

# **ALDEHYDE METABOLIC REPROGRAMMING IN OESOPHAGEAL ADENOCARCINOMA**

**Stefan Samad Antonowicz MBChB MRCS (Eng)**

Department of Surgery and Cancer, Imperial College London  
St Mary's Hospital, London, UK

**2017**

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

**Supervisor**

**Professor George B Hanna PhD FRCS**

Department of Surgery and Cancer, Imperial College London,  
St Mary's Hospital, London, UK

## **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.

20<sup>th</sup> April 2017

Stefan Antonowicz

## **COPYRIGHT DECLARATION**

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

## ABSTRACT

Oesophageal adenocarcinoma (OAC) has unmet clinical needs as the UK five-year survival is 14%. Efforts to enhance early diagnosis uncovered enriched volatile aldehydes in OAC patients' breath, although their origins and fate are unknown.

Following comprehensive bioinformatics analyses, it was hypothesised that detoxification loss enriches aldehydes in the transforming lower oesophagus. Pursuing this biology could help refine OAC breath testing, deepen understanding of oncogenesis and uncover therapeutic susceptibilities. This PhD aimed to describe OAC aldehyde metabolism, its genetic framework, and its oncogenic effects.

A bespoke ultra-performance liquid chromatography tandem mass spectrometry (UPLC-MS/MS) method was validated to unambiguously quantify 43 aldehydes and ketones in tissue samples. Multiple aldehyde species were enriched in OAC tissues, suggesting active carbonyl stress, field effects, and a requirement for competent defences.

Genetically, aldehyde oxidoreductase expression loss defined OAC tissues, compared to normally resident tissue. Five aldehyde dehydrogenase isoenzymes were consistently and significantly depleted ( $P < 10^{-8}$  to  $^{-20}$ ); these findings were validated at the RNA ( $n = 67$ ) and protein ( $n = 412$ ) levels in clinical samples. In particular, loss of *ALDH3A2* was associated with disease progression and independently predicted poorer survival (OR = 1.64, 95% C.I. 1.13 – 2.39,  $P = 0.01$ ).

To explore the effects of aldehyde metabolic rewiring, a second UPLC-MS/MS method was developed, which suggested that aldehyde-DNA adducts are also enriched in OAC tissues. Mechanistic studies *in vitro* revealed that ALDH inhibition is sufficient to enrich metabolic aldehyde in OAC cells. Finally, stable perturbation of *ALDH3A2* in OAC cells highlighted a potential tumour suppressor role for this gene, as CRISPR-Cas9 mediated knockout enhanced cell growth through cell cycle shunting and affected redox control.

These data highlight genetically deregulated aldehyde metabolism as a feature of OAC, which may contribute to carcinogenesis. Clinical implications and future research directions are discussed.

# CONTENTS PAGE

<b>THESIS DECLARATION</b>	<b>2</b>
<b>COPYRIGHT DECLARATION</b>	<b>2</b>
<b>ABSTRACT</b>	<b>3</b>
<b>CONTENTS PAGE</b>	<b>4</b>
<b>LIST OF FIGURES</b>	<b>8</b>
<b>LIST OF TABLES</b>	<b>10</b>
<b>LIST OF ABBREVIATIONS</b>	<b>11</b>
<b>ACKNOWLEDGEMENTS</b>	<b>12</b>
<b>CHAPTER 1 - INTRODUCTION</b>	<b>13</b>
<b>1.1 Oesophageal adenocarcinoma</b>	<b>14</b>
1.1.1 Clinical background & project motivation	14
1.1.2 Pathophysiology & genetic landscape	15
1.1.3 Metabolic landscape	17
<b>1.2 Normal aldehyde metabolism</b>	<b>21</b>
1.2.1 Reactive aldehydes	21
1.2.2 General sources of aldehydes	22
1.2.3 Sources of specific aldehydes	23
1.2.4 Genetic mediators of aldehyde metabolism	24
<b>1.3 Aldehyde metabolism in cancer</b>	<b>27</b>
1.3.1 Reactive aldehydes as biomarkers of cancer	27
1.3.2 Aldehyde genes and cancer	29
1.3.3 Consequences of deregulated aldehyde metabolism	30
1.3.4 Aldehyde metabolism the oesophagus	32
<b>1.4 Gaps in the literature</b>	<b>34</b>
<b>1.5 Thesis hypothesis and objectives</b>	<b>35</b>
<b>CHAPTER 2 – ALDEHYDES IN OESOPHAGEAL ADENOCARCINOMA</b>	<b>36</b>
<b>Summary</b>	<b>37</b>
<b>2.1 Methodological rationale</b>	<b>38</b>
2.1.1 Choice of analytical technique	38
2.1.2 Method goals & selection of targets	43
<b>2.2 Hypothesis and aims</b>	<b>44</b>

<b>2.3 Aldehyde method development</b>	<b>45</b>
2.3.1 Materials	45
2.3.2 Human and murine biospecimens	45
2.3.3 Instrumentation	47
2.3.4 Initial settings	47
2.3.5 Aldehyde precautions	48
2.3.6 MS/MS development	49
2.3.7 Chromatography development	52
2.3.8 Injector settings	58
2.3.9 Sample preparation development.	58
2.3.10 Optimised sample preparation method	67
2.3.11 Calculations and statistics	69
2.3.12 Aldehyde method validation	70
2.3.13 Method application rationale	73
<b>2.4 Results: Application of UPLC-MS/MS method to samples.</b>	<b>74</b>
2.4.1: Aldehydes in OAC and control tissue samples	74
2.4.2: Aldehyde concentrations in murine OAC xenografts	79
2.4.3: Aldehyde concentrations in OAC urine.	81
<b>2.5 Discussion</b>	<b>83</b>
2.5.1 A UPLC-ESI-MS/MS method for quantifying carbonyls from biospecimens.	83
2.5.2 Aldehydes in normal and malignant oesophageal tissues	85
2.5.3 Aldehydes in OAC xenografts	86
2.5.4 Aldehydes in OAC and control urine	87
<b>CHAPTER 3 – GENETIC DRIVERS OF ALDEHYDE METABOLIC REPROGRAMMING IN OESOPHAGEAL ADENOCARCINOMA</b>	<b>89</b>
<b>Summary</b>	<b>90</b>
<b>3.1 Methodological rationale</b>	<b>92</b>
<b>3.2 Hypothesis and aims</b>	<b>93</b>
<b>3.3 Methods</b>	<b>94</b>
3.3.1 Bioinformatics discovery of candidate drivers of aldehyde reprogramming	94
3.3.2 Patient samples	95
3.3.3 Quantitative reverse transcriptase polymerase chain reaction (qPCR)	96
3.3.4 Immunohistochemistry	98
3.3.5 Cell culture	101
3.3.6 Copy number analysis	102
3.3.7 Immunoblotting	102
3.3.8 Immunofluorescence	103
3.3.9 Statistics	103
<b>3.4 Results</b>	<b>105</b>
3.4.1 Candidate discovery using bioinformatics	105

3.4.2 Patients	108
3.4.3 Candidate <i>ALDH</i> expression in at mRNA level – qPCR study	109
3.4.4 Candidate <i>ALDH</i> expression at protein level – immunohistochemistry studies	111
3.4.5 <i>ALDH</i> expression in OAC <i>in vitro</i> and <i>in vivo</i> models	118
3.4.6 Regulation of <i>ALDH3A1</i> and <i>-3A2</i> and association with 17p LOH	120
<b>3.5 Discussion</b>	<b>122</b>
<b>CHAPTER 4 – EFFECTS OF ALDEHYDE METABOLIC REPROGRAMMING IN OESOPHAGEAL ADENOCARCINOMA</b>	<b>127</b>
<b>Summary</b>	<b>128</b>
<b>Hypothesis &amp; aims</b>	<b>130</b>
<b>4.1 Aldehyde-nucleoside adducts</b>	<b>131</b>
4.1.1 Methodological rationale	131
4.1.2 Adducts method development	135
4.1.3 Final adducts method	142
4.1.4 Method validation	144
4.1.5 Results	147
4.1.6 Discussion	149
<b>4.2 <i>ALDH</i> perturbation of normal and malignant oesophageal cell models</b>	<b>153</b>
4.2.1 Methodological rationale	153
4.2.2 Methods	154
4.2.3 Results	155
4.2.4 Discussion	159
<b>4.3 Metabolic and non-metabolic functions of <i>ALDH3A2</i></b>	<b>161</b>
4.3.1 Methodological rationale	161
4.3.2 Methods 1: Stable overexpression of <i>ALDH3</i> by lentiviral transduction	162
4.3.3 Methods 2: <i>ALDH3A2</i> knockout by CRISPR-Cas9 editing	164
4.3.4 Methods 3: Endpoint assays	168
4.3.5 Results	169
4.3.6 Discussion	174
<b>CHAPTER 5 – THESIS CONCLUSIONS &amp; FUTURE WORK</b>	<b>177</b>
<b>5.1 Conceptual gains from this thesis</b>	<b>178</b>
5.1.1 New paradigms in OAC cancer biology, and new therapeutic opportunities	178
5.1.2 Demonstrate biologically distinct, clinically relevant subgroups	180
5.1.3 Breath test refinement	181
<b>5.2 Technical gains from this thesis</b>	<b>183</b>

<b>5.3 Future research</b>	<b>184</b>
5.3.1 Short term goals	184
5.3.2 Longer term projects	185
<b>BIBLIOGRAPHY</b>	<b>187</b>
<b>APPENDIX 1: PRISMA FLOW CHART OF SYSTEMATIC REVIEW</b>	<b>214</b>
<b>APPENDIX 2: ETHICAL APPROVALS FOR USE OF HUMAN BIOMATERIAL</b>	<b>215</b>
<b>APPENDIX 3: OLIGONUCLEOTIDE SEQUENCES</b>	<b>219</b>
<b>APPENDIX 4: GENE EXPRESSION OF ALDH GENES NOT QUANTIFIED BY QPCR</b>	<b>220</b>
<b>APPENDIX 5: RELATED GENE EXPRESSION (WANG ET AL)</b>	<b>221</b>
<b>APPENDIX 6: PLASMID MAPS OF PACKAGING VECTORS</b>	<b>222</b>

## List of figures

Figure 1: Human cancers ranked by mutation burden. ....	16
Figure 2: Methodological quality of included OAC metabolomics studies .....	18
Figure 3: Aldehyde metabolism .....	21
Figure 4: Schematics of general principles of GC, LC, and MS/MS. ....	40
Figure 5: Dinitrophenylhydrazine reactions with selected aldehyde subtypes.....	42
Figure 6: Major fragments of mono-DNP-hydrazones.....	51
Figure 7: Total ion chromatograms of separated DNP-hydrazone isomers/isobars. ....	54
Figure 8: Final chromatographic method .....	55
Figure 9: Example calibration curves.....	56
Figure 10: Derivatisation development. ....	60
Figure 11: Sample preparation development.....	62
Figure 12: Aldehyde matrix effects .....	70
Figure 13: Aldehyde concentrations in OAC tissues and controls.....	76
Figure 14: Aldehyde concentrations in subcutaneous tumours.....	80
Figure 15: Urinary aldehyde concentrations .....	82
Figure 16: Candidate-based discovery of drivers of aldehyde reprogramming.....	91
Figure 17: Representative immunostaining of ALDHs .....	100
Figure 18: ALDH candidate discovery and IPA analysis. ....	107
Figure 19: qPCR biopsy study.....	110
Figure 20 Candidate ALDH expression in the discovery immunohistochemistry....	112
Figure 21: <i>ALDH3A2</i> expression loss occurs with progression.....	113
Figure 22 Kaplan-Meier survival curves fitted to <i>ALDH</i> expression.....	114
Figure 23: <i>ALDH3A1</i> and <i>3A2</i> expression in the validation OCCAMS cohort .....	115
Figure 24: <i>ALDH</i> expression in validated oesophageal models .....	119
Figure 25: Copy number correlates with <i>ALDH3A1</i> and <i>3A2</i> expression. ....	121
Figure 26: Normal nucleosides.....	133
Figure 27: Selected modified nucleosides for study.....	134
Figure 28: Nucleoside chromatography using various stationary phases.....	139
Figure 29: Nucleoside adduct quantification in clinical samples.....	148
Figure 30: Basal and ALDH-inhibited aldehyde phenotypes of oesophageal cell models. ....	156
Figure 31: Effects of individual and combination siRNA knockdown on aldehyde metabolic phenotypes. ....	158
Figure 32: Production of <i>ALDH3A2</i> -IRES-GFP lentivirus.....	163
Figure 33: Development of <i>ALDH3A2</i> knockout FLO1 cells using Cas9 editing.....	166



Figure 34: Functional phenotyping of <i>ALDH3A2</i> <i>in vitro</i> .....	168
Figure 35: Influence of <i>ALDH3A2</i> on cancer-specific phenotypes in OAC.....	171
Figure 36: Free and bound aldehyde metabolic phenotypes in <i>ALDH3A2</i> -constrained FLO1 cells.....	173

## List of tables

Table 1: Findings of systematic review into oesophageal cancer metabolomics studies.....	19
Table 2: Examples of selected aldehyde subtypes .....	22
Table 3: Aldehyde dehydrogenase isoenzymes and selected properties.....	26
Table 4: Aldehyde materials .....	46
Table 5: Initial instrument settings .....	48
Table 6: Monitored ion transitions and fragmentation energies.....	50
Table 7: Differentiation of ketones and alkanals on fragmentation pattern.....	51
Table 8: Comparison of calibration properties of ESI and APCI ion sources.....	57
Table 9: Method validation report for aldehyde UPLC-MS/MS method .....	72
Table 10: Demographics of UPLC-MS/MS clinical phenotyping .....	75
Table 11: Aldehyde concentrations (pg/mg) in selected subgroups of AdT samples	78
Table 12: Characteristics of <i>bona fide</i> OAC cell lines .....	101
Table 13: Summary of GSEA findings in three datasets .....	105
Table 14: The ten highest ranked GO v5 geneset discoveries from GSEA of Wang et al.....	106
Table 15: Clinico-demographic features of the <i>ALDH</i> expression cohorts .....	108
Table 16: Associations of <i>ALDH3A1/2</i> staining to metadata in 2 <sup>nd</sup> IHC cohort (OCCAMS).....	116
Table 17: Independent predictors of death in OAC identified using a Cox model...	117
Table 18: Authentic external and internal standard sources and abbreviations.....	135
Table 19: Initial instrument settings .....	136
Table 20: Fragmentation patterns and energies for target compounds.....	137
Table 21: Linear range of quantification of external standards.....	140
Table 22: Method validation data for nucleoside-adducts .....	145
Table 23: Method validation data for normal nucleosides .....	146

## List of abbreviations

( <sup>1</sup> H)-NMR	(proton)-nuclear magnetic resonance
(HP)LC	(high performance) liquid chromatography
(UP)LC	(ultra performance) liquid chromatography
AdT	Adenocarcinoma tissue
ALDH	Aldehyde dehydrogenase
APCI	Atmospheric pressure chemical ionisation
BE	Barretts tissue
C1, C2...	Alkanals (formaldehyde, acetaldehyde...)
cdA	8-chlorodeoxyadenosine
cdG	8-chlorodeoxyguanosine
CrodG	Crotondeoxyguanosine
DNPH	dinitrophenylhydrazine
edA	Ethenodeoxyadenosine
edC	Ethenodeoxycytidine
ESI	Electrospray ionisation
EtdG	Ethyldeoxyguanosine
GC	gas chromatography
GSEA	geneset enrichment analysis
HNE	4-hydroxy-2-nonenal
ICGC	International Cancer Genome Consortium
IHC	immunohistochemistry
IPA	Ingenuity Pathway Analysis
ISTD	stable isotope labelled internal standard
LLE	liquid-liquid extraction
m/z	mass-to-charge ratio
MDA	malondialdehyde
MS	mass spectrometry
MS/MS	triple quadrupole mass spectrometry
MW	molecular weight
OAC	oesophageal adenocarcinoma
OdG	Oxodeoxyguanosine
ONE	4-oxo-2-nonenal
OSCC	oesophageal squamous cell carcinoma
PCA	principle component analysis
PTR	proton transfer reaction
QC	quality control
qRT-PCR	(quantitative reverse transcriptase) polymeras chain reaction
RSD	relative standard deviation
SIFT	selected ion flow tube
SPE	solid phase extraction
SqN	Proximal squamous mucosa for endoscopically normal oesophagus
SqT	Proximal squamous mucosa for OAC-bearing oesophagus
TCGA	The Cancer Genome Atlas
VOC	Volatile organic compound

## **Acknowledgements**

Primarily, I would like to thank my supervisor and mentor Professor George Hanna for this three-year journey. His unfailing principles have become the bedrock of my academic and clinical training, and I hope I can similarly inspire my trainees with his philosophies in the years to come.

My sincere thanks go to Prof Nagy Habib and all the World of Surgery staff. I would like to acknowledge Prof Rob Goldin, Dr Hiromi Kudo, Dr Francesca Rosini, and Dr Mahrokh Nohadani for their help with histopathology matters; Mr Haonan Lu, Dr Chiara Recchi and Prof Hani Gabra for molecular matters; Dr Shona McRae and Prof Rebecca Fitzgerald for facilitating OCCAMS access and providing project advice. My thanks also go to Tom Wiggins, Sheraz Markar, Nima Abbassi-Ghadi, and Sacheen Kumar.

Particular thanks also go to Dr Zsolt Bodai, for guiding me with everything MS. By extension, my thanks go to Prof Zoltan Takats for facilitating this collaboration.

Most importantly my deep gratitude to my family, in particular my wife Aurelia, who have supported me without really ever seeing me; I love you and I owe you everything.

# **CHAPTER 1 - INTRODUCTION**

## 1.1 Oesophageal adenocarcinoma

### 1.1.1 Clinical background & project motivation

In 2012, oesophageal cancer was the 6th most common cause of UK cancer mortality, and internationally responsible for 400,200 deaths(1). There are two main histological subtypes – oesophageal adenocarcinoma (OAC) and squamous cell cancer (OSCC). Worldwide, there are 52,000 new OAC cases each year, and established risk factors include gastro-oesophageal reflux and obesity (1–3). OSCC is more common worldwide – 398,000 new cases in 2012 – and is more prevalent in the Middle East and East Asia. In the last 40 years, the incidence of OAC has increased 600% in the west, making it among the fastest growing of all cancer incidences (2–4). The UK has the highest national burden of oesophageal cancer with an annual incidence of 7.2 per 100,000 men, some 10x the global average(3).

The current standard for treatment for oesophageal cancer varies with histological subtype and disease stage at presentation (5). OSCC tumours often respond dramatically to chemoradiotherapy, and surgery may be reserved only for those which residual, refractory or relapsing disease after treatment. OAC tumours are less sensitive to non-surgical treatment, and resection remains the primary curative treatment. For early stage tumours, surgery alone offers excellent survival prospects at 1 year (95% for T1)(5). Chemotherapy or chemoradiotherapy may be offered in a “neo-adjuvant” setting, with the intention of eliminating micrometastasis and potentially down-staging disease prior to resection; both have shown survival advantage (6,7). The additional advantage conferred by radiotherapy is being investigated (8). Further post-operative oncological treatment may be recommended depending on histopathological features of risk.

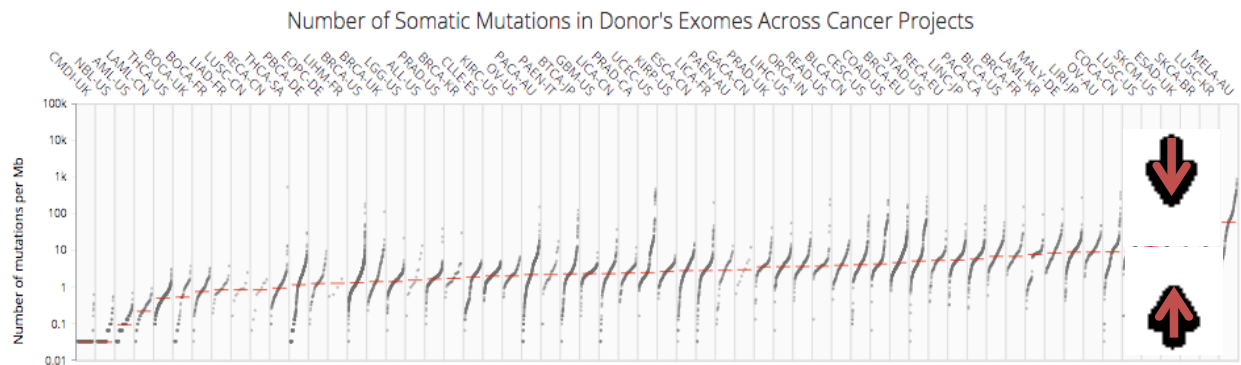
The UK OAC short-term survival has been transformed over the last 25 years through service centralisation, neo-adjuvant treatments, and enhanced recovery after surgery (9). Thus, short-term mortality after oesophagectomy since 1980s has reduced 30-fold to <1%, and 5-year survival in surgically treated patients has improved to 50-55% (5,10,11). However, the overall 5-year survival remains low at 10-14% (12,13). This is explained by the pattern of presentation – 70% of patients will present with incurably advanced disease (5,14). Thus, there is an urgent unmet

need for new strategies to detect preclinical disease, so these safe and effective treatments can be provided.

A novel, non-invasive means of OAC diagnosis has been developed, using trace metabolites - volatile organic compounds (VOCs) - in exhaled breath (15,16). The diagnostic model supporting this test is broadly comprised of certain saturated aldehydes (in particular, acetaldehyde, butanal, pentanal, nonanal and decanal) among other molecular classes (15–18). However, the mechanism of this enrichment is not understood, and it was hypothesised that volatile aldehydes (VAs) signpost relevant biological events in OAC carcinogenesis. Thus, the strategic rationale for investigating these biomarkers mechanistic basis is to: (i) uncover new paradigms in OAC cancer biology, and potentially novel therapeutic opportunities, (ii) demonstrate biologically distinct, clinically relevant patient subgroups, supporting precision medicine through VOC testing (iii) inform breath test refinement thorough mechanistic understanding, and so both improve analytics and strengthen the argument for clinical implementation.

### **1.1.2 Pathophysiology & genetic landscape**

The oesophagus is a muscular tube that conveys liquids and food from the pharynx to the stomach. A requirement is to convey luminal contents *safely*, protecting itself and the wider body from temperature extremes and toxic elements of the undigested food bolus. Relevant protective adaptations include a pre-epithelial alkaline mucous barrier, a stratified epithelium lining the entire tube, a robust expression of a large panel of detoxification genes, including glutathione transferases (GSTs etc)(19), aldo-keto reductases (AKR)(20), etc, and post-epithelial buffering system of high intracellular bicarbonate ions(21). Further protective adaptations within the stratified squamous epithelium includes partial keratinisation, very tight cell-cell adherence, and lipid metabolic reprogramming to support a particularly inert plasma membrane composed of ceramides (21,22). However, compared to the gastroduodenal epithelium, the pre-epithelial barrier is not well developed, and DNA toxicity may occur even in the healthy oesophagus (23,24).



**Figure 1: Human cancers ranked by mutation burden.** Arrows indicate the oesophageal adenocarcinoma dataset (ICGC whole genome sequencing data) (25)

Oesophageal adenocarcinoma (OAC) typically arises in the lower third of the oesophagus and gastro-oesophageal junction (GOJ). It is associated with pro-inflammatory risk factors such as gastro-oesophageal reflux, smoking and obesity(26–28), and has a well-defined carcinogenic sequence beginning with the replacement of the normal stratified squamous epithelium with columnar cells (Barrett’s metaplasia). There is considerable debate regarding the origin of Barrett’s cells; hypotheses include a direct metaplasia from squamous cells, out-growth of submucosal glands, or even expansion of remnant embryonic cells residing in the lower oesophagus and GOJ (29,30). Recent sequencing data suggest that the Barrett’s “neopithelium” copes poorly with genotoxic stress, and acquires transforming mutations (31–34). Consequently, 0.12-0.16% of Barrett’s patients will progress to invasive OAC each year (35,36). Comparison of pan-cancer whole genome sequencing data reveals that OAC is among the most mutated of all human malignancies (see Figure 1)(37). OAC and BM share a unique mutational signature characterised by  $A/T > C/G$  transversions within certain trinucleotide contexts (31,33), which cannot be fully explained by DNA repair defects or age(37). Apart from *TP53*, which is mutated in >80% of OAC cases, there are no recurrently mutated genes with a prevalence >15% (31–34). Structural variant analyses suggest frequent yet untargeted chromosomal rearrangements, amplifications, deletions and chromothripsis, consistent with profound genome instability(31,38). Taken together, this pattern of locus-indiscriminate, base-specific genotoxicity suggests that OAC is a prototypical carcinogen-driven malignancy. Recent OAC whole genome sequencing data have proposed mutational subsets based on age, DNA-damage repair dysfunction, and “acid mutagenesis” (34), although the precise mutagenic drivers are not yet determined.



### 1.1.3 Metabolic landscape

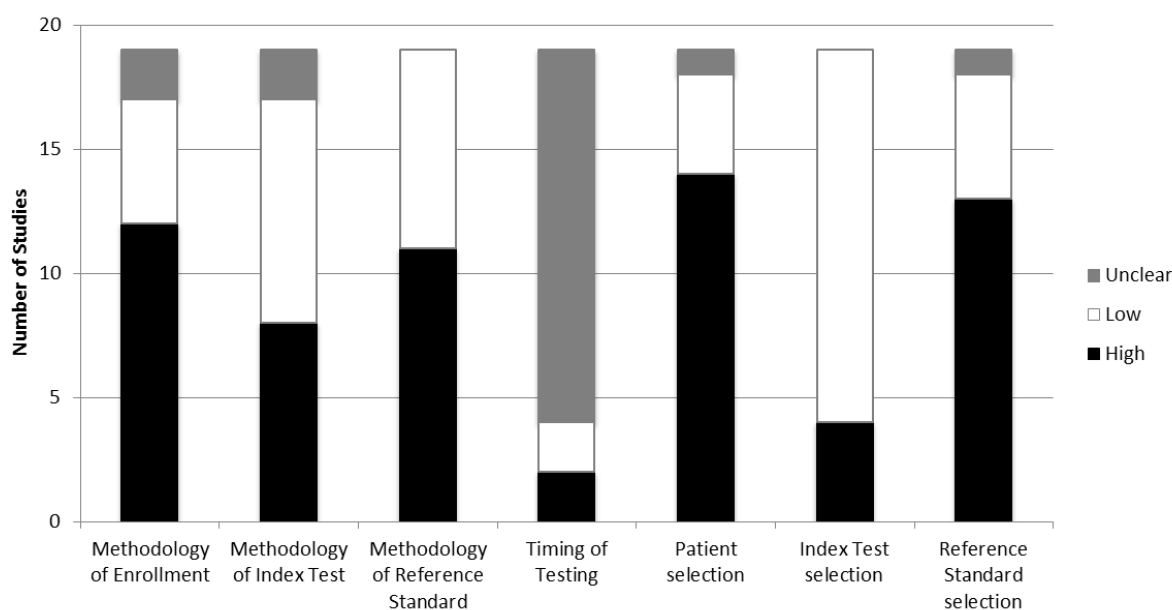
To help develop an investigative strategy for volatile biomarker mechanisms, a survey of empirical data regarding oesophageal cancer metabolism was undertaken.

A systematic review of quantitative data was performed of Embase, Medline and Web of Science database, using the search terms (“metabo\*omics” OR “metabolic profiling”) AND (“oesophageal cancer”) and the limits (“2000-2016” and “human”). Only metabolomic studies comparing cancers subjects with relevant controls were included. Titles, abstracts, and full-texts were processed using a PRISMA algorithm (see Appendix 1). Studies were selected if they used nuclear magnetic resonance spectroscopy (NMR) or mass spectrometry (MS) to compare metabolites from OAC biospecimens. For the purpose of quality assessments, endoscopic diagnosis was selected as a reference standard for both controls and cases. The Candidate and a collaborator (T.W., see acknowledgements) performed independent searches and 19 studies were agreed for data extraction. The primary extraction item was a qualitative description of significantly altered metabolites. Secondary extraction items included methods, and methodological and reporting quality.

Methodological quality was summarised using the QUADAS-2 tool(39) (see Figure 2). In biofluids studies, a recurrent limitation was that controls often were not endoscoped, and therefore may have occult endoluminal pathology. Metabolomic results of most studies were interpreted with supervised analyses, i.e. the analysis was not blinded. Reporting quality was assessed using the STARD checklist and found to be moderate-to-poor (median score 11/25) (40). Given the obvious range in study quality, a post-hoc Bonferroni correction was applied to all studies not accounting for multiplicity, using a reporting cut-off of  $\alpha = 0.05/n$  where  $n$  is the number of reported features. This methodological approach has been published (41).

The results of this review are provided in Table 1. Studies were generally exploratory in nature and featured case-controlled design. One group has repeated work in experimentally separate cohorts, with some validation of their previous findings(15,16). Studies involving serum/plasma were the most common. Changes in amino acids and sugars were most frequently reported metabolite, which is unsurprising given the nature of the methods (i.e. NMR and untargeted MS methods).

**Figure 2: Methodological quality of included OAC metabolomics studies (assessed by QUADAS-2 tool)**



In plasma studies, consistently enriched metabolite types included ketone bodies (e.g. acetone, b-hydroxybutyrate), short chain carboxylic acids (formate, acetate), long chain fatty acids (palmitic and myristic acid), and fuels (glutamate, glucose, creatine, lactate). Essential amino acids were generally depleted in OAC/OSCC plasma, although lysine was reported enriched in more than one study. In volatile metabolomic series in breath and urine, aldehydes, short-chain carboxylic acids and phenolics were enriched in gastro-oesophageal malignancies. Non-volatile metabolomics series in urine reported enrichment of short-chain carboxylic acids, a branched chain amino acid (leucine) and related catabolites (3-hydroxyvalerate). OAC tissue studies have only been carried out using NMR (two studies) and revealed enrichment of branched chain amino acids (leucine, isoleucine, valine, among other essential amino acids), phospholipid metabolites, unsaturated lipids, short- and long- chain fatty acids, and ketone bodies.

All of these compounds are specific metabolic precursors or products of reactive aldehydes, although no study had a specific methodology designed to quantify aldehydes themselves. Other major phenotypes suggested by these findings include amino acid consumption, glycolytic switch, de novo lipogenesis, and phospholipid perturbations.

**Table 1: Findings of systematic review into oesophageal cancer metabolomics studies**

Ref.	Study	Year	Type	n (Ca)	Platform	Findings
<i>Studies investigating plasma</i>						
(42)	Djukovic	2010	OAC	26 (14)	UPLC-TQMS	UP: uridine; DOWN: 1-methyladenosine, 2,2,di-methylguanosine, N2-methylguanosine, cytidine
(43)	Zhang	2011	OAC	118 (68)	<sup>1</sup> H-NMR	See study 3
(44)	Zhang	2012	OAC	113 (67)	LC-MS & <sup>1</sup> H-NMR	UP: lactate, carnitine, b-hydroxybutyrate, citrate, lysine, creatine, glucose; DOWN: Valine, leucine/isoleucine, methionine, tyrosine, tryptophan, 5-hydroxytryptophan, myristic acid, linolenic acid, linoleic acid
(45)	Ikeda	2011	OAC	27 (15)	GC-MS	UP: lactate
(46)	Zhang	2013	"EC"	50 (25)	UPLC-diode & <sup>1</sup> H-NMR	UP: aspartate, cysteine, leucine, phenylalanine, lysine, b-hydroxybutyrate, creatine, creatinine, lactate, glutamate, glutamine, histine; DOWN; methionine, tryptophan, LDL/VLDL, unsaturated lipids, acetate, a-glucose, tyrosine
(47)	Sanchez-Espiridion	2015	OAC	652 (321)	LC-MS/MS	UP: beta-hydroxybutyrate, d-mannose DOWN: L-proline
(48)	Liu	2013	OSCC	152 (72)	UPLC-ESI-TOFMS	UP: phosphatidylinositol, lithocholytaurine, phosphatidic acid, l-urobilinogen, phosphatidylcholine, phosphatidyl ethanolamine, sphingosine 1-phosphate, phosphatidylserine (16:0/14:0), lithocholate 3-o-glucuronide, DOWN: desmosine/isodesmosine,
(49)	Xu	2013	OSCC	228 (124)	RRLC/ESI-MS	UP: Lactate; DOWN: LysoPC(14:0), LysoPC(20:3) UP: Lactic acid, b-hydroxybutyric acid, hypotaurine, aspartic acid, b-alanine, 3-hydroxybutyric acid, myristic acid, palmitic acid, oleic acid, linoleic acid, palmitelaidic acid, 1-monooleoylglycerol, ribose, maltose, lactose, creatinine.
(50)	Jin	2014	OSCC	110 (80)	GC-MS	DOWN: Glucose, alanine, glutamine, citric acid, fumaric acid, 1,5-anhydroglucitol, valine, 2-ketoisovaleric acid, 2-ketoisocaproic acid, 3-methyl-2oxovaleric acid, tryptophan, indolelactic acid, iminodiacetic acid, phosphoethanolamine, g-aminobutyric acid, glycolic acid, cysteine, methylcysteine, a-tocopherol, g-tocopherol, threonine, uric acid, erythritol, inositol, myo-inositol-1-phosphate, cholestrol, hydroxylamine

(51)	Ma	2014	OSCC	111 (51)	HPLC-diode	UP: aspartate; DOWN: glutamate, glycine, histidine, threonine, taurine, alanine, methionine, isoleucine, leucine, phenylalanine
(52)	Mir	2015	OSCC	80 (40)	LC-ESI-TOFMS	<i>P</i> values not reported; 652 deregulated features of which 101 were phosphocholines
<i>Studies investigating breath</i>						
(16)	Kumar	2012	GO	53 (18)	SIFT-MS	UP: hexanoic acid, phenol, methyl phenol
(15)	Kumar	2015	GO	210 (89)	SIFT-MS	UP: hexanoic acid, phenol, methyl phenol, ethyl phenol, butanal, pentanal, hexanal, heptanal, octanal, nonanal, decanal
<i>Studies investigating urine</i>						
(18)	Huang	2012	GO	51 (17)	HS-SIFT-MS	UP: acetaldehyde, acetone, acetic acid, hexanoic acid, hydrogen sulphide, methanol
(53)	Davis	2012	OAC	91 (66)	<sup>1</sup> H-NMR	UP: urea, acetate, pantothenate, 3-hydroxyvalerate, acetone, formate, leucine, succinate
(54)	Xu	2016	OSCC	124 (62)	LC-ESI-TOFMS	UP: pyroglutamic acid, uric acid, deoxycytidine, phenylacetylglutamine, cGMP DOWN: carnitines C9:0
<i>Studies investigating tissue</i>						
(55)	Wu	2009	EC	40 (20)	GC-MS	UP: valine, isoleucine, tyrosine, asparagine, alanine; arabinofuranoside, tetradecanoic acid, hexadecanoic acid, naphthalene, 1-butanamine, aminoquinolone, myo-inositol, phosphoric acid, pyrimidine nucleoside DOWN: L-altrose, d-galactofuranoside, arabinose, bisethane
(56)	Yakoub	2010	OAC	122 (35)	<sup>1</sup> H-MAS-NMR	UP: phospho-choline, myo-inositol, glutamine, inosine, adenosine, uridine
(57)	Wang	2013	OAC	105 (89)	<sup>1</sup> H-MAS-NMR	UP: glutamate, valine, leucine/isoleucine, L-tyrosine, methionine, phenylalanine, GABA, phenylacetylglutamine, taurine; unsaturated lipids, short-chain fatty acids, phosphocholine; glycoprotein, acetone, malonate, acetoacetate, acetate, trimethylamine, uracil, ATP, NAC, DOWN: Creatine, glycine, glutamine, creatinine; myo-inositol, choline, glucose, ethanol, AMP, nicotinamide

## 1.2 Normal aldehyde metabolism

### 1.2.1 Simple aldehydes

Simple aldehydes are low-mass organic compounds that have a terminal –CHO group, with a double bond between the terminal carbon and oxygen (see Figure 3). The C=O relationship allows the oxygen to glean a disproportionate quantity of the shared electron cloud, causing a partial charge gradient across the bond. The partial positive charge on the “carbonyl” carbon renders it susceptible to attack from nucleophilic spare electron pairs on e.g. thiol or amine groups. These nucleophile subgroups are biologically ubiquitous, and therefore aldehydes are widely bioactive. Accordingly, all living systems have broad-spectrum detoxification systems, which work to convert aldehydes to non-toxic products by manipulating the carbonyl carbon. An overview of human aldehyde metabolism is given in Figure 3, and the sources and genetic influences are described below. Examples of selected aldehyde subtypes are given in Table 2.

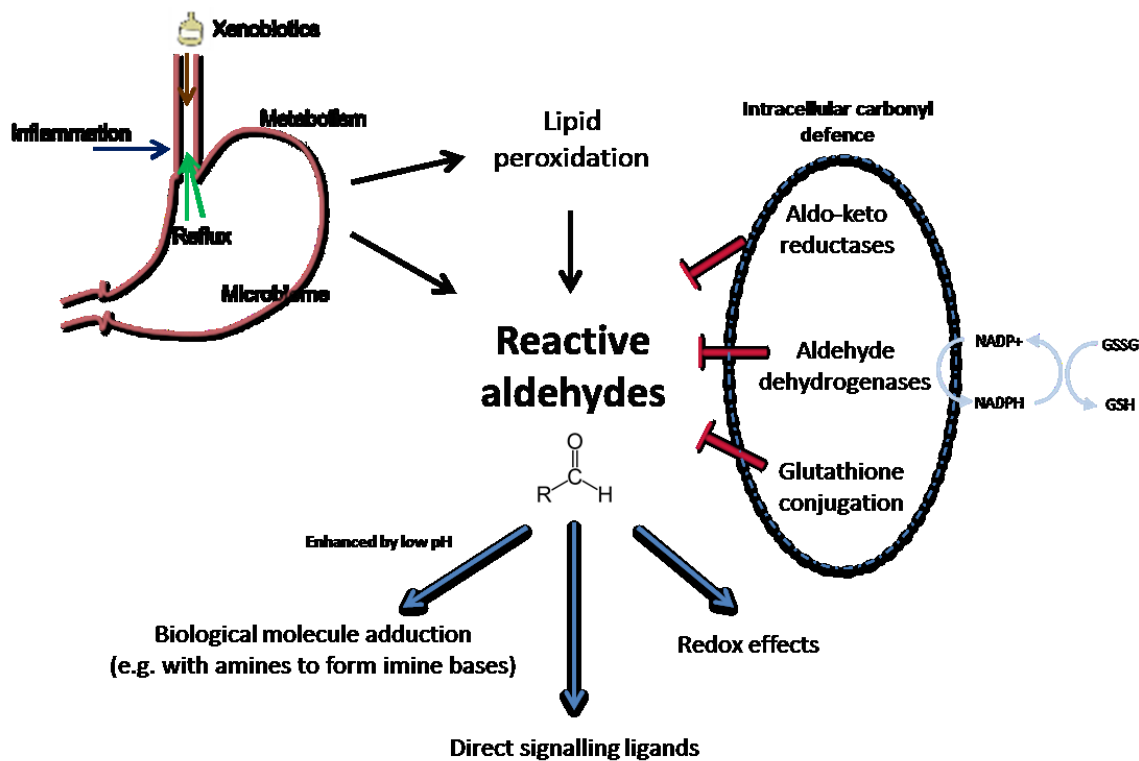

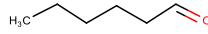
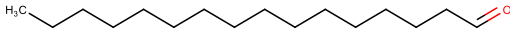
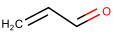
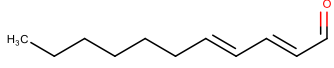

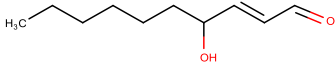
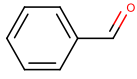


Figure 3: Aldehyde metabolism (in the context of selected oesophageal stressors)

**Table 2: Examples of selected aldehyde subtypes**

Class	Examples	Source	Structure	
Short chain alkanals	C1-C5*	Xenobiotics/environmental; sugar, amino acid, lipid metabolism; LP/LO		Formaldehyde (C1)
Medium chain alkanals	C6-C13	LP/LO		Hexanal (C6)
Fatty alkanals	>C14	Sphingolipid metabolism; plasmalogen catabolism; LP/LO		Hexadecanal (C16)
Enals	Acrolein, crotonaldehyde	Xenobiotics and environmental pollution		Acrolein
Dienal	2,4-decadienal	LP/LO		2,4-decadienal
Dialdehydes	Glyoxal, malondialdehyde	Sugar metabolism, LP/LO		Glyoxal
Bifunctional enals	4-Hydroxy-2-nonenal, Oxo-2-nonenal	LP/LO		4-Hydroxy-2-nonenal
Aromatics	Benzaldehyde, Cinnamaldehyde	Xenobiotics and environmental pollution		Benzaldehyde

\*Cx refers to an alkanal with carbon chain length x throughout this thesis. LP/LO - lipid peroxidation and auto-oxidation

## 1.2.2 General sources of aldehydes

The most important process for diverse aldehyde production is a chaotic phenomenon called “lipid peroxidation” - a chain reaction of radical-initiated hydrocarbon decomposition in lipidic structures such as plasma membranes – which occurs in all cells to some degree (58,59). Depending on the structure of the decomposing lipid, more than 200 aldehyde end-products can be generated (60), and it is potentiated by any process that generates radicalised oxygen or hydroxyl species. Similarly, “lipid auto-oxidation” can produce diverse aldehydes at ambient conditions from simple reactions with oxygen, for example edible oil and animal fat rancidity (61). The presence of a metal catalyst can expedite these reactions, by generating radicalised species through Fenton chemistry (62). The common odour of blood and metal is in fact certain pungent volatile carbonyls (of which 1-octen-3-one, nonanal and decanal are the most prominent) generated by the action of metallic or haem iron on tissue and skin-surface lipids (63). In human pathophysiology, the metal iron storage disorders Wilson’s disease (copper) and haemochromatosis (iron)

are associated with aldehyde-DNA adducts, presumably arising through enhanced Fenton chemistry(64).

Aldehydes are common environmental pollutants. Formaldehyde, acetaldehyde, glyoxal, methylglyoxal, acrolein and aromatic aldehydes are present in petrol and diesel combustion exhaust(65). Industrial incinerators and forest fires also provide a significant environment carbonyl burden. Dialdehydes such glyoxal and methylglyoxal can be produced through photo-chemical reactions on atmospheric hydrocarbons, explaining afternoon ambient spikes of these compounds in cities(66).

Ingested materials can also provide a significant carbonyl stress. In smoke from a reference tobacco cigarette there is 3.23 mg of total carbonyls, of which there is milligram quantities of acetaldehyde, and microgram quantities of diacetyl, formaldehyde, acrolein, propanal, malondialdehyde, glyoxal and methylglyoxal (67). Volatile low molecular weight aldehydes are important aroma compounds(68), for example decanal (meaty), heptanal (fresh-cut grass), cinnamaldehyde (cinnamon), and isovaleraldehyde (nutty). Accordingly they are found in diverse beverages and foods, in particular whiskeys (volatile aldehydes), plant waxes (fatty aldehydes) and charred/smoked meats (aromatic, enals, bifunctional aldehydes). Lastly, oral microbiota are known to locally produce aldehydes in the presence of ethanol (69,70), although the carbonyl metabolism of the oesophageal microbiome has not been assessed.

### **1.2.3 Sources of specific aldehydes**

Myriad metabolic pathways feature specific aldehyde intermediates. Formaldehyde (C1) links the enzymatic production of formate from gastrointestinal methane(65). Acetaldehyde (C2) is free product of the enzymatic decarboxylation of pyruvate, which channels this glycolytic product to the citrate cycle by the formation of acetyl CoA (71). Acetaldehyde additionally participates in amino acid and ethanolamine metabolism, and is produced in great quantity in the catabolism of ethanol and is responsible for the effects of “hangover”. Propanal (C3) is formed in the conversion of glycolytic products to propanoyl CoA, a precursor to mesaconic acid for vitamin B12 cofactor synthesis (65). The catabolism of valine and lysine yields isobutanal (C4). Propenal (acrolein) is produced from spermine metabolism.

The dialdehydes glyoxal and methylglyoxal are particularly important as the presence of two carbonyl carbons means they are both highly reactive and able to cross-link proteins. Glyoxal is an end-product of non-enzymatic oxidative damage to lipids, DNA and glucose(65). Methylglyoxal is formed enzymatically from glycolysis products, ketone bodies or threonine (72). Both are abundant metabolites, and typical plasma concentrations are in the micromolar range(72). These are among the most reactive of all aldehydes and form nucleoside adducts in a few seconds at room temperature(73). Accordingly, these aldehydes have dedicated oxidation enzymes – the glyoxalase system – which is constitutively expressed in all cells. Long chain “fatty” aldehydes have different specific sources, as identified by studies of patients with Sjogren-Larsson syndrome, a specific inborn metabolic error that leads to fatty aldehyde accumulation (see below)(74). Aside from lipid peroxidation, breakdown products of sphingosine metabolism and plasmalogens can produce specific fatty alkanals and alkenals(75). Fatty acid synthesis and beta-oxidation can also yield fatty aldehydes as intermediates.

#### 1.2.4 Genetic mediators of aldehyde metabolism

Aldehydes from these diverse sources are detoxified by three predominant routes: (i) **oxidation to carboxylic acids** (main route, aldehyde dehydrogenase (ALDH) superfamily), (ii) **reduction to alcohols and alkanes** (aldo-keto reductases) (iii) **conjugation to glutathione** (58,76,77) (see Figure 3).

**Aldehyde oxidation** In 2005 the human *ALDH* nomenclature was agreed by international consensus to include 19 isoenzymes split in to 13 classes (summarised in Table 1.2), with class members being grouped by sequence homology (78). All *ALDHs* contain a catalytic site, an NAD(P)<sup>+</sup> cofactor binding site, and homodimerization domain (79,80). Their chief function is to oxidise carbonyl side-groups to carboxylic acids by cleaving water and reducing the cofactor to NAD(P)H. A subset has evolved specialised, specific detoxification functions with a low  $K_m$  and high  $V_{max}$  for their substrates (e.g. *ALDH2* and acetaldehyde, *ALDH5A1* and gamma aminobutyric acid metabolism, etc, see Table 3). Given the enormous potential range of molecules bearing a carbonyl moiety, most *ALDH* isoenzymes have acquired a wide range of specificity, and there is redundancy across the *ALDH* superfamily and



with other detoxification routes (79). They are expressed in all tissues and most highly expressed in the liver (81,82).

Inborn errors in several *ALDH* isoenzymes give rise to specific metabolic syndromes, however all have tissue-limited clinical manifestations, implying non-redundancy is generally context-dependent (see Table 3). For instance, the neurocutaneous Sjogren-Larsson syndrome arises specifically from mutations in *ALDH3A2*, and characterised by spasticity, mental retardation, and squamous hyperplasia manifesting as ichthyosis(83). *ALDH* genes play significant role in several regulatory and signalling networks, including ligand formation (e.g. by producing retinoic and phytanic acid, or removing methylglyoxal and hydroxynonenal), redox control and potentially direct non-metabolic effects (see below, and Table 3). For example, molecular studies of *ALDH3A1* suggest it has structural roles, is a UV-scavenger, and causes cell-cycle braking, in addition to its usual aldehyde detoxification properties(84–86).

**Other aldehyde detoxification pathways** There are 125 known aldo-keto reductases which perform general and specific aldehyde reduction in a tissue-specific manner (for example, recycling of aldehyde neurotransmitters)(87). Additionally, there are 22 glutathione-s-transferases, including the main detoxification routes for glyoxal and methylglyoxal detoxification, glyoxalases I and II. Stable isotope tracing experiments assessing 4-hydroxy-2-nonenal (HNE) detoxification revealed that ALDH oxidation is the preferred metabolic route under normoxic conditions, with alternative pathways reserved for hypoxic conditions(88). This exemplifies the challenging nature of aldehyde biochemistry – the dominant metabolic phenotype will depend on (i) the specific aldehyde (ii) tissue-specific gene expression (iii) microenvironmental pressures, with enormous potential diversity at each step.

**Table 3: Aldehyde dehydrogenase isoenzymes and selected properties**

Name	Locus	Cell localisation	Specificity	Substrate/Function	Clinical relevance
ALDH1A1	9q21	Cytosol	Broad	Retinal, DOPAL, Acetaldehyde, LPOP	?CSC marker
ALDH1A2	15q21	Cytosol	Broad	Retinal, DOPAL, Acetaldehyde, LPOP	?CSC marker
ALDH1A3	15q23	Cytosol	Broad	Retinal, DOPAL, Acetaldehyde, LPOP	Silenced in gastric cancer; ?CSC marker
ALDH1B1	9p13	Mitochondria	Broad	LPOP	Highly expressed in colon cancer
ALDH1L1	3q21	Cytoplasm	Specific	10-formyltetrahydrofolate	Implicated in astrocytomas; Role in cancer progression
ALDH1L2	12q23	Mitochondria	Specific	10-formyltetrahydrofolate	
ALDH2	12q24	Mitochondria	Broad	Acetaldehyde (very low Km); LPOP	Inactivating polymorphisms increase risk of cancer
ALDH3A1	17p11.2	Cytosol, nucleus	Broad	LPOP (high Km for medium and aromatic)	Smith-Magenis syndrome; role in cell cycle
ALDH3A2	17q11.2	Microsome, membranes	Broad	LPOP (high Km for 'fatty' aldehydes)	Sjogren-Larsson syndrome
ALDH3B1	11q13	Cytosol	Broad	LPOP (high Km for long-chain)	
<i>ALDH3B2</i>	<i>11q13</i>	?	?	Unknown	Maybe pseudogene as codon 17 is a stop.
ALDH4A1	1p36	Mitochondria	Broad	Proline degradation, LPOP (short to medium chain)	Type II hyperproliferation; induced by p53
ALDH5A1	6p22	Mitochondria	Specific	GABA catabolism	Mutations cause 4-hydroxybutyricaciduria
ALDH6A1	14q24	Mitochondria	Specific	Valine and pyrimidine catabolism	Mutations enrich hydroxy-carboxylic acid
ALDH7A1	5q23	Mitochondria	Specific	Lysine catabolism	Implicated in prostate cancer
ALDH8A1	6q23	Cytosol	Broad	9-cis-retinal, LPOP (Km increasing with chain length)	
ALDH9A1	1q24	Cytosol	Broad	g-aminobutyraldehyde, betaine, DOPAL, acetaldehyde	?Non-alcoholic steatohepatitis
ALDH16A1	19q13	?	?	Unknown	
ALDH18A1	10q24	Mitochondria	Specific	Glutamate catabolism	Progeria-like neurocutaneous syndrome

CSC, cancer stem cell; LPOP, lipid peroxidation products; DOPAL, 3,4-dihydroxyphenylacetaldehyde

## 1.3 Aldehyde metabolism in cancer

### 1.3.1 Reactive aldehydes as biomarkers of cancer

Aldehydes uncommonly feature in untargeted metabolomics analyses as their reactivity and wide range of polarity is not typically accounted for in routine MS profiling methods and their hydrocarbon structure is poorly resolved by routine <sup>1</sup>H-NMR spin-echo sequences. An exception is the profiling of “volatile organic compounds” (VOC) in breath or in the headspace above biofluids using direct MS, as many aldehydes are volatile. Thus, the majority of evidence for aldehydes in cancer comes either from untargeted analyses in volatile matrices, or from targeted analyses in non-volatile matrices following a specific method development. All are single-centre and of exploratory character, and a selection is briefly reviewed.

**Gaseous biosamples** Lung cancer was among the first malignancies to be subject to breath analysis. In 1999 a Lancet paper described the breath analysis of 108 patients with abnormal chest radiographs found hexanal (C6) and heptanal (C7) among a panel of 22 VOCs that were discriminant for cancer(89). Twelve of the remaining compounds were alkanes or alkenes, i.e. redox partners of aldehydes. This study reported 100% sensitivity and 83% specificity for the detection of cancer; on later cross-validation in independent work, these indices fell to 85% and 81% respectively(90). A 2010 Italian study found average breath alkanals from C3-C9 to be 2-3 fold higher across 40 non-small-cell lung cancer patients, compared to matched controls(91). Fuch et al described differences in C1-C10 alkanals in lung cancer patients' breath, and found C5, C6, C9, and C10 to be significantly enriched, including a median 10 fold increase for C9 and C10 (92). However, the concentrations of >40% of samples were below the limit of detection, implying analytical refinement is required.

In a 2006 report studying 51 patients with breast cancer using sorbent-trap-GC/MS, the Philips group also reported several discriminatory breath VOCs including C7 (93). A more recent report determined that nonanal and decanal in addition to hexadecane and butanone were enriched in the breath of patients with ovarian cancer, which is striking as the aetiological paradigm for this disease is epigenetic rather than carcinogen-mediated (94). The same group have also reported a breath study in gastric cancer featuring 484 unique participants, with eight VOCs identified to be

enriched(95). These included the aldehyde furfural in addition to hexadecane and butanone. Spanel and co-workers identified C1 to be significantly increased in the urine headspace of patients with prostate cancer(96). A recent report in colorectal cancer by Altomare et al using thermal-desorption (TD)-GC-MS described 15 discriminating VOCs including the aldehyde C10.

In four papers increased concentrations of multiple aldehydes measured by SIFT-MS were detected in the breath and biofluids of patients with oesophageal and gastric malignancies(15–18); these are considered below in Chapter 1.4.

**Liquid biosamples** In 1997 Yazdanpah and colleagues validated a GC-MS method for aldehyde quantitation in biofluids, and used this to assess carbonyl profiles 27 cases of childhood cancer(97). A range of 27 different aldehydes showed disease-specific profiles, although this work was not further validated. Guadagni and colleagues used solid phase microextraction GC-MS to assess urine headspace hexanal and heptanal in 10 patients with lung cancer, and noted significantly increased concentrations of both compounds (98). These aldehydes were found to be similarly enriched in lung cancer blood headspace, measured with SPME-GC/MS. Xue et al applied the same technique to liver cancer blood headspace, and found sensitivities and specificities of 94.7% and 100% for hexanal, 84.2% and 100% for 1-octen-3-ol, and 89.5% and 100% for octane (99).

**Tissue biosamples** A study from Poland in 2007 used a dinitrophenylhydrazine LC-MS/MS approach to measured lipid peroxidation aldehydes in astrocytomas and found concentrations of lipid peroxidation products such as hydroxyhexanal and hydroxynonenal to be significantly increased in higher grade disease(100). A recent Austrian study measured a variety of alkanes, alcohols, ketones, and aldehydes, in the headspace of surgical lung cancer specimens using TD-GC/MS. C1, C2, C3, C4, C8, C10 and acrolein were all easily quantifiable and increased in cancer, although only C2 was significantly increased in these samples (101). These were the only studies identified which profiled aldehyde analysis in tissue.

This evidence indicates that the biochemistry of reactive aldehydes and their redox partner compounds (alkanes/alkenes, alcohols, and carboxylic acids) is likely to be fundamental to cancer, although the molecular mechanisms and clinical utility are not established. The first large-scale studies supported by sensitive analytics, clinically-orientated analyses and appropriate validation are beginning to be reported.

### 1.3.2 Aldehyde genes and cancer

***Aldehyde metabolising genes as oncosuppressors*** Recently, two large-scale genome-wide association studies indicated that polymorphisms near *ALDH1A2* increase risk of future Barrett's oesophagus(102,103). Methylation silencing of *ALDH1A3* is associated with gastric cancer(104). Inactivating variants in *ALDH2* are the most common of all human inherited non-silent somatic polymorphisms, and they predispose to oropharyngeal and oesophageal malignancies(105,106). These data have been extended to show that acetaldehyde-DNA adducts are enriched in Japanese alcoholics harbouring *ALDH2* risk alleles (107). Further epidemiological series implicate molecular aldehydes (e.g. acetaldehyde, benzaldehyde) in oesophageal squamocellular carcinogenesis(71,108), in addition to other malignancies(70,109). Methylation silencing of glutathione pathways typify Barrett's oesophagus and progression to dysplasia(110).

From a mechanistic perspective, two groups have demonstrated enriched acetaldehyde-DNA adducts in the oesophageal mucosa of an *Aldh2* knockout mouse that was fed ethanol (23,111). One group went on to show similar DNA adducts in gastric and hepatic samples from the same model (112,113). In other work in haematopoiesis, loss of *Aldh2* and a specific DNA repair pathway (*Fandc2*) generated a particular susceptibility to ingested ethanol, and was sufficient to generate leukaemias in double-knockout mice, reaffirming the carcinogenic nature of acetaldehyde (114,115). Conversely, a model of *ALDH2* over-expression reported attenuated acetaldehyde-mediated cytotoxicity and genotoxicity (116).

Inborn metabolic errors consequent to ALDH variants can provide some insights into how these genes might contribute to malignancy. Typically, substrate accumulation causes specific neurological or neurocutaneous syndromes (e.g. *ALDH4A1* and hyperprolinaemia; *ALDH3A2* and fatty alcohols in Sjogren-Larsson syndrome), implying that only certain tissue are sensitive to *ALDH* loss-of-function. For example, loss of *ALDH3A2* is associated with keratinocyte hyperplasia. These syndromes are also extremely rare, suggesting that inactivating somatic traits are only compatible with life given appropriate collateral genetics. In recent OAC/OSCC sequencing studies there are no recurrent mutations in aldehyde genes (although only *TP53* is recurrently mutated in >10% of either malignancy (33,34,117)), although ALDH effectors affect proliferation and differentiation (see Table 3). Thus, ALDH expression can influence oncogenic phenotypes, but only in specific contexts and poised states.

**Aldehyde metabolising genes as oncogenes** There has been much focus on aldehyde metabolising genes in the context of normal and cancer stem cells, owing to an assay which identifies subpopulations with high tumorigenic qualities based on their ability to metabolise an aldehyde-bearing fluorescent pro-dye (Aldefluor™)(118). This property has traditionally been attributed to *ALDH1A1*, although more recently *ALDH1A3* (119) and *ALDH2* (115). *ALDH1A1* is thought to identify a tumorigenic subpopulation of gastric adenocarcinoma cells(120). Several groups have also noted high *ALDH1A1* expression in OSCC tumorigenic subpopulations, and high tissue immunoreactivity is associated with aggressive and advanced tumours (121–123). In non-small cell lung cancer, *ALDH1A1* and *-3A1* expression selects for cancer stem cells and highly expressing tumour are associated with adverse outcomes(124). In prostate cancer, strong expression of *ALDH7A1* is associated with disease progression and aggression. However, there are no reports of *ALDH* over-expression in OAC (in fact there are no convincing descriptions of OAC cancer stem cells by any method). In summary several aldehyde-detoxifying genes have some cancer association, although these relationships are not uniform, often contradictory, and seem context specific.

### 1.3.3 Consequences of deregulated aldehyde metabolism

Aldehydes participate in diverse chemical reactions at ambient temperatures. From a toxicity perspective, the most important are nucleophilic-condensation reactions with thiol and amine nucleophiles, which are widespread subgroups on biological molecules. An imine conjugate is formed (“aldehyde adduct”) with loss of a water molecule. Thus, the partner molecule’s structure is altered with potential significance for function. Evidence suggests that 1 in 10 proteins will be carbonylated at some point, rising to 1 in 3 in the last third of life(125,126). One example is protein glycation in diabetic hyperglycaemic states, which is mediated by the carbonyl-C1 carbon of linear monosaccharaides(127). Another is protein adducts of lipid peroxidation products such as 4-hydroxy-2-nonenal (HNE), malondialdehyde and glyoxal. These bifunctional adducts can lead to protein crosslinking and accumulation, and may form the basis for multiple deposition-based diseases(125,126). Interestingly target binding can be specific; HNE adducts have been shown to selectively impair glutathione metabolism and proteasome activity, activating feedback loops that propagate protein dysfunction(128). Tools for

assessing protein carbonyl based on polyclonal antibodies have been available since the 1950s, and thus a significant literature has developed, assessing these adducts in disease ranging from ischaemic heart disease to Alzheimer's and cancer.

Aldehydes can also react with DNA amine groups. This can alter the macromolecular structure, affecting how polymerases, histones and repair proteins interact with the double-helix(65). Worse still, if the modified amine normally participates in Watson-Crick base-pairing, the distorting adduct could interfere with normal transcription, leading to polymerase arrest, polymerase mismatch (i.e. mutation), and even double strand breaks. For example, in a bacterial model, site-specific mutagenesis from an acetaldehyde-adenosine adduct reveals transversion preference at a rate of 2-7%(129). Thus, multiple aldehydes have thus been designated suspected or confirmed mutagens by the International Agency for the Research on Cancer (IACR), including formaldehyde, acetaldehyde, malondialdehyde, furfural, glyoxal, methylglyoxal, acrolein, crotonaldehyde and 4-hydroxy-2-nonenal (130). The carcinogenicity of these compounds (in contrast to linear sugars, fatty aldehydes etc) is explained by their small size, which allows them to infiltrate the nucleus and then DNA grooves, permitting exposure of their carbonyl carbon to relevant DNA amine. As with protein carbonyls, aldehyde-nucleoside adducts are stable compared to free aldehydes, and therefore appropriate targets for genotoxicity analyses (58,60,131).

The carcinogenicity of several aldehydes have been confirmed in long-term animal studies(130). In humans, formaldehyde is linked to the development of leukaemias (reviewed in 70). There is now considerable evidence linking acetaldehyde from ingested ethanol to numerous cancers (reviewed in 50,76). Observational series have shown increases in aldehyde-DNA adducts in leucocyte DNA of alcoholics (134). In oesophageal cancers, epidemiological series have linked excess alcohol with both squamous cell and adenocarcinoma of the oesophagus (135,136).

Certain aldehydes and aldehyde-adducts can also act as direct signalling ligands. With relevance to the oesophagus, acetaldehyde has been shown to stimulate hypertrophy and hyperplasia of rat oesophageal keratinocytes(71). HNE rapidly activates JNK and nf-kB oncogenic pathways, with measurable signalling cascades within 30 minutes of treatment(137) Aldehydes can also modify the cysteine residues of Keap1, allowing stabilisation and nuclear translocation of the pro-survival transregulator nrf2(138,139). Inhaled 2,4, decadienal can induce cell proliferation in bronchial epithelial cell, etheno-DNA adducts and murine lung tumours(140–142).

However, all aldehydes induce apoptosis at high concentrations, typically through BCL2 and c-Jun pathways. There is also evidence that aldehyde-modified nucleotides can influence purinergic receptor signalling(143), which are important in oesophageal inflammation and motility(144).

#### **1.3.4 Aldehyde metabolism the oesophagus**

The oesophagus additionally undergoes a number of specific aldehyde stress processes. The undigested food bolus may contain free aldehydes, either native to the food, from the cooking method (e.g. the Maillard reaction of meat searing), or from the actions of oral microbiota (69,70,109). Tobacco smoke also contains numerous carbonyl and radical species, which can be swallowed either directly (aerophagy), solubilised in saliva, or delivered haematologically to the oesophagus. Aldehydes may be produced *in situ*, most importantly through the enzymatic oxidation of consumed alcohols (135,145). Additionally, lipids of epithelial cells will undergo auto-oxidation or peroxidation (e.g. with ionic iron from ingested red meat); the recent findings of peroxidation-prone lipidomes in Barrett's metaplasia and adenocarcinoma raise the possibility of corresponding characteristic aldehyde flux (146,147).

In the refluxing oesophagus, local processes add to carbonyl stress. Reflux of bile salts solubilise pre-epithelial hydrophobic protectants, exposing the apical membrane to chemical attack(148). Cell necrosis from low pH will generate inflammatory cytokines, leading to neutrophil activation, the production of superoxide radicals, and enhanced peroxidative phenotypes(126). Most importantly, carbonyl addition reactions require protonation as a rate-limiting step, and therefore low pH greatly enhances aldehydes' ability to react with nucleophilic partners. This is uniquely important in the oesophagus, as the poorly developed pre-epithelial alkaline barrier affords little protection compared to the stomach.

So far, metabolomics studies in oesophageal malignancies have reported aldehyde enrichment in breath and urine, although none have specifically assessed carbonyls in oesophageal tissues or blood (see section 1.1.3). However, numerous precursors of aldehyde metabolism have been shown to be enriched in oesophageal cancer tissues and biofluids, including branched-chain essential amino acids, sphingosine-1-phosphate, and unsaturated lipids. Additionally, numerous products of aldehyde



metabolism have been shown to be enriched, including aliphatic carboxylic acids from C1-C16, branched-chain carboxylic acids and alkanes.

In addition, genomic, methylomic and epidemiological studies (see 1.3.2) indicate that impaired aldehyde detoxification pathways predispose to and are associated with Barrett's metaplasia and OAC. In particular, glutathione transferases are widely methylation silenced (19), and glutathione itself is generally and sequentially depleted through disease progression(149). Additionally, there is evidence of *de novo* activated fatty acid synthesis in OAC (147), which can place enormous pressures on redox capacity and impair alternative detoxification routes for aldehydes (a single palmitate molecule requires 14 reducing equivalents for *de novo* synthesis(150)). Thus, there is the potential for a "perfect storm" of increased aldehyde production with enhanced carbonyl reactivity and impaired detoxification (through genetically-determined "metabolic reprogramming"), which is unique to the refluxing oesophagus, and which is entirely unaddressed by the literature.

## 1.4 Gaps in the literature

These connected gaps in the literature were identified:

1. No data regarding aldehyde quantitation in oesophageal tissue, despite evidence that in OAC (i) aldehydes are present in breath and urine samples (ii) the oesophagus undergoes mechanisms of and is uniquely susceptible to aldehyde stress (iii) detoxification systems are impaired (iv) metabolomics evidence reports enriched aldehyde precursors and redox products (v) extreme and untargeted genotoxic burden cause by unknown mutagens
2. Very limited aldehyde-biofluids data in any cancer, including OAC.
3. No assessment of the genetic basis of aldehyde detoxification in OAC
4. No quantitative data regarding aldehyde-nucleotide interactions in OAC despite (i) evidence of aldehyde-DNA adducts in animal models of oesophageal cancer and clinical material from related cancers (ii) evidence of severe and indiscriminate DNA damage by unknown mutagens

***Rationale for OAC disease focus*** Aldehyde metabolism may be important in both OAC and OSCC, which is consistent with the view that the structure and function of this organ is a key determinant of oesophageal aldehyde stress. It was decided to focus the work on OAC, because: (i) OAC formed the case majority in the index non-invasive biomarker studies (ii) the changes in aldehyde detoxification pathways in OAC were more pronounced in the exploratory transcriptomic screens (see Chapter 3) (iii) these are highly heterogeneous diseases, and for the purposes of mechanistic dissection, it is important to rationalise the studied biology as much as possible (iv) the OAC disease burden is higher in the UK.

## 1.5 Thesis hypothesis and objectives

This thesis addresses these gaps by pursuing the following hypothesis: ***Metabolic reprogramming enriches aldehydes in the OAC microenvironment and this contributes to oncogenesis.*** Thus three interconnected objectives were planned.

1. Assess aldehydes metabolic profiles in OAC tissue samples
  - i. Develop and validate an analytical method to unambiguously quantify common aldehydes in tissue samples and other biospecimens
  - ii. Use this method to quantify aldehydes in OAC tissue samples and appropriate controls
  - iii. Also use this method to determine aldehyde metabolic profiles in OAC murine-xenografts and OAC biofluids.
  
2. Assess the genetic basis for aldehyde metabolic reprogramming in OAC
  - i. Use a candidate-based approach to discover candidate drivers of aldehyde reprogramming, initially using archived expression datasets
  - ii. Validate leading candidates by direct expression analysis in clinical material
  - iii. Establish the clinical significance of candidate gene deregulation
  - iv. Position candidate gene deregulation within OAC transformation
  - v. Discover upstream coordination of candidate perturbation
  - vi. Define aldehyde gene expression in normal and malignant oesophageal models, and compare this to aldehyde phenotypes
  
3. Assess the effects of aldehyde metabolic reprogramming in OAC
  - i. Develop and validate a second analytical method to measure aldehyde-nucleotide adducts in DNA hydrolysates, and use this to quantify aldehyde DNA damage in OAC DNA samples and controls
  - ii. Provide *in vitro* demonstration that candidate reprogramming driver genes are sufficient to generate metabolic flux
  - iii. Provide *in vitro* dissection of the non-metabolic oncogenic effects of aldehyde reprogramming arising from key genetic drivers.

## **CHAPTER 2 – ALDEHYDES IN OESOPHAGEAL ADENOCARCINOMA**

## Summary

Biologically relevant reactive aldehydes pose analytical challenges owing to volatility, reactivity, and variation in polarity. Thus, numerous quantitative approaches have been attempted including ranging from colorimetric and immunoreactive assays to modern methods based on mass spectrometry. Derivatisation to dinitrophenylhydrazine (DNPH) prior to liquid chromatography triple quadrupole mass spectrometry (LC-MS/MS) has proven a popular analytical choice as sub-ng/ml concentrations can be unambiguously determined over a broad range of chain length and configurations. However, analytics can be complicated by other biologically relevant DNPH-reactive isomers/isobars, which may be significantly more abundant in biological samples. For example, aliphatic ketones, alkanals and saturated dialdehydes are isomers at the same carbon chain length (C-1 for dialdehydes). In this chapter, an LC-MS/MS method was developed to measure 43 carbonyl compounds comprising 35 aldehydes (18 aliphatic, 2 aromatics, 9 enals, 6 bifunctional aldehydes), and 8 ketones. The method was sensitive to the low ng/ml range and gave accurate results from biofluids and tissue samples. The method was applied to the measurement of aldehydes in human and murine-xenograft OAC tumours, OAC patient biofluids, and relevant controls.

This section was undertaken in collaboration with Dr Zsolt Bodai, who closely supervised the student in the initial characterisation of the method.

## 2.1 Methodological rationale

### 2.1.1 Choice of analytical technique

**Analytical challenges** Compared to their redox partners fatty acids, aldehydes are uncommonly measured metabolites (151). Reasons for this include(79,152–155):

- (i) High reactivity of the electrophilic carbonyl group in biological matrices leading to analyte instability and necessitating derivatisation
- (ii) A range of volatility, necessitating derivatisation
- (iii) A range of polarity and solubility with increasing chain length.
- (iv) Some aldehydes (acetaldehyde (C<sub>2</sub>), glyoxal) are ubiquitous, and can contaminate analytical glassware and solvents increasing background
- (v) Large range of specific, aldehyde-active enzymes in biological matrices
- (vi) Less likely to be identified in untargeted metabolomic screens, in particular nuclear magnetic resonance and non-derivatised MS

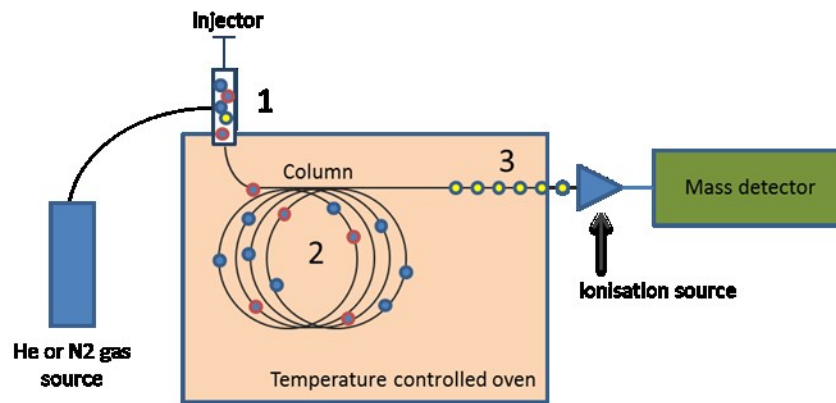
This diverse set of analytical challenges means that no single analytical platform is perfectly suitable for measuring aldehydes in liquid or solid samples, and thus a large number of techniques have been reported. The first techniques included colorimetric assays, gas-liquid chromatography, and thin layer chromatography. More recently, mass spectrometry coupled to a separation technique (e.g. gas chromatography, liquid chromatography) or a selection technique (selected ion flow tube or proton transfer reaction mass spectrometry) has become the standard for unambiguous quantitation (131,152,156).

**Early techniques and non-specific detectors** Brady and Elsmie (1926) are credited with first promoting the use of dinitrophenylhydrazine for the derivatization of aldehydes and ketones to stable, non-polar dinitrophenylhydrazones (157). The different hydrazones were initially described by their characteristic crystallisation properties. With the advent of ultraviolet spectroscopy, the yellow-orange hydrazones could be determined by their characteristic absorption spectra, and this was the mainstay of aldehyde quantitation until the 1980s together with thin layer chromatography(156,158,159). Other carbonyl-reactive dyes have been described, which form colorimetric or fluorimetric detectable products, and which are suitable for summative aldehyde quantitation in the context of a rapid assay. The most

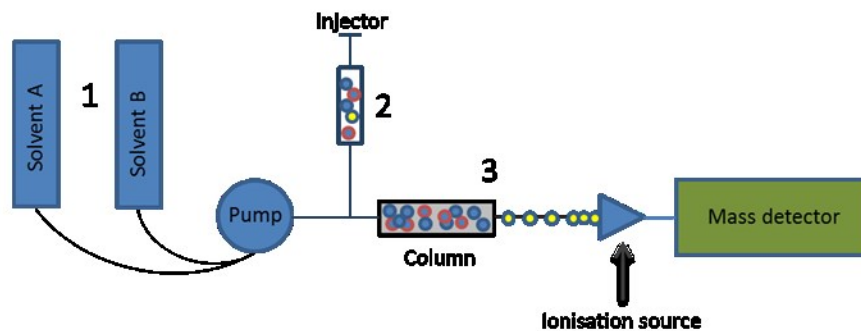
commonly used was the “thiobarbituric acid reactive substances” (TBARS) assay, which was thought to be specific for malondialdehyde but in fact was widely reactive with hydroperoxides and conjugated aldehydes(160). These inherently limited methodologies were widely superseded by hyphenated mass spectrometry over 30 years ago.

**Gas chromatography with mass spectrometry (GC-MS)** Biological matrices’ extraordinary complexity complicates targeted quantification of individual components. Within the remit of mass spectrometry, the selectivity of the mass spectrometer is limited by mass resolution, and so different compounds of equal mass-to-charge ratio will be indistinguishable in the detector. Separation techniques seek to overcome this, and including pre-analytical sample preparation, followed by gas or liquid chromatography. Gas chromatography separates molecules using thermal energy (see Figure 4). The sample is evaporated to high temperature and applied to a low temperature column, and is retained. The column then is slowly heated, and individual compounds elute from the column, depending on boiling point, polarity, column properties etc. In addition, the gaseous nature of samples facilitates sample pre-concentration (e.g. thermal desorption, needle-trap device, solid phase microextraction, and solid phase extraction), thus improving sensitivity. Deng et al described blood aldehyde quantification by solid phase microextraction (SPME) - GC-MS, and commented that the “volatility and activity” of aldehydes made quantification from the liquid phase very challenging (161). These properties will be enhanced at high temperature, and so prior modification to a stable state is necessary, most commonly derivatisation to pentafluorobenzylhydrazine (PFBHA) (162–164).

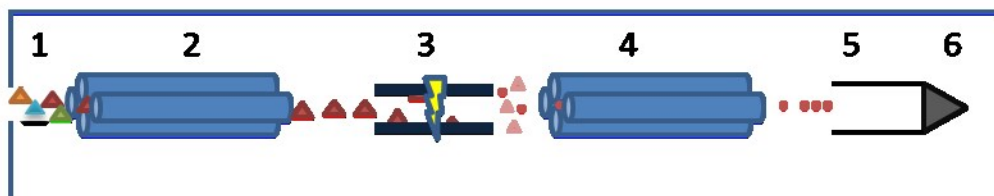
In the context of cancer, GC-MS aldehyde applications has included blood, tissue, urine, breath, and breast milk(162,165–167). However, GC-MS carries notable issues in the context of measuring aldehydes in biological matrices. Primarily, the GC-MS system is not well suited to non-volatile compounds, especially carbon lengths >C12 which may be unstable at desorption temperatures(151). Additionally, GC-MS systems are less robust to biological contamination compared to LC-MS systems, and thus stringent sample preparation is needed. Lastly, run times are often x2-3 longer than UPLC-MS methods. With the requirement for additional sample preparation, these issues limit high-throughput analysis and expose the method to drift and batch effects.



**Gas chromatography -MS:** 1. A high temperature complex sample is injected to a low temperature column, and analytes condense onto the column 2. The oven temperature is slowly increased 3. Different compounds elute off the column at different temperatures, allowing individual quantitation in the detector



**Liquid chromatography -MS:** 1. The system is flushed with a particular ratio of two solvents, usually water and an organic solvent such as methanol or acetonitrile 2. The sample is injected into the solvent stream, becoming absorbed onto the column packing through ionic and weak interactions. 3. As the ratio of solvents is slowly changed, analytes elute off the column and are individually ionised and then quantified in the mass spectrometer



**MS/MS mass detector:** 1. Analyte ions from the ion source enter the vacuum housing of the mass spectrometer and pass to the first quadrupole (Q1). 2. Q1 filters the ions according to  $m/z$ . Only specific 'parent ions' enter the collision cell (Q2). 3. Parent ions are fragmented. 4. Daughter ions are selected in the second mass filter (Q3). 5. A stream of pure daughter ions are exchanged for electrons and amplified. 6. The electrons are counted as the mass detector. Electron count is thus proportional to the parent ions entering the mass spectrometer.

Figure 4: Schematics of general principles of GC, LC, and MS/MS.

MS, mass spectrometry. He, helium; N2, nitrogen;  $m/z$  mass-to-charge.

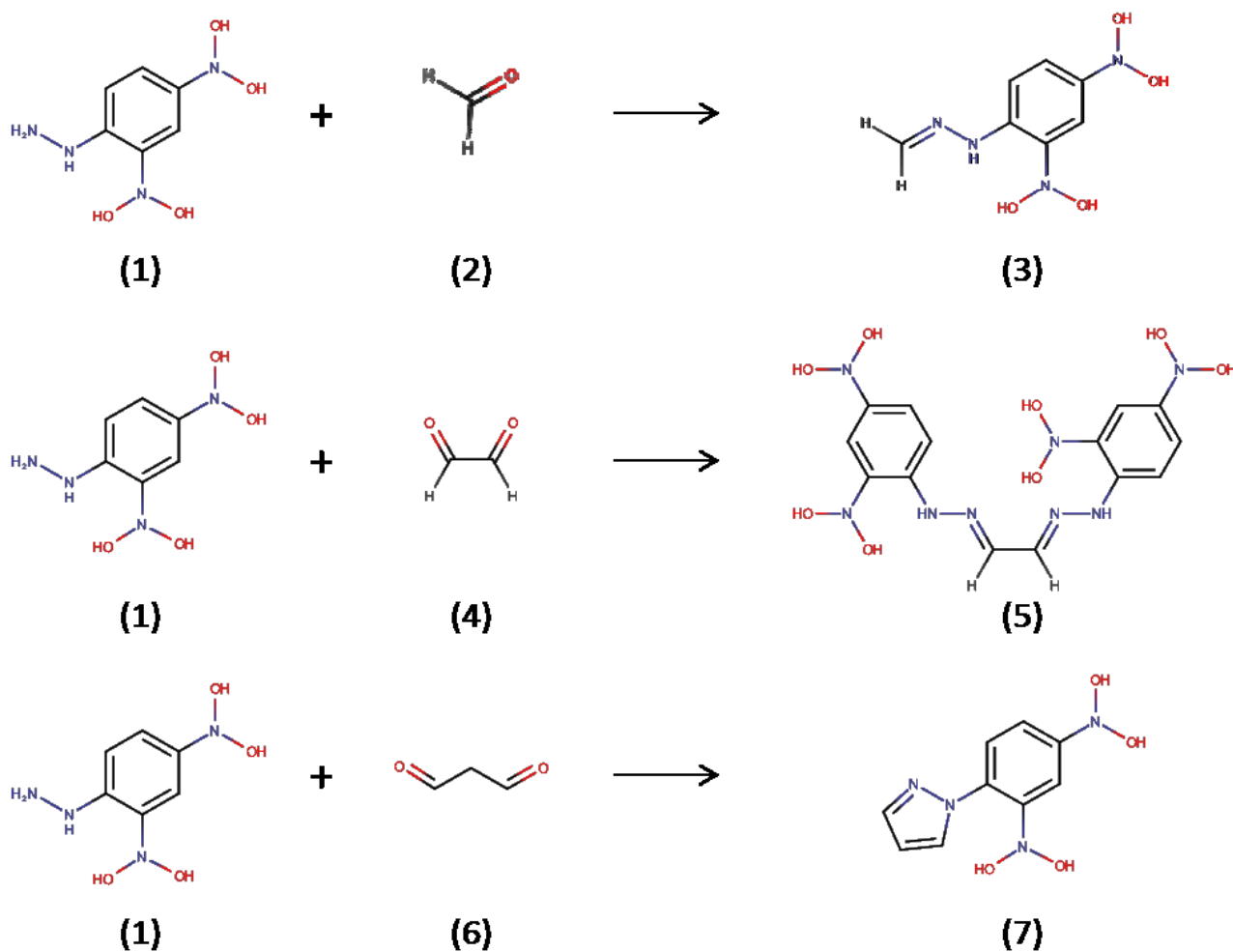


**Liquid chromatography mass spectrometry (LC-MS)** Liquid chromatography separates compounds by how they interact with an adsorbent solid, using a solvent stream of varying dissociative strength (see Figure 4). Thus, samples are prepared and injected in the liquid phase, and so volatile and/or unstable compounds are poorly suited to LC-MS quantification in their native state. Aldehydes can be measured directly by LC-MS, but better sensitivity, recovery and separations can be achieved by derivatization. Owing to its heritage, dinitrophenylhydrazine (DNPH) was among the first derivatisation agents used for this purpose (see Figure 5 for example reactions)(168); there have been recent success with dansylhydrazine, cyclohexanedione, and 4-APEBA (169–172). However, DNPH remains the most popular, and combined with LC-MS remains the choice approach for quantifying aldehydes in the methods of Western environmental standards agencies (e.g. Environmental Protection Agency EPA 8315A (SW-486) (152,173)). The ‘DNP-hydrazone’ products are stable, non-volatile, and non-polar across the aldehyde mass range, and thus are suitable for reverse phase liquid chromatography and sample preparation techniques such as liquid-liquid or solid phase extraction.

Recent publications in aldehyde quantitation have trended toward quantifying many target aldehydes in a single analytical run, including facility for quantifying and identifying unknown carbonyls (169,171,174). This indicates that aldehydes’ are increasingly being recognised as a biologically related group, and that a helpful strategy is to measure particular target aldehydes with others simultaneously. From the technical perspective, there are a number of debated issues in quantifying aldehydes with LC-MS, including derivatisation, sample clean-up (169,175), choice of ionisation source (176–178), dealing with background contamination (145,153), and the implementation of technological advances in instrumentation. For example, the Waters “ultra-performance” liquid chromatography (UPLC) systems utilise smaller particle sizes and higher pressures to provide enhanced selectivity, yet there are only a small number of reports applying this to simple aldehyde quantitation (174,179). Aldehyde LC-MS method literature is further discussed in the Method Development Section 2.3 below.

**Other mass spectrometry methods for aldehyde quantitation** As discussed in section 1.3.2, volatile aldehydes are frequently measured in breath or headspace analyses using direct forms of mass spectrometry (SIFT-MS, PTR-MS) or indirect methods (SPME-GC-MS). The chief advantage of these techniques is that aldehydes’ volatility provides natural separation from most interferences. However,

by its nature, these techniques cannot measure non-volatile aldehydes, and reproducible tissue-headspace analysis of unstable compounds is a challenge. Given that key biomarker aldehydes are poorly volatile (e.g. decanal), it was decided to use DNPH-LC-MS as the full range of aldehydes could be assessed without compromise.



**Figure 5: Dinitrophenylhydrazine reactions with selected aldehyde subtypes**

DNPH (1) forms different products with aldehydes depending on the number and position of carbonyl carbons. Products include (i) Mono-derivatisation, e.g. with formaldehyde (2) forming formaldehyde-dinitrophenylhydrazone (3). (ii) Di-derivatisation, e.g. with the dialdehyde glyoxal (4) forming glyoxal-(1,2)-di-(dinitrophenylhydrazone). (iii) Special derivatization e.g. with dialdehyde malondialdehyde (6), which forms a unique closed ring hydrazone derivative (7)

### 2.1.2 Method goals & selection of targets

The primary objective for this section is to quantify aldehyde concentrations in tissue samples from the normal and malignant oesophagus, with appropriate analytical confidence. Secondary goals are to generalise quantitation to plasma, urine, and cell and animal samples. Thus, the robust method would have to identify low concentrations of highly reactive compounds in complex matrices with potentially strong but inconstant interferences, in particular from lipids. Ultra-high performance liquid chromatography was favoured over gas chromatography as analysis is better suited to high throughput acquisitions, is more robust to complex matrices, and because of the wide range of polarity and boiling points of the target compounds.

The following issues were projected:

1. Use of ESI or APCI ion source
2. Best parent ion to measure bifunctional and dialdehydes.
3. Disambiguation of isomers/isobars
4. DNPH reaction time, pH; adjustment for background and stereoisomerisation
5. Analyte loss due to reactivity of aldehydes during sample preparation
6. Sample preparation suited to high throughput processing
7. Method validation (FDA protocol was selected (180))

A total of 43 carbonyl species were selected for analysis (full list provided in Table 4). These included all available alkanals (18), a selection of enals (9), aromatic aldehydes (2), dialdehydes (4) and other biologically relevant lipid peroxidation products (2), and a selection of ketones (8). The rationale for this large group of targets was to provide a comprehensive description of aldehyde phenotypes, and in particular capture key substrates for suggested drivers of aldehyde reprogramming (e.g. medium chain alkanals and aromatics for *ALDH3A1*, fatty alkanals for *ALDH3A2*; see Chapter 3). Ketones were additionally included as these DNPH-reactive molecules are isomers of alkanals in both free and DNP-hydrazone form. Therefore, the unambiguous determination of alkanals requires the knowledge that they are not ketones.

## 2.2 Hypothesis and aims

Chapter hypothesis: There are significant and consistent differences in aldehyde concentration in OAC tissue and biofluids compared to relevant controls.

The objective of this section was to develop and validate a quantitative LC-MS method for a panel of aldehydes and their DNPH-reactive isomers/isobars, and apply this to understand local and systemic aldehyde dynamics in oesophageal adenocarcinoma. Specifically, this section aimed to:

1. Develop an LC-MS method to unambiguously identify all of the major aldehyde classes, satisfying the analytical goals set out in 2.1.2.
2. Validate the method & determine appropriate biosample-types for aldehyde phenotyping
3. Use this method to determine aldehyde metabolic profiles in the malignant oesophagus and appropriate controls
4. Use this method to determine aldehyde metabolic profiles in OAC patients' biofluids
5. Use this method to compare aldehyde metabolic profiles in patient and subcutaneous murine xenograft tumours.

## **2.3 Aldehyde method development**

### **2.3.1 Materials**

External and selected isotope-labelled standards (ISTDs) for the intended targets were purchased (see Table 4). Analytical grade (UPLC) water, acetonitrile, sodium chloride, formic acid and acetic acid were all purchased from Sigma, as was 2, 4-dinitrophenylhydrazine (0.2M, in 70% phosphoric acid). Malondialdehyde (MDA) was synthesised from the precursor tetraethoxypropane (TEP) by hydrolysing 10 $\mu$ l TEP in 0.1% HCl dissolved in 10ml UPLC water for 10 minutes at 90°C (181). MDA-d2 was similarly prepared from the fully deuterated TEP precursor.

### **2.3.2 Human and murine biospecimens**

Patient samples were accessed following Imperial College Healthcare Tissuebank Review approvals R14097, R15097 and R16018 (see Appendix 2). Samples were collected under standardised protocols designed to preserve metabolic information without artefact formation (182–184). Single biopsies were placed directly into a pre-labelled cryovial (Nunc) and snap frozen in liquid nitrogen, and then moved to -80°C for storage.

Biofluid processing followed similarly standardised procedures(184–187). Urine samples were collected as a mid-stream catch into polypropylene 40mL vials and placed on ice. Aliquots of 500 $\mu$ l were snap frozen as soon as practically possible to -80°C (<3 minutes). For plasma samples, 10 mL whole blood samples were taken by antecubital venepuncture using the direct Vacutainer system (BD Biosciences) into EDTA, mix by inversion 10 times, iced, and centrifuged at 4000 RPM for 3 minutes at 4 °C. The plasma supernatant was removed without disturbing the cellular layers, and frozen in 500 $\mu$ l aliquots at -80°C. The buffy coat was saved for DNA extraction (see Section 5.1).

All animal experiments were conducted with full Home Office approval, under project license 70/7997 (P.I. Prof H Gabra, ICL), and Establishment License 70/2722

**Table 4: Aldehyde materials**

Common Name	Abbrev.	-R	MW	CAS	Supplier
<b>Aldehyde External Standards</b>					
Formaldehyde	C1	H	30.03	50-00-0	Sigma
Acetaldehyde	C2	CH3	44.05	75-07-0	Sigma
Propanal	C3	C2H5	58.08	123-38-6	Sigma
Butanal	C4	C3H7	72.11	50-89-5	Sigma
Pentanal	C5	C4H9	86.13	110-62-3	Sigma
Hexanal	C6	C5H11	100.16	66-25-1	Sigma
Heptanal	C7	C6H13	114.18	111-71-7	Sigma
Octanal	C8	C7H15	128.21	124-13-0	Sigma
Nonanal	C9	C8H17	142.24	124-19-6	Sigma
Decanal	C10	C9H19	156.27	112-31-2	Sigma
Undecanal	C11	C10H21	170.29	112-44-7	Sigma
Dodecanal	C12	C11H23	184.32	112-54-9	Toronto Research Chemicals
Tridecanal	C13	C12H25	198.34	10486-19-8	Sigma
Tetradecanal	C14	C13H27	212.37	124-25-4	Sigma
Pentadecanal	C15	C14H29	226.4	2765-11-9	Tokyo Chemical Industries
Hexadecanal	C16	C15H31	240.43	629-80-1	Sigma
Heptadecanal	C17	C16H33	<b>254.45</b>	629-90-3	Tokyo Chemical Industries
Octadecanal	C18	C17H35	268.49	638-66-4	Tokyo Chemical Industries
Benzaldehyde	Benz	Benzene ring	106.12	100-52-7	Sigma
Cinnamaldehyde	Cinn	Benzene-C2H4-	132.16	104-55-2	Sigma
Acrolein	Acr	C2H3	56.06	107-02-8	Sigma
Crotonaldehyde	Cro	C3H5	70.09	4170-30-3	Sigma
Trans-2-pentenal	C5=	C4H7	84.12	1576-87-0	Sigma
Trans-2-hexenal	C6=	C5H9	98.14	6728-26-3	Sigma
Trans-2-heptenal	C7=	C6H11	112.17	18829-55-5	Sigma
Trans-2-octenal	C8=	C7H13	126.2	2548-87-0	Sigma
Trans-2-nonenal	C9=	C8H15	140.22	18829-56-6	Sigma
2,4-nonadienal	C9==	C8H13	138.22	5910-87-2	Sigma
2,4-decadienal	C10==	C9H15	152.21	25152-84-5	Sigma
4-Hydroxy-2-nonenal	HNE	C8H14OH	156.22	29343-52-0	Cayman Chemicals
4-oxo-2-nonenal	ONE	C8H13O	154.21	103560-62-9	Cayman Chemicals
Tetraethoxypropane (for MDA)	MDA	CH2*	72.06	542-78-9	Sigma
Glyoxal	Gly	.*	58.04	107-22-2	Sigma
Methylglyoxal	MGLy	CH3*	72.06	78-98-8	Sigma
Glutaraldehyde	Glut	C2H5*	100.117	111-30-8	Sigma
<b>Ketone External Standards</b>					
Acetone	K3	n/a	58.08	67-64-1	Sigma
2-Butanone	K4	n/a	72.11	78-93-3	Sigma
2-Pentanone	K5	n/a	86.13	107-87-9	Sigma
2-Hexanone	K6	n/a	100.16	591-78-6	Sigma
2-Heptanone	K7	n/a	114.18	110-43-0	Sigma
2-Octanone	K8	n/a	128.21	111-13-7	Sigma
2-Nonanone	K9	n/a	142.24	821-55-6	Sigma
2-Decanone	K10	n/a	156.27	693-54-9	Sigma
<b>Internal Standards</b>					
Tetraethoxypropane-d2 (for MDA-d2)	MDA-d2	n/a	74.06		Santa Cruz
Acetaldehyde-d4	C2-d4	n/a	48.05		Sigma
Hexanal-d12	C6-d12	n/a	112.16		Sigma
Hexadecanal-d5	C16-d5	n/a	245.43		Santa Cruz
Hydroxynonenal-d3	HNE-d3	n/a	159.22		Cayman Chemicals
Oxononenal-d3	ONE-d3	n/a	157.21		Cayman Chemicals

\*These are dialdehydes, i.e. two functional aldehyde groups, separated by the indicated atoms. MDA, malondialdehyde; MW, molecular weight; CAS, chemical abstracts service registry number.

(Central Biomedical Services, Imperial College London). The Candidate obtained a personal license for the work, Home Office ref: ICFFBE6F8. For xenograft experiments, FOX-P1<sup>nu/nu</sup> mutant mice (athymic nudes) were purchased (Envigo, Surrey, UK) at 6-8 weeks age and allowed to settle for one week. The mice were fed sterilised water and chow *ad libetum*.

At the start of xenografting experiments, 4 million OAC cells were injected into one flank only (5 mice per cell line) in a 50:50 mix of Matrigel® (Sigma) and culture media (see Chapter 4). The original intention was to take paired tissue and plasma samples at the end of the experiment. The mice were weighed daily and checked for adverse local and constitutional sequelae of tumorigenesis. Tumours were allowed to develop until the project license limit (200mm<sup>3</sup>) at which point the mice were culled by lethal intra-peritoneal injection of Euthatol™ (Envigo), with secondary confirmation by femoral laceration. This termination route was selected the other available methods (cervical dislocation, CO<sub>2</sub> asphyxiation, intravenous lethal injection) could all compromise adequate blood sampling. The samples were frozen as dissected and the whole procedure was completed <3 minute after confirmation.

### **2.3.3 Instrumentation**

All aldehyde experiments were conducted using a Waters (Acquity) UPLC binary solvent manager and autosampler interfaced to a Waters Xevo TQ-micro (MS/MS) detector, equipped with either APCI or ESI ion sources depending on the experiment. The proprietary Masslynx software (SCN 909) was used for all experiments.

### **2.3.4 Initial settings**

The Waters automatic optimisation Intellistart™ program was used to define initial source and optimum compound-dependent mass spectrometry, using a sample infusion. The outcome of this process is given in Table 5. Optimisation standards were prepared as 1 mL aliquots in 50:50 water:acetonitrile, a derivatised with 20µl 80mM DNPH for 1 hour at 25°C without further pH modifications (it was found these conditions fully derivatised the aldehyde stock solutions, following the optimisation experiments described in 2.3.9).

**Table 5: Initial instrument settings****Sample manager**

Injection volume ( $\mu\text{l}$ )	1-5
Temperature ( $^{\circ}\text{C}$ )	4
Column	Cortecs C18 1.6 $\mu\text{m}$
Precolumn filter frit size ( $\mu\text{m}$ )	0.2
Column Temperature ( $^{\circ}\text{C}$ )	35

**Binary solvent manager**

Flow rate (ml/min)	0.5
Mobile phase A/Weak needle wash	Water
Mobile phase B/Strong needle wash	Acetonitrile
Seal wash	Water
Aqueous modifier (concentration, mM)	None

**Source settings**

Type	Electrospray ionisation (pos + neg mode)
Source temperature ( $^{\circ}\text{C}$ )	150
Desolvation temperature ( $^{\circ}\text{C}$ )	400
Desolvation Gas Flow (l/Hr)	650
Cone Gas Flow (l/h)	200
Capillary voltage (V)	2500
Cone voltage (V)	10

**2.3.5 Aldehyde precautions**

As detailed in the EPA protocol and others, the ubiquity of aldehydes can lead to unacceptable background interference if appropriate precautions are not taken. Thus, all glassware was scrupulously cleaned immediately after use and left in a 70 $^{\circ}\text{C}$  oven for at least 3 hours to evaporate contaminants(152,153,188). Background values in all reagents were assessed prior to experiments. A dedicated pipette was used for liquid handling of DNPH. All external standards at any concentration were handled in a Class 1 environment. All non-derivatised aldehydes were handled on ice or dry ice depending on the solvent and experiment, including biological specimens. Samples were moved to an unused 2mL glass or 0.3mL polypropylene vial prior to analysis, and sealed with a Teflon lined cap.



### 2.3.6 MS/MS development

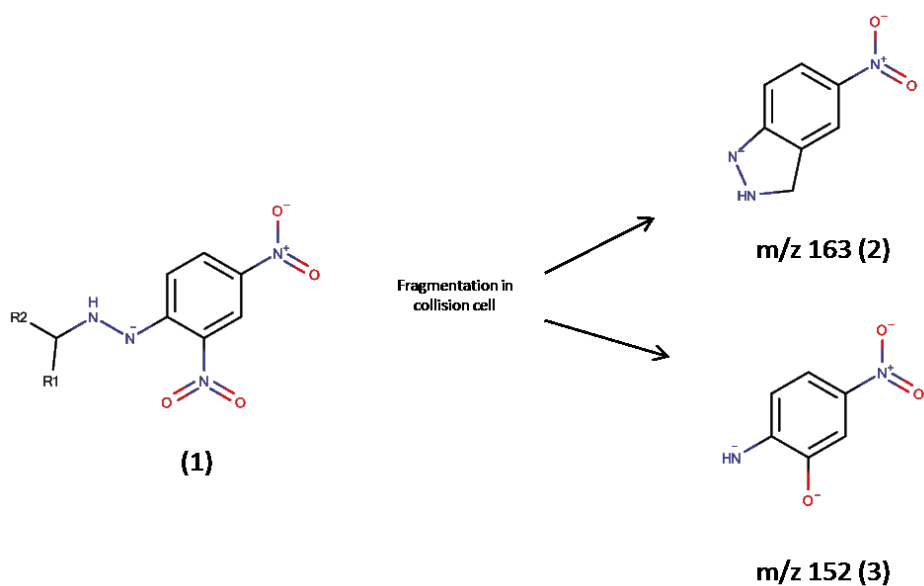
Ionisation was initially carried out in the negative mode, using the  $[M-H]^-$  predicted parent ion mass of the mono-derivatised DNP-hydrazone (169,176–179). This strategy worked well for most targets, although for MDA and other bifunctionals, the  $[M-H]^-$  intensity was weak. A literature evaluation revealed that MDA preferentially forms a unique cyclic DNPH derivative, which only ionises in the positive mode (see Figure 5). Additionally, bifunctional aldehydes including 4-oxo-2-nonenal and the dialdehydes form di-DNPH derivatives. Derivatisation studies revealed the ratios and reaction times for the corresponding mono- and di- DNPH products of targets with two carbonyls; this is presented in Section 2.3.8. These experiments led to a complete list of preferred parent ions for fragmentation optimisation (see Table 6). Daughter ion scans were performed using individual 10mcg/mL stock solution of each selected target, and the most intense daughter ions were selected for manual fragmentation optimisation. For this, a combined 10mcg/mL stock solution of all targets was analysed at the preferred parent ion mass and most intense daughters, with incremental increases in collision energy (CE) and then cone voltage (CV). On the basis of these experiments, the optimised transition masses and energies were selected (Table 6).

Mono-derivatised DNPH hydrazones gave 163, 152, and 151 daughter ions, with the 163 being typically the most intense (these are fragments of the DNPH, hence their commonality between different aldehydes, see Figure 6)(175,189). It was seen in parallel chromatography experiments (see later) that ketones and alkanals isomers co-elute, and therefore disambiguation on fragmentation pattern was required. Comparative analysis of a ketone mix and an alkanal mix revealed that co-eluting aldehyde/ketone isomers give similar intensities of 152 and 151 ions. However, ketones do not give significant 163 fragments at any carbon number (see Table 7)(175). The 163 ion is also the most intense aldehyde fragment, and so is suitable for unambiguous alkanal quantification in the presence of ketones. In the absence of a 163 peak, either the 152 or 151 daughters can be used for ketone quantification. In mixed samples, the ketone gives an early peak in the 152 transition, which can be used for unambiguous quantification (albeit with relatively high limit of quantification as the intensity of this minor peak is low, see Figure 7). The  $m/z$  182 daughter ion was the most intense for all bifunctional aldehydes.

**Table 6: Monitored ion transitions and fragmentation energies**

Compound	MW	Hydrazone-MW	Mode	Parent ion	Quantitative daughter ion			Other daughter(s)
				[M-H] <sup>+</sup>	Daughter MW	CE	CV	
<b>Aldehydes</b>								
Formaldehyde	30.0	210.0	Neg	209.0	163	10	10	151, 178
Acetaldehyde	44.1	224.0	Neg	223.0	163	20	10	152, 151
Propanal	58.1	238.1	Neg	237.1	163	20	10	152, 151
Butanal	72.1	252.1	Neg	251.1	163	20	10	152, 151
Pentanal	86.1	266.1	Neg	265.1	163	10	10	152, 151
Hexanal	100.2	280.1	Neg	279.1	163	15	10	152, 151
Heptanal	114.2	294.2	Neg	293.2	163	20	12	152, 151
Octanal	128.2	308.2	Neg	307.2	163	15	12	152, 151
Nonanal	142.2	322.2	Neg	321.2	163	15	14	152, 151
Decanal	156.3	336.2	Neg	335.2	163	15	20	152, 151
Undecanal	170.3	350.3	Neg	349.3	163	15	15	152, 151
Dodecanal	184.3	364.3	Neg	363.3	163	20	15	152, 151
Tridecanal	198.3	378.3	Neg	377.3	163	20	22	152, 151
Tetradecanal	212.4	392.3	Neg	391.3	163	20	25	152, 151
Pentadecanal	226.4	406.4	Neg	405.4	163	20	25	152, 151
Hexadecanal	240.4	420.4	Neg	419.4	163	20	25	152, 151
Heptadecanal	<b>254.5</b>	434.4	Neg	433.4	163	20	25	152, 151
Octadecanal	268.5	448.5	Neg	447.5	163	20	25	152, 151
Benzaldehyde	106.1	286.1	Neg	285.1	163	15	15	152, 151
Cinnamaldehyde	132.2	312.1	Neg	311.1	163	15	20	152, 151
Acrolein	56.1	236.0	Neg	235.0	163	15	12	152, 151
Crotonaldehyde	70.1	250.1	Neg	249.1	163	20	10	152, 151
Trans-2-pentenal	84.1	264.1	Neg	263.1	163	15	12	152, 151
Trans-2-hexenal	98.1	278.1	Neg	277.1	163	15	15	152, 151
Trans-2-heptenal	112.2	292.1	Neg	291.1	163	10	10	152, 151
Trans-2-octenal	126.2	306.2	Neg	305.2	163	15	12	152, 151
Trans-2-nonenal	140.2	320.2	Neg	319.2	163	15	14	152, 151
2,4-nonadienal	138.2	318.2	Neg	317.2	163	15	15	152, 151
2,4-decadienal	152.2	332.2	Neg	331.2	163	20	20	152, 151
4-Hydroxy-2-nonenal	156.2	336.2	Neg	335.2	163	15	12	152, 151
4-oxo-2-nonenal*	154.2	514.7	Neg	513.7	182	15	30	167, 317
Malondialdehyde*	72.1	234.1	Pos	235.1	159	15	20	143, 188
Glyoxal*	58.0	418.6	Neg	417.6	182	15	30	234
Methylglyoxal*	72.1	432.2	Neg	431.2	182	10	10	122
Glutaraldehyde*	100.1	460.0	Neg	459.0	182	15	20	163
<b>Ketones</b>								
Acetone	58.1	238.1	Neg	237.1	152	15	15	151
2-Butanone	72.1	252.1	Neg	251.1	152	15	15	151
2-Pentanone	86.1	266.1	Neg	265.1	152	15	15	151
2-Hexanone	100.2	280.1	Neg	279.1	152	15	15	151
2-Heptanone	114.2	294.2	Neg	293.2	152	15	15	151
2-Octanone	128.2	308.2	Neg	307.2	152	15	20	151
2-Nonanone	142.2	322.2	Neg	321.2	152	15	20	151
2-Decanone	156.3	336.2	Neg	335.2	152	20	20	151
<b>Internal Standards</b>								
Malondialdehyde-d2	74.1	236.1	Pos	237.1	161.1	15	20	190.9
Acetaldehyde-d4	48.1	228.0	Neg	227.0	163	20	10	152, 151
Hexanal-d12	112.2	292.1	Neg	291.1	163	20	10	152, 151
Hexadecanal-d5	245.4	425.4	Neg	424.4	163	20	25	152, 151
Hydroxynonenal-d3	159.2	339.2	Neg	338.2	163	15	12	152, 151
Oxononenal-d3	157.2	517.7	Neg	516.7	182	15	30	167, 317

\*These dialdehydes or oxo-aldehydes form di-DNPH derivatives. \*\*Malondialdehyde's unique conformation allows the formation of a unique close-ring derivative



**Figure 6: Major fragments of mono-DNP-hydrazones.**

Mono-DNP-hydrazones (1) form characteristic daughter ions at  $m/z$  163 (2) and 152 (3), both of which entirely originate from the DNP moiety. Thus, these daughters are constant for all chain lengths.

**Table 7: Differentiation of ketones and alkanals on fragmentation pattern**

Carbon	m/z 163			m/z 152		
	Alkanal	Ketone	k:a ratio (%)	Alkanal	Ketone	k:a ratio (%)
3	5835	46	0.8	1711	3629	212.1
4	5929	79	1.3	5909	870	14.7
5	6880	3	0.0	8135	1072	13.2
6	6574	23	0.3	5673	1940	34.2
7	5748	21	0.4	4507	1591	35.3
8	5334	2	0.0	6357	2483	39.1
9	6708	0	0.0	5845	3178	54.4
10	3127	6	0.2	3425	1889	55.2

These data were used to populate a “scheduled multiple reaction monitoring” method for the detector. The MRM concept greatly enhances signal-to-noise ratios by fixing the detector at specific  $m/z$  values, rather than scanning across the mass range. This increases the dwell time at the target mass, which in turn increases the evidence by which the detector populates a peak. Typically the detector can cope with a modest number of different transitions, and thus the MRM method is scheduled to match the detectors’ settings to the targets which are eluting at a given point in the method.

Towards the end of method development, detector dwell times were further enhanced by rationalising the final aldehyde MS/MS method to two (alkanals) or three (others) transitions. Ketones were also removed from the method after it was clearly possible to distinguish them from alkanals on fragment ions (see below). Including internal standards this led to a total of 156 monitored transitions, which were split into groups of 10-20, scheduled according to column elution.

### **2.3.7 Chromatography development**

For initial experiments, the stationary phase (C18) and mobile phases (A, aqueous phase: water; B, organic phase: acetonitrile) were selected on the available literature (153,169,176–179). A starting ratio of 70% A to 30% B was selected as this gave good retention and provides optimal ionisation of the more polar hydrazones(176). The Waters ultra-performance chromatographic system is designed to operate at high pressures using bonded-phase columns with small particle sizes. This means high flow rates across the column can be used, which decreases the time for beginning to end of elution, leading to narrow peak width and higher intensities in a shorter overall run. The C18 column selected for this study had a particularly low particle size (1.6 $\mu$ m, “Cortecs” Waters brand), and this study was the first to apply this technology to aldehyde quantitation. Mobile phase modification by pH or volatile buffer did not significantly improve signal intensity in the MRM method, and therefore the mobile phases were left unchanged.

A critical chromatography goal is to separate target isomers/isobars. If they co-elute, the detector will not distinguish them on parent  $m/z$ , and may not be able to distinguish them on daughter  $m/z$ . Thus, the mobile phase gradient was adjusted until separation of target isomers/isobars was achieved (see Figure 7):

(i) Acetone, propanal, glyoxal (MW = 58): Mono-DNP-hydrazones (m/z 238) of acetone and propanal separated by 1 minute. The di-DNPH derivative of glyoxal eluted much later and has a different parent (m/z 418). The minor mono-DNPH derivative of glyoxal (m/z 238) eluted at 1.44 minutes and was thus well separated.

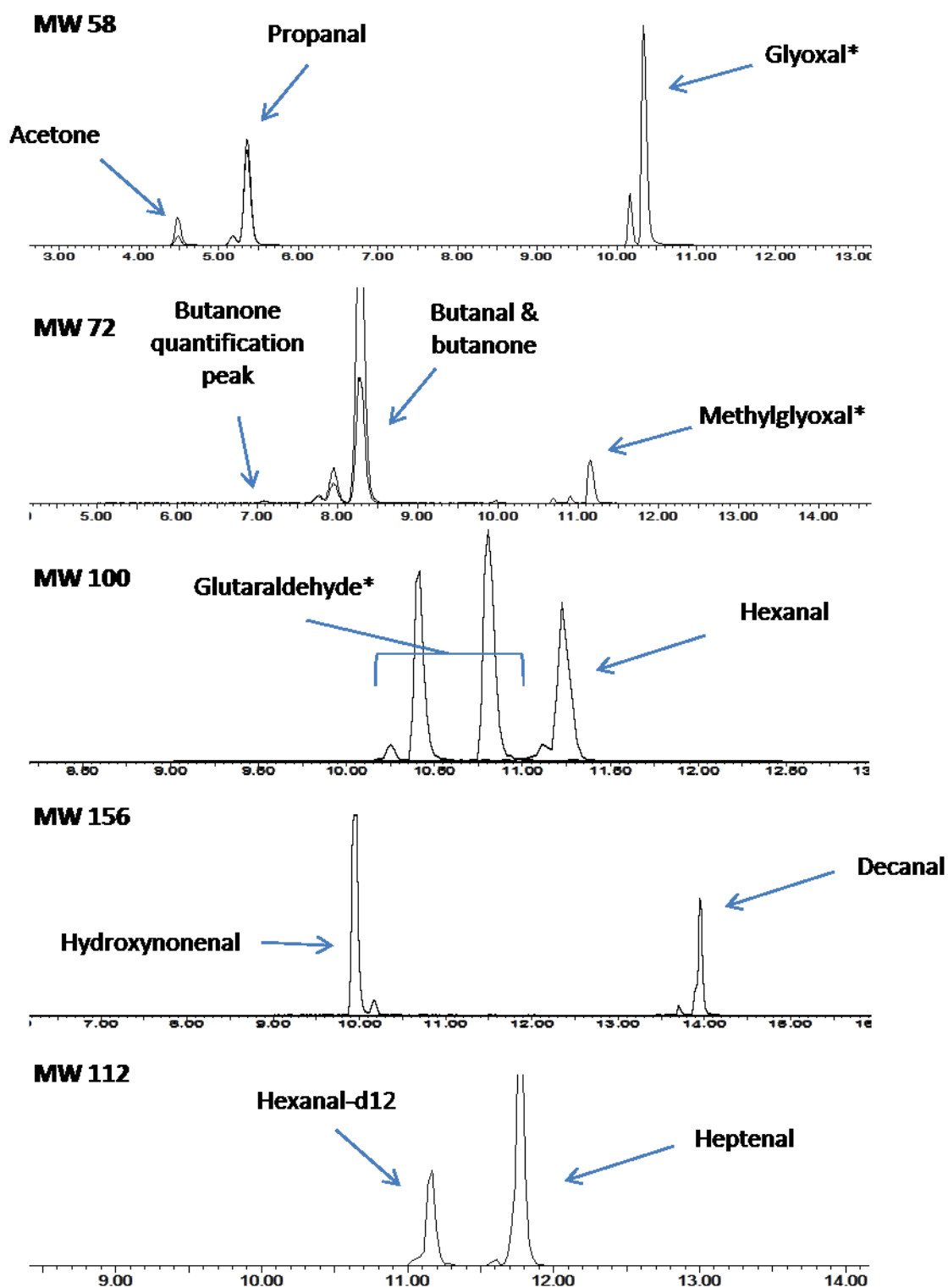
(ii) Butanone, butanal, methylglyoxal (MW = 72): Mono-DNP-hydrazones (m/z 252) could not be separated by chromatography (co-eluting at 8 minutes), but can be distinguished by aldehyde/ketone fragmentation pattern (see 2.2.6 and Table 2.2.4 as for all heavier aldehyde/ketone pairs). Methylglyoxal gave an abundant di-DNPH derivative (m/z 432) eluting 3 minutes after butanal and butanone. It also gave a less abundant mono-DNPH derivative (m/z 252), which eluted at 1.71 and 2.5 minutes, thus separated from the isomer DNP-hydrazones by 6 minutes.

(iii) Glutaraldehyde and hexanal (MW = 100): separated by 20 seconds, and also because glutaraldehyde formed di-DNPH-derivative

(iv) Hydroxynonenal and decanal (MW = 156) separated by 3 minutes

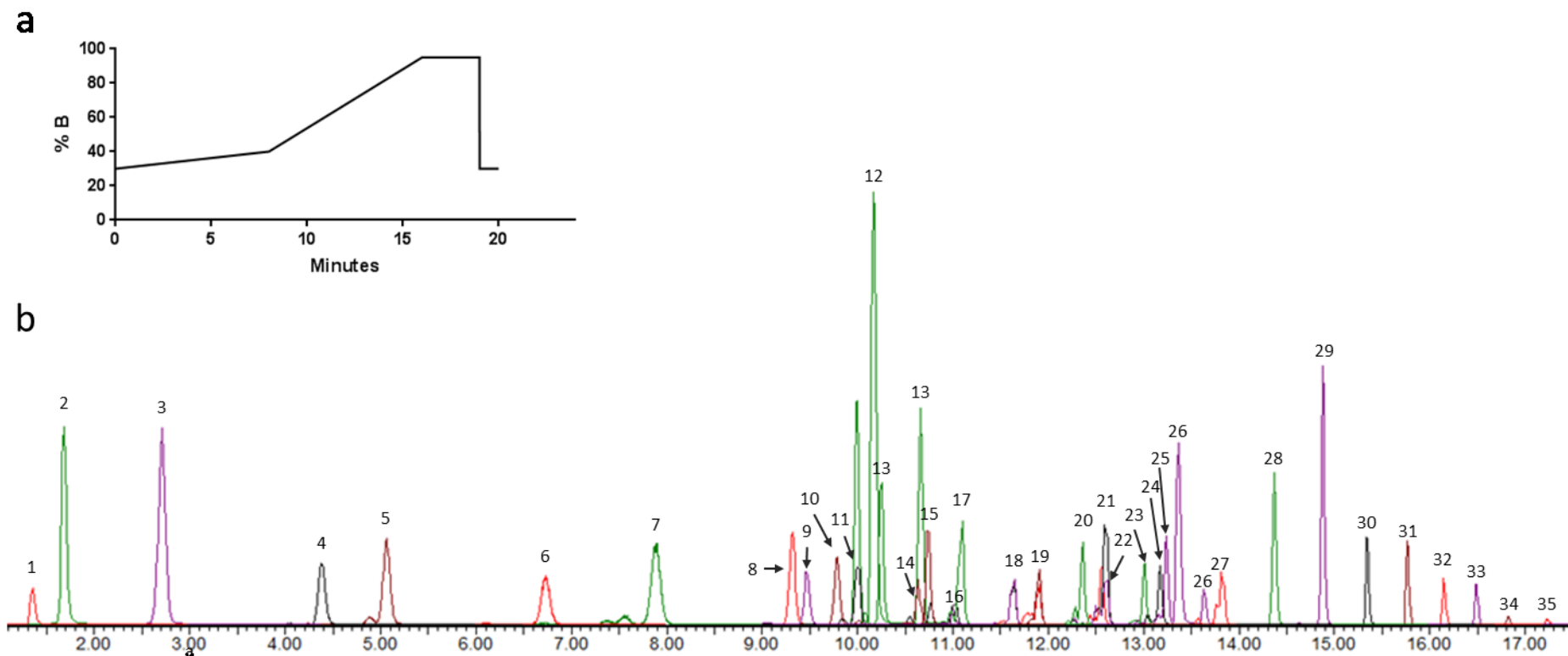
(v) Hexanal-d12 and trans-2-heptenal (MW = 112) separated by 30 seconds.

Thus, all isomer/isobar target compounds and ISTDs could be unambiguously determined. The finished chromatographic method is illustrated in Figure 8.



**Figure 7: Total ion chromatograms of separated DNP-hydrazone isomers/isobars.**

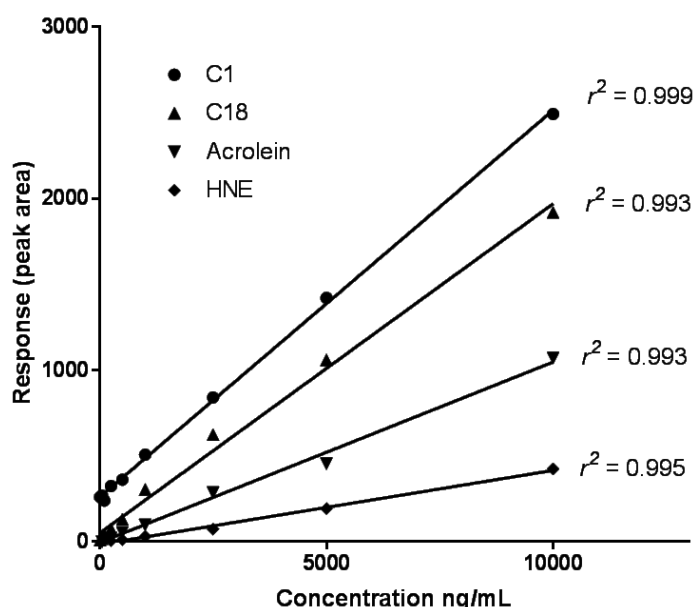
All of the pre-determined isomers groups could be separated by chromatography except aldehydes/ketones pairs above C3. However, these could be distinguished by their characteristic daughter ions (see 2.2.6). \*Indicates di-DNPH derivative – these were therefore separated by chromatography *and* by parent/daughter *m/z*.



**Figure 8: Final chromatographic method**

Panel (a) Schematic of mobile phase gradient. Panel (b) Chromatogram of the 35 aldehydes in the final method (ketones removed for clarity; external standards at 100 ng/mL, ESI ion source in negative mode except MDA). 1 Malondialdehyde; 2 Acetaldehyde; 3 Formaldehyde; 4 Acrolein; 5 Propanal; 6 Crotonaldehyde; 7 Butanal; 8 Benzaldehyde; 9 Pentenal; 10 Hydroxynonenal; 11 Pentanal; 12 Glyoxal; 13 Glutaraldehyde; 14 Cinnamaldehyde; 15 Hexenal; 16 Methylglyoxal; 17 Hexanal; 18 Heptenal; 19 Heptanal; 20 Octenal; 21 Octanal; 22 Nonadienal; 23 Nonenal; 24 Decadienal; 25 Nonanal; 26 Oxononenal; 27 Decanal; 28 Undecanal; 29 Dodecanal; 30 Tridecanal; 31 Tetradecanal; 32 Pentadecanal; 33 Hexadecanal; 34 Heptadecanal; 35 Octadecanal.

Each target compound could now be confidently quantified from a mixture of target compounds, dissolved in neat mobile phase. Measuring range was tested next. Calibration curves were fitted, plotting response against serial dilutions of standards. This allows the “linear range” of the detector to be determined (i.e. the concentration range at which a given response of the unknown sample is likely to yield an accurate concentration), in addition to a preliminary understanding of the method sensitivity. The lower limit of quantification (LLOQ) was defined for each target compound as the lowest calibration point in which the analytical precision (%RSD) was <5%. Linear regression can then be used to provide line-of-best fit (see Figure 9), using the  $r^2$  coefficient to describe how well model describes the measured calibration points (>0.99 is acceptable, see Table 8).



**Figure 9: Example calibration curves (electrospray ionisation results provided)**

There is considerable debate regarding the best ionisation source for aldehyde quantitation in biological matrices (169,177–179). The available choices for this project were atmospheric pressure chemical ionisation (APCI) and electrospray ionisation (ESI). Aldehydes’ range of polarity and structural similarity to lipids renders them susceptible to ion suppression when extracted from complex samples. This can happen in both APCI and ESI sources, although the ESI source may be more susceptible (190,191). One paper has discussed their use for aldehyde quantitation



in the setting of the Waters UPLC system (191), and this found the ESI sources to be more stable and sensitive. Both were examined for this method at the point of calibration testing, and ESI was found to be more precise, stable, and with better limit of detection across the calibration range (see Table 8). Potential matrix effects arising from ionisation suppression due to co-eluting interferences were examined in the next section.

**Table 8: Comparison of calibration properties of ESI and APCI ion sources**

Hydrazones of	t <sub>R</sub>	APCI			ESI			
		r <sup>2</sup>	LLOQ*	%RSD	r <sup>2</sup>	LLOQ*	%RSD	Linear range*
Formaldehyde	1.72	0.9950	5.0	12.5	0.9989	0.10	4.7	0.1 - 100
Acetaldehyde	2.79	0.9826	1.0	10.8	0.9984	1*	2.7	1* - 100
Propanal	5.21	0.9597	1.0	8.1	0.9993	0.10	6.2	0.1 - 100
Butanal	8.1	0.9618	5.0	11.2	0.9994	0.10	4.1	0.1 - 100
Pentanal	10.09	0.9976	5.0	17.3	0.9992	0.10	1.8	0.1 - 100
Hexanal	11.11	0.9510	1.0	23.7	0.9964	0.10	5.4	0.1 - 100
Heptanal	11.98	0.9840	5.0	21.1	0.9804	0.10	5.9	0.1 - 100
Octanal	12.68	0.9895	2.5	15.4	0.9950	0.05	0.9	0.05 - 100
Nonanal	13.31	0.9861	2.5	14.7	0.9980	0.05	1.8	0.05 - 100
Decanal	13.91	0.9943	1.5	11.2	0.9592	0.05	7.0	0.05 - 100
Undecanal	14.46	0.9912	8.0	40.7	0.9967	0.05	3.8	0.05 - 100
Dodecanal	14.96	0.9941	5.0	11.8	0.9962	0.05	7.5	0.05 - 200
Tridecanal	15.43	0.9944	2.5	26.4	0.9973	0.05	0.6	0.05 - 200
Tetradecanal	15.85	0.9886	10.0	10.5	0.9931	0.05	3.1	0.05 - 500
Pentadecanal	16.22	0.9729	10.0	17.2	0.9953	0.05	2.7	0.05 - 500
Hexadecanal	16.57	0.9822	10.0	21.3	0.9954	0.10	1.8	0.1 - 500
Heptadecanal	16.93	0.9533	2.5	31.9	0.9978	0.10	3.7	0.1 - 500
Octadecanal	17.37	0.9797	1.9	26.4	0.9939	0.10	5.3	0.1 - 500
Benzaldehyde	9.41	0.9970	1.0	20.7	0.9978	0.10	3.1	0.1 - 100
Cinnamaldehyde	10.72	0.9742	10.0	26.2	0.9924	0.50	5.5	0.5 - 100
Acrolein	4.51	0.9958	1.0	7.2	0.9934	0.05	5.3	0.05 - 100
Trans-2-pentenal	9.56	0.9758	2.0	15.6	0.9913	0.25	5.8	0.25 - 100
Trans-2-hexenal	10.82	0.9801	1.0	11.1	0.9875	0.50	5.5	0.5 - 100
Trans-2-heptenal	11.72	0.9689	1.0	22.3	0.9824	0.25	4.9	0.25 - 100
Trans-2-octenal	12.45	0.9608	2.5	15.5	0.9399	0.25	5.6	0.25 - 100
Trans-2-nonenal	13.08	0.9935	5.0	9.9	0.9875	0.50	0.6	0.5 - 100
2,4-nonadienal	12.64	0.9855	2.5	11.6	0.9920	0.10	9.0	0.1 - 100
2,4-decadienal	13.26	0.9926	2.5	17.4	0.9968	0.50	1.2	0.5 - 100
Crotonaldehyde	6.86	0.9840	2.5	10.6	0.9976	0.05	4.1	0.05 - 100
4-Hydroxy-2-nonenal	9.86	0.9974	5.0	24.6	0.9950	0.25	0.1	0.25 - 100
Malondialdehyde	1.37	0.9842	<0.1	24.1	0.0574	5.00	7.2	5 - 100
Glyoxal	10.25	0.8158	2.5	27.6	0.9594	1*	6.3	1* - 100
Methylglyoxal	11.09	0.9581	2.5	33.1	0.9732	0.10	2.7	0.1 - 100
Glutaraldehyde	10.73	0.9937	0.5	16.3	0.9996	0.05	4.3	0.05 - 100
4-oxo-2-nonenal	13.45	0.9841	1.0	21.5	0.9897	0.50	2.7	0.5 - 100

\*Lower limit of quantification (LLOQs, see text) and linear range given in ng/mL. C2 and glyoxal were detectable at much lower values, although background typically precluded quantitation below this value. t<sub>R</sub> = retention time (minutes); %RSD, relative standard deviation, n = 5 at the LLOQ. APCI, atmospheric pressure chemical ionisation; ESI, electrospray ionisation. All samples prepared in 50:50 acetonitrile/water mix with no pH correction, 1 hour derivatisation with standard DNPH protocol.

### 2.3.8 Injector settings

The highly non-polar nature of DNP-hydrazones may cause a problem of carry-over between injections, usually because of sampling needle contamination in the injector system. Thus, mitigation of carry-over is a key item for aldehyde method development. Using the mobile phases as the wash buffers, carryover with DNP-hydrazones of fatty aldehydes was >20% between samples. Thus the wash cycles for the needle were manipulated to improve carry-over. It was eventually found that changing the washes to 2 mL of 100% isopropanol followed by 4mL of 100% acetonitrile improved the carry-over to acceptable limits <2% (Table 9), without contaminating the injector with isopropanol.

Different sample injection sizes were trialled (1, 2, 5, 7.5, 10, and 20 $\mu$ l). There was evidence of column overload after 7.5 $\mu$ l, and thus 5 $\mu$ l injection volume was chosen. Samples were maintained at 4°C immediately prior to injection in the UPLC system. Stability experiments confirmed acceptable variation to >24 hours when maintained at this temperature (see Table 9). Storage for >24 hours was at -80°C.

### 2.3.9 Sample preparation development.

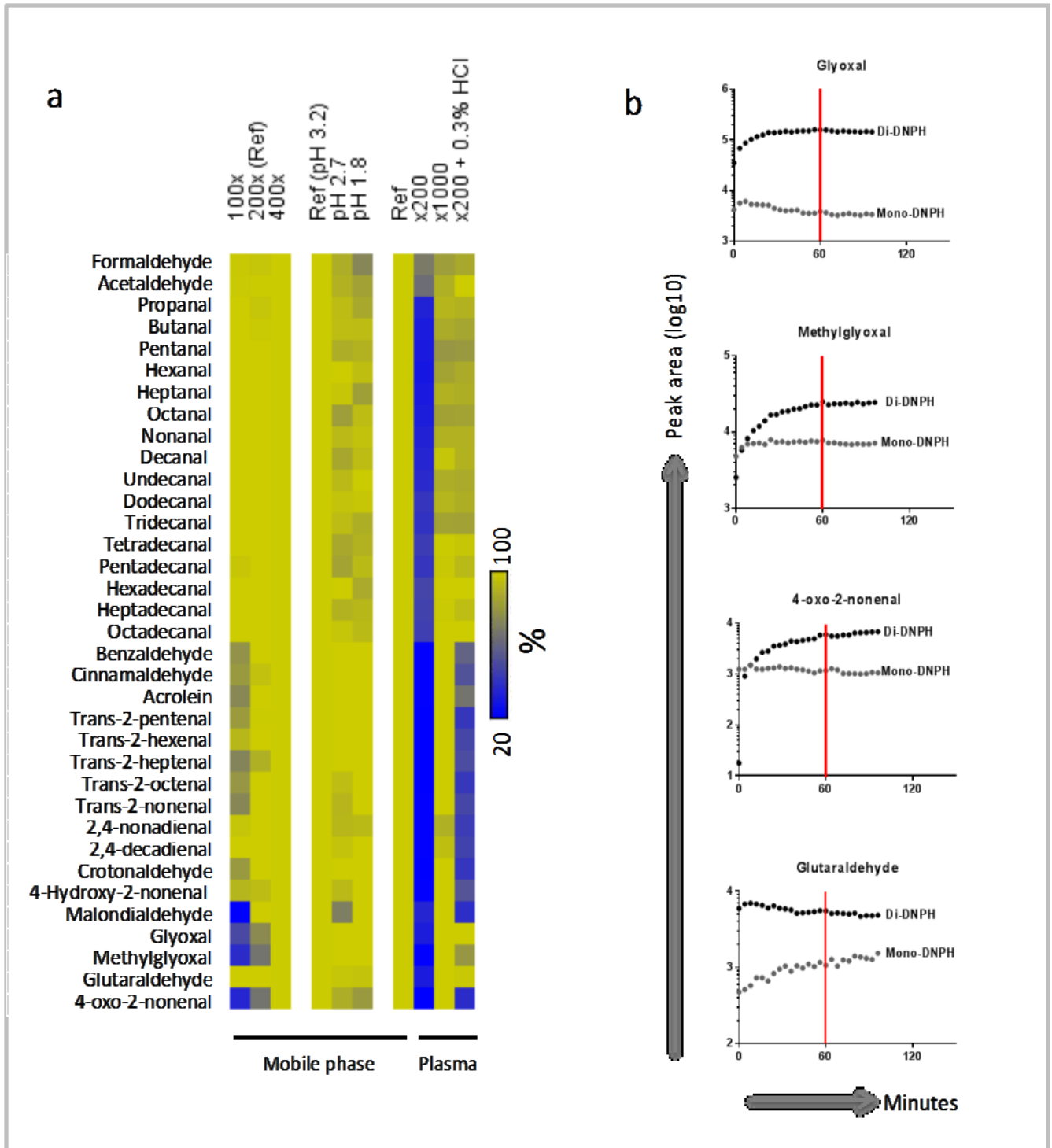
**Derivatization** DNPH derivatization was selected as it is convenient, stable, available, well validated and is the method of choice of major standards agencies (152,176,178,192,193). However, the requirement for molar excess of DNPH, low pH, reaction time, and subsequent extraction of the non-polar derivatives from the native sample introduces several opportunities for error, and thus these variables must be individually optimised to maximise recovery and combat inconsistency from partial derivatization.

The first step of DNPH/aldehyde reaction is protonation of the hydrazine side-group, and therefore acidic conditions are required for the reaction (generally a pH of 2-3, even as high as 5 (145)). Solid pure DNPH is explosive, and so it is supplied as a 0.2M solution in 70% phosphoric acid, (11.6M, pH 0.14). At least a x100 molar excess of DNPH is required for complete derivatization (145,176,178,189,193). Using 100 ng/mL spiking experiments (thought to be toward the upper end of natural biological concentrations), it was shown that a x200 molar excess of DNPH in 11.6mM phosphoric acid in 50:50 aqueous acetonitrile yielded similar intensities to

x400/23.2mM conditions, in keeping with previous reports (see Figure 10a)(145,189). Further pH modification of the reaction mix with 0.1% or 1% formic acid (measured pH 2.7 and 1.8 respectively) mildly impaired responses. The x200 molar excess without further pH modification was selected for further optimisation studies (measured pH 3.2).

These results were taken forward to biofluids. Using plasma samples in which the protein had been removed with an equal volume of acetonitrile, the measured intensities of a 100 ng/mL spike dropped to <10% for nearly all targets (see Figure 10a). Biofluids are robust buffers and may be rich in unknown carbonyl moieties, and thus both insufficient acidity and insufficient DNPH could both contribute to poor recoveries. To estimate their relative effects on the reaction, the same reaction was supplemented with 0.3% HCl, which facilitated improved recoveries of most targets by 30-100%. However, supplementing the reaction with a greater amount of DNPH – to x1000 molar excess (116mM phosphoric acid, measured pH 2.1) - generally improved recoveries to 80-120% of the reference. Additionally in the HCl supplemented replicates, there was a 170% recovery of acetaldehyde, suggesting adventitious liberation of the target by acid hydrolysis of the sample or column, as suggested by previous reports (145,194). On the basis of these experiments, the x1000 molar excess conditions were selected for derivatization of biosamples (it should be noted that the data presented here are repeats of the initial derivatization experiments, following optimisations presented below regarding reaction time, acetonitrile deproteination, and clean-up with liquid-liquid extraction).

Special attention was given to the rate kinetics of dialdehyde-hydrazones and MDA-DNPH formation under test conditions, by serially quantifying concentrations every four minutes from the initiation of the reaction. It was found that peak and stable intensities were reached after 1 hour derivatization at ambient temperature (see Figure 10b).



**Figure 10: Derivatisation development.**

Panel a, Heatmaps of response for DNPH reactions with 100ng/mL external standard mixes in different solvents, pH, and molar excess of DNPH. Columns 1-3: Molar excess of DNPH (x100, x200, x400); 4-6: Different pH in mobile phase Columns 7-10 Optimisation in plasma

**Background** Formaldehyde, acetaldehyde, glyoxal and methylglyoxal are ubiquitous environmental aldehydes, leading to high background and potentially masking natural variation in biosamples (145,152,153). This is illustrated in Figure 11a, where the concentrations of the aldehydes were measured after simply adding 40ul DNPH to 960ul week-old acetonitrile. This represents contamination of DNPH, glassware and/or solvent with lab air aldehydes, the effect will be additive each time the bottle is opened. The background concentrations were higher still in a DNPH-blank sample containing OCT, a common sectioning media for frozen tissue samples, and in a DNPH-blank sample also containing HPLC water and sodium chloride (see Figure 11a). Thus, steps to reduce this background as fully as possible were undertaken, including strict adherence to EPA recommendations for clean glassware (152), hexane extraction of the DNPH stock solution, which is made in a large batch using fresh solvents and tested before use (145,152), and finally baking sodium chloride, ceramic, plastic and all glassware in a 70°C oven for at least three hours prior to use. This led to the final background concentration of <1ng/mL for acetaldehyde, glyoxal and methylglyoxal (“Best salt”); the normal range for these metabolites in human biofluids was 2 to 25 ng/mL, and thus considered acceptable background. The background level of all other targets was either undetectable or <10% of the normal lower limit in natural samples.

**Clean-up** The purpose of sample clean-up is to preserve the natural targets while removing interferences (195,196). GC-MS systems are vulnerable to interference contamination and thus multiple clean-up techniques have been described, including headspace analysis, solid phase micro-extraction and single-drop microextraction sample clean-up have been reported (165,197,198). LC-MS systems are more robust to contamination, but ESI and to a lesser extent APCI ionisation efficiency can be sensitive to co-eluting non-target compounds (“matrix effects”). Liquid-liquid extraction (LLE) is a relatively simple clean-up technique which takes advantage of a differing solubility in immiscible solvents to separate solutes, and has been widely used for DNP-hydrazone clean-up as it is simple, cheap and effective (169,175). Solid phase extraction is sometimes favoured as it suits the non-polar nature of DNP-hydrazones and several groups have reported excellent clean-up and pre-concentration (153,169,179). However, it is more involved than LLE, and is thus less suited to high-throughput sample preparation, a pre-determined goal of this method, and thus LLE was selected.

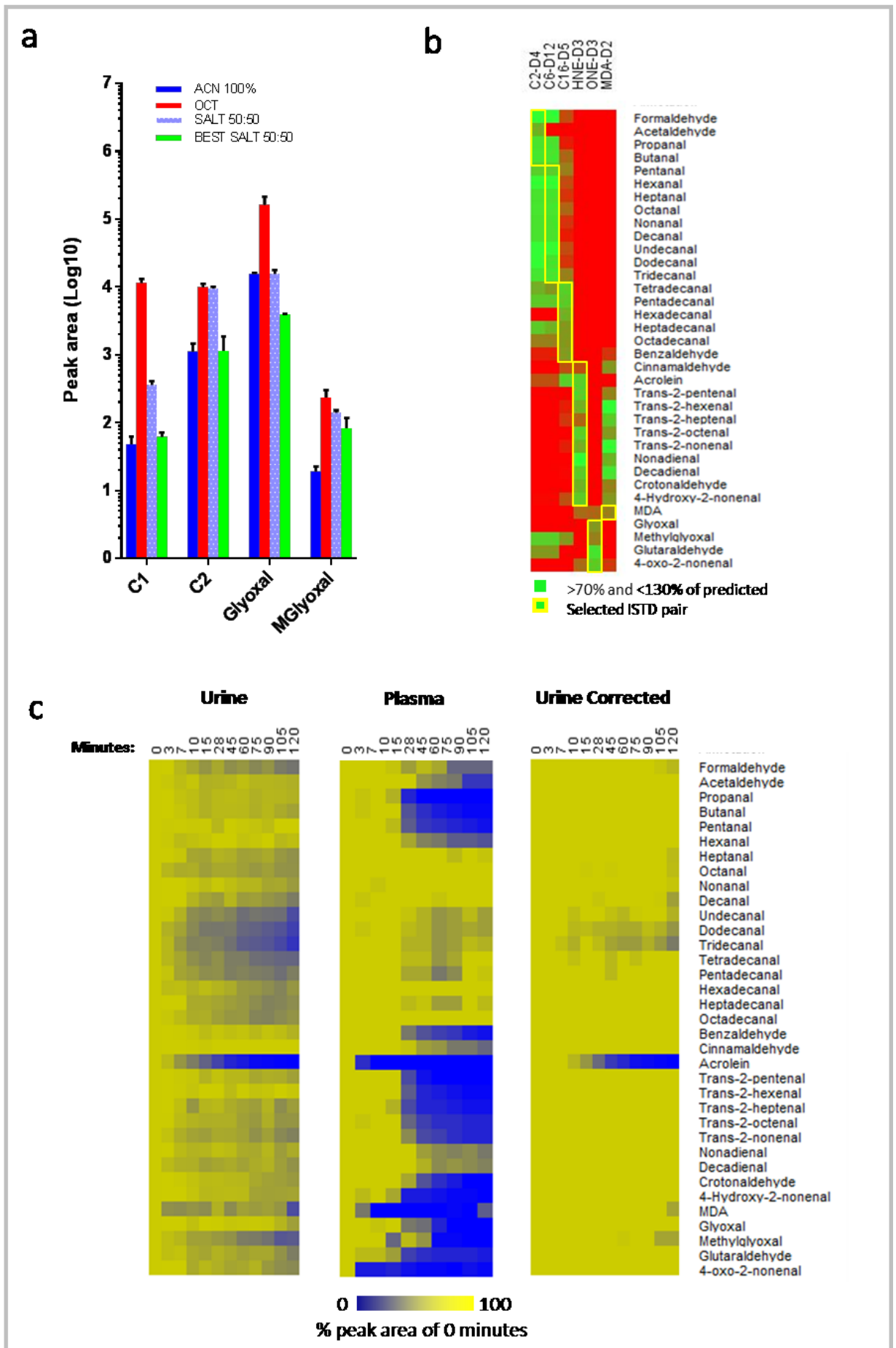
In measuring aldehydes in biosamples, proteins are a particular problem as they can sequester or even metabolise aldehydes, in addition to direct covalent reaction (see *Loss Experiments*, below)(154). DNP-hydrazones could also be sequestered to hydrophobic protein regions. Thus, several groups report sample deproteination prior to derivatisation (153,169,193). Published alternatives for deproteination include ultrafiltration or precipitation with methanol, perchloric acid or acetonitrile (145,153,199). Ultrafiltration was not favoured as the spin columns are expensive and single use. Methanol and perchloric acid were also discounted as they may cause adventitious formaldehyde production (152,200).

Conveniently, after protein precipitation, acetonitrile becomes immiscible in water if saturated with sodium chloride. It was established that a 50:50 mix of acetonitrile to biofluid was sufficient to fully deproteinate all samples (including plasma, the most proteinaceous biosample tested), by vortexing the mixture for 10 seconds, and then centrifuging at 14,500 rpm for 3 minutes at 4°C. The supernatant was then removed to a fresh vial, derivatised by adding DNPH as above, and then phase separated (LLE) by saturating with sodium chloride (100mg / 1mL total volume). The organic phase containing the extracted DNP-hydrazones was then injected into the UPLC-MS/MS system.

The efficiency of this process (DNP-hydrazone extraction) was checked by spiking plasma samples with known concentrations of pre-derivatised external standards dissolved in acetonitrile; these results are given in the Validation section (2.3.12)

**Figure 11: Sample preparation development (next page).**

Panel a, Initial background levels of selected aldehyde in the indicated solutions (1mL) with or without sodium chloride (100mg). The optimised background is presented, following all precautions to minimise background contamination ('best salt'). Panel b, Internal standard pairing by accuracy of concentration correction to predicted in plasma ( $n = 5$ ). Yellow highlight indicates the best ISTD-target pair used for phenotyping experiments. Panel c, Rate of loss (minutes, at top) of 100ng/mL aldehyde spike in urine, before (right) and after correction with internal standard (middle). Also provided is urine loss corrected by loss of paired internal standard (right). Each compound is normalised to the measured concentration at 0 min. ACN, acetonitrile; OCT – optimal cutting temperature compound; MGlyoxal, methylglyoxal; other aldehyde abbreviations as previously.



**Internal standard correction** Despite extensive efforts to control variables and simplify method elements, sample preparation variation is inevitable, and necessitates internal control procedures. Stable isotope dilution is a powerful means of controlling variation introduced by sample processing, with a 75-year heritage (195,196,201). This involves spiking a known concentration of target compound labelled with a stable isotope (herein required to as an 'internal standard' – ISTD). The ISTD should then be exposed to the same variation as the native target, react identically, and yet still be distinguishable in the mass spectrometer. Hence ISTD loss should be equal with target compounds losses, and therefore individual sample variation can be accurately corrected. Sample processing error is often correctable by a single ISTD for all targets. However, ionisation mechanism interferences ("matrix effects") may disproportionately affect targets, and be different between samples. If clean-up is inadequate and variable matrix effects are active, then ISTDs for each target compound is required to ensure accuracy. However, this requires pragmatism, as isotope-labelled standards are expensive and may not be available.

The following internal standards were selected for initial testing: (i) acetaldehyde-deuterium 4 (C2-d4) for polar alkanals (ii) hexanal-d12 for medium chain alkanals (iii) hexadecanal-d5 for fatty alkanals (iv) hydroxynonenal-d3 for enals and co-eluting aromatics (v) ONE-d3 for dialdehydes (vi) MDA-d2 for malondialdehyde as ionisation and derivatization was unique to this compound. To test the performance of these ISTDs in correcting analyte loss during sample preparation, triplicate urine samples were spiked with external standards (final concentration 100ng/mL) and ISTDs (50ng/mL), and then immediately derivatised using the optimal conditions presented previously (the following data is the result of a late experiment which also used the optimised deproteination and LLE technique described in subsequent sections).

The measured concentrations of the external standards were then corrected by each ISTD, and compared to the measured concentration of a 100ng/mL calibration point made in acetonitrile, and the accuracies were calculated as:  $((\text{measured concentration in spiked} - \text{measured concentration in blank}) / \text{expected concentration}) \times 100$ . The results are presented in Figure 11b. Accurate correction (80-120% expected (196) was observed across the intended pairing, and so a 5 mcg/mL ISTD 100x stock mix was made in volume and divided into multiple single use aliquots. Although lower concentrations of ISTD would still be easily detectable and more economical, this higher concentration was selected to support accurate analysis in the event of a significant fall in recovery.



**Loss studies** Literature accounts of the instability of free aldehydes in biospecimens are relatively sparse (e.g. Tomono et al (169)). Therefore it was decided to test this for the pre-determined target compounds in two biofluids of different protein content. Thirty millilitre samples of urine and plasma (fetal bovine serum in this case (F9665, Sigma)) were spiked to a starting concentration of 100ng/mL of the target compounds. The ISTDs were also added to the same concentration, and the reaction proceeded on ice (to reduce volatile losses). Triplicate samples were then prepared at regular time intervals using the optimised derivatisation and extraction conditions described above and injected into the LC-MS system.

Both biofluid types exhibited analyte loss (see Figure 11c), but aldehydes were generally less stable in plasma than in urine. Acrolein in particular was undetectable in plasma after 7 minutes, and after 45 minutes in urine. Dialdehydes and bifunctionals were also rapidly lost, again in plasma more than urine. Interestingly alkanals C11-C15 seemed to be mildly unstable in urine but not plasma; this was verified in a repeat of this experiment. The efficacy of the ISTDs in correcting these losses was assessed (see Figure 11c, right panel). Accuracies were between 80-105% for all targets and time-points, except acrolein after 10 minutes. This was because the spiked acrolein disappeared below the limit of quantification after 10 minutes, and thus the ISTD correction could not be applied. This exemplifies the importance of combining ISTD correction with at least reasonable analyte recovery.

These data suggest that free aldehydes' stability in a biofluid is inversely proportional to the protein content. Thus, perfect recovery of *in situ* concentrations from fluids such as plasma is likely to be very difficult, especially with the requirement to remove red cells before derivatisation. This should not preclude aldehydes' measurement altogether, in particular the less labile species, although it underlines the importance of taking samples on ice, working quickly, and adding ISTDs, deproteinating and derivatising the as soon as practically possible, with great care in handling all samples uniformly from the instance of biological disconnection.

**Sample collection** With these principles in mind, sample collection was considered for each biospecimens-type. For urine samples, mid-catch urine was aliquotted in 200ul samples and directly frozen at -80°C. One aliquot was sent for creatinine concentration (measured by contract with Imperial College Healthcare Biochemistry Department, under contract, and using clinical grade analytics), and another was

tested for the presence of infection (Siemens Multistix 10SG, Siemens) The samples were thawed on ice with regular vortexing and as soon as practically possible 200 $\mu$ l aliquots were mixed with 160 $\mu$ l acetonitrile containing 62.5 ng/mL ISTDs, and then deproteinated and derivatised as described. In the case of whole blood, rapid and complete removal of red blood cells is required before deproteination to remove erythrocyte reducing agents (i.e. haem) without disrupting the red cell membrane, and this is achieved by taking the sample in EDTA on ice, and centrifugation at 4000 rpm at 4°C for 3 minutes in line with international metabolomics standards (184,202). The plasma supernatant was then processed as for urine.

Quantifying aldehydes from tissue had slightly different requirements, although the principles of analyte preservation, deproteination and derivatization were similar. Tissue samples were snap frozen in liquid nitrogen within a few seconds of anatomical disconnection, and thus the aldehydes were less vulnerable to confounding reactions as in biofluids. However, these solid-phase analytes must be extracted to a liquid for measurement in the MS system, and this must be done quickly and completely without losing target compounds. Options of this include cryogenic grinding, rotor-stator homogenisation, or bead-beating. Following published recommendations for metabolomics analysis of tissue samples (183,203)), the final method involved grinding the frozen tissue in a ceramic pestle and mortar pre-chilled to -80°C, which was kept on dry ice reduced the tissue particle size to <1mm (usually <1 minute). The still-frozen tissue powder is then weighed on a microgram scale to allow for input normalisation of the final measured concentration, and then extracted in bead-beater (Mixer Mill MM 200, Retsch, Dusseldorf, Germany) for 60 seconds in a 50:50 water-acetonitrile solvent at 30 cycles per second, with a single 5mm stainless steel ball-bearing. The solvent and the tube rack were maintained at -20°C. Debris and protein was then pelleted by centrifugation at 14,500 rpm for 3 minutes and the sample was then derivatised and analysed as per biofluids. This method was favoured to avoid adventitious aldehyde formation through heating from excessive bead-beating.

After each use, the ceramic grinding equipment was scrubbed with Virkon, thoroughly rinsed with 18-ohm water and then UPLC water, baked at 70°C overnight, wrapped in foil and then stored at -80°C. Microdissection was not undertaken as the OCT mounting medium contained crotonaldehyde, acetaldehyde, butanone and some fatty aldehydes, and alternative methods were not possible without thawing the specimen (see Figure 11a).

### 2.3.10 Optimised sample preparation method

**Preparation of derivatization agent:** The derivatization stock solution was prepared by extracting 1ml of the 0.2M DNPH in 70% phosphoric acid stock solution in 5mL hexane and vigorously agitating for 10 minutes on an orbital shaker (300rpm). The phases were separated by centrifugation at 4000 rpm for 5 mins, and the extraction was repeated four times. One millilitre of extracted DNPH was diluted in 19 mL UPLC-grade acetonitrile from a fresh bottle. This diluted stock solution (about 10mM DNPH) was tested for background by injection into the UPLC-MS/MS system, and sealed as 1mL aliquots in glass vials with Teflon caps. One aliquot was used to check pH 1-1.2 with a pH meter.

**Preparation of internal standards** A mix of all six internal standards (concentration 5 mcg/mL) was prepared, aliquotted into Teflon sealed glass vials, stored at -80°C and used once. The final spiking concentration was 50 ng/mL.

**Biofluids:** Frozen liquid samples were thawed as rapidly as possible. As soon as possible aliquots of 200µl were mixed with 160µl ice-cold acetonitrile spiked with ISTDs to a final concentration of 62.5 ng/ml. The mixture was vortexed on a strong setting for 5 seconds, and centrifuged at 14,500 RPM in a pre-chilled centrifuge for 1 minute. The supernatant was removed to a fresh pre-chilled 1.5mL tube, and 40µl of the diluted DNPH was added from a fresh aliquot per experiment. The mixture was vortexed every 10 minutes for 1 hour at ambient temperature. Then, the mixture was saturated with sodium chloride (approximately 20mg, baked overnight at 70°C), vortexed, and centrifuged at 14,500 RPM for 5 minutes. At least 100µL of the yellow organic layer was moved to a polypropylene vial, sealed with a Teflon cap, and stored at -80°C until analysis by UPLC-ESI-MS/MS. All samples were prepared or analysed in a single bench effort.

**Tissue and murine xenograft samples:** Tissue samples were dissected on dry ice to 8 to 20mg (if necessary), and freeze-homogenised in a ceramic pestle-and-mortar, moved to a fresh pre-chilled 2mL tube and re-weighed. This second weight was used to correct the final measured concentration to biomass input. The powder was extracted in 180µl dry-ice-cold acetonitrile:water mixture containing ISTDs (62.5ng/mL, as above) for 1 minute in a Reitsch oscillator (MultiMix 200, 25 cycles/sec) with pre-chilled tube rack, and then centrifuged at 14,500 rpm for 1

minute in a rotor pre-chilled to -20°C. The supernatant was aspirated as fully as possible to a fresh 2ml tube without disturbing the pellet, and mixed with 40µl DNPH. The pellet was re-extract with a further 180µl, centrifuged and combined with the first supernatant. The derivatization reaction was allowed to proceed at ambient temperature for 1 hour, and then phase separated, stored and analysed as for biofluids. All samples were prepared or analysed in a single bench effort.

***In vitro samples:*** The specified time-point the cell media (extracellular aldehydes) was aspirated from the experiment well, centrifuged for 1 minute at 200 RPM to pellet debris and the supernatant was processed as for plasma and urine. For intracellular aldehydes, the adherent cells were quenched in 180µl dry-ice-cold 50:50 HPLC water:acetonitrile containing 50ng/mL ISTD and incubated on ice for 3 minutes. Without disturbing the cleared cell remnants, the extract was moved to a fresh 1.5mL prechilled tube and the extraction was repeated. The combined extracts were derivatised and further processed as for biofluid samples.

***DNP-hydrazone quantitation:*** Liquid chromatography was conducted using a 5cm C18 Cortecs column (particle size 1.6 µm, internal pore 2.1 µm), UPLC-grade water (aqueous phase A) and HPLC-grade acetontirile (organic phase B). Five microlitres of sample was injected using the integrated Waters autosampler, under initial conditions of 70% A and 30% B. The organic composition was then changed as follows: increased from 0.20 to 8 minutes to 40%, then 8 to 16 minutes to 95%, held at 95% for 3 minutes, and then reduced to 30% for the final minute to a total run of 20 minutes. Flow rate throughout was 0.5ml/min; the column was maintained at 40°C, and the sample vials at 4°C. Column eluate was ionised in the electrospray ion source in the negative mode for all targets except MDA, which used positive ionisation for 1 to 3 minutes. Source settings were as follows: source temperature 150°C, capillary voltage 2.5 kV, cone voltage 10 V, cone gas flow rate 200L/hr, desolvation gas temperature 400°C, desolvation gas flow rate 650L/hr. For mass spectrometry, analytes were measured in the multiple reaction-monitoring (MRM) mode, monitoring transitions, collision energies and cone voltages as in Table 2.2.3. Samples were analysed in one continuous run for each experiment. Injection order was randomised to prevent a batch effect. A calibration point was injected every 5 samples to ensure precision, immediately followed by a blank to check for carryover. For quantitation of patient or model samples, three calibration points reflecting low, medium and high concentrations were included to control drift across the linear range.

### 2.3.11 Calculations and statistics

Responses were initially inspected in the Masslynx operating software for basic quality control such as peak shape and retention. To generate peak areas, a Targetlynx (Waters, SCN855) integration method was constructed for the established transitions occurring at predicted retention times. The software performs an automated integration after two peak smooths, however in practice the integration was often incomplete or inaccurate, and therefore all peaks were manually inspected and adjusted if necessary. The final integrations were exported to Excel via a custom indexing R program (RStudio version 0.99.484, script author Dr Z Bodai), to produce a single matrix featuring samples in rows and transitions in columns.

For time-dependent optimisations and other relative experiments, raw peak areas were used. Otherwise, absolute concentrations were calculated against an external standard calibration. For ISTD controlled experiments, the peaks areas of both the external standard calibration and the unknown sample were divided by the peak area of the assigned ISTD. Tissue and cell aldehyde experiments were normalised to the biomass input, and therefore were expressed as either pg/mg tissue or pg/mg protein respectively (protein calculated using BCA see 3.3.7). To assist the interpretation of absolute tissue concentrations, an assay lower limit of quantification (aLLOQ) was set at 20 pg/mg, given that 1ng/mL was easily quantifiable for all targets, and that the minimum tissue input was strictly 10mg, and that a 0.2mL extraction volume was used. Variation in urine concentration was normalised by urinary creatinine (183).

The data was visualised using heatmaps generated with the Java package GENE-E (Broad Institute). The 'global' setting was used to illustrate the most abundant aldehydes, and the 'relative' function was used to row-normalise the data, thus visualising which samples expressed the most of a particular target compound. Principle component analysis was undertaken using Metaboanalyst 3.0 (<http://www.metaboanalyst.ca/>). To remove irrelevant effects from disparity in the mean metabolite concentrations within each sample, the data was log transformed and mean-centred. The first two principle components were used to visualise clustering. For univariate analysis, Wilcoxon, Kruskal-Wallis and 2-way ANOVA were used as appropriate, using false discovery rate (FDR) correction for multiplicity error ( $q < 0.05$ ).

### 2.3.12 Aldehyde method validation

The Food and Drugs administration guidelines were selected to guide method validation (180), including the assessment of matrix effects, recovery, accuracy, precision, and carry-over.

**Matrix effects** The sample preparation method was designed to minimise potential ionisation interferences; protein and aqueous solutes are removed, and lipids are poorly miscible in acetonitrile. To check that 'matrix effects' (ionisation interferences) were controlled, 5 mL samples of plasma and tissue (100mg powdered pig esophagus dissolved in UPLC water) were deproteinated with an equal volume of acetonitrile. Triplicate 99 $\mu$ l aliquots were then spiked with 1 $\mu$ l pre-derivatised DNP-hydrazone stock solutions to a final concentration of about 5, 10, 25, and 50 ng/mL. Responses from tissue solution and plasma were compared to acetonitrile dilutions. As seen in Figure 12, there were similar responses ( $\pm$ <5%) across all different chain lengths and functional groups, indicating that matrix effects in these samples were minimal.

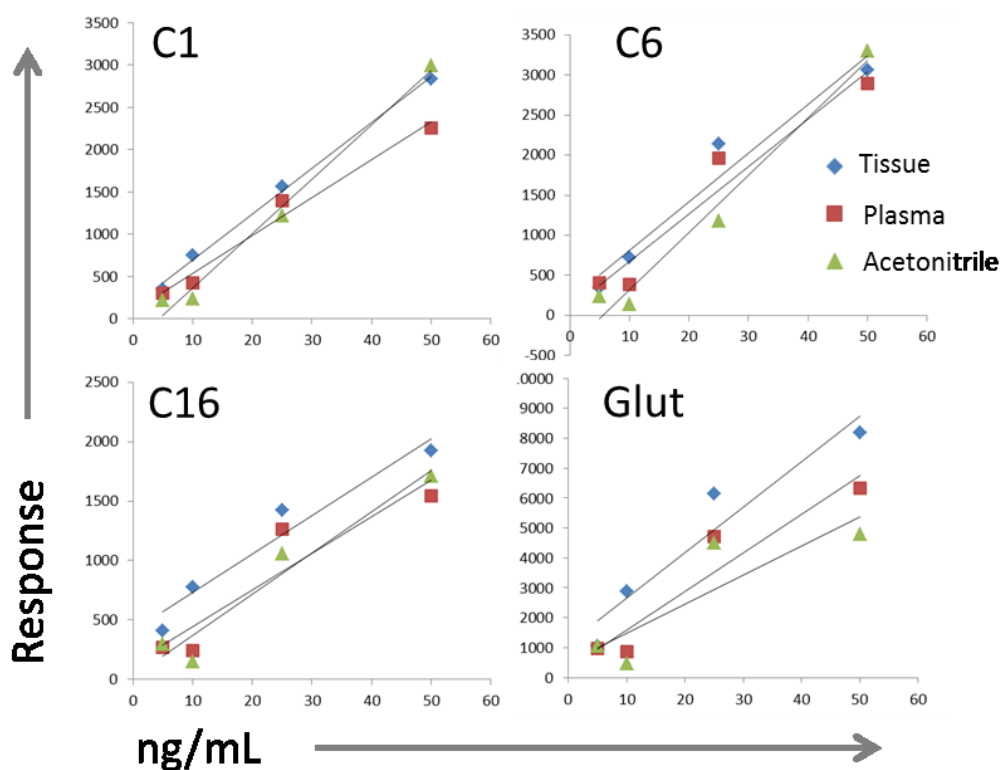


Figure 12: Aldehyde matrix effects (ng/mL indicates spiking concentration)

**Recovery** Working internal standards should correct target loss during the analytical method. However, it is still important to ensure good recovery, as sensitivity will be compromised, and a step-wise approach was taken to assess potential sources of analyte loss.

To test the efficiency of the LLE, quintuplet 200ul plasma samples were combined with an equal volume of pre-derivatised DNP-hydrazones in acetonitrile (i.e. calibration points, at 5ng/mL and 100ng/mL) in a 1:1 ratio by volume. The LLE was carried out as stated above, and the organic phase was injected into the UPLC-MS/MS system. LLE recovery (%) was calculated as (average measured concentration – average blank concentration)/calibration point concentration x 100. The results are detailed in Table 9; generally, target DNP-hydrazone completely extracted with the LLE method.

To estimate the recovery of whole process, bulk samples of media, urine, plasma and tissue-extract were spiked with underderivatised external standards to a final concentration of 100ng/mL, as well as ISTDs to 50ng/mL. The tissue recovery was estimated by extracting 400mg powdered pig esophagus in 10 mL 50:50 acetonitrile water, and spiking the slurry with the same concentrations as above prior to deproteination (this methodology – dissolving and then spiking – was felt to generate more interpretable data than trying to spike in the solid phase). Working quickly, the samples were then processed according to the optimised method and the recoveries calculated as above. Extra plasma sample were prepared owing to concerns of analyte stability. As seen in Table 9, recoveries from tissue and urine were acceptable (>50%). From more proteinaceous fluid overall recoveries were generally less good, consistent with the loss studies reported in Figure 10.

**Accuracy** Was tested by correcting the measured concentrations of the overall recovery experiments by the ISTD responses, as detailed in section 2.2.12. As seen in Table 9, the selected ISTD pairing worked well across all biospecimens, accurately correcting the target compound responses towards their expected value.

**Precision** Instrument precision was calculated as the percentage relative standard deviation (%RSD) across multiple injections of the same sample, and was shown in Section 2.2.5 to be 3-6%. Overall precision was calculated as %RSD of the spiked preparative replicates in each of the different matrices. For biofluids this was satisfactory, as the %RSD across biological triplicates was just a few percentage

**Table 9: Method validation report for aldehyde UPLC-MS/MS method**

	Plasma		Media			Urine			Plasma			Tissue			Carryover	
	5ng/mL	100ng/mL	100ng/mL spike			100ng/mL spike			100ng/mL spike			100ng/mL spike			Isopropanol SWV	
	LLE recovery (%)		RE (%)	ACC (%)	RSD (%)	RE (%)	ACC (%)	RSD (%)	RE (%)	ACC (%)	RSD (%)	RE (%)	ACC (%)	RSD (%)	0.5mL	2mL
<i>n</i> =	5	5	3	3	3	3	3	3	5	5	5	3	3	3	5	5
Formaldehyde	87	102	27	137	5	53	78	6	39	110	2	111	129	7	1	0
Acetaldehyde	77	94	27	144	8	55	117	5	44	263	4	110	81	6	4	1
Propanal	69	71	38	156	8	62	104	10	61	154	4	77	103	12	1	0
Butanal	91	104	33	135	11	55	107	8	54	145	1	79	86	0	1	0
Pentanal	89	96	34	131	1	52	92	7	55	127	5	108	93	7	0	0
Hexanal	87	101	30	105	7	53	126	4	48	137	3	122	79	7	1	0
Heptanal	97	126	30	96	9	49	110	18	56	151	3	110	106	8	1	0
Octanal	115	129	33	107	7	55	118	5	54	139	4	137	90	7	1	0
Nonanal	58	106	32	104	13	49	82	10	48	151	2	112	85	18	1	0
Decanal	107	109	31	86	10	44	87	6	51	167	2	110	94	12	3	0
Undecanal	118	113	33	68	7	37	111	10	53	147	3	121	95	12	3	0
Dodecanal	125	133	40	61	6	40	111	6	56	154	4	117	96	12	3	0
Tridecanal	116	135	40	71	8	45	130	8	53	140	2	119	94	11	3	1
Tetradecanal	117	131	43	77	8	51	120	7	61	113	1	125	96	2	3	1
Pentadecanal	100	115	39	76	9	50	145	9	49	118	3	118	114	7	4	1
Hexadecanal	81	95	20	89	9	51	125	5	18	132	2	107	85	9	4	2
Heptadecanal	140	127	42	87	9	58	105	4	41	169	5	119	102	3	6	1
Octadecanal	93	113	30	89	7	53	112	5	22	191	2	91	105	12	8	3
Benzaldehyde	90	97	36	108	7	55	93	7	59	139	3	91	118	8	2	0
Cinnamaldehyde	84	102	40	109	5	56	94	2	96	149	4	110	107	13	3	0
Acrolein	98	97	18	113	4	57	79	8	39	104	4	100	105	10	0	0
Trans-2-pentenal	96	105	32	108	6	54	94	7	56	100	7	96	125	2	0	0
Trans-2-hexenal	127	93	35	119	7	62	104	5	67	103	3	116	119	14	1	1
Trans-2-heptenal	85	119	34	115	11	59	102	7	61	118	5	110	117	12	4	0
Trans-2-octenal	124	108	33	123	9	64	93	9	65	104	6	108	132	17	2	0
Trans-2-nonenal	127	126	36	104	10	55	101	8	57	106	4	122	123	9	3	0
2,4-nonadienal	119	128	38	102	6	58	98	6	64	85	6	113	116	13	2	0
2,4-decadienal	137	133	30	128	8	56	96	5	59	92	3	107	110	12	2	0
Crotonaldehyde	103	103	42	105	10	63	86	6	74	107	4	102	113	15	0	0
4-Hydroxy-2-nonenal	107	118	31	120	7	48	99	8	67	97	7	105	110	13	3	0
Malondialdehyde	80	89	23	141	9	60	111	10	43	103	2	114	85	12	0	1
Glyoxal	104	137	34	104	7	48	106	6	55	183	9	48	86	18	8	0
Methylglyoxal	94	123	37	108	9	56	153	7	63	158	12	49	94	4	2	1
Glutaraldehyde	62	78	36	52	9	26	150	3	28	103	14	56	107	15	2	0
4-oxo-2-nonenal	122	127	6	91	9	47	110	11	43	114	1	61	85	11	5	1

LLE, liquid-liquid recovery; RE, recovery (%); ACC, accuracy (%); %RSD, relative standard deviation (technical replicates); SWV, volume of isopropanol in the strong needle wash. Carryover values are % of previous in a blank



points higher than the instrument precision, if anything (see Table 9). For tissue it was slightly higher again at around 10%. This could be due to the inherently inhomogeneous nature of the tissue-extract slurry, and is an accepted limitation of the method.

### **2.3.13 Method application rationale**

The loss/recovery experiments described in sections 2.3.9 and 2.3.12 identified that aldehydes are increasingly unstable in biofluids of higher protein content. Spiking experiments with plasma and media revealed poorer recoveries that were likely explained by greater protein interactions with the free analytes rather than preparative loss of the DNP-hydrazones. Within the constraints of the spiking experiments, these losses were correctable with ISTDs for all matrices. However, these limitations must be considered with the nature of sample acquisition:

- Tissue samples had good recoveries and accuracies, and the nature of sample acquisition meant metabolites were generally protected owing to the few seconds between anatomical disconnection and snap freezing. The method therefore applied to quantify aldehydes in tissue samples.
- Plasma had poor recoveries and good accuracy. However, the need to fractionate whole blood will mean an unprotected period of 5-10 minutes. In light of the loss experiments, this was felt to be prohibitive for too many compounds, and therefore aldehydes were not quantified in plasma samples.
- Media experiments had relatively poor recoveries and good accuracies, however the sample could be processed in a few seconds from the well, and therefore this application was selected (see Chapter 4).
- Urine aldehydes had reasonable recoveries, accurate ISTD correction, and were reasonably stable over 30 minutes, and so OAC urine aldehydes were measured.

The method was primarily applied to quantify aldehydes in OAC tissue and relevant control tissue. To assess whether aldehyde enrichment maybe a cell-autonomous phenotype, the method was then applied to quantify aldehydes in human OAC tumours established subcutaneously in nude mice (cell line xenografts). Finally, the method was applied to quantify OAC urinary aldehydes in a pilot cohort.

The method was also applied to quantify aldehydes in *in vitro* experiments; these are discussed in Chapter 4.

## 2.4 Results: Application of UPLC-MS/MS method to samples.

### 2.4.1: Aldehydes in OAC and control tissue samples

**Patient characteristics** To test the hypothesis that aldehydes are deregulated in OAC tissue, aldehydes were extracted and quantified from OAC biopsies (AdT) according to the optimised method (2.3.10). Two control tissues were selected: proximal squamous mucosa biopsies from the same patients (i.e. matched sample, SqT) and also age/sex-matched squamous mucosa from patients with endoscopically normal upper gastrointestinal tracts (SqN samples). During the study 35 AdT/SqT and 10 SqN patients were recruited (there is no published data on free aldehyde tissue concentrations to inform a power calculation). The ten SqN patients underwent endoscopy for a variety of reasons, including upper abdominal pain (5), dysphagia (2), black stools (2), anaemia (1), although no endoluminal cause was found in any patient. Samples from one OAC participant were excluded as the patient was taking a known ALDH inhibitor disulphiram (Antabuse). Demographics and tumour characteristics are given in Table 10. There were no significant differences between the groups, although less SqN participants took acid suppression medication.

**Aldehyde metabolic phenotypes across groups** As mentioned previously, a cautiously high universal LLOQ of 20 pg/mg tissue as applied across the target compounds to assist interpretation. Instrument stability was checked by injecting quality controls every 5 to 7 samples at low (2.5ng/mL), medium (25ng/mL) and high concentration (100ng/mL) and the measured %RSDs were 3.8%, 1.4% and 1.3% respectively.

As seen in Figure 13, most aldehydes were detectable above this concentration. Alkanals were the most abundant, with short chain (C1, C2) and fatty alkanals (C9 and C15-18) in particular being the most concentrated. From C12, there was an increasing abundance with chain length, particularly with even number carbon chains. In the most concentrated samples, C16 and C18 were close to the upper limit of quantification. Reactive species such as enals, dienals, dialdehydes and bifunctional aldehydes featured variably and less prominently; in particular, trans-2-pentenal, trans-2-hexenal, and the two dienals were rarely detected. Generally however, a complex mix of aldehydes at fairly high concentrations was observed.

Table 10: Demographics of UPLC-MS/MS clinical phenotyping

	Tissue		<i>P</i> † =	Urine		<i>P</i> † =
	AdT/SqT	SqN		OAC	Con	
<b><i>n</i> =</b>	34	10		17	23	
<b>Age*</b>	67 (58 - 63)	67 (43 - 81)	0.81	66 (58 - 73)	64 (58 - 71)	0.93
<b>Gender</b>	31 (91%)	9 (90%)	0.85	14 (82%)	17 (74%)	0.91
<b>Smoking History</b>			1			0.66
Never	19 (56%)	5 (50%)		12 (71%)	19 (83%)	
Ex	12 (35%)	3 (30%)		4 (24%)	3 (13%)	
Current	3 (9%)	2 (20%)		1 (6%)	1 (4%)	
<b>Alcohol history†</b>			0.17			0.53
Within limits	29 (86%)	6 (60%)		11 (65%)	17 (74%)	
Excess of limits	5 (14%)	4 (40%)		6 (35%)	6 (26%)	
<b>Reflux symptoms</b>	12 (35%)	4 (40%)	1	9 (53%)	9 (39%)	0.54
<b>Acid suppression</b>	23 (67%)	3 (30%)	0.064	9 (53%)	9 (39%)	0.54
<b>Local stage</b>						
T1	5 (15%)	-		1 (6%)	-	
T2	2 (6%)	-		2 (12%)	-	
T3	19 (56%)	-		12 (71%)	-	
T4	8 (24%)	-		2 (12%)	-	
<b>Nodal stage</b>						
N0	9 (26%)	-		7 (41%)	-	
N1	19 (56%)	-		8 (47%)	-	
N2	5 (15%)	-		1 (6%)	-	
N3	1 (3%)	-		1 (6%)	-	
<b>Metastasis</b>						
M0	33 (97%)	-		17 (100%)	-	
M1	1 (3%)	-		0	-	
<b>Differentiation</b>						
Well	4 (12%)	-		3 (18%)	-	
Moderate	16 (41%)	-		12 (71%)	-	
Poor	14 (47%)	-		2 (12%)	-	
<b>Post-chemo sample</b>	14 (41%)	-		0	-	

\*Age is presented as median (interquartile range) otherwise all other values are n (%) †21 units/week for male, 14 units per week for female ‡Mann-Whitney U-test used for age otherwise Chi-squared/Fisher's exact test. Reflux symptoms - currently reporting reflux symptoms such as heartburn, waterbrash. Acid suppression - patients taking a proton pump inhibitor or histamine antagonist

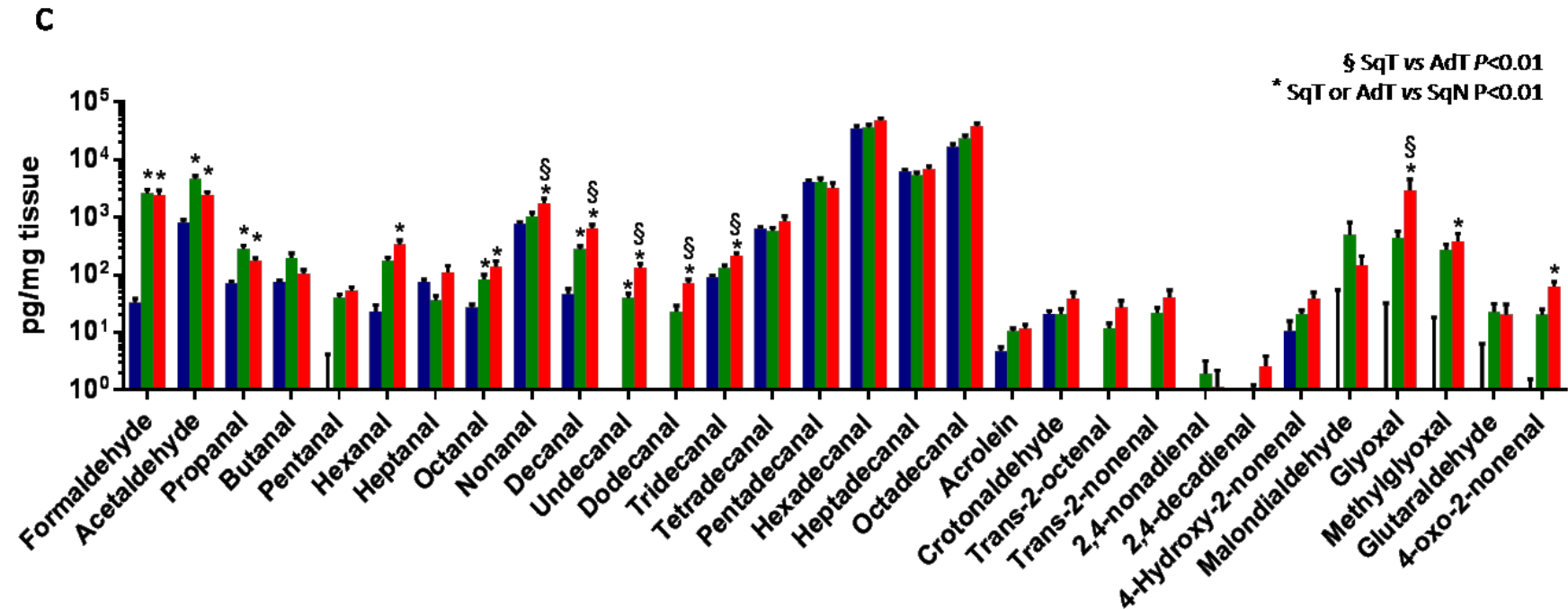
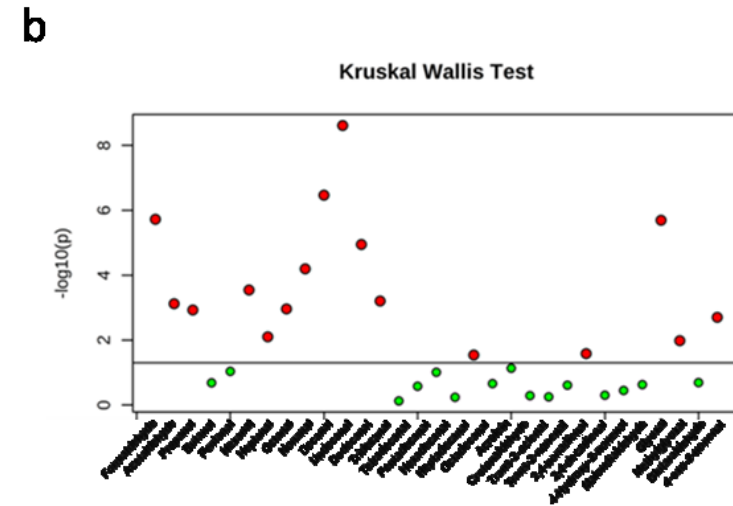
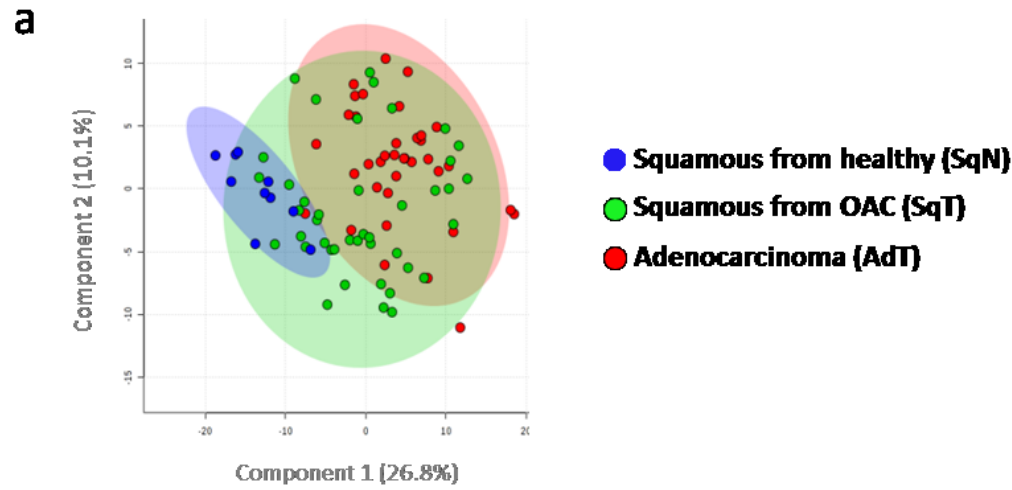
**Aldehyde metabolic differences between groups** Principle component analysis was chosen to explore aldehyde patterns between sample types (see Figure 13a). This multivariate technique assesses relationships in reduced dimensional space without pre-determined specification of sample groups (i.e. statistically unsupervised). The data was normalised by weight, and then log transformed and mean-centred to harmonise the weight each variable brought to the model despite large differences in their concentration. The first two components (of 12) explained 26.8% and 10.1% of the observed variation between the samples. In this unsupervised model, AdT clustered separately from SqN, suggesting exclusive aldehyde metabolic phenotypes. However, SqT clustered over both AdT and SqN, indicating shared phenotypes and potentially metabolic field effects.

To probe for discriminating metabolites, the Kruskal-Wallis test was applied to each target compound, with a Bonferroni correction (see Figure 13b and Section 2.3.11). There was enrichment of three classes of aldehydes in both SqT and AdT compared to SqN: very short chain aldehydes ( $P < 0.001$ ), medium-to-fatty aldehydes ( $P < 0.01-0.00001$ ), and dicarbonyls ( $P < 0.01$ ). Notably, C2 and C6 were both around 10x more concentrated in AdT and SqT, and C1 was nearly 100x more concentrated, compared to SqN. Between normal and malignant tissue from OAC (i.e. SqT and AdT), there was only significant differences in C9, C10, C11, C12 and C13 (C9-C13) concentrations, all of which were increased in OAC tissue. C11 and C12 were relatively trace compounds and the concentration of both in SqN was below the universal LLOQ (20 pg/mg); however for SqT and AdT, the measured concentrations were at least 4x higher than the LLOQ.

A pragmatic choice was to acquire OAC samples by surgery and endoscopy. To check whether aldehyde metabolic differences could be explained by differences in sampling methodology, a subgroup analysis was performed on OAC samples, which found no significant differences between the groups. A second subgroup analysis was performed by acid suppressing medication use as there seem to be more OAC patients taking these. Again, there were no differences in any target compound.

**Figure 13: Aldehyde concentrations in OAC tissues and controls (next page).**

Panel a, Principle component analysis of aldehyde phenotypes in oesophageal tissues (colours explained. Shaded areas represent 95% confidence intervals). Panel b, Significance of aldehyde differences (log<sub>10</sub> transformed  $P$ -values) in SqN, SqT and AdT, tested by Kruskal-Wallis test. Here red dots indicate that there were significant differences between the groups, green dots indicate no significant difference. Panel c, Univariate analysis of aldehyde differences between the indicated tissue-types (log<sub>10</sub> transformed concentrations in tissue).



**Table 11: Aldehyde concentrations (pg/mg) in selected subgroups of AdT samples**

	Sampling technique			Use of PPI/H2 antagonist		
	Endoscopy	Surgery	<i>P</i> =	No	Yes	<i>P</i> =
<i>n</i> =	<b>20</b>	<b>14</b>		<b>11</b>	<b>23</b>	
<b>Formaldehyde</b>	2322	2223	0.93	1637	2590	0.43
<b>Acetaldehyde</b>	2832	1754	0.12	1674	2729	0.15
<b>Propanal</b>	184	146	0.37	171	167	0.92
<b>Butanal</b>	124	63	0.11	109	94	0.71
<b>Pentanal</b>	56	56	0.99	66	51	0.46
<b>Hexanal</b>	319	396	0.59	371	341	0.84
<b>Heptanal</b>	157	65	0.23	52	151	0.21
<b>Octanal</b>	185	80	0.20	84	169	0.33
<b>Nonanal</b>	2319	1190	0.17	1210	2162	0.27
<b>Decanal</b>	768	519	0.31	620	687	0.79
<b>Undecanal</b>	171	88	0.12	104	153	0.38
<b>Dodecanal</b>	86	54	0.23	60	79	0.50
<b>Tridecanal</b>	245	183	0.21	242	209	0.54
<b>Tetradecanal</b>	820	1030	0.64	749	982	0.62
<b>Pentadecanal</b>	2600	4332	0.31	2536	3685	0.53
<b>Hexadecanal</b>	51270	44521	0.60	50369	47593	0.84
<b>Heptadecanal</b>	6700	7396	0.74	7147	6910	0.91
<b>Octadecanal</b>	41274	36392	0.64	39824	38996	0.94
<b>Benzaldehyde</b>	25	0	0.44	0	22	0.68
<b>Cinnamaldehyde</b>	0	0	n/a	0	0	n/a
<b>Acrolein</b>	0	0	n/a	0	0	n/a
<b>Crotonaldehyde</b>	42	39	0.88	21	50	0.28
<b>Pentenal</b>	0	0	n/a	0	0	n/a
<b>Hexenal</b>	0	0	n/a	0	0	n/a
<b>Heptenal</b>	0	0	n/a	0	0	n/a
<b>Octenal</b>	32	24	0.68	27	30	0.89
<b>Nonenal</b>	47	40	0.82	32	49	0.59
<b>2,4-nonadienal</b>	0	0	n/a	0	0	n/a
<b>2,4-decadienal</b>	0	0	n/a	0	0	n/a
<b>Malondialdehyde</b>	147	173	0.86	145	164	0.91
<b>Glyoxal</b>	4347	1118	0.40	1304	3837	0.53
<b>Methylglyoxal</b>	495	290	0.52	265	480	0.52
<b>Glutaraldehyde</b>	21	24	0.90	31	0	0.62
<b>Oxononenal</b>	58	74	0.57	95	50	0.13

P-values calculated with Mann-Whitney U-test. PPI, proton pump inhibitor; H2 antagonist, H2 receptor blocking antacid medication.

#### 2.4.2: Aldehyde concentrations in murine OAC xenografts

**Model characteristics** The purpose of this experiment was to measure aldehyde concentrations subcutaneous OAC tumours, to assess whether cell-autonomous factors are sufficient to enrich aldehydes. Single-flank subcutaneous xenografts were established for three OAC cell lines (OE33, FLO1, and ESO51,  $n = 5$  mice each). These lines were selected as they express *ALDH* genes relatively poorly (see Chapter 3). Xenograft tumours were harvested when they reached the license limit of 200mm<sup>3</sup>.

**Aldehyde metabolic phenotypes across samples** The same universal LLOQ (20 pg/mg) was applied to this quantitative series. As only 15 samples were run, a single medium concentration quality control was serially injected at random, with an analytical precision of 0.9%. A similar pattern of trace and abundant aldehydes was observed in the murine xenografts compared to OAC aldehyde phenotypes *in situ* (see Figure 14). The most abundant were fatty alkanals; the least abundant were medium chain alkanals, enals, and dienals.

**Aldehyde metabolic differences between groups** Aldehyde concentrations of each xenograft quintuplet were compared to metabolic phenotypes *in situ* (AdT) using two-way ANOVA, with false discovery rate (FDR) correction for multiplicity error ( $q$  set at 0.05; this different test was selected as binary comparisons across multiple groups was being undertaken, rather than assessing differences across three groups). Fatty alkanals C13-C18 were significantly more enriched in the subcutaneous xenograft compared to the patient samples. Mean concentrations of C3, C5, C6, C9, and several bifunctional aldehydes were also higher in the xenograft tumours, although not significantly so after multiplicity correction.

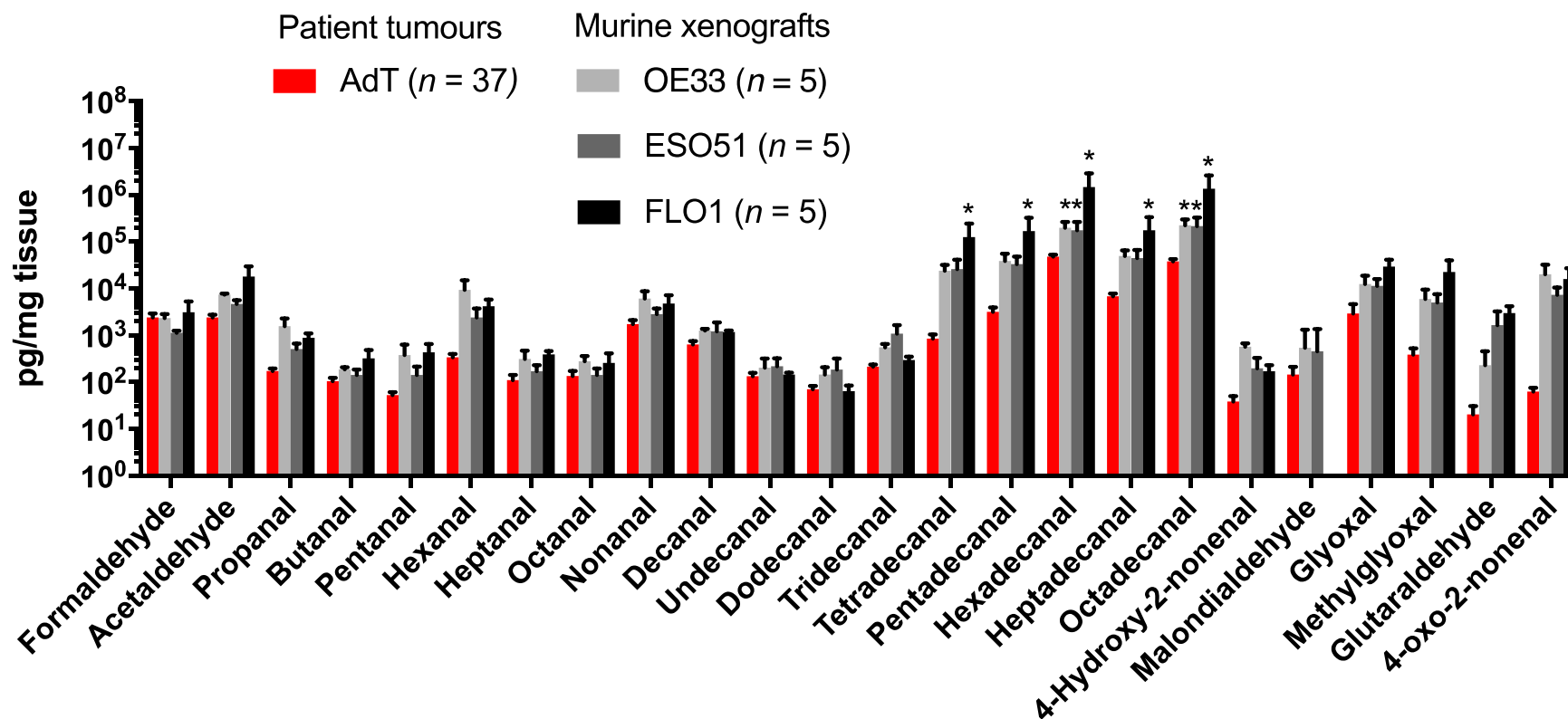


Figure 14: Aldehyde concentrations in subcutaneous murine xenograft tumours compared to OAC tumours *in situ*. \*FDR  $q < 0.05$ .



### 2.4.3: Aldehyde concentrations in OAC urine.

**Patient characteristics** To test whether circulating aldehyde concentrations were modulated in OAC, comparative studies were undertaken using urine samples from OAC patients and relevant controls. To maximise potential effects in this preliminary experiment, two cohorts only were selected: healthy volunteers with endoscopically normal upper GI tracts (HV), and patients with oesophageal adenocarcinoma (OAC). The only previous work investigating aldehyde concentrations in OAC urine headspace showed significant differences with 10-15 in each group (18). Using this as a guide, it was decided to recruit 20 participants to each group. Control patients were referred for endoscopy for the following reasons: dyspepsia (6) persistent reflux (4) bloating/cramps (4) black stools (2) ulcer follow-up (2) dysphagia (2), although no endoscopic diagnoses accounted for their complaints. Baseline variables and tumour characteristics are given in Table 10. There were no differences in baseline variables between the two cohorts.

**Aldehyde metabolic phenotypes across samples** A universal LLOQs for urine were taken as 20 pg/ $\mu$ mol creatinine, given the previous analytical LLOQ (around 200pg/mL) and an average creatinine concentration of 10 $\mu$ mol/mL across the measured samples. Acetaldehyde, nonanal, hexadecanal, octadecanal, glyoxal and methylglyoxal were among the most abundant aldehydes in urine (see Figure 15), although the striking enrichment of fatty alkanals seen in tissue was not observed. As before, aromatics, enals and dienals were of trace concentration, frequently below the LLOQ, and thus were not included in comparative analysis.

**Aldehyde metabolic differences between groups** To assess how the two groups clustered in reduced dimensional space, PCA analysis was again performed. The first and second principle components (PCs) of the 13-component model explained 20.8% and 9.1% of the variation respectively (Figure 15a). There was no separation of the groups in any combination of the components. Similarly on univariate analysis (using Wilcoxon test), only hexanal was significantly different (OAC 480 pg/ $\mu$ mol vs control 321 pg/ $\mu$ mol,  $P=0.032$ ), although not after multiplicity correction (Figure 15b). These values are displayed for each measurable target compound (Figure 15c).

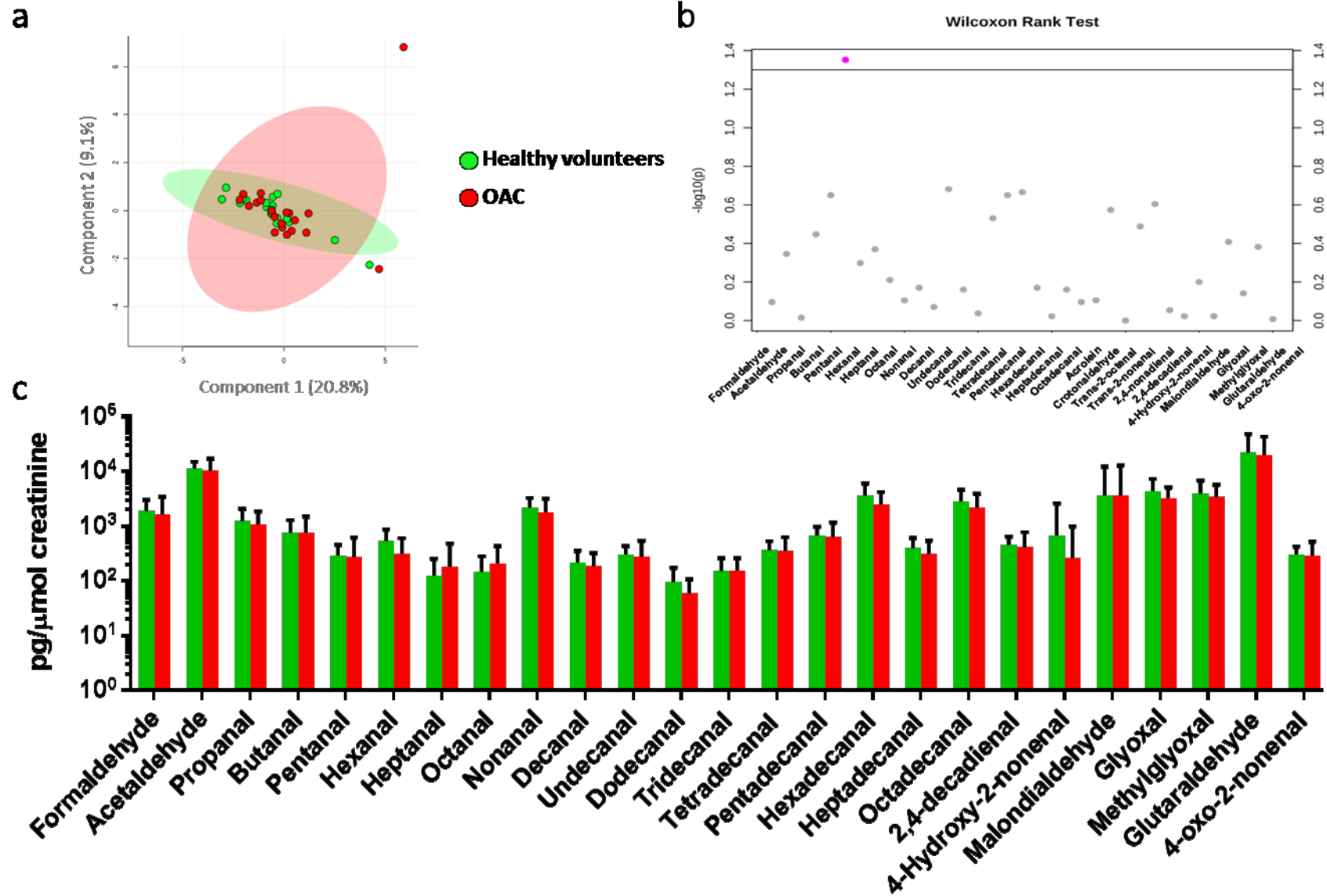


Figure 15: Urinary aldehyde concentrations in OAC compared to healthy volunteers (abbrev as above)

## 2.5 Discussion

### 2.5.1 A UPLC-ESI-MS/MS method for quantifying carbonyls from biospecimens.

This is the first quantitative method that sought to capture this variety of aldehyde and ketones (43 compounds), with unambiguous identification for all targets. Although LC-MS quantitation of special DNPH derivatives like MDA and glyoxal have been independently reported (153,178), none have captured these different products in a single method with two ionisation modes. Other strengths of this method include: (i) rapid, facile sample processing with minimised matrix effects (ii) a relatively fast chromatography method for 43 compounds (iii) improved sensitivity limits of compared to most recent reports (169,178,179) (iv) reasonable recoveries in multiple biosample types (v) excellent accuracies across all sample types. The end result was a practical method for the confident measurement of common aldehydes in tissue, urine and model samples.

A number of concepts regarding aldehyde quantitation were revisited and expanded. These results will be discussed following the projected issues set out in Chapter 2.1

*Ionisation technique* The Waters ESI source was found to offer an order of magnitude better sensitivity for all target derivatives except the special MDA derivative, which was >50x less sensitive than the APCI source (a potential solution for measuring trace MDA levels is simply to repeat analytical runs after ESI acquisition with APCI, with a shorter chromatography run). The ESI source was also more stable (RSDs all <10%), and did not experience significant matrix effects with the saline-acetonitrile LLE (190,191). This was in keeping with a previous report of measuring DNP-hydrazones with the Waters ESI source (179). From a practical perspective, the ESI source allowed >75% targets to be quantified at native concentrations.

*Best parent ion to measure bifunctional and dialdehydes and disambiguation of isomers/isobars* The DNPH derivatization chemistry of common aldehyde subgroups has been extensively investigated, although typically in studies of a few target compounds or one subgroup. In the context of biological samples, alkanals, ketones and dialdehydes are all abundant compounds, are isomers/isobars, and all interact with DNPH, and therefore the unambiguous determination of these compounds

critically relies on a collective understanding of their DNPH reactions, UPLC elution and fragmentation pattern (see Figure 7). Some clear rules emerged from the present study: (i) Alkanals (and simple aldehydes more generally) uniquely give an abundant 163 ion (mentioned in (175)) (ii) in the absence of a 163 ion, the abundant 153 or 152 ion can be used to quantify ketones (iii) in the presence of a 163 ion, an early ketone-specific minor peak can be for quantification (iv) bifunctional carbonyl usually give a more abundant di-DNPH derivative, which is best quantified on the 182 daughter ion (mentioned in (153)) (v) MDA has a special derivative that give unique daughters (mentioned in (178)).

*Derivatisation conditions* Utilising DNPH in biospecimens is complicated by the presence of strong buffers (which interfere with the pH-dependent reaction) and also non-target carbonyls (which decrease the availability of DNPH to react with target compounds). This was illustrated in Figure 10, wherein the optimised DNPH concentration for external standards in acetonitrile failed to recover the same standards from a plasma sample. The final method used a 1000x molar excess of DNPH, which seemed sufficient to recover compounds competently. This also contained an excess of phosphoric acid (pH ~2.1) with benefits of minimising adventitious stereoisomer formation (193) overcoming biospecimens buffers, as well as providing adequate acidity to overcome native buffers.

*Background* In keeping with several reports (153), background concentrations of C1, C2, glyoxal, and methylglyoxal were high if appropriate precautions regarding DNPH, solvents and glassware were not taken. The levels were reduced to acceptable values ( i.e. <10% of the lowest native concentration) by following published recommendations (145,152,153), in particular extracting DNP-hydrazone from DNPH stock solutions with hexane, using only the freshest analytical grade solvents, and baking analytical glass and plastic.

*Analyte loss due to reactivity of aldehydes during sample preparation* Aldehydes' stability in biosamples decreased with increasing matrix complexity. Low molecular weight and/or bifunctional aldehydes were the most vulnerable to losses; acrolein and malondialdehyde were lost in just a few minutes in ice-cold plasma, in keeping with previous analyses(169). This highlights the importance of introducing ISTDs and derivatising as soon as practically possible after sample processing, and indicates that plasma aldehyde analysis is probably not suitable for the more reactive species (and questions the validity of a large number of studies investigating plasma-MDA

concentrations with TBARS assays over the last 40 years (e.g. 201–203), and several recent studies of plasma-aldehydes using LC-MS (169,170,194)). Others have looked at more direct measurement techniques to measure blood aldehydes, for example whole blood headspace analysis using solid phase microextraction (SPME)–GC-MS (99,165); clearly much more work is needed to overcome the conflicting priorities of collating samples, fractionating whole blood, and preserving unstable compounds. Aldehyde recoveries from tissue were better than for biofluids, presumably as the protein was precipitated within a few seconds of the tissue being extracted, and thus the opportunity to interact with biogenic nucleophiles was much less.

*Method validation* The UPLC-MS/MS method performed adequately in accuracy experiments in all the test biospecimens, and ultimately this is the critical parameter to satisfy. Additionally the sensitivity of the method was adequate despite a cautiously high universal LLOQ. Recovery of target compounds from biofluids was less than ideal although correctable, and this stressed the essential requirement for robust ISTDs in the method.

### **2.5.2 Aldehydes in normal and malignant oesophageal tissues**

This was the first report of free tissue aldehyde concentrations in oesophageal tissue. The measured concentrations were in the same range as previously described in astrocytomas, one of the few other reports in any other tissue using an MS technology (100). Isomer/isobar metabolites such as propanal, acetone and glyoxal were abundant. This underlines the importance of developing a sufficiently discriminant method for the proposed application.

Aldehyde concentrations were remarkably stable across the different samples and tissue-types, with the same compounds tending to be abundant (C2, C6, C9, C14-18) or rare. However, the most notable finding was tissue-specific aldehyde profiles. The most striking differences were between samples from OAC patients (either AdT or SqT) and squamous samples from healthy patients (SqN). Several alkanal species and three bifunctional aldehydes were significantly increased OAC samples. In the context of these diffusible and pervasive metabolites, it is unsurprising most metabolic differences were at an organ level rather tissue-specific; this confirms previous metabolic (56) and other (207,208) studies proposing “field effects”

throughout the malignant oesophagus. Similar effects have been noted in OSCC(209). The tissue PCA (Figure 13a) is suggestive of such field change, with the 95% confidence intervals for SqN and AdT being effectively exclusive, and that of SqT sharing characteristics of both. The only differences between SqT and AdT were alkanals C9 to C13 and glyoxal. This fascinating set of sequential carbon lengths is suggestive of a common origin, or a common catabolic impairment, with the non-polar, semi-volatile nature precluding diffusion between tissues.

The nature of tissue collection may also part explain the similar metabolic phenotypes of SqT and AdT. Endoscopic biopsy forceps' jaws are 3mm<sup>3</sup> and thus will collect the most superficial cells of the oesophageal wall with a sample of mucous. In the RNA study (see Section 3), histological control of the tissue samples revealed a misclassification rate was 7%. However, similar control procedures were not possible in this cohort given the instability of aldehydes in thawed samples, and thus a similar error rate must be assumed for the AdT cohort. Additionally, a small volume of each sample will be composed of the pre-epithelial mucous layer, which is continuous throughout the oesophagus.

The most concentrated aldehydes were fatty alkanals, in particular C16 and C18. There was no difference in concentration between the tested tissue-types, indicating that these high concentrations are related to normal oesophageal processes. Palmitic (16:0) and stearic (18:0) acids are the most abundant saturated fatty acids in blood and diverse tissues (210), and the analogous concentration of the corresponding aldehydes suggests a parallel metabolic process. Other explanations could be (i) features of unexplored metabolism (ii) DNPH capture of bound acyl-carbonyl intermediates during fatty acid synthesis (iii) reactivity of the derivatization agent towards non-carbonyl species e.g. fatty acyl-CoA conjugates or the fatty acids themselves. However, there are no reports of carboxylic acids reacting with DNPH under physiological conditions (in fact, DNPH is traditionally thought of as a selective agent for determining carbonyl concentrations in a fatty acid mixture). This was verified by a simple bench experiment of mixing neat DNPH with palmitic acid (data not shown). Similarly, there are no reports of acyl-CoA species reacting with DNPH.

### **2.5.3 Aldehydes in OAC xenografts**

Subcutaneous OAC tumours developed in immunocompromised nude mice form complex tumours with a microvasculature and stromal support. They have been

criticised as relevant models of human OAC tumours, as they (i) are relatively clonal (ii) undergo attenuated immune cell interactions (iii) do not undergo luminal stress. From a mechanistic perspective, this model offers a convenient way of controlling these variables, as an observation of the expected metabolic phenotypes could indicate that extrinsic factors were not essential i.e. enriched aldehydes are “cell-autonomous”. Additionally, they are fundamentally more relevant for the study of toxins such as aldehydes as compared to standard *in vitro* cell culture, as these techniques inherently avoids cellular stress.

Most aldehydes' concentration was similar or enriched across the three xenograft cell line types compared to patients' tumours. These results may suggest that luminal and immune stress may be less influential in aldehyde production compared to cell-autonomous factors. However, there is much work required to fully explore this; xenograft tumours have a purer cellularity, grow more quickly, have a highly exaggerated tumour-to-host mass ratio, and recent findings suggest they are can be more hypoxic than native tumours(211). All of these factors may diminish the interpretable value of the observed effects, although nullifying these potential confounders with better controls is difficult. It may be more helpful to look at aldehyde concentrations within an OAC animal model with the capacity for longitudinal sampling through disease transformation, for example the EBV-IL1 mouse model of Barrett's adenocarcinoma, or the surgical oesophago-duodenal anastomotic reflux rat model (212,213). Interesting, previous groups have potentiated tumour production in a oesophago-duodenal anastomosis set my enriching the chow with iron, and noted that several markers of oxidative stress including TBARS and protein carbonyls all enriched (214).

#### **2.5.4 Aldehydes in OAC and control urine**

Urine contained C2, C9 and dialdehydes among lower concentrations of other aldehydes, but had relatively less fatty aldehydes. Of the 25 aldehydes which were consistently above the LLOQ, only five (C1, C2, C6, C10, and methylglyoxal) have been identified in a recent comprehensive survey of urinary metabolites using eight analytical platforms and a comprehensive literature review (215)., underlining the requirement of bespoke methodology for aldehyde quantitation. However, there was no separation of normal and OAC samples on the basis of aldehyde concentrations. A previous study has assessed urinary aldehydes using SIFT-MS and found significantly increased C2 in OAC patient samples (18), however this study failed to

normalise target concentrations to the overall water content of the urine. In this study, urinary creatinine concentration was used to normalise the concentrating effects of the kidney (216). Thus the data presented here may be less prone to error and more representative of the natural values in urine. Patients coming for endoscopy or surgery will be starved for >6 hours at the point of sampling. For the metabolic point-of-view, this offers standardisation of interferences from active food digestion and absorption, but it also means that urine samples are more likely to be stored in the bladder, at 37°C, for lengthy and unstandardized periods, which may not be suitable to analyse reactive aldehydes (see *Loss Studies* section 2.3.9). Perhaps a better approach would be to measure freshly passed urine in catheterised patients, as the latent time from kidney to sample point would be <1 minute.



**CHAPTER 3 – GENETIC DRIVERS OF  
ALDEHYDE METABOLIC  
REPROGRAMMING IN OESOPHAGEAL  
ADENOCARCINOMA**

## Summary

The purpose of this Experimental Section was to use a candidate-based approach to discover potential genetic influences of aldehyde metabolic reprogramming in OAC (see Figure 16).

Initially, historical expression datasets were explored for candidates using two complementary *in silico* mining techniques. These analyses suggested that several “aldehyde oxidoreductase” genes were consistently and profoundly down-regulated in OAC compared to normal oesophageal mucosa. For a subset of candidates (8 isoenzymes of aldehyde dehydrogenase, ALDH), these findings were validated experimentally using quantitative polymerase chain reaction on RNA extracted from endoscopic biopsies (OAC  $n = 67$ ). Then, using immunohistochemistry, expression trends were further evaluated across the tumour microenvironment for the most promising candidates *ALDH3A1* and *ALDH3A2*, and compared to clinical metadata (OAC  $n = 412$ ). Low expression of both isoenzymes occurred in 90% and 76% of OAC cases respectively, with a transition point occurring between Barrett’s dysplasia and invasive cancer. Low expression of *ALDH3A2* was significantly associated with higher stage, more positive lymph nodes, poor differentiation and poorer overall survival, and was independently predictive on Cox regression (patients with low *ALDH3A2* were 64% more likely to die of their disease,  $P = 0.01$ ). These survival trends were verified in The Cancer Genome Atlas dataset.

Moving upstream, a thorough informatics review was then undertaken to search for explanations of *ALDH3A1/2*-expression loss in OAC. These genes were not commonly mutated or methylated. However, they co-locate to 17p, a locus which frequently undergoes loss of heterozygosity in OAC, resulting in unfavourable prognosis. Using the TCGA dataset, a genome-wide analysis found that expression of *ALDH3A1/2* was highly correlated with several of its telomeric neighbours, suggestive that the trans-regulatory event is loco-regional. Finally, there was a trend between *ALDH3A2* copy number and expression.

Using immunoblotting, representative *ALDH* expression patterns were observed in cell models of OAC and squamous keratinocytes, further validating primary patient data and identifying appropriate systems for further functional experiments.

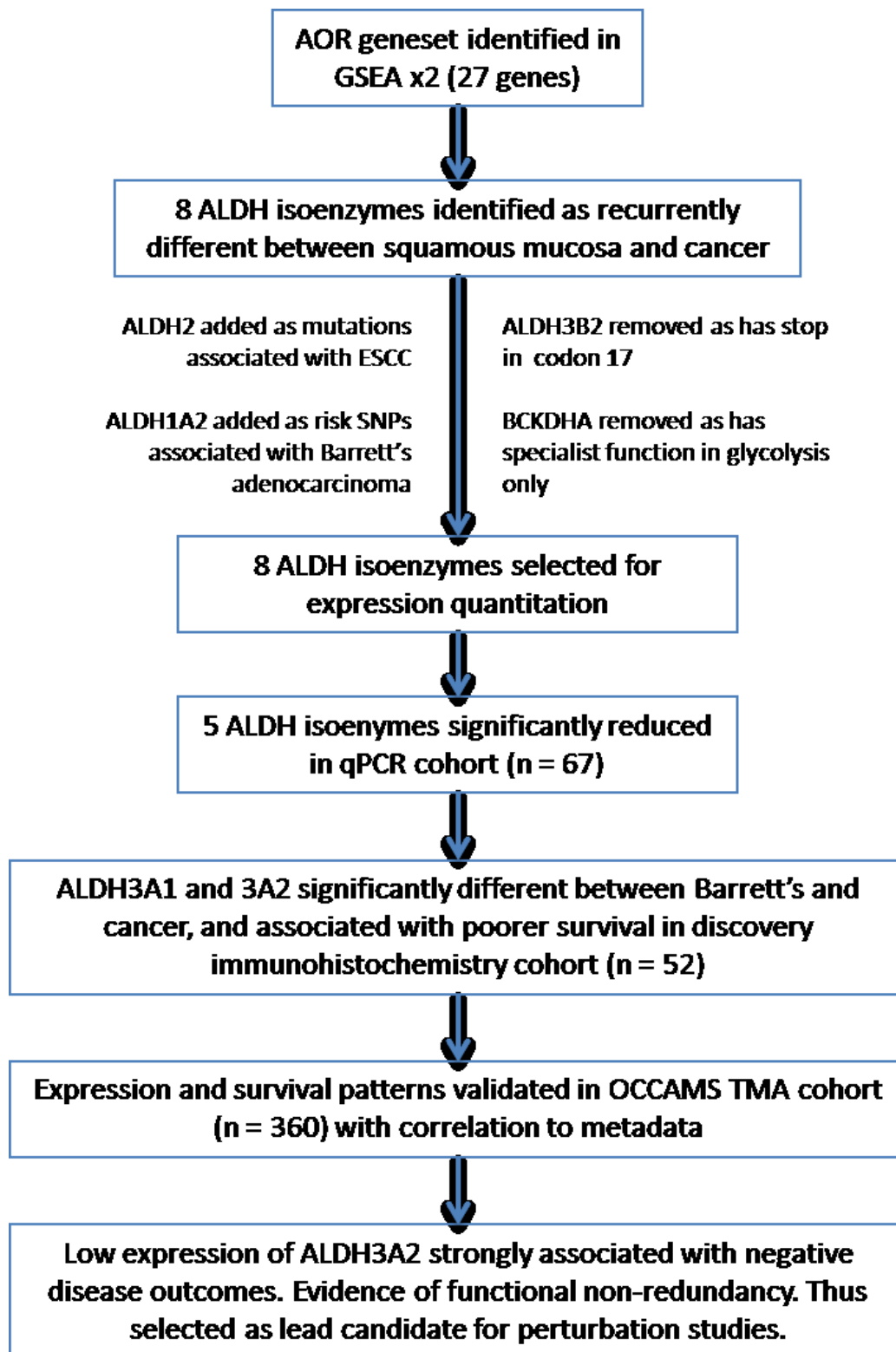


Figure 16: Summary of candidate-based discovery of drivers of aldehyde reprogramming in OAC

### 3.1 Methodological rationale

Since the introduction of microarray-based gene expression profiling technique in the 1990s, journals have stipulated public access to associated datasets as a condition of publication. More recently, next generation sequencing technologies have permitted high-resolution molecular characterisation of multiple molecular information sources at relatively low cost, including somatic code (whole genome/exome sequencing), gene expression, micro RNAs and splice variants (RNA-seq), and so on (ATAC-seq etc.). Thus, international collaborations have been established to comprehensively characterise the molecular framework of cancer, with the data freely distributed to facilitate evidence-based hypothesis testing (e.g. The Cancer Genome Atlas, International Cancer Genome Consortium etc).

These rich datasets must be approached cautiously, as their complexity renders univariate analyses prone to false discovery. Helpfully, informatics resources have emerged which aggregate dataset elements into functionally related subgroups, which are then ranked according to a particular characteristic (e.g. how large or stable a change in expression is). This simplifies the multiplicity of the dataset while accentuating biological trends, giving a more powerful understanding of the dominant functional differences between investigated groups. For added power, two contrasting algorithms were used in the present analysis: gene-set enrichment analysis(217), which collapses data to user-defined functional divisions (e.g. KEGG pathways or Gene Ontology groups), and Ingenuity Pathway Analysis, which collapses data to highly-evidenced pathways in a massive curated collection.

The results of these *in silico* analyses must then be validated empirically. This can be achieved at the RNA level with quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR, or qPCR), or the protein level with immunohistochemistry (IHC). With adequate primer design, qPCR offers a highly selective and quantitative expression analysis of any gene, with splice variant resolution, but is limited by cell contamination in the initial input. Immunohistochemistry offers expression analysis across the microenvironment by using antibodies to target staining of histological sections, and thus offers functional or mechanistic insights by assessing expression by environmental context. Paraffin-embedded archived material is suitable for this methodology, facilitating well-powered analyses of uncommon diseases such as OAC through national resource collaborations. These samples may also have comprehensive clinical metadata with lengthy follow-up for detailed subgroup

analyses. However, IHC is limited by the availability of specific antibodies, and is semi-quantitative and relatively subjective. The strengths of both qPCR and IHC largely offset the weaknesses of the other, and thus in this work they have been used together as they were synergistic and complementary. Subgroup analyses stratified by baseline metadata was undertaken to provide clinical significance of the measured gene expression changes. Informatics searches using TCGA data were also undertaken to search for upstream regulatory clues, and to verify patterns in the primary data.

*ALDH* expression was tested in a large panel of oesophageal cell models using Western blotting, with the intention of further assessing cancer/normal *ALDH* expression patterns with this third empirical technique, and also to define models for the next phases of the work. The single commercial “normal” oesophageal model – the SV-40 immortalised HET-1A cell line - expressed none of the expected *ALDH* isozymes, and therefore better “normal” models were established using primary cultures of oesophageal keratinocytes (the questionable validity of the HET-1A line precluded it from further study). These were validated as representative models using epithelial and keratinocyte markers, both with immunofluorescence and further immunoblotting, prior to *ALDH* quantitation.

### **3.2 Hypothesis and aims**

Chapter hypothesis: stereotyped genetic influences contribute to aldehyde metabolic reprogramming in oesophageal adenocarcinoma

Aims:

- i. To discover candidate drivers of aldehyde reprogramming using informatics analyses of archived expression datasets
- ii. To validate leading candidates by direct gene expression analysis in clinical material
- iii. To establish the clinical significance of candidate gene deregulation
- iv. To understand candidate gene deregulation through OAC transformation
- v. To discover upstream coordination of candidate perturbation
- vi. To define aldehyde gene expression in normal and malignant oesophageal models

## 3.3 Methods

### 3.3.1 Bioinformatics discovery of candidate drivers of aldehyde reprogramming

To generate hypotheses and inform study design, a thorough informatics survey was undertaken. This had the following tasks:

1. Identify candidate genetic coordinators of metabolic reprogramming in OAC compared to proximal normal epithelium (PNE) and Barrett's metaplasia (BM) through *in silico* expression analysis
2. Validate candidates in all relevant datasets
3. Correlate candidate expression to clinical metadata
4. Suggest regulatory mechanisms of candidate metabolic genes

The Gene Expression Omnibus (National Centre for Biotechnology Information, USA, <http://www.ncbi.nlm.nih.gov/gds/>) and ArrayExpress (European Molecular Biology Laboratory, Wellcome Genome Campus, Cambridgeshire, UK, <https://www.ebi.ac.uk/arrayexpress>) (date both last accessed: 02/02/2016) were searched with the following terms: "\$esophagus" OR "\$esophageal cancer" OR "\$esophageal adenocarcinoma" OR "Barretts". Any study publishing transcriptomic data (microarray or RNA-sequencing) and comparing normal squamous with either Barretts or oesophageal adenocarcinoma were included. No limitations were placed on tissue specification techniques, although a minimum sample size of 10 in each group was selected. Studies were identified from the following countries: Australia (GSE39491), Germany (GSE26886), USA (GSE13898), and UK (GSE34619) (49,218–220)

Geneset Enrichment Analysis (GSEA)(217) was primarily chosen to probe transcriptome libraries. For each analysed sample, GSEA compartmentalises related genes into pre-curated genesets, which are then ranked according to overall gene enrichment between two sample-groups. It thus provides an unbiased, functionally orientated and holistic overview of expression datasets. The following settings were used, as recommended in the index publication: (i) weighted signal enrichment (ii) "Signal2Noise" ranking metric (iii) exclude sets > 500 and <100 (iv) normalise to "meandiv". The analysis was repeated on all publically available expression libraries containing at least 10 each of OAC and PNE samples, with a false discovery rate

(FDR)  $q$  value set at 0.05. To verify candidate phenotypes, Ingenuity Pathway Analysis (Qiagen) was also undertaken on the lead dataset, using the standard settings in the “Core Analysis”. To focus analysis on clear differences, thresholds were set at fold change of more than 2, and a significance level of  $1 \times 10^{-6}$ .

Next, candidate geneset expression was meta-analysed using univariate analysis (Mann-Whitney U-test) using the following comparisons: squamous vs adenocarcinoma, Barrett’s vs adenocarcinoma, squamous vs Barrett’s, Barrett’s vs gastric, normal vs gastric. A significance threshold of  $1 \times 10^{-7}$  was applied to account for multiplicity error.

Several other informatics sources were mined for further information relating to candidate gene expression and regulation. The Cancer Genome Atlas (TCGA) (221) was additionally interrogated for further functional and regulatory information regarding the lead candidates. Of the available 185 cases of esophageal cancer (last date accessed – 10/03/2017), 56 with histologically confirmed OAC were included in the analysis. RNA-seq, copy number variation (CNV), methylation and clinical metadata were analysed. Macrodeletion events were assessed using CNV data with the algorithm GISTIC 2.0 in the GenePattern environment (Broad Institute)(222). To assess neighbour co-expression patterns, RNA-seq fragments-per-kilobase-per-megabase reads were correlated against the lead candidate FPKM with Spearman test in Excel. To test the effects of gene expression on survival, the expression cohort was dichotomised about the median expression value into “high” or “low”, and survival curves were fitted each group. The statistical significance of differences in survival was assessed with the log-rank test. To assess somatic mutation frequency of these genes, whole genome sequencing data from two International Cancer Genome Consortium (ICGC) substudies (33,34) was selectively analysed for the candidate genes of interest.

### **3.3.2 Patient samples**

Ethical approval for the use of patient biospecimens and clinical data was obtained under Imperial Tissuebank committee approval number R14067, R15047, and R14018 under NRES Tissuebank ethical approval 14/LO/0742. For extraction of RNA and DNA, tissue samples were retrieved endoscopically or immediately after oesophageal disconnection during oesophagectomy, snap frozen in liquid nitrogen,

and stored at -80°C. A standardised sample processing procedure was followed for all samples. The following definitions and abbreviations were used:

- SqN: normal squamous epithelium proximal from an endoscopically normal patient (>5cm from GOJ)
- SqT: normal squamous epithelium proximal from patients with cancer (>5cm from tumour)
- BM: Barrett's metaplasia (pseudostratified columnar epithelium in the oesophagus with goblet cells)
- BD: Barrett's dysplasia (IHC only): BM with gland distortion/crowding, nuclear atypia, hyperchromatism, abnormal mitoses (223)
- AdT: oesophageal adenocarcinoma

A frozen 2-4 mm<sup>2</sup> specimen was weighed, freeze-mounted flat in OCT medium (Thermo-Life, UK), and flank cryosectioned. The flank sections were stained with haematoxylin and eosin, and used to define microdissection on a cold anvil with cold clean scalpel to a cell purity of >90%. For fractionation of RNA and DNA from tissue samples, the microdissected specimen was homogenised in Trizol using a three-step process (this was necessary to achieve good yields from the fibrous normal epithelium, see Figure 19a)(224). First, biopsies were mashed with a hand pestle to <0.5mm pieces, followed by a Reitsch bead-beater for fine blending (30 cycles/sec for 1 minute, as in Section 2), and finally a Qiashreder (Qiagen) to homogenise nucleic acid fragment lengths. Nucleic acids were precipitated from the chloroform-extracted phases and cleaned using the RNeasy and DNeasy silica spin-column kits (both Qiagen), according to the manufacturer's instructions. Strict RNA isolation precautions were rigorously adhered to, including use of RNase removal agents, fresh tips, certified plastic, and regular glove changing. Purity and concentration was assessed using ultraviolet absorbance (ND1000, Thermo-Life); a 260:280nm >1.9 for RNA and >1.8 for DNA was considered acceptable.

### **3.3.3 Quantitative reverse transcriptase polymerase chain reaction (qPCR)**

The initial 26 genes in the AOR GO geneset were rationalised to promising candidates by excluding pseudogenes, functionally irrelevant genes, and those that were not discriminatory. *ALDH1A2* and *ALDH2* were added to the list of candidates owing to strong epidemiological links with OAC. This left eight genes for experimental analysis: *ALDH1A1*, *ALDH1A2*, *ALDH1A3*, *ALDH2*, *ALDH3A1*, *ALDH3A2*, *ALDH4A1*,



and *ALDH9A1*. A sample size was calculated as 67 matched cancer-normal pairs based expression distributions extracted from an *in silico* dataset (Wang et al)(225). This calculation took an alpha = 0.05, power = 0.8, sample attrition = 15%, and returned sample sizes of 5 (*ALDH4A1*) to 67 (*ALDH1A1*) matched pairs.

To generate templates for qPCR, 1µg of RNA was converted to complementary DNA (cDNA) using the Superscript III First Strand Supermix system (Life Technologies, Thermo Scientific), and a GeneAmp® PCR System 9700 (PE Applied Biosystems), according to the manufacturer's instructions. Relative quantitative PCR was undertaken using the SybrSelect mastermix (Thermo-Life) and a 7900 HT thermal cycler (Applied Biosystems) according to the manufacturer's instructions. The MIQE standards checklist was used to ensure methodological and reporting quality for a qRT-PCR experiment(226,227), including the selection of reference genes from a panel of ten (*GAPDH* was the most stable in an initial screen of ten samples, although it is also an AOR, so *HPRT1* was co-analysed as a second reference (110,228,229). Primers were selected through PrimerBank and BLASTed to ensure specificity (see Appendix 3 for sequences)(230,231). Reactions were run in groups of four pairs of samples in 384-well plates using 10µl reaction volume, with template- and reaction- master mixes being mixed in each well. Each plate therefore had 8 template mixes across 10 reaction mixes (8 target + 2 control), with each combination triplicated. Each well had 5ul SybrSelect, 0.5ul of each primer (final concentration 10nM), 2ul of water and 2ul of 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, and products were checked on an agarose gel and sequenced. To ensure analytical uniformity between plates, an inter-plate control was included consisting of a reference pool of the first 20 cDNAs. If any plate was >1 standard deviation away from the mean *GAPDH* cycle number it was repeated.

Relative gene expression was given as fold change and was calculated using the ddCT method, where the number of cycles needed to reach a threshold intensity value is compared between sample-pairs, having normalise template input to the reference gene intensity (*GAPDH*). For multi-cohort comparisons, the copies-per-thousand *GAPDH* value was calculated.

### 3.3.4 Immunohistochemistry

Immunohistochemistry was initially carried out on archival paraffin embedded tissue, either whole sections (21 OAC cases) or using a tissue microarray fabricated in house using local material (31 OAC cases, 5 with matched PNE, a gift from Prof R Goldin). The evidence for antibody selection was provided by The Human Protein Atlas (<http://www.proteinatlas.org/>). Antibodies were validated as recognising proteins of appropriate size by immunoblotting, and also predicted immunostaining specificity. Specific antibodies could be determined for four targets highlighted in the qPCR study: ALDH3A1 (HPA051150, Sigma), ALDH3A2 (HPA014769, Sigma), ALDH4A1 (1A12-A5, Abnova), and ALDH9A1 (HPA010873, Sigma). No ALDH1A3-targeting antibody could be convincingly optimised for immunohistochemistry, although the HPA046271 antibody detected appropriate bands in appropriate samples on immunoblotting.

Sections were de-wax, hydrated, and subjected to heat-induced epitope retrieval by microwaving at 900W for 10 minutes in citrate buffer (pH 6). Then, slides were incubated with hydrogen peroxide to block endogenous peroxidases, and non-specific antibody binding was inhibited using a protein block (Dako). Primary antibodies were then incubated in 1% bovine serum albumin overnight at 4°C, followed by a species-specific secondary antibody conjugated to horseradish peroxidase, and developed using diaminobenzidine (Sigma). All steps were followed by PBS washes.

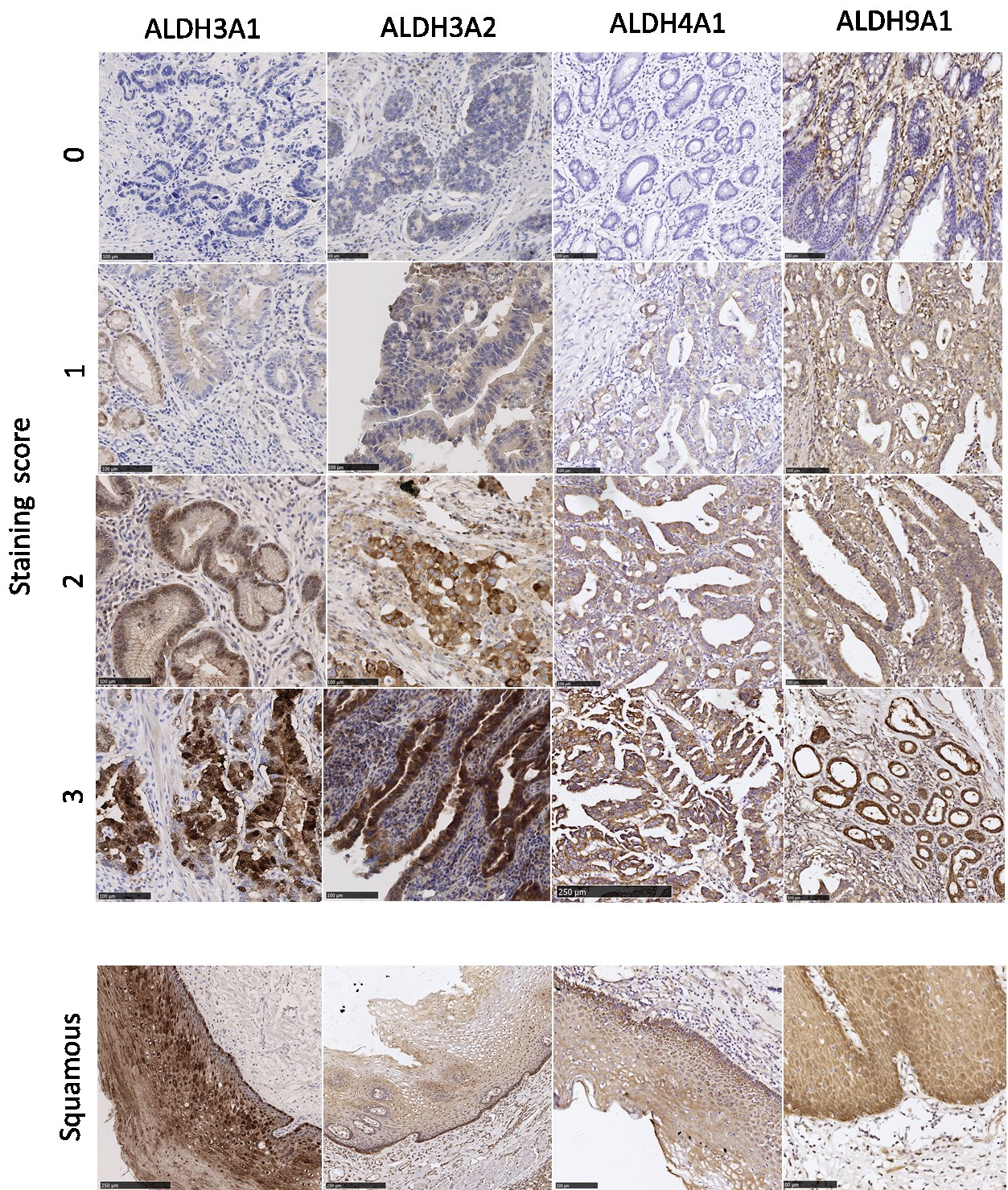
A second cohort of 360 OAC cases in triplicate across 18 tissue microarrays from 6 UK centres was a generous gift of the OCCAMS collaboration (P.I. Prof R Fitzgerald, University of Cambridge). This cohort was immunostained by Ms H Kudo (P.I. Prof R Goldin, ICL), using the Leica Bond™ system. Antigen retrieval solution 2 was used for both antibodies, using the same primary dilutions as for manual staining. Sections were imaged with a NanoZoomer (2.0-HT, Hamamatsu).

For whole mount sections SqT, BE and AdT regions were scored in five random high-powered areas according 0-3 on a basis of staining intensity (see Figure 17, 0 = no staining or <50% mild staining; 1 = > 50% mild staining, no moderate staining; 2 = any moderate staining, <50% strong staining; 3 = >50% strong staining) with a single average score per tissue type per patient used for comparative analysis (after several ALDH immunohistochemistry reports (119,121,122,232,233)). For correlation

to metadata, the immunoscore was dichotomised to negative or positive, with a cut-off of >1 being positive. For tissue microarrays, replicate cores for each patient were provided in quintuplet (ICL cohort) or triplicate (OCCAMS cohort). Scoring was on the same 0-3 basis and also undertaken by a second independent assessor (Dr F. Rosini, histopathology fellow). Disagreements in scoring were resolved with a consultant oesophago-gastric pathologist (Prof R. Goldin).

Typical scores for each ALDH isozyme are given in Figure 17. *ALDH3A1*, *-3A2*, and *-9A1* were scored on a cytoplasmic staining pattern, and *ALDH4A1* was scored on mitochondrial staining. Any nuclear staining for *ALDH3A1* and *-3A2* was also noted.





**Figure 17: Representative immunostaining of ALDH3A1, ALDH3A2, ALDH4A1 and ALDH9A1 in whole-mount OAC and squamous sections**

Bar represents 100  $\mu$ M.



### 3.3.5 Cell culture

The SV-40 immortalised normal squamocellular line HET-1A, and the oesophageal adenocarcinoma cell lines ESO-26, ESO53, FLO-1, KYAE-1, OE19, OE33, and SK-GT-4 were purchased from Public Health England and maintained in the recommended media and conditions. These were the only eight lines verified as being bona fide OAC lines in a recent investigation (see Table 12)(234).

**Table 12: Characteristics of *bona fide* OAC cell lines**

<i>Cell line</i>	<i>Derived</i>	<i>Patient characteristics</i>	<i>Tumour characteristics</i>	<i>Culture mode</i>	<i>P53 status</i>	<i>Tumorigenic in nude mice</i>
FLO-1	1991	68y white male	Distal adenocarcinoma	Adherent	Mutant	Yes
OACM5.1	2001	47y white female	Distal adenocarcinoma Line derived from lymph node metastasis	Adherent/ Suspension	Mutant	Yes
ESO26	2000	56y white male	Distal oesophagus and GOJ Nodal and distant metastases	Suspension	Mutant	Yes
KYAE-1	2001	60y asian male	Distal adenocarcinoma; pretreated with chemo-radiotherapy	Adherent	Mutant	Yes
OE33	1993	73y female	Distal adenocarcinoma arising in Barrett's; stage IIA, poorly differentiated	Adherent	Mutant	Yes
ESO51	2000	74y white male	Distal adenocarcinoma arising in Barretts	Suspension	Mutant	Yes
OE19	1993	72y male	GOJ adenocarcinoma; stage III moderately differentiated	Adherent	Mutant	Yes
SK-GT-4	1989	89y white male	Distal adenocarcinoma arising in Barrett's; T2N1, well differentiated	Adherent	Mutant	Yes

For primary culture of oesophageal keratinocytes, an optimised method derived from two recent publications was developed(235,236). A single 2-3mm<sup>2</sup> specimen of normal oesophageal mucosa was obtained at endoscopy or post-oesophagectomy and incubated overnight in Advanced DMEM (Gibco) supplemented with 10% fetal bovine serum, 2mM glutamine, 20mg/ml gentamicin, 250µg/ml amphotericin, 100 units/ml penicillin, and 100µg/ml streptomycin (all Sigma). The epithelial sheet was then dissected from the submucosa, washed at least ten times in PBS, minced, and incubated in collagenase III for one hour. The dissociated cells were gently passed through a 70 µm mesh and seeded onto plates coated with 0.01mg/ml fibronectin, 0.03mg/ml collagen, and 0.01 mg/ml albumin (all Sigma), and cultured in Advanced DMEM (Gibco), 2 mM glutamine (Sigma), antibiotics as above, and 10 µm Y27632 (a

Rho kinase (ROCK) inhibitor, Stem Cell Technologies). This reversibly immortalises cells in a 'basal' –like, state(236). After one week, gentamicin and amphotericin was removed and the cultures were mycoplasma tested. The ROCK inhibitor was removed one week before any phenotyping or perturbation experiment. Established models expanded rapidly and were amenable to long-term cold storage. However, continuous culture was not possible for longer than 4-6 months as the keratinocytes terminally differentiated and stopped dividing.

### **3.3.6 Copy number analysis**

Copy number analysis was undertaken using a qPCR approach on a DNA template (rather than cDNA). The *ALDH3A2* Taqman copy number assay was purchased from Life Technologies and used according to the manufacturer's instructions. The DNA template (50ng) was taken from patient DNA co-extracted with the RNA samples in the *ALDH* expression analysis study (following clean-up from Trizol using the DNeasy kit, Qiagen, according to the manufacturer's instruction).

### **3.3.7 Immunoblotting**

For protein experiments in 6-well format, cells were lysed in 250µl of RIPA buffer containing fresh 1x phosphatase inhibitor (Cocktail 3, Sigma) and protease inhibitors (Complete Mini™, Roche) at 4°C for 30 minutes. Residual debris was collected using a cell scraper. The lysate was then homogenised using a sonicator for two cycles of 10 seconds on ice, and then centrifuged at 20,000g for 10 minutes. The supernatant was collected and stored at -80°C until use. Protein was isolated by lysing cells in RIPA buffer (50mM Tris, 150mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP-40, 1x complete protease inhibitor cocktail (Roche), 1x complete phosphatase inhibitor cocktail 3 (Sigma)) and sonicated for 15 seconds to shear DNA.

Sonicated RIPA protein lysates were centrifuged for 10 minutes at 13,000g to pellet debris. The supernatant was then boiled for 3 minutes with 10% mercaptoethanol to reduce thiol bridges, and electrophoresed in 6-12% polyacrylamide gels containing sodium dodecyl sulphate according to standard descriptions (237). Primary structure polypeptides were then transferred onto polyvinylidene difluoride (PVDF, Biorad) membranes using a semi-dry blotting system (Biorad) according to the manufacturer's instructions. Membranes were then blocked in 5% non-fat milk/tris-

buffered saline/0.01% Tween 20 (M-TBST; all Sigma), following by primary antibody hybridisation overnight at 4°C. Unbound antibody was then removed with three washes in TBST, followed by second hybridisation for one hour to species-specific IgG conjugated to horse radish peroxidase. Binding was then visualised with enhanced chemiluminescence (Pierce) and hyperfilm (Amersham).

For keratinocyte validation, clinically-validated mouse monoclonal antibodies targeting keratinocyte-specific keratins CK5/6 (Clone D5/16 B4; Merck), the nuclear factor p63 (4A4; Abcam) the mesenchymal factor Vimentin (V9; Sigma), and the epithelial marker E-cadherin (NCH38; Dako) were utilised. In addition, the anti-phospho-Histone H<sub>2AX</sub> (Ser139) (20E3) and anti-p53 antibody (DO-1) were used to assess DNA health (Cell Signalling Technologies). The hydroxynonenal-protein adduct (HNEJ, Abcam, 1:100) antibody was used to detect protein carbonylation. The same *ALDH* isoenzymes were used as described in Section 3.3.4, in addition to *ALDH2* (ab108306, Abcam). The bicinchoninic acid assay (Sigma) was used to quantify protein concentration. Typically 30µg was loaded into 5mm wells; equality of protein loading was assessed using an antibody directed against a-tubulin (CST, multi-cell line experiments) or b-actin (CST, single cell line experiments).

### **3.3.8 Immunofluorescence**

Immunofluorescence microscopy (IFM) was used to immunophenotype in vitro models. Cells were grown on coverslips to ~30% confluence and fixed in 4% paraformaldehyde at 4°C for 3 hours. Cells were washed in saponin and stained for one hour with antibodies to CK5/6, E-cadherin, p63, and *ALDH3A1*, using optimised concentrations. Visualisation was actuated using species-specific secondary antibodies conjugated to fluorophores (either goat anti-mouse IgG (Alexa 488 conjugated) or donkey-anti rabbit IgG (Alexa 546 conjugated), both Life Technologies), and visualised using an SP5 confocal system (Leica, UK).

### **3.3.9 Statistics**

For comparing expression data from two cohorts, the Mann-Whitney U-test was selected as the data was considered non-parametric. For comparing more than two expression datasets the Kruskal-Wallis test was used with Bonferroni correction. Analyses of *ALDH* expression in disease strata were undertaken in pre-determined

clinically relevant subgroups if the expected number of cases was in each group were ten or more, and expected to be relevant to aldehyde metabolism. Thus the following groups were selected: prior neo-adjuvant chemotherapy (yes/no), local invasion (<T3, >T2), nodal status (N0, N>0), differentiation (well/mod/poor), taking proton pump inhibitor (yes/no). In the expanded analysis using OCCAMS data, these subgroups were appropriately expanded (see Results). Interactions were tested with chi-squared tests. Twelve cases did not have complete survival data, leading to a total of 400 with matched survival, metadata and expression data. Kaplan-Meier survival curves were used to fit survival data to dichotomised expression indices. A Cox proportional hazards model was fitted to test the independence of prognostic variables. These were performed in SPSS (version 23, IBM) or Prism (version 7, Graphpad)



## 3.4 Results

### 3.4.1 Candidate discovery using bioinformatics

The geneset set 'Gene Ontology' (GO v5) was selected to collapse the micro-array as it features well-defined metabolic genesets. In all tested datasets, the "aldehyde oxidoreductase acting on the aldehyde or oxo group of donors" geneset was among the most prominently enriched in squamous mucosa compared to adenocarcinoma tissue, featuring 2<sup>nd</sup> of 1005 in the two microarray analyses (see Table 13). This geneset incorporates the 27 genes that encode proteins which act to oxidise aldehydes to carboxylic acids. A later analysis using RNA-seq data from the OCCAMS collaboration (adenocarcinoma, n = 27; squamous mucosa, n = 5) found a similar result, although the geneset featured less prominently, and none of the geneset differences reach discovery thresholds. Other genesets which ranked highly were generally those characteristic of a squamous epithelium, including terminal differentiation and epithelial markers, and cell-cell connectivity (see Table 14).

**Table 13: Summary of GSEA findings in three datasets**

Study	<i>n</i>	Tissue isolation	Profiling technique	AOR Rank (FDR <25%)	q =
Wang	OAC 21 Norm Sq 19	Laser capture	Microarray	2 <sup>nd</sup> (70)	<0.0001
Kim	OAC 69 Norm Sq 28	Macrodissection	Microarray	2 <sup>nd</sup> (2)	0.008
OCCAMS	OAC 27 Norm Sq 5	Macrodissection	RNA-seq	23 <sup>rd</sup> (0)	NS

AOR, aldehyde oxidoreductase geneset; FDR, false discovery rate

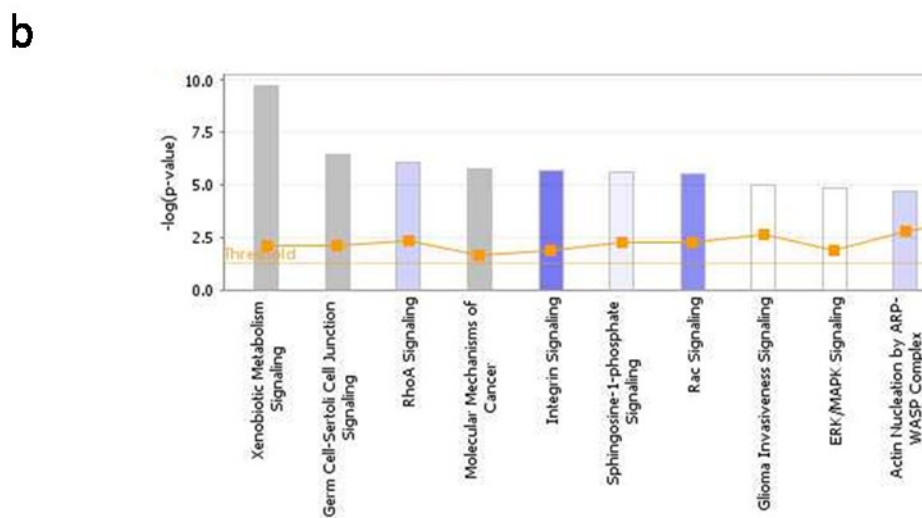
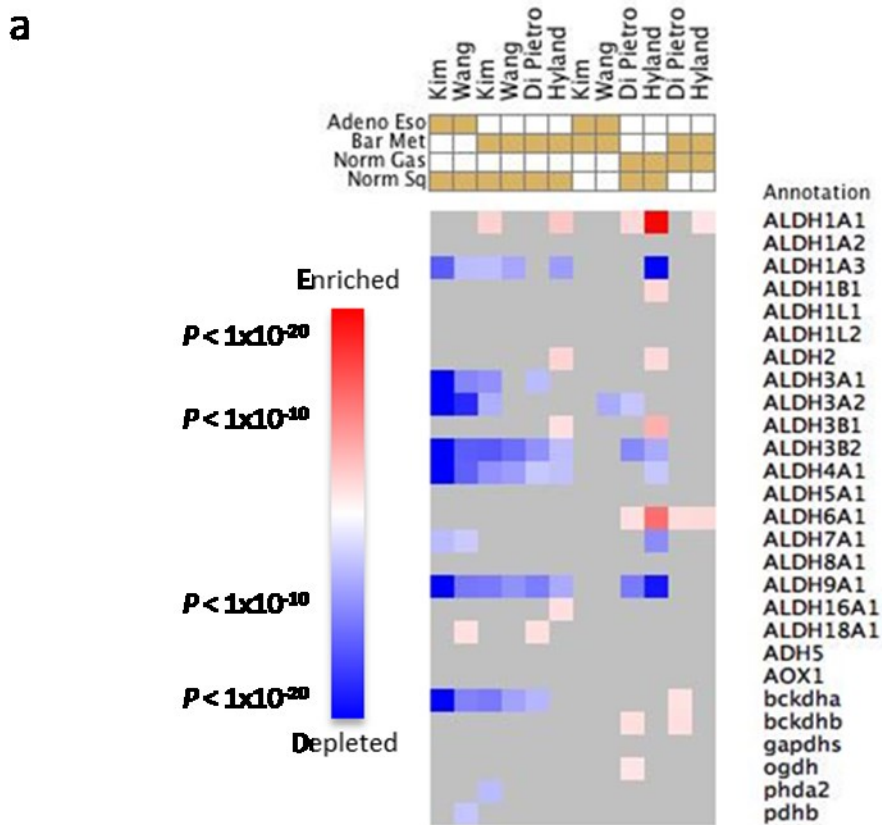
**Table 14: The ten highest ranked GO v5 geneset discoveries from GSEA of Wang et al**

Rank	Geneset details (Gene Ontology classification)	FDR q value
1	Tissue development	<0.0001
2	Oxidoreductase acting on the aldehyde or oxo group of donors	<0.0001
3	Ectoderm development	<0.0001
4	Epidermis development	<0.0001
5	Protein binding bridgeing	<0.0001
6	Vitamin metabolic process	<0.0001
7	Endosome	<0.0001
8	Oxidoreductase activity acting on the aldehyde or oxo group of donors (NAD/NADP binders)	<0.0001
9	Intercellular junction	<0.0001
10	Morphogenesis of an epithelium	<0.0001

Univariate analysis of each AOR geneset constituent was performed to identify the key drivers of the phenotype, across all datasets reporting Barrett's or OAC transcriptomic data with a relevant tissue control. As seen in Figure 18a, there were consistent and highly significant differences in several isoenzymes of ALDH, as well as a more specialised AOR involved in glycolysis. In particular, *ALDH1A3*, *-3A1*, *-3A2*, *-3B2*, *-4A1*, and, *9A1* were convincingly and consistently suppressed in adenocarcinoma compared to squamous mucosa (all  $P < 1 \times 10^{-10}$ ), and all but *-3A1* and *-3A2* were also suppressed in Barrett's metaplasia and gastric tissue. Between Barrett's and adenocarcinoma, the only isoenzyme to pass significance thresholds was *ALDH3A2* (expression reduced in cancer,  $P < 1 \times 10^{-11}$ ). *ALDH1A1* was the only isoenzyme in which expression was enriched in non-squamous tissue, in particular in Barrett's compared to normal squamous mucosa.

*ALDH3B2* was excluded from further analysis as it has a stop in codon 17 and thought to be a pseudogene, as was *BCKDH*, as it has a highly specialised role in glycolysis and is not thought to contribute to wider aldehyde metabolism. Owing to its association with Barrett's adenocarcinoma, *ALDH1A2* was added to the validation cohort, as was *ALDH2*, which is associated with OSCC when mutated. Thus, eight candidates were identified for experimental validation.

To extend these findings, a second approach was undertaken to collapse univariate gene expression to functionally relevant groups, using IPA. As seen in Figure 18b, the most significantly altered pathway was "xenobiotic metabolism signalling", of which various ALDHs form a significant part.



**Figure 18: ALDH candidate discovery and IPA analysis.**

Panel a. Candidate discovery and initial validation. Panel A: significance of enrichment or depletion in dichotomous univariate analysis of candidate *ALDH* expression in four different tissue types, parsed from microarray expression datasets as indicated. Panel b. Top ten most significantly modulated core analysis pathways in Ingenuity Pathway Analysis, using Wang et al dataset.

### 3.4.2 Patients

Clinico-demographic features comparing the phenotyping cohorts are given in Table 15. The groups were age-sex matched, and generally tumour characteristics were similar (tendency to T>2 N>0 presentation). There were significant differences in the tumour invasion depth and nodal metastases, and a non-significant difference in neo-adjuvant chemotherapy use, between cohort 1/2 and 3. Additionally, the five year survival was significantly poorer in the 2<sup>nd</sup> IHC cohort.

**Table 15: Clinico-demographic features of the *ALDH* expression cohorts**

	qPCR cohort	1st IHC cohort	2nd IHC cohort	<i>P</i> =
<b><i>n</i></b>	67	52	360	
<b>Age*</b>	64	63	66	
<b>Male</b>	53 (79%)	40 (76%)	288 (80%)	0.872
<b>Local stage</b>				<0.001
T1	13 (19%)	12 (23%)	25 (7%)	
T2	10 (15%)	9 (17%)	67 (19%)	
T3	34 (50%)	26 (46%)	251 (70%)	
T4	10 (16%)	5 (10%)	17 (5%)	
<b>Nodal stage</b>				<0.001
N0	24 (36%)	22 (42%)	106 (30%)	
N1	32 (48%)	10 (19%)	238 (66%)	
N2/3	11 (16%)	18 (35%)	15 (4%)	
<b>Differentiation</b>				0.008
Well/Moderate	35 (52%)	29 (56%)	136 (38%)	
Poor	32 (48%)	23 (44%)	224 (62%)	
<b>Post-chemo sample</b>	37 (55%)	26 (50%)	154 (43%)	0.131
<b>5 yr survival</b>	n/a	28 (54%)	86 (24%)	0.002

All *P* values calculated with chi-squared tests. Staging characteristics according to TNM 7

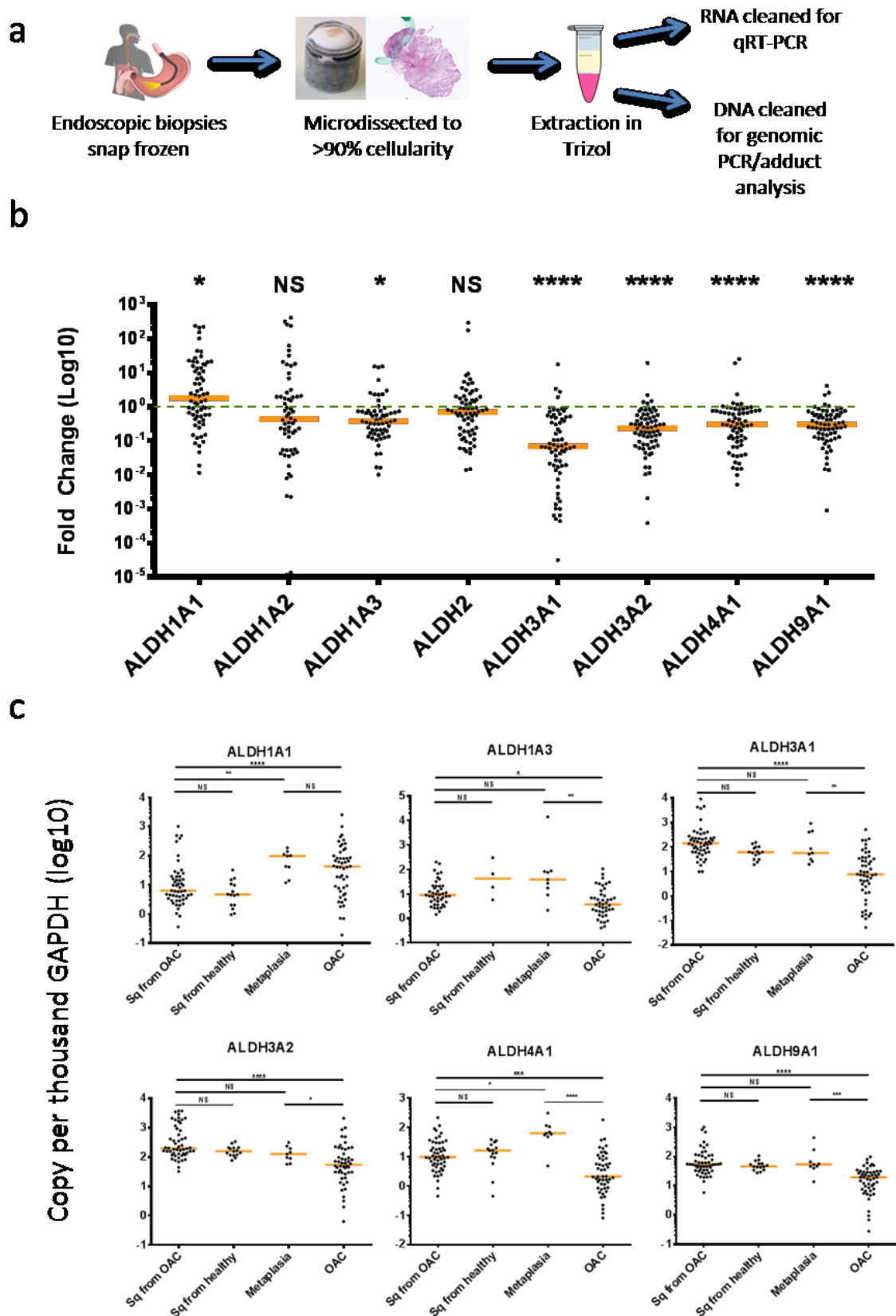
### 3.4.3 Candidate *ALDH* expression in at mRNA level – qPCR study

The sample processing scheme is provided in Figure 19a. Of 182 samples prepared, 26 (14%) required microdissection to achieve appropriate cellularity, and 12 (7%) contained no target cell-type and required replacement.

Consistent and significant decreased expression was noted in five *ALDH* subtypes (*ALDH1A3*, *-3A1*, *-3A2*, *-4A1*, and *-9A1*) in AdT compared to matched SqT. This ranged from a *median* 10 to 40 fold (90-98%) reduction in relative RNA message (Figure 19b, note  $\log_{10}$  transformed axis). This pattern of expression was identical to that observed in the microarray expression libraries. Of all, *ALDH3A1* was most strikingly suppressed – in some cases there was a thousand-fold reduction in the expression of this gene. One *ALDH* isoform, *ALDH1A1*, was significantly increased, albeit in a wide distribution of expression.

*ALDH1A2* transcripts were frequently undetectable or required high (>37) CT values with wide standard deviations to reach intensity thresholds in both SqT and AdT. Such high cycling values must be treated with caution as accuracy is degraded. It was considered that this gene is not expressed in oesophageal tissues, and it was thus excluded from further studies.

*ALDH* expression was additionally quantified in SqN (i.e. squamous from healthy volunteers) and Barrett's metaplasia (BE) samples (see Figure 19c). There were no differences in *ALDH* expression in squamous mucosa from the normal and malignant oesophagus. In keeping with earlier bioinformatics analyses, expression of *ALDH1A1* was also increased in BE compared to Sq samples, and expression of *ALDH3A1* and *-3A2* were not different. Expression of *ALDH4A1* and *9A1* were both significantly higher in BE compared to AdT. For *ALDH4A1*, expression was even higher in BE than SqT, which was inconsistent with in the findings in silico.



**Figure 19: qPCR biopsy study**

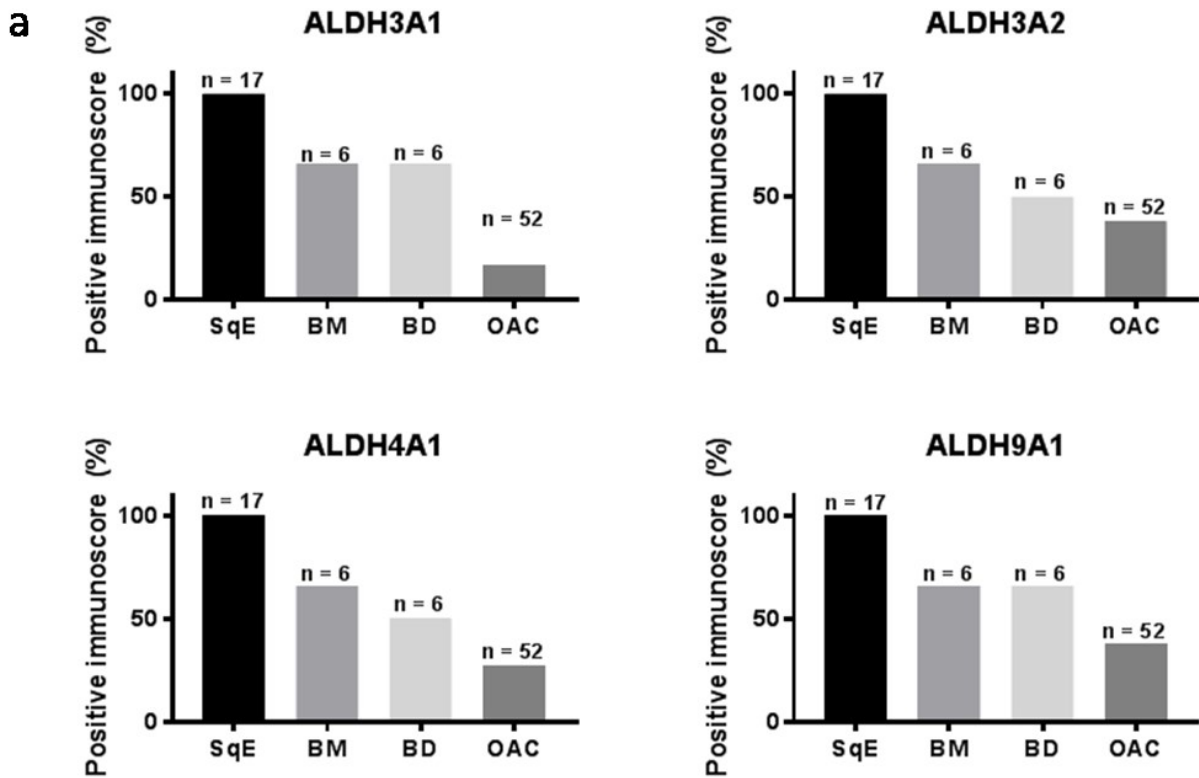
Panel A. Schematic of sampling strategy Panel B. Fold change in RNA transcripts between matched proximal squamous mucosa and adenocarcinoma tissue from 67 patients with OAC. Green line indicates normalised value for squamous mucosa. P-values calculated by copies per thousand GAPDH (CPKG) analysis. Panel C. CPKG *ALDH* quantitation of indicated tissue types (SqT  $n = 67$ , SqN = 10, Metaplasia = 10, OAC = 67). \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$

### 3.4.4 Candidate *ALDH* expression at protein level – immunohistochemistry studies

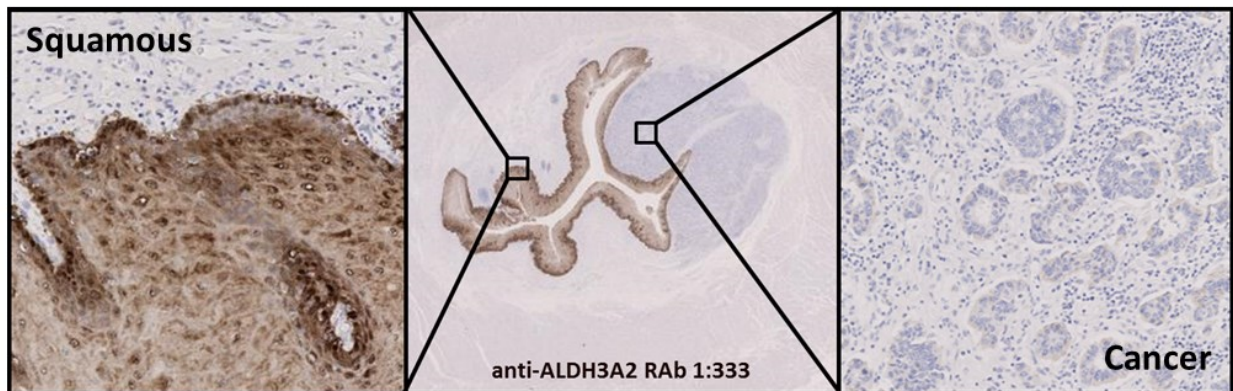
***Tissue expression patterns in Imperial discovery cohort*** Findings at the RNA level were extended using immunophenotyping of archived paraffin-embedded post-surgical tissue with the following objectives: (i) to verify *ALDH* protein expression (ii) to establish topological and subcellular expression patterns (iii) to correlate expression to survival and other metadata (which was not yet possible for the prospectively collected biopsy cohort).

In keeping with *in silico* and PCR data, all oesophageal squamous epithelia (SqE) broadly expressed the measured *ALDH* isoenzymes (see Figure 20a). Particularly strong staining for *ALDH3A1* and *-3A2* was noted in the basal squamous layer, followed by reduced expression in the next layers, followed by increased expression (see Figure 20b). The absolute basal, middle and top layers also showed frequent nuclear staining for the isoenzymes. In contrast, there was a heterogeneous expression for adenocarcinoma with only a minority showing moderate or strong immunostaining (*ALDH3A1*, 8.9%; *ALDH3A2*, 27.1%; *ALDH4A1* 26%, *ALDH9A1* 38%). Typically, both normal and tumour-adjacent mesenchymal, lymphatic and vascular cells did not express these enzymes (see Figure 20a). Representative staining for each candidate is given in Figure 17 and a whole mount example of *ALDH3A2* with both SqE and AdT areas on the same section is given in Figure 20b.

The embedded tissue blocks for this analysis were selected by the presence of OAC, however, in ten whole mount cases, areas of non-dysplastic (six cases) and dysplastic Barrett's epithelium (six cases) were available for comparison. There was moderate or strong expression of all *ALDH* isoenzymes in 3 or 4 of the 6 cases of BE and BD (see Figure 20a). Broadly, there was similar expression between both metaplasia and dysplastic columnar change in the oesophagus, particularly for *ALDH3A1* and *ALDH3A2*. The whole mount sections contained additional information regarding *ALDH* expression and malignant behaviour. These include evidence of (i) no evidence of expression gradients in peritumour tissue (ii) strong SqT and BE *ALDH3A2* staining immediately adjacent to weak AdT staining (iii) reduced expression of *ALDH3A2* in a lymph node metastasis compared to the index tumour in a single slide (see Figure 21). Important, differential staining between tissue-types was extremely crisp, implying that transregulation is cell-autonomous rather than a reaction to microenvironment stimuli.



**b**



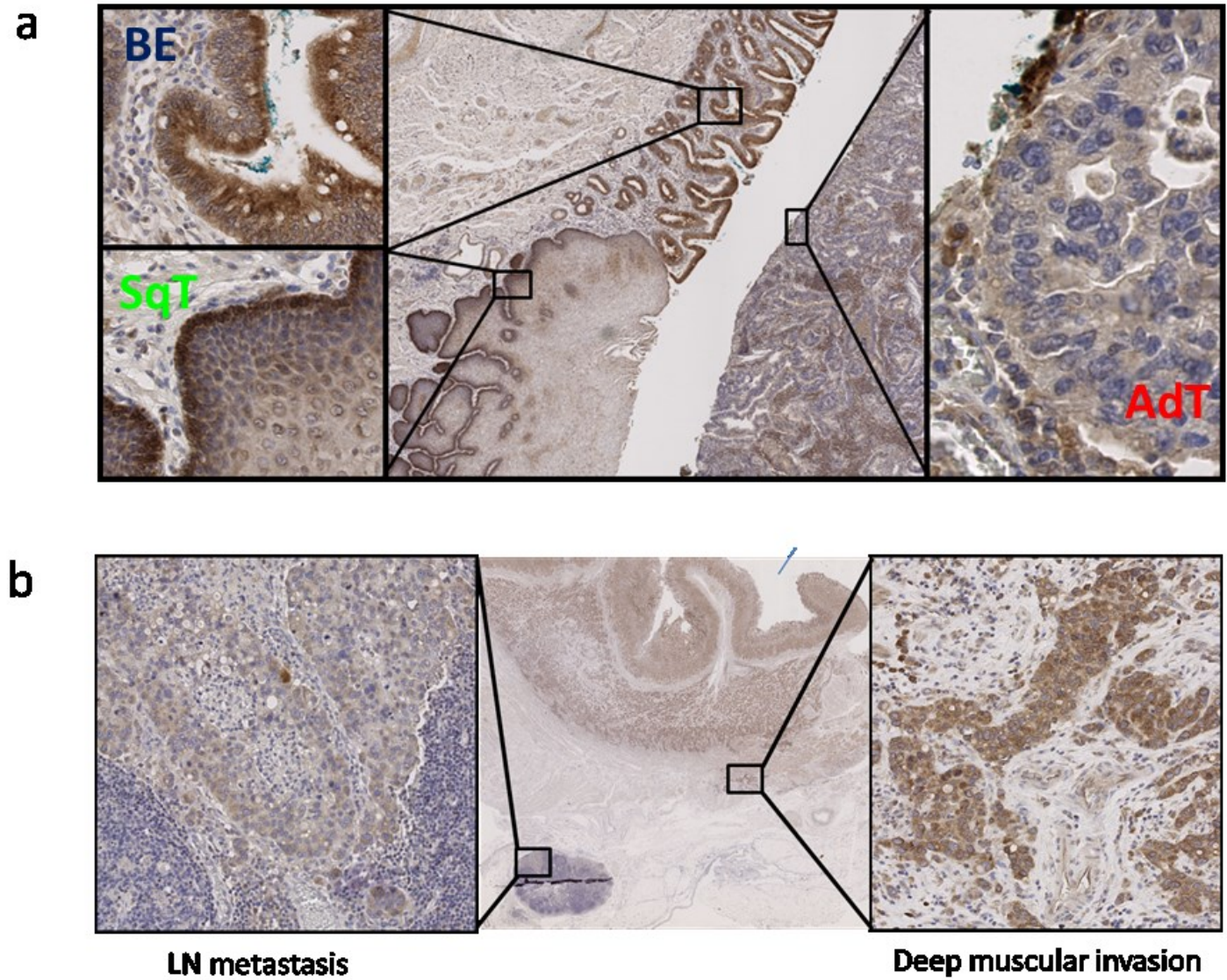
**Figure 20 Candidate *ALDH* expression in the discovery immunohistochemistry cohort (Imperial patients, n = 52)**

Panel a; An axial section through an oesophagus containing a submucosal adenocarcinoma, and stained for *ALDH3A2*. Panel b; Candidate *ALDH* expression patterns across different tissues in the combined cohort.

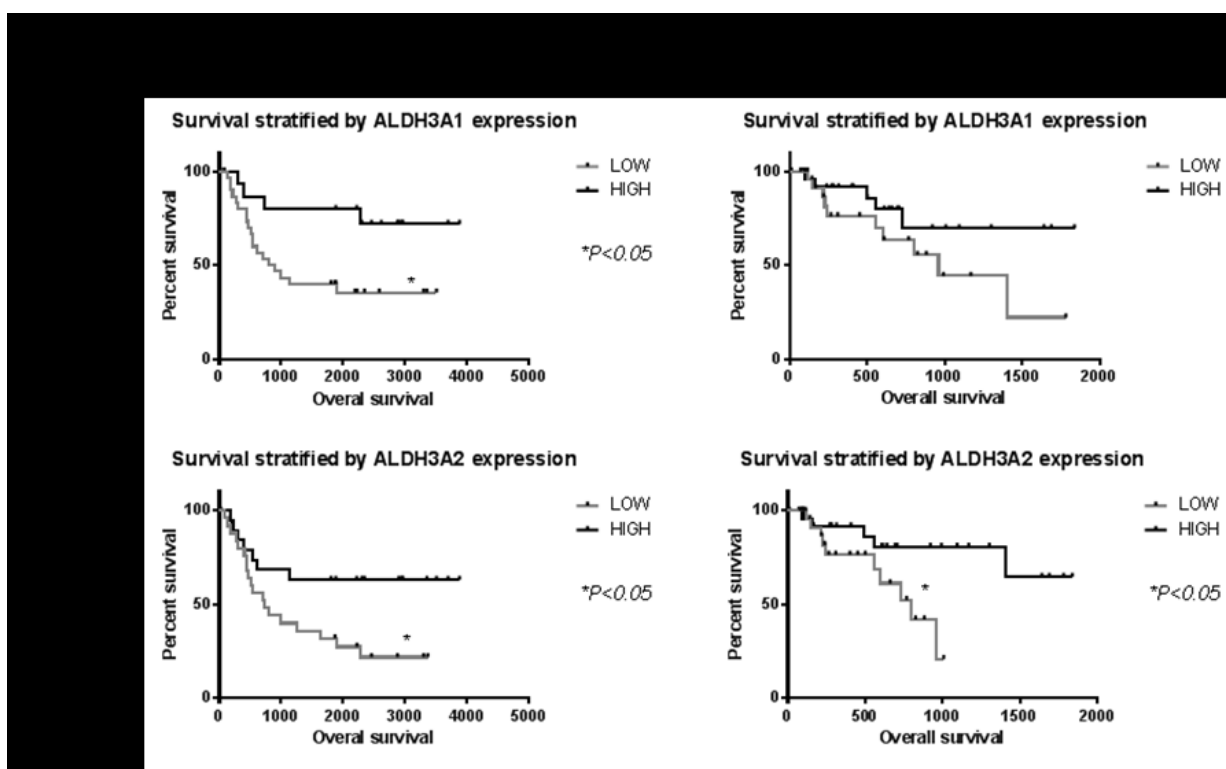


**Figure 21: *ALDH3A2* expression loss occurs with progression**

Panel a, whole mount section showing strong staining for SqT (basal layer) and BE and weak staining for AdT within close proximity. Panel b, whole mount section showing differential *ALDH3A2* expression between deeply invasive tumour and a metastatic lymph node deposit, associating loss of expression with progression.

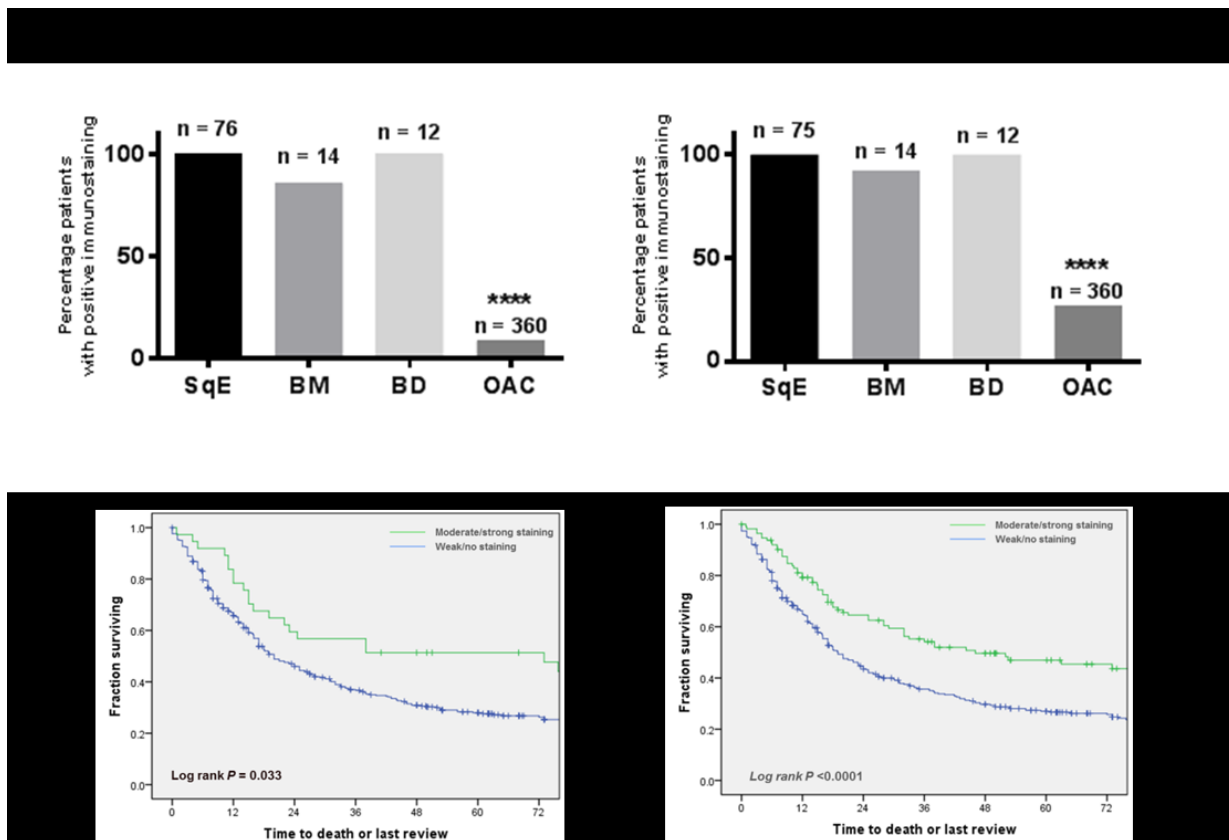


**Expression studies supervised by clinical metadata in the discovery (ICL cohort)** Molecular studies using archived material permits subgroup analyses based on long-term clinical outcomes. In the Imperial IHC cohort, there was no difference in *ALDH* expression in any pre-determined clinical subgroup, including a history of neoadjuvant chemotherapy (data not shown). However, stratification of expression to survival revealed significantly poorer overall survival in tumours with low expression of *ALDH3A1* and *-3A2* (see Figure 22, left column). These findings were verified using RNA-sequencing (RNA-seq) data from The Cancer Genome Atlas. Candidate *ALDH* RNA-seq fragments-per-kilobase-per-megabase (FPKM) read counts were used to stratify matched survival. The FPKM values were simply partitioned into 'high' and 'low' cohorts using the median as a cut-off. This strategy yielded the same significantly poorer survival pattern for low expressors of *ALDH3A2*, and a non-significant pattern of poorer survival for *ALDH3A1* (see Figure 22).



**Figure 22 Kaplan-Meier survival curves fitted to *ALDH* expression using ICL immunophenotyping data & The Cancer Genome Atlas RNA-seq data** Significance assessed with Mantel-Cox test ( $P < 0.05$ ). (Imperial IHC, *ALDH3A1*-high  $n = 5/52$ , *ALDH3A2*-high  $n = 14/52$ ; TCGA  $n = 56$ , cohorts divided by median expression value)

**Expression studies in the OCCAMS TMA validation cohort** The stability of these survival patterns suggested that the metadata association studies in the Imperial cohort were likely underpowered, and so a second IHC cohort was analysed with significantly expanded numbers. Tissue microarrays from a UK national resource collaboration for OAC research were thus stained for these genes (OAC  $n = 360$ ; only *ALDH3A1* and *ALDH3A2* were selected, owing to the survival associations). For *ALDH3A1*, only 20 of 360 cases (5.6%) showed moderate or strong expression, in contrast to nearly all SqT and BE; for *ALDH3A2*, 73 of 360 (20.2%) showed moderate or strong expression, in contrast to nearly all SqT and BE (see Figure 23). This clear expression loss in OAC for both markers was highly significant (two-tailed  $\chi^2$  test,  $P < 0.0001$ ), compared to SqN, BM, and BD. These results are summarised in Figure 23. In this third survival cohort, the same significantly poorer survival was noted for weak/absent expressors of both *ALDH3A1* (log rank,  $P < 0.033$ ) and *ALDH3A2* (log-rank,  $P < 0.001$ )



**Figure 23: Summary of ALDH3A1 and 3A2 expression patterns and survival correlation in the validation OCCAMS TMA cohort ( $n = 360$ )**

**Table 16: Associations of ALDH3A1/2 staining to metadata in 2<sup>nd</sup> IHC cohort (OCCAMS)**

	ALDH3A1 staining			ALDH3A2 staining		
	-ve	+ve	P =	-ve	+ve	P =
<i>n</i>	340	20		287	73	
<b>Patient</b>						
Age (median years)†	65.0	62.0	0.437	65	64	0.437†
Pre-op chemotherapy (%)‡	43%	22%	0.088	37%	55%	0.008**
Sex (Male, %)‡	81%	70%	0.244	75%	78%	0.424
<b>Tumour</b>						
Differentiation (%)			0.453			0.039
<i>Well</i>	16%	10%		16%	14%	
<i>Moderate</i>	22%	20%		19%	36%	
<i>Poor/Undifferentiated</i>	62%	70%		65%	51%	
Siewert (%)			0.59			0.593
<i>Not recorded</i>	6%	5%		6%	4%	
1	77%	75%		78%	74%	
2	12%	15%		11%	16%	
3	5%	5%		5%	5%	
Post op T-Stage(%)			0.191			0.006**
1	6%	15%		5%	16%	
2	18%	25%		18%	21%	
3	70%	60%		74%	55%	
4	5%	0%		4%	7%	
Post op N-Stage (%)			0.576			0.101
0	29%	40%		27%	40%	
1	67%	60%		69%	56%	
2	1%	0%		1%	4%	
3	3%	0%		3%	0%	
Number of positive nodes (median)†	5.0	3.0	0.948	5.1	2.8	0.005**
Post op M-Stage (% M1)‡	5%	10%	0.132	5%	3%	0.747
Vascular or Neural invasion (+ve)‡	52%	40%	0.435	53%	46%	0.395
IHC +ve for other ALDH3A (%)	21%	45%	.023*	20%	40%	.033*

All P values generated with chi squared tests, apart from †Mann-Whitney U-test, and ‡Fisher's exact test.

\*P <0.05, \*\*P <0.01, \*\*\*P <0.001

In the OCCAMS dataset, low expression of *ALDH3A2* was significantly associated with deeper local invasion, more positive lymph nodes, poorer disease differentiation, and having not had pre-operative chemotherapy (see Table 16). Given that these factors can all contribute to survival, a Cox proportional hazards model was fitted to assess whether *ALDH3A2* expression was an independent predictor of overall survival or confounded by these other factors. As can be seen in Table 17, low expression of *ALDH3A2* independently predicted death, was associated with a 64% increased likelihood of mortality ( $P = 0.01$ ). This effect was equal to that levied by being >70 years old, although not as profound as lateT or N stage. There was also significant co-suppression of the two genes, although these were relatively weak associations

Even in this expanded cohort, *ALDH3A1* expression was not associated with any baseline characteristic apart from co-suppression with *ALDH3A2*; this may be explained by relatively rarity of *ALDH3A1* positive tumours, and thus under-populated subgroups.

**Table 17: Independent predictors of death in OAC identified using a Cox model**

	Odds Ratio	95% C.I.	P
<b>Age (&gt;70y)</b>	1.61	1.21 - 2.13	0.001
<b>ALDH3A1-low</b>	1.09	0.55 - 2.15	0.81
<b>ALDH3A2-low</b>	1.64	1.13 - 2.39	0.01
<b>T2</b>	2.01	0.76 - 5.30	0.16
<b>T3</b>	4.70	1.87 - 11.82	0.001
<b>T4</b>	9.38	2.43 - 36.23	0.001
<b>Poor differentiation</b>	1.28	0.75 - 2.19	0.37
<b>N1</b>	2.30	1.59 - 3.31	<0.001
<b>N2</b>	3.58	1.55 - 8.29	0.003
<b>N3</b>	13.42	3.92 - 45.99	<0.001

### 3.4.5 *ALDH* expression in OAC *in vitro* and *in vivo* models

Primary keratinocyte cultures were validated as non-keratinising squamous epithelial monolayers by checking keratinocyte markers (cytokeratin 5/6), epithelial markers (p63, E-cadherin), and the absence of a mesenchymal marker (vimentin, all see Figure 24a). In addition, the genetic stability of these lines was verified by the absence of *TP53* expression and H<sub>2</sub>A<sub>X</sub> phosphorylation. Epithelial-to-mesenchymal transition (EMT) status was also defined for the OAC lines as this gives an understanding of disease stage of the model. ESO26, OE33, OE19 and SK-GT4 expressed high E-Cadherin and low vimentin, indicating an epithelial phenotype. In contrast, OACM5.1, and Eso51 expressed low E-cadherin and high vimentin, indicating mesenchymal status. FLO1 expressed both markers and was considered intermediate.

The purity of the keratinocyte cultures was further verified using fluorescent immunostaining of a candidate culture (199N, see Figure 24b). Cultures were visual verified as pure, as all cells stained strongly for cytokeratin 5/6 and for epithelial markers. Notably *ALDH3A1* seemed to localise the nuclei or peri-nuclear cytoplasm of this keratinocyte culture.

All tested keratinocyte cultures broadly expressed *ALDH* isozymes (see Figure 24a, right immunoblot). This was in contrast to the commercial OAC lines, in which expression was highly variable. FLO-1 and OACM5.1 did not express any *ALDH* isozymes except *ALDH9A1*, and were thus considered “*ALDH-lo*”. Eso51 and OE19 expressed all of the isozymes, including some over-expression of *ALDH3*, and were considered *ALDH-hi* (*ALDH3A1* is known to be amplified in COSMIC data for OE19 (238)). *ALDH3A1* and *-3A2* were co-expressed in the same cell lines. Additionally, *ALDH2* and *-9A1* appeared to be co-expressed.

Finally all the OAC cell lines seemed to have a relatively higher burden of hydroxynonenal-adducted proteins, compared to the four tested keratinocyte cultures.



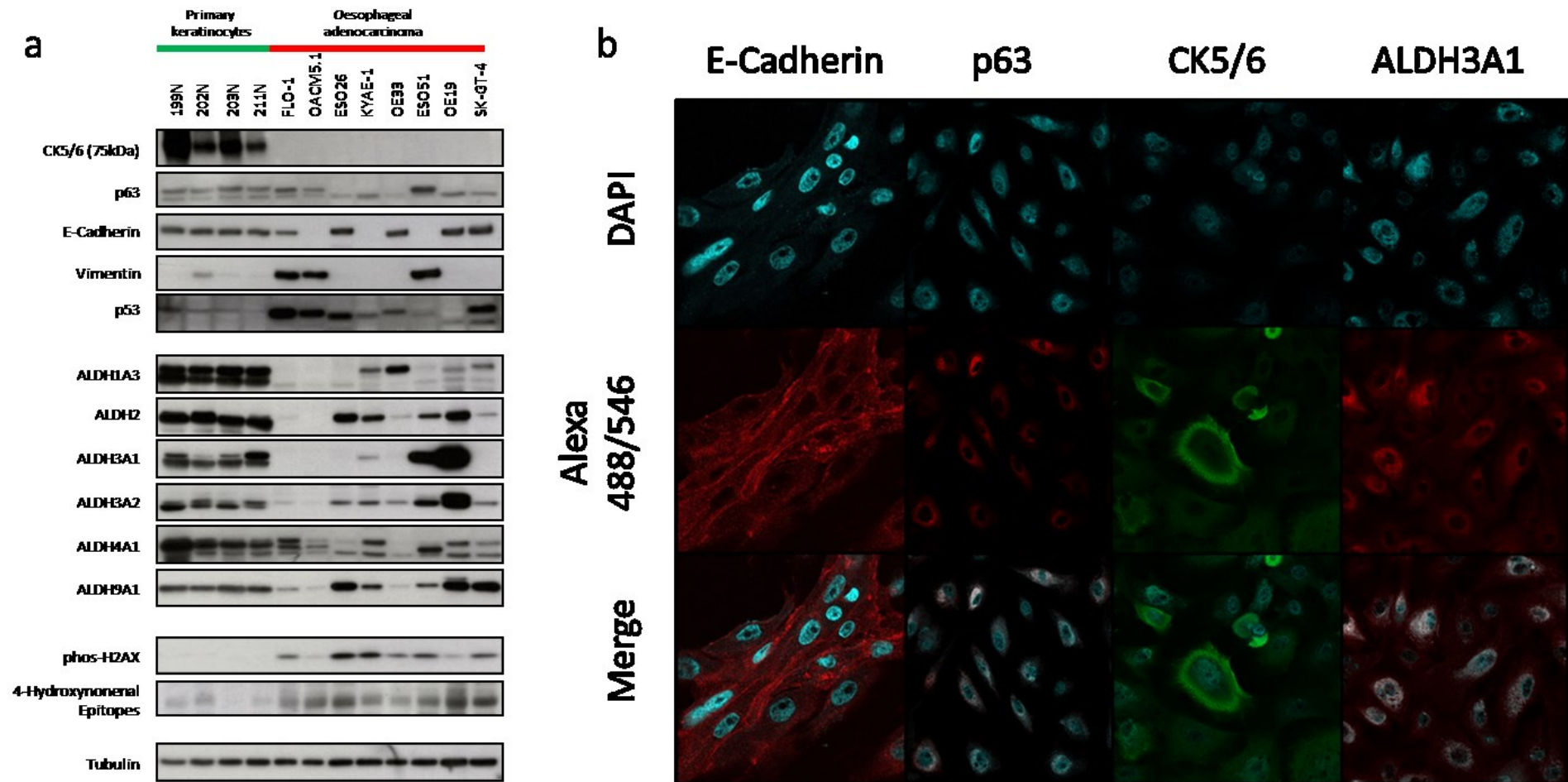


Figure 24: *ALDH* expression in validated oesophageal models of normal squamous and adenocarcinoma cells.

Panel a. oesophageal cell panel immunoblot results (Top cluster), Keratinocyte, epithelial, and mesenchymal markers (Middle) *ALDH* candidate isozymes (Bottom) genotoxicity surrogates & loading control. Panel B, Representative immunofluorescence studies of cultured keratinocytes (x40)

### 3.4.6 Regulation of *ALDH3A1* and *-3A2* and association with 17p LOH

To search for regulatory clues regarding loss of *ALDH3A* activity, legacy datasets with multi-parametric molecular phenotyping were systemically reviewed. A 'surgical sieve' approach was taken, building on the hypothesis that loss of expression can either be mediated by genetic, epigenetic or trans-regulatory phenomena. Searching a recent OAC sequencing catalogue (33) revealed that both *ALDH3A1* and *ALDH3A2*, as with other *ALDH* isozymes, were seldom mutated in OAC (Figure 25a), and rarely in coding regions. Similarly, analysis of TCGA 450K methylomic data suggested that *ALDHs* are rarely hypermethylated in cancer (see Figure 25b). Given that 17p is frequently targeted for loss-of-heterozygosity events in OAC (207,208,239), it was hypothesised that copy number changes related to macro-chromosomal events could contribute to *ALDH3A1/3A2* expression loss. To test this, single nucleotide polymorphism (SNP) data from the TCGA was aggregated using the GISTIC2.0 algorithm, to provide a high level overview of regional amplifications and deletions. As summarised in Figure 25c, >60% of patients had 17p copy loss, which included the *ALDH3A* locus at 17p11.2. To illustrate how structural variants may influence *ALDH3A2*, the expression of this gene was correlated to all other genes using TCGA RNA-seq data, and then ranked by Spearman's coefficient (see Figure 25d). Of the top 20 most correlated genes, 11 mapped to 17p11.2, and all were telomeric of the candidates of interest, strongly indicating that a macro- event was driving expression changes at this locus. Furthermore, correlating *ALDH3A2* CNV status with expression using TCGA data in cBioportal (240), there was a trend towards copy status affecting transcript number (see Figure 25e).

An experimental validation was attempted by measuring *ALDH3A2* copy number in DNA co-extracted for the endoscopic biopsy *ALDH* expression study. In modest cohorts of tumour and normal tissue, that was no significant difference in *ALDH3A2* copy number (see Figure 25f).





### 3.5 Discussion

In this Chapter, a screen for potential genetic influences on OAC metabolism revealed that loss of aldehyde detoxification was a defining phenotype. For highly discriminating genes, these patterns were validated in three cross-validating experimental datasets using two complementary molecular techniques. Expression loss of five *ALDH* isoenzymes – with a median expression fold change of 90-98% at the mRNA level – was a defining feature of AdT and similar effects were seen in other non-squamous lower oesophageal tissue. In one informatics and two experimental datasets collectively involving matched expression and survival data on 456 OAC patients, the majority had low *ALDH3A1* and *-3A2* expression, and this was repeatedly associated with poorer survival. In particular, both *in silico* and in validation, there was significantly less *ALDH3A2* expression in OAC compared to metaplasia. This may indicate a novel tumour suppressor function for these genes in OAC, or that this is ‘passenger’ expression loss is a passenger in a wider oncosuppressor loss-of-function event (241). Database reviews suggested that copy number changes and/or signalling likely coordinate *ALDH* trans-repression, and in fact low expression of *ALDH3A* genes may be related to a long-known 17p loss-of-heterozygosity event in OAC tumorigenesis (208). Finally, further evidence for *ALDH* expression flux was provided by immunoblotting oesophageal cell lysates. This revealed strong, consistent and uniform expression of all tested *ALDH* isoenzymes at expected molecular weights in oesophageal keratinocyte cultures. This is in keeping with an extensive literature regarding their protective adaptations (see 1.1.2), and in keeping with *ALDH* expression data in keratinocytes in other organs (242,243). *ALDH* expression in OAC lines was heterogeneous, and seemed to be associated with enhanced hydroxynonenal-protein adduction and genomic instability.

The predominant strengths of this Chapter are as follows:

- (i) Through unbiased, complementary informatics, identification that reduced aldehyde detoxification gene expression is a defining feature of OAC (and to a lesser extent, Barrett’s tissues) compared to normal squamous mucosa. (Table 13 and Table 14)
- (ii) Rationalisation of candidates based on a systematic re-analysis of archived expression data (Figure 18)
- (iii) Extensive molecular validation of these findings using two molecular techniques in whose strengths counteract the weaknesses of the other (Figure 19-21)

- (iv) Validation of the initial ICL IHC findings in a large independent cohort, featuring cases from 6 UK centres, a different staining technique and a different immunostainer (Figure 22)
- (v) Further validation using a third complementary technique (immunoblotting) using in vitro samples, reaffirming the phenotype and adding further validation of predicted molecular weights (Figure 24)
- (vi) Robust sample sizes for all studies:
  - qPCR – a study of 67 cases designed with a power calculation
  - IHC – two studies with a combined 412 cases
  - Immunoblot – featuring all eight available *bona fide* OAC cancer cell lines and four normal models derived through a novel methodology
- (vii) Extensive comparison to clinical metadata, identifying tumour suppressor roles for *ALDH3A1* and *-3A2* in OAC carcinogenesis, and confirming their profound influence on disease behaviour in three separate survival cohorts.

Thus, the workflow was strictly evidence-based, using data from each molecular level to inform in the next study. This approach identified the reduced expression of a number of ALDHs in non-squamous tissue, a finding which complements the wider argument set out in Section 1.3.4, in which loss of aldehyde defences and potentiated aldehyde sources are defining features of the transforming lower oesophagus (in fact, these informatics exercises formed the decisive evidence on which to focus the whole PhD project on aldehydes, rather than other aspects of volatile chemistry).

A potential counter-argument to these findings is that the observed ALDH loss merely represents different expression patterns of glandular tissue, and thus is not important to oncogenesis. Looking at the data comparing SqN to normal gastric mucosa in Figure 18, for *ALDH9A1*, this is may be correct, but for *ALDH3A1*, *3A2*, and *4A1* there is was not correct. For *ALDH3A1* and *-3A2*, incremental differences associated with disease progression were seen both *in silico* and experimentally (see Figure 19), and the extensive metadata relationships suggests that either expression-loss of *ALDH3A* genes, or the process governing expression-loss, fundamentally alters these tumours' behaviour.

At any rate, it could be argued that defining transforming roles for *ALDH3A1* or *3A2* is distracting, as the key point is that these ALDH-lo cells (i.e. on the metaplasia to adenocarcinoma spectrum) have replaced the normal ALDH-high cells, which

suggests that they are less able to cope with oesophagus' unique exposure and susceptibility to aldehyde stress. A critical future experiment for this work will compare the aldehyde-detoxifying function of keratinocytes and OAC cells, either as protein lysates, cultured cells or ex vivo explants.

The weaknesses of the qPCR study (e.g. lack of spatially-resolved expression, potential for cell contamination) have been offset in this project by the strengths of the IHC studies, and vice versa (IHC weaknesses: potential for non-specific binding, lack of specific antibodies for some targets). There was one discrepancy between informatics, qPCR, and IHC data, for example *ALDH4A1*, which was seemingly over-expressed in BE compared to OAC in the qPCR study. This phenotype emerged from a relatively small group (n = 10), suggesting the need for more extensive studies. Metaplasia epithelium is one cell thick, and thus tighter histological control (e.g. with laser capture microdissection) may be needed to overcome cell contamination error. Accordingly, in these results, IHC is better suited to assessing BE expression than qPCR, and these findings were generally in keeping with informatics data. Importantly, stroma, muscle, inflammatory cells and blood vessels showed negative to weak staining if any for all markers in both normal and malignant tissue samples, confirming that the microarray and qPCR expression differences are due to epithelial phenotypes.

Oesophageal adenocarcinoma is a less common malignancy with roughly 100-150 operable cases per UK cancer network per year (5). This prevalence encumbers efforts to conduct large molecular studies at single centres; although both the *ALDH* expression studies (RNA level) was adequately powered with local material, the subgroup association studies (IHC) necessitated pooled resources from a historical national collaboration (qPCR cohort: 2013-15, 1<sup>st</sup> IHC 2002-2010; 2<sup>nd</sup> IHC 1990-2005). It is therefore not surprising that several baseline characteristics were different across the three experimental expression datasets. Neo-adjuvant treatment was popularised in the UK after the OEO2 neo-adjuvant trial was published in 2002 (244). Additionally, surgical and pathology practices have shifted to more radical lymphadenectomy and closer attention to lymph node yields respectively, and together these changes could account for the observed differences. Despite these features *ALDH* expression patterns were nonetheless stable.

Expression of *ALDH1A2* and *ALDH2* were both not particularly discriminatory in informatics search, but they were included in the PCR study owing to highly cited

epidemiological data suggesting inactivating variants confer OAC risk (103,245). *ALDH1A2* was not expressed in any measured tissue including SqN, and this suggests that risk polymorphisms' effect must happen at a very early stage in disease (perhaps during oesophageal development), if there is a genuine connection at all. *ALDH2* was expressed in all tissues with a distribution that indicated tight regulation. It will be of interest to take the same experimental designs to OSCC, in which the risk polymorphisms were first described.

*ALDH3A1* was the most significantly suppressed gene. Given the underlying genotoxic hypothesis for *ALDH*-metabolic reprogramming in OAC, this finding is critical, as this is the only *ALDH* gene which is thought localises to the nucleus (84,86,242). In this work, strong *ALDH3A1* nuclear staining was noted, both *in situ* (see Figure 17), and *in vitro* (Figure 24). With *ALDH3A2*, another potentially suppressed gene, *ALDH3A1* is the only *ALDH* thought to influence cell cycle progression, particularly in keratinocytes (74,84). Notably, expression was highest in the slow-cycling basal keratinocytes. This gene did seem to influence survival in OAC, although *ALDH3A1* positive tumours were so rare that subgroups were underpopulated even in the validation IHC cohort. *ALDH3A1* does not have a specific inborn metabolic error, suggesting its metabolic properties are redundant (79). However, in the present work, all of its homologous *ALDH*s (*ALDH3A2*, *-3B1*, *-3B2*) were found to be poorly expressed in OAC, either in informatics datasets and/or experimentally (see Figure 18, Figure 23, and Appendix 3). The potential for this to lead to measurable changes in metabolic phenotypes is addressed in Section 4.

*ALDH3A2* expression was associated with prior use of neo-adjuvant chemotherapy in OAC. Over-expression of several *ALDH*s has been associated with chemotherapy resistance in breast, lung, ovarian and colorectal epithelial cancers (79,232,246,247), so this could represent a tumour response to chemotherapy. However, *ALDH3A2* expression was also associated with favourable stage, differentiation and survival, implying that if expression was an adaptation to treatment, it did not interfere with treatment efficacy (in contrast to *ALDH3A1* and non-small cell lung cancer, in which it is thought to confer chemoresistance (124)).

Further informatics experiments indicated that expression loss of *ALDH3A1* and *-3A2* may be associated with loss of heterozygosity (LOH) at 17p11. Macrodeletion events in OAC involving the whole of 17p are well described (207,208,239). More recently, higher resolution studies have shown the *ALDH3A1/3A2* locus at 17p11.2 to be

specifically targeted for deletion in both OAC and OSCC (248,249). A complicating factor is that *TP53* lies a few megabases telomeric of this locus, raising the possibility that *ALDH3A1/3A2* copy changes are a passenger effect to the *TP53* LOH “second-hit” (the first hit being the point mutations that occurs in >80% of OAC cases)(31,33,241). The finding that OE19 cells are both amplified and over-express *ALDH3A1* implies that expression is highly copy-dependent. Despite these various lines of evidence, experimental validation of whether 17p copy loss is sufficient to drive *ALDH3A1/3A2* expression loss has not yet been achieved, and could make an attractive next step (e.g. fluorescence in-situ hybridisation correlated to an expression analysis).

*ALDH4A1* and *-9A1* were similarly hypoactive in the majority of OAC cases. Loss of these markers showed a non-significant trend towards poorer survival; increasing the sample sizes may discern significantly different patterns. *ALDH4A1* is known to be a p53 target<sup>74</sup>. Given that >80% of OAC are *TP53* mutant, this may contribute to the observed *ALDH4A1* hypoactivity. Loss of *ALDH9A1*, a cytosolic isoenzyme, would interfere with diverse pathways including acetaldehyde, dopamine, and putrescine catabolism, although this seemed to be a glandular rather than neoplastic phenotype.

**CHAPTER 4 – EFFECTS OF ALDEHYDE  
METABOLIC REPROGRAMMING IN  
OESOPHAGEAL ADENOCARCINOMA**

## Summary

The clinical phenotyping experiments in Chapters 2 and 3 set out to describe aldehyde metabolism in normal and malignant oesophageal tissues, and genetic factors that may influence it. These data suggested that aldehydes are enriched in the malignant oesophagus, and that this occurs in the context of reduced ALDH expression. The effects of this metabolic reprogramming were next to be investigated. The sum effects are likely to be vast, given the broad reactivity and bio-activity of these compounds. However a number of specific questions were prioritised for investigation: In OAC, (i) Do aldehydes interact with DNA? (ii) Is ALDH derangement sufficient to enrich measurable aldehyde? (iii) Does *ALDH3A2* have *bona fide* tumour suppressor functions?

As discussed in Chapter 1, several aldehydes form exocyclic and other adducts on the base pairing face of DNA nucleotides, implying molecular distortion and the potential for mismatch during a polymerase read. Free aldehydes were enriched in OAC tissues, and so second UPLC-MS/MS method was developed to quantify selected aldehyde-nucleoside adducts in the same target tissues. This found that adducts of acetaldehyde and hydroxynonenal were enriched in OAC SqT and AdT compared to SqN and leucocyte DNA, although these results have certain limitations.

The phenotyping studies in Chapters 2 and 3 offer snapshots of OAC aldehyde metabolism from a metabolic and genetic perspective, respectively. However, to dissect causality, mechanistic studies are required by controlling candidate gene expression in models of OAC and measuring metabolic and non-metabolic outputs. The importance of this approach in the context of metabolism science has recently been reviewed (208). Thus, two ALDH perturbation strategies were selected: first, pharmacological inhibition with a global ALDH inhibitor (diethylaminobenzaldehyde, DEAB), and then precision perturbation with single and combinatorial knockdown with RNA interference (RNAi). There was little difference in aldehyde concentrations in a panel of normal and malignant oesophageal cell cultures under normal culture conditions, although all cell lines displayed enhanced aldehyde phenotypes with DEAB, emphasising how aldehydes are constitutively produced and accumulate without competent defences. However, convincing specific aldehyde phenotypes could not be determined following RNAi, suggesting that functional redundancy or inadequate knockdown impair aldehyde accumulation in this model.



The final sub-study explored the potential tumour suppressor functions of *ALDH3A2*. This built on the expression analysis in Chapter 2, which revealed a strong association of low *ALDH3A2* with adverse disease features and poorer survival. Moreover, *ALDH3A2* gives rise to an in-born metabolic error of which keratinocyte hyperplasia is a prominent feature, suggesting functional non-redundancy and potential non-metabolic influences in cell cycle control. Thus, OAC cell lines stably perturbed for *ALDH3A2* were established and the effects on relevant metabolic and non-metabolic phenotypes were assessed.

## Hypothesis & aims

Chapter hypothesis: aldehyde metabolic reprogramming contributes to carcinogenesis. Three effects will be assessed in detail (i) Aldehyde-nucleotide adducts in OAC (ii) ALDH-aldehyde mechanistic modelling (iii) Tumour suppressor functions of *ALDH3A2*

### 1. Adducts:

- Develop and validate UPLC-MS/MS method to quantify aldehyde-nucleotide adducts in DNA hydrolysates.
- In particular, to address three specific issues pertaining to nucleoside analytics: (i) extreme differences in constituent target concentrations (ii) efficient and reproducible nuclease hydrolysis (iii) analyte recovery
- Quantify candidate aldehyde-nucleotide adduct concentrations in OAC DNA samples and relevant controls

### 2. Specific metabolic effects of *ALDH* loss:

- Test whether low *ALDH* expression is sufficient to enrich aldehydes in *in vitro* oesophageal cell models
- Test whether chemical and genetic *ALDH* suppression is sufficient to enrich aldehydes in *in vitro* oesophageal cell models

### 3. Oncotypic phenotypes of *ALDH3A2* loss.

- Develop *ALDH3A2*-constrained OAC models using CRISPR-Cas9 engineering and lentiviral ORF transduction
- Validate these models metabolically
- Measure explanatory non-metabolic phenotypes for the putative tumour suppressor roles of *ALDH3A2*, including growth, cell cycle, and redox control

## 4.1 Aldehyde-nucleoside adducts

### 4.1.1 Methodological rationale

Aldehydes are thought to exercise their toxic effects predominantly through biological molecule adduction (58,131). For example, electrophilic aldehydes readily attack amine groups to form Schiff bases, including the sidegroups of nitrogenous bases in DNA, including Watson-Crick base-pairing participants. This can distort these molecular features and potentially program error (i.e. mutations) into subsequent cell division, if not repaired. Thus, the International Agency for Research on Cancer lists eight common aldehydes as potential or known mutagens, including formaldehyde, acetaldehyde, glyoxal, methylglyoxal, malondialdehyde, furfural and propanal (130).

Aldehyde-nucleotide adducts form attractive analytical targets as they are more stable than free carbonyls and give mechanistic insights into pathophysiology (250–252). In the context of oesophageal carcinogenesis, the presence of aldehyde-nucleotide adducts could (i) confirm general and/or specific roles for aldehydes in mutagenesis (ii) compare this role to other mutagenic processes e.g. halo-adduction (iii) potentially highlight defects in DNA repair machinery. In addition, the potential permutations of aldehyde adducts with nucleotides (4 partners) will be less than digested peptides (20+) or lipids (100+), and DNA lesions are more likely to be heritable. Together, this suggests a given aldehyde-nucleotide adduct has a higher probability of being detectable by MS, and a logical first line of investigation.

Techniques for studying nucleoside modifications include antibody-based, <sup>32</sup>P post-labelling, and LC-MS. Antibody-based techniques such as ELISA and immunohistochemistry are restricted by the availability of specific antibodies, which are difficult to generate with specificity for these relatively small immunogenic motifs. One 1993 immunohistochemical study has been reported of oesophageal cancer, in which strong neoplastic staining was noted for an antibody reputed to bind methyladenosine and pseudouridine(253).

<sup>32</sup>P post-labelling is a relatively old thin-layer chromatography technique that relies on labelling digested nucleosides with phosphorus-32. This is a strong beta-emitter with a low half-life, and therefore this technique is somewhat limited by the need for robust radioactivity precautions (but this also makes the technique exquisitely

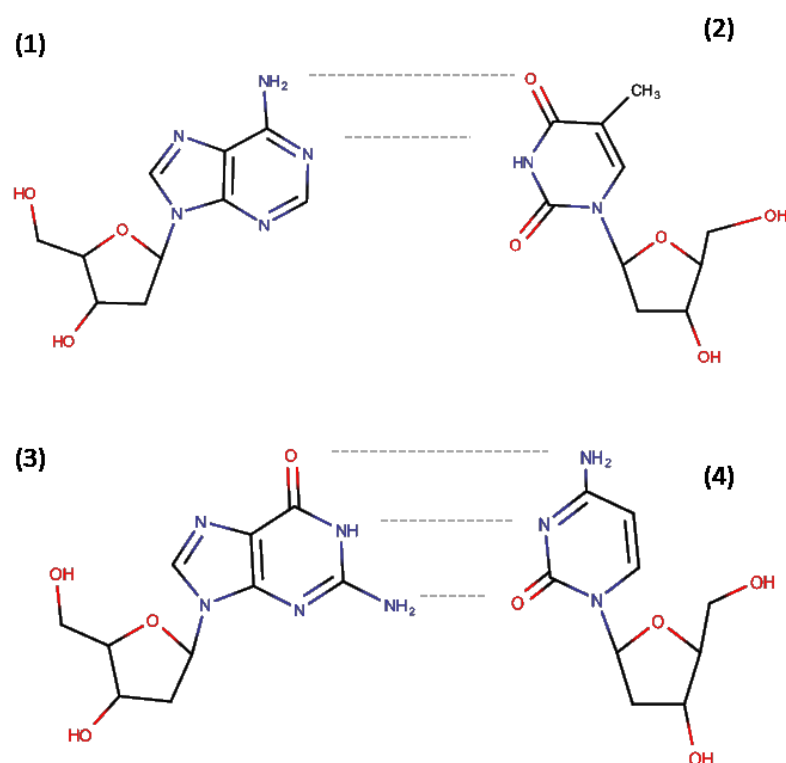
sensitive)(254). However, in the absence of alternatives in the 1980s and 1990s, <sup>32</sup>P post-labelling has been used to describe nucleotide was used to describe nucleoside adducts in gastric cancers, such as N7-methyldeoxyguanosine (255), and in studying the DNA adduct lesions caused by bile in familial polyposis (256,257).

In recent years LC-MS/MS has gained traction for studying nucleoside adducts, because it is fast, sensitive, does not require radioactive labelling, is suitable for high-throughput analysis and offers lucid compound identification through column retention, parent mass-to-charge ratio, and fragmentation pattern (250,252). In general cancer research, an enormous diversity of nucleoside modification has been investigated with this technique, including base oxidation (the most common change), methylation (the most often investigated), deamination, amination, peroxidation, halogenation and bulky adduct formation (excellently reviewed in (250)). In upper GI malignancy, LC-MS/MS has been used to study lipid peroxidation adducts such as those caused by acetaldehyde, HNE, ONE and 4-oxo-2-hexenal in gastric cancers (24,112,258,259). The role of acetaldehyde in OSCC tumorigenesis has been studied in the context of ethanol-fed *ALDH2* *-/-* knockout mice (23,260), although no previous reports have investigated any nucleoside adduct in oesophageal adenocarcinoma. Given the severe genetic injury associated with OAC, this is a clear gap in the current adduct literature.

The technique consists of isolating a very pure DNA sample, using heat or enzymes to hydrolyse the DNA, dephosphorylating the nucleosides to prevent re-polymerisation, then removing all protein elements, and finally injecting into an electrospray source and analysing by LC-MS/MS. By far the most abundant daughter over a wide range of fragmentation energies is usually the 116 neutral-loss ion, given by the ionised nitrogenous base ring(s) which has lost the ribose ring. Specific analytical issues that have been described include the 6-8 orders of magnitude difference between the relatively rare modified nucleosides and the four unmodified bases, the relatively high DNA input requirement (50-100mcg), the need for inhibitors that prevent adventitious nucleoside modification (e.g. from endogenous deaminases), the challenge of acquiring a complete hydrolysis without adventitious base formation, and the lack of available isotope labelled internal standards (250–252,261–263).

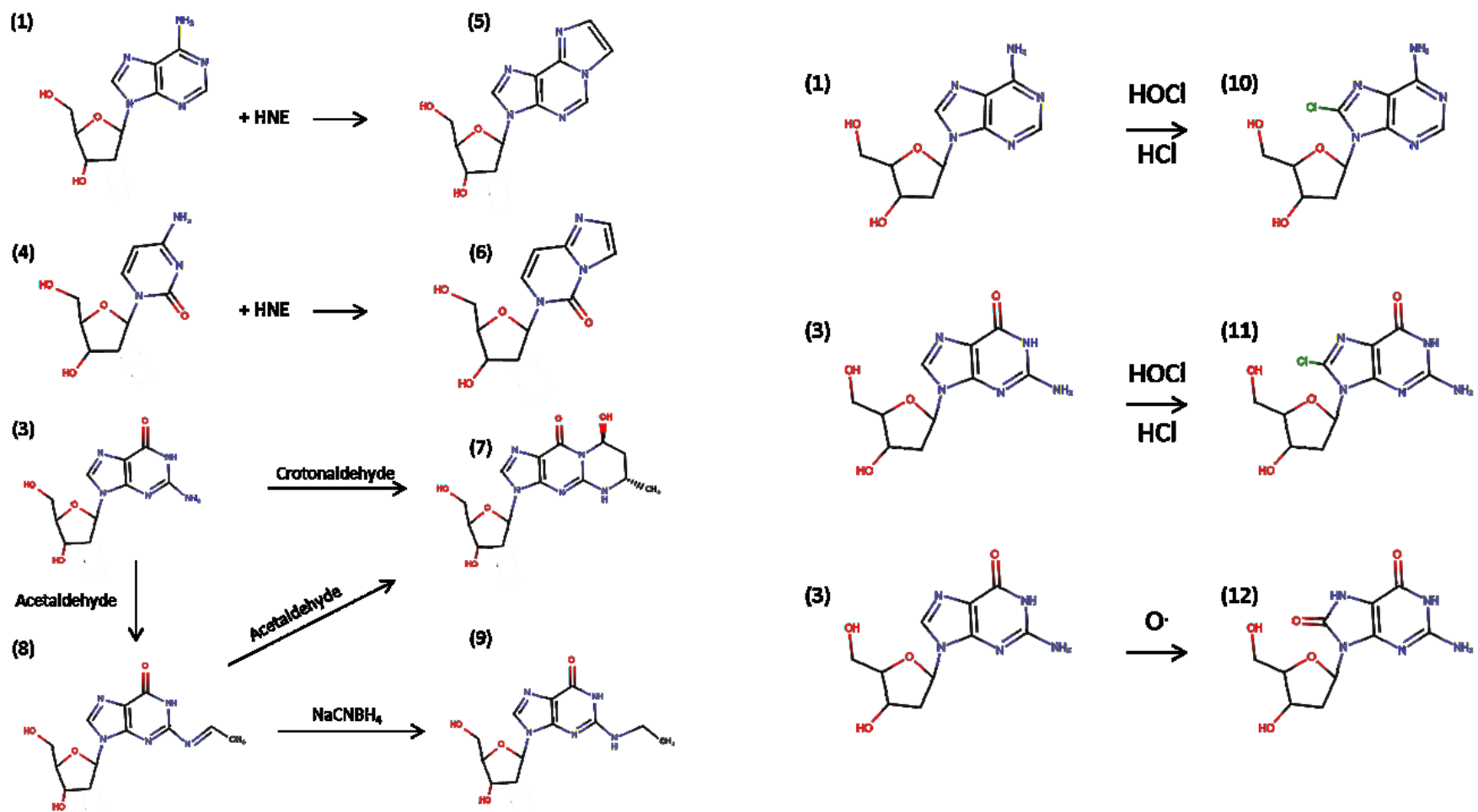
Some groups have described an “adductome” approach, taking advantage of the 116 neutral loss to screen for unknown modified nucleosides (24,264). Using accurate

mass or simply  $m/z$  values, the authors offer initial structural identification, and then validate their findings as much as possible with synthesised standards. It was decided that for this proof-of-concept, a targeted analysis would be preferable, and therefore seven targets were selected for initial characterisation. This included two base adducts each of two aldehydes – acetaldehyde and hydroxynonenal – which are generated through different biology and have the most evidence for genotoxicity in cancer generally (252,261,265,266). In addition, the potential for base halogenation was investigated, taking the hypothesis that HCl reflux and hypochlorous acid from neutrophil could potentiate genotoxicity (267,268), as there is a recent trend toward an “acid signature” in OAC sequencing studies (34). Lastly, 8-oxo-deoxyguanosine, the most common modified nucleoside and which arises for oxidative stress, was also selected, such that these three mechanisms of base modification could be compared in OAC. The structure of normal nucleosides is provided in Figure 26, and the reaction schemes of the modified nucleosides in provided in Figure 27.



**Figure 26: Normal nucleosides**

Left, normal nucleosides: Deoxyadenosine (dA, 1) pairs with deoxythymidine (dT, 2); deoxyguanosine (dG, 3) pairs with deoxycytidine (dC, 4).



**Figure 27: Selected modified nucleosides for study**

Left, aldehyde-nucleotide adducts: ethenodeoxyadenosine (edA, 5) and ethenodeoxycytidine (edC, 5) are formed from a complex reaction involving the epoxidation of HNE. Crotondeoxyguanosine (CrodG, 7) is formed by reaction of dG with crotonaldehyde or two acetaldehydes. Acetaldehyde also gives an unstable intermediate (8) that can be converted to ethyl deoxyguanosine (EtdG, 9) by reduction. Right, halo/oxo adducts. dA and dG can be halogenated by HOCl or HCl to give 8-chlorodeoxyadenosine (cdA, 10) or 8-chlorodeoxyguanosine (cdG, 11). dG can also be oxidised by singlet oxygen, most commonly giving 8-oxo-2-deoxyguanosine (odG, 12).

#### 4.1.2 Adducts method development

Dr Z Bodai is supervised the Candidate in the development of this method

**UPLC-MS/MS materials & biospecimens** All standards and solvents were of the highest purity available. HPLC-grade water and methanol were purchased from Sigma. Ultra-performance liquid chromatography stationary phase columns (see below) were purchased from Waters. 0.2 micron pre-column mesh filters were used routinely. Unlabelled and stable isotope-labelled standard nucleosides were obtained as in Table 18. Micrococcal nuclease, spleen phosphodiesterase and alkaline phosphatase were purchased from Worthington. The following inhibitors were purchased: tetrahydrouridine (Calbiochem/Merck), pentostatin, deferoxamine, butylated hydroxytoluene (all Sigma). Salts and buffers outlined below were purchased from Sigma. This work was undertaken using an Acuity Sample Manager/Injector/Binary pump coupled to a Waters TQS triple quadrupole mass spectrometer. Biospecimens were accessed through the same IRB and HO approvals as set out in Section 2.3.2.

**Table 18:** Authentic external and internal standard sources and abbreviations

Common Name	Abbreviation	CAS	Supplier
<b>Normal nucleosides</b>			
Deoxyadenosine	dA	958-09-8	Sigma
Deoxycytidine	dC	951-77-9	Sigma
Deoxyguanosine	dG	961-07-9	Sigma
Deoxythymidine	dT	50-89-5	Sigma
<b>Adducted nucleosides</b>			
1,N6-Etheno-2'-deoxyadenosine	edA	68498-25-9	Santa Cruz
3,N4-Etheno-2'-deoxycytidine	edC	68498-26-0	Carbosynth
8-Chloro-2'-deoxyguanosine	cdG	437715-62-3	Carbosynth
8-Chloro-2'-deoxyadenosine	cdA	85562-55-6	Carbosynth
$\alpha$ -Methyl- $\gamma$ -hydroxy-1,N2-propano-2'-deoxyguanosine	CrodG	132014-87-0	TRC
N2-ethyl-deoxyguanosine	EtdG	101803-03-6	Sigma
8-Oxo-2'-deoxyguanosine	OdG	88847-89-6	Sigma
<b>Labelled nucleotides</b>			
8-Oxo-2'-deoxyguanosine-d <sub>3</sub>	OdG-d3	n/a	Sigma
$\alpha$ -Methyl- $\gamma$ -hydroxy-1,N2-propano-2'-deoxyguanosine-d <sub>3</sub>	CrodG-d3	n/a	TRC
Deoxyadenosine-N <sup>13</sup> <sub>5</sub>	dA-N5	n/a	CIL

CAS, Chemical Abstracts Service; TRC, Toronto Research Chemicals; CIL, Cambridge Isotope Laboratories

**MS/MS development** Standards were dissolved 50:50 in acetonitrile and water to 1mg/mL solutions. Volumetric flasks and a nanogram scale ensured accuracy. Aliquots were stored at -80°C. The electrospray ionisation source was selected owing to the polar nature of the target compounds, operating in the positive mode, using initial settings detailed in Table 19. Using single 10 mcg/ml working solutions infused directly the mass spectrometer, product ion scans were undertaken to identify lead transitions and optimise fragmentation energies from the calculated  $[M+1]^+$  ion. At low fragmentation energies, the  $[M+1-116]^+$  daughter response was typically highest by 2-3 orders of magnitude, with smaller fragments becoming more abundant with higher energies. The collision energy and cone voltage were optimised to maximise intensity of the most abundant fragments, and these details are provide in Table 20. A preliminary multiple reaction monitoring MS method was thus establish for these target compounds. The detector dwell time was set to 'auto' in the Waters software and gave >10 points-per-peak. The ribose neutral loss was always at least 50-fold more intense than any other fragment even with optimised fragmentation.

**Table 19:** Initial instrument settings

<b>Sample manager</b>		
Injection (uL)		10
injection mode		Partial loop with needle overfill
Temperature (°C)		4
Column		Phenyl-hexyl 1.7um particle size x 2.1mm x 50mm
Precolumn filter frit size (um)		0.2um
Column Temperature		35
<b>Binary solvent manager</b>		
Flow rate (mL/min)		0.3
Mobile phase A/Weak needle wash		Water with 5mM ammonium formate
Mobile phase B/Strong needle wash		Methanol
Seal wash		Water
<b>Source settings</b>		
Type		Electrospray ionisation in positive mode
Source temperature (°C)		150
Desolvation temperature (°C)		400
Desolvation Gas Flow (L/Hr)		400
Cone Gas Flow (L/h)		200
Capillary voltage (V)		2500
Cone voltage (V)		10



**Table 20:** Fragmentation patterns and energies for target compounds.

Compound	MW	Parent ion	Ribose (116) loss			Other daughter(s)
		[M+H] <sup>+</sup>	Daughter m/z	CE	CV	
edA	275.1	276.1	160.1	10	16	133, 117, 106, 81
edC	251	252	136	15	10	109, 81
cdA	285	286	170	15	15	117, 73
cdG	301	302	186	15	20	117, 73
EtdG	295	296	180	15	10	110, 135
CrodG	337	338	222	15	10	178, 204
OdG	283	284	168	10	10	57, 60
dA	251.2	252.2	136	10	70 <sup>a</sup>	118.8, 98.9
dC	227.2	228.2	112	10	70 <sup>a</sup>	117, 94.8, 68.8
dG	267.2	268.2	152	10	70 <sup>a</sup>	134.8, 116.9, 109.8
dT	242.2	243	127.2	10	70 <sup>a</sup>	
edA(N <sup>15</sup> <sub>5</sub> )	280.1	281.1	165.1	10	16	138, 81
OdG(d <sub>3</sub> )	340	341	225	10	10	57, 60
CrodG(d <sub>3</sub> )	283	284	168	15	10	181

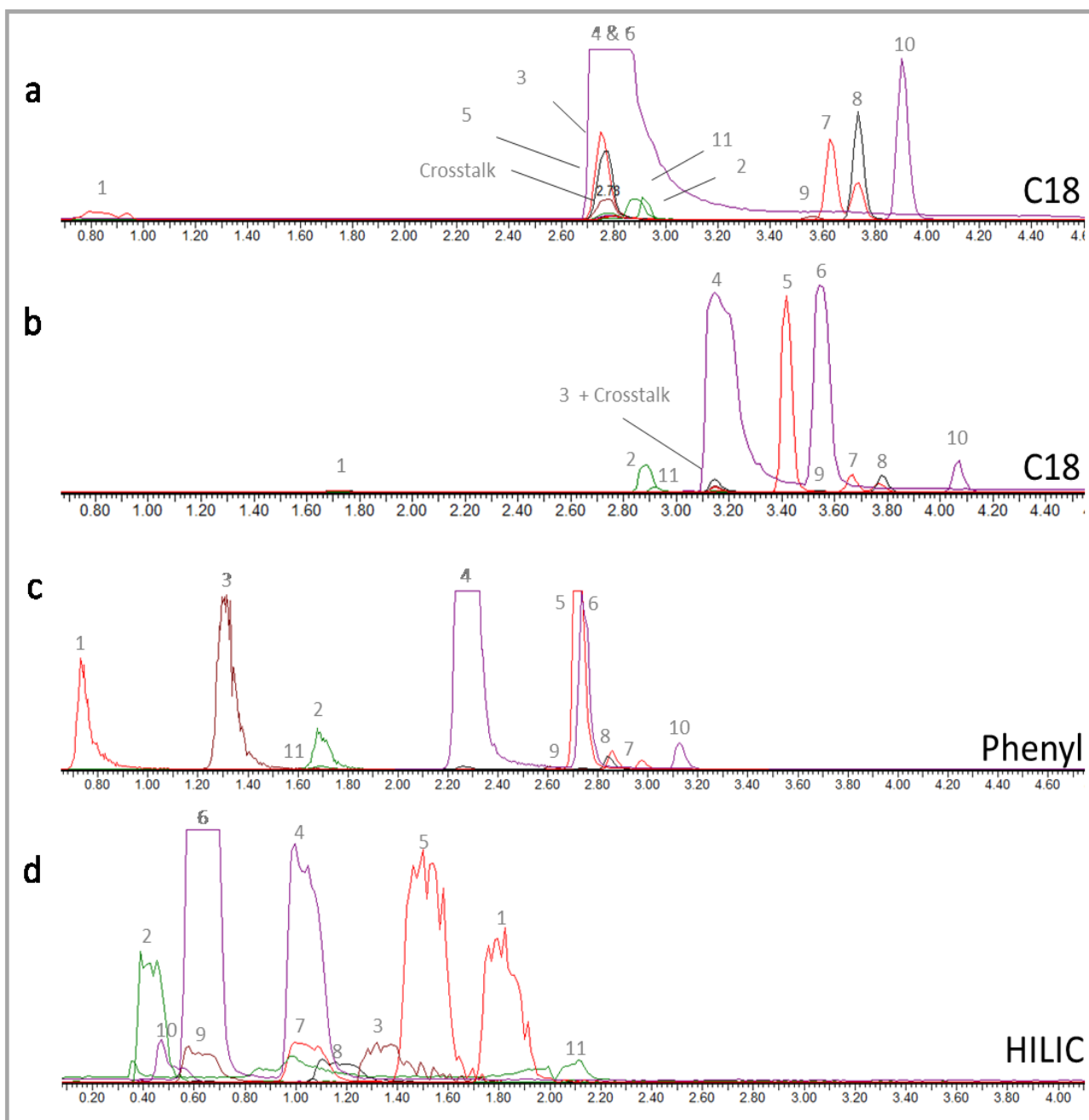
<sup>a</sup>Purposefully unoptimised cone voltages were used to control basic nucleotide ionisation. MW, molecular weight; [M + H]<sup>+</sup> protonated primary ion; CE, collision energy; CV, cone voltage; Compound abbreviations as previously.

**Chromatography development** The ratio of the target adducts to the four normal nucleosides in DNA hydrolysates from living systems can be 1:10<sup>6</sup>-10<sup>8</sup> (24,251,252). Thus, effective separation of the normal bases from the adducted bases is required, as (i) the linearity of the detector in modern MS system is only a few orders of magnitude, indicating the need to measure each sample at different dilutions (269) (ii) ion suppression caused by co-eluting high concentration compounds (e.g. unmodified nucleosides) may reduce target compound responses below the limit of quantification (190,191) (iii) if the MS/MS system is overloaded by high concentrations of a particular ion, a “crosstalk” phenomenon can occur leading to erroneous “ghost” peaks, i.e. false positives (269). Various sample preparation techniques have tried to remove normal nucleotides from others, including solid phase extraction and preparative-scale high performance liquid chromatography (251,270). However, removing unmodified nucleosides precludes accurately quantifying the *proportion* of base modification, as only the total DNA input could then be used to calculate adduct prevalence, which relies on a perfect and consistent digest that internal standards will not control. Separating nucleoside fractions by preparative HPLC will preserve all targets but inherently result in analyte loss, require

additional method steps, and extra solvents. Thus, it would be advantageous to measure unmodified and modified nucleotides in a single quantitative method.

An innovative solution is ultra-performance liquid chromatography (UPLC). As discussed in Section 2.2, UPLC offers improved peak resolution compared to older systems, which should counteract crosstalk from near-eluting high concentration compounds, as well as improved signal/noise ratios and therefore the LOD. Thus, various combinations of stationary phase (phenyl, C18/HSS-T3, HILIC) and mobile phase (methanol, acetonitrile) were checked to establish adequate selectivity. Five microliters of a 1 mcg/mL mix of all unlabeled nucleoside standards was serially injected into the MS system, initially under isocratic conditions. Representative chromatograms are given in Figure 28. The phenyl-hexyl reverse phase column was the only stationary phase option that segregated normal nucleosides from modified nucleosides, albeit with one overlap – OdG (see Figure 28c). The measured dead volume of this column was 0.65 minutes, therefore deoxycytidine,  $R_t$  0.7, is probably unretained (however with hydrophilic interaction chromatography (HILIC), deoxycytidine is well-retained and therefore quantifiable if required, see Figure 28d). Mobile phase gradients were optimized to maximally separate the target compounds over a 6 minute run (with a final 2 minute restitution phase). The pH limit of this column is 8; pH modification below this enhanced ionization but sacrificed retention and resolution, and therefore no pH modification was selected. Modest modification of the aqueous phase with ammonium formate improved ionization of all targets, and therefore was included in the aqueous phase at a concentration of 5mM.

The rationale for accepting the OdG overlap was as follows. First, an isotope-labelled standard was available for this compound, which should correct all ionization issues related to normal nucleoside co-elution. Second, to minimize issues of crosstalk, extreme fragmentation energies were used for the normal nucleosides (see Table 20); only deoxyadenosine continued to provide crosstalk owing to its shared transition with edC, and this eluted after OdG. Third, aldehyde-nucleotide adducts were the priority targets, and these were all well separated. Forth, validation experiments (see below) proved that OdG was accurately and precisely quantifiable. The relatively high limit of detection was comparably less important as this is the most common nucleoside modification (251), and therefore sufficiently sensitive.



**Figure 28: Nucleoside chromatography using various stationary phases.**

Panel a, HSS-T3 C18 column and water/methanol *gradient 1*. Panel b, HSS-T3 C18 column and water/methanol *gradient 2*. Panel c, phenyl-hexyl column and *gradient 2*. Panel d, HILIC column and HILIC mobile phases, *gradient 1*. 1-4 Normal nucleosides: 1, dC; 2, dT; 3, dG; 4, dA. 5-11 Adducts: 5, edA; 6, edC; 7, CrodG; 8, EtdG; 9, cdG; 10, cdA; 11, odG

Unlabelled ('external') standard calibrations were established by serial dilutions in the presence of 50mcg/mL unmodified nucleosides, to emulate potential ion suppression from the four normal nucleosides. Linearity for all compounds ranged at the pg to ng/ml scale with correlation coefficient  $r^2 > 0.995$ . Limits of detection (defined as signal-to-noise ratio  $> 3$ ) and of quantification (defined as the lowest concentration where the relative standard deviation of measurement was  $< 10\%$ ) were established; the upper limit of quantification was also characterized as the highest concentration before linearity was reduced below 0.995. The Waters system was able to provide linear quantification across at least 5 orders of magnitude for all nucleosides, however the expected concentration of unmodified nucleosides was expected to be up  $> 100x$  higher still. Thus, it was determined that all analytical samples would be measured in duplicate at two concentrations: neat, and 1:1000 dilution. This second dilution brought the concentration of the unmodified nucleosides within the linear range of the respective calibrations.

**Table 21:** Linear range of quantification of external standards

	$t_R$	LOD <sup>a</sup>	LLOQ		Calibration range	Linearity, $r^2$	LLOQ <sup>b</sup>		LLOQ <sup>c</sup>	
			pg/ml				per $10^8$ normal nucs		fmol	
<b>edA</b>	2.79	5	25	5 - 2500	0.9997	7	0.91			
<b>edC</b>	2.79	5	5	5 - 2500	0.9992	1	0.20			
<b>cdA</b>	3.14	10	10	5 - 2500	0.9998	3	0.35			
<b>cdG</b>	2.65	500	750	500 - 10000	0.9981	183	24.92			
<b>EtdG</b>	2.91	10	25	5 - 2500	0.9993	6	0.85			
<b>CrodG</b>	3	25	50	5 - 2500	0.9994	11	1.48			
<b>OdG</b>	1.7	50	250	5 - 2500	0.9997	65	8.83			
			ng/ml							
<b>dA</b>	2.2	0.5	50	0.25 - 100	0.9979	-	-			
<b>dC</b>	0.8	0.5	25	0.25 - 100	0.9995	-	-			
<b>dG</b>	1.2	0.5	50	0.25 - 100	0.9966	-	-			
<b>dT</b>	1.4	5	500	1 - 500	0.9999	-	-			
<b>edA-N5</b>	2.79	<i>As per unlabelled compound - spiking concentration 1000pg/mL</i>								
<b>CrodG</b>	3									
<b>OdG</b>	1.7									

<sup>a</sup>In neat mobile phase n=12

<sup>b</sup>Calculated given a standard 50 mcg DNA input leading to a 40 nmol dA yield

<sup>c</sup>Calculated as total injected quantity in a 10ul injection

**Sample preparation development** The final element for optimization is the isolation of pure nucleosides from biological samples for quantitation. A review of the literature indicates three critical components to this: (i) Pure isolation of >50mcg DNA quantities without adventitious adduct formation (ii) Complete DNA hydrolysis and dephosphorylation to nucleosides, without adventitious modification (iii) Complete deproteination after enzymatic digest (24,113,251,265).

Using DNA extracted from porcine stomach and oesophagus, three isolation techniques were tested (i) silica spin columns (DNAeasy columns, Qiagen) (ii) Trizol extraction with urea-based back extraction (Invitrogen) (iii) a liquid-liquid extraction kit (Gentra Puregene kit, Qiagen). It was found that the silica spin column had a maximum binding capacity well below the minimum limit suggested by the literature (50mcg), and there was considerable chloroform and phenol contamination with Trizol (data not shown). The Gentra kit provided good tissue lysis, complete DNA recovery, and was relatively simple and flexible, and was thus chosen for clinical phenotyping. In keeping with previous reports, it was also found that the use of inhibitors to control against deamination (dA – pentostatin, dC – tetrahydrouridine), oxidation (butylated hydroxytoluene), and peroxidation (deferoxamine) increased yield considerably, in particular the concentration of dA.

DNA hydrolysis was a critical component for optimisation, as accurate quantitation depends on individual nucleotides being liberated from DNA and then dephosphorylated to nucleosides, without further modification to unknown structures. Thus, a range of different incubation conditions were attempted, focussing on edA and dA as model targets, and it was found that hydrolysis and dephosphorylation increased intensities to 3 hours and then reduced. With these optimisations, the hydrolysis yield was calculated as 28%. Testing different DNA input amounts revealed consistency of this yield, although it was decided that the input quantity for samples should be consistent, and defined as the minimum amount necessary to ensure the natural concentrations were consistently above the limit of detection (50mcg of DNA).

To control for variation introduced by these and further unknown factors, isotope-labelled internal standards were selected. Two standards were available commercially (trideuterated CrodG (Crod3) and OdG (Od3)). These adequately corrected the early eluting OdG and late eluting CrodG and EtdG, but not the middle region in which edA, edC and cdA elute soon after dA. Thus, N<sup>15</sup>-labelled edA

(edAN<sup>15</sup><sub>5</sub>) was made according to published protocols by reacting N<sup>15</sup>-labelled dA with an excess of chloroacetaldehyde (Sigma) (251,259), and purity checked by quantifying edA and edAN<sup>15</sup><sub>5</sub> in serial dilutions (no dAN<sup>15</sup><sub>5</sub> was quantifiable in any sample). The concentration of edAN<sup>15</sup><sub>5</sub> in the reaction mix was estimated by comparing the dilutions to standard calibrations for edA, and 1 mg/mL stock solutions were made in 50:50 water:acetonitrile (as for Crod3 and Od3).

#### 4.1.3 Final adducts method

**DNA isolation:** Snap frozen human, pig or murine tissue samples were weighed on ice and microdissected (see Chapter 3) to achieve a cellular purity >90%. The samples were then processed according to the Gentra kit instructions: approximately 30mg (cancer) to 40mg (squamous) specimens were homogenised in a pestle and mortar under liquid nitrogen to fine granules. The samples were then incubated in the kit cell lysis buffer with proteinase K 10mcg/mL for 3 hours at 55°C until completely dissolved. RNA was then removed with a 20 minute incubation with Rnase A. Protein was precipitated and removed, and then the DNA was precipitated with isopropanol and washed in ethanol. The Pure DNA pellet was then rehydrated and quantified using absorption spectroscopy at 260:280nm (Nanodrop™, Thermo-Life) and verified using dye-intercalation (Qubit™, Invitrogen). For cell samples, approximately 3 million cells were collected with a rubber policeman, washed in PBS, and resuspended in 100µl PBS. The DNA was then extracted using the Gentra kit as above. For whole blood DNA samples, 10ml of whole blood was centrifuged at room temperature for 10 minutes at 3000rpm. The buffy coat intermediate phase was collected and frozen. The red cells were then lysed with red cell lysis buffer (Sigma), and the leucocytes pelleted, resuspended, lysed and further processed as above.

**Nuclease hydrolysis:** DNA samples (10-100µg depending on experiment) of DNA were then lyophilised under vacuum (MiVac Speedvac) and resuspended in 54µl digestion buffer (17mM sodium citrate, 8mM calcium chloride, pH 6.0, 33 units micrococcal nuclease, 0.27 units spleen phosphodiesterase). This buffer also contained: 100µM butylated hydroxytoluene, 50µM deferoxamine, 10µg/ml tetrahydrouridine, and 10µM pentostatin. The mixture was incubated at 37°C for 3 hours. Phosphate groups were then removed to prevent re-polymerisation, by adding 147µl phosphatase buffer (comprising 3U alkaline phosphatase, 0.1M Tris-Cl pH 8.5, 0.2mM zinc sulphate), and allowing the reaction to proceed at 37 °C for 3 hours. The

hydrolysate was then lyophilised to about 20µl. The enzymes were removed by twice extracting with 100µl 70% methanol. The supernatants were combined, lyophilised, and resuspended in 100µl UPLC grade water prior to reverse phase UPLC-MS/MS.

**Nucleoside quantification:** Liquid chromatography was conducted using a Acquity UPLC® system (Waters) equipped with a 5cm phenyl-hexyl column (particle size 1.7µm, internal pore 2.1 µm), 5mM ammonium formate aqueous phase (A) and 100% methanol organic phase (B). 10µl of sample was injected using the integrated Waters autosampler, under initial conditions of 95% A and 5% B. The organic composition was then changed as follows: increased from 1 to 2.5 minutes to 30% B, then 2.5 to 3.5 to 90% B, then 3.5 to 4 to 5% B, and then 4-6 minutes maintained at 5%. Flow rate throughout was 0.3ml/min; the column was maintained at 40°C, and the sample vials at 4°C. Column eluate was ionised in the electrospray ion source in the positive mode. Source settings were as follows: capillary voltage 2 kV, cone voltage 10 V, cone gas flow rate 100L/hr, desolvation gas temperature 400°C, desolvation gas flow rate 200L/hr. For mass spectrometry, analytes were measured in the multiple reaction-monitoring (MRM) mode, with key transitions, collision energies and cone voltages as in Table 20.

Samples were analysed in one continuous run for each experiment. Injection order was randomised to prevent a batch effect. A blank sample was injected every 5 samples to check for carry-over, immediately followed by a suitable calibration point to ensure precision. Peak responses were converted to concentrations by comparing unknowns to calibration curves corrected with the ISTDs (i.e. an internal standard calibration). The overall formula was  $((T1/IS1)-I11)/C1$ , where T1 is the target compound response in the sample, IS1 is the internal standard response in the in sample, I11 is the intercept calculated for the internal standard calibration, and C1 is the coefficient of the internal standard calibration. To provide adduct concentration as a function of total nucleotides, unmodified nucleotides were also quantified using 100x diluted matched samples. The final adduct concentration was expression as adducts per  $10^8$  nucleotides using the formula:  $[\text{adducts } (\mu\text{M})/\text{dA (mM)}]/0.29/10$ .

#### 4.1.4 Method validation

Food and Drug Administration (FDA) guidelines for bioanalytical method validation were used to inform nucleoside method validation(180).

**Recovery, matrix effects, accuracy of nucleoside adducts:** To test recoveries for each nucleoside through the digest process, a bulk porcine DNA extract was split into multiple 50mcg triplicates and spiked with a range of external standard concentrations reflecting the expected natural concentration of the targets, before or after enzymatic hydrolysis. The samples were prepared as per protocol, quantified by UPLC-MS/MS. Recoveries were calculated as:  $([\text{pre-digest spike}] \div [\text{post-digest spike}]) * 100$ . Matrix effects were calculated as:  $(([\text{post-digest spike}] - [\text{unspiked}]) \div [\text{expected concentration}]) * 100$ . Preparative precision was also calculated as the relative standard deviation of the measured intensities across biologically quintuplet preparations, by the formula  $(\text{standard deviation of replicates} \div \text{average of replicates}) * 100$ . Analytical precision was defined by the same formula, measuring the same sample five times. Accuracies were produced by correcting the measured target compound concentrations with the paired internal standard, using an internal standard calibration as in Section 2. The pairings were: edA, edC, cdA, and EtdG to edAN5; CrodG to Crod3; OdG to Od3. Analytics were precise if %RSD was <10%, accurate if +/-20% of expected, and well recovered if +/- 30% expected.

As seen in Table 22, the spiked compounds were well recovered from the enzymatic hydrolysis. For edA, edC, and cdA, there was mild ion suppression in the source, although always >50% of the expected value. CrodG and EtdG experienced mild ion enhancement, but on balance these five compounds were relatively stable with good preparative and analytical precision, and accuracies. OdG experienced variable matrix effects, however the sensitivity, recovery and accuracy for this compound were acceptable. In contrast, cdG performed the least well. This compound ionised poorly, with a 10x higher limit of detection. Despite compensating by increasing the spiking concentrations 10 fold, preparative precision was relatively poor, and matrix effects and accuracy were variable. It was concluded that the poor sensitivity and relatively high error for cdG were prohibitive, and thus it was excluded from clinical phenotyping (in addition, this halogenated adduct was a pre-determined lower analytical priority than the carbonyl adducts).



**Table 22:** Method validation data for nucleoside-adducts

		Recovery (%)						
spike, pg/mL	n	edA	edC	cdA	cdG	EtdG	CrodG	OdG
250.0	5	128	89	95	110	122	105	98
500.0	5	92	73	83	92	96	108	125
		Matrix Effects (%)						
spike, pg/mL	n	edA	edC	cdA	cdG	EtdG	CrodG	OdG
250.0	5	72	64	75	142	79	168	78
500.0	5	94	54	66	86	92	120	132
		Preparative precision (%RSD)						
spike, pg/mL	n	edA	edC	cdA	cdG	EtdG	CrodG	OdG
250.0	5	8.5	8.8	3.7	10.3	7.4	12.9	11.4
500.0	5	9.5	5.6	7.2	12.3	6.6	9.3	10.3
		Analytical precision (%RSD)						
spike, pg/mL	n	edA	edC	cdA	cdG	EtdG	CrodG	OdG
250.0	5	0.6	6.5	2.0	2.2	3.8	5.4	1.4
500.0	5	8.7	5.1	4.0	0.9	2.7	0.3	7.8
		Accuracy (% of expected 250 pg/mL spike)						
Tissue	n	edA	edC	cdA	cdG	EtdG	CrodG	OdG
Thymus	3	111	106	103	62	101	109	109
Omentum	3	98	86	72	47	97	96	101
Small Bowel	3	82	66	78	96	101	120	139
Stomach	3	96	113	102	112	155	125	142
Liver	3	81	87	86	106	163	117	159
Cell line	3	92	108	69	25	116	123	115
Whole Blood	3	101	147	108	29	139	117	180
		Stability, % of 0h						
In mobile phase	n	edA	edC	cdA	cdG	EtdG	CrodG	OdG
12h	10	91	86	79	131	107	97	95
96h	10	109	111	93	85	119	122	42
In sample matrix	n	edA	edC	cdA	cdG	EtdG	CrodG	OdG
12h	10	115	112	116	89	113	111	93
96h	10	95	114	143	142	128	119	53

NB: cdG spiked were ten fold quoted. %RSD, percent relative standard deviation

**Stability** Intra-day and inter-day stability was assessed by re-injecting external standard calibration samples (n = 10) and spiked DNA digest samples (n = 10) at 12h and 96h, having been stored at 4°C in the autosampler storage rack. The instrument was dedicated for this experiment throughout. A single batch of mobile phase sufficient for the entire experiment was used. It was shown that the target compounds were stable at both time-points, in both neat mobile phase and enzymatic hydrolysates.

**Validation of normal nucleoside quantitation** A method goal was to express adduct burden per normal nucleosides. Thus, the performance of the method at quantifying normal nucleosides was also tested (dC was excluded as its early elution questioned its quantitative reliability). A pig DNA sample was split into 50mcg aliquots, digested as usual, and independently spiked with a mix of the four normal nucleosides to final concentration of 25mcg/mL (dA, dG) or 100mcg/mL (dT). The hydrolysates were then diluted 1:1000 and injected into the UPLC-MS/MS system. Using an external standard calibration, the accuracies for all three tested nucleosides were acceptable (see Table 23), with reasonable analytical precision and intra-day stability. Interestingly dA increased significantly to 96h, both in mobile phase samples and hydrolysates; presumably this represents degradation of oligomers.

**Carryover** was assessed using water blanks after all low, medium and high concentration QCs and samples, and found to be <0.1% for all target compounds including the high concentration unmodified nucleosides.

**Table 23:** Method validation data for normal nucleosides

Accuracy (% expected)				
Measured (mcg/mL)	<i>n</i>	dA	dG	dT
Unspiked	3	38	11	39
25 mcg/mL spike	3	66	33	-
100 mcg/mL spike	3	-	-	144
Subtraction		28	22	105
<b>Accuracy</b>				
%	3	112	85	105
Analytical precision (25mcg/mL spike)				
	<i>n</i>	dA	dG	dT
%RSD	3	7	10	19
Stability (% of 0h)				
In mobile phase	<i>n</i>	dA	dG	dT
12	3	80	115	80
96	3	175	101	69
In matrix				
12	3	92	103	103
96	3	246	101	82

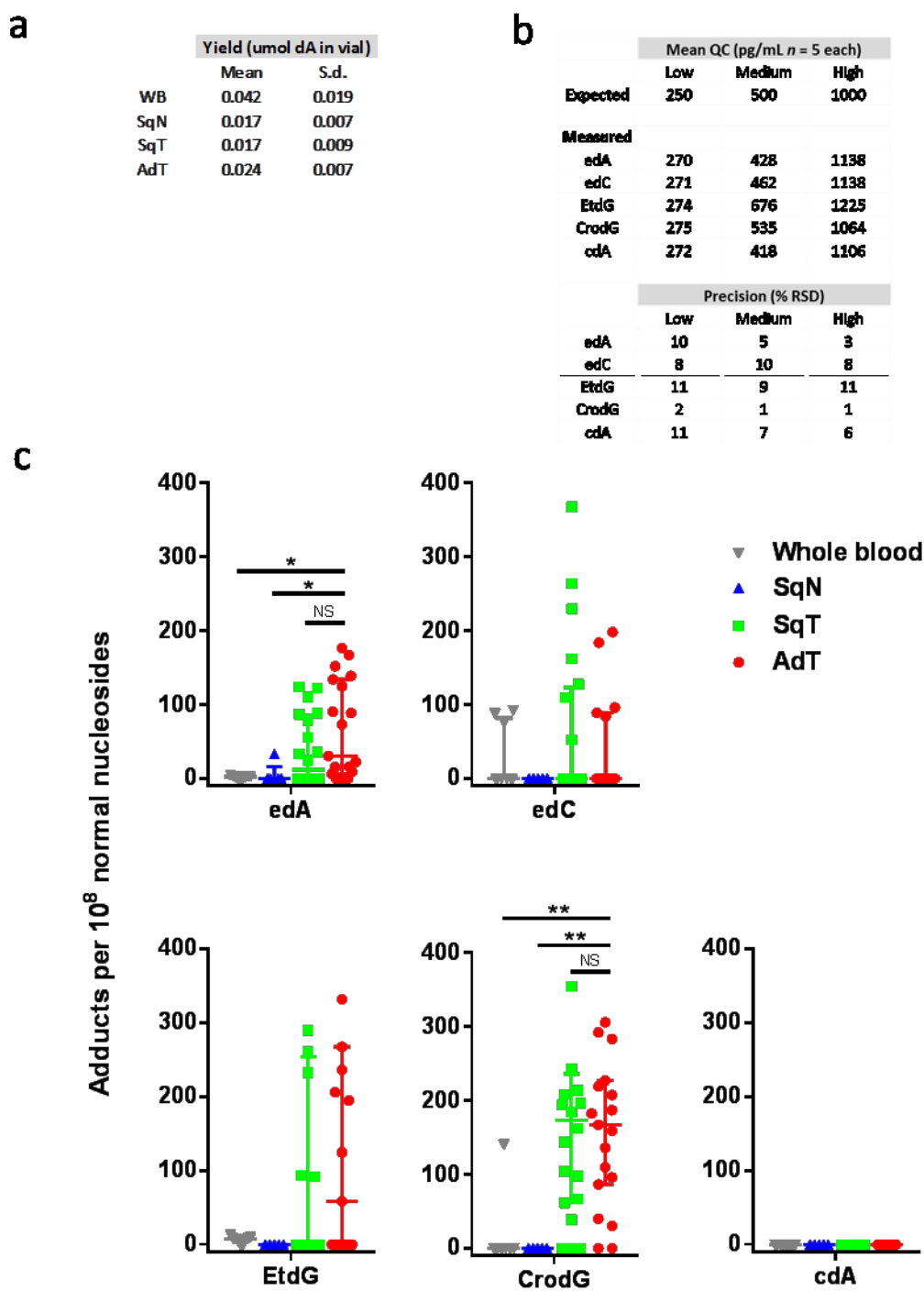
#### 4.1.5 Results

The validated UPLC-ESI-MS/MS method for nucleoside quantitation was applied to clinical samples to test the hypothesis that the five candidate modified nucleosides are enriched in malignant oesophageal tissues. No data was available to generate a power calculation; recent series investigating adducts in similar tissue have ten or less in each group (24,258–260). However, the need for statistical power needs to be balanced against the bulk tissue requirement, which is large (50-100mg) and thus prone to multi-clonality and inadvertent cell contamination.

Thus, in the main comparison – SqT vs AdT - sample sizes of twenty in each group were selected. Further controls were selected to control against field effects, including pooled biopsies of SqN (5 pools of 8-10 biopsies from different patients, to a total of 50mg per pool) and DNA extracted from whole blood from healthy volunteers (WB,  $n = 7$ ).

The samples were run in random order over two days with a fresh calibration at the start of the day. Low, medium and high concentration spiked quality controls were included after every fifth samples, following analytical guidelines. As detailed in Figure 29, the method performed stably throughout, with reasonable accuracy and precision values at these three concentrations.

All modified nucleosides were frequently above the limit of detection in DNA hydrolysates from both SqT and AdT, with the exception of cdA, which was not detected above the limit of detection in any sample. edA was detected in all tested tissue-types, although it was significantly enriched in AdT and SqT. Similarly, CrodG was the most commonly detected adduct, and again was significantly enriched in AdT compared to SqN and WB. CrodG was similarly enriched between AdT and SqT. edC and EtdG were enriched in a minority of SqT and AdT samples; there were no significant differences between the cohorts in these sample-sets.



**Figure 29: Nucleoside adduct quantification in clinical samples.**

Panel a, average quantity of deoxyadenosine in by sample type. Panel b, QC accuracy and precision for analytical run. Panel c, adduct quantity across the tissue-types. Sample and adduct abbreviations as previous. P-values calculated with Kruskal-Wallis test with Dunnett's multiple comparisons test, using AdT as the reference category.

#### 4.1.6 Discussion

The nucleoside-adduct quantitative method was demonstrated to be accurate and precise for the measurement of 5 targets to a relevant sensitivity in oesophageal tissues. The method took advantage of the selectivity and speed of the UPLC system in separating modified from unmodified nucleosides, removing the need for prior preparative HPLC separation. Through comparison to co-quantified normal nucleosides rather than input DNA amount, the method was able to present adduct burden in a way that accounted for variation in hydrolysis. The method was then applied to quantify nucleoside-adducts in benign and malignant oesophageal tissues. Concentrations of four of the five candidates were generally above the LOQ in the main sample-classes, and were significantly higher in normal squamous and adenocarcinoma tissue from the malignant oesophagus, compared to normal squamous or whole blood from patients with an endoscopically normal upper GI tract.

**Method development** A major goal of DNA-adduct quantitation is improved sensitivity(271). This aids accurate quantification in low burden samples, but more importantly rationalises the requirement for DNA input. Most established protocols call for 100µg DNA inputs(24,251,259,265), necessitating 50-100mg of oesophageal tissue, which is a huge amount (whole genome sequencing calls for 5µg). Larger tissue samples are more likely to contain contaminating cell types, be oligoclonal, and have heterogeneous proximity to influential factors (e.g. perfusion, or the oesophageal lumen). They are also practically difficult from a sampling perspective, especially in the oesophagus, where tissue is harder to access, carries significant risk to healthy patients, and where untreatable cancers can be relatively small. This is particularly important in the context of early disease, a critical subgroup for early diagnosis and transformation studies.

With this in mind, the study achieved excellent sensitivity for edA, edC, EtdG, CrodG, cdA, and odG, and was the best reported for EtdG, CrodG, and cdA, despite using 25 to 50mcg of DNA. One recent report(262) highlighted the use of UPLC-MS/MS for the sensitive quantification of edA and edC, reporting LOQs of 1.4 fmol and 1.2 fmol injection amounts respectively. These were similar to or five-fold less sensitive (edC) than the present work, and it should be noted that the LOQ definitions were also different (S/N 3 instead of 5). Moreover, these authors' chromatographic method was twice as long, and they did not present their data in the context of the normal

genome. Impressively, Chen et al achieved attomol sensitivity for edA and edC using nanoLC-nanoelectrospray-MS/MS(272), however their “assay LOQs” were similar to those established here. Both papers also reported similar adduct concentrations in whole blood to this study. Thus, the presented method matches practicality with versatility and excellent sensitivity,

Another reported limitation of the nucleoside-adduct approach to DNA toxicity studies is the relatively low throughput(271). A method development goal was to improve the assay lead time, which was accomplished by establishing a selective UPLC chromatographic run of just 6 minutes rather than 14-45 minutes(251,262), using the UPLC to separate unmodified from modified nucleosides during the analytical run rather than a two-step pre-analytical preparative HPLC fractionation(251,265), and opting for liquid-liquid extraction(259) rather than more labour-intensive solid phase extraction (271). This easily brought the DNA-to-peak time below 12 hours, and comfortably allowed 48 samples per day to be processed. DNA extraction time could not be improved upon and was a bottleneck - a high yield completely devoid of protein requires patience and care and remains a limitation of this and any other technique that needs bulk DNA input.

Despite establishing MS/MS parameters for all 7 targets initially identified, two were excluded from clinical phenotyping – cdG, which ionised poorly, performed badly in validation and could not be quantified in any biological samples, and odG, which ionised very well and was easily quantifiable in tissue samples, yet whose pre-determined isotope labelled internal standard Od3 had a co-eluting isomer in clinical samples, precluding its use in the derived method (the other internal standards performed inadequately in validation studies). Nonetheless, accurate quantification of the four aldehyde-nucleoside adducts, and one halo-adduct, was established to excellent sensitivity and selectivity, and this is the first report of combining these targets.

***Clinical phenotyping*** Nucleoside-adducts were subsequently quantified in hydrolysates of DNA isolated from oesophageal tissue samples, and whole blood. All four aldehyde adducts were significantly more concentrated in normal and malignant oesophageal tissue compared to blood DNA or SqN, which is consistent with the free aldehyde data presented in Chapter 2. The similar concentration across SqT and AdT types may also be consistent with the free aldehyde metabolomics study,

implying a metabolic field effect across the malignant oesophagus, as other studies have determined(56). As before, explanations include passive diffusion of these reactants from their true source, or a common reaction to a prevailing stimulus (i.e. germline genetics, gastro-oesophageal reflux, poor dental hygiene, smoking, alcohol etc).

The chief limitation of this study is that the sampling techniques for SqT and AdT (post-surgical tissue) were different from SqN (endoscopic biopsy), raising the possibility that cohort differences are artefactual. Surgical tissue may be ischaemic for ~3-5 minutes prior to snap-freezing. However, it is argued that this would not result in measurable adducts, as (i) there was no difference in free aldehyde concentrations in endoscopic or surgically-acquired tissue (see Table 11) (ii) in isotope tracing studies in models of ischaemia-reperfusion injury, authors have reported that aldehydes are not enriched by hypoxia, although their metabolism may be rewired away from oxidation to carboxylic acids(273). Nonetheless, this view requires confirmation by further comparative studies based on single sampling routes and more tissue controls, perhaps supported by relevant *in vitro* OAC models of hypoxic DNA damage.

The timing of genotoxicity is an important consideration in interpreting these studies. Prior evidence supporting this work suggested that *ALDH2* germline variants(105) and *ALDH1A2* polymorphisms predispose to oesophageal malignancies(103), indicating that aldehyde toxicity may occur very early in pathogenesis. In Chapter 2, it was shown that aldehydes are similarly enriched in Barrett's metaplasia tissue. In Chapter 3, it was determined that the majority of ALDH isoenzymes are not expressed in non-squamous tissues, consistent with the hypothesis that aldehyde exposure occurs early and may be a very prolonged – decades – before malignancy emerges. When OAC patients present to cancer services, they may have since altered their pathogenic habits, for example by losing weight, having stopped smoking or drinking alcohol, and by taking acid suppression medications. Thus the adduct burden would be less than at the time of transformation – yet in this study, aldehyde-adducts were still prominent lesions.

These data offer preliminary insights into the role of nucleoside-adduct analysis in the context of genotoxicity studies in OAC. To take these findings further from the analytical perspective, an interesting approach would be adductome analysis using a high resolution high sensitivity technique such as time-of-flight or ion trap mass

spectrometry(24,258,264), taking advantage of the [M-116]<sup>+</sup> transition to identify occult modified nucleosides in these samples. In this work, the one halo-adduct under investigation – 8-chloro-2-deoxyadenosine(274) – was not quantifiable in any sample, although this clearly is not sufficient to exclude halogenation as a mutagenic mechanism in OAC. Nonetheless, this highlights how adduct analysis can offer insights into the mechanics of disease-specific genotoxicity.

From the clinical perspective, the immediate extension of this work would be to verify that the identified aldehyde-nucleoside interactions are not spurious. If aldehydes can be confirmed as specific mutagens to OAC, manipulating their metabolism may offer novel therapeutic opportunities(275), or even as a novel means of prophylaxis against cancer for patients with Barrett's oesophagus. Additionally it will be of value to determine if aldehyde-based genotoxicity defines a clinically relevant subset of disease (34).

From a mechanistic perspective, it will be of interest to further dissect the contribution of each ALDH enzyme to genotoxicity, and further perturbation studies in vitro may be helpful, especially with better simulation of the microenvironment (e.g. conditioned media, hypoxic culture, co-culture with fibroblasts and macrophages etc). However a more attractive solution may be *C. elegans*, whose relatively small genome, rapid replication, and established variants lacking DNA repair machinery make direct dissection of aldehyde-based mutational patterns more achievable (276).



## 4.2 ALDH perturbation of normal and malignant oesophageal cell models

### 4.2.1 Methodological rationale

The simplest vital eukaryotic model systems are cells grown *in vitro*, as they are flexible, readily available, relatively cheap, and amenable to a growing toolkit of transient and stable genetic perturbation strategies. Currently, there are nine immortalised oesophageal cell lines; one derived from normal oesophagus (HET-1A) or oesophageal adenocarcinoma tissue (eight, see Table 12). The HET-1A normal line has been extensively criticised (209), was found to express p53 and  $\gamma$ -H<sub>2</sub>AX (data not shown), and was thus precluded from further analysis. Consultation of the literature, and participation in a national oesophageal cell model workshop revealed current trends for 2D or organotypic primary keratinocyte culture for normal models, and thus this was pursued. To compromise between cost, time, practicality and flexibility, a new application of a validated culture technique was selected that relies on the primary culture of keratinocytes freshly harvested for endoscopic or surgical resection tissue. This adapted a protocol defined for the culture of skin cells, which uses a Rho kinase inhibitor and mitotically-incompetent fibroblasts to preserve keratinocytes in a stem-like state, thus delaying terminal differentiation for up to 6 months (rather than 2-4 weeks). This “conditionally reprogrammed culture (CRC)” technique is effectively and conveniently reversed with withdrawal of the feeder layer and Rho kinase inhibitor, permitting experiments in a representative normal model system. There are no publications describing the application of this technique to oesophageal keratinocytes, and thus it was necessary to validate the first four cultures established with this technique using western blotting and immunofluorescence (shown in Figure 24).

For perturbation studies, first a pharmacological approach was selected using diethylaminobenzaldehyde (DEAB). This has been shown to broadly inhibit ALDH and stimulate aldehyde enrichment (277,278), and is thus a convenient tool to limit interferences from enzymatic redundancy when validating endpoint analytics. Next, to assess how each candidate ALDH may influence aldehyde phenotypes more precisely, several genes had to be suppressed in a facile and flexible method, transient transfection with small interfering RNA (RNA interference) was selected. The four cytoplasmic ALDH isoenzymes (-1A3, -3A1, -3A2, -9A1) that were

convincingly altered in OAC compared to squamous epithelium were selected for combinatorial silencing. The cell line OE33 was selected as it expresses relatively low levels ALDH isoenzymes (in particular *ALDH4A1*), decreasing the likelihood that inter-*ALDH* redundancy would mask metabolic phenotypes consequent to specific perturbation.

#### 4.2.2 Methods

**Cell lines and chemicals** The same cell models were used as in Chapter 3. Diethylaminobenzaldehyde (DEAB) was purchased from Sigma, diluted in dimethylsulphoxide (DMSO, Sigma) to a 1000x working solution, aliquotted and stored at -80°C. Toxicity experiments were conducted to determine the maximum tolerable doses of DEAB in each tested cell line. The final DMSO concentration was 0.1% in all wells. All experiments included unconditioned and DMSO-only controls.

**Transient ALDH knockdown using siRNA** Small interfering RNAs directed against *ALDH1A3*, *3A1*, *3A2*, and *9A1* were purchased from Thermo-Life (Silencer Select™ system), together with scrambled control RNA (Silencer Select™ Negative Control 1). As the intention was to undertake combinatorial knockdown (see below), only a single oligonucleotide was selected against each target. Cells were transfected using the Oligofectamine™ (Thermo-Life) lipofection reagent, following the manufacturer's instructions. Concentrations of siRNA and OFM were optimised in single perturbation experiments; typically, 5nM concentration with 6µl OFM in a 6-well format was sufficient to achieve a knockdown of >90% by qRT-PCR. Typically, cells were transfected at 30-40% confluency and harvested at 72 hours (RNA experiments) or 96 hours (protein and metabolite experiments). Cells were soaked in antibiotic-free Optimem™ media (Thermo-Life) for 20 minutes prior to transfection. The transfection mixture was replaced with normal complete media after 6 hours.

The relative difference in siRNA between single and multiple knockdown conditions was balanced using scrambled siRNA, so that the final concentration of total siRNA oligonucleotide was always 50nM. Given that the RNA-induced silencing complex (RISC) will have different specificities to different mRNA:siRNA duplexes, conditions optimised for single gene silencing may not exhibit the same knockdown efficiency when combined. Thus, siRNA concentrations were titrated experimentally until a silencing efficiency >90% was achieved for all targets.

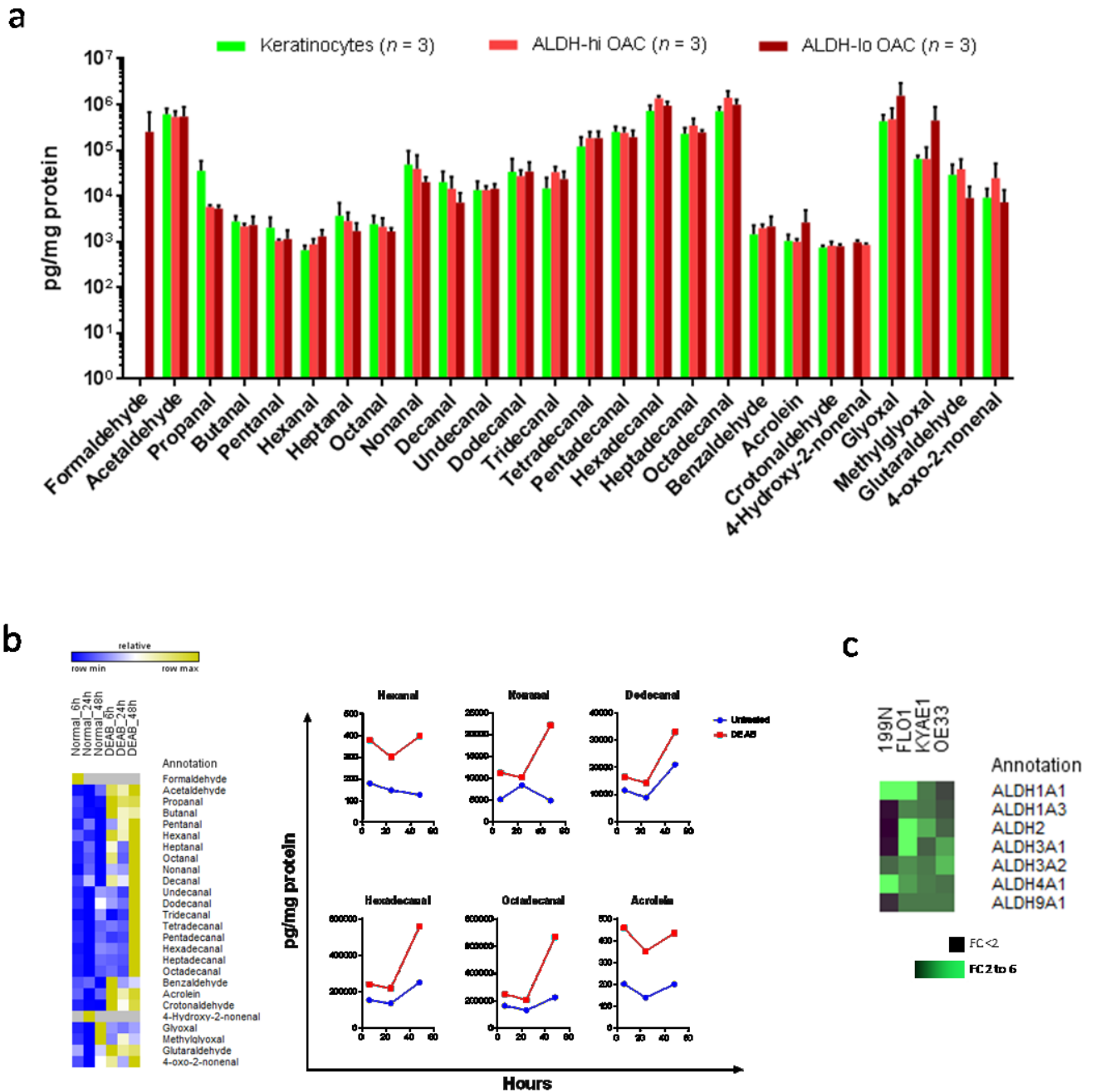
RNA for gene expression analysis was isolated using the RNeasy kit (Qiagen). Experiments were usually in 6 well plates; at the specified time-point, cells were washed three times in PBS, lysed directly in the well using 350µl of Buffer RLT from the RNeasy Spin Column kit (Qiagen). RNA was then precipitated, washed and eluted using the RNeasy kit according to the manufacturer's specifications, and then analysed as in Chapter 3. Protein experiments were undertaken as in Chapter 3. Metabolite quantitation was by the optimised UPLC-MS/MS method, as set out in Chapter 2.

### 4.2.3 Results

To begin *in vitro* dissection of ALDH-aldehyde relationships, the concentration of aldehydes were measured in a panel of nine lines, separated in three groups: primary keratinocytes (199N, 202N, 216N), ALDH-hi OAC lines (OE19, ESO51, ESO26), and ALDH-lo (FLO-1, OE33, OACM5.1). Cells were seeded at 40% confluence, grown for strictly 48 hours in fresh media, and then metabolically quenched and aldehydes quantified (see Chapter 2.2.11). Differences in biomass at 48h were accounted for by normalising aldehyde concentrations to total protein input. Typically, each well contained 300-400µg of protein per well, and so all limits of quantification were <1000pg/mg protein (apart from MDA, 2.5ng/mg protein)

Nearly all aldehydes were consistently above the limit of quantification (see Figure 30). As in primary tissue samples, concentrations of C2, C9, C15-18 and glyoxal were the most abundant. In these intracellular extracts, mid-chain aldehydes were also present in robust quantities. Compared to keratinocytes and ALDH-hi OAC lines, there was 2-5 fold increased concentrations of acrolein, glyoxal and methylglyoxal in ALDH-lo lines, although this was not significant after multiplicity correction. Hydroxynonenal was not detectable in keratinocytes, and present in OAC lines, although the concentrations were close to the LLOQ.

The global ALDH inhibitor diethylaminobenzaldehyde (DEAB) has been shown to enrich aldehydes in *in vitro* systems (278,279). This experiment was repeated with the present validated UPLC-MS/MS aldehyde method, to first test whether the technique is sensitive to known mediator of aldehyde flux (i.e. a biological method validation) and secondly whether pharmacological ALDH inhibition is sufficient to



**Figure 30: Basal and ALDH-inhibited aldehyde phenotypes of oesophageal cell models.**

Panel a, concentration of all intracellular aldehydes above the LLOQ in a series of 9 cell models divided into three ADLH-defined groups. Panel b, Effect of 200uM diethylaminobenzaldehyde (DEAB) on OE33 cells over 48 hours (row normalised heatmap at left, selected compounds at right). Panel c Effect of DEAB on *ALDH* expression in some oesophageal cell models

enrich aldehydes in a relevant OAC model. The OE33 line was selected as it is a low *ALDH3A1* expressor, which is the only ALDH isoenzyme shown to metabolise DEAB(277). Toxicity experiments showed 200uM DEAB to be the maximum non-toxic dose in this cell line, in keeping with the previous reports(278,279). As seen in Figure 30b, several aldehydes became enriched after only 6 hours in 200uM DEAB; all target compounds increased in concentration by 2 to 5 fold after 48 hours.

This data suggests that the target compounds are in continuous production even under normal metabolic states, or at least that DEAB provokes their production. To test whether cells are sensitive to these changes, *ALDH* expression in a two normal and two malignant cell lines was checked after 48h 200uM DEAB. As seen in Figure 30c, the OAC lines displayed enhanced expression of several *ALDH* isoenzymes of interest, including *ALDH3A1/2*, *-4A1* and *-9A1*, although these were subtle (2-5 fold changes). The tested keratinocytes only increased expression of *ALDH1A1* and *4A1*, indicating context-dependent expression in which cell-autonomous mediators are dominant compared to metabolic sensors.

Finally, an attempt was made to assess how specific *ALDH* isoenzymes influence overall aldehyde phenotypes, using a combinatorial RNAi strategy. Robust siRNA-mediated knockdown of the transcript targets was noted both at an RNA and protein level, both for individual genes and in combination (see Figure 31a&b). Expression of non-silenced target genes was relatively stable, indicating that target silencing did not lead to rebound expression of functionally related *ALDHs*. Both individual and combinatorial knockdown of *ALDH1A3* and *3A1* seemed to be associated with increased H<sub>2</sub>A<sub>X</sub> phosphorylation, however protein carbonylation was particularly high in this experiment and no differences could be elicited.

Consequences of *ALDH* silencing on aldehyde metabolites were complex and in places contradictory (see Figure 31c). Knockdown of *ALDH3A1* individually resulted in a 2.5 fold increase in butanal (8975 pg/mg protein vs 2952 pg/mg protein) and benzaldehyde (2561 pg/mg vs 986 pg/mg), although similar changes were not apparent when *ALDH3A1* was silenced in combination with other target genes. However, for acrolein, crotonaldehyde, glyoxal, methylglyoxal and glutaraldehyde, the highest concentrations were observed when all four targets were silenced, with consistent and lower concentrations observed in lesser combinations, potentially indicating genuine metabolic effects for *ALDH* loss (see Figure 31c).

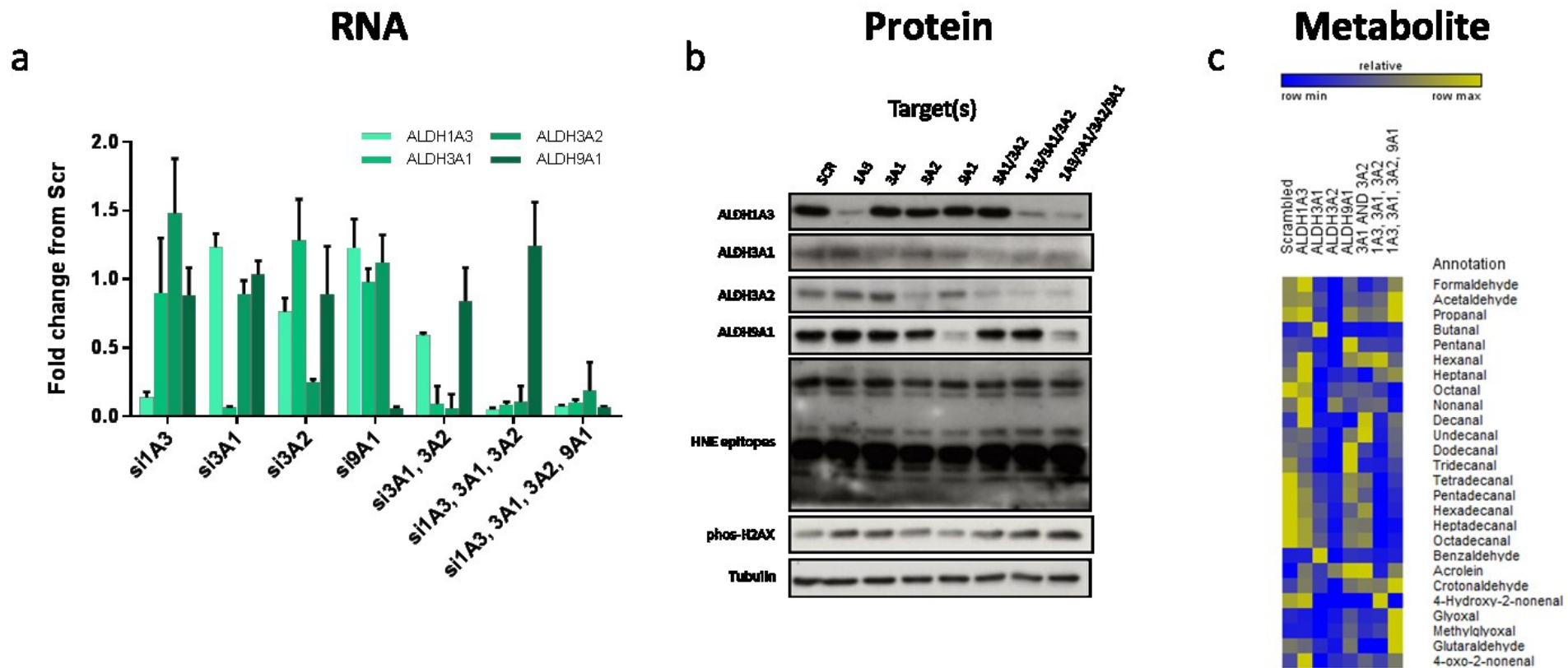


Figure 31: Effects of individual and combination siRNA knockdown on aldehyde metabolic phenotypes.

Panel a siRNA-mediated knockdown efficiency at RNA level using qPCR. Panel b, siRNA-mediated knockdown efficiency at protein level quantified using immunoblotting. Panel c, row-normalised relative aldehyde concentrations in the respective ALDH-perturbed wells (intracellular metabolites), measured by UPLC-MS/MS

#### 4.2.4 Discussion

The purpose of these experiments was to use the validated free aldehyde method to explore the interplay between the candidate *ALDH* isoenzymes and aldehyde metabolic phenotypes. Under optimal growth conditions, keratinocytes and adenocarcinoma cells displayed strikingly similar aldehyde concentrations, which is probably unsurprising as glutathione and other redox co-factors are added in excess. The most enriched carbonyls – C2, C9, C15 to C18 – were as in the primary tissue analysis in Chapter 2, indicating that the relative aldehyde metabolic landscape is remarkably stable in eukaryotic cells.

Pharmacological inhibition of *ALDH* in OE33 cells with DEAB successfully enriched the tested aldehydes at 48 hours, in keeping with previous reports in lung cancer cell lines (278). Thus, this data serves several purposes: (i) adequate *ALDH* inhibition is sufficient to enrich aldehydes in oesophageal adenocarcinoma cells (ii) aldehyde metabolic flux is above the LLOQ for the present UPLC-MS/MS method (iii) the assayed aldehydes are continuously produced, and may build up if metabolising systems are sufficiently impaired.

DEAB may be actively metabolised by at least one *ALDH* isoenzyme (*ALDH3A1*)(277), suggesting it may be an imperfect inhibitor. Additionally, it may act through indirect pathways to enrich aldehydes – for example, by depleting co-factors such as GSH and NAD(P)<sup>+</sup>. To overcome these limitations, a precision perturbation strategy (siRNA) was selected to assay the importance of some candidate *ALDH* isoenzymes to aldehyde metabolism in OE33 cells. Although there may have been genuine enrichment of enals and dialdehydes when all four candidates were silenced, metabolic profiles arising from these perturbations were not convincing, presumably owing to redundancy between these and unmeasured isoenzymes, in addition to alternative detoxification routes. In this experiment ideal culture conditions were used including 21% oxygen, fresh co-factors and glutathione after transfection, and appropriate cell seeding. Thus, it is unlikely that this constitutes a valid representation of *in situ* aldehyde metabolism, and it will be of interest to repeat these experiments with culture conditions which simulate increased carbonyl stress. The degree of silencing may have been insufficient to reveal metabolic effects; this may also be addressed by depleting anti-oxidant reserves with conditioning experiments, or an alternative silencing strategy such as small hairpin RNA.

There are few reports of measuring aldehydes with mass spectrometry after targeted *ALDH* perturbation *in vitro* (278), although several groups have reported *in vitro* aldehyde phenotyping studies of basal states (181,280–282). These studies fail to convincingly replicate clinically-defined biomarkers, implying that context-specific effects are essential (although there are methodological concerns regarding these reports – for example, measuring metabolites from a 2L suspension of normally adherent cells over 2 hours is unlikely to generate representative data (280,283)).



## 4.3 Metabolic and non-metabolic functions of ALDH3A2

### 4.3.1 Methodological rationale

In Section (4.2), it was shown that *ALDH* isoenzymes must be widely suppressed to enrich aldehyde phenotypes. Nonetheless, it was decided to explore the function of *ALDH3A2* in more depth using more sophisticated perturbation strategies, following several notable reasons. First, low expression of *ALDH3A2* was clearly associated with adverse clinical outcomes, including disease progression and poorer survival. Second, mutations in this gene cause the neuro-cutaneous disease Sjogren-Larsson syndrome, which is characterised by fits, spasticity, mental retardation, and most intriguingly, hyperkeratosis. The finding of a specific clinical syndrome attributed to this gene suggests functional non-redundancy. Third, *ALDH3A2* has specificity to fatty aldehydes, which are the most abundant in the studied samples and systems and may be of functional importance.

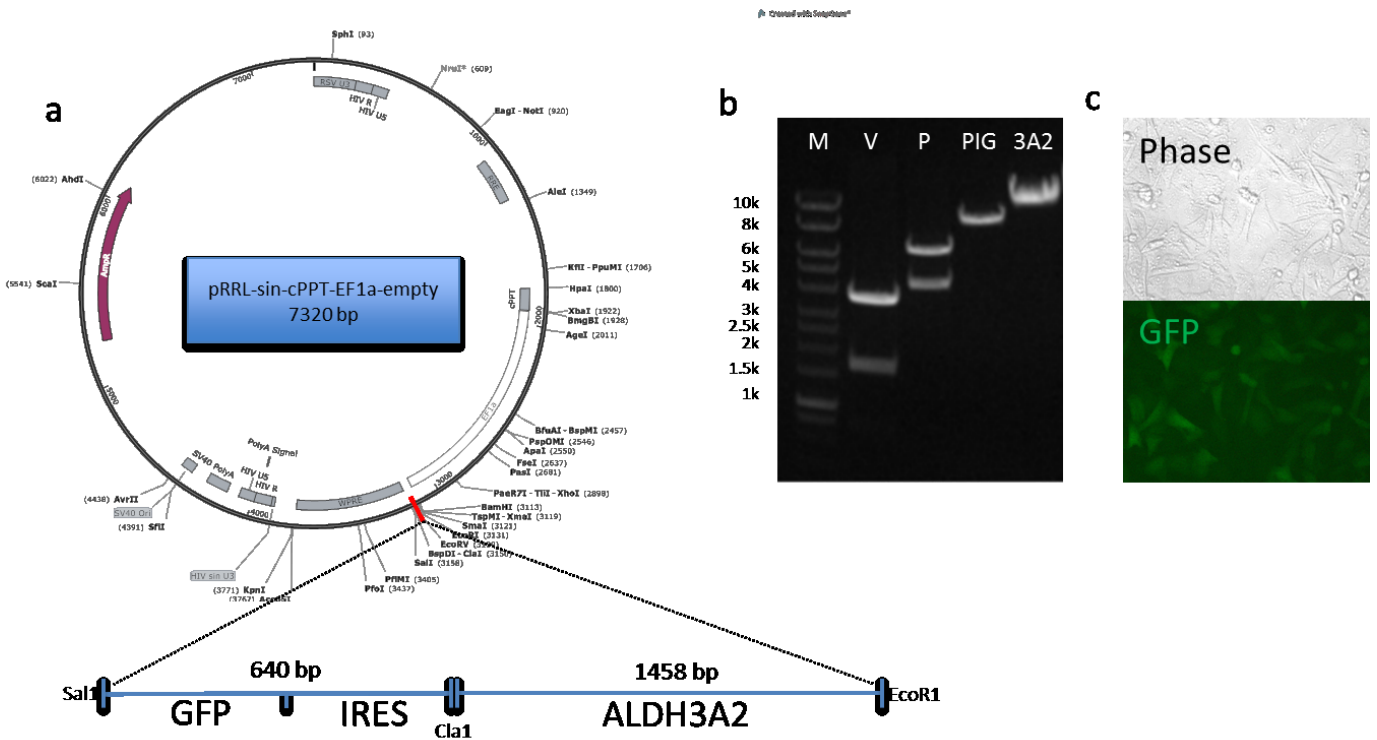
Given the difficulties encountered in the previous section, it was decided that the base model for this work should be relatively ALDH-lo, to maximise the potential for measurable functional endpoints to emerge. Thus the FLO1 OAC line was selected, as this grows relatively rapidly and poorly expresses all *ALDH* genes except for *ALDH3A2* and *-4A1*. To knockout *ALDH3A2*, CRISPR-Cas9 editing was selected (284). *ALDH3A2* has 26 transcripts of which at least 10 encode proteins, and fortunately all transcripts >100 amino acids share the same start ATG. This ATG is surrounded by high specificity cut sites for Cas9 editing, and thus a clean edit to comprehensively disrupt this gene with zero residual expression is easily achievable. However, the requirement for isogenic cloning means that further control conditions are needed to ensure the intended edit is responsible for measured phenotypes, and this can be easily done by reintroducing the *ALDH3A2* open reading frame (ORF) with a promoter using a lentivirus. Functional endpoints were tailored to test why *ALDH3A2* may influence survival and are described below.

#### 4.3.2 Methods 1: Stable overexpression of *ALDH3A2* by lentiviral transduction

**Plasmid cloning:** The viral coding plasmids pRRL-sin-cPPT-EF1a-multi clone site (MCS), the viral packaging plasmids psPAX2 and VSVG (see Appendix 6), and the internal ribosome entry site-green fluorescent protein sequence (IRES-GFP), were kindly provided by Prof Nagy Habib (Division of Surgery, Imperial). These plasmids were transformed into TOP10 chemically competent cells (Thermo-Life), maxi-prepped (Qiagen), and checked by restriction digest using EcoR1 (New England Biolabs, NEB), agarose gel electrophoresis for predicted fragment size, gel extracted and column purified (Qiagen), and Sanger sequenced.

The pRRL-sin-cPPT-EF1a-*ALDH3A2*-IRES-GFP viral plasmid was made by sequential addition of *ALDH3A2* and then IRES-GFP (see Figure 32a). Initially, the IRES-GFP fluorescent selection cassette was amplified using primers tagged with appropriate sticky ends to enable Cla1 and Sal1 restriction digest (NEB) (see Appendix 3 for oligonucleotide sequences). The pRRL-sin-cPPT-EF1a-MCS plasmid was linearised using the same two digest enzymes, and treated with calf intestinal phosphatase (NEB) to inhibit re-circularisation. Both the digested plasmid and insert were then separated from undigested counterparts by 1.5% agarose gel electrophoresis, and column purified (Gel Extraction kit, Qiagen). The column eluates were then quantified using absorbance at 260:280nm, and the insert was ligated into the vector in a 3:1 molar ratio, using the quick ligase kit (NEB). The recombinant plasmid was immediately transformed into ultra-competent cells (Alpha select GOLD, Bionline), cloned on ampicillin-agar plates, mini-prepped (Genejet, Thermo Fisher), redigested as above, and checked on an agarose gel. Clones of interest were then Sanger sequenced

This process was repeated to insert the *ALDH3A2* open reading frame (ORF) between the EF1a promoter and the IRES. The full length *ALDH3A2* ORF was amplified from reverse transcribed (SuperScript III, as before) mRNA of the 199N primary keratinocyte culture, with primers tagged with appropriate sticky ends for Cla1 and EcoR1 restrict digest (NEB). This amplicon was then digested and purified, ligated into pRRL-sin-cPPT-EF1a-IRES-GFP. The complete plasmid was then cloned, prepped, checked by restriction digest (see Figure 32b) and Sanger sequenced. A clone expressing the correctly annealed and orientated complete plasmid was then Maxi-prepped ready for production of lentivirus.



**Figure 32: Production of ALDH3A2-IRES-GFP lentivirus.**

Panel a, cloning plan of the mammalian lentivirus encoding plasmid in bacteria. A larger pRRR map is given in Append 2. Panel b, agarose gel separation of EcoR1 digests of VSVG (V, 1651 and 4173 base pairs), psPAX (P, 4374 and 6329), pRRR-sin-cPPT-IRES-GFP (PIG, 7960 bp), and the complete plasmid (3A2, 9418 bp). Panel c, brightfield and green fluorescent images of GFP-sorted FLO1 cells transduced with PIG-3A2.

**Virus production and eukaryotic transduction:** All viral work was subjected to enhanced biosafety approvals in line with College rules for infective bioengineering (approval reference: GMIC-5645). Strict viral safety precautions were used throughout. To make lentiviral particles, the pRRR-sin-cPPT-EF1a-ALDH3A2-IRES-GFP, psPAX2 and VSVG plasmids were then all transfected (Lipofectamine 3000, Life) into 50% confluent HEK-293T packaging cells in a ratio of 15:4:1, in the presence of antibiotics. These cells carry the large T antigen, which allows the transfected plasmids to be expressed at high copy number by binding to and inactivating p53 and p105-Rb. IRES-GFP control lentiviral particles (i.e. without ALDH3A2 ORF) were also made. Transfection conditions were optimised using different combinations of total DNA and lipofectamine reagents, according to the manufacturer's instructions. The lentivirus-containing media was then collected at 48-60 hours post-transfection. Debris and cells were removed by passing the virus-containing media through a 100 nm cellulose syringe filter. The multiplicity of infection (MOI) was estimated by transducing target cells with serial dilutions of

lentiviral particles. Stable ALDH3A2-GFP and GFP-only expressing mammalian cell lines were established by transducing 40% confluent cells with sufficient virus to achieve an MOI of 1 at 96 hours, as assessed by visual inspection of the GFP-expressing cells. The transduced population was then cultured for a further week. A GFP-low population was then enriched using fluorescence-activated cell sorting, taking the view that the least fluorescent population would have an average integration ratio closer to 1 (See Figure 32c).

#### 4.3.3 Methods 2: *ALDH3A2* knockout by CRISPR-Cas9 editing

See Figure 33a for a schematic of the CRISPR-Cas9 workflow.

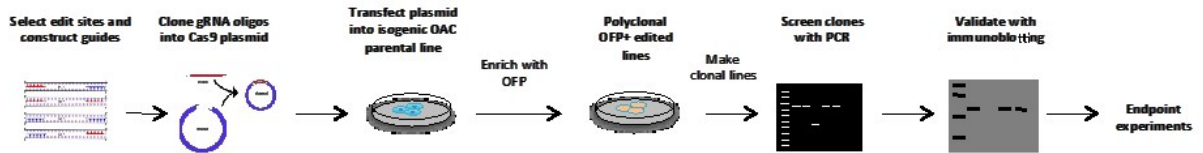
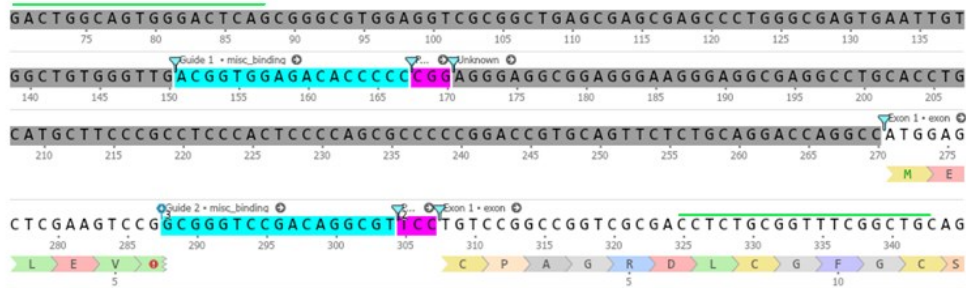
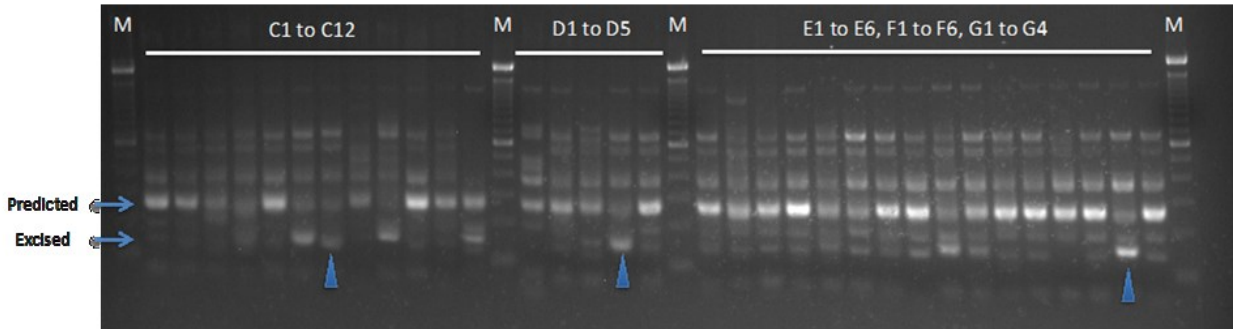
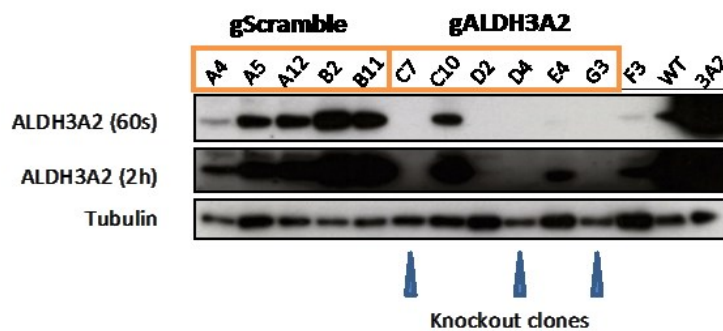
**Edit design:** Since gene editing using the CRISPR-Cas9 system was first described in 2012(284), it has evolved very quickly with already high-sensitivity third generation enzymes being produced. The experiments for this project were designed in 2014 using the best available technology at the time(285). Guide RNA design and targeting was conditional on three factors: (i) minimisation of off-target effects (ii) predictable editing (iii) definitive gene disruption. Thus, the GeneArt single-vector CRISPR-Cas9 system was selected for editing experiments, using established guide design tools (<http://crispr.mit.edu/>)(285). Edit sites were selected such that all functional splice variants would be affected, and predicted off target were always in inter-genic regions and no more than 80% homologous to the targeting guide. For *ALDH3A2*, one guide targeting a CRISPR site 20 bases 5' of the start ATG and a second guide targeting exon 1 were selected as they displayed a selectivity >99.1%. This meant the start codon for all 15 recognised splice variants is excised with this strategy. Alternative reading frames have no start-to-stop coding regions greater than 20 amino acids. This two-guide strategy was chosen to (i) facilitate screening of isogenic clones by predictable amplicon modification (ii) ensure a complete absence of translation. Prior to editing, targets regions were checked for predicted sequence by genomic PCR and Sanger sequencing. See Figure 33b for sequence topology. In addition to *ADLH3A2*-targetting guides, another plasmid with scrambled guide RNA was made as an editing control (supplied with kit).

**Construction of CRISPR plasmid, cloning, sequencing:** The Geneart CRISPR-Cas9 single vector kit (Thermo Fisher) was used, according to the manufacturer's

instructions. Briefly, sticky-ended oligonucleotides matching the selected guide sequences were ligated into a linearised vector containing tracrRNA and Cas9 enzyme coding regions, with relevant promoters, antibiotic selection regions etc. The circularised plasmid also contains an orange fluorescent protein (OFP) reporter gene for enrichment of edited cells. The circularised plasmid is then used to transform competent bacteria, which are cloned, expanded, prepped for plasmid DNA, and sequenced for successful ligation and directionality. A colony expressing the correct plasmid is then further expanded.

**Optimisation of CRISPR transfection conditions** Initially, the Lipofectamine 2000 reagent was chosen to transfect the CRISPR-Cas9 plasmids into mammalian cells. Optimisation of transfection efficiency was attempted by titrating concentrations of plasmid DNA and lipofection reagent to maximum OFP-expressing cells (checked with flow cytometry, FACSCalibur, BD, according to manufacturer's instructions), with <10% cell death as assessed morphologically. However, the transfection efficiency was poor (optimally 10%), and therefore the Lipofectamine 3000 reagent was used and similarly titrated. Optimal transfection efficiency was improved to 40% with this approach; it is thought the improvement is due to transfection of this larger sized plasmid (10kb) being facilitated by this agent's supercoiling step.

**Bulk transfection, fluorescence activated cell sorting and establishment of isogenic clones:** In addition to checking transfection efficiency, the OFP reporter allows for enrichment of transfected cells by fluorescence activated cell sorting (FACS). This work was undertaken at the Clinical Sciences Centre Flow Facility (Hammersmith Hospital Campus, Imperial College), using centre protocols. Bulk transfection for enrichment was undertaken using optimised transfection conditions at 10cm dish scale. The experiment was planned around a FACS appointment 48-72 hours after transfection. One million cells were seeded and grown to 60% in the presence of antibiotics. Lipofection was carried out in the morning, again in the presence of antibiotics, and the media was replaced at 6h. Shortly before sorting, cells were trypsinised, washed once in sterile PBS, and resuspended in cell sorting buffer. Cells were collected into usual media, supplemented with glutamine, gentamicin, and amphotericin, and 20% fetal calf serum (these broad-spectrum antimicrobials are selected owing to their relative thermostability). Enriched cells were cultured for one week in the same media, replaced fresh every other day, and then changed to normal culture conditions.

**a****b****c****d**

**Figure 33: Development of *ALDH3A2* knockout FLO1 cells using Cas9 editing.**

Panel a, Workflow schematic of CRISPR-Cas9 technique; Panel b, *ALDH3A2* edit design at Exon 1. The selected guide sequences (blue) and NGG palindromic PAM sequence (pink) are given, as well the sites of the check primers for the region (green). Note that guide 2 is anti-sense. Panel c, clone screening also PCR of edit sites. Candidate edited clones are indicated (C7, D4, G3). Panel d, functional validation of candidate clones by immunoblotting. The successfully edited clones are indicated.

Once the enriched population had grown to approximately 5 million, it was checked for mycoplasma, stocks were frozen, and isogenic clones created. Single cell colonies were generated by serially diluting a triple-counted population (Muse Cell Analyser, BD) to a concentration that gave a final colony efficiency of 20% of the wells (typically 4-6 cells per ml). This titration reflected how well lines tolerated trypsinisation and isogenic cloning, and the colony efficiency was kept low to minimise multiclonality. This final suspension was used to seed five 96 well plates per edit (150µl per well), and the colonies were established over 2-4 weeks. Media was replaced every third day. The outer wells of the plate contained solely PBS, to protect the remaining wells from dehydration. Established isogenic clones were appropriately passaged until 60 stable populations were growing in a single 96 well plate.

**Edit screening** The single 96 well plate was grown to 60-90% confluence, and split 1:6, such that four plates were frozen whole as stock plate, one plate used for passage, and one for edit screening. Typically, 30µl of 0.025% trypsin/EDTA per well was used. For clone screening, 5µl of the trypsin-cell suspension was added to 25µl of DirectPCR lysis buffer (Viagen Biotech) containing 0.2mg/ml proteinase K (Sigma). Cells were then lysed at 55°C for one hour, followed by 85°C for 45 minutes to denature the proteinase K.

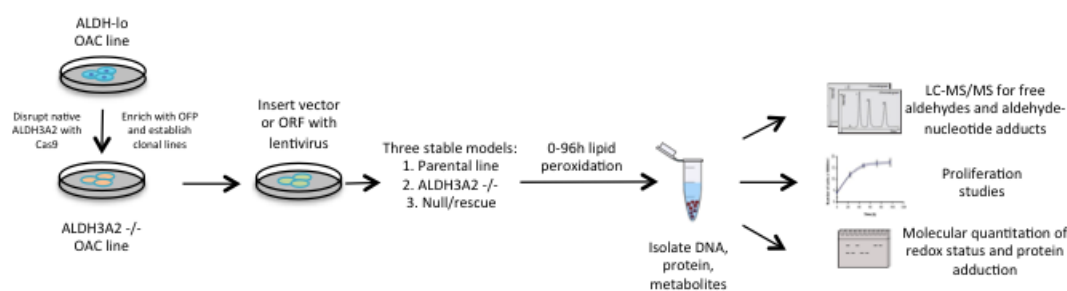
This lysate was used as a template for the generation of amplicons using the Phusion Phire master mix (Thermo-Fisher). Master mixes were made in the ratio 10µl :6:1:1 for polymerase master mix : water : forward primer (10uM stock) : reverse primer (10uM stock), and aliquoted into a 96 well PCR plate using a multichannel pipette. Two microlitres of template was then added to each well, and mixed by pipetting 10 times. Thermal cycling was then carried out according the Phusion instructions, and with the calculated annealing temperatures of the custom oligonucleotides. The PCR products were then analysed by agarose electrophoresis using a 2% gel on a 25cm gel tank (Biorad), using the 48 well comb and a 100 bp DNA size marker (TrackIT DNA ladder, Thermo-Fisher)

**Edit validation** Six to ten candidate clones were then validated in two ways. Functional gene disruption was confirmed using immunoblotting (see Section 3.3.7 for methods and

Figure 33d for results). Actual gene disruption, with allelic discrimination, was verified using topoisomerase-based cloning (Zero Blunt TOPO cloning kit, Thermo-Fisher). This system uses a linearised cloning vector with covalently attached topoisomerase enzyme at both blunt ends. This construct can be attacked by a blunt-ended PCR product, releasing the topoisomerase and circularising the vector; the TOPO vector cannot be circularised without an insert. Therefore, this system enables both rapid and high efficiency cloning. The Zero Blunt vector instructions were carefully followed. Briefly, a ligation reaction was carried out in the vector:insert:salt:water ratio 1µl:1:1:3. The ligation mixture was then used to transform competent bacteria, which were then spread onto ampicillin plates, picked, expanded, plasmid purified and Sanger sequenced using a high-throughput 96 well plate format. At least five clones per transformation reaction (i.e. per clone) were checked, to generate ratio to estimate allelic balance.

#### 4.3.4 Methods 3: Endpoint assays

To gain functional insights into the tumour suppressor qualities of *ALDH3A2*, the constrained OAC model was assessed using a panel of metabolic and non-metabolic endpoint experiments. Figure 34 highlights the work undertaken.



**Figure 34: Functional phenotyping of *ALDH3A2* in vitro**

**Cell viability (& count)** Cells were seeded in a 96 well plates in sextuplet for each condition at a density of 30% (typically 1000 cells for FLO1 and OE33 cells). At the relevant timepoint, media was carefully removed and the cells then incubated in 0.3mg/ml (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) (MTT) in Optimem media (both Thermo-Life) for 3 hours. The media was again aspirated very carefully, and the undisturbed formazan crystals dissolved in 100µl DMSO. The absorbance at 575nm was then measured using a colorimeter.



**Cell apoptosis and cell cycle assays** Both were measured on a Muse personal cell analyser, using the brand's platform-specific kits. Briefly, cells were fixed, permeabilised, stained, washed and read by flow cytometry. For cell cycle, cells were serum starved for 24 hours to synchronise cycling; the kit's staining reagents were propidium iodide and a fluorescently labelled antibody targeting annexin V.

**Immunoblotting:** Was as described in Chapter 3.3.7. Additional primary antibodies targeting the phosphorylated histone marker H2AX (Cell Signalling, 1:2000) and the hydroxynonenal-protein adduct (HNEJ, Abcam, 1:100) were also used.

**Redox status:** The ratios of the reduced and oxidised forms of the aldehyde dehydrogenase co-factor NADP<sup>+</sup> (NADP-Glo™ kit, Promega), and the aldehyde-binding detoxification factor glutathione (GSH-Glo™, Promega), were quantified with colorimetric kits following the included protocol.

**Free cell aldehydes:** Cell samples were incubated for 0-96 hours with or without pro-peroxidation systems, typically in 6-well format. For FLO1 cells and derivatives, the typical seeding density was 80,000 cells per well. For extracellular metabolites, an aliquot of media was directly removed from the well at a specified timepoint and processed as per the optimised and validated technique detailed in Chapter 2.3. Final aldehyde concentrations were normalised to the total protein of the well, quantified using the bicinchoninic acid assay as in Chapter 3.3.7. For intracellular metabolites of adherent cells, 200µl of dry-ice cold 50:50 acetonitrile:water was added directly to the well, and the well contents were scraped using a lifter. The solvent was immediately removed to a pre-chilled microcentrifuge tube and spun at 14500 rpm for 3 minutes in a pre-chilled centrifuge. The supernatant was either immediately frozen at -80 or directly decanted to a fresh tube containing DNPH as per Chapter 2.3. Final aldehyde concentrations were normalised to the total protein content of a parallel well.

To simulate lipid peroxidation, cells were incubated with a mixture of ferrous sulphate (to generate free radicals) and ascorbic acid (to recycle the Fe (III)), as previously described (286).

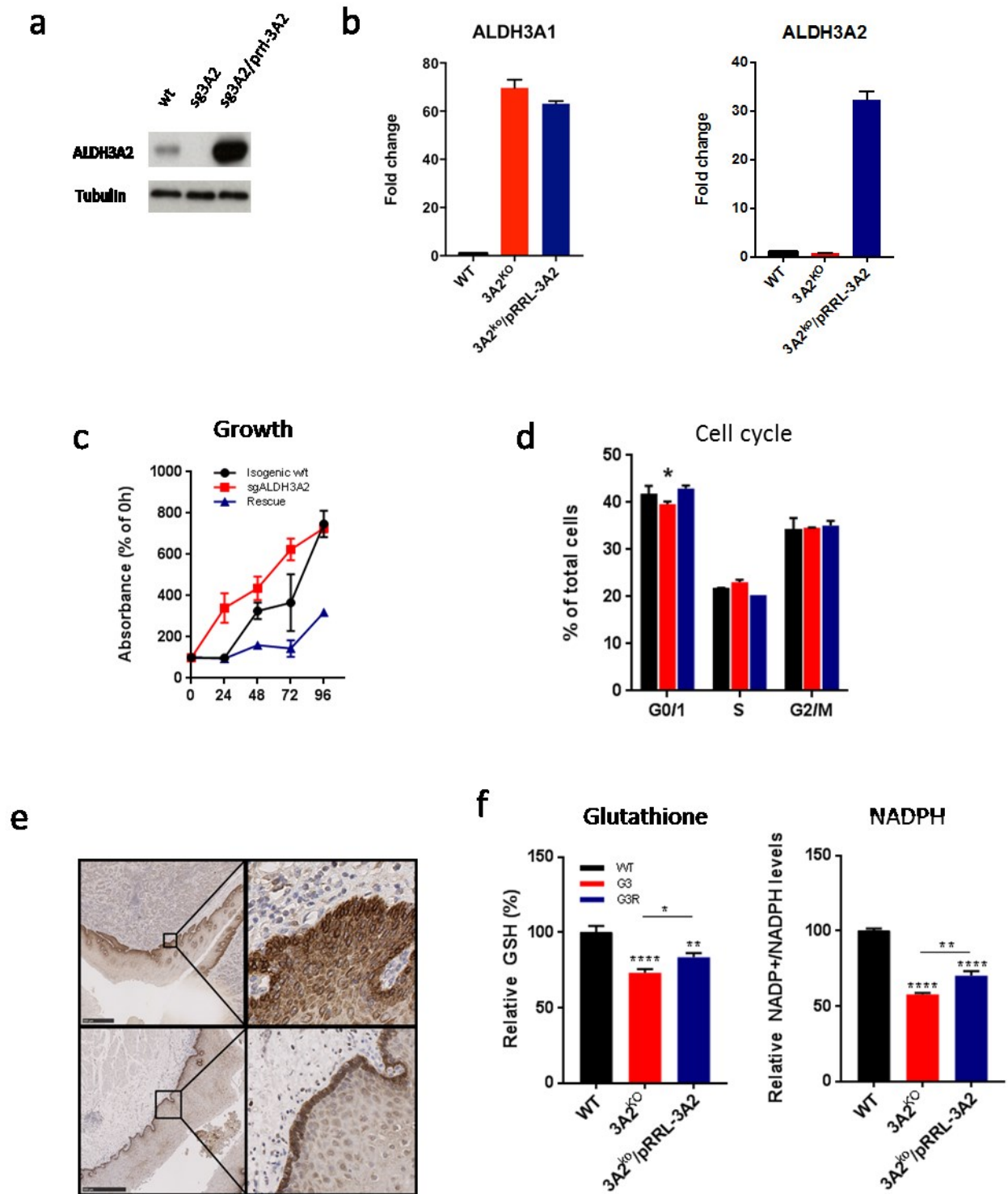
#### **4.3.5 Results**

CRISPR-Cas9 editing to remove the *ALDH3A2* start ATG in isoclinal FLO1 cells produced at least three clones with successful complete knockout of *ALDH3A2*, verified at a protein level.

Given that low *ALDH3A2* expression was associated with adverse clinical outcomes, it was decided to check whether this gene affects cancer-specific phenotypes. An isogenic FLO1 parental line, a FLO1<sup>3A2-/-;PRRL-EV</sup> line (G3-EV) and FLO1<sup>3A2-/-;PRRL-3A2</sup> rescue line (G3-3A2) were selected for these experiments, and *ALDH3A2* expression was validated at the protein and RNA level (see **Figure 35a&b**). Interestingly, there was rebound expression of *ALDH3A1* in the knockout line. The G3-EV null line grew significantly faster than either the parental or rescue lines, as measured by MTT assay (see **Figure 35c**). After 96 hours, MTT activity equalised between parental and G3-EV cells, owing to cell confluency in the wells.

To check whether the different growth rates were due to increased replication or decreased cell loss, cell cycle and apoptosis markers were checked using flow cytometry. There was no difference in the apoptotic cell populations between the three lines, but G3-EV cells were less likely to be in the G0/1 phase of replication, suggesting cell cycle activation explains the growth characteristics of these cells (see **Figure 35d**). As described in Section 3, the most intense staining occurred *in situ* in basal keratinocyte, thought to contain the slow-cycling replenishing cell population (see **Figure 35e**)(287).

To further assess the metabolic consequences of stable *ALDH3A2* knockout, same three lines were assessed for redox status (see **Figure 35f**). *ALDH3A2* exerts redox influence through its co-factor NADP<sup>+</sup>, and is thus a source of reducing agents by generating NADPH. In turn, cells recycle NADP<sup>+</sup> by reducing glutathione to its reduced form GSSG. Thus, the ratio of these species is a surrogate of the relative redox status of these *ALDH3A2*-controlled lines. As seen in **Figure 35f**, G3-EV cells had significantly decreased anti-oxidant signature, characterised by lower NADPH/NADP<sup>+</sup> and GSH/GSSG ratios. These effects were partially reversed by transduction of exogenous *ALDH3A2*.

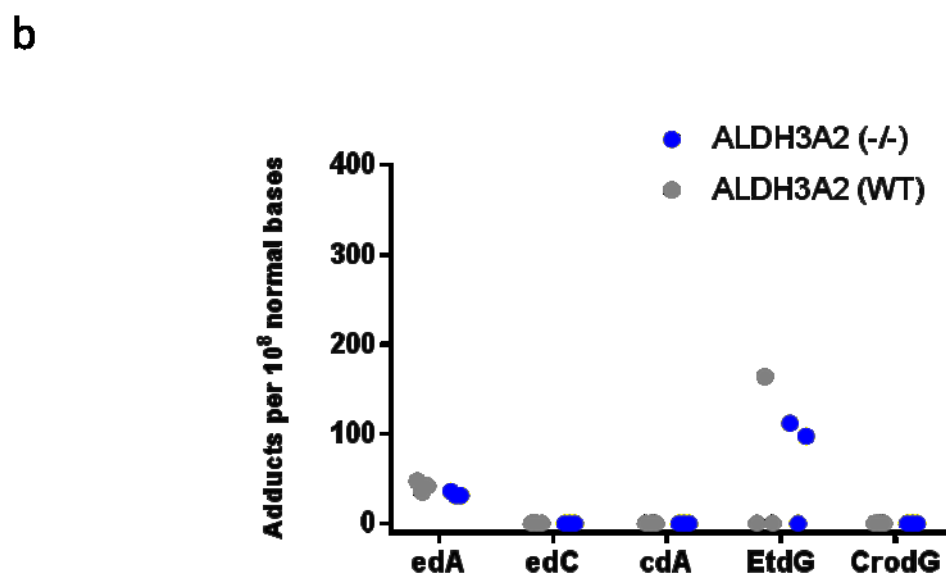
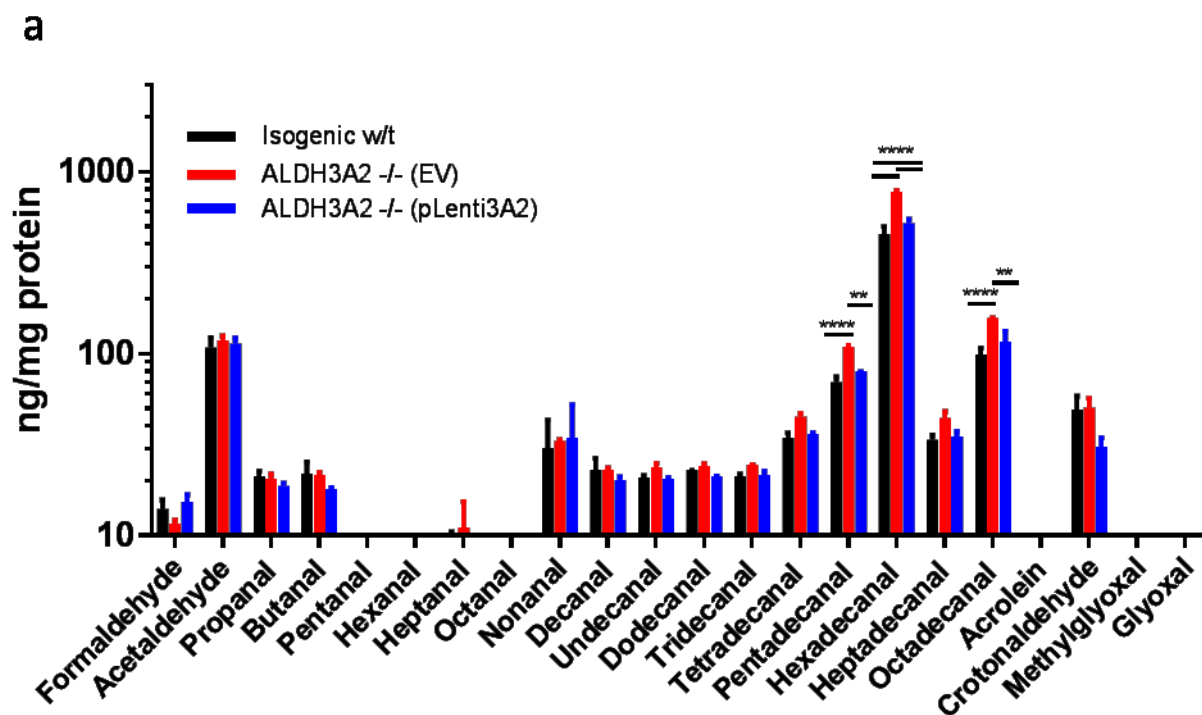


**Figure 35: Influence of *ALDH3A2* on cancer-specific phenotypes in OAC.**

Panel a, validation of *ALDH3A2*-constrained FLO1 cell models Panel b, expression of *ALDH3A1* and *3A2* at the RNA level in the same models. Panel c & d, growth and cell cycle analysis of *ALDH3A2* constrained FLO1 cells. Panel e re-analysis of IHC data for nuclear staining of *ALDH3A2*. Panel f, GSH and NADP<sup>+</sup>/NADPH in *ALDH3A2*-constrained FLO1 cells

The effects metabolic effects of controlling *ALDH3A2* expression were assessed using UPLC-MS/MS to quantify aldehydes. As seen in Figure 36a, the same alkanal species C1, 2, 9, 16, 18 featured most prominently, although there was no significant *ALDH3A2*-specific differences under standard conditions. However, when peroxidative conditions were simulated using the Fe(II)/Ascorbate system, there was a global increase in multiple aldehyde species, in particular C2-C4, C10 to C18, and dienals. In particular, there was more C15 - C18 in G3-EV cells compared to G3-R or parental cells.

Finally, the effect of *ALDH3A2* perturbation on aldehyde-nucleoside adducts was assessed using UPLC-MS/MS (see Figure 36b). For this experiment, the relatively large DNA requirement and the slow growth of the G3-3A2 cells precluded a three-way comparison. To add extra controls, the three *ADLH3A2* edit clone: D4, C7 and G3, were selected and compared to three cultures of the parental isogenic FLO1 line. Under standard culture conditions, there were measurable quantities of edA and EtdG in both parental and *ALDH3A2* null FLO1 lines. However, there was no difference in the load of adducts between these cell types.



**Figure 36: Free and bound aldehyde metabolic phenotypes in *ALDH3A2*-constrained FLO1 cells.**

Panel a, Aldehyde metabolic phenotypes quantified using UPLC-MS/MS (see Section 2) and normalised to total protein input. Panel b, aldehyde-nucleotide adducts in wildtype and *ALDH3A2* knockout FLO1 cells.

#### 4.3.6 Discussion

In this section, the effects of *ALDH3A2* on neoplastic phenotypes was assessed, to help interpret the clinical observation that low *ALDH3A2* expression is associated with disease aggression, progression and poorer survival (see Chapter 3). It was observed that *ALDH3A2*<sup>-/-</sup> FLO1 cells grow at a much faster rate than their parental or *ALDH3A2* transduced controls, possibly through cell cycle progression. Mutations in *ALDH3A2* cause Sjogren-Larsson syndrome (SLS), in which hyperkeratosis is universally severe and implies that *ALDH3A2* can non-redundantly influence cell growth and division in certain contexts (75,288,289). Given that the oesophagus is normally lined by keratinocytes, these contexts may be relevant to OAC. Reviewing the *ALDH3A2* oesophageal immunostaining revealed particularly strong expression in the basal keratinocyte layer, which comprises slow-cycling cells responsible for stratified epithelial replenishment (290). In 2005, Haug et al showed that transducing SLS keratinocytes with functional *ALDH3A2* enzyme reduced hyperkeratosis in a 3D air-medium stratified keratinocyte model (291). Kihara et al recently described phenotypes of a *Aldh3a2* knockout mouse and revealed increased skin keratinocyte growth both *in vivo* and *ex vivo* (292). In human cells, the close neighbour and paralog *ALDH3A1* is a known negative cell cycle regulator (84,122,293). This study extends these findings by reporting *ALDH3A2* impaired growth, possibly through cell cycle braking in an *in vitro* model of OAC.

The mechanism by which *ALDH3A2* exerts cell cycle control is not yet understood, although several hypotheses have been proposed. Sphingosine-1-phosphate (S1P) is a lipid signalling molecule that is potently cytoprotective and anti-apoptotic in keratinocytes(22) and in a wide range of cancers. Sphingosine kinase (SPHK1) has been shown to be widely upregulated in cancer. In gastro-oesophageal cell lines, chemotherapy resistance was strongly associated with SPHK1 expression (294), implying metabolic switch to S1P formation. *ALDH3A2* is responsible for converting the sphingosine-1-phosphate metabolites hexadecanal and 2-hexadecenal to corresponding carboxylic acids, implying that *ALDH3A2* loss of this may permit S1P accumulation. Indeed, increased S1P concentration has been noted in OSCC plasma (see Table 1)(48). Additionally, loss of *ALDH3A2* may be associated with an oxidative stress-phenotype (292) in particular through activation of nrf2 (*NFE2L2*), which stabilises in the absence of its ubiquitination partner *KEAP1*. Constitutive Nrf2 activation has been associated with proliferation in many cancers(138). In the Wang et al legacy dataset(225), *KEAP1* and *NFE2L2* expression was low in Barretts and

OAC compared to keratinocytes, indicating that Nrf2 protein is able to stabilise in the cytoplasm and is probably enriched (see Appendix 5). Lastly, *ALDH3A2* directly metabolises phytal to phytanic acid, a natural ligand for peroxisome-proliferator activated receptor alpha (PPAR-alpha) a potent regulator of lipid metabolism which has been shown to affect tumorigenesis and tumour progression in humans and animal models(74,295). These mechanisms require urgent molecular dissection in the context of OAC as they are all druggable with available specific inhibitors.

Comparing aldehyde metabolism between the *ALDH3A2*-constrained FLO1 cells revealed that differences in fatty aldehyde metabolism only emerged after simulated peroxidation. Fatty aldehyde formation is poorly understood, but thought to involve the action of reactive oxygen species and enzymatic cleavage of plasmalogens and polyunsaturated fatty acids, as well as specific pathways such as sphingosine metabolism (74,75). Recent mass spectrometry imaging data suggest that lipidome constituents are profoundly different in OAC (146,147), and an important extension of this work should test the contribution of these phenotypes to altered aldehyde metabolism. The influence of *ALDH3A2* on nucleoside adduct formation was also piloted and there were no effects on etheno- adducts. This suggests that the non-redundant properties of *ALDH3A2* loss may be limited to fatty aldehydes in this context, but it is the first direct evidence that specific aldehyde phenotypes can be traced to specific genetic lesions.

It was also noted that loss of *Aldh3a2* stimulated oxidative stress-response genes in knockout mouse model keratinocytes (292). In this work, loss of *ALDH3A2* was significantly associated with a diminished anti-oxidant signature under normal conditions. Taking together, this suggests deployment of alternative pathways (e.g. glutathione conjugation, or reduction to alcohol/alkane (273)) in response to the accumulation of *ALDH3A2* substrates. The Kihara group also reported parallel up-regulation of other potentially collateral *Aldh3* genes in a tissue specific manner(233,292). There was little evidence of this in human OAC in the genes tested in this study (see Chapter 3), or corresponding microarray expression data for untested *ALDH* genes (see Appendix 3), suggesting that *ALDH3A2* loss may be particularly influential in this context. Finally, Kihara et al reported that *Aldh3a2*<sup>-/-</sup> keratinocytes are less adept at metabolising fatty aldehydes in conditioned media experiments, and a limitation of the present work is that similar functional assays have not yet been undertaken with the available models.

Sjogren-Larsson syndrome is an exceptionally rare disease, with the biggest case series being <20 (75,296). OAC-associated mutations in *ALDH3A2* are also very rare (see Chapter 3), and together this suggests complete loss of *ALDH3A2* is only life-compatible with the availability of specific collateral metabolism. Indeed, multiple attempts in this project to establish *ALDH3A2*<sup>-/-</sup> FLO1 xenografts in either nude or SCID mice were not successful (data not shown). However, reduced expression of *ALDH3A2* was clearly demonstrated in numerous datasets and models, and thus suggests that a “poised state” of hypo-functionality exists in OAC.

There is no perfect model of OAC and as such, these data are subject to limitations. CRISPR-Cas9 editing involves nuclease treatment of genomic DNA, and thus risks introducing unintended (“off-target”) permanent edits. This project addressed this, using (i) highly selective guides (ii) transient Cas9 expression (iii) multiple edit clones, although it would be helpful to verify selectivity by either whole genome sequencing, or targeted amplicon sequencing of predicted potential off-target edit sites. New ultra-selective CRISPR tools have since emerged, including a dual-guide, single-strand approach, and lower affinity Cas9 enzymes which only bind to perfectly matching target sequences. A future research priority will be to deploy these and alternative approaches (e.g. small hairpin (sh) RNA) to verify the observed *ALDH3A2*-dependent OAC-specific phenotypes. Another limitation is that transduction of the *ALDH3A2* ORF only partially restored redox effects despite a lot more protein, however others have found the same patterns with exogenous expression(291).

It will also be of great interest to determine the role of *ALDH3A2* in oesophageal biology in complex models, such as the *Aldh3a2* conditional knockout mouse, or even SLS patients. SLS descriptive studies tend to focus on the neuro-cutaneous manifestations which define the disease and none have specifically assessed gastrointestinal complications (297); this may well be the perfect system to study the specific effects of *ALDH3A2* hypoactivity in human oesophageal oncobiology.



## **CHAPTER 5 – THESIS CONCLUSIONS & FUTURE WORK**

## 5.1 Conceptual gains from this thesis

The specific research gains can be subdivided according to the strategic rationale for the project, set out in Chapter 1.1.1.

### 5.1.1 New paradigms in OAC cancer biology, and new therapeutic opportunities

*Aldehyde metabolism is deregulated in oesophageal adenocarcinoma* Previously, free aldehydes were detected in OAC patients' breath (15,16). In Chapter 1, it was argued that aldehyde metabolic profiles should be enriched in OAC tissues as aldehyde sources are widely potentiated, and detoxification systems are impaired. In Chapter 2, the first evidence was enriched aldehyde in the OAC tissue was presented, taking advantage of a bespoke quantitative method utilising unambiguous analytics and validated internal control procedures. Among many others, C2 was over 10 times more concentrated in samples from the malignant oesophagus compared to samples from the endoscopically normal oesophagus. In Chapter 3, an unbiased systematic review of all archived transcriptomic analyses in Barrett's and OAC confirmed that loss of *ALDH* expression is among the most defining features of squamous mucosa replacement, and this was extensively validated using clinical phenotyping experiments. In Chapter 4, the interaction between DNA and the candidate mutagens C2 and HNE was tested using a second bespoke UPLC-MS/MS method. Again there was clear enrichment of HNE- and acetaldehyde- based DNA adducts, both verifying the findings of Chapter 2 and extending these by providing a mutagenic mechanism for OAC simply based on the geometry of the quantified targets. Moreover, global inhibition of ALDHs was sufficient to enrich aldehydes, implying that cellular aldehyde governance relies on competent *ALDH* defences rather than source control.

This evidence is consistent with the leading project hypothesis that aldehyde detoxification is sufficiently impaired to enrich mutagenic aldehydes in OAC tissues, which then participate in carcinogenic biochemistry. Additionally, the finding of OAC expression loss of *ALDH3A2*, a non-redundant metabolic gene, begins to suggest that non-invasive molecular traits can be traced to specific genetic lesions, which may have profound implications for stratified medicine and accurate therapeutic

monitoring (i.e. providing the right treatment to the right patient, at the right time, rather than a probability of success).

These preliminary data require validation to test the stability of these phenotypes across populations. Metabolomics studies can be confounded by diet and lifestyle factors (298), and an important next step will test aldehyde phenotypes in international OAC cohorts where demographics and co-morbidities may be different. It will also be interesting to expand these analyses to other disease-states, including OSCC and reflux oesophagitis. Lastly this work has provided only the first glimpses of the relationship between *ALDH* genes and metabolic and non-metabolic phenotypes in OAC, and much more work is needed in representative models.

*Aldehyde-directed therapy* Expression profiling suggested that attenuated aldehyde detoxification occurs relatively early in the metaplasia-neoplasia continuum, and an early research priority will test the timing of aldehyde metabolic reprogramming through this sequence. If aldehyde enrichment specifically contributes to disease progression, this may offer a new therapeutic paradigm in 'dysplasia prophylaxis' – i.e. measures to prevent genotoxicity in cases with metaplasia. Rationally designed small molecule activators of ALDHs already exist because of the *ALDH2\*2* alcohol flushing syndrome market(275), in particular Alda-89, a specific activator of *ALDH3A1* and *-3A2*(299). In addition, the luminal and pre-gastric nature of Barretts epithelium potentially renders it amenable to topical treatment by swallowed transactivating agents (e.g. small activating RNA)(300), or by available aldehyde scavenging agents such as stearylamine or NS2 (2-[3-amino-6-chloro-quinolin-2-yl]-propan-2-ol)(74).

*ALDH3A2 as a candidate tumour suppressor gene which may be linked to 17p deletion* The candidate-based approach to discover and validate clinically relevant genetic mediators of aldehyde metabolism identified low expression of *ALDH3A2* to be significantly associated with aggressive and progressive disease, and poorer overall survival. This lies in the 17p11.2 region which frequently undergoes loss-of-heterozygosity events and has been a known poor prognostic marker for two decades (207,208). Previous work has linked this to *TP53* deletion (208), although emerging data suggests synchronous deletion of 17p neighbour genes can offer synergistic pro-cancer phenotypes (241). The results of section 4.3 demonstrated that a permanent deletion of *ALDH3A2* enhanced growth characteristics in an OAC cell model seemingly through cell cycle shunting, and will be of interest to validate

this in other models and with other perturbation techniques. How *ALDH3A2* may influence cell cycling is not established, although the leading hypotheses set out in Chapter 4.3.3 indicate readily druggable candidate mechanisms.

### **5.1.2 Demonstrate biologically distinct, clinically relevant subgroups**

Low expression of *ALDH3A2* was significantly associated with poor prognostic features and worse outcomes. Efforts are underway to validate this in a second international cohort, and to test whether these effects are independent of 17p LOH and/or *TP53* status. Microarray-based markers (301) and other molecular-based prognostic models (239,302) have also shown promise for precision decision-making and are undergoing validation, and it will be important to test whether *ALDH3A2* further complements these panels. However, on its own, *ALDH3A2* has two characteristics that set it apart (i) the functional data in Chapter 4.3 suggest there is an inherent mechanistic basis for this marker's clinical relevance (ii) through its non-redundant metabolic properties, changes in *ALDH3A2* expression may prove to be non-invasively detectable.

Apart from two subgroup analysis designed to check against potential confounders, patient stratification by metabolic phenotypes to clinical metadata was not undertaken in this series as the numbers of cases were felt to be too low to provide meaningful analysis. For example, the majority of patients have neo-adjuvant treatment, are T3, have local nodal involvement (N1), and are relatively old, and hence other classes would be weakly populated in associated studies and thus prone to bias. Clearly it is a priority to test aldehyde profiles against clinical indices, in particular for hypothesised pathophysiological mechanisms – e.g. the presence of reflux, history of medical acid suppression, tobacco smoking, ethanol ingestion, evidence of inflammation on histology etc. The aforementioned Barretts aldehyde phenotyping with strict histological supervision may also indicate risk populations, supporting personalised prophylactic measures.

### 5.1.3 Breath test refinement

A strength of the aldehyde UPLC-MS/MS method was that it measured a large set of aldehydes with robust isomeric differentiation. The original OAC breath studies (15) utilised SIFT-MS, which is a form of ambient mass spectrometry that identifies target compounds on the basis of characteristic product ions formed in a drift tube and quantified in a single detector quadrupole resolving integer  $m/z$  values. Thus, the responses of alkanal, ketone and dialdehyde isomers/isobars – e.g. propanal/acetone/glyoxal (MW 58) or butanal/butanone/methylglyoxal (MW 72) will not be discriminated. Others will compound complexity. These compounds represent different biology (see Chapter 1), and were all abundant in tissue and urine. Unambiguous compound identification is therefore essential, and future OAC breath analytics must provide appropriate selectivity (e.g. using accurate mass technology such as time-of-flight mass spectrometry).

The reduced expression of several ALDH isoenzymes suggests that Barretts and OAC cells may rely on alternate metabolic pathways to detoxify aldehydes, and an important study from the mechanistic and diagnostic perspectives will measure aldehyde concentrations in the context of these metabolic partners at a tissue level (e.g. alcohol, alkanes, fatty acids and aldehyde-glutathione conjugates). Alkanes in particular are extremely volatile and highly suited to breath analysis, and breath method development should consider including these metabolic relatives.

An ultimate goal of this work was to provide a mechanistic basis for the appearance of trace volatile aldehydes in OAC patients' breath and urine. A recent review of haematological metabolomic biomarkers in gastrointestinal cancer (41) revealed that systemic metabolic traits of cancer are detectable even at the earliest stages of transformation. The obvious explanation for breath and urine biomarkers is that tumour metabolites enter the blood and are expelled in the breath and urine. There are several issues with this however: (i) the liver is an efficient detoxification organ and aldehydes will pass through it from the oesophagus to the lungs (ii) aldehydes are unstable in plasma (see Chapter 2) (iii) effects of other systemic influences such as the colonic microbiome, nutrition, etc, are not understood. The biofluid experiments in this work did not reveal obvious discriminating features although were likely subject to recovery limitations (as explained in Chapter 2.4). If these technical aspects can be accounted for in revised analytics, and no haematological link is found, alternative solutions must be sought.

One alternative explanation is that inherited variants (e.g. in *ALDH1A2* or *ALDH2* (103,105)) are actually responsible for VOC phenotypes, and there has yet to be a VOC analysis stratified by genotype. Notably, *ALDH1A2* was poorly expressed in any oesophageal tissue, including healthy controls. A more likely explanation is that the candidate volatile aldehydes are produced in or diffuse into the pre-epithelial mucous layer, which is continuous with the respiratory and oral mucous membranes. Several lines of evidence support this: (i) Compared to keratinocytes, Barretts and OAC cells display very different lipid profiles featuring more peroxidation-prone desaturations and differential chain lengths which could support specific carbonyl profiles (146,147); (ii) carbonyl stress may be strongest in the most luminal cells (iii) the local microbiome will be most influential in the most luminal cells (iv) this direct route bypasses blood and the liver. This hypothesis can be tested by further aldehyde profiling studies in using pre-epithelial mucous in the mouth, pharynx, and along the oesophagus. Nonetheless, more work is required to establish the relative contribution of body-wide VOC sources to breath biomarkers, with attention to liver pathology, diet/nutrition, and the distal gut microbiome.

## 5.2 Technical gains from this thesis

Two UPLC-MS/MS methods covering a total of 43 free carbonyls and 11 normal and modified nucleosides were developed and validated for accurate quantitation. These methods were applied to multiple sample types, looking at this biology in different and complementary ways. However, a key strength of this work is the parallel investigation of the genetic basis for this metabolic reprogramming. This strategy helped shape method development by focussing on the critical emerging metabolite-groups (e.g. fatty aldehydes for *ALDH3A2*), emphasising the requirement of tackling biochemistry problems by equally investing in biology and chemistry simultaneously. A number of cell line-VOC efforts by analysts have found conflicting results (e.g. 247,248). By incorporating cell biology good practice and manipulating culture conditions, expected phenotypes can emerge, and there is much scope to expand these efforts for the study of aldehyde metabolism.

The developed UPLC-MS/MS methods provided confidence regarding compound identification and target concentration, although it is inherently blinkered to non-target compounds which may be of clinical or biological relevance. Thus an extension to this work may be parallel metabolomic studies as follows:

- More compounds to the targeted method: e.g. branched chain, alpha-keto-alkanals, phytal, retinal, betaine aldehyde, succinaldehyde etc
- Parallel targeted method for biochemical partners: lipids, fatty acids, alkanes, amino acids, glutathione conjugates, aldehyde-lipid conjugates
- Untargeted nucleoside-adduct analysis, utilising 116 neutral loss and accurate mass analysis, followed by compound elucidation with NMR or MS/MS

An additional strength of this work was to marry these methods with models of oesophageal cancer. Lastly, this was the first experiment to test aldehydes in cell lines perturbed with CRISPR-Cas9 editing, although more are likely to come.

## 5.3 Future research

### 5.3.1 Short term goals

The immediate succession goals have already been discussed at the end of each experimental Chapter. The purpose of these experiments is to produce a complete description of OAC aldehyde metabolism from source to fate, beyond the remit of the present project. In summary, these include:

#### ***Free aldehydes in oesophageal tissue:***

- Validation of tissue phenotyping in separate cohorts, with larger sample sizes to allow subgroup analyses by clinical factors;
- Linked comparative studies with related metabolites (e.g. glutathione-aldehyde conjugates; protein carbonyls; alkanes/alcohols);
- Expanded applications of the UPLC-MS/MS method, including measurement of aldehydes from lipid tissue extracts and oesophageal mucous samples;
- Revised biofluid methodology for urine and blood looking at temporising measures to improve analyte stability;
- Correlative studies with breath.

#### ***Genetic framework for aldehyde metabolism:***

- Further validation of *ALDH* expression patterns and survival to explore prognostic utility, especially in the context of chemotherapy;
- Correlation studies to establish the role of 17p deletion on *ALDH3A* expression using fluorescence in situ hybridisation and IHC;
- Upstream analyses to identify *ALDH* transregulators using transcriptomics, especially in the context of genetic determinants of the keratinocyte lineage.

#### ***Effects of aldehyde deregulation:***

- Functional experiments to assess whether oesophageal keratinocyte cultures display different aldehyde detoxification to OAC cultures;
- Verification of the *ALDH3A2* functional effects with alternative silencing methods and/or alternative cell models
- Transcription profiling to determine how *ALDH3A2* affects cell cycling;



- *In vitro* isotope tracing studies to assess where aldehydes are chiefly generated, and what is their predominant fate.

### 5.3.2 Longer term projects

Wider projects arising from these results include:

***Aldehyde metabolism in other OG cancers*** Tissue-aldehyde measurements were a success in this project, although there are very few prior reports of tissue aldehyde profiling in literature, and only one using an MS methodology (162). Thus there is huge scope to apply this to other diseases, and OSCC will be an early goal. This is because (i) the shared genetic and anatomical risks will make aldehyde deregulation likely (ii) there is a clearer progression for the index normal tissue, facilitating the supervision of phenotyping and in particular model development (iii) it is fundamentally important from the clinical perspective to know if this cancer is detectable by the same molecular means, and discernable from OAC.

Assessing this metabolism in gastric cancer is also an urgent research priority; the clinical need is similarly pressing, the risk factors for aldehyde deregulation are similarly present, and there is already empirical metabolic evidence supporting the hypothesis that aldehydes are deregulated in this disease (15,112,259).

***Aldehyde metabolism in Barrett's transformation*** Understanding aldehyde metabolic reprogramming through Barrett's transformation will also be of interest, not least to understand if there is opportunity for carcinogenic intervention at a pre-malignant stage, and whether this may be non-invasively detectable. There is an urgent unmet need for a cheap, safe and acceptable screening tool for OAC, either for patients with known Barrett's, on patients with non-specific upper GI symptoms (303). Currently only one promising non-invasive test is being assessed in non-case-controlled prospective multi-centre studies. This involves regurgitating a small sponge and may be effective from a clinical perspective, although patient acceptability is not clear (304). At any rate, unpicking aldehyde metabolism with similar methodologies will require robust histology to account for dysplastic influences, and how to marry this with aldehydes inherent instability will require development. Additionally, there is evidence of volatile aldehyde enrichment in a

panel of other malignancies; and the present research model is readily transferrable to address these other clinical questions.

***Refined non-invasive diagnostics*** As discussed, a research priority will be to revisit OAC breath studies to more specifically assess aldehydes, ketones and related redox partners, with technology that is sufficiently selective to permit unambiguous compound identification (i.e. accurate mass); these innovations should greatly improve diagnostic performance. With this enhanced biomarker research will come several related projects in validation, prognostication, treatment response and disease surveillance. It will also be of interest to explore other related modalities, such as saliva.

***Therapeutic manipulation of aldehyde metabolism*** If aldehyde deregulation is verified as a carcinogenic mechanism in Barrett's metaplasia, there are a number of options for manipulating aldehyde metabolism, including through the use of aldehyde scavenger agents, ALDH-recruiting drugs, and even gene therapies, and these may offer a new concept in medical Barrett's prophylaxis.

## Bibliography

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015 Mar;65(2):87–108.
2. Enzinger PC, Mayer RJ. Esophageal cancer. *N Engl J Med*. 2003;349(23):2241–52.
3. Arnold M, Soerjomataram I, Ferlay J, Forman D. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut*. 2015 Mar;64(3):381–7.
4. Bosetti C, Levi F, Ferlay J, Garavello W, Lucchini F, Bertuccio P, et al. Trends in oesophageal cancer incidence and mortality in Europe. *Int J Cancer*. 2008;122(5):1118–29.
5. Chadwick G, Varaganam M, Brand C, Cromwell D, Maynard N, Riley S CT. National Oesophago-Gastric Cancer Audit 2016.
6. van Hagen P, Hulshof MCCM, van Lanschot JJB, Steyerberg EW, van Berge Henegouwen MI, Wijnhoven BPL, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med*. 2012 May 31;366(22):2074–84.
7. Allum WH, Stenning SP, Bancewicz J, Clark PI, Langley RE. Long-term results of a randomized trial of surgery with or without preoperative chemotherapy in esophageal cancer. *J Clin Oncol*. 2009 Oct 20;27(30):5062–7.
8. Niamh Keegan, Fionnuala Keane, Sinead Cuffe, Moya Cunningham, Narayanasamy Ravi, Geraldine Lee, M. John Kennedy, William Grogan, Brian O'Neill, Derek G. Power, Gregory Leonard, Maccon M. Keane, Cormac Small, Imelda Parker, Brian Moulton, Seamus O'Reilly, IGG. ICORG 10-14: Neo-AEGIS: A randomized clinical trial of neoadjuvant and adjuvant chemotherapy (modified MAGIC regimen) versus neoadjuvant chemoradiation (CROSS protocol) in adenocarcinoma of the esophagus and esophagogastric junction. *J Clin Oncol*. 2014;32(5):TPS4145.
9. Allum WH, Blazeby JM, Griffin SM, Cunningham D, Jankowski JA, Wong R. Guidelines for the management of oesophageal and gastric cancer. *Gut*. 2011;60(11):1449–72.
10. Earlam R, Cunha-Melo JR. Oesophageal squamous cell carcinoma: I. A critical review of surgery. *Br J Surg*. 1980;67(6):381–90.
11. Jamieson GG, Mathew G, Ludemann R, Wayman J, Myers JC, Devitt PG.

- Postoperative mortality following oesophagectomy and problems in reporting its rate. *Br J Surg*. 2004;91(8):943–7.
12. Geographic Patterns of Cancer Survival in England: Patients followed up to 2011. Office for National Statistics. 2011.
  13. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. 2011;61(2):69–90.
  14. Vakil N, Moayyedi P, Fennerty MB, Talley NJ. Limited value of alarm features in the diagnosis of upper gastrointestinal malignancy: systematic review and meta-analysis. *Gastroenterology*. 2006;131(2):360–90.
  15. Kumar S, Huang J, Abbassi-Ghadi N, Mackenzie HA, Veselkov KA, Hoare JM, et al. Mass Spectrometric Analysis of Exhaled Breath for the Identification of Volatile Organic Compound Biomarkers in Esophageal and Gastric Adenocarcinoma. *Ann Surg*. 2015;262(6):981–90.
  16. Kumar S, Huang J, Abbassi-Ghadi N, Patrik S, Smith D, Hanna GB, et al. Selected ion flow tube mass spectrometry analysis of exhaled breath for volatile organic compound profiling of esophago-gastric cancer. *Anal Chem*. 2013;85(12):6121–8.
  17. Kumar S, Huang J, Cushnir JR, Spanel P, Smith D, Hanna GB, et al. Selected ion flow tube-MS analysis of headspace vapor from gastric content for the diagnosis of gastro-esophageal cancer. *Anal Chem*. 2012;84(21):9550–7.
  18. Huang J, Kumar S, Abbassi-Ghadi N, Spanel P, Smith D, Hanna GB, et al. Selected ion flow tube mass spectrometry analysis of volatile metabolites in urine headspace for the profiling of gastro-esophageal cancer. *Anal Chem*. 2013;85(6):3409–16.
  19. Peng DF, Razvi M, Chen H, Washington K, Roessner A, Schneider-Stock R, et al. DNA hypermethylation regulates the expression of members of the Mu-class glutathione S-transferases and glutathione peroxidases in Barrett's adenocarcinoma. *Gut*. 2009;58(1):5–15.
  20. Peters CJ, Rees JR, Hardwick RH, Hardwick JS, Vowler SL, Ong CA, et al. A 4-gene signature predicts survival of patients with resected adenocarcinoma of the esophagus, junction, and gastric cardia. *Gastroenterology*. 2010;139(6):1995–2004 e15.
  21. Orlando RC. Esophageal mucosal defense mechanisms. *GI Motil online, Publ online* 16 May 2006; | doi101038/gimo15. 2006;
  22. Manggau M, Kim D-S, Ruwisch L, Vogler R, Schäfer-Korting M, Kleuser B, et al. 1 $\alpha$ ,25-Dihydroxyvitamin D<sub>3</sub> Protects Human Keratinocytes from Apoptosis by the Formation of Sphingosine-1-Phosphate. *J Invest Dermatol*. 2001

- Nov;117(5):1241–9.
23. Amanuma Y, Ohashi S, Itatani Y, Tsurumaki M, Matsuda S, Kikuchi O, et al. Protective role of ALDH2 against acetaldehyde-derived DNA damage in oesophageal squamous epithelium. *Sci Rep.* 2015/09/17. 2015;5(August):14142.
  24. Kanaly RA, Matsui S, Hanaoka T, Matsuda T. Application of the adductome approach to assess intertissue DNA damage variations in human lung and esophagus. *Mutat Res.* 2007;625(1–2):83–93.
  25. Lawrence MS, Stojanov P, Polak P, Kryukov G V., Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013 Jul 11;499(7457):214–8.
  26. Lagergren J, Bergstrom R, Lindgren A, Nyren O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med.* 1999;340(11):825–31.
  27. Lagergren J. Controversies surrounding body mass, reflux, and risk of oesophageal adenocarcinoma. *Lancet Oncol.* 2006;7(4):347–9.
  28. Nilsson M, Johnsen R, Ye W, Hveem K, Lagergren J. Obesity and estrogen as risk factors for gastroesophageal reflux symptoms. *JAMA.* 2003;290(1):66–72.
  29. Leedham SJ, Preston SL, McDonald SAC, Elia G, Bhandari P, Poller D, et al. Individual crypt genetic heterogeneity and the origin of metaplastic glandular epithelium in human Barrett's oesophagus. *Gut.* 2008;57(8):1041–8.
  30. Wang X, Ouyang H, Yamamoto Y, Kumar PA, Wei TS, Dagher R, et al. Residual embryonic cells as precursors of a Barrett's-like metaplasia. *Cell.* 2011;145(7):1023–35.
  31. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet.* 2013 May;45(5):478–86.
  32. Ross-Innes CS, Becq J, Warren A, Cheetham RK, Northen H, O'Donovan M, et al. Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat Genet.* 2015;47(9):1038–46.
  33. Weaver JM, Ross-Innes CS, Shannon N, Lynch AG, Forshew T, Barbera M, et al. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat Genet.* 2014 Aug;46(8):837–43.
  34. Secrier M, Li X, de Silva N, Eldridge MD, Contino G, Bornschein J, et al. Mutational signatures in esophageal adenocarcinoma define etiologically

- distinct subgroups with therapeutic relevance. *Nat Genet.* 2016 Sep 5;
35. Bhat S, Coleman HG, Yousef F, Johnston BT, McManus DT, Gavin AT, et al. Risk of malignant progression in Barrett's Esophagus patients: Results from a large population-based study. *J Natl Cancer Inst.* 2011;103(13):1049–57.
  36. Hvid-Jensen F, Pedersen L, Drewes AM, Sorensen HT, Funch-Jensen P. Incidence of adenocarcinoma among patients with Barrett's esophagus. *N Engl J Med.* 2011;365(15):1375–83.
  37. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio S a JR, Behjati S, Biankin A V, et al. Signatures of mutational processes in human cancer. *Nature.* 2013 Aug 22;500(7463):415–21.
  38. Nones K, Waddell N, Wayte N, Patch A-M, Bailey P, Newell F, et al. Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nat Commun.* 2014;5:5224.
  39. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529–36.
  40. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Br Med J.* 2003;326(7379):41–4.
  41. Antonowicz S, Kumar S, Wiggins T, Markar SR, Hanna GB. Diagnostic Metabolomic Blood Tests for Endoluminal Gastrointestinal Cancer-A Systematic Review and Assessment of Quality. *Cancer Epidemiol Biomarkers Prev.* 2016;25(1):6–15.
  42. Djukovic D, Baniyadi HR, Kc R, Hammoud Z, Raftery D. Targeted serum metabolite profiling of nucleosides in esophageal adenocarcinoma. *Rapid Commun Mass Spectrom.* 2010;24(20):3057–62.
  43. Zhang J, Liu L, Wei S, Nagana Gowda GA, Hammoud Z, Kesler KA, et al. Metabolomics study of esophageal adenocarcinoma. *J Thorac Cardiovasc Surg.* 2011;141(2):469–75, 475-4.
  44. Zhang J, Bowers J, Liu L, Wei S, Gowda GAN, Hammoud Z, et al. Esophageal cancer metabolite biomarkers detected by LC-MS and NMR methods. *PLoS One.* 2012 Jan;7(1):e30181.
  45. Ikeda A, Nishiumi S, Shinohara M, Yoshie T, Hatano N, Okuno T, et al. Serum metabolomics as a novel diagnostic approach for gastrointestinal cancer. *Biomed Chromatogr.* 2012;26(5):548–58.
  46. Zhang X, Xu L, Shen J, Cao B, Cheng T, Zhao T, et al. Metabolic signatures of esophageal cancer: NMR-based metabolomics and UHPLC-based focused

- metabolomics of blood serum. *Biochim Biophys Acta*. 2013;1832(8):1207–16.
47. Sanchez-Espiridion B, Liang D, Ajani JA, Liang S, Ye Y, Hildebrandt MAT, et al. Identification of Serum Markers of Esophageal Adenocarcinoma by Global and Targeted Metabolic Profiling. *Clin Gastroenterol Hepatol*. 2015;13(10):1730–1737.e9.
  48. Liu R, Peng Y, Li X, Wang Y, Pan E, Guo W, et al. Identification of Plasma Metabolomic Profiling for Diagnosis of Esophageal Squamous-Cell Carcinoma Using an UPLC/TOF/MS Platform. *Int J Mol Sci*. 2013;14(5):8899–911.
  49. Xu J, Chen Y, Zhang R, Song Y, Cao J, Bi N, et al. Global and targeted metabolomics of esophageal squamous cell carcinoma discovers potential diagnostic and therapeutic biomarkers. *Mol Cell Proteomics*. 2013;12(5):1306–18.
  50. Jin H, Qiao F, Chen L, Lu C, Xu L, Gao X. Serum Metabolomic Signatures of Lymph Node Metastasis of Esophageal Squamous Cell Carcinoma. *J Proteome Res*. 2014;13(9):4091–103.
  51. Ma H, Hasim A, Mamtimin B, Kong B, Zhang H-PP, Sheyhidin I. Plasma free amino acid profiling of esophageal cancer using high-performance liquid chromatography spectroscopy. *World J Gastroenterol*. 8653;26(14):8653–9.
  52. Mir SA, Rajagopalan P, Jain AP, Khan AA, Datta KK, Mohan S V, et al. LC–MS-based serum metabolomic analysis reveals dysregulation of phosphatidylcholines in esophageal squamous cell carcinoma ☆. *J Proteomics*. 2015;127:96–102.
  53. Davis VW, Schiller DE, Eurich D, Sawyer MB. Urinary metabolomic signature of esophageal cancer and Barrett's esophagus. *World J Surg Oncol*. 2012;10:271.
  54. Xu J, Chen Y, Zhang R, He J, Song Y, Wang J, et al. Global metabolomics reveals potential urinary biomarkers of esophageal squamous cell carcinoma for diagnosis and staging. *Sci Rep*. 2016 Oct 11;6:35010.
  55. Wu H, Xue R, Lu C, Deng C, Liu T, Zeng H, et al. Metabolomic study for diagnostic model of oesophageal cancer using gas chromatography/mass spectrometry. *J Chromatogr B Anal Technol Biomed Life Sci*. 2009;877(27):3111–7.
  56. Yakoub D, Keun HC, Goldin R, Hanna GB. Metabolic profiling detects field effects in nondysplastic tissue from esophageal cancer patients. *Cancer Res*. 2010 Nov 15;70(22):9129–36.
  57. Wang LS, Young M, Kuo CT, Arnold M, Martin E, Sardo C, et al. Metabolomic profiling reveals a protective modulation on fatty acid metabolism in colorectal

- cancer patients following consumption of freeze-dried black raspberries. *Cancer Res.* 2013;Conference:104th Annual Meeting of the American Association f.
58. Eckl PM, Ortner A, Esterbauer H. Genotoxic properties of 4-hydroxyalkenals and analogous aldehydes. *Mutat Res.* 1993;290(2):183–92.
  59. Esterbauer H, Eckl P, Ortner A. Possible mutagens derived from lipids and lipid precursors. *Mutat Res.* 1990;238(3):223–33.
  60. Esterbauer H. Estimation of peroxidative damage. A critical review. *Pathol Biol.* 1996;44(1):25–8.
  61. Frankel EN. *Lipid oxidation.* Oily Press; 2005. 470 p.
  62. Fenton HJ. Oxidation of tartaric acid in the presence of iron. *J Chem Soc.* 1894;65:899–910.
  63. Glindemann D, Dietrich A, Staerk H-J, Kusch P. The Two Odors of Iron when Touched or Pickled: (Skin) Carbonyl Compounds and Organophosphines. *Angew Chemie Int Ed.* 2006 Oct 27;45(42):7006–9.
  64. Nair U, Bartsch H, Nair J. Lipid peroxidation-induced DNA damage in cancer-prone inflammatory diseases: A review of published adduct types and levels in humans. *Free Radic Biol Med.* 2007;43(8):1109–20.
  65. O'Brien PJ, Siraki AG, Shangari N. Aldehyde Sources, Metabolism, Molecular Toxicity Mechanisms, and Possible Effects on Human Health. *Crit Rev Toxicol.* 2005 Jan 10;35(7):609–62.
  66. Grosjean E, Grosjean D, Fraser MP, Cass GR. Air Quality Model Evaluation Data for Organics. 2. C1–C14 Carbonyls in Los Angeles Air. 1996;
  67. Fujioka K, Shibamoto T. Determination of toxic carbonyl compounds in cigarette smoke. *Environ Toxicol.* 2006 Feb;21(1):47–54.
  68. Fahlbusch K-G, Hammerschmidt F-J, Panten J, Pickenhagen W, Schatkowski D, Bauer K, et al. *Flavors and Fragrances.* In: Ullmann's Encyclopedia of Industrial Chemistry. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2003.
  69. Homann N, Tillonen J, Salaspuro M. Microbially produced acetaldehyde from ethanol may increase the risk of colon cancer via folate deficiency. *Int J Cancer.* 2000;86(2):169–73.
  70. Homann N, Tillonen J, Meurman JH, Rintamaki H, Lindqvist C, Rautio M, et al. Increased salivary acetaldehyde levels in heavy drinkers and smokers: a microbiological approach to oral cavity cancer. *Carcinogenesis.* 2000;21(4):663–8.
  71. Homann N, Karkkainen P, Koivisto T, Nosova T, Jokelainen K, Salaspuro M.



- Effects of acetaldehyde on cell regeneration and differentiation of the upper gastrointestinal tract mucosa. *J Natl Cancer Inst.* 1997;89(22):1692–7.
72. Thornalley PJ. Pharmacology of methylglyoxal: formation, modification of proteins and nucleic acids, and enzymatic detoxification--a role in pathogenesis and antiproliferative chemotherapy. *Gen Pharmacol.* 1996 Jun;27(4):565–73.
  73. Olsen R, Molander P, Øvrebø S, Ellingsen DG, Thorud S, Thomassen Y, et al. Reaction of Glyoxal with 2'-Deoxyguanosine, Thymidine, and Calf Thymus DNA: Identification of DNA Adducts. 2005;(November 2015):730–9.
  74. Rizzo WB. Genetics and prospective therapeutic targets for Sjögren-Larsson Syndrome. *Expert Opin orphan drugs.* 2016 Apr;4(4):395–406.
  75. Rizzo WB. Fatty aldehyde and fatty alcohol metabolism: review and importance for epidermal structure and function. *Biochim Biophys Acta.* 2014 Mar;1841(3):377–89.
  76. Ichihashi K, Osawa T, Toyokuni S, Uchida K. Endogenous formation of protein adducts with carcinogenic aldehydes: implications for oxidative stress. *J Biol Chem.* 2001;276(26):23903–13.
  77. Secretan B, Straif K, Baan R, Grosse Y, El Ghissassi F, Bouvard V, et al. A review of human carcinogens--Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. *Lancet Oncol.* 2009;10(11):1033–4.
  78. Sophos NA, Vasiliou V. Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chem Biol Interact.* 2003/02/27. 2003;143–144:5–22.
  79. Marchitti SA, Brocker C, Stagos D, Vasiliou V. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin Drug Metab Toxicol.* 2008;4(6):697–720.
  80. Vasiliou V. Aldehyde dehydrogenase genes. *Adv Exp Med Biol.* 1997;414:595–600.
  81. Consortium GTe. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (80- ).* 2015;348(6235):648–60.
  82. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. *Science (80- ).* 2015 Jan 23;347(6220):1260419–1260419.
  83. Sjögren T, Larsson T. A neuro-cutaneous syndrome in Swedish families. *Acta Psychiatr Scand.* 1957 Sep;32(S113):3–3.
  84. Pappa A, Brown D, Koutalos Y, DeGregori J, White C, Vasiliou V. Human aldehyde dehydrogenase 3A1 inhibits proliferation and promotes survival of

- human corneal epithelial cells. *J Biol Chem*. 2005;280(30):27998–8006.
85. Lassen N, Pappa A, Black WJ, Jester J V, Day BJ, Min E, et al. Antioxidant function of corneal ALDH3A1 in cultured stromal fibroblasts. *Free Radic Biol Med*. 2006;41(9):1459–69.
  86. Pappa A, Estey T, Manzer R, Brown D, Vasiliou V. Human aldehyde dehydrogenase 3A1 (ALDH3A1): biochemical characterization and immunohistochemical localization in the cornea. *Biochem J*. 2003;376(Pt 3):615–23.
  87. Jez JM, Penning TM. The aldo-keto reductase (AKR) superfamily: an update. *Chem Biol Interact*. 2001 Jan 30;130–132(1–3):499–525.
  88. Baradat M, Jouanin I, Dalleau S, Taché S, Gieules M, Debrauwer L, et al. 4-hydroxy-2(E)-nonenal metabolism differs in *Apc* +/+ Cells and in *Apc* Min/+ cells: It may explain colon cancer promotion by heme iron. *Chem Res Toxicol*. 2011;24(11):1984–93.
  89. Phillips M, Gleeson K, Hughes JM, Greenberg J, Cataneo RN, Baker L, et al. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *Lancet*. 1999 Jun 5;353(9168):1930–3.
  90. Phillips M, Cataneo RN, Cummin AR, Gagliardi AJ, Gleeson K, Greenberg J, et al. Detection of lung cancer with volatile markers in the breath. *Chest*. 2003;123(6):2115–23.
  91. Poli D, Goldoni M, Corradi M, Acampa O, Carbognani P, Internullo E, et al. Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatisation SPME-GC/MS. *J Chromatogr B Anal Technol Biomed Life Sci*. 2010;878(27):2643–51.
  92. Fuchs P, Loeseken C, Schubert JK, Miekisch W. Breath gas aldehydes as biomarkers of lung cancer. *Int J Cancer*. 2010;126(11):2663–70.
  93. Phillips M, Cataneo RN, Ditkoff BA, Fisher P, Greenberg J, Gunawardena R, et al. Prediction of breast cancer using volatiles in the breath. *Breast Cancer Res Treat*. 2006 Sep;99(1):19–21.
  94. Amal H, Shi D-Y, Ionescu R, Zhang W, Hua Q-L, Pan Y-Y, et al. Assessment of ovarian cancer conditions from exhaled breath. *Int J Cancer*. 2015 Mar 15;136(6):E614–22.
  95. Amal H, Leja M, Funka K, Skapars R, Sivins A, Ancans G, et al. Detection of precancerous gastric lesions and gastric cancer through exhaled breath. *Gut*. 2016 Mar;65(3):400–7.
  96. Spanel P, Smith D, Holland TA, Al Singary W, Elder JB. Analysis of formaldehyde in the headspace of urine from bladder and prostate cancer

- patients using selected ion flow tube mass spectrometry. *Rapid Commun Mass Spectrom.* 1999;13(14):1354–9.
97. Yazdanpanah M, Luo X, Lau R, Greenberg M, Fisher LJ, Lehotay DC. Cytotoxic aldehydes as possible markers for childhood cancer. *Free Radic Biol Med.* 1997;23(6):870–8.
  98. Guadagni R, Miraglia N, Simonelli A, Silvestre A, Lamberti M, Feola D, et al. Solid-phase microextraction-gas chromatography-mass spectrometry method validation for the determination of endogenous substances: urinary hexanal and heptanal as lung tumor biomarkers. *Anal Chim Acta.* 2011;701(1):29–36.
  99. Xue R, Dong L, Zhang S, Deng C, Liu T, Wang J, et al. Investigation of volatile biomarkers in liver cancer blood using solid-phase microextraction and gas chromatography/mass spectrometry. *Rapid Commun Mass Spectrom.* 2008 Apr 30;22(8):1181–6.
  100. Zajdel A, Wilczok A, Slowinski J, Orchel J, Mazurek U. Aldehydic lipid peroxidation products in human brain astrocytomas. *J Neurooncol.* 2007 Jul 23;84(2):167–73.
  101. Filipiak W, Filipiak A, Sponring A, Schmid T, Zelger B, Ager C, et al. Comparative analyses of volatile organic compounds (VOCs) from patients, tumors and transformed cell lines for the validation of lung cancer-derived breath markers. *J Breath Res.* 2014 May 27;8(2):27111.
  102. Levine DM, Ek WE, Zhang R, Liu X, Onstad L, Sather C, et al. A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett's esophagus. *Nat Genet.* 2013;45(12):1487–93.
  103. Palles C, Chegwidden L, Li X, Findlay JM, Farnham G, Castro Giner F, et al. Polymorphisms near TBX5 and GDF7 are associated with increased risk for Barrett's esophagus. *Gastroenterology.* 2015;148(2):367–78.
  104. Yamashita S, Tsujino Y, Moriguchi K, Tatematsu M, Ushijima T. Chemical genomic screening for methylation-silenced genes in gastric cancer cell lines using 5-aza-2'-deoxycytidine treatment and oligonucleotide microarray. *Cancer Sci.* 2006;97(1):64–71.
  105. Yokoyama A, Kato H, Yokoyama T, Tsujinaka T, Muto M, Omori T, et al. Genetic polymorphisms of alcohol and aldehyde dehydrogenases and glutathione S-transferase M1 and drinking, smoking, and diet in Japanese men with esophageal squamous cell carcinoma. *Carcinogenesis.* 2002;23(11):1851–9.
  106. Yokoyama A, Watanabe H, Fukuda H, Haneda T, Kato H, Yokoyama T, et al. Multiple cancers associated with esophageal and oropharyngolaryngeal

- squamous cell carcinoma and the aldehyde dehydrogenase-2 genotype in male Japanese drinkers. *Cancer Epidemiol Biomarkers Prev.* 2002;11(9):895–900.
107. Yukawa Y, Muto M, Hori K, Nagayoshi H, Yokoyama A, Chiba T, et al. Combination of ADH1B\*2/ALDH2\*2 polymorphisms alters acetaldehyde-derived DNA damage in the blood of Japanese alcoholics. *Cancer Sci.* 2012;103(9):1651–5.
  108. Homann N, Jousimies-Somer H, Jokelainen K, Heine R, Salaspuro M. High acetaldehyde levels in saliva after ethanol consumption: methodological aspects and pathogenetic implications. *Carcinogenesis.* 1997;18(9):1739–43.
  109. Homann N, Tillonen J, Rintamaki H, Salaspuro M, Lindqvist C, Meurman JH. Poor dental status increases acetaldehyde production from ethanol in saliva: a possible link to increased oral cancer risk among heavy drinkers. *Oral Oncol.* 2001;37(2):153–8.
  110. Peng D, Belkhir A, Hu T, Chaturvedi R, Asim M, Wilson KT, et al. Glutathione peroxidase 7 protects against oxidative DNA damage in oesophageal cells. *Gut.* 2012;61(9):1250–60.
  111. Yu HS, Oyama T, Matsuda T, Isse T, Yamaguchi T, Tanaka M, et al. The effect of ethanol on the formation of N<sup>2</sup>-ethylidene-dG adducts in mice: implications for alcohol-related carcinogenicity of the oral cavity and esophagus. *Biomarkers.* 2012;17(3):269–74.
  112. Nagayoshi H, Matsumoto A, Nishi R, Kawamoto T, Ichiba M, Matsuda T. Increased formation of gastric N<sup>(2)</sup>-ethylidene-2'-deoxyguanosine DNA adducts in aldehyde dehydrogenase-2 knockout mice treated with ethanol. *Mutat Res.* 2009;673(1):74–7.
  113. Matsuda T, Matsumoto A, Uchida M, Kanaly RA, Misaki K, Shibutani S, et al. Increased formation of hepatic N<sup>2</sup>-ethylidene-2'-deoxyguanosine DNA adducts in aldehyde dehydrogenase 2-knockout mice treated with ethanol. *Carcinogenesis.* 2007;28(11):2363–6.
  114. Langevin F, Crossan GP, Rosado I V, Arends MJ, Patel KJ. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature.* 2011;475(7354):53–8.
  115. Garaycochea JI, Crossan GP, Langevin F, Daly M, Arends MJ, Patel KJ. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature.* 2012;489(7417):571–5.
  116. Li SY, Li Q, Shen JJ, Dong F, Sigmon VK, Liu Y, et al. Attenuation of acetaldehyde-induced cell injury by overexpression of aldehyde

- dehydrogenase-2 (ALDH2) transgene in human cardiac myocytes: role of MAP kinase signaling. *J Mol Cell Cardiol.* 2006;40(2):283–94.
117. Lin DC, Hao JJ, Nagata Y, Xu L, Shang L, Meng X, et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nat Genet.* 2014;46(5):467–73.
118. Ma I, Allan AL. The role of human aldehyde dehydrogenase in normal and cancer stem cells. *Stem Cell Rev.* 2011;7(2):292–306.
119. Marcato P, Dean CA, Pan D, Araslanova R, Gillis M, Joshi M, et al. Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. *Stem Cells.* 2011;29(1):32–45.
120. Katsuno Y, Ehata S, Yashiro M, Yanagihara K, Hirakawa K, Miyazono K. Coordinated expression of REG4 and aldehyde dehydrogenase 1 regulating tumorigenic capacity of diffuse-type gastric carcinoma-initiating cells is inhibited by TGF-beta. *J Pathol.* 2012;228(3):391–404.
121. Wang Y, Zhe H, Gao P, Zhang N, Li G, Qin J. Cancer stem cell marker ALDH1 expression is associated with lymph node metastasis and poor survival in esophageal squamous cell carcinoma: a study from high incidence area of northern China. *Dis Esophagus.* 2012;25(6):560–5.
122. Yang L, Ren Y, Yu X, Qian F, Bian BS, Xiao HL, et al. ALDH1A1 defines invasive cancer stem-like cells and predicts poor prognosis in patients with esophageal squamous cell carcinoma. *Mod Pathol.* 2014;27(5):775–83.
123. Huang D, Gao Q, Guo L, Zhang C, Jiang W, Li H, et al. Isolation and Identification of Cancer Stem-Like Cells in Esophageal Carcinoma Cell Lines. *Stem Cells Dev.* 2009 Apr;18(3):465–74.
124. Sullivan JP, Spinola M, Dodge M, Raso MG, Behrens C, Gao B, et al. Aldehyde dehydrogenase activity selects for lung adenocarcinoma stem cells dependent on notch signaling. *Cancer Res.* 2010;70(23):9937–48.
125. Stadtman ER, Levine RL. Protein oxidation. *Ann N Y Acad Sci.* 2000;899:191–208.
126. Dalle-Donne I, Rossi R, Giustarini D, Milzani A, Colombo R. Protein carbonyl groups as biomarkers of oxidative stress. *Clin Chim Acta.* 2003 Mar;329(1–2):23–38.
127. Wang Y, Yu H, Shi X, Luo Z, Lin D, Huang M. Structural mechanism of ring-opening reaction of glucose by human serum albumin. *J Biol Chem.* 2013 May 31;288(22):15980–7.
128. Sayre LM, Lin D, Yuan Q, Zhu X, Tang X. Protein Adducts Generated from

- Products of Lipid Oxidation: Focus on HNE and ONE. *Drug Metab Rev.* 2006 Jan 9;38(4):651–75.
129. Levine RL, Yang IY, Hossain M, Pandya GA, Grollman AP, Moriya M. Mutagenesis induced by a single 1,N6-ethenodeoxyadenosine adduct in human cells. *Cancer Res.* 2000;60(15):4098–104.
  130. Monographs on the Evaluation of Carcinogenic Risk to Humans [Internet]. International Agency of Research on Cancer. World Health Organisation; Available from: [http://monographs.iarc.fr/ENG/Classification/latest\\_classif.php](http://monographs.iarc.fr/ENG/Classification/latest_classif.php)
  131. Esterbauer H, Schaur RJ, Zollner H. Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. *Free Radic Biol Med.* 1991;11(1):81–128.
  132. Zhang L, Freeman LE, Nakamura J, Hecht SS, Vandenberg JJ, Smith MT, et al. Formaldehyde and leukemia: epidemiology, potential mechanisms, and implications for risk assessment. *Env Mol Mutagen.* 2010;51(3):181–91.
  133. Seitz HK, Stickel F. Acetaldehyde as an underestimated risk factor for cancer development: role of genetics in ethanol metabolism. *Genes Nutr.* 2010;5(2):121–8.
  134. Yu HS, Oyama T, Isse T, Kitagawa K, Pham TT, Tanaka M, et al. Formation of acetaldehyde-derived DNA adducts due to alcohol exposure. *Chem Biol Interact.* 2010;188(3):367–75.
  135. Seitz HK, Meier P. The role of acetaldehyde in upper digestive tract cancer in alcoholics. *Transl Res.* 2007;149(6):293–7.
  136. Salaspuro M. Interactions of alcohol and tobacco in gastrointestinal cancer. *J Gastroenterol Hepatol.* 2012;27 Suppl 2:135–9.
  137. Amma H, Naruse K, Ishiguro N. Involvement of reactive oxygen species in cyclic stretch- induced NF-  $\kappa$ B activation in human fibroblast cells. *Br J.* 2005;
  138. Mitsuishi Y, Motohashi H, Yamamoto M. The Keap1–Nrf2 system in cancers: stress response and anabolic metabolism. *Front Oncol.* 2012;2:200.
  139. Vasiliou V, Qamar L, Pappa A, Sophos NA, Petersen DR. Involvement of the electrophile responsive element and p53 in the activation of hepatic stellate cells as a response to electrophile menadione. *Arch Biochem Biophys.* 2003;413(2):164–71.
  140. Carvalho VM, Asahara F, Di Mascio P, Campos IPD, Cadet J, Medeiros MHG. Novel 1,N-6-etheno-2'-deoxyadenosine adducts from lipid peroxidation products. *Chem Res Toxicol.* 2000;13(5):397–405.
  141. Lin P, Lee H-L, Cheng H-I, Chen C-Y, Tsai M-H, Liu H-J. Metabolomic profiling of mice urine and serum associated with trans-trans 2, 4-decadienal induced

- lung lesions by liquid chromatography-mass spectrometry. *Anal Bioanal Chem.* 2014 Jul 28;406(17):4287–97.
142. Chang LW, Lo W-S, Lin P. Trans, trans-2,4-decadienal, a product found in cooking oil fumes, induces cell proliferation and cytokine production due to reactive oxygen species in human bronchial epithelial cells. *Toxicol Sci.* 2005 Oct;87(2):337–43.
  143. Emidio Camaioni †, José L. Boyer ‡, Arvind Mohanram ‡, T. Kendall Harden ‡ and, Kenneth A. Jacobson\* †. Deoxyadenosine Bisphosphate Derivatives as Potent Antagonists at P2Y1 Receptors. 1998;
  144. Clave P, Shaker R. Dysphagia: current reality and scope of the problem. *Nat Rev Gastroenterol Hepatol.* 2015 May;12(5):259–70.
  145. Kozutsumi D, Arita M, Kawashima A, Adachi M, Takami M. An improved method for acetaldehyde determination in blood by high-performance liquid chromatography and solid-phase extraction. *J Chromatogr Sci.* 2002;40(9):477–82.
  146. Abbassi-Ghadi N, Veselkov K, Kumar S, Huang J, Jones E, Strittmatter N, et al. Discrimination of lymph node metastases using desorption electrospray ionisation-mass spectrometry imaging. *Chem Commun.* 2014;50(28):3661.
  147. Abbassi-Ghadi N, Golf O, Kumar S, Antonowicz S, McKenzie JS, Huang J, et al. Imaging of Esophageal Lymph Node Metastases by Desorption Electrospray Ionization Mass Spectrometry. *Cancer Res.* 2016 Oct 1;76(19):5647–56.
  148. Hills BA. Oesophageal surfactant: evidence for a possible mucosal barrier on oesophageal epithelium. *Aust N Z J Med.* 1994 Feb;24(1):41–6.
  149. Sihvo EIT, Salminen JT, Rantanen TK, Rämö OJ, Ahotupa M, Färkkilä M, et al. Oxidative stress has a role in malignant transformation in Barrett's oesophagus. *Int J Cancer.* 2002 Dec 20;102(6):551–5.
  150. Pavlova NN, Thompson CB. Perspective The Emerging Hallmarks of Cancer Metabolism. *Cell Metab.* 2016;23(1):27–47.
  151. Berdyshev E V. Mass spectrometry of fatty aldehydes. *Biochim Biophys Acta - Mol Cell Biol Lipids.* 2011 Nov;1811(11):680–93.
  152. Determination of Carbonyl Compounds. *Environ Prot Agency.* 1996;100(December):1–34.
  153. Ojeda AG, Wrobel K, Escobosa ARC, Garay-Sevilla ME, Wrobel K. High-performance liquid chromatography determination of glyoxal, methylglyoxal, and diacetyl in urine using 4-methoxy-o-phenylenediamine as derivatizing reagent. *Anal Biochem.* 2014;449(1):52–8.

154. Tan Y, Siebert KJ. Modeling bovine serum albumin binding of flavor compounds (alcohols, aldehydes, esters, and ketones) as a function of molecular properties. *J Food Sci.* 2008;73(1).
155. Olivares A, Dryahina K, Navarro JL, Flores M, Smith D, Spanel P. Selected ion flow tube-mass spectrometry for absolute quantification of aroma compounds in the headspace of dry fermented sausages. *Anal Chem.* 2010;82(13):5819–29.
156. Esterbauer H, Zollner H. Methods for determination of aldehydic lipid peroxidation products. *Free Radic Biol Med.* 1989;7(2):197–203.
157. Brady OL, Elsmie G V. The use of 2:4-dinitrophenylhydrazine as a reagent for aldehydes and ketones. *Analyst.* 1926;51(599):77.
158. O'Brien JS, Sampson EL. Fatty acid and fatty aldehyde composition of the major brain lipids in normal human gray matter, white matter, and myelin. *J Lipid Res.* 1965;6:545–51.
159. Gilbertson JR, Ferrell WJ, Gelman RA. From Rat , Dog , and Bovine Heart Muscle. 1967;8.
160. Wheatley RA. Some recent trends in the analytical chemistry of lipid peroxidation. *TRAC.* 2634:24–10.
161. Deng C, Zhang X. A simple, rapid and sensitive method for determination of aldehydes in human blood by gas chromatography/mass spectrometry and solid-phase microextraction with on-fiber derivatization. *Rapid Commun Mass Spectrom.* 2004 Aug 15;18(15):1715–20.
162. Luo XP, Yazdanpanah M, Bhooi N, Lehotay DC. Determination of aldehydes and other lipid peroxidation products in biological samples by gas chromatography-mass spectrometry. Vol. 228, *Analytical biochemistry.* 1995. p. 294–8.
163. Kouremenos KA, Johansson M, Marriott PJ. Advances in gas chromatographic methods for the identification of biomarkers in cancer. *J Cancer.* 2012;3(1):404–20.
164. Chen T, Cao Y, Zhang Y, Liu J, Bao Y, Wang C, et al. Random Forest in Clinical Metabolomics for Phenotypic Discrimination and Biomarker Selection. *Evidence-Based Complement Altern Med.* 2013;
165. Li N, Deng C, Yin X, Yao N, Shen X, Zhang X. Gas chromatography-mass spectrometric analysis of hexanal and heptanal in human blood by headspace single-drop microextraction with droplet derivatization. *Anal Biochem.* 2005 Jul 15;342(2):318–26.
166. Chen F, Wang C, Zhang M, Zhang X, Liu Y, Ye J, et al. Sensitive



- determination of endogenous hexanal and heptanal in urine by hollow-fiber liquid-phase microextraction prior to capillary electrophoresis with amperometric detection. *Talanta*. 2014;119:83–9.
167. Elisia I, Kitts DD. Quantification of hexanal as an index of lipid oxidation in human milk and association with antioxidant components. *J Clin Biochem Nutr*. 2011;49(3):147–52.
  168. Fung K, Grosjean D. Determination of nanogram amounts of carbonyls as 2,4-dinitrophenylhydrazones by high-performance liquid chromatography. *Anal Chem*. 1981;53(2):168–71.
  169. Tomono S, Miyoshi N, Ohshima H. Comprehensive analysis of the lipophilic reactive carbonyls present in biological specimens by LC/ESI-MS/MS. *J Chromatogr B Anal Technol Biomed Life Sci*. 2015;988:149–56.
  170. Eggink M, Wijtmans M, Kretschmer A, Kool J, Lingeman H, De Esch IJP, et al. Targeted LC-MS derivatization for aldehydes and carboxylic acids with a new derivatization agent 4-APEBA. *Anal Bioanal Chem*. 2010;397(2):665–75.
  171. Tie C, Hu T, Jia Z-X, Zhang J-L. Derivatization Strategy for the Comprehensive Characterization of Endogenous Fatty Aldehydes Using HPLC-Multiple Reaction Monitoring. *Anal Chem*. 2016;88(15):7762–8.
  172. Williams TI, Lovell MA, Lynn BC. Analysis of derivatized biogenic aldehydes by LC tandem mass spectrometry. *Anal Chem*. 2005 May 15;77(10):3383–9.
  173. International Organization for Standardization. Indoor air -- Part 3: Determination of formaldehyde and other carbonyl compounds in indoor air and test chamber air -- Active sampling method [Internet]. ISO 16000. 2011 [cited 2016 Sep 30]. p. 3. Available from: [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=51812](http://www.iso.org/iso/catalogue_detail.htm?csnumber=51812)
  174. Siegel D, Meinema AC, Permentier H, Hopfgartner G, Bischoff R. Integrated Quantification and Identification of Aldehydes and Ketones in Biological Samples. *Anal Chem*. 2014 May 20;86(10):5089–100.
  175. Mathew S, Grey C, Rumpunen K, Adlercreutz P. Analysis of carbonyl compounds in sea buckthorn for the evaluation of triglyceride oxidation, by enzymatic hydrolysis and derivatisation methodology. *Food Chem*. 2011;126(3):1399–405.
  176. Zwiener C, Glauner T, Frimmel FH. Method optimization for the determination of carbonyl compounds in disinfected water by DNPH derivatization and LC-ESI-MS-MS. *Anal Bioanal Chem*. 2002 Mar 19;372(5–6):615–21.
  177. Nagy K, Pollreisz F, Takáts Z, Vékey K. Atmospheric pressure chemical ionization mass spectrometry of aldehydes in biological matrices. *Rapid*

- Commun Mass Spectrom. 2004;18:2473–8.
178. Andreoli R, Manini P, Corradi M, Mutti A, Niessen WMA. Determination of patterns of biologically relevant aldehydes in exhaled breath condensate of healthy subjects by liquid chromatography/atmospheric chemical ionization tandem mass spectrometry. *Rapid Commun Mass Spectrom.* 2003;17(7):637–45.
  179. Miller JH, Gardner WP, Gonzalez RR. UHPLC separation with MS analysis for eight carbonyl compounds in mainstream tobacco smoke. *J Chromatogr Sci.* 2010;48(January):12–7.
  180. FDA. Analytical Procedures and Methods Validation for Drugs and Biologics. *Guid Ind.* 2015;(July):1–15.
  181. Shestivska V, Antonowicz SS, Dryahina K, Kubista J, Smith D, Spanel P. Direct detection and quantification of malondialdehyde vapour in humid air using selected ion flow tube mass spectrometry supported by gas chromatography/mass spectrometry. *Rapid Commun Mass Spectrom.* 2015;29(11):1069–79.
  182. Want EJ, Masson P, Michopoulos F, Wilson ID, Theodoridis G, Plumb RS, et al. Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat Protoc.* 2012 Dec 6;8(1):17–32.
  183. Liang X, Ubhayakar S, Liederer BM, Dean B, Ran-Ran Qin A, Shahidi-Latham S, et al. Evaluation of homogenization techniques for the preparation of mouse tissue samples to support drug discovery. *Bioanalysis.* 2011 Sep;3(17):1923–33.
  184. Members MSIB, Sansone SA, Fan T, Goodacre R, Griffin JL, Hardy NW, et al. The metabolomics standards initiative. *Nat Biotechnol.* 2007;25(8):846–8.
  185. Beckonert O, Keun HC, Ebbels TMD, Bundy JG, Holmes E, Lindon JC, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc.* 2007;2(11):2692–703.
  186. Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, et al. Global metabolic profiling procedures for urine using UPLC-MS. *Nat Protoc.* 2010 Jun 6;5(6):1005–18.
  187. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc.* 2011;6(7):1060–83.
  188. Sargent M. Guide to achieving reliable quantitative LC-MS measurements.

- 2013.
189. Kölliker S, Oehme M, Dye C. Structure Elucidation of 2,4-Dinitrophenylhydrazone Derivatives of Carbonyl Compounds in Ambient Air by HPLC/MS and Multiple MS/MS Using Atmospheric Chemical Ionization in the Negative Ion Mode. *Anal Chem.* 1998;70(9):1979–85.
  190. Gosetti F, Mazzucco E, Zampieri D, Gennaro MC. Signal suppression/enhancement in high-performance liquid chromatography tandem mass spectrometry. *J Chromatogr A.* 2010;1217(25):3929–37.
  191. King R, Bonfiglio R, Fernandez-Metzler C, Miller-Stein C, Olah T. Mechanistic investigation of ionization suppression in electrospray ionization. *J Am Soc Mass Spectrom.* 2000;11(11):942–50.
  192. Richardson SD, Caughran T V., Poiger T, Guo Y, Crumley FG. Application of DNPH Derivatization with LC/MS to the Identification of Polar Carbonyl Disinfection Byproducts in Drinking Water. *Ozone Sci Eng.* 2000 Dec;22(6):653–75.
  193. Uchiyama S, Inaba Y, Kunugita N. Derivatization of carbonyl compounds with 2,4-dinitrophenylhydrazine and their subsequent determination by high-performance liquid chromatography. *J Chromatogr B Anal Technol Biomed Life Sci.* 2011;879(17–18):1282–9.
  194. Xu H, Lv L, Hu S, Song D. High-performance liquid chromatographic determination of hexanal and heptanal in human blood by ultrasound-assisted headspace liquid-phase microextraction with in-drop derivatization. *J Chromatogr A.* 2010;1217(16):2371–5.
  195. Snyder LR, Kirkland JJ, Glajch JL. *Practical HPLC Method Development.* Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1997.
  196. Snyder LR, Kirkland JJ, Dolan JW. *Introduction to Modern Liquid Chromatography.* Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2009.
  197. Jeong H-S, Chung H, Song S-H, Kim C-I, Lee J-G, Kim Y-S. Validation and Determination of the Contents of Acetaldehyde and Formaldehyde in Foods. *Toxicol Res.* 2015 Sep;31(3):273–8.
  198. Sugaya N, Sakurai K, Nakagawa T, Onda N, Onodera S, Morita M, et al. Development of a headspace GC/MS analysis for carbonyl compounds (aldehydes and ketones) in household products after derivatization with o-(2,3,4,5,6-pentafluorobenzyl)hydroxylamine. *Anal Sci.* 2004 May;20(5):865–70.
  199. Korchazhkina O, Exley C, Andrew Spencer S. Measurement by reversed-phase high-performance liquid chromatography of malondialdehyde in normal

- human urine following derivatisation with 2, 4-dinitrophenylhydrazine. *J Chromatogr B*. 2003;794(2):353–362.
200. Bicking MKL, Marcus Cooke W, Kawahara FK, Longbottom JE. Effect of pH on the reaction of 2,4-dinitrophenylhydrazine with formaldehyde and acetaldehyde. *J Chromatogr A*. 1988;455(C):310–5.
201. Rittenberg, D; Foster GL. A new procedure for quantitative analysis by isotope dilution, with applicaiton to the determination of amino acids and fatty acids. *J Biol Chem*. 1940;134:653.
202. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007;3(3):211–21.
203. Lin CY, Wu H, Tjeerdema RS, Viant MR. Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. *Metabolomics*. 2007 Mar 14;3(1):55–67.
204. Lovell MA, Ehmann WD, Butler SM, Markesbery WR. Elevated thiobarbituric acid-reactive substances and antioxidant enzyme activity in the brain in Alzheimer's disease. *Neurology*. 1995;45(8):1594–601.
205. Keller JN, Schmitt FA, Scheff SW, Ding Q, Chen Q, Butterfield DA, et al. Evidence of increased oxidative damage in subjects with mild cognitive impairment. *Neurology*. 2005;64(7):1152–6.
206. Kolanjiappan K, Manoharan S, Kayalvizhi M. Measurement of erythrocyte lipids, lipid peroxidation, antioxidants and osmotic fragility in cervical cancer patients. *Clin Chim Acta*. 2002;326(1):143–9.
207. Galipeau P, Prevo L, Sanchez C. Clonal expansion and loss of heterozygosity at chromosomes 9p and 17p in premalignant esophageal (Barrett's) tissue. *Natl Cancer ....* 1999;
208. Galipeau P, Cowan D. 17p (p53) allelic losses, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. *Proc*. 1996;
209. Chuang S-C, Hashibe M, Scelo G, Brewster DH, Pukkala E, Friis S, et al. Risk of Second Primary Cancer among Esophageal Cancer Patients: a Pooled Analysis of 13 Cancer Registries. *Cancer Epidemiol Prev Biomarkers*. 2008;17(6).
210. Hodson L, Skeaff CM, Fielding BA. Fatty acid composition of adipose tissue and blood in humans and its use as a biomarker of dietary intake. *Prog Lipid Res*. 2008;47(5):348–80.
211. Tian L, Goldstein A, Wang H, Ching Lo H, Sun Kim I, Welte T, et al. Mutual

- regulation of tumour vessel normalization and immunostimulatory reprogramming. *Nature*. 2017 Apr 3;
212. Dall'Olmo L, Fassan M, Dassie E, Scarpa M, Realdon S, Cavallin F, et al. Role of Proton Pump Inhibitor on Esophageal Carcinogenesis and Pancreatic Acinar Cell Metaplasia Development: An Experimental In Vivo Study. Mukaisho K, editor. *PLoS One*. 2014 Nov 21;9(11):e112862.
  213. Quante M, Bhagat G, Abrams JA, Marache F, Good P, Lee MD, et al. Bile Acid and Inflammation Activate Gastric Cardia Stem Cells in a Mouse Model of Barrett-Like Metaplasia. *Cancer Cell*. 2012 Jan 17;21(1):36–51.
  214. Chen X, Ding YW, Yang G y, Bondoc F, Lee MJ, Yang CS. Oxidative damage in an esophageal adenocarcinoma model with rats. *Carcinogenesis*. 2000 Feb;21(2):257–63.
  215. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, et al. The Human Urine Metabolome. Dzeja P, editor. *PLoS One*. 2013 Sep 4;8(9):e73076.
  216. Warrack BM, Hnatyshyn S, Ott K-H, Reily MD, Sanders M, Zhang H, et al. Normalization strategies for metabonomic analysis of urine samples. *J Chromatogr B*. 2009 Feb;877(5–6):547–52.
  217. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
  218. Hyland PL, Hu N, Rotunno M, Su H, Wang C, Wang L, et al. Global changes in gene expression of Barrett's esophagus compared to normal squamous esophagus and gastric cardia tissues. *PLoS One*. 2014/04/10. 2014;9(4):e93219.
  219. Kim SM, Park YY, Park ES, Cho JY, Izzo JG, Zhang D, et al. Prognostic biomarkers for esophageal adenocarcinoma identified by analysis of tumor transcriptome. *PLoS One*. 2010;5(11):e15074.
  220. di Pietro M, Lao-Sirieix P, Boyle S, Cassidy A, Castillo D, Saadi A, et al. Evidence for a functional role of epigenetically regulated midcluster HOXB genes in the development of Barrett esophagus. *Proc Natl Acad Sci U S A*. 2012;109(23):9077–82.
  221. TCGA. Home - The Cancer Genome Atlas - Cancer Genome - TCGA [Internet]. [cited 2016 Jun 30]. Available from: <http://cancergenome.nih.gov/>
  222. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*.

- 2011;12(4):R41.
223. Schlemper RJ, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM, et al. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut*. 2000 Aug;47(2):251–5.
  224. Pena-Llopis S, Brugarolas J. Simultaneous isolation of high-quality DNA, RNA, miRNA and proteins from tissues for genomic applications. *Nat Protoc*. 2013;8(11):2240–55.
  225. Wang Q, Ma C, Kemmner W. Wdr66 is a novel marker for risk stratification and involved in epithelial-mesenchymal transition of esophageal squamous cell carcinoma. *BMC Cancer*. 2013/03/22. 2013;13:137.
  226. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*. 2009;55(4):611–22.
  227. Bustin SA, Benes V, Garson J, Hellemans J, Huggett J, Kubista M, et al. The need for transparency and good practices in the qPCR literature. *Nat Methods*. 2013;10(11):1063–7.
  228. Alvi MA, Liu X, O'Donovan M, Newton R, Wernisch L, Shannon NB, et al. DNA methylation as an adjunct to histopathology to detect prevalent, inconspicuous dysplasia and early-stage neoplasia in Barrett's esophagus. *Clin Cancer Res*. 2013;19(4):878–88.
  229. Hong J, Resnick M, Behar J, Wang LJ, Wands J, DeLellis RA, et al. Acid-induced p16 hypermethylation contributes to development of esophageal adenocarcinoma via activation of NADPH oxidase NOX5-S. *Am J Physiol Gastrointest Liver Physiol*. 2010;299(3):G697-706.
  230. PrimerBank [Internet]. [cited 2017 Apr 10]. Available from: <https://pga.mgh.harvard.edu/primerbank/>
  231. BLAST: Basic Local Alignment Search Tool [Internet]. [cited 2017 Apr 10]. Available from: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
  232. Marchitti SA, Orlicky DJ, Brocker C, Vasiliou V. Aldehyde Dehydrogenase 3B1 (ALDH3B1): Immunohistochemical Tissue Distribution and Cellular-specific Localization in Normal and Cancerous Human Tissues. *J Histochem Cytochem*. 2010;58(9):765–83.
  233. Kitamura T, Naganuma T, Abe K, Nakahara K, Ohno Y, Kihara A. Substrate specificity, plasma membrane localization, and lipid modification of the aldehyde dehydrogenase ALDH3B1. *Biochim Biophys Acta - Mol Cell Biol Lipids*. 2013;1831(8):1395–401.
  234. Boonstra JJ, Van Marion R, Beer DG, Lin L, Chaves P, Ribeiro C, et al.

- Verification and unmasking of widely used human esophageal adenocarcinoma cell lines. *J Natl Cancer Inst.* 2010;102(4):271–4.
235. Kalabis J, Wong GS, Vega ME, Natsuzaka M, Robertson ES, Herlyn M, et al. Isolation and characterization of mouse and human esophageal epithelial cells in 3D organotypic culture. *Nat Protoc.* 2012/01/14. 2012;7(2):235–46.
236. Chapman S, Liu X, Meyers C, Schlegel R, McBride AA. Human keratinocytes are efficiently immortalized by a Rho kinase inhibitor. *J Clin Invest.* 2010/06/03. 2010;120(7):2619–26.
237. Burnette W. “Western blotting”: electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with. *Anal Biochem.* 1981;
238. COSMIC: Catalogue of Somatic Mutations in Cancer - Home Page [Internet]. [cited 2017 Apr 6]. Available from: <http://cancer.sanger.ac.uk/cosmic>
239. Reid B, Prevo L, Galipeau P, Sanchez C. Predictors of progression in Barrett’s esophagus II: baseline 17p (p53) loss of heterozygosity identifies a patient subset at increased risk for neoplastic progression. *Am J.* 2001;
240. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013 Apr 2;6(269):p11.
241. Liu Y, Chen C, Xu Z, Scuoppo C, Rillahan CD, Gao J, et al. Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. *Nature.* 2016 Mar 16;531(7595):471–5.
242. Pappa A, Chen C, Koutalos Y, Townsend AJ, Vasiliou V. Aldh3a1 protects human corneal epithelial cells from ultraviolet- and 4-hydroxy-2-nonenal-induced oxidative damage. *Free Radic Biol Med.* 2003;34(9):1178–89.
243. Black W, Chen Y, Matsumoto A, Thompson DC, Lassen N, Pappa A, et al. Molecular mechanisms of ALDH3A1-mediated cellular protection against 4-hydroxy-2-nonenal. *Free Radic Biol Med.* 2012;52(9):1937–44.
244. Medical Research Council Oesophageal Cancer Working Group, Party MRCW, Medical Research Council Oesophageal Cancer Working Party\*. Surgical resection with or without preoperative chemotherapy in oesophageal cancer: a randomised controlled trial. *Lancet.* 2002 May 18;359(9319):1727–33.
245. Yokoyama A, Muramatsu T, Ohmori T, Yokoyama T, Okuyama K, Takahashi H, et al. Alcohol-related cancers and aldehyde dehydrogenase-2 in Japanese alcoholics. *Carcinogenesis.* 1998;19(8):1383–7.
246. Deng S, Yang X, Lassus H, Liang S, Kaur S, Ye Q, et al. Distinct expression

- levels and patterns of stem cell marker, aldehyde dehydrogenase isoform 1 (ALDH1), in human epithelial cancers. *PLoS One*. 2010;5(4).
247. van den Hoogen C, van der Horst G, Cheung H, Buijs JT, Lippitt JM, Guzman-Ramirez N, et al. High aldehyde dehydrogenase activity identifies tumor-initiating and metastasis-initiating cells in human prostate cancer. *Cancer Res*. 2010;70(12):5163–73.
248. Huang J, Hu N, Goldstein AM, Emmert-Buck MR, Tang ZZ, Roth MJ, et al. High frequency allelic loss on chromosome 17p13.3-p11.1 in esophageal squamous cell carcinomas from a high incidence area in northern China. *Carcinogenesis*. 2000/11/04. 2000;21(11):2019–26.
249. Wang X, Li X, Cheng Y, Sun X, Sun X, Self S, et al. Copy number alterations detected by whole-exome and whole-genome sequencing of esophageal adenocarcinoma. *Hum Genomics*. 2015;9(1):22.
250. Liu S, Wang Y. Mass spectrometry for the assessment of the occurrence and biological consequences of DNA adducts. *Chem Soc Rev*. 2015;44(44):7829–54.
251. Taghizadeh K, L MJ, Pang B, Sullivan M, Dong M, Plummer E, et al. Quantification of {DNA} damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry. *Nat Protoc*. 2008;3(8):1287–98.
252. Lonkar P, Dedon PC. Reactive species and DNA damage in chronic inflammation: Reconciling chemical mechanisms and biological fates. *Int J Cancer*. 2011;128(9):1999–2009.
253. Masuda M, Nishihira T, Itoh K, Mizugaki M, Ishida N, Mori S. An immunohistochemical analysis for cancer of the esophagus using monoclonal antibodies specific for modified nucleosides. *Cancer*. 1993 Dec 15;72(12):3571–8.
254. Phillips DH, Arlt VM. The <sup>32</sup>P-postlabeling assay for DNA adducts. *Nat Protoc*. 2007 Nov;2(11):2772–81.
255. Kim D-Y, Cho M-H, Yang H-K, Hemminki K, Kim J-P, Jang J-J, et al. Detection of Methylation Damage in DNA of Gastric Cancer Tissues Using <sup>32</sup>P-Postlabelling Assay. *Japanese J Cancer Res*. 1999 Oct;90(10):1104–8.
256. Scates DK, Spigelman AD, Phillips RK., Venitt S. The use of <sup>32</sup>P-postlabelling in studies of the nature and origin of DNA adducts formed by bile from patients with familial adenomatous polyposis and from normal patients. *Mutat Res Mol Mech Mutagen*. 1997;378(1):113–25.



257. Scates DK, Spigelman, Phillips RKS. Differences in the levels and pattern of DNA-adduct labelling in human cell lines MCL-5 and CCRF, proficient or deficient in carcinogen-metabolism, treated in vitro with bile from familial adenomatous polyposis patients and from unaffected controls. *Carcinogenesis*. 1996;17(4):7–7.
258. Chou PH, Kageyama S, Matsuda S, Kanemoto K, Sasada Y, Oka M, et al. Detection of lipid peroxidation-induced DNA adducts caused by 4-Oxo-2(E)-nonenal and 4-Oxo-2(E)-hexenal in human autopsy tissues. *Chem Res Toxicol*. 2010;23(9):1442–8.
259. Matsuda T, Tao H, Goto M, Yamada H, Suzuki M, Wu Y, et al. Lipid peroxidation-induced DNA adducts in human gastric mucosa. *Carcinogenesis*. 2013;34(1):121–7.
260. Yukawa Y, Ohashi S, Amanuma Y, Nakai Y, Tsurumaki M, Kikuchi O, et al. Impairment of aldehyde dehydrogenase 2 increases accumulation of acetaldehyde-derived DNA damage in the esophagus after ethanol ingestion. *Am J Cancer Res*. 2014;4(3):279–84.
261. Blair IA. DNA adducts with lipid peroxidation products. *J Biol Chem*. 2008;283(23):15545–9.
262. H. L, S. C, S. W, X. J, S. Z, R. Z, et al. Ultrasensitive UPLC-MS/MS method for analysis of etheno-DNA adducts in human white blood cells. *Free Radic Res*. 2015;49(9):1049–54.
263. Zhang S, Villalta PW, Wang M, Hecht SS. Analysis of Crotonaldehyde- and Acetaldehyde-Derived 1, N<sup>2</sup> -Propanodeoxyguanosine Adducts in DNA from Human Tissues Using Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometry. *Chem Res Toxicol*. 2006;19(10):1386–92.
264. Ishino K, Kato T, Kato M, Shibata T, Watanabe M, Wakabayashi K, et al. Comprehensive DNA adduct analysis reveals pulmonary inflammatory response contributes to genotoxic action of magnetite nanoparticles. *Int J Mol Sci*. 2015;16(2):3474–92.
265. Pang B, Zhou X, Yu H, Dong M, Taghizadeh K, Wishnok JS, et al. Lipid peroxidation dominates the chemistry of DNA adduct formation in a mouse model of inflammation. *Carcinogenesis*. 2007;28(8):1807–13.
266. Chen HJ, Chung FL. Formation of etheno adducts in reactions of enals via autoxidation. *Chem Res Toxicol*. 7(6):857–60.
267. Masuda M, Suzuki T, Friesen MD, Ravanat JL, Cadet J, Pignatelli B, et al. Chlorination of guanosine and other nucleosides by hypochlorous acid and

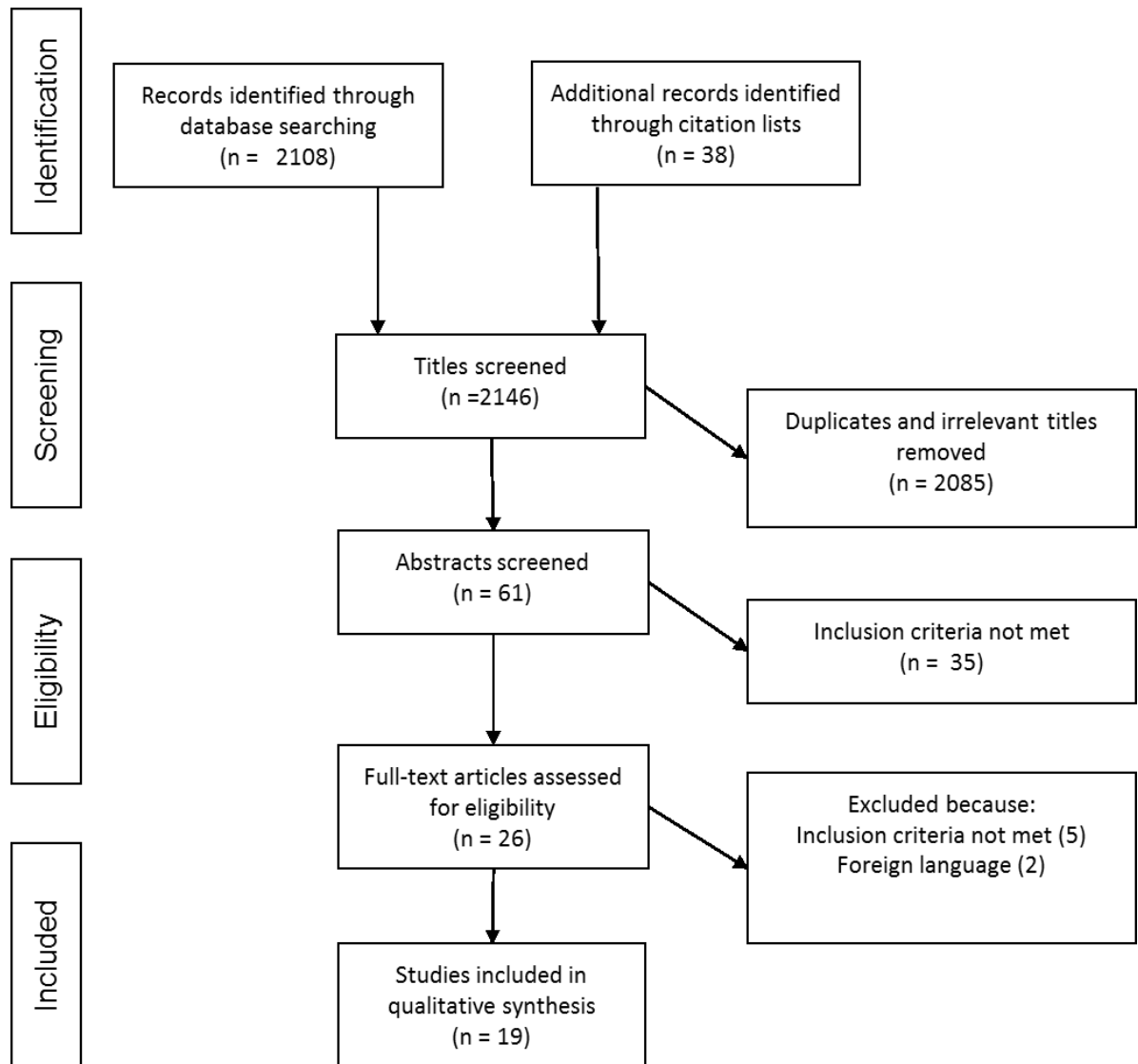
- myeloperoxidase of activated human neutrophils: Catalysis by nicotine and trimethylamine. *J Biol Chem.* 2001;276(44):40486–96.
268. Asahi T, Kondo H, Masuda M, Nishino H, Aratani Y, Naito Y, et al. Chemical and immunochemical detection of 8-halogenated deoxyguanosines at early stage inflammation. *J Biol Chem.* 2010;285(12):9282–91.
269. Detection. In: *Introduction to Modern Liquid Chromatography.* Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010. p. 147–97.
270. Rodríguez-Gonzalo E, Herrero-Herrero L, García-Gómez D. Development, validation and application of a fast analytical methodology for the simultaneous determination of DNA- and RNA-derived urinary nucleosides by liquid chromatography coupled to tandem mass spectrometry. *J Chromatogr B Anal Technol Biomed Life Sci.* 2016;1019:132–9.
271. Churchwell MI, Beland FA, Doerge DR. Quantification of Multiple DNA Adducts Formed through Oxidative Stress Using Liquid Chromatography and Electrospray Tandem Mass Spectrometry. 2001;
272. Chen H-JC, Lin G-J, Lin W-P. Simultaneous Quantification of Three Lipid Peroxidation-Derived Etheno Adducts in Human DNA by Stable Isotope Dilution Nanoflow Liquid Chromatography Nanospray Ionization Tandem Mass Spectrometry. *Anal Chem.* 2010 Jun;82(11):4486–93.
273. Hill BG, Awe SO, Vladykovskaya E, Ahmed Y, Liu S-Q, Bhatnagar A, et al. Myocardial ischaemia inhibits mitochondrial metabolism of 4-hydroxy-trans-2-nonenal. *Biochem J.* 2009 Jan 15;417(2):513–24.
274. Chen LS, Bahr MH, Sheppard TL. Effects of 8-chlorodeoxyadenosine on DNA synthesis by the klenow fragment of DNA polymerase I. Vol. 13, *Bioorganic & Medicinal Chemistry Letters.* 2003.
275. Chen CH, Ferreira JC, Gross ER, Mochly-Rosen D. Targeting aldehyde dehydrogenase 2: new therapeutic opportunities. *Physiol Rev.* 2014;94(1):1–34.
276. Meier B, Cooke SL, Weiss J, Bailly AP, Alexandrov LB, Marshall J, et al. *C. elegans* whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. *Genome Res.* 2014 Oct;24(10):1624–36.
277. Morgan CA, Parajuli B, Buchman CD, Dria K, Hurley TD. N,N-diethylaminobenzaldehyde (DEAB) as a substrate and mechanism-based inhibitor for human ALDH isoenzymes. *Chem Biol Interact.* 2014/12/17. 2015;234:18–28.
278. Chippendale TW, Hu B, El Haj AJ, Smith D. A study of enzymatic activity in

- cell cultures via the analysis of volatile biomarkers. *Analyst*. 2012;137(20):4677–85.
279. Rutter A V, Chippendale TW, Yang Y, Spanel P, Smith D, Sule-Suso J. Quantification by SIFT-MS of acetaldehyde released by lung cells in a 3D model. *Analyst*. 2013;138(1):91–5.
  280. Filipiak W, Sponring A, Filipiak A, Ager C, Schubert J, Miekisch W, et al. TD-GC-MS analysis of volatile metabolites of human lung cancer and normal cells in vitro. *Cancer Epidemiol Biomarkers Prev*. 2010;19(1):182–95.
  281. Mochalski P, Sponring A, King J, Unterkofler K, Troppmair J, Amann A. Release and uptake of volatile organic compounds by human hepatocellular carcinoma cells (HepG2) in vitro. *Cancer Cell Int*. 2013;13(1):72.
  282. Brunner C, Szymczak W, Hollriegl V, Mortl S, Oelmez H, Bergner A, et al. Discrimination of cancerous and non-cancerous cell lines by headspace-analysis with PTR-MS. *Anal Bioanal Chem*. 2010;397(6):2315–24.
  283. Sponring A, Filipiak W, Ager C, Schubert J, Miekisch W, Amann A, et al. Analysis of volatile organic compounds (VOCs) in the headspace of NCI-H1666 lung cancer cells. *Cancer Biomark*. 2010;7(3):153–61.
  284. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* (80- ). 2012;337(6096).
  285. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013 Oct 24;8(11):2281–308.
  286. Yara S, Lavoie J-C, Beaulieu J-F, Delvin E, Amre D, Marcil V, et al. Iron-Ascorbate-Mediated Lipid Peroxidation Causes Epigenetic Changes in the Antioxidant Defense in Intestinal Epithelial Cells: Impact on Inflammation. Tsuji Y, editor. *PLoS One*. 2013 May 22;8(5):e63456.
  287. Barbera M, Fitzgerald RC. Cellular origin of Barrett's metaplasia and oesophageal stem cells. *Biochem Soc Trans*. 2010;38(2):370–3.
  288. Keller MA, Zander U, Fuchs JE, Kreutz C, Watschinger K, Mueller T, et al. A gatekeeper helix determines the substrate specificity of Sjogren-Larsson Syndrome enzyme fatty aldehyde dehydrogenase. *Nat Commun*. 2014;5:4439.
  289. Willemsen MAAP, IJlst L, Steijlen PM, Rotteveel JJ, de Jong JGN, van Domburg PHMF, et al. Clinical, biochemical and molecular genetic characteristics of 19 patients with the Sjögren–Larsson syndrome. *Brain*. 2001;124(7).

290. Barbera M, di Pietro M, Walker E, Brierley C, Macrae S, Simons BD, et al. The human squamous oesophagus has widespread capacity for clonal expansion from cells at diverse stages of differentiation. *Gut*. 2014;0(1):1–9.
291. Haug S, Braun-Falco M. Restoration of fatty aldehyde dehydrogenase deficiency in Sjögren–Larsson syndrome. *Gene Ther*. 2006 Jul 9;13(13):1021–6.
292. Naganuma T, Takagi S, Kanetake T, Kitamura T, Hattori S, Miyakawa T, et al. Disruption of the Sjögren-Larsson Syndrome Gene *Aldh3a2* in Mice Increases Keratinocyte Growth and Retards Skin Barrier Recovery. *J Biol Chem*. 2016 May 27;291(22):11676–88.
293. Stagos D, Chen Y, Cantore M, Jester J V, Vasiliou V. Corneal aldehyde dehydrogenases: multiple functions and novel nuclear localization. *Brain Res Bull*. 2010;81(2–3):211–8.
294. Hama H. Fatty acid 2-Hydroxylation in mammalian sphingolipid biology. *Biochim Biophys Acta*. 2010;1801(4):405–14.
295. Tachibana K, Yamasaki D, Ishimoto K, Doi T. The Role of PPARs in Cancer. *PPAR Res*. 2008;2008:102737.
296. Willemsen MA, IJlst L, Steijlen PM, Rotteveel JJ, de Jong JG, van Domburg PH, et al. Clinical, biochemical and molecular genetic characteristics of 19 patients with the Sjögren-Larsson syndrome. *Brain*. 2001;
297. Fuijkschot J, Theelen T, Seyger MMB, van der Graaf M, de Groot IJM, Wevers RA, et al. Sjögren–Larsson syndrome in clinical practice. *J Inherit Metab Dis*. 2012 Nov 26;35(6):955–62.
298. Nicholson JK, Holmes E, Kinross JM, Darzi AW, Takats Z, Lindon JC. Metabolic phenotyping in clinical and surgical environments. *Nature*. 2012;491(7424):384–92.
299. Xiao N, Cao H, Chen C-H, Kong CS, Ali R, Chan C, et al. A Novel Aldehyde Dehydrogenase-3 Activator (Alda-89) Protects Submandibular Gland Function from Irradiation without Accelerating Tumor Growth. *Clin Cancer Res*. 2013 Aug 15;19(16):4455–64.
300. Reebye V, Saetrom P, Mintz PJ, Huang K-W, Swiderski P, Peng L, et al. Novel RNA oligonucleotide improves liver function and inhibits liver carcinogenesis in vivo. *Hepatology*. 2014 Jan;59(1):216–27.
301. Peters CJ, Hardwick RH, Vowler SL, Fitzgerald RC, Oesophageal Cancer C, Molecular Stratification Study G. Generation and validation of a revised classification for oesophageal and junctional adenocarcinoma. *Br J Surg*. 2009;96(7):724–33.

302. di Pietro M, Boerwinkel DF, Shariff MK, Liu X, Telakis E, Lao-Sirieix P, et al. The combination of autofluorescence endoscopy and molecular biomarkers is a novel diagnostic tool for dysplasia in Barrett's oesophagus. *Gut*. 2015;64(1):49–56.
303. di Pietro M, Chan D, Fitzgerald RC, Wang KK. Screening for Barrett's Esophagus. *Gastroenterology*. 2015;148(5):912–23.
304. Kadri SR, Lao-Sirieix P, O'Donovan M, Debiram I, Das M, Blazeby JM, et al. Acceptability and accuracy of a non-endoscopic screening test for Barrett's oesophagus in primary care: cohort study. *BMJ*. 2010;341:c4372.

## Appendix 1: PRISMA flow chart of OAC metabolomics systematic review



## Appendix 2: Ethical approvals for use of human biomaterial



**Imperial College Healthcare Tissue Bank**  
**Department of Surgery and Cancer,**  
**Hammersmith Hospital,**  
**Du Cane Road,**  
**London W12 0HS**  
**Tel: +44 7711 701382**  
**Email: gerry.thomas@imperial.ac.uk**

**ICHTB HTA licence: 12275**  
**REC Wales approval: 12/WA/0196**

Professor George Hanna  
Academic Surgical Department  
10<sup>th</sup> Floor  
EQM Building  
St. Mary's Hospital  
London  
W2 1NY

9th July 2014

Dear George

**Re: Tissue Bank application number Project R14067; Drivers of volatile aldehyde modulation in oesophageal adenocarcinoma**

I am pleased to confirm that the Tissue Management Committee of the ICHTB has approved your application for access.

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 Gerry Thomas  
DI, Hammersmith HTA Research Licence.



**Imperial College Healthcare Tissue Bank  
Department of Surgery and Cancer,  
Hammersmith Hospital,  
Du Cane Road,  
London W12 0HS  
Tel: +44 7711 701382  
Email: gerry.thomas@imperial.ac.uk**

**ICHTB HTA licence: 12275  
REC Wales approval: 12/WA/0196**

Professor George Hanna  
Academic Surgical Department  
10<sup>th</sup> Floor  
QEOM Building  
St. Mary's Hospital  
London  
W2 1NY

26<sup>th</sup> September 2014

Dear George,

**Re: Tissue Bank application number Project R14097; Drivers of volatile aldehyde modulation in oesophageal adenocarcinoma**

I am pleased to confirm that the Tissue Management Committee of the ICHTB has approved your application for access.

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

Yours sincerely,

Professor Gerry Thomas  
DI, Hammersmith HTA Research Licence.





**Imperial College Healthcare Tissue Bank  
Department of Surgery and Cancer,  
Hammersmith Hospital,  
Du Cane Road,  
London W12 0HS  
Tel: +44 7711 701382  
Email: gerry.thomas@imperial.ac.uk**

**ICHTB HTA licence: 12275  
REC Wales approval: 12/WA/0196**

George Hanna  
Academic Surgical Department  
10<sup>th</sup> Floor  
QEQM Building  
St. Mary's Hospital  
London  
W2 1NY

21 March 2016

Dear George,

**Re: Tissue Bank application number R16018;  
Project title: Aldehyde reprogramming in oesophageal adenocarcinoma**

I am pleased to confirm that the Tissue Management Committee of the ICHTB has approved your application for access.

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 Gerry Thomas  
DI, Hammersmith HTA Research Licence.

11 July 2016

Professor George Hanna  
Imperial College Healthcare NHS Trust  
St Mary's Hospital  
Praed Street  
London  
W2 1NY

Dear Prof Hanna

**RE: Approval of Substantial Amendment 7**

**Project Title: Oesophageal Cancer Clinical and Molecular Stratification Study (OCCAMS)- A prospective Multicentre Trial of a revised clinical and molecular staging system for Oesophageal and Junctional Adenocarcinoma**

**Joint Research Compliance Office Reference number: 13SM1472**

**Ethics reference number: 10/H0305/1**

**Principal Investigator: Prof G Hanna**

Thank you for forwarding the details for the amendment to the original study documentation for the above study. The Trust is happy to approve the amendments with respect to risk assessment, research governance and cost implications.

The documents reviewed and approved were:

Document	Version	Date
Notification of Substantial Amendment form		15 March 2016
REC Approval Letter		12 April 2016
Divisional Approval		23 May 2016
Protocol	9	02 March 2016
Patient Consent Form	8	02 March 2016
PIS Adenocarcinoma	7	02 March 2016
PIS early lesions	7	02 March 2016
Covering Letter		14 March 2016

Before you commence your research, please note that you must be aware of your obligations to comply with the minimum requirements for compliance with the Research Governance indicators 17 (Data Protection); 25 (Health and Safety) and 22 (Financial Probity). Details of the requirements to be met can be found in the on the Trust Research Governance Framework available on [www.dh.gov.uk](http://www.dh.gov.uk).

I wish you well in your research and look forward to seeing regular annual progress reports, funding details as well as information about any publications resulting from the study.

Yours sincerely

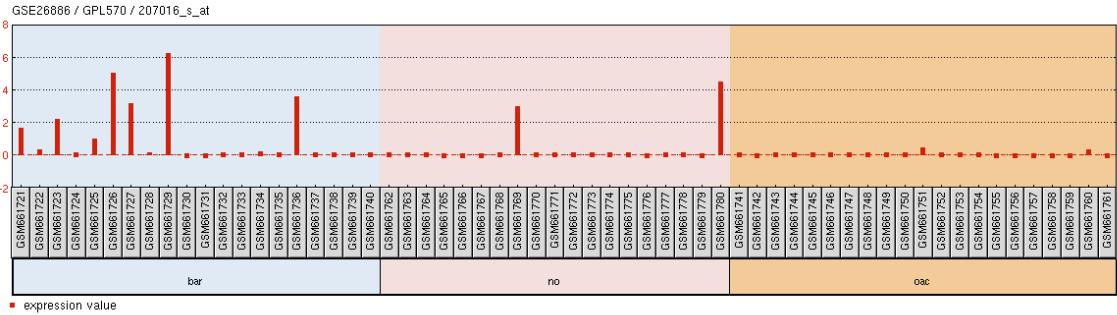


## Appendix 3: Oligonucleotide sequences

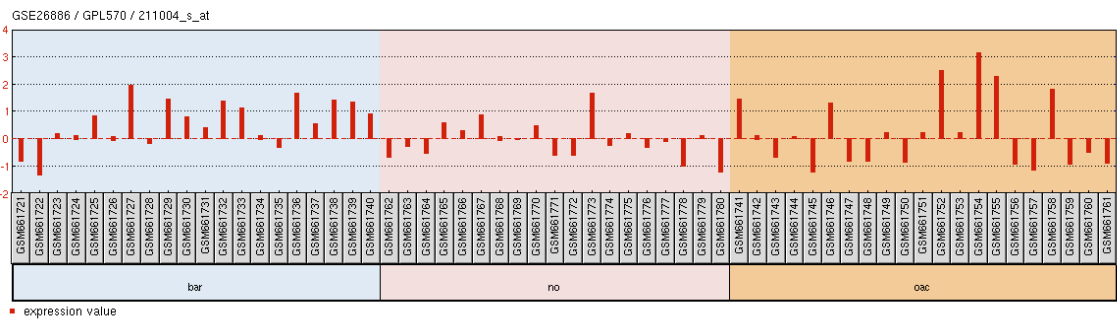
Name	Sequence	Product size
<b>qPCR target genes</b>		
ALDH1A1F	CTGCTGGCGACAATGGAGT	89
ALDH1A1R	CGCAATGTTTTGATGCAGCCT	89
ALDH1A2F	TTGCAGGGCGTCATCAAAAC	121
ALDH1A2R	ACACTCCAATGGGTTCATGTC	121
ALDH1A3F	TGAATGGCACGAATCCAAGAG	100
ALDH1A3R	CACGTCGGGCTTATCTCCT	100
ALDH2F	ATGGCAAGCCCTATGTCATCT	94
ALDH2R	CCGTGGTACTTATCAGCCCA	94
ALDH3A1F	CTCTGTGACCCCTCGATCCA	77
ALDH3A1R	GCATCTTCCCGTAGAACTCTT	77
ALDH3A2F	CTTGGAATTACCCCTTCGTCTC	145
ALDH3A2R	TCCTGGTCTAAATACTGAGGGAG	145
ALDH4A1F	CCATCTCGCCCTTAACTTAC	89
ALDH4A1R	ACTGGGCTTCCATAGGACCA	89
ALDH9A1F	TGGAGTCAAAAATCTGGCATGG	87
ALDH9A1R	AGTAGCAATTTTCATCTCCCGT	87
<b>qPCR internal control genes</b>		
GAPDHF	GGCCAAGGTCATCCATGACAA	101
GAPDHR	AGGGGCCATCCACAGTCTTCT	101
HPRT1F	CCTGGCGTCGTGATTAGTGAT	104
HPRT1R	AGACGTTCAAGTCCGTCCATAA	104
ACTBF	CTGGAACGGTGAAGGTGACA	140
ACTBR	AAGGGACTTCCTGTAACAATGCA	140
B2MF	TGCTGTCTCCATGTTTGATGTATCT	86
B2MR	TCTCTGCTCCCCACCTCTAAGT	86
HMBSF	GGCAATGCGGCTGCAA	155
HMBSR	GGGTACCCACGCGAATCAC	155
SDHAF	TGGGAACAAGAGGGCATCTG	86
SDHAR	CCACCACTGCATCAAATTCATG	86
UBCF	ATTGGGTCGCGGTTCTTG	133
UBCR	TGCCTTGACATTCTCGATGGT	133
YWHAZF	ACTTTTGGTACATTGTGGCTTCAA	94
YWHAZR	CCGCCAGGACAAACCAGTAT	94
<b>Lentivirus cloning</b>		
CLA1-IRESGFP-F	GTACATGCATATGGTGAGCAAGGGCGAGGAG	652
IRESGFP-SAL1-R	GTACCAGCTGTCACTTGTACAGCTCGT	652
ECOR1-3A2-F	GTACGAATTCATGGAGCTCGAAGTCCG	1468
3A2-CLA1-R	CGAGATCGATTTATGAGTCTTCTGAGAGATTTTAC	1468
<b>CRISPR cloning</b>		
g3A2_Ex1_#9_sen	ACGGTGGAGACACCCCCGGGTTTT	n/a
g3A2_Ex1_#9_comp	CCGGGGGGTGTCTCCACCGTCGGTG	n/a
g3A2_Ex1_#4_anti	GGAACGCCTGTCCGACCCCGTTTT	n/a
g3A2_Ex1_#4_comp	GCGGGTCCGACAGGCGTCCCGGTG	n/a
g3a2_Ex1_checkF	GACTGGCAGTGGGACTCAG	251
g3a2_Ex1_checkR	GCAGCCGAAACCGCAGAG	251

## Appendix 4: Gene expression of ALDH genes not quantified by qPCR

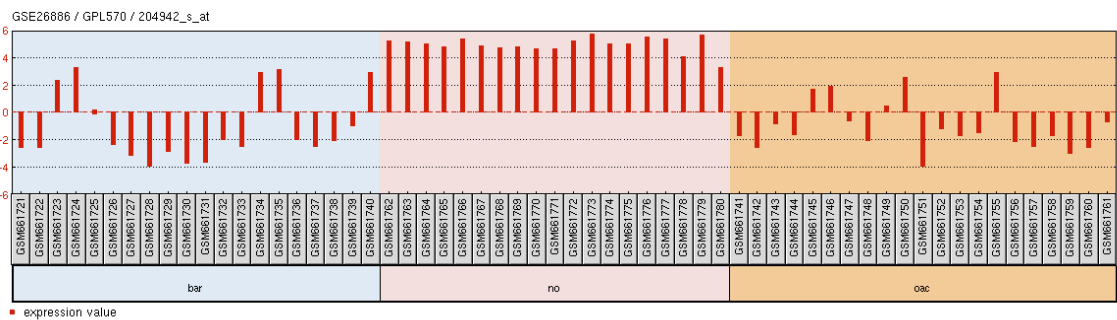
### ALDH1A2



### ALDH3B1

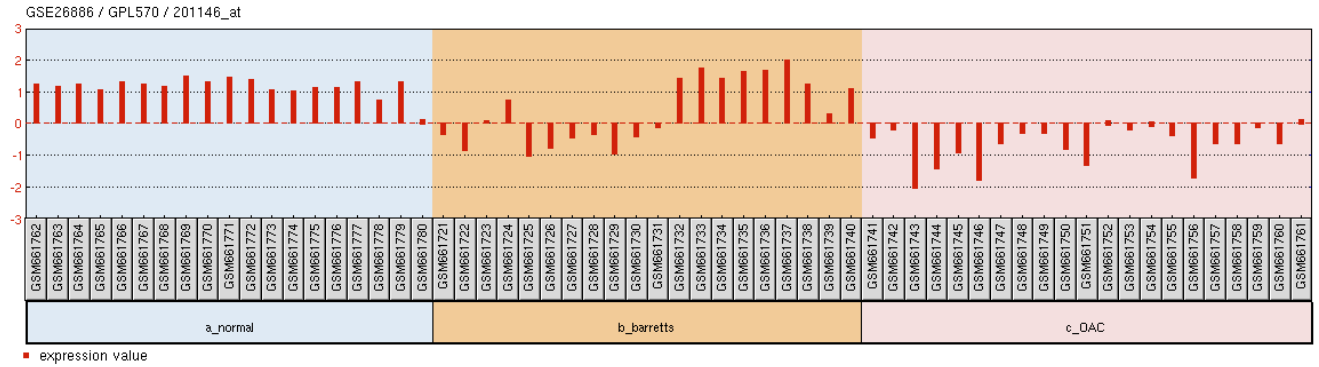


### ALDH3B2



## Appendix 5: Related gene expression (Wang et al)

### NFE2L2 (nrf2)



### KEAP1

