to simultaneously detect all  $m/z$  ions enables faster sample analysis and suits high-throughput settings and quick separation methods. TOF devices also provide superior sensitivity and accuracy, with precise flight time measurements essential for identifying unknowns and detailing

molecular structures. Additionally, their broad dynamic range facilitates the detection of ions at varying concentrations within a single run, vital for quantifying compounds in mixed samples. Overall, these characteristics make TOF-MS the preferred choice for untargeted VOC analysis.

#### **1.3.4 Mechanisms of VOC production**

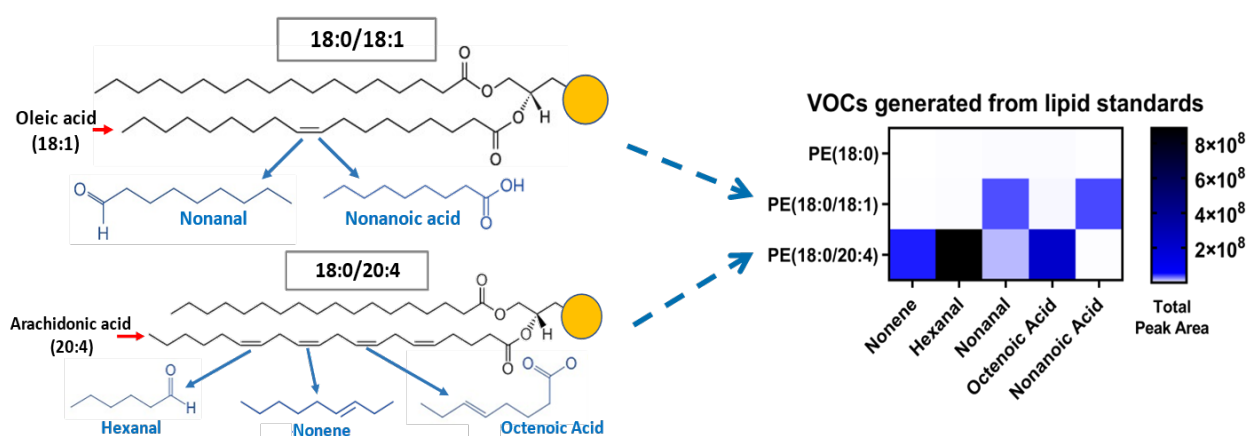
Despite promising results for cancer detection, the precise origin of cancer-associated VOCs remains poorly understood. Understanding the mechanisms of VOC production could refine existing diagnostic models, strengthen the case for clinical implementation, and broaden the translational value to areas such as patient stratification and treatment monitoring. In the setting of cancer, the discriminating VOCs may originate from the tumour, microbiome, host-microbiome interactions, inflammation, or as yet unidentified systemic processes<sup>116–118</sup>.

The majority of cancer-associated VOCs in exhaled breath are alcohols, aldehydes, carboxylic acids, esters, ketones, and alkanes<sup>119</sup>. A major source of these VOCs is believed to be reactive oxygen species (ROS)-mediated peroxidation of cell membrane phospholipids containing unsaturated fatty acyl chains<sup>120</sup>. Unsaturated fatty acids are highly prone to oxidation, leading to the creation of hydroperoxides even under ambient non-enzymatic conditions<sup>121</sup>. The process initiates with the formation of lipid radicals through the extraction of a hydrogen atom from a fatty acid molecule by a radical<sup>122</sup>. These lipid radicals are exceptionally reactive and readily engage with other radicals leading to the generation of LOO• radicals. These radicals can then seize hydrogen atoms from other lipids, producing a lipid hydroperoxide (LOOH) and generating another lipid radical. This new lipid radical continues the cycle by reacting with oxygen, forming LOO•, which in turn attacks another lipid, resulting in the creation of a lipid hydroperoxide. This chain reaction perpetuates the accumulation of lipid hydroperoxide as it progresses, which then degrades to produce a wide range of VOCs, including aldehydes, alkanes, and carboxylic acids<sup>123</sup>.

Abbassi-Ghadi et al used a spatially resolved lipid analysis technique (desorption electrospray ionization-mass spectrometry imaging, DESI-MSI) to interrogate phospholipid profiles across histologically pure OAC tissue sections<sup>124</sup>. Acyl chains of OAC glycerophospholipids were found to have more desaturations and longer lengths than adjacent normal tissue, providing candidate source compounds for VOCs detected in breath. Antonowicz et al provided the first mechanistic link between exhaled breath VOCs and tumour VOCs in OAC<sup>125</sup>. This study demonstrated that loss of aldehyde detoxification is an OAC hallmark and is sufficient to lead to the accumulation of endogenous

aldehydes, particularly when additional stress such as hypoxia was present. It was demonstrated that the enriched aldehydes cause genotoxicity and interfere with the DNA damage response, offering a new adjuvant therapeutic opportunity. The aldehyde accumulation also demonstrated a metabolic field effect in the malignant oesophagus, suggestive of an endoluminal route for OAC-associated exhaled aldehydes. Experimental data from the Hanna Lab has demonstrated that diagnostic VOCs can be derived from liposomes generated from phospholipid standards containing unsaturated acyl chains incubated under ambient conditions (Figure 3). Future work will utilise stable isotope tracing to monitor the kinetics of VOC production to exhaled breath.

In summary, lipids may be a major source of tumour-derived VOC production in OAC as the OAC phospholipidome is exceptionally prone to lipid peroxidation and aldehyde enrichment.



**Figure 3. Volatile products can be traced to the lipidome**  
Liposomes were incubated for 48 hours under ambient conditions (37°C) and headspace VOCs quantified using GC-MS  
(Image and data courtesy of Philip Leung)

### 1.3.5 Metabolic reprogramming and redox homeostasis under hypoxia

Hypoxia is a profound environmental stressor, and tumour cells display several metabolic adaptations to survive in the hypoxic microenvironment. These adaptations radically alter carbon fate and redox homeostasis<sup>126</sup>. The molecular pathways that underpin these adaptations are mediated primarily through the interactions of hypoxia-inducible factor (HIF), mammalian target of rapamycin (mTOR) and the unfolded protein response (UPR)<sup>127</sup>. These pathways are the critical survival mechanisms that promote tolerance of hypoxia.

HIFs are a family of heterodimeric transcription factors which are critical to the metabolic reprogramming of tumour cells in response to hypoxia. HIF-1α is the major metabolic hub in the

regulation of the hypoxic response. Under normoxic conditions, HIF-1 $\alpha$  is hydroxylated by a family of oxygen-dependent prolyl-hydroxylases and targeted for proteasomal degradation. Under hypoxic conditions, this process is inhibited, HIF-1 $\alpha$  accumulates and forms the active dimer with HIF-1 $\beta$ . HIF binds to numerous hypoxia-responsive elements in promoter regions of target genes to activate their expression. HIF induction reprograms metabolism by upregulating glycolytic genes, inactivating pyruvate dehydrogenase, enhancing fatty acid synthesis and enhancing ROS detoxification systems<sup>126,128</sup>. Due to decreased pyruvate oxidation, hypoxic tumour cells increasingly rely on alternative carbon sources for energy such as glutamine and exogenous acetate<sup>129,130</sup>.

mTOR is a serine/threonine protein kinase that forms the catalytic subunit of mTOR complexes (mTORC1/2)<sup>131</sup>. mTORC1 plays a central role in regulating energy metabolism in response to environmental conditions. Several studies have identified mTOR as a positive regulator of the biosynthesis of lipids, proteins, amino acids and nucleotides<sup>127</sup>. Hypoxia downregulates mTOR via a negative feedback loop involving HIF-regulated proteins REDD1, REDD2 and BNIP3<sup>132</sup>. Cycling hypoxia, however, can upregulate mTOR synthesis suggesting the interaction between hypoxia and mTOR may be context dependent<sup>133</sup>.

Severe hypoxia can interfere with protein synthesis and lead to the accumulation of misfolded proteins in the endoplasmic reticulum (ER stress)<sup>134</sup>. This activates UPR signalling pathways which act to attenuate protein translation and induce metabolic adaptations that enhance survival under hypoxia<sup>135</sup>. For example, the UPR signalling proteins ATF4 and ATF6 regulate amino acid and lipid metabolism respectively<sup>134,136</sup>. The UPR-PERK-eIF2 pathway induces uptake of cysteine, glutathione synthesis and protection against ROS produced during periods of cycling hypoxia<sup>137</sup>.

Mitochondrial ROS production paradoxically increases in hypoxia and is required for the activation of HIF- $\alpha$ <sup>138,139</sup>. Molecular oxygen (O<sub>2</sub>) is highly electrophilic and acts as a terminal acceptor of electrons in the mitochondrial electron transport chain. With reduced levels of molecular oxygen to act as a final acceptor, electrons leak from complex I and III to interact with residual oxygen and form superoxide radicals<sup>139</sup>. Hypoxia-induced alterations in the mitochondrial electron transport chain, including electron leakage, reverse electron transport, and changes in the redox state of electron carriers, contribute to an increase in ROS (superoxide, hydrogen peroxide and hydroxyl radical) production. Hypoxic signalling pathways act to suppress mitochondrial respiration and induce antioxidant enzymes which produce NADPH and GSH. Recent data has confirmed that total ROS and hydrogen peroxide levels increase, even in response to severe hypoxia (<0.1% O<sub>2</sub>)<sup>140</sup>. Excessive

hypoxia-induced ROS, however, can exceed the antioxidant defences and react with lipids in cell membranes to initiate lipid peroxidation<sup>141,142</sup>.

In summary, at least three pathways mediate the hypoxic response and play a crucial role in maintaining redox homeostasis. These widely influence a large number of genes to mediate the hypoxic response, and the precise combination of changes is context specific. Hypoxia-induced ROS production can overwhelm the cellular antioxidant defence mechanisms, initiating lipid peroxidation, leading to the degradation of lipids in cell membranes. This process contributes significantly to oxidative stress-mediated cellular damage and dysfunction observed in hypoxic conditions.

### 1.3.6 Lipid metabolism under hypoxia

Activation of hypoxic signalling pathways has a major impact on cellular lipid metabolism<sup>143</sup>. Cell growth, proliferation, and survival entail the synthesis of new membranes, which necessitates the production of fresh lipids with the appropriate molecular composition. The biochemical process responsible for converting acetyl coenzyme A into fatty acids (FA) is referred to as *de novo* lipogenesis. At the core of this pathway lies the conversion of saturated FA into mono-unsaturated FA, which play pivotal roles in mammalian membranes, cell signalling, and serve as a reservoir of metabolic energy. This crucial rate-limiting oxidation step is facilitated by stearoyl-CoA desaturase (SCD1), an oxygen-dependent enzyme responsible for maintaining the equilibrium between saturated and unsaturated FA. SCD1 carries out this vital function by transforming both externally absorbed and internally synthesized saturated FA into (mono)-unsaturated FA. Although alternative desaturase pathways have been discovered, including a FADS2-dependent pathway overexpressed in several cancer types<sup>144</sup>, *de novo* lipogenesis via SCD-1 is the major pathway.

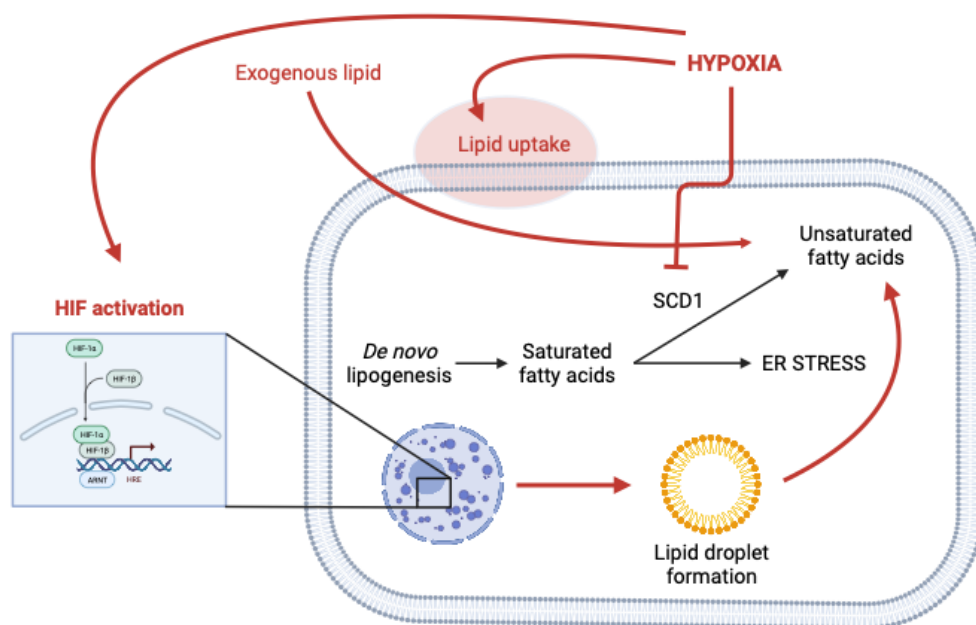
Since SCD1 is oxygen-dependent, a build-up of saturated FA precursors can occur under hypoxia as *de novo* lipogenesis is repressed. Colorectal cancer cells, for example, displayed a reduced C18:1/C18:0 ratio upon exposure to hypoxia (1% O<sub>2</sub>) indicating reduced SCD1 activity<sup>145</sup>. Many studies, however, have reported a paradoxical increase in the unsaturation index of lipid classes under hypoxia. Yu et al. reported that in HeLa cells, the unsaturation index of lysophosphatidylcholines, phosphatidylcholines, phosphoethanolamines and lysophosphatidylethanolamines increased after hypoxic stimulation<sup>146</sup>. Tobias et al. evaluated lipidomic properties across colorectal cancer spheroids and found that the hypoxic core regions contained higher levels of PUFA-containing TGs and long chain fatty acids<sup>147</sup>. Tamiya-Koizumi et al. also reported an increase in PUFA-containing plasmalogens in colorectal cancer



cells (Caco2) after 72 hours of hypoxic culture<sup>148</sup>. Finally, Zou et al. discovered that the HIF-2 $\alpha$ -HILPDA axis induces a metabolic program in clear cell renal carcinoma cells that selectively enriched PUFA-containing phospholipids<sup>149</sup>.

Cancer cells must therefore use alternative routes to maintain lipid homeostasis and sustain the desaturated lipid pool required for membrane formation and signalling. This is vital as the build-up of saturated FAs can lead to endoplasmic reticulum (ER) stress and eventual apoptosis<sup>150–152</sup>. Experiments by Young et al. revealed that hypoxic and serum-starved conditions reduced cancer cell viability and activated the unfolded protein response (UPR), but this could be rescued by the addition of exogenous desaturated lipids<sup>151</sup>. This critical dependence on exogenous desaturated lipids under hypoxic stress highlights a key metabolic vulnerability among many cancer cell types<sup>150,151,153</sup>. Both hypoxic and Ras-transformed cells can support growth by scavenging unsaturated fatty acids from the serum<sup>150</sup>. The precise molecular mechanisms that mediate and control the uptake of fatty acids from the environment remain under investigation. Matschke et al, for example, found that the rescue effect of exogenous desaturated lipids on hypoxic (0.2% O<sub>2</sub>) lung adenocarcinoma cells could be reduced by SGK1-inhibition, suggesting a key role for SGK1 in fatty acid uptake and survival of these cells under severe hypoxia<sup>154</sup>.

In addition to fatty acid uptake, hypoxia leads to significant lipid droplet (LD) accumulation, mediated by the HIF signalling pathway<sup>155,156</sup>. The downstream effectors that are critical for hypoxia-mediated LD formation include *HILPDA*-mediated suppression of lipid turnover<sup>157</sup>, *PLIN*-mediated repression of fatty acid metabolism<sup>158</sup> and *FABP3/FABP7*-dependent fatty acid uptake<sup>156,159</sup>. LDs have an antioxidant role by sequestering and protecting unsaturated FAs from oxidation<sup>156,160</sup>. The hydrolysis of unsaturated TGs within lipid droplets can also supply unsaturated FAs as substrates for phospholipid synthesis, which maintains lipid homeostasis during hypoxia to counteract a build-up of toxic saturated lipids<sup>161</sup>. Therefore LDs appear to allow ‘fine tuning’ of the saturated and unsaturated fatty acid distribution to maintain cellular lipid homeostasis. A summary of this process is illustrated in Figure 4.



**Figure 4. Maintaining lipid homeostasis under hypoxia**  
(Image created on BioRender and adapted from Ackerman et al<sup>162</sup>)

Additionally, Kim et al. showed that cytosolic NAD<sup>+</sup>/NADH is reduced under hypoxia when the electron transport chain is impaired but can be effectively restored by enhanced production of highly unsaturated fatty acids (HUFA) synthesised by desaturase enzymes encoded by *FADS1* and *FADS2*<sup>163</sup>. This increase in HUFAs provides a mechanism for glycolytic NAD<sup>+</sup> recycling. This pattern of increase in HUFAs under hypoxia was evident among cholesterol esters, phosphatidylcholines, diacylglycerols, and most noticeably, triglycerides. This relatively novel finding is an additional mechanism that could alter lipid homeostasis under hypoxia but has not been investigated in cancer cells.

In summary, hypoxic cancer cells adopt a lipogenic phenotype as a result of increased lipid scavenging and lipid droplet formation required to maintain membrane lipid homeostasis<sup>145</sup>. This adaptation appears to be predominantly driven by HIF and UPR signalling. The role of individual genetic mutations, nutrient deprivation, and tumour-environment crosstalk in modulating tumour lipid metabolism *in vivo* in response to hypoxia remains under investigation.

### 1.3.7 Hypoxia and VOCs

Despite the ubiquitous presence of hypoxia in the tumour microenvironment, there are no published *in vivo* studies that have investigated the influence of tumour hypoxia on VOC production. There are also no experimental studies which evaluated the influence of hypoxia on VOC production from OG cancer cells *in vitro*. Three studies have evaluated *in vitro* hypoxia-associated VOCs in other cancer

types. Taware *et al.* cultured breast cancer cell lines MCF-7 and MDA-MB-231 under normoxic (21% O<sub>2</sub>) and hypoxic (1% O<sub>2</sub>) conditions in serum-free media for 48 hours<sup>164</sup>. Cell culture supernatant was extracted and analysed using HS-SPME-GC-MS. Twenty VOCs were significantly altered after hypoxic exposure, including alcohols, aldehydes, fatty acids, and ketones. Issit *et al.* cultured MDA-MB-231 breast cancer cells in a static headspace chamber for >24h under low oxygen (~1% O<sub>2</sub>) conditions for direct headspace sampling and GC-MS analysis<sup>165</sup>. The authors reported a reduction in hexane and increase in styrene from cells after hypoxic culture. Finally, Furuhashi *et al.* cultured A549 lung cancer cells under low oxygen (~1.5% O<sub>2</sub>) conditions for 72 hours for direct headspace analysis using HS-GC-MS<sup>166</sup>. Hypoxia was found to increase the production of trans-2-hexenol, a product of lipid peroxidation, especially under low glucose and high lactate conditions. The authors concluded that lipid peroxidation was enhanced by hypoxia. Despite methodological limitations of these experiments (e.g. the precision of oxygen control, variable timepoints and sampling techniques), these studies provide limited data that hypoxia can generate an altered VOC signature in cancer cells.

## 1.4 Gaps in the literature

In summary, OAC continues to have extremely poor all-stage survival and more effective treatments are needed. Intratumoural hypoxia is a cause of treatment resistance and is clinically actionable, however available biomarkers are either not clinically realistic, or not suited to serial measurement. Hypoxia gene expression signatures have demonstrated utility as a biomarker of hypoxia in other cancers, albeit requiring a tissue biopsy. Exhaled breath analysis is an attractive potential solution, permitting non-invasive treatment stratification and monitoring. Growing evidence suggests that lipids may be a major source of tumour-derived VOC production in OAC. Hypoxia profoundly alters lipid metabolism and is a source of oxidative stress, and thus may generate a volatile signature amenable to non-invasive detection.

Several gaps in the literature have been identified:

1. No studies have evaluated the clinical utility of hypoxia gene signatures in OAC despite (i) hypoxia appearing to be prevalent in OAC and (ii) value demonstrated in other hypoxic cancers.
2. No volatilomic or lipidomic data describing hypoxic adaptation in OAC.
3. No data using volatilomics to stratify HTT precision medicine trials in OAC.

## 1.5 Thesis hypothesis and objectives

The leading project hypothesis is that ***OAC tumour hypoxia influences exhaled VOCs and can be used to stratify and monitor treatment response***. A rationally selected hypoxia gene signature will be used to define a hypoxia VOC model for therapeutic response prediction and monitoring.

1. Selection of an empirically validated hypoxia gene signature for OAC
  - a) Evaluate published hypoxia gene signatures using a stepwise *in silico* validation package.
  - b) Conduct a retrospective analysis of breath VOC data stratified by a literature gene signature.
  - c) Derive and validate a prognostic OAC-specific hypoxia gene signature using a network approach.
2. Define candidate hypoxia-associated exhaled VOCs in OAC.
  - a) Develop and apply a laser capture microdissection and RNA sequencing workflow for OAC tissue.
  - b) Apply a quality-controlled GC-MS analytical pipeline to analyse breath samples from OAC patients.
  - c) Conduct a pilot prospective analysis of exhaled VOC data from OAC patients stratified by hypoxic gene expression to identify candidate hypoxia-responsive VOCs.
3. Evaluate possible mechanisms of hypoxia-driven VOC production using cell models.
  - a) Describe metabolic adaptation to hypoxia in OAC by analysing cell line gene expression datasets.
  - b) Develop an analytical pipeline for non-targeted headspace GC-MS analysis of cell culture extracts.
  - c) Describe the lipidome and volatilome of cell culture extracts derived from hypoxic OAC cell lines.
  - d) Evaluate the impact of hypoxia on redox homeostasis in OAC cells.
4. Identify hypoxia-sensitive drugs for a VOC-stratified drug trial.
  - a) Select drugs which selectively target hypoxic OAC cells using an *in silico* approach.
  - b) Validate the predicted response of hypoxic OAC cells to candidate drugs.
  - c) Establish VOC monitoring in a clinical trial of hypoxia-targeted therapy.

**CHAPTER 2 – A HYPOXIA GENE EXPRESSION  
SIGNATURE FOR OESOPHAGEAL  
ADENOCARCINOMA**

## Summary

The aim of this chapter was to select a suitable hypoxia gene signature to act as a reference test for a hypoxia-stratified VOC discovery study. First, a review of literature hypoxia gene expression signatures was performed. This included an analysis of gene signature expression in cell lines exposed to hypoxia, inter-signature correlation, signature quality and prognostic utility. The best performing literature gene signature (Toustrup) was used to perform a retrospective analysis of exhaled VOC data stratified by hypoxia status. This demonstrated significant VOC differences between patients with hypoxia-high and hypoxia-low tumours.

Although the Toustrup signature performed best in this *in silico* review, current evidence suggests a need to derive tumour-specific signatures due to cell-type specific transcriptional responses to hypoxia. RNA-seq was used to identify genes significantly up-regulated under hypoxia in three OAC cell lines. Identified genes were used for network analysis and refinement in a training cohort, to generate a final 12-gene novel OAC hypoxia-associated gene signature (OAC-Hyp). An analysis of signature quality confirmed that the OAC-Hyp signature had robust expression and statistical metrics. Univariate and multivariate analysis confirmed that the final gene signature was prognostic in independent datasets, outperforming the Toustrup signature. Molecular analyses confirmed that hypoxic tumours identified by the OAC-Hyp signature were enriched with hypoxia-regulated genes and characterised by aggressiveness and immune dysfunction.

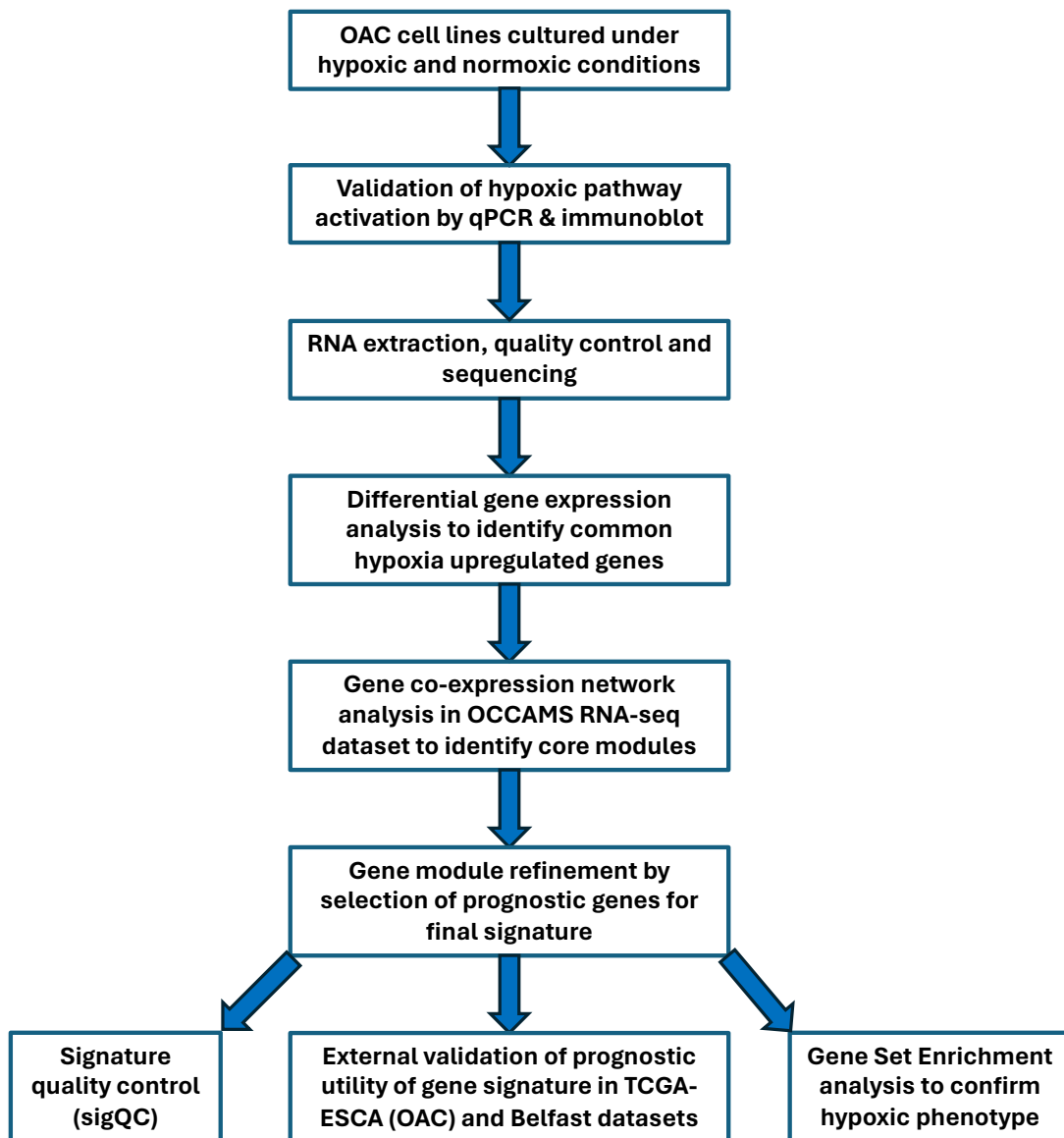


Figure 5. Summary of OAC hypoxia-associated gene signature derivation method

## 2.1 Methodological rationale

### 2.1.1 Using a gene expression signature as a reference test for tumour hypoxia

Designing a discovery study for hypoxia-associated VOCs required a suitable reference test that was scalable, pragmatic, and clinically relevant. Many of the available hypoxia biomarkers and their respective strengths and limitations have already been discussed previously in Section 1.2.3. Polarographic needle electrodes can provide direct measurements of tumour oxygenation but is technically challenging to apply in the oesophagus<sup>167</sup>. Imaging (PET-CT and MRI with hypoxia tracers) is an alternative option but there were no trials evaluating this technology in OAC. Pimonidazole can be safely administered to patients but is limited by the practical challenges with drug administration.

A hypoxia gene expression signature was therefore chosen as the most suitable reference test for hypoxia as this could be easily applied prospectively and retrospectively when tissue biopsies were available. In addition to their demonstrated prognostic and predictive utility in other cancers, they have been used widely in the literature in the molecular characterisation of hypoxic tumours. For example, Thienpont et al used a hypoxia gene signature in TCGA and discovered that hypoxic tumours were more likely to be hypermethylated<sup>168</sup>. This led to the discovery that hypoxia induces a loss of a key regulator of DNA methylation. Bhandari et al applied a hypoxia signature to multiple cancer cohorts and found a strong negative correlation between hypoxia and expression of miR-133a-3p. In controlled *in vitro* models of prostate cancer, miR-133a-3p was reduced under hypoxia and introducing a miR-133-3p mimic decreased cell viability and invasiveness. Finally, Ye et al calculated hypoxia scores in TCGA tumours and discovered strong correlations between hypoxia and the computationally predicted responses to a number of FDA-approved drugs<sup>169</sup>. These associations were subsequently validated *in vitro* using drug sensitivity experiments under hypoxic and normoxic conditions. These data demonstrate that hypoxia gene signatures can be used to identify clinically relevant hypoxia-associated features which can be validated *in vitro*.

Selection of an appropriate gene signature was the next challenge. Attempts have been made to identify a universal or 'core' hypoxia gene signature that could be applied to several or even all cancers. Buffa et al derived a hypoxia 'metagene' from breast, lung and head and neck cancer data that is the most widely cited hypoxia signature in the literature<sup>170</sup>. Lombardi et al recently reported 48 HIF target genes that were highly conserved in six cancer cell lines<sup>171</sup>. While there may be overlap between some cancers, the transcriptional response to hypoxia is highly variable between cell



types<sup>172</sup>. In a review of hypoxia gene signatures, the expression patterns of 65 top genes common to 32 hypoxia signatures were found to be highly variable between cancer types<sup>50</sup>. Recent data has revealed that variable cell type DNA methylation patterns established under normoxic conditions is the underlying mechanism of cancer cell-type specific transcriptional responses to hypoxia<sup>173</sup>. The practical implication is that hypoxia gene signatures that can be prognostic or predict treatment response in one cancer may have no prognostic or predictive utility in other cancers<sup>49</sup>. Therefore, while literature hypoxia gene signatures may have utility in some cancers from which they were not derived, there was an evident need to derive a tumour site specific hypoxia signature.

### 2.1.2 Signature derivation and validation

Several methods have been used to: (i) derive gene signatures using computational approaches, (ii) generate a combined hypoxia score and (iii) establish cut-offs to define 'hypoxia-high' and 'hypoxia-low' tumours. However, there is no consensus on the optimal approach or methodological studies comparing the performance of different methods.

Early studies used only aggregated expression data from cell lines exposed to normoxic and hypoxic conditions to generate a signature that could be directly applied to patient samples. This is controversial, as cell line responses to hypoxia may be unrepresentative of the *in vivo* hypoxic phenotype. More recent studies have combined *in vitro* data with an analysis of gene expression in patient samples<sup>170</sup>. This approach begins with an investigation into the expression and regulation of hypoxia-associated 'seed' genes in patient samples. Further steps then use computational techniques to aggregate and refine the signature prior to generating a hypoxia score. This method was selected because it has a proven track record of producing reliable and robust predictive signatures<sup>53,170,174,175</sup>.

For signature scoring, a common approach involves standardising the log<sub>2</sub> gene expression for each gene and dichotomizing samples into >50% expression (+1) and <50% expression for each gene (-1)<sup>170,176</sup>. Cut-off values for a binary classification of hypoxia vs normoxia can then be defined by using an external method (e.g. oxygen electrodes<sup>47</sup>, pimonidazole, imaging markers), unsupervised clustering techniques<sup>177</sup> or dichotomization by median hypoxia score<sup>49,174,178</sup>. Finally, efforts were made to validate that the gene signature truly identifies clinically-relevant hypoxic cohorts including: molecular characterisation to confirm hypoxia-associated pathways are up-regulated in hypoxic samples and demonstration of prognostic utility in multiple independent datasets<sup>56</sup>.

### 2.1.3 Choice of oxygen concentration

A key consideration in hypoxia research is the selection of relevant oxygen concentrations for *in vitro* experiments to model chronic hypoxia in the tumour microenvironment. The majority of literature gene signatures were derived using expression data from cells cultured at 1% O<sub>2</sub>. At this oxygen concentration, hypoxic pathways such as HIF and occasionally UPR are reliably induced in the majority of cell types, including OAC (Figure 6). As discussed in section 1.2.1, however, tumours are characterized by a wide range of absolute oxygen tensions. Literature data suggests that cells cultured at oxygen levels much lower than 1% tend to exhibit the highest rates of radio-resistance ('radiobiological' hypoxia)<sup>179</sup>. Chronic exposure (> 24 hours) to these lower concentrations of oxygen (0.2-0.5% O<sub>2</sub>) also selectively down-regulates pathways related to genetic instability and induces sensitivity to drugs targeting DNA repair<sup>36,180,181</sup>. Thus, given the clinical relevance of lower oxygen tensions, a concentration of 0.2% O<sub>2</sub> was chosen for all *in vitro* experiments related to signature derivation, metabolomic profiling and drug testing throughout this work.

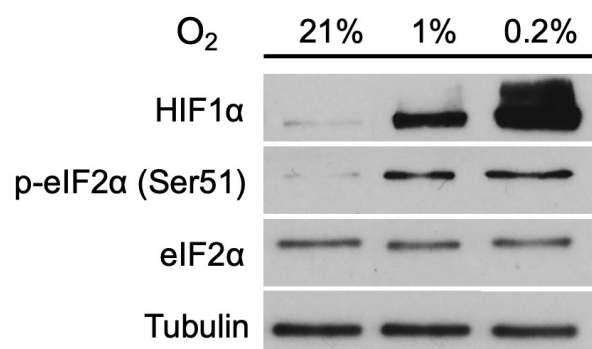


Figure 6. Western blot analysis of hypoxia pathways in FLO1 cells cultured at 1% O<sub>2</sub> and 0.2% O<sub>2</sub> for 48 hours

## 2.2 Hypothesis and aims

Chapter hypothesis: Hypoxia gene signatures can stratify OAC tumours by hypoxic status.

The objective of this section was to systematically evaluate literature hypoxia gene signatures and derive a novel OAC-specific hypoxia-associated gene signature that could be applied in a prospective clinical study to identify non-invasive surrogates of tumour hypoxia. Specifically this section aimed to:

1. Evaluate published hypoxia gene signatures using a stepwise *in silico* validation package.
2. Conduct a retrospective analysis of breath VOC data stratified by a literature gene signature.
3. Derive and validate a prognostic OAC-specific hypoxia-associated gene signature using a network approach.

## 2.3 Methods

### 2.3.1 Review of literature gene signatures

To identify a candidate hypoxia gene signature from the literature, hypoxia gene signatures (n = 23) were downloaded from *RadiationSigDB* (Table 2)<sup>182</sup>. While not acting as an exhaustive list of all hypoxia gene signatures, this database contains the most widely used hypoxia gene signatures in the literature and uses standardised annotation, which is a major challenge in summarising and combining signatures). All gene signatures contain genes upregulated by hypoxia, but vary in curation methods, scoring methods and included genes.

| Author                     | Origin         | Genes |
|----------------------------|----------------|-------|
| Buffa (Full)               | HNSCC & Breast | 51    |
| Buffa (Reduced)            | HNSCC & Breast | 15    |
| Eustace                    | HNSCC          | 26    |
| Fjeldbo                    | Cervix         | 6     |
| Ghazoui                    | Breast         | 70    |
| Hu                         | Breast         | 13    |
| Lendahl                    | Multiple       | 30    |
| Ragnum                     | Prostate       | 32    |
| Seigneuric (Early 0%)      | Multiple       | 71    |
| Seigneuric (Early 2%)      | Multiple       | 34    |
| Seigneuric (Late 0% Short) | Multiple       | 66    |
| Seigneuric (Late 2% Short) | Multiple       | 30    |
| Sørensen                   | SCC            | 27    |
| Starmans Cluster 1         | Breast         | 69    |
| Starmans Cluster 2         | Breast         | 246   |
| Starmans Cluster 3         | Breast         | 156   |
| Starmans Cluster 4         | Breast         | 93    |
| Starmans Cluster 5         | Breast         | 162   |
| Starmans Cluster 6         | Breast         | 14    |
| Starmans Cluster 7         | Breast         | 28    |
| Toustrup                   | HNSCC          | 15    |
| Winter                     | HNSCC          | 99    |
| Yang (Bladder)             | Bladder        | 24    |

**Table 2. Hypoxia gene signatures from RadiationSigDB**

A method of hypoxia gene signature validation was adapted from a previous study where gene signature performance was evaluated in ten microarray expression datasets from cell lines exposed

to hypoxia and normoxia<sup>169</sup>. For the adapted the method, an additional five datasets were downloaded from Gene Expression Omnibus (GEO)<sup>183</sup> to increase coverage across epithelial and glandular cell types and hypoxic conditions, although no published OAC hypoxia datasets were identified (Table 3).

| Dataset   | Cancer type       | n  | Hypoxic conditions        |
|-----------|-------------------|----|---------------------------|
| GSE71280  | Gastric cancer    | 1  | 48h @ 1% O <sub>2</sub>   |
| GSE64710  | Gastric cancer    | 1  | 48h @ 1% O <sub>2</sub>   |
| GSE33521  | Cervical cancer   | 2  | 6h @ 1% O <sub>2</sub>    |
| GSE43608  | Colon cancer      | 2  | 72h @ 1% O <sub>2</sub>   |
| GSE70051  | HNSCC             | 2  | 48h @ 0.1% O <sub>2</sub> |
| GSE125177 | Gastric cancer    | 3  | 48h @ 1% O <sub>2</sub>   |
| GSE3188   | Breast cancer     | 3  | 16h @ 1% O <sub>2</sub>   |
| GSE73556  | Glioblastoma      | 3  | 12h @ 1% O <sub>2</sub>   |
| GSE79069  | Cervical cancer   | 3  | 20h @ 1% O <sub>2</sub>   |
| GSE3051   | Cervical cancer   | 3  | 24h @ 1% O <sub>2</sub>   |
| GSE55935  | Prostate cancer   | 4  | 48h @ 1% O <sub>2</sub>   |
| GSE75034  | Cervical cancer   | 4  | 48h @ 0.2% O <sub>2</sub> |
| GSE77606  | Various           | 7  | 48h @ 1% O <sub>2</sub>   |
| GSE67549  | Pancreatic cancer | 9  | 24h @ 1% O <sub>2</sub>   |
| GSE30979  | NSCLC             | 10 | 72h @ 1% O <sub>2</sub>   |

**Table 3. Microarray expression datasets from GEO Datasets**

In an attempt to standardize scoring methods between gene signatures for comparison, hypoxia scores were generated using the R package *gsva*<sup>184</sup>. This technique aims to quantify the activity of functionally related genes within each sample in a dataset. It calculates a sample-wise enrichment score for each gene set/pathway, indicating the overall activity level of genes within that set for each sample. The enrichment score is derived by comparing the expression levels of genes within a set to those in the entire dataset, assessing whether genes in the set are consistently up- or down-regulated in a sample. GSEA generates a matrix of enrichment scores, where rows represent gene sets/pathways and columns represent different samples. Positive enrichment scores indicate increased expression, and vice versa.

The selection process for a candidate literature hypoxia gene expression signature included the following analyses:

- i. *Stability*: GSVA hypoxia scores were generated from 15 *in vitro* expression datasets (Table 3) to determine which gene signature was consistently upregulated under hypoxic conditions.
- ii. *Representativeness*: Correlation between signatures and established hypoxia gene sets<sup>185</sup> in TCGA-ESCA (OAC) was calculated using Spearman's rank correlation coefficient in order to identify the signature most representative of other hypoxia signatures and hypoxic signalling.
- iii. *Quality*: Signature quality in TCGA-ESCA was then assessed using the R package *sigQC* which provides a comprehensive assessment of signature expression, variability, and structure<sup>186</sup>.
- iv. *Clinical correlation*: Hypoxia status in TCGA-ESCA (OAC) was determined by unsupervised hierarchical clustering using the best performing gene signature (*ward.D* function in the R *stats* package)<sup>168</sup>. The top 3 sub-clusters were assigned as hypoxia score-high, intermediate, and low<sup>168,169</sup>. Kaplan-Meier survival curves were plotted and compared using the log-rank test to compare hypoxia score-high and low clusters.

### 2.3.2 Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR)

For the preparation of qPCR templates, 1µg of RNA underwent conversion to complementary DNA (cDNA) utilizing the Superscript III First Strand Supermix system (Thermo Fisher Scientific, UK). The process was carried out following the manufacturer's guidelines using a GeneAmp® PCR System 9700 (Applied Biosystems). Relative quantitative PCR (qRT-PCR) was performed using the SYBR Select master mix (Thermo Fisher Scientific, UK) and a 7900 HT thermal cycler (Applied Biosystems), following the respective manufacturers' instructions. Adherence to the MIQE standards checklist ensured methodological and reporting quality throughout the qRT-PCR experiment<sup>187</sup>. The Toustrup gene signature was quantified in clinical samples using qRT-qPCR and normalized to the geometric mean of three reference genes as described in the original paper<sup>188</sup>. Primers were selected through PrimerBank and run through Primer-BLAST to ensure specificity (see Appendix 3 for sequences). Gene expression was normalized to the geometric mean of three reference genes (*RPL37A*, *ACTR3*, and *NDFIP1*), previously validated in OAC cell lines<sup>188</sup>. Reactions were run in groups of four samples per 384-well plate using 10µl reaction volumes. Each well had 5µl SYBR Select Master Mix, 0.5µl of each primer (final concentration 10nM), 2µl of water and 2µl of cDNA template. Thermal cycling was as follows: 94°C for 3 minutes, then [94°C for 15 seconds, 60°C for 2 minutes] repeated 40 times. Melt curves were used to check for primer-dimers. Relative gene expression was calculated using the delta-delta C<sub>t</sub> method.

### 2.3.3 Retrospective gene-VOC analysis

VOC data (SIFT-MS), tissue UPLC-MS/MS aldehyde data and archival matched tumour tissue RNA (n = 22) were previously generated from a multi-centre clinical trial of breath testing in OG cancer (courtesy of Dr Sheraz Markar and Dr Stefan Antonowicz)<sup>109</sup>. RNA was retrieved under Imperial College Healthcare Tissue Bank Approval R20037. The VOC and LC-MS datasets were divided equally by median gene signature expression into hypoxia-high and hypoxia-low tumours to identify differentially expressed VOCs ( $p < 0.05$ ).

### 2.3.4 Cell culture

The following OAC cell lines were purchased from Public Health England: FLO1, OE19 and OE33. All cell lines have been previously authenticated using short tandem repeat profiling<sup>189</sup>. COSMIC data<sup>190</sup> confirms that all harbour TP53 mutations and are BRCA1/2 wild-type (Table 4). FLO1 was cultured in DMEM (high glucose) medium (Thermo Fisher Scientific, UK) supplemented with 10% foetal bovine serum (FBS), 2mM glutamine and penicillin-streptomycin. OE19 and OE33 were cultured in RPMI-1640 medium (Thermo Fisher Scientific, UK) supplemented with 10% FBS, 2mM glutamine and penicillin-streptomycin. Cells were serially passaged using trypsin-EDTA 0.05% and were not maintained beyond passage 20. Mycoplasma testing was carried out monthly (MycoAlert™ Mycoplasma Detection Kit, Lonza). Cells were seeded in 6-well plates at densities based on growth curves to achieve a target confluency of 60% after 48 hours incubation under normoxia.

| Cell line   | Location   | Site of origin | Phenotype <sup>191</sup> | pTNM    | TP53 mutation |
|-------------|------------|----------------|--------------------------|---------|---------------|
| <b>FLO1</b> | Distal OAC | Primary tumour | Mesenchymal              | pT2N1M0 | c.830G>T      |
| <b>OE19</b> | Cardia AC  | Primary tumour | Epithelial               | pT3N1Mx | c.929dupA     |
| <b>OE33</b> | Distal OAC | Primary tumour | Epithelial               | pT3N0Mx | c.404G>A      |

Table 4. Authenticated OAC cell lines used in this work

### 2.3.5 Hypoxia treatment

After incubation overnight to allow adherence, plates were transferred to an H35 Hypoxystation (Don Whitley Scientific). The media was changed, and cells were incubated at 0.2% O<sub>2</sub> for 48 hours. Control plates were incubated in a normoxic incubator. After 48 hours, the media was removed, and cells were

washed with PBS prior to protein and RNA extraction. All media and PBS used for the processing of hypoxia-treated cells were initially 'de-gassed' by incubating at 0.2% O<sub>2</sub> overnight.

### **2.3.6 Immunoblot**

For experiments using the 6-well plate format, cells were washed twice using ice-cold PBS and lysed using 250µl of RIPA Lysis and Extraction Buffer (ThermoFisher) containing 1X protease (cOmplete™ Protease Inhibitor Cocktail) and 1X phosphatase inhibitors (Phosphatase Inhibitor Cocktail 3, Sigma Aldrich) at 4°C for 30 minutes. Cell lysate was collected into an Eppendorf tube using a cold cell scraper. The lysate was sonicated for two cycles of 10 seconds on ice to complete cell lysis and shear genomic DNA to reduce sample viscosity. Samples were then centrifuged at 16,000xg for 20 minutes at 4°C to pellet debris, and supernatant was carefully transferred to a new Eppendorf tube. Samples were stored at -80°C or used immediately. Protein lysate samples were quantified using a bicinchoninic acid (BCA) assay kit (Sigma Aldrich) and concentrations were normalized to enable equal protein loading. Samples were mixed with Laemmli protein sample buffer (Bio-Rad) containing 2-mercaptoethanol and boiled at 95°C for 5 minutes to disrupt disulphide bonds. Samples were then centrifuged for 1 minute at 16,000xg. Proteins were separated using SDS-PAGE by loading equal amounts (20µg) into the wells of hand cast gels (8-12%) in a Bio-Rad blotting chamber. Electrophoresis was run at 120V for 100 minutes prior to transfer onto polyvinylidene difluoride (PVDF) membranes (Bio-Rad) using either a semi-dry blotting system or overnight wet tank transfer. Gel percentage and transfer conditions were optimized for specific protein targets. Membranes were blocked in 5% non-fat milk in tris-buffered saline/0.01% Tween 20 (TBS-T) (Sigma) for 1 hour at room temperature, followed by incubation in primary antibody overnight at 4°C. Membranes were then washed three times using TBS-T, followed by incubation in secondary antibody conjugated to horse radish peroxidase for 1 hour at room temperature. Membranes were washed again three times using TBS-T. Enhanced chemiluminescence substrate (Sigma) was added to the membrane for 5 minutes. Chemiluminescence was visualised using an iBright imaging system (ThermoFisher). Primary antibodies and dilutions used for hypoxia validation were: HIF-1α (1:1000) (BD Biosciences #610958), eIF2α (Cell Signaling Technology #9722) and phospho-eIF2α (Ser51) (Cell Signaling Technology #9722).

### **2.3.7 RNA extraction**

At the specified timepoint, cells were washed twice with PBS prior to the addition of 350µl of Buffer RLT (Qiagen) for cell lysis. Lysate was then added to a QIAshredder column and centrifuged at full



speed for 2 minutes. RNA was then precipitated, washed, and eluted from this lysate using the RNeasy Mini Kit (Qiagen) according to manufacturer instructions. Extracted RNA was quantified using absorption spectroscopy (Nanodrop, ThermoFisher) and RNA integrity was verified using electrophoresis (RNA 6000 Pico Kit with the 2100 Bioanalyzer, Agilent). The RNA sample quality was assessed based on the fragment size distribution, specifically noting two peaks that corresponded to 18S and 28S ribosomal RNAs, along with a signal in the small molecular weight range indicative of small RNAs. This distribution was cross-referenced with a comprehensive database of RNA samples to calculate an RNA Integrity Number (RIN). The resulting electropherograms were used to evaluate both the integrity and concentration of the RNA samples. The RIN, calculated by the instrument, assigned a value where 10 signifies intact RNA, while a value of 1 indicates complete degradation.

### **2.3.8 RNA sequencing**

Only RNA with an RNA integrity number > 8, 260/280nm ratio > 1.8 and 260/230nm ratio > 1.8 were taken forward for sequencing by Novogene. Library preparation was performed using the Illumina TruSeq® Stranded mRNA assay. Poly-A tail RNA sequencing at 20 million paired-end reads was conducted in three separate biological replicates for each of the three OAC cell lines, utilizing the Illumina Hi-Seq4000 platform. Upon removal of sequence adapters, Trimmomatic (v0.36)<sup>192</sup> was employed for quality trimming of the reads. Subsequently, alignment against the human reference genome (GRCH38) was performed, and the counts per gene were calculated using STAR (v 2.5.3)<sup>193</sup> with annotation sourced from GENCODE version 27.

### **2.3.9 Clinical gene expression datasets**

Human OAC mRNA expression data and clinical metadata were acquired from the following sources and pre-processed for gene signature derivation and validation.

*The Cancer Genome Atlas (TCGA)*<sup>194</sup>: Pre-chemotherapy OAC tissue samples were collected between 2011-2013 from multiple North American centres. Tumour samples underwent histological analysis and samples with ≥60% tumour nuclei and ≤20% necrosis submitted for RNA extraction and sequencing using an Illumina HiSeq2000. Reads were aligned to a human reference genome (hg19/GRCh37), filtered and annotated in ENSEMBL (v59). Level 3 RNA-seq data in Expectation Maximisation (RSEM) format were downloaded from Broad GDAC Firehose and  $\log_2(n+1)$

transformed. The TCGA-ESCA dataset was filtered to include only OAC tumours (n = 88, of which all had survival data and 71 had matched demographic data).

*Oesophageal Cancer Classification and Molecular Stratification (OCCAMS) consortium*<sup>195</sup>: This data was kindly provided by the OCCAMS collaboration after a prize competition and steering committee review. OAC samples were obtained from OCCAMS sites (2010 – 2019) during surgical resection or by biopsy at endoscopic ultrasound, snap-frozen, and sent for cellularity review by two expert pathologists. Cancer samples with cellularity  $\geq 70\%$  underwent RNA extraction and sequencing using an Illumina HiSeq4000. RNA-seq raw counts (n = 263, of which 195 had survival data and matched demographic data) were filtered to remove low expressing genes (counts < 10) and normalized using a variance stabilizing transformation using DESeq2. This was used as the training dataset for gene signature generation as it was possible to process the raw data. Demographic data included age, sex, smoking history, tumour stage, chemotherapy type, response and survival.

*Belfast cohort* (kindly provided by Dr Richard Turkington at Queen's University, Belfast)<sup>196</sup>: The Belfast dataset was generated from FFPE pre-chemotherapy endoscopic biopsies retrieved from 273 patients with resectable OAC from four UK centres between 2003 to 2014 (Northern Ireland Cancer Centre, n = 170; University of Cambridge, n = 92; University of Edinburgh, n = 23; University of Southampton, n = 7). Tumour samples containing >50% tumour underwent RNA extraction and gene expression analysis using microarray-based technology (Almac Diagnostics Xcel array, Almac, Craigavon, UK). Data was provided processed from raw CEL files. Where multiple probesets were linked to a single gene, outlier probesets were excluded and the probeset with the highest median absolute deviation was selected. Patients in this cohort were treated with neoadjuvant chemotherapy followed by surgical resection. All patients underwent follow-up according to local institutional guidelines and were followed up for at least 2 years. Response data was described using the Mandard scale, with a responder defined as TRG  $\leq 2$ . Matched survival data was available for 271/273 patients.

### **2.3.10 Development and validation of an OAC mRNA hypoxia-associated gene signature**

A seed-based co-expression network method was used to generate an OAC hypoxia-associated gene signature, as this technique has successfully been used to derive robust prognostic hypoxia gene signatures in other cancers<sup>53,170,178</sup>. The R package *DESeq2* was used to process RNA-seq data for differential expression analysis and identify hypoxia up-regulated protein coding genes (FC > 1.5) in each OAC cell line. The Benjamini–Hochberg method was used for multiple testing correction with P

value cut-off specified at 0.05. Genes that were up-regulated by hypoxia in all three OAC cell lines were used as hypoxia seed genes. Using the R package *igraph*, a gene co-expression network was then constructed in the OCCAMS training dataset to provide a system level view of the interactions between established hypoxia seed genes. Gene-gene interactions were filtered by a threshold value of correlation (Spearman's  $\rho > 0.5$ ). The gene network was partitioned using the Louvain method into non-overlapping gene modules<sup>197</sup>. Genes that were not connected to the main component were removed. Individual gene modules were then refined, selecting genes that were prognostic (Cox regression  $P < 0.05$  and  $HR > 1$ ), based on the evidence that tumour hypoxia is associated with aggressive tumour phenotype and poor prognosis.

Hypoxia gene signature scores for patient samples were then constructed using a standardized method<sup>176,198,199</sup> (Figure 7). For each gene of the signature, patients with the top 50% expression for each gene were given a score of +1, and bottom 50% were given a score of -1. This was repeated for each gene of the signature, and the sum total was the hypoxia score for each patient. High scores indicated the tumour was hypoxic, and low scores indicated normoxia. The derived signatures were then tested in independent validation cohorts (TCGA-ESCA (OAC) and the Belfast cohort). Patients were stratified into hypoxia-high and hypoxia-low groups based on an optimum cut-off selected by the *surv\_cutpoint* function of *survminer*, or median cohort hypoxia score. For a multivariate analysis, the hypoxic score was adjusted for age of diagnosis, gender, tumour stage and smoking status (available for the TCGA and OCCAMS datasets only). Meta-analysis was conducted using the *meta* package, using a fixed effects model.

$$\sum_{n=1}^N gene_{expr,n}$$

$N$  = number of genes in the signature;  $gene_{expr,n}$  is the median dichotomized value for the gene expression of the  $n$ th gene in the signature compared to the expression levels of that gene from all samples. If the level of the  $n$ th gene is above the median for all samples then  $gene_{expr,n}$  is 1, otherwise -1

**Figure 7. Hypoxia scoring equation**  
Described in Fox et al<sup>176</sup>

### 2.3.11 Differential gene expression analysis

Unsupervised hierarchical clustering of Z-scores of signature gene expression was used to identify functionally relevant sub-groups using the *ward.D* function in the R *stats* package<sup>168</sup>. The top 3 sub-clusters were assigned as hypoxia score-high, intermediate and low. Differential expression analysis between hypoxia-high and hypoxia-low groups was performed using *DESeq2* and Gene Set Enrichment Analysis (GSEA) was performed using *fgsea*. Pathway analysis was conducted using HALLMARK gene sets, defining enriched gene sets with a P value < 0.05.

### 2.3.12 Differential immune infiltration analysis

RNA-seq data from OCCAMS (non-log transformed FPKM) was uploaded to the CIBERSORT tool<sup>200</sup>. The required input is RNA-seq data in FPKM/TPM format or Affymetrix CEL raw files, therefore this was only performed with the OCCAMS dataset. CIBERSORT was executed using the LM22 signature gene set in both relative and absolute modes, employing 100 permutations while disabling quantile normalization, following the recommended approach for RNA-seq data. The CIBERSORT tool produced absolute immune fraction scores for 22 distinct immune cell populations within the datasets. Only patients exhibiting statistically significant deconvolution results across all subsets were included for analysis. Wilcoxon rank-sum tests were computed for each immune population within both the hypoxia-high and hypoxia-low groups in the OCCAMS dataset.

### 2.3.13 Statistical analysis

Continuous data has been presented as mean (standard deviation, SD) or median (interquartile range, IQR). Categorical data has been presented as frequency (percentage). Parametric data were compared using Student's T test or ANOVA, non-parametric data were compared using Wilcoxon or Kruskal-Wallis rank sum test, and categorical data were compared using Pearson's Chi-squared test or Fisher exact test. All data analysis was conducted using R version 4.3.2 unless otherwise indicated. The main packages used were: *dplyr* (data manipulation); *DESeq2* and *fgsea* (RNA sequencing); *survminer* (survival analysis including Kaplan-Meier and Cox multivariate regression models); *igraph* (network analysis); *ggplot2* (data visualization). Data visualization was also performed using GraphPad Prism version 9.

## 2.4 Results

### 2.4.1 Review of literature hypoxia gene signatures

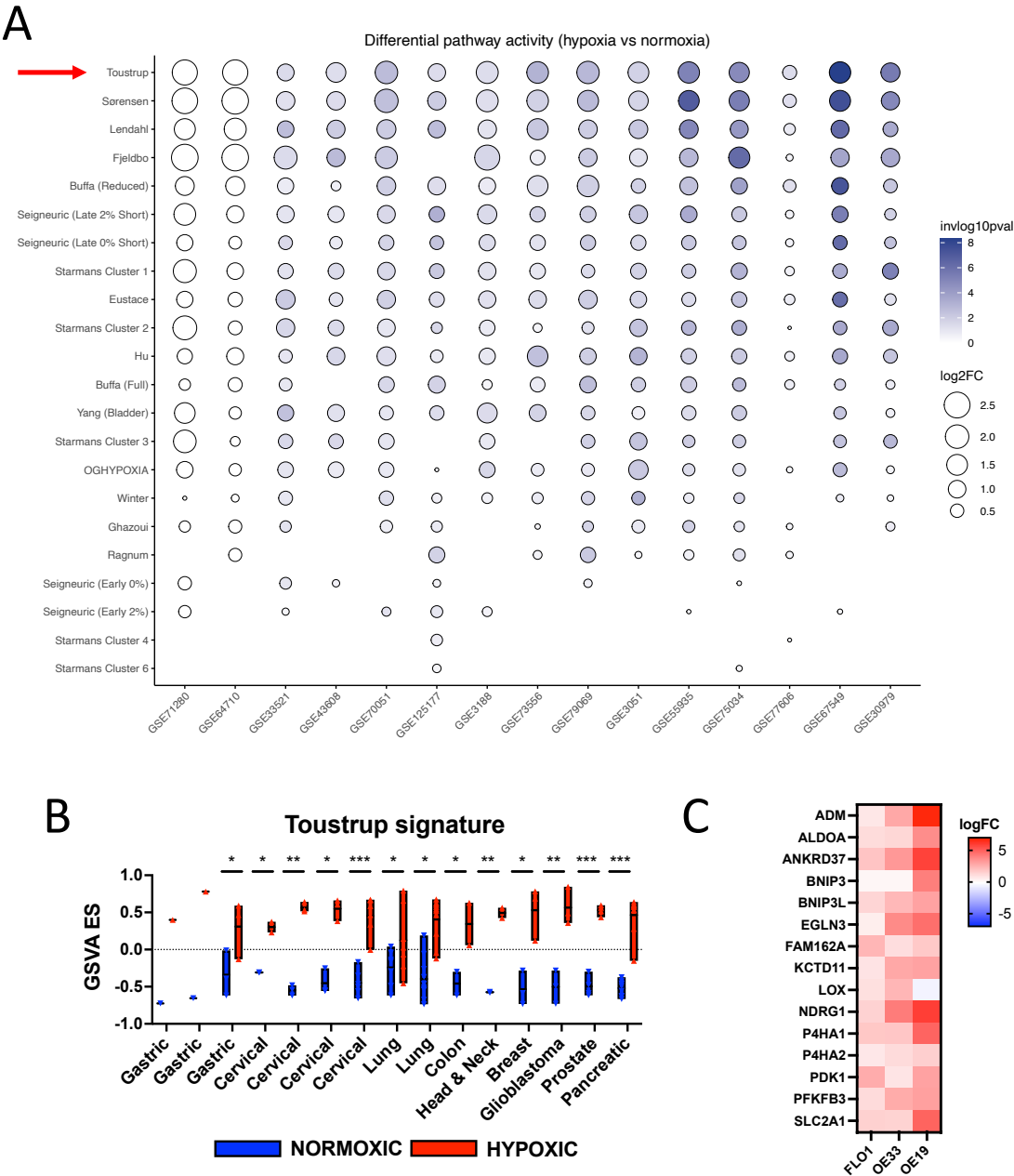
The performance and quality of 23 literature hypoxia gene signatures was reviewed systematically. Most gene signatures were enriched in microarray expression datasets of non-OAC cell lines exposed to hypoxia (Figure 8A). In particular, nearly all cell lines undergoing hypoxia were enriched for the Toustrup signature (Figure 8B), suggesting that this signature was independent of biological context. The Toustrup signature genes were also strongly up-regulated in all three OAC cell lines under evaluation (Figure 8C).

Next, the correlation between gene signatures and established hypoxia gene sets was evaluated. This showed that the Toustrup signature correlated strongly with all hypoxia gene sets in OAC suggesting it was highly representative of hypoxic signalling in OAC (Figure 9A). Analysis of inter-signature correlation revealed that no one signature was representative of all other signatures (Figure 9B). Instead, gene signatures clustered with other signatures with similar curation methods. Signatures including the Toustrup and Buffa signatures correlated strongly with most of the included signatures, suggesting these signatures represent similar biology.

Signature quality was then assessed using *sigQC* in both TCGA-ESCA OAC and OSCC. Implementation of this package is used to evaluate three quality metrics of a gene signature: (i) signature transportability (does the signature retain high expression and variability across datasets?), (ii) signature biological integrity (does the signature produce clusters with distinct patterns of expression?) and (iii) signature suitability (is the signature robust and produce similar classification irrespective of scoring method?). Additionally, it can also be used to evaluate whether the dataset is suitable for use with the signature. The various metrics that reflect signature transportability, biological integrity, suitability, and dataset suitability are summarised in a radar plot (Figure 9C). The area encompassed by the radar reflects a summary score of overall quality. This shows that the Toustrup signature displays consistent high quality in both OAC and OSCC datasets, suggesting signature transportability. For comparison, the most widely used hypoxia signature in the literature (Buffa) slightly under-performed against the Toustrup signature, in terms of gene variability and expression.

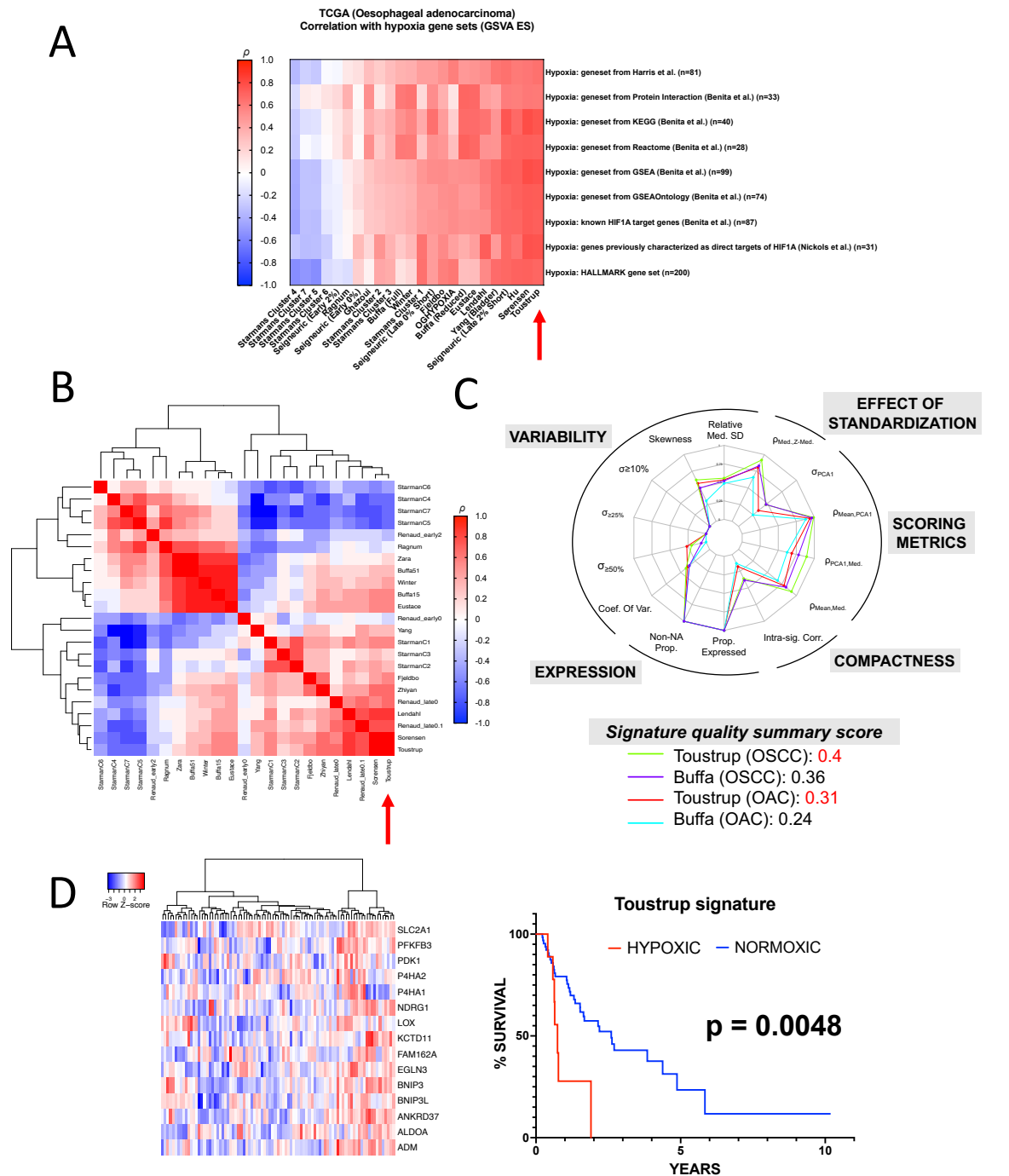
Since the Toustrup signature appeared to be the most robust, representative, and highest quality gene signature from the preceding analysis, this signature was applied to the TCGA datasets. Unsupervised

hierarchical clustering revealed strong clusters of low and high expression, with distinct prognostic outcomes (Figure 9D).



**A**, Bubble plot demonstrating fold change (bubble size) and significance (bubble shading) of GSVA score fold change (hypoxia/normoxia) in each cell line GEOdataset. **B**, Bar chart showing GSVA scores in normoxia and hypoxia from cell line data in GEOdatasets. **C**, Heatmap showing log fold change (RT-qPCR data) of Toustrup signature genes in OAC cell lines cultured under hypoxia (0.2% O<sub>2</sub>) and normoxia for 48 hours.

**Figure 8. Performance of literature hypoxia signatures in vitro**

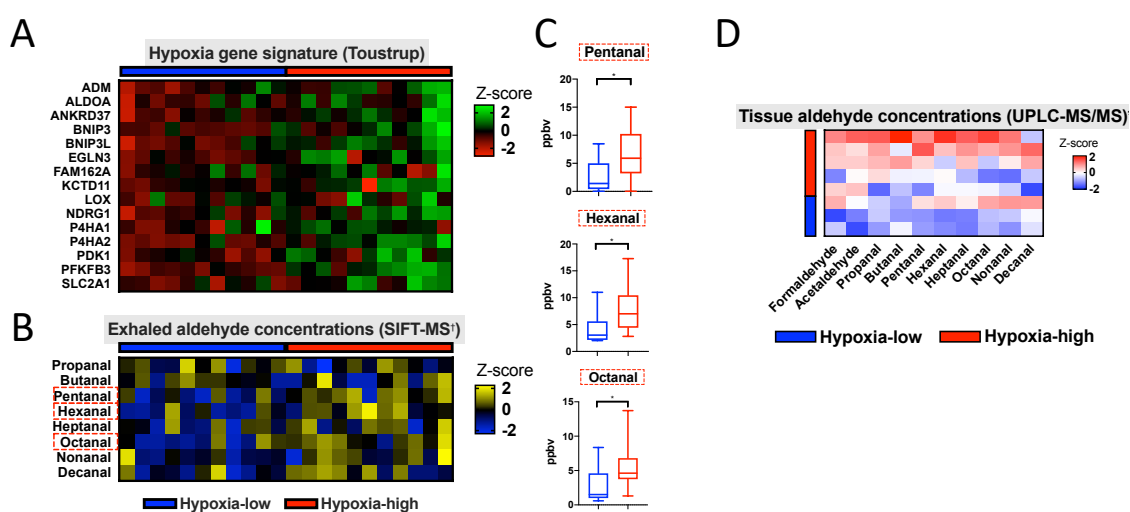


**A**, Correlation between GSVA enrichment scores for published/*in-silico*-derived hypoxia gene signatures and well-defined hypoxia gene sets in TCGA-ESCA (OAC). **B**, Heatmap showing inter-signature GSVA score correlation in TCGA-ESCA (OAC). **C**, Radar plot from *sigQC* summarizing signature quality metrics in TCGA-ESCA using Toustrup and Buffa hypoxia signatures. **D**, Unsupervised hierarchical clustering of TCGA OAC gene expression data stratified by Toustrup signature (hypoxia-high, intermediate and low) and Kaplan-Meier plot of overall survival in TCGA OAC cohort comparing hypoxia-high and low clusters.

**Figure 9. Performance of literature gene signatures in vivo**

## 2.4.2 Retrospective VOC analysis stratified by hypoxia

Since it was unknown whether there was an association between tumour hypoxia and exhaled VOCs in OAC, an exploratory analysis of archival RNA samples and VOC data was performed using the best performing literature gene signature (Figure 10). Targeted VOC data (acquired using SIFT-MS), tissue LC-MS data and matched tumour tissue RNA (n = 22) were retrieved from a multi-centre clinical trial of breath testing in OG cancer<sup>109</sup>. The Toustrup gene signature was quantified in mRNA derived from clinical OAC samples using qRT-PCR and normalized to the geometric mean of three reference genes as described in the original paper. The VOC and LC datasets were divided equally by median gene signature expression into hypoxia-high and hypoxia-low tumours to identify differentially expressed VOCs. A stratified analysis revealed that breath and tissue from patients with hypoxia-high tumours appeared to be relatively enriched with volatile aldehydes, particularly medium-chain alkanals: pentanal, hexanal and octanal ( $P < 0.05$ ). No difference was found for fatty acids and phenol compounds.



**A**, Toustrup gene expression (RT-qPCR) and **B**, breath VOC data from OAC patients (n=22). Heatmap columns represent gene expression and breath VOC data from individual patients with OAC. **C**, Box and whisker plots of significantly upregulated VOCs in hypoxia-high cohort. **D**, Sub-group analysis of OAC tissue aldehyde concentrations normalized to sample mass. Rows represent individual tumours with available LC-MS data. \* $P < 0.05$ .

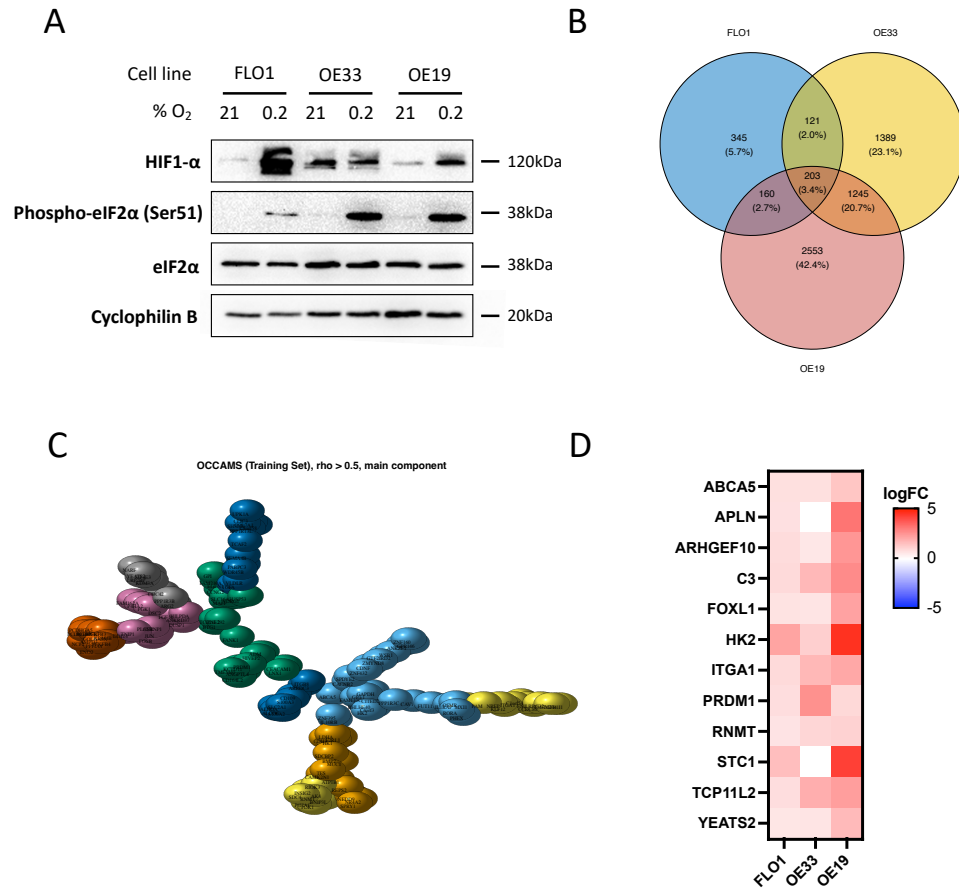
**Figure 10. Retrospective analysis of VOC data stratified by the Toustrup hypoxia signature**



### 2.4.3 Development of a novel OAC hypoxia-associated gene signature

To identify a core set of hypoxia up-regulated genes, OAC cell lines were cultured under normoxic and hypoxic (0.2% O<sub>2</sub>) conditions prior to RNA sequencing. Immunoblot confirmed up-regulation of markers of HIF and UPR upon hypoxic exposure which validated the experimental conditions had induced key hypoxic pathways, although interestingly HIF1- $\alpha$  was also expressed under normoxic conditions in OE33 (Figure 11A). The QC check prior to RNA-seq confirmed that all samples had an RIN 10 (highest quality) (Appendix 4). Differential gene expression analysis revealed that 203 protein-coding genes were up-regulated (FC > 1.5, FDR P < 0.05) across all three OAC cell lines (Figure 11B). Co-expression network analysis was performed in the OCCAMS dataset using this core set of genes. This identified a highly correlated gene module enriched with hypoxia-related genes (Figure 11C) which was refined to a compact signature (OAC-Hyp) by selecting prognostic genes (HR > 1, Cox P < 0.05 (Table 5).

Signature quality of the OAC-Hyp signature was compared against the best performing literature gene signature using *sigQC*, with the final results summarised in the radar plot (Figure 12A). The OAC-Hyp signature out-performed the Toustrup signature in the OCCAMS and TCGA-ESCA (OAC) datasets, but showed less favourable quality metrics in the Belfast dataset. Signature gene expression and variability was similar for both signatures. Intra-signature correlation (indicating signature compactness) and scoring metric correlations were similar for both gene signatures, except in the Belfast dataset where the Toustrup signature performed slightly better (Figure 12B). The plots in Figure 12C show correlations between median metrics using non-standardized (absolute expression) and standardized (Z-score) data. This shows similar performance between datasets. The strength of correlation was variable between datasets, from strong (OCCAMS) to weak (Belfast).

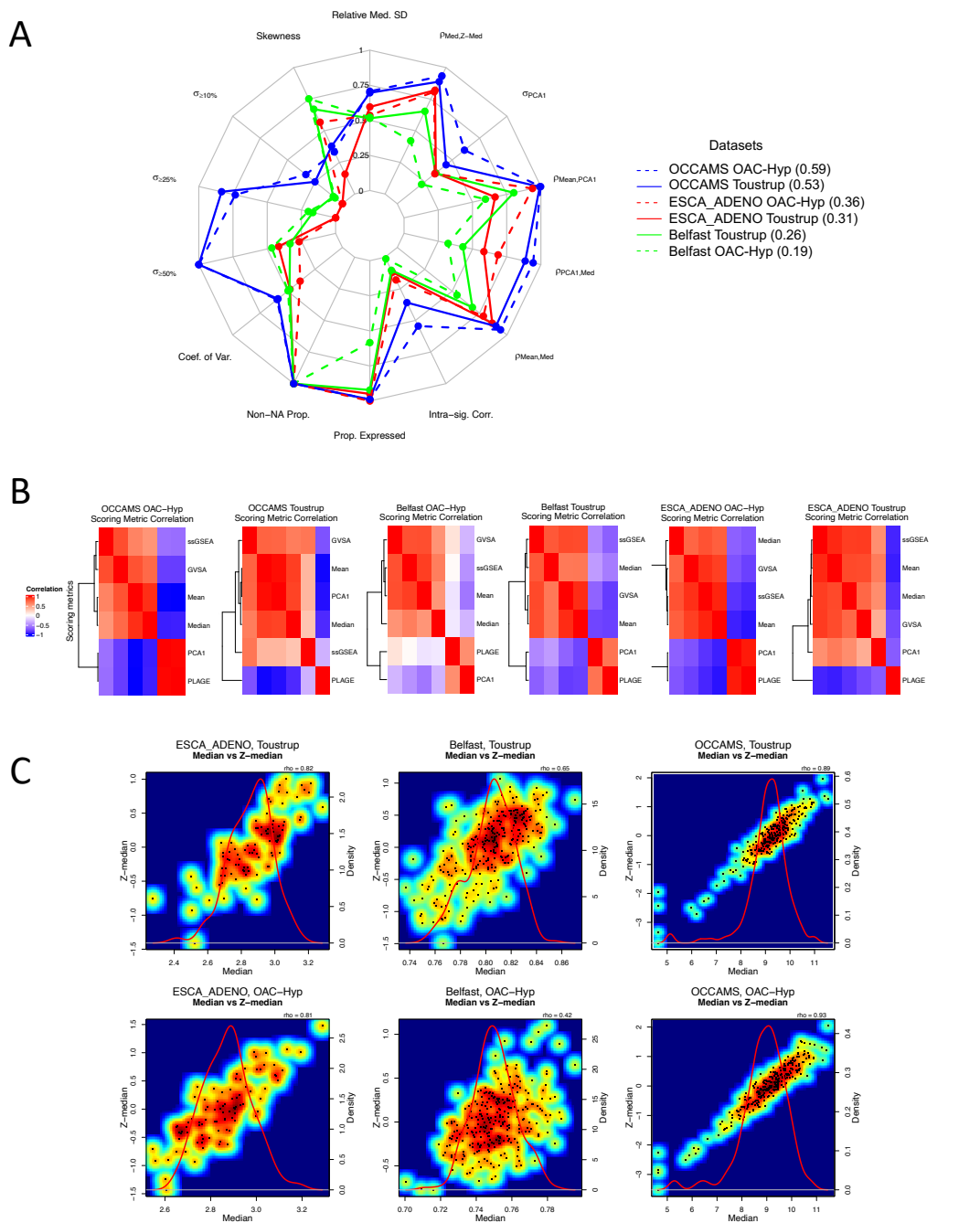


**A**, Western blot analysis of OAC cells following incubation at 0.2% O<sub>2</sub> for 48 hours for HIF and UPR activation markers. **B**, Venn diagram showing the overlap of hypoxia up-regulated genes (FC > 1.5, FDR p < 0.05) between OAC cell lines. **C**, Hypoxia gene co-expression network (Spearman's  $\rho$  > 0.5) of candidate signature genes in OCCAMS. **D**, Heatmap showing log fold change (RNA-seq data) of final novel signature gene signature in OAC cell lines cultured under hypoxia (0.2% O<sub>2</sub>) and normoxia for 48 hours.

**Figure 11. Derivation of the OAC-Hyp novel hypoxia-associated gene signature**

| Gene symbol     | Protein                                   | Function         |
|-----------------|-------------------------------------------|------------------|
| <i>ABCA5</i>    | ATP binding cassette subfamily A member 5 | Solute transport |
| <i>APLN</i>     | Apelin                                    | Angiogenesis     |
| <i>ARHGEF10</i> | Rho guanine nucleotide exchange factor 10 | Proliferation    |
| <i>C3</i>       | Complement C3                             | Immune           |
| <i>FOXL1</i>    | Forkhead box L1                           | Proliferation    |
| <i>HK2</i>      | Hexokinase 2                              | Glycolysis       |
| <i>ITGA1</i>    | Integrin subunit alpha 1                  | Metastasis       |
| <i>PRDM1</i>    | PR/SET domain 1                           | Immune           |
| <i>RNMT</i>     | RNA guanine-7 methyltransferase           | Proliferation    |
| <i>STC1</i>     | Stanniocalcin 1                           | Metastasis       |
| <i>TCP11L2</i>  | T-complex 11 like 2                       | Proliferation    |
| <i>YEATS2</i>   | YEATS domain containing 2                 | Proliferation    |

**Table 5. Novel hypoxia-associated gene signature (OAC-Hyp)**



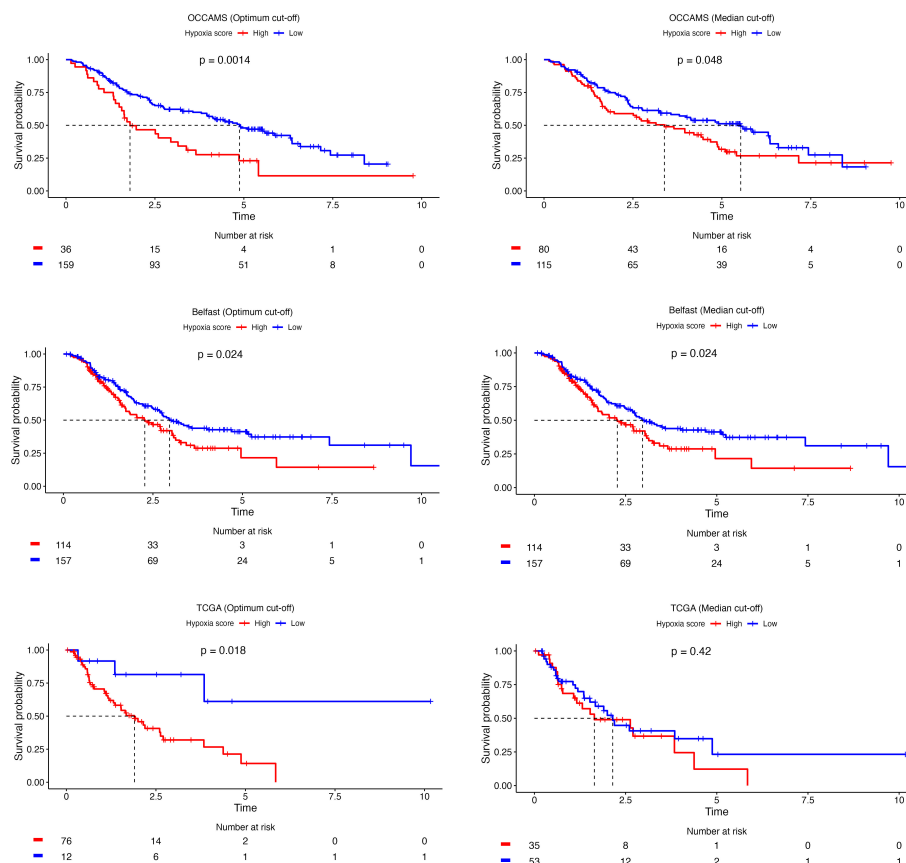
**A**, Radar plot summarizing signature quality metrics to evaluate compactness, expression of signature genes and variability of signature genes. **B**, Correlation between six gene signature scoring metrics (mean, median, PCA1, ssGSEA, GSVA, and PLAGE). **C**, Correlation between median and Z-score median scores. The colour scheme refers to the density of points in that region of the scatterplot. Correlation metric = Spearman's  $\rho$ .

**Figure 12. Quality evaluation of OAC-Hyp and Toustrup gene signatures**

## 2.4.4 Prognostic and predictive utility

The prognostic utility of the OAC-Hyp gene signature was evaluated using a Kaplan-Meier analysis with log-rank testing in the OCCAMS training cohort, and two independent datasets (TCGA-ESCA (OAC) and Belfast cohorts) (Figure 13). Cohorts were dichotomized by either median signature score, or an optimum cut-off assigned by the *surv\_cutpoint* function of *survminer*. Using an optimum cut-off, the signature was prognostic in OCCAMS ( $p = 0.0014$ ), Belfast ( $p = 0.024$ ) and TCGA-ESCA (OAC) cohorts. Using the median signature score, the signature was prognostic in OCCAMS ( $p = 0.048$ ), Belfast ( $p = 0.024$ ) but not TCGA-ESCA (OAC) cohorts ( $p = 0.42$ ). In contrast, the Tousttrup signature was not prognostic in OCCAMS or Belfast cohorts using either an optimum cut-off or median scores (Appendix 5). The Tousttrup signature was prognostic in TCGA-ESCA (OAC) using an optimum cut-off ( $p = 0.0051$ ) but was not prognostic using a median cut-off ( $p = 0.12$ ).

### OAC-HYP SIGNATURE



Kaplan-Meier plots showing survival curves for hypoxia-high and hypoxia-low subgroups defined by either an optimum cut-off score selected by *survminer* or median hypoxia score. A log-rank test was used to compare survival curves, and calculated P values are shown.

**Figure 13. Prognostic utility of OAC-Hyp signature for overall survival in independent datasets**

This was further evaluated by incorporating hypoxia scores in a Cox regression analysis for univariate and multivariate analysis in OCCAMS (Table 6) and TCGA-ESCA (OAC) (Table 7) datasets. There were no baseline differences between hypoxia-high and hypoxia-low cohorts stratified by median hypoxia score (Appendix 6). The OAC-Hyp signature was prognostic in the OCCAMS dataset on univariate (HR 1.17, 95% CI 1.05 – 1.29,  $p = 0.004$ ) and multivariate analysis (HR 1.13, 95% CI 1.02 – 1.26,  $p = 0.024$ ). On univariate analysis, the OAC-Hyp signature was also prognostic in the Belfast dataset (HR 1.04 (95% CI 1.01 – 1.08,  $p = 0.021$ )). The OAC-Hyp signature was prognostic in TCGA-ESCA (OAC) on univariate (HR 1.10, 95% CI 1.01 – 1.19,  $p = 0.025$ ) but not multivariate analysis (1.11, 95% CI 0.98 - 1.27,  $p = 0.10$ ). Meta-analysis revealed an HR of 1.06 (95% CI 1.03-1.09) on univariate and 1.12 (95% CI 1.03-1.22) on multivariate analysis (Figure 14 and Figure 15).

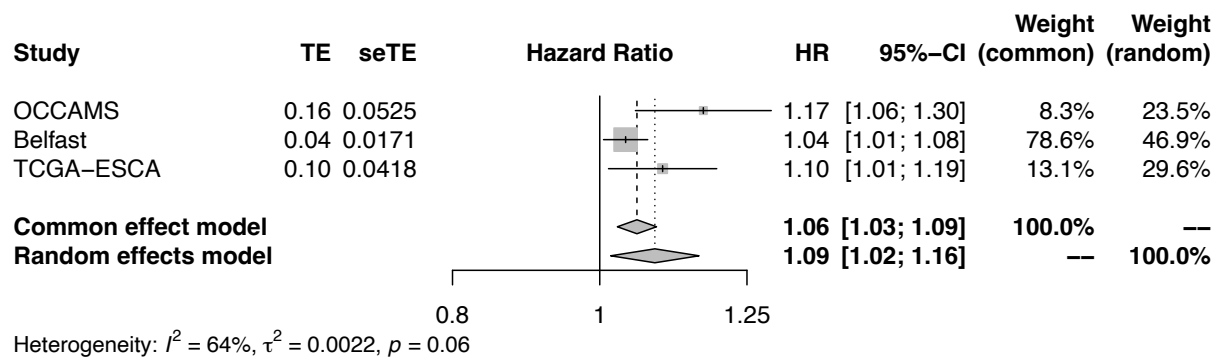


Figure 14. Forest plot of univariate hazard ratios (OAC-Hyp)

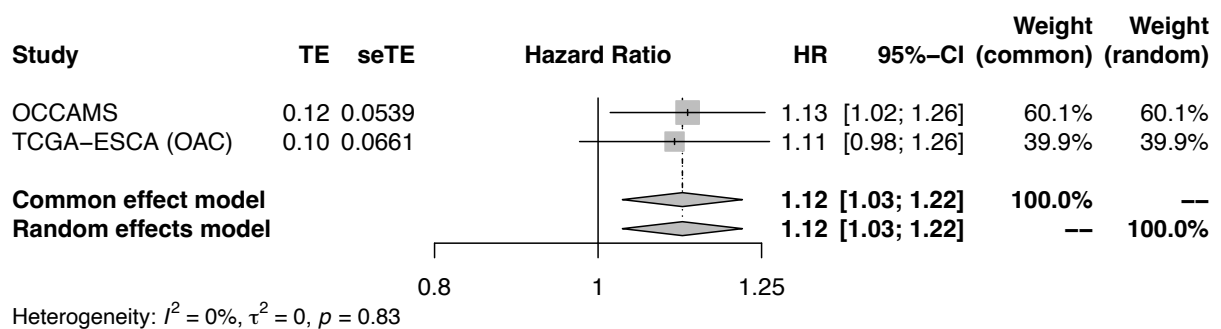


Figure 15. Forest plot of multivariate hazard ratios (OAC-Hyp)

In contrast, on univariate analysis, the Toustrup signature was not prognostic in the OCCAMS dataset (HR 0.98 (95% CI 0.91 – 1.06,  $p = 0.6$ )) or the Belfast dataset (HR 1.00 (95% CI 0.95 – 1.04,  $p = 0.9$ )). For TCGA, the Toustrup signature was prognostic on univariate (HR 1.08 (95% CI 1.02 – 1.15,  $p = 0.011$ )) but not multivariate analysis (HR 1.11 (95% CI 0.98 – 1.25,  $p = 0.10$ )).

| Characteristic                                           | Univariate analysis |                     |         | Multivariate analysis |                     |         |
|----------------------------------------------------------|---------------------|---------------------|---------|-----------------------|---------------------|---------|
|                                                          | HR <sup>1</sup>     | 95% CI <sup>1</sup> | p-value | HR <sup>1</sup>       | 95% CI <sup>1</sup> | p-value |
| Age                                                      | 1.03                | 1.01, 1.05          | 0.006   | 1.04                  | 1.02, 1.06          | <0.001  |
| Gender                                                   |                     |                     |         |                       |                     |         |
| Female                                                   | —                   | —                   |         | —                     | —                   |         |
| Male                                                     | 1.74                | 0.98, 3.11          | 0.060   | 1.59                  | 0.85, 2.97          | 0.15    |
| Smoking                                                  |                     |                     |         |                       |                     |         |
| Current                                                  | —                   | —                   |         | —                     | —                   |         |
| Former                                                   | 1.03                | 0.57, 1.85          | >0.9    | 0.72                  | 0.38, 1.37          | 0.3     |
| Never                                                    | 0.92                | 0.48, 1.77          | 0.8     | 0.89                  | 0.45, 1.75          | 0.7     |
| Unknown                                                  | 0.82                | 0.39, 1.75          | 0.6     | 0.72                  | 0.33, 1.61          | 0.4     |
| Stage                                                    |                     |                     |         |                       |                     |         |
| Stage 1                                                  | —                   | —                   |         | —                     | —                   |         |
| Stage 2A                                                 | 2.65                | 0.56, 12.5          | 0.2     | 4.22                  | 0.86, 20.7          | 0.076   |
| Stage 2B                                                 | 1.91                | 0.76, 4.76          | 0.2     | 1.82                  | 0.72, 4.59          | 0.2     |
| Stage 3                                                  | 2.49                | 1.18, 5.24          | 0.016   | 2.61                  | 1.23, 5.54          | 0.012   |
| Stage 4A                                                 | 2.59                | 1.09, 6.13          | 0.031   | 2.79                  | 1.17, 6.66          | 0.020   |
| Stage 4B                                                 | 7.48                | 3.16, 17.7          | <0.001  | 7.12                  | 2.90, 17.5          | <0.001  |
| Unknown                                                  | 2.12                | 0.45, 10.0          | 0.3     | 2.23                  | 0.46, 10.7          | 0.3     |
| OAC-Hyp hypoxia score                                    | 1.17                | 1.05, 1.29          | 0.004   | 1.13                  | 1.02, 1.26          | 0.024   |
| <sup>1</sup> HR = Hazard Ratio, CI = Confidence Interval |                     |                     |         |                       |                     |         |

Table 6. Hazard ratios (OS) from Cox proportional hazards model (OCCAMS)

| Characteristic                                           | Univariate analysis |                     |         | Multivariate analysis |                     |         |
|----------------------------------------------------------|---------------------|---------------------|---------|-----------------------|---------------------|---------|
|                                                          | HR <sup>1</sup>     | 95% CI <sup>1</sup> | p-value | HR <sup>1</sup>       | 95% CI <sup>1</sup> | p-value |
| Age                                                      | 0.99                | 0.96, 1.01          | 0.4     | 0.99                  | 0.93, 1.05          | 0.7     |
| Gender                                                   |                     |                     |         |                       |                     |         |
| Female                                                   | —                   | —                   |         | —                     | —                   |         |
| Male                                                     | 0.81                | 0.28, 2.36          | 0.7     | 0.50                  | 0.10, 2.48          | 0.4     |
| Smoking                                                  |                     |                     |         |                       |                     |         |
| Lifelong non-smoker                                      | —                   | —                   |         | —                     | —                   |         |
| Current smoker                                           | 1.84                | 0.37, 9.19          | 0.5     | 1.34                  | 0.12, 14.8          | 0.8     |
| Current reformed smoker > 15 years                       | 1.19                | 0.44, 3.19          | 0.7     | 1.79                  | 0.44, 7.31          | 0.4     |
| Current reformed smoker < 15 years                       | 1.47                | 0.37, 5.78          | 0.6     | 0.73                  | 0.07, 7.45          | 0.8     |
| Stage                                                    |                     |                     |         |                       |                     |         |
| Stage 1                                                  | —                   | —                   |         | —                     | —                   |         |
| Stage 2                                                  | 2.30                | 0.63, 8.37          | 0.2     | 1.65                  | 0.27, 10.0          | 0.6     |
| Stage 3                                                  | 4.95                | 1.22, 20.1          | 0.025   | 2.27                  | 0.39, 13.3          | 0.4     |
| OAC-Hyp Hypoxia score                                    | 1.10                | 1.01, 1.19          | 0.025   | 1.11                  | 0.98, 1.27          | 0.10    |
| <sup>1</sup> HR = Hazard Ratio, CI = Confidence Interval |                     |                     |         |                       |                     |         |

Table 7. Hazard ratios (OS) for Cox proportional hazards model (TCGA-ESCA (OAC))

Two datasets also contained response data for neoadjuvant platinum-based chemotherapy. Cohorts were stratified by median hypoxia score. In the OCCAMS dataset, there were no significant differences between the proportions of non-responders (TRG 4/5) comparing hypoxia-high vs hypoxia-low groups (35% vs. 36%; p = 0.54) although 47% of patients had missing/unknown response data. In the Belfast dataset, there were also no significant differences between the proportions of non-responders

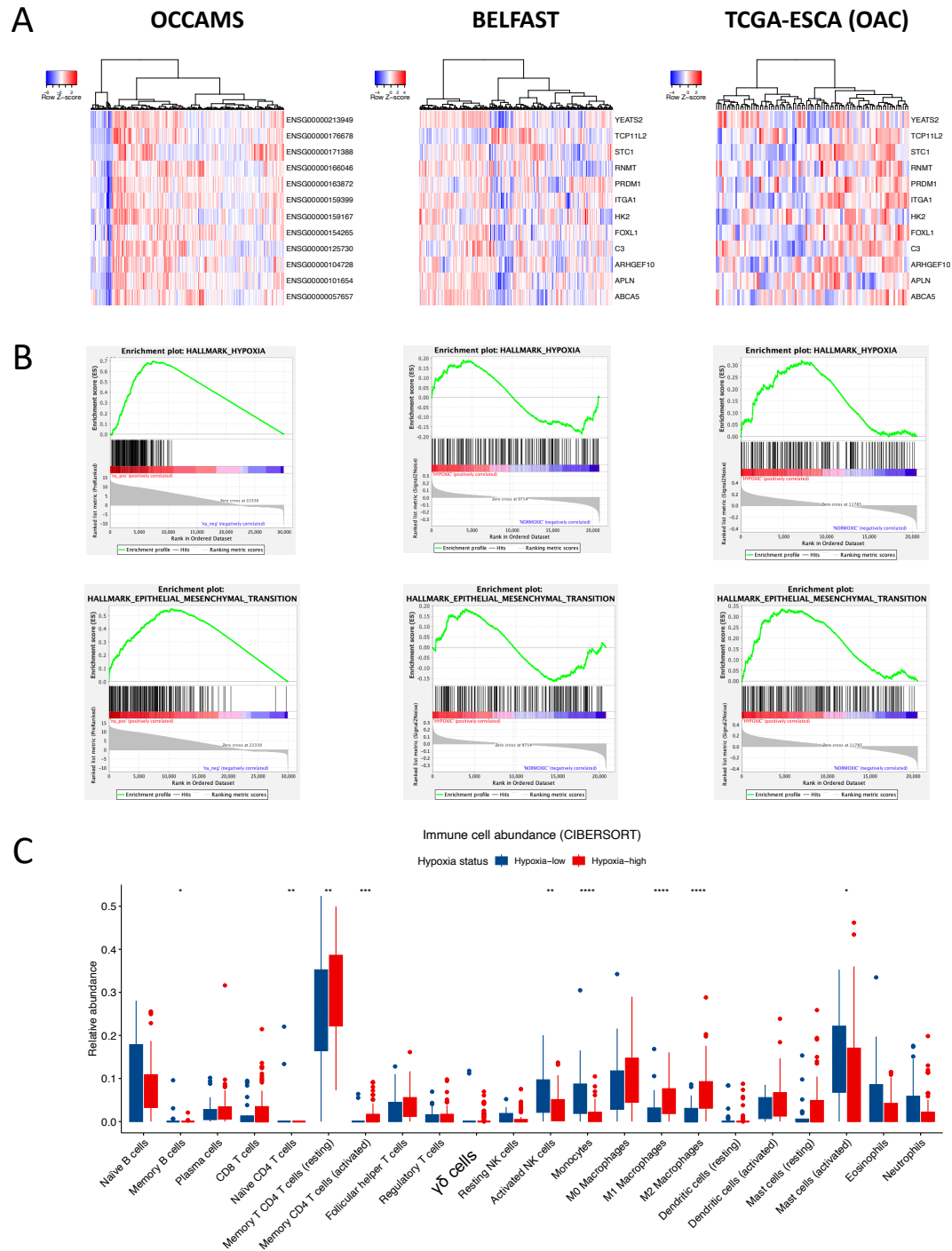
comparing hypoxia-high vs hypoxia-low groups (88% vs. 88%;  $p > 0.99$ ) although 25% had missing/unknown response data.

#### **2.4.5 Differential gene expression analysis**

To identify functional sub-groups for differential gene expression analysis, unsupervised hierarchical clustering was used to cluster datasets into hypoxia-high, -intermediate and -low sub-groups (Figure 16A). Gene Set Enrichment Analysis was used to identify pathways upregulated in the hypoxia-high cohort relative to the hypoxia-low cohort to determine whether the gene signature identified tumours with distinct molecular phenotypes. This revealed that the HALLMARK\_HYPOXIA gene set was consistently and strongly up-regulated in the hypoxia-high cohorts ( $P < 0.05$ ) (Figure 16B). The HALLMARK\_EPITHELIAL\_MESENCHYMAL\_TRANSITION was also consistently up-regulated in all hypoxia-high groups in all three cohorts. Additional HALLMARK gene sets up-regulated in at least 2 out of 3 datasets were: angiogenesis, reactive oxygen species, glycolysis, E2F targets, interferon gamma response, IL2-STAT signalling, and TNF signalling.

#### **2.4.6 Immune response signature**

Since it was noted that the hypoxia-high cohort was enriched with pathways involved in immune responses on GSEA, a further analysis was performed. The CIBERSORT tool is an *in-silico* tool used to deconvolve bulk gene expression data and characterize immune cell type proportions of 22 functionally defined human immune subsets (Figure 16C). This revealed that the hypoxia-high cohort in the OCCAMS dataset had an increased infiltration of: memory CD4 T cells (resting and activated) and macrophages (M1 and M2). Conversely, there was a decreased infiltration of: memory B cells, naïve CD4 T cells, activated NK cells, monocytes and activated mast cells.



**A**, Unsupervised hierarchical clustering of datasets using the OAC-Hyp signature, into hypoxia-high, -intermediate, and -low cohorts. **B**, Enrichment plots of HALLMARK gene sets up-regulated in hypoxia-high clusters. **C**, CIBERSORT analysis of immune cell fractions in the OCCAMS dataset, comparing hypoxia-high and -low clusters..

**Figure 16. Molecular and immune phenotypes of hypoxic OAC tumours**



## 2.5 Discussion

In this chapter, a novel, prognostic OAC-specific hypoxia-associated gene expression signature was derived using a network-based approach.

GSEA analysis confirmed that hypoxia-high cohorts identified by OAC-Hyp had a distinct phenotype. Induction of pathways such as EMT and angiogenesis suggested that hypoxic tumours are associated with invasiveness and metastasis at a molecular level. This is consistent with data suggesting that hypoxic signalling may be a key promoter of metastasis<sup>201</sup>. The immune signature of hypoxia-high tumours also revealed a suppression of immune responses, most strongly indicated by reduced infiltration of activated NK cells and naïve CD4 T cells. This is corroborated by literature data which suggests that hypoxia can suppress immune responses through multiple mechanisms and may promote resistance to immunotherapy<sup>202</sup>. Overall, this indicates hypoxia-high tumours identified by OAC-Hyp had aggressive molecular characteristics and a tendency to suppress anti-tumour immune responses, which may partly explain the poor long-term survival.

The Toustrup signature performed best in a global review of published hypoxia gene signatures. This signature was originally derived to predict hypoxic modification in head and neck squamous cell carcinoma and is the only gene signature validated using direct tissue pO<sub>2</sub> measurements<sup>47</sup>. Toustrup signature genes were induced in a wide range of cancer cell lines<sup>188</sup>, correlated with other signatures, and had high quality statistical metrics. This suggests that it could be used to classify hypoxic status in different cancer types, although the OAC-specific gene signature appeared to perform better as a prognostic biomarker in OAC.

A key strength of this Chapter was a detailed analysis of signature quality. The standardized output generated by *sigQC* establishes whether the favourable expression patterns and statistical properties of genes in a signature are maintained across multiple independent datasets. This is useful in evaluating whether a gene signature has translational utility but does not evaluate how well the signature performs in detecting relevant biology. The signature quality metrics confirmed that both OAC-Hyp and Toustrup signatures perform well in terms of expression, variability, and signature compactness. OAC-Hyp performed better than Toustrup signature in two out of three datasets. It was notable that there was a large variability in the strength of median and Z-score scoring between datasets which may be an effect of variable data pre-processing. Previous work has noted that variable data-pre-processing can alter the performance of the signatures<sup>176</sup>. Critical future work will be to

explore the utility of an ensemble analyses which involve combining multiple pre-processing techniques to evaluate signature quality.

The retrospective stratified VOC analysis provided encouraging evidence that candidate hypoxia-responsive VOCs could be identified using a hypoxia gene signature. However, the data is limited by the small sample size, small range of target VOCs and a lack of chromatography which limits the ability of SIFT-MS to differentiate isomers confidently (e.g. small aldehydes and ketones).

The novel gene signature was found to be prognostic on multivariate analysis in two datasets, suggesting that tumour hypoxia may be an independent prognostic factor in OAC. A limitation of the present work, however, is a lack of additional covariate data (e.g. performance status) which was not recorded or missing in these datasets. Additionally, no dataset had matched tumour tissue  $pO_2$  to enable selection of a cut-off hypoxic score to classify tumours as hypoxic, therefore the median score was used instead. Hypoxic status was not associated with the response to neoadjuvant chemotherapy in OAC in this analysis. This may be partly explained by missing data or related to the fact that non-microenvironmental factors may play a more important role in drug resistance than hypoxia<sup>203</sup>. An additional explanation may be inadequate sampling of tumours given invasive edge was not characterised. Therefore, the relationship between hypoxia and chemo-radiotherapy response requires further evaluation.

In summary, a rational and unbiased informatics approach was used to derive an OAC-specific hypoxia-associated gene signature for hypoxia classification in OAC that identified clinically relevant subgroups.

## **CHAPTER 3 – HYPOXIA-ASSOCIATED EXHALED VOCS IN OESOPHAGEAL ADENOCARCINOMA**

## Summary

The aim of this chapter was to perform a pilot clinical phenotyping study of exhaled VOCs from OAC patients stratified by hypoxic gene expression (OAC-Hyp). A laser capture microdissection workflow was applied to patient samples to extract tumour RNA for sequencing. Pre-operative breath analysis was performed with a sorbent-based breath capture technique combined with TD-GC-TOF-MS analysis. Breath contains a complex mixture of analytes range in polarity and volatility, therefore two complementary MS methods (mid-polar and polar) were used for comprehensive breath VOC profiling on breath samples. A sequential re-collection workflow was also validated which enabled analysis of the same breath samples with GCxGC-TOF-MS to improve compound identification in untargeted workflows. Quality control procedures were integral to both RNA sequencing and breath analysis. However, technical challenges with extracting high-quality RNA from tumour tissue yielded a significant attrition rate. Both targeted (quantitative) and untargeted (tile-based) GC-MS data analysis techniques were employed in the final cohort (n = 30). Candidate VOC biomarkers suitable for future validation studies were identified from both analyses with the ability to accurately discriminate hypoxic classes.

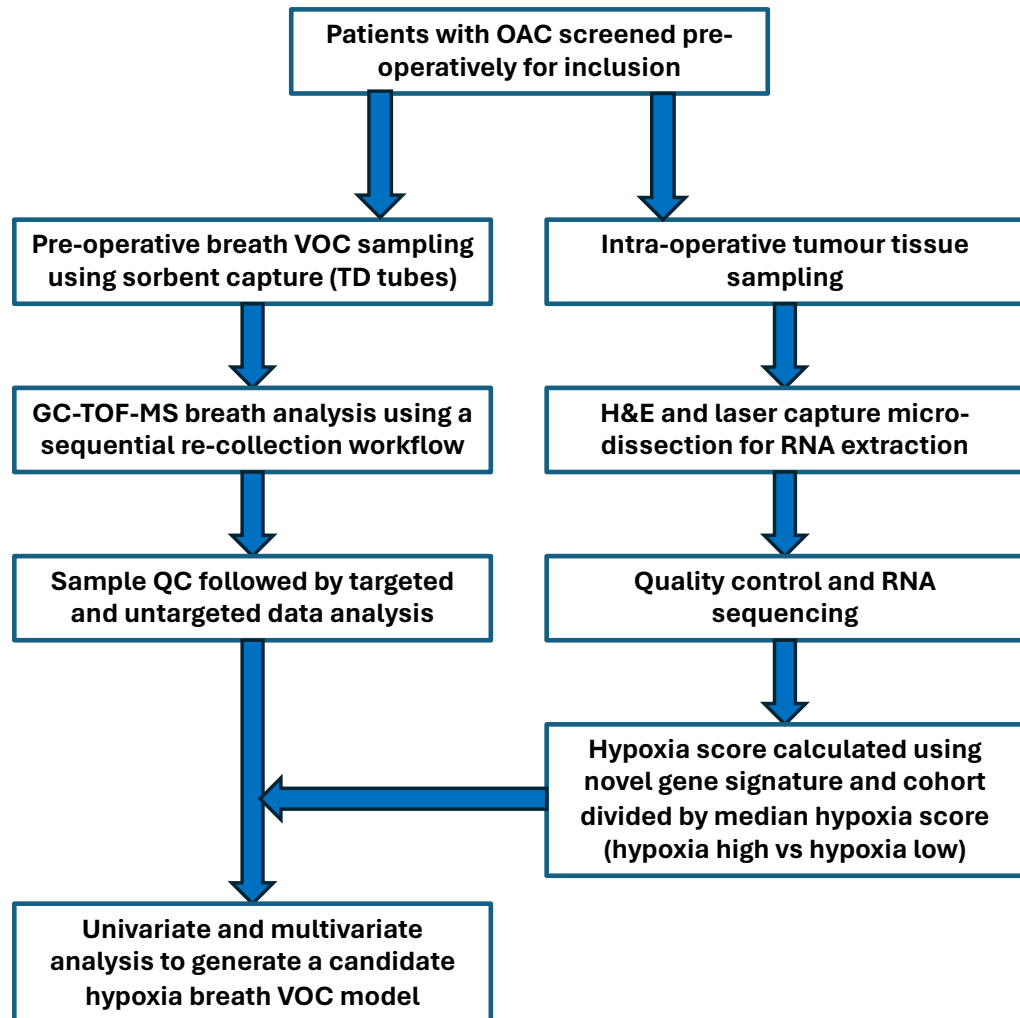


Figure 17. Summary of the pilot clinical phenotyping study

## **3.1 Methodological rationale**

### **3.1.1 Pilot study design**

The primary objective for this section was to identify exhaled VOCs with significant association with tumour hypoxia, defined by a validated OAC-specific hypoxia gene expression signature. Since this is the first study design of its kind, a pilot clinical study was performed to generate data on sample attrition and methodology in addition to possible VOC biomarkers of hypoxia. It is envisaged that this pilot data will be applied in a larger multi-centre prospective clinical study to build a robust hypoxia VOC model, ready for evaluation in clinical trials of HTT.

There are several advantages of a candidate-based or 'rational' biomarker approach proposed in this work, rather than only using untargeted global profiling in clinical trials of HTT to build a predictive model. Untargeted analysis often fails in isolation due to a bias towards abundant compounds and multiplicity error. Quantification is relative and there are significant challenges with compound identification and batch effects<sup>204</sup>. A bespoke targeted method permits the unambiguous and absolute quantification of selected analytes. Finally, linking a proposed biomarker with the underlying pathophysiology is important for clinical endorsement.

### **3.1.2 Laser capture micro-dissection**

To quantify gene signature expression in patient samples, a suitable tissue RNA extraction technique was required. Within whole tumour biopsies, malignant cells can be intermixed with variable amounts of stromal, immune, and normal epithelial cells. This background interference can significantly limit the sensitivity and specificity of molecular analyses. Therefore laser capture micro-dissection (LCM) was chosen as the preferred method of isolating pure tumour cell populations for RNA sequencing. Spatial transcriptomics was considered as an alternative option, but due to technical constraints has a lower depth of transcript coverage compared to the comprehensive coverage offered by RNA-seq.

The instrument used in this work was a Zeiss PALM MicroBeam which employs a 355nm pulsed frequency tripled solid-state laser integrated into a microscope. Importantly, this wavelength does not adversely affect biomolecules, enabling preservation of molecular data. The beam energy and diameter can be adjusted, focusing it to under 0.6µm for precise small-scale manipulation based on

objective magnification and numerical aperture. The energy from the laser catapults the dissected specimen off the slide into the adherent cap of a specially designed microfuge tube.

### **3.1.3 Choice of breath analysis technique**

Breath is a complex mixture of analytes ranging in polarity and volatility, with many VOCs present at very low (sub-parts per trillion, ppt) concentrations<sup>205</sup>. The diagnostic VOCs in OAC patients, for example, are predominantly fatty acids and aldehydes<sup>107,109</sup>. The ideal analytical pipeline for breath analysis trials would therefore permit the accurate separation, identification, and quantification of all potentially relevant VOCs in breath.

Previous OAC studies utilised direct (online) breath sampling analysed by selected ion flow tube mass spectrometry (SIFT-MS)<sup>107,109</sup>. This method relies on soft chemical ionization of breath VOCs by reagent ions within a flow tube, which produces characteristic product ions detected by a quadrupole mass spectrometer. This is a simple and sensitive method which can offer real-time analysis. However limited mass resolution, separation, and selectivity make it challenging to differentiate between isomeric compounds that have similar mass-to-charge ratios and ionization/fragmentation behaviours.

Gas chromatography-mass spectrometry (GC-MS) remains the gold standard for VOC analysis as this technique allows a high level of confidence in compound identification through chromatographic separation. Pre-concentration of breath onto sorbent-packed thermal desorption (TD) tubes prior to GC-MS analysis represents a convenient and robust solution for sample collection, transport, and storage of breath VOCs. With gas chromatography, the main methodological challenge is selecting a column phase that is suitable for the majority of expected analytes. A mid-polar cyanopropyl-type column is ideal for most compounds in breath such as alkanes, aldehydes, ketones etc. but can give poor performance with volatile fatty acids. Conversely, phases that are optimized for volatile fatty acids, such as wax type columns tend to give poor performance with low polarity compounds.

### **3.1.4 Sequential re-collection principle**

Even when using two separate column phases to analyse a sample the complexity of breath means compound identification can be difficult due to co-elutions. Two-dimensional (2D) gas chromatography (GCxGC-MS) can offer much greater chromatographic resolution but generally

requires slow temperature ramps which can reduce throughput. Hence, breath analysis using two high throughput one-dimensional (1D) methods with pooling of samples for analysis by GCxGC-MS to identify poorly resolved compounds of interest would seem optimum. However, offline breath analysis using TD tubes is usually considered a 'one shot' technique as the loss of sample after thermal desorption prevents the opportunity for re-analysis using alternative techniques.

Automated split flow re-collection technology was therefore used in this study as this allows a portion of every breath sample to be directed onto a TD tube after desorption. Breath samples can be split and re-collected at different stages of the GC injection process which can be tailored to the requirements of the assay - after tube desorption (inlet split), trap desorption (outlet split) or both (double split). This provides the opportunity to enhance sample characterization using different GC instruments or methods without the need to collect additional samples. Since single tube re-collection can lead to reduced analyte recovery with every re-collection cycle as only a percentage of the original sample is re-collected, sequential re-collection to pool breath samples can also be performed to improve VOC detection. This involves re-collecting the split from several sample replicates onto one tube. This is particularly advantageous for trace level compounds in breath at the thresholds of detectability (sub-ppt).



## 3.2 Hypothesis and aims

Chapter hypothesis: Exhaled VOCs can act as non-invasive surrogates of OAC tumour hypoxia.

The objective of this section was to utilise optimised RNA sequencing and GC-MS analytical techniques to identify candidate exhaled VOCs of tumour hypoxia in a pilot prospective clinical phenotyping study. Specifically this section aimed to:

1. Develop and apply a laser capture microdissection and RNA sequencing workflow for OAC tissue.
2. Apply a quality-controlled GC-MS analytical pipeline to analyse breath samples from OAC patients.
3. Conduct a pilot prospective analysis of exhaled VOC data from OAC patients stratified by hypoxic gene expression to identify candidate hypoxia-responsive VOCs.

### 3.3 Methods

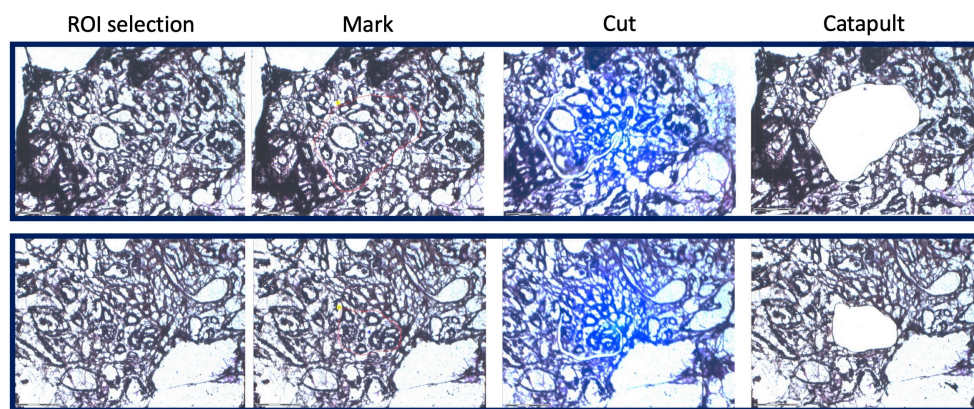
#### 3.3.1 Protocol for the pilot prospective phenotyping study

All consecutive adult patients (aged 18-90) formally diagnosed with OAC on histology (including Siewert 3 GOJ tumours) who were attending Imperial College NHS Healthcare Trust for surgery (either staging laparoscopy or surgical resection) between September 2022 and September 2023 were eligible for inclusion. After informed consent on the morning of surgery, patients underwent breath collection and tissue sampling following standardized protocols (below). Clinical data was collected on a standardized proforma, and included time of last oral intake, smoking status, neoadjuvant chemotherapy status, medical history, and medication history. The patient exclusion criteria for this study included: OSCC, active infection, previous oesophageal or gastric resection, synchronous cancer, or patients unable to provide consent or a breath sample. This study received ethical approval from the Imperial College Tissue Bank committee (REC: 17/WA/0161; Project: R20037) (Appendix 7). As this was a pilot study, a formal sample size calculation was not performed.

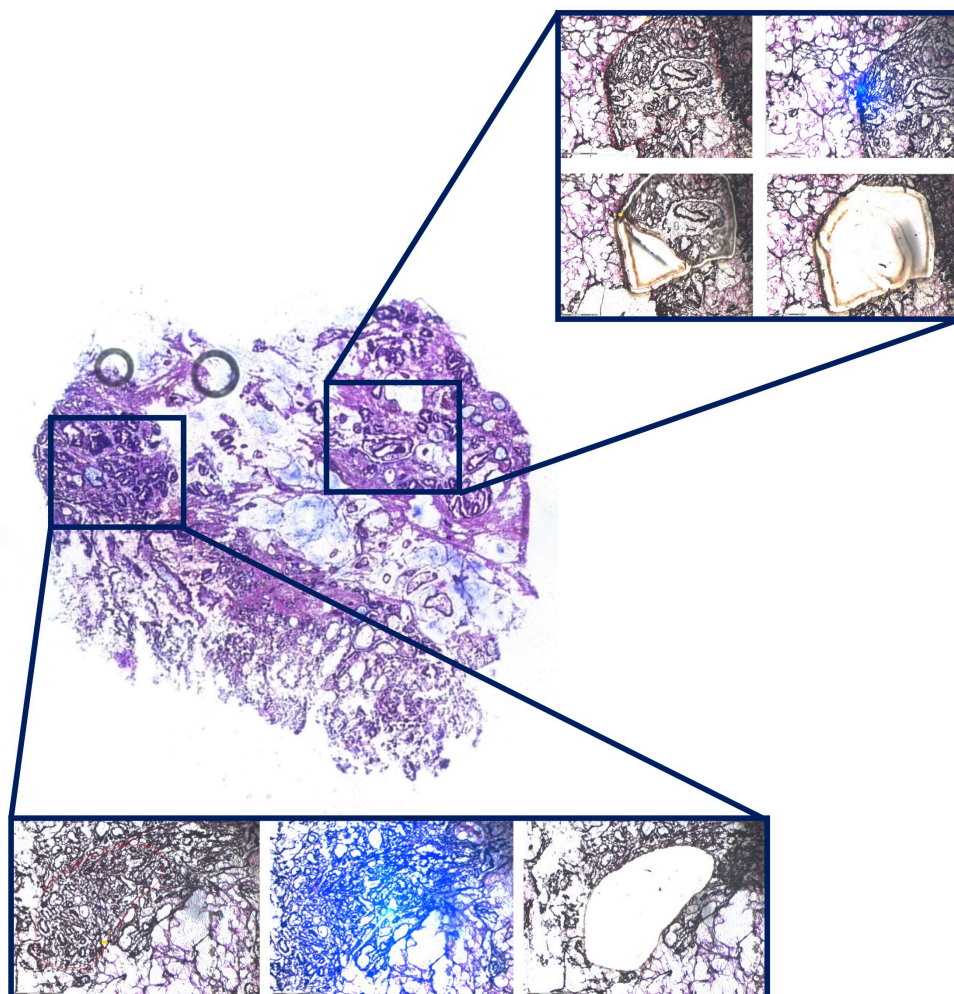
#### 3.3.2 Laser capture microdissection (LCM)

LCM best practice was followed as per the original Nature Protocol<sup>206</sup>. At the time of endoscopy or immediately after oesophageal disconnection during oesophagectomy, tissue biopsies were retrieved and snap frozen using a dry ice/ethanol bath and stored at -80°C. Frozen tissue biopsies were mounted in nuclease-free water and cryosectioned in a microtome set to -20°C. Tissue sections of 10µm thickness were cut and mounted onto plain glass slides for haematoxylin and eosin (H&E) staining, and nuclease-free polyethylene naphthalate (PEN) membrane glass slides (Leica Microsystems). Slides were stored at -80°C prior to LCM. Tissue sections on plain glass slides were stained using an H&E kit (Abcam #ab245880) as per manufacturer instructions and imaged using a widefield microscope (Zeiss Axio Observer). Tissue sections on PEN membrane slides were stained using the Histogene™ LCM Frozen Section Staining Kit as per manufacturer instructions (Thermo Fisher Scientific, UK). Employing the H&E image as a guide, precise cutting outlines were meticulously traced around cancer cells to mitigate any risk of normal tissue contamination (Figure 18 and Figure 19). Laser-captured cells were collected on LCM AdhesiveCaps 200 (Carl Zeiss™) using a Zeiss PALM MicroBeam LCM system (Carl Zeiss™). Multiple tumour regions were obtained from at least three different regions of the tumour. All microdissected samples from a single biopsy were acquired within 30 minutes. Buffer RLT with supplemental β-mercaptoethanol was added to sample tubes and tubes were immediately inverted.

Strict RNA isolation precautions were rigorously adhered to, including use of RNase removal agents, fresh tips, certified plastic, and regular glove changing.



**Figure 18. Steps of laser capture microdissection**



**Figure 19. Laser capture microdissection can be used to isolate pure tumour cell populations**

### 3.3.3 RNA extraction

Lysate was then added to a QIAshredder column and centrifuged at full speed for 2 minutes. RNA was then precipitated, washed, and eluted from this lysate using the RNeasy Mini Kit (Qiagen) according to manufacturer instructions. Extracted RNA was quantified using absorption spectroscopy (Nanodrop, ThermoFisher) and RNA integrity was verified using electrophoresis (RNA 6000 Pico Kit with the 2100 Bioanalyzer, Agilent). The RNA sample quality was assessed based on the fragment size distribution, specifically noting two peaks that corresponded to 18S and 28S ribosomal RNAs, along with a signal in the small molecular weight range indicative of small RNAs. This distribution was cross-referenced with a comprehensive database of RNA samples to calculate an RNA Integrity Number (RIN). The resulting electropherograms were used to evaluate both the integrity and concentration of the RNA samples. The RIN, calculated by the instrument, assigned a value where 10 signifies intact RNA, while a value of 1 indicates complete degradation.

### 3.3.4 RNA sequencing

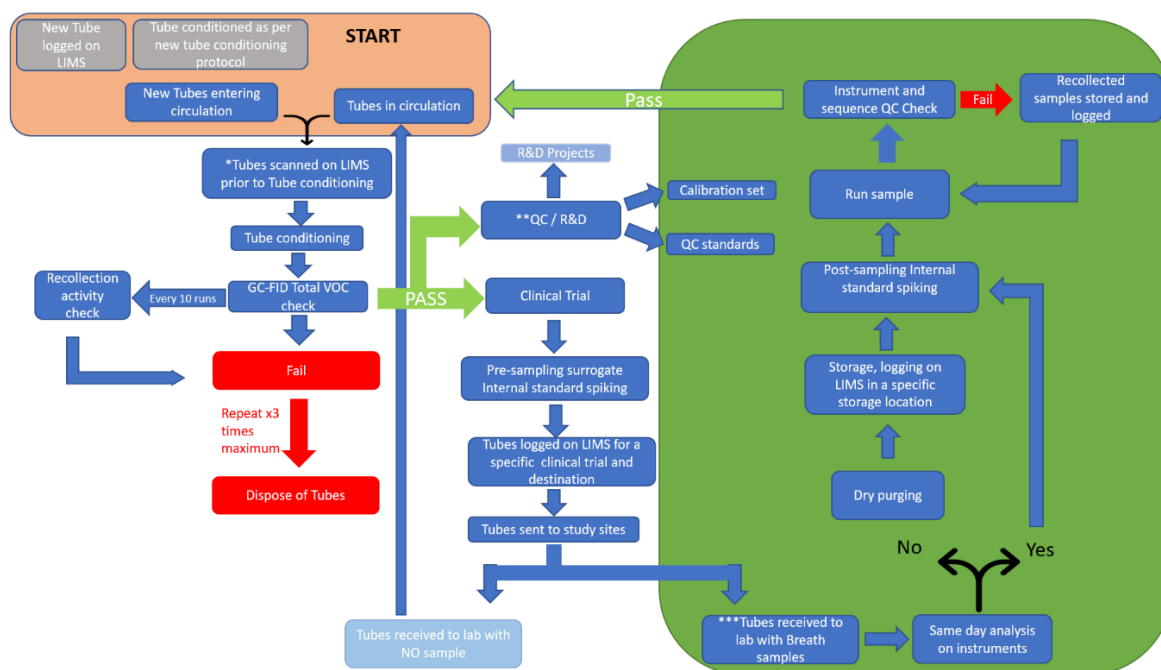
Only RNA with an RNA integrity number  $\geq 4$  with a smooth base line, 260/280nm ratio  $> 1.8$  and 260/230nm ratio  $> 1.8$  were taken forward for sequencing by Novogene. Library preparation was performed using the Illumina TruSeq® Stranded mRNA assay. Poly-A tail RNA sequencing at 20 million paired-end reads was conducted in three separate biological replicates for each of the three OAC cell lines, utilizing the Illumina Hi-Seq4000 platform. Upon removal of sequence adapters, Trimmomatic (v0.36)<sup>192</sup> was employed for quality trimming of the reads. Subsequently, alignment against the human reference genome (GRCH38) was performed, and the counts per gene were calculated using STAR (v 2.5.3)<sup>193</sup> with annotation sourced from GENCODE version 27. Hypoxia gene signature scores for patient samples were then constructed using a standardized method<sup>176,198,199</sup> (Figure 7). For each gene of the signature derived in Chapter 2, patients with the top 50% expression for each gene were given a score of +1, and bottom 50% were given a score of -1. This was repeated for each gene of the signature, and the sum total was the hypoxia score for each patient. High scores indicated the tumour was hypoxic, and low scores indicated normoxia. Patients were stratified into hypoxia-high and hypoxia-low groups based on the median cohort hypoxia score. Clinical demographics and breath VOC data were compared between hypoxia-high and hypoxia-low groups.

### **3.3.5 Principles of thermal desorption for breath VOC capture**

Thermal desorption (TD) tubes contain highly concentrating sorbents (Tenax TA/Carbograph-5TD) that are capable of capturing VOCs. The VOCs captured on the sorbents are then thermally desorbed without damaging the TD tube so it can be re-used multiple times. Their efficiency in capturing compounds also means they can capture exogenous contaminants. The sorbent materials themselves can gradually degrade resulting in artefact peaks. It is therefore essential to have a dedicated conditioning protocol for TD tubes, and accessories (i.e. caps) that come into contact with the analytical flow path.

### **3.3.6 Quality assurance**

TD tube preparation and analysis was performed within a bespoke laboratory (VOC Laboratory, Hammersmith Hospital Campus, Imperial College London) designed for the analysis of VOCs. The complete workflow is displayed in Figure 20. The laboratory has been designed to minimise the risk of environmental VOCs contaminating samples. Specific design features include: positive pressure air flow, active VOC air filtration and the exclusive use of inert materials (e.g. stainless steel, glass, copper, VOC-free paint). Zero air, hydrogen and nitrogen were produced continuously from standalone generators and filtered using hydrocarbon traps (Supelco). Ultra-pure helium was purchased separately (BOC) and was also filtered using a hydrocarbon trap. Long-term storage caps, robot caps and diffusion locking caps were cleaned daily in a vacuum oven at a temperature of 200°C at <20mbar pressure. Only individuals who had been trained in the laboratory quality assurance procedures were allowed to enter these facilities. The status of every TD tube was logged by barcode scanning during processing and information entry into a Laboratory Information Management System (LIMS). Tubes were handled with gloved hands to prevent contamination with skin VOCs and lab members were not permitted to wear heavily scented cosmetic products. Cleaning products were also not permitted in areas used for tube preparation.



**Figure 20. Complete TD tube workflow**  
Produced by Aaron Parker (VOC Laboratory Manager)

### 3.3.7 Preparation of TD tubes for breath sampling

TD tube preparation followed a Standard Operating Procedure developed by the VOC laboratory and implemented by trained individuals only. Regular audits were maintained by the lab to ensure individuals adhered to this SOP. TD tubes were conditioned (cleaned) using a TC-20 TD tube conditioning unit (Markes International Ltd, Llantrisant, UK). An individual unit could condition 20 tubes simultaneously. The flow rate was checked with a flow meter to ensure N<sub>2</sub> flow was at 100ml/min. Tubes were inserted into the manifold with the sampling end (grooved end) facing away from the device. The manifold with tubes was then placed in the heating block. Tubes were conditioned for 2 hours at 310°C under continuous nitrogen flow. The manifold was then removed from the heating block to allow the tubes to cool for 5 minutes. Inert diffusion-locking caps (DiffLok™, Markes International Ltd, Llantrisant, UK) were placed on the sampling ends of the tubes, and stainless-steel caps were placed on the non-sampling ends, minimizing the time the tubes were open to the atmosphere. These tubes were then placed in TD autosampler trays.

### 3.3.8 GC-FID method

Conditioned TD tubes were subsequently checked for cleanliness by analysis using a gas chromatography-flame ionization detection (GC-FID) instrument. TD tubes were loaded onto a TD100-xr thermal desorption unit and sequentially pre-purged with helium (50 mL/min) for 30 seconds then

heated to 280°C for 5 minutes. VOCs were desorbed onto a focusing trap (U-T12ME-2S, Markes International) with a 5.5:1 outlet split, which was then heated to 280°C for 5 min to release VOCs into the GC column. The GC system used was an 8890 (Agilent Technologies) coupled to an FID. All data were acquired and pre-processed with ChromSpace (V2.1.6, Markes International). The focusing trap was set at 30°C and a fused silica capillary column (HP-5ms, 30 m x 250µm x 0.25µm, Agilent) was used with hydrogen carrier gas. The GC oven was initially set at 60°C for 1 min, then increased to 300°C (50 °C/min with a 2 min hold). Following Top-hat background subtraction (peak width set at 20 seconds), peaks were integrated using the 'curve fitting' peak detection algorithm within the ChromSpace software, using pseudo-Gaussian smoothing at 3 points. The total area of all identified peaks was calculated and compared against previously set QC limits displayed in a Levey-Jennings plot (Figure 21). These limits had been set by analyzing the total peak area in 500 conditioned TD tubes. Tubes with a total peak area > 1 s.d. of the mean were classed as a QC failure and were re-conditioned with new caps. Conversely, a total peak area < 1 s.d. could indicate a trap blank had run due to a tube pressurization failure (e.g. from worn O rings within the caps). These tubes were also re-conditioned with new caps. Tubes passing the QC check on the GC-FID progressed to surrogate addition.

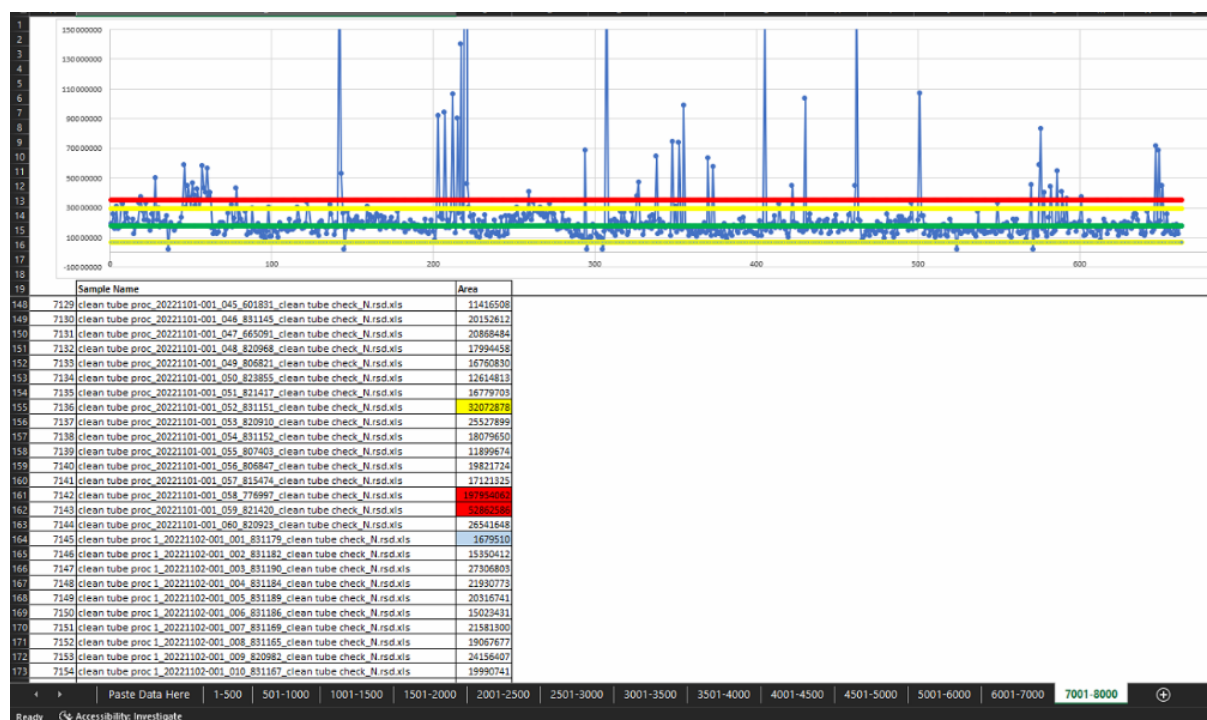


Figure 21. GC-FID QC log file

### **3.3.9 Surrogate addition**

A surrogate is an analyte added to samples prior to extraction, or in the case of breath, an analyte added to TD tubes prior to breath collection. Surrogate addition monitors all the steps from tube preparation to final analysis. Low surrogate recovery during data analysis would constitute a QC failure. Neat surrogate standard (naphthalene-d8) was purchased from Sigma (product code: 176044-1G). Stock surrogate (1000mg/L) and working surrogate (5mg/L) were prepared in GC-MS grade methanol using calibrated balance and pipettes. Surrogate was stored at -10°C and had an expiry of 1 year. After passing the GC-FID check, TD tubes were capped with robot caps in place of the inert Difflok caps and loaded onto a PAL3 Series III Dual Head Sample Handling Robot. The sample handling robot spiked 1µl of working surrogate and dry purged each conditioned tube. The sequence included a syringe check to ensure the syringe needle was patent. Once complete, the caps were replaced with long-term storage caps, and the tubes were ready for sampling.

### **3.3.10 Breath sampling procedure**

The breath sampling procedure followed a Standard Operating Procedure developed by the VOC laboratory (SOP002) and was implemented by trained individuals only. Regular audits were maintained by the lab to ensure individuals at internal and external sites adhered to this SOP. This technique captured VOCs in mixed expiratory breath and acquired all phases of exhaled air. Participants were asked to fast prior to breath sampling for a minimum of 6 hours for solid food and 2 hours for liquids to minimize the contribution of exogenous VOCs derived from diet. Breath was collected using custom-made, single-use Nalophan™ (PET-polyethylene terephthalate) bags with a 2000 mL capacity (Ansheng (HK) Medical Trading Limited). Nalophan was an inert material that provided compound stability for up to 12 hours. These bags had a mouthpiece port and two ports for TD tubes (Figure 22), both made from inert polypropylene.





**Figure 22. Breath collection bag**

Two TD tubes were connected to the ports on the breath collection bag via the grooved end of the tubes. These tubes were then connected to a custom-made Gas Sampling Pump (GSPs) (Ansheng (HK) Medical Trading Limited) connected to a mains power adaptor (Figure 23). GSPs were calibrated every month using a manual flow meter to ensure that flow rates remained within  $\pm 10\%$  of the programmed flow rate (200ml/min).



**Figure 23. Correct setup of the breath collection bag, TD tubes and gas sampling pump**

After switching the device on, the flow rate was set to 200ml/min, and the sampling time was set to 240 seconds, which enabled the collection of 800 mL breath per tube. Participants were asked to take a deep breath in, remove the cap from the mouthpiece, and breathe out into the bag until they felt a slight resistance indicating the bag was full. The cap was replaced onto the mouthpiece, and the device began to sample 800 mL of breath per tube at the indicated flow rate. Once the sampling process had been completed, the TD tubes were removed and immediately re-capped with long-term storage caps.

This process was repeated a second time with the patient to obtain a further 2 sample tubes. After the collection of 4 sample tubes, 2 final room air (blank) samples were collected by attaching 2 clean TD tubes to the GSP without attaching a breath collection bag. With each sampling episode, the following information was collected using a source data worksheet: date of collection, patient study ID, TD tube numbers, sample type (breath sample or room air), and file numbers. All 6 TD tubes from the sampling episode were placed in a dedicated tube box (1 box per patient), which was labelled with the site code, patient study ID, and date of collection before being returned to the VOC laboratory. For external sites (SOLAR trial – Chapter 5), samples were returned to the VOC laboratory by courier on either the same day or the next day. Tube boxes were kept in a 4°C refrigerator if stored overnight.

### **3.3.11 Sample preparation for storage**

Upon arrival at the laboratory, TD tubes were logged using the LIMS system. Prior to storage, TD tubes had to be 'dry purged' before being stored in the -80°C freezer. Since the relative humidity of breath was close to 100%, the hydrophilic nature of the graphitized carbon black secondary sorbent could lead to significant water retention. Aside from generating ice crystals, which could damage the sorbent and increase background VOCs, the presence of water in TD tubes could have several negative effects on downstream TD-GC-MS analysis. Water could expand rapidly on the trap during firing, for example, leading to an irreproducible split during injection into the GC. Condensed water generated by oven ramp temperatures below 100°C could also inhibit the interaction of analytes with the stationary phase and increase column bleed. Therefore, it has been shown to be crucial to dry purge tubes before storage to prevent degradation of breath samples and improve reproducibility<sup>207</sup>. TD tubes were dry purged using the TC-20 TD tube conditioning units. The flow rate was checked with a flow meter to ensure N<sub>2</sub> flow was at 100ml/min, and then the flow was switched off at the regulator. TD tubes were then inserted with the grooved ends into the manifold block (opposite to the position used for conditioning). The flow was then switched on for 3 minutes to remove any condensed water. The flow was switched off, and TD tubes were re-capped. Subsequently, TD tubes were stored in a dedicated tube storage -80°C freezer, with the location noted on the LIMS. The long-term storage caps on TD tubes were retightened 1 hour after placing them in the -80°C freezer to ensure there was no loosening that could result in sample loss or contamination. This freezer was continuously monitored using a T-scan system.

### **3.3.12 Preparation of calibration mixes**

Calibration curves were constructed by analysing standards of known concentrations of target analytes. This enabled absolute quantification of target compounds as well as served as a quality control tool, ensuring the accuracy and reproducibility of analysis. Calibration curves were generated at a minimum of once a week, whenever the calibration quality check failed the limits of a system suitability check or after maintenance. Mid-polar and polar calibration mixes were custom products manufactured by Sigma. The full list of compounds for each calibration mix is listed in Appendix 8 & 9. The compounds within these calibration curve mixes had been shown to be associated with OG cancer, from both published and unpublished data from the Hanna lab. Additionally, the compounds covered the majority of compound classes of biologically relevant breath VOCs<sup>103</sup>. Calibration mixes were supplied at a stock concentration of 100µg/mL. Clean TD tubes were loaded onto a PAL3 Series III Dual Head Sample Handling Robot. The sample handling robot prepared working concentrations of the calibration mix, spiked 1µl of each working calibration standard mix, and dry purged each tube. The sequence included a syringe check to ensure the syringe needle was patent. The final amounts for each calibration tube (Cal 1 to Cal 8) were 0.5, 1, 2, 4, 6, 8, 10, and 12ng. An additional two clean tubes were spiked with Cal 5 (6ng) and were used as calibration checks at the beginning and end of a batch. Once complete, the robot caps were replaced with inert diffusion-locking caps, and the tubes were ready for analysis.

### **3.3.13 Internal standard addition**

Internal standards of known quantities were added to all TD tubes (samples and standards) before analysis to correct for variations in instrument response. Neat internal standards (Table 8) were purchased from Sigma. Stock internal standard mix (1000mg/L) and working internal standard mix (5mg/L) were prepared in GC-MS grade methanol. Surrogate was stored at -10°C and had an expiry of 1 year. Sample TD tubes were retrieved from the -80°C freezer and allowed to equilibrate at room temperature for at least 30 minutes. After storage caps were replaced with robot caps, TD tubes were loaded onto a PAL3 Series III Dual Head Sample Handling Robot. The sample handling robot spiked 1µl of working internal standard (5ng) and dry purged each tube. The sequence included a syringe check to ensure the syringe needle was patent. Once complete, the robot caps were replaced with inert diffusion-locking caps, and the tubes were ready for analysis.

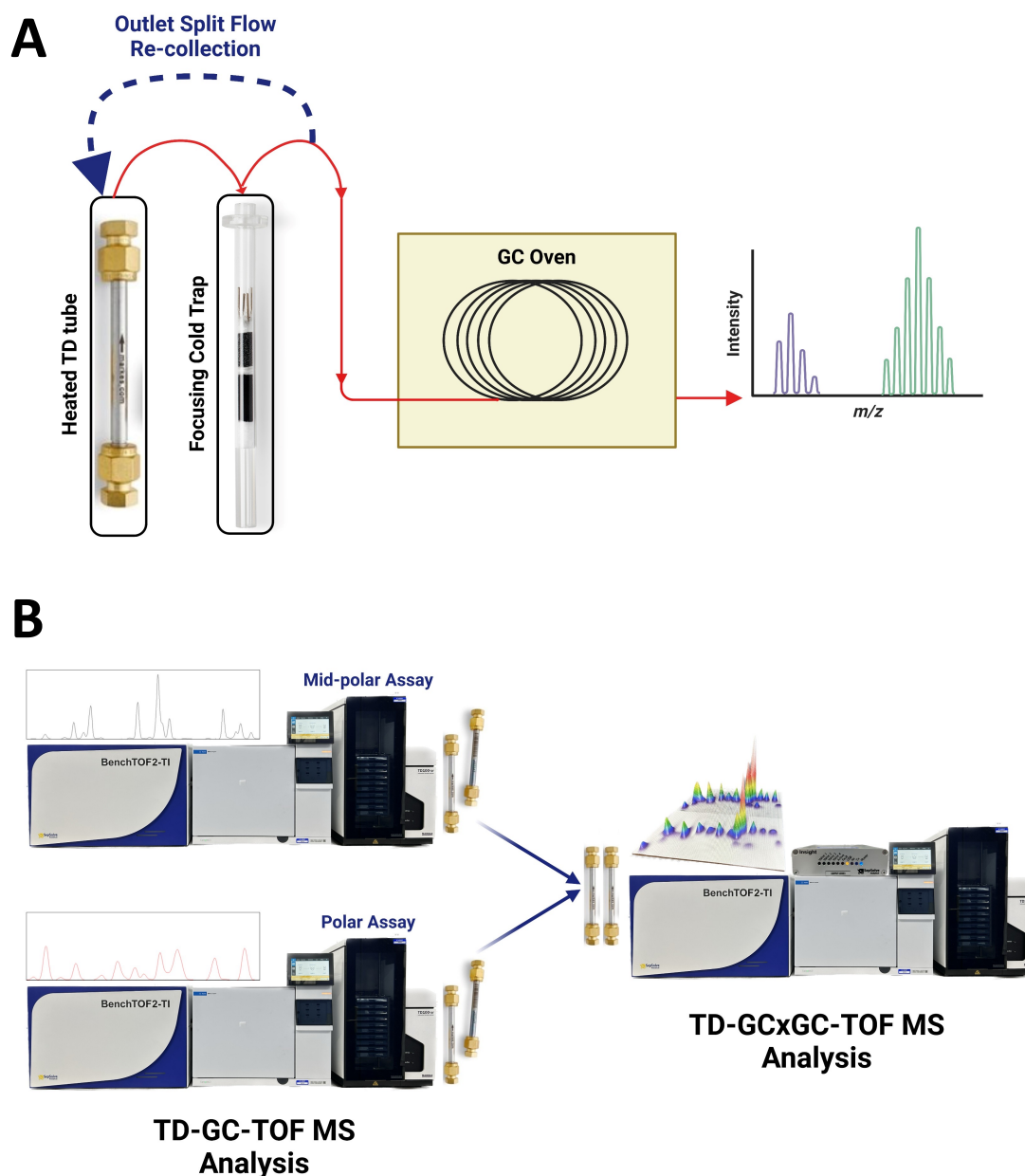
| Compound        | Part number | CAS No.    |
|-----------------|-------------|------------|
| Propanol-d8     | 175897-5G   | 22739-76-0 |
| Acetophenone-d8 | 296732-1G   | 19547-00-3 |
| Octane-d18      | 151971-1G   | 17252-77-6 |
| Toluene d8      | 434388-5G   | 2037-26-5  |
| Benzaldehyde-d6 | 485551-5G   | 17901-93-8 |

**Table 8. Stable isotope-labelled internal standards**

### 3.3.14 TD-GC-MS methods

**MS instrumentation.** First, a system suitability check was performed. This included an air/water check to ensure there were no leaks in the system, and a monthly auto-optimisation to adjust the filament voltage. TD tubes were then loaded onto a TD100-xr thermal desorption unit and sequentially pre-purged with helium (50 mL/min) for 1 minute then heated to 280°C for 8 minutes. VOCs were desorbed onto a focusing trap (U-T12ME-2S, Markes International) with a 5.3:1 outlet split for sample recollection (6 mL/min split flow, 1.4 mL/min column flow), which was then heated to 300°C for 3 minutes to release VOCs into the GC column. The GC-MS system used was an 8890 (Agilent Technologies) coupled to a Bench-TOF2 (Markes International). TOF measurement was continuous with mass range set between 35-500 m/z and ionization energy -70 eV. All data were acquired and pre-processed with ChromSpace (version 2.1.6, Markes International).

**Sequential re-collection workflow.** For validation experiments, breath was collected from healthy volunteers after informed consent and accessed under Imperial College Healthcare Tissue Bank Review Board Approval R21028 (REC approval: 17/WA/0161). From four breath tubes collected per patient, two tubes were analysed using a TD-GC-ToF-MS mid-polar method and two tubes were analysed using a TD-GC-ToF-MS polar method. After sequential 2-1 re-collection on each instrument, the final two tubes were analysed using TD-GCxGC-ToF-MS (Figure 24).



**Figure 24. Sequential re-collection workflow**  
**A, Focusing trap sample re-collection. B, 1D to 2D re-collection workflow**

**TD-GC-ToF-MS methods.** For the mid-polar method, the focusing trap was set at 10°C and a fused silica capillary column (Rxi-624Sil, 30 m x 0.250 mm, 0.14 µm, Restek) was used with helium carrier gas. The GC oven was initially set at 40°C for 1 min, then increased to 280°C (10°C/min with 10 min hold). For the polar assay, the focusing trap was set at 10°C and a Stabilwax-DA, 30 m x 0.25 mm x 0.25 µm (Restek Corporation, Bellefonte, USA) was used with helium carrier gas. The GC oven was set at 40°C for 1 min, then increased to 250°C (10°C/min with 10 min hold). TOF data acquisition rate was set at 6 Hz.

**TD-GCxGC-ToF-MS method.** TD tubes were desorbed to 280°C for 8 minutes, with focusing trap temperature set at 30°C. An INSIGHT flow modulator (2nd Gen, SepSolve Analytical) was used for parallel split with valve period of 4s with 100ms flush. VOCs were separated using a BD-Heavy wax column in the first dimension (20 m x 0.18 mm x 0.18 µm), and a VF-200ms column in the second dimension (5 m x 0.25mm x 0.10 µm). The GC oven was initially programmed at 50°C for 3 min, then increased to 260°C (5°C/min with 5 min hold). TOF data acquisition rate was set at 50 Hz.

### 3.3.15 GC-MS data analysis

**GC-MS data pre-processing.** All LSC data files underwent dynamic baseline compensation (DBC) with peak width at 6s for 1D methods and 0.2s for the 2D method followed by peak alignment using ChromSpace software.

**GC-MS data processing – targeted workflow.** The targeted analysis has previously undergone method validation to meet the standards of ISO 17025 / UKAS accreditation. For every batch, the calibration check file was used to manually set retention times and integration windows ( $\pm 0.03$ s) for all compounds within the calibration standard check before running an analysis sequence on ChromSpace (Figure 25). The 'Quant Results' module within the ChromSpace software was used to evaluate the calibration curve and samples in each batch (Figure 26). Compound peaks in all standard and sample files were manually checked to ensure that peaks had been integrated correctly. Calibration curves were checked to ensure the linearity ( $R^2$ ) was  $> 0.99$  for each compound. The software automatically calculates the absolute amount of analyte per tube using the calibration curves. This data was then exported for statistical analysis.

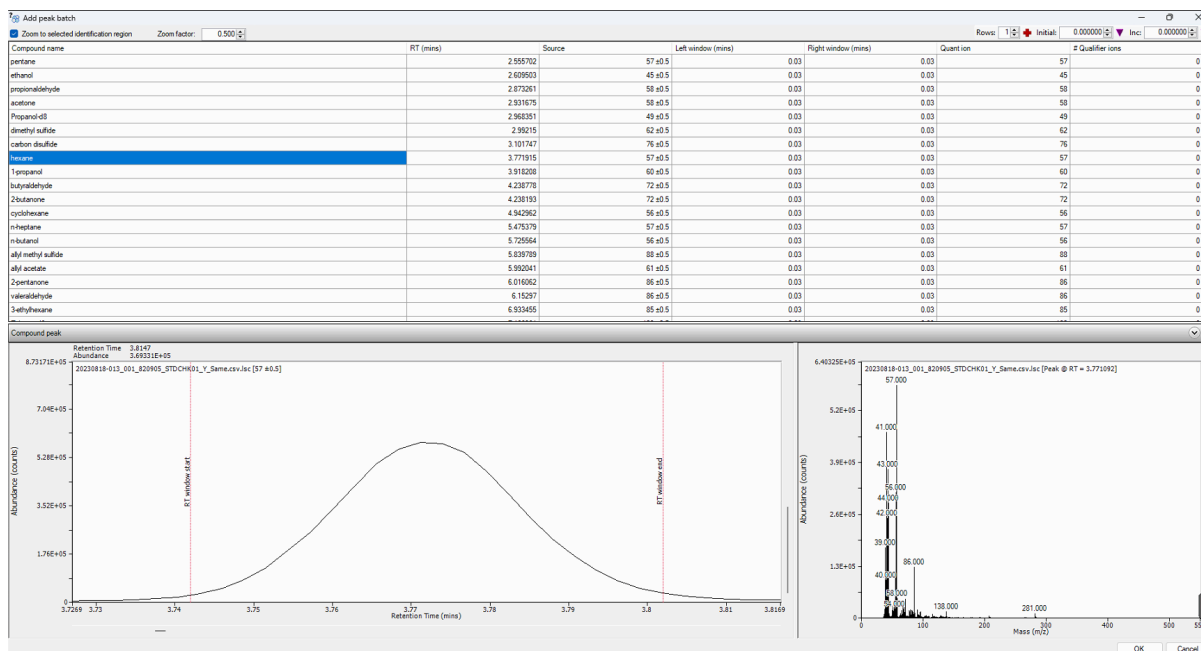


Figure 25. Manually setting retention times and integration windows for the targeted analysis

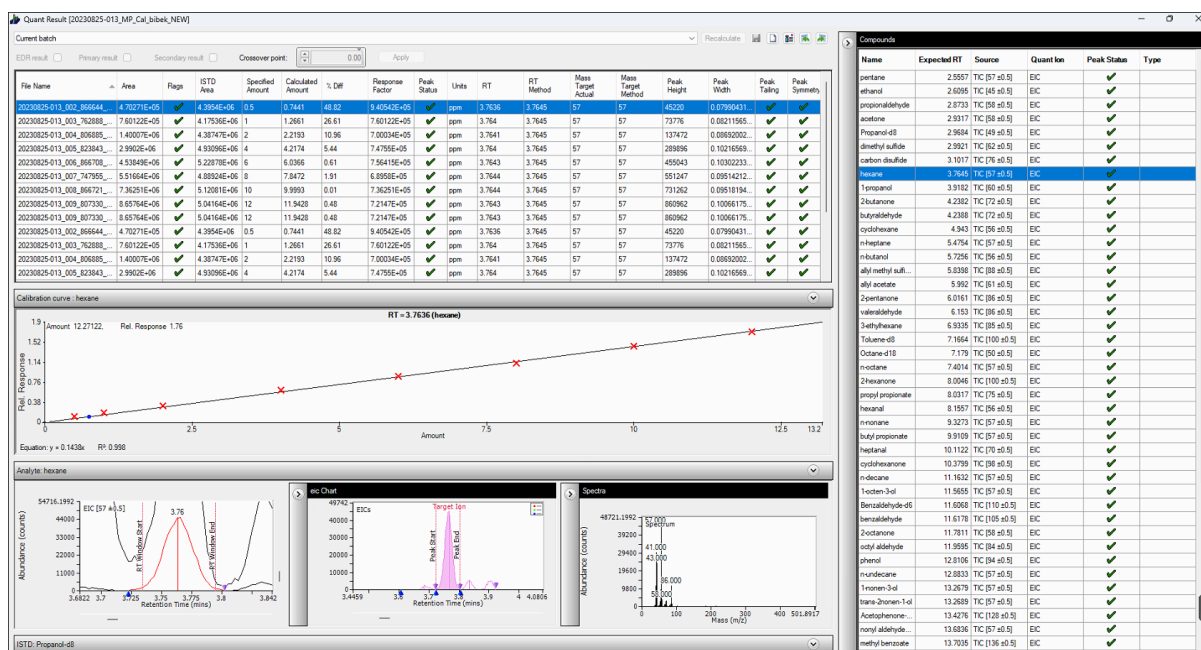


Figure 26. Quant Result browser for manual peak checking and quantification

**System suitability (post-processing).** Following the targeted analysis, an additional system suitability check was performed to ensure the integrity of the run, with any failures reported to the QC manager. The calibration check standards at the start and end of the batch were reviewed, ensuring relative response variance <30%. The expected recovery of surrogate within each tube was 70-130%. Blank tubes were also run with every batch and checked to monitor for background contamination.

**GC-MS & GCxGC-MS data processing – untargeted workflow.** For untargeted feature discovery, a tile-based analysis was used<sup>208–210</sup>. The traditional peak table and deconvolution approach was found to be highly sensitive to chromatographic misalignment even after software-based peak alignment, and in early attempts at untargeted analysis, led to several errors with integration. This is a known limitation of automated deconvolution when samples are complex and contain many co-eluting analytes<sup>211</sup>. The tile-based algorithm bins regions of the chromatogram into tiles rather than considering individual peaks, with the user inputting a custom tile width and overlap. Each tile-based feature is reported as {retention time}\_{ion}. The tile dataset is then subjected to supervised or unsupervised dimension reduction techniques (e.g. PCA, PLS-DA) and feature discovery. LSC files underwent integration with the 'Tile Sum' algorithm in ChromSpace. RT window width was set to 6 seconds and RT overlap to 25%. All peak filters were initially set to '0'. After running the data processing sequence, the completed data was initially opened in the 'ChromCompare' module of the ChromSpace software. The initial filtering used was a peak area of >100,000 and label frequency of >80%. All features were normalised to Toluene-d8 to correct for instrument response. The data was then exported for univariate and multivariate analysis. Tiles of interest were manually reviewed on chromatograms to identify their corresponding peaks. This peak list was then used for targeted integration.

### 3.3.16 Statistical analysis

Targeted and untargeted feature data from samples that passed QC for both RNA and GC-MS analysis was used for feature selection.

Multivariate data analysis was conducted using SIMCA<sup>®</sup> v17 (Sartorius). Prior to principal component analysis (PCA) and orthogonal partial least squares discriminate analysis (OPLS-DA), the dataset underwent log transformation (base 10) and UV scaling. PCA, an unsupervised technique, was utilized to observe inherent clustering patterns and identify potential outliers within the dataset. In contrast, OPLS-DA, a supervised method, involved modelling the data to distinguish hypoxia-high versus hypoxia-low samples.

For OPLS-DA, the assessment of model quality involved internal cross-validation ( $Q^2Y\text{-hat}$  value) and permutation testing. Permutation testing compared the true  $Q^2Y\text{-hat}$  value with 999 models generated by randomly permuting class memberships. Valid and robust models were characterized by a positive  $Q^2Y\text{-hat}$  value and a permutation p-value below 0.05. The Variable Importance in



Projection (VIP) score was utilized to identify features with the greatest potential for discrimination. Features with a VIP score exceeding 1.35 were deemed significant for differentiating between groups. To confirm the reliability of the OPLS-DA model, permutation testing and Cross-Validation Analysis of Variance (CV-ANOVA) were applied. Since individual classifiers may generate spurious results, an ensemble approach was used next, combining important features common to different classifiers to identify a core set of features most important for class differentiation.

Data was uploaded to Metabonalalyst v5.0 web portal ( <https://www.metaboanalyst.ca/> ), log transformed (base 10) and underwent auto-scaling. For multivariate analysis, the 'Biomarker analysis' module was used, with hypoxia status (hypoxia-high and hypoxia-low) as the primary phenotype label. Using the Multivariate Exploratory ROC Analysis tool, three classification and feature rank methods were selected to identify robust class-differentiating features and a compact signature that excluded highly correlating variables: Support Vector Machine (SVM), Partial least squares-discriminant analysis (PLS-DA) and Random Forest (RF). ROC curves were generated by Monte-Carlo cross validation (MCCV) using balanced sub-sampling and significant features were ranked by average importance. Features which were common to all three models were selected for final evaluation. The final model was then evaluated using the ROC curve-based model evaluation. The Area Under the Curve represents the discriminatory power of the model, with 1 indicating perfect classification.

Continuous data has been presented as mean (standard deviation, SD) or median (interquartile range, IQR). Categorical data has been presented as frequency (percentage). Parametric data were compared using Student's T test or ANOVA, non-parametric data were compared using Wilcoxon or Kruskal-Wallis rank sum test, and categorical data were compared using Pearson's Chi-squared test or Fisher exact test. All data analysis was conducted using R version 4.3.2 unless otherwise indicated. The main packages used were: *dplyr* (data manipulation); *ggplot2* (data visualization). Data visualization was also performed using GraphPad Prism version 9.

## 3.4 Results

### 3.4.1 Sample QC and cohort selection

A total of 51 patients were recruited for this pilot study. All patients underwent breath analysis and tissue sampling according to standardized methods outlined previously. QC procedures for GC-MS and RNA sequencing were performed first to select only high quality samples for downstream multivariate analysis.

First, GC-MS system and sample suitability checks were performed by quantifying surrogate and internal standard (ISTD) abundance per tube. Batches were checked to ensure surrogate/ISTD in samples and standards had an RSD consistently < 30%. Batches were excluded if RSD was above 30% in all surrogate/ISTD which could indicate instrument malfunction. Figure 27 and Figure 28 show within-sample variance for mid-polar and polar batches respectively. Individual internal standards such as benzaldehyde-d6 and propanol-d8 were noted to have a higher variance than other internal standards. No outlier batches were identified where all internal standards and surrogate had an RSD > 30% indicating possible instrument malfunction, therefore no batches were excluded.

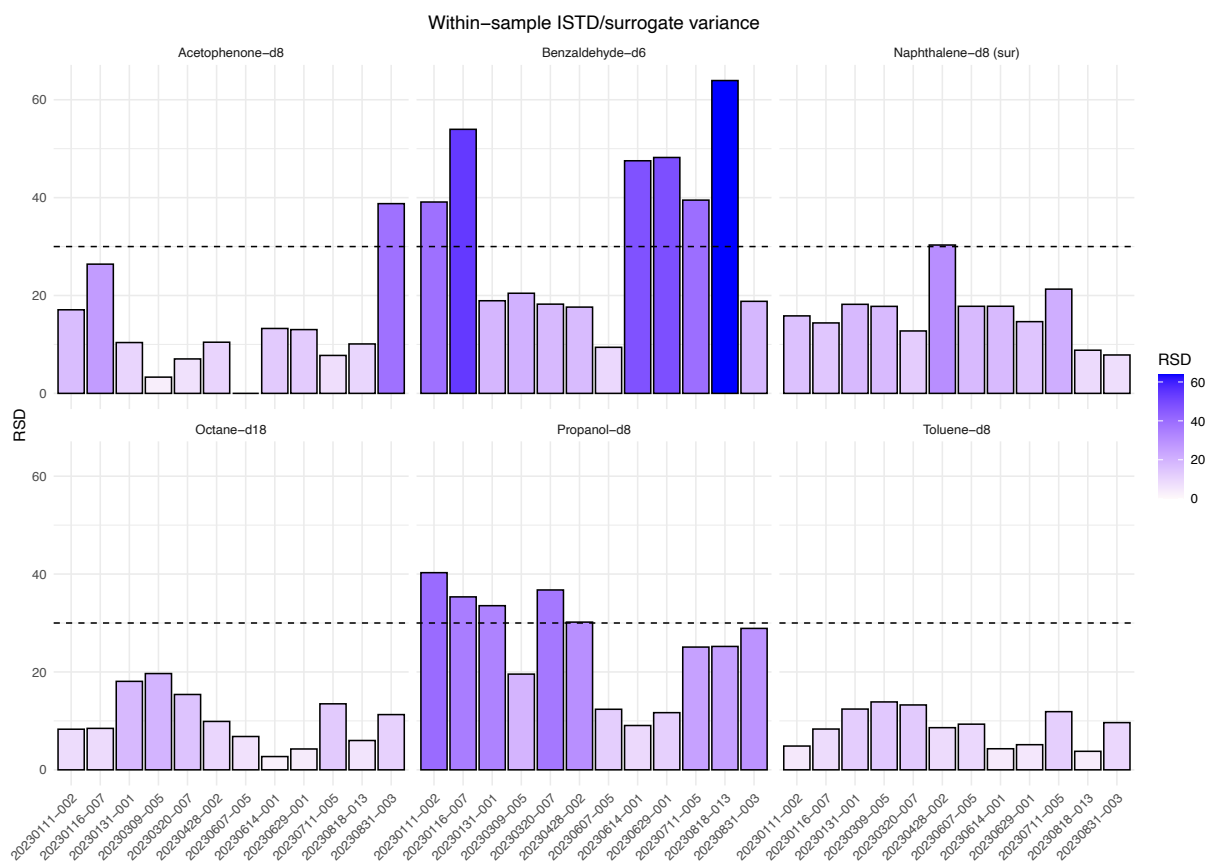


Figure 27. Mid-polar inter-sample ISTD/surrogate variance

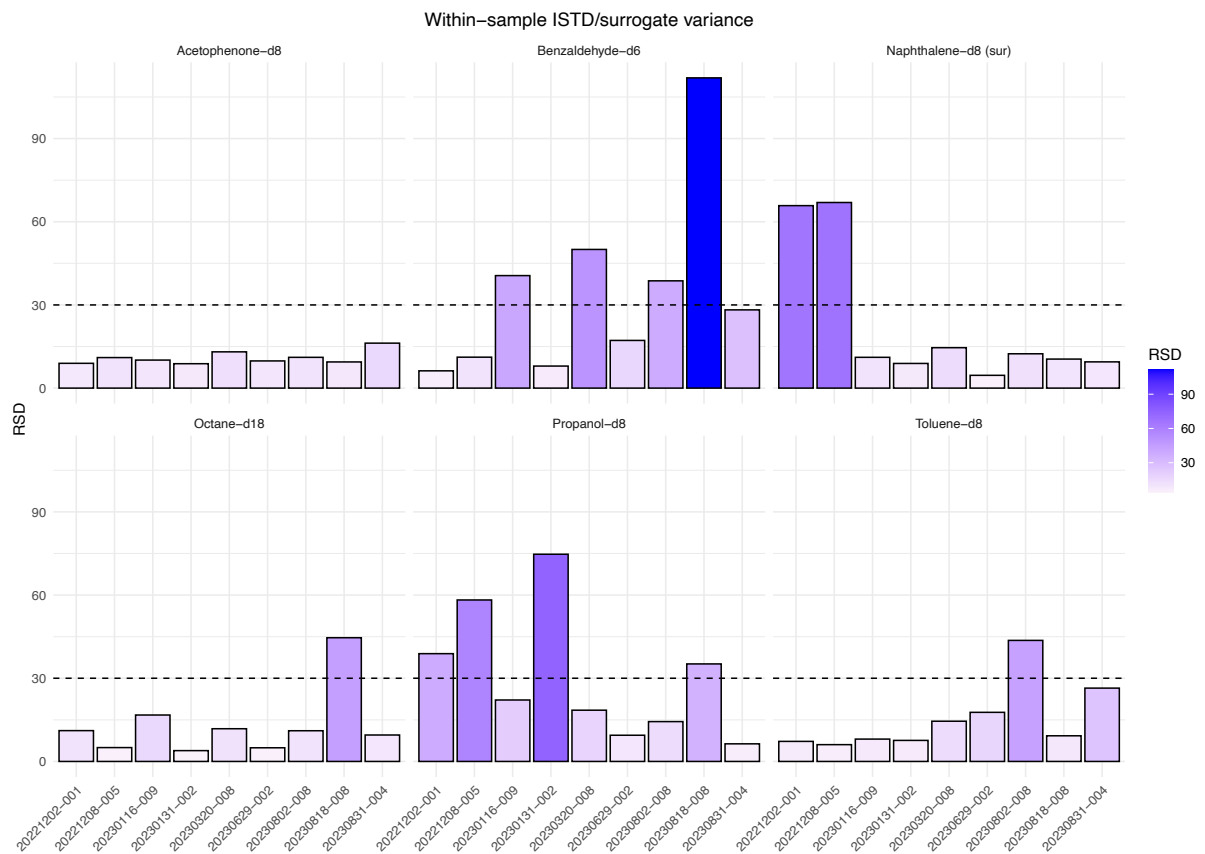
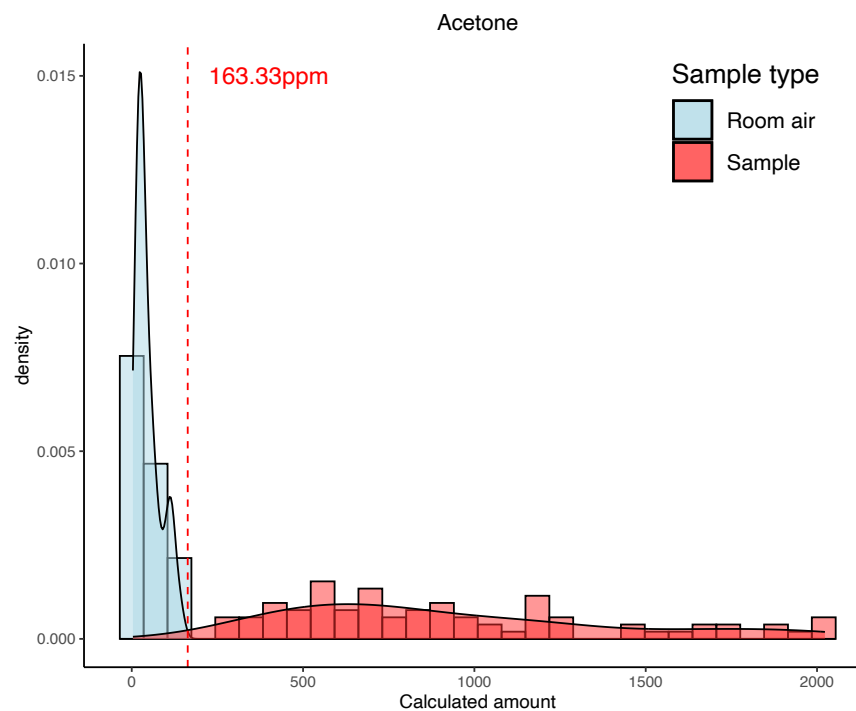


Figure 28. Polar inter-sample ISTD/surrogate variance

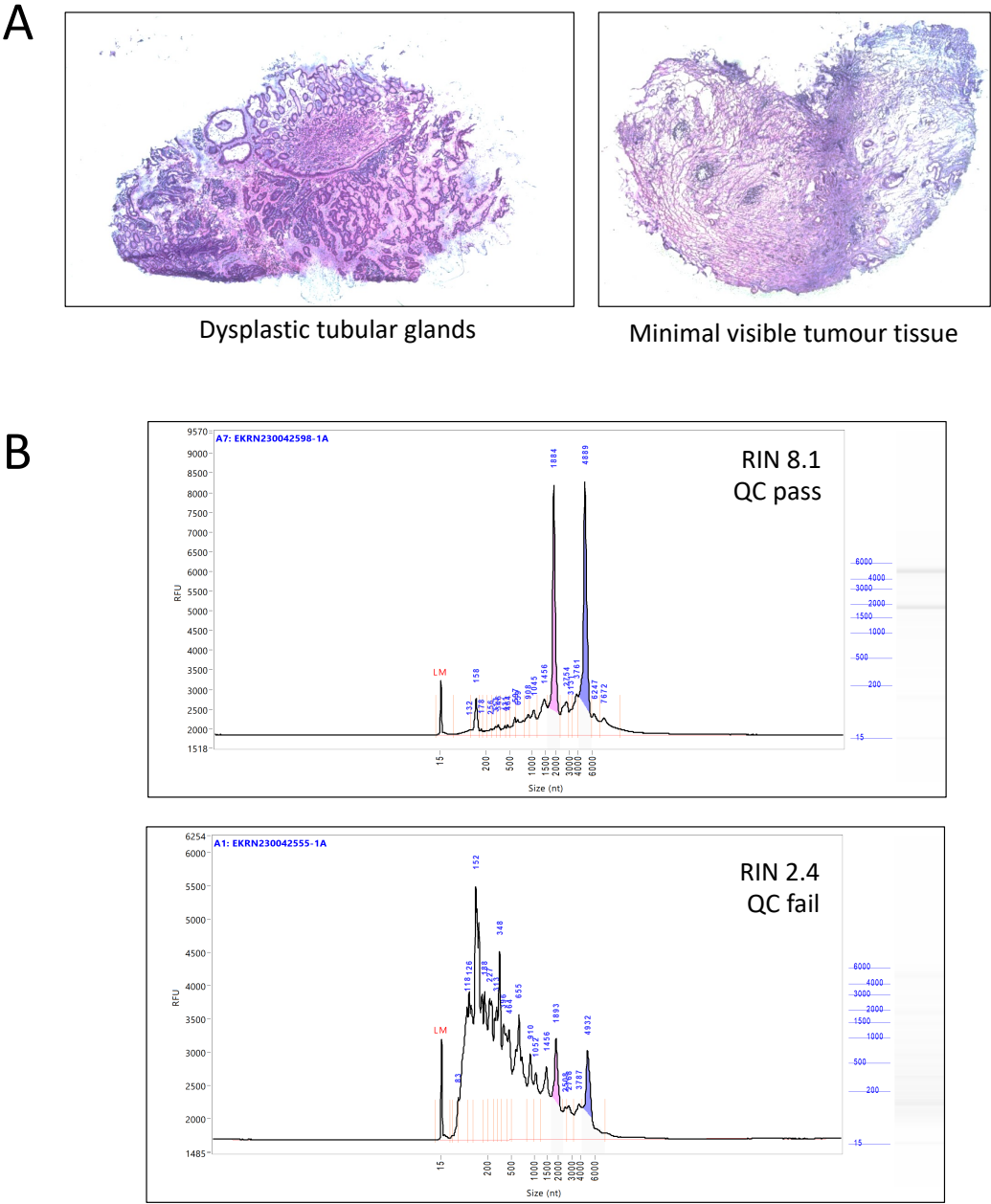
Individual samples were excluded if surrogate/ISTD had a peak area  $< 1 \times 10^6$  on a tube. Samples from one patient were excluded as surrogate was not detected in any of the sample tubes on both mid-polar and polar analyses which may have indicated a blocked syringe during tube spiking. Acetone has traditionally been used as a marker to indicate an adequate breath sample<sup>212</sup>. Although thresholds have not been defined in the current breath workflow, the density plot in Figure 29 indicated a distinct separation between included breath samples and room air samples analysed with the mid-polar method, with 99% of breath samples containing  $> 163$ ppm acetone.



**Figure 29. Density plot of acetone abundance on breath sample and room air TD tubes**

Figure 30 shows the quality control procedures used for the LCM RNA-seq workflow. First, H&E sections were checked to ensure that biopsies contained a sufficient amount of tumour tissue amenable to micro-dissection. This often revealed highly variable amounts of tumour tissue. If tumour biopsies contained only necrotic or normal tissue, they were excluded ( $n = 5$ ) (Figure 30A). RNA extraction following LCM generated yields ranging from 34.96 to 386.22 ng/ $\mu$ l RNA. Extracted RNA samples underwent integrity analysis using an Agilent Bioanalyzer. Electrochromatograms were inspected to identify two prominent peaks representing the 18S and 28S rRNA subunits with a flat baseline. The software also provided an RNA integrity number (RNA) ranging from 1 (degraded RNA) to 10 (fully intact RNA). Figure 30B shows an example of a QC pass (RIN 8.1) and fail (RNA 2.4) with

the QC fail demonstrating a markedly noisy baseline indicating degraded RNA. A total of 15 samples were excluded on the basis of RNA QC failure, leading to a final cohort of 30 patient samples which underwent RNA sequencing (Figure 31).



**A**, Representative H&E sections showing sections with high and low cellularity. **B**, Bioanalyzer electrochromatograms from LCM-extracted RNA samples showing both a QC pass and fail.

**Figure 30. Quality control for RNA extraction and sequencing**

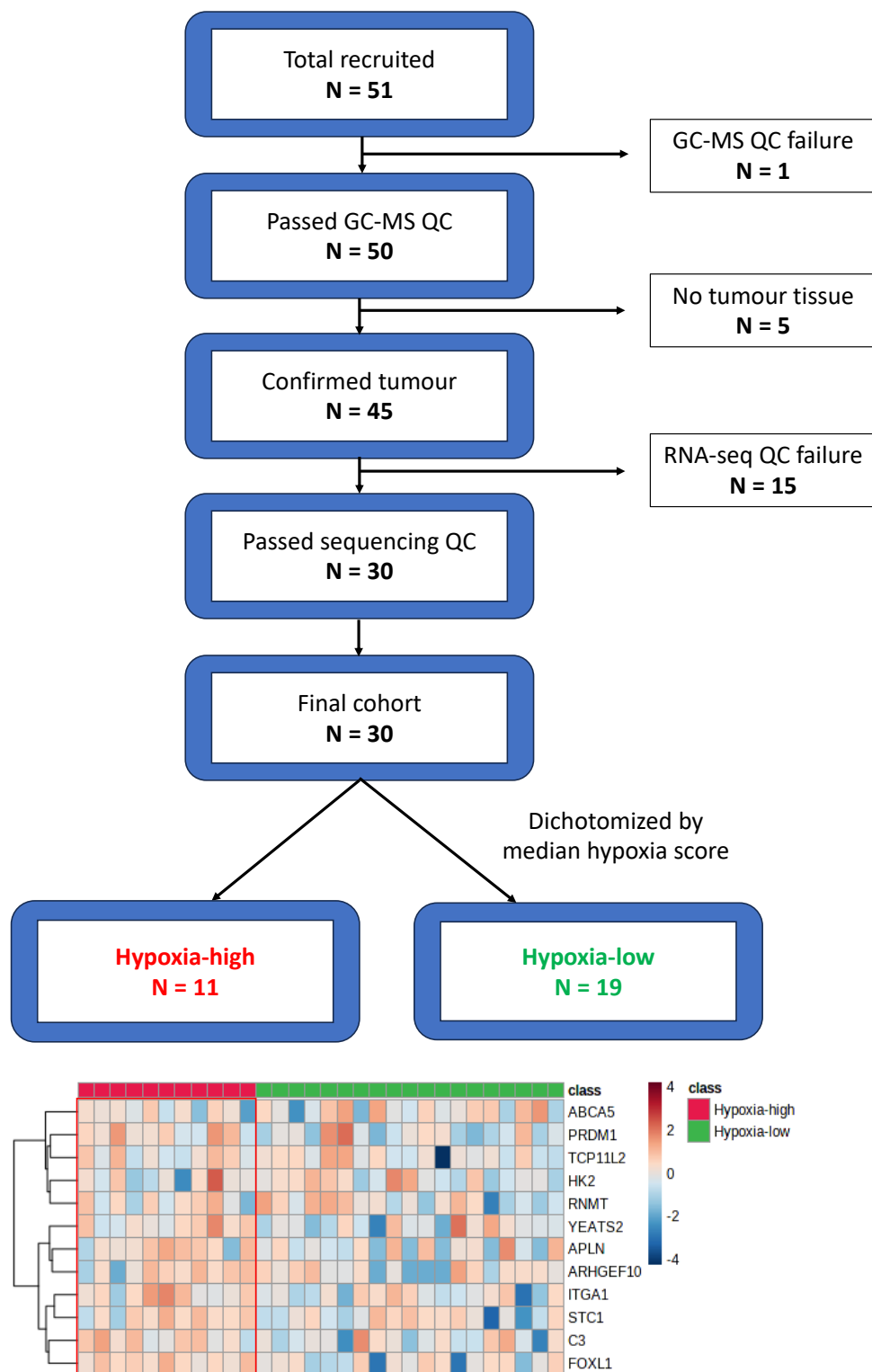


Figure 31. Summary of patient selection for the pilot clinical phenotyping study

Hypoxia scores were calculated using the OAC-Hyp gene signature and the cohort was dichotomized by median hypoxia score. This resulted in two groups for comparison, hypoxia-high (n = 11) and hypoxia-low (n = 19). Baseline characteristics of included patients are shown in Table 9. There were no significant differences in age, gender, smoking status, oncological and pathological outcomes between hypoxia-high and hypoxia-low groups. There was also an equal proportion of patients sampled before and after neoadjuvant chemotherapy.

| Hypoxia status               | Overall, N = 30 <sup>1</sup> | High, N = 11 <sup>1</sup> | Low, N = 19 <sup>1</sup> | p-value <sup>2</sup> |
|------------------------------|------------------------------|---------------------------|--------------------------|----------------------|
| <b>Age</b>                   | 75 (66, 79)                  | 77 (68, 80)               | 73 (66, 77)              | 0.31                 |
| <b>Gender</b>                |                              |                           |                          | >0.99                |
| Female                       | 4 (13%)                      | 1 (9.1%)                  | 3 (16%)                  |                      |
| Male                         | 26 (87%)                     | 10 (91%)                  | 16 (84%)                 |                      |
| <b>Current smoker</b>        | 1 (3.3%)                     | 0 (0%)                    | 1 (5.3%)                 | >0.99                |
| <b>Planned operation</b>     |                              |                           |                          | 0.86                 |
| Extended total gastrectomy   | 3 (10%)                      | 1 (9.1%)                  | 2 (11%)                  |                      |
| Oesophagectomy               | 17 (57%)                     | 7 (64%)                   | 10 (53%)                 |                      |
| Staging laparoscopy          | 10 (33%)                     | 3 (27%)                   | 7 (37%)                  |                      |
| <b>Sampling time</b>         |                              |                           |                          | 0.92                 |
| Post-chemo                   | 16 (53%)                     | 6 (55%)                   | 10 (53%)                 |                      |
| Pre-chemo                    | 14 (47%)                     | 5 (45%)                   | 9 (47%)                  |                      |
| <b>Type</b>                  |                              |                           |                          | 0.33                 |
| Distal oesophageal           | 20 (67%)                     | 6 (55%)                   | 14 (74%)                 |                      |
| GOJ (Type I)                 | 1 (3.3%)                     | 1 (9.1%)                  | 0 (0%)                   |                      |
| GOJ (Type II)                | 3 (10%)                      | 2 (18%)                   | 1 (5.3%)                 |                      |
| GOJ (Type III)               | 6 (20%)                      | 2 (18%)                   | 4 (21%)                  |                      |
| <b>AJCC Stage</b>            |                              |                           |                          | 0.51                 |
| I                            | 3 (10%)                      | 0 (0%)                    | 3 (16%)                  |                      |
| IIA                          | 2 (6.7%)                     | 1 (9.1%)                  | 1 (5.3%)                 |                      |
| IIB                          | 1 (3.3%)                     | 0 (0%)                    | 1 (5.3%)                 |                      |
| III                          | 23 (77%)                     | 9 (82%)                   | 14 (74%)                 |                      |
| IVA                          | 1 (3.3%)                     | 1 (9.1%)                  | 0 (0%)                   |                      |
| <b>Neoadjuvant treatment</b> |                              |                           |                          | 0.42                 |
| None                         | 8 (27%)                      | 4 (36%)                   | 4 (21%)                  |                      |
| Yes                          | 22 (73%)                     | 7 (64%)                   | 15 (79%)                 |                      |
| <b>R status</b>              |                              |                           |                          | 0.61                 |
| 0                            | 21 (70%)                     | 9 (82%)                   | 12 (63%)                 |                      |
| 1                            | 5 (17%)                      | 1 (9.1%)                  | 4 (21%)                  |                      |
| Unknown                      | 4 (13%)                      | 1 (9.1%)                  | 3 (16%)                  |                      |
| <b>Mandard</b>               |                              |                           |                          | 0.28                 |
| 1                            | 1 (3.3%)                     | 0 (0%)                    | 1 (5.3%)                 |                      |
| 2                            | 2 (6.7%)                     | 1 (9.1%)                  | 1 (5.3%)                 |                      |
| 3                            | 7 (23%)                      | 5 (45%)                   | 2 (11%)                  |                      |
| 4                            | 3 (10%)                      | 0 (0%)                    | 3 (16%)                  |                      |
| 5                            | 7 (23%)                      | 2 (18%)                   | 5 (26%)                  |                      |
| Unknown                      | 10 (33%)                     | 3 (27%)                   | 7 (37%)                  |                      |
| <b>Lymph node ratio</b>      | 0.01 (0.00, 0.05)            | 0.01 (0.00, 0.05)         | 0.01 (0.00, 0.06)        | >0.99                |
| <b>Differentiation</b>       |                              |                           |                          | 0.81                 |
| Moderate                     | 20 (67%)                     | 7 (64%)                   | 13 (68%)                 |                      |
| Poor                         | 9 (30%)                      | 4 (36%)                   | 5 (26%)                  |                      |
| Well                         | 1 (3.3%)                     | 0 (0%)                    | 1 (5.3%)                 |                      |
| <b>Perineural invasion</b>   |                              |                           |                          | >0.99                |
| No                           | 15 (50%)                     | 6 (55%)                   | 9 (47%)                  |                      |
| Unknown                      | 4 (13%)                      | 1 (9.1%)                  | 3 (16%)                  |                      |
| Yes                          | 11 (37%)                     | 4 (36%)                   | 7 (37%)                  |                      |
| <b>Vascular invasion</b>     |                              |                           |                          | 0.66                 |
| No                           | 8 (27%)                      | 4 (36%)                   | 4 (21%)                  |                      |
| Unknown                      | 4 (13%)                      | 1 (9.1%)                  | 3 (16%)                  |                      |
| Yes                          | 18 (60%)                     | 6 (55%)                   | 12 (63%)                 |                      |

<sup>1</sup>Median (IQR); n (%)

<sup>2</sup>Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test

**Table 9. Baseline characteristics of patients in the pilot phenotyping study**

### 3.4.4 Sequential re-collection pipeline method development

As highlighted, biomarker discovery in untargeted TD-GC-MS workflows is complicated by the challenge of compound identification, particularly due to co-elutions. Automated split flow re-collection offers the opportunity to performed additional analysis on the same breath sample with orthogonal techniques but can reduce analyte recovery as a portion of sample is lost with every re-collection cycle. Pooling replicate samples onto one tube (sequential re-collection) prior to downstream analysis could address this problem and obviate the need to collect additional sample tubes. As part of method development for this study, a 2-1 re-collection protocol was evaluated.

Split flow re-collection was first evaluated by spiking TD tubes with calibration mixes and quantifying analyte loss over consecutive desorptions and re-collection cycles of the same tube on the 1D GC-MS instruments (Figure 32). On both instruments, analytes showed stable decay curves with each re-collection cycle (all  $R^2 > 0.99$  on regression modelling). There was a 50% reduction in response ( $t_{1/2}$ ) on average over every 2.47 re-collection cycles ( $R^2 = 0.97$ ). This confirmed that single tube re-collection is reproducible but detrimental to the recovery of trace compounds. Next, the impact of sequential desorption and re-collection cycles on the 1D GC-TOF-MS instrument was evaluated by re-collecting flow from two tubes onto one tube with every cycle. For both assays, analytes showed stable exponential growth with sequential re-collection cycles (all  $R^2 > 0.99$ ) On average, there was a doubling in response ( $T_d$ ) every 1.45 re-collection cycles ( $R^2 = 0.98$ ). This equates to a 73.6% average increase in response by re-collecting two tubes onto one tube. Re-collection from the 1D to the 2D instruments was then evaluated by comparing the response on the 2D GC-MS after single tube (1-1) compared with sequential tube (2-1) re-collection. Sequential re-collection (2-1) generated an average 2.34-fold increase in response in all compound classes (Figure 33).



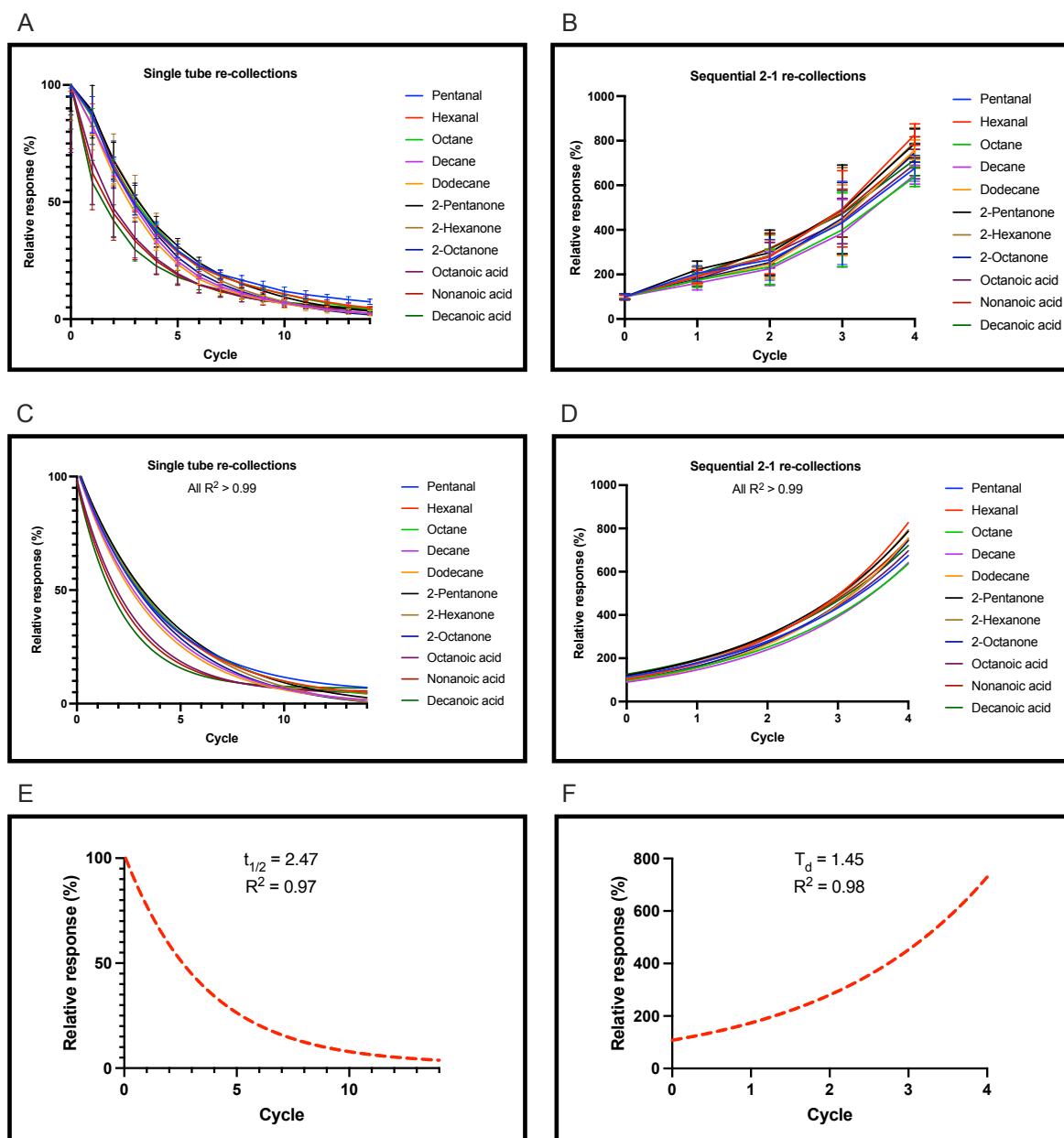


Figure 32. Relative responses after single tube or sequential re-collection cycles (raw, regression model and average)

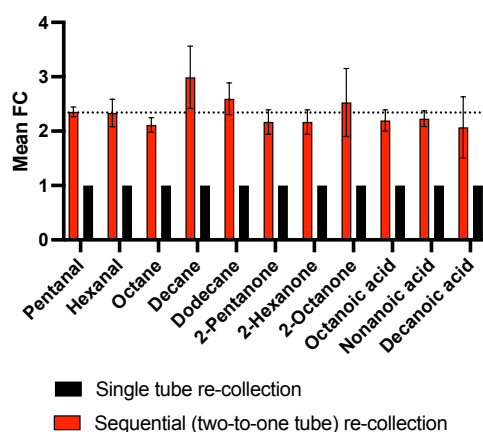


Figure 33. Relative responses of analytical standards on TD-GC-TOF-MS after re-collection on TD-GC-TOF-MS

Finally, breath was collected from healthy volunteers ( $n = 3$ ) and a selection of compounds present in all breath samples were selected to compare re-collection protocols (Figure 34). Sequential re-collection (2-1) generated an average 2.99-fold increase in response in all compound classes (Figure 2). Figure 35 is an example of a co-eluting breath VOC (decane and 2-butyl-1-octanol) separated on GCxGC-TOF-MS which could not be resolved by either mid-polar or polar 1D assays.

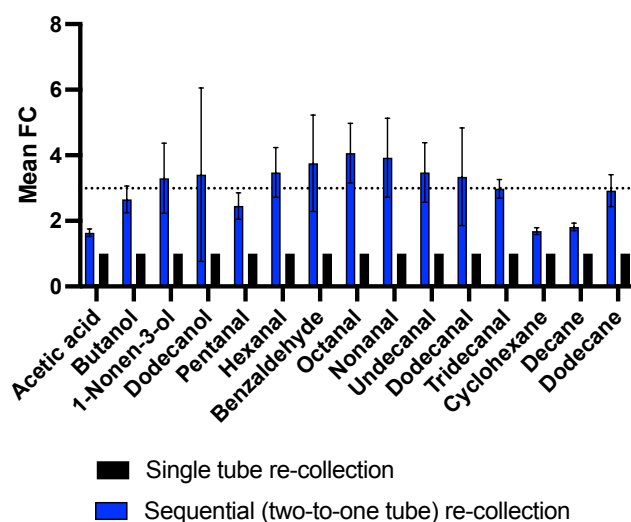


Figure 34. Relative responses of breath VOCs from healthy individuals on GCxGC-TOF-MS after re-collection on GC-MS

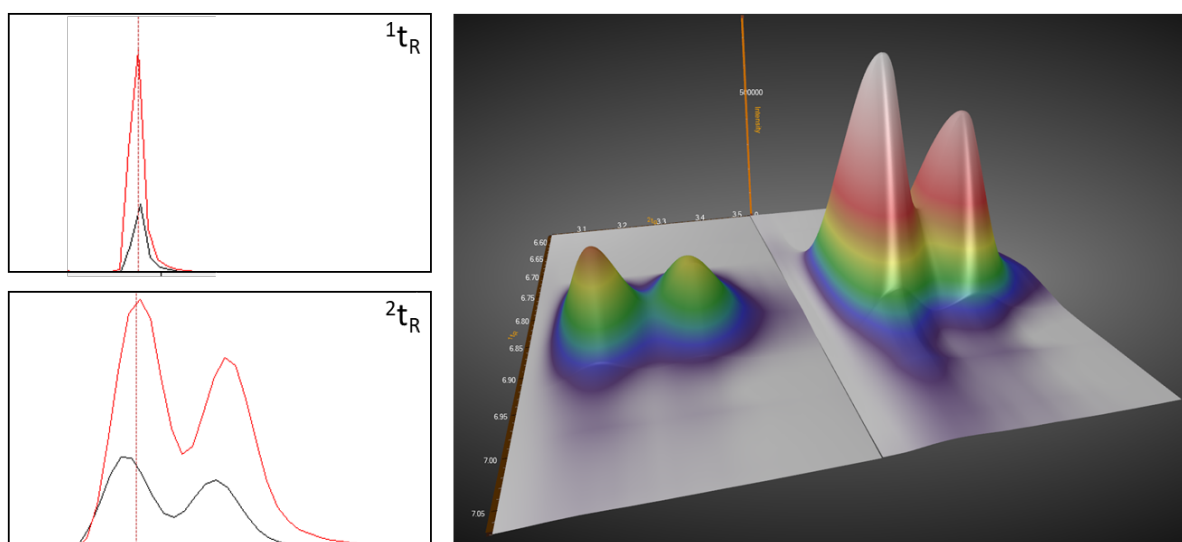


Figure 35. Resolution of co-eluting VOCs in breath after single (1-1) (black) and sequential (2-1) (red) re-collection

### 3.4.2 Targeted GC-MS analysis

Targeted GC-MS analysis enabled absolute quantification with unambiguous compound identification of compounds within a custom calibration mix. Compounds were selected on the basis of potential diagnostic utility from previous OG cancer discovery studies. Representative calibration curves for the 52 mid-polar compounds and 10 polar compounds are shown in Figure 36 and Figure 37. The majority of compounds showed strong linearity in relative response ( $R_2 > 0.9$ ) between 0.5 – 12 ppm. For the mid-polar method, cyclohexanone and valeraldehyde (pentanal) had consistently poor linearity and were excluded from further analysis. For the polar method, acetic acid also had consistently poor linearity and was excluded from further analysis.

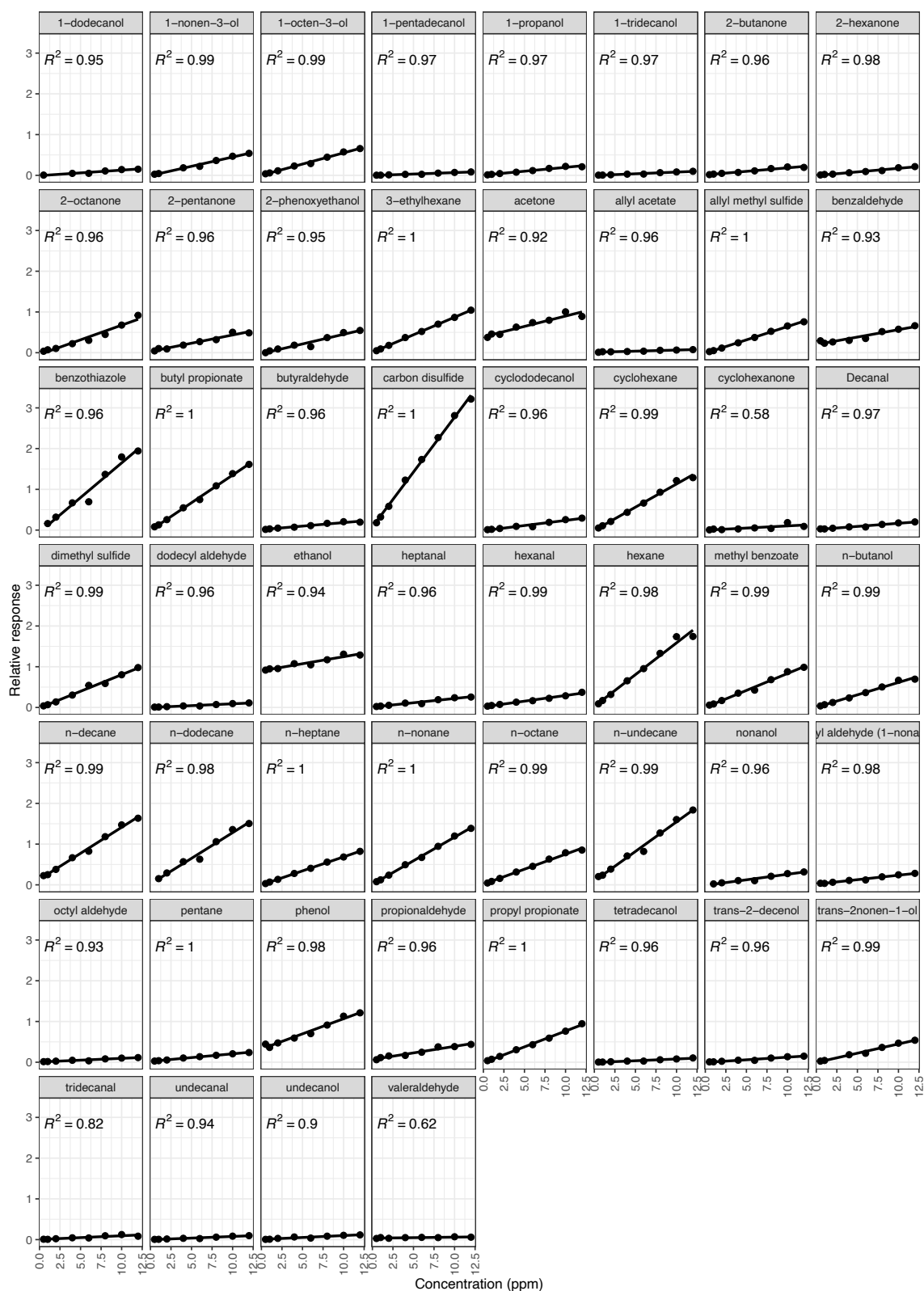
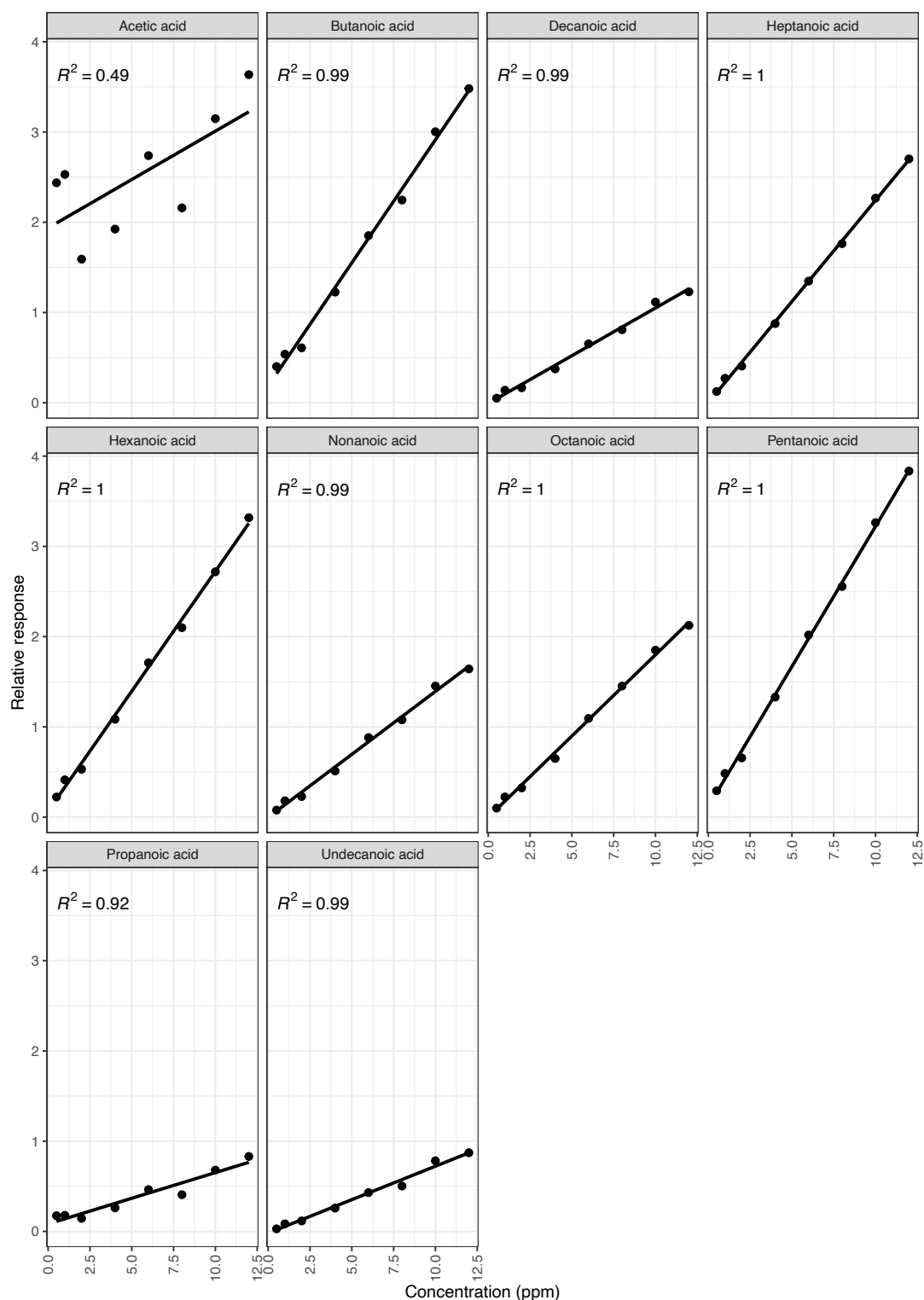


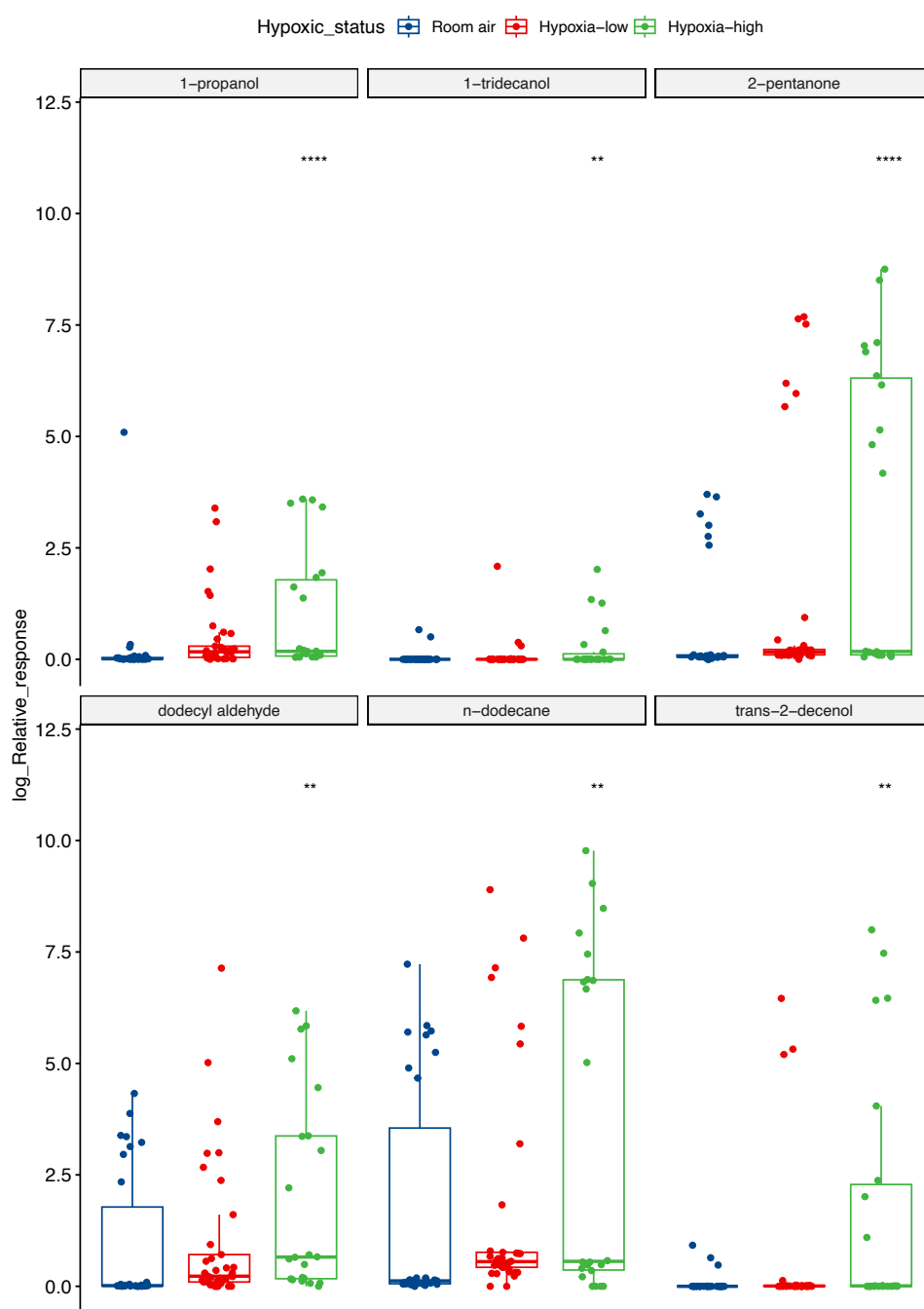
Figure 36. Calibration curves for the targeted mid-polar method



**Figure 37. Calibration curves for the targeted polar method**

For the targeted analysis, univariate multi-group analysis was performed comparing room air, hypoxia-low and hypoxia-high classes to identify VOCs which could significantly discriminate hypoxia-high classes. This revealed six VOCs (1-propanol, 1-tridecanol, 2-pentanone, dodecyl aldehyde, n-

dodecane and trans-2-decenol that were significantly enriched in the breath of hypoxia-high OAC patients (Figure 38). However, in a ROC curve analysis, these VOCs alone had relatively poor performance in discriminating hypoxic classes (AUROC 0.683; 95% CI 0.518 – 0.841). No polar compounds could significantly discriminate breath samples from room air samples or between hypoxic classes.



Comparisons performed using a Kruskal-Wallis test.

\*\*\* P < 0.0005, \*\* P < 0.005, \* P < 0.05

**Figure 38. Significantly different VOCs between hypoxic classes and room air with a targeted analysis**

### 3.4.3 Untargeted GC-MS analysis

A tile-based methodology was used for untargeted GC-MS analysis due to the significant challenges of automated software-based alignment and deconvolution algorithms highlighted previously. For the mid-polar tile dataset (Figure 39), unsupervised analysis with PCA showed clear separation between breath samples and room air blanks but no separation between hypoxic classes. Supervised analysis with OPLS-DA showed separation between hypoxia-high and hypoxia-low groups ( $R^2Y$  0.991 and  $Q^2$  0.563), although the lower  $Q^2$  relative to  $R^2Y$  suggested a degree of model overfitting. This was validated from permutation analysis (999 permutations) and CV-ANOVA p value with a p-value of  $1 \times 10^{-4}$ . VIP analysis identified 50 tile features with a  $VIP > 1.35$ . For the polar tile dataset (Figure 40), unsupervised analysis with PCA showed clear separation between breath samples and room air blanks but no separation between hypoxic classes. Supervised analysis with OPLS-DA did not show separation between hypoxia-high and hypoxia-low groups ( $R^2Y$  0.297 and  $Q^2$  0.0462) with no significant features on VIP analysis.

Since the polar method did not show separation on OPLS-DA, only the mid-polar data was taken forward for further multivariate analysis. Ensemble analyses were performed using SVM, PLS-DA and random forest algorithms to identify robust class-differentiating features. This identified 42 features which were common to all three models. These features were tentatively identified by manually interrogating 1D mid-polar chromatograms to exclude siloxane or sorbent contaminants, overlapping ions mapping to the same compound, and inorganic analytes, while the re-collected 2D chromatograms were used to screen for potential co-elutions (Figure 41). This was refined to a final eight VOC model which was re-run on the mid-polar dataset using a semi-targeted peak table method and evaluated using a ROC analysis (Figure 42). The final model showed good discriminatory performance (AUC 0.888 95% CI 0.77 – 0.975). The tentative identities of the eight VOCs are listed in Table 10.

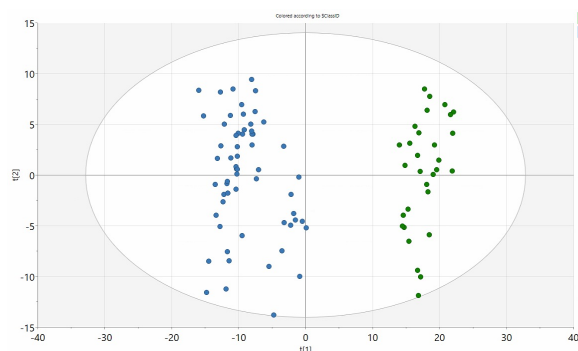
| Feature tile | Direction in Hypoxia-high | Direction in Hypoxia-low | Compound name         | CAS No.   | Class    |
|--------------|---------------------------|--------------------------|-----------------------|-----------|----------|
| 9.350_87     | ↓                         | ↑                        | Hexane, 2,4-dimethyl- | 589-43-5  | Alkane   |
| 9.875_84     | ↓                         | ↑                        | Octane, 2,7-dimethyl- | 1072-16-8 | Alkane   |
| 14.525_113   | ↓                         | ↑                        | Tridecane             | 629-50-5  | Alkane   |
| 11.900_98    | ↓                         | ↑                        | 1-Hexanol, 2-ethyl-   | 104-76-7  | Alcohol  |
| 8.825_43     | ↓                         | ↑                        | Octane, 3-methyl-     | 2216-33-3 | Alkane   |
| 13.175_91    | ↓                         | ↑                        | Undecane, 2-methyl-   | 7045-71-8 | Alkane   |
| 9.575_42     | ↑                         | ↓                        | 1-Hexanol             | 111-27-3  | Alcohol  |
| 10.925_130   | ↑                         | ↓                        | Butanal, 2-ethyl-     | 97-96-1   | Aldehyde |

Table 10. Candidate exhaled VOCs for OAC hypoxia classification

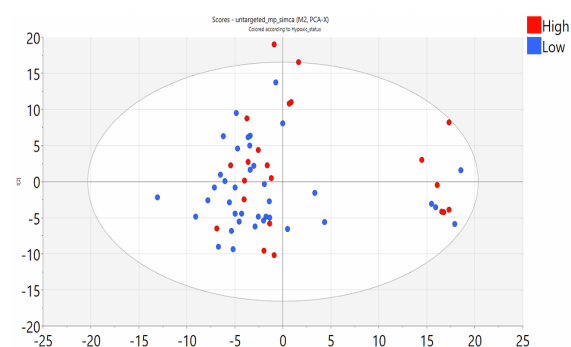
## MIDPOLAR UNTARGETED ANALYSIS

**A**

**PCA – Blank vs Sample**

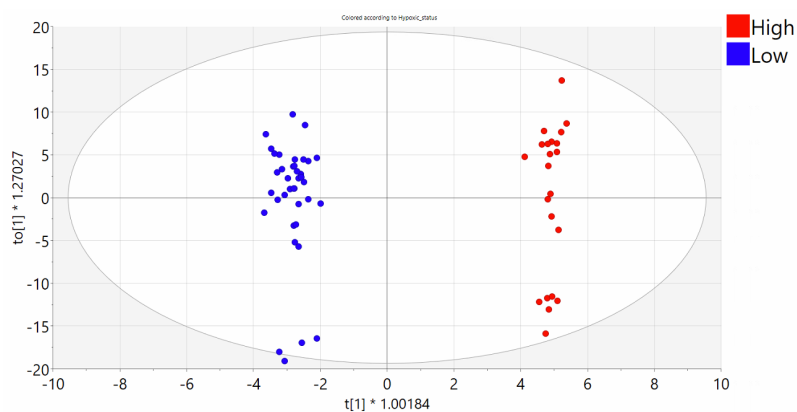


**PCA – Hypoxia-high vs Hypoxia-low**

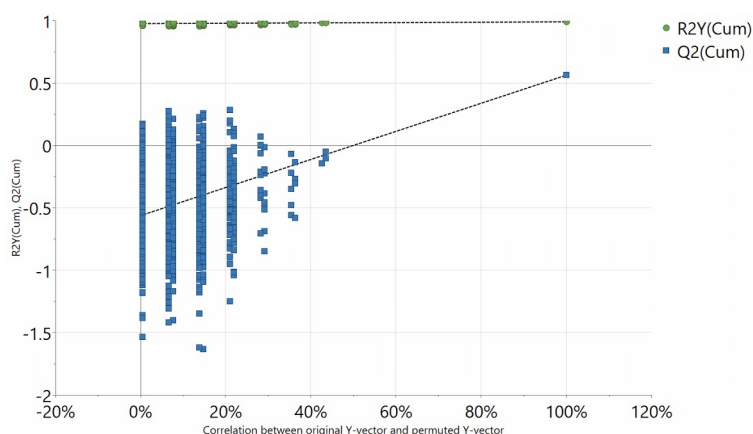


**B**

**OPLS-DA – Hypoxia-high vs Hypoxia-low**



**C**



**A**, PCA plots of the separation between samples and blanks, and hypoxic classes. **B**, Score plot of the OPLS-DA model generated on the untargeted GC-TOF-MS tile data set, indicating hypoxic class separation. **C**, Permutation testing – 999 repeats: The Y-axis represents the  $R^2Y$  and  $Q^2$  values of each model. The X-axis shows the correlation coefficient between the “real” Y and the permuted Y.  $R^2Y$ , 0.991;  $Q^2$ , 0.564.

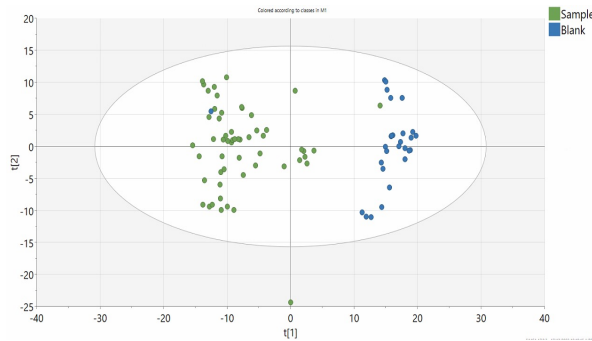
**Figure 39. TD-GC-TOF-MS (mid-polar) untargeted analysis**



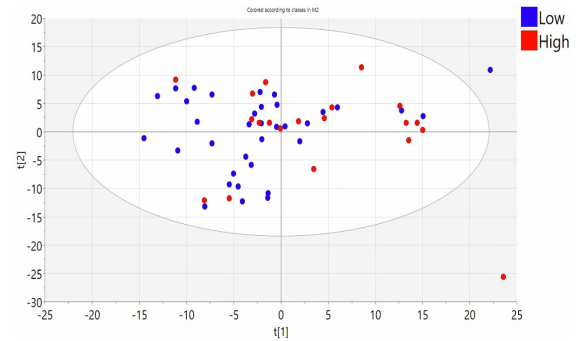
## POLAR UNTARGETED ANALYSIS

**A**

PCA – Blank vs Sample

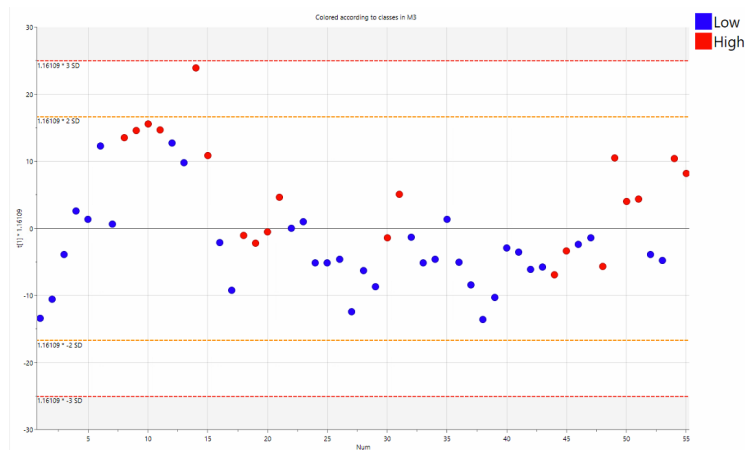


PCA – Hypoxia-high vs Hypoxia-low

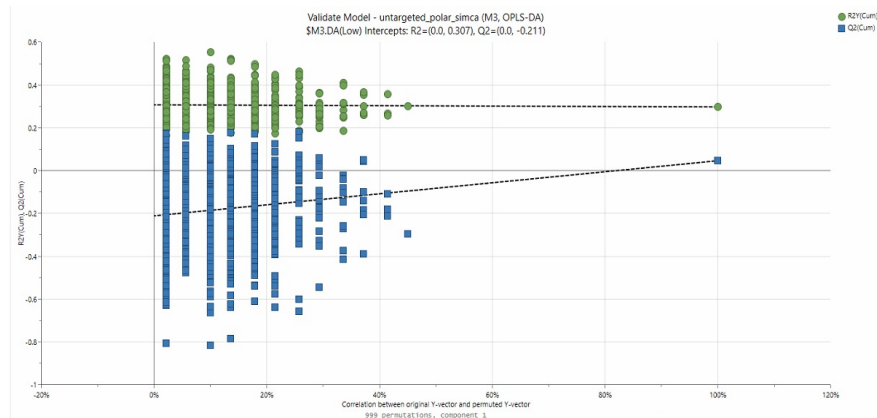


**B**

OPLS-DA – Hypoxia-high vs Hypoxia-low

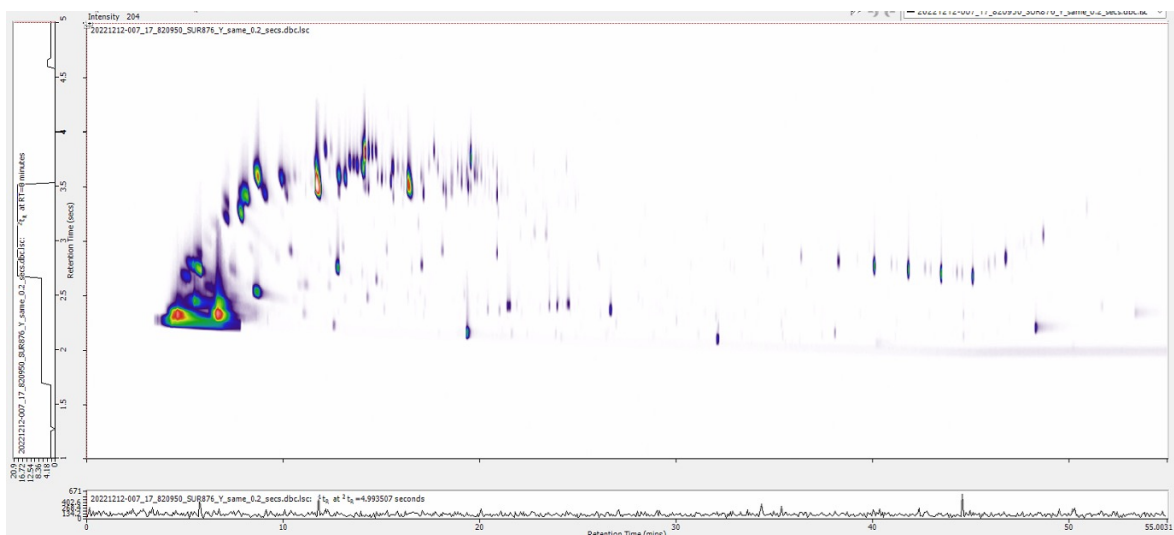


**C**

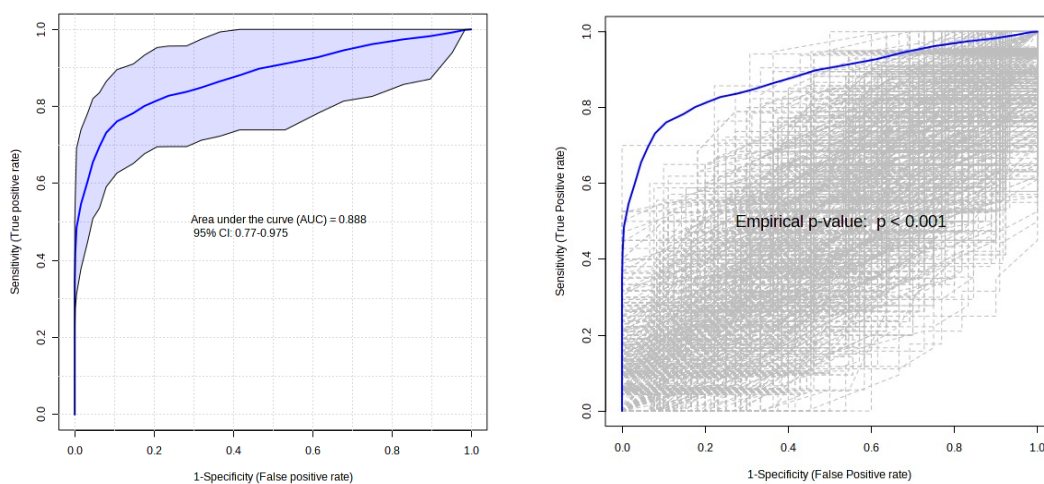


**A**, PCA plots of the separation between samples and blanks, and hypoxic classes. **B**, Score plot of the OPLS-DA model generated on the untargeted GC-TOF-MS tile data set, indicating no hypoxic class separation. **C**, Permutation testing – 999 repeats: The Y-axis represents the  $R^2Y$  and  $Q^2$  values of each model. The X-axis shows the correlation coefficient between the “real” Y and the permuted Y.  $R^2Y$ , 0.297;  $Q^2$ , 0.0462.

**Figure 40. TD-GC-TOF-MS (mid-polar) untargeted analysis**



**Figure 41. Representative GCxGC-TOF-MS chromatogram**



ROC analysis of the 8 VOC model using a Random Forests algorithm. **A**, ROC curve after 100 cross-validations. **B**, Permutation testing (1000 permutations).

**Figure 42. ROC analysis of the 8 VOC model using a Random Forests algorithm**

### 3.5 Discussion

In this section, a validated hypoxic gene signature was used to perform a stratified analysis of exhaled VOCs in OAC patients. The main finding from this pilot cohort study was the identification of candidate VOCs which defined hypoxic OAC tumours. Diagnostic performance assessment of the VOC model from untargeted GC-MS analysis revealed that this VOC signature acted as a reliable surrogate of hypoxic gene expression. The implication is that oncological molecular phenotypes may be reflected in the volatilome which could permit non-invasive treatment stratification.

A key strength of this Chapter was the quality control used at every stage of breath analysis. This was vital for ensuring the accuracy, reliability, and consistency of measurements. An additional strength of this work was validation of an efficient sequential TD re-collection pipeline, which was shown to improve sensitivity for downstream TD-GC-TOF-MS analysis to help improve compound identification. Future work is ongoing to establish thresholds for sample suitability in large patient cohorts, using either single or multiple compounds to confirm 'adequate' breath samples have been collected.

Overall, LCM was a useful technique that enabled the isolation of pure tumour populations. The LCM workflow highlighted a significant limitation associated with extracting RNA from bulk tissue: the amount of tumour tissue obtained from tissue biopsies can vary greatly, which may result in inconsistent RNA yield and quality. However, there was a significant attrition rate (41.2%) attributable to low RNA yields and RNA degradation during tissue processing. Low yields are expected from micro-dissected specimens and can be increased by increasing the amount of tissue available for sectioning and extraction. RNase enzymes, however, are ubiquitous in the environment and are difficult to inactivate as they are stable and do not require co-factors. Stringent measures are required at every stage from tissue harvesting to RNA elution to mitigate their effects. Attempts at minimizing RNase-induced RNA degradation for this study included: snap freezing tissue at the time of retrieval, cleaning the bench tops and instruments with RNase Zap®, utilising RNase-free reagents and plasticware, and limiting micro-dissection time per slide to 30 minutes. Cryosectioning, however, could have been improved by using single-use heat-treated microtome blades cleaned with RNase Zap®. Since tissue harvesting is a vital step that can influence RNA stability due to endogenous RNases, an additional procedure that could have been performed at the time of tissue retrieval was to use an RNA stabilizing solution such as *RNA/later*®. Actively perfusing *RNA/later*® into tissue has been shown to significantly improve RNA yields compared to immersion, presumably by exposing more of the surface area of the tissue to the stabilization solution<sup>213</sup>. The reason RNA preservation solutions were not used in this

work was due to potential for significant tissue disruption, causing histological artefacts and preventing accurate pathological assessment for LCM<sup>213</sup>. Another critical step is the staining and micro-dissection step. Ethanol used as a solvent in all preparatory steps could significantly reduce RNA degradation<sup>214</sup> as well as adding RNase inhibitors to staining solutions<sup>215</sup>. Downstream of sample preparation, alternative next generation sequencing techniques could also be used to maximise utility of low-quality samples for large scale gene expression profiling studies compared to conventional poly-A-based techniques. These include mRNA-seq, focusing on the coding transcriptome only, RNA sequencing utilising ribosomal RNA depletion or random priming instead of poly-A selection, or RNA exome capture sequencing, all of which can be used on FFPE or highly degraded samples<sup>216</sup>. Therefore additional steps could be evaluated to improve RNA stability in future LCM protocols alongside alternative sequencing platforms.

The compounds in the final targeted and untargeted models were alkanes, aldehydes, ketones, and alcohols. These are all compound classes that could theoretically originate from lipid peroxidation<sup>120</sup>. It is interesting to note that six out of eight VOCs from the untargeted analysis were down-regulated in the hypoxic sub-group while the compounds identified on targeted analysis were increased. The scarcity of detailed research into the metabolism of VOCs in humans means that the role of the detected metabolites and the relevance of their increase/decrease in disease states is still mostly unknown. Among the VOCs identified from untargeted analysis, only one (1-Hexanol, 2-ethyl-) has been previously reported in the human volatilome on searching the Human Breathomics Database<sup>103</sup> in articles unrelated to cancer. Future work will need to confirm the identity of compounds identified on the untargeted analysis using authentic standards and validate their performance in a larger population using a targeted method with absolute quantification. However this method of analysis cannot conclusively say whether OAC tumour hypoxia directly enriches these VOCs in breath. It would be of great interest to perform metabolic and volatilomic profiling of tumour tissue directly to bridge the gap between exhaled VOCs and hypoxic gene expression. Spatial lipidomics in particular (e.g. using DESI-MS) paired with spatial transcriptomics could also be a powerful technique to link lipid profiles with corresponding hypoxic transcriptional responses.

The tile base algorithm was a powerful feature discovery technique but was also found to generate a significant number of false positives due to siloxane contaminants and redundant features (overlapping ions). This can be mitigated by fine tuning initial noise filtering thresholds at a cost of reduced power in feature discovery. Use of an ensemble model by combining three classification and regression algorithms was felt to strengthen model building by reducing the chance of detecting

spurious classifiers. This technique was a simplified version of a method used recently to develop a breath VOC model for colorectal cancer, which utilised six machine learning techniques with multiple permutations<sup>212</sup>. A limitation of the existing method is the absence of weight differentiation among classifiers, considering that some classifiers may outperform others. A future research priority will be to optimize this pipeline by automating tile-based feature discovery and combining the output with weighted multivariate classifiers. Recent validation of tile-based Random Forest analysis for GCxGC-ToF-MS analysis suggests this is an area of active interest<sup>210</sup>.

There was a lack of corroboration between the compounds identified in the retrospective analysis (2.4.2) which were mainly aldehydes and the VOCs identified in targeted analysis in this study. This could be due to the small sample sizes, different gene signatures and improvement in quality control over time. However, SIFT-MS and GC-MS also operate on fundamentally different principles which make direct comparison challenging. In SIFT-MS, VOCs are ionized using precisely controlled ion-molecule reactions in a flow tube. This permits direct, real-time, quantitative analysis without sample preparation or prior calibration, relying on known reaction rates between analytes and reagent ions. GC-MS, on the other hand, separates VOCs based on their interactions with a stationary phase prior to ionization and measurement using MS. This key separation step provides more detailed information about individual VOCs within a mixture, including structural insights based on fragmentation patterns which can be compared against a reference spectral library. The reliance of SIFT-MS on a kinetic database established through ion-molecule reactions with selected reagent/precursor ions for VOC verification may result in ambiguous identification for isobaric and isomeric compounds. Thus, the fundamental differences between SIFT-MS and GC-MS in terms of separation, detection, and ionization techniques result in data that cannot be directly compared<sup>217</sup>.

This data is also subject to limitations. The final sample size was small which limited the ability to perform internal validation by splitting the dataset into training and test cohorts, investigate the influence of demographic or tumour factors (e.g. size) on VOC profiles, and risks model overfitting. Spatial heterogeneity (e.g. superficial vs core vs invasive edge) was not directly assessed by taking multi-region biopsies. Establishing the within-tumour and inter-tumour variance of signature genes is an important quality assurance step in validating a gene signature for clinical use<sup>51</sup>. Combining pre- and post-chemotherapy samples may have generated artefacts as the short ischaemic time of surgical tissue could influence hypoxic gene expression as well as a latent effect of chemotherapy. This requires further analysis in a larger pre-chemotherapy cohort. Directly correlating gene expression-

based hypoxia scores with levels of intra-tumoural oxygen is challenging, as hypoxic signalling can differ from the actual hypoxic conditions of the cellular surroundings (microenvironmental hypoxia).

In conclusion, this study offers preliminary insights and methodology into defining hypoxia-associated VOCs in OAC patients. An immediate extension of this work would be to verify the identity of candidate VOCs with authentic standards, refine the LCM extraction and sequencing process, and conduct a larger prospective study of exhaled VOCs with tissue-level metabolic phenotyping.

# **CHAPTER 4 – METABOLIC EFFECTS OF HYPOXIC REPROGRAMMING IN OESOPHAGEAL ADENOCARCINOMA**

## Summary

The objective of this Experimental Section was to explore the mechanisms of hypoxia-associated VOC signatures through a descriptive phenotyping study of OAC cell lines cultured under hypoxic conditions. Transcriptomic data revealed a co-ordinated transcriptional response to hypoxia in OAC cells characterised by enrichment of glycolysis and EMT-related genes. Two genes related to lipid metabolism, *HILPDA* and *ANGPTL4*, were strongly up-regulated in all OAC cell lines under hypoxia ( $\log_{2}FC > 1.5$ , FDR  $P < 0.05$ ). Global lipidomic profiling revealed a common set of lipids induced in all OAC cell lines that was predominantly composed of phosphatidylcholines. Headspace VOC analysis was performed on extracted lipids using high capacity sorptive TD-GC-TOF-MS. An automated data filtering and normalization pipeline was developed and applied to the analysis of volatilomic data. This identified nine VOCs that were up-regulated by hypoxia which correlated with a subset of lipids. Finally, redox analysis suggested that hypoxia induces a state of high oxidative stress and conditions favouring lipid peroxidation in OAC cells.

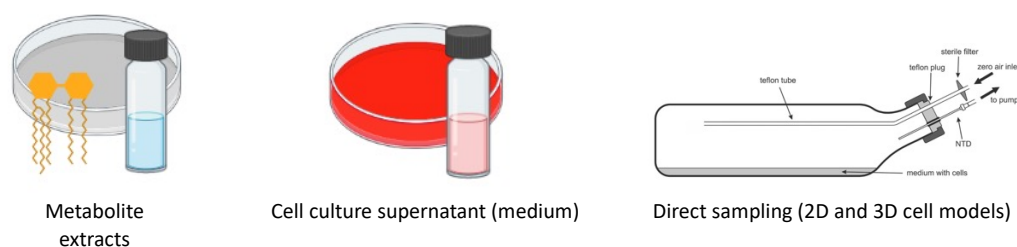


## 4.1 Methodological rationale

### 4.1.1 In vitro VOC analysis

For this section, the underlying mechanisms for hypoxia-associated VOC production were explored. Several options exist to extract VOCs from cell models for quantitative analysis.

VOCs can be extracted directly from the headspace of cell cultures in closed systems or from cell culture extracts using headspace or liquid-phase analytics (Figure 43). Direct sampling has notable advantages over indirect methods, including the ability to perform dynamic VOC sampling in a non-destructive manner. However there are significant technical challenges with designing a system for headspace collection that maximises analyte recovery of trace VOCs and maintains optimal culture conditions. As a consequence, although attempts at development are ongoing<sup>218</sup>, there are no standardized systems for headspace collection. Cell culture medium VOC analysis is another option but is limited by high background VOC emissions. Given the prevailing hypothesis for a lipid origin for VOCs in OAC (Chapter 1.3), an offline approach focussed on modelling lipid-to-VOC dynamics from purified cell lipid extracts was adopted.



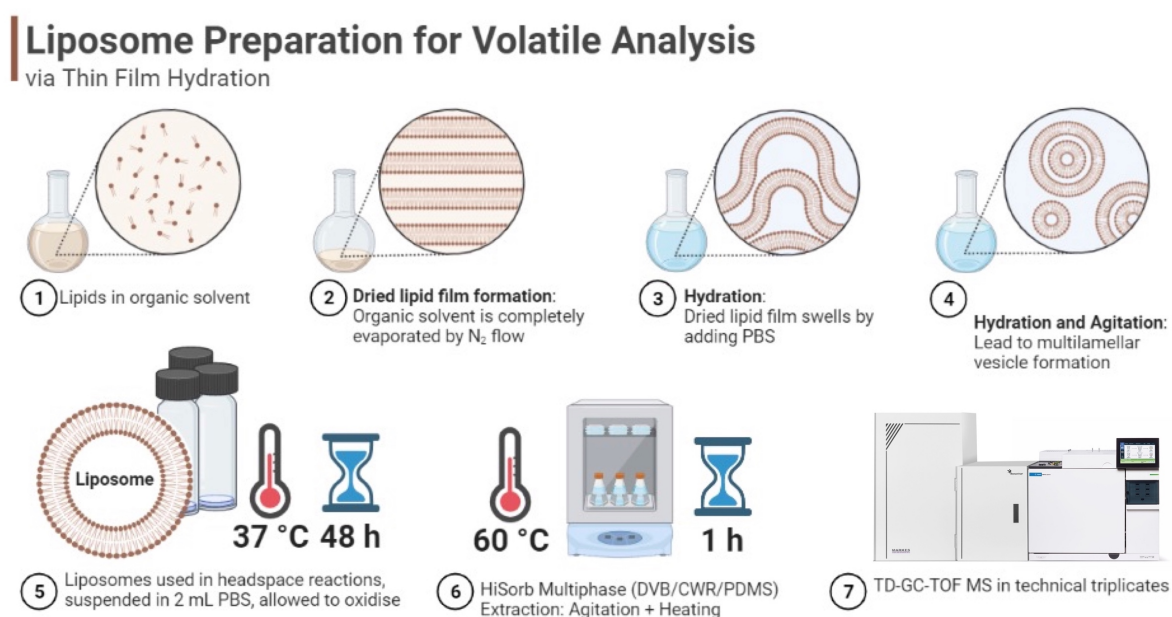
**Figure 43. Choice of sample matrix for *in vitro* VOC analysis**  
(Created using BioRender)

### 4.1.2 Modelling lipid to VOC dynamics

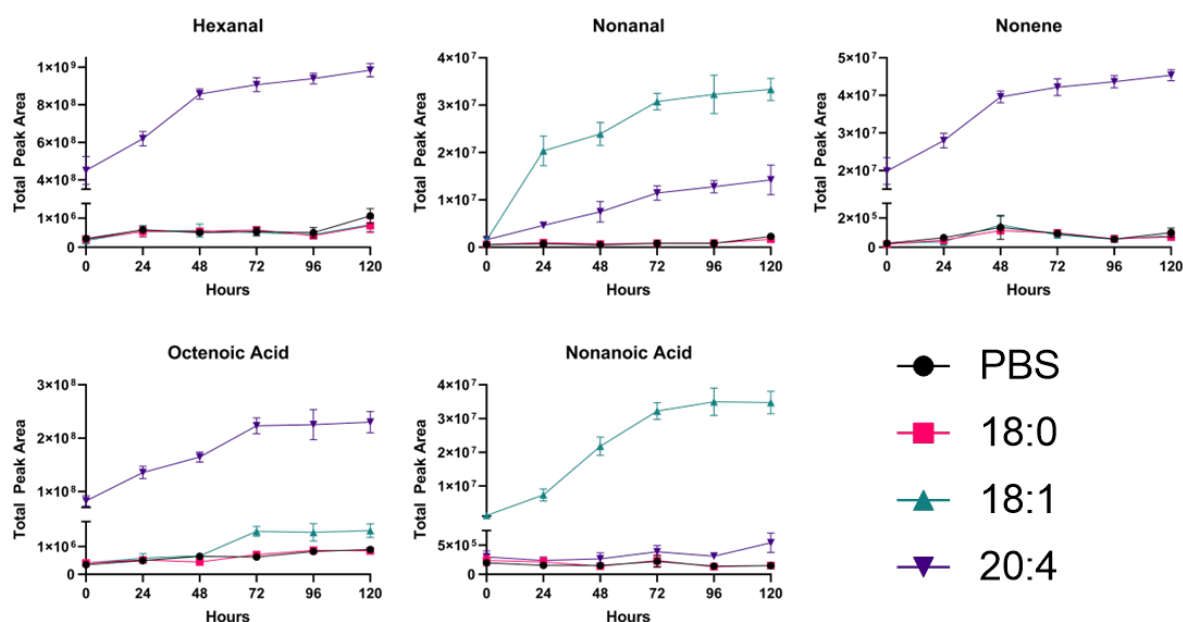
This method relies on the analysis of VOCs produced from simple bilayer vesicles generated from cellular lipid extracts. Shestivska et al previously used synthetic membranes in the form of large unilamellar vesicles (LUVs) to study VOC emissions resulting from lipid peroxidation<sup>219</sup>. In this study, LUVs were prepared from phospholipid standards and exposed to Fenton's reagent (ferrous iron) for real-time monitoring of volatile aldehyde production of SIFT-MS. Several short- and medium-chain aliphatic aldehydes were increased in the headspace of the LUVs upon peroxidative stress,

demonstrating the utility of this model system in investigating how the composition of the lipidome translates to a VOC signature.

An in-house optimised pipeline for the analysis of VOCs derived from cellular lipid extracts was applied to identify lipidomic and volatilomic changes upon exposure to hypoxia. A phase separation technique was used to extract the lipid fraction from cell lysates into an organic solvent phase, which was then evaporated off under inert gas. These lipid samples were split to enable paired untargeted lipidomic and VOC analysis. For VOC analysis, liposomes (LUVs) were generated by gentle sonication in PBS which acted to mimic cellular membranes. The liposomes were then exposed to ambient oxidative conditions for prior to undergoing headspace VOC analysis (Figure 44). Previous data has shown that 72 hours was an optimum timepoint to monitor the production of VOCs from lipids containing unsaturated acyl chains (Figure 45).



**Figure 44. Liposome preparation for volatilomic analysis**  
(Created using BioRender by Philip Leung and reproduced with consent)



**Figure 45. Volatile products can be traced to the lipidome (dynamic profiling)**  
 Analysis of VOC production from LUVs generated from standards revealed that 72 hours was the optimum timepoint  
 (Data courtesy of Philip Leung and reproduced with consent)

#### 4.1.3 Headspace VOC analysis

Headspace GC-MS analysis is well-suited for the analysis of VOCs. This technique extracts only the gaseous phase above a matrix which simplifies sample preparation, minimizes matrix interference and protects the GC column from the introduction of non-volatile contamination. A static headspace sampling technique using high-capacity sorptive extraction (HiSorb®) was chosen as this has been shown to provide high sensitivity and selectivity for trace analytes over conventional solid phase microextraction (SPME) (Figure 46). These are inert steel probes with a sorbent phase that are first exposed to sample headspace to capture VOCs and then placed within TD tubes for GC-MS analysis. A multiphase (PDMS/CAR/DVB) sorbent was chosen over single phase (PDMS) as previous data published by the Hanna lab showed that the multiphase HiSorb outperformed the single phase in terms of analyte recovery and reproducibility<sup>220</sup>.



**Figure 46. High capacity sorptive extraction-thermal desorption method for static headspace extraction (Images from Markes International)**

#### **4.1.4 Designing an automated pipeline for GC-MS data analysis**

The analysis of untargeted GC-MS data is a complex process. Peak overlaps from compound co-elution complicates accurate compound identification. Additionally, baseline noise, peak tailing, drift, and artifacts hinder precise peak detection and quantification. Variations in instrument performance and methodologies introduce data variability, making standardization difficult. The computational demands for processing large or intricate datasets are high, demanding significant resources and time. Crucial pre-processing steps such as noise reduction and peak detection are time consuming and susceptible to human error.

Automated deconvolution is the process to separate overlapping peaks in a mixture, enabling better identification and quantification. Initial attempts at automated deconvolution using both software and Opensource web servers revealed that these algorithms were computationally intensive and highly error prone, especially with complex samples. Minor variance in peak alignment, peak shape, and baseline noise was found to generate false-positive (missing real peaks) and false-negative (detecting non-existent peaks) results in each sample. Additionally, variability in pre-processing filtering steps was found to generate significantly different conclusions.

To address this challenge, a semi-automated peak table method for GC-MS data processing with a comprehensive filtering and normalisation pipeline was developed using R programming language. The aim was to enable standardisation, transparency, and reproducibility of data pre-processing of biofluid VOC data.

## 4.2 Hypothesis and aims

Chapter hypothesis: Hypoxia generates a stereotypical volatile response in OAC.

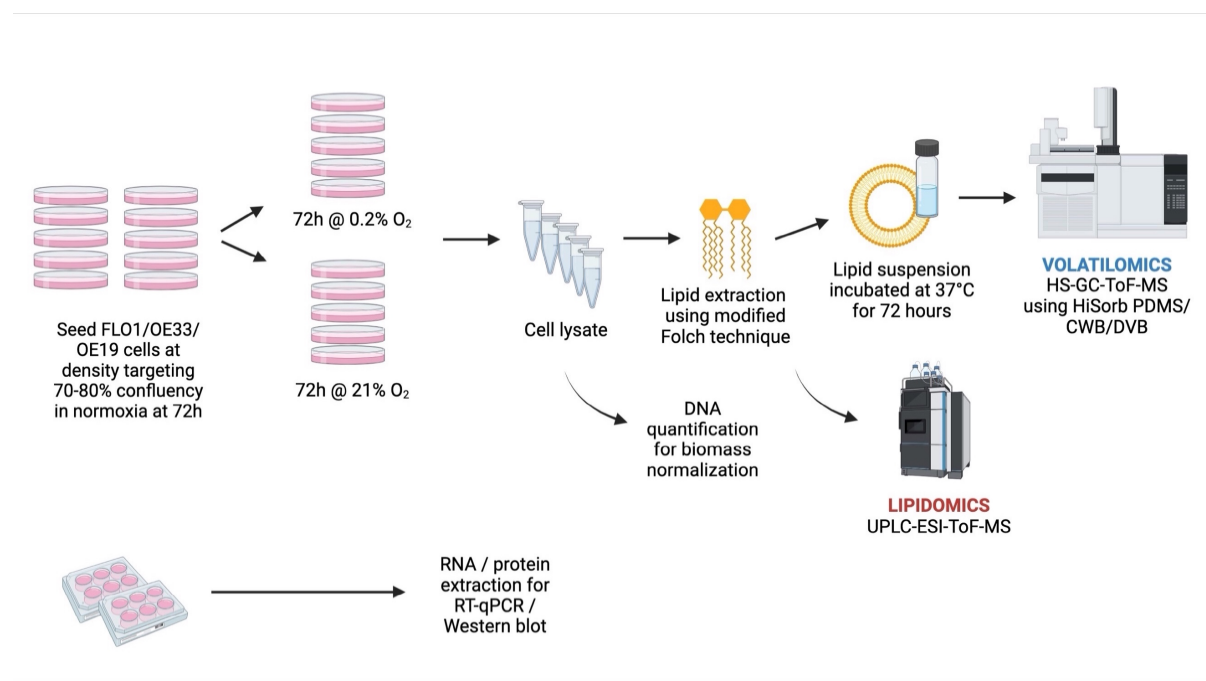
The objective of this section was to conduct a phenotyping study of OAC cell models to describe hypoxia-induced lipid and VOC signatures in OAC. Specifically, the aims of this section were to:

1. Describe metabolic adaptation to hypoxia in OAC by analysing cell line gene expression datasets.
2. Develop an analytical pipeline for non-targeted headspace GC-MS analysis of cell culture extracts.
3. Describe the lipidome and volatilome of cell culture extracts derived from hypoxic OAC cell lines.
4. Evaluate the impact of hypoxia on redox homeostasis in OAC cells.

## 4.3 Methods

### 4.3.1 Cell culture

Methodology for this chapter has been reported in line with experimental design and reporting standards for metabolomics studies of mammalian cell lines<sup>221</sup>. For metabolic profiling, FLO1, OE19 and OE33 cells were cultured as previously described (Chapter 2.3.4) and seeded in 10 cm dishes. Five replicates for each condition (normoxia and hypoxia) were generated and dishes with cell-free media were also incubated under the same conditions. For RNA extraction and immunoblot, cells were seeded on 6-well plates with three biological replicates. The final experimental workflow is described in Figure 47.



**Figure 47. Final experimental lipid-VOC workflow**  
(Created using BioRender)

Additionally, for flow cytometry assays, cells were seeded on 10cm dishes with three biological replicates. For fluorescence microscopy assays, cells were seeded in 8-well chamber plates, one plate per condition (Nunc™ Lab-Tek™ II Chamber Slide™ System, Thermo Fisher Scientific). For antioxidant co-factors assays, cells were seeded in 96-well white optical-bottom plates (Nunc MicroWell, Thermo Fisher Scientific) required for luminescence measurements with six replicates per plate. Cells were seeded at densities based on growth curves to achieve a target confluency of 70-80% after 72 hours incubation under normoxia.

#### **4.3.2 Hypoxia treatment**

After incubation overnight to allow adherence, plates were transferred to an H35 Hypoxystation (Don Whitley Scientific). The media was changed, and cells were incubated at 1% and/or 0.2% O<sub>2</sub> for 72 hours. Control plates were incubated in a normoxic incubator. All media and PBS used for the processing of hypoxia-treated cells were initially 'de-gassed' by incubating in the hypoxic incubator overnight.

#### **4.3.3 Immunoblot**

The method for Western blot has been previously described (Chapter 2.3.6). Primary antibodies used at 1:000 dilution were:  $\alpha$ -HIF-1 $\alpha$  (BD Biosciences #610958),  $\alpha$ -HIF-2 $\alpha$  (1:1000) (Novus Biologicals #NB100-122),  $\alpha$ -eIF2 $\alpha$  (Cell Signaling Technology #9722) and  $\alpha$ -phospho-eIF2 $\alpha$  (Ser51) (Cell Signaling Technology #9722). Secondary antibodies used at 1:2000 dilution were:  $\alpha$ - $\beta$ -tubulin (CST #2146) and  $\alpha$ -cyclophilin B (CST #43603).

#### **4.3.4 RNA extraction and sequencing**

The methods for RNA extraction and sequencing have been previously described (Chapter 2.3.7 and 2.3.8).

#### **4.3.4 Differential gene expression analysis**

The R package *DESeq2* was used to process RNA-seq counts data for differential expression analysis and identify hypoxia up- and down-regulated genes in each OAC cell line. The Benjamini–Hochberg method was used for multiple testing correction with P value cut-off specified at 0.05. The R package *fgsea* was used to perform GSEA, with gene ranking based on the DESeq2 test statistic and MSigDB v7.2 Hallmark, KEGG, Reactome and Gene Ontology gene sets as the reference gene lists. The R package *gsva* was used to generate sample-wise enrichment scores for lipid gene sets regulated unsaturated fatty acid metabolism.

#### **4.3.7 Redox assays (flow cytometry)**

For ROS measurement, cells were resuspended in 1 $\mu$ M dihydrorhodamine 123 (DHR) (Thermo Fisher Scientific) in PBS and incubated for 30 min at 37°C. This probe is non-fluorescent but can be oxidized to cationic rhodamine 123 which localizes to mitochondria and exhibits green fluorescence. Cells were pelleted by centrifugation and resuspended in 500 $\mu$ l PBS. For lipid peroxidation measurement<sup>222</sup>, cells were incubated with 500nM BODIPY™ 581/591 C11 (Lipid Peroxidation Sensor) (Thermo Fisher Scientific) for 30 minutes at 37°C. This fluorescent probe localizes to cell membranes in live cells and contains a polyunsaturated butadienyl element which can be oxidized, resulting in a shift of the fluorescence excitation/emission peaks from red (non-oxidized, 581/591 nm) to green (oxidized, 488/510 nm). Cells were washed with PBS and detached non-enzymatically with Versene solution (Thermo Fisher Scientific). The cell suspension was washed twice with PBS and resuspended in 500 $\mu$ l PBS. All PBS wash was retained and collected with the detached cells. Cells stained for ROS and lipid peroxidation were transferred to flow cytometry tubes via a 40 $\mu$ m nylon mesh prior to flow cytometry. Flow cytometry was performed using a FACS Canto II instrument (BD). Cells were gated according to forward and side scatter values, including a doublet gate to exclude cell clusters. Data was processed using FlowJo™ version 10.

#### **4.3.8 Redox assays (fluorescence microscopy)**

Live cells in 8-well chamber plates were incubated with 500nM BODIPY™ 581/591 C11 (Lipid Peroxidation Sensor) (Thermo Fisher Scientific) for 30 minutes. The media was replaced with Live Cell Imaging Solution (Invitrogen) and chamber slides were immediately imaged using a Stellaris 8 confocal microscope (Leica) as per the original *Nature Protocol*<sup>223</sup>. Images were processed using Fiji software and a custom macro toolset developed by Stephen Rothery at Imperial College Facility for Imaging by Light Microscopy (FILM).

#### **4.3.9 Redox assays (antioxidant co-factors)**

Quantification of antioxidant co-factors was performed using NADP/NADPH-Glo™ and GSH/GSSG-Glo™ assays (Promega) as per manufacturer instructions.



#### **4.3.10 Sample preparation for paired lipidomics and volatilomics**

All reagents were UPLC-grade unless specified otherwise. From each 10cm dish, the cell media was removed, centrifuged for 2 minutes at 4000rpm to pellet debris, and stored at -80°C. Cells were then washed twice with ice-cold 150mM ammonium acetate. After addition of 500µl ice-cold UPLC-grade water to each 10cm dish for both quenching and extraction, cells were scraped into microfuge tubes and stored at -80°C. An aliquot of cell homogenate was used for DNA extraction and quantification (DNeasy Blood & Tissue Kit, Qiagen) for biomass normalization<sup>224</sup>. A modified Folch method was subsequently used for lipid extraction. After thawing samples on ice, ice-cold methanol/chloroform (2:1) was added to microfuge tubes, vortexed for 4 minutes and incubated for 20 minutes on ice. A further 200µl of ice-cold chloroform was added to microfuge tubes, followed by vortexing for 1 minute and incubation for 30 minutes on ice. Samples were then centrifuged for 15 minutes at 11,000rpm at 4°C. The lower organic phase was collected in amber glass vials. The extraction process was then repeated (double extraction). Each extraction included extraction blanks (i.e. water alone) as controls. Samples were then split equally into two amber glass vials (for paired lipidomics and volatilomics), before being dried under a gentle nitrogen stream.

#### **4.3.11 Ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) method**

All lipid samples were sent to the MRC-NIHR BRC National Phenome Centre (NPC) for analysis. The following is a description of the NPC SOP for lipidomic analysis. Samples were re-suspended in water/isopropanol with a mixture of lipid internal standards. Separation via liquid chromatography was performed using an Acquity UPLC system (Waters Corporation, USA) with injections for positive and negative electrospray ionization (ESI). An Acquity UPLC BEH column (C8, 2.1 × 100 mm, 1.7 µm) was employed for this purpose. The mobile phases comprised water/isopropanol/acetonitrile (in a ratio of 2:1:1 v:v:v) with the addition of 5mM ammonium acetate, 0.05% acetic acid, and 20 µM phosphoric acid for Mobile Phase A. Mobile Phase B was isopropanol:acetonitrile (in a ratio of 1:1 v:v) with the addition of 5mM ammonium acetate and 0.05% acetic acid. The flow rate stood at 0.6 ml/min with a total runtime of 15 minutes. The gradient began at a starting condition of 1% mobile phase B for 0.1 minutes, rose to 30% mobile phase B between 0.1 and 2 minutes, and reached 90% mobile phase B from 2 to 11.5 minutes. It remained at 99.99% mobile phase B from 12 to 12.55 minutes before reverting to the initial condition for re-equilibration.

For mass spectrometry detection, a Xevo G2-S QToF mass spectrometer (Waters MS Technologies, UK) was used, with data acquisition in both positive and negative modes. Specific MS settings: capillary voltage of 2.0 kV for positive mode and 1.5 kV for negative mode, a sample cone voltage of 25V, a source offset of 80, source temperature at 120 °C, desolvation temperature at 600 °C, desolvation gas flow at 1000 L/h, and cone gas flow at 150 L/h. The data collection occurred in centroid mode, scanning within a range of 50 – 2000 m/z, and with a scan time of 0.1s. LockSpray mass correction was implemented for mass accuracy. This correction utilized a 600 pg/μL leucine enkephaline solution in water/acetonitrile (1:1, v/v) at a flow rate of 15 μL/min, targeting m/z 556.2771 in ESI+ and m/z 554.2615 in ESI- for accurate calibration.

Pooled Quality Control samples (QC) were generated by combining aliquots from all samples and were injected once every 10 samples. Spectral deconvolution was performed using *xcms*. Metabolic features were filtered, retaining only those exhibiting a relative coefficient of variation (CV) below 15% across all QC samples. Additionally, any metabolic features that did not demonstrate a correlation coefficient greater than 0.9 in a serial dilution series of QC samples were excluded from the analysis. Targeted peak annotation was performed using *peakPanther*<sup>225</sup>.

#### **4.3.12 Liposome preparation for headspace analysis**

Dried lipid samples were resuspended with 600μL PBS and vortexed briefly every 15 minutes for 1 hour to generate a liposomal (large unilamellar vesicles) suspension (Figure 44). Pooled QC samples were generated by combining 150μL aliquots from all samples and also used to generate a serial dilution series. PBS (1550μL) and re-suspended sample (450μL) were added to amber glass crimp-top headspace vials (20ml volume) (Markes International). After addition of internal standards (propanol-d8, octane-d18, toluene-d8, acetophenone-d8 and benzaldehyde-d6), the vials were sealed with caps containing silicone/PTFE septa (Markes International). Vials were incubated at 37°C for 72 hours to permit ambient lipid oxidation and release of VOCs (Figure 44).

#### **4.3.13 Preparation of sorbent probes for headspace sampling**

Headspace VOC extraction was performed using short high capacity sorptive extraction probes (HiSorb®) containing DVB/CWR/PDMS sorbent phase (Markes International). This sorbent phase was chosen to capture the widest range of VOCs. Sorbent probes were placed within TD tubes and were conditioned (cleaned) using a TC-20 TD tube conditioning unit (Markes International Ltd, Llantrisant,

UK). An individual unit could condition 20 tubes simultaneously. The flow rate was checked with a flow meter to ensure N<sub>2</sub> flow was at 100ml/min. Tubes were inserted into the manifold which was then placed in the heating block. Tubes were conditioned for 2 hours at 260°C under continuous nitrogen flow. The manifold was then removed from the heating block to allow the tubes to cool for 5 minutes. Inert diffusion-locking caps (DiffLok™, Markes International Ltd, Llantrisant, UK) were placed on the sampling ends of the tubes, and stainless-steel caps were placed on the non-sampling ends, minimizing the time the tubes were open to the atmosphere. These tubes were then placed in TD autosampler trays.

#### **4.3.14 GC-FID method**

Conditioned sorbent probes in TD tubes were subsequently checked for cleanliness by analysis using a GC-FID instrument, similar to sorbent tubes used for breath. TD tubes were loaded onto a TD100-xr thermal desorption unit and sequentially pre-purged with helium (50 mL/min) for 30 seconds then heated to 260°C for 10 minutes. VOCs were desorbed onto a focusing trap (U-T12ME-2S, Markes International) with a 5.5:1 outlet split, which was then heated to 300°C for 3 min to release VOCs into the GC column. The GC system used was an 8890 (Agilent Technologies) coupled to an FID. All data were acquired and pre-processed with ChromSpace (V2.1.6, Markes International). The focusing trap was set at 30°C and a fused silica capillary column (HP-5ms, 30 m x 250µm x 0.25µm, Agilent) was used with hydrogen carrier gas. The GC oven was initially set at 60°C for 1 min, then increased to 300°C (50 °C/min with a 2 min hold). Following Top-hat background subtraction (peak width set at 20 seconds), peaks were integrated using the 'curve fitting' peak detection algorithm within the ChromSpace software, using pseudo-Gaussian smoothing at 3 points. The total area of all identified peaks was calculated and compared against previously set QC limits displayed in a Levey-Jennings plot. These limits had been set by analysing the total peak area in 500 conditioned sorbent probes in TD tubes. Tubes with a total peak area > 1 s.d. of the mean were classed as a QC failure and were re-conditioned with new caps. Conversely, a total peak area < 1 s.d. could indicate a trap blank had run due to a tube pressurization failure (e.g. from worn O rings within the caps). These tubes were also re-conditioned with new caps. Tubes passing the QC check on the GC-FID progressed to headspace analysis.

#### 4.3.15 Headspace extraction

At the specified timepoint, sorbent probes were introduced into vials via the septa to expose the sorbent material to the headspace VOCs. Vials were agitated at 300rpm at 60°C on a HiSorb agitator (Markes International) for 1 hour. The sorbent probes were then placed into empty steel TD tubes which were capped with inert diffusion locking caps. All batches included PBS blanks, pooled QC samples every 10 samples and a serial dilution series using PBS.

#### 4.3.16 TD-GC-MS methods

**MS instrumentation.** First, a system suitability check was performed. This included an air/water check to ensure there were no leaks in the system, and a monthly auto-optimisation to adjust the filament voltage. TD tubes were then loaded onto a TD100-xr thermal desorption unit and sequentially pre-purged with helium (50 mL/min) for 1 minute then heated to 260°C for 15 minutes. VOCs were desorbed onto a focusing trap (U-T12ME-2S, Markes International) with a 5.3:1 outlet split for sample recollection (6 mL/min split flow, 1.4 mL/min column flow), which was then heated to 300°C for 3 minutes to release VOCs into the GC column. Split flow from HiSorb® probes was re-collected onto standard sorbent TD tubes (Tenax/Carbograph-5TD) normally used for breath analysis for further analysis using TD-GCxGC-ToF-MS. The GC-MS system used was an 8890 (Agilent Technologies) coupled to a Bench-TOF2 (Markes International). TOF measurement was continuous with mass range set between 35-500 m/z and ionization energy -70 eV. All data were acquired and pre-processed with ChromSpace (version 2.1.6, Markes International).

**TD-GC-ToF-MS method.** A mid-polar method alone was used for HiSorb® analysis. The focusing trap was set at 10°C and a fused silica capillary column (Rxi-624Sil, 30 m x 0.250 mm, 0.14 µm, Restek) was used with helium carrier gas. The GC oven was initially set at 40°C for 1 min, then increased to 280°C (10°C/min with 10 min hold). TOF data acquisition rate was set at 6 Hz.

**TD-GCxGC-ToF-MS method.** TD tubes were desorbed to 280°C for 8 minutes, with focusing trap temperature set at 10°C. An INSIGHT flow modulator (2nd Gen, SepSolve Analytical) was used for parallel split with valve period of 4s with 100ms flush. VOCs were separated using a BD-Heavy wax column in the first dimension (20 m x 0.18 mm x 0.18 µm), and a VF-200ms column in the second dimension (5 m x 0.25mm x 0.10 µm). The GC oven was initially programmed at 50°C for 3 min, then increased to 260°C (5°C/min with 5 min hold). TOF data acquisition rate was set at 50 Hz.

#### 4.3.17 GC-MS data analysis

**GC-MS data pre-processing.** All LSC data files underwent dynamic baseline compensation (DBC) with peak width at 6s for 1D methods and 0.2s for the 2D method followed by peak alignment using ChromSpace software.

**GC-MS data processing method using a semi-automated peak table integration method.** Pooled QC LSC files were exported to CSV files, and a custom R script (*compiler.R*, Appendix 10) was developed to create a composite chromatogram file from all QC files. This single file was then converted back to an LSC file. This composite chromatogram file containing data from all QC files underwent automated deconvolution using the following settings: peak area 100,000, minimum ion count 2000, and feature annotation using the NIST mass spectral library (reverse match factor > 800). The peak table file was exported and underwent further processing (*composite\_file\_processing.R*, Appendix 11) to remove: features with retention time < 1.2 minute (the minimum time required for analytes from samples to reach the detector), siloxanes (using a mass filter), signal to noise < 3 and known contaminants from the sorbent or transfer line (e.g. fluorinated compounds released from the Tenax sorbent). After filtering, the feature names, retention times, and primary ions (M1) were used to develop a targeted integration method. Every peak was reviewed again on the composite chromatogram file to manually confirm retention times, integration windows ( $\pm 0.03$ s), source and quantitative ions were optimal. This method was used to analyse all samples in the batch. The 'Quant Results' module within the ChromSpace software was then used review all sample files to ensure that peaks had been integrated correctly. This data was then exported for filtering and statistical analysis.

**GC-MS data processing – automated normalization and filtering.** An automated R script (*biofluids\_normalization\_filtering.R*, Appendix 12) was developed with a standardised sequential filtering strategy which involved:

1. Blank filtering: Features which had a blank/QC ratio < 30% were retained to exclude background contaminants.
2. Linearity filtering: Only features which had a Spearman's  $\rho > 0.7$  in the serial dilution series were retained to exclude non-linear features.
3. Internal standard selection: Features were normalised against each of the five internal standards. For each feature, the internal standard which generated the lowest CV between pooled QC samples was selected.

4. CV filtering: After internal standard normalization, only reproducible features which had a CV < 30% across QC samples were retained.

The final data pre-processing pipeline is displayed in Figure 48.

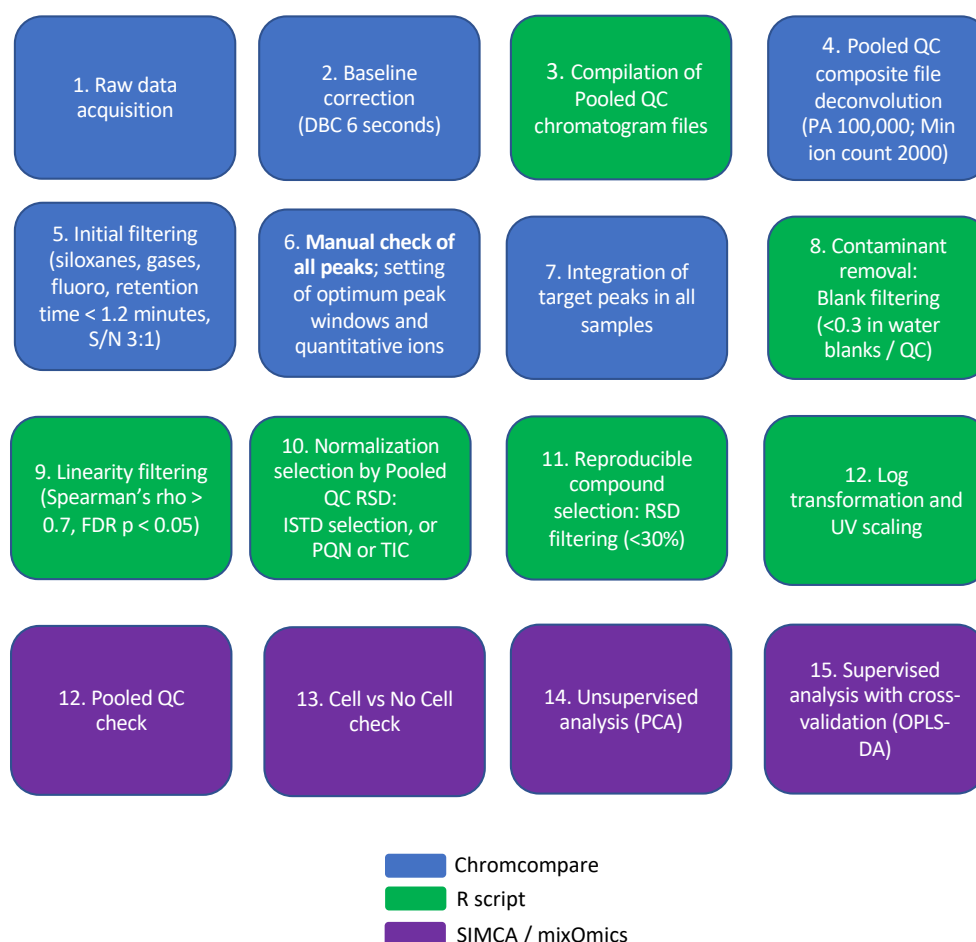


Figure 48. Final standardized GC-MS data pre-processing pipeline

#### 4.3.18 Statistical analysis

Multivariate data analysis was conducted using SIMCA® v17 (Sartorius). Prior to PCA and OPLS-DA, the dataset underwent biomass (DNA concentration) normalization, log transformation (base 10) and UV scaling. PCA was utilized to observe inherent clustering patterns and identify potential outliers within the dataset and OPLS-DA involved modelling the data to distinguish hypoxic versus normoxic samples.

For OPLS-DA, the assessment of model quality involved internal cross-validation ( $Q^2Y$ -hat value) and permutation testing. Permutation testing compared the true  $Q^2Y$ -hat value with 999 models

generated by randomly permuting class memberships. Valid and robust models were characterized by a positive  $Q^2Y$ -hat value and a permutation p-value below 0.05. The metabolic features responsible for class differentiation were identified by examining the corresponding S-plot. A cut-off threshold of 0.05 was applied in this analysis, which involved a scatter plot depicting model loadings and their correlation to class differentiation. Since individual classifiers may generate spurious results, an ensemble approach was used next, combining important features common to different classifiers to identify a core set of features most important for class differentiation.

Data was then uploaded to the Metaboanalyst v5.0 web portal ( <https://www.metaboanalyst.ca/> ) using the Statistical Analysis (metadata table) module. Both the data file and metadata file (containing cell type and oxygen conditions as categorical covariates) were uploaded. Heatmaps were generated using the 'heatmap visualization' tool, setting distance measure to 'Euclidean' and clustering algorithm to 'Ward'. Samples were arranged by cell type and oxygen conditions. Univariate analysis was performed using the 'Linear models with covariate adjustments' tool. This uses *limma* method for group comparison. Primary metadata was set to 'Oxygen', reference group 'Normoxia' and covariates (control for) was 'Cell type'. In order to define a compact metabolic signature of hypoxia that excluded highly correlated features, the 'Biomarker analysis' module was used, with oxygen conditions as the primary phenotype label. Three classification and feature rank methods were selected to identify robust class-differentiating features: Support Vector Machine (SVM), Partial least squares-discriminant analysis (PLS-DA) and Random Forest (RF). ROC curves were generated by Monte-Carlo cross validation (MCCV) using balanced sub-sampling and significant features were ranked by average importance. Features which were important (top 100 ranked by importance) to all three models were selected for final evaluation. The final model was then evaluated using the ROC curve-based model evaluation. The Area Under the Curve represents the discriminatory power of the model, with 1 indicating perfect classification.

Continuous data has been presented as mean (standard deviation, SD) or median (interquartile range, IQR). Categorical data has been presented as frequency (percentage). Parametric data were compared using Student's T test or ANOVA, non-parametric data were compared using Wilcoxon or Kruskal-Wallis rank sum test, and categorical data were compared using Pearson's Chi-squared test or Fisher exact test. All data analysis was conducted using R version 4.3.2 unless otherwise indicated. The main packages used were: *dplyr* (data manipulation); *ggplot2* (data visualization). Data visualization was also performed using GraphPad Prism version 9.

## 4.4 Results

### 4.4.1 Transcriptomics

Transcriptomic data derived from cell line experiments described in Chapter 2 was reviewed to evaluate hypoxic metabolic reprogramming in greater detail. Differential gene expression analysis was performed using *DESeq2*, which quantifies fold change using negative binomial generalized linear models<sup>226</sup> (Figure 49). At a threshold of  $\log_2FC > 1.5$  and  $FDR P < 0.05$ , there were 55, 2265 and 920 genes up-regulated in FLO1, OE33 and OE19 respectively upon hypoxic exposure. At a threshold of  $\log_2FC < -1.5$  and  $FDR P < 0.05$ , there were 5, 1791 and 315 genes down-regulated in FLO1, OE33 and OE19 respectively upon hypoxic exposure. There were 18 up-regulated genes which were common to all three cell lines, of which 14 were protein-coding (Table 11 and Figure 50A). Figure 50B shows a STRING-generated protein interaction network. This contains a subcluster of highly interconnected proteins with overrepresentation of metabolic genes regulating glycolysis and lipid metabolism. Importantly, two genes in this network, *HILPDA* and *ANGPTL4*, are key regulators of lipid metabolism.

Gene Set Enrichment Analysis was performed using *fgsea* using the HALLMARK, KEGG, REACTOME and GENE ONTOLOGY gene sets (Appendix 13) to identify common co-ordinated transcriptional responses to hypoxia. All hypoxic cells were characterized by enrichment of hypoxia, EMT, glycolysis and inflammatory response gene sets in all four databases, which is consistent with *in vivo* GSEA results reported in Chapter 2. A targeted analysis of pathways regulating unsaturated fatty acid metabolism was also performed using Gene Set Variation Analysis (Figure 51). This revealed significant down-regulation of the fatty acid metabolism under hypoxic conditions. A detailed analysis of this pathway revealed that this was driven mainly by down-regulation of genes regulating lipid turnover and aldehyde detoxification.



## DIFFERENTIAL GENE EXPRESSION ANALYSIS (48H HYPOXIA (0.2% O<sub>2</sub>) VS 48H NORMOXIA (21% O<sub>2</sub>))

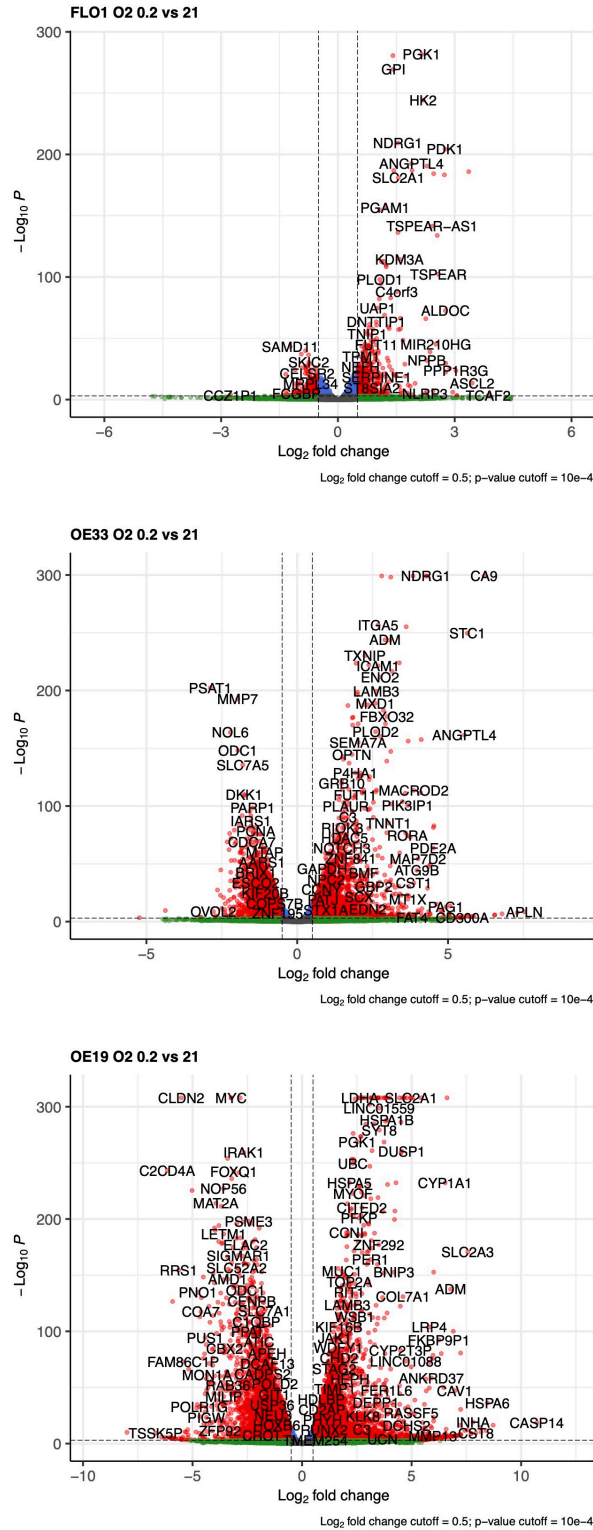
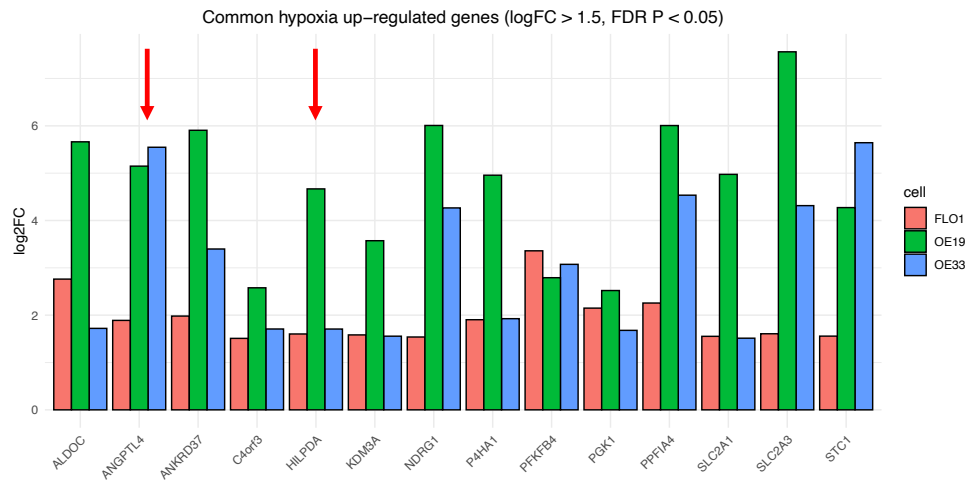


Figure 49. Volcano plots demonstrating differential gene expression analysis of OAC cell lines exposed to hypoxia

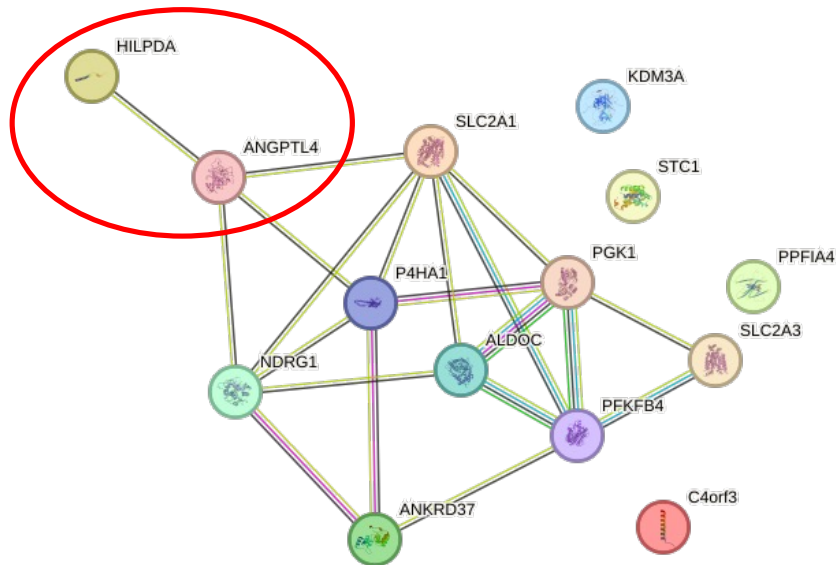
| <b>Gene</b>    | <b>Protein</b>                                                    | <b>Function</b>       |
|----------------|-------------------------------------------------------------------|-----------------------|
| <i>ALDOC</i>   | Aldolase, fructose-bisphosphate C.                                | Glycolysis            |
| <i>ANGPTL4</i> | Angiopoietin-related protein 4                                    | Lipid metabolism      |
| <i>ANKRD37</i> | Ankyrin repeat domain 37                                          | Unknown               |
| <i>C4orf3</i>  | Uncharacterized protein C4orf3                                    | Unknown               |
| <i>HILPDA</i>  | Hypoxia-inducible lipid droplet-associated protein                | Lipid storage         |
| <i>KDM3A</i>   | Lysine-specific demethylase 3A                                    | Epigenetic regulation |
| <i>NDRG1</i>   | N-myc downstream regulated 1                                      | Stress response       |
| <i>P4HA1</i>   | Prolyl 4-hydroxylase subunit alpha-1                              | Collagen synthesis    |
| <i>PFKFB4</i>  | 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 4            | Glycolysis            |
| <i>PGK1</i>    | Phosphoglycerate kinase 1                                         | Glycolysis            |
| <i>PPFIA4</i>  | Liprin-alpha-4                                                    | Signal transduction   |
| <i>SLC2A1</i>  | Solute carrier family 2, facilitated glucose transporter member 1 | Glucose transporter   |
| <i>SLC2A3</i>  | Solute carrier family 2, facilitated glucose transporter member 3 | Glucose transporter   |
| <i>STC1</i>    | Stanniocalcin-1                                                   | Calcium homeostasis   |

**Table 11. Common hypoxia up-regulated genes and protein products**

**A**

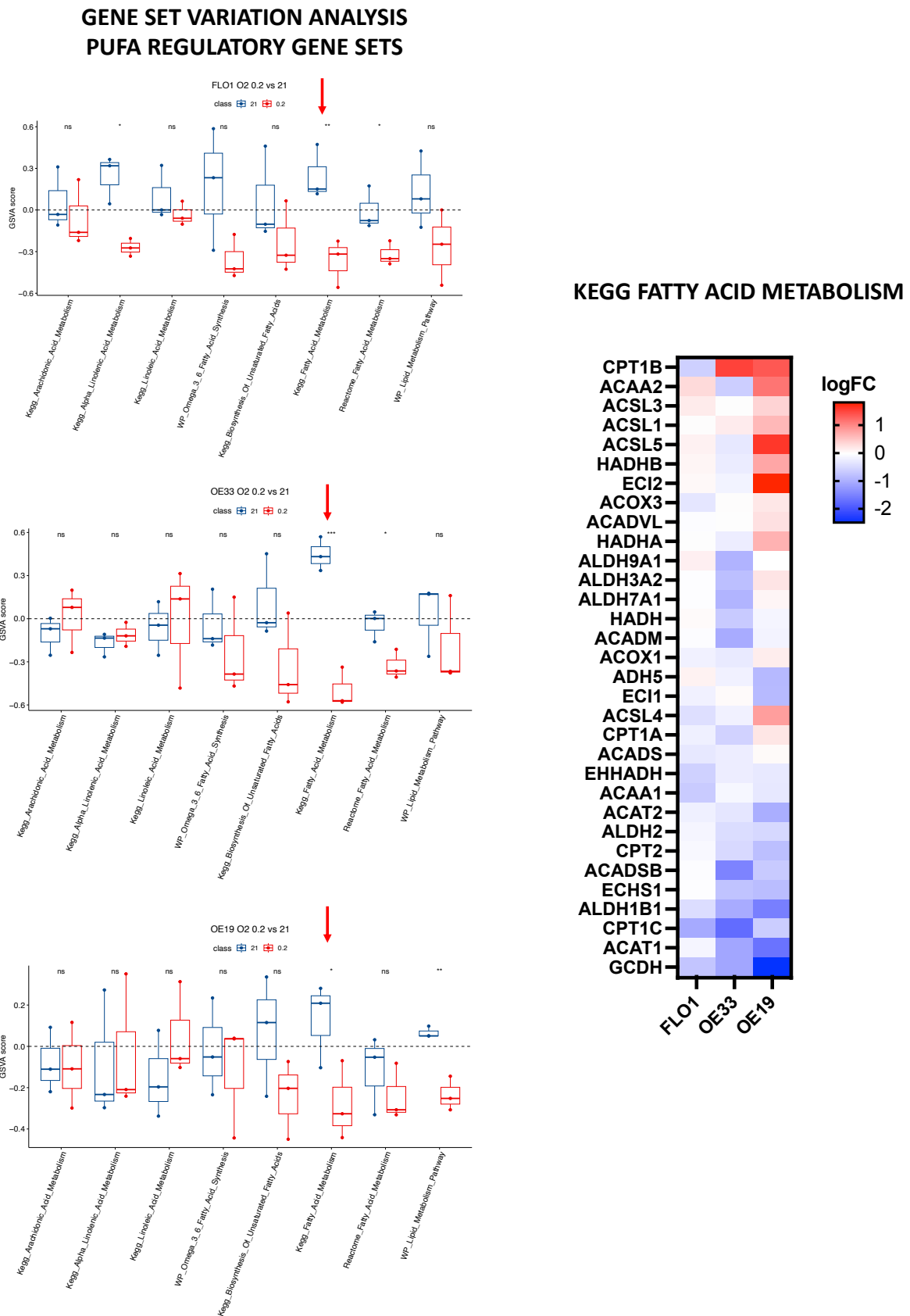


**B**



**A**, Boxplot showing common hypoxia up-regulated genes in all three OAC cell lines upon exposure to hypoxia (0.2% O<sub>2</sub>). **B**, Protein-protein association network for the 14 hypoxia up-regulated genes. *HILPDA* and *ANGPTL4* are highlighted.

**Figure 50. Common hypoxia up-regulated genes and STRING protein association network**



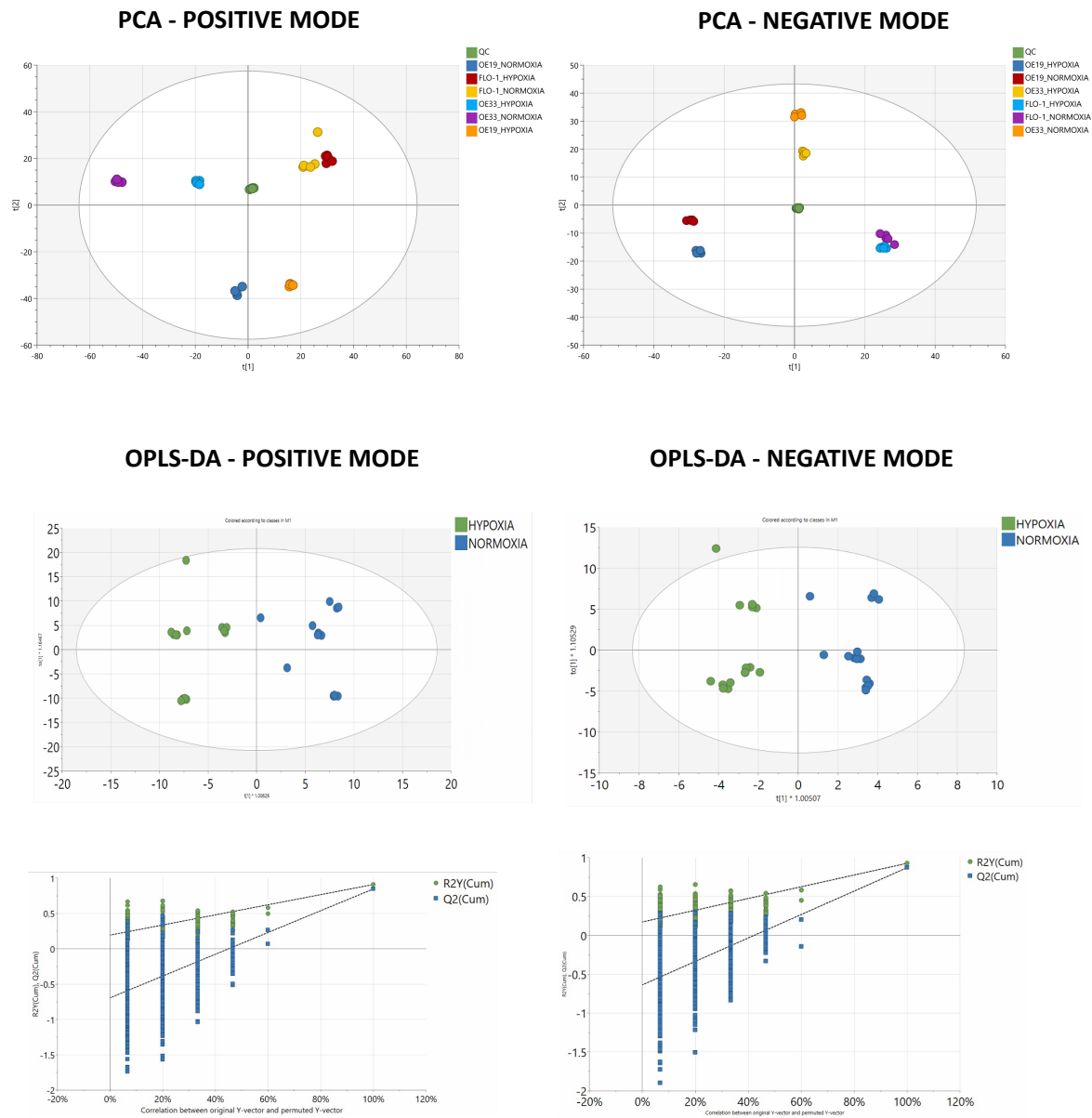
**Figure 51. GSEA analysis of regulators of unsaturated fatty acid metabolism**  
**Boxplots of GSEA scores of lipid metabolic pathways in OAC cells under normoxic and hypoxic conditions**  
**Heatmap showing relative fold change of fatty acid metabolism genes under hypoxia**

#### 4.4.2 Lipidomics

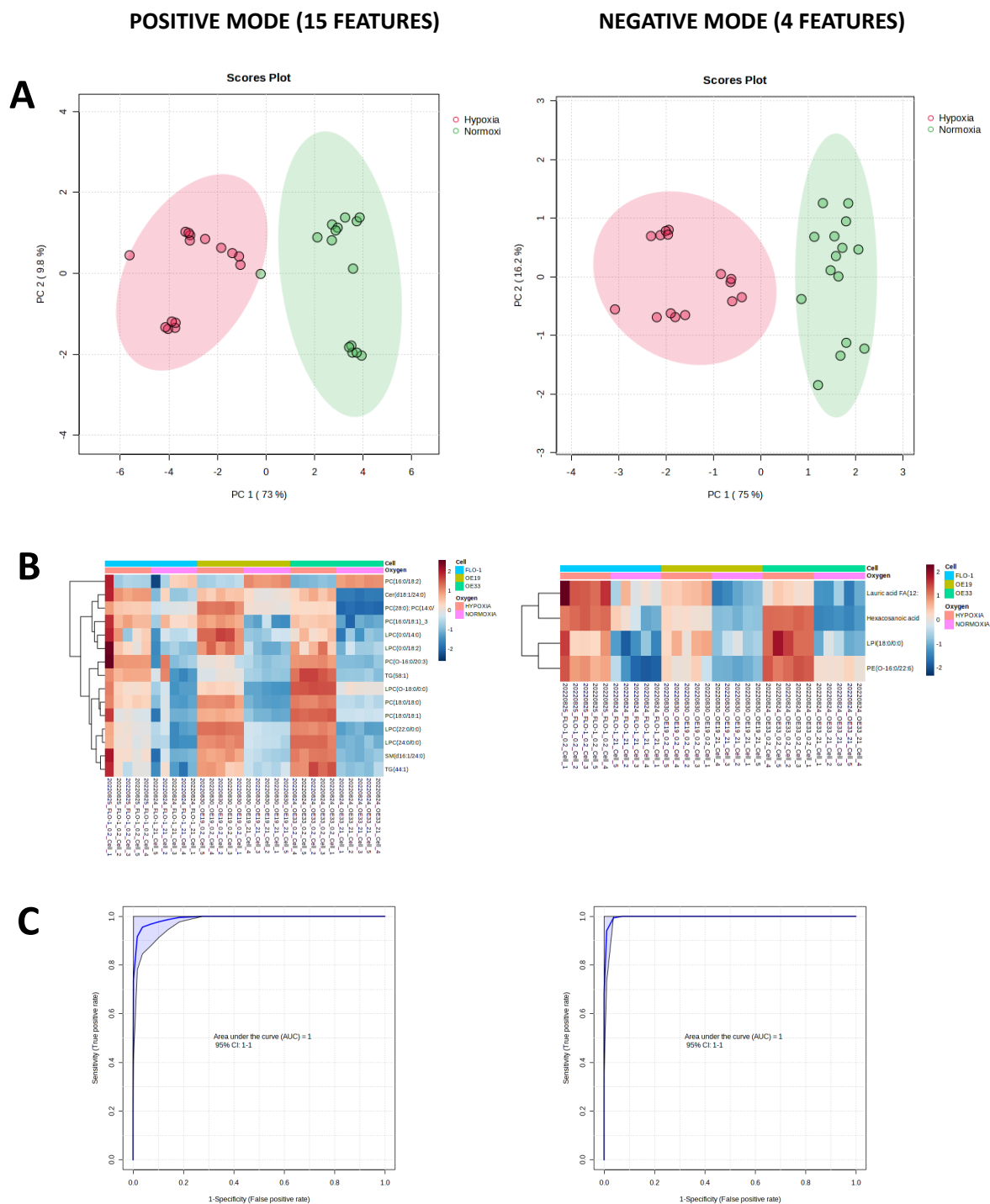
To identify a lipid signature that could reflect a core hypoxic response in OAC, lipidomic profiling was performed for all OAC cell lines exposed to hypoxia (0.2% O<sub>2</sub>) and normoxia. For both positive mode and negative mode data, unsupervised analysis with PCA showed strong separation by cell type suggesting cell type-specific lipid phenotypes (Figure 52). Separation by hypoxic status was also evident. QC samples were tightly clustered indicating minimal instrumental drift during acquisition.

For the positive mode data, supervised analysis with OPLS-DA showed clear separation between hypoxic and normoxic groups ( $R^2Y$  0.909 and  $Q^2$  0.847), which was validated from permutation analysis (999 permutations) and CV-ANOVA p value with a p-value of  $2.56 \times 10^{-8}$ . VIP analysis identified 28 annotated lipids with VIP scores > 1.35. For the negative mode data, supervised analysis with OPLS-DA showed clear separation between hypoxic and normoxic groups ( $R^2Y$  0.928 and  $Q^2$  0.872), which was validated from permutation analysis (999 permutations) and CV-ANOVA p value with a p-value of  $3.47 \times 10^{-9}$ . VIP analysis identified 11 annotated lipids with VIP scores > 1.35.

The lipidomic data was taken forward for further multivariate analysis. Ensemble analyses were performed using SVM, PLS-DA and Random Forest algorithms to identify robust class-differentiating features. This identified 19 features which were common to all three models and showed high accuracy in discriminating normoxic from hypoxic samples (AUC = 1) (Figure 53 and Table 12). This core hypoxic lipid signature was composed of 19 lipids, of which 18 were up-regulated in all OAC cell lines upon hypoxic exposure. This included 10 unsaturated lipids, of which 9 were up-regulated and 1 was down-regulated by hypoxia and 9 saturated lipids, all of which were up-regulated. The major lipid class was PC (phosphatidylcholine). Further analysis revealed that these lipids were highly correlated, suggestive of highly regulated core response to hypoxia (Figure 54).



**Figure 52. Unsupervised and supervised analyses of cell line lipidomic data**



**A**, Class separation on PCA using the core hypoxia-associated lipid signature. **B**, Heatmap showing relative change in lipid abundance upon exposure to hypoxia. **C**, ROC curve analysis of the discriminatory performance of the core hypoxia-associated lipid signature.

**Figure 53. Core OAC hypoxia-associated lipid signature**

| Lipid (PeakPantherR annotation)            | Mode     | Class | Saturation  | Direction in hypoxia |
|--------------------------------------------|----------|-------|-------------|----------------------|
| TG(44:1)                                   | Positive | TG    | Unsaturated | Up                   |
| TG(58:1)                                   | Positive | TG    | Unsaturated | Up                   |
| SM(d16:1/24:0)                             | Positive | SM    | Unsaturated | Up                   |
| PE(O-16:0/22:6)                            | Negative | PE    | Unsaturated | Up                   |
| PC(18:0/18:0)                              | Positive | PC    | Saturated   | Up                   |
| PC(18:0/18:1)                              | Positive | PC    | Unsaturated | Up                   |
| PC(16:0/18:1)_3                            | Positive | PC    | Unsaturated | Up                   |
| PC(O-16:0/20:3)                            | Positive | PC    | Unsaturated | Up                   |
| PC(28:0); PC(14:0/14:0)  <br>PC(16:0/12:0) | Positive | PC    | Saturated   | Up                   |
| PC(16:0/18:2)                              | Positive | PC    | Unsaturated | Down                 |
| LPI(18:0/0:0)                              | Negative | LPI   | Saturated   | Up                   |
| LPC(24:0/0:0)                              | Positive | LPC   | Saturated   | Up                   |
| LPC(22:0/0:0)                              | Positive | LPC   | Saturated   | Up                   |
| LPC(0:0/18:2)                              | Positive | LPC   | Unsaturated | Up                   |
| LPC(O-18:0/0:0)                            | Positive | LPC   | Saturated   | Up                   |
| LPC(0:0/14:0)                              | Positive | LPC   | Saturated   | Up                   |
| Hexacosanoic acid FA(26:0)                 | Negative | FA    | Saturated   | Up                   |
| Lauric acid FA(12:0)                       | Negative | FA    | Saturated   | Up                   |
| Cer(d18:1/24:0)                            | Positive | Cer   | Unsaturated | Up                   |

Table 12. Core OAC hypoxia-associated lipid signature

#### CORRELATION ANALYSIS

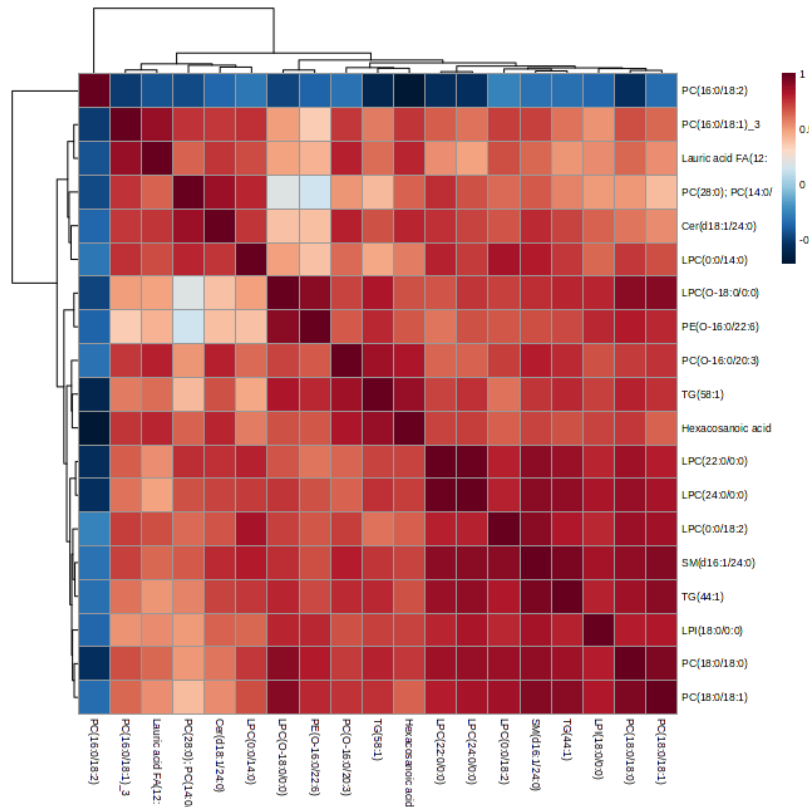


Figure 54. Correlation analysis of core lipids associated with the hypoxic response in OAC cells



#### 4.4.3 Volatilomics

Lipid samples, including 3 blanks, 6 pooled QC and 3 QC dilutions, were then incubated under ambient oxidation conditions for 72 hours. Following this, all samples underwent headspace GC-TOF-MS analysis in one batch. The standardised pipeline developed for VOC biofluids described in Section 4.3.17 was applied for this analysis and the raw code is included in Appendix 10 – 12. A summary of the pipeline was described previously in Figure 48. The data was filtered to remove contaminants and unreproducible, non-linear peaks, and normalized using internal standards to correct for instrument response and DNA concentration to correct for biomass. Pre- and post- filtering/normalization PCA plots are shown in Figure 55, showing an improvement in QC clustering and blank separation. One outlier sample was noted and excluded from further analysis.

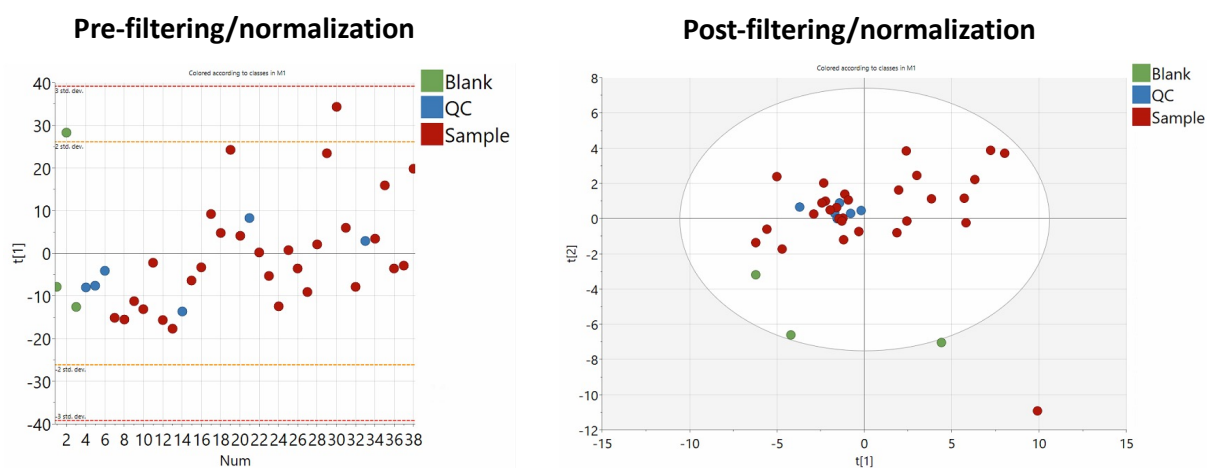
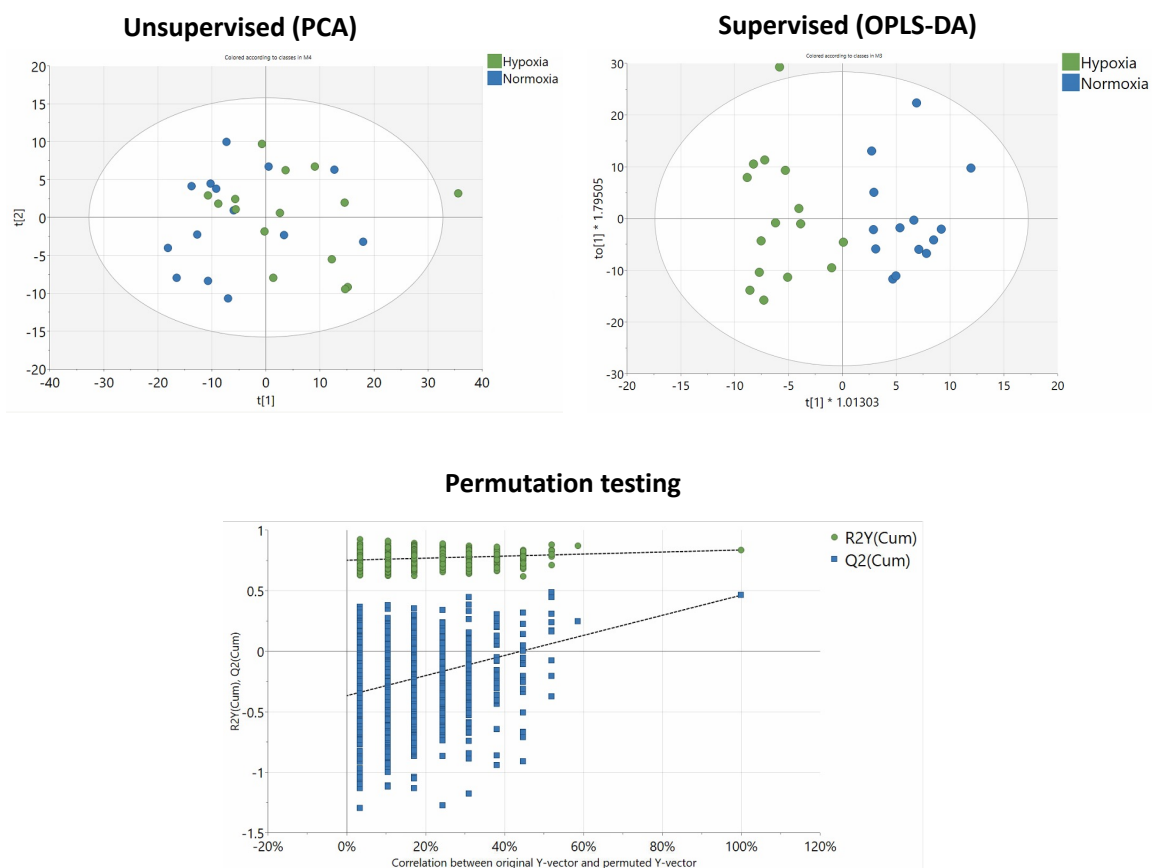


Figure 55. PCA score plots before and after GC-MS data pre-processing

Unsupervised analysis using PCA did not show separation between hypoxic and normoxic samples (Figure 56). Supervised analysis with OPLS-DA showed separation between hypoxic and normoxic groups ( $R^2Y$  0.836 and  $Q^2$  0.463), which was validated from permutation analysis (999 permutations) and CV-ANOVA p value with a p-value of 0.031. VIP analysis identified 33 VOCs with VIP scores  $> 1.35$ .



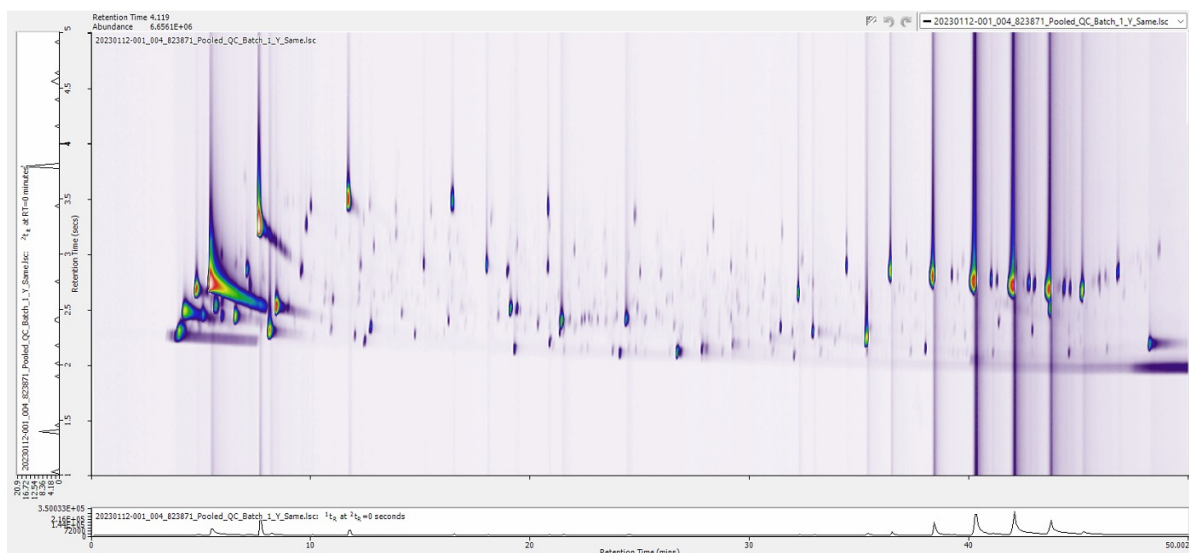
**Figure 56. PCA, OPLS-DA and permutation plots for VOC data**

Further multivariate analysis was performed using ensemble analyses with SVM, PLS-DA and Random Forest algorithms. This analysis identified nine class-separating VOCs including alcohols, ketones, an alkane, and a fatty ester (Table 13). All VOCs were up-regulated in hypoxic samples. Four VOCs had match factors < 800 indicating low confidence with compound identification based on spectral similarity to the NIST database. Analysis of GCxGC-ToF-MS chromatograms from re-collected pooled QC samples did not identify any co-elutions (Figure 57). On univariate analysis, six of these VOCs were significantly increased ( $p < 0.05$ ) in hypoxic samples relative to normoxic samples (Figure 58). ROC analysis showed moderate discriminatory performance of these nine VOCs in class separation (AUC 0.772; 95% CI 0.483 – 0.984).

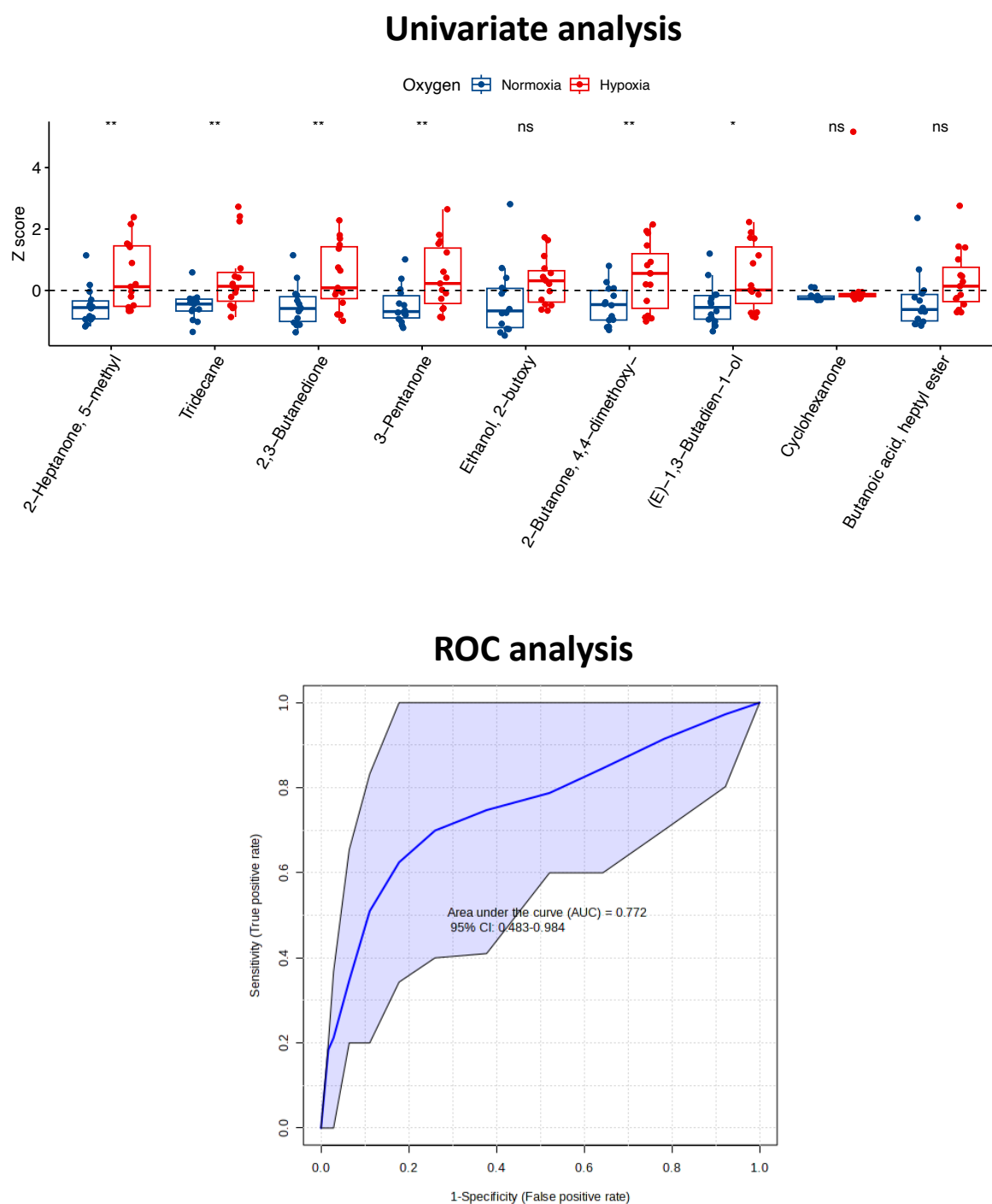
| NIST Library Hit            | Class            | CAS No.    | Retention time | Primary ion | FMF | RMF |
|-----------------------------|------------------|------------|----------------|-------------|-----|-----|
| (E)-1,3-Butadien-1-ol       | Alcohol          | 70411-98-2 | 5.36           | 75          | 627 | 908 |
| 2-Heptanone, 5-methyl       | Ketone           | 18217-12-4 | 5.47           | 43          | 662 | 857 |
| 2,3-Butanedione             | Ketone           | 431-03-8   | 5.82           | 43          | 987 | 987 |
| 3-Pentanone                 | Ketone           | 96-22-0    | 5.96           | 57          | 759 | 820 |
| 2-Butanone, 4,4-dimethoxy-  | Ketone           | 5436-21-5  | 8.34           | 75          | 735 | 735 |
| Ethanol, 2-butoxy           | Alcohol          | 111-76-2   | 10.06          | 57          | 708 | 880 |
| Cyclohexanone               | Ketone           | 108-94-1   | 10.13          | 55          | 920 | 938 |
| Tridecane                   | Alkane           | 629-50-5   | 11.96          | 57          | 691 | 823 |
| Butanoic acid, heptyl ester | Fatty acid ester | 5870-93-9  | 17.70          | 71          | 572 | 743 |

**Table 13. Significant VOCs separating hypoxic classes on multivariate analysis**

Abbreviations: CAS, Chemical Abstracts Service; FMF, forward match factor; NIST, National Institute of Standards and Technology; RMF, reverse match factor.



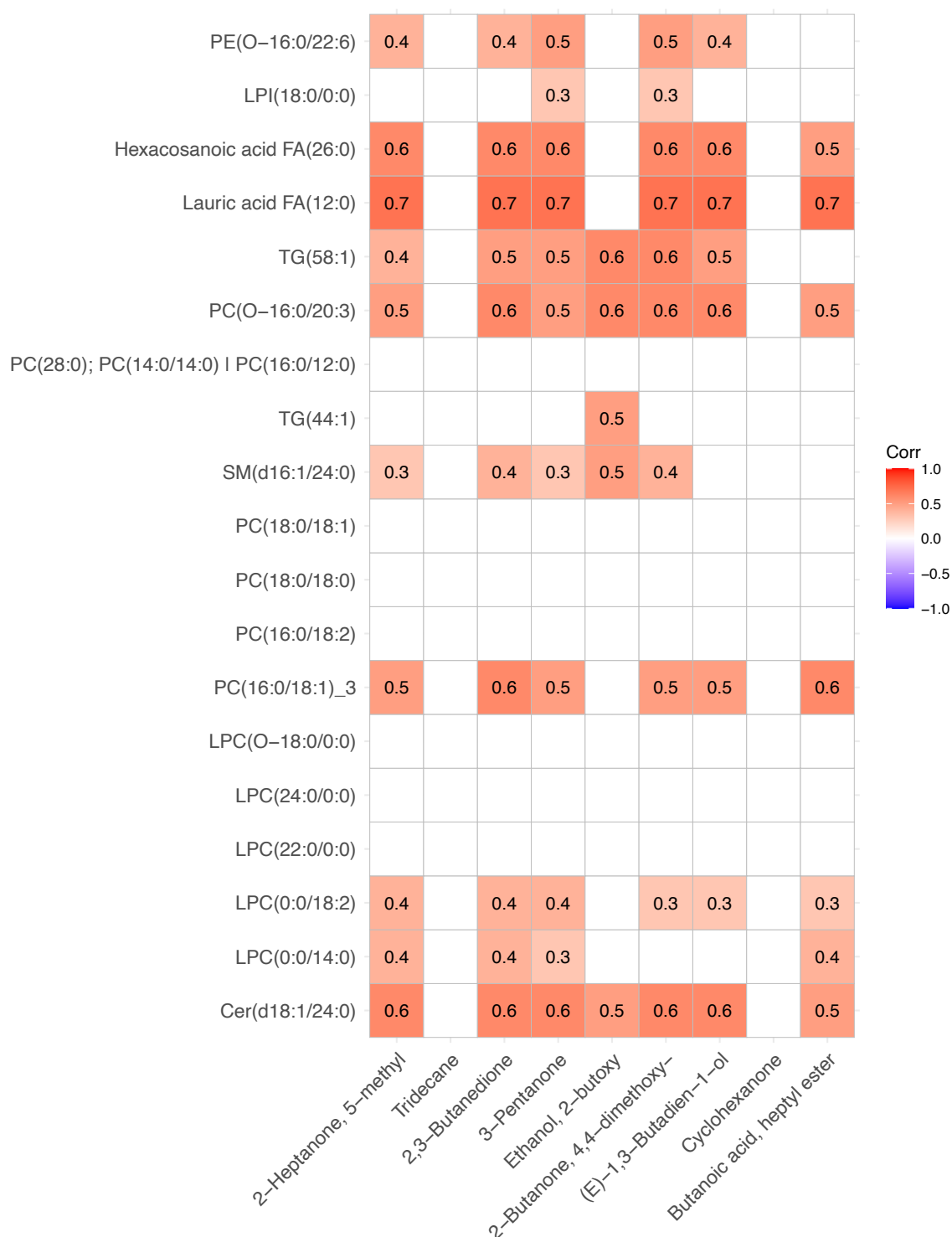
**Figure 57. Representative GCxGC-TOF-MS chromatogram from a pooled QC sample**



**Figure 58. Univariate and ROC curve analysis of VOCs separating hypoxic and normoxic cells**

Finally, correlation analysis was performed to identify significant correlations (Spearman's  $\rho > 0.3$ ,  $p < 0.05$ ) between the core lipid and VOC signatures related to hypoxia in OAC samples (Figure 59). This revealed that 10 lipid metabolites (5 saturated, 5 unsaturated) were correlated with a subset of VOCs. Seven lipids and two VOCs did not show any significant correlation.

## Lipid-VOC correlation analysis

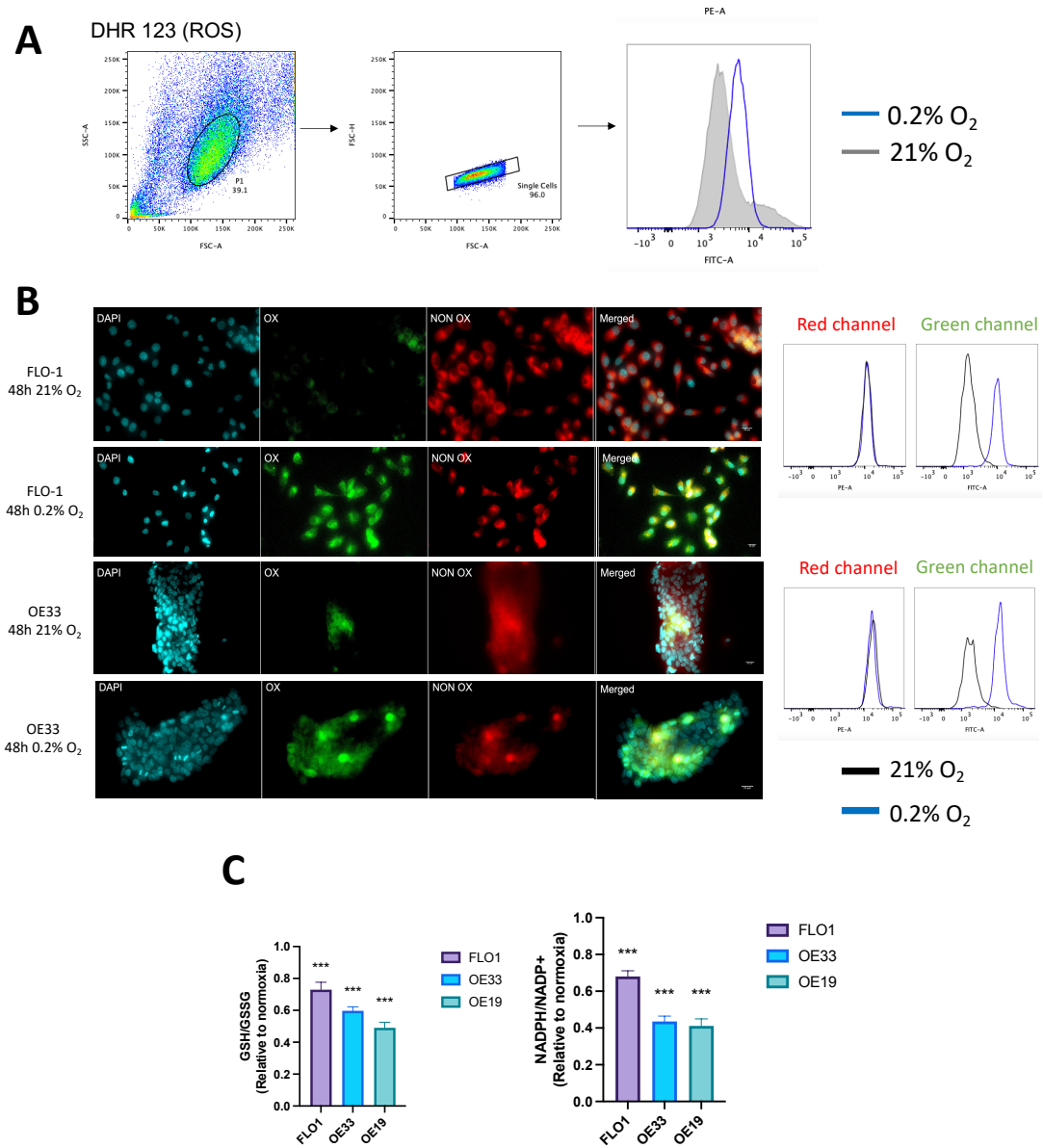


Only significant (unadjusted  $P < 0.05$ ) correlation are shown

Figure 59. Lipid-VOC correlation analysis

#### 4.4.4 Redox analysis

To further assess the metabolic consequences of hypoxia, the redox status of OAC cell lines was assessed (Figure 60). ROS levels measured using DHR increased between 1.5-2-fold in FLO1 cells upon exposure to hypoxia. Examination of FLO1 and OE33 cells using confocal microscopy and flow cytometry after staining with the lipid peroxidation marker BODIPY 581/591 C11 revealed that hypoxia also increased lipid peroxidation. Glutathione (GSH) is a ROS scavenger and is oxidised to GSSG whereas NADPH supplies electrons required for the regeneration of glutathione. The relative ratios of GSH to GSSG and NADPH to NADP<sup>+</sup> can therefore be used as measures of oxidative stress. Hypoxic cells had a significantly decreased antioxidant signature, characterised by lower NADPH/NADP<sup>+</sup> and GSH/GSSG ratios.



**A**, Relative changes in DHR fluorescence (ROS marker) on flow cytometry in FLO1 cells cultured under normoxic and hypoxic conditions. **B**, Relative changes in C11 BODIPY 581/591 (lipid peroxidation marker) fluorescence on confocal microscopy in FLO1 and OE33 cells under normoxic and hypoxic conditions. (Ox = Oxidized lipids; Non-ox = Non-oxidized lipids) **C**, Relative changes in antioxidant co-factor concentration in OAC cell lines.

**Figure 60. Redox state of OAC cells under hypoxia**

## 4.5 Discussion

The purpose of experiments in this Chapter was to explore the interplay between hypoxic transcriptional signalling and metabolic phenotypes in OAC cells.

Hypoxia induced profound metabolic reprogramming in all OAC cells, with induction of gene sets regulating glycolysis and EMT. Fatty acid metabolism was consistently and profoundly down-regulated in all cells which may favour lipid accumulation under hypoxia<sup>155</sup>. Importantly, *HILPDA* and *ANGPTL4* have emerged as major candidates regulating lipid metabolic programming in response to hypoxia OAC. *HILPDA* over-expression has been associated with poor prognosis in several cancers<sup>227</sup>. The protein product (HIG2) promotes neutral lipid droplet formation by inhibiting ATGL-mediated lipolysis<sup>228</sup> and can mediate radio-resistance<sup>229</sup>. In renal cancer, *HILPDA* was found to selectively enrich unsaturated lipids including phosphatidylcholines and phosphatidylethanolamines<sup>149</sup>. The secreted *ANGPTL4* protein is an angiopoietin that regulates lipid metabolism by inhibiting lipoprotein lipase in a PPAR-dependent manner<sup>228,230</sup>. Its expression is essential for tumour growth and mediates resistance in gastric cancer<sup>231</sup>. Therefore, these genes are prime candidates for genetic perturbation and stable isotope tracing studies to investigate their role in regulating the VOC response to hypoxia in OAC.

Interestingly, the transcriptional response to hypoxia was less pronounced in FLO1 compared to OE33 and OE19. FLO1 has a different biological phenotype, displaying mesenchymal markers (vimentin) compared to OE33 and OE19 which express epithelial markers (E-cadherin)<sup>191</sup>. This may also be artefactual and related to nutrient availability and pH, as FLO1 required a different culture medium (DMEM with high glucose) compared to OE33 and OE19. It was noted that the pH of the DMEM medium rarely became acidic. It is known that many genes related to hypoxia overlap with acidosis, and acid can directly activate HIF1- $\alpha$  signalling independent of hypoxia. An intriguing area of future research would be to explore the cross-talk between hypoxic and acid signalling and their relationship to lipid/volatile phenotypes.

The lipidomic data suggested a regulated and stable response to hypoxia in OAC, characterized by an enrichment of both saturated and unsaturated lipids which were predominantly phosphatidylcholines. This is a class of glycerophospholipid composed of a phosphate group, choline group and two fatty acyl chains. In prostate cancer, hypoxia was shown to induce choline kinase via HIF-1 which directly binds the *ChoK* promoter<sup>232</sup>. Phosphatidylcholines play a crucial role in how cancer



cells manage stress and may contribute to the development of drug resistance<sup>233</sup>. Although this response could be an essential mechanism for the survival of OAC cells, the biological motivation for this stereotyped lipid phenotype is an ongoing line of investigation.

Analysis of the GC-MS data also revealed a volatile signature which partially correlated with the lipid signature, indicating direct or indirect routes to production. The lower  $Q^2$  of OPLS-DA and wide confidence interval of the AUROC indicated the prediction model was less reliable than the lipidomic prediction model and a larger difference between  $R^2Y$  and  $Q^2$  indicated a degree of overfitting. This was likely due to the small dataset and wider variance in gas phase analytes between replicates compared to lipid metabolites which are inherently more stable. Similar to Chapter 3, since there is a relative lack of detailed research into the metabolism of VOCs in humans, the relevance of the detected metabolites is unknown. Among the VOCs identified from untargeted analysis, only two (butane-2,3-dione and cyclohexanone) have been previously reported in the human volatilome on searching the Human Breathomics Database<sup>103</sup> in articles unrelated to cancer. Future work will need to confirm the identity of compounds identified on the untargeted analysis using authentic standards and validate their performance in a larger population using a targeted method with absolute quantification. A targeted analysis was not performed as part of this Experimental work as calibration standards could not be loaded directly onto HiSorb© probes. Ongoing method development on a headspace-to-TD tube method may address this challenge and be applicable in future experiments.

The redox findings suggest that chronic hypoxia in OAC generates redox dysfunction and oxidative stress. This is consistent with experimental data in other cancers which show that hypoxia paradoxically increases ROS and lipid peroxidation<sup>138,142,234</sup>. It would be of value to determine whether this occurs *in vivo*, for example, by co-localizing oxidative stress markers (e.g. 4-HNE) with pimonidazole-stained tissue. Tumour cells may also adapt to chronic hypoxia by up-regulating survival pathways such as *NRF2*<sup>235</sup>. This could be explored further by quantifying ROS levels in OAC cell lines cultured under long term-hypoxia to determine whether these responses persist.

A key strength of this Chapter is the analytical pipeline developed to automate and streamline complex GC-MS data pre-processing. It will be of interest to evaluate the pooled composite method against (or together with), existing methods in peak identification and integration, including the tile-based analysis used earlier. Future work will aim to incorporate a peak alignment algorithm into the code based on batch calibration standard mixes. This may be superior to software-based alignment which often fails when retention times have shifted significantly.

There are several limitations to this data. Lipid and VOC signatures will require external validation with additional cell models as only three cell lines have been used. Immortalized cell lines may no longer resemble the tumour from which they originate and may not serve as accurate models of metabolic responses *in vivo*. Using more advanced models in the future which recapitulate tumour heterogeneity (e.g. patient-derived organoids) may resemble the *in vivo* phenotype more accurately. A weakness of the applied liposome model is an inability to distinguish between cellular compartments and the availability of lipids for peroxidation. Hypoxia-induced lipid droplet formation may serve to maintain lipid homeostasis but also sequester reactive PUFAs to prevent lipid peroxidation. Liberating these lipids through the extraction process therefore may not faithfully recapitulate true cellular biochemistry. Additionally, the current lipidomic workflow is also limited as only a subset of the dataset was annotated by the *peakPanther* software which limits the conclusions of this data. For example, the correlation of VOCs with saturated lipids (which cannot generate volatiles) are likely to indicate that the saturated lipids have a strong correlation with other unsaturated lipids not reported in the annotated dataset. This prevents accurate and quantitative estimates of the global unsaturation index of the cell as a significant fraction of the lipidome remains unannotated.

In summary, these experiments suggest that hypoxia can induce lipid reprogramming with a resultant volatile signature in OAC cells *in vitro*. An immediate extension of this work would be to perform genetic perturbation and stable isotope tracing experiments to unpick the mechanisms of hypoxia-associated VOC signatures, in addition to ongoing validation work in more representative cancer models.

## **CHAPTER 5 – TRANSLATIONAL RELEVANCE OF A HYPOXIC VOC SIGNATURE**

## Summary

The aim of this final Experimental Section was to identify and validate candidate anti-cancer therapeutics with increased sensitivity against hypoxic OAC cells. This would provide the rationale to evaluate the predictive performance of VOC hypoxia biomarkers in a clinical trial. An *in silico* screening approach was first applied to OAC gene expression datasets, which predicted sensitivity of hypoxic tumours to PARP inhibition. An analysis of cell line and clinical expression datasets revealed an association between increased hypoxia and down-regulation of DNA damage repair pathways. Western blot analysis confirmed down-regulation of homologous recombination repair proteins in OAC cells under chronic hypoxia. Drug testing experiments confirmed that OAC cells had reduced clonogenic survival after PARP inhibitor treatment and this was associated with increased DNA damage. The incorporation of breath analysis into an active clinical trial of olaparib (SOLAR) is discussed.

## 5.1 Methodological rationale

### 5.1.1 Hypoxia and response to therapy

Hypoxia can contribute to both resistance and sensitivity to anticancer therapies. Hypoxia-targeting therapeutics (HTT) and their mechanisms of action have been reviewed in 1.2.2. A clinical trial of a suitable HTT in OAC was therefore desired in order to establish the proof-of-principle of VOC monitoring of HTT.

Bioinformatic techniques were used first to identify candidate HTT. This involved imputing drug response from a large database of gene expression and drug response data from cell lines, into clinical expression datasets stratified by hypoxic status. This was followed by drug sensitivity experiments on selected drugs under hypoxic conditions to validate the *in silico* findings. The primary candidate for evaluation was a class of therapeutics known as PARP inhibitors.

### 5.1.2 PARP inhibitors and contextual synthetic lethality

DNA damage in cancer cells can be induced by both endogenous and exogenous factors (e.g. ROS, toxins, radiation, aldehydes). Single-strand DNA breaks are detected by the N-terminal domains of PARP resulting in DNA binding (PARP trapping) and enzyme activation<sup>236,237</sup>. PARP1 initiates the transfer of poly(ADP-ribose) (PAR) chains from NAD<sup>+</sup> to target proteins, including PARP1 itself (auto-PARylation). This recruits DNA damage repair proteins (e.g. XRCC1) to DNA for the repair of single-strand breaks.

PARP inhibition (e.g. olaparib, rucaparib, veliparib) results in the accumulation of single-strand breaks, which can lead to replication fork collapse and the formation of double-strand breaks<sup>238</sup>. These require homologous recombination (HR) repair for repair, a complex process involving many proteins including RAD51, BRCA1 and BRCA2<sup>239</sup>. This is the least error prone form of double-strand break DNA repair and requires a template sister chromatid in G2/S phase of the cell cycle. An alternative pathway is the non-homologous end joining (NHEJ) which is error prone but active at all stages of the cell cycle<sup>240</sup>. Mutations in HR genes (e.g. BRCA1/2) leads to a functional loss of HR and exquisite sensitivity to PARP inhibition (synthetic lethality)<sup>241</sup>. Combining PARP inhibition with other DNA damage repair inhibitors or cytotoxic agents has shown synergism in pre-clinical studies<sup>242</sup>.

Experimental studies have shown that chronic severe hypoxia can downregulate HR proteins (RAD51, BRCA1 and BRCA2) leading to a functional loss of HR<sup>36,180,181</sup>. As a consequence, HR-defective hypoxic can be sensitized to PARP inhibition, described as ‘contextual’ synthetic lethality (Figure 61)<sup>38</sup>. Alternative studies, however, have reported conflicting data, suggesting that mild hypoxia may increase, not decrease, resistance of cancer cells to PARP inhibitors<sup>243</sup>. Therefore the response to PARP inhibition under hypoxia may be context- and cell-type specific and warranted further investigation in OAC.

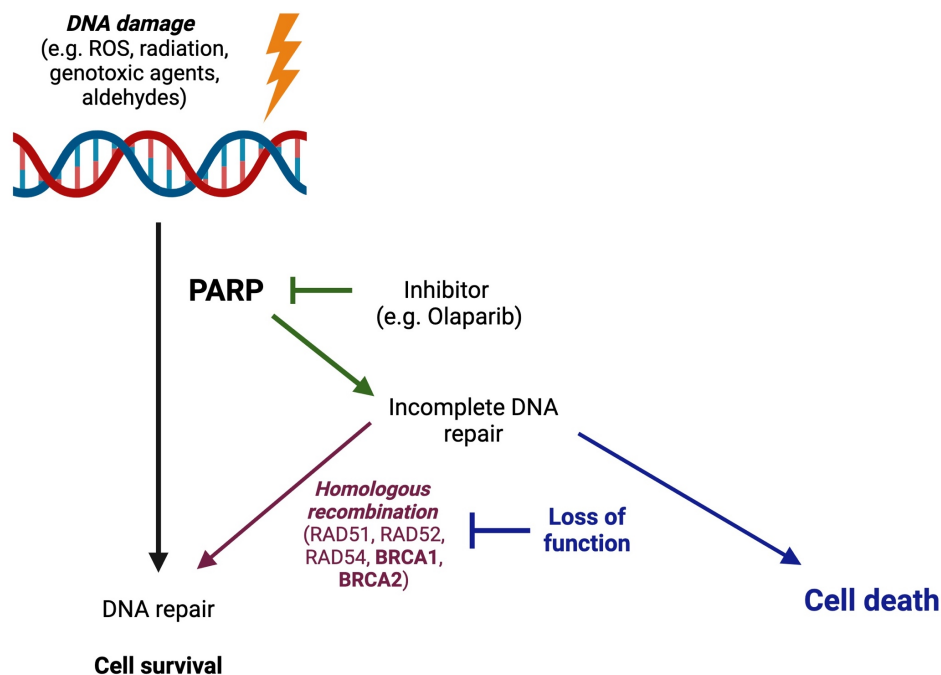


Figure 61. PARP inhibitors and synthetic lethality

## 5.2 Aims

The objective of this chapter was to identify and validate hypoxia-sensitive drugs for a VOC-stratified drug trial. The specific aims of this section were to:

1. Select drugs which selectively target hypoxic OAC cells using an *in silico* approach.
2. Validate the predicted response of hypoxic OAC cells to candidate drugs.
3. Establish VOC monitoring in a clinical trial of hypoxia-targeted therapy.

## 5.3 Methods

### 5.3.1 In silico prediction of drug response

Gene expression and drug response data were downloaded from the Genomics of Drug Sensitivity in Cancer (GDSC) database ( <https://www.cancerrxgene.org/> ). This dataset containing drug response data for 519 compounds in 970 cell lines. Expression data for OAC patients are described in Section 2.3.9. The R package *oncoPredict* was used to impute drug response in patient samples using expression and drug response data from cell lines. The R package *gsva* was used to generate sample-wise enrichment scores for the OAC hypoxia-associated gene signature and MSigDB v7.2 KEGG pathways (homologous recombination, HR and non-homologous end joining, NHEJ). The correlation between the imputed drug response and hypoxia-associated mRNA expression (GSVA score) was calculated using Spearman's rank correlation, considering Spearman's  $\rho > +0.2$  and FDR < 0.05 as significant and demonstrating potential drug sensitivity in hypoxic cells.

### 5.3.2 Cell culture

FLO1, OE19 and OE33 cells were cultured as previously described (Chapter 2.3.4). COSMIC data<sup>190</sup> has confirmed these cell lines are wild-type for BRCA1, BRCA2 and ATM. For RNA extraction and immunoblot, cells were seeded on 6-well plates with three biological replicates. For flow cytometry assays, cells were seeded on 10 cm dishes with three biological replicates. For fluorescence microscopy assays, cells were seeded in 8-well chamber plates, one plate per condition (Nunc™ Lab-Tek™ II Chamber Slide™ System, Thermo Fisher Scientific). For sulforhodamine B assays, cells were seeded in 96-well plates with six biological replicates. Cells were seeded at densities based on growth curves to achieve a target confluency of 70-80% after 72 hours incubation under normoxia. Seeding parameters for clonogenic assays are described separately below.

### 5.3.3 Therapeutics

Olaparib, rucaparib, veliparib and AZD6738 were purchased from Selleckchem. Drugs were re-constituted in DMSO, aliquoted and stored at -80°C. Each aliquot was thawed for use once and discarded after use. All drug treatments included an untreated sample, DMSO blank and concentrations ranging from 0.1µM to 10µM, reflecting expected tumour tissue concentrations



(normal to high dose oral formulations) from pharmacokinetic analyses<sup>244,245</sup>. PARP enzyme activity was assessed using the Universal Chemiluminescent PARP Assay Kit (Trevigen).

#### **5.3.4 Hypoxia treatment**

After incubation overnight to allow adherence, plates were transferred to an H35 Hypoxystation (Don Whitley Scientific). The media was changed, and cells were incubated at 0.2% O<sub>2</sub> for 72 hours. Control plates were incubated in a normoxic incubator. All media and PBS used for the processing of hypoxia-treated cells were initially 'de-gassed' by incubating in the hypoxic incubator overnight.

#### **5.3.5 Immunoblot**

The method for Western blot has been previously described (Chapter 2.3.6). Primary antibodies used at 1:1000 dilution were:  $\alpha$ -HIF-1 $\alpha$  (BD Biosciences #610958),  $\alpha$ -BRCA1 (Santa Cruz #sc-6954),  $\alpha$ -BRCA2 (Merck #OP95),  $\alpha$ -RAD51 (Merck #PC130),  $\alpha$ -Phospho-Histone H2A.X (Ser139) (20E3) (Cell Signaling Technology #9718) and  $\alpha$ -PARP1 (Abcam, #ab32138). Secondary antibodies used at 1:2000 dilution were:  $\alpha$ - $\beta$ -tubulin (CST #2146) and  $\alpha$ -cyclophilin B (CST #43603).

#### **5.3.6 RNA extraction and sequencing**

The methods for RNA extraction and sequencing have been previously described (Chapter 2.3.7 and 2.3.8).

#### **5.3.7 Differential gene expression analysis**

The R package *DESeq2* was used to process RNA-seq counts data for differential expression analysis and identify hypoxia up- and down-regulated genes in each OAC cell line. The Benjamini–Hochberg method was used for multiple testing correction with P value cut-off specified at 0.05.

#### **5.3.8 Cell cycle analysis (flow cytometry)**

Cell cycle analysis was performed to evaluate OAC cell models and their proliferative potential under severe hypoxia, since progression through mitosis is required for PARP inhibitor cytotoxicity<sup>246</sup>. At the indicated timepoint, bromodeoxyuridine (BrdU) was added to the culture medium to a final

concentration of 20 $\mu$ M for 2 hours in the dark. Cells were fixed with 70% ethanol followed by treatment with 2N hydrochloric acid for 30 minutes and 0.1M sodium tetraborate for 30 minutes. BrdU staining was performed with  $\alpha$ -BrdU (1:50) (Sigma Aldrich) for 1 hour at room temperature followed by a 30-minute incubation with goat anti-mouse IgG FITC (Sigma Aldrich). After a final wash, cells were re-suspended in 500 $\mu$ l PBS containing 10 $\mu$ g/ml RNase A and 20 $\mu$ g/ml propidium iodide. The cell suspension was transferred to a flow cytometry tube via a 40 $\mu$ m nylon mesh prior to flow cytometry. Flow cytometry was performed using a FACS Canto II instrument (BD). Cells were gated according to forward and side scatter values, including a doublet gate to exclude cell clusters. Data was processed using FlowJo™ version 10.

### **5.3.9 Immunofluorescence**

Cells were treated and fixed using 4% paraformaldehyde in PBS. Cell membranes were permeabilised with 0.1% Triton X-100 for 10 min, and blocked with 1% BSA, 1% TBS, 0.1% Tween-20 for 30 min. Cells were incubated with anti-phospho-H2AX (Ser 139) (20E3) primary antibody (Cell Signaling #9718) with 1:400 dilution in 1% BSA, 1% TBS, 0.1% Tween-20 for 2 h, followed by anti-Rabbit IgG secondary antibody (Sigma Aldrich #F0382) with 1:80 dilution in 1% BSA for 2 h in darkness. Cells were washed with TBS and counterstained with 0.67  $\mu$ g/mL of DAPI stain. Images of 3 random areas (420 $\times$ 560  $\mu$ m) in each well were captured using EVOS FL Cell Imaging System (Thermo Fisher Scientific) at 20X magnification (70% brightness) under DAPI and GFP channels.

### **5.3.10 Sulforhodamine B (SRB) assay**

At the specified timepoint, 100 $\mu$ l of ice-cold 10% trichloroacetic acid was added to each well and plates were incubated on ice for 1 hour. Plates were washed four times with slow-running tap water and left to dry. To each well, 100 $\mu$ l of 0.057% SRB was added and plates were left for 30 minutes at room temperature. Plates were rinsed four times with 1% acetic acid to remove unbound dye and left to dry at room temperature. Finally, 200 $\mu$ l of 10mM Tris base solution was added, and plates were shaken on a gyratory shaker for 5 minutes to solubilize the protein-bound dye. The optical density (OD) was measured on a spectrophotometer plate reader at 510nm.

### 5.3.11 Clonogenic assay

Cells were synchronized in G1 phase using the mitotic shake off technique<sup>247</sup> prior to seeding. Cells were seeded in 6-well plates at a density of 1000 cells per well and allowed to attach for 6 hours. At this point, medium was removed, and cells were cultured with media containing the indicated doses of PARP inhibitor under normoxia or hypoxia for 72 hours, followed by a 10-14 day culture in the absence of an inhibitor. Cell media was changed every 3 days. At the end of the experiment, the medium was removed, and cells were fixed with 400 $\mu$ l 100% methanol for 20 minutes at room temperature. The methanol was then removed and replaced with 1ml of 0.5% crystal violet in 25% methanol for 5 minutes. After staining, the crystal violet solution was removed, and plates were washed with deionized water and left to dry. Plates were imaged using a scanner and colonies (>50 cells) were counted using the Particle Analyzer function of ImageJ.

## 5.4 Results

### 5.4.1 In silico analysis

To identify candidate hypoxia-sensitive therapeutics, a computational screening approach was used. The *calcPhenotype* function of the R package *oncoPredict*, builds a ridge regression model built from cell line gene expression and drug screening experiments applied to tumour gene expression data to impute predicted clinical responses (IC50). A similar approach has been used successfully to identify hypoxia-sensitive therapeutics previously<sup>169</sup>. Cell line drug response data from GDSC1 and GDSC2 datasets was compiled to generate a dataset of 519 unique drugs. Correlation analysis was then performed between OAC-Hyp hypoxia scores generated in patient OAC gene expression datasets and imputed drug response. Significantly negative correlation between hypoxia score and predicted drug sensitivity (IC50) was interpreted as indicating a candidate hypoxia-sensitive therapeutic. In all three OAC datasets, hypoxic tumours had predicted sensitivity to PARP inhibitors, with olaparib identified in two out of three datasets (Table 14 and Figure 62). Targeted gene set analysis using GSVA revealed suppression of key DNA repair pathways (homologous recombination and non-homologous end joining pathway activity) related to PARP inhibitor sensitivity upon exposure of cells to hypoxia (0.2% O<sub>2</sub>) (Figure 63). This was also identified in TCGA-ESCA (OAC) stratified by the OAC-Hyp signature (Figure 64).

| Database          | GDSC drug   | Spearman's $\rho$ | Unadjusted P value    | Adjusted P value      |
|-------------------|-------------|-------------------|-----------------------|-----------------------|
| TCGA-ESCA (Adeno) | PARP 9495   | -0.35             | $8.94 \times 10^{-4}$ | $1.05 \times 10^{-2}$ |
| Belfast           | Olaparib    | -0.35             | $2.00 \times 10^{-9}$ | $1.20 \times 10^{-8}$ |
| Belfast           | PARP 9482   | -0.24             | $4.82 \times 10^{-5}$ | $1.02 \times 10^{-4}$ |
| OCCAMS            | Olaparib    | -0.27             | $6.47 \times 10^{-6}$ | $1.64 \times 10^{-5}$ |
| OCCAMS            | Talazoparib | -0.19             | $2.18 \times 10^{-3}$ | $3.41 \times 10^{-3}$ |
| OCCAMS            | PARP 9482   | -0.14             | $2.33 \times 10^{-2}$ | $4.12 \times 10^{-2}$ |

**Table 14. Significant hypoxia score-imputed drug response correlations for PARP inhibitors**

This analysis also identified 12 hypoxia-sensitive drugs ( $p < 0$ , FDR  $p < 0.05$ ) and 19 hypoxia-resistant drugs ( $p > 0$ , FDR  $p < 0.05$ ) common to all datasets (Appendix 14). Hypoxia-sensitive drugs included agents targeting genome integrity (KU-55933), protein stability (MG-132, bortezomib), EK-MAPK signalling (refametinib) and RTK signalling (AZD1332, Dasatinib, EphB4).

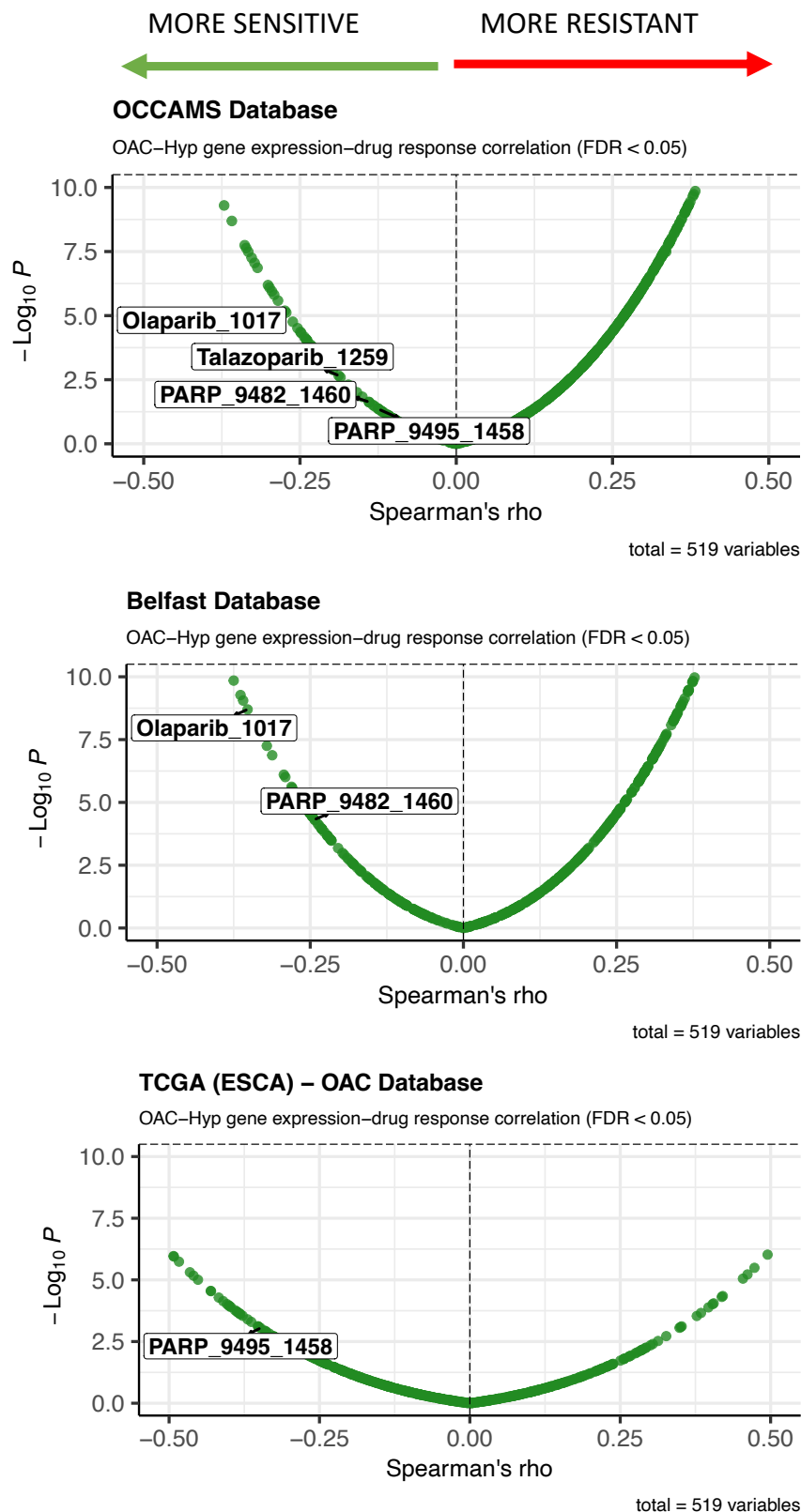


Figure 62. Volcano plots showing hypoxia score-imputed drug response correlations in OAC datasets

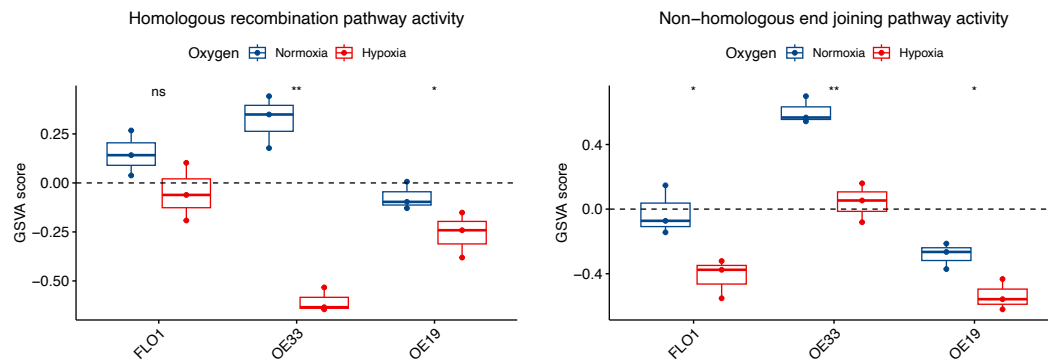


Figure 63. DNA repair pathway activity in cell lines under hypoxia

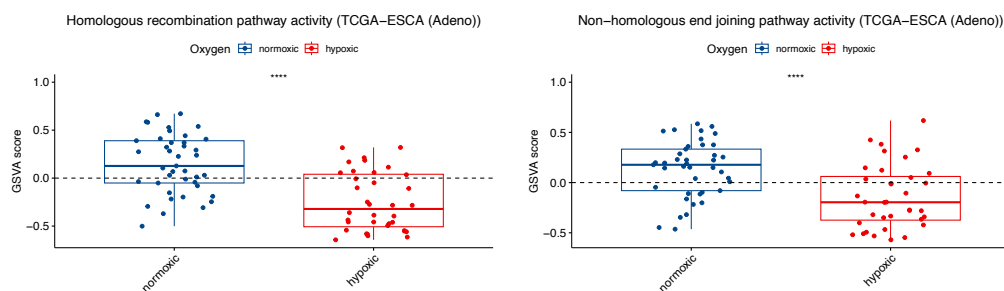


Figure 64. DNA repair pathway in TCGA-ESCA (OAC) stratified by the OAC-Hyp gene signature

#### 5.4.2 Evaluation of PARP inhibitors in vitro

To further explore the findings from the *in silico* analysis, the expression of homologous recombination proteins in OAC cell lines cultured under hypoxic conditions was evaluated using Western blot (Figure 65). This revealed that BRCA2 and BRCA1 were down-regulated in all OAC cell lines, and RAD51 was down-regulated in 2/3 cell lines by hypoxia.

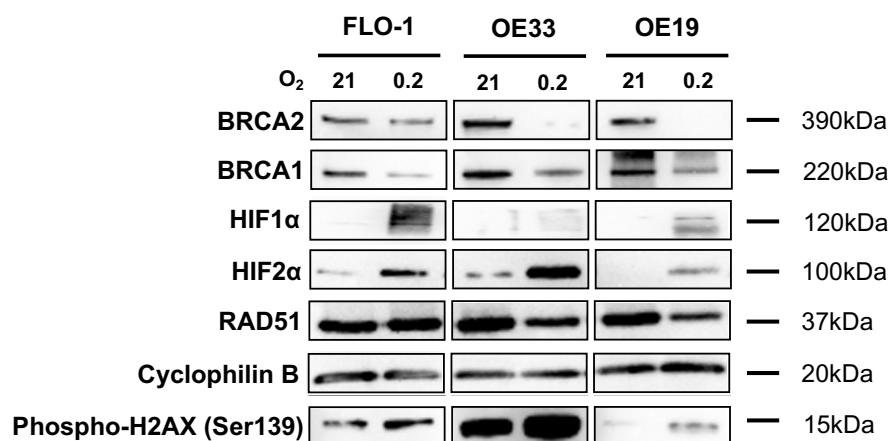
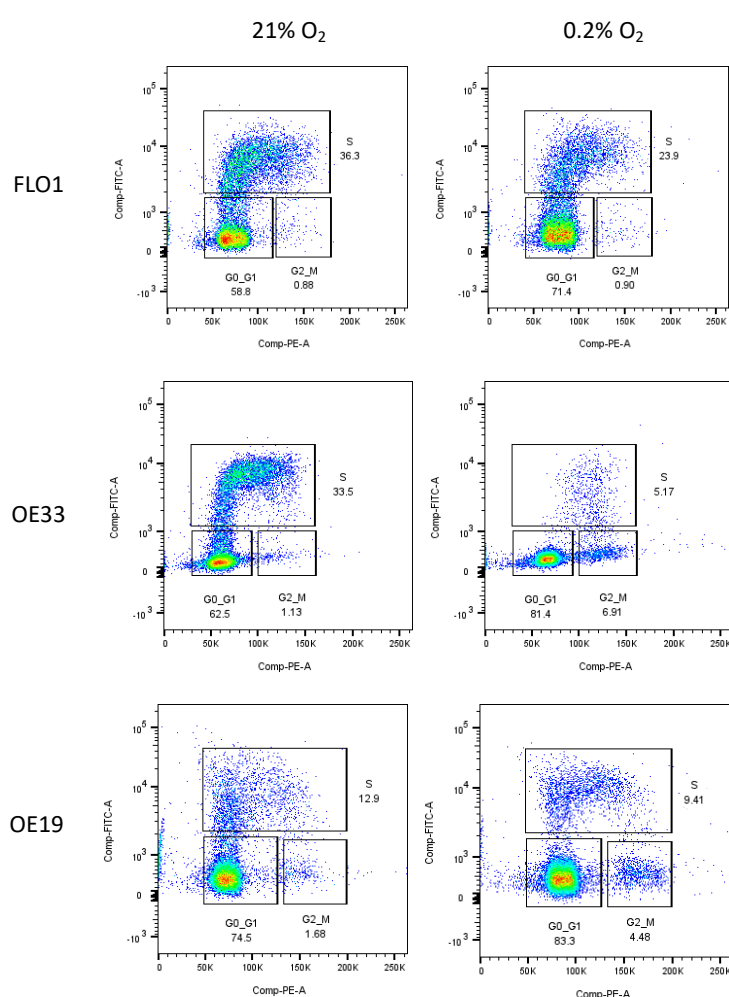


Figure 65. Western blot analysis of key DNA repair proteins under hypoxia in OAC cells

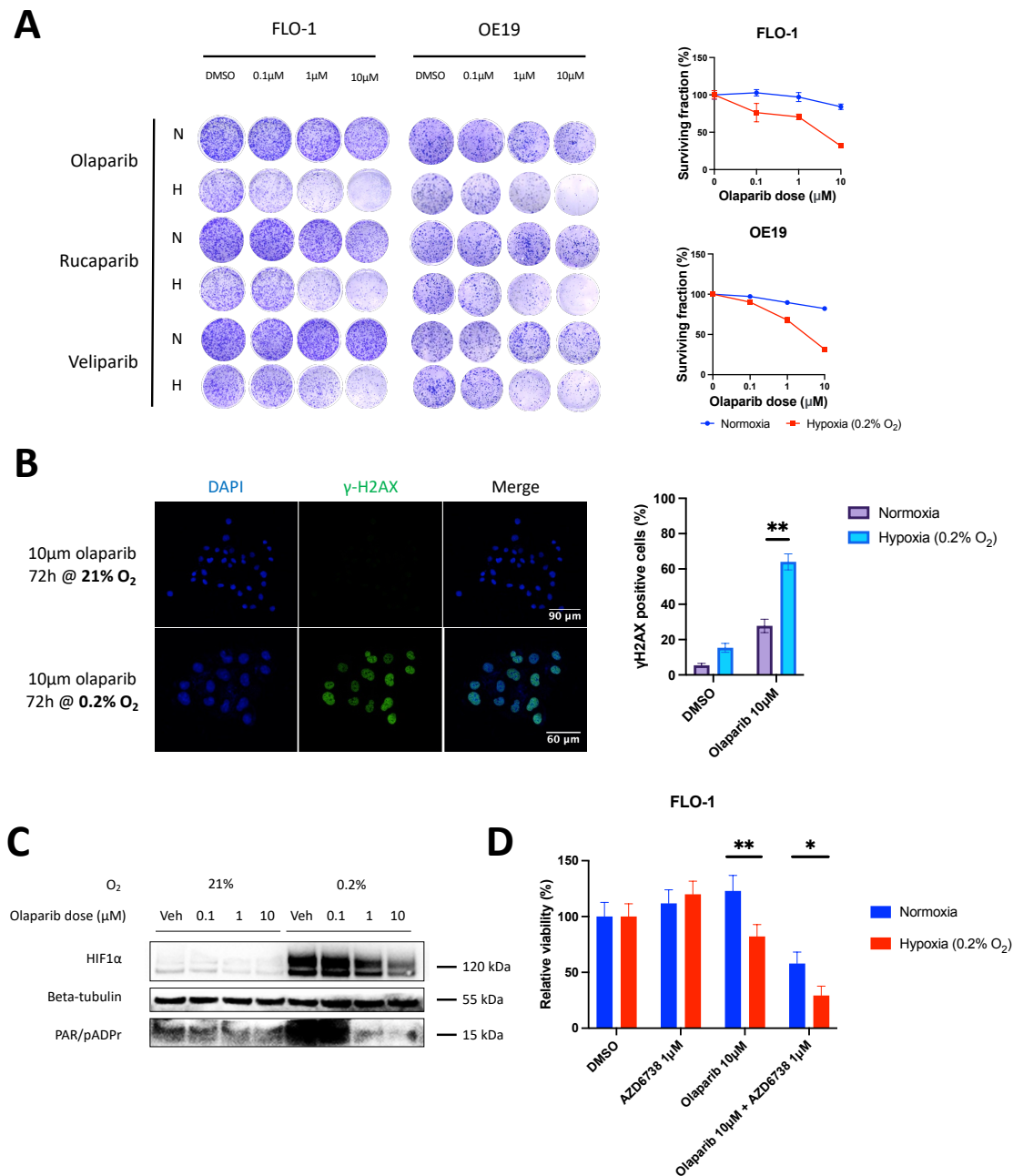
Hypoxia can induce G1/S phase arrest in non-transformed cells<sup>248</sup>, but many cancer cells can sustain replication even under severe hypoxia. Since PARP inhibitors only target actively proliferating cells<sup>246</sup>, cell cycle analysis with BrDU-PI was performed to evaluate whether OAC cells continued to replicate after 72 hours of hypoxia (0.2% O<sub>2</sub>). The data suggested that OE33 underwent replication arrest at 72 hours, indicated by the marked reduction in the number of BrDU+ (S phase) cells and increase in G1/S phase ratio. In contrast, FLO1 and OE19 cells continued to sustain proliferation albeit at a slower rate, indicated by the only slight increase in G1/S phase ratio and reduction in BrDU+ cells. As OE33 cells did not form colonies and underwent replication arrest under hypoxia, this cell line was not selected for further drug testing.



**Figure 66. Cell cycle analysis using BrDU-PI of OAC cells under normoxic and hypoxic conditions**

FLO1 and OE19 cells then underwent drug testing with three PARP inhibitors under clinical evaluation. Clonogenic survival was reduced in FLO-1 and OE19 cells treated with all 3 PARP inhibitors under hypoxia (Figure 67A) and this was associated with increased DNA damage ( $\gamma$ -H2AX foci) (Figure 67B).

Western blot analysis confirmed that PAR polymers were reduced upon treatment with olaparib (Figure 67C). Olaparib was found to be synergistic with an ATR inhibitor but this combination markedly increased cytotoxicity, reducing viability of normoxic cells to 55% (Figure 67D).



**A**, Clonogenic assay of OAC cell lines treated with PARP inhibitors under normoxia or hypoxia (0.2% O<sub>2</sub>) for 72 hours, followed by culture without the inhibitor for 10-14 days. **B**, Confocal microscopy of nuclear H2AX staining in FLO1 cells after treatment with olaparib under normoxia and hypoxia. **C**, Western blot of PAR polymers after olaparib treatment in FLO1. **D**, SRB assay of FLO1 cells treated with olaparib and an ATR inhibitor for 72 hours under normoxia and hypoxia.

**Figure 67. Evaluation of PARP inhibitors in OAC cells under normoxic and hypoxic conditions**



## 5.5 Discussion

In this Chapter, it was discovered that hypoxia confers sensitivity of OAC cells to PARP inhibition, likely through transcriptional suppression of homologous recombination. Bioinformatic and *in vitro* analysis confirmed that DNA repair pathways are suppressed in hypoxic OAC cells. This data is consistent with previous *in vitro* work which showed hypoxia down-regulates homologous recombination, sensitizing cancer cells to PARP inhibitors through contextual synthetic lethality<sup>38</sup>. Additional evidence has emerged in clinical trials, where, for example, olaparib efficacy improved when combined with an anti-angiogenesis drug that can induce hypoxia<sup>249</sup>.

Genotoxic aldehydes can induce DNA damage in OAC<sup>125</sup> and may also play a direct role in modulating the response to PARP inhibition. Exogenous aldehydes have been shown to lead to a selective dose-dependent depletion of BRCA2 by activating proteasomal degradation<sup>250</sup>. A loss of HR proteins was also found to make cancer cells susceptible to exogenous aldehydes and aldehyde dehydrogenase inhibition<sup>251</sup>. Given OAC cells have an inherent lack of aldehyde detoxification and hypoxia has been shown to enrich endogenous aldehydes *in vitro*<sup>125</sup>, this requires further exploration.

A key limitation is understanding how these findings will relate to clinical practice. Resistance mechanisms to PARP inhibitors are complex and an area of active research, occurring even in *BRCA2/BRCA1* mutant tumours which are infrequent in OAC<sup>252</sup>. Inadequate tumour perfusion is a limiting factor common to all anti-cancer therapeutics including PARP inhibitors. One potential solution may be hypoxia-prodrug PARP inhibitors which could increase the effective dose in hypoxic tumour tissue while minimizing systemic toxicity<sup>253</sup>. Clinical and translational trials will be required to explore these findings further.

## 5.6 SOLAR study

The SOLAR trial<sup>254</sup> is a phase II single-arm study of single agent olaparib in the treatment of advanced OG cancer (IRAS project ID: 234206). The Sponsor is Royal Marsden NHS Foundation Trust (RMH), and Principal Investigator is Dr Naureen Starling. The full trial protocol is available on ClinicalTrials.gov (ID: NCT03829345). This study was selected for an opportunistic 'proof-of-principle' study of VOC-based therapeutic monitoring, integrated with a comprehensive package of potential predictive biomarkers.

Patients with locally advanced or metastatic OG cancer that has progressed on first- or second-line treatment are eligible for this trial and are recruited at two RMH sites (Chelsea and Sutton). Participants (n = 27) will be administered olaparib orally twice daily (BD) at 300mg continuously for each 28-day cycle. Response will be evaluated with PET-CT at the end of treatment, with  $\geq 18.7\%$  patients expected to achieve disease control (data from Stage I of SOLAR). The primary outcome is the disease-control rate at 8-weeks defined by the RECIST v1.1 criteria<sup>255</sup>. Secondary outcomes are: overall response rate, duration of response, overall survival, progression-free survival, time to radiological progression and safety. Serial breath analysis (Figure 68) using the standardised techniques described previously (Chapter 3.3) is performed at screening (pre-treatment), 2-, 4- and 8- week encounters (details of amendment AMD2103/15 for breath collection included in Appendix 15). Multiple candidate molecular biomarkers of response are also being evaluated during this study, including RNA-seq and whole exome sequencing. Secondary analyses will include: (i) Comparison of the model against other evaluated biomarkers (whole exome, RNA-seq etc.) and (ii) Sensitivity analysis to evaluate the robustness of the model, which will help inform further validation studies. The trial is due for completion in 2024.

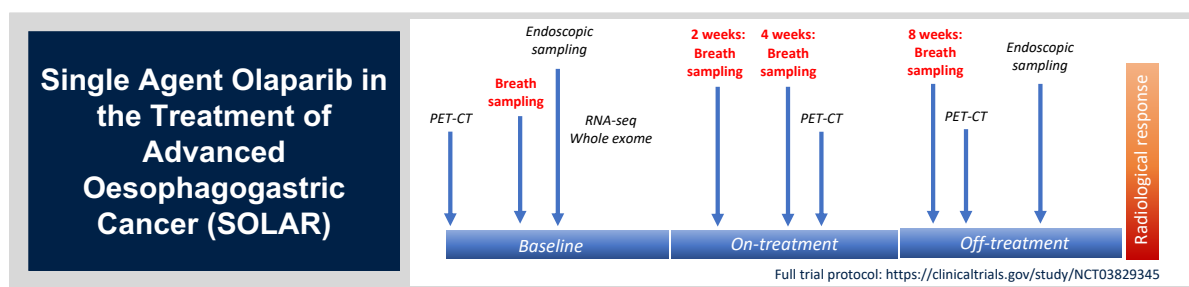


Figure 68. Summary of the SOLAR trial protocol

## **CHAPTER 6 – THESIS CONCLUSIONS AND FUTURE WORK**

## 6.1 Conceptual gains from this thesis

The specific research gains can be subdivided according to the strategic rationale for the project, set out in Chapter 1.

*Hypoxic gene expression defines biological subgroups of OAC.* An initial systematic literature review suggested that hypoxia was clinically relevant in OAC but there was a scarcity of research on the topic. A novel, OAC-specific, hypoxia-associated gene signature was then developed to stratify OAC patients by hypoxic status. This was developed by identifying a core set of hypoxia-regulated genes from OAC cell lines with refinement using patient expression data. Hypoxic tumours were associated with significantly worse overall survival in multiple independent datasets and characterised by an aggressive molecular phenotype with immune dysfunction. This finding is significant as data from previous pan-cancer analyses suggest that hypoxia leads to genomic instability and mutagenesis which can influence cancer evolution<sup>199</sup>. More work is required to understand the role that hypoxia may play in the evolutionary trajectory of OAC.

*Hypoxia is associated with a distinct metabolic phenotype in OAC.* The requirement for tumour hypoxia biomarkers in OAC and the rationale for breath analysis was discussed in the opening Chapter. It was argued that the current evidence, which points to a lipid origin for VOCs in OAC, supports the notion that tumour hypoxia may produce a detectable exhaled volatile signature. The previously derived hypoxic gene signature was used to classify OAC patients and compare their exhaled breath phenotypes in a pilot cohort study. This preliminary analysis provided a VOC model that could accurately discriminate between hypoxic classes, suitable for further validation. This implied that OAC molecular phenotypes may be reflected in the volatilome. Hypoxic metabolic reprogramming was investigated further by profiling OAC cell lines cultured under hypoxia which included the use of bespoke data pre-processing techniques. This revealed core transcriptional, lipid and volatile signatures related to hypoxia in OAC associated with redox stress.

*Hypoxia is a therapeutic target in OAC.* The translational relevance of tumour hypoxia was explored in the final Experimental chapter. From an *in silico* screen, PARP inhibitors emerged as candidate hypoxia-targeting therapeutics. Bioinformatic and experimental data was consistent with literature data suggesting hypoxia induces sensitivity to PARP inhibitors through contextual synthetic lethality. This is a new finding in OAC, and provides pre-clinical rationale for a proof-of-principle study of breath analysis performed alongside a clinical trial (SOLAR).

Overall, this provides evidence to support the leading project hypothesis that OAC tumour hypoxia influences exhaled VOCs and could be used to stratify and monitor treatment response. These initial findings need to be confirmed by evaluating the stability of VOC signatures in a larger cohort, given the molecular heterogeneity of OAC, and confounding of metabolomic data with diet and lifestyle factors<sup>256</sup>. This work has also provided a description of hypoxia-associated lipid and volatile profiles *in vitro* but does not clearly explain how this is regulated or relates to the *in vivo* phenotype. A closer investigation of key regulators such *HILPDA* is needed along with tracing experiments to unpick these mechanisms in representative models. Not only could this shed new insights into the origins of VOCs in OAC, but also critical survival pathways in hypoxic OAC cells amenable to therapeutic intervention.

## 6.2 Technical gains from this thesis

In Chapter 2, a method of gene signature evaluation, derivation and validation was developed and incorporated several previously reported techniques in a unified approach. This strategy was designed to build a gene signature with biological and clinical relevance, that was also robust and high quality across multiple expression datasets. The final derived signature could identify distinct biological subtypes of OAC and there is scope to expand this methodology in other cancers.

In Chapter 3, a method of uniting parallel 1D GC-MS methods with 2D-GC-MS using split flow re-collection was described. Alongside stringent quality assurance procedures, this technique provides comprehensive and efficient coverage of the exhaled volatilome and improves confidence in compound identification in untargeted workflows.

In Chapter 4, a standardised pre-processing pipeline for biofluid VOC data was written into code with the aim to improve analytical reproducibility and efficiency. Many previous cell VOC analyses have found conflicting results<sup>257</sup>, and a major factor can be variable data pre-processing. This pipeline gives confidence in the data quality through comprehensive filtering to exclude contaminating, non-linear and unreproducible compounds. Filtering and normalization settings can also be easily modified in the code and reported to improve transparency.

## 6.3 Future research

In addition to the immediate succession goals outlined at the end of each chapter, the following projects are planned and will aim to produce a complete description of the role of hypoxia in the volatile metabolism of OAC, outside of the scope of this thesis. This will support the ultimate aim of developing a non-invasive breath test for tumour hypoxia in OAC.

### 6.3.1 Genetic framework for hypoxia

Further experiments will aim to unpick the direct contribution of key hypoxic pathways to the volatile signature. Transcriptomic work in this thesis, for example, has identified *HILPDA*, a downstream target of the HIF signalling pathway, as a prime candidate for regulating the lipid-VOC response. Stable knockdown models of OAC, silencing hypoxia-related pathways, will be used for functional analyses such as cell cycle analysis, lipid droplet formation and redox homeostasis under hypoxia. Metabolic characterisation will include lipidomic and headspace VOC quantification. Fate mapping with stable isotope tracing will then be used to confirm the origin of hypoxia-associated VOCs. If these adaptations to hypoxia are verified as metabolic dependencies, it will also be of interest to explore therapeutic options which directly inhibit or silence these pathways since selectively targeting hypoxic cancer cells is an urgent research priority.

### 6.3.2 Description of the hypoxic volatilome of OAC organoid models

Patient-derived organoid (PDO) models may offer a more faithful representation of human OAC compared to immortalised cell lines. Fully characterised OAC PDOs will be cultured under hypoxic and normoxic conditions and undergo direct headspace VOC analysis using a microchamber sampling device. This will permit the study of the tumour-specific component of cancer-associated VOCs, bridging the gap between cell and human models.

### 6.3.3 Hypoxia-Driven Response in Oesophageal Adenocarcinoma (HYDRA) study

This study is an HRA approved study (REC 23/PR/0723) where patients (n = 10) with OAC will receive the hypoxia-labelling agent pimonidazole prior to surgery. The aims of this study are to: (i) describe the aberrant hypoxic lipidome in OAC using hypoxia-labelled tissue, (ii) perform pathway analysis using matched transcriptomic data, (iii) generate a hypoxic gene signature in OAC, (iv) derive a breath

model to stratify tumour hypoxia in OAC. Patients will undergo intra-operative tumour tissue sampling, with multiple biopsies taken from the periphery and core to ensure representation of global tumour hypoxia. Tumour biopsies will be stained for hypoxia (pimonidazole), proliferation, vasculature, and epithelial markers with support from an expert histopathologist. Spatial lipidomic analysis will be performed using desorption electrospray ionisation mass spectrometry (DESI-MS) comparing normoxic and hypoxic regions. Spatial transcriptomic analysis will be performed to compare normoxic and hypoxic regions. Transcriptomic data will be used for pathway analysis and to derive an OAC-specific hypoxic gene signature. By applying the hypoxic gene signature to existing paired transcriptomic-breath datasets (AROMA-Bioresource (n=150) (REC 21/HH/7100) & Upper GI Imperial Tissue Bank (n=100) (REC 17/WA/0161), candidate VOC hypoxia biomarkers will be identified.

#### **6.3.4 Stakeholder analysis on the breath test-assisted treatment pathway**

The aim of this study will be to map out barriers to adoption and develop risk mitigation strategies for scaling up a hypoxia breath test to commercialisation. Working with NIHR London IVD, the opinion of different stakeholders will be solicited on the utility of the breath test, barriers to implementation and the acceptable cost point to reach commercialisation through Decision Conferencing. Pathways will be modelled by using a decision tree wherein separate branches would represent existing versus breath test-assisted pathways. Alignment with other non-invasive tests will also be considered. Overall, this will identify and provide potential solutions to the challenges that may exist in translating the test to clinical practice.



## References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *Ca Cancer J Clin* **70**, 7–30 (2020).
2. Hvid-Jensen, F., Pedersen, L., Drewes, A. M., Sørensen, H. T. & Funch-Jensen, P. Incidence of Adenocarcinoma among Patients with Barrett's Esophagus. *New Engl J Med* **365**, 1375–1383 (2011).
3. Offman, J., Pesola, F. & Sasieni, P. Trends and projections in adenocarcinoma and squamous cell carcinoma of the oesophagus in England from 1971 to 2037. *Brit J Cancer* **118**, 1391–1398 (2018).
4. Oesophageal cancer survival statistics | Cancer Research UK. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/survival>.
5. Al-Batran, S.-E. *et al.* Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial. *Lancet* (2019) doi:10.1016/s0140-6736(18)32557-1.
6. Hagen, P. van *et al.* Preoperative chemoradiotherapy for esophageal or junctional cancer. *New Engl J Medicine* **366**, 2074–84 (2012).
7. Donahue, J. M. *et al.* Complete Pathologic Response After Neoadjuvant Chemoradiotherapy for Esophageal Cancer Is Associated With Enhanced Survival. *Ann. Thorac. Surg.* **87**, 392–399 (2009).
8. Oesophageal cancer incidence statistics | Cancer Research UK. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/incidence>.
9. Recommendations | Oesophago-gastric cancer: assessment and management in adults | Guidance | NICE. <https://www.nice.org.uk/guidance/ng83/chapter/Recommendations#palliative-management-2>.
10. Shah, M. A. *et al.* Efficacy and Safety of Pembrolizumab for Heavily Pretreated Patients With Advanced, Metastatic Adenocarcinoma or Squamous Cell Carcinoma of the Esophagus. *Jama Oncol* **5**, 546–550 (2019).
11. Wu, S.-G. *et al.* Sites of metastasis and overall survival in esophageal cancer: a population-based study. *Cancer Manag. Res.* **9**, 781–788 (2017).
12. Vaupel, P., Höckel, M. & Mayer, A. Detection and Characterization of Tumor Hypoxia Using pO<sub>2</sub> Histograms. *Antioxid Redox Sign* **9**, 1221–1236 (2007).
13. Vaupel, P. & Mayer, A. Hypoxia in cancer: significance and impact on clinical outcome. *Cancer Metast Rev* **26**, 225–239 (2007).
14. Carmeliet, P. & Jain, R. K. Angiogenesis in cancer and other diseases. *Nature* **407**, 249–257 (2000).
15. Höckel, M. & Vaupel, P. Tumor Hypoxia: Definitions and Current Clinical, Biologic, and Molecular Aspects. *Jnci J National Cancer Inst* **93**, 266–276 (2001).
16. Grimes, D. R. & Partridge, M. A mechanistic investigation of the oxygen fixation hypothesis and oxygen enhancement ratio. *Biomed Phys Eng Express* **1**, 045209 (2015).
17. Lequeux, A. *et al.* Impact of hypoxic tumor microenvironment and tumor cell plasticity on the expression of immune checkpoints. *Cancer Lett* **458**, 13–20 (2019).

18. Luoto, K. R., Kumareswaran, R. & Bristow, R. G. Tumor hypoxia as a driving force in genetic instability. *Genome Integr* **4**, 5 (2013).
19. Muz, B., Puente, P. de la, Azab, F. & Azab, A. K. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia* **3**, 83–92 (2015).
20. Nordsmark, M. *et al.* Measurements of hypoxia using pimonidazole and polarographic oxygen-sensitive electrodes in human cervix carcinomas. *Radiother Oncol* **67**, 35–44 (2003).
21. Nordsmark, M. *et al.* Prognostic value of tumor oxygenation in 397 head and neck tumors after primary radiation therapy. An international multi-center study. *Radiother Oncol* **77**, 18–24 (2005).
22. Chopra, S. *et al.* Comparing oxygen-sensitive MRI (BOLD R2\*) with oxygen electrode measurements: A pilot study in men with prostate cancer. *Int J Radiat Biol* **85**, 805–813 (2009).
23. Jing, X. *et al.* Role of hypoxia in cancer therapy by regulating the tumor microenvironment. *Mol Cancer* **18**, 157 (2019).
24. Brizel, D. M. *et al.* Tumor oxygenation predicts for the likelihood of distant metastases in human soft tissue sarcoma. *Cancer Res* **56**, 941–3 (1996).
25. Erler, J. T. *et al.* Lysyl oxidase is essential for hypoxia-induced metastasis. *Nature* **440**, 1222–1226 (2006).
26. Chan, D. A. & Giaccia, A. J. Hypoxia, gene expression, and metastasis. *Cancer Metast Rev* **26**, 333–339 (2007).
27. Singleton, D. C., Macann, A. & Wilson, W. R. Therapeutic targeting of the hypoxic tumour microenvironment. *Nat Rev Clin Oncol* 1–22 (2021) doi:10.1038/s41571-021-00539-4.
28. Tharmalingham, H. & Hoskin, P. Clinical trials targeting hypoxia. *Br J Radiology* **92**, 20170966 (2018).
29. Song, Y. P., Mistry, H., Choudhury, A. & Hoskin, P. Long-term outcomes of hypoxia modification in bladder preservation: Update from BCON trial. *J Clin Oncol* **37**, 356–356 (2019).
30. Overgaard, J. *et al.* A randomized double-blind phase III study of nimorazole as a hypoxic radiosensitizer of primary radiotherapy in supraglottic larynx and pharynx carcinoma. Results of the Danish Head and Neck Cancer Study (DAHANCA) Protocol 5-85. *Radiother Oncol* **46**, 135–146 (1998).
31. Wardman, P. *et al.* Radicals from one-electron reduction of nitro compounds, aromatic N-oxides and quinones: the kinetic basis for hypoxia-selective, bioreductive drugs. *Biochem Soc Symp* **61**, 171–194 (1995).
32. Sun, J. D. *et al.* Selective Tumor Hypoxia Targeting by Hypoxia-Activated Prodrug TH-302 Inhibits Tumor Growth in Preclinical Models of Cancer. *Clin Cancer Res* **18**, 758–770 (2012).
33. Spiegelberg, L. *et al.* Evofosfamide sensitizes esophageal carcinomas to radiation without increasing normal tissue toxicity. *Radiother Oncol* (2019) doi:10.1016/j.radonc.2019.06.034.
34. Hammond, E. M., Dorie, M. J. & Giaccia, A. J. Inhibition of ATR Leads to Increased Sensitivity to Hypoxia/Reoxygenation. *Cancer Res* **64**, 6556–6562 (2004).
35. Leszczynska, K. B. *et al.* Preclinical testing of an Atr inhibitor demonstrates improved response to standard therapies for esophageal cancer. *Radiotherapy Oncol J European Soc Ther Radiology Oncol* **121**, 232–238 (2016).
36. Chan, N. *et al.* Chronic Hypoxia Decreases Synthesis of Homologous Recombination Proteins to Offset Chemoresistance and Radioresistance. *Cancer Res* **68**, 605–614 (2008).

37. Bindra, R. S. *et al.* Hypoxia-Induced Down-regulation of BRCA1 Expression by E2Fs. *Cancer Res* **65**, 11597–11604 (2005).
38. Chan, N. *et al.* Contextual synthetic lethality of cancer cell kill based on the tumor microenvironment. *Cancer Res* **70**, 8045–54 (2010).
39. Olcina, M., Lecane, P. S. & Hammond, E. M. Targeting hypoxic cells through the DNA damage response. *Clin Cancer Res Official J Am Assoc Cancer Res* **16**, 5624–9 (2010).
40. Dewhirst, M. W. A potential solution for eliminating hypoxia as a cause for radioresistance. *Proc National Acad Sci* **115**, 10548–10550 (2018).
41. Secomb, T. W., Hsu, R., Ong, E. T., Gross, J. F. & Dewhirst, M. W. Analysis of the Effects of Oxygen Supply and Demand on Hypoxic Fraction in Tumors. *Acta Oncol* **34**, 313–316 (1995).
42. Zannella, V. E. *et al.* Reprogramming Metabolism with Metformin Improves Tumor Oxygenation and Radiotherapy Response. *Am Assoc Cancer Res* **19**, 6741–6750 (2013).
43. Benej, M. *et al.* Papaverine and its derivatives radiosensitize solid tumors by inhibiting mitochondrial metabolism. *Proc National Acad Sci* **115**, 10756–10761 (2018).
44. Skinner, H. D. *et al.* Metformin use and improved response to therapy in esophageal adenocarcinoma. *Acta Oncol* **52**, 1002–1009 (2012).
45. Hunter, F. W., Wouters, B. G. & Wilson, W. R. Hypoxia-activated prodrugs: paths forward in the era of personalised medicine. *Brit J Cancer* **114**, 1071 (2016).
46. Spiegelberg, L. *et al.* Hypoxia-activated prodrugs and (lack of) clinical progress: The need for hypoxia-based biomarker patient selection in phase III clinical trials. *Clin Transl Radiat Oncol* **15**, 62–69 (2019).
47. Toustrup, K. *et al.* Development of a Hypoxia Gene Expression Classifier with Predictive Impact for Hypoxic Modification of Radiotherapy in Head and Neck Cancer. *Cancer Res* **71**, 5923–5931 (2011).
48. Tatum, J. L. *et al.* Hypoxia: Importance in tumor biology, noninvasive measurement by imaging, and value of its measurement in the management of cancer therapy. *Int J Radiat Biol* **82**, 699–757 (2006).
49. Eustace, A. *et al.* A 26-Gene Hypoxia Signature Predicts Benefit from Hypoxia-Modifying Therapy in Laryngeal Cancer but Not Bladder Cancer. *Clin Cancer Res* **19**, 4879–4888 (2013).
50. Harris, B. H. L., Barberis, A., West, C. M. L. & Buffa, F. M. Gene Expression Signatures as Biomarkers of Tumour Hypoxia. *Clin Oncol* **27**, 547–560 (2015).
51. Lukovic, J. *et al.* Intratumoral heterogeneity and hypoxia gene expression signatures: Is a single biopsy adequate? *Clin Transl Radiat Oncol* **19**, 110–115 (2019).
52. Betts, G. N. J. *et al.* Prospective technical validation and assessment of intra-tumour heterogeneity of a low density array hypoxia gene profile in head and neck squamous cell carcinoma. *Eur J Cancer* **49**, 156–165 (2013).
53. Yang, L. *et al.* A Gene Signature for Selecting Benefit from Hypoxia Modification of Radiotherapy for High-Risk Bladder Cancer Patients. *Clin Cancer Res* **23**, 4761–4768 (2017).
54. Toustrup, K. *et al.* Validation of a 15-gene hypoxia classifier in head and neck cancer for prospective use in clinical trials. *Acta Oncol* **55**, 1091–1098 (2016).

55. Thomson, D. *et al.* NIMRAD – A Phase III Trial to Investigate the Use of Nimorazole Hypoxia Modification with Intensity-modulated Radiotherapy in Head and Neck Cancer. *Clin Oncol* **26**, 344–347 (2014).
56. Yang, L. & West, C. M. Hypoxia gene expression signatures as predictive biomarkers for personalising radiotherapy. *Br J Radiology* **92**, 20180036 (2018).
57. Semenza, G. L. Hypoxia, Clonal Selection, and the Role of HIF-1 in Tumor Progression. *Crit Rev Biochem Mol* **35**, 71–103 (2008).
58. Wykoff, C. C. *et al.* Hypoxia-inducible expression of tumor-associated carbonic anhydrases. *Cancer Res* **60**, 7075–83 (2000).
59. Semenza, G. L. Targeting HIF-1 for cancer therapy. *Nat Rev Cancer* **3**, 721–732 (2003).
60. Kuijk, S. J. A. van *et al.* Prognostic Significance of Carbonic Anhydrase IX Expression in Cancer Patients: A Meta-Analysis. *Frontiers Oncol* **6**, 69 (2016).
61. Yu, M. *et al.* The prognostic value of GLUT1 in cancers: a systematic review and meta-analysis. *Oncotarget* **8**, 43356–43367 (2017).
62. Shin, K. H. *et al.* Detecting changes in tumor hypoxia with carbonic anhydrase IX and pimonidazole. *Cancer Biol Ther* **6**, 70–75 (2007).
63. Varia, M. A. *et al.* Pimonidazole: A Novel Hypoxia Marker for Complementary Study of Tumor Hypoxia and Cell Proliferation in Cervical Carcinoma. *Gynecol Oncol* **71**, 270–277 (1998).
64. Kaanders, J. H. A. M. *et al.* Pimonidazole binding and tumor vascularity predict for treatment outcome in head and neck cancer. *Cancer Res* **62**, 7066–74 (2002).
65. Le, Q.-T. *et al.* Identification of osteopontin as a prognostic plasma marker for head and neck squamous cell carcinomas. *Clin Cancer Res Official J Am Assoc Cancer Res* **9**, 59–67 (2003).
66. Overgaard, J. *et al.* Plasma osteopontin, hypoxia, and response to the hypoxia sensitiser nimorazole in radiotherapy of head and neck cancer: results from the DAHANCA 5 randomised double-blind placebo-controlled trial. *Lancet Oncol* **6**, 757–764 (2005).
67. Lin, J. *et al.* Osteopontin (OPN/SPP1) isoforms collectively enhance tumor cell invasion and dissemination in esophageal adenocarcinoma. *Oncotarget* **6**, 22239–22257 (2015).
68. Zhao, H. *et al.* The role of osteopontin in the progression of solid organ tumour. *Cell Death Dis* **9**, 356 (2018).
69. Lund, S. A., Giachelli, C. M. & Scatena, M. The role of osteopontin in inflammatory processes. *J Cell Commun Signal* **3**, 311–22 (2009).
70. Chao, K. S. C. *et al.* A novel approach to overcome hypoxic tumor resistance: Cu-ATSM-guided intensity-modulated radiation therapy. *Int J Radiat Oncol Biology Phys* **49**, 1171–1182 (2001).
71. Zimny, M. *et al.* FDG—a marker of tumour hypoxia? A comparison with [18F]fluoromisonidazole and pO<sub>2</sub>-polarography in metastatic head and neck cancer. *Eur J Nucl Med Mol I* **33**, 1426–1431 (2006).
72. Mortensen, L. S. *et al.* Accessing radiation response using hypoxia PET imaging and oxygen sensitive electrodes: A preclinical study. *Radiother Oncol* **99**, 418–423 (2011).

73. Dubois, L. J. *et al.* Preclinical evaluation and validation of [18F]HX4, a promising hypoxia marker for PET imaging. *Proc National Acad Sci* **108**, 14620–14625 (2011).
74. Tao, R. *et al.* Hypoxia imaging in upper gastrointestinal tumors and application to radiation therapy. *J Gastrointest Oncol* **9**, 1044–1053 (2018).
75. Walsh, J. C. *et al.* The Clinical Importance of Assessing Tumor Hypoxia: Relationship of Tumor Hypoxia to Prognosis and Therapeutic Opportunities. *Antioxid Redox Sign* **21**, 1516–1554 (2014).
76. Baudalet, C. & Gallez, B. How does blood oxygen level-dependent (BOLD) contrast correlate with oxygen partial pressure (pO<sub>2</sub>) inside tumors? *Magnet Reson Med* **48**, 980–986 (2002).
77. Cooper, R. A. *et al.* Tumour oxygenation levels correlate with dynamic contrast-enhanced magnetic resonance imaging parameters in carcinoma of the cervix. *Radiother Oncol* **57**, 53–59 (2000).
78. O'Connor, J. P. B. *et al.* Oxygen-Enhanced MRI Accurately Identifies, Quantifies, and Maps Tumor Hypoxia in Preclinical Cancer Models. *Cancer Res* **76**, 787–95 (2015).
79. Shao, Q. *et al.* In vivo photoacoustic lifetime imaging of tumor hypoxia in small animals. *J Biomed Opt* **18**, 076019 (2013).
80. Esipova, T. V. *et al.* Two new “protected” oxyphors for biological oximetry: properties and application in tumor imaging. *Anal Chem* **83**, 8756–65 (2011).
81. Gallez, B., Baudalet, C. & Jordan, B. F. Assessment of tumor oxygenation by electron paramagnetic resonance: principles and applications. *Nmr Biomed* **17**, 240–262 (2004).
82. Lee, C. P. *et al.* A phase I study of the nitroimidazole hypoxia marker SR4554 using 19F magnetic resonance spectroscopy. *Brit J Cancer* **101**, 1860–1868 (2009).
83. Matsumoto, S. *et al.* Simultaneous imaging of tumor oxygenation and microvascular permeability using Overhauser enhanced MRI. *Proc National Acad Sci* **106**, 17898–17903 (2009).
84. Le, Q.-T. & Courter, D. Clinical biomarkers for hypoxia targeting. *Cancer Metast Rev* **27**, 351 (2008).
85. Kappler, M. *et al.* Immunohistochemical Detection of HIF-1 $\alpha$  and CAIX in Advanced Head-and-Neck Cancer: Prognostic Role and Correlation with Tumor Markers and Tumor Oxygenation Parameters. *Strahlenther Onkol* **184**, 393–399 (2008).
86. England, P. H. Index of Grey Literature and Alternative Sources and Resources. [https://libguides.kcl.ac.uk/ld.php?content\\_id=31075973](https://libguides.kcl.ac.uk/ld.php?content_id=31075973) (2020).
87. Liberati, A. *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Bmj* **339**, b2700–b2700 (2009).
88. McShane, L. M. *et al.* REporting recommendations for tumour MARKer prognostic studies (REMARK). *Brit J Cancer* **93**, 387–391 (2005).
89. Hayden, J. A., Windt, D. A. van der, Cartwright, J. L., Côté, P. & Bombardier, C. Assessing bias in studies of prognostic factors. *Ann Intern Med* **158**, 280–6 (2013).
90. Blayney, J. K. *et al.* Glucose transporter 1 expression as a marker of prognosis in oesophageal adenocarcinoma. *Oncotarget* **9**, 18518–18528 (2018).

91. Mirza, A. "A prospective study to evaluate pimonidazole as a marker of tumour hypoxia in gastro-oesophageal cancer" in *Markers of Treatment Response for Gastro-Oesophageal Cancers*. (2012).
92. Griffiths, E. A. *et al.* Hypoxia-inducible factor-1 $\alpha$  expression in the gastric carcinogenesis sequence and its prognostic role in gastric and gastro-oesophageal adenocarcinomas. *Brit J Cancer* **96**, 95–103 (2006).
93. Griffiths, E. A. *et al.* Increasing expression of hypoxia-inducible proteins in the Barrett's metaplasia–dysplasia–adenocarcinoma sequence. *Brit J Cancer* **96**, 1377–1383 (2007).
94. Huhta, H. *et al.* Intratumoral lactate metabolism in Barrett's esophagus and adenocarcinoma. *Oncotarget* **8**, 22894–22902 (2017).
95. Driessen, A. *et al.* Expression of Carbonic Anhydrase IX (CA IX), a Hypoxia-Related Protein, Rather Than Vascular-Endothelial Growth Factor (VEGF), a Pro-Angiogenic Factor, Correlates With an Extremely Poor Prognosis in Esophageal and Gastric Adenocarcinomas. *Ann Surg* **243**, 334–340 (2006).
96. Stahl, A. *et al.* FDG PET imaging of locally advanced gastric carcinomas: correlation with endoscopic and histopathological findings. *Eur J Nucl Med Mol I* **30**, 288–295 (2003).
97. Klaassen, R. *et al.* Feasibility and repeatability of PET with the hypoxia tracer [18F]HX4 in oesophageal and pancreatic cancer. *Radiother Oncol* **116**, 94–99 (2015).
98. Larue, R. T. H. M. *et al.* A phase 1 'window-of-opportunity' trial testing evofosfamide (TH-302), a tumour-selective hypoxia-activated cytotoxic prodrug, with preoperative chemoradiotherapy in oesophageal adenocarcinoma patients. *Bmc Cancer* **16**, 644 (2016).
99. Ling, F. C. *et al.* HIF-1 $\alpha$  mRNA is not associated with histopathological regression following neoadjuvant chemoradiation in esophageal cancer. *Anticancer Res* **26**, 4505–9 (2006).
100. Smit, J. K. *et al.* Prediction of response to radiotherapy in the treatment of esophageal cancer using stem cell markers. *Radiotherapy Oncol J European Soc Ther Radiology Oncol* **107**, 434–41 (2013).
101. Daly, C. S. *et al.* Hypoxia modulates the stem cell population and induces EMT in the MCF-10A breast epithelial cell line. *Oncol Rep* **39**, 483–490 (2017).
102. Kim, H., Lin, Q., Glazer, P. M. & Yun, Z. The hypoxic tumor microenvironment in vivo selects the cancer stem cell fate of breast cancer cells. *Breast Cancer Res* **20**, 16 (2018).
103. Kuo, T.-C. *et al.* Human Breathomics Database. *Database* **2020**, baz139 (2020).
104. Schiffman, J. D., Fisher, P. G. & Gibbs, P. Early Detection of Cancer: Past, Present, and Future. *Am. Soc. Clin. Oncol. Educ. Book* 57–65 (2015) doi:10.14694/edbook\_am.2015.35.57.
105. Woodfield, G. *et al.* Feasibility and acceptability of breath research in primary care: a prospective, cross-sectional, observational study. *Bmj Open* **11**, e044691 (2021).
106. Abela, J. E., Skeldon, K. D., Stuart, R. C. & Padgett, M. J. Exhaled ethane concentration in patients with cancer of the upper gastrointestinal tract - a proof of concept study. *Biosci. trends* **3**, 110–4 (2009).
107. Kumar, S. *et al.* Mass Spectrometric Analysis of Exhaled Breath for the Identification of Volatile Organic Compound Biomarkers in Esophageal and Gastric Adenocarcinoma. *Ann Surg* **262**, 981–990 (2015).
108. Kumar, S. *et al.* Selected Ion Flow Tube Mass Spectrometry Analysis of Exhaled Breath for Volatile Organic Compound Profiling of Esophago-Gastric Cancer. *Anal. Chem.* **85**, 6121–6128 (2013).

109. Markar, S. R. *et al.* Assessment of a Noninvasive Exhaled Breath Test for the Diagnosis of Oesophagogastric Cancer. *Jama Oncol* (2018) doi:10.1001/jamaoncol.2018.0991.
110. Su, R. *et al.* Mass spectrometry for breath analysis. *TrAC Trends Anal. Chem.* **158**, 116823 (2023).
111. Demeestere, K., Dewulf, J., Roo, K. D., Wispelaere, P. D. & Langenhove, H. V. Quality control in quantification of volatile organic compounds analysed by thermal desorption–gas chromatography–mass spectrometry. *J. Chromatogr. A* **1186**, 348–357 (2008).
112. Harshman, S. W. *et al.* Evaluation of a standardized collection device for exhaled breath sampling onto thermal desorption tubes. *J. Breath Res.* **14**, 036004 (2020).
113. Sani, S. N. *et al.* LC-MS/MS Based Volatile Organic Compound Biomarkers Analysis for Early Detection of Lung Cancer. *Cancers* **15**, 1186 (2023).
114. Brais, C. J., Ibañez, J. O., Schwartz, A. J. & Ray, S. J. RECENT ADVANCES IN INSTRUMENTAL APPROACHES TO TIME-OF-FLIGHT MASS SPECTROMETRY. *Mass Spectrom. Rev.* **40**, 647–669 (2021).
115. Li, C. *et al.* Towards Higher Sensitivity of Mass Spectrometry: A Perspective From the Mass Analyzers. *Front. Chem.* **9**, 813359 (2021).
116. Miekisch, W., Schubert, J. K. & Noeldge-Schomburg, G. F. E. Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clin Chim Acta* **347**, 25–39 (2004).
117. Boots, A. W. *et al.* The versatile use of exhaled volatile organic compounds in human health and disease. *J. Breath Res* **6**, 027108 (2012).
118. Bos, L. D. J., Sterk, P. J. & Schultz, M. J. Volatile Metabolites of Pathogens: A Systematic Review. *Plos Pathog* **9**, e1003311 (2013).
119. Hanna, G. B., Boshier, P. R., Markar, S. R. & Romano, A. Accuracy and Methodologic Challenges of Volatile Organic Compound–Based Exhaled Breath Tests for Cancer Diagnosis: A Systematic Review and Meta-analysis. *Jama Oncol* e182815–e182815 (2018) doi:10.1001/jamaoncol.2018.2815.
120. Ratcliffe, N. *et al.* A mechanistic study and review of volatile products from peroxidation of unsaturated fatty acids: an aid to understanding the origins of volatile organic compounds from the human body. *J. Breath Res* **14**, 034001 (2020).
121. Gunstone, F. D. Reaction of oxygen and unsaturated fatty acids. *J. Am. Oil Chem. Soc.* **61**, 441–447 (1984).
122. Aikens, J. & Dix, T. A. Hydrodioxyl (Perhydroxyl), Peroxyl, and Hydroxyl Radical-Initiated Lipid Peroxidation of Large Unilamellar Vesicles (Liposomes): Comparative and Mechanistic Studies. *Arch. Biochem. Biophys.* **305**, 516–525 (1993).
123. Frankel, E. N. Lipid oxidation: Mechanisms, products and biological significance. *J. Am. Oil Chem. Soc.* **61**, 1908–1917 (1984).
124. Abbassi-Ghadi, N. *et al.* De novo lipogenesis alters the phospholipidome of esophageal adenocarcinoma. *Cancer Res* canres.4035.2019 (2020) doi:10.1158/0008-5472.can-19-4035.
125. Antonowicz, S. *et al.* Endogenous aldehyde accumulation generates genotoxicity and exhaled biomarkers in esophageal adenocarcinoma. *Nat Commun* **12**, 1454 (2021).

126. Eales, K. L., Hollinshead, K. E. R. & Tennant, D. A. Hypoxia and metabolic adaptation of cancer cells. *Oncogenesis* **5**, e190 (2016).
127. Wouters, B. G. & Koritzinsky, M. Hypoxia signalling through mTOR and the unfolded protein response in cancer. *Nat Rev Cancer* **8**, 851–864 (2008).
128. Fukuda, R. *et al.* HIF-1 Regulates Cytochrome Oxidase Subunits to Optimize Efficiency of Respiration in Hypoxic Cells. *Cell* **129**, 111–122 (2007).
129. Kamphorst, J. J., Chung, M. K., Fan, J. & Rabinowitz, J. D. Quantitative analysis of acetyl-CoA production in hypoxic cancer cells reveals substantial contribution from acetate. *Cancer Metabolism* **2**, 23 (2014).
130. Metallo, C. M. *et al.* Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **481**, 380–384 (2011).
131. Saxton, R. A. & Sabatini, D. M. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* **168**, 960–976 (2017).
132. Dunlop, E. A. & Tee, A. R. Mammalian target of rapamycin complex 1: signalling inputs, substrates and feedback mechanisms. *Cell Signal* **21**, 827–35 (2009).
133. Yuan, G., Nanduri, J., Khan, S., Semenza, G. L. & Prabhakar, N. R. Induction of HIF-1 $\alpha$  expression by intermittent hypoxia: Involvement of NADPH oxidase, Ca<sup>2+</sup> signaling, prolyl hydroxylases, and mTOR. *J Cell Physiol* **217**, 674–685 (2008).
134. Bartoszewska, S. & Collawn, J. F. Unfolded protein response (UPR) integrated signaling networks determine cell fate during hypoxia. *Cell Mol Biol Lett* **25**, 18 (2020).
135. Hetz, C. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nat Rev Mol Cell Bio* **13**, 89–102 (2012).
136. Harding, H. P. *et al.* An Integrated Stress Response Regulates Amino Acid Metabolism and Resistance to Oxidative Stress. *Mol Cell* **11**, 619–633 (2003).
137. Rouschop, K. M. *et al.* PERK/eIF2 signaling protects therapy resistant hypoxic cells through induction of glutathione synthesis and protection against ROS. *Proc National Acad Sci* **110**, 4622–4627 (2013).
138. Guzy, R. D. *et al.* Mitochondrial complex III is required for hypoxia-induced ROS production and cellular oxygen sensing. *Cell Metab* **1**, 401–408 (2005).
139. Guzy, R. D. & Schumacker, P. T. Oxygen sensing by mitochondria at complex III: the paradox of increased reactive oxygen species during hypoxia. *Exp. Physiol.* **91**, 807–819 (2006).
140. Ma, T. S. *et al.* Hypoxia-induced transcriptional stress is mediated by ROS-induced R-loops. *Nucleic Acids Res.* **51**, 11584–11599 (2023).
141. Negre-Salvayre, A., Coatrieux, C., Ingueneau, C. & Salvayre, R. Advanced lipid peroxidation end products in oxidative damage to proteins. Potential role in diseases and therapeutic prospects for the inhibitors. *Brit J Pharmacol* **153**, 6–20 (2008).
142. Saliakoura, M. *et al.* PLC $\gamma$ 1 suppression promotes the adaptation of KRAS-mutant lung adenocarcinomas to hypoxia. *Nat Cell Biol* **22**, 1382–1395 (2020).



143. Munir, R., Lisec, J., Swinnen, J. V. & Zaidi, N. Lipid metabolism in cancer cells under metabolic stress. *Brit J Cancer* **120**, 1090–1098 (2019).
144. Vriens, K. *et al.* Evidence for an alternative fatty acid desaturation pathway increasing cancer plasticity. *Nature* **566**, 403–406 (2019).
145. Valli, A. *et al.* Hypoxia induces a lipogenic cancer cell phenotype via HIF1 $\alpha$ -dependent and -independent pathways. *Oncotarget* **6**, 1920–1941 (2014).
146. Yu, Y., Vidalino, L., Anesi, A., Macchi, P. & Guella, G. A lipidomics investigation of the induced hypoxia stress on HeLa cells by using MS and NMR techniques. *Mol Biosyst* **10**, 878–890 (2014).
147. Tobias, F. & Hummon, A. B. Lipidomic comparison of 2D and 3D colon cancer cell culture models. *J. Mass Spectrom.* **57**, e4880 (2022).
148. Tamiya-Koizumi, K. *et al.* Cellular concentrations of plasmalogen species containing a polyunsaturated fatty acid significantly increase under hypoxia in human colorectal cancer, Caco2 cells. *Biochem. Biophys. Res. Commun.* **611**, 1–7 (2022).
149. Zou, Y. *et al.* A GPX4-dependent cancer cell state underlies the clear-cell morphology and confers sensitivity to ferroptosis. *Nat Commun* **10**, 1617 (2019).
150. Kamphorst, J. J. *et al.* Hypoxic and Ras-transformed cells support growth by scavenging unsaturated fatty acids from lysophospholipids. *Proc National Acad Sci* **110**, 8882–8887 (2013).
151. Young, R. M. *et al.* Dysregulated mTORC1 renders cells critically dependent on desaturated lipids for survival under tumor-like stress. *Gene Dev* **27**, 1115–1131 (2013).
152. Ariyama, H., Kono, N., Matsuda, S., Inoue, T. & Arai, H. Decrease in Membrane Phospholipid Unsaturation Induces Unfolded Protein Response\*. *J Biological Chem* **285**, 22027–22035 (2010).
153. Ta, N. L. & Seyfried, T. N. Influence of Serum and Hypoxia on Incorporation of [14C]-d-Glucose or [14C]-L-Glutamine into Lipids and Lactate in Murine Glioblastoma Cells. *Lipids* **50**, 1167–1184 (2015).
154. Matschke, J., Wiebeck, E., Hurst, S., Rudner, J. & Jendrossek, V. Role of SGK1 for fatty acid uptake, cell survival and radioresistance of NCI-H460 lung cancer cells exposed to acute or chronic cycling severe hypoxia. *Radiat. Oncol.* **11**, 75 (2016).
155. Mylonis, I. *et al.* Hypoxia causes triglyceride accumulation by HIF-1-mediated stimulation of lipin 1 expression. *J Cell Sci* **125**, 3485–93 (2012).
156. Bensaad, K. *et al.* Fatty Acid Uptake and Lipid Storage Induced by HIF-1 $\alpha$  Contribute to Cell Growth and Survival after Hypoxia-Reoxygenation. *Cell Rep.* **9**, 349–365 (2014).
157. VandeKopple, M. J. *et al.* HILPDA Regulates Lipid Metabolism, Lipid Droplet Abundance, and Response to Microenvironmental Stress in Solid Tumors. *Mol Cancer Res* **17**, 2089–2101 (2019).
158. Qiu, B. *et al.* HIF2 $\alpha$ -Dependent Lipid Storage Promotes Endoplasmic Reticulum Homeostasis in Clear-Cell Renal Cell Carcinoma. *Cancer Discov.* **5**, 652–667 (2015).
159. Mattijssen, F. *et al.* Hypoxia-inducible lipid droplet-associated (HILPDA) is a novel peroxisome proliferator-activated receptor (PPAR) target involved in hepatic triglyceride secretion. *J Biol Chem* **289**, 19279–93 (2014).

160. Bailey, A. P. *et al.* Antioxidant Role for Lipid Droplets in a Stem Cell Niche of *Drosophila*. *Cell* **163**, 340–353 (2015).
161. Ackerman, D. *et al.* Triglycerides Promote Lipid Homeostasis during Hypoxic Stress by Balancing Fatty Acid Saturation. *Cell Reports* **24**, 2596-2605.e5 (2018).
162. Ackerman, D. & Simon, M. C. Hypoxia, lipids, and cancer: surviving the harsh tumor microenvironment. *Trends Cell Biol* **24**, 472–478 (2014).
163. Kim, W. *et al.* Polyunsaturated Fatty Acid Desaturation Is a Mechanism for Glycolytic NAD<sup>+</sup> Recycling. *Cell Metab.* **29**, 856-870.e7 (2019).
164. Taware, R. *et al.* Extracellular volatilomic alterations induced by hypoxia in breast cancer cells. *Metabolomics* **16**, 21 (2020).
165. Issitt, T., Reilly, M., Sweeney, S. T., Brackenbury, W. J. & Redeker, K. R. GC/MS analysis of hypoxic volatile metabolic markers in the MDA-MB-231 breast cancer cell line. *Frontiers Mol Biosci* **10**, 1178269 (2023).
166. Furuhashi, T., Matsumoto, Y., Ishii, R., Sugasawa, T. & Ota, S. Hypoxia and lactate influence VOC production in A549 lung cancer cells. *Front. Mol. Biosci.* **10**, 1274298 (2023).
167. Hammond, E. M. *et al.* The Meaning, Measurement and Modification of Hypoxia in the Laboratory and the Clinic. *Clin Oncol* **26**, 277–288 (2014).
168. Thienpont, B. *et al.* Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* **537**, 63–68 (2016).
169. Ye, Y. *et al.* Characterization of hypoxia-associated molecular features to aid hypoxia-targeted therapy. *Nat Metabolism* **1**, 431–444 (2019).
170. Buffa, F. M., Harris, A. L., West, C. M. & Miller, C. J. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *Brit J Cancer* **102**, 428 (2010).
171. Lombardi, O. *et al.* Pan-cancer analysis of tissue and single-cell HIF-pathway activation using a conserved gene signature. *Cell Rep.* **41**, 111652 (2022).
172. Chi, J.-T. *et al.* Gene Expression Programs in Response to Hypoxia: Cell Type Specificity and Prognostic Significance in Human Cancers. *Plos Med* **3**, e47 (2006).
173. D’Anna, F. *et al.* DNA methylation repels binding of hypoxia-inducible transcription factors to maintain tumor immunotolerance. *Genome Biol* **21**, 182 (2020).
174. Yang, L. *et al.* Validation of a hypoxia related gene signature in multiple soft tissue sarcoma cohorts. *Oncotarget* **9**, 3946–3955 (2017).
175. Lane, B., Khan, M. T., Choudhury, A., Salem, A. & West, C. M. L. Development and validation of a hypoxia-associated signature for lung adenocarcinoma. *Sci Rep-uk* **12**, 1290 (2022).
176. Fox, N. S., Starmans, M. H., Haider, S., Lambin, P. & Boutros, P. C. Ensemble analyses improve signatures of tumour hypoxia and reveal inter-platform differences. *Bmc Bioinformatics* **15**, 170 (2014).
177. Fardin, P. *et al.* A biology-driven approach identifies the hypoxia gene signature as a predictor of the outcome of neuroblastoma patients. *Mol. Cancer* **9**, 185 (2010).

178. Yang, L. *et al.* Development and Validation of a 28-gene Hypoxia-related Prognostic Signature for Localized Prostate Cancer. *Ebiomedicine* **31**, 182–189 (2018).
179. Mistry, I. N., Thomas, M., Calder, E. D. D., Conway, S. J. & Hammond, E. M. Clinical Advances of Hypoxia-Activated Prodrugs in Combination With Radiation Therapy. *Int. J. Radiat. Oncol.Biol.Phys.* **98**, 1183–1196 (2017).
180. Bindra, R. S. *et al.* Down-Regulation of Rad51 and Decreased Homologous Recombination in Hypoxic Cancer Cells. *Mol Cell Biol* **24**, 8504–8518 (2004).
181. Meng, A. X. *et al.* Hypoxia down-regulates DNA double strand break repair gene expression in prostate cancer cells. *Radiother Oncol* **76**, 168–176 (2005).
182. Manem, V. S. & Dhawan, A. RadiationGeneSigDB: a database of oxic and hypoxic radiation response gene signatures and their utility in pre-clinical research. *Br J Radiology* **92**, 20190198 (2019).
183. Clough, E. & Barrett, T. The Gene Expression Omnibus Database. *Methods Mol Biology Clifton N J* **1418**, 93–110 (2016).
184. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *Bmc Bioinformatics* **14**, 7 (2013).
185. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
186. Dhawan, A. *et al.* Guidelines for using sigQC for systematic evaluation of gene signatures. *Nat Protoc* **14**, 1377–1400 (2019).
187. Bustin, S. A. *et al.* The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clin Chem* **55**, 611–622 (2009).
188. Sørensen, B. S. *et al.* The usability of a 15-gene hypoxia classifier as a universal hypoxia profile in various cancer cell types. *Radiother Oncol* **116**, 346–351 (2015).
189. Boonstra, J. J. *et al.* Verification and Unmasking of Widely Used Human Esophageal Adenocarcinoma Cell Lines. *Jnci J National Cancer Inst* **102**, 271–274 (2010).
190. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, gky1015- (2018).
191. Maacha, S., Hong, J., Lersner, A. von, Zijlstra, A. & Belkhiri, A. AXL Mediates Esophageal Adenocarcinoma Cell Invasion through Regulation of Extracellular Acidification and Lysosome Trafficking. *Neoplasia* **20**, 1008–1022 (2018).
192. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
193. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
194. Network, C. G. A. R. *et al.* Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169 (2017).
195. Frankell, A. M. *et al.* The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat Genet* **51**, 506–516 (2019).

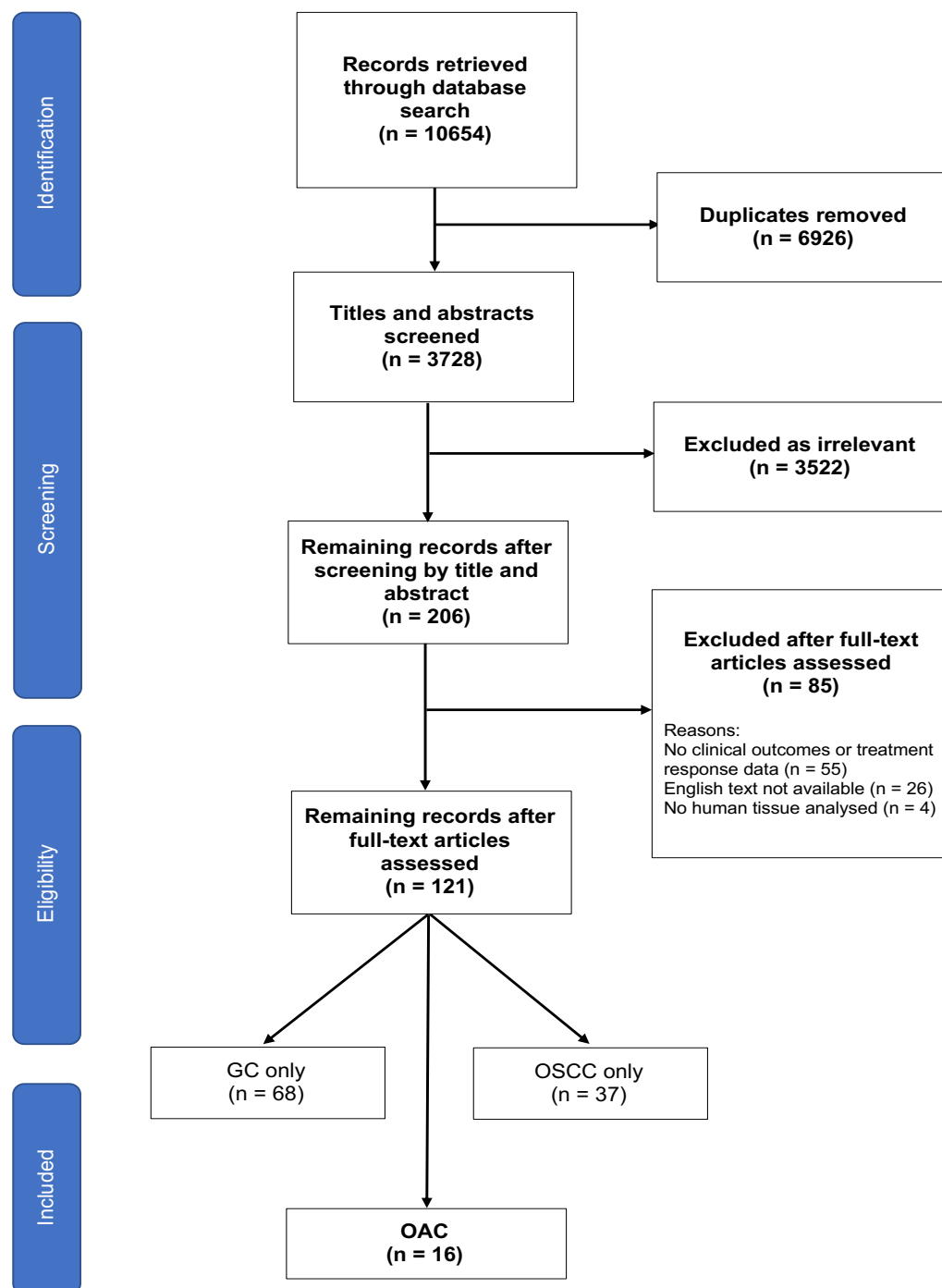
196. Turkington, R. C. *et al.* Immune activation by DNA damage predicts response to chemotherapy and survival in oesophageal adenocarcinoma. *Gut* **gutjnl-2018-317624** (2019) doi:10.1136/gutjnl-2018-317624.
197. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.: Theory Exp.* **2008**, P10008 (2008).
198. Bhandari, V. *et al.* Molecular landmarks of tumor hypoxia across cancer types. *Nat Genet* **51**, 308–318 (2019).
199. Consortium, P., Bhandari, V., Li, C. H., Bristow, R. G. & Boutros, P. C. Divergent mutational processes distinguish hypoxic and normoxic tumours. *Nat Commun* **11**, 737 (2020).
200. Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Cancer Systems Biology, Methods and Protocols. *Methods Mol. Biol.* **1711**, 243–259 (2018).
201. Rankin, E. B., Nam, J.-M. & Giaccia, A. J. Hypoxia: Signaling the Metastatic Cascade. *Trends Cancer* **2**, 295–304 (2016).
202. Fu, Z., Mowday, A. M., Smail, J. B., Hermans, I. F. & Patterson, A. V. Tumour Hypoxia-Mediated Immunosuppression: Mechanisms and Therapeutic Approaches to Improve Cancer Immunotherapy. *Cells* **10**, 1006 (2021).
203. Lohan-Codeço, M. *et al.* Molecular mechanisms associated with chemoresistance in esophageal cancer. *Cell. Mol. Life Sci.* **79**, 116 (2022).
204. Stavropoulos, G. *et al.* Implementation of quality controls is essential to prevent batch effects in breathomics data and allow for cross-study comparisons. *J. Breath Res.* **14**, 026012 (2020).
205. Drabinska, N. *et al.* A literature survey of volatiles from the healthy human breath and bodily fluids: the human volatilome. *J Breath Res* (2021) doi:10.1088/1752-7163/abf1d0.
206. Espina, V. *et al.* Laser-capture microdissection. *Nat. Protoc.* **1**, 586–603 (2006).
207. Wilkinson, M., White, I. R., Goodacre, R., Nijsen, T. & Fowler, S. J. Effects of high relative humidity and dry purging on VOCs obtained during breath sampling on common sorbent tubes. *J. Breath Res.* **14**, 046006 (2020).
208. Parsons, B. A. *et al.* Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC–TOFMS) Data Using a Null Distribution Approach. *Anal Chem* **87**, 3812–3819 (2015).
209. Marney, L. C. *et al.* Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data. *Talanta* **115**, 887–895 (2013).
210. Gaida, M., Cain, C. N., Synovec, R. E., Focant, J.-F. & Stefanuto, P.-H. Tile-Based Random Forest Analysis for Analyte Discovery in Balanced and Unbalanced GC × GC–TOFMS Data Sets. *Anal. Chem.* **95**, 13519–13527 (2023).
211. Pinkerton, D. K., Reaser, B. C., Berrier, K. L. & Synovec, R. E. Determining the Probability of Achieving a Successful Quantitative Analysis for Gas Chromatography–Mass Spectrometry. *Anal. Chem.* **89**, 9926–9933 (2017).
212. Woodfield, G. *et al.* Diagnostic Performance of a Noninvasive Breath Test for Colorectal Cancer: COBRA1 Study. *Gastroenterology* **163**, 1447–1449.e8 (2022).

213. Griffin, M., Abu-El-Haija, M., Abu-El-Haija, M., Rokhlina, T. & Uc, A. Simplified and versatile method for isolation of high-quality RNA from pancreas. *BioTechniques* **52**, 332–334 (2012).
214. 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).
215. Zhang, X. *et al.* Robust Acquisition of Spatial Transcriptional Programs in Tissues With Immunofluorescence-Guided Laser Capture Microdissection. *Front. Cell Dev. Biol.* **10**, 853188 (2022).
216. Cieslik, M. *et al.* The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* **25**, 1372–1381 (2015).
217. Gould, O., Drabińska, N., Ratcliffe, N. & Costello, B. de L. Hyphenated Mass Spectrometry versus Real-Time Mass Spectrometry Techniques for the Detection of Volatile Compounds from the Human Body. *Molecules* **26**, 7185 (2021).
218. Traxler, S., Bischoff, A.-C., Trefz, P., Schubert, J. K. & Miekisch, W. Versatile set-up for non-invasive in vitro analysis of headspace VOCs. *J Breath Res* **12**, 041001 (2018).
219. Shestivska, V. *et al.* Evaluation of lipid peroxidation by the analysis of volatile aldehydes in the headspace of synthetic membranes using selected ion flow tube mass spectrometry. *Rapid Commun Mass Sp* **32**, 1617–1628 (2018).
220. Myridakis, A. *et al.* Global Urinary Volatolomics with (GC×)GC-TOF-MS. *Anal. Chem.* **95**, 17170–17176 (2023).
221. Hayton, S., Maker, G. L., Mullaney, I. & Trengove, R. D. Experimental design and reporting standards for metabolomics studies of mammalian cell lines. *Cell Mol Life Sci* **74**, 4421–4441 (2017).
222. Martinez, A. M., Kim, A. & Yang, W. S. Immune Mediators in Cancer, Methods and Protocols. *Methods Mol. Biol.* **2108**, 125–130 (2020).
223. Jonkman, J., Brown, C. M., Wright, G. D., Anderson, K. I. & North, A. J. Tutorial: guidance for quantitative confocal microscopy. *Nat. Protoc.* **15**, 1585–1611 (2020).
224. Silva, L. P. *et al.* Measurement of DNA Concentration as a Normalization Strategy for Metabolomic Data from Adherent Cell Lines. *Anal. Chem.* **85**, 9536–9542 (2013).
225. Wolfer, A. M. *et al.* peakPanther, an R package for large-scale targeted extraction and integration of annotated metabolic features in LC–MS profiling datasets. *Bioinformatics* **37**, btab433 (2021).
226. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
227. Liu, C., Zhou, X., Zeng, H., Wu, D. & Liu, L. HILPDA Is a Prognostic Biomarker and Correlates With Macrophage Infiltration in Pan-Cancer. *Front. Oncol.* **11**, 597860 (2021).
228. Zhang, X. *et al.* Inhibition of intracellular lipolysis promotes human cancer cell adaptation to hypoxia. *Elife* **6**, e31132 (2017).
229. Zhang, Y. *et al.* HILPDA-mediated lipidomic remodelling promotes radiotherapy resistance in nasopharyngeal carcinoma by accelerating mitophagy. *Cell. Mol. Life Sci.* **80**, 242 (2023).

230. Dijk, W. & Kersten, S. Regulation of lipoprotein lipase by Angptl4. *Trends Endocrinol. Metab.* **25**, 146–155 (2014).
231. Baba, K. *et al.* Hypoxia-induced ANGPTL4 sustains tumour growth and anoikis resistance through different mechanisms in scirrhous gastric cancer cell lines. *Sci. Rep.* **7**, 11127 (2017).
232. Glunde, K. *et al.* Hypoxia Regulates Choline Kinase Expression through Hypoxia-Inducible Factor-1 $\alpha$  Signaling in a Human Prostate Cancer Model. *Cancer Res.* **68**, 172–180 (2008).
233. Saito, R. de F., Andrade, L. N. de S., Bustos, S. O. & Chammas, R. Phosphatidylcholine-Derived Lipid Mediators: The Crosstalk Between Cancer Cells and Immune Cells. *Front. Immunol.* **13**, 768606 (2022).
234. Dada, L. A. *et al.* Hypoxia-induced endocytosis of Na,K-ATPase in alveolar epithelial cells is mediated by mitochondrial reactive oxygen species and PKC- $\zeta$ . *J. Clin. Investig.* **111**, 1057–1064 (2003).
235. Küper, A. *et al.* Overcoming hypoxia-induced resistance of pancreatic and lung tumor cells by disrupting the PERK-NRF2-HIF-axis. *Cell Death Dis.* **12**, 82 (2021).
236. Caldecott, K. W. Protein ADP-ribosylation and the cellular response to DNA strand breaks. *DNA Repair* **19**, 108–113 (2014).
237. Kim, Y. *et al.* Reverse the Resistance to PARP Inhibitors. *Int. J. Biol. Sci.* **13**, 198–208 (2017).
238. Murai, J. *et al.* Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. *Cancer Res.* **72**, 5588–5599 (2012).
239. Saleh-Gohari, N. *et al.* Spontaneous Homologous Recombination Is Induced by Collapsed Replication Forks That Are Caused by Endogenous DNA Single-Strand Breaks. *Mol. Cell. Biol.* **25**, 7158–7169 (2005).
240. Patel, A. G., Sarkaria, J. N. & Kaufmann, S. H. Nonhomologous end joining drives poly(ADP-ribose) polymerase (PARP) inhibitor lethality in homologous recombination-deficient cells. *Proc National Acad Sci* **108**, 3406–3411 (2011).
241. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
242. Kim, H. *et al.* Combining PARP with ATR inhibition overcomes PARP inhibitor and platinum resistance in ovarian cancer models. *Nat Commun* **11**, 3726 (2020).
243. Mehibel, M. *et al.* Eliminating hypoxic tumor cells improves response to PARP inhibitors in homologous recombination-deficient cancer models. *J Clin Invest* **131**, e146256 (2021).
244. Bruin, M. A. C., Sonke, G. S., Beijnen, J. H. & Huitema, A. D. R. Pharmacokinetics and Pharmacodynamics of PARP Inhibitors in Oncology. *Clin. Pharmacokinet.* **61**, 1649–1675 (2022).
245. Liston, D. R. & Davis, M. Clinically Relevant Concentrations of Anticancer Drugs: A Guide for Nonclinical Studies. *Clin Cancer Res* **23**, 3489–3498 (2017).
246. Schoonen, P. M. *et al.* Progression through mitosis promotes PARP inhibitor-induced cytotoxicity in homologous recombination-deficient cancer cells. *Nat Commun* **8**, 15981 (2017).
247. Lieberman, H. B. & Fox, M. H. Cell Cycle Checkpoint Control Protocols. *Methods Mol Biology Clifton N J* **241**, 11–16 (2003).

248. Pettersen, E. O. & Lindmo, T. Inhibition of cell-cycle progression by acute treatment with various degrees of hypoxia: Modifications induced by low concentrations of misonidazole present during hypoxia. *Br. J. Cancer* **48**, 809–817 (1983).
249. Lheureux, S. *et al.* Evolve: A post PARP inhibitor clinical translational phase II trial of cediranib-olaparib in ovarian cancer—A Princess Margaret Consortium – GCIg Phase II Trial. *J. Clin. Oncol.* **37**, 5521–5521 (2019).
250. Tan, S. L. W. *et al.* A Class of Environmental and Endogenous Toxins Induces BRCA2 Haploinsufficiency and Genome Instability. *Cell* **169**, 1105–1118.e15 (2017).
251. Tacconi, E. M. *et al.* BRCA1 and BRCA2 tumor suppressors protect against endogenous acetaldehyde toxicity. *Embo Mol Med* **9**, 1398–1414 (2017).
252. Li, H. *et al.* PARP inhibitor resistance: the underlying mechanisms and clinical implications. *Mol Cancer* **19**, 107 (2020).
253. Dickson, B. D., Wong, W. W., Wilson, W. R. & Hay, M. P. Studies Towards Hypoxia-Activated Prodrugs of PARP Inhibitors. *Molecules* **24**, 1559 (2019).
254. Cartwright, E. *et al.* SOLar: A translational phase II study of single-agent olaparib in the treatment of advanced esophagogastric cancer. *J Clin Oncol* **38**, TPS471–TPS471 (2020).
255. Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**, 228–247 (2009).
256. Nicholson, J. K. *et al.* Metabolic phenotyping in clinical and surgical environments. *Nature* **491**, 384–392 (2012).
257. Filipiak, W. *et al.* A Compendium of Volatile Organic Compounds (VOCs) Released By Human Cell Lines. *Curr Med Chem* **23**, 2112–2131 (2016).

## Appendix 1: PRISMA Flow Chart





## Appendix 2: REMARK tool to assess reporting criteria of prognostic biomarkers

| Section                                                                       |                                                                                                                                                              |                                                                      | Criteria                                                                                                                                                                                                                                                     | %     |
|-------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| Introduction                                                                  |                                                                                                                                                              |                                                                      | Marker stated                                                                                                                                                                                                                                                | 100.0 |
|                                                                               |                                                                                                                                                              |                                                                      | Objective stated                                                                                                                                                                                                                                             | 100.0 |
|                                                                               |                                                                                                                                                              |                                                                      | Pre-specified hypothesis stated                                                                                                                                                                                                                              | 42.9  |
| Materials and Methods                                                         | Patients                                                                                                                                                     | 2. Characteristics                                                   | Source of patients                                                                                                                                                                                                                                           | 100.0 |
|                                                                               |                                                                                                                                                              |                                                                      | Inclusion criteria (i.e. stage of cancer)                                                                                                                                                                                                                    | 64.3  |
|                                                                               |                                                                                                                                                              | Exclusion criteria                                                   | 35.7                                                                                                                                                                                                                                                         |       |
|                                                                               | 3. Treatment details                                                                                                                                         | Details of treatment (e.g. chemo, surgery)                           | 78.6                                                                                                                                                                                                                                                         |       |
|                                                                               |                                                                                                                                                              | Timing of therapy relative to specimen collection                    | 71.4                                                                                                                                                                                                                                                         |       |
|                                                                               | 4. Specimen characteristics                                                                                                                                  |                                                                      | Methods of preservation                                                                                                                                                                                                                                      | 71.4  |
|                                                                               |                                                                                                                                                              |                                                                      | Storage                                                                                                                                                                                                                                                      | 21.4  |
|                                                                               |                                                                                                                                                              |                                                                      | Time between time of storage and time of marker assay                                                                                                                                                                                                        | 0.0   |
|                                                                               | 5. Assay methods                                                                                                                                             |                                                                      | Amount of specimen required to perform the assay                                                                                                                                                                                                             | 0.0   |
|                                                                               |                                                                                                                                                              |                                                                      | Strategies employed to reduce the measurement error                                                                                                                                                                                                          | 7.1   |
|                                                                               |                                                                                                                                                              |                                                                      | Blinding of the person making the marker assessment to clinical outcomes                                                                                                                                                                                     | 21.4  |
|                                                                               | Study design                                                                                                                                                 | 6. Type of study, stratification/matching, time period and follow up | Time period cases were taken                                                                                                                                                                                                                                 | 100.0 |
|                                                                               |                                                                                                                                                              |                                                                      | The end of follow-up period                                                                                                                                                                                                                                  | 35.7  |
|                                                                               |                                                                                                                                                              |                                                                      | Median follow-up time                                                                                                                                                                                                                                        | 28.6  |
|                                                                               |                                                                                                                                                              |                                                                      | Marker measurements were extracted retrospectively from existing records, assays were newly performed using stored specimens, or assays were performed in real time using prospectively collected specimens.                                                 | 92.9  |
|                                                                               |                                                                                                                                                              | 7. Clinical endpoints                                                | The endpoint should be defined precisely                                                                                                                                                                                                                     | 100.0 |
|                                                                               |                                                                                                                                                              | 8. Variables examined                                                | Fully define all variables                                                                                                                                                                                                                                   | 100.0 |
|                                                                               | Statistical analysis methods                                                                                                                                 | 9. Sample size rationale                                             | Either sample size calculation, or effect size calculation given the pre-determined sample size                                                                                                                                                              | 7.1   |
|                                                                               |                                                                                                                                                              | 10. Statistical methods specified                                    | Describe statistical methods with sufficient detail for verification                                                                                                                                                                                         | 100.0 |
|                                                                               |                                                                                                                                                              | 11. How marker values were handled                                   | Must state that "all data was accounted for" or "no missing data occurred"                                                                                                                                                                                   | 0.0   |
|                                                                               |                                                                                                                                                              |                                                                      | For continuous variables: clarify whether the data were kept on the original scale or log transformed, and indicate whether the relationship was modeled as linear or non-linear. For categorized variables: specify the cutpoints and how they were chosen. | 85.7  |
| Results                                                                       | Data                                                                                                                                                         | 12. All patients accounted                                           | The study must show either a flow diagram (e.g. CONSORT), or a study profile diagram                                                                                                                                                                         | 0.0   |
|                                                                               |                                                                                                                                                              | 13. Descriptive statistics reported                                  | Report the number of patients and the number of events                                                                                                                                                                                                       | 85.7  |
|                                                                               |                                                                                                                                                              |                                                                      | Distributions of basic demographic variables and standard prognostic variables                                                                                                                                                                               | 100.0 |
|                                                                               | Analysis and presentation                                                                                                                                    | 14. Relation of marker to prognostic variable                        | Description of the distribution of the marker of interest should also be provided                                                                                                                                                                            | 100.0 |
|                                                                               |                                                                                                                                                              |                                                                      | The association of the tumor marker with standard prognostic variables                                                                                                                                                                                       | 100.0 |
|                                                                               |                                                                                                                                                              | 15. Univariate analysis                                              | The relation between a categorical marker and outcome                                                                                                                                                                                                        | 100.0 |
|                                                                               |                                                                                                                                                              |                                                                      | Confidence intervals                                                                                                                                                                                                                                         | 21.4  |
|                                                                               |                                                                                                                                                              |                                                                      | P-value                                                                                                                                                                                                                                                      | 100.0 |
|                                                                               |                                                                                                                                                              | 16. Multivariate analysis                                            | The relation between a categorical marker and outcome                                                                                                                                                                                                        | 42.9  |
|                                                                               |                                                                                                                                                              |                                                                      | Confidence intervals                                                                                                                                                                                                                                         | 14.3  |
|                                                                               |                                                                                                                                                              |                                                                      | P-value                                                                                                                                                                                                                                                      | 50.0  |
|                                                                               |                                                                                                                                                              | 17. Estimated effects and CI                                         | The study must evaluate whether the new marker maintains some association with clinical outcome after accounting for these standard prognostic variables                                                                                                     | 50.0  |
|                                                                               |                                                                                                                                                              |                                                                      | Confidence intervals                                                                                                                                                                                                                                         | 14.3  |
| P-value                                                                       | 50.0                                                                                                                                                         |                                                                      |                                                                                                                                                                                                                                                              |       |
| Discussion and explanation of how these standard variables have been selected | 21.4                                                                                                                                                         |                                                                      |                                                                                                                                                                                                                                                              |       |
| 18. Further investigations                                                    | Must test their model with one of the following: test their assumption, sensitivity analysis, or internal validation analyses or external validation studies | 7.1                                                                  |                                                                                                                                                                                                                                                              |       |
|                                                                               | Discussion                                                                                                                                                   | 19. Interpretation of results                                        | One of the following: acknowledgment of any biases or inconsistencies in the data, limitations of the assay methods, or limitations of the design or data analysis methods                                                                                   | 42.9  |
|                                                                               |                                                                                                                                                              |                                                                      | Comment on whether the results are consistent with, or differ from, the general tendency in previous studies and offer potential explanations for differences.                                                                                               | 85.7  |
| 20. Future directions                                                         |                                                                                                                                                              | A discussion if the biomarker is clinically useful                   | 85.7                                                                                                                                                                                                                                                         |       |
|                                                                               | Future research plans                                                                                                                                        | 21.4                                                                 |                                                                                                                                                                                                                                                              |       |

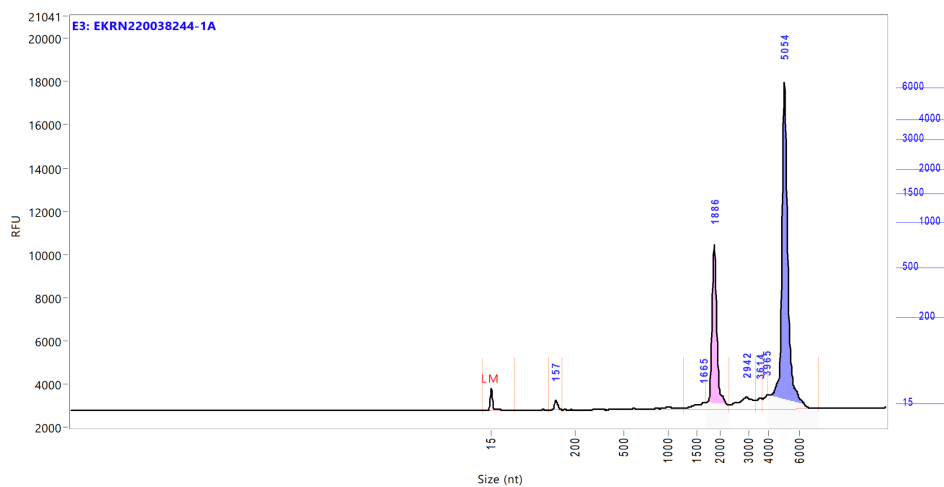
### Appendix 3: Primer sequences for the Toustrup hypoxia gene signature

| Target          | Sequence                |
|-----------------|-------------------------|
| ADM-Forward     | CCTGCCCAGACCCTTATTCG    |
| ADM-Reverse     | CCGGACTGCTGTCTTCGG      |
| ALDOA-Forward   | ATGCCCTACCAATATCCAGCA   |
| ALDOA-Reverse   | GCTCCCAGTGGACTCATCTG    |
| ANKRD37-Forward | TGCTCAGACAGAAACGGAGTC   |
| ANKRD37-Reverse | AACCCACGTGACATCAGCAC    |
| BNIP3-Forward   | CAGGGCTCCTGGGTAGAACT    |
| BNIP3-Reverse   | CTACTCCGTCCAGACTCATGC   |
| BNIP3L-Forward  | TTGGATGCACAACATGAATCAGG |
| BNIP3L-Reverse  | TCTTCTGACTGAGAGCTATGGTC |
| EGLN3-Forward   | GGCGTCTCCAAGCGACAC      |
| EGLN3-Reverse   | GCCACCATTGCCTTAGACCT    |
| KCTD11-Forward  | GAGGGGAGCCCACATTTTCA    |
| KCTD11-Reverse  | GAGTCTAGTCGGAAGCCGTG    |
| LOX-Forward     | TGCGCTCAGTAACCTGCGG     |
| LOX-Reverse     | CACCAGGTCTGGGAGACCGTA   |
| NDRG1-Forward   | CGCTGAGGTGAAGCCTTTGG    |
| NDRG1-Reverse   | CCTGCTCCTGGACATCAAACCTC |
| P4HA1-Forward   | TGACCACAGCACAGTACAGAGT  |
| P4HA1-Reverse   | TCTGGCTCATCTTCCGTGCAA   |
| P4HA2-Forward   | CAAACCTGGTGAAGCGGCTAAA  |
| P4HA2-Reverse   | GCACAGAGAGGTTGGCGATA    |
| PDK1-Forward    | CTGTGATACGGATCAGAAACCG  |
| PDK1-Reverse    | TCCACCAAACAATAAAGAGTGCT |
| PFKFB3-Forward  | ATTGCGGTTTTCGATGCCAC    |
| PFKFB3-Reverse  | GCCACAACCTGTAGGGTCGT    |
| SLC2A1-Forward  | TCTGGCATCAACGCTGTCTTC   |
| SLC2A1-Reverse  | CGATACCGGAGCCAATGGT     |
| ACTR3-Forward   | TTGAGTGGTGGTAGATTGAAGC  |
| ACTR3-Reverse   | CCAAACTGCATATCGCTGCAT   |
| NDFIP1-Forward  | GGTGATGCTCCTCCACCTTAC   |
| NDFIP1-Reverse  | TCCTCATCTCTCCAGGAACCA   |
| RPL37A-Forward  | AGAAAGTCGGGATCGTCGGT    |
| RPL37A-Reverse  | CGTCTCTTCATCTTGGTTTTGCC |

## Appendix 4: Quality control for cell line RNA sequencing

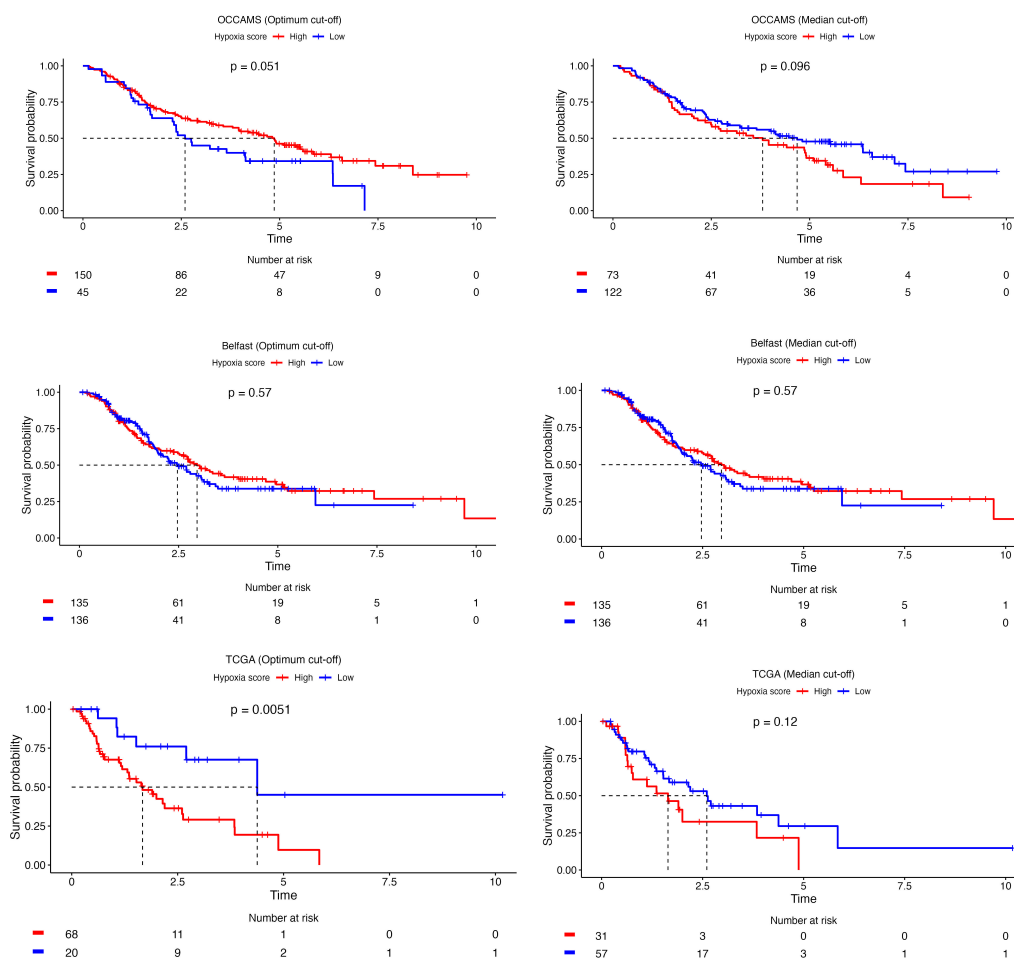
## RNA INTEGRITY ANALYSIS (AGILENT BIOANALYZER)

| No. | Sample Name | Nucleic Acid ID  | Concentration(ng/ul) | Volume(ul) | Total amount(ug) | RIN | Sample QC Results | Sample QC Memo |
|-----|-------------|------------------|----------------------|------------|------------------|-----|-------------------|----------------|
| 1   | FLO1_Nor_1  | EKRN220038244-1A | 192.06               | 12         | 2.30468          | 10  | Pass              | N/A            |
| 2   | FLO1_Nor_2  | EKRN220038245-1A | 158.26               | 15         | 2.37386          | 10  | Pass              | N/A            |
| 3   | FLO1_Nor_3  | EKRN220038246-1A | 87.56                | 12         | 1.05071          | 10  | Pass              | N/A            |
| 4   | FLO1_Hyp2_1 | EKRN220038247-1A | 39.70                | 13         | 0.51615          | 10  | Pass              | N/A            |
| 5   | FLO1_Hyp2_2 | EKRN220038248-1A | 152.43               | 12         | 1.82918          | 10  | Pass              | N/A            |
| 6   | FLO1_Hyp2_3 | EKRN220038249-1A | 160.97               | 10         | 1.60972          | 10  | Pass              | N/A            |
| 7   | OE33_Nor_1  | EKRN220038250-1A | 621.18               | 14         | 8.69653          | 10  | Pass              | N/A            |
| 8   | OE33_Nor_2  | EKRN220038251-1A | 828.61               | 13         | 10.77196         | 10  | Pass              | N/A            |
| 9   | OE33_Nor_3  | EKRN220038252-1A | 841.33               | 15         | 12.62002         | 10  | Pass              | N/A            |
| 10  | OE33_Hyp2_1 | EKRN220038253-1A | 769.41               | 12         | 9.23292          | 10  | Pass              | N/A            |
| 11  | OE33_Hyp2_2 | EKRN220038254-1A | 352.19               | 11         | 3.87411          | 10  | Pass              | N/A            |
| 12  | OE33_Hyp2_3 | EKRN220038255-1A | 402.78               | 13         | 5.23617          | 10  | Pass              | N/A            |
| 13  | OE19_Nor_1  | EKRN220038256-1A | 648.93               | 14         | 9.08503          | 10  | Pass              | N/A            |
| 14  | OE19_Nor_2  | EKRN220038257-1A | 795.70               | 15         | 11.93552         | 10  | Pass              | N/A            |
| 15  | OE19_Nor_3  | EKRN220038258-1A | 740.01               | 14         | 10.36008         | 10  | Pass              | N/A            |
| 16  | OE19_Hyp2_1 | EKRN220038259-1A | 505.78               | 13         | 6.57519          | 10  | Pass              | N/A            |
| 17  | OE19_Hyp2_2 | EKRN220038260-1A | 467.21               | 16         | 7.47542          | 10  | Pass              | N/A            |
| 18  | OE19_Hyp2_3 | EKRN220038261-1A | 398.02               | 15         | 5.97034          | 10  | Pass              | N/A            |



## Appendix 5: Prognostic utility of the Toustrup hypoxia gene signature

### TOUSTRUP SIGNATURE



Kaplan-Meier plots showing survival curves for hypoxia-high and hypoxia-low subgroups defined by either an optimum cut-off score selected by *survminer* or median hypoxia score. A log-rank test was used to compare survival curves, and calculated P values are shown/

## Appendix 6: Baseline characteristics of patients in OCCAMS and TCGA-ESCA (OAC) datasets

| OCCAMS                   |                               |                           |                           |                      |
|--------------------------|-------------------------------|---------------------------|---------------------------|----------------------|
| Baseline Characteristics |                               |                           |                           |                      |
| Hypoxia status           | Overall, N = 195 <sup>1</sup> | High, N = 80 <sup>1</sup> | Low, N = 115 <sup>1</sup> | p-value <sup>2</sup> |
| <b>Age</b>               | 67 (59, 74)                   | 68 (60, 75)               | 66 (59, 73)               | 0.52                 |
| <b>Gender</b>            |                               |                           |                           | 0.45                 |
| Female                   | 34 (17%)                      | 12 (15%)                  | 22 (19%)                  |                      |
| Male                     | 161 (83%)                     | 68 (85%)                  | 93 (81%)                  |                      |
| <b>Smoking</b>           |                               |                           |                           | 0.70                 |
| Current                  | 26 (13%)                      | 8 (10%)                   | 18 (16%)                  |                      |
| Former                   | 97 (50%)                      | 41 (51%)                  | 56 (49%)                  |                      |
| Never                    | 48 (25%)                      | 20 (25%)                  | 28 (24%)                  |                      |
| Unknown                  | 24 (12%)                      | 11 (14%)                  | 13 (11%)                  |                      |
| <b>Stage</b>             |                               |                           |                           | 0.93                 |
| Stage 1                  | 21 (11%)                      | 7 (8.8%)                  | 14 (12%)                  |                      |
| Stage 2A                 | 3 (1.5%)                      | 1 (1.3%)                  | 2 (1.7%)                  |                      |
| Stage 2B                 | 22 (11%)                      | 11 (14%)                  | 11 (9.6%)                 |                      |
| Stage 3                  | 98 (50%)                      | 41 (51%)                  | 57 (50%)                  |                      |
| Stage 4A                 | 30 (15%)                      | 13 (16%)                  | 17 (15%)                  |                      |
| Stage 4B                 | 17 (8.7%)                     | 6 (7.5%)                  | 11 (9.6%)                 |                      |
| Unknown                  | 4 (2.1%)                      | 1 (1.3%)                  | 3 (2.6%)                  |                      |
| <b>Tumour grade</b>      |                               |                           |                           | 0.96                 |
| Moderate                 | 63 (32%)                      | 24 (30%)                  | 39 (34%)                  |                      |
| Moderate to poor         | 29 (15%)                      | 12 (15%)                  | 17 (15%)                  |                      |
| Moderate to well         | 3 (1.5%)                      | 1 (1.3%)                  | 2 (1.7%)                  |                      |
| Poor                     | 43 (22%)                      | 20 (25%)                  | 23 (20%)                  |                      |
| Unknown                  | 53 (27%)                      | 21 (26%)                  | 32 (28%)                  |                      |
| Well                     | 4 (2.1%)                      | 2 (2.5%)                  | 2 (1.7%)                  |                      |

<sup>1</sup>Median (IQR); n (%)

<sup>2</sup>Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

| TCGA-ESCA (OAC)                    |                              |                           |                          |                      |
|------------------------------------|------------------------------|---------------------------|--------------------------|----------------------|
| Baseline Characteristics           |                              |                           |                          |                      |
| Hypoxia status                     | Overall, N = 71 <sup>1</sup> | High, N = 28 <sup>1</sup> | Low, N = 43 <sup>1</sup> | p-value <sup>2</sup> |
| <b>Age</b>                         | 71 (59, 77)                  | 73 (56, 77)               | 70 (60, 77)              | 0.70                 |
| <b>Gender</b>                      |                              |                           |                          | 0.20                 |
| Female                             | 12 (17%)                     | 7 (25%)                   | 5 (12%)                  |                      |
| Male                               | 59 (83%)                     | 21 (75%)                  | 38 (88%)                 |                      |
| <b>Smoking</b>                     |                              |                           |                          | 0.70                 |
| Lifelong non-smoker                | 20 (33%)                     | 7 (29%)                   | 13 (35%)                 |                      |
| Current smoker                     | 7 (11%)                      | 2 (8.3%)                  | 5 (14%)                  |                      |
| Current reformed smoker > 15 years | 25 (41%)                     | 10 (42%)                  | 15 (41%)                 |                      |
| Current reformed smoker < 15 years | 9 (15%)                      | 5 (21%)                   | 4 (11%)                  |                      |
| <b>Stage</b>                       |                              |                           |                          | 0.12                 |
| Stage 1                            | 12 (21%)                     | 2 (8.3%)                  | 10 (29%)                 |                      |
| Stage 2                            | 21 (36%)                     | 11 (46%)                  | 10 (29%)                 |                      |
| Stage 3                            | 25 (43%)                     | 11 (46%)                  | 14 (41%)                 |                      |

<sup>1</sup>Median (IQR); n (%)

<sup>2</sup>Wilcoxon rank sum test; Fisher's exact test

## Appendix 7: Ethical approval for biosample collection



Imperial College  
London

Imperial College Healthcare Tissue Bank  
Department of Surgery and Cancer,  
Charing Cross Hospital  
Fulham Palace Road  
London W6 8RF  
Tel: 020 3311 7173  
Email: [tissuebank@imperial.ac.uk](mailto:tissuebank@imperial.ac.uk)

**ICHTB HTA licence: 12275**  
**REC approval: 17/WA/0161**

Professor George Hanna  
QEOM Building  
Imperial College London  
St Mary's Campus  
London

21<sup>st</sup> August 2020

Dear Professor Hanna,

**Re: Application for Access Approval. R20037 – Clinical Applications Of Volatile Hypoxia Signatures In Oesophageal Adenocarcinoma**

This is to confirm that the above project has been approved by ICHTB. REC3 Wales has confirmed that approval via ICHTB Application Review Panel provides you with deemed ethics approval for your research project. Details of your research project will be provided to REC 3 Wales through ICHTB's annual report.

ICHTB is supported by funding from the NIHR Imperial Biomedical Research Centre (BRC). Acknowledgement of this infrastructural report is required in any publication that results from your research project, and you have agreed to this by signing our MDTA as part of your application. The recommended wording is as follows: "Infrastructure support for this research was provided by the NIHR Imperial Biomedical Research Centre (BRC) and ICHTB". Further information on Imperial's BRC can be found here <https://imperialbrc.nihr.ac.uk/about-us/acknowledgement/>.

We would be grateful if you would send copies of or doi links for any publication that results from this study to the Tissue Bank ([tissuebank@imperial.ac.uk](mailto:tissuebank@imperial.ac.uk)). Your support with this helps us to keep the Tissue Bank running as a research resource for all.

If you have any questions, please see contact details above.

Yours sincerely,

Professor Iain McNeish  
DI, HTA Research Licence.



**Imperial College Healthcare Tissue Bank**  
**Department of Surgery and Cancer**  
**Charing Cross Hospital**  
**Fulham Palace Road**  
**London W6 8RF**  
**Tel: 0203 311 7173**  
**Email: [tissuebank@imperial.ac.uk](mailto:tissuebank@imperial.ac.uk)**

**ICHTB HTA licence: 12275**  
**REC Wales approval: 17/WA/0161**

Professor George Hanna  
QEQM Building Imperial College London  
St Mary's Campus  
London

Tuesday 13<sup>th</sup> of October 2020

Dear Professor Hanna

**Re: Tissue Bank application number –R20037**  
**Project title: Clinical Applications of Volatile Hypoxia Signatures in Oesophageal Adenocarcinoma**

I am pleased to confirm that the Imperial College Healthcare Tissue Bank has received and approved the following amendments;

Amendment 1A-

- Retrieval of paired breath samples in addition to tissue.
- Clarifying up to six biopsies per patient will be retrieved. The sample size remains the same.
- Mention of processing tissue for direct mass spectrometry analysis.

In order to satisfy HTA tracking requirements, would you please ensure the samples you use are reported to the tissue bank via the online database and sub-collection, along with any publications you produce, that should contain the acknowledgment wording as per the signed MTA.

Yours sincerely,

Professor Iain McNeish  
DI, HTA Research Licence.





**Imperial College Healthcare Tissue Bank**  
**Department of Surgery and Cancer**  
**Charing Cross Hospital**  
**Fulham Palace Road**  
**London W6 8RF**  
**Tel: 0203 311 7173**  
**Email: [tissuebank@imperial.ac.uk](mailto:tissuebank@imperial.ac.uk)**

**ICHTB HTA licence: 12275**  
**REC Wales approval: 17/WA/0161**

22<sup>nd</sup> October 2021

Dear Professor George Hanna,

**Re: Project Amendment Application Approval**  
**R20037-2A: Clinical Applications Of Volatile Hypoxia Signatures In Oesophageal**  
**Adenocarcinoma (Expires: Sep 2023)**

I am pleased to confirm that the Imperial College Healthcare Tissue Bank has received and approved the following amendment:

Amendment 2A:

- Time extension to 30/09/23

In order to satisfy HTA tracking requirements, please ensure that the samples you use are reported to the Tissue Bank via the online database and Sub-Collection. Please also make sure that any publications you produce, contain the acknowledgment wording as per the signed MTA.

Yours sincerely,

Professor Iain McNeish  
DI, HTA Research Licence.



## Appendix 8: Targeted quantification settings for the TD-GC-ToF-MS mid-polar method

| Compound                   | t <sub>R</sub> | Quantitative fragment ion | ISTD        |
|----------------------------|----------------|---------------------------|-------------|
| Pentane                    | 2.42           | 57                        | Propanol-d8 |
| Ethanol                    | 2.44           | 45                        | Propanol-d8 |
| Propionaldehyde            | 2.69           | 58                        | Propanol-d8 |
| Acetone                    | 2.75           | 58                        | Propanol-d8 |
| Propanol-d8                | 2.80           | 49                        | ISTD        |
| Dimethyl Sulfide           | 2.82           | 62                        | Propanol-d8 |
| Carbon Disulfide           | 2.93           | 76                        | Propanol-d8 |
| Hexane                     | 3.56           | 57                        | Propanol-d8 |
| 1-Propanol                 | 3.69           | 60                        | Propanol-d8 |
| Butyraldehyde              | 3.99           | 72                        | Toluene-d8  |
| 2-Butanone                 | 4.13           | 72                        | Toluene-d8  |
| Cyclohexane                | 4.68           | 56                        | Toluene-d8  |
| N-Heptane                  | 5.21           | 57                        | Toluene-d8  |
| N-Butanol                  | 5.43           | 56                        | Toluene-d8  |
| Allyl Methyl Sulfide       | 5.54           | 88                        | Toluene-d8  |
| Allyl Acetate              | 5.54           | 61                        | Toluene-d8  |
| 2-Pentanone                | 5.70           | 86                        | Toluene-d8  |
| Valeraldehyde              | 5.84           | 86                        | Toluene-d8  |
| 3-Ethylhexane              | 6.63           | 85                        | Toluene-d8  |
| Toluene-d8                 | 6.84           | 100                       | ISTD        |
| Octane-d18                 | 6.86           | 50                        | ISTD        |
| N-Octane                   | 7.09           | 57                        | Toluene-d8  |
| 2-Hexanone                 | 7.66           | 100                       | Toluene-d8  |
| Propyl Propionate          | 7.70           | 75                        | Toluene-d8  |
| Hexanal                    | 7.80           | 56                        | Toluene-d8  |
| N-Nonane                   | 9.00           | 57                        | Toluene-d8  |
| Butyl Propionate           | 9.56           | 57                        | Toluene-d8  |
| Heptanal                   | 9.75           | 70                        | Toluene-d8  |
| Cyclohexanone              | 10.00          | 98                        | Toluene-d8  |
| N-decane                   | 10.82          | 57                        | Toluene-d8  |
| 1-Octen-3-ol               | 11.09          | 57                        | Toluene-d8  |
| Benzaldehyde-d6            | 11.21          | 110                       | ISTD        |
| Benzaldehyde               | 11.23          | 105                       | Toluene-d8  |
| 2-Octanone                 | 11.40          | 58                        | Toluene-d8  |
| Octyl Aldehyde             | 11.58          | 84                        | Toluene-d8  |
| Phenol                     | 12.43          | 94                        | Toluene-d8  |
| n-Undecane                 | 12.54          | 57                        | Toluene-d8  |
| 1-Nonen-3-ol               | 12.90          | 57                        | Toluene-d8  |
| Trans-2-nonen-1-ol         | 12.90          | 57                        | Toluene-d8  |
| Acetophenone-d8            | 13.03          | 110                       | ISTD        |
| Nonyl Aldehyde (1-Nonanal) | 13.30          | 57                        | Toluene-d8  |
| Methyl Benzoate            | 13.31          | 136                       | Toluene-d8  |
| N-Dodecane                 | 14.14          | 57                        | Toluene-d8  |
| Nonanol                    | 14.39          | 56                        | Toluene-d8  |
| Decanal                    | 14.91          | 57                        | Toluene-d8  |
| Naphthalene-d8             | 15.08          | 136                       | Toluene-d8  |
| Benzothiazole              | 15.69          | 135                       | Toluene-d8  |
| 2-Phenoxyethanol           | 15.76          | 94                        | Toluene-d8  |
| Trans-2-decenol            | 15.88          | 57                        | Toluene-d8  |
| Undecanal                  | 16.41          | 82                        | Toluene-d8  |
| Undecanol                  | 17.32          | 55                        | Toluene-d8  |
| Dodecyl Aldehyde           | 17.82          | 57                        | Toluene-d8  |
| 1-Dodecanol                | 18.65          | 55                        | Toluene-d8  |
| Tridecanal                 | 19.15          | 82                        | Toluene-d8  |
| 1-Tridecanol               | 19.64          | 83                        | Toluene-d8  |
| Cyclododecanol             | 20.19          | 82                        | Toluene-d8  |
| Tetradecanol               | 21.11          | 83                        | Toluene-d8  |
| 1-Pentadecanol             | 22.26          | 83                        | Toluene-d8  |

## Appendix 9: Targeted quantification settings for the TD-GC-ToF-MS polar method

| Compound        | t <sub>R</sub> | Quantitative fragment ion | ISTD       |
|-----------------|----------------|---------------------------|------------|
| Octane-d18      | 2.00           | 50                        | ISTD       |
| Propanol-d8     | 2.85           | 49                        | ISTD       |
| Toluene-d8      | 4.09           | 100                       | ISTD       |
| Acetic acid     | 9.51           | 60                        | Toluene-d8 |
| Propanoic acid  | 10.55          | 45                        | Toluene-d8 |
| Benzaldehyde-d6 | 10.66          | 110                       | ISTD       |
| Butanoic acid   | 11.60          | 60                        | Toluene-d8 |
| Acetophenone-d8 | 12.05          | 110                       | ISTD       |
| Pentanoic acid  | 12.85          | 60                        | Toluene-d8 |
| Naphthalene-d8  | 13.54          | 136                       | Toluene-d8 |
| Hexanoic acid   | 14.01          | 60                        | Toluene-d8 |
| Heptanoic acid  | 15.12          | 60                        | Toluene-d8 |
| Octanoic acid   | 16.16          | 60                        | Toluene-d8 |
| Nonanoic acid   | 17.19          | 60                        | Toluene-d8 |
| Decanoic acid   | 18.13          | 60                        | Toluene-d8 |
| Undecanoic acid | 19.07          | 60                        | Toluene-d8 |

## Appendix 10: compiler.R

```
library(tidyverse)

#input name of folder in your working directory with csv files to compile
nameFolder <- "compile"

#run compilation code - this may take a few minutes to process with many files
vectorFiles <- list.files(nameFolder, full.names = TRUE)
listDf <- lapply(vectorFiles, function(fileName){
  dfFile <- read.csv(fileName, header=T)[-1]
  return(dfFile)})

all_rows = unique(unlist(lapply(listDf, rownames)))
all_cols = unique(unlist(lapply(listDf, colnames)))
final_mat = matrix(NA, nrow = length(all_rows), ncol = length(all_cols))
rownames(final_mat) = all_rows
colnames(final_mat) = all_cols
for(i in seq_along(listDf)) {
  final_mat[rownames(listDf[[i]]), colnames(listDf[[i]])[is.na(final_mat[rownames(listDf[[i]]),
colnames(listDf[[i]])])] = 0
  final_mat[rownames(listDf[[i]]), colnames(listDf[[i]])] = final_mat[rownames(listDf[[i]]),
colnames(listDf[[i]])] + as.matrix(listDf[[i]])
}
final_mat <- as.data.frame(final_mat)
listDf2 <- lapply(vectorFiles, function(fileName){
  dfFile <- read.csv(fileName, header=T)
  return(dfFile)})
df1<-listDf2[[1]]
RT<-df1$RT
compiled<-data.frame(RT, final_mat)
compiled$TIC[is.na(compiled$TIC)] <- ""
names<-colnames(compiled)
names<-str_remove(names, "X")
colnames(compiled)<-names

#run below to write final file - adjust output file name as required - this will be deposited in the
working directory
write.csv(compiled, "compiled.csv", row.names=F)
```

## Appendix 11: composite\_file\_processing.R

```
library(tidyverse)

df1 <- read.csv('batch1_composite.csv',header=T, na.strings = c("", "NA"))
df1 <- df1 %>% filter(!is.na(Compound)) #Removes any unidentified compounds (RMF < 800)
df1 <- df1 %>% filter(RT > 1.2) #Removes retention time < 1.2
df1 <- df1 %>% filter(Siloxane == "0") #Removes siloxanes
df1 <- df1 %>% filter(Signal.to.Noise > 3) #Removes signal-to-noise < 3
df1 <- df1 %>% filter(!grepl('fluor|chlor|sil|Sil|deriv|osane',Compound)) #Removes known
contaminants
df1 <- df1 %>% select(Compound, RT, M1)
write.csv(df1, "batch1_composite_peaklist.csv",row.names=F)
```

## Appendix 12: biofluids\_normalization\_filtering.R

```
library(ggpubr)
library(corrplot)
library(Hmisc)
library(tidyverse)
library(stats)
library(stringr)

### Data prep

df1 <- read.csv("batch1.csv", header=F, sep = '\t')
df1<-as.data.frame(df1[-(1:3),])
colnames(df1)[1]<-"X"
df1 <- as.data.frame(str_split_fixed(df1$X,"", 19))
colnames(df1)<-df1[1,]
df1<-df1[-1,]
df1<-df1%>%dplyr::select(`Compound name`, `File name`, Area)
ID <- data.frame(do.call('rbind', strsplit(as.character(df1$`File name`),'_',fixed=TRUE)))
ID <- ID %>% dplyr::select(X4,X5)
ID <- ID %>% unite('ID',X4:X5, sep = "_", remove = TRUE)
df1 <- cbind(df1, ID)
df1 <- df1 %>% group_by(`File name`, ID, `Compound name`) %>% summarise(`Area` = max(`Area`))
df1$Area <- as.numeric(df1$Area)
df1 <- df1 %>% pivot_wider(names_from = "Compound name", values_from = "Area")
df1 <- as.data.frame(df1)
write.csv(df1, 'batch2_semitargeted.csv',row.names=F)

#### Blank filtering

istd <- c('Toluene-D8',
          'Benzaldehyde-d6',
          'Octane-d18',
          'Propanol-d8',
          'Acetophenone d8',
          'Naphthalene-d8')

df1 <- read.csv('batch2_semitargeted.csv',header=F)
colnames(df1) <- df1[1,]
df1 <- df1[-1,]
df1.qc <- df1 %>% filter(ID == "Pooled_QC")
df1.qc <- t(df1.qc)
colnames(df1.qc) <- df1.qc[1,]
df1.qc <- df1.qc[-1,]
df1.qc <- df1.qc[-1,]
a<-length(colnames(df1.qc))
df1.qc <- as.data.frame(df1.qc)
df1.qc[,c(1:a)]<-sapply(df1.qc[,c(1:a)],as.numeric)
df1.qc$qc_mean <- apply(df1.qc,1,mean)
df1.blank <- df1 %>% filter(ID == "PBS_Blank")
df1.blank <- t(df1.blank)
```

```

colnames(df1.blank) <- df1.blank[1,]
df1.blank <- df1.blank[-1,]
df1.blank <- df1.blank[-1,]
a<-length(colnames(df1.blank))
df1.blank <- as.data.frame(df1.blank)
df1.blank[,c(1:a)]<-sapply(df1.blank[,c(1:a)],as.numeric)
df1.blank$blank_mean <- apply(df1.blank,1,mean)
df1.qc$blank_mean <- df1.blank$blank_mean
df1.qc <- df1.qc %>% drop_na(qc_mean)
df1.qc$blank_ratio <- df1.qc$blank_mean/df1.qc$qc_mean
df1.qc <- df1.qc %>% filter(blank_ratio < 0.6)
blankfiltered_compounds <- row.names(df1.qc)

```

#### #### Linearity filtering

```

data<-read.csv("batch2_semitargeted.csv",header=F)
data <- t(data)
colnames(data) <- data[1,]
data <- as.data.frame(data)
data <- data %>% dplyr::select(1:10)
data <- data %>% dplyr::filter(`File name` %in% blankfiltered_compounds)
data.corr <- data %>% add_row(`File name` = "Concentration", .before = 1)
names <- data.corr$`File name`
data.corr$`File name` <- NULL
row.names(data.corr) <- NULL
data.corr[1,1:9]<-c("0","0","0","0.25","0.50","0.75","1","1","1")
data.corr<-t(data.corr)
data.corr <- matrix(as.numeric(data.corr), ncol = ncol(data.corr))
corr_mat<-rcorr(data.corr,type="spearman")
mat<-corr_mat$r
mat<-as.data.frame(mat)
pval<-corr_mat$p
pval<-as.data.frame(pval)
mat <- mat %>% dplyr::select(V1)
colnames(mat)[1]<-"rho"
pval <- pval %>% dplyr::select(V1)
colnames(pval)[1]<-"p"
mat$p <- pval$p
mat$compounds<-names
mat<-mat[-1,]
mat$padj<-p.adjust(mat$p, method = "fdr")
mat <- mat %>% dplyr::filter(!compounds %in% istd)
mat <- mat %>% filter(padj < 0.05)
mat <- mat %>% filter(rho > 0.7)
linear_compounds<-mat$compounds

```

#### ### ISTD Selection, Normalization and CV filtering

```

base <- read.csv("batch1.csv", header=F, sep = '\t')
base<-as.data.frame(base[-(1:3),])
colnames(base)[1]<-"X"

```

```

base <- as.data.frame(str_split_fixed(base$X,"", 19))
colnames(base)<-base[1,]
base<-base[-1,]
base<-base%>%dplyr::select(`Compound name`, `File name`, Area)
base<-base%>%pivot_wider(names_from = "Compound name", values_from = "Area", values_fn =
max)
names<-colnames(base)
base <- data.frame(lapply(base, function(x) {gsub("n/a", NA, x)}))
colnames(base)<-names
z <- length(colnames(base))
base[,c(2:z)]<-sapply(base[,c(2:z)],as.numeric)
base <- base %>% dplyr::filter(grepl('20220106-001', `File name`)) #Selection of date of batch run
base_QC <- base %>% dplyr::filter(grepl('Pooled_QC', `File name`))
cv <- function(x, na.rm = FALSE) sd(x, na.rm = na.rm)/mean(x, na.rm = na.rm)

```

```

df1 <- base_QC
df1.istd <- df1 %>% dplyr::select('Toluene-D8')
df1 <- df1 %>% dplyr::select(!'Acetophenone d8')
df1 <- df1 %>% dplyr::select(!'Benzaldehyde-d6')
df1 <- df1 %>% dplyr::select(!'Octane-d18')
df1 <- df1 %>% dplyr::select(!'Propanol-d8')
df1 <- df1 %>% dplyr::select(!'Toluene-D8')
df1 <- df1 %>% dplyr::select(!"File name")
df1 <- cbind(df1, df1.istd)
a <- length(colnames(df1))
df1 <- df1/df1[,a]
df1[a]<- NULL
df1$rep <- c('1','2','3')
Toluene_d8_r2 <- df1 %>%
  pivot_longer(-rep, names_to="Compound", values_to = "area") %>%
  group_by(Compound) %>%
  summarise(cv = cv(area))
colnames(Toluene_d8_r2)[2]<-"Toluene-D8"

```

```

df1 <- base_QC
df1.istd <- df1 %>% dplyr::select('Acetophenone d8')
df1 <- df1 %>% dplyr::select(!'Acetophenone d8')
df1 <- df1 %>% dplyr::select(!'Benzaldehyde-d6')
df1 <- df1 %>% dplyr::select(!'Octane-d18')
df1 <- df1 %>% dplyr::select(!'Propanol-d8')
df1 <- df1 %>% dplyr::select(!'Toluene-D8')
df1 <- df1 %>% dplyr::select(!"File name")
df1 <- cbind(df1, df1.istd)
a <- length(colnames(df1))
df1 <- df1/df1[,a]
df1[a]<- NULL
df1$rep <- c('1','2','3')
Acetophenone_d8_r2 <- df1 %>%
  pivot_longer(-rep, names_to="Compound", values_to = "area") %>%
  group_by(Compound) %>%
  summarise(cv = cv(area))

```

```

colnames(Acetophenone_d8_r2)[2]<-"Acetophenone d8"

df1 <- base_QC
df1.istd <- df1 %>% dplyr::select('Benzaldehyde-d6')
df1 <- df1 %>% dplyr::select(!'Acetophenone d8')
df1 <- df1 %>% dplyr::select(!'Benzaldehyde-d6')
df1 <- df1 %>% dplyr::select(!'Octane-d18')
df1 <- df1 %>% dplyr::select(!'Propanol-d8')
df1 <- df1 %>% dplyr::select(!'Toluene-D8')
df1 <- df1 %>% dplyr::select(!"File name")
df1 <- cbind(df1, df1.istd)
a <- length(colnames(df1))
df1 <- df1/df1[,a]
df1[a]<- NULL
df1$rep <- c('1','2','3')
Benzaldehyde_d6_r2 <- df1 %>%
  pivot_longer(-rep, names_to="Compound", values_to = "area") %>%
  group_by(Compound) %>%
  summarise(cv = cv(area))
colnames(Benzaldehyde_d6_r2)[2]<-"Benzaldehyde-d6"

```

```

df1 <- base_QC
df1.istd <- df1 %>% dplyr::select('Octane-d18')
df1 <- df1 %>% dplyr::select(!'Acetophenone d8')
df1 <- df1 %>% dplyr::select(!'Benzaldehyde-d6')
df1 <- df1 %>% dplyr::select(!'Octane-d18')
df1 <- df1 %>% dplyr::select(!'Propanol-d8')
df1 <- df1 %>% dplyr::select(!'Toluene-D8')
df1 <- df1 %>% dplyr::select(!"File name")
df1 <- cbind(df1, df1.istd)
a <- length(colnames(df1))
df1 <- df1/df1[,a]
df1[a]<- NULL
df1$rep <- c('1','2','3')
Octane_d18_r2 <- df1 %>%
  pivot_longer(-rep, names_to="Compound", values_to = "area") %>%
  group_by(Compound) %>%
  summarise(cv = cv(area))
colnames(Octane_d18_r2)[2]<-"Octane-d18"

```

```

df1 <- base_QC
df1.istd <- df1 %>% dplyr::select('Propanol-d8')
df1 <- df1 %>% dplyr::select(!'Acetophenone d8')
df1 <- df1 %>% dplyr::select(!'Benzaldehyde-d6')
df1 <- df1 %>% dplyr::select(!'Octane-d18')
df1 <- df1 %>% dplyr::select(!'Propanol-d8')
df1 <- df1 %>% dplyr::select(!'Toluene-D8')
df1 <- df1 %>% dplyr::select(!"File name")
df1 <- cbind(df1, df1.istd)
a <- length(colnames(df1))
df1 <- df1/df1[,a]

```



```

df1[a]<- NULL
df1$rep <- c('1','2','3')
Propanol_d8_r2 <- df1 %>%
  pivot_longer(-rep, names_to="Compound", values_to = "area") %>%
  group_by(Compound) %>%
  summarise(cv = cv(area))
colnames(Propanol_d8_r2)[2]<-"Propanol-d8"

df1 <- base_QC
df1 <- df1 %>% dplyr::select(!'Acetophenone d8')
df1 <- df1 %>% dplyr::select(!'Benzaldehyde-d6')
df1 <- df1 %>% dplyr::select(!'Octane-d18')
df1 <- df1 %>% dplyr::select(!'Propanol-d8')
df1 <- df1 %>% dplyr::select(!'Toluene-D8')
df1 <- df1 %>% dplyr::select(!"File name")
df1$rep <- c('1','2','3')
Raw_r2 <- df1 %>%
  pivot_longer(-rep, names_to="Compound", values_to = "area") %>%
  group_by(Compound) %>%
  summarise(cv = cv(area))
colnames(Raw_r2)[2]<-"Raw"

list_df = list(Toluene_d8_r2, Acetophenone_d8_r2, Benzaldehyde_d6_r2, Octane_d18_r2,
Propanol_d8_r2, Raw_r2)
df2 <- list_df %>% reduce(inner_join, by='Compound')
df3 <- df2 %>%
  group_by(Compound) %>%
  dplyr::summarise_each(funs = (sum)) %>%
  mutate(`Best ISTD`=apply(.,2:7, 1, function(x) names(x)[which.min(x)]))

df3_Td8 <- df3 %>% filter(`Best ISTD` == "Toluene-D8")
'Td8_compounds' <- df3_Td8$Compound
df1 <- base
row.names(df1) <- df1$`File name`
df1 <- df1 %>% dplyr::select('Toluene-D8', all_of(Td8_compounds))
df1 <- df1/df1[,1]
df1$`Toluene-D8`<- NULL
df1_Td8 <- df1

df3_Ad8 <- df3 %>% filter(`Best ISTD` == "Acetophenone d8")
'Ad8_compounds' <- df3_Ad8$Compound
df1 <- base
row.names(df1) <- df1$`File name`
df1 <- df1 %>% dplyr::select('Acetophenone d8', all_of(Ad8_compounds))
df1 <- df1/df1[,1]
df1$`Acetophenone d8`<- NULL
df1_Ad8 <- df1

df3_Od18 <- df3 %>% filter(`Best ISTD` == "Octane-d18")
'Od18_compounds' <- df3_Od18$Compound
df1 <- base

```

```

row.names(df1) <- df1$`File name`
df1 <- df1 %>% dplyr::select('Octane-d18', all_of(Od18_compounds))
df1 <- df1/df1[,1]
df1$`Octane-d18` <- NULL
df1_Od18 <- df1

df3_Pd8 <- df3 %>% filter(`Best ISTD` == "Propanol-d8")
'Pd8_compounds' <- df3_Pd8$Compound
df1 <- base
row.names(df1) <- df1$`File name`
df1 <- df1 %>% dplyr::select('Propanol-d8', all_of(Pd8_compounds))
df1 <- df1/df1[,1]
df1$`Propanol-d8` <- NULL
df1_Pd8 <- df1

df3_Bd6 <- df3 %>% filter(`Best ISTD` == "Benzaldehyde-d6")
'Bd6_compounds' <- df3_Bd6$Compound
df1 <- base
row.names(df1) <- df1$`File name`
df1 <- df1 %>% dplyr::select('Benzaldehyde-d6', all_of(Bd6_compounds))
df1 <- df1/df1[,1]
df1$`Benzaldehyde-d6` <- NULL
df1_Bd6 <- df1

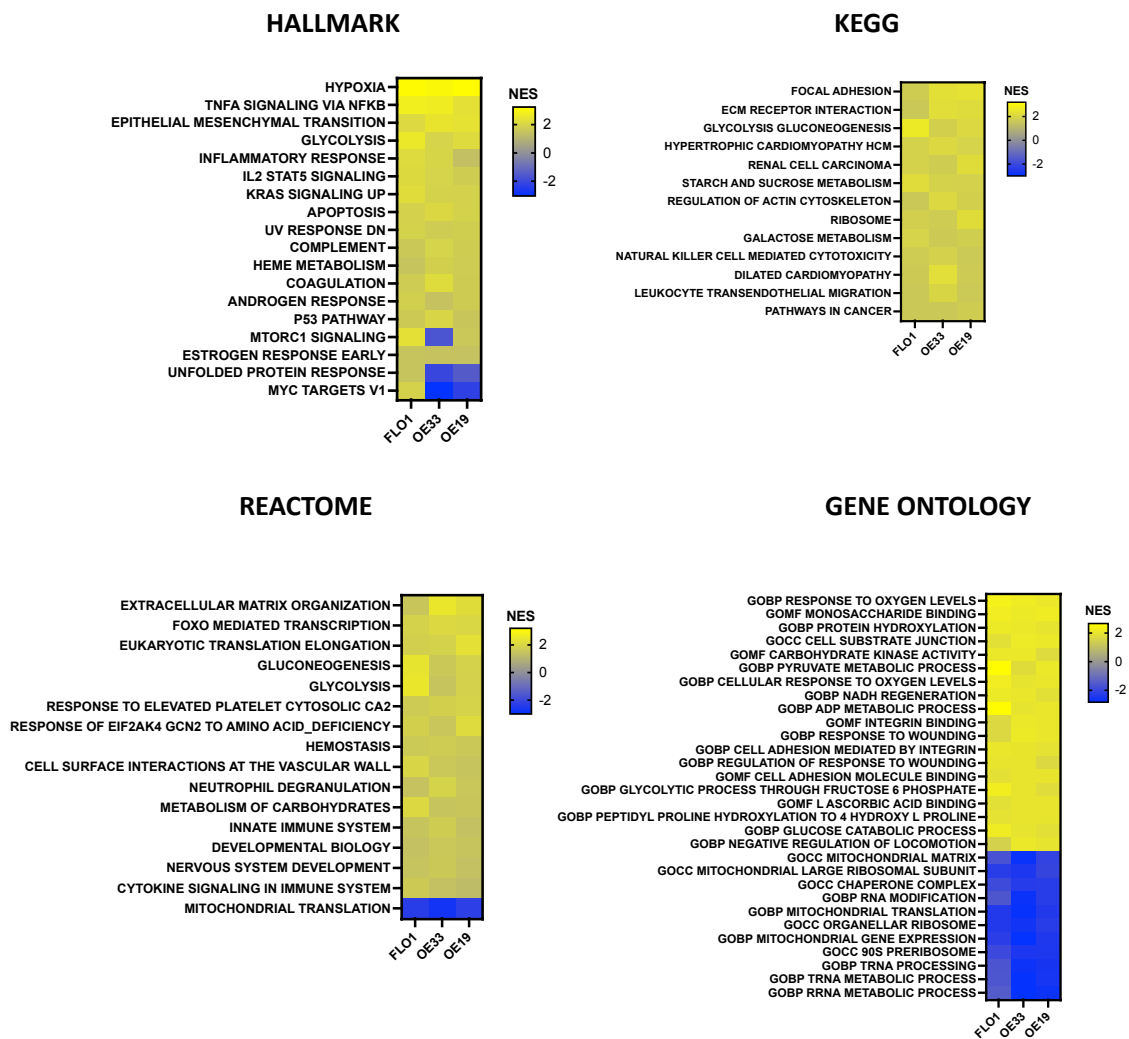
df3_Raw <- df3 %>% filter(`Best ISTD` == "Raw")
'Raw_compounds' <- df3_Raw$Compound
df1 <- base
row.names(df1) <- df1$`File name`
df1 <- df1 %>% dplyr::select(all_of(Raw_compounds))
df1_Raw <- df1

df1_all <- cbind(df1_Td8, df1_Ad8, df1_Od18, df1_Pd8, df1_Bd6, df1_Raw)
df1_all <- df1_all %>% mutate(File_name = row.names(df1_all), .before = 1)
row.names(df1_all) <- NULL
df1_all_qc <- df1_all %>% dplyr::filter(grepl('Pooled_QC', File_name))
df1_cv <- df1_all_qc %>%
  pivot_longer(-File_name, names_to="Compound", values_to = "area") %>%
  group_by(Compound) %>%
  summarise(cv = cv(area))
df1_cv <- df1_cv %>% dplyr::filter(cv < 0.30)
reproducible_compounds <- df1_cv$Compound

final_compounds <- intersect(blankfiltered_compounds, reproducible_compounds)
final <- df1_all %>% dplyr::select(File_name, all_of(reproducible_compounds))
ID <- as.data.frame(str_split_fixed(final$File_name, "_", 19))
ID <- ID %>% unite('ID', 4:5, sep = "_", remove = TRUE)
final <- final %>% mutate(Class = ID$ID, .after = 1)
final <- final %>% dplyr::filter(!grepl('LSD', Class))
colnames(final)[1] <- "Samples"
colnames(final)[2] <- "Label"
write.csv(final, 'final_batch1.csv', row.names=F)

```

Appendix 13: Gene Set Enrichment Analysis of OAC cell lines – Common Hypoxia Regulated Pathways



NES = Normalized Enrichment Score

## Appendix 14: oncoPredict drug predictions common to all datasets

| Predicted hypoxia-sensitive | Target pathway                    | Predicted hypoxia-resistant | Target pathway          |
|-----------------------------|-----------------------------------|-----------------------------|-------------------------|
| AZD1332                     | RTK signalling                    | AZD8055                     | PI3K/MTOR signalling    |
| KU-55933                    | Genome integrity                  | GSK269962A                  | Cytoskeleton            |
| RO-3306_1052                | Cell cycle                        | PF-4708671                  | PI3K/MTOR signalling    |
| MG-132                      | Protein stability and degradation | QS11                        | ARFGAP1                 |
| Dasatinib                   | RTK signalling                    | Sunitinib                   | RTK signalling          |
| CAP-232, TT-232, TLN-232    | Metabolism                        | PHA-665752                  | RTK signalling          |
| Refametinib                 | ERK MAPK signalling               | Phenformin                  | Biguanide               |
| TGX221                      | PI3K/MTOR signalling              | BMS-345541                  | IKK                     |
| Bortezomib                  | Protein stability and degradation | Sorafenib                   | RTK signalling          |
| AZD6482                     | PI3K/MTOR signalling              | CGP-082996                  | Cell cycle              |
| EphB4                       | RTK signalling                    | CAY10566                    | Stearoyl-CoA desaturase |
| AZD8186                     | PI3K/MTOR signalling              | Pemetrexed                  | DNA replication         |
|                             |                                   | Pyrimethamine               | Dihydrofolate reductase |
|                             |                                   | Seliciclib                  | Cell cycle              |
|                             |                                   | Ribociclib                  | Cell cycle              |
|                             |                                   | TAF1                        | TAF1                    |
|                             |                                   | Dihydrorotenone             | Unknown                 |
|                             |                                   | OF-1                        | BRPF1B, BRPF2           |
|                             |                                   | BMS-754807                  | RTK signalling          |



## Appendix 15: Ethical approval for breath sampling during the SOLAR trial

| Amendment Tool                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | For office use              |                                                                          |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|--------------------------------------------------------------------------|
| v1.5 25 Mar 2021                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | QC: No                      |                                                                          |
| <b>Section 1: Project information</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                             |                                                                          |
| Short project title*:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | SOLAR                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                             |                                                                          |
| IRAS project ID* (or REC reference if no IRAS project ID is available):                                                                                                                                                                                                                                                                                                                                                                                                                                           | 234206                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                             |                                                                          |
| Sponsor amendment reference number*:                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | AM2103/15                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                             |                                                                          |
| Sponsor amendment date* (enter as DD/MM/YY):                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 18 December 2020                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                             |                                                                          |
| Briefly summarise in lay language the main changes proposed in this amendment. Explain the purpose of the changes and their significance for the study. If the amendment significantly alters the research design or methodology, or could otherwise affect the scientific value of the study, supporting scientific information should be given (or enclosed separately). Indicate whether or not additional scientific critique has been obtained (note: this field will adapt to the amount of text entered)*: | <p>Protocol updated to reflect administrative updates re Version and date as well as small punctuation, spelling and syntax changes.</p> <p>Pg3: R.Kalaitzaki removed as trial statistician. Amina Tran added. Pg5-6: Co-investigators updated</p> <p>Protocol synopsis updated to include the evaluation of the role of hypoxia-associated volatile organic compounds in predicting response to Olaparib Pg14: Protocol synopsis updated to add breath sampling to summary of procedures. Sec3: Treatment scheme updated to reflect breath sampling in screening and on treatment.</p> <p>Sec4: Study glossary updated.</p> <p>Sec 5.2 Treatment background: Polyadenosine 5'-diphosphoribose [poly-(ADP-ribose)] polymerisation (PARP) and Olaparib (PARP 1 inhibitor) has been updated in accordance with the latest Olaparib IB.</p> <p>Sec6.3 Exploratory aims have been updated to reflect evaluation of hypoxia-associated volatile organic compounds.</p> <p>Sec 10.2 Screening assessments has been updated to include collection of a breath sample.</p> <p>Sec 10.3 On trial assessments has been updated to revise the window for performing PETCT during the COVID-19 pandemic and to include breath sampling.</p> <p>Sec 10.5 Sample collection has been updated to clarify requirement for archival tumour tissue collection and to include breath sampling.</p> <p>Sec 10.6 has been updated to include breath samples.</p> <p>Sec10.8 Table 1 has been updated to revise the window for performing PETCT during the Covid-19 pandemic and to include breath sampling. Sec 12.1 PV AE definition updated to reflect that in general, abnormal laboratory findings without clinical significance (based on the investigator's judgement) should not be recorded as adverse events; however, laboratory value changes requiring therapy or adjustments in prior therapy are considered adverse events.</p> |                             |                                                                          |
| Project type (select):                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | <input checked="" type="radio"/> Specific study<br><input type="radio"/> Research tissue bank<br><input type="radio"/> Research database                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                             |                                                                          |
| Has the study been reviewed by a UKECA-recognised Research Ethics Committee (REC) prior to this amendment?:                                                                                                                                                                                                                                                                                                                                                                                                       | <input checked="" type="radio"/> Yes <input type="radio"/> No                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                             |                                                                          |
| What type of UKECA-recognised Research Ethics Committee (REC) review is applicable? (select):                                                                                                                                                                                                                                                                                                                                                                                                                     | <input checked="" type="radio"/> NHS/HSC REC<br><input type="radio"/> Ministry of Defence (MoDREC)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                             |                                                                          |
| Is all or part of this amendment being resubmitted to the Research Ethics Committee (REC) as a <b>modified amendment</b> (i.e. a substantial amendment previously given an unfavourable opinion)?                                                                                                                                                                                                                                                                                                                 | <input type="radio"/> Yes <input checked="" type="radio"/> No                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                             |                                                                          |
| Where is the NHS/HSC Research Ethics Committee (REC) that reviewed the study based?:                                                                                                                                                                                                                                                                                                                                                                                                                              | <input checked="" type="radio"/> England                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <input type="radio"/> Wales | <input type="radio"/> Scotland<br><input type="radio"/> Northern Ireland |
| Was the study a clinical trial of an investigational medicinal product (CTIMP) OR does the amendment make it one?:                                                                                                                                                                                                                                                                                                                                                                                                | <input checked="" type="radio"/> Yes <input type="radio"/> No                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                             |                                                                          |
| EudraCT number*:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 2018-000454-22                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                             |                                                                          |
| Was this clinical trial of an investigational medicinal product (CTIMP) processed under the Combined Ways of Working (CWoW) pilot?:                                                                                                                                                                                                                                                                                                                                                                               | <input type="radio"/> Yes <input checked="" type="radio"/> No                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                             |                                                                          |
| Did the study receive Pharmacy Assurance?:                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | <input type="radio"/> Yes <input checked="" type="radio"/> No                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                             |                                                                          |
| Was the study a clinical investigation or other study of a medical device OR does the amendment make it one?:                                                                                                                                                                                                                                                                                                                                                                                                     | <input checked="" type="radio"/> Yes <input type="radio"/> No                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                             |                                                                          |
| Does this clinical investigation or other study of a medical device require a Notice of No Objection from MHRA Devices?:                                                                                                                                                                                                                                                                                                                                                                                          | <input type="radio"/> Yes <input checked="" type="radio"/> No                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                             |                                                                          |
| Did the study involve the administration of radioactive substances, therefore requiring ARSAC review, OR does the amendment introduce this?:                                                                                                                                                                                                                                                                                                                                                                      | <input type="radio"/> Yes <input checked="" type="radio"/> No                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                             |                                                                          |