

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

University of Southampton

Faculty of Medicine

School of Cancer Sciences

**Investigating the tumour immune microenvironment as a determinant of antigen
presentation and immune response in oesophageal cancer**

by

William Boyce Pratt BSc. Hons, MRes

ORCID ID <https://orcid.org/0000-0002-3214-6552>

Thesis for the degree of Doctor of Philosophy February 2024

Abstract

University of Southampton

Faculty of Medicine, School of Cancer Sciences

Doctor of Philosophy

Investigating the Tumour Immune Microenvironment as a Determinant of Antigen

Presentation and Immune Response in Oesophageal Cancer

by William Boyce Pratt

Oesophageal cancer (OC), the 12th most common cancer (UK), presents with high clinical need and poor survival. OC is histologically divided into adenocarcinoma (OAC) and squamous cell carcinoma. Antigen presentation machinery (APM) is crucial for eliciting anti-tumoral immune responses by cancer-antigen recognition. However, the genomic landscape of APM genes and dysregulation of expression in OAC is unknown.

A bioinformatic approach has been employed to elucidate the landscape of genomic defects in APM genes using large multi-omics datasets (TCGA and OCCAMS). Digital cytometry methods were applied to determine the impact of APM expression on immune composition within the tumour immune microenvironment (TiME). Additionally, immunohistochemistry (IHC) of a large tissue microarray series for candidate APM genes and immune cell markers (HLA-A/B/C/E/Class 2, CD3, CD4, CD8, Foxp3) was performed with digital histopathology analysis to validate findings. Lastly, cell modelling and single-cell RNA sequencing were implemented to determine the role of CSDE1 in regulating the expression of MHC class I genes in OAC cell lines.

Genomic landscape analysis found APM mutation incidence to be infrequent and copy-number segments complex over the HLA locus of chromosome 6; these results did not inform survival analysis. However, the expression of APM genes showed a significant impact on overall survival (OS), with the expression of 12 out of 18 MHC class I, 12/20 MHC class II and 4/7 APM regulators were associated to altered OS. Following this analysis, the impact of APM gene expression on the TiME identified 10/20 MHC class I, 7/20 MHC class II and 3/7 APM gene expression regulators associated with altered immune composition. Using these approaches together, identified 17/45 assessed APM genes possessed both an association with survival outcomes and altered immune composition, including *CSDE1*, *HLA-E*, *ERAP2*, *CD74*, *HLA-DRB1* and *HLA-DRB5*, among others.

IHC demonstrated that the low expression of HLA-E and HLA class II and the high expression of TAP1 and CSDE1 corresponded to shorter OS. T cell density alone did not associate with OS. Yet the high expression of HLA-A/B/C combined with the low abundance of CD8+ T cells resulted in shorter OS compared to high HLA-A/B/C protein expression and greater CD8+ T cell density. CD3/8+ T-cell density correlated with greater HLA-ABC and HLA class 2 protein expression and a correspondingly increased T-cell infiltrate.

Knockdown of *CSDE1* expression produced greater JAK/STAT1 signalling and increased expression of HLA-A/B/C transcripts, suggesting a significant negative regulator role in the expression of MHC class I HLAs. *CSDE1* was upregulated in single cancer cells compared to healthy comparative cells, with evidence suggesting overexpression may be acquired in premalignancy over increasing dysplasia (i.e., Barrett's oesophagus).

This work has translational value for OAC patients. Firstly, APM expression presents a potential marker of altered survival and immune composition, which has biomarker utility for immunotherapy implementation. Secondly, *CSDE1* may form an actionable pathway to improve immune responses as a potential future target for vaccine therapies.

Table of Contents

ABSTRACT	I	
TABLE OF CONTENTS	I	
TABLE OF TABLES	V	
TABLE OF FIGURES	VII	
RESEARCH THESIS: DECLARATION OF AUTHORSHIP	IX	
ACKNOWLEDGEMENTS	XI	
DEFINITIONS AND ABBREVIATIONS	XIII	
CHAPTER 1	GENERAL INTRODUCTION	15
1.1	STUDY RATIONALE.....	15
1.2	INCIDENCE OF OESOPHAGEAL CANCER	17
1.3	RISK FACTORS OF OESOPHAGEAL CANCER	17
1.3.1	Sex and age.....	17
1.3.2	Gastro-oesophageal reflux disease (GORD) and Barrett's oesophagus	18
1.3.3	Smoking and alcohol consumption.....	20
1.3.4	Obesity.....	20
1.3.5	Medication.....	20
1.3.6	Genomic associations	21
1.4	MEDICAL INTERVENTION	22
1.4.1	Diagnosis.....	22
1.4.2	Staging and grading	22
1.4.3	Treatment – the current standard of care in the UK	24
1.5	HISTOLOGY.....	26
1.6	OAC EVOLUTION, DISEASE PROGRESSION AND METASTASIS.....	27
1.7	IMMUNOTHERAPY TRIALS	28
1.8	THE TUMOUR MICROENVIRONMENT	31
1.9	ANTIGEN PRESENTATION MACHINERY (APM) AND PATHWAYS	34
1.9.1	The MHC I Pathway	36
1.9.2	The MHC II Pathway	38
1.9.3	Cross-presentation Pathways.....	40
1.9.4	Regulation of APM gene expression.....	42
1.10	CANDIDATE APM GENES.....	44
1.11	THE GENOMICS OF OAC & OSCC.....	49
1.12	LITERATURE REVIEW OF MOLECULAR PATHWAY DEFECTS OF APM IN OAC AND ESCC	52
1.13	RESEARCH HYPOTHESES, AIMS AND OBJECTIVES.	55
CHAPTER 2	GENERAL MATERIALS AND METHODS	58
2.1	DOWNLOAD AND FILTER TCGA DATA.	58
2.1.1	Clinical.....	58
2.1.2	Mutation	58
2.1.3	Copy-number	59
2.1.4	RNA Expression.....	59
2.2	DOWNLOAD AND FILTER ICGC DATA.	60
2.3	DOWNLOAD AND FILTERING OCCAMS DATA.....	60
CHAPTER 3	THE LANDSCAPE OF GENOMIC DEFECTS OF ANTIGEN PRESENTATION MACHINERY GENES	
IN OAC	63	
3.1	INTRODUCTION	63
3.1.1	Antigen presentation and known pathway defects in OAC.....	63
3.1.2	MHC I endogenous pathway	64
3.1.3	MHC II Pathway	66
3.1.4	Alternative processing of cancer antigens.....	67
3.1.5	Hypothesis and research objectives	68
3.2	METHOD	69

Table of Contents

3.2.1	Datasets	69
3.2.2	Power and sample size	70
3.2.3	Mutation/copy number analysis.....	70
3.2.4	Differential expression analysis	70
3.2.5	Survival analysis (mRNA)	70
3.2.6	Methylation analysis.....	71
3.2.7	Statistical Analysis.....	71
3.3	RESULTS.....	72
3.3.1	Description of the study cohorts.....	72
3.3.2	APM genomic aberrations	75
3.3.3	APM dysregulated mRNA expression.	84
3.3.4	APM genes associated with clinical outcomes and disease progression.....	87
3.3.5	APM regulators and survival associations	88
3.3.6	Correlation analysis identifies correspondent expression between APM genes and APM gene expression regulators.....	89
3.3.7	MHC class I APM gene candidates and Survival Associations.....	91
3.3.8	MHC class II APM gene Candidates and survival associations.....	93
3.3.9	APM epi-genomic aberrations.....	101
3.4	DISCUSSION.....	105
3.4.1	Mutation incidence of APM genes in OAC is infrequent.	105
3.4.2	Copy number events are not informative in OAC but are frequent among APM genes of the HLA-locus.....	106
3.4.3	APM gene candidates and differential expression in OAC.	106
3.4.4	Expression of APM gene candidates associates with overall, cancer-specific, and disease-free survival in OAC.	107
3.4.5	Transcription regulation by genome methylation alters the APM landscape of mRNA expression via regulation of known APM gene expression regulators.	117
3.4.6	Limitations	119
3.4.7	Conclusions.....	120
CHAPTER 4	INVESTIGATING THE RELATIONSHIP BETWEEN IMMUNE COMPOSITION ESTIMATES AND PROGNOSTIC APM GENE DEFECTS.	121
4.1	INTRODUCTION	121
4.1.1	Deconvolution methods	121
4.1.2	Limitations of deconvolution methodology.....	123
4.1.3	Validation of deconvolution output	124
4.1.4	The Tumour Immune Microenvironment (TIME) of OAC	124
4.1.5	B lymphocytes.....	125
4.1.6	CD8 T cells.....	126
4.1.7	CD4 T cells.....	127
4.1.8	T cells regulatory (Tregs).....	128
4.1.9	Natural killer cells (NK)	129
4.1.10	Monocytes	129
4.1.11	Macrophages	130
4.1.12	Dendritic cells (DCs)	130
4.1.13	Mast Cells.....	131
4.1.14	Eosinophils.....	131
4.1.15	Neutrophils	132
4.2	HYPOTHESIS AND RESEARCH OBJECTIVES.....	132
4.3	METHODOLOGY	134
4.3.1	Samples and data.....	134
4.3.2	CIBERSORTx	134
4.3.3	Statistical analysis of deconvolution data.....	135
4.4	RESULTS.....	137
4.4.1	Validation of CIBERSORTx analysis	137
4.4.2	A description of the TIME in OAC	138
4.4.3	Correlation of prognostically significant APM genes and immune subpopulations in OAC	143
4.4.4	The impact of APM gene expression on immune effector/Treg cell populations in the OAC TIME.	149

4.4.5	The impact of APM gene expression on myeloid cell populations in the OAC TIME.....	153
4.4.6	The impact of APM gene expression on B lineage cell populations in the OAC TIME.....	156
4.5	DISCUSSION.....	160
4.5.1	The TIME of OAC is diverse and is broadly defined into four phenotypes.	160
4.5.2	Effector cell populations are increased in cases with high expression of MHC class I in OAC	162
4.5.3	Effector cell populations are greater in cases with characteristically high expression of MHC class II in OAC.....	164
4.5.4	Myeloid cell populations are altered in cases because of MHC class I expression in OAC.	166
4.5.5	Myeloid cell populations are altered in cases with high MHC class II expression in OAC. 168	
4.5.6	B lineage cell populations in the OAC TIME are altered by the expression of CALR and CD1D in OAC	170
4.5.7	B lineage cell populations in the OAC TIME are increased by the expression of MHC class II genes	170
4.5.8	Effector cell populations are altered in cases because of APM gene expression regulators in OAC.	171
4.5.9	Limitations	172
4.5.10	Conclusions.....	173
CHAPTER 5	CLINICAL AND IMMUNOHISTOCHEMICAL VALIDATION OF PROGNOSTIC APM GENES.	174
5.1	INTRODUCTION	174
5.1.1	Immunohistochemistry for scientific and clinical applications.....	174
5.1.2	Immunohistochemistry investigation of OAC immunity and antigen presentation machinery	175
5.1.3	Limitation of immunohistochemistry	178
5.1.4	Hypothesis and chapter aim and objectives.....	179
5.2	METHODOLOGY	180
5.2.1	TMA samples and clinical data	180
5.2.2	Staining	180
5.2.3	Imaging	181
5.2.4	Automated analysis	181
5.2.5	Statistical analysis.....	183
5.2.6	Power and sample size calculations	183
5.3	RESULTS	184
5.3.1	TMA quality control and stain scoring optimisation.....	184
5.3.2	Exploratory data analysis reveals high heterogeneity in APM protein expression in OAC.	189
5.3.3	Exploratory data analysis reveals high heterogeneity in immune cell densities in OAC. ..191	
5.3.4	Correlation analysis identifies relationship between APM protein expression and immune composition in OAC.	193
5.3.5	Quantile APM protein expression analysis demonstrates altered immune composition in OAC.	195
5.3.6	Validation of deconvolution by immunohistochemistry	201
5.3.7	Univariate survival analysis demonstrates significant impact of APM protein expression and immune composition on patient outcomes.	204
5.3.8	Multivariate survival analysis identifies independent prognostic APM protein expression and immune composition in OAC.	205
5.4	DISCUSSION.....	209
5.4.1	APM gene expression in the OAC TMAs presents with high heterogeneity and clusters in five distinct groups.	209
5.4.2	T cell density reflects known immunophenotypes from deconvolution analysis.	211
5.4.3	APM protein expression correlates with T cell density in OAC.	211
5.4.4	APM protein expression associates to altered survival outcomes in OAC.	213
5.4.5	T cell density associates to altered univariate survival outcomes in OAC.....	213
5.4.6	Limitations and future work	214
5.4.7	Conclusion	214
CHAPTER 6	CELLULAR MODELLING OF MHC CLASS I EXPRESSION REGULATION BY CSDE1 IN OAC.	216
6.1	INTRODUCTION	216

Table of Contents

6.1.1	Known immune evasion strategies used by cancer and current knowledge in OAC.....	216
6.1.2	The regulatory signalling pathways of MHC class I expression.	218
6.1.3	The novel role of CSDE1 in antigen presentation machinery gene expression regulation	221
6.1.4	Biological functions of CSDE1	223
6.1.5	CSDE1 As a target for therapy in cancer	224
6.1.6	Manipulating gene expression in vitro experiments	225
6.1.7	Cell models of OAC	228
6.1.8	Hypothesis and chapter aim, and objectives.....	230
6.2	METHODOLOGY	231
6.2.1	Cell expression profiling.....	231
6.2.2	General principles of cell culture	231
6.2.3	CSDE1 knockdown by siRNA transfection.....	233
6.2.4	CSDE1 overexpression by nucleofection	233
6.2.5	RNA isolation and purification	233
6.2.6	Nucleic acid quantification	234
6.2.7	Quantitative Real-Time PCR (qRT-PCR) by TaqMan.....	235
6.2.8	Western blotting.....	237
6.2.9	Single-cell RNA-seq analysis.....	240
6.2.10	Statistical analyses	240
6.3	RESULTS.....	241
6.3.1	Cell model selection and characteristics	241
6.3.2	Validation of CSDE1 knockdown in OAC cell model	243
6.3.3	APM gene expression in OAC cancer cells is downregulated by the expression of CSDE1.	244
6.3.4	APM protein expression in OAC cancer cells may be downregulated by the expression of CSDE1 as a modulator of the STAT1/JAK pathway.	246
6.3.5	Attempted CSDE1 overexpression in OE33 cells.	249
6.3.6	The expression of CSDE1 is high in single cancer cells from human OAC tumours.	250
6.4	DISCUSSION.....	263
6.4.1	CSDE1 expression is a key regulator of MHC class I expression by modulating STAT1 signalling in FLO-1 OAC cells.	263
6.4.2	Localisation of CSDE1 mRNA expression in patient tumour samples	264
6.4.3	Limitations/future work.....	266
6.4.4	Conclusion	267
CHAPTER 7	GENERAL DISCUSSION	268
7.1	GENOMIC DISRUPTION OF APM IN OAC AND ITS IMPACT ON IMMUNE RESPONSE AND CLINICAL PROGNOSIS.	271
7.2	POTENTIAL MECHANISM OF DYSREGULATED APM/ MHC IN OAC – THERAPEUTIC AXIS?.....	277
7.3	CONCLUSION	281
7.4	FUTURE DIRECTIONS	281
APPENDIX A	APM H SCORES AND IMMUNE COMPOSITION HEATMAP	285
APPENDIX B	WESTERN BLOTS OF CSDE1 KNOCKDOWN EXPERIMENT.....	286
APPENDIX C	PERCENTAGE DIFFERENCE QPCR OF CSDE1 KNOCKDOWN EXPERIMENT.....	289
APPENDIX D	CSDE1 EXPRESSION IN OAC CANCER CELLS COMPARED TO HEALTHY COMPARATIVE CELLS	290
GLOSSARY OF TERMS.....		291
LIST OF REFERENCES		293

Table of Tables

TABLE 1 TABLE OF TMN STAGE CATEGORIES WITH DESCRIPTIONS SPECIFIC TO OAC. ADAPTED FROM RICE ET AL., 2017 (PERMISSION OBTAINED) 48.....	24
TABLE 2 HISTOLOGICAL GRADES FOR TUMOUR SECTIONS (52).....	24
TABLE 3 PREOPERATIVE CHEMOTHERAPY DRUGS USED IN THE UK FOR THE TREATMENT OF OAC AND ESCC.	25
TABLE 4 KEY APM GENES IN THE MHC I PATHWAY WITH THEIR ASSOCIATED APM PROCESSES AND PUBLISHED ROLES IN CANCERS.....	44
TABLE 5 KEY APM GENES IN THE MHC II PATHWAY WITH THEIR ASSOCIATED APM PROCESSES AND PUBLISHED ROLES IN CANCERS.....	47
TABLE 6 COHORT CHARACTERISTICS ACROSS TCGA, OCCAMS AND ICGC, PLUS COMBINATIONS OF TCGA/OCCAMS AND TCGA/ICGC.	74
TABLE 7 B2M MUTATION DESCRIPTIONS IN OAC CASES WITHIN THE TCGA-ESCA DATASET, LABELLED TO FIGURE 15.	78
TABLE 8 HLA-B MUTATION DESCRIPTIONS IN ORDERED BY CANCER TYPE WITHIN THE TCGA-ESCA DATASET, LABELLED TO FIGURE 16, ORDERED BY DISEASE TYPE.....	80
TABLE 9 A TABLE OF DIFFERENTIAL EXPRESSION RESULTS FOR OUR APM CANDIDATE GENES WITH LOG FOLD CHANGE OF GENE EXPRESSION COMPARING OAC TO NORMAL STOMACH TISSUE.....	85
TABLE 10 MHC I CANDIDATE GENE EXPRESSION ASSOCIATION WITH SURVIVAL OUTCOMES AMONG THE OAC COHORTS (TCGA AND OCCAMS). SIGNIFICANCE RESPECTIVE TO OUTCOME.	93
TABLE 11 MHC II CANDIDATE GENE EXPRESSION ASSOCIATION WITH SURVIVAL OUTCOMES AMONG THE OAC COHORTS (TCGA AND OCCAMS). SIGNIFICANCE RESPECTIVE TO OUTCOME.	96
TABLE 12 CoxPH OVERALL SURVIVAL ANALYSIS IN UNIVARIATE AND MULTIVARIATE MODEL.....	98
TABLE 13 CoxPH CANCER SPECIFIC SURVIVAL ANALYSIS IN UNIVARIATE AND MULTIVARIATE MODEL.....	99
TABLE 14 CoxPH DISEASE-FREE SURVIVAL ANALYSIS IN UNIVARIATE AND MULTIVARIATE MODEL.	100
TABLE 15 SIGNIFICANT DIFFERENTIALLY METHYLATED CPGs OF THE HLA-LOCUS ON CHROMOSOME 6.....	101
TABLE 16 METHYLATION-EXPRESSION (mRNA:TMM) CORRELATIONS FROM APM CANDIDATES WITH DIFFERENTIAL METHYLATION IN OAC TUMOUR TISSUE AND NORMAL OESOPHAGUS.	104
TABLE 17 SUMMARY OF CELL DENSITIES IN OAC FROM STEIN ET AL, 2017 (340).....	124
TABLE 18 CoxPH SURVIVAL ANALYSIS COMPARING THE HEATMAP CLUSTERS OF IMMUNE PHENOTYPES DERIVED FROM FIGURE 27.	142
TABLE 19 SUMMARY OF SIGNIFICANT IMMUNE CELL CONTENT DIFFERENCE BETWEEN UPPER AND LOWER APM GENE EXPRESSION QUANTILES.....	159
TABLE 20: TABLE OF ANTIBODIES USED IN IHC STAINING.	181
TABLE 21 CoxPH MULTIVARIATE OS FOR TMA IHC APM PROTEIN EXPRESSION (H SCORE).	205
TABLE 22 CoxPH MULTIVARIATE OS FOR TMA IHC APM PROTEIN EXPRESSION (PERCENTAGE POSITIVITY).....	206
TABLE 23 CoxPH MULTIVARIATE OS FOR TMA IHC T CELL DENSITY.	207
TABLE 24 CoxPH MULTIVARIATE OS FOR TMA IHC HLA-ABC H SCORE:CD8 T CELL DENSITY.	208
TABLE 25 COMMONLY USED OAC CELL LINES IN LITERATURE.....	229
TABLE 26 ANTIBODIES USED IN IMMUNOBLOTTING OF TARGET PROTEINS IN WESTERN BLOT ANALYSIS.....	239

Table of Equations

EQUATION 1: H SCORING EQUATION FOR SCORING DIFFUSE IHC STAINS.....	182
EQUATION 2: STANDARD CURVE AND mRNA QUANTIFICATION EQUATIONS FOR TAQMAN ANALYSIS.	236
EQUATION 3: PROTEIN QUANTIFICATION EQUATION FOR BCA ASSAY.	238

Table of Figures

FIGURE 1 DIAGRAM REPRESENTING THE ANATOMICAL LOCATION OF GASTROESOPHAGEAL CARCINOMA SUBTYPES FROM THE PROXIMAL OESOPHAGUS TO THE DISTAL STOMACH.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 2 DEPICTION OF GORD-OAC CANCER PATHOGENESIS.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 3 TMN STAGING 8TH EDITION. REPRESENTATION OF DIFFERENT TNM STAGES OF OAC IN RELATION TO HUMAN ANATOMY.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 4 HISTOLOGY OF THE HUMAN OESOPHAGUS.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 5: MECHANISM OF ACTION OF IMMUNE CHECKPOINT INHIBITORS.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 6: CYTOKINES FORMS AN ESSENTIAL INTRACELLULAR COMMUNICATION NETWORK WHICH MAY ACT TO SUPPRESS ANTI-TUMOUR IMMUNE RESPONSE VIA SEVERAL DIVERSE MECHANISMS.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 7: MHC PATHWAYS IN HUMANS.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 8 MHC I PATHWAY OVERVIEW.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 9 MHC II PATHWAY OVERVIEW.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 10 CROSS-PRESENTATION CELLULAR MECHANISMS.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 11 THE DRIVER GENE LANDSCAPE OF OAC.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 12 A BAR CHART REPRESENTING LITERATURE SEARCH.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 13: MUTATION/COPY NUMBER ONCOLOTS FOR OAC IN THE COMBINED TCGA-ESCA/ICGC-ESAD COHORT.. ERROR! BOOKMARK NOT DEFINED.
FIGURE 14 MUTATION BURDEN BETWEEN SAMPLES POSSESSING A MUTATION IN APM GENES AND THOSE WITHOUT	ERROR! BOOKMARK NOT DEFINED.
FIGURE 15 LOLLIPOP PLOT OF B2M MUTATIONS IN OAC CASES WITHIN THE TCGA-ESCA DATASET; MUTATION DESCRIPTIONS FOUND IN TABLE 3.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 16 LOLLIPOP PLOT OF HLA-B MUTATIONS IN OAC CASES WITHIN THE TCGA-ESCA DATASET; MUTATION DESCRIPTIONS FOUND IN TABLE 3.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 17 MUTATION ONCOLOT FOR DRIVER GENES IDENTIFIED IN THE 551 FRANKELL ET AL PAPER (2019).	ERROR! BOOKMARK NOT DEFINED.
FIGURE 18: REPRESENTATIVE COPY NUMBER SEGMENTS IN THE HLA-LOCUS ON CHROMOSOME 6 (6P21). ..	ERROR! BOOKMARK NOT DEFINED.
FIGURE 19 VOLCANO PLOT OF DIFFERENTIAL EXPRESSION ANALYSIS (DEA). COMPARING APM GENE EXPRESSION BETWEEN TCGA/OCCAMS TUMOUR SAMPLES AND GTEx NORMAL STOMACH	ERROR! BOOKMARK NOT DEFINED.
FIGURE 20 VOLCANO PLOT OF DIFFERENTIAL EXPRESSION ANALYSIS (DEA).....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 21 CORRELATION HEATMAP OF MHC CLASS I EXPRESSION WITH APM REGULATORS.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 22 CORRELATION HEATMAP OF MHC CLASS II EXPRESSION WITH APM REGULATORS.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 23 UNIVARIATE COXPH FOREST PLOT OF APM GENE EXPRESSION IN THE TCGA/OCCAMS COHORT USING MAXIMALLY RANKED STATISTICS FOR OPTIMAL CUT POINTS BETWEEN LOW/HIGH EXPRESSION GROUPS.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 24 METHYLATION OF APM CANDIDATE GENES COMPARED BETWEEN NORMAL OESOPHAGUS AND OAC SAMPLES USING BETA-METHYLATION VALUES.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 25 SCATTER PLOT OF SIGNIFICANT METHYLATION TO mRNA EXPRESSION (TMM) CORRELATIONS OF APM CANDIDATES WHICH POSSESSED SIGNIFICANTLY DIFFERENT METHYLATION IN OAC COMPARED TO NORMAL TISSUE.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 26: SCATTER PLOTS WITH LINEAR REGRESSION COMPARING METHYLICBERSORT WITH VARIATIONS OF CIBERSORTX ANALYSIS.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 27 Z-SCORED HEATMAP OF CIBERSORT SIMPLIFIED DATA IN THE ENTIRE OAC COHORT	ERROR! BOOKMARK NOT DEFINED.
FIGURE 28 STACKED BAR CHART OF CIBERSORT FRACTIONAL VALUES BY HEATMAP CLUSTER DISPLAYED IN FIGURE 27. .	ERROR! BOOKMARK NOT DEFINED.
FIGURE 29 BOX AND WHISKER PLOTS OF IMMUNE CHARACTERISTICS COMPARED BETWEEN HEATMAP CLUSTERS OF IMMUNE PHENOTYPES DISPLAYED IN FIGURE 27.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 30 FOREST PLOT OF COXPH SURVIVAL ANALYSIS COMPARING THE HEATMAP CLUSTERS OF IMMUNE PHENOTYPES DERIVED FROM FIGURE 27.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 31 CORRELATION HEATMAP OF APM mRNA EXPRESSION TO CIBERSORT ABSOLUTE SCORES FOR IMMUNE CELL SUBPOPULATIONS.....	ERROR! BOOKMARK NOT DEFINED.
FIGURE 32 CORRELATION HEATMAP OF APM mRNA EXPRESSION TO CIBERSORT ABSOLUTE SCORES FOR IMMUNE CELL SUBPOPULATIONS WITH HEATMAP CLUSTERS IDENTIFIED IN FIGURE 27.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 33 BOXPLOT OF EFFECTOR/TREG CELL CIBERSORT ABSOLUTE SCORES BY QUANTILE EXPRESSION FOR MHC CLASS I AND ALTERNATIVE APM GENES.	ERROR! BOOKMARK NOT DEFINED.

Table of Figures

FIGURE 34 BOXPLOT OF EFFECTOR CELL CIBERSORT ABSOLUTE SCORES BY QUANTILE EXPRESSION FOR MHC CLASS II GENES.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 35 BOXPLOT OF EFFECTOR CELL CIBERSORT ABSOLUTE SCORES BY QUANTILE EXPRESSION FOR APM REGULATOR GENES.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 36 BOXPLOT OF MYELOID CELL CIBERSORT ABSOLUTE SCORES BY QUANTILE EXPRESSION FOR MHC CLASS I GENES.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 37 BOXPLOT OF MYELOID CELL CIBERSORT ABSOLUTE SCORES BY QUANTILE EXPRESSION FOR MHC CLASS II GENES.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 38 BOXPLOT OF B LINEAGE CELLS CIBERSORT ABSOLUTE SCORES BY QUANTILE EXPRESSION FOR MHC CLASS I GENES.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 39 BOXPLOT OF B LINEAGE CELLS CIBERSORT ABSOLUTE SCORES BY QUANTILE EXPRESSION FOR MHC CLASS II GENES.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 40 CONSORT DIAGRAM OF TMA SAMPLE SIZE AVAILABLE POST IHC STAINING	ERROR! BOOKMARK NOT DEFINED.
FIGURE 41 IHC APM STAINING AND H SCORING.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 42 BLAND-ALTMAN PLOT COMPARING PERCENTAGE POSITIVITY TO H SCORES.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 43 IHC IMMUNE CELL STAINING AND DETECTION.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 44 SUPER-PIXEL STRUCTURE PREDICTION.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 45 Z-SCORED HEATMAP OF APM PERCENTAGE POSITIVE STAINING (N = 88)	ERROR! BOOKMARK NOT DEFINED.
FIGURE 46 Z-SCORED HEATMAP OF APM H SCORES (N = 88)	ERROR! BOOKMARK NOT DEFINED.
FIGURE 47 Z-SCORED HEATMAP OF IMMUNE CELL DENSITY (N = 126)	ERROR! BOOKMARK NOT DEFINED.
FIGURE 48 CORRELATION HEATMAP OF APM STAINING SCORES AND IMMUNE DENSITY	ERROR! BOOKMARK NOT DEFINED.
FIGURE 49 BOXPLOTS OF T CELL DENSITY BETWEEN APM PROTEIN EXPRESSION GROUPS (HIGH VS LOW).	...ERROR!	BOOKMARK NOT DEFINED.
FIGURE 50 BOXPLOTS OF T CELL DENSITY BETWEEN QUANTILE APM PROTEIN EXPRESSION GROUPS BY PERCENTAGE POSITIVITY (UPPER VS LOWER).	ERROR! BOOKMARK NOT DEFINED.
FIGURE 51 BOXPLOTS OF T CELL DENSITY BETWEEN QUANTILE APM PROTEIN EXPRESSION GROUPS BY H SCORE (UPPER VS LOWER).	ERROR! BOOKMARK NOT DEFINED.
FIGURE 52 SCATTER PLOT WITH LINEAR REGRESSION COMPARING PREDICTED TUMOUR PERCENTAGE AND CSDE1 H SCORES IN OAC TMAs.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 53 SCATTER PLOT WITH LINEAR REGRESSION COMPARING IHC T CELL DENSITY WITH CIBERSORT ABSOLUTE SCORE.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 54 SCATTER PLOT WITH LINEAR REGRESSION COMPARING TCGA-ESCA OAC LYMPHOCYTE PERCENTAGE WITH TILS CIBERSORT ABSOLUTE SCORE.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 55 FOREST PLOT OF CoxPH UNIVARIATE OS FOR TMA IHC APM PROTEIN EXPRESSION AND T CELL DENSITY.ERROR!	BOOKMARK NOT DEFINED.
FIGURE 56 TYPE I, II AND III INTERFERON SIGNALLING PATHWAYS FOR THEIR RESPECTIVE RECEPTORS...	ERROR!	BOOKMARK NOT DEFINED.
FIGURE 57 THE MHC CLASS I PROMOTOR REGION.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 58: CSDE1 AS A NEGATIVE REGULATOR OF MHC CLASS I EXPRESSION VIA THE JAK/STAT PATHWAY.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 59 ONCOLYTIC TRAP AND AMBUSH STRATEGY.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 60 HEATMAP OF TMM NORMALISED EXPRESSION OF CSDE1, MHC CLASS I GENE EXPRESSION REGULATORS, MHC CLASS I GENES AND HMGA2 FOR AVAILABLE OAC CELL LINES. Z-SCORED, MINKOWSKI DISTANCE, WARD.D2 LINKAGE.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 61 WESTERN BLOT OF CSDE1 AND GAPDH EXPRESSION FOLLOWING CSDE1 KNOCKDOWN IN FLO-1 CELLS.	...ERROR!	BOOKMARK NOT DEFINED.
FIGURE 62 PROTEIN QUANTIFICATION OF CSDE1 NORMALISED TO GAPDH FOLLOWING siRNA KNOCKDOWN IN FLO-1 CELLS.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 63 mRNA QUANTIFICATION OF CSDE1, HLA-A, HLA-B AND HLA-C EXPRESSION FOLLOWING CSDE1 KNOCKDOWN IN FLO-1 CELLS.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 64 WESTERN BLOT OF CSDE1, HLA-ABC, pSTAT1, STAT1 AND GAPDH EXPRESSION FOLLOWING CSDE1 KNOCKDOWN IN FLO-1 CELLS.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 65 PROTEIN QUANTIFICATION OF CSDE1, HLA-ABC, pSTAT1, STAT1 AND GAPDH EXPRESSION FOLLOWING CSDE1 KNOCKDOWN IN FLO-1 CELLS.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 66 CORRELATION ANALYSIS OF PROTEIN EXPRESSION (CSDE1, HLA-ABC AND pSTAT1/STAT1) IN FLO-1 OAC CELLS.	ERROR! BOOKMARK NOT DEFINED.
FIGURE 67 FLUORESCENCE MICROSCOPY OF GFP NUCLEOFECTED CELLS	ERROR! BOOKMARK NOT DEFINED.
FIGURE 68 UMAP OF SINGLE CELL CSDE1 GENE EXPRESSION IN OAC AND GASTRIC TUMOURS.	ERROR! BOOKMARK NOT DEFINED.

- FIGURE 69 UMAP OF SINGLE CELL *CSDE1* AND *HLA-A* GENE CO-EXPRESSION IN OAC AND GASTRIC CANCER TUMOURS.. ERROR! BOOKMARK NOT DEFINED.
- FIGURE 70 UMAP OF SINGLE CELL *CSDE1* AND *HLA-A* GENE CO-EXPRESSION IN OAC CELLS..... ERROR! BOOKMARK NOT DEFINED.
- FIGURE 71 VIOLIN PLOT OF *CSDE1*:*HLA-A* mRNA RATIO IN OAC CANCER CELLS WITH COMPARATIVE NORMAL HEALTHY GASTRIC CARDIA CELLS..... ERROR! BOOKMARK NOT DEFINED.
- FIGURE 72 DOT PLOT OF *CSDE1*, MHC CLASS I, *PTPN2*, AND INTERFERON SIGNALLING RESPONSE GENES FOR HIGH/LOW *CSDE1*: *HLA-A* mRNA RATIO..... ERROR! BOOKMARK NOT DEFINED.
- FIGURE 73 HEATMAP OF SINGLE-CELL RNA-SEQ TPM NORMALISED EXPRESSION OF *CSDE1*, MHC CLASS I GENES AND INTERFERON RESPONSE GENES..... ERROR! BOOKMARK NOT DEFINED.
- FIGURE 74 HEATMAP OF SINGLE CELL RNA-SEQ (PATIENT LEVEL) TPM NORMALISED EXPRESSION OF *CSDE1*, MHC CLASS I GENES AND INTERFERON RESPONSE GENES. R..... ERROR! BOOKMARK NOT DEFINED.
- FIGURE 75 VIOLIN PLOT OF *CSDE1* mRNA EXPRESSION IN PRECANCEROUS LESIONS (BARRETT'S OESOPHAGUS)..... ERROR! BOOKMARK NOT DEFINED.

Research Thesis: Declaration of Authorship

Print name: WILLIAM BOYCE PRATT

Title of thesis: Investigating the Tumour Immune Microenvironment as a Determinant of Antigen Presentation and Immune Response in Oesophageal Cancer

I declare that this thesis and the work presented in it are my own and have been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. Except for such quotations, this thesis is entirely my work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signature: Date:

Acknowledgements

I want to acknowledge my supervisory team for their support and insight into this project (Dr Matthew Rose-Zerilli, Profs Tim Underwood, Tim Elliot, Paul Skipp, Jacek Brodzki), as well as members of the Underwood lab (Dr Robert Walker and Dr Saqib Rahman) for their support in data curation. I would also like to thank Mr Samuel Timms for his assistance in R programming and Dr Benjamin Sharpe, Mr Ben Grace, Miss Anna Broadbent, and Dr Margaret Ashton-Key for their assistance in the TMA experiment. Additionally, I would like to acknowledge Dr Tim Fenton for providing methylation deconvolution data for benchmarking analysis. Lastly, I would like to acknowledge my partner, Jasmine McGregor, for her support throughout this project.

Definitions and Abbreviations

APM.....	Antigen presentation machinery
BE.....	Barrett's oesophagus
BMI.....	Body Mass Index
CAR-T.....	Chimeric antigen receptor therapy
CIN.....	Chromosomal instability
CTLs.....	Cytotoxic T lymphocytes
DACO.....	Data Access Compliance Office
DCs.....	Dendritic Cells
GORD.....	Gastro-Oesophageal reflux disease
HLA.....	Human leukocyte antigen
IHC.....	Immunohistochemistry
LES.....	Lower oesophageal sphincter
MHC.....	Major Histo-compatibility complex
MIIC.....	MHC II compartment
NK.....	Natural killer cells
OAC.....	Oesophageal adenocarcinoma
OC.....	Oesophageal cancer
OSCC.....	Oesophageal squamous cell carcinoma
PLC.....	Protein loading complex.
PPI.....	Proton pump inhibitor
PTM.....	Post-translational modification.
RMSE.....	Root mean square error
scRNA-seq.....	Single-cell RNA-sequencing
TIICs.....	Tumour infiltrating inflammatory cells.
TILs.....	Tumour infiltrating cells.
TIME.....	Tumour immune microenvironment
WES.....	Whole exome sequencing
WGS.....	Whole genome sequencing

Chapter 1 General Introduction

1.1 Study rationale

Oesophageal cancer (OC) is the seventh most common cancer, globally attributed to 572,034 new cases being reported in 2018 and 508,585 deaths. In 2014-2016, 9,100 new cases of OC were reported in the UK, placing OC as the 14th most common cancer type nationally (1, 2). OC is split into two histological subtypes: adenocarcinomas (OAC) and squamous cell carcinomas (ESCC). Although these diseases share the same primary site, they possess several distinct risk factors, incidence trends, histology, and genomic characteristics (3).

OAC refers to cancers originating from mucus-secreting cells in the lower third of the oesophagus. OAC is positively associated with Barrett's oesophagus (a precursor to OAC) and is characterised by chromosomal instability (CIN) (**Figure 1**). Other risk factors of OAC include obesity, smoking, age, use of non-steroidal anti-inflammatory drugs (NSAIDs) and proton pump inhibitors (PPIs), and race. In contrast, OSCC refers to cancer of squamous epithelia in the upper and middle third of the oesophagus. OSCC is characterised with amplification of CCND1, TP63/SOX2 and deletion of KDM6A (**Figure 1**). The significant risk factors of OSCC include smoking and alcohol intake; although smoking is a risk factor for OAC and OSCC, the associated risk in OSCC is far greater (OR of 2.77 vs 5.63 at 60+ years packed smoking; OAC: OSCC) (4).

Antigen presentation is key in eliciting immune responses towards cancer, yet cancers often somatically aberrate the components of this pathway to reduce or prevent immune response towards tumours. The landscape and impact of aberrations in antigen presentation machinery (APM) are well-detailed in other cancers such as breast, lung, colorectal and even OSCC. However, research into the impact of APM aberrations in OAC is severely lacking (5-7). It is important to address this knowledge gap in OAC as it will significantly impact current and future efforts in treating OAC with immunotherapeutics, including immune checkpoint blockade therapy, CAR-T cell therapy and adoptive T cell therapy (8).

To address this, we will focus on datamining publicly available datasets of OAC to discover and landscape APM defects, determining their clinical impact on survival and the immune cell subpopulations in the tumour immune microenvironment (TIME). Furthermore, we shall validate these findings *in situ* using immunohistochemistry (IHC)

methodology, whilst exploring the spatial expression of APM proteins and immune cell spatial distributions. The implementation of single-cell RNA-seq (scRNA-seq) methodology will allow us to examine how APM gene candidates are expressed in different cell types. Through this, we can investigate the role of antigen presentation in editing the immune microenvironment and immune suppression.

Overall, we expect several APM components that have been reported to impact TIME and overall survival to possess a similar impact on our OAC cohorts. We are particularly interested in MHC I machinery, such as HLA-A, which are known instigators of CD8+ cytotoxic T lymphocyte anti-tumoral activity (9).

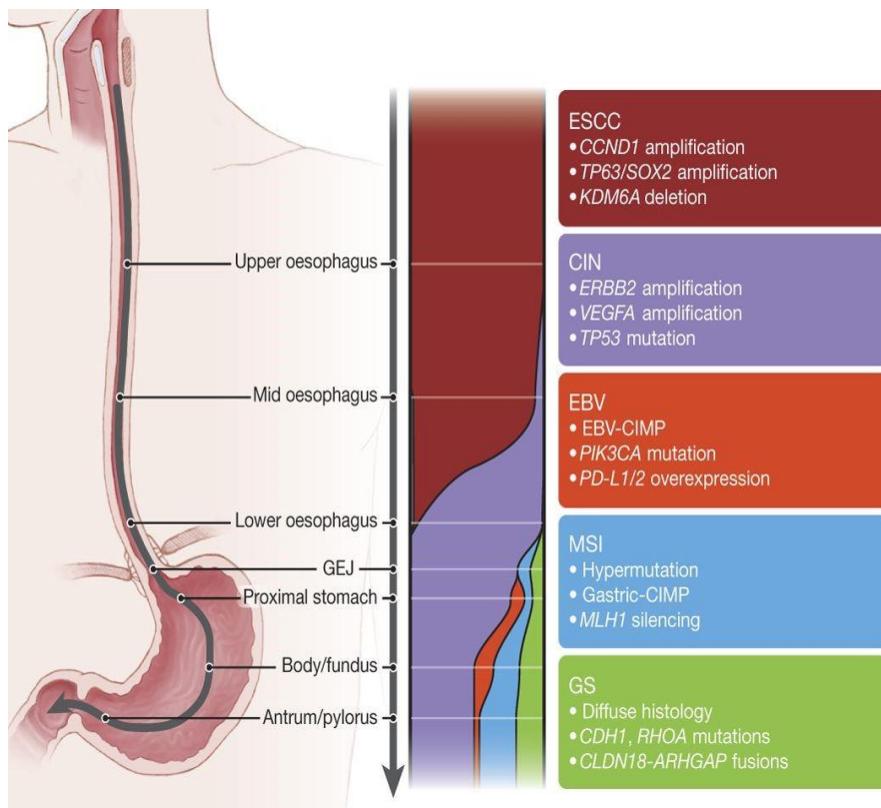


Figure 1 Diagram representing the anatomical location of gastroesophageal carcinoma subtypes from the proximal oesophagus to the distal stomach. Colour bandwidth represents the proportion of each subtype within the anatomic region. Notable subtypes features are stated in each subtype heading (adapted from The Cancer Genome Atlas Research Network) (3).

1.2 Incidence of Oesophageal cancer

OAC occurs predominantly in males, in the US, the male: female incidence ratio (9:1) shows a significant trend towards males (1, 10, 11). Similarly, in the UK, age standardised (AS) rates displays males as the predominant sex in OAC, scoring an AS rate of 22.6 compared to females scoring 8.3 in 2016 (1). Male: female incidence ratios of OAC differ geographically, with the highest ratios found in North America (7.6:1), Oceania (6.2:1), Europe (6:1); lowest ratios located to Asia (4.4:1), Latin America/Caribbean (3.9:1) and Africa (1:1) (12). OAC incidence also differs by geographical location. The highest rates (OAC cases per 100,000) are found in Northern/Western Europe (3.4), North America (3.5) Oceania (3.1) (12)(13). The lowest rates of OAC are located to South/Central America/Caribbean (1.2), North/West Africa (0.7), Sub-Saharan Africa (0.4), Eastern Europe (0.8) and Asia (0.6) (12)(13). Notably, the UK has the highest incidence of OAC globally. This may be due to the obesity epidemic and other risk factors such as GORD and Barrett's oesophagus. The incidence rates are 7.2 for males and 2.5 for females.

1.3 Risk factors of Oesophageal cancer

1.3.1 Sex and age

As noted, prior sex is a prominent risk factor for OAC, occurring primarily in men, with high male to female ratios for OAC are exhibited in Northern America (7.6:1), Oceania (6.2:1), and Europe (6.0:1), with lower ratios in Asia (4.4:1), Latin America and the Caribbean (3.9:1), and Africa (1:1); these ratios have remained steady, but recent research shows a steady increase in sex ratios in the UK and Netherlands (13, 14). The risk of OAC increases with age; in the UK, the highest OAC incidence age group is 80+ years old (84.1 OAC cases per 100,000 and the lowest incidence age group is <49 years old (up to 3.1 OAC cases per 100,000) (15).

Interestingly, all age groups (80+, 65-79, 50-69 and 40-49 years old) display an increase in incidence since 1970; for instance, the incidence of OAC in 80+-year-old men increased from 23.0 to 84.1 per 100,000, most likely relating to the obesity epidemic and GORD incidence (15). Similarly, OSCC incidence increases with age with 24.9 OAC cases per

100,000 in 80+-year-olds (UK); contrasting to OAC, the incidence of OSCC has decreased since 1970 from 42.3 to 25.1 for 80 + year old men (15).

The difference in incidence between men and women in the UK and elsewhere is likely because of compounded risk factor differences between the sexes, UK examples include; obesity where 67% of men and 62% of women were classed as overweight; smoking statistics from 2015 indicate 19.3% of men and 15.3% of women smoked cigarettes, however, the difference between the incidence of smoking has been steadily falling since 1945 (16-18).

1.3.2 Gastro-oesophageal reflux disease (GORD) and Barrett's oesophagus

One of the primary associated risk factors for OAC is Gastro-oesophageal reflux disease. GORD is a globally prevalent disorder characterised by the chronic reflux of stomach acid into the lower oesophagus and is significantly more prevalent in western populations (18 to 28%) compared to eastern populations (2 to 8%), characterised as the most significant risk factor for OAC incidence (19).

Barrett's oesophagus (BE) is a histological and genetic precursor to OAC, referring to the formation of a metaplastic columnar epithelium with a crypt-like structure in the oesophagus (resembling the intestinal epithelia) because of damage of the squamous epithelium. Often, the damage caused to the squamous epithelium is caused by GORD and stomach acid reflux; in response, specialised columnar cells form the oesophageal epithelia and secrete mucins to protect the oesophagus from the refluxed acid (**Figure 2**).

The exact pathogenesis of BE was elusive, with two hypotheses arising. Firstly, in trans-differentiation, a squamous epithelial cell first dedifferentiates into a transitional cell that then differentiates into a columnar cell, this process may occur under a reflux-induced inflammatory microenvironment, upregulating CDX2, pSMAD1, pSMAD5 and/or pSMAD8 expression (20-22). Alternatively, trans-commitment of progenitor cells may explain the pathogenesis of BE; in this mechanism, pluripotent stem cells within oesophageal submucosal glands undergo trans-commitment to form a columnar epithelium (23, 24). A recent study has suggested a two-phase model for the development of BE with the first phase possessing positive selection for homozygous mutation in CDKN2A and TP53. These lineages will spread disproportionately across the segment, while mutations in ARID1A and SMARCA4 may be positively selected for excess of functional mutations, but

do not see large-scale spreading across the segment (25, 26). The second phase shows stabilisation of the lineages in both likely cancer outcome and non-cancer outcome lineages, with the cancer lineages possessing one or more risky lineages. These are typically marked by TP53 mutations which undergo chromosomal instability, copy number variation and structural alterations, and genome doubling with local expansion of the unstable clones attributed to colonization during wound repair in the epithelial layer (25, 26). A striking study in 2021 aimed to molecularly characterize all putative cell origins for BE and explain whether all OAC subtypes originate from BE; within this study it was determined BE originates from gastric cardia via c-MYC and HNF4A-driven transcriptional programs with OAC most likely arising from undifferentiated Barrett's oesophagus cell types (27).

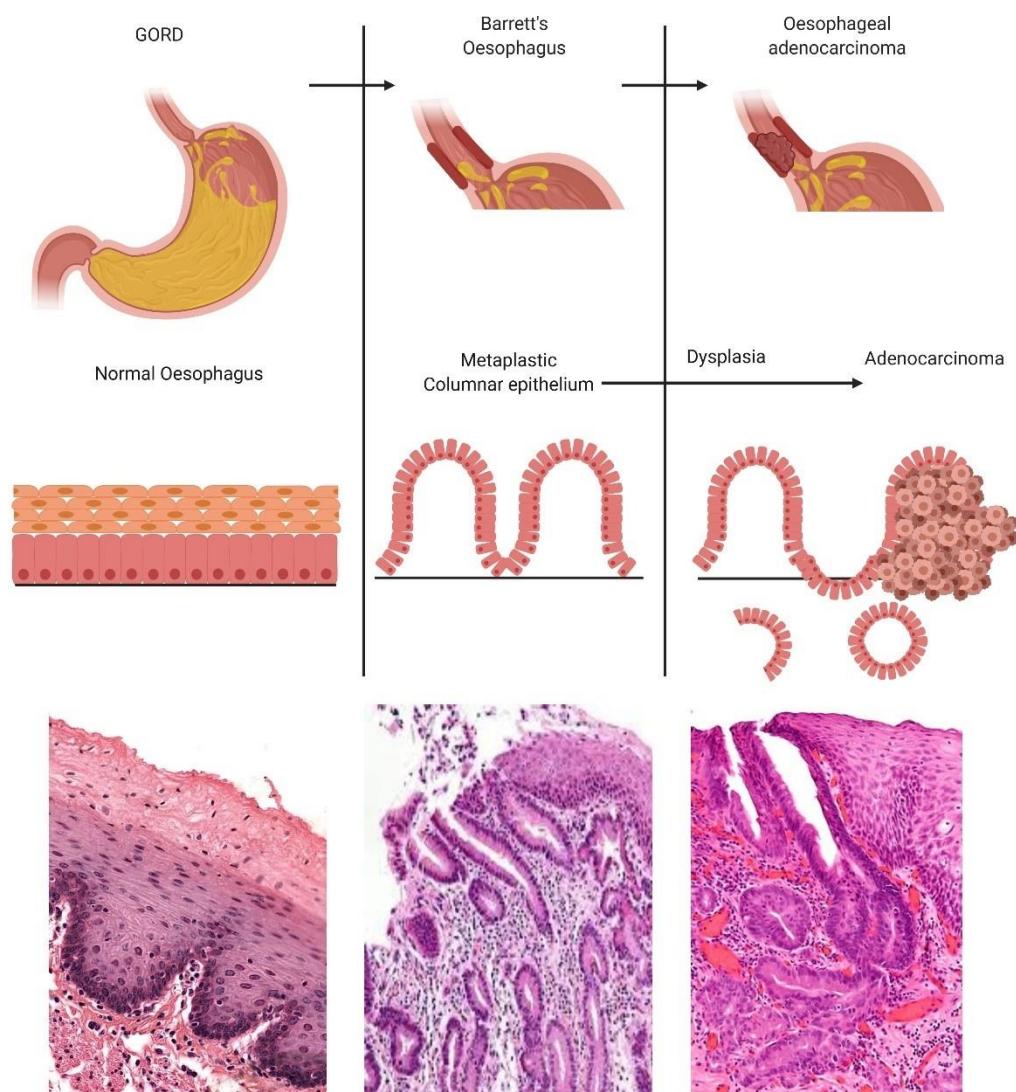


Figure 2 Depiction of GORD-OAC cancer pathogenesis. Left: Gastric acid refluxes into the oesophagus, damaging the normal squamous epithelium; Middle: Adaptive response leads to the formation of a metaplastic columnar epithelium (BE), secreting mucus to

protect the oesophagus; Right: Over a process of dysplasia, columnar epithelial cells gain cancer driver mutations (TP53, CDKN2A, SMAD4, ARID1A and PIK3CA), forming an adenocarcinoma. Created with Biorender.com.

1.3.3 Smoking and alcohol consumption

A strong association between cigarette smoking and OAC (OR of 1.96); the length of smoking cessation is associated with decreased risk (<10 years of smoking cessation: OR = 0.82 and ≥10 years of smoking cessation: OR = 0.71) (28). In a recent study, ESCC also displayed an association with smoking, though to a lesser degree than OAC with an OR of 1.49 for current smokers and 1.04 for ex-smokers ($P = <0.001$) (29).

However, when adjusting this data for alcohol consumption, the difference in risk diminishes (OR of 1.12 for current smokers and 0.83 for ex-smokers; $P = 0.191$) (29). following up this finding alcohol consumption was investigated; OR rates adjusted for smoking highlighted alcohol consumption as a critical risk factor, with current drinkers being at the highest risk (OR of 2.24) and ex-drinkers possessing lesser risk (OR of 1.51) (29).

1.3.4 Obesity

Rising incidence of OAC has been correlated to the obesity epidemic faced by western nations, with studies identifying this trend (30-33). Interestingly, the abdominal diameter has been more explicitly correlated increase in OAC risk than BMI, with studies associating abdominal obesity with OAC risk (31, 32). Obesity may be linked to OAC via the incidence of BE with several studies identifying this link; obesity has showed a significant and independent association to oesophageal inflammation and BE; it is important to note that these studies focus on specifically abdominal obesity rather than BMI as these show differing risk of BE and OAC (34-36).

1.3.5 Medication

Several studies have implicated the use of specific medications in OAC risk. This includes proton pump inhibitors (PPIs) and Lower oesophageal sphincter (LES)-relaxing medications. One study identified around half of the patients with OAC also take PPIs; despite decreasing the acidity of refluxed stomach acid further investigation showed that

the reflux of bile salts alone may contribute Barrett's oesophagus (described later) and OAC transformation (37, 38). Whereas LES-relaxing medication increased the patient-recorded incidence of reflux symptoms and increasing the risk of developing OAC (39). Adjusting the data for reflux symptoms, diminished the positive relationship between LES-relaxing medication and OAC, suggesting the relationship between LES-relaxing medication and OAC is mostly because of the risk of developing gastro-oesophageal reflux symptoms (38, 39).

1.3.6 Genomic associations

Despite the sparse number of publications in OAC on APM defects, a few papers describe the impact of HLA genes on the risk of developing BE and subsequently OAC; the findings of these papers are detailed below.

One study identified loss of HLA-A/B/C expression in BE epithelia compared to normal epithelia by immunostaining (BE: 50% -positive, Normal: 68.3% -positive); whereas HLA-DR staining was substantially higher in BE than normal tissue (BE: 51.6% -positive, Normal: 11.7 % -positive); Expression of HLA-DP/DQ/DR was markedly upregulated in BE compared to normal tissue BE: 73.4% -positive, Normal: 18.3% -positive) (39). Further analysis discovered a total loss of HLA-A/B/C with a gain of HLA-DP/DQ/DR was clear in 37.5% of BE patients but none in the control patients; this MHC I loss/MHC II acquire phenotype was also positively associated with dysplasia (39).

Remarkably, polymorphisms within the MHC locus have been associated with BE risk, predisposition, and risk of OAC. One study identified variants in two loci associated to BE risk. Firstly, chromosome 6p21, rs9257809 within the MHC locus (OR 1.21) and chromosome 16q24, rs9936833 closest to the protein-coding gene is FOXF1 (OR 1.14), implicated in oesophageal development and structure; similar findings were identified in two external studies which associated these variants BE, reflux-symptoms and OAC risk (40-42).

1.4 Medical intervention

1.4.1 Diagnosis

Most patients are diagnosed with OC after reporting symptoms of progressive dysphagia, weight loss and anaemia (43). This is typically followed up with upper gastrointestinal endoscopy with biopsies, examining oesophagus to identify obstructions in the lumen, the wall of the oesophagus and outside the oesophagus compressing the lumen (43). Unfortunately, OC is often diagnosed at a late-stage, because of the oesophageal anatomy allowing for symptomless development (44). One study shows longer hospital delays result in poorer patient short-term outcomes (higher overall morbidity and mortality rates); however, long-term outcomes did not change (45). In the US, upon diagnosis only approximately 25% of patients present with localised disease, limiting available treatment options; this is reflected in the UK as 60-70% of patients present with late-stage disease, and are deemed not suitable for treatment with curative intent (46-48). However, the impact of the late diagnosis on survival is disputed; one study found time to diagnosis (TTD) was associated to patient malnutrition but did not affect resectability and both short- and long-term outcomes (44).

1.4.2 Staging and grading

OAC tumours are staged in two main systems in the UK, The TNM (T, primary tumour; N, regional lymph node involvement; M, distant metastatic spread) classification system (8th edition) and the number system (49). The TNM system allows staging to be represented using distinct categories, displayed in **Figure 3** and **Table 1**; the stratification of tumours allows for stratification of patient survival, by patient survival decreasing with increasing TNM stage group (numbers system; **Table 2**). T is used to classify the tumours local invasion, which, in combination with metastases and the presence of cancer cell in lymph nodes classifications depicts the tumours overall progression (50). notably, the edition version is vital as continual improvements are made regularly to improve the system. For instance, the transition from the 6th edition to the 7th edition resulted in better prognostic stratification of overall survival in oesophageal cancers and was especially performed greater in OAC (51). Secondly, histological grading is performed by a pathologist to describe the differentiation of the cancer cells within a tumour section

(**Table 2**); in oesophageal cancer histological grade is important as increasing histological grade is associated with decreased survival for early-stage cancers and used to distinguish stage I and stage IIA cancers by G1/G2 (well differentiated and moderately differentiated; stage I) from G3 (poorly differentiated; stage IIA) (**Table 1 & 2**) (50).

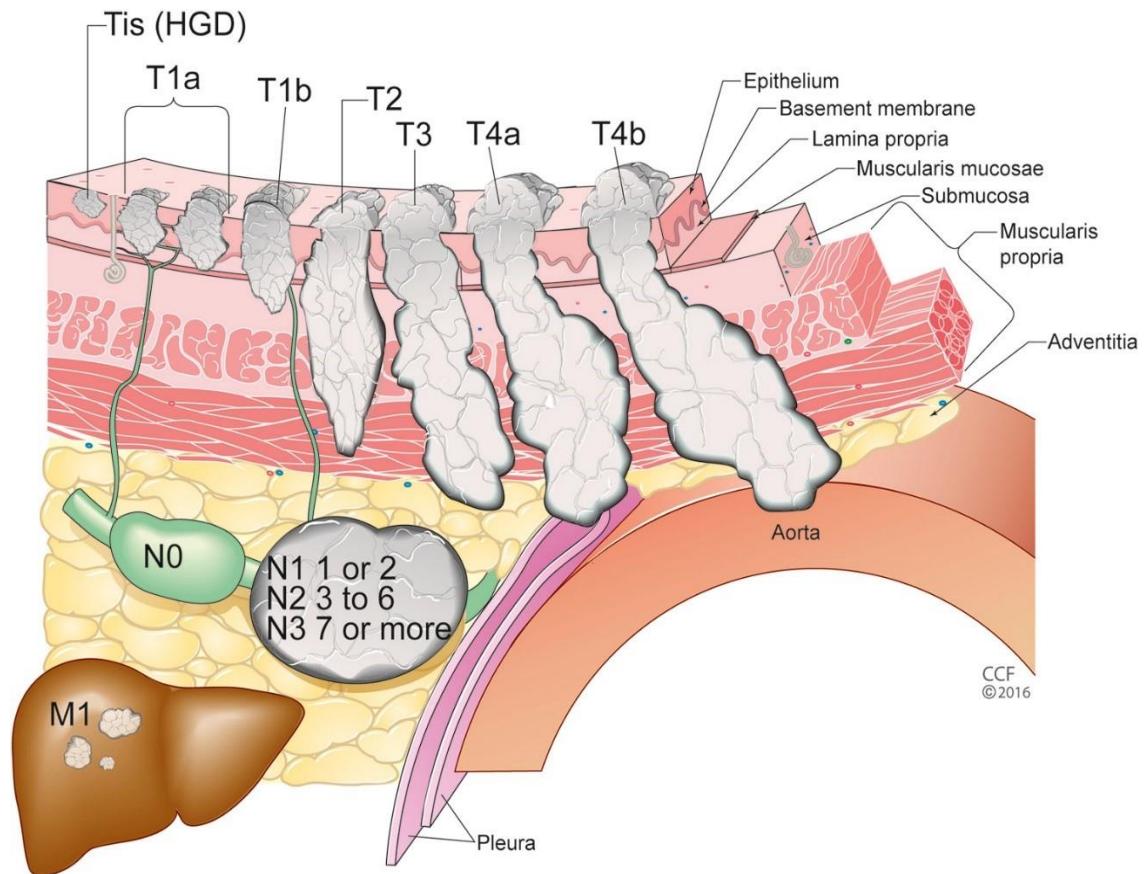


Figure 3 TMN staging 8th edition. Representation of different TNM stages of OAC in relation to human anatomy. Tis, high-grade dysplasia (HGD); T1, cancer invades lamina propria, muscularis mucosae, or submucosa; T2, cancer invades muscularis propria; T3, cancer invades adventitia; T4a, resectable cancer invading adjacent structures such as pleura, pericardium, or diaphragm; and T4b, unresectable cancer invading other adjacent structures, such as the aorta, vertebral body, or trachea. The N classifications are as follows: N0, no regional lymph node metastasis; N1, regional lymph node metastases involving one to two nodes; N2, regional lymph node metastases involving three to six nodes; and N3, regional lymph node metastases involving seven or more nodes. M is classified as follows: M0, no distant metastasis; and M1, distant metastasis. Adapted from Rice et al., 2017 (Permission obtained) (52).

Table 1 Table of TMN stage categories with descriptions specific to OAC. Adapted from Rice et al., 2017 (Permission obtained) 48.

T CATEGORY		N CATEGORY		M CATEGORY	
T0	No evidence of primary tumour	N1	Metastasis in 1–2 regional lymph nodes	M1	Distant metastasis
Tis	High-grade dysplasia, defined as malignant cells confined by the basement membrane	N2	Metastasis in 3–6 regional lymph nodes		
T1	Tumour invades the lamina propria, muscularis mucosae, or submucosa	N3	Metastasis in ≥ 7 regional lymph nodes		
T1a	Tumour invades the lamina propria or muscularis mucosae	NX	Regional lymph nodes cannot be assessed		
T1b	Tumour invades the submucosa				
T2	Tumour invades the muscularis propria				
T3	Tumour invades the adventitia				
T4	Tumour invades adjacent structures				
T4a	Tumour invades the pleura, pericardium, azygos vein, diaphragm, or peritoneum				
T4b	Tumour invades other adjacent structures, such as the aorta, vertebral body, or trachea				

Table 2 Histological grades for tumour sections (52).

GRADE	DEFINITION
GX	Grade cannot be assessed stage grouping as G1
G1	Well-differentiated
G2	Moderately differentiated
G3	Poorly differentiated
G4	Undifferentiated stage grouping as G3 squamous

1.4.3 Treatment – the current standard of care in the UK

After staging, a management plan is devised, tailored to the patient's stage; co-morbidities, WHO performance status, considering the patient's and family's wishes (52). In the UK, neo-adjuvant therapy followed by surgery represents the standard of care in the UK (52). Patients that receive treatment with curative intent for OAC are advocated to undergo preoperative chemotherapy or radiation therapy followed by surgery (53). Several UK-based trials have promoted the preoperative use of cisplatin, 5-fluorouracil and epirubicin in OC (Table 3), producing a significant survival advantage over surgery alone (54, 55).

Table 3 Preoperative chemotherapy drugs used in the UK for the treatment of OAC and ESCC.

DRUG NAME	CHEMICAL NAME	MECHANISM OF ACTION
Cisplatin	cis-diamminedichloroplatinum (II)	Binds purine residues and causes deoxyribonucleic acid (DNA) damage in cancer cells; oxidative stress (55, 56).
5-fluorouracil	fluoropyrimidine 5-fluorouracil (5-FU)	Metabolites of 5-FU inhibit nucleotide synthetic enzyme thymidylate synthase, and mis-incorporates fluoronucleotides into RNA and DNA, causing DNA and RNA damage (56).
Epirubicin	Epirubicin hydrochloride	Interferes with DNA, RNA and protein synthesis via the intercalation of DNA, inhibition of topoisomerase II activity, generation of oxygen and drug-free radicals (57).

Unfortunately, only 37.6% of patients are deemed suitable for treatment with curative intent. The low eligibility is because of the majority of patients presenting with late-stage disease, the US SEER (Surveillance, Epidemiology, and End results program) demonstrates this phenomenon with approximately 40% of patients are diagnosed with distant metastatic disease, 32% with regional disease, and only 18% of individual have localised disease at the point of diagnosis (58-60).

1.5 Histology

The histology of OAC typically displays the columnar epithelium compared to the squashed squamous epithelium found in healthy oesophageal tissue; while normal tissue shows a well-differentiated structure, OAC shows mass disruptions to structures such as the epithelium and glandular structures (Figure 4) (61-63). Notably, there are several three histologic types of columnar metaplasia which may cause adenocarcinoma including cardiac-type columnar mucosa composed of only mucosal cells; fundic-type columnar mucosa composed of mucosal cells, acid secreting parietal cells and chief cells; and intestinal-type columnar mucosa characterised by the presence of goblet cells (66).

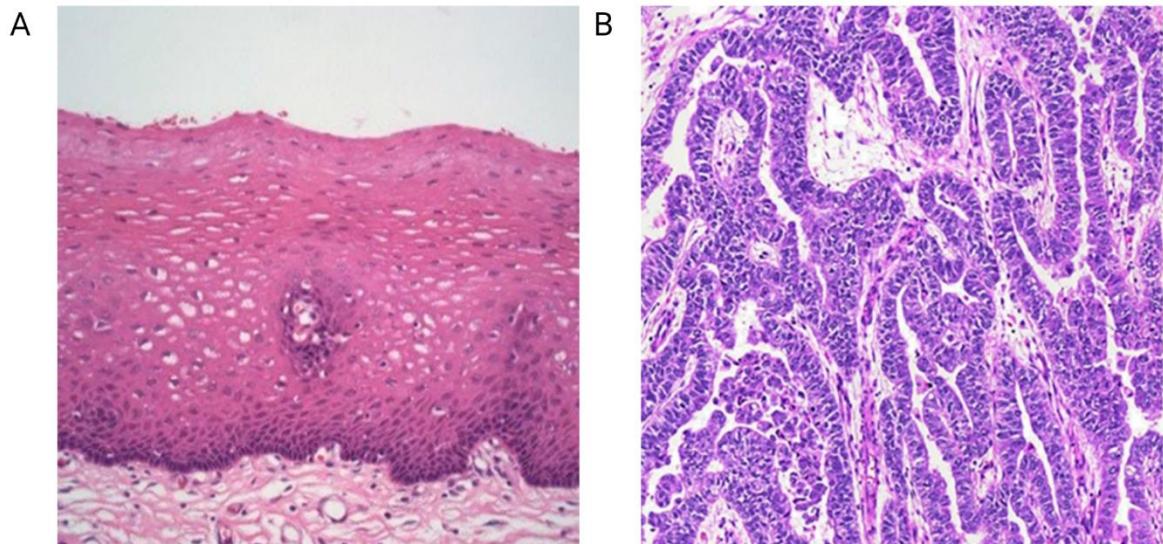


Figure 4 Histology of the human oesophagus. (A) Histology of normal squamous epithelium of the oesophagus, displaying a squashed epithelium towards the lumen with the lamina propria beneath divided by the basement membrane (65). (B) Histology of OAC, characteristically exhibits a columnar epithelium with malformed glands beneath, large irregular hyperchromatic nuclei may be found invading the submucosa (64). Permissions obtained.

1.6 OAC evolution, disease progression and metastasis

Evolution: As stated earlier, OAC is highly associated with GORD and Barrett's oesophagus. Occurring through a defined pathogenic pathway (**Figure 2**). However, the exact mechanism behind this transformation is still unknown. Studies have suggested that transdifferentiation may be a crucial part of the pathogenesis of Barrett's oesophagus; this process involves fully differentiated squamous epithelium cells changing into differentiated columnar cells, with or without mitotic division (65, 66). However, this mechanism is highly criticised as lineage-tracing studies have yielded no transdifferentiation events which result in columnar cells (66). Other studies suggest transcommitment; a process in which immature progenitor cells are reprogrammed to alter their patterns of cellular differentiation may allow for the squamous-columnar epithelial transition. The exact cell of origin for this process was elusive, yet recent literature suggests BE originating from gastric cardia via c-MYC and HNF4A-driven transcriptional programs with OAC arising from undifferentiated Barrett's oesophagus cell types (67, 68). Notably, somatically acquired genomic copy number alterations increase as OAC develops with the median percentage of the genome with a diploid copy number being 99.7% in Barrett's oesophagus compared to the OAC median diploid copy number at 62.4%. The most frequently affected candidate cancer genes among the 37.6% included GATA4, KLF5, MYB, PRKCI, CCND1, FGF3, FGF4, FGF19 and VEGFA (69). Despite this notable change in copy number, the DNA mutational signatures overlap between Barrett's oesophagus, early and late adenocarcinoma. Only a small amount of difference between early and late adenocarcinoma was noted, with 6.9% of the detected mutations being OAC-unique. This may indicate that the progression from BE to OAC driven by common mutagens and copy-number aberration rather than mutation. The overlap between BE and OAC suggests they are genetically similar to each other, and it has been further noted that 25% of OAC diagnosis occur within a year of BE diagnosis. This has previously been interpreted to be OAC present but missed by endoscopy, however, this may also be evidence of rapid progression of BE to OAC the pre-existing genetically unstable lineages (69).

During the development of OAC, tumour cells begin to preferentially express PD-L2 over PD-L1, potentially because of IL4/IL13 cytokines from CD4+ Th2 cells within the inflammatory environment produced during cancer development; also, preferential

expression of PD-L1 in tumour infiltrating inflammatory cells (TIIICs) over tumour cells was detected in OAC, compared to the ESCC with the expression of PD-L1 in both tumour cells and TIIICs (70). Preferential expression of PD-L1 and PD-L2 is a recurrent immune evasion mechanism found in OAC and other cancers. Specifically, Small-cell lung cancer (71), squamous cell carcinoma of the oral cavity (72), cervical cancer (73), ovarian cancer, breast cancer, melanoma, bladder cancer, head and neck cancer, soft tissue sarcoma and prostate cancer (74, 75) have showed frequent copy-number increases of chromosome 9p24 where the PD-L1 gene (CD274) is located.

Disease progression: Deconvolution analysis comparing BE to OAC found the largest expression changes in BE progression occurred in M2 macrophages, pro-B cells, and eosinophils, whereas immunohistochemical analysis found OAC tumour sections were relatively immune poor, with a rise in PD-L1 expression and loss of CD8+ T cells (76).

Metastasis: A recent study of 388 samples of OAC found that 90% of patients possessed subclones from the primary tumour which spread rapidly from the primary site to form multiple metastasis sites, including the lymph nodes and liver, supporting a clonal diaspora model for metastases of OAC. Further analysis revealed high L1 transposon activity in metastatic OAC compared to the primary site relating to a larger proportion of structural variants in the metastatic samples, suggesting an increase in genomic instability in later disease stages and potentially metastasis (77).

1.7 Immunotherapy trials

Interestingly, recent US trials focusing on immunotherapy have showed moderate responses to checkpoint blockade inhibitors, Pembrolizumab and Nivolumab in OAC (78, 79). Pembrolizumab, an inhibitor of PD-1, became the first approved immune checkpoint inhibitor in 2014 after a phase Ib trial in advanced melanoma (KEYNOTE-001) (80, 81). Nivolumab, a fully human IgG4 monoclonal antibody targeting PD-1 with high affinity for PD-1 inhibiting PDL1/PD-L2-PD-1 binding (**Figure 5**), received approval for treatment of metastatic melanoma, non-small cell lung cancer, and renal cell carcinoma (82). Strikingly, Nivolumab demonstrated positive responses in the treatment of chemotherapy-refractory gastric and gastroesophageal junction in a double-blind phase 3

trial for patients who have received two or more lines of chemotherapy and have not been selected for PD-L1 expression (ATTRACTION-2) (83).

Ipilimumab is a monoclonal antibody targeted to CTLA-4, a transmembrane receptor on T cells, capable of activating T cells in cancer settings (**Figure 5**). A recent study (CheckMate-032) employed the usage of ipilimumab plus nivolumab in the treatment of advanced oesophageal cancer, confirming dual treatment with ipilimumab and nivolumab was superior to nivolumab monotherapy (84). The CheckMate-032 trial demonstrates combination therapy of anti-CTLA-4 and Anti-PD-1/PD-L1 may be more effective than monotherapies; this avenue of combination treatment options is being investigated in several cancers advanced melanoma, metastatic osteosarcoma, colorectal cancer, recurrent glioblastoma, and renal cell carcinoma with favourable responses to these combination therapies (85-92).

CAR-T (Chimeric antigen receptor) cells may also provide an effective measure against OAC in future with guanylyl cyclase C (GUCY2C) targeting CAR-T cells typically used to target human colorectal cancer may apply to OAC, with several studies indicating GUCY2C is highly expressed by OAC particularly in moderately to well-differentiated tumours (93-95).

Overall, most immunotherapeutic efforts in OAC are focused on the use of checkpoint blockade therapeutics including Pembrolizumab, Nivolumab and ipilimumab in monotherapies or in combined therapies (79, 84, 96-103). A current literature search (15/08/2021) could not identify any trials or publications using adoptive T cell therapies in OAC; however, adoptive T cell therapy in combination with low-dose Nivolumab has been trialled in two patients with OSCC which may indicate adoptive T cell therapies may also present a future treatment for OAC (8).

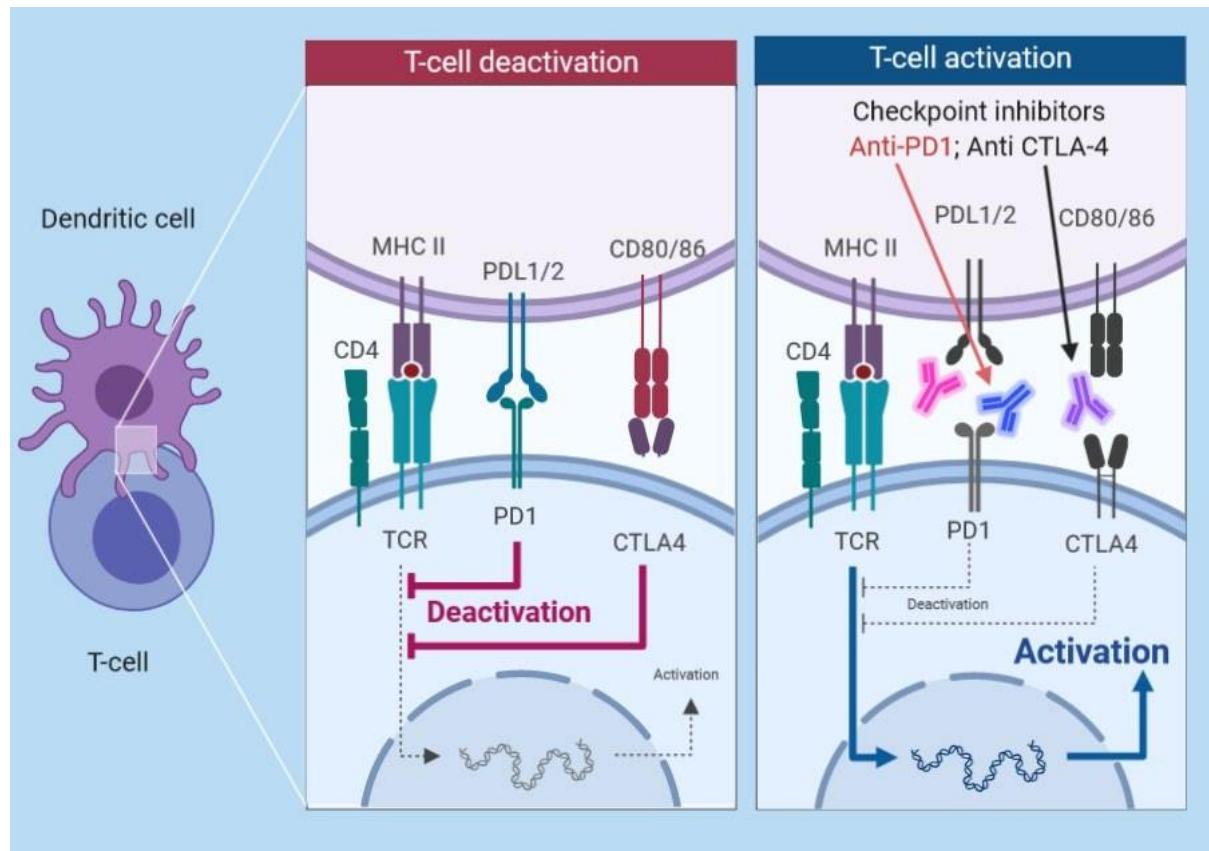


Figure 5: Mechanism of action of immune checkpoint inhibitors. (Left) PD-1 is expressed on activated T cells which may bind to PD-L1 expressed on tumour cells, leading to T cell exhaustion. CTLA-4 competes with the co-stimulatory T cell molecule CD28 for B7 ligands, upon bonding CTLA-4 decreases T cell proliferation and cytotoxic activities. (Right) Blockade of CTLA-4 by anti-CTLA-4 antibodies and PD-1 with anti-PD-1 or anti-PD-L1 antibodies stimulates T cell activity and inhibits tumour cell's ability to suppress T cell activity through the expression of PD-1 or CTLA-4 (104). (Created with Biorender.com).

1.8 The tumour microenvironment

The tumour immune microenvironment (TIME) is the network of interacting cancer and immune cells found within a tumour. Notably, the mechanisms of immunity with these niches differ significantly from those outside of the tumour microenvironment niche, because of cancer's ability to manipulate and obstruct normal immune functions as a mechanism to avoid anti-tumoural immunity. For its significant role in the development and progression of tumours, the TIME is deemed as an emerging hallmark of cancer (105).

Specifically, three general classes of TIME have been defined infiltrated-excluded, infiltrated-inflamed and infiltrated-T lymphocytes. Infiltrated-excluded is defined by the exclusion of cytotoxic T lymphocytes (CTLs) from the tumour core; instead, CTLs may be found on the periphery of the tumour interacting with tumour-associated macrophages or stuck in fibrotic nests; comparatively infiltrated-inflamed TIMEs are characterised by the high abundance of PD-L1 expression on tumour and myeloid-derived cells, and possess CTLs high in the expression of *GZMB*, *IFNG* and *PDCD1*; finally, infiltrated-T lymphocytes possess tumour infiltrating lymphocytes (TILs) localised to immune cell compositions similar to those found in lymph nodes.

Overall, the TIME is defined by the immunosuppressive role it performs to achieve anti-tumoural immune evasion. However, fully characterising the TIME can be difficult because of the number of cells and molecules involved in these processes. For example, the cytokine network which forms the inter-cellular immune communication between specialised immune cells forms a complex network, which may act to either suppress or promote immune destruction of cancer cells (**Figure 6**).

Immune profiling of OAC in 47 samples identified high expression of checkpoint markers including LAG3, TIM3, CTLA4 and CD276 compared to PD1/PD-L1, besides upregulated CD38 and LILRB1 expression demonstrating multiple immune checkpoint mechanisms suggesting immune checkpoint therapies should be used in combinations for the most effective treatment (106).

Immunochemical and deconvolution analysis have characterised the OAC microenvironment by poor cytotoxic effector cell infiltration and increased immune inhibitory signalling; between BE and OAC T cells no significant increases in CD4, CD8, or Treg populations or subpopulations was detected, however, Th1 and Th2 cells were found increased in OAC (77).

Whilst the myeloid cell populations appeared unchanged overall, there was a difference in the balance of M1/M2 macrophages with M2 macrophages found to be significantly increased in OAC compared to BE, pro-B cells were also found increased in OAC as compared to BE with the remaining B cell populations being unchanged (77). Overall, these analyses identify much of the immune population remains similar to BE, however, some distinct differences can be identified in the immune subpopulations; these notable differences in the immune subpopulation balances expected and measured differences in cytokines be found to modulate and impede immune functions such as IL6 and CXCL8 (77, 107).

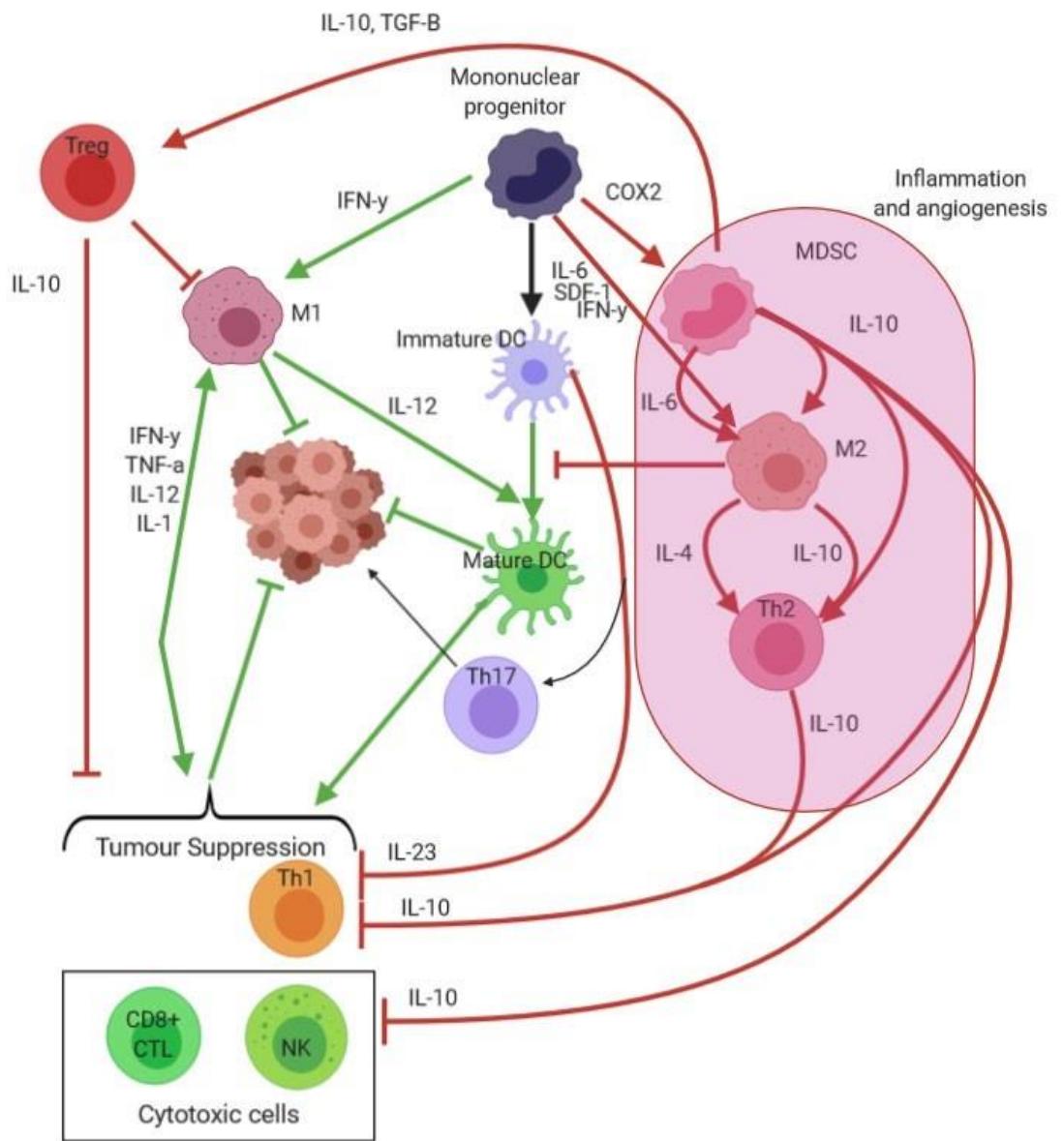


Figure 6: Cytokines forms an essential intracellular communication network which may act to suppress anti-tumour immune response via several diverse mechanisms.

One key pathway of immunosuppression (Red lines) is located via MDSCs inducing M2 macrophage polarisation via secretion of IL-6 and IL-10 and Treg activity via IL-10 and TGF-B secretion, this results in a greater concentration of IL-10 within the TME inhibiting CD8+ and NK cell activity while M2 macrophages hinder the maturation of DCs reducing presentation of tumour antigens to T cells. Conversely, tumour suppression (Green lines) is also dependant on the usage of the cytokine network with M1 macrophages communicating with CD8+, NK and Th1 T cells allowing tumour directed immune rejection to occur.

1.9 Antigen Presentation Machinery (APM) and pathways

Antigen presentation functions within host immunity for the recognition of self and non-self. T cells may act to destroy infected host cells through the specific recognition of antigens presented on antigen presentation molecules on the cell surface. Upon specific recognition and with the co-stimulatory signal, the T cell may activate and direct cytotoxicity towards the infected cell, aiding in the clearance of infection.

Antigen presentation may be functionally divided into endogenous and exogenous pathways (**See Figure 7**). The endogenous pathway, also known as MHC I functions to display intracellular peptides, such as self-proteins or incorporated viral proteins, whereas exogenous pathways (MHC II) function to display extracellular peptides, engulfed by a professional antigen presenting cell (108). Interestingly, cancer cells may also present cancer-specific antigens on the cell surface, which may be recognised by cancer-specific T-cells, acting to suppress tumour formation by destroying cancer cells (109).

The MHC I pathway is an antigen presentation mechanism characterised by displaying endogenous peptides on MHC I molecules to CD8+ T cells (**Figure 7A**). MHC I pathways are expressed in most cell types in the body, excluding non-nucleated cells such as red blood cells which rely on CD47 markers for immune recognition and tolerance (110).

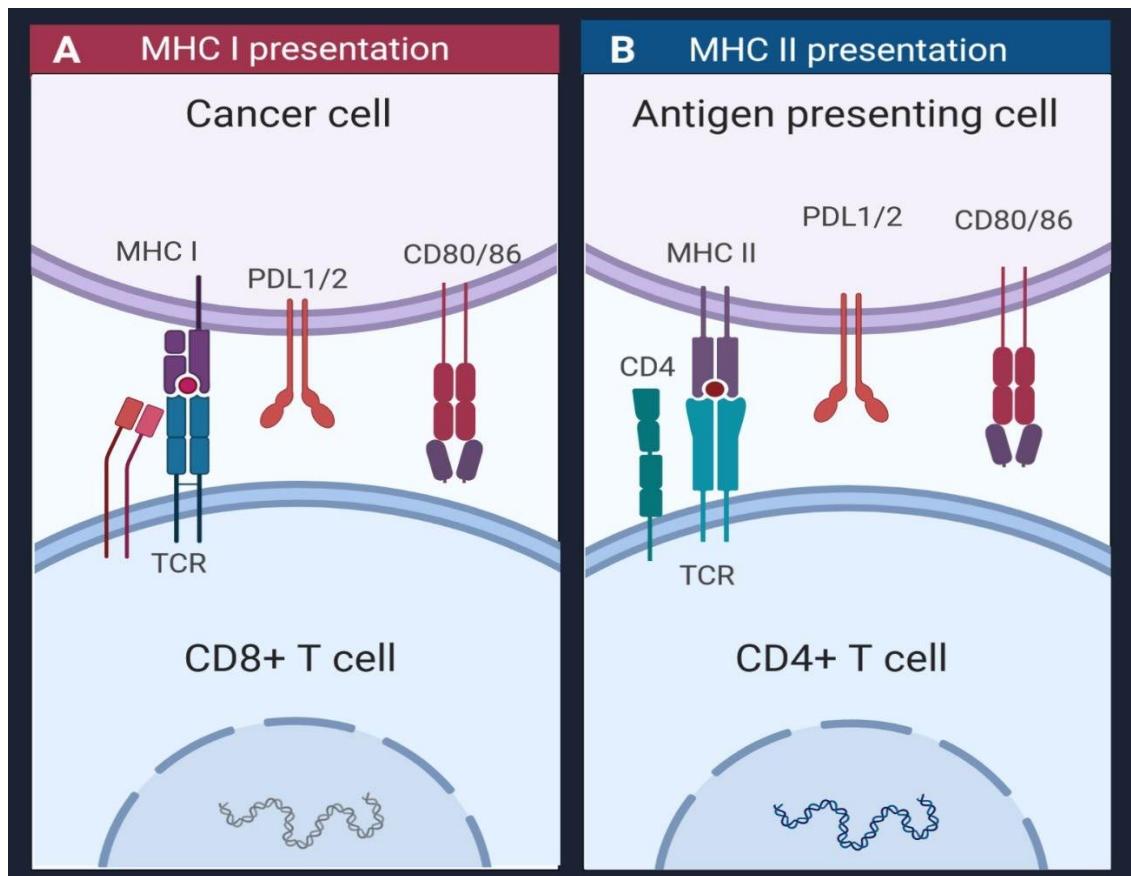


Figure 7: MHC pathways in humans. Cellular interface of the Antigen Presentation Pathways (A) MHC I presentation; endogenous peptides are presented to CD8+ T cells, which specifically recognise the antigen with a T cell receptor (TCR), provided the T cell is supplied with a recognisable antigen, co-stimulatory signals, and a lack of co-inhibitory signals, the CD8+ T cell may activate and destroy the cell. B) MHC II presentation; exogenous peptides are presented to CD4+ T cells, specifically recognising the antigen with a TCR, provided the T cell is supplied with a recognisable antigen, co-stimulatory signals, and a lack of co-inhibitory signals, the CD4+ T cell may activate and function to stimulate/support an immune response. (Created with Biorender.com).

1.9.1 The MHC I pathway

The MHC I pathway is an antigen presentation mechanism characterised by displaying endogenous peptides on MHC I molecules to CD8+ T cells (**Figure 7A**). MHC I pathways are expressed in most cell types in the body, excluding non-nucleated cells such as red blood cells which rely on CD47 markers for immune recognition and tolerance (110). Endogenous peptides (normal, cancer-specific or pathogenic) are produced from proteins within the cytosol which are degraded by the proteasome into peptide fragments; several peptide fragments are transported into the endoplasmic reticulum (ER) lumen through the ER membrane translocon, transporter associated with antigen processing (TAP1/2) (110) (**See Figure 8**). The nascent heavy chain of the MHC I molecule is co-translationally transported into the ER lumen by the translocon docking to the ribosome on the cytoplasmic face of the ER membrane, then binds to calnexin until a dimer with Beta-2-microglobulin (B2M) is formed. The MHC I- B2M complex is then released from calnexin, binding to the chaperonin proteins calreticulin and ERp57, which transport the complex to TAP, where the complex binds to TAP via Tapasin forming the peptide-loading complex (PLC) (110).

At this stage, peptide fragments can now bind to the peptide-binding cleft of the MHC I molecule, provided these fragments are 8-10 amino acids (AA) in length. However, peptides longer than 8-10 AA in size may be trimmed further by ERAP1 and ERAP2. After successful binding of a stable peptide to the peptide-binding cleft of an MHC I molecule, the MHC-peptide complex leaves the PLC, effectively passing the ER quality control process, allowing the MHC-peptide complex to be transferred to the Golgi apparatus network, transporting the complex to the cell surface (**Figure 8**) (110).

The MHC I-peptide complex may then be specifically recognised by a CD8+ T cell receptor (TCR), providing this T cell is supplied with a non-self/non-healthy MHC I presented antigen, sufficient co-stimulatory signals (E.g. CD28-B7 interaction) and lack of co-inhibitory signals (E.g. CD80/86-CTLA4 interaction) (110). Notably, MHC I molecules also function as an inhibitory signal to NK cells which actively target cells which lose MHC I expression (111).

To evade immune destruction, cancer targets MHC I genes for somatic mutation, effectively abrogating the MHC I pathway to hinder the presentation of cancer antigens on MHC I molecules to the immune system. For example, one study identified B2M somatic mutations early in cancer development, which are attributed to the loss of function and reduced MHC I surface expression, thus forming a cancer immune evasion mechanism (112). Further examples of somatically acquired APM defects have been observed in the HLA class I genes including HLA-A/B/C, displaying roles in cancer immune evasion by producing antigen loss variants that undergo selective immune pressure during cancer development (113). Overall APM genes represent an inviting target for somatically acquired aberration in cancer development and should be a key topic of research as specific APM aberrations may impact the efficacy of immunotherapy (113).

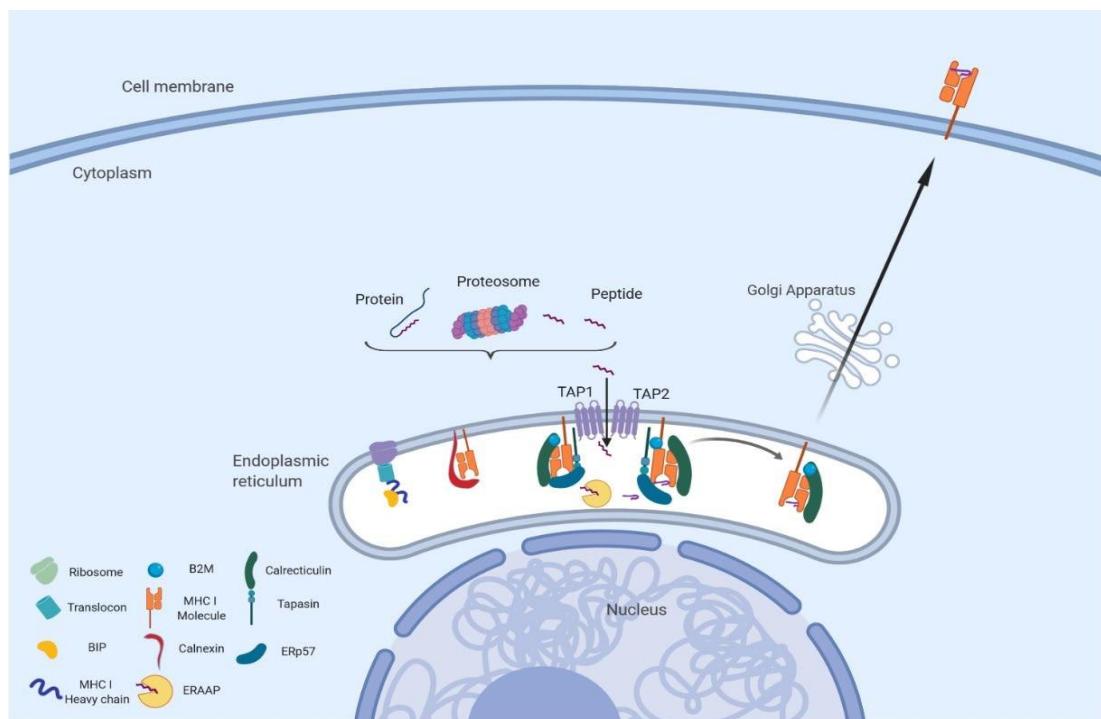


Figure 8 MHC I pathway overview. MHC I molecules are transcribed by ribosomes on the ER, folded, then transported by chaperones to Tapasin (TAP1/2). Endogenous peptides are degraded by proteasomes into peptide fragments then transported into the ER lumen by Tapasin, peptide fragments exceeding 8-10 AA in size are degraded further by ERAAP. In the ER lumen, antigen peptide fragments form complexes with MHC I molecules which are then transported to the cell surface via the Golgi apparatus for display. (Created with Biorender.com).

1.9.2 The MHC II pathway

The MHC II pathway is characterised by the display of exogenous peptides on MHC II molecules to CD4+ T cells (**Figure 7B**). MHC II pathways are primarily expressed by professional antigen presenting cells, such as dendritic cells and macrophages, as well as some endothelial cells, thymic epithelial cells, and B cells. MHC gene evolution has been a key topic of research over the past five decades, with only a few species being used to make direct comparisons between MHC I and II. A few studies have identified pathogenic selective pressure and the variation of MHC II alleles has played a key role in the evolutionary history of MHC II genes; however, the evolutionary origin of these antigen presenting cells is mostly unknown, meaning little evidence exists to determine the exact origins of MHC II genes (114, 115). Notably, MHC II expression is present in epithelial cells of the lung, intestine, and stomach; specifically gastric epithelial cells have demonstrated to upregulate MHC II expression in chronic inflammatory conditions, such as autoimmune conditions and may extend to gastric cancers (116).

Exogenous peptides carriers, such as pathogenic bacteria are engulfed into phagosomes which fuse with lysosomes forming phagolysosomes, these structures degrade exogenous peptides with proteases into peptide fragments between 13 and 17 AA in size (**Figure 9**). Meanwhile, MHC II molecules are transcribed in the ER alongside the invariant chain (li) which form a complex together, the complex is then transported to the MHC II compartment (MIIC) via the Golgi apparatus (110).

Following peptide degradation, the phagolysosome fuses with the MIIC releasing degraded exogenous peptides and proteases into the MIIC; proteases CTSS or CTSL act to cleave li into CLIP. CLIP is exchanged with an exogenous peptide fragment, catalysed, and stabilised by HLA-DM, which itself is regulated by HLA-DO. Finally, the MHC II peptide complex is transported to the cell surface for display via the endosomal network. Alternatively, MHC II-li complexes can be transported directly from the ER to the cell surface, which may be recycled back into the MIIC for antigen loading (110).

The displayed MHC II peptide-complex may stimulate and activate CD4+ T cells, which act as helper cells supporting the immune response towards pathogen clearance, such as secretion of immunostimulatory cytokines; specifically, the secretion of IL-1, IL-2, IL-12, IFN- γ and TNF- α from Th1 cells play a positive role in anti-tumoral immunity, bolstering CD8+ and NK cell activity, survival and expansion (109, 117).

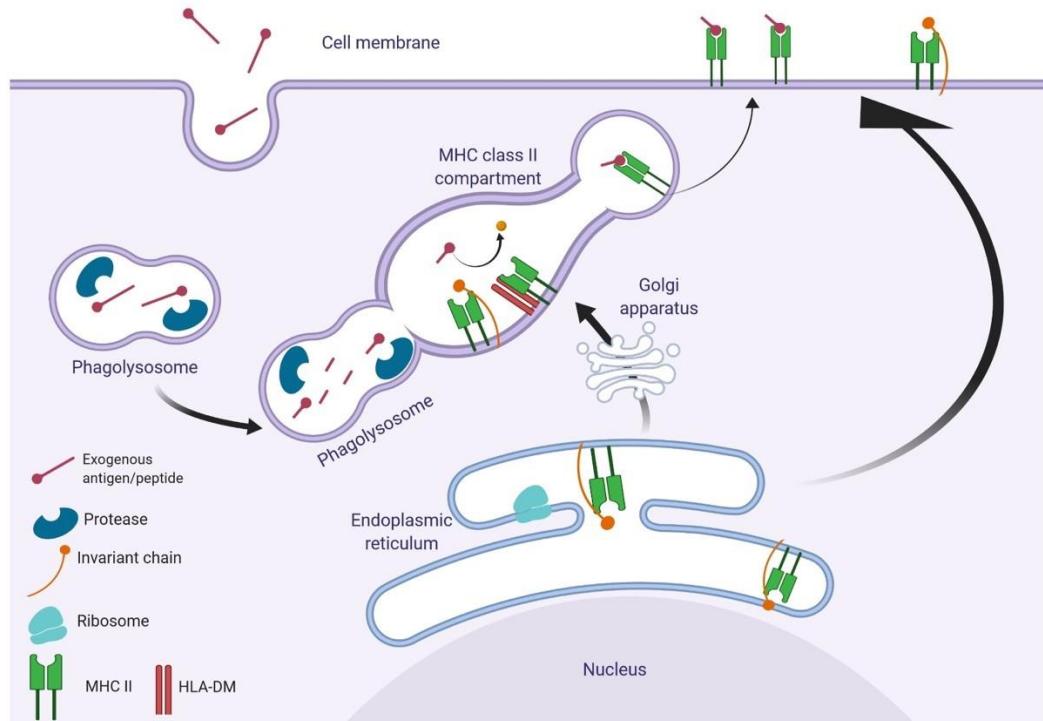


Figure 9 MHC II pathway overview. MHC II molecules and invariant chains are transcribed by ribosomes on the ER, folded, bound, then transported to the MHC II compartment (MIIC) via the Golgi apparatus. Exogenous peptides are internalised and degraded by proteases in phagolysosomes, phagolysosomes then fuse with the MHC II compartment, releasing proteases which cleave invariant chains on MHC II molecules into CLIP; CLIP is then exchanged with antigen peptides assisted by HLA-DM. MHC II-antigen complexes are then transported by the endocytic network to the cell surface. (Created with Biorender.com).

1.9.3 Cross-presentation pathways

The mechanisms of presenting exogenous antigen on MHC I molecules are coined as cross-presentation pathways. These pathways have been discovered predominantly in dendritic cells (DCs), being tied to important functions including the generation of an immune response directed towards viruses and tumours, immunisation and in the induction of immune tolerance (118).

Two general pathways exist within this domain; the cytosolic pathway, where internalised antigens are transported to the cytosol from the endosomal compartments, exogenous cytosolic peptides are then degraded by the proteasome, transported into the ER by TAP, allowing for the loading of exogenous derived peptides in the MHC I pathway; the vacuolar pathway, internalised antigens are degraded by lysosomal proteases and then loaded onto MHC I molecules within the endocytic compartment (**Figure 10**) (118). Interestingly, Cathepsin S, not Cathepsin L and B have been shown to play a key role in antigen degradation within the vacuolar pathway, demonstrated within DCs of TAP deficient mice (119).

In cancer, cross-presentation of tumour antigens can occur via phagocytosis of dead cancer cells; as the local tumour microenvironment can be pro-inflammatory cell death of tumour cells can be significant, resulting in cellular debris (120). The debris from dead cells can be engulfed into APCs then presented on MHC I molecules via the cross-presentation pathways (121). Interestingly, cross-presentation of tumour antigens can be enhanced by tumour burden and may be conducted by CD1c dendritic cells and macrophages, which may cause CD8+ anti-tumoral T cell responses (122, 123).

Finally, the cross-presentation mechanism may be exploited for future immunotherapy efforts; several studies have identified immunotherapies focusing on enhancing DCs or introducing tumour antigens to DCs could allow for greater T cell priming, as tumour antigens would be presented to both CD4+ and CD8+ T cell, inducing an anti-tumoral response (124-126).

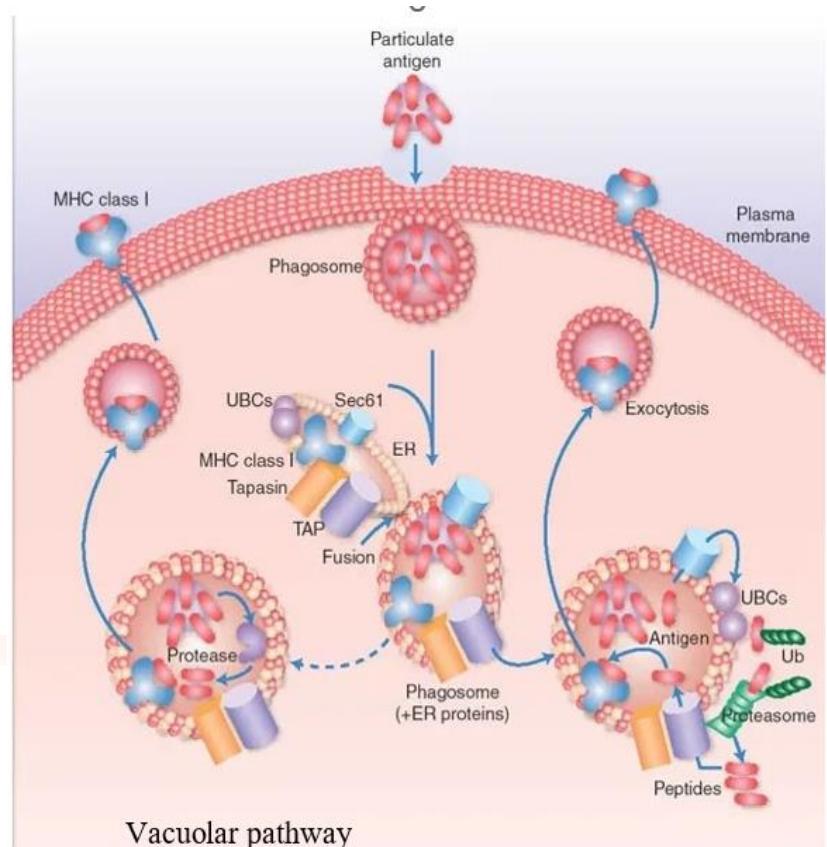


Figure 10 Cross-presentation cellular mechanisms. In both the vacuolar and cytosolic pathway, exogenous peptides are engulfed by the cell and internalised into a phagosome, which then fuses to the endoplasmic reticulum (ER), allowing it to obtain antigen presentation machinery proteins such as TAP, Tapasin, MHC I molecules and Sec61 translocons. In the cytosolic pathway the phagosome Some internalised exogenous peptides are transported out to the cytosol, potentially via ubiquitination by ubiquitin (Ub)-conjugating enzymes (UBCs) resulting in translocation of the peptides through the Sec61 translocon; peptides translocated to the cytosol can be degraded by proteasomes, allowing for the cleaved peptides to be transported back into the fused phagosome through TAP permitting peptide binding to MHC I molecules using antigen presentation machinery gained from the ER. In contrast, proteases degrade the vacuolar pathway internalised peptides within the ER-fused phagosome by proteases, allowing Tapasin the load the cleaved peptides onto MHC I molecules, then transported to the cell surface. Adapted from Rock, K., 2003. (107).

1.9.4 Regulation of APM gene expression

The activation of MHC class I genes except for *HLA-G* is mediated by multiple conserved *cis*-acting regulatory promoter elements including ISRE (interferon-stimulated response element) and the SXY-module (127). Interestingly, these regulatory promoter elements are further involved in the transcriptional activation of the *B2M* promoter, but not *TAP* and immunoproteasome promoters (127-129). Within this pathway for MHC class I promotion, the enhancer A is bound by the nuclear factor NF-KB and ISRE is bound by interferon regulatory factor (IRF) family members. Specifically, transcription factors NF- κ B and IRF-1 function as mediators of the TNF α and IFN- γ pathway gene activation which result in induction of MHC class I transcription. Notably, IRF-1 inactivation has been uncovered in cancers. In leukaemia and pre-leukemic myelodysplasia (MDS) loss of chromosome 5 or deletion within the long arm where the *IRF-1* gene is located occurred in 30% of MDS cases, 15% of de novo acute myelogenous leukaemia (AML) cases, 50% of cases of secondary AML arising from MDS (130-132). Following the discovery of IRF-1 loss in AML and MDS further research identified 50% of gastric tumours and 32% of invasive breast carcinomas exhibit LOH at the 5q region implying a critical contribution of IRF-1 to the development of stomach carcinoma and invasive disease in breast cancers (133, 134). Specifically, to oesophageal carcinoma, 5q31.1 was reported as the smallest commonly deleted region in 57% of the specimens tested (35/61 tumours); a further study found expression of IRF-1 to be decreased and IRF-2 increased in OSCCs compared with matched normal oesophageal tissue (135, 136). As prior mentioned, *HLA-G* expression is not regulated by NF- κ B or IRF-1, instead the *HLA-G* promoter can be transactivated by the cyclic-AMP response element binding protein (CREB)-1. Whereas, Ras-responsive binding protein-1 (RREB-1) acts as a transcriptional repressor of *HLA-G* via chromatin remodelling of the *HLA-G* locus (137).

A Further regulator of MHC class I expression was discovered in 2010 by ChIP (Chromatin Immunoprecipitation) analysis revealing NLRC5 associated to MHC class I promotor regulating its transcriptional activity (138). NLRC5 functions via forming a MHC class I enhanceosome together with the RFX-complex (RFX5, RFXAP, RFXANK), ATF1/CREB and the NFY-complex, thus acting as a key co-activator for MHC class I expression (139). In a pan-cancer study, preferential methylation, copy number and somatic mutations of NLRC5 associated to impaired cytotoxic T-lymphocyte activity and decreased MHC class I

expression in breast, skin, liver, ovarian, bladder, lung, and prostate cancers (140). The pan-cancer study further identified the expression of NLRC5 was correlated with survival of cancer patients with skin, rectal, bladder, uterine, cervical, and head and neck tumours (140).

Class II transactivator (CIITA) is a transcriptional coactivator that regulates γ -interferon-activated transcription of MHC class I and II genes. CIITA structure possess a series of regulatory domains including an activation domain, an acetyltransferase domain, a proline/serine/threonine (PST) domain, a GBD (GTP binding domain) and a canonical LRR (C-terminal leucine rich region) domain common to NLR (Nod-like receptor) proteins (141). Notably, deficiency of CIITA results in aberrant MHC gene expression and has been associated to autoimmune diseases such as Type II bare lymphocyte syndrome. Functionally, CIITA regulates MHC gene expression via two distinct mechanisms, as a transcriptional activator and as a general transcription factor. In CIITA's transcriptional activator role CIITA nucleates an enhanceosome comprising the DNA binding transcription factors RFX, cyclic AMP response element binding protein, and NF-Y. As a transcription factor CIITA functionally replaces TAF1 a TFIID component. Similarly, to TAF1, CIITA possesses acetyltransferase (AT) and kinase activities and actively contributes to both MHC class I and II transcription.

Recently, the role of CSDE1 as a regulator of MHC class I expression was discovered, functioning via the stabilisation of TCPTP a tyrosine phosphatase which dephosphorylates STAT1 in the IFN- γ pathway. Within these studies, the greater expression of CSDE1 in a melanoma cell line resulted in high dephosphorylation of STAT1 preventing translocation of STAT1 into the cell nucleus downstream of the JAK/STAT pathway (142). This function inherently prevents STAT1 from binding to the γ -activated sequence (GAS), this inhibits the promotion of IRF1 which handles MHC class I promotion by binding to ISRE a prior discussed transcriptional promoter of MHC class I expression, decreasing the expression of MHC class I genes. Further investigation yielded the regulation of CSDE1 occurs via SMYD3 trimethylation of CSDE1 locus. A further study identified CSDE1 may form a target for therapy via oncolytic viruses, creating a selection pressure towards an escape-associated tumour antigen via mutation of CSDE1 (C-T) which may be a target for cancer vaccines, adoptive T-cell and CAR-T cell therapies (143).

1.10 Candidate APM genes

Using prior literature several APM genes have been reported as possessing an impact on survival and immune responses in other cancers and diseases, by assessing the known APM defects in other cancers which impact prognostic values a list of defined APM gene candidates for analysis in OAC was determined. These candidates and their respective impact in other cancers are stated below in **Tables 4 and 5**.

Table 4 Key APM genes in the MHC I pathway with their associated APM processes and published roles in cancers

APM PROTEIN; MHC I	APM GENE SYMBOL; MHC I	GENE FUNCTION	EVIDENCE FROM LITERATURE
TAPASIN	<i>TAPBP</i>	Assembly of the MHC class I loading complex.	Tapasin is downregulated in multiple OSCC cell lines. Tapasin downregulation is associated with poorer outcomes in triple-negative breast cancer (144, 145). Loss of Tapasin correlates with diminished CD8+ T-cells and disease progression in colorectal cancer (149).
TAPBPR	<i>TAPBPL</i>	Assembly of the MHC class I loading complex; binds peptide-free MHC I molecules (permitting peptide editing)	TAPBPR isoforms alter association with MHC I molecules; long isoforms may down-regulate surface expression of MHC I molecules (150).
HLA-A	<i>HLA-A</i>	Assembly of the MHC class I loading complex. Antigen presenting molecule.	The HLA-A *11 allele is strongly associated to OSCC risk (151). HLA-A expression is regulated by the HER2 signalling pathway in OSCC and gastric cancers (152).
HLA-B	<i>HLA-B</i>	Assembly of the MHC class I loading complex. Antigen presenting molecule.	The HLA-B*46 allele is strongly associated to OSCC risk (151). HLA-B expression is downregulated in gastric cancer and in lymphatic metastasis (153).
HLA-C	<i>HLA-C</i>	Assembly of the MHC class I loading complex. Antigen presenting molecule.	HLA-C expression is down-regulated in CRC (154).

HLA-E	<i>HLA-E</i>	Assembly of the MHC class I loading complex. Antigen presenting molecule.	HLA-E+ gastric tumours carry poorer 5-year survival than HLA- tumours (155).
HLA-G	<i>HLA-G</i>	Assembly of the MHC class I loading complex. Antigen presenting molecule.	HLA-G expression correlates with poor prognosis in gastric carcinomas (156).
BETA-2 MICROGLOBULIN	<i>B2M</i>	Assembly of the MHC class I loading complex. Antigen presentation	B2M mutations reduce the overall levels of cell surface MHC-I molecules; B2M mutant tumours possessed elevated cytotoxicity (124). B2M loss of heterozygosity in metastatic melanoma immunotherapy is enriched. threefold in non-responders (~30%) compared to responders (~10%) and associated with poorer overall survival (157).
ERP57	<i>PDIA3</i>	Assembly of the MHC class I loading complex.	PDIA3high gastric tumours have improved prognosis over PDIA3low tumours (158).
CALRETICULIN	<i>CALR</i>	Assembly of the MHC class I loading complex.	CALR expression is associated with infiltration of T-cells; strong CALR expression with high CD3+ and CD45RO+ T-cell infiltration is associated with higher 5-year survival in colon cancer (159).
CALNEXIN	<i>CANX</i>	Assembly of the MHC class I loading complex.	Calnexin downregulated in ~25% of primary laryngeal squamous cell carcinoma lesions, number of infiltrating CD8+ T cells correlated with calnexin expression (160).
ERAP1	<i>ERAP1</i>	Assembly of the MHC class I loading complex. Degradation of peptides in the ER	ERAP1 down-regulation suffices to stimulate the cytotoxic activity of NK cells and CTLs (against an unknown cancer antigen), resulting in tumour growth arrest (161, 162).
ERAP2	<i>ERAP2</i>	Assembly of the MHC class I loading complex. Degradation of peptides in the ER.	ERAP2 expression lost in 17 out of 26 liver carcinoma samples. Loss of ERAP1/2 expression is associated with a lack of surface HLA class I molecules (163).
TAP1	<i>TAP1</i>	Peptide transport into the ER lumen.	High expression of TAP1 by OAC correlated with significantly shorter overall survival times in OAC (164).
TAP2	<i>TAP2</i>	Peptide transport into the ER lumen.	Increased levels of MIR125a-5p and MIR148a-3p reduces levels of TAP2 in OAC, associated to poor patient outcomes (164).

Chapter 1

PROTEASOME BETA SUBUNIT BETA1I (INDUCIBLE)	<i>PSMB9</i>	Peptide generation and trimming.	Overexpression of PSMB8 and PSMB9 correlates to better survival and improved response to immune checkpoint inhibitors of melanoma patients (165).
PROTEASOME BETA SUBUNITS BETA2I (INDUCIBLE)	<i>PSMB10</i>	Peptide generation and trimming.	LMP10 nuclear expression in HPV positive and LMP10 cytoplasmic expression in the HPV-negative tonsillar and base of tongue squamous cell carcinoma patients correlated to better clinical outcome (166).
PROTEASOME BETA SUBUNITS BETA5I (INDUCIBLE)	<i>PSMB8</i>	Peptide generation and trimming.	Overexpression of PSMB8 and PSMB9 correlates to better survival and improved response to immune checkpoint inhibitors of melanoma patients (165).

Roles in cancer collected by literature search on PubMed

(<https://www.ncbi.nlm.nih.gov/pubmed/>) using the following search terms (1) *Gene symbol* OR *Protein name* AND “Antigen presentation” AND “Cancer” (2) *Gene symbol* OR *Protein name* AND “Antigen presentation” (3) (1) *Gene symbol* OR *Protein name* AND “Cancer”.

Table 5 Key APM genes in the MHC II pathway with their associated APM processes and published roles in cancers.

APM PROTEIN; MHC II	APM GENE SYMBOL; MHC II	GENE FUNCTION	EVIDENCE FROM LITERATURE
HLA-DM	<i>Alpha: HLA-DMA</i> <i>Beta: HLA-DMB</i>	Stabilisation and loading of the MHC II complex	Decreased expression in B cell lymphoma (167); Lack of HLA-DM in head and neck squamous cell carcinoma (168).
HLA-DO	<i>Alpha: HLA-DOA</i> <i>Beta: HLA-DOB</i>	Antigenic peptide chaperoning and loading into MHC II molecules.	Methylated promoters in RFX5-negative B-lymphoma cells (169).
HLA-DP	<i>Alpha: HLA-DPA1</i> <i>Beta: HLA-DPB1</i>	Assembly of the MHC class II loading complex. Antigen presentation.	Low expression of HLA-DPA1, and HLA-DPB1 is associated with poor prognosis in paediatric adrenocortical tumours (170).
HLA-DQ	<i>Alpha: HLA-DQA1, HLA-DQAA2</i> <i>Beta: HLA-DQB1, HLA-DQB2</i>	Assembly of the MHC class II loading complex. Antigen presentation.	HLA-DQA1*03 defined as a risk allele in lung adenocarcinoma (171).
HLA-DR	<i>Alpha: HLA-DRA</i> <i>Beta: HLA-DRB1, HLA-DRB4, HLA-DRB5</i>	Assembly/Stabilisation and loading of the MHC II loading complex; antigen presentation.	Methylated HLA-DR promoters in RFX5-negative B-lymphoma cells (169). Low expression of HLA-DRA is associated with poor prognosis in paediatric adrenocortical tumours (170). The HLA-DRB1*1601 allele is strongly associated with gastric cancer development (172).
INVARIANT CHAIN LI/CLIP	<i>CD74 (cell surface form)</i>	Assembly/Stabilisation and loading of the MHC II loading complex; antigen presentation; MHC II complex transport. Antigen presentation.	Novel somatic gene fusion, CD74-NRG1 identified in lung adenocarcinoma (173). CD74 interacts with CD44 and enhances tumorigenesis and metastasis in breast cancer (174).

CATHEPSIN S	<i>CTSS</i>	Degradation of Ii into CLIP (B cells, DCs, CD74 expressing cells); peptide generation.	CTSS expression mediates gastric cancer cell migration (175). Down-regulation of cathepsin S suppresses triple-negative breast cancer growth and metastasis (176).
CATHEPSIN L	<i>CTSL</i>	Degradation of Ii into CLIP (Thymic cortical epithelial cells); peptide generation.	Expression contributes to breast tumour angiogenesis (177).
SPPL2A	<i>SPPL2A</i>	CD74 turnover from the MIIC.	Knockout of SPPL2a results in build-up of invariant chain N-terminal fragments, impaired endosomal trafficking of MHC II molecules in B cells (178).
AEP	<i>LGMD</i>	Peptide generation.	miR-3978 inhibits gastric carcinoma and metastatic progression in a mice model by regulating legumain protein expression; confirmed in 20 human gastric carcinoma patients (179).
GILT	<i>IFI30</i>	Peptide generation.	GILT expression is required for optimal presentation of the melanoma antigen TRP1 (180). Loss of GILT expression correlates with poor survival and disease-free progression in diffuse large B cell lymphoma and breast cancer (181, 182).

Roles in cancer collected by literature search on PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>) using the following search terms (1) *Gene symbol* OR *Protein name* AND “Antigen presentation” AND “Cancer” (2) *Gene

symbol* OR *Protein name* AND “Antigen presentation” (3) (1) *Gene symbol* OR *Protein name* AND “Cancer”.

1.11 The genomics of OAC & OSCC

Several studies to date have investigated the genomic alterations which may drive the progression of OAC from BE through the Dysplasia-BE-OAC pathway, discussed in the ‘OAC evolution, disease progression and metastasis’ section. In 2013, Dulak et al. analysed the genome sequencing of 149 OAC patient samples, with 15 sample being analysed using whole genome sequencing; from this investigation, Dulak and colleagues noted several genomic features of OAC. First, they established that a the highest rate of A>C transversions in OAC was present in the noncoding areas, while within the coding region a there was a distinct overrepresentation of under-expressed genes; further investigation identified 26 frequently mutated key genes including *TP53* (72%), *ELMO1* (25%), *DOCK2* (12%), *CDKN2A* (12%), *ARID1A* (9%), *SMAD4* (8%) and *PIK3CA* (6%); these findings were further validated by a study in 551 OAC samples reporting similar driver mutation events (**Figure 11**) (183, 184).

Beyond genomic mutation, this study also identified frequent amplification of oncogenes *KRAS* (21%), *HER2* (19%), *EGFR* (16%), *CND1* (10%) and *MET* (6%), and deletion of *SMAD4* (34%), *CDKN2A* (32%) and *ARID1A* (10%) (183). Among these, *TP53* aberration is the most frequent; this a key tumour suppressor gene coding for the P53 protein, which function to arrest the cell cycle upon the detection of DNA damage, allowing enactment of DNA repair pathways or programmed cell death; this gene is frequently disrupted in many cancers because of its key role in inhibiting carcinogenesis (185). Interestingly, *ELMO1* and *DOCK2* encode dimerization partners and intracellular mediators of the Rho family GTPase, *RAC1*; the frequency of alteration of these gene suggests aberrant *RAC1* activation contributes to malignant transformation, mainly by enhancing cellular motility, supported by studies of other cancers (186-190).

Finally, the loss of *SMAD4* and *ARID1A* expression often observed in OAC cases have a significant impact on the genome. *SMAD4* acts within TFG-B signalling as the central mediator of the pathway which results in the recruitment of transcriptional factors to transactivate or repress target genes. Because of this function *SMAD4* play a significant

role as a tumour suppressor by inducing cell cycle arrest and apoptosis within the G1 phase (191, 192).

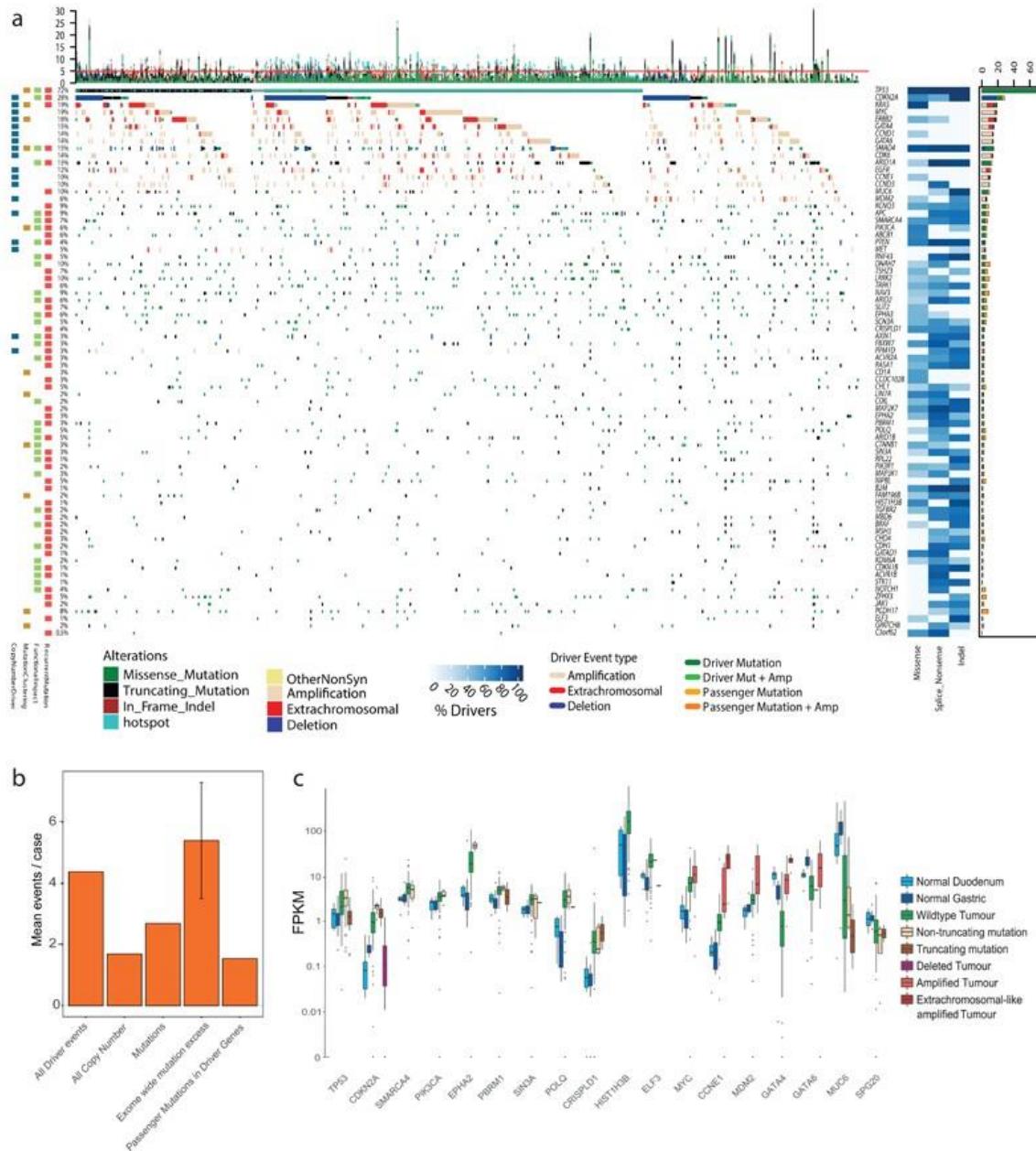


Figure 11 The driver gene landscape of OAC. (A) Driver mutations or CNVs are shown for each patient of 551 OACs. Amplification is defined as >2 copy number adjusted ploidy (2x ploidy of that case) and extrachromosomal amplification as >10 copy number adjusted ploidy (10x ploidy for that case). Driver associated features for each driver gene are displayed to the left. On the right, the percentages of different mutation and copy number changes are displayed, differentiating between driver and passenger mutations using dNdScv, and the % of predicted drivers by mutation type is shown. Above the plot are the number of driver mutations per sample with an indication of the mean (red line = 5). (B) Mean driver events per case in 551 OACs and comparison to exome-wide excess of mutations generated by dNdScv. (C) Expression changes in OAC driver genes

compared to normal intestinal tissues in RNA matched samples (n=116). Only genes with expression changes of note are shown (184).

Founding investigations of ESCC using whole-genome sequencing methodology have provided a solid basic understanding of the biology underlying the condition including the alleles predisposing to ESCC. Lin et al. conducted whole-exome and targeted deep sequencing of 139 paired ESCC cases, and analysed SCNVs from over 180 ESCC cases; from these experiments Lin et al. reported frequent mutations in *AT1*, *FAT2*, *ZNF750*, and *KMT2D* besides known commonly mutated genes (*TP53*, *PIK3CA*, and *NOTCH1*) (193).

Song et al. conducted a comprehensive genomic analysis of 158 ESCC cases (WGS of 17 cases, WES of 71, 53 WES case plus 70 cases underwent array comparative genomic hybridization analysis) as part of the International Cancer Genome Consortium research project (ICGC) (194). Using this methodology, they identified 8 significantly mutated genes including 6 known (*TP53*, *RB1*, *CDKN2A*, *PIK3CA*, *NOTCH1*, and *NFE2L2*), 1 tumour-associated gene (*ADAM29*), and mutations in *FAM135B* not previously linked to cancer (194). The *FAM135B* was mutated in 6.8% of cases (6 of 88) and was associated to poorer prognosis ($p = 0.026$, log-rank test); further analysis implicated *FAM135B* in promotion of malignancy in ESCC. Exome sequencing on the 113 tumour-control pairs identified a mean of 82 non-silent mutation per ESCC tumour, exhibited a mutational profiling resembling other squamous cell carcinoma of tissue, but was distinct from OAC's mutational profile. Mutations in genes involve in cell cycle and apoptotic regulation were mutated in 99% of the ESCC cases which underwent exome sequencing (*TP53*, 93%; *CCND1*, 33%; *CDKN2A*, 20%; *NFE2L2*, 10%; and *RB1*, 9%) (194).

Research has explored the mutational pathways of ESCC, elucidating mutation in epigenetic modulatory genes. With Song et al. identifying frequent non-silent mutations in 48 genes which function in histone modification mechanism (53.4% of ESCC cases), including *KMT2D*, *ASH1L*, *KMT2C*, *SETD1B*, *CREBBP*, and *EP300*; pathway analysis found these somatic aberrations were mainly involved in the Wnt, cell cycle and Notch signalling pathways (194). Gao et al. reported cell cycle, apoptosis and DNA damage control pathways were ubiquitously dysregulated, owing mainly to *TP53* mutation and chromatin modification pathway genes (*Notch*, *phosphoinositide 3-kinase*, and *Ras* pathways) to a lesser degree (194). Histone-modification related genes were frequently

mutated including *KMT2D* (19%), *KMT2C* 6%), *KDM6A* (7%), *EP300* (10%), and *CREBBP* (6%); also, Hippo and Notch pathways were frequently dysregulated by mutations in *FAT1*, *FAT2*, *FAT3*, or *FAT4* (27%) or *AJUBA* (JUB; 7%) and *NOTCH1*, *NOTCH2*, or *NOTCH3* (22%) or *FBXW7* (5%) (194). Notably, cases possessing mutations in *EP300* had dismal survival ($P = 0.0032$) (194).

1.12 Literature review of molecular pathway defects of APM in OAC and ESCC

OAC presents with high expression potential for neo-antigens with high affinity for HLA class I binding (Median of 20 neo-antigens expressed per sample in the mutagenic subgroup of OAC). Especially in cases with dominant T>G mutational patterns, suggesting the use of immunotherapy may be effective in these cases (195).

Despite the known high neo-antigen loading of OAC, a few studies have highlighted potential defects in MHC pathways of OAC; firstly, a study identified an increase HLA-DR expression within the inflammation to cancer progression sequence (196). Furthermore, the multivariate analysis demonstrated low expression of HLA-DR was correlated with poor survival, suggesting HLA-DR may be useful as an independent prognostic indicator in OAC patients (196). Another study identified increased levels of micro-RNAs, MIR125a-5p and MIR148a-3p in OAC cell lines, reducing the level of TAP2 and MHC I expression, which correlates to markers of adaptive immune response and overall shorter survival (164).

Two studies have indicated MHC variants may play a significant role in predisposition to Barrett's oesophagus and the risk of OAC; finding a specific MHC single nucleotide variant (rs9257809 A>G allele; 9% in Caucasian) was strongly associated with a predisposition to Barrett's oesophagus, finding an OR of 1.38 in males and 1.11 in females; interestingly, the susceptibility to OAC also increases in individuals homozygous for rs9257809 with an OR of 1.12 (Sex corrected) against heterozygous wildtype-variant with an OR of 0.79 (Sex corrected) (197, 198)

Overall, research using multi-omics datasets, into APM defects in OAC is lacking, and many core components of the MHC I and II pathways are yet to be fully investigated for defects in OAC. One aim of this research is to address this.

In ESCC, defects in APM pathways are significantly more elucidated than in OAC. Within ESCC, studies have explored the role of APM defects in risk, immune response, prognosis, biological regulation, treatment, progression, and carcinogenesis. For example, a few studies have implicated high expression of *HLA-G* in poorer patient outcomes, while another study associated a high expression of *HLA-I* and *HLA-F* with cancer risk and poorer overall survival (199-201). Other APM defects have also been associated with ESCC carcinogenesis. Interestingly, Human Papilloma Virus (HPV) infection has been associated with *HLA-DRB1* alleles relating to carcinogenesis; HPV has also been shown to regulate the methylation and expression of *HLA-DQB1* further associated with ESCC cancer risk and poor patient outcomes (202-204).

Overall, the literature investigating APM defects in ESCC is significantly more extensive than OAC, demonstrated by a basic literature search (**Figure 12**); these research papers have discovered several prognostically relevant defects, highlighting a need for further investigation in the APM defects of OAC.

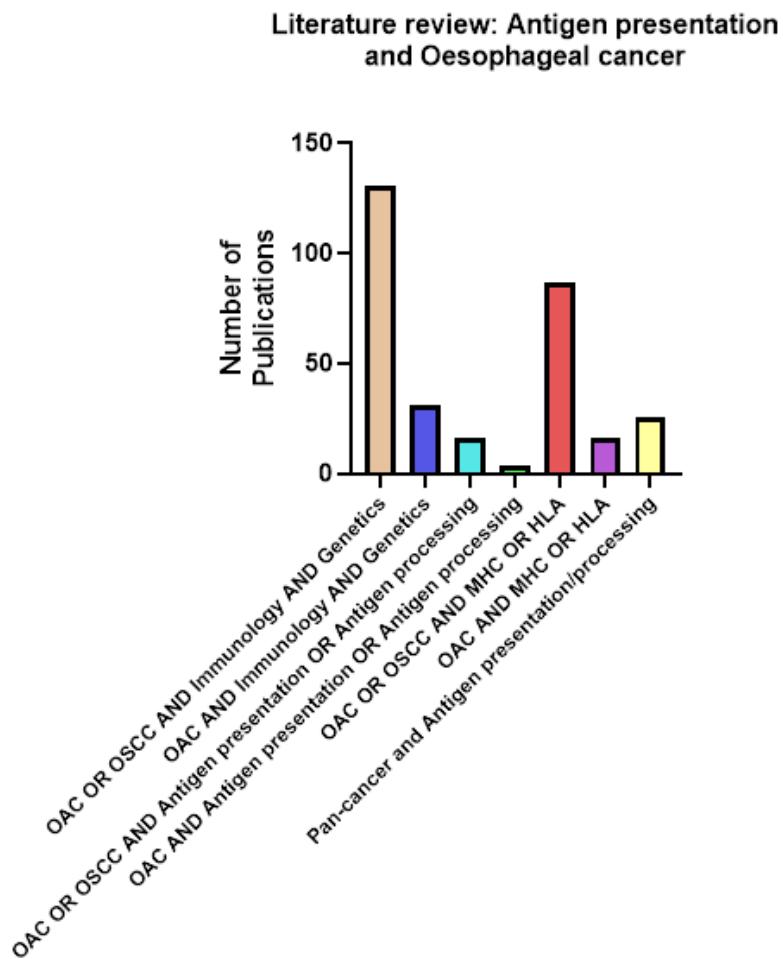


Figure 12 A bar chart representing literature search. Conducted (Date: 27/02/2023) on the PubMed database (<https://www.ncbi.nlm.nih.gov/pubmed/>) using search terms for OAC and ESCC (“Oesophageal Adenocarcinoma” OR “Oesophageal Squamous Cell Carcinoma” OR “Oesophageal cancer” AND (Antigen presentation” OR MHC OR HLA). The hits were filtered to primary research articles and reviewed individual to affirm relevance to the antigen presentation and histological types of oesophageal cancer. The findings of this literature search are displayed above.

1.13 Research hypotheses, aims and objectives.

Prior research has indicated a role of APM alterations in patient outcomes in several cancers, including gastric and OSCC; however, the current literature has not explored APM alterations in OAC and their impact on patient outcomes.

Overall hypothesis: *'The tumour micro-environment is a determinant of antigen presentation and immune response in oesophageal adenocarcinoma.'*

Thereby, forms the following aims and objectives:

Aim 1 (Results Chapter 3): Characterisation of the genetic and transcriptomic landscape of Antigen Processing Machinery (APM) and MHC I & II pathway abnormalities ('defects') in OAC and their clinical relevance within two published genomic datasets; The Cancer Genome Atlas (TCGA) and Oesophageal cancer clinical and molecular stratification (OCCAMS) programs.

Objective 1a: TCGA and ICGC datasets will be analysed to measure the frequency of APM candidate mutation, copy-number changes, and differential expression from normal oesophageal mucosa; describing known somatic APM genomic events via COSMIC, regarding the background literature and survival outcomes.

Objective 1b: TCGA and OCCAMS datasets will be analysed using maximally selected rank statistics optimal cut-offs to measure the impact of APM candidate expression on overall survival in univariate and multivariate analysis.

Objective 1c: TCGA and OCCAMS datasets will be analysed using correlation analysis to determine the key regulatory factors associated to APM candidate expression.

Aim 2 (Results Chapter 4): Investigate immune infiltrate and activity in OAC and/ or recurrent molecular defects in APM/ MHC I & II pathways by digital cytometry utilising deconvolution analysis of bulk transcriptomic data.

Objective 2a: Characterise the immune cell subpopulations in OAC with our cohorts using CIBERSORT deconvolution and compare to known published immune populations in OAC.

Objective 2b: Assess the impact of antigen processing machinery defects identified by the chapter 3 datamining analysis on the immune cell subpopulations our OAC cohorts (TCGA and OCCAMS).

Objective 2c: Determine the prognostic value of antigen processing machinery defects in combination with immune cell distributions.

Aim 3 (Results Chapter 5): Validate the clinical significance of APM, MHC I & II expression and immune cell infiltrate in oesophageal adenocarcinoma by immunohistochemistry.

Objective 3a: Immune and APM protein staining data on a Southampton OAC TMA series, will be analysed to measure the level OAC T cell infiltrate in tumour cores, determining the variance in marker expression and to validate my observations from Aims 1 & 2.

Objective 3b: APM staining scores and immune density data will be collated to identify significant correlations between APM protein expression and T cell density.

Objective 3c: APM staining scores and immune density data will be analysed with maximally selected rank statistics optimal cut-offs to measure the impact of APM protein expression on patient outcomes (constructing Univariate and Multivariate models for Overall survival, disease-free survival, and cancer-specific survival and important clinical co-variates).

Aim 4 (Results Chapter 6): Model the regulatory role of *CSDE1* expression on prognostically significant APM genes in OAC cancer cell lines by transfection knockdown/overexpression of *CSDE1*.

Objective 4a: Knockdown/overexpression of *CSDE1* will be performed via siRNA and validated at the mRNA and protein level in OAC cell lines.

Objective 4b: MHC class I mRNA and protein expression levels will be measured in knockdown/overexpression *CSDE1* cell models, determining the effect of *CSDE1* expression on MHC class I expression.

Objective 4c: phosphorylation status of STAT1 protein levels will be measured in OAC cell lines to determine the effect of altered *CSDE1* expression on the activation of the JAK/STAT signalling pathway.

Objective 4d: *CSDE1* mRNA expression will be repartitioned into single cell populations from single-cell RNA sequencing data generated from primary human tumour tissue to determine whether *CSDE1* is over expressed in OAC cancer cells compared to normal cells.

Chapter 2 General Materials and Methods

2.1 Download and filtering TCGA data.

Data was selected and downloaded directly from the TCGA repository (<https://portal.gdc.cancer.gov/>) using the TCGAbiolinks R package to specifically access the TCGA-ESCA project for mutation, copy-number variation and expression counts data; the data was filtered to cases of OAC or OSCC by querying subtypes acquire patient barcodes for each subtype (205). Data was queried and downloaded as a summarised experiment with attached relevant clinical data. File formats for downloaded TCGA data were: mutation as MAF format, expression counts as summarised experiment, copy-number variation as summarised experiment.

2.1.1 Clinical

Clinical data was quality controlled to determine the completeness of the data, observing patient follow-up data, vital status, staging, sex, age, co-morbidity (Weight, smoker status, alcohol history), grading/differentiation status; cases without this required data was excluded from analysis.

2.1.2 Mutation

Mutation data was first quality controlled to remove duplicated mutations and filtered to the cases of OAC using the patient barcodes from the filtered clinical data. Further quality inspection compared the known mutational landscape of OAC from literature to a summarised MAF using a TiTv plot which classifies Single Nucleotide Variants into Transitions and Transversions and top mutated genes. Mutation data was then processed using MAFtools into subsetted MAF files comprised of the MAF data for our gene candidate list and known driver genes in OAC for ease of handling. Using the summarised MAF data the mutations per sample was extracted to highlight hypermutated cases. Mutation data was visualised using Oncoplot for our subsetted MAF files and

annotated with the mutation per sample per patient column extracted from the earlier summarise MAF.

2.1.3 Copy-number

Copy-number variation was quality controlled by TCGAs SNP 6.0 pipeline quality control metrics at two stages using signal/noise ratio and number of segments, samples with outlier values were excluded by the TCGA study. Further quality control was conducted by comparing the summarised copy-number calls to known copy number alterations in OAC. Copy number data was then paired with clinical annotations for stage, BMI category, smoking, and alcohol history, ethnicity, and gender. Visualisation of copy number data was conducted using Oncoprints for absolute copy number calls with annotated clinical features mention above. Further visualisation focused on viewing the copy number segment data using IGV viewer to observe the complexity of copy-number segments on chromosome 6.

2.1.4 RNA Expression

Expression data was quality controlled by visualising library sizes, by filtering out non-expressed genes, visualising count distributions using DEseq2 both before and after TMM (Trimmed Mean of M) normalisation of the RNA-seq counts. Expression data was processed as a summarised experiment using EdgeR to normalise the samples by TMM. Next the TMM normalised counts were processed by filtering the count to our candidate list and known drivers, these filtered counts were processed for clustering and PCA analysis; expression data was combined with GTex normal oesophageal mucosa samples and normalised to allow for differential expression analysis visualised by volcano plots. Further processing mated the clinical data to the TMM counts to allow for survival analysis using survminer() R program and maximally Selected Rank Statistics (206). Visualisation was conducted using heatmaps to visualise the clustering of expression of APM genes across the cohort; survival data was visualised in forest plot for CoxPH analysis and Kaplan-Meier plots for survival analysis. Scripts for data mining analysis of TCGA data are made available on GitHub (<https://github.com/wp1g19/OAC-Thesis-Antigen-processing-machinery-and-the-immune-microenvironment>).

2.2 Download and filtering ICGC data.

2.2.1 Download

ICGC ESAD project data (release 28) was directly downloaded from the ICGC portal (<https://dcc.icgc.org/>). Data downloaded included simple somatic mutation files, copy number somatic mutation files and donor clinical files.

- Clinical

Patient clinical was quality controlled for complete clinical data on follow up times, vital times, staging, gender, age, smoking, and alcohol history.

- Mutation

Mutation data first processed from the simple somatic mutation file format to the MAF format using MAFTools on iridis 4 (script available at <https://github.com/wp1g19/OAC-Thesis-Antigen-processing-machinery-and-the-immune-microenvironment>). After MAF format conversion the MAF file was quality controlled to remove duplicate mutations and summarised for quality inspection comparing the known mutational landscape of OAC from literature to a summarised MAF using a TiTv plot which classifies Single Nucleotide Variants into Transitions and Transversions and top mutated genes. Further processing extracted the mutation count per sample from the summarised MAF file. Mutation data was visualised into an Oncoplot in combination with copy-number variation data and annotated with total mutation count per patient to identify hypermutable cases. Copy-number quality control of copy-number data was conducted by removing duplicated absolute copy number calls and by comparing the summarised copy-number calls to known copy number alterations in OAC. Copy-number data as copy number somatic mutation files were collated with the MAF mutation file using MAFTools then visualised by Oncoplot with annotated total mutation count per patient sample.

2.3 Download and filtering OCCAMS data.

OCCAMS expression RNA-seq data was retrieved from the Southampton Underwood group filestore in RNA count file format, access to this data can be obtained directly from ICGC under a DACO application (<https://daco.icgc.org/>). Clinical data was also obtained from the Underwood group filestore but can be openly accessed via the ICGC portal.

Firstly, clinical data was filtered for tumour samples with complete clinical data on follow up times, vital status, staging, age, gender, smoking, and alcohol history. RNA-seq count files from OCCAMs were collated together using R cbind() functions and filtered to the quality controlled clinical data. Expression data was quality controlled by visualising library sizes, by filtering out non-expressed genes, visualising count distributions using DEseq2 both before and after TMM (Trimmed Mean of M) normalisation of the RNA-seq counts.

- Clinical

OCCAMS Patient clinical obtained via the Underwood laboratory then quality controlled for complete clinical data on follow up times, vital times, staging, gender, age, smoking, and alcohol history.

- RNA Expression

RNA-seq expression data from OCCAMS was combined with TCGA expression data after TMM normalisation using Combat-seq was processed using PCa analysis and hierachal clustering with Euclidean distance which was visualised by heatmaps (207). Further processing combined OCCAMS/TCGA expression data with GTex normal oesophageal mucosa samples for differential expression analysis using EdgeR. For survival analysis the clinical data was integrated into the normalised counts and analysed using CoxPH analysis besides survival analysis with maximally Selected Rank Statistics optimal cut points visualised by forest plots and Kaplan-Meier plots, respectively. Multivariate analysis was conducted using a CoxPH model with known prognostic clinical factors ascertained from the 551 multivariate model (differentiation status, sex, age, and treatment response).

2.4 Survival analysis

Survival analysis was conducted using R (4.2.1) employing Maximally ranked statistics from the Maxstat package (determines the lowest p value comparing survival between two groups with minimum proportions to maintain power) before conducting CoxPH

univariate and multivariate survival using the `finalfit()` package in R (208, 209). This approach was employed to formulate:

- Overall survival (OS) time (Months) which calculates the survival time of a patient within the cohort.
- Cancer-specific survival (CSS) which excludes the impact of death events not related to the cancer such as surgical complications.
- Disease-free survival (DFS) which details the length of time between treatment and disease recurrence/progression.

Multivariate CoxPH analysis clinical model is constructed using key clinical features including TNM staging, Age and Sex as used in prior studies employing the backwards elimination model (210).

Chapter 3 Survival differences via expression of curated APM gene candidates was determined using the landscape of genomic defects of antigen presentation machinery genes in OAC.

3.1 Introduction

Exploring the published literature identified a significant knowledge gap surrounding the genomic landscape of APM defects in OAC and their association to altered TIME immune distributions and clinical outcomes. To address the knowledge gap established in my literature review I set an experimental hypothesis: '*Genomic defects in antigen processing machinery of oesophageal adenocarcinoma are dysregulated and associate with clinical outcomes.*' To address this hypothesis, I selected two large OAC genomic datasets (TCGA and ICGC/OCCAMS) to datamined the landscape of APM genomic was performed to ascertain the incidence of genomic events in my APM gene candidates and their respective impact on overall survival.

Current literature on the landscape of APM component defects in OAC is severely lacking compared to other cancers, this leaves a knowledge gap on the details of APM components and their associations to the TIME immune cell distributions and overall survival. Using external references which demonstrate APM defects impact the TIME, this study hypothesises that 'The tumour micro-environment is a determinant of antigen presentation and immune response in oesophageal adenocarcinoma.' To address our primary hypothesis, we will first data mine publicly available datasets to discover the landscape of APM defects and their associations to the TIME immune cell distributions and overall survival.

3.1.1 Antigen presentation and known pathway defects in OAC.

Antigen presentation is a key function within host immunity for the recognition of cancer cells as T cells may act to destroy infected host cells through the specific recognition of antigens presented on antigen presentation molecules on the cell surface. Upon specific recognition and with the co-stimulatory signal, the T cell may activate and direct cytotoxicity towards the infected cell, assisting in the clearance of infection. Because of

its significant role in immune surveillance and its potential for anti-tumoral immunity, antigen presentation pathway components are often somatically targeted by cancer to evade immune destruction, these immune evasion mechanisms are discussed below categorised by pathway and function in context to OAC and other cancers.

3.1.2 MHC I endogenous pathway

This pathway is crucial for T cell recognition of cancer, presenting endogenous peptides to CD8+ T lymphocytes for recognition of cancer antigens which results in T cell activation, expansion and anti-tumoral immune response. This pathway is composed of several key components which can be divided into differing functions allowing the process to present peptides on HLA molecules to CD8+ T lymphocytes.

Firstly, several genes participate in the assembly of the MHC I loading complex, these include the major HLA genes (*HLA-A/B/C/E/G*) as well as *CALR*, *CANX*, *B2M*, *ERp57*, *ERAP1*, *ERAP2*, *TAPBPL* and *TAPBP*. Genes involved in the assembly of the MHC I loading complex are important for transport of the HLA molecule to the TAP transporter in the endoplasmic reticulum preparing the molecule for peptide binding and eventual travel to the cell surface via the Golgi Apparatus.

Defects (expression, mutation, copy number alteration) among genes required for MHC I loading complex assembly can significantly impact the pathway impeding peptide loading to the HLA peptide binding site. Defects in this subgroup of MHC I genes in OAC have not been explored in the literature, with a singular reference to B2M as a novel driver gene in OAC in the 551 OCCAMS study. This mutation was specifically associated with hypermutated cases noting a three-way association among hypermutation, Wnt activation and loss of immune-signalling genes including B2M which has been prior linked to immune escape in colorectal cancer. This suggests an acquired immune evasion mechanism to prevent immune surveillance for hypermutated tumours (176, 198-200). The 551 OCCAMS study findings suggest our data mining should aim to explore the prevalence of hypermutable cases with B2M mutations as this could be a recurrent immune evasion mechanism implemented by OAC tumours.

Other cancers have identified defects in MHC I genes involved in MHC I loading complex assembly. For example, TAPBP downregulation has been found in OSCC cell lines and triple-negative breast cancer, with the impact on the immune microenvironment

demonstrated in colorectal cancer indicating loss of Tapasin correlates with diminished CD8+ T cell populations and disease progression (140-142).

Interestingly, B2M mutation defects has been identified outside of OAC with a pan-cancer study using The Cancer Genome Atlas (TCGA) identified *B2M* mutations reduced the overall level of cell surface MHC I molecules and reduced levels of cytotoxicity; *B2M* mutation demonstrates defects in APM components which function to assemble the MHC I loading complex may act to reduce antigen presentation of cancer antigens via impedance of MHC I loading complex assembly, preventing the surface expression of MHC I antigen presentation molecules.

The next category of MHC I genes are classified as genes functioning in peptide transport into the endoplasmic reticulum lumen. This category encompasses two key genes *TAP1* and *TAP2*, these genes functions together to form the Transporter associated with antigen processing protein complex which acts to transport cytosolic endogenous peptides into the lumen of the endoplasmic reticulum in proximity to the MHC I loading complex. Defects in *TAP1* have been noted in OAC with high expression of *TAP1* correlating to significantly shorter survival times in a study of 51 patients (189). *TAP2* has also been noted to possess clinical prognostic value in OAC with reduced expression of *TAP2* due to increased MIR125a-5p and MIR148a-3p expression associated to poor patients' outcomes (189). The findings from the existing literature suggests that the RNA expression of *TAP1* and *TAP2* should be explored to investigate the potential prognostic value in our large datasets.

The final sub-category of MHC I genes are defined by their function in peptide generation and trimming, these include *ERAP1*, *ERAP2*, *PSMB9*, *PSMB10*, *PSMB8*; these genes in prior publications have demonstrated a significant impact on the peptide repertoire available for antigen presentation, thereby they may be somatically targeted to reduce cancer antigen presentation. There is a lack of knowledge of the impact of defects within this sub-category in OAC, conversely, in other cancers these genes have been associated to prognostic value for overall survival. For example, *ERAP2* loss of expression was identified in liver carcinomas and was associated to lack of MHC I surface molecules, whereas *PSMB9* over-expression in melanoma patients has been correlated to greater survival and improved response to immune-checkpoint inhibitors (157, 201). Combined, these findings suggest exploring the expression profile of these genes involved in peptide generation may have prognostic value in OAC.

Overall, the MHC I endogenous pathway is somatically downregulated in other cancers (0-93%), however, there is a significant knowledge gap in the prevalence and prognostic values of MHC I APM defects in OAC, forming a research goal to identify the incidence of defects previously identified in OAC and other cancers and assess them for clinical prognostic value (211).

3.1.3 MHC II pathway

This pathway is also known as the exogenous pathway is characterised by the display of exogenous peptides on MHC II molecules to CD4+ T cells. MHC II pathways are primarily expressed by professional antigen presenting cells, such as dendritic cells and macrophages, as well as some endothelial cells, thymic epithelial cells, and B cells. Targeting cancer cells via MHC II is less direct than the MHC I pathway, however, a known mechanism which professional antigen presentation cells (APCs) may help target cancer cells occurs via APCs engulfing peptides from fragments of dead cancer cells or secreted tumour specific antigens (202). MHC II components can be divided into sub-category based on their similar functions which will be explored below.

Firstly, genes which function solely in the assembly, stabilisation and loading of the MHC II loading complex, which includes *HLA-DR*, *HLA-DQ*, *HLA-DP*, *CD74*, *HLA-DM* and *HLA-DO*. Exploration of the literature identified several defects possessing prognostic value in other cancers yet yielded no results on the prognostic value of these defects in OAC. An example of these defects is found in paediatric adrenocortical tumours where low expression *HLA-DR* and *HLA-DP* results in poor prognosis (162). A further example is found in *HLA-DM* which is found under-expressed in B-Cell lymphoma and head and neck squamous cell carcinoma cell which showed impeded immune recognition in B-cell lymphomas (159, 160).

The second sub-category includes genes which function in peptide generation, including *LGMDN*, *IFI30*, *CTSS* and *CTSL*. The prognostic role of these genes in cancer immunity is more ambiguous because of the multiple functional roles these genes. For example, *IFI30* expression was associated with enhanced leucocyte mediated immune and inflammatory response in Glioblastoma. However, within the same study *IFI30* was found to activate IL6-STAT6 signal pathway which has been linked to increased migration and invasion and metastasis in colorectal cancers (203). Another example is found is *GILT* expression which was crucial for optimal presentation of the melanoma antigen TRP1 in one study. In

another absence of *GILT* was associated to poor patient survival in breast cancer patients. However, another dual function arises as *GILT* also performs proliferation inhibiting functions with the absence of *GILT* was positively correlated with adverse characteristics of breast cancers, such as histological type, tumour size, lymph nodes status, and pTNM stage (174).

These genes highlight the need for quality control and cautious evaluation of prognostic value, as the impact on clinical outcome may not strictly relate to defective antigen presentation.

Overall, the impact of MHC II defects in cancer in the literature appear to be less impactful on clinical outcomes and more infrequent, with cancer mostly somatically targeting candidates in the MHC I pathway. MHC II defects have demonstrated impact in other cancers yet the impact of said defects in OAC is currently unknown forming a key knowledge gap which this study aims to address.

3.1.4 Alternative processing of cancer antigens

Antigen presentation mechanisms are not strictly restricted to the MHC I endogenous and MHC II exogenous pathways, with alternative processing pathways which may allow for cancer antigen recognition. One alternative pathway involves the *MR1* gene which has been demonstrated to be capable of presenting cancer antigens to MR1 restricted T cells able to recognise pan-cancer (204). Despite the keen interest in MR1-restricted T cell immunotherapies the expression of *MR1* in cancers especially in OAC is not documented. Therefore, understanding the landscape of *MR1* defects plus expression may function as both a prognostic and stratification tool for determining ideal candidates for MR1-restricted T cell therapies in future.

An additional antigen presentation pathway centres around the *CD1* family of genes which present lipid molecules as opposed to the protein peptide fragments presented by the MHC I and II pathways. Interestingly, natural killer cells (NKT) are CD1D-restricted cells capable of recognising cancer-specific lipids bound to CD1D molecules, and alter the TIME by killing TAMs and altering the activity of effects of CD1d+ myeloid-derived suppressor cells (MDSCs) mediated immune suppression preventing immune suppression (205-207).

Alternative antigen processing pathways have demonstrated impact in the literature from current NKT cell responses reshaping the immunosuppressive microenvironment via lipid antigen recognition to a potential pan-cancer therapy using MR1-restricted T cell therapies. From the literature the expression and genomic defects of both *CD1* family and *MR1* genes will be explored as they may possess impact on clinical outcomes in OAC, as they have demonstrated in other cancers.

3.1.5 **Hypothesis and research objectives**

The literature search (**in Section 2.2**) identified a specific knowledge gap in the landmark of APM gene defects in OAC by expression, mutation, and copy-number, whereas the in OSCC several published papers were found in HLA molecules (**See Figure 12**). Furthermore, the few publications focusing on APM defects in OAC did not focus on their prognostic value and utility, leaving specific knowledge gaps which this chapter will explore to address the following hypothesis, aims and objectives.

Hypothesis: '*Antigen processing machinery of oesophageal adenocarcinoma are dysregulated and associate with clinical outcomes.*'

Aim 1: Characterisation of the genetic and transcriptomic landscape of Antigen Processing Machinery (APM) and MHC I & II pathway abnormalities ('defects') in OAC and their clinical relevance within two published genomic datasets; The Cancer Genome Atlas (TCGA) and Oesophageal cancer clinical and molecular stratification (OCCAMS) programs.

Objective 1a: TCGA and ICGC datasets will be analysed to measure the frequency of APM candidate mutation, copy-number changes, and differential expression from normal oesophageal mucosa; describing known somatic APM genomic events via COSMIC, regarding the background literature and survival outcomes.

Objective 1b: TCGA and OCCAMS datasets will be analysed using maximally selected rank statistics optimal cut-offs to measure the impact of APM candidate expression on overall survival in univariate and multivariate analysis.

Objective 1c: TCGA and OCCAMS datasets will be analysed using correlation analysis to determine the key regulatory factors associated to APM candidate expression.

3.2 Methodology

3.2.1 Datasets

To address the assessment of the incidence and clinical prognostic value of defects in APM genes of OAC large genomic datasets were employed to data mine the genomic information to determine the incidence of defects within a cohort of OAC patients. Datasets were selected based on availability, cohort size and quality of clinical data, from this criterion two OAC datasets, TCGA and ICGC/OCCAMS and normal stomach tissue dataset, GTEx was selected.

The TCGA project is a publicly available pan-cancer data repository which has molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types, generating genomic, epigenomic, transcriptomic, and proteomic data (212). For our interests, the TCGA-ESCA dataset will allow us to explore OAC and OSCC mutation (MAF: Mutect2), copy-number variation, methylation (Affymetrix array) and mRNA expression (Raw counts) data from 187 patients (81 OAC; 90 OSCC) .

The International Cancer Genome Consortium (ICGC) is also a pan-cancer data repository which has collated clinical and molecular data from over 100,000 cancer patients participating in therapeutic clinical trials from 84 worldwide cancer projects, including OCCAMs project (184, 213). ICGC data is partially available to the public with mutation and absolute copy number calls being open access, however, access to expression RNA-seq data is closed access and requires application through DACO (<https://daco.icgc.org/>). For this study we will access the ESAD-UK dataset within the ICGC project comprised of 409 OAC patient samples from UK sites, matching our available RNA-seq expression data from 152 OCCAMS patient samples.

GTEx standing for Genotype-Tissue Expression project, collects post-mortem gene RNA-seq expression data from multiple regions of the human body, including organs such as heart, lung, liver, brain, and among others. From the analysis presented here, GTEx stomach tissue comprising 324 normal stomach tissue samples with mRNA sequencing

data as raw counts (Open Access available from <https://gtexportal.org/home/tissue/Stomach?tissueSelect=Stomach>).

3.2.2 Power and sample size

I used power and sample size calculator (<http://powerandsamplesize.com/Calculators/>) to assess the power of discovery in the cohorts using a prior publication exploring TCGA pan cancer power for mutation (214). The datasets were quality controls for complete clinical characteristics of age, sex and pTNM staging data, then combined in R to achieve power of discovery for mutation, copy-number, and expression data.

3.2.3 Mutation/copy number analysis

Mutation data from TCGA in Mutation Annotation Format (MAF) format was queried and downloaded into an R environment using TCGAbiolinks, then read imported into MAFtools (215). Meanwhile, ICGC ESDA project simple-somatic-mutation data was downloaded from the ICGC portal (Release 28) and converted into MAF format before merging the two OAC dataset MAF files with MAFtools. Mutation frequency was visualised using Oncoplot whilst mutation burden of OAC samples with/without APM mutations was exported to GraphPad prism from visualisation and statistic comparison using Mann-Whitney U tests.

3.2.4 Differential expression analysis

To achieve differential expression analysis (DEA), RNA-seq data was obtained via an additional cohort of normal gastro-oesophageal junction samples (n=375) from the GTEx dataset within the Recount2 project. The normal RNA-seq data was imported into R using the Recount package, batch corrected to the TCGA/OCCAMS RNA-seq data using ComBat in TMM format; batch corrected RNA-seq data was used for DEA analysis using the DESeq2 package and visualised by EnhancedVolcano (207, 216-219). Batch corrected OAC and OSCC data were also used in a DEA using the methods above.

3.2.5 Survival analysis (mRNA)

mRNA expression counts from TCGA and OCCAMS were converted to TMM then batch corrected using ComBat (selected after comparing batch correction between limma and ComBat via PCA), before mating with clinical data complete for Overall survival time

(Months), survival status, cancer-specific survival, recurrence status, recurrence time (Months), TNM staging, Age and Sex (207). Survival differences via expression of curated APM gene candidates was determined using Maximally ranked statistics from the Maxstat package (determines the lowest p value comparing survival between two groups with minimum proportions to maintain power) before conducting CoxPH univariate and multivariate survival in a backwards elimination model via the finalfit() package in R (208, 209).

3.2.6 Methylation analysis.

Methylation data was only obtainable from the TCGA OAC dataset, this data was obtained using the Shiny Methylation Analysis Resource Tool (SMART) (Available from: <http://www.bioinfo-zs.com/smартapp/>) for the entire TCGA-ESCA dataset; this tool was employed because of its ability to integrate multi-omics and clinical data with DNA methylation while avoiding the computation and time cost in aligning methylation data from TCGA (220). Methylation analysis target genes were selected from APM candidate genes and APM gene expression regulators possessing clinical survival associations at the mRNA expression level. CpGs were selected on a per gene basis for CpG with significantly different methylation compared to normal oesophageal tissue and with relevance to their respective gene's expression (i.e., CpG island within a proximal promoter of the gene of interest).

Methylation of each gene was visualised in GraphPad prism 10 (<https://www.graphpad.com/features>) comparing OAC to normal tissue. Additionally, differentially methylated CpGs within the HLA locus of chromosome 6 were queried in SMART and presented in table form.

3.2.7 Statistical analysis

Chi-squared and Anova were performed in Graphpad prism 10 to compare the cohorts characteristics including Age, sex and pTNM staging to determine whether cohorts possess any significant differences between the datasets implemented in the analysis. T tests were implemented using Graphpad prism 10 to compare the mutation rate between APM mutated and non-APM mutated groups, to compare the methylation values for APM genes.

3.3 Results

3.3.1 Description of the study cohorts.

Firstly, I explored the cohort characteristics for my analysis across age, sex and pTNM clinical features. These clinical features were selected to describe the cohort based upon matching available characteristics across the different projects and their associated value within clinical practice. Cohorts were combined to provide the greatest sample size dependant on analysis. These combined cohorts included TCGA/OCCAMS for mRNA gene expression analysis, TCGA/ICGC employed for mutation/copy number analysis (**Table 6**). Additionally, individual cohorts were assessed for each analysis type (expression, copy-number, mutation and methylation, **Table 6**).

For age, an average of 65-67 years were present across the cohorts and possessed a wide range at their extremes of 27-87 years, using an unpaired t-test confirmed there was no statistical difference in age among in comparing the cohorts and combined cohorts ($p = 0.65$) (**Table 6**).

The youngest patient among the cohorts was 27 years old present in the TCGA-ESCA dataset and was of interest as OAC presents as a disease of age. This specific case presents as a 27-year-old male diagnosed with Stage IIIA OAC which received curative chemoradiotherapy, however, exploring the new tumour event data identified a distant metastasis in the liver 160 days into treatment.

To address the early onset of disease comparative to the rest of my cohort and external cohort I observed the patient possessed a history of reflux disease and was diagnosed prior with Barrett's oesophagus which may indicate the reason for the early onset of OAC.

Sex in the incidence of OAC has prior been reported as a male dominant disease and my cohorts exhibited this similar male dominance, among the OAC cohort the Male: Female ratio varied minorly between the OAC cohorts from 79.4-87%:13-20.6% (Male: Female), using a Chi-squared analysis these minor difference were determined to be non-significant ($p = 0.96:0.92$; Male: Female) (221) (**Table 6**).

OAC tumours are most frequently diagnosed in pT 3, denoted in several studies, my OAC cohorts confirmed this trend with the pT2-4 being the most frequent reported diagnosis (pT2-4: 49.5-82.9%) compared to the reported diagnosis of pT0-1 (pT0-1: 5.1-24.6%) and unknown pT-stage (pTNM: 12-30.9%), among the three pT stage categories both pT0-1

and pT2-4 approached a significant difference ($p=0.086;0.076$) but were not significantly different according to Chi-Squared analysis (222) (**Table 6**).

OAC tumours are often diagnosed with positive regional nodal spread, according to the literature, within my OAC cohorts tumour positive lymph nodes were present in majority of cases with a positive frequency (pN1-3) ranging from 40.2-63.2%, negative nodal involved case frequency ranging from 23.2-29.9% and unknown lymph node involved cases ranging from 9-26.1%, no statistically significant differences between the cohorts was observed in Chi-Squared analysis (pN0 $p=0.99$; pN1-3 $p=0.33$; pNNA $p=0.14$) (222) (**Table 6**).

OAC is often diagnosed in the later stages often resulting in a relatively high metastasis frequency of ~30% compared to other diseases. Among my selected OAC cohorts the rate of metastasis is infrequent compared to external studies (pM1:0.1-7.3%), however the frequency of non-metastasised OAC cases remains like external studies in three cohorts (pM0 34.7-77%) (**Table 6**). The low non-metastasized OAC case frequency is particularly observed in the OCCAMS cohort (pM0:37.4%) also impacting the TCGA/OCCAMS combined cohort, these differences in metastasis rates among the OAC cohorts was determine as statistically significant by Chi-Squared analysis in pM0 and pM-NA reported cases (pM0: $p=0.0045$; pMNA: $p=0.0008$), this suggests a potential input error in the OCCAMS dataset for metastasis at pathological diagnosis (**Table 6**).

Table 6 Cohort characteristics across TCGA, OCCAMS and ICGC, plus combinations of TCGA/OCCAMS and TCGA/ICGC.

PARAMETER	TOTAL/%	TOTAL/%	TOTAL/%	TOTAL/%	TOTAL/%	CHI-SQUARE
	(TCGA) N = 69	(OCCAMS) N = 107	(ICGC) N = 409	(COMBINED) TCGA/OCCAMS N = 176	(COMBINED) TCGA/ICGC N = 478	(X ²)/ANOVA P VALUE
AGE (YEARS)	67.2 (Range 28-87)	66.5 (Range 45.1-87.7)	65.9 (Range 36-87)	66.8 (Range 27-87.7)	66.6 (Range 27-87)	0.65
SEX						
MALE	60 (87%)	85 (79.4%)	351	145 (82.4%)	411 (86%)	0.26 (0.99)
FEMALE	9 (13%)	22 (20.6%)	(85.8%) 58 (14.2%)	31 (17.6%)	67 (14%)	0.96 (0.92)
PT						
0-1	17 (24.6%)	21 (19.6%)	21 (5.1%)	38 (21.6%)	38 (7.9%)	8.2 (0.086)
2-4		53 (49.5%)	339	91 (51.7%)	377 (78.9%)	8.5 (0.076)
NA	38 (55.1%)	33 (30.9%)	(82.9%) 49 (12%)	47 (26.7%)	63 (13.2%)	5.0 (0.29)
	14 (20.3%)					
PN						
0	16 (23.2%)	32 (29.9%)	109 (26.7%)	48 (27.3%)	125 (26.2%)	0.36 (0.99)
1-3		43 (40.2%)		82 (46.6%)	302 (63.2%)	4.6 (0.33)
NA	39 (56.5%)	32 (29.9%)	263 (64.3%)	46 (26.1%)	51 (10.6%)	7.0 (0.14)
	14 (20.3%)		37 (9%)			
PM						
0	37 (53.6%)	40 (37.4%)	331	77 (43.8%)	368 (77%)	15.0
1	5 (7.3%)	1 (0.1%)	(80.9%)	6 (3.4%)	28 (5.9%)	(0.0045)
NA	27 (39.1%)	66 (61.7%)	23 (5.6%)	93 (52.8%)	82 (17.1%)	5.3 (0.25)
			55 (13.5%)			19.0 (0.0008)

Sex, pT, pN and pM data categorized into discrete values and compared between cohorts using χ^2 analysis. Age data compared between cohort using ANOVA. Difference in N number in combined cohorts compared to individual cohort N numbers is due to non-matching clinical features.

3.3.2 APM genomic aberrations.

An extremely low incidence of APM mutation occurred within my OAC cohort (ICGC-ESDA and TCGA-ESCA combined) with 75/502 cases containing an APM mutation across all 46 candidate APM genes (See **Figure 13**). The highest incidence of mutation occurring in CD1A/D/C at 2-3% of the entire cohort, though mostly comprised as missense mutations which were not previously reported in the cosmic database (**Figure 13**). Although the noted low mutation rate of APM gene candidates in the combined TCGA-ESCA/ICGC dataset, several mutations may have a significant impact on the formation of APM component proteins required for antigen presentation. Additionally, the mutational burden of OAC tumour samples was significantly greater in samples containing an APM mutation than those without (**See Figure 14**).

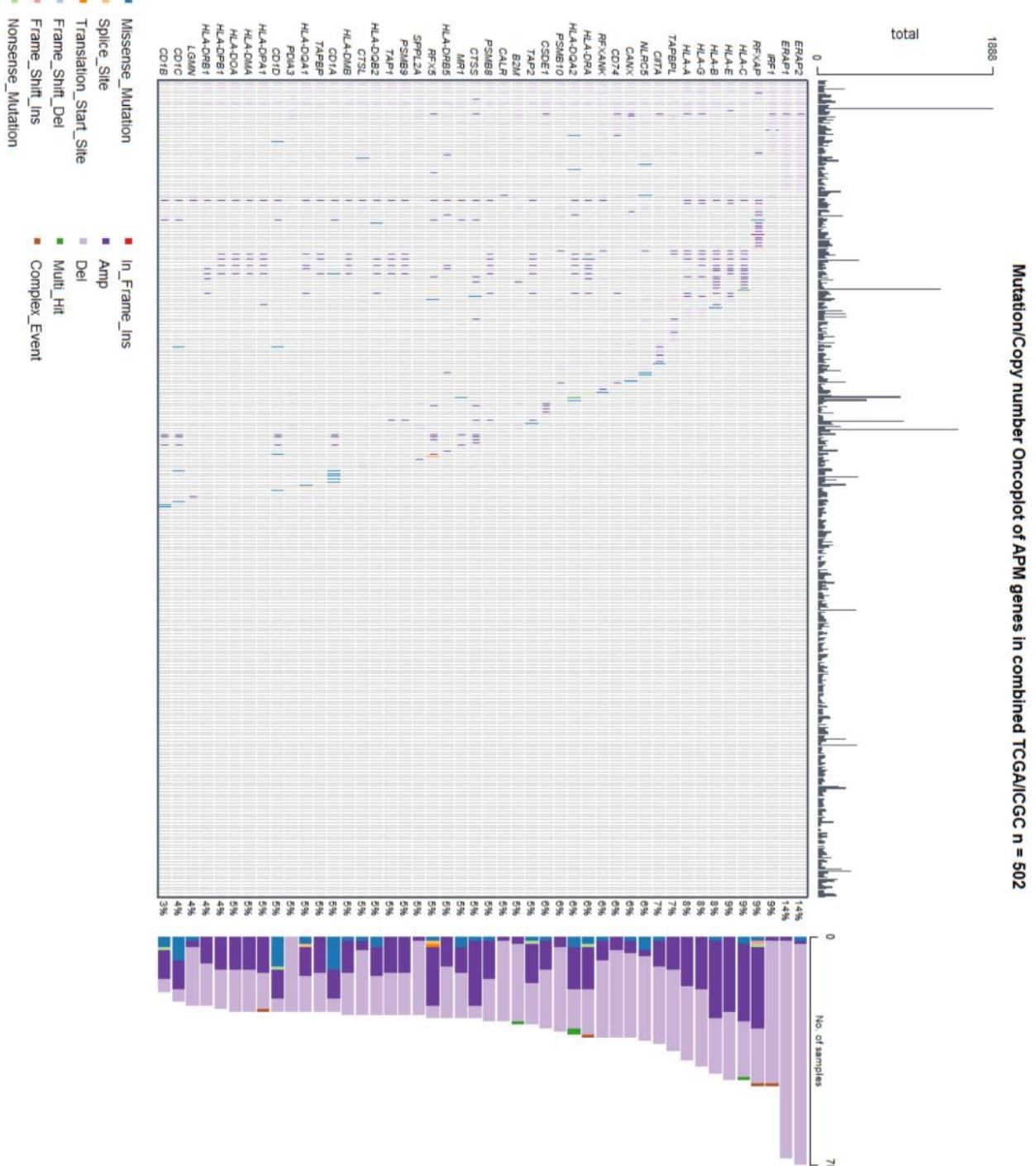


Figure 13 : Mutation/Copy number Oncoplot for OAC in the combined TCGA-ESCA/ICGC-ESAD cohort.

Depicting the frequency of mutation and type of mutation with the APM candidate genes is paired with a mutation count above each sample. Mutations plotted to patients (columns) and genes (rows) with mutation rate display left of each Oncoplot in percentages, total mutation count of all genes annotated above each patient sample as a bar chart labelled Mutcount.

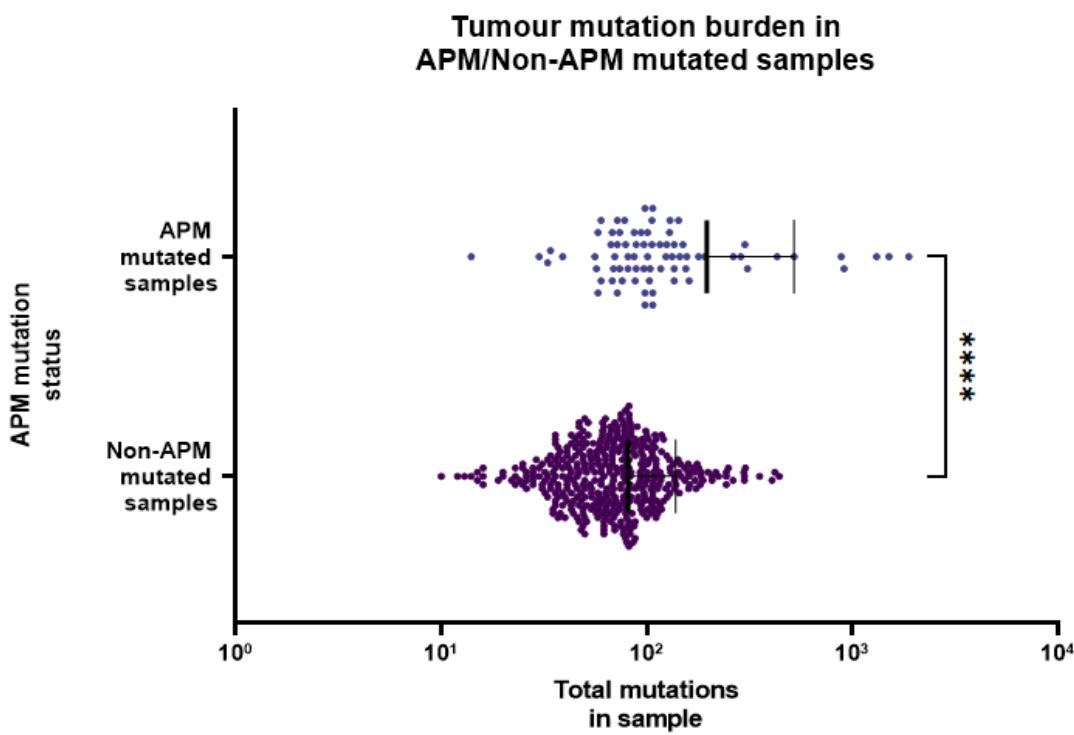


Figure 14 Mutation burden between samples possessing a mutation in APM genes and those without, statistical comparison via Mann-Whitney U test ($p < 0.0001$).

B2M mutations ($n=4$) found in two OAC patients may have a significant impact on the *B2M* protein and its function. As displayed in **Figure 15 and Table 7**, *B2M* mutations within one sample (TUMOR SAMPLE BARCODE: TCGA-L5-A40I-01) demonstrated a mutation within the translation start-site; similar mutations found within this site have previously been identified in several cancers including haematopoietic and lymphoid cancers, lung, ovarian, pancreatic, soft tissue, and gastrointestinal cancer (223-234); additionally, mutations within this site are predicted a high pathogenicity FATHMM score of 0.86. it is important to note 1 of the 2 samples possessing *B2M* mutations also presented with high mutational burden. Within the cohort two multi-hit mutation events within two ICGC patient samples are found with high mutation burden, including a nonsense mutation and three frame shift deletions within the C1 domain of the *B2M* gene. Furthermore, a frameshift deletion mutation identified within a separate sample (TCGA-V5-AASX-01) occurs within an S-nitrosylation site likely impacting post-translational modification (**Figure 15; Table 7**).

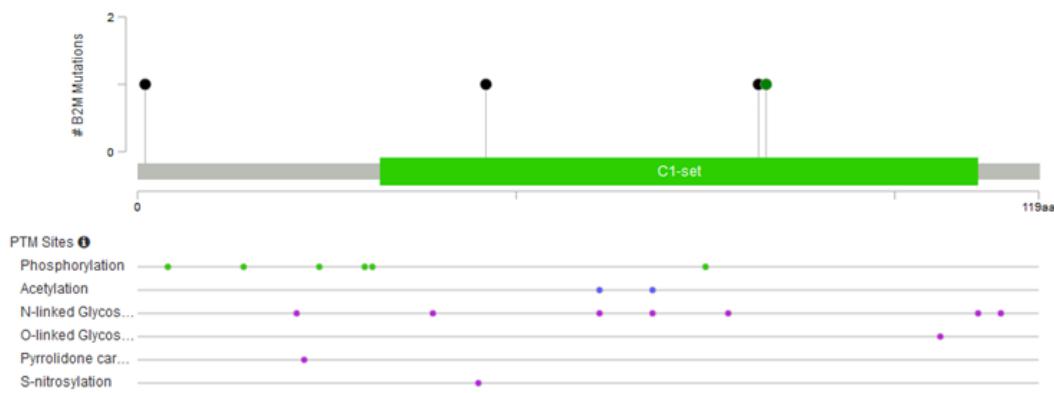


Figure 15 Lollipop plot of B2M mutations in OAC cases within the TCGA-ESCA dataset; mutation descriptions found in table 7. Plots retrieved from cBioportal.

Table 7 B2M mutation descriptions in OAC cases within the TCGA-ESCA dataset, labelled to Figure 15.

LABEL	SAMPLE ID	CANCER TYPE	PROTEIN CHANGE	MUTATION TYPE	TOTAL MUTATIONS IN SAMPLE
A	<u>TCGA-V5-AASX-01</u>	Oesophageal Adenocarcinoma	Y46Cfs*10	FS del	327
B	<u>TCGA-L5-A4OI-01</u>	Oesophageal Adenocarcinoma	Y83Sfs*19	FS del	2043
C	<u>TCGA-L5-A4OI-01</u>	Oesophageal Adenocarcinoma	M1?	Stop Codon	2043
D	<u>TCGA-L5-A4OI-01</u>	Oesophageal Adenocarcinoma	Y83F	Missense	2043

HLA-B mutations were more frequent than B2M mutations in our datasets (**Figure 16**; **Table 8**), with 4/8 mutations predicted to result in loss of function (nonsense and frameshift), this relates to the literature where in OSCC mutations in the immunoglobulin C1-set domain effect function by abrogating HLA-B-Tapasin interactions; interestingly, a single known mutation in *HLA-B* (c.343+2T>C) was previously identified in colorectal and stomach cancers being associated to positive selection in immune rich TIMEs (225, 235). Mutation among APM gene expression regulators is infrequent among my cohort with the highest frequency occurring in *RFXAP* with 4 patients consisting of nonsense, missense, in Frame insertion and a frame shift insertion mutation; unfortunately, because of the encoding available in the ICGC cohort the nonsense and in Frame insertion cannot be evaluated qualitatively.

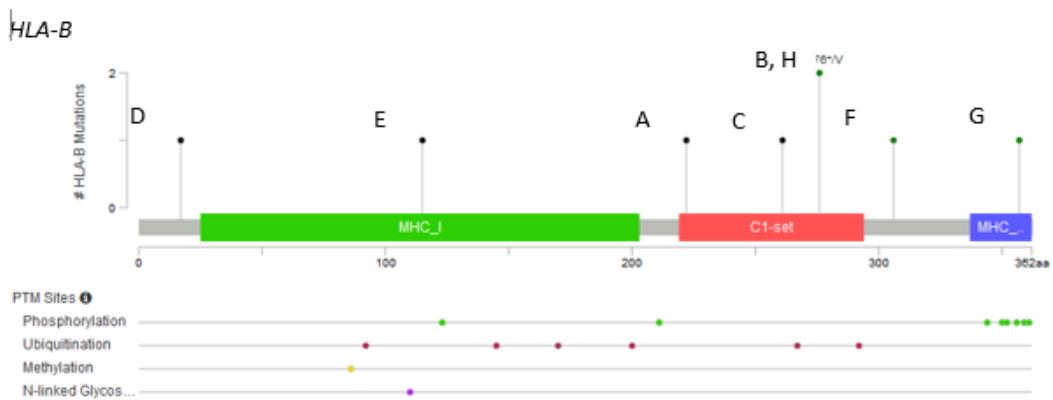


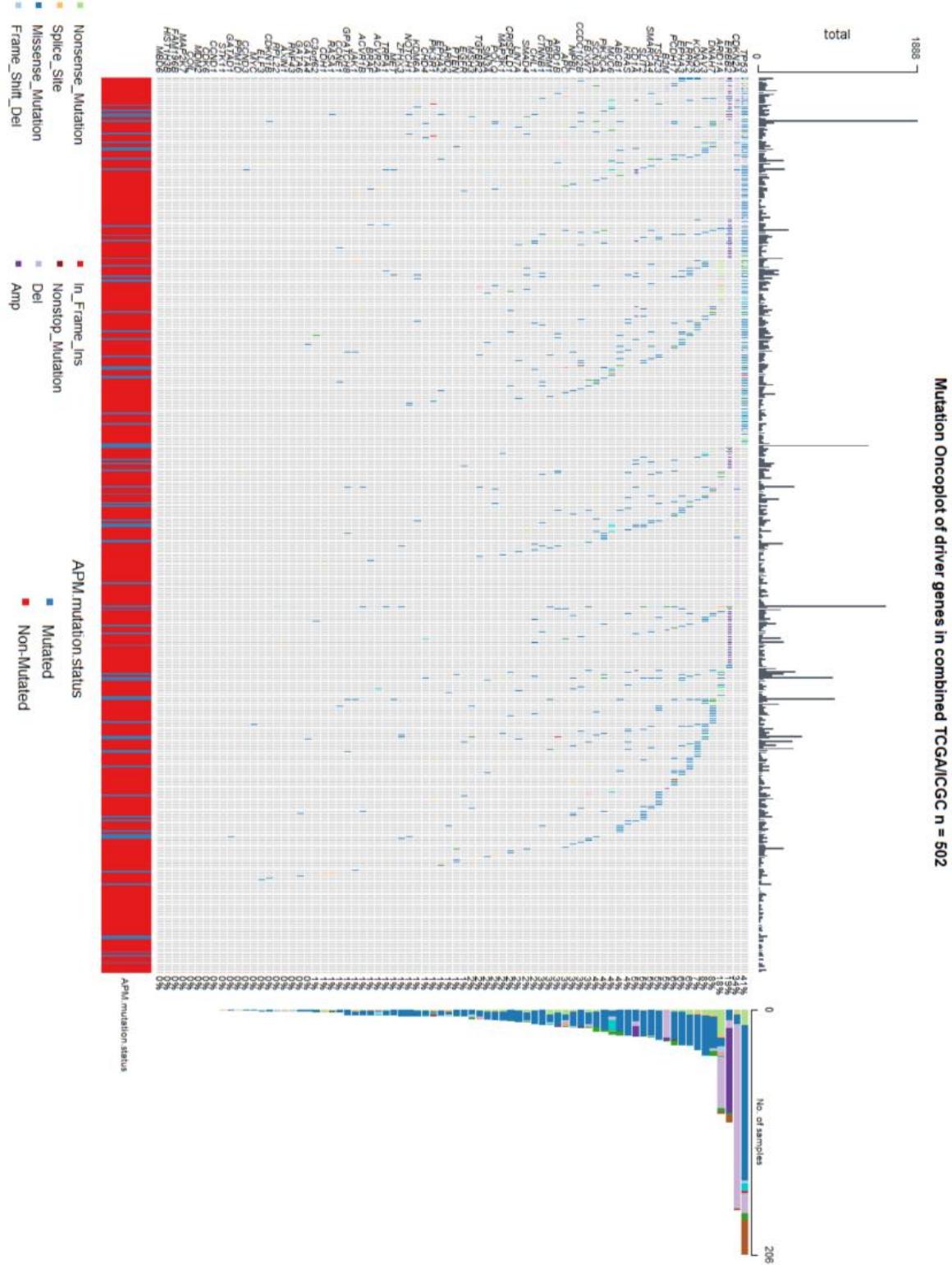
Figure 16 Lollipop plot of HLA-B mutations in OAC cases within the TCGA-ESCA dataset; mutation descriptions found in table 8. Plots retrieved from cBioportal.

Table 8 HLA-B mutation descriptions in ordered by cancer type within the TCGA-ESCA dataset, labelled to Figure 16, ordered by disease type.

LABEL	SAMPLE ID	CANCER TYPE	PROTEIN CHANGE	MUTATION TYPE	# MUT IN SAMPLE
B	<u>TCGA-L5-</u> <u>A8NJ-01</u>	Oesophageal Adenocarcinoma	G276*	Nonsense	297
C	<u>TCGA-V5-</u> <u>AASX-01</u>	Oesophageal Adenocarcinoma	G261*	Nonsense	327
G	<u>TCGA-L5-</u> <u>A4OP-01</u>	Oesophageal Adenocarcinoma	D357A	Missense	122
H	<u>TCGA-L5-</u> <u>A8NJ-01</u>	Oesophageal Adenocarcinoma	G276V	Missense	297
A	<u>TCGA-JY-</u> <u>A93F-01</u>	Oesophageal Squamous Cell Carcinoma	E222*	Nonsense	168
D	<u>TCGA-VR-</u> <u>A8EP-01</u>	Oesophageal Squamous Cell Carcinoma	L17Rfs*82	FS ins	117
E	<u>TCGA-Z6-</u> <u>AAPN-01</u>	Oesophageal Squamous Cell Carcinoma	X115_splice	Splice	369
F	<u>TCGA-JY-</u> <u>A6FG-01</u>	Oesophageal Squamous Cell Carcinoma	V306F	Missense	277

The missense mutation in *RFXAP* occurred within the c-terminal binding domain and possesses a high PolyPhen score suggesting the mutation is probably damaging (0.998) but has not been prior described in the COSMIC database; the frame shift insertion occurs in exon 1 of the *RFXAP* gene, however, the mutation does not appear in the COSMIC database. Only 1 missense mutation was detected in *IRF1* and *CITA* with the *CITA* mutation labelled as benign by PolyPhen (Score of 0.02), the *IRF1* mutation could not be scored because of lack of annotation within the ICGC dataset. *NLR5* 4 missense mutations occur within the ICGC datasets but could not be qualitatively assessed because of lack of annotation. *RFXANK* 1 missense mutation occurs in c-terminal binding domain within the TCGA dataset and scores as benign by PolyPhen (0 score). Only a single *CSDE1* missense mutation occurred within the ICGC datasets, lacking annotation for further analysis. 3 mutations occurred in *RFX5* 1 splice intron (ICGC), 1 missense mutation scoring as possibly damaging (0.544), and 1 translation start site mutation (ICGC).

Comparing the tumour driver genes identified in the Frankell *et al* article, with my cohort analysis found several of the associated driver gene mutations were similar in abundance except for *B2M* mutation, which did not achieve similar representation (**Figure 17**) (236).



mutation burden (Top) and APM mutation status

Within the copy number analysis, I observed high genome complexity in copy number segments in the HLA-locus on chromosome 6 (6p21) with many of the amplified segments overlapping with the *VEGFA* gene a known recurrent amplification event in OAC (**Figure 18**).

The complexity of copy number segment over the locus was further identified in cases with copy number segments spanning the HLA-locus displaying co-amplification/deletion across multiple MHC class I & II genes present in 7.6% of cases within the TCGA/ICGC cohort (n = 38/502). Because of the complexity in the HLA-locus with the majority of copy-number events being described as an APM co-amplification/deletion group. The clinical significance of APM gene copy number events could not be deconvoluted to the single gene level, meaning survival association were based upon the amplification or deletion of a group of APM genes; survival analysis of these co-copied groups did not find any significance differences in overall survival. Next, I assessed the regulators of APM gene expression, these gene are not located on chromosome 6 avoiding the associated genomic complexity. Firstly, *IRF1* displayed the highest frequency of copy-number events among the candidate APM gene expression regulators with 46 patients exhibiting *IRF1* copy number events of which only a sole case displayed an amplification event with the remainder possessing *IRF1* deletion. The second highest frequency of copy-number event among the APM gene expression regulators was found in *RFXAP* with 43 patients possessing a copy-number event 25 of which exhibited amplification and 17 deletions. *CITA* copy number events were present in 32 patients with 8 amplification and 24 deletion events. *RXANK* copy number events occurred in 30 patients of which 24 were deletion events and 6 amplifications. *NLRC5* copy number events were present in 28 patients among these were 26 deletion and 2 amplification events. Within the cohort 27 patients possessed a copy number event in *CSDE1* consisting of 18 deletion and 9 amplification events. Finally, 22 patients demonstrated *RFX5* copy number events with 4 deletions and 18 amplifications.

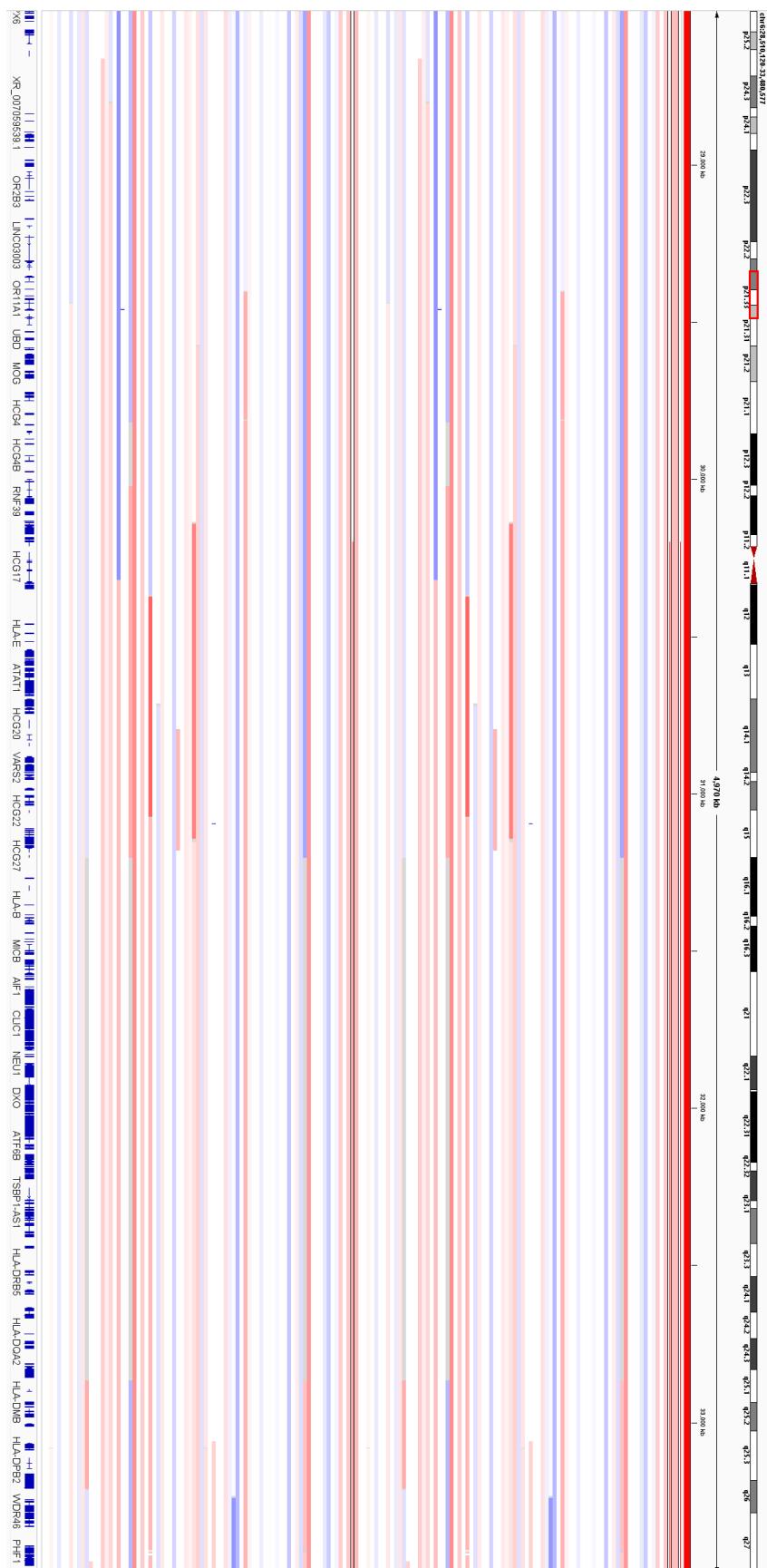


Figure 18: Representative copy number segments in the HLA-locus on chromosome 6 (6p21)
blue and red represents loss and gain respectively, each line on y-axis represents individual patients.

3.3.3 APM dysregulated mRNA expression.

Differential expression analysis of the TCGA/OCCAMS bulk RNA-seq data (n=176) identified none of my APM gene candidates were differentially expressed when comparing to GTEx normal stomach (n=359). Oesophageal mucosa tissue was selected using prior literature assessing the top varying genes, reduced features from principal component analysis, and encoded features from an autoencoder neural network (237). Among the APM gene candidates which were only small fold changes in differential expression were detected with the greatest increase of expression found in CTSS and IRF1 (4.07E-02 and 3.19E-02 respectively), however both genes did not achieve significance; similarly the greatest decrease in expression identified in TAP2 and HLA-DPA1 (-1.45E-02 and -1.31E-02 respectively) was not significant in my differential expression analysis (**Table 9, Figure 19**).

Unfortunately, differential expression analysis between OAC and OSCC did not yield any significant fold change in the expression of our APM gene candidates, neither achieving a Log2 fold change greater than +1/-1 nor the p value threshold (**See Figure 20**).

Table 9 A table of differential expression results for our APM candidate genes with log fold change of gene expression comparing OAC to normal stomach tissue.

GENE	BASEMEAN	LOG2FOLDCHANGE	LOG2 FOLD CHANGE SE	WALD STATISTIC	PVALUE
<i>CTSS</i>	44.983196	4.07E-02	0.109	0.374	0.708
<i>IRF1</i>	43.761558	3.19E-02	0.0791	0.403	0.687
<i>SPPL2A</i>	45.669112	2.40E-02	0.0370	0.648	0.517
<i>CANX</i>	365.66515	2.36E-02	0.0634	0.372	0.710
<i>RFXAP</i>	4.4176	2.31E-02	0.0568	0.406	0.685
<i>CSDE1</i>	504.49757	2.19E-02	0.0509	0.431	0.666
<i>RFX5</i>	37.967755	1.96E-02	0.0404	0.485	0.628
<i>NLRC5</i>	45.122318	1.64E-02	0.1036	0.158	0.874
<i>ERAP1</i>	59.02741	1.38E-02	0.0413	0.333	0.739
<i>CITA</i>	40.130435	1.17E-02	0.0661	0.177	0.859
<i>CTSL</i>	57.124888	1.16E-02	0.0815	0.143	0.887
<i>PDIA3</i>	200.54399	9.96E-03	0.0561	0.178	0.859
<i>CD1D</i>	1.002181	9.58E-03	0.142	0.0673	0.946
<i>MR1</i>	22.019528	8.45E-03	0.0255	0.331	0.741
<i>TAPBPL</i>	21.283693	7.81E-03	0.0513	0.152	0.879
<i>TAPBP</i>	157.91175	5.31E-03	0.0455	0.117	0.907
<i>LGMN</i>	62.518148	3.85E-03	0.0466	0.0827	0.934
<i>HLA-C</i>	425.26405	2.12E-03	0.0601	0.0353	0.972
<i>HLA-A</i>	393.5236	2.02E-03	0.0531	0.0380	0.970
<i>HLA-E</i>	481.5501	7.32E-04	0.0426	0.0172	0.986
<i>RFXANK</i>	24.719202	-9.42E-05	0.0356	0.00264	0.998
<i>HLA-B</i>	534.29728	-7.38E-04	0.0647	0.0114	0.991
<i>CD74</i>	441.40637	-1.52E-03	0.0854	0.0178	0.986
<i>HLA-G</i>	6.649073	-2.30E-03	0.0720	-0.0320	0.974
<i>HLA-DMA</i>	25.072684	-3.12E-03	0.0865	0.0360	0.971
<i>B2M</i>	955.77353	-3.46E-03	0.0593	0.0582	0.954
<i>PSMB10</i>	17.868314	-5.32E-03	0.0473	-0.112	0.911
<i>PSMB9</i>	15.42342	-6.70E-03	0.0864	0.0775	0.938
<i>PSMB8</i>	39.852858	-6.76E-03	0.0587	0.115	0.908
<i>CALR</i>	319.23215	-7.02E-03	0.0549	0.127	0.898
<i>TAP1</i>	73.765951	-7.41E-03	0.0715	0.104	0.918
<i>HLA-DRA</i>	138.40034	-8.20E-03	0.143	0.0573	0.954
<i>HLA-DRB1</i>	95.535384	-8.66E-03	0.103	0.0844	0.933
<i>HLA-DPB1</i>	43.172828	-1.09E-02	0.0900	-0.121	0.904
<i>HLA-DMB</i>	13.371499	-1.31E-02	0.0803	0.164	0.870
<i>TAP2</i>	48.822086	-1.45E-02	0.0682	0.213	0.833
<i>HLA-DPA1</i>	66.933144	-2.03E-02	0.124	-0.164	0.870
<i>ERAP2</i>	25.460888	-4.01E-02	0.162	0.248	0.804
<i>HLA-DQB2</i>	3.754546	-4.50E-02	0.148	0.303	0.762
<i>HLA-DQA1</i>	18.632014	-6.43E-02	0.160	0.402	0.688

HLA-DQA2	4.119687	-6.46E-02	0.166	0.389	0.697
HLA-DRB5	39.667806	-1.54E-01	0.197	0.778	0.437
HLA-DOA	7.489424	-1.68E-01	0.177	-0.954	0.340

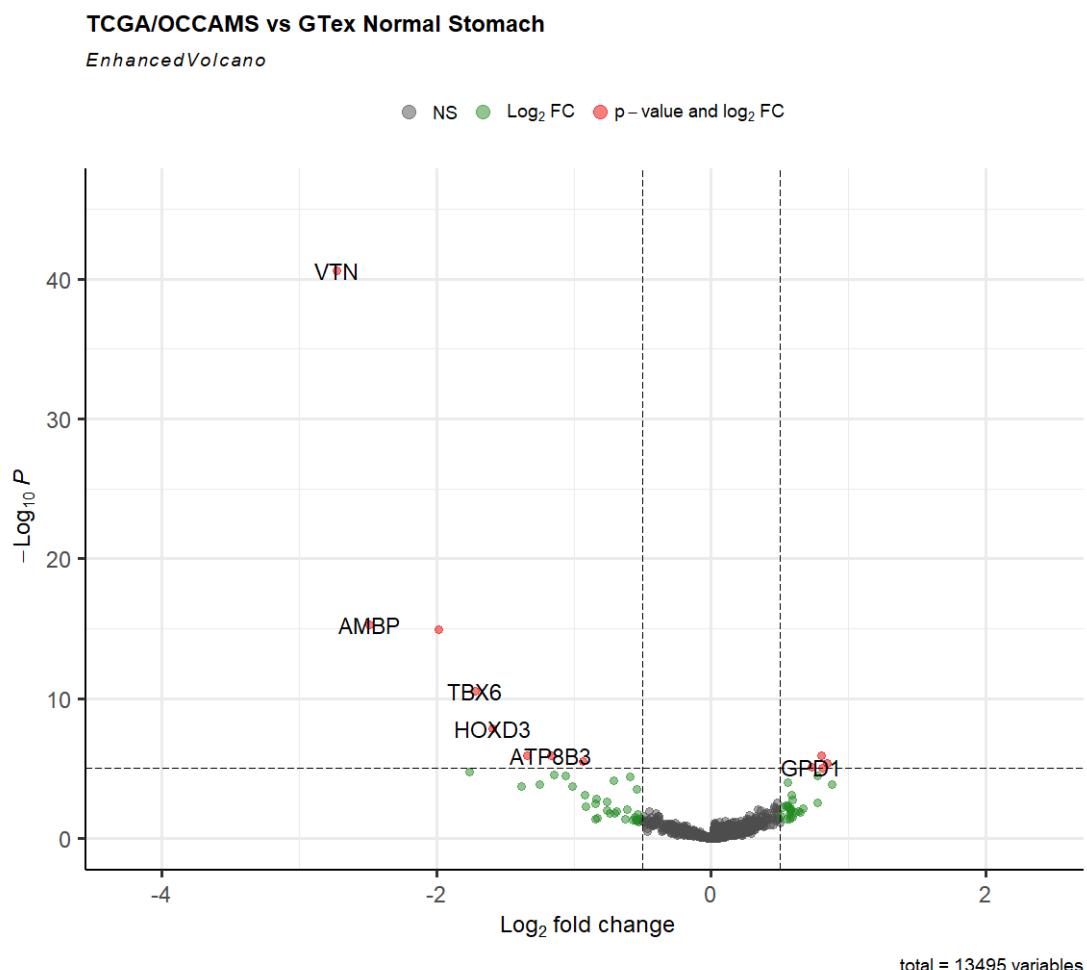


Figure 19 Volcano plot of differential expression analysis (DEA). Comparing APM gene expression between TCGA/OCCAMS tumour samples and GTEx Normal stomach samples. log fold change, over -log10 p value, cut-offs: Log2 fold change 0.5, -log10 p value 10e-5.

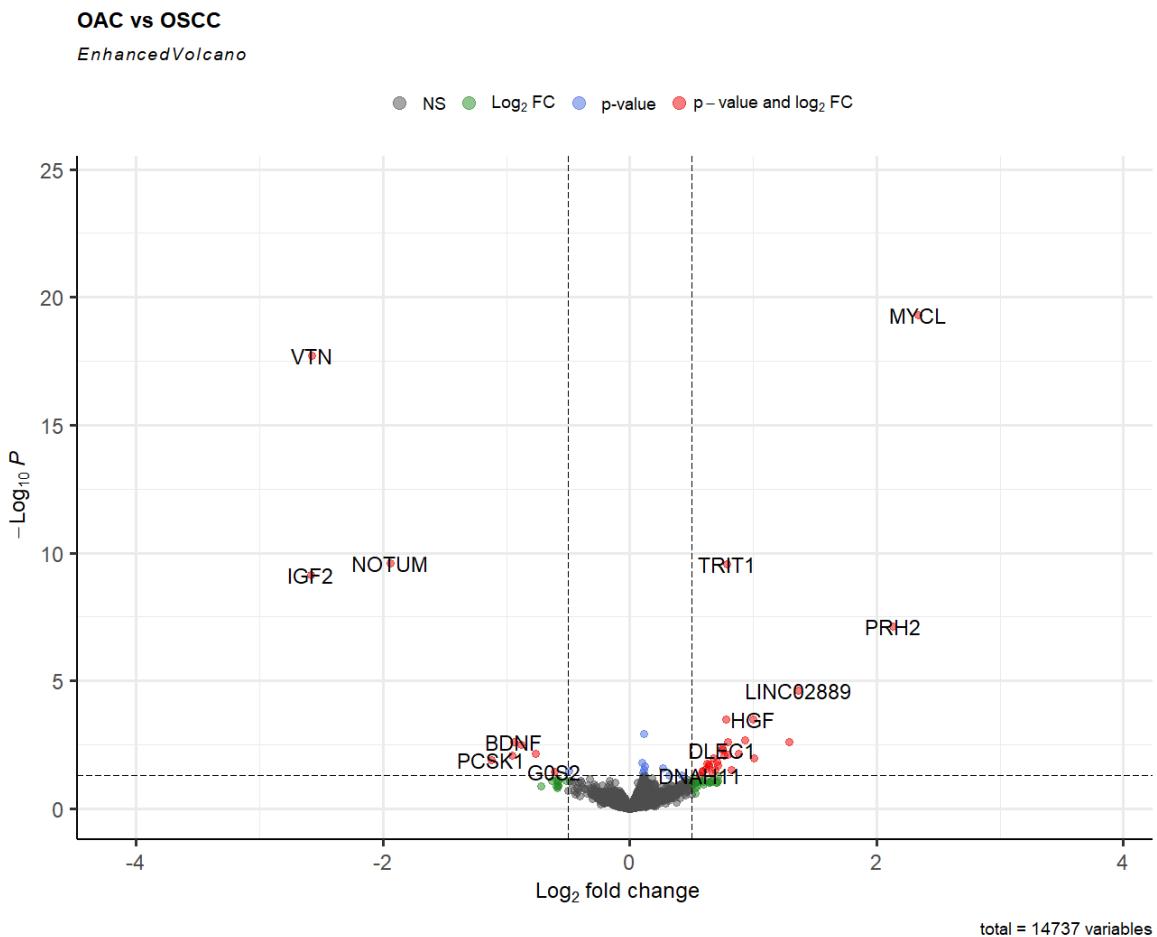


Figure 20 Volcano plot of differential expression analysis (DEA). Comparing APM gene expression between TCGA/OCCAMS tumour samples and TCGA OSCC samples. log fold change, over -log10 p value, cut-offs: Log2 fold change 0.5, -log10 p value 10e-5.

3.3.4 APM genes associated with clinical outcomes and disease progression.

It is unknown whether the level of expression of my APM gene candidates possessed prognostically significant associations with overall survival. Therefore, I sought to assess the maximal differences in survival (overall survival, cancer-specific survival, and disease-free survival) because of APM gene expression, which would allow me to filter out APM gene expression profiles which are potentially non-informative in OAC. To achieve this, I explored overall survival with maximally ranked statistics to select high and low survival expression groups for each APM gene candidate for comparison, producing this analysis yielded significant differential survival dependent on APM expression. Furthermore, the significant associations in overall survival, cancer-specific survival and disease-free survival were evaluated in a multivariate model to determine if the prognostic significance was independent of prognostic clinical features.

3.3.5 APM regulators and survival associations

Firstly, four prior mentioned regulators of APM gene expression were associated to different survival outcomes in my OAC cohort. The high expression group of *CSDE1* (n=130) possessed a significant association to poor overall survival in univariate survival analysis (HR: 1.74, p=0.015). In multivariate survival analysis the significance of the association between high *CSDE1* expression and risk for poor overall survival grew (HR: 2.90, p=0.008) demonstrating itself as an independent prognostic factor (**Figure 23, Table 12**).

In univariate cancer specific survival analysis (CSS), only the high expression group of *CSDE1* (n=97) associated with poor prognosis with borderline significance (HR: 1.70, p=0.051), but non-significant by multivariate survival analysis (**Table 13**). Univariate DFS analysis identified the low expression group of *CSDE1* (n=18) was associated with increased risk of disease recurrence with high significance (HR: 2.81, p=0.003), multivariate analysis this association retained high significance and increased in risk 1.6-fold (HR: 4.39, p=0.015) (**Table 14**).

The *RFX5* gene, an enhanceosome RFX-complex regulator of MHC class I and class II expression also possessed association to overall survival. The low expression group of *RFX5* (n=138) associated to poor overall survival (HR: 2.10, p=0.013). In univariate CSS analysis, the low expression group of *RFX5* (n=109) associated with poor prognosis (HR: 2.81, p=0.017). Under multivariate analysis this association lost significance (**Table 13**). Univariate disease-free survival (DFS) analysis identified the low expression group of *RFX5* (n=108) was associated with increased risk of disease recurrence with high significance (HR: 2.81, p=0.024), in multivariate analysis this association possessed borderline significance with similar risk (HR: 2.65, p=0.053) (**Table 14**).

Additionally, the high expression group of *RFXAP* (n=105) associated to poorer overall survival (HR:1.63, p=0.022); in multivariate analysis the only RFX-complex gene to retain its significance was *RFX5* (HR:2.78, p=0.028) with *RFXAP* expression losing its prognostic significance (HR:1.01, p=0.987); (**Figure 23, Table 12**).

The fourth APM regulator uncovered as prognostically significant was *CITTA*, a gene responsible for expression of MHC class II genes in professional APCs, of which the low expression group was associated to poor survival in univariate analysis (HR:1.65, p=0.016), yet did not remain significant in multivariate survival analysis (HR:1.40, p=0.323). In univariate CSS analysis the low expression group of *CITTA* (n=64) associated with poor prognosis with borderline significance (HR: 1.68, p=0.058), unfortunately multivariate analysis this association lost significance and inverted the association with risk (HR: 0.71, p=0.405) (**Table 13**).

3.3.6 Correlation analysis identifies correspondent expression between APM genes and APM gene expression regulators.

To explore the potential mechanism between APM gene expression regulators and survival outcomes I performed a correlation analysis between their expression and the expression of MHC class I and II genes. Firstly, within the MHC class I genes *CSDE1* presents a significant negative correlation of *HLA-A/B/C* (p = 0.032, 0.028, 0.014; see **Figure 21**), *PSMB9* (R = -0.182; p = 0.015; see **Figure 21**) and *TAPBPL* (R = -0.176; p = 0.02; see **Figure 21**), but a positive correlation with *CANX* (R = 0.197; p = 0.0086; see **Figure 21**). Interestingly, *CSDE1* expression also negatively correlates with other APM regulators including *NLRG5*, *IRF1* and *CITTA* (p = 0.0046, 0.05, 0.05; see **Figure 21**).

NLRG5 expression positively correlated to *HLA-A/B/C/E*, *TAP1*, *TAP2*, *TAPBP*, *TAPBPL*, *PSMB9/10* expression (p <0.01), and negatively correlated with *CANX* and *CALR* expression (Respectively, R = -0.247, -0.216; p = 0.0009, 0.004; see **Figure 21**).

IRF1 expression correlated to MHC class I HLAs including *HLA-A/B/C/E/G* (p < 0.001; see **Figure 21**), the Tapasin translocon components *TAP1*, *TAP2*, *TAPBP* and *TAPBPL* (p < 0.0001; see **Figure 21**), *B2M* (R = 0.349; p < 0.0001; see **Figure 21**), and finally peptide generation genes *ERAP1*, *ERAP2* and *PSMB8/9/10* (p < 0.0001; see **Figure 21**).

CITTA expression correlated with MHC class I HLAs *HLA-A/B/C/E/G* (p < 0.01; see **Figure 21**), Tapasin translocon components *TAP1*, *TAP2*, *TAPBP* and *TAPBPL* (p < 0.0001; see **Figure 21**), *B2M* (R = 0.226; p = 0.003; see **Figure 21**) and peptide generation genes *ERAP2* and *PSMB9/10* (p < 0.01; see **Figure 21**).

The APM gene expression regulators which found correlation with MHC class II APM expression included *IRF1*, *RFX5*, *RFXAP* and *CIITA*. *CIITA* possessed the strongest positive correlations with MHC class II expression in *HLA-DMA/DMB*, *HLA-DOA*, *HLA-DPA1/B1*, *HLA-DRA*, *HLA-DQA1/A2/B2*, and *HLA-DRB1/B5* ($p < 0.0001$; see **Figure 22**).

RFX5 also positively correlated with MHC class II HLA expression in *HLA-DMA/DMB*, *HLA-DOA*, *HLA-DPA1/B1*, *HLA-DRA*, *HLA-DQA1/A2/B2*, and *HLA-DRB1/B5* ($p < 0.05$; see **Figure 22**), as well as, *CD74* ($R = 0.344$; $p < 0.001$; see **Figure 22**), and peptide generation genes *CTSS*, *SPPL2A*, and *PDI A3* ($p < 0.05$; see **Figure 22**). Conversely, *RFXAP* possessed a negative correlation with *HLA-DMA* and *LGMN* (Respectively, $R = -0.160$, -0.152 ; $p < 0.05$; see **Figure 22**).

Finally, *IRF1* displayed a positive correlation with MHC class II HLA expression in *HLA-DMA/DMB*, *HLA-DOA*, *HLA-DPA1/B1*, *HLA-DRA*, *HLA-DQA1/A2/B2*, and *HLA-DRB1/B5* ($p < 0.001$; see **Figure 22**), *CD74* ($R = 0.386$; $p < 0.0001$; see **Figure 22**), and peptide generation genes *CTSS* and *CTSL* (Respectively, $R = 0.308$, 0.210 ; $p < 0.01$; see **Figure 22**).

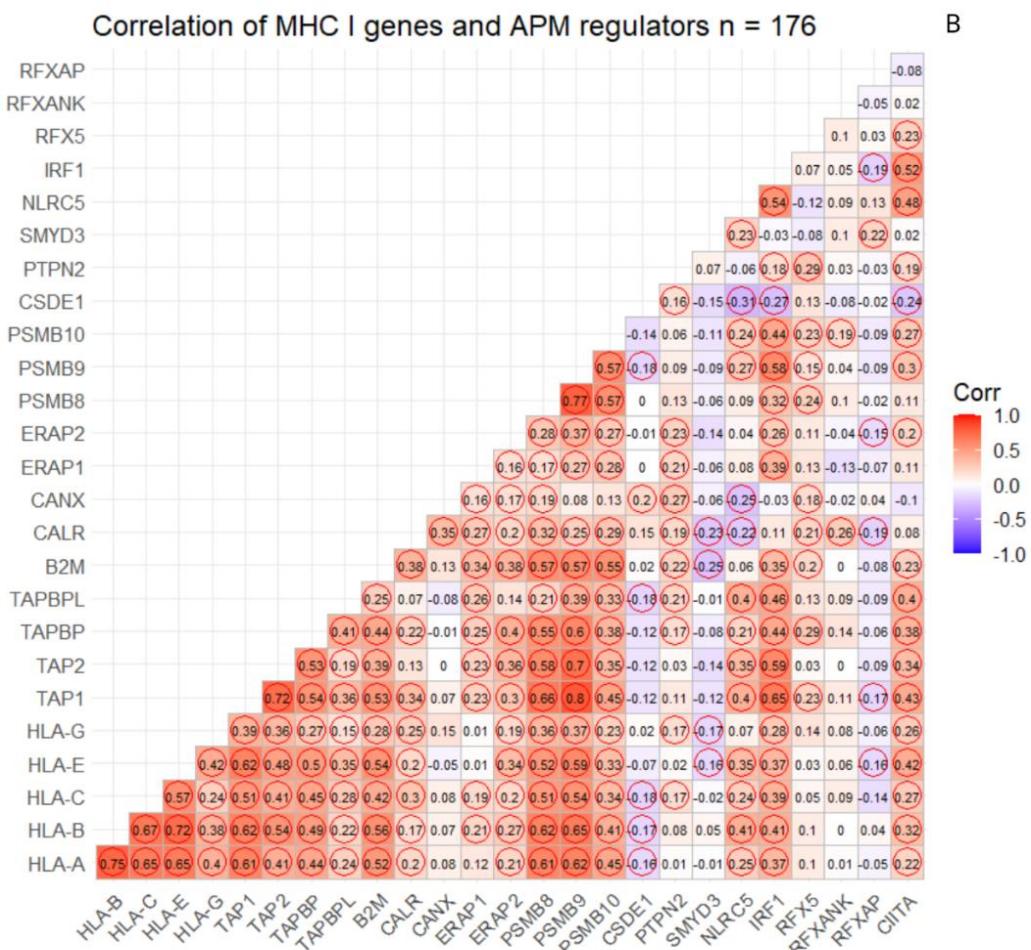


Figure 21 Correlation heatmap of MHC class I expression with APM regulators. Red circles represent significant correlations (Pearson's correlation: $p < 0.05$). A: MHC class I expression with APM regulators. B: MHC class II expression with APM regulators.

Correlation of MHC II genes and APM regulators n = 176

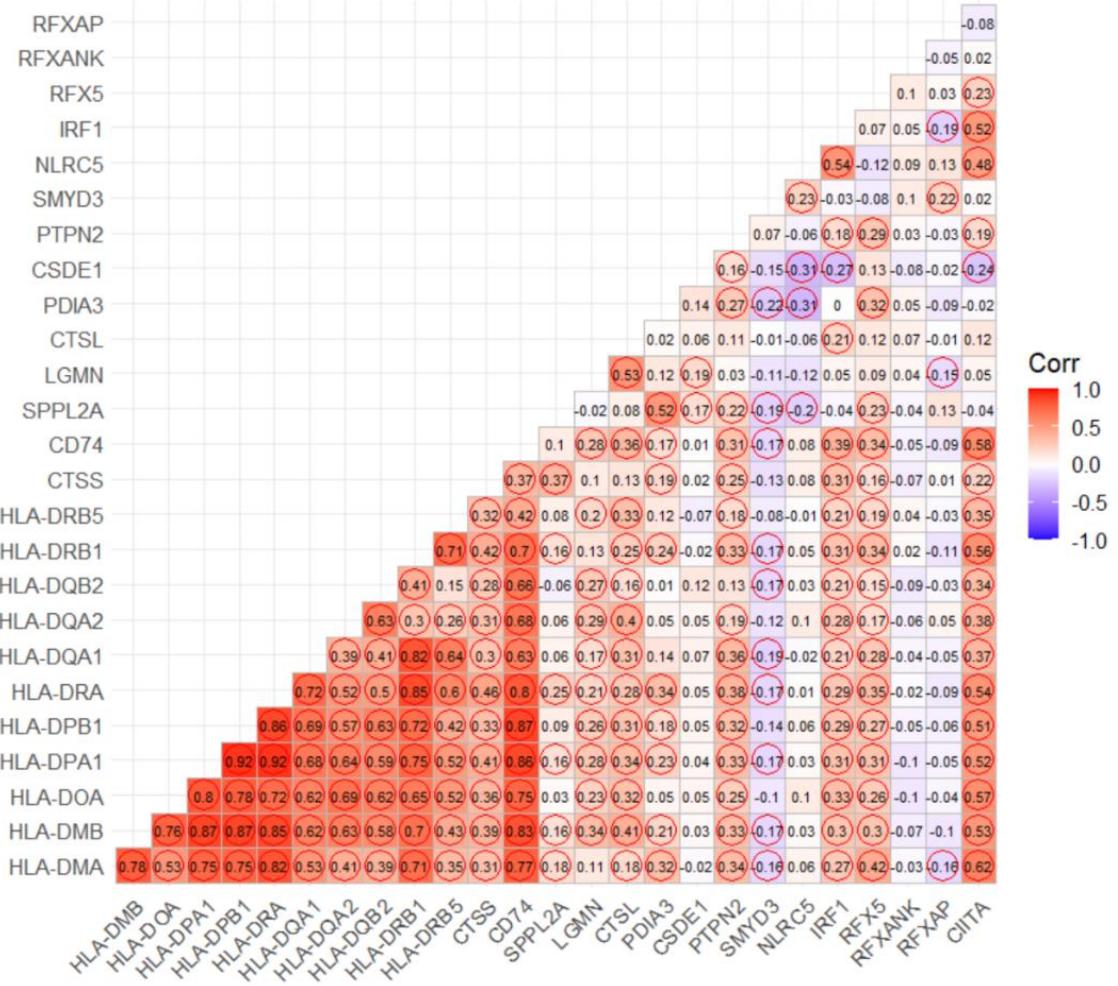


Figure 22 Correlation heatmap of MHC class II expression with APM regulators. Red circles represent significant correlations (Pearson's correlation: $p < 0.05$). A: MHC class I expression with APM regulators. B: MHC class II expression with APM regulators.

3.3.7 MHC class I APM gene candidates and survival associations

Out of the five examined MHC class I loading complex candidate genes only low *HLA-A* expression was associated with altered OS (HR:1.80, $p=0.015$), CSS (HR: 2.06, $p=0.023$) and DFS (HR: 2.25, $p=0.021$) in the OAC cohort (See Figure 23, Table 12 & 13). However, the association between poor survival outcomes and low *HLA-A* expression did not withstand multivariate model analysis (See Table 12 & 13). Additionally, *HLA-B/-E/-G* demonstrated an association with altered CSS (HR: 1.77, $p=0.033$; HR: 2.09, $p=0.036$; HR: 1.77, $p=0.031$) within the OAC cohort (See Table 13). Despite these univariate CSS findings within *HLA-B/-E/-G* expression groups, when applied to multivariate survival analysis the significance was not retained. Furthermore, none of the survival outcomes were associated with *HLA-C* expression.

Examining the six candidate genes known to participate in the assembly of the MHC class I loading complex, only low *CALR* expression was associated with both altered OS (HR: 1.91, p=0.018) and CSS (HR: 2.11, p=0.039), these findings retained significance in multivariate modelling for OS, but not CSS analysis (**See Table 12 & 13**). Both low expression of *TAPBPL* and *TAPBP* were associated with altered CSS (HR:2.00, p=0.009; HR:2.20, p=0.014) and DFS (HR:2.05, p=0.036; HR:2.16, p=0.011); only low *TAPBP* expression retained significance in multivariate CSS analysis (**See, Table 13 & 14**). The remaining three gene candidates participating in MHC class I assembly (*B2M*, *PDIA3* and *CANX*) did not possess any significant survival associations with OS, CSS or DFS.

Of the seven MHC class I gene candidates which function to generate peptides and transport them into the endoplasmic reticulum, singularly the high expression group of *PSMB10* associated with altered OS (HR: 1.61, p=0.025), CSS (HR: 2.71, p=0.001), and DFS (HR: 1.79, p=0.040) (**See Table 12, 13 & 14**). Despite the interesting univariate significance of the high *PSMB10* expression group, these results did not remain significant in multivariate survival analysis. Among this subset of MHC class I genes, the low expression group of *ERAP2* demonstrated an association with poor OS (HR: 1.72, p=0.007) and CSS (HR: 2.05, p=0.023) (**See Table 12 & 13**); multivariate survival analysis only association with OS were retained. The high expression groups of *ERAP1*, *PSMB8* and *PSMB9* displayed an association with shorter CSS (HR:2.55, p=0.030; HR:2.32, p=0.016; HR:2.76, p=0.003) (**See Table 13**). Despite finding significantly altered CSS from these genes in univariate analysis, application of the multivariate model did not yield significance. Finally, the expression groups of Tapasin translocon, *TAP1* and *TAP2* did not possess an association with any survival measure (OS/CSS/DFS). Overall, twelve out of the eighteen MHC class I gene candidates demonstrated an association with patients' outcomes, with only four of the twelve associations retaining significance within the multivariate survival analysis (**Table 10**).

Table 10 MHC I candidate gene expression association with survival outcomes among the OAC cohorts (TCGA and OCCAMS). Significance respective to outcome.

APM GENE SYMBOL; MHC I	EXPRESSION-SURVIVAL	SIGNIFICANCE	
		ASSOCIATION	UNIVARIATE
<i>TAPBP</i>	LOW EXPRESSION – SHORTER DFS	*	*
<i>TAPBPL</i>	LOW EXPRESSION – SHORTER OS, CSS	NS, **	NS, NS
	HIGH EXPRESSION – SHORTER DFS	*	NS
<i>HLA-A</i>	LOW EXPRESSION – SHORTER OS, CSS, DFS	*, *, *	NS, NS, NS
<i>HLA-B</i>	LOW EXPRESSION – SHORTER CSS	*	NS
<i>HLA-C</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>HLA-E</i>	LOW EXPRESSION – SHORTER CSS	*	NS
<i>HLA-G</i>	LOW EXPRESSION – SHORTER CSS	*	NS
<i>B2M</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>PDIA3</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>CALR</i>	LOW EXPRESSION – SHORTER OS, CSS	*, *	*, NS
<i>CANX</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>ERAP1</i>	LOW EXPRESSION – SHORTER CSS	*	*
<i>ERAP2</i>	LOW EXPRESSION – SHORTER OS, CSS	*	*
<i>TAP1</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>TAP2</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>PSMB9</i>	LOW EXPRESSION – SHORTER CSS	*	NS
<i>PSMB10</i>	HIGH EXPRESSION – SHORTER OS, CSS, DFS	*, **, *	NS, NS, NS
<i>PSMB8</i>	LOW EXPRESSION – SHORTER CSS	*	NS

P values: NA = Non-applicable, NS = non-significant, * <0.05 , ** <0.01 , *** <0.001 .

3.3.8 MHC class II APM gene candidates and survival associations

Findings with HLAs

Out of the fifteen examined MHC class II loading complex/assembly candidate genes eight demonstrated associations with altered survival outcomes (OS, CSS, DFS) (**Table 11**). Firstly, the low expression of CD74 was associated to shorter OS approaching significance (HR:1.67, p=0.054); and achieving significance with a reduced CSS and DFS time (HR: 2.88, p=0.006; HR: 2.04, p=0.050) (**See Figure 23, Table 12, 13 & 14**). However, these associations are not retained in the multivariate clinical (**See Figure 23, Table 12, 13 & 14**). Low expression of the *HLA-DMA* gene transcript demonstrated the most significant association with poor overall survival (HR:1.77, p=0.005; **see Table 12**).

Despite possessing the highest significance within the univariate model, under multivariate this association have borderline significance at the 5% level (HR:1.93, p=0.051; **see Table 12**).

The next significant survival association observed was with lower expression of *HLA-DRA* (OS HR:1.67, p=0.013; CSS HR: 2.07, p=0.01; DFS HR: 2.07, p=0.01); only the association with shorter CSS remained significant in multivariate analysis (HR:2.07, p=0.027; **see Table 12, 13 & 14**).

For *HLA-DPA1* low expression was related to poorer OS, CSS and DFS (HR:1.53, p=0.035; HR: 2.37, p=0.025; HR: 2.54, p=0.001; **See Figure 23, Table 12, 13 & 14**), but only poorer OS and DFS, in the multivariate model (**see Table 12 & 14**). The low expression of *HLA-DRB5* only related to increased risk of recurrence of disease with high significance (DFS HR: 2.52, p=0.001) in univariate tests (**see Table 14**). Decreased *HLA-DQA1* expression demonstrated an association with CSS and DFS (HR: 1.85, p=0.054; HR: 2.35, p=0.003); but neither remained significant in multivariate analysis (**see Table 13 & 14**).

Additionally, OAC patients with lower *HLA-DRB1* expression corresponded with poorer CSS (HR: 2.71, p=0.002) and an increased risk of recurrence (DFS HR: 2.34, p=0.006), but only in univariate analysis (**see Table 13 & 14**). Decreased expression of *HLA-DOA* in OAC patients corresponded with increased recurrence of disease (DFS HR: 2.20, p=0.013), yet failed to withstand multivariate analysis (**see Table 14**). The final significant relationship between MHC Class II genes and survival outcome was between low *HLA-DQA2* expression and recurrence of disease (DFS HR: 2.26, p=0.015), however in multivariate analysis, this relationship lost significance (**see Table 14**)

Peptide generation

Within, this class of candidate genes three out of five candidate demonstrate association with survival outcomes (**See Table 11**), high expression of *SPPL2A* (n=104) was associated to poorer survival outcomes (OS HR:1.52, p=0.045), however this finding did not stand up to multivariate testing against the clinical model (OS HR:1.07, p=0.831). In univariate CSS analysis the high expression group of *SPPL2A* associated with poor prognosis with significance (CSS HR: 1.52, p=0.045), multivariate analysis demonstrated a non-significant trend with increased in risk (HR: 2.73, p=0.190) (**Table 13**).

The low expression group of *LGMN* (n=60) was associated to lower patient overall survival (HR:1.95, p=0.001), despite the high significance of this association the significance is lost in multivariate testing against the clinical model (OS HR:1.66, p=0.165). In univariate CSS analysis, the low expression group of *LGMN* (n=32) associated with poor prognosis with great significance (HR: 2.55, p=0.001), in multivariate analysis this association lost significance (HR: 1.33, p=0.625) (**Table 13**). Univariate DFS analysis identified the low expression group of *LGMN* (n=68) was associated with increased risk of disease recurrence with great significance (HR: 1.93, p=0.015), multivariate analysis failed to yield significance.

Finally, the low expression group of CTSS corresponded with shorter CSS (HR:1.99, p=0.010), but did not retain significance within the multivariate model (**Table 13**).

Table 11 MHC II candidate gene expression association with survival outcomes among the OAC cohorts (TCGA and OCCAMS). Significance respective to outcome.

APM GENE SYMBOL; MHC II	EXPRESSION-SURVIVAL ASSOCIATION	SIGNIFICANCE	
		UNIVARIATE	MULTIVARIATE
<i>HLA-DMA</i>	LOW EXPRESSION – SHORTER OS	**	NS
<i>HLA-DMB</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>HLA-DOA</i>	LOW EXPRESSION – SHORTER DFS	*	NS
<i>HLA-DOB</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>HLA-DPA1</i>	LOW EXPRESSION – SHORTER OS, CSS & DFS	* , * , **	* , NS, NS
<i>HLA-DPB1</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>HLA-DQA1</i>	LOW EXPRESSION – SHORTER CSS & DFS	NS, **	NS, NS
<i>HLA-DQA2</i>	LOW EXPRESSION – SHORTER DFS	*	NS
<i>HLA-DQB1</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>HLA-DQB2</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>HLA-DRA</i>	LOW EXPRESSION – SHORTER OS, CSS & DFS	* , * , *	* , NS, NS
<i>HLA-DRB1</i>	LOW EXPRESSION – SHORTER CSS	**	NS
<i>HLA-DRB4</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>HLA-DRB5</i>	LOW EXPRESSION – SHORTER DFS	***	NS
<i>CD74 (cell surface form)</i>	LOW EXPRESSION – SHORTER OS, CSS & DFS	NS, **. *	NS, NS, NS
<i>CTSS</i>	LOW EXPRESSION – SHORTER CSS	**	NS
<i>CTSL</i>	NO SURVIVAL ASSOCIATION	NA	NA
<i>SPPL2A</i>	HIGH EXPRESSION – SHORTER OS & CSS	* , *	NS, NS
<i>LGMN</i>	LOW EXPRESSION – SHORTER OS, CSS & DFS	*** , ***. **	NS, NS, NS
<i>IFI30</i>	NO SURVIVAL ASSOCIATION	NA	NA

P values: NA = Non-applicable, NS = non-significant, * <0.05, ** <0.01, *** < 0.001.



Figure 23 Univariate CoxPH forest plot of APM gene expression in the TCGA/OCCAMS Cohort using maximally ranked statistics for optimal cut points between low/high expression groups. Only significant/approximate significant results shown $p < 0.55$. * Represents genes which greater expression associates with greater risk.

Table 12 CoxPH overall survival analysis in univariate and multivariate model.

VARIABLE	LEVELS	N (%)	HR (UNIVARIABLE) N = 176	HR (MULTIVARIABLE) N = 78
Age	Mean (SD)	66.8 (11.1)	0.99 (0.98-1.01, p=0.515)	1.01 (0.98-1.05, p=0.440)
Sex	Female	31 (17.6)	-	-
	Male	145 (82.4)	1.65 (0.90-3.01, p=0.105)	0.77 (0.31-1.91, p=0.576)
pT	0-1	38 (29.5)	-	-
	2-4	91 (70.5)	1.39 (0.78-2.48, p=0.266)	0.48 (0.17-1.37, p=0.172)
pN	0	48 (36.9)	-	-
	1-3	82 (63.1)	2.56 (1.45-4.52, p=0.001)	3.83 (1.41-10.45, p=0.009)
pM	0	77 (92.8)	-	-
	1	6 (7.2)	7.07 (2.75-18.19, p<0.001)	8.04 (2.99-21.61, p<0.001)
CSDE1	low	130 (74.4)	-	-
	high	44 (25.6)	1.74 (1.11-2.71, p=0.015)	2.90 (1.31-6.40, p=0.008)
RFX5	high	138	-	-
	low	36	2.10 (1.17-3.76, p=0.013)	2.78 (1.12-6.91, p=0.028)
SPPL2A	low	69	-	-
	high	105	1.52 (1.01-2.30, p=0.045)	1.07 (0.56-2.07, p=0.831)
RFXAP	low	69	-	-
	high	105	1.63 (1.07-2.49, p=0.022)	1.01 (0.52-1.93, p=0.987)
CD74	high	40	-	-
	low	134	1.67 (0.99-2.82, p=0.054)	1.12 (0.53-2.34, p=0.769)
CIITA	high	86	-	-
	low	88	1.65 (1.10-2.47, p=0.016)	1.40 (0.72-2.71, p=0.323)
PSMB10	low	121	-	-
	high	53	1.61 (1.06-2.44, p=0.025)	1.30 (0.65-2.61, p=0.456)
LGMN	high	114	-	-
	low	60	1.95 (1.30-2.93, p=0.001)	1.66 (0.81-3.41, p=0.165)
TAPBPL	high	124	-	-
	low	50	1.48 (0.99-2.24, p=0.058)	1.00 (0.48-2.06, p=0.991)
CALR	high	42	-	-
	low	132	1.91 (1.12-3.28, p=0.018)	2.52 (1.01-6.32, p=0.048)
ERAP2	high	109	-	-
	low	65	1.72 (1.16-2.56, p=0.007)	3.15 (1.58-6.27, p=0.001)
HLA-DPA1	high	89	-	-
	low	85	1.53 (1.03-2.27, p=0.035)	1.98 (1.04-3.75, p=0.037)
HLA-DMA	high	98	-	-
	low	76	1.77 (1.18-2.64, p=0.005)	1.93 (1.00-3.73, p=0.051)
HLA-DRA	high	81	-	-
	low	93	1.67 (1.12-2.49, p=0.013)	2.07 (1.08-3.94, p=0.027)
HLA-DRB5	high	104	-	-
	low	70	1.45 (0.98-2.16, p=0.065)	1.70 (0.88-3.30, p=0.116)
HLA-A	high	140	-	-
	low	34	1.80 (1.12-2.88, p=0.015)	1.30 (0.61-2.73, p=0.497)

Number in data frame = 138, Number in model = 65, Missing = 73, Number of events = 26, Concordance = 0.822 (SE = 0.034), R-squared = 0.363 (Max possible = 0.942), Likelihood ratio test = 29.321 (df = 6, p = 0.000). Determining significance independent of the clinical model, top four variables form the clinical model each gene is assessed against.

Table 13 CoxPH cancer specific survival analysis in univariate and multivariate model.

VARIABLE	LEVELS	N (%)	HR (UNIVARIABLE) N = 176	HR (MULTIVARIABLE) N = 78
Age	Mean (SD)	66.8 (11.1)	0.99 (0.98-1.01, p=0.515)	1.01 (0.98-1.05, p=0.440)
Sex	Female	31 (17.6)	-	-
	Male	145 (82.4)	1.65 (0.90-3.01, p=0.105)	0.77 (0.31-1.91, p=0.576)
pT	0-1	38 (29.5)	-	-
	2-4	91 (70.5)	1.39 (0.78-2.48, p=0.266)	0.48 (0.17-1.37, p=0.172)
pN	0	48 (36.9)	-	-
	1-3	82 (63.1)	2.56 (1.45-4.52, p=0.001)	3.83 (1.41-10.45, p=0.009)
pM	0	77 (92.8)	-	-
	1	5 (6.1)	7.07 (2.75-18.19, p<0.001)	8.04 (2.99-21.61, p<0.001)
<i>CSDE1</i>	low	41 (74.4)	-	-
	high	97 (25.6)	1.70 (1.00-2.91, p=0.051)	2.06 (0.78-5.43, p=0.143)
<i>HLA-E</i>	high	121 (87.7)	-	-
	low	17 (12.3)	2.09 (1.05-4.15, p=0.036)	1.09 (0.28-4.31, p=0.899)
<i>SPPL2A</i>	low	34 (24.6)	-	-
	high	104 (75.4)	1.52 (1.01-2.30, p=0.045)	2.73 (0.61-12.26, p=0.190)
<i>HLA-DPA1</i>	high	123 (89.1)	-	-
	low	15 (10.9)	2.37 (1.11-5.05, p=0.025)	1.15 (0.18-7.23, p=0.878)
<i>PSMB9</i>	high	122 (88.4)	-	-
	low	16 (11.6)	2.76 (1.42-5.36, p=0.003)	0.55 (0.14-2.25, p=0.409)
<i>HLA-B</i>	high	88 (63.8)	-	-
	low	50 (36.2)	1.77 (1.05-2.97, p=0.033)	1.22 (0.48-3.09, p=0.672)
<i>HLA-DRA</i>	high	102 (73.9)	-	-
	low	36 (26.1)	2.07 (1.19-3.61, p=0.010)	0.75 (0.23-2.47, p=0.632)
<i>PSMB8</i>	high	122 (88.4)	-	-
	low	16 (11.6)	2.32 (1.17-4.62, p=0.016)	0.43 (0.12-1.58, p=0.204)
<i>HLA-DRB5</i>	high	115 (83.3)	-	-
	low	23 (16.7)	2.08 (1.04-4.16, p=0.038)	1.39 (0.43-4.45, p=0.579)
<i>HLA-DQA1</i>	high	111 (80.4)	-	-
	low	27 (19.6)	1.85 (0.99-3.46, p=0.054)	0.88 (0.35-2.22, p=0.790)
<i>HLA-DRB1</i>	high	116 (84.1)	-	-
	low	22 (15.9)	2.71 (1.42-5.15, p=0.002)	1.05 (0.35-3.13, p=0.927)
<i>TAPBP</i>	high	116 (84.1)	-	-
	low	22 (15.9)	2.20 (1.18-4.11, p=0.014)	2.02 (0.73-5.62, p=0.176)
<i>HLA-G</i>	high	85 (61.6)	-	-
	low	53 (38.4)	1.77 (1.05-2.98, p=0.031)	1.64 (0.72-3.76, p=0.240)
<i>HLA-A</i>	high	116 (84.1)	-	-
	low	22 (15.9)	2.06 (1.11-3.85, p=0.023)	0.62 (0.18-2.17, p=0.452)
<i>ERAP2</i>	high	118 (85.5)	-	-
	low	20 (14.5)	2.05 (1.03-4.08, p=0.040)	2.15 (0.64-7.21, p=0.214)
<i>ERAP1</i>	high	29 (21.0)	-	-
	low	109 (79.0)	2.55 (1.09-5.94, p=0.030)	1.38 (0.45-4.25, p=0.577)
<i>CALR</i>	high	35 (25.4)	-	-
	low	103 (74.6)	2.11 (1.04-4.31, p=0.039)	1.51 (0.55-4.18, p=0.425)
<i>TAPBPL</i>	high	92 (66.7)	-	-
	low	46 (33.3)	2.00 (1.19-3.37, p=0.009)	1.00 (0.44-2.29, p=0.994)
<i>CTSL</i>	high	31 (22.5)	-	-
	low	107 (77.5)	2.08 (0.98-4.42, p=0.056)	1.76 (0.57-5.39, p=0.324)
<i>LGMN</i>	high	106 (76.8)	-	-
	low	32 (23.2)	2.55 (1.48-4.40, p=0.001)	1.33 (0.42-4.19, p=0.625)
<i>PSMB10</i>	high	112 (81.2)	-	-
	low	26 (18.8)	2.71 (1.52-4.86, p=0.001)	1.63 (0.64-4.12, p=0.305)
<i>CIITA</i>	high	64 (46.4)	-	-
	low	74 (53.6)	1.68 (0.98-2.87, p=0.058)	0.71 (0.31-1.60, p=0.405)
<i>CD74</i>	high	125 (90.6)	-	-
	low	13 (9.4)	2.88 (1.36-6.12, p=0.006)	1.66 (0.20-13.84, p=0.639)
<i>MR1</i>	high	85 (61.6)	-	-
	low	53 (38.4)	1.68 (1.00-2.82, p=0.051)	1.45 (0.59-3.54, p=0.417)
<i>CTSS</i>	high	86 (62.3)	-	-
	low	52 (37.7)	1.99 (1.18-3.35, p=0.010)	1.47 (0.58-3.72, p=0.415)
<i>RFX5</i>	high	29 (21.0)	-	-
	low	109 (79.0)	2.81 (1.21-6.55, p=0.017)	2.33 (0.78-6.97, p=0.132)

Number in data frame = 138, Number in model = 65, Missing = 73, Number of events = 26, Concordance = 0.822 (SE = 0.034), R-squared = 0.363 (Max possible = 0.942), Likelihood ratio test = 29.321 (df = 6, p = 0.000). Determining significance independent of the clinical model, top five variables form the clinical model each gene is assessed against.

Table 14 CoxPH disease-free survival analysis in univariate and multivariate model.

Determining significance independent of the clinical model, top five variables form the clinical model each gene is assessed against.

VARIABLE	LEVELS	N (%)	HR (UNIVARIABLE) N = 176	HR (MULTIVARIABLE) N = 78
Age	Mean (SD)	66.8 (11.1)	0.99 (0.98-1.01, p=0.515)	1.01 (0.98-1.05, p=0.440)
Sex	Female	31 (17.6)	-	-
	Male	145 (82.4)	1.65 (0.90-3.01, p=0.105)	0.77 (0.31-1.91, p=0.576)
pT	0-1	38 (29.5)	-	-
	2-4	91 (70.5)	1.39 (0.78-2.48, p=0.266)	0.48 (0.17-1.37, p=0.172)
pN	0	48 (36.9)	-	-
	1-3	82 (63.1)	2.56 (1.45-4.52, p=0.001)	3.83 (1.41-10.45, p=0.009)
pM	0	77 (92.8)	-	-
	1	6 (7.2)	7.07 (2.75-18.19, p<0.001)	8.04 (2.99-21.61, p<0.001)
<i>RFX5</i>	high	30 (21.7)	-	-
	low	108 (78.3)	2.48 (1.13-5.47, p=0.024)	2.65 (0.99-7.13, p=0.053)
<i>CTSS</i>	high	27 (19.6)	-	-
	low	111 (80.4)	2.27 (0.97-5.28, p=0.058)	0.95 (0.29-3.12, p=0.927)
<i>CD74</i>	high	125 (90.6)	-	-
	low	13 (9.4)	2.04 (1.00-4.17, p=0.050)	1.34 (0.36-5.05, p=0.663)
<i>PSMB10</i>	low	100 (72.5)	-	-
	high	38 (27.5)	1.79 (1.03-3.13, p=0.040)	0.91 (0.37-2.22, p=0.878)
<i>LGMN</i>	high	70 (50.7)	-	-
	low	68 (49.3)	1.93 (1.14-3.28, p=0.015)	1.44 (0.64-3.23, p=0.375)
<i>TAPBPL</i>	low	121 (87.7)	-	-
	high	17 (12.3)	2.05 (1.04-3.85, p=0.036)	1.96 (0.44-8.33, p=0.375)
<i>HLA-DPA1</i>	high	125 (90.6)	-	-
	low	13 (9.4)	2.54 (1.27-5.09, p=0.008)	1.87 (0.63-5.57, p=0.261)
<i>HLA-DOA</i>	high	119 (86.2)	-	-
	low	19 (13.8)	2.20 (1.18-4.11, p=0.013)	0.63 (0.22-1.83, p=0.401)
<i>HLA-DQA2</i>	high	122 (88.4)	-	-
	low	16 (11.6)	2.26 (1.17-4.38, p=0.015)	2.03 (0.63-6.52, p=0.232)
<i>HLA-DRA</i>	high	126 (91.3)	-	-
	low	12 (8.7)	2.56 (1.24-5.26, p=0.011)	1.67 (0.50-5.59, p=0.406)
<i>HLA-DRB5</i>	high	113 (81.9)	-	-
	low	25 (18.1)	2.52 (1.43-4.44, p=0.001)	1.77 (0.74-4.24, p=0.199)
<i>HLA-DQA1</i>	high	114 (82.6)	-	-
	low	24 (17.4)	2.35 (1.33-4.14, p=0.003)	0.97 (0.42-2.22, p=0.940)
<i>HLA-DRB1</i>	high	117 (84.8)	-	-
	low	21 (15.2)	2.34 (1.27-4.31, p=0.006)	0.76 (0.28-2.07, p=0.591)
<i>TAPBP</i>	high	112 (81.2)	-	-
	low	26 (18.8)	2.16 (1.20-3.92, p=0.011)	2.35 (1.03-5.41, p=0.044)
<i>HLA-A</i>	high	126 (91.3)	-	-
	low	12 (8.7)	2.25 (1.13-4.46, p=0.021)	0.83 (0.28-2.45, p=0.740)
<i>CSDE1</i>	high	120 (87.0)	-	-
	low	18 (13.0)	2.67 (1.40-5.09, p=0.003)	4.39 (1.33-14.53, p=0.015)

Number in data frame = 138, Number in model = 65, Missing = 73, Number of events = 26, Concordance = 0.822 (SE = 0.034),

R-squared = 0.363 (Max possible = 0.942), Likelihood ratio test = 29.321 (df = 6, p = 0.000)

3.3.9 APM epi-genomic aberrations.

Survival analysis suggests several genes in both MHC pathways as well as regulators of their expression are associated to differential survival outcomes, however, it is unknown what mechanism OAC may be leveraging to alter the expression of these genes. Therefore, I have explored the epigenetics (Methylation) of genes of interest among my candidate genes within the TGCA-ESCA dataset. Firstly, I observed the methylation across the HLA locus, finding ten CpGs linked to MHC genes which exhibit differential methylation between normal oesophageal tissue and OC. Six of these differentially methylated CpGs are associated to HLA class II genes *HLA-DPA1/B1* with the remain four CpGs linked to *HLA-G* and *HLA-F* genes, all of these sites are significantly hyper-methylated compared to normal tissue (p<0.001; see **Table 15**).

Table 15 significant differentially methylated CpGs of the HLA-locus on chromosome 6.

PROBE	LOG2(FOLD CHANGE)	P.VALUE	ADJ.PVALUE	START	END	GENE
CG18914211	1.951	7.20E-06	0.000209	29828038	29828039	HLA-G
CG03521696	1.476	2.00E-06	0.000111	29827818	29827819	HLA-G
CG19990651	0.921	0.00017943	0.00163	33080778	33080779	HLA-DPA1; HLA-DPB1
CG09234582	0.743	0.000391344	0.00286	33080509	33080510	HLA-DPA1; HLA-DPB1
CG01132696	0.878	2.63E-05	0.000444	33080781	33080782	HLA-DPA1; HLA-DPB1
CG06437840	1.10	0.000254439	0.00209	33080752	33080753	HLA-DPA1; HLA-DPB1
CG26645432	0.622	0.000384841	0.00283	33080725	33080726	HLA-DPA1; HLA-DPB1
CG00126638	0.772	1.66E-06	0.000103	29827806	29827807	HLA-G
CG20617328	-0.847	0.000242639	0.00203	33072833	33072834	HLA-DPA1
CG24177217	-0.709	4.95E-06	0.000171	29734276	29734277	HLA-F; HLA-F-AS1

Next, I moved forward to assess the differential methylation of candidate genes which possessed an association with survival via mRNA expression. Within the MHC class I candidate genes five genes were selected for analysis based on their impact on survival.

Among the HLAs of the MHC pathway only *HLA-A* and *HLA-G* demonstrated a significant increase in methylation value comparing OAC to normal tissue (*HLA-A*: Normal B-value mean 0.1972; Tumour B-value 0.4814; $p < 0.0001$. *HLA-G*: Normal B-value mean 0.2147; Tumour B-value 0.5560; $p < 0.0001$. See **Figure 24**); *HLA-B* was examined for differential methylation between normal and OAC finding no significant differences. HLA-complex assemble gene *CALR* and peptide generation gene *ERAP2* was also assessed for methylation differences in normal and OAC tissue finding a significant increase in *CALR* methylation among tumour samples, but no significant methylation difference in *ERAP2* (*CALR*: Normal B-value mean 0.7394; Tumour B-value 0.8472; $p = 0.0173$. *ERAP2*: Normal B-value mean 0.9187; Tumour B-value 0.8881; $p = 0.4994$. See **Figure 24**).

Two MHC class II HLAs, *HLA-DPA1* and *HLA-DPB1* were assessed for differential methylation between normal oesophagus and OAC tissue finding a significant increase in methylation among OAC samples (*HLA-DPA1*: Normal B-value mean 0.3223; Tumour B-value 0.7085; $p < 0.0001$. *HLA-DPB1*: Normal B-value mean 0.3324; Tumour B-value 0.7511; $p < 0.0001$. See **Figure 24**).

Four APM gene expression regulators were assessed including a negative regulator *CSDE1* and positive regulators *CITA* and *RFX5*. The negative regulator, *CSDE1*, demonstrated a substantial significant decrease in methylation of the proximal promoter for expression (*CSDE1*: Normal B-value mean 0.5736; Tumour B-value 0.3550; $p < 0.0001$. See **Figure 24**). The positive regulator of MHC class II HLA expression, *CITA*, displayed a significant decrease in methylation in OAC samples compared to normal tissue (*CITA*: Normal B-value mean 0.4355; Tumour B-value 0.2936; $p = 0.0017$. See **Figure 24**), whereas the positive regulator of MHC class I HLA expression, *RFX5*, demonstrated an increased methylation score in OAC compared to normal oesophagus (*RFX5*: Normal B-value mean 0.5685; Tumour B-value 0.6628; $p = 0.0372$. See **Figure 24**). Observing methylation over cancer stage, did not yield any significant differences with tumour progression.

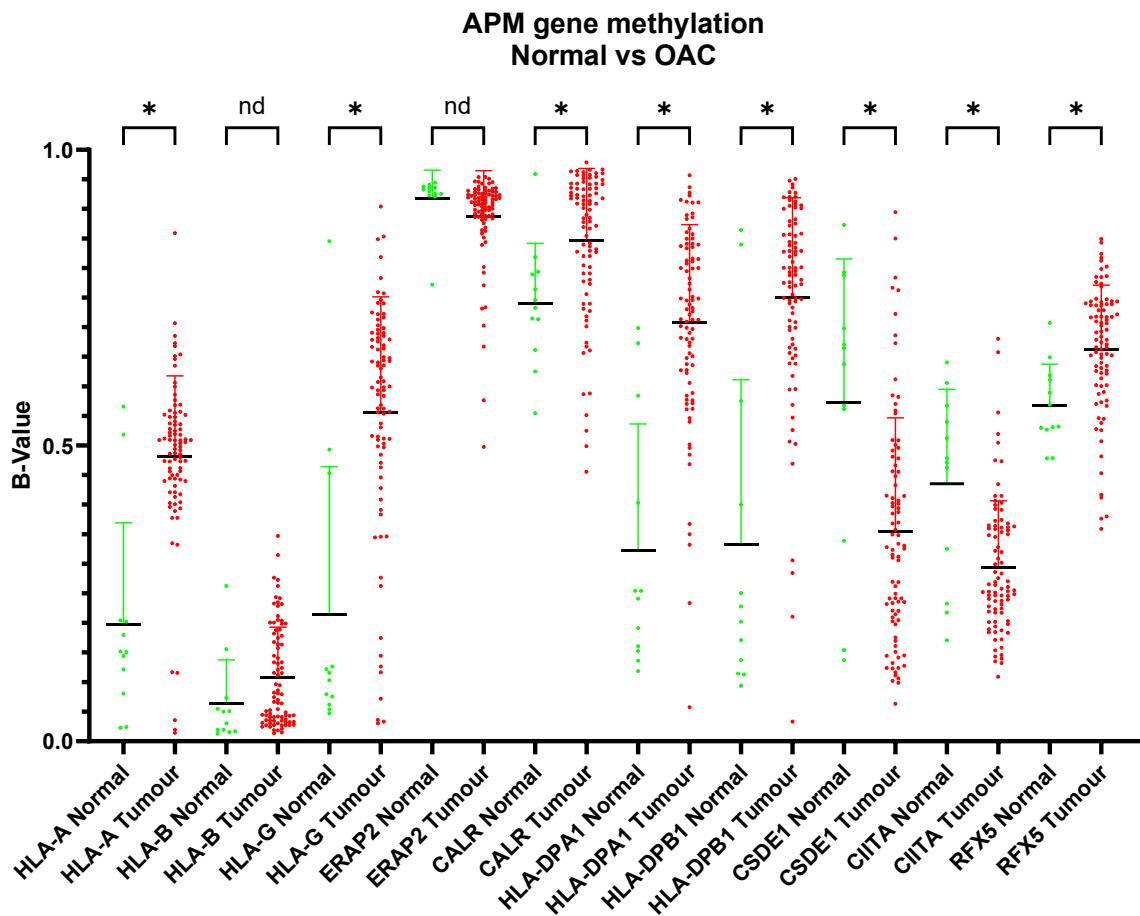


Figure 24 Methylation of APM candidate genes compared between normal oesophagus and OAC samples using Beta-methylation values. Statistical test represented are Kruskal-Wallis tests with FDR correction (ns = non-significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$).

Finally, I explored the correlation between methylation of the APM gene candidates with significantly different methylation values compared to normal tissue presented in **Figure 24**, and their respective mRNA expression. Using this analysis, I found two APM gene expression regulators, *CIITA* and *CSDE1*, possessed a negative correlation between methylation and gene expression (*CIITA*: $R = -0.25$, $p = 0.027$. *CSDE1*: $R = -0.27$, $p = 0.016$. See **Figure 25A/B** and **Table 16**). Among MHC class I HLAs, *HLA-B* possessed an unexpected positive correlation between methylation and expression, whereas *HLA-G* possess the expected negative methylation to expression correlation (*HLA-B*: $R = 0.22$, $p = 0.049$. *HLA-G*: $R = -0.29$, $p = 0.0095$. See **Figure 25C/D** and **Table 16**).

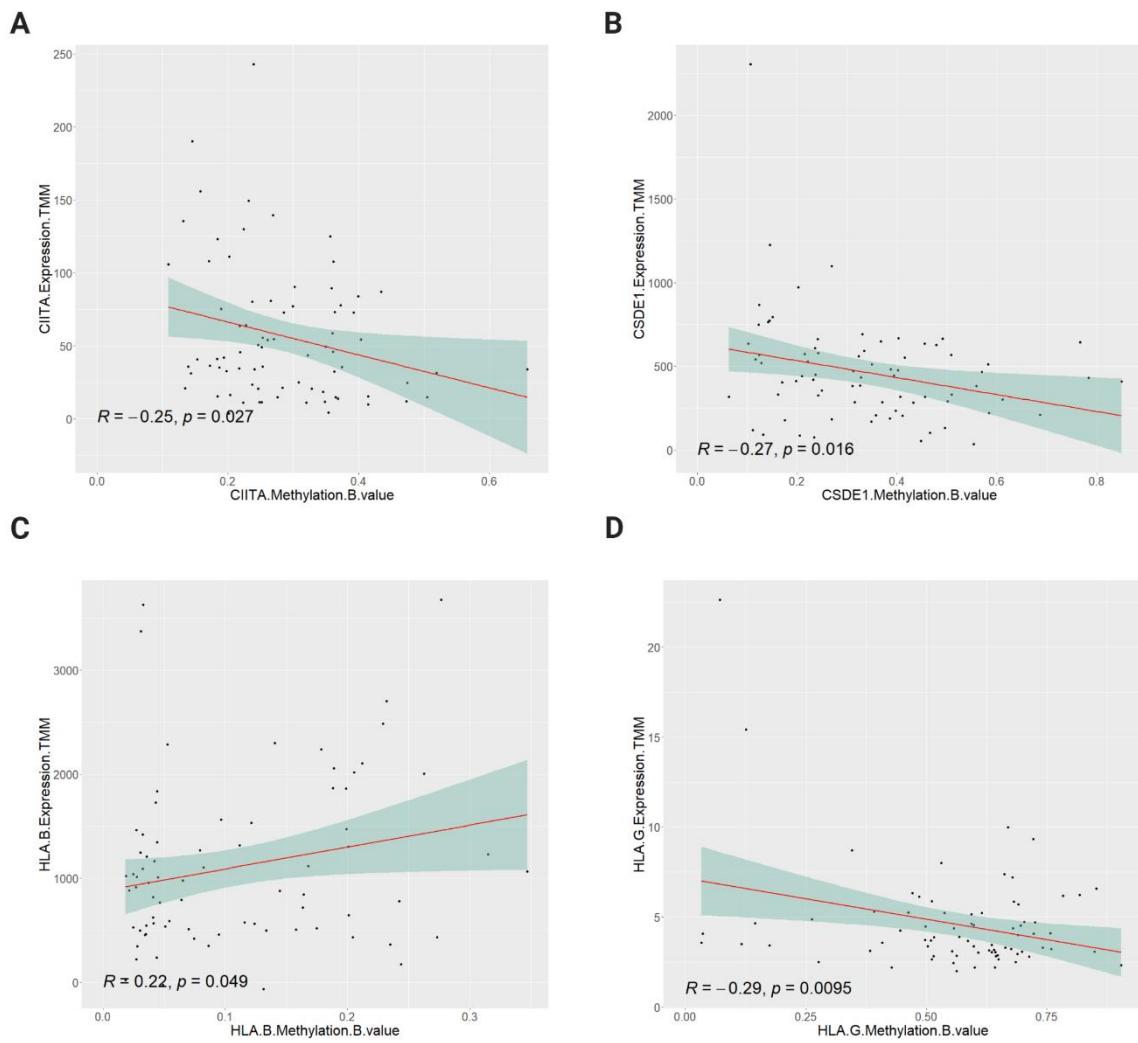


Figure 25 Scatter plot of significant methylation to mRNA expression (TMM) correlations of APM candidates which possessed significantly different methylation in OAC compared to normal tissue. A: CIITA. B: CSDE1. C: HLA-B. D: HLA-G.

Table 16 methylation-expression (mRNA: TMM) correlations from APM candidates with differential methylation in OAC tumour tissue and normal oesophagus.

GENE	R	P VALUE
HLA-A	0.1	0.37
HLA-B	0.22	0.049
HLA-G	-0.29	0.0095
ERAP2	-0.078	0.5
CALR	0.02	0.86
HLA-DPA1	0.02	0.86
HLA-DPB1	0.03	0.8
CSDE1	-0.27	0.016
CIITA	-0.25	0.027
RFX5	-0.1	0.39

3.4 Discussion

3.4.1 Mutation incidence of APM genes in OAC is infrequent.

In this Chapter, the work aimed to identify somatic APM genomic alterations with clinical value. Although mutation analysis did not yield many recurrent APM mutations, there were several of interest in this study. The *B2M* mutations may be of interest as mutations in *B2M* may result in loss of MHC class I expression, which would benefit hyper-mutated tumours through the reduction in cancer antigen presentation. This hypothesis appears to fit with our analysis showing a much higher mutation count a the *B2M* mutated case. One study specifically identifies a *B2M* mutation found in our analysis resulting in a Y46Cfs*10 reporting a lack of *B2M* expression (Immunohistochemistry) in colorectal cancer patient samples harbouring this mutation (238). This may represent a cancer immune evasion mechanism as *B2M* functions in the assembly of MHC I molecules on the ER, lack of *B2M* expression would then restrict cancer antigen presentation to tumour infiltrating lymphocytes preventing an anti-tumoral response. Additionally, cases containing an APM mutation had increased Tumour mutational burden, this could suggest cancer with higher mutational burden seek somatic mutation of APM components to elicit immune evasion or higher mutational burden simply results in a statistically higher probability of APM mutation. Furthermore, we identified mutations of the immunoglobulin C1-set domain in *HLA-B* of OAC cases; specifically, a mutation in the 220–227 α3 loop, E222 has been shown to abrogate MHC I and Tapasin interaction, resulting in loss of function (239). Conversely, mutation in *HLA-B* and *B2M* have been associated to poorer prognosis and further could negatively impact the outcome for immunotherapies as somatic mutations of these APM genes within the TCGA dataset (Pan-cancer) have been associated to high mutation burden, increased neoantigen loading, and higher NK cell infiltrate with CD8+ T cell infiltration; these effects are associated with immune evasion mechanisms. *B2M* mutations in this instance reduces the overall level of surface MHC-I molecules while mutations in *HLA* molecules disturb the overall composition of the MHC-I complex landscape, aiding immune surveillance escape (240). However, this study did not explore OAC or OSCC, instead opting to analyse colorectal, ovarian and lung cancer among others, which in the case of colorectal cancer, T cell infiltrate is associated to higher tissue inflammation which is further associated to poor prognosis (241).

3.4.2 Copy number events are not informative in OAC but are frequent among APM genes of the HLA-locus.

Copy-number analysis yielded two interesting subgroups, an amplification group, and a deletion group, which share the copy-number events across a gene subset in the TCGA/ICGC cohort which was present in approximately 7% of cases. Segment analysis identified genes within a gene set of interest, within the minimally affected genomic region, were located proximally to each other, and shared the same copy-number segments, furthermore, these genes were proximally located to *VEGFA* a previously identified frequently amplified gene in OAC (242).

However, the incidence of these groups was much higher in OAC than OSCC, and the number of copy-number events in total was found to be significantly higher in OAC; this greater number of events may be due to the complexity of copy-number segments found in OAC compared to OSCC resulting from OACs characteristic chromosomal instability (243). Overall, the amplification and deletion groups may be of interest for further study to determine the impact on immune populations via deconvolution analysis.

The APM gene expression regulator candidates possessed a significant number of copy-number events with the highest frequency being *IRF1* consisting of mostly deletions. This is interesting as loss of *IRF1* could result in low APM expression, conversely *IRF1* has also been reported to inhibit anti-tumoral responses via the upregulation of PD-L1 in tumour cells (244, 245).

3.4.3 APM gene candidates and differential expression in OAC.

In the expression analysis of the TCGA/OCCAMS OAC cohort the APM gene candidates did not possess significant differential expression compared to normal stomach samples. However, this lack of finding may be due to methodological factors such as using unmatched normal samples from the GTEx dataset, or the lack of an appropriate normal tissue. Unlike OSCC where normal squamous oesophageal epithelium forms an appropriate normal control, cells which derive into OAC are more like that of stomach epithelia, leading to difficulty in producing an accurate DEA. Additionally, the known downregulation of APM machinery in OSCC has mostly been associated via immunohistochemistry methodology rather than RNA-seq methods making comparisons difficult between the histological subtypes (148).

3.4.4 Expression of APM gene candidates associates with overall, cancer-specific, and disease-free survival in OAC.

From assessment of the literature performed in the general introduction (Chapter 1), a comprehensive analysis of the APM gene expression and their respective associations with clinical outcome had not previously been published, therefore forming a substantial knowledge gap, in OAC. This was addressed, by elucidating the association between the list of curated APM genes and clinical outcomes including overall survival, CSS and DFS in both a univariate and multivariate model with important co-variates used in the clinic (Age, Sex, pT, pN and pM) and in biomarker studies in the literature.

Firstly, some of the regulators of MHC expression possessed a significant association with outcome. For instance, high *CSDE1* expression was significantly associated with overall survival and DFS in both univariate and multivariate analysis. This observation fits with the known role of *CSDE1* downregulating MHC class I expression via stabilising TCPTP, a phosphatase kinase enzymes, which in turn dephosphorylates STAT1 in the JAK/STAT pathway downstream of IFN- γ signalling (145). The IFN- γ signalling pathway is a significant driver of MHC class I expression, ultimately activating the ISRE regulatory promoter elements allowing for further downstream promotion of MHC class I expression by the known trans-activators of MHC class I expression (*RFX5*-family, *NLRCS5* and *IRF-1*) (130, 142).

Prior publications have demonstrated *CSDE1* acts as a master regulator of MHC class I expression, in cancer, the greater expression of *CSDE1* in a melanoma cell line resulted in high dephosphorylation of STAT1 preventing translocation of STAT1 into the cell nucleus downstream of the JAK/STAT pathway as prior described (145).

Thereby, my analysis and literature suggest *CSDE1* expression is OAC significantly associates with poor outcomes, likely via downregulating MHC class I expression resulting in immune evasion. Interestingly, these observations present an opportunity for targeting *CSDE1* in OAC. A recent paper presented a study of oncolytic viruses, where the authors were successful in creating a selection pressure towards an escape-associated tumour antigen via infection with a mutant form of *CSDE1* which was highly expressed and could be a potential target for cancer vaccines, adoptive T-cell and CAR-T cell therapies (146). This may have potential in OAC, as *CSDE1* is not a known target gene

in OAC, nor a recurrently targeted passenger gene in our data (only 1 of 502 cases with a missense mutation).

Expression of RFX5-family genes including *RFX5* and *RFXAP* in the data analysis, identified association with overall survival, CSS and DFS. The *RFX5*-family genes play a significant role in the regulation of MHC class I expression forming the RFX-complex (*RFX5*, *RFXAP*, *RFXANK*) within the MHC class I enhanceosome together with *NLRC5* trans-activator, *ATF1/CREB* and the NFY-complex which bind to the SXY-module for promoting MHC class I expression (142). *RFX5* has prior been published as a prognostic biomarker and associated with immune infiltration in stomach adenocarcinoma. Within that study, expression of *RFX1*, *RFX3*, *RFX4*, *RFX5*, *RFX7* and *RFX8* was significantly elevated in STAD tissue versus adjacent normal tissue. Moreover, patients with high *RFX5* and *RFX7* expression had a better overall survival, first progression, post-progression survival and significantly associated with the abundance of immune cells, the expression of immune biomarkers and tumour mutational burden (246).

Evidence of the role of *RFXAP* in cancer is sparse, a single study identified *RFXAP* expression correlated with tumour stage and poor prognosis, but associated these findings with *RFXAP* overexpression upregulating *KDM4A* and attenuated methylation of H3K36, thereby impairing DNA repair and enhancing the DNA damage induced by fisetin (247). Considering my findings and the literature, I suggest *RFX5* associates to survival outcomes by regulating factor of MHC class I expression in OAC, this finding is justified by the recent publication of its function in gastric cancers as prognostic biomarker associated with immune infiltration (248). Further investigation of the expression of *RFX5* and immune abundance may identify parallels with the stomach adenocarcinoma study and OAC, the regulatory role of this APM regulator on MHC class I expression will be assessed among the other APM regulators in the chapter 4 and 6. On the other hand, *RFXAP* could not be validated by the literature and its association with survival outcomes in OAC identified here may be only representative of the role of the MHC class I enhanceosome or potentially alternative roles beyond APM gene regulation.

Lower *CITA* expression was associated to poorer overall survival and CSS only in univariate analysis. *CITA* is a transcriptional coactivator that regulates γ -interferon-activated transcription of MHC class I and II genes, of which mutation has been prior

associated to autoimmune diseases such as Type II bare lymphocyte syndrome (249, 250). In lung adenocarcinomas, loss of *CIITA* was associated to decreased MHCII expression, low T cell infiltration and converted tumours from anti-PD-1 sensitive to anti-PD-1 resistant (251). Additionally, high *CIITA* expression has been prior associated to high MHC-II expression in Epstein–Barr virus-associated stomach adenocarcinomas, potentially contributing to the highly immunogenic tumour microenvironment noted in this specific tumour subset (252). Combining the known literature and my findings could suggest *CIITA* plays a significant role in the immunogenicity of OAC tumours. Thereby, cases with poor *CIITA* expression may also exhibit low anti-tumoral immunity leading to poorer patient outcomes. Further analysis to observe the distribution of immune cells within the OAC microenvironment among cases of low/high *CIITA* expression may yield further insight into the overarching role in the immunogenicity of tumours to identify parallels with stomach adenocarcinomas (See chapter 4).

Within my analysis *IRF-1* and *NLRCS5* expression did not appear as a prognostic indicator of survival outcome. This provides an interesting insight into the function of these genes in OAC may not be intrinsic to modulation of MHC class I expression. Whereas, the prior mentioned *CSDE1* gene may be functionally responsible for said modulation as a master regulator, thereby it is of import to assess the correlation between the known APM expression regulators and APM expression. This is especially important as brief examination of the OE19 OAC cell line demonstrates a loss in *IRF-1* correlating to substantially lower MHC class I expression in comparison to other OAC cell lines (OE33 and FLO-1) (253).

Overall, these findings support the premise of APM expression regulators may form a prognostic marker of survival outcome in OAC, yet whether these genes may form a therapeutic target in OAC is yet to be elucidated. *CSDE1* is of interest due to its ‘master regulator’ role in MHC class I expression, which I will follow up in modulating *CSDE1* expression in OAC cell lines.

MHC class I gene expression have demonstrated significant associations with survival outcomes in other diseases including OSCC and gastric cancers, and link to the upregulation of adaptive anti-tumoral immune responses. Furthermore, MHC class I gene expression downregulation has prior been associated with cancer pathogenesis and alluded to be important for the efficacy of immune checkpoint blockade therapy.

Within my analysis I found many parallels in the importance of MHC class I expression in survival outcomes reported in the literature of other cancers.

The consequence of MHC class I HLA-antigen loading complex assembly/antigen presentation gene expression prior published mirrors my finding reported in my OAC cohort expression. Specifically, the low expression of *HLA-A/B/E/G* displayed a correspondence with poorer survival outcomes, similar to a prior meta-analysis publication in gastric cancers where HLA class I overexpression possessed a significant positive association with OS (254). However, the same study notes negligible impact of HLA class I overexpression on DFS, opposed to my findings in *HLA-A* which demonstrated a link between low expression and shorter DFS in univariate analysis. This could indicate minor differences in the importance of HLA class I expression between OAC and gastric cancers in survival outcomes, however, the concordance between OS and HLA class I expression in gastric cancer, the most similar cancer to OAC, lends confidence to my findings in the expression survival associations highlighted in my analysis of OAC.

These findings among HLA class I molecules suggest CD8 T cell recognition via HLA in cancers may have a profound impact on overall anti-tumoral immunity in OAC, however this must be explored further to identify modulation of T cell subpopulations in OAC due to HLA class I expression, which will be explored via digital cytometry in chapter 4.

The TAPASIN genes (*TAP1*, *TAP2*, *TAPBP*, *TAPBPL* and *TAPBPR*) have prior been associated to survival outcomes and immune modulation in cancers including colorectal, malignant melanoma, head and neck squamous cell carcinoma (HNSCC), renal cell carcinoma, colorectal carcinoma, glioblastoma, lung carcinoma, and neuroblastoma (255-257). Downregulation of TAPASIN genes has been closely tied to decreased HLA class I presentation, restriction in neo-antigen repertoire and diminished CD8+ T cell responses in cancer (255, 256). Within my data analysis only *TAPBP* and *TAPBPL* demonstrated altered survival outcomes within OS, DFS and CSS; *TAPBP* functionally acts to localise HLA molecules to the TAPASIN translocon in the ER allowing for optimal antigen loading, low expression of *TAPBP* has been linked to poor patient survival in a pan-cancer analysis, concordant with my OAC cohort findings of low *TAPBP* resulting in shorter DFS (258).

TAPBPL presents a more complex survival interaction with low expression associating to shorter OS and CSS, but high expression linking to poorer DFS. Examining the literature higher *TAPBPL* expression has been prior correlated to improved OS in breast cancer, forming part of a prognostic model; however, *TAPBPL* has also been implicated as a novel regulator within T cells resulting in inhibition of T cell responses and proliferation, unfortunately, as my OAC cohort contains only bulk-RNA I could not determine the distribution of *TAPBPL* expression across the cellular subtypes, leaving open the possibility *TAPBPL* expression may represent a positive prognostic factor when expressed in cancer cells (displayed in altered OS and CSS), but a negative prognostic factor when expressed in T cell populations (259, 260). Exploration of differential immune subpopulations due to *TAPBPL* expression via digital cytometry may indicate whether this is a possibility and worth further investigation, this is explored in **Chapter 4**.

Within my analysis of the OAC cohort low expression of *CALR* resulted in shorter OS and CSS; *CALR* functions in the early assembly of HLA class I molecules prior to association to the TAPASIN translocon, several publications have associated the decreased expression of *CALR* with either poorer or improved prognosis depended on the disease type. For example, within colorectal, AML, glioblastoma, NSCLC, ovarian, urothelial cancers, high expression of *CALR* associates to improved survival outcomes; conversely, in gastric, bladder, MCL, neuroblastoma and pancreatic cancers, the increased expression of *CALR* corresponds to poorer patient outcomes (261-276). Thereby, my analysis yields a striking result as despite the similarity between gastric, pancreatic cancer and OAC, increased *CALR* expression forms a positive prognostic indicator of survival; this result finds validation with published analysis of OAC samples from the TCGA dataset demonstrating improved relapse free survival with greater *CALR* expression (277). This analysis in combination with the literature could support *CALR* as positive prognostic factor in OAC, however, it is important to note the roles *CALR* can perform outside of antigen presentation, including epithelial-to-mesenchymal transition (EMT) and metastasis, reported in breast and gastric cancer, which is not assessed in my analysis (278, 279). Importantly, assessing the impact of *CALR* on immune cell subpopulation distribution may indicate whether the function of *CALR* in improved OAC patient outcomes is immune related (presented in chapter 4).

Peptide generation is incredibly important in cancer antigen presentation to produce immunogenic neoantigens which can optimally bind to the HLA molecule peptide binding cleft. This importance is reflected in the literature, for example, the ERAP genes (*ERAP1* & *2*) function within the ER to cleave peptides into the appropriate length for binding to MHC class I HLAs (8-11). In cancer *ERAP1* expression can alter the neo-epitope repertoire, as demonstrated by *ERAP1* inhibitors/knockout models which produce increased neoantigen peptide length and greater cytotoxic T cell infiltrate (279-281). Unfortunately, this does not find concordance with my analysis with low *ERAP1* expression corresponding with shorter CSS, however, this cannot reflect inhibition or knockout of *ERAP1* as there is still functional quantity of *ERAP1* in these cases. *ERAP2* similarly to *ERAP1* functions to cleave peptides for optimal HLA binding; in Squamous cell lung carcinoma, high *ERAP2* expression was identified as an independent positive prognostic factor, this is concordant low *ERAP2* expression shorter OS, CSS (282). Conversely, *ERAP2* inhibition in leukaemia has resulted in the presentation of novel epitopes which could enhance the immunogenicity of tumour cells (283, 284). Interestingly, approximately 25% of individuals possess a SNP which results in lack of *ERAP2* expression, which may simulate the altered epitope described in the leukaemia publication, however assessing the samples in the OAC cohort which lacked *ERAP2* expression did not indicate any altered survival outcomes in OAC (284). Further investigation into the immune cell landscape with *ERAP1/2* expression may yet yield interesting findings beyond survival analysis, explored in **Chapter 4**.

The immunoproteasome is vital for peptide generation, degrading proteins in the cytosol before peptide translocation to the ER. Within my mRNA expression analysis three immunoproteasome component genes demonstrated an association with altered survival. Specifically, the low expression of *PSMB8/9* corresponded with shorter CSS, whilst conversely greater *PSMB10* correlated to shorter OS, CSS and DFS. This presents an interesting dichotomy, between the impact of the immunoproteasome components on survival in OAC; exploring the literature reinforces a known dichotomy due to the multiple functions of the immunoproteasome. The functions of the immunoproteasome include peptide generation, maintenance of protein homeostasis and promoting tumorigenic cytokine expression. While playing a positive role in antigen presentation, these components may possess differing roles, for example, *PSMB8/9* expression is

known to associate to improved prognosis due to its role in improved neo-antigen and TAAs presentation. This has been demonstrated on melanoma cells, with downregulation presenting a major mechanism accounting for MHC class I antigen loss in colorectal cancer (284, 285).

Interestingly, the prognostic value of *PSMB10* varies between disease type, for example, greater *PSMB10* expression in pancreatic cancer corresponds to longer OS, whereas, in gastric cancer greater *PSMB10* expression results in poorer OS (286). Without further analysis it is difficult to infer which role each immunoproteasome component is performing; however, this analysis suggests *PSMB8/9* may be improving antigen presentation by producing immunogenic peptides for presentation, whilst *PSMB10* may either be functioning to improve cancer cell survival by protein homeostasis or by promoting tumorigenic cytokine expression. Further analysis observing the immune cell distribution between prognostic expression groups of *PSMB8/9/10* may yield further insight into the function of these immunoproteasome components in OAC.

MHC class II genes are exclusively expressed by professional antigen presenting cells (APCs) including dendritic, macrophages and B cells. Typically, the presence of APCs in the immune microenvironment forms a positive prognostic factor, though the specific subtype of macrophage possesses differing outcomes with M1 macrophages deemed a positive prognostic feature in cancer and M2 a negative factor. In respect to my analysis assessing the impact of MHC class II on survival outcomes may in be indicative markers of two factors, the presence/lack of APCs in the tumour immune microenvironment by aggregate MHC class II expression or differences immune composition. Justifying the later point, prior publications have demonstrated M1 macrophages express higher levels of MHC class II molecules due to macrophages MHC class II expression being inducible rather than constitutive like dendritic cells. Thereby, identifying the significant survival associations within MHC class II gene expression should be paired with digital cytometry to deconvolute the specific cell of origin potentially responsible for altered survival outcomes (287-289).

Firstly, among the HLA-antigen loading complex assembly/antigen presentation genes, nine out of fifteen candidate gene's expression corresponded with altered survival outcomes. Consistently, the low expression of these nine candidates possessed a similar negative impact on survival outcomes, this could support the potential argument that

this gene expression profile represents cases of high or low presence of APCs in OAC tumours or indicate the polarisation state of macrophages in the OAC TIME. This finding will be further explored in chapter 4 via digital cytometry methodology to characterise APCs distribution related to differential expression of MHC class II genes.

Exploring the literature finds support for the survival associations identified by my analysis of MHC class II gene expression in OAC; for instance HLA-DR of which *HLA-DRA* is a component of has been prior identified as an independent positive prognostic indicator in tumour epithelium of OAC. This may provide insight into the reasoning behind OAC being characterised as an immune hot tumour type, as HLA-DR is reported to possess the greatest correlation with cytotoxicity markers in cancer . Additional publications highlight HLA-DR correlates to increased T cell infiltrate in lung adenocarcinoma and forms a marker of immune hot tumours predicting responses to checkpoint blockade therapy in NSCLC (196, 290, 291). In combination with published literature, the expression of *HLA-DRA* in OAC may represent immune hot tumours which could form exceptional targets for immunotherapy. These studies findings can be further extrapolated into my findings in *HLA-DRB1/5* which when lowly expressed associated to poor patient outcomes. importantly, the subtype of APC which is responsible for this outcome is unknown, further analysis conducted in chapter 4 will explore whether high *HLA-DRA* expression in OAC can be attributed to a specific APC subtype.

The literature surrounding the importance of *HLA-DMA* expression in cancer survival is sparely reported, studies so far have focused on the downregulation of these genes via external mechanisms including C-MYC, however, two studies of interest identified downregulation of *HLA-DMA*; firstly, a pan-cancer analysis demonstrated *HLA-DMA* correlates with cytotoxicity markers in cancer, though possess the smallest correlation with cytotoxic activity; the second study demonstrates *HLA-DMA* is downregulated over stage in cutaneous melanoma with higher expression corresponding to greater OS (292, 293). Aligning my findings with these publications suggests *HLA-DMA* may form a component of immune response in OAC resulting in greater survival outcomes via eliciting responses from CD4+ T cells, assessment using digital cytometry presented in chapter 4 will explore this interaction.

According to a recent publication *HLA-DPA1* expression possesses the third greatest correlation with cytolytic behaviour, thereby within the scope of OAC the expected result would form an association between increased *HLA-DPA1* expression and improved survival outcomes due to improved immune suppression of OAC tumours. My analysis demonstrates low *HLA-DPA1* within OAC tumours corresponds with shorter OS, CSS and DFS finding concordance with the known biological role of *HLA-DPA1* in eliciting CD4+ T cell responses, to be confirmed in digital cytometry analysis in chapter 4.

In my analysis I uncovered low *HLA-DQ* gene expression in the form of *HLA-DQA1/2* in OAC associated to shorter survival outcomes, aligning this finding to research in others cancers identifies *HLA-DQ* genes highly correlates with cytolytic activity in tumours (293). Despite this, publications in OSCC how linked the expression of *HLA-DQA1* with poor survival outcomes corresponding to increased immune evasion (294, 295). My findings presented here highlight further differences in the TIME of OAC as compared to OSCC and may clarify the role of *HLA-DQ* genes in OAC compared of OSCC, yet further analysis into predicting the cells expressing these genes must be conducted (see chapter 4).

Lastly among the HLA class II molecules, low expression of *HLA-DOA* corresponded to shorter DFS; current literature places *HLA-DOA* as the least immunogenic among its peers, despite this, *HLA-DOA* has been identified as a candidate within a 7-gene signature which predicts improved DFS in Hepatocellular Carcinoma (296).

Decreased expression of *CD74* within the OAC cohort demonstrated a correlation with shorter OS, CSS and DFS, this provides an interesting perspective of the potential role of *CD74* in OAC. Notably, prior literature surrounding *CD74* in cancer either demonstrates the genes role as a positive or negative prognostic factor dependant on the disease type and the role *CD74* plays either within antigen presentation or non-antigen presentation processes such as macrophage migration or mesenchymal epithelial transition (MET). For example in gastric and colon cancers, fibroblasts which expressed *CD74* increase gastric cancer cell proliferation and drove MET of normal gastro-intestinal cells; this finding is in opposition to my analysis (297). Though supporting arguments can be found in breast cancer research which indicate *CD74* expression may be a positive prognostic feature of basal-like breast cancer and prostate cancer (286, 298). *CD74* expression also correlated to high mean density of CD8, CD4 and CD68 TILs in basal-like breast cancer,

and increased macrophages, activated dendritic cells, and neutrophils in gliomas (298, 299). Examination of *CD74* expression survival group via digital cytometry will be used in chapter 4, to assess whether the expression of *CD74* mirrors the increased infiltrate of TILs seen in basal-like breast cancer.

Examination of *SPPL2A* expression within the OAC cohort found high expression corresponds to shorter OS and CSS, this is a particularly interesting result following the *CD74* analysis, as *SPPL2A* functions to cleave *CD74* in the MHC class II compartment. Thereby, by extension the role of *SPPL2A* in OAC may mirror that of *CD74*, where high cleavage prevents the optimal effect of *CD74* in improving survival outcomes; observing the literature find scarce detail on the function of *SPPL2A* in cancers especially within the scope of antigen presentation. However, exploring the human protein database does find in gastric cancers high expression results in longer OS, whereas high expression in pancreatic cancer corresponds to shorter OS (286). This could potentially relate to either *SPPL2A* functions producing an inflammatory microenvironment, altered DC differentiation or the reported negative role of *SPPL2A* in the nuclear translocation of the *ODZ1*, a protein associated to poor patient outcome in glioblastoma (300-302). Thereby, in my further analysis of *SPPL2A* I will determine whether the impact on survival outcomes in OAC is related to any immunological differences in immune cell composition regarding DC differentiation or by the association of the roles of *CD74* and *SPPL2A* in immune cell distributions of OAC (see chapter 4).

The remaining two MHC class II genes, *LGMN* and *CTSS* are both proteases involved in peptide generation for MHC class II HLA loading, interestingly they both share the same trend with survival with low *LGMN* expression corresponding to shorter OS, CSS, DFS and low *CTSS* expression associating to shorter CSS. Mining the literature surrounding *CTSS* and *LGMN* in cancer highlights a discordance between my OAC cohort and other cancers. For example, high expression of *CTSS* in gastric and lung cancer relates to gastric cancer cell migration, invasion and poor survival outcomes in both diseases. However, the role of *CTSS* in these cancers has been associated to ECM remodelling, cell migration and producing an inflammatory microenvironment, rather than its antigen presentation role, even forming a potential treatment vector through inhibition (303-307). Conversely, *CTSS* in the case of OAC may bolster immune responses as shown in other cancers where the

expression of *CTSS* correlates to increased immunogenicity of tumours and immune infiltrate including oesophageal, gastric and pancreatic cancers, further correlating to TMB (308). This opens an interesting question on the role of *CTSS* as a positive prognostic factor in OAC and the potential that the high TMB of OAC is responsible for the disparity of survival outcomes between OAC and other cancers. Mirroring the literature of *CTSS*, *LGMN* is also established as a negative prognostic factor in cancers including gastric, colorectal and breast cancers, leveraging this effect by promoting tumour development and TAMs stimulated cell proliferation, migration, and invasion (309-311). For these two key proteases it is key to follow up the survival analysis with digital cytometry to explore the alteration of the TIME due to their expression, this could potentially unravel the roles *CTSS* and *LGMN* are playing in OAC which produces this dichotomy in survival outcomes between OAC and other cancers.

The final APM gene candidate to exhibit altered survival outcomes with expression belonged to an alternative APM pathway, high *MR1* expression in OAC presents as a positive factor in survival outcome, with only recent publications beginning to explore the function role of *MR1* in tumour immune. One such study explored the novel role *MR1* possesses in eliciting responses from MAIT cells to produce a specific anti-tumoral response in a pan-cancer cell panel, suggesting *MR1* is capable of presenting highly immunogenic lipids to MAIT cells to target cancer (312). Interestingly, the survival impact of *MR1* expression in OAC has not yet been explored in the literature, thereby my analysis suggests *MR1* may be capable of eliciting immune responses in OAC, unfortunately, by current deconvolution methodology MAIT cells cannot be deconvoluted, yet digital cytometry analysis will endeavour to identify any other adaptions in the TIME due to *MR1* expression (See chapter 4).

3.4.5 Transcription regulation by genome methylation alters the APM landscape of mRNA expression via regulation of known APM gene expression regulators.

To explore the mechanism potentially behind cancer regulation of APM genes which possessed a significant association I conducted an epigenetic analysis of ten APM genes comprised of five MHC class I, two MHC class II HLAs and three APM gene expression regulators. Notably, the methylation of CpG sites associated to eight of the ten gene explored genes were differentially methylated in comparison to normal oesophagus.

Of the eight differentially methylated genes, six demonstrated significantly higher methylation including MHC class I components *HLA-A/G* and *CALR*, *HLA-DPA1/B1* of the MHC class II HLAs and APM gene expression regulator *RFX5*. Significantly reduced methylation of CpG probes associated to the expression of *CITA* and *CSDE1* was also identified with the remaining genes *HLA-B* and *ERAP2* not reaching a significant association.

Exploring the literature yielded interesting parallels in the methylation of the explored genes, firstly, the methylation of CpG site for *HLA-A* identified in my analysis (cg09803951) has been prior identified as possessing associated with reduced *HLA-A* expression in pre-invasive squamous cell lung cancer (313). However, then assessing the correlation between this CpGs site (cg09803951) methylation and expression within my OAC cohort I did not find a significant correlation. Interestingly, the methylation of *HLA-G* in OAC appeared significantly lower compared to normal tissue, with literature demonstrating similarly events in *HLA-G* de-methylation in other cancers to exploit *HLA-G*'s ability to elicit immune tolerance towards cancer cells and perturbed T/NK cell infiltrate (314, 315). This argument is further supported by the analysis presented displaying a statistically significant correlation between the methylation CpG associated with *HLA-G* and its respective mRNA expression. Despite these results finding concordance with the wider literature, the survival association between low *HLA-G* expression and poorer survival still provides a confusing outlook on the function of *HLA-G* in OAC, highlighting a further need to explore the dynamics of immune subpopulations due to *HLA-G* expression (See chapter 4).

The final MHC class I gene investigated, *CALR*, demonstrates an increase methylation status in OAC tissue, in the current literature *CALR* expression is presented as a negative prognostic characterise of multiple cancers including OAC, thereby, the increase in methylation was unexpected, however the CpG site which displayed the greatest differential methylation in this case did not correlate to expression.

Both MHC class II *HLA-DPA1/B1* possessed increased methylation status in OAC tissue compared to normal oesophageal tissue, however the same methylation site did not correspond to gene expression. Notably, this result finds concordance with CRC and serrated polyposis syndrome where *HLA-DPA1* being hypermethylated compared to normal tissue, conversely, the CpG site the publication differs from that selected within my analysis (cg12858166) (316). This alternative CpG site within my analysis did not yield

either a correlation with expression or differential methylation from normal tissue, which could demonstrate key differences in the regulation of MHC class II expression in OAC compared to CRC (316). Interestingly, the methylation of HLA-DP (including alpha and beta variants) has been prior associated with poorer overall survival in gastric cancer as part of a four-gene methylation signature, which could suggest other CpG sites for *HLA-DP* genes could have an impact on OAC survival due to the similarity between the diseases (317).

Among the APM gene expression regulators, *CIITA* and *CSDE1* demonstrated altered methylation as compared to normal oesophageal tissue. *CIITA*, which functions as a trans-activator of MHC class II expression and has been prior identified as a somatic target for epigenetic silencing in cancer associating to poorer survival and disease progression, demonstrated in rhabdomyosarcoma, haematopoietic tumour cells and gastrointestinal cancers (318, 319). This provides a striking contrast to my OAC cohort results (TCGA only) where methylation of *CIITA* decreased in OAC tumours compared to normal tissue. Furthermore, the methylation of the CpG site associated to *CIITA* did correlate to *CIITA* expression in OAC suggesting APCs may upregulate *CIITA* expression in OAC tumours as a positive immune response or a response to inflammatory cytokines present in OAC tumours. Yet *CIITA* upregulation is not understood adenocarcinomas (320). Reduced methylation of *CSDE1* in my OAC cohort compared to normal oesophagus mirrors a recent publication, which demonstrates cancer modifies methylation at the *CSDE1* locus via a methyltransferase *SYMD3* which mediates H3K4 trimethylation of *CSDE1* locus in mouse melanoma cancer models with a more recent study replicating this finding in oral squamous cell carcinoma (321, 322).

3.4.6 Limitations

During my data mining investigation a few notable limitations arose which require explanation. Firstly, the clinical data could seek improvements in accurate reporting, for instance, several samples did not accurately report the TNM staging within both the TCGA and OCCAMS datasets, requiring these samples to be filtered out, decreasing the overall sample size. This leads to a further issue as although the cohort did possess power of detection for 20% incidence of genomic events (copy number, mutation, mRNA expression). Study of the complete OCCAMS dataset (n=551) did detect B2M mutations as a somatic driver mutation, this was not reflected in my OAC cohort potentially due to the smaller sample size (242). Furthermore, issues in clinical data effected the results of the

multivariate analysis due to the lack of metastasis reporting in the TNM staging resulting in a reduced multivariate model N number.

Unfortunately, there is lack of methylation data within the OCCAMS dataset which severely reduced the number of OAC samples to $n = 87$, which could not afford a decent power of detection. An additional issue involves the complexity of copy number segments over chromosome 6 preventing any attempt to deconvolute the impact of copy number at the single gene level.

3.4.7 Conclusions

This data mining experiment has formed an excellent discovery for the landscape APM genomic defects among OAC patients, despite limitations presented above. Overall, the results of my data analysis suggest somatic mutations is infrequent among OAC patients meaning APM machinery proteins are intact in OAC upon expression. Somatic copy number in OAC over the HLA locus presented a complex issue where multiple HLAs of both the MHC class I and II were co-copied on the same segment, furthermore this may be driven by selection of *VEFGA*, a well understood copy number amplification in OAC.

This led to exploring the mRNA expression, finding the expression of APM gene candidates in OAC associates to survival outcomes. Additionally, the regulation of APM gene expression was explored, finding *CSDE1*, *IRF1*, *NLRC5*, *RFX5* and *CIITA* all formed a component in APM gene expression regulation via correlation analysis, with *CSDE1* and *RFX5* demonstrating association with survival outcomes. Importantly, the *CSDE1* survival association withstood multivariate testing, suggesting *CSDE1* may be an independent prognostic marker of survival. To explore the regulation of APM gene expression regulators I observed the methylation status of these genes, with the most striking result demonstrating demethylation of the *CSDE1* locus in OAC compared to normal oesophageal tissue. I have characterised the OAC APM genomic landscape, finding APM gene expression and revealed several targets to analyse further, including *HLA-A/B*, *TAPASIN* genes, and *CSDE1* based on my findings and the wider literature. This research also presents a holistic landscape of APM genomics in OAC which has not prior been presented in such depth, unlike OSCC which has been greatly explored for APM genomic defects. Excitingly, there is merit to further investigate APM gene candidate expression in OAC and its impact on the OAC TIME. The next step in my analysis is to observe the impact of APM gene expression which possessed prognostic significance on the

distributions of immune cell subpopulations using digital cytometry (See Chapter 4), this should potentially further elucidate the role of APM gene expression in the TIME and anti-tumoral immunity.

Chapter 4 Investigating the relationship between immune composition estimates and prognostic APM gene defects.

4.1 Introduction

From the prior chapter, it is understood defects in APM gene expression within OAC tumours can associate with patient overall survival, however, the mechanism by which this survival impact is come about is not yet understood.

This chapter will focus on elucidating whether prognostically informative APM gene expression (determined by OS association from Chapter 3 results) is tied to changes in the tumour immune microenvironment (TIME), estimated by digital cytometry deconvolution analysis to identify the proportions of immune cell subpopulations from the combined TCGA and OCCAMS RNA-seq data.

4.1.1 Deconvolution methods

Determining immune cell compositions in tumour tissue relies on several conventional methods including immunohistochemistry and flow cytometry, however, these techniques often use single markers for cell types such as a CD8+ antibody to detect CD8+ T cells. Enumerating immune subsets using these methods suffer from limitations in phenotypic markers and can be challenging to practically implement and standardise (323) . With recent advancements in in silico methods and computational modelling (i.e., machine learning) allow for predicting fractions of multiple cell types in gene expression profiles in sample mixtures, these methods have been coined as ‘immune profiling by signature matrix deconvolution’ or ‘deconvolution’ for short. Multiple deconvolution computation models have since been developed since the advent of more powerful computation hardware including OLS, nnls, RLR, FARDEEP, and CIBERSORT (324-327). Deconvolution analysis methodology can be divided into two broad categories, reference-based (supervised) methods, which function by using *a priori* defined reference matrix made of expected gene expression profiles of cell types. Mathematically, a reference

matrix is represented as $Hm \times k$, m representing expected value for markers for a particular cell type and k representing cell types known to be present in the sample; using the reference matrix a deconvolution methodology can use non-negative or constrained linear regression methods to dissect cell types based upon the samples gene expression matrix (328).

Reference-free (unsupervised) approaches contrast to referenced-based approaches by using *ad hoc* feature selection, with the variability of each feature determining how informative a feature is in the sample mixture; selected features are then used to dissect cell types (329). A literature search of deconvolution methodology found the top deconvolution methods used in publication (>100 articles) are CIBERSORT and Xcell.

CIBERSORT and the latest version CIBERSORTx are supervised deconvolution methods which infer cell type abundance and cell-type-specific gene expression from RNA profiles of intact tissues; this methodology was first reported in 2015 with CIBERSORTx being reported in 2019 (327, 330). Outside of the initial report, CIBERSORT was first published in a pan-cancer analysis of the TCGA dataset identifying an association between 22 distinct leukocyte subsets and cancer survival (331). CIBERSORTx also provides two distinct reports on immune abundance, namely absolute and fractional values. Fractional scores represent the proportion of a cell type targeted by the analysis, for example the proportion of LM22 (Immune cells) which are CD4+ T helper cells, whereas absolute scores scale cellular fractions to produce a score reflecting each cell type's absolute proportion allowing for direct comparison between samples and cell types (332).

Next Xcell deconvolution, first reported in 2017, outperformed the previous version of CIBERSORT by not being reliant on Affymetrix microarray studies, instead integrating single sample gene set enrichment analysis with deconvolution approaches and could identify 64 cell types opposed to the 22 cell types CIBERSORT can deconvolute.

Importantly CIBERSORTx carries several advantages including estimation of sub-population and total immune content using its absolute setting (333, 334). A recent study benchmarked a number of deconvolution methodologies, assessing them with both bulk and single-cell RNA transcriptomic reference data to determine the accuracy of each using pseudo-bulk datasets with known cellular proportions using RMSE (root mean square error) as a measure of accuracy (335). For the analysis of bulk RNA the top five performing bulk deconvolution methods were determined to be OLS, nnls, CIBERSORTx, RLR, and FARDEEP (in running order) due to their similarly low RMSE (335). Assessment of

deconvolution methods in scRNA-seq data identified the DWLS, MuSiC, SCDC as the top three as they achieved median RMSE values lower than 0.05, penalised regression approaches such as lasso, ridge, elastic net regression, and DCQ performed worse in scRNA-seq data achieving a median RMSE of approximately 0.1 (335). An external review of deconvolution methods in immune-oncology examined the use of CIBERSORT, EPIC, MCPcounter, quanTlseq, TIMER and Xcell; from this report xCell was suggested as it performed best with the lowest RMSE; however, results of this analysis were much in concordance with the prior benchmarking paper placing CIBERSORT and Xcell as the most accurate deconvolution methods (332, 333, 336-340). Using the available literature CIBERSORTX provides high accuracy, the second highest cell type distinction with the 22 LM immune panel (A validated signature matrix containing 547 genes that distinguish 22 human hematopoietic cell phenotypes), can be used to infer total immune content of a sample and is by far the most published (A PUBMED literature search of “CIBERSORT” and “Xcell” terms, resulted in: 761 vs 182 research articles). Thereby, by consulting the literature the deconvolution analysis for our project will focus on the use of CIBERSORTX to identify the proportions of the immune cells, whilst allowing analysis to determine total immune content in OAC samples.

4.1.2 Limitations of deconvolution methodology

Despite the usefulness of deconvolution methodology in determining the immune cell subpopulations in cancer there are several limitations associate with these techniques. Firstly, the spatial location of the immune cell types cannot be determined by deconvolution; the spatial location of immune cells is important as this can help elucidate the mechanistic effects of these cells in the TME. For example, the location of CD8+ T cells can help determine the immunophenotype of a specific tumour, with exclusion of CD8+ T cells from the intratumorally space being a marker of an immunosuppressive TME (210). Secondly, only a specified number of immune cell phenotypes can be classified by deconvolution methodology. For example, within the LM22 signature matrix the signatures for CD4+ cells are CD4 naïve, CD4 resting and CD4 activated, which does not provide granular data on the immunophenotype of CD4 cells which can significantly impact the TME. This includes Th1/Th2 balance, requiring further exploration using methodologies such as gene set enrichment analysis, as demonstrated in other immunological studies (341, 342). Finally, RMSE values produced in deconvolution analysis are significantly impacted by *a priori* knowledge, with sub-setting of known

markers of cells prior to deconvolutions significantly lowering the RMSE, suggesting sensitivity to input data. This sensitivity may mean that different datasets may not be directly comparable; furthermore, this sensitivity was explored by removing a single immune cell type from pancreatic samples finding that removing a cell type from the reference matrix made MuSiC, NNLS, and CIBERSORT results less accurate (343).

4.1.3 Validation of deconvolution output

Despite deconvolution addressing the limitation of phenotypical markers presented in immunohistochemistry (IHC), methodologies such as RNA-scope, spatial transcriptomics, and flow cytometry should be employed to validate the findings of deconvolution analysis (344, 345).

4.1.4 The Tumour Immune Microenvironment (TIME) of OAC

The TIME is the network of interacting cancer and immune cells found within a tumour. Notably, the mechanisms of immunity within these niches are significantly different from those outside of the tumour microenvironment niche, due to cancer's ability to manipulate and obstruct normal immune functions as a mechanism to avoid anti-tumoural immunity. For its significant role in the development and progression of tumours, the TIME is deemed as a hallmark of cancer (346). A previous study of the TIME of 111 primary OAC resections (mean age 65; sex Male: Female, 96:15; T category T1: 33, T2: 10, T3:65, T4:3; Lymph node metastasis: 52; distant metastasis: 4; Tumour grade G1-2: 64, G3-4: 47) depict a high range of CD3+, CD8+, FoxP3+ TILs distribution. See key results summarised in **Table 17**.

Table 17 Summary of cell densities in OAC from Stein et al, 2017 (210).

IMMUNE CELL MARKER	TUMOURAL LOCATION	CELL DENSITY (Intratumoural/Peritumoural)
CD3	INTRATUMOURAL	1-231/0.849 mm ²
CD3	PERITUMOURAL	1-220/0.849 mm ²
CD8	INTRATUMOURAL	1-130/0.849 mm ²
CD8	PERITUMOURAL	1-145/0.849 mm ²
FOXP3	INTRATUMOURAL	0-73/0.849 mm ²
FOXP3	PERITUMOURAL	0-35/0.849 mm ²

This study also identified the TILs counts for each T cell subtype (CD3+, CD8+ and FoxP3+) correlated with each other and further correlated with the total inflammatory reaction

(210). Another study of 354 OAC tissue microarrays identified that OAC seems to preferentially express PD-L2 over PD-L1, detecting PD-L2 in 51.7% of OACs compared to the 2% of cases with PD-L1 epithelial expression (73). Moreover, PD-L1 is expressed in both tumour cells and tumour infiltrating immune cells in OSCC, but there is preferential expression of PD-L1 in TIICs (Tumour infiltrating immune cells) rather than in tumour cells in OAC (73). Notably, interleukin-6 (IL-6) is highly expressed in some OAC cases, being especially expressed in cancer-associated fibroblasts (CAFs) (347). High secretion of IL-6 in OAC patients may impact the differentiation of monocytes from dendritic cells to macrophages, effecting cell fate of monocytes in the TIME identifiable by deconvolution analysis (347, 348). One study used the estimation of Stromal and Immune Cells in Malignant Tumour Tissues Using Expression Data (ESTIMATE) algorithm, which calculates a score based on the expression of immune signature genes (141 genes) including markers of T cells; analysis of the TIME of OAC found the median immune score was higher in female as compared to male patients and was correlated with tumour-node-metastasis stage (349, 350).

Important cellular players of the TIME in OAC are described across the following sections below.

4.1.5 B lymphocytes

B lymphocytes are cells that express clonally diverse cell surface immunoglobulin (Ig) receptors that can recognise specific antigenic epitopes. These cells mediate multiple functions essential for immune homeostasis; B cells are essential in the activation of T cells demonstrated in mice depleted of B cells at birth using anti-IgM antiserum (351). Other functions B cells perform in host immunity include antigen presentation to T cells, cytokine secretion, Th1/Th2 cytokine balance and the regulation of dendritic cells (352-354). In oesopho-gastric adenocarcinomas B cells were increased in tumour samples and subset-analyses of TILs showed increased proportions of differentiated and activated B cells and an enrichment for follicular T helper cells. These tumour-associated B cells (TABs) in OAC were mainly organized in tertiary lymphoid structures (TLS), which are similar structurally to secondary lymphoid organs (355). Additionally, B cells were decreased in tumours with high expression of PD-L1 or impaired MHC I HLA expression (355). From the literature, low proportions of B cells in OAC tumours from deconvolution could link to high PD-L1 expression within prognostically significant APM expression

groups, this is due to B cell infiltrate decreasing in tumours with HLA-loss and PD-L1 expression (355).

4.1.6 CD8 T cells

Naive CD8+ T cells specifically recognise antigens presented on APCs (antigen presenting cells) MHC I molecules activating and maturing into cytotoxic T lymphocytes (CTLs). CTLs provide immune defence against intracellular bacteria, viral, protozoa infections, in addition to providing anti-tumoral responses(356). CD8+ T cells can specifically target cancer cells, thus forming the majority of immune anti-tumoral responses, directing cytotoxicity towards cancer cells by secreting cytotoxic granules. Cytotoxic granules contain perforin, plus granzymes A and B, these respectively act to form pores in cell membranes and induce apoptosis through activation of a caspase cascade establishing an apoptosome (357). As CD8+ T cells can target cancer cells, their mechanisms of engagement can be exploited by immunotherapies to boost anti-tumoral responses. Immunotherapies such as anti-CTLA-4 has shown promise in keeping T-cells activated by preventing tumour cell B7 ligand from binding to CTLA-4 receptors. This blockade of the immune checkpoint allows T-cells to direct cell cytotoxicity towards cancer cells (358). These therapies are currently being explored in OAC and gastric cancers, with the presence of CD8+ TILs being a major factor in response to checkpoint blockade therapies; thereby, APM expression which impacts the infiltration of CD8+ TILs in OAC measured by deconvolution may be a prognostic tool in future to assess the likelihood of patient response to checkpoint blockade in OAC (359). Interestingly, within CD8+ T cells there can be found a CD8+ Treg population, this subset notably is capable of secreting inhibitory cytokines and chemokines; including IL-10, transforming growth factor (TGF)- β , IL16, IFN- γ (360). Remarkably, CD8 $^+$ CD28 $^-$ Tregs can render APCs tolerogenic, through upregulation of immunoglobulin-like transcript (ILT)3 and ILT4 expression (361). CD8+ TILs correlate with improved survival in patients with OAC, with CD8+ TILs forming an independent prognostic factor and were associated to significant pathological response to neoadjuvant chemotherapy. Within the same study multivariate analysis increased levels of CD4 $^+$ and CD8 $^+$ TILs were associated with significant local tumour regression, lymph node downstaging and improved cancer specific survival (362).

4.1.7 CD4 T cells

CD4+ T cells often coined, 'T helper' cells due to their support functions in assisting most aspects of the adaptive immune system from helping activate B cells to secrete antibodies, macrophages to destroy ingested microbes and even assist in activating CD8+ cytotoxic T lymphocytes to kill cancer cells. CD4+ T cells can be divided into multiple cell types; firstly, Th₁ cells participate in responding to intracellular pathogen and cancerous cells. In the presence of IFN- γ and IL-12, and absence of IL-4 and IL-10, naive CD4⁺ T cells are driven to differentiate into Th₁ cells (363). This maturation towards the Th₁ immunophenotype is further driven by a positive feedback loop, resulting from Th₁ cells expressing IFN- γ in the microenvironment, also suppressing generation of Th₂, Th₁₇, and Tregs (364-366). Moreover, secretion of IL-2 from Th₁ cells promotes the proliferation of lymphocytes, enhancing the activity of NK cells by upregulating IL-12 receptor and STAT4 expression; this interaction leads to increased secretion of IFN- γ and TNF from NK cells (367). Furthermore, IFN- γ and IL-12 secreted from Th₁ cells may increase CD40 expression, thereby promoting macrophage functions, such as phagocytosis and presentation of exogenous antigens (368). Th₂ cells participate in type-II immune responses towards parasites, but also play significant roles in allergies and atopic illnesses (369). TCR activation and cytokine-mediated signalling are important during Th₂ cell differentiation. TCR activation leading to NFAT, NF- κ B and AP-1 activation results in upregulation of *IRF4* expression, this in combination with upregulation of *GATA3* leads to Th₂ differentiation. IL-4 and IL-13 can also promote Th₂ differentiation through binding to IL-4 receptors, resulting in the upregulation of *GATA3* via STAT6 (370). Interestingly, squamous esophagitis to Barrett's oesophagus progression is accompanied by a transition from a Th₁ to Th₂ immune response which is hypothesised to induce PD-L2 through the Th₂ cytokines IL-4 and IL-13 in the progression from squamous esophagitis to Barrett's oesophagus, contributing to immune evasion of OAC (73). Thereby, detection of a Th₂ dominant Th₁/Th₂ balance in deconvolution analysis could relate to immune suppression of TIME via PD-L2 induction. Th₁₇ cells are proinflammatory cells, these cells mediate host defence against bacteria and fungi by secreting IL-17. Th₁₇ cells characteristically express; ROR γ T, ROR α , IL-17A, and IL-17F (371, 372). Interestingly, Th₁₇ cells are polarised in the presence of TGF- β and IL-6 (371). However, unlike their Th₁ and Th₂ counterparts, Th₁₇ do not secrete IL-6, thereby do not possess a positive feedback loop for the generation of Th₁₇ cells. Instead, IL-21 is key to Th₁₇ expansion, promoting

Th₁₇ expansion via an IL-6 and TCR independent mechanism (373). In OAC IL-17 secreted by Th₁₇ cells promotes cell invasiveness and migration through ROS-dependent, NF-κB-mediated MMP-2/9 activation (374). Furthermore, in reflux oesophagitis (RO) and BE an increased ratio of Th₁₇/Treg was observed when compared to normal controls, and the proportion of Th₁₇/Treg in BE was also increased in comparison RO patients (375). Expression levels of ROR γ T, IL-17, IL-6, and TGF- β were elevated, while the levels of Foxp3 and IL-10 were reduced in patients when compared to the controls, validating the increase in Th₁₇/Treg ratio (375).

4.1.8 T cells regulatory (Tregs)

CD4+ Tregs act to suppress immune functions, prevent autoimmune disease and to promote self-tolerance by secreting specific cytokines such as IL-10 and TGF- β . Tregs were first described as immunosuppressive CD4⁺CD25⁺ T cells in 1995, prior to identifying the roles of IL-2 and TGF- β in maintenance and development of T cells, and FoxP3 as a master regulator of Treg functions (376). Notably, Tregs secrete immunosuppressive cytokines including IL-10, IL-35, and TGF- β . These cytokines play additional roles promoting tTreg (thymus Treg) and pTreg (peripheral site Treg) proliferation and function, as well as, exerting suppression of other effector cells. TGF- β secreted from Tregs produces several biological responses, these include inhibiting transcription of IL-2 and cyclin D expression, preventing effector cell proliferation. Furthermore, TGF- β inhibits Tbet and GATA3, suppressing Th₁, Th₂, and CD8+ T cell responses. TGF- β also induces FoxP3 expression on other Tregs in the microenvironment, whilst promoting T cell survival by downregulating C-MYC and FasL expression, needed in activation-induced cell death of lymphocytes (ACID) (377). In OAC and BE, an increase in the frequency of Tregs compared with normal controls with patients possessing higher Treg counts in the centre of the tumour associated with lower stage of disease and were positively correlated with the density of CD8⁺ CTLs (378). However, the presence of Tregs was not an independent prognostic factor to CD8+ T cells in OAC suggesting the increase in Treg numbers is related to increase the host defence mechanism overall. The literature on Tregs suggest that Tregs could present a marker for overall immune host defence mechanisms in our OAC cohort.

4.1.9 Natural killer cells (NK)

NK cells are innate lymphocyte cells that mediate anti-tumour and anti-viral responses; by secreting cytokines and chemokines NK cells can regulate immune responses, for example, CC-chemokine ligand 5 (CCL5), XC-chemokine ligand 1 (XCL1) and XCL2 secreted from NK cells promotes the recruitment of dendritic cells into solid tumours, this event correlates with improved survival (379). Activated NK cells mediate cell cytotoxicity through the release of granzyme B, inducing a cascade of caspases cleavages, resulting in cell apoptosis (380). NK cells work synergistically with professional antigen presenting cells, in secreting cytokines required to induce inflammation and recruit lymphocytes to sites of inflammation, allowing the foundation of cellular adaptive response.

Furthermore, activated NK cells express HLA-DR that can initiate MHC II-dependent CD4+ T-cell proliferation (381). Importantly, NK cells are capable of targeting cancer cells which undergo loss of MHC I cell surface molecules directing cytotoxic activity towards the cancer cells (382). In OAC, a high concentration of NK cells was associated to prolonged overall survival independently from clinical and other immune cell proportions; specifically, the best prognosis from this study was identified in patients high density of either CD20⁺ B cells or IgKC⁺ plasma cells combined with high density of either high CD³⁺, CD8+, FoxP3⁺ or NKp46⁺ lymphocytes, with high density of NKp46⁺ cells being significantly associated with a lower tumour stage (383). From the literature, a high fraction of NK cell in the TIME is a positive prognostic marker to investigate in the deconvolution of our cohorts, furthermore the impact of APM gene expression on NK cell subpopulation density can be explored; the evidence for this includes analysis of CD56^{dim}CD16⁺ NK and CD56^{bright}CD16⁻ NK cells in OAC which relates to disease progression (384).

4.1.10 Monocytes

Monocytes perform a range of functions in cancer at distinct stages of tumour growth and progression; protumour functions include differentiation in tumour associated macrophages, metastatic cell seeding, suppression of T cell function and remodelling of the extra cellular matrix (385-387). Antitumoral function of monocytes include tumour cytotoxicity, prevention of metastasis, engulfment of tumour material, recruitment of/correlation with NK cells and inhibition of Tregs (388-394). In deconvolution analysis of our cohorts, the presence of monocytes may indicate either immune suppression or

cytotoxic activity, observing the cytotoxic activity in high monocyte populations may indicate their role in the OAC microenvironment.

4.1.11 Macrophages

Macrophages are crucial mediators of tissue homeostasis in the TIME, the function of these cells are often distorted by tumours to stimulate proliferation, angiogenesis and metastasis, the macrophages exploited by tumours are known as tumour associated macrophages (TAMs). TAMs are often described in tumours as one of two phenotypes M1 or M2 macrophages; M1 TAMs are pro-inflammatory and are activated by IFN- γ and lipopolysaccharide and promote tumour immunity by expressing elevated levels of tumour necrosis factor (TNF) and inducible nitric oxide synthase. Conversely, M2 TAMs are anti-inflammatory and pro-tissular, expressing MHC class II molecules and pro-tumorigenic when they express high levels of arginase 1, IL-10, CD163, CD204 or CD206 (395). In cancer, M1 macrophages are often polarised to the M2 phenotype by IL-4, IL-10 and TGF- β cytokines secreted into the TIME (396). In OAC prevalence of the M2 subtype was found to predict lymph node metastasis, cellular invasion, and poor prognosis (397-401).

4.1.12 Dendritic cells (DCs)

DCs are a specialised diverse group of professional APCs which play crucial roles in the initiation and regulation of both innate and adaptive responses. Importantly, in the TIME, DCs acquire and process tumour antigens presenting them on MHC II molecules with two other stimulatory signals, providing co-stimulation and stimulatory soluble factors, shaping T cell responses to cancer (402, 403). Tumours often modulate the function of DCs to achieve immune evasion including the inhibition of differentiation, exclusion from the TIME, disrupting activation, direct inactivation, impaired handling of the antigen presentation, metabolic stress and reducing viability (402). The presence of DCs has been noted to increase from BE to OAC, this could suggest in OAC may impair the function of these DCs shifting the DCs to an anti-inflammatory phenotype by exploiting the IL-6/JAK/STAT3 pathway. This could produce functionally incompetent dendritic cells as previously described in head and neck squamous cell carcinoma, non-small cell carcinoma and breast adenocarcinoma (402, 404, 405). Overall, the population of DCs should remain rare in comparison to other immune cell types, with resting DC

being a potential marker of reduced anti-tumoral immunity via the aforementioned IL-6/STAT3 pathway, my deconvolution analysis will aim to confirm this in OAC (406).

4.1.13 Mast Cells

Mast cells are long-lived secretory cells, belonging to the granulocyte family which possess a multitude of functions in angiogenesis, homeostasis, innate and adaptive immunity, and cytokine/chemokine release. In cancers mast secrete pro-angiogenic factors, such as VEGF, bFGF, TGF-beta, TNF-alpha, and IL-8, supporting tumour growth (407). Mast cells can act as either pro-tumoral or antitumoral; for example, mast cells can secrete TNF- α and increasing antigen presentation by dendritic cells, promoting pro-inflammatory T cell responses and monocyte/macrophage activation, or conversely, mast cells can secrete IL-10 and thus block T cell proliferation (408, 409). The abundance of Mast cells in OAC is known to correlate to improved OS, especially in cases with lymph node metastasis (410). Additionally, pancreatic ductal adenocarcinoma (PDAC) Mast cell concentration within the intertumoral space, but not the peritumoral or the intratumorally space, was associated to poor prognosis (411). Due to the genomic similarities between PDAC and OAC, there may be similarities between the concentration of mast cells and prognosis in OAC as seen in PDAC. Thereby, in our analysis assessing the impact of expression of our prognostically significant APM genes on the population of mast cells by deconvolution analysis may elucidate the role of APM genes in mast cell prevalence in OAC, suggesting potential crosstalk between specific APM genes and mast cells or providing a biomarker for their presence.

4.1.14 Eosinophils

Eosinophils are innate immune cells, belonging to the granulocyte family which secrete a diverse range of cytokines (412). In cancer, eosinophils can act to support anti-tumoral responses by normalizing tumour vessels and enhancing infiltration of CD8+ T cells by chemoattractant recruitment (413). The impact of eosinophils on survival has been investigated in OAC, demonstrating eosinophil positive tumour possess longer overall survival compare to eosinophil negative tumours in cases only treated with surgery (414). Thereby, investigating the impact of APM gene expression on eosinophil subpopulations in OAC by deconvolution may yield insights by which APM gene expression impacts the innate immune response and overall survival.

4.1.15 Neutrophils

Neutrophils are granulocytes which form most abundant leukocytes in the circulation and function to capture and destroy invading microorganisms using phagocytosis and intracellular degradation, release of granules, and by forming neutrophil extracellular traps (415). In tumours, neutrophils are known as tumour-associated neutrophils and can have antitumour and pro-tumour functions; pro-tumoral function include the released of reactive oxygen species (ROS), secretion of pro-tumour cytokines and chemokines (TGF- β , HGF, CCL4, CXCL8, IL17) (416-418). In OAC patients with elevated levels of TANs (Tumour associated neutrophils) were associated to poorer prognosis. This suggests the deconvolution analysis should aim to identify any significant difference in neutrophil populations for each candidate gene. Differences in neutrophil populations associated to APM gene expression may indicate APM gene expression could be affecting survival through impacting the number, function, and phenotype of neutrophils. Interestingly, Mast cell positive tumour possess longer overall survival compare to Mast cell negative tumours (414).

4.2 Hypothesis and research objectives

The prior datamining investigation of the landscape of genomic defects in APM genes identified the expression of several APM genes was prognostically significant in our cohorts, however, the impact of APM gene expression on the TIME immune cell subpopulations is unknown. To explore how APM gene expression impacts the immune subpopulations of OAC, deconvolution analysis will be conducted to elucidate the proportions of TIME immune cell subpopulations and thereby investigate the immune response associated with altered APM gene expression.

Hypothesis 2: The expression of prognostically significant antigen presentation machinery genes defines the immunophenotype of OAC by editing the distribution of immune cell subpopulations in the TIME.

Aim 2 (Results Chapter 4): Investigate immune infiltrate and activity in OAC and/ or recurrent molecular defects in APM/ MHC I & II pathways by digital cytometry utilising deconvolution analysis of bulk transcriptomic data.

Objective 2a: Characterise the immune cell subpopulations in OAC with our cohorts using CIBERSORT deconvolution and compare to known published immune populations in OAC.

Objective 2b: Assess the impact of antigen processing machinery defects identified by the chapter 3 datamining analysis on the immune cell subpopulations our OAC cohorts (TCGA and OCCAMS).

Objective 2c: Determine the prognostic value of antigen processing machinery defects in combination with immune cell distributions.

4.3 Methodology

4.3.1 Samples and data

Samples for deconvolution analysis were extracted from the prior data mining analysis for 176 samples from the combined TCGA-OCCAMS cohort. The data used for deconvolution included RNA-seq counts normalised using TPM, batch corrected between the two datasets (Combat-seq) and clinical data parsed to include survival times, survival events and optimal cut point classifiers for each APM gene (207).

4.3.2 CIBERSORTx

Deconvolution functions on a linear model which focus on a gene expression profile as a linear equation of pre-defined signatures of immune genes at different ratios. These linear regression models are applied to estimate gene coefficients used to infer the immune cell abundances (323). CIBERSORTx specifically, is a deconvolution method using v-support vector regression method to estimate the immune cell proportions (relative) from a gene expression profile (323).

Firstly, the prerequisite mixture files for TCGA and OCCAMs were prepared for analysis. The mixture file contains the gene expression profiles of the samples, this forms a table where the first row consists of column headers containing the sample labels and the first column consists of row headers containing the gene name or symbols, with the data points occupying the remainder of the table. For our analysis, the mixture files were produced from the TPM normalised RNA-seq counts batch corrected using ComBat-seq produced in a tab-delimited (.txt) format. Next the mixture files were uploaded into to the CIBERSORT file storage (<https://cibersort.stanford.edu/upload.php>), then input into the CIBERSORT basic configuration parameters on the CIBERSORT webpage (<https://cibersortx.stanford.edu/>); parameters used for each run are as follows:

- Job type: Impute Cell Fractions
- Signature matrix file: LM22.update-gene-symbols.txt
- Mixture file: MIXTURE.txt
- Batch correction: enabled
- Batch correction mode: B-mode
- Disable quantile normalization: false
- Run mode (relative or absolute): absolute
- Permutations: 100

These parameters were selected based on advice provided in the CIBERSORTx publication, permutations set to 100 to produce “meaningful p values” (i.e. significant level down to p-value 0.01), Batch correction disabled as data was prior batch corrected, quantile normalisation disabled as data in RNA-seq format, finally, absolute mode used as these results provide comparable results between samples and cell types (323).

4.3.3 Statistical analysis of deconvolution data

For the combined dataset, CIBERSORT analysis for all immune cell phenotypes were simplified into the broad categories by adding together absolute values of immune cells (B cells, Plasma cells, T cells CD8, T cells CD4, T cells regulatory (Tregs), NK cells, Macrophages, Dendritic cells, Granulocytes, Monocytes) then z-score scaled prior to passing data to pheatmap function (pheatmap package) in R version 4.0.2, clustered using Euclidean distance with ward.d2 linkage (419). Heatmaps were annotated with the TPM expression of *GZMA*, *GZMB*, *PRF1*, the absolute TILs score and CYTscore (A score representing immune cytotoxicity via degranulation) calculated from the geometric mean of *GZMA* and *PRF1* expression ($\sqrt{GZMA} \times PRF1$) (420).

To confirm the accuracy and optimise the CIBERSORTx analysis, variations on the CIBERSORTx analysis were conducted and compared to Methyl-CIBERSORT data (orthogonal standard/ calibrator method using paired DNA methylation data with high accuracy to the flow cytometry gold standard) for matching TCGA samples (gifted by Dr Tim Fenton with permission to re-use). Variations in data sources and batch corrected data types tested included TIMER 2.0 TPM, CIBERSORTx B-mode batch corrected TPM (only TCGA samples), CIBERSORTx B-mode batch corrected TPM (combined OCCAMS and TCGA samples) and finally CIBERSORTx ComBat-seq batch corrected TPM (combined OCCAMS and TCGA samples) (327, 421).

Using the RNA-seq data from **Chapter 3** results I produced a correlation analysis, correlating the expression of prognostically significant APM genes identified in **Chapter 3** with the immune cell subpopulation CIBERSORTx absolute values produced using methods explained above, these results were placed in a correlation heatmap using the R `ggcorplot()` function edited to identify significant correlations (Alpha set at the 5% level) (422).

To assess differences in immune cell distributions due to APM gene expression, upper/lower mRNA expression quantile cut points (<0.25 and >0.75 quantiles) for APM genes were introduced to dichotomise the CIBERSORT analysis, then passed to GraphPad prism 9 for statistical analysis and visualisation using boxplots (423). The normality of the CIBERSORT results in each expression group was assessed using Shapiro-Wilk tests, then central tendencies (means or medians) for each cell phenotype between the high/low APM gene expression groups were directly compared using unpaired T-tests (Parametric data) or Mann-Whitney U tests (non-parametric) based on data normality. The prior dichotomisation was repeated though using the upper and lower quantiles of APM gene expression to divide the CIBERSORT results, then passed to GraphPad prism 10 repeating the analysis described above. Results were corrected for false discovery using “Two-stage set-up method of Benjamini, Krieger and Yekutieli” in GraphPad prism.

4.4 Results

4.4.1 Validation of CIBERSORTx analysis

Firstly, I correlated the sample-level variation of CIBERSORTx analysis to MethylCIBERSORT data to select the most reliable batch correction method for CIBERSORTx using the CD8+ T cell estimates as a benchmark. CD8+ estimates were selected as they had the highest performance of CIBERSORTx estimated cell fractions in line with flow cytometry ($R^2 = 0.86$) (330).

CD8+ T cell scores derived from the publicly available TIMER 2.0 CIBERSORT analysis (using the original CIBERSORT algorithm) yielded a $R^2 = 0.54$ (see **Figure 26A**) with MethylCIBERSORT, with CD8+ T cell scores derived from CIBERSORTx using B-mode (TCGA samples only) displayed an improvement in correlation with a $R^2 = 0.58$. However, once combining the TCGA and OCCAMS datasets with B-mode CIBERSORTx analysis demonstrated a significant drop in correlation for TCGA cases, producing a $R^2 = 0.39$. This limitation was addressed by using ComBat-seq to batch correct RNA-seq counts between TCGA and OCCAMS before producing TPM data for input into CIBERSORTx analysis with B-mode enabled (**Figure 26D**).

These CD8+ Rho values are in line with flow cytometry derived cell type fraction versus CIBERSORT fraction sample correlation (Rho) values (0.5-0.86) from Newman et al. (**Figure 3** in Newman, et al. 2019) (330). Using this analysis, I was able to determine that using TPM normalised RNA-seq counts which was batch corrected using ComBat-seq would be sufficient to make comparisons between TCGA and OCCAMS samples and was used for all downstream analyses in this chapter.

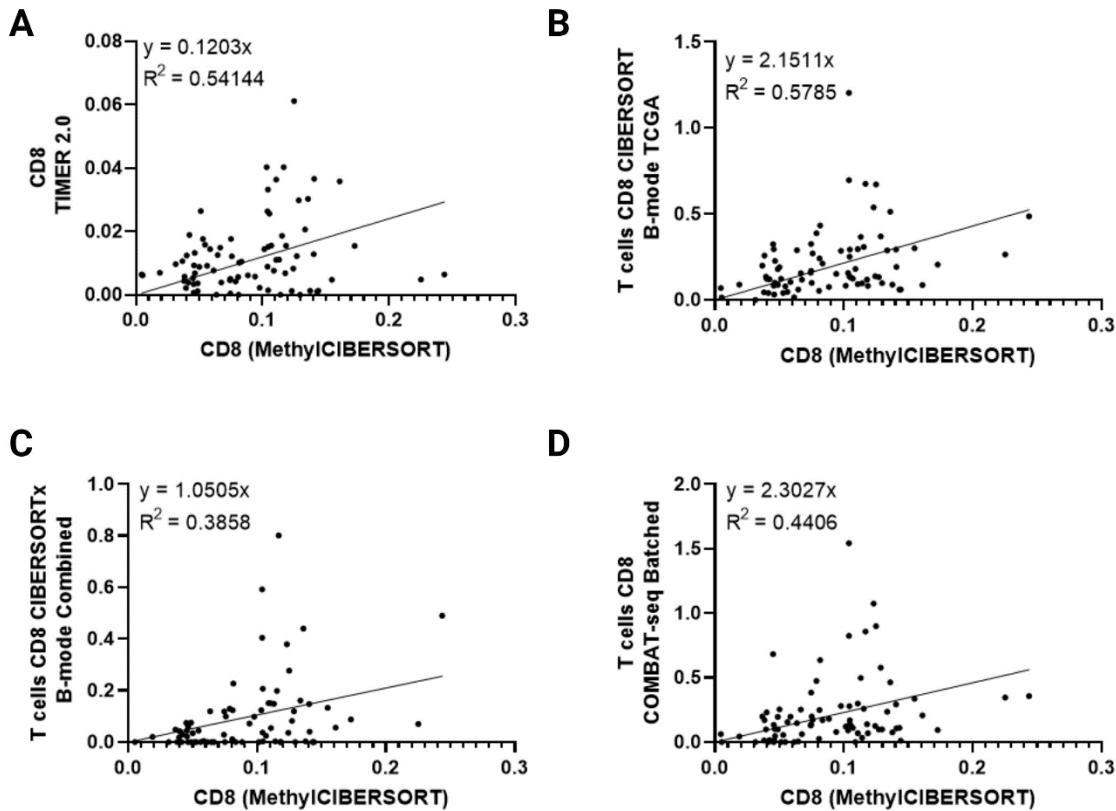


Figure 26: Scatter plots with linear regression comparing MethylCIBERSORT with variations of CIBERSORTx analysis. A: CD8 MethylCIBERSORT Absolute score vs CD8 Absolute score TIMER 2.0. B: CD8 MethylCIBERSORT Absolute score vs CD8 Absolute score CIBERSORTx B-mode (TCGA only). C: CD8 MethylCIBERSORT Absolute score vs CD8 Absolute score CIBERSORTx B-mode (combined OCCAMS and TCGA samples) D: CD8 MethylCIBERSORT Absolute score vs CD8 Absolute score CIBERSORTx ComBat-seq batch corrected (combined OCCAMS and TCGA samples).

4.4.2 A description of the TIME in OAC

I moved to elucidate the TIME of OAC by clustering the CIBERSORTx estimated immune cell subtype distributions. This analysis identified four distinct clusters, The first cluster was driven by the presence of monocytes and granulocytes, the second represents an immune hot group with excessive amounts of effector cells including CD8+ T, CD4+ T and NK cells, with high macrophages (See Figure 27 & 28). A third cluster presented as immune cold relatively lacking a high subset of immune cell types (See Figure 27 & 28). Finally, the fourth cluster was characterised by B cells, however a subset of these samples does also have high plasma cells estimates (See Figure 27 & 28).



Figure 27 Z-scored heatmap of CIBERSORT simplified data in the entire OAC cohort (n = 176), clustered using Euclidean distance with ward.d2 linkage, four distinct clusters of immune phenotypes, the expression of GZMB and PRF1 (TPM, Z-scored) and absolute TILs labelled in the top annotation.

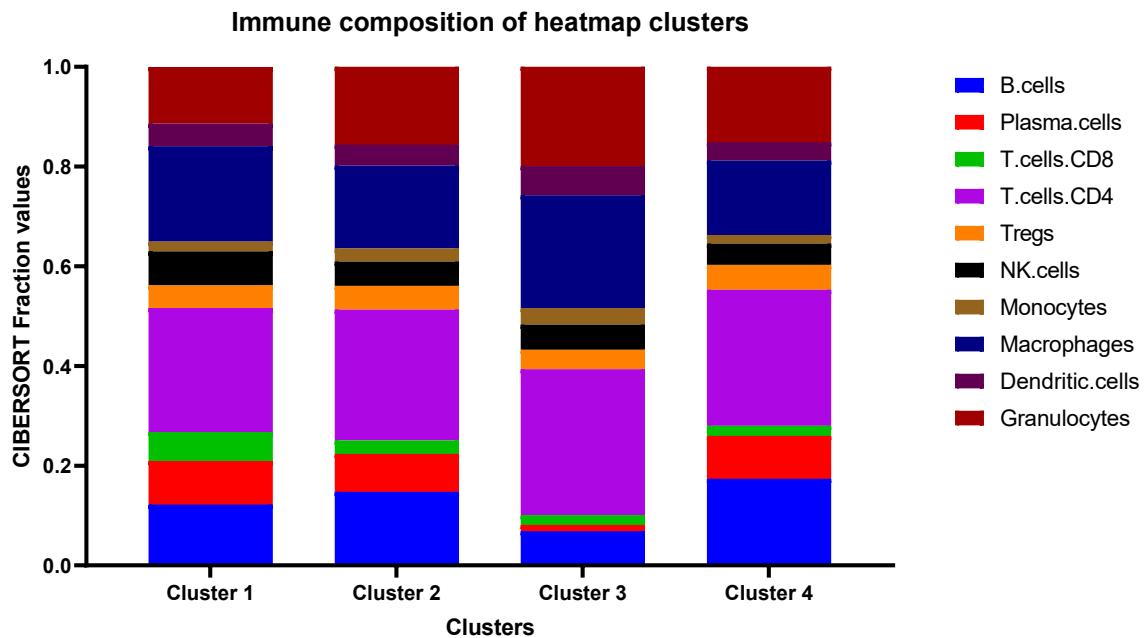


Figure 28 Stacked bar chart of CIBERSORT fractional values by heatmap cluster displayed in Figure 27.

Interestingly, comparing the absolute TIL between the heatmap clusters found a significant difference between all clusters except cluster 1 & 2 (See Figure 29A). Cluster 3 possessed a significantly lower TILs compared to all other clusters, while cluster 2 possessed a significantly greater TILs absolute score to all other clusters excepting cluster 1 (See Figure 29A). Expanding the comparisons between the clusters I observed cluster 3 also possessed a significantly lower expression of granzyme-B (GZMB) compared to cluster 2, of which cluster 2 also possessed a significantly greater GZMB expression compared to cluster 4 (See Figure 29B). Similarly, the expression of perforin-1 (PRF1) was significantly reduced in cluster 3 compared to cluster 2 and 4 (See Figure 29C). These results place cluster 1 and 2 with the most potential cytotoxicity, followed by cluster 4 with the remaining cluster 3 with the least potential cytotoxicity (See Figure 29C). Calculating the CYT score demonstrated a significantly lower cytotoxicity potential of cluster 3 compared to cluster 2 (See Figure 29D).

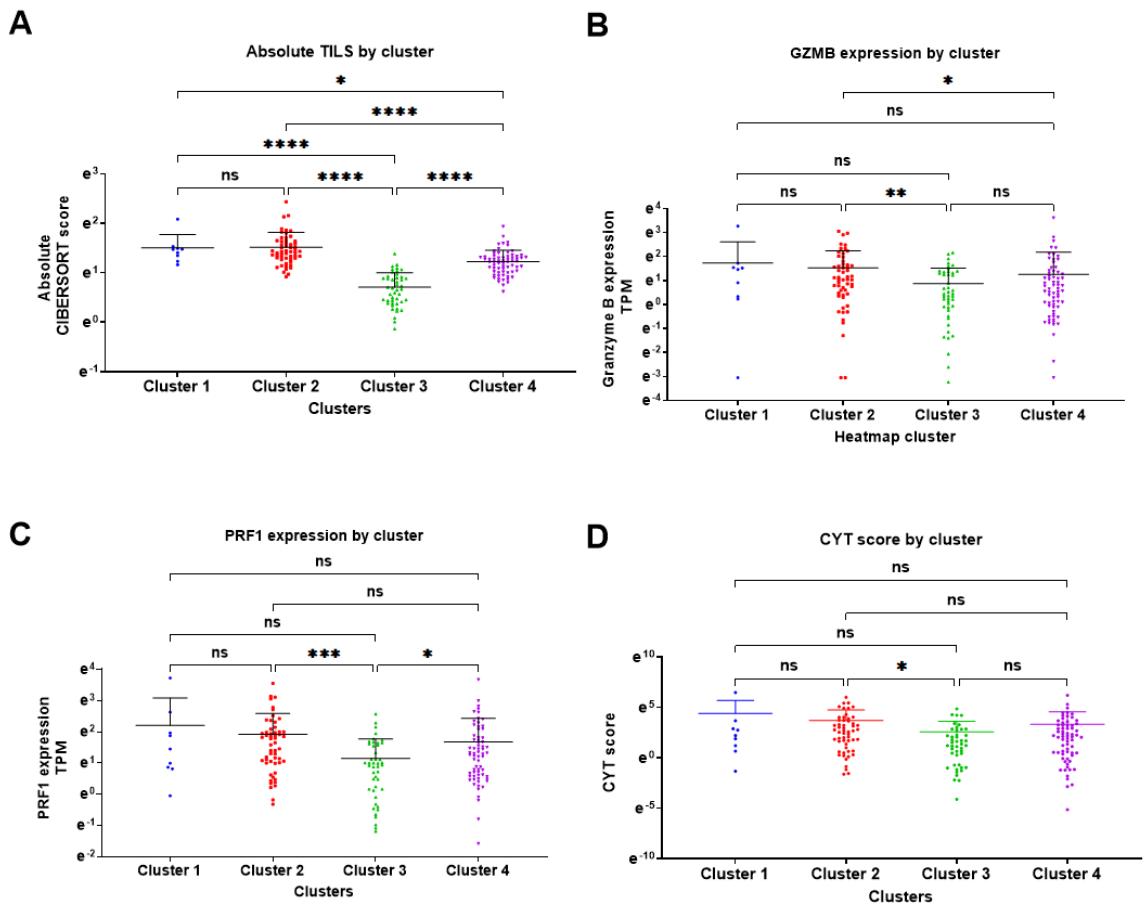


Figure 29 Box and Whisker plots of immune characteristics compared between heatmap clusters of immune phenotypes displayed in Figure 27. A: The absolute TILs score derived from CIBERSORTx by heatmap cluster. B: The expression of GZMB by heatmap cluster. C: The expression of PRF1 by heatmap cluster. D: CYT score by heatmap cluster. Statistical test Mann-Whitney U test with FDR correction, p values * <0.05 , ** <0.01 , *** <0.001 , NS = non-significant.

Lastly, in my cluster analysis I assessed the impact of my four immunophenotype clusters (IP-cluster) on survival outcomes. In CoxPH overall survival analysis did not identify any significance differences between the immunophenotype clusters in both univariate and multivariate analysis, though cluster 3 did possess the highest, but non-significant hazards ratio among the clusters (See Figure 30 & Table 18).

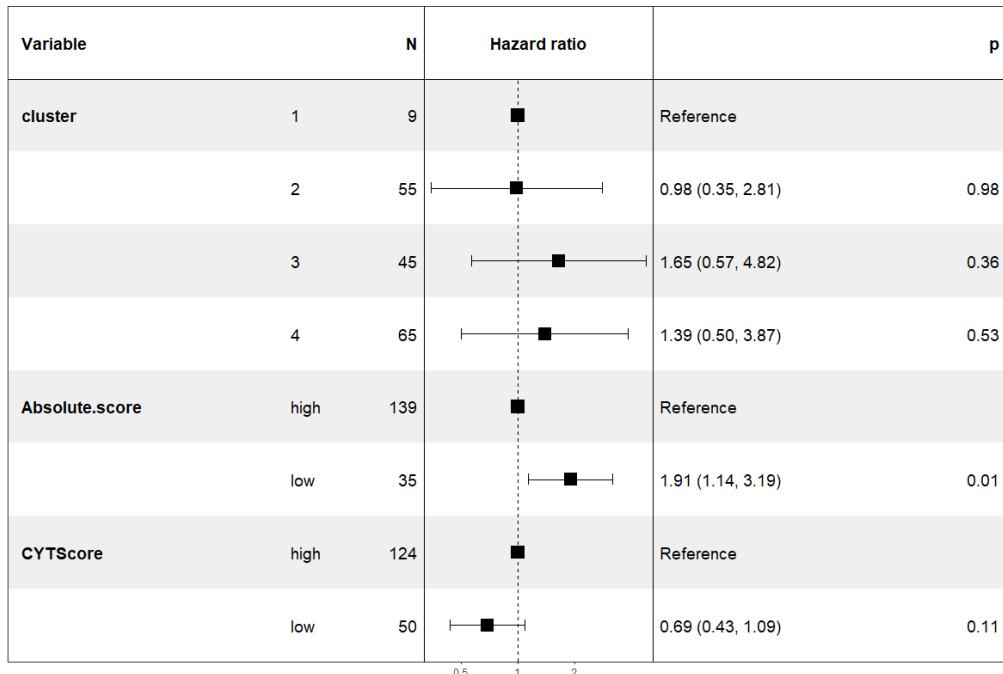


Figure 30 Forest plot of CoxPH survival analysis comparing the heatmap clusters of immune phenotypes derived from Figure 27.

Table 18 CoxPH survival analysis comparing the heatmap clusters of immune phenotypes derived from Figure 27.

LABEL	LEVELS	ALL	HR (UNIVARIABLE)	HR (MULTIVARIABLE)
CLUSTER	1	9 (5.2)	-	-
	2	55 (31.6)	0.98 (0.35-2.81, P=0.977)	0.47 (0.06-3.90, P=0.487)
	3	45 (25.9)	1.65 (0.57-4.82, P=0.358)	0.34 (0.04-3.28, P=0.352)
	4	65 (37.4)	1.39 (0.50-3.87, P=0.529)	0.60 (0.08-4.72, P=0.627)
ABSOLUTE SCORE	HIGH	139 (79.9)	-	-
	LOW	35 (20.1)	1.91 (1.14-3.19, P=0.014)	3.16 (1.00-10.03, P=0.051)
CYTSCORE	HIGH	124 (71.3)	-	-
	LOW	50 (28.7)	0.69 (0.43-1.09, P=0.110)	0.75 (0.36-1.56, P=0.436)
AGE	MEAN (SD)	66.7 (11.0)	0.99 (0.98-1.01, P=0.515)	1.01 (0.98-1.04, P=0.691)
SEX	FEMALE	31 (17.8)	-	-
	MALE	143 (82.2)	1.65 (0.90-3.01, P=0.105)	0.84 (0.34-2.08, P=0.708)
PT	0-1	38 (29.9)	-	-
	2-4	89 (70.1)	1.39 (0.78-2.48, P=0.266)	0.46 (0.19-1.12, P=0.087)
PN	0	48 (37.5)	-	-
	1-3	80 (62.5)	2.56 (1.45-4.52, P=0.001)	4.20 (1.63-10.77, P=0.003)
PM	0	97 (89.0)	-	-
	1	12 (11.0)	2.91 (1.32-6.43, P=0.008)	6.26 (2.39-16.40, P<0.001)

Number in data frame = 174, Number in model = 91, Missing = 83, Number of events = 50, Concordance = 0.729 (SE = 0.041), R-squared = 0.250 (Max possible = 0.982), Likelihood ratio test = 26.175 (df = 8, p = 0.001)

4.4.3 Correlation of prognostically significant APM genes and immune subpopulations in OAC

Moving forward, I sought to determine if APM gene expression (mRNA) associated with altered OS correlated to individual immune cell subpopulation CIBERSORT absolute scores in OAC. This analysis would allow for further filtering of APM genes for downstream analysis by refining APM gene candidates which correlate with altered immune cell compositions.

This process resulted in significant correlations between eleven MHC class I genes and immune populations, twenty-three for MHC class II, nine of alternative APM (A non-conventional antigen presentation gene outside the typical MHC class I/II pathways) and two finally for APM gene expression regulators for all samples (**See Figure 31**).

However, assessing individual immune phenotype clusters reveals further significant correlations; starting in cluster one, correlations are observed include five between MHC class I genes and immune subpopulations, five correlations between MHC class II genes and immune subpopulations, two between alternative APM genes and immune subpopulations, and five between APM gene expression regulators and immune subpopulations (**See Figure 32**). Cluster 2 reveals few significant correlations between APM genes and immune subpopulations consisting of five MHC I class correlations, seven MHC class II correlations, eight alternative APM correlations and a lack of APM gene expression regulators with a significant correlation (**See Figure 32**). Cluster 3 demonstrates significant correlations between APM genes and immune subpopulations including seven MHC I class correlations, thirteen MHC class II correlations, three alternative APM correlations and a single APM gene expression regulator with a significant correlation (**See Figure 32**). Finally, cluster 4 possessed significant APM gene correlations with immune subpopulations for five MHC I class correlations, nine MHC class II correlations, and two APM gene expression regulators with a significant correlation (**See Figure 32**). The following results subsections will break down these correlations.

Correlation of MHC genes and APM regulators n = 176

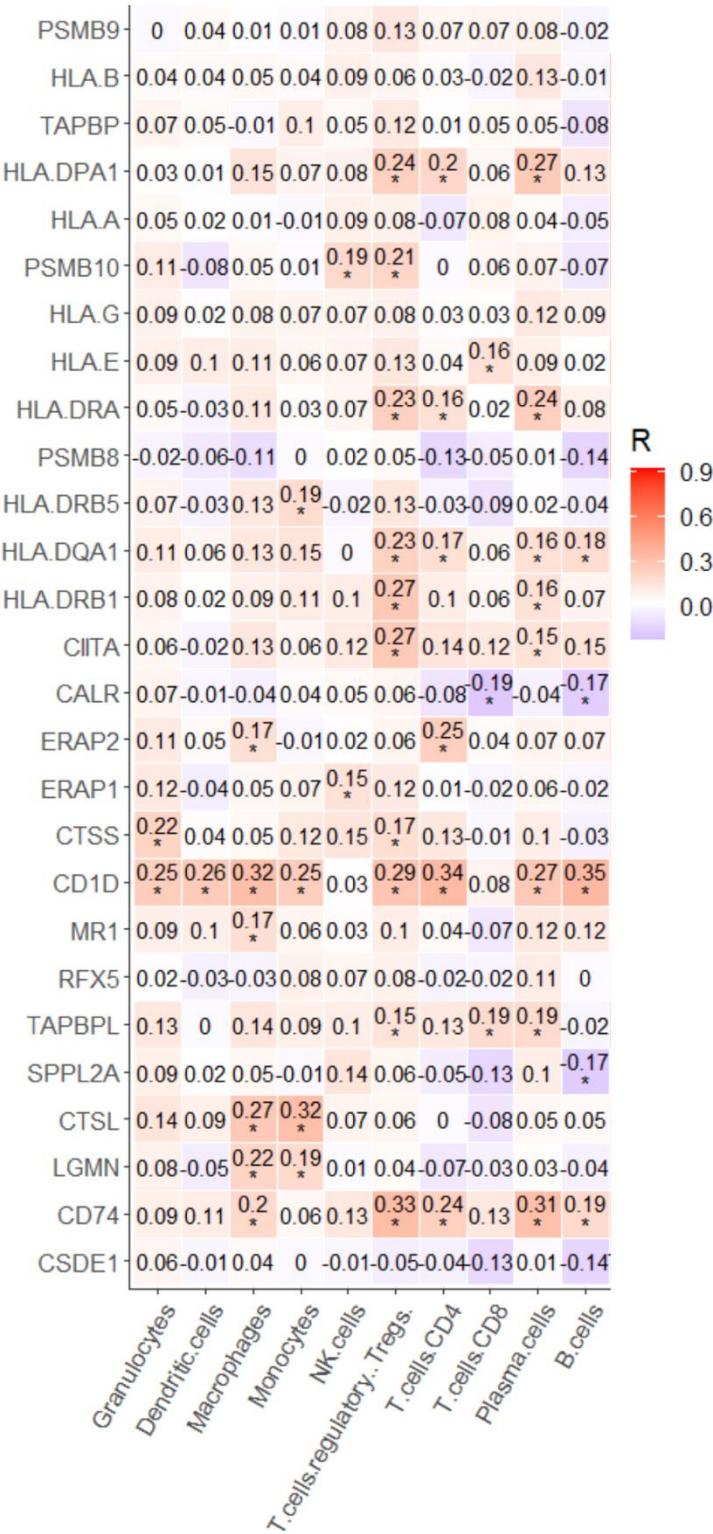
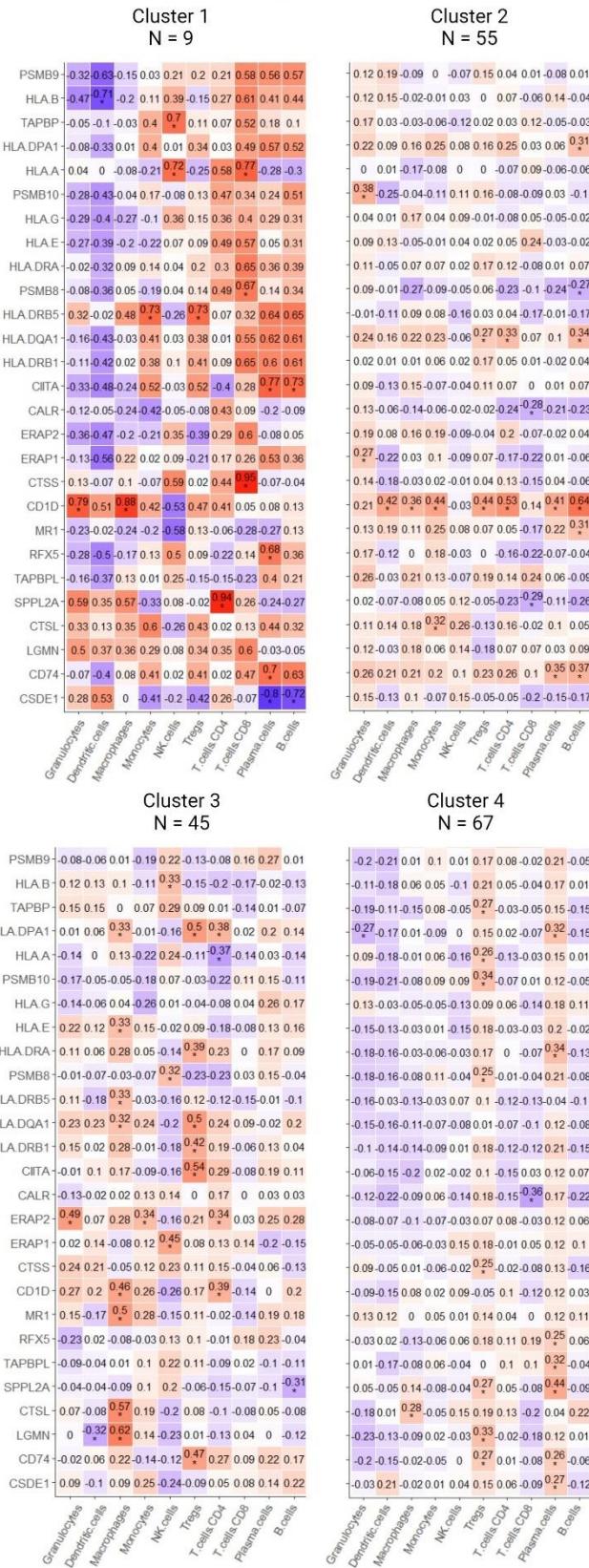


Figure 31 Correlation heatmap of APM mRNA expression to CIBERSORT ABSOLUTE scores for immune cell subpopulations, with negative values representing a negative correlation and positive values a positive correlation. * = p < 0.05. Drawn using ggcorrplot in R 4.0.2.



4.4.3.1 Effector and Treg populations

Among effector cells, several APM genes demonstrated a significant correlation to immune cell subpopulations including CD8+, CD4+, NK and Treg cells.

Firstly, MHC class I (*HLA-A/ B/ C*) and APM expression of *CALR*, *TAPBPL* and *HLA-E* correlated significantly to increased CD8+ T cell scores (Respectively: $R = -0.19, 0.19, 0.16$; $p < 0.05$; **See Figure 31**). Breaking down the clusters, further correlations between MHC class I genes and CD8 T cells were observed including *HLA-A* and *PSMB10* in cluster 1 (Respectively: $R = 0.77, 0.67$; $p < 0.05$; **See Figure 32**) and *CALR* in both cluster 2 and 4 (Respectively: $R = -0.28, -0.36$; $p < 0.05$; **See Figure 32**). The expression of MHC class II genes did not demonstrate any correlations with CD8+ T cell scores, though *CD74* did trend towards a correlation ($R = 0.13$; $p = 0.09$). Lastly, for CD8+ T cells, the APM regulator, *CSDE1* correlated with CD8+ T cell ABSOLUTE scores approaching significance ($R = -0.13$; $p = 0.09$; **See Figure 31**). Observing individual clusters yielded two further correlations with *CTSS* in cluster 1 ($R = 0.95$; $p < 0.0001$; **See Figure 32**) and *SPPL2A* in cluster 2 ($R = -0.29$; $p < 0.05$; **See Figure 32**).

For CD4+ T cell – APM gene expression correlations, a single MHC class I (*ERAP2*) and alternative APM gene (*CD1D*) found correlation to CD4+ T cell populations (Respectively: $R = 0.25, 0.34, 0.16$; $p < 0.01$; **See Figure 32**). A further correlation was observed in cluster 2 with *CD1D* ($R = 0.53$; $p < 0.0001$; **See Figure 32**) and an additional three correlations in cluster 3 with *CD1D*, *HLA-A*, *ERAP2* (Respectively: $R = 0.39, -0.37, 0.34$; $p < 0.05$; **See Figure 32**). A larger number of CD4+ T cell correlations were found with MHC class II gene expression including *CD74*, *HLA-DQA1*, *HLA-DRA* and *HLA-DPA1* (Respectively: $R = 0.24, 0.17, 0.16, 0.1$; $p < 0.05$; **See Figure 31**). No significant correlation was observed between APM gene expression regulators and CD4+ T cell abundance. Cluster 1 displays a further correlation is observed in *SPPL2A* ($R = 0.34$; $p < 0.0001$; **See Figure 32**), cluster 2 demonstrates a correlation in *HLA-DQA1* ($R = 0.33$; $p < 0.05$; **See Figure 32**) and finally a single correlation in cluster 3 with *HLA-DPA1* ($R = 0.38$; $p < 0.05$; **See Figure 32**).

Only two APM genes found correlation of NK cell populations, both belonging to the MHC class I pathway *ERAP1* and *PSMB10* (Respectively: $R = 0.15, 0.19$; $p < 0.05$; **See Figure 31**). Cluster 1 revealed a further two correlations in *HLA-A* and *TAPBP* (Respectively: $R = 0.72$,

0.70; $p < 0.05$; **See Figure 32**); cluster 3 demonstrates a further four correlation with *ERAP1*, *HLA-B*, and *PSMB8* (Respectively: $R = 0.45, 0.33, 0.32$; $p < 0.05$; **See Figure 32**).

Treg populations correlation with APM genes including two MHC class I genes *TAPBPL* and *PSMB10* (Respectively: $R = 0.15, 0.21$; $p < 0.05$; **See Figure 31**), and one alternative APM gene, namely *CD1D* ($R = 0.29$; $p < 0.01$; **See Figure 31**). Six MHC class II genes correlated to Treg populations including *CD74*, *CTSS*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DRA* and *HLA-DPA1* (Respectively: $R = 0.33, 0.17, 0.27, 0.23, 0.23, 0.21$; $p < 0.05$; **See Figure 31**).

Finally, an APM gene expression regulator, *CITA*, positively correlated to Treg populations ($R = 0.27$; $p < 0.01$; **See Figure 31**). Tregs demonstrate further correlations *HLA-DRB5* in cluster 1 ($R = 0.73$; $p < 0.05$; **See Figure 32**), *CD1D* and *HLA-DQA1* in cluster 2 (Respectively: $R = 0.44, 0.27$; $p < 0.05$; **See Figure 32**). Cluster 3 correlations with Tregs includes *CITA*, *HLA-DQA1*, *HLA-DPA1*, *CD74*, *HLA-DRB1* and *HLA-DRA* (Respectively: $R = 0.54, 0.5, 0.5, 0.47, 0.42, 0.39$; $p < 0.01$; **See Figure 32**); cluster 4 correlations with Tregs includes *PSMB10*, *LGMN*, *CD74*, *SPPL2A*, *TAPBP*, *HLA-A*, *PSMB8* and *CTSS* (Respectively: $R = 0.34, 0.33, 0.27, 0.27, 0.27, 0.26, 0.25, 0.25$; $p < 0.05$; **See Figure 32**)

4.4.3.2 Myeloid cell populations

Among the myeloid cell populations several APM genes with OS associations found correlation to subpopulation density in OAC including macrophages, DCs, granulocytes and monocytes.

Macrophage populations displayed correlation with MHC class I/II and alternative APM genes within my OAC cohort. Firstly, only one MHC class I gene *ERAP2* showed a significant positive correlation to macrophage populations ($R = 0.17$; $p < 0.05$; **See Figure 31**). MHC class II genes possessed a positive correlation to macrophage absolute scores including *CD74*, *LGMN* and *CTSL* (Respectively: $R = 0.2, 0.22, 0.27$; $p < 0.01$; **See Figure 31**). Finally, for Macrophages, alternative APM genes *CD1D* and *MR1* displayed a positive correlation to macrophage population (Respectively: $R = 0.32, 0.17$; $p < 0.05$; **See Figure 31**). Within individual clusters several correlations between APM genes and macrophages were observed, namely, in cluster 1 a correlation with *CD1D* ($R = 0.88$; $p < 0.01$; **See Figure 32**); in cluster 2 *CD1D* correlates with macrophages ($R = 0.36$; $p < 0.01$; **See Figure 32**). A further eight correlations with macrophages are observed in cluster 3 including *LGMN*,

CTSL, *MR1*, *CD1D*, *HLA-DRB5*, *HLA-E*, *HLA-DPA1* and *HLA-DQA1* (Respectively: $R = 0.62$, 0.57 , 0.5 , 0.46 , 0.33 , 0.33 , 0.33 , 0.32 ; $p < 0.05$; **See Figure 32**); a single additional correlation was identified in *CTSL* within cluster 4 ($R = 0.28$; $p < 0.01$; **See Figure 32**).

Among DCs only *CD1D* demonstrated a correlation to their score within my OAC cohort (Respectively: $R = 0.26$; $p < 0.001$; **See Figure 31**). A further three correlations were identified within the heatmap clusters including *HLA-B* in cluster 1 ($R = -0.71$; $p < 0.01$; **See Figure 32**), *CD1D* in cluster 2 ($R = 0.42$; $p < 0.01$; **See Figure 32**), and finally *LGMM* in cluster 3 ($R = -0.32$; $p < 0.01$; **See Figure 32**).

The granulocyte population positively correlated to the expression of one MHC class II gene, *CTSS* ($R = 0.22$; $p < 0.01$; **See Figure 31**), and one alternative APM gene, *CD1D* ($R = 0.25$; $p < 0.001$; **See Figure 31**). Further correlations within clusters included *CD1D* in cluster 1 ($R = 0.79$; $p < 0.001$; **See Figure 32**), *PSMB10* and *ERAP1* in cluster 2 (Respectively: $R = 0.38$, 0.27 ; $p < 0.05$; **See Figure 32**), *ERAP2* in cluster 3 ($R = 0.49$; $p < 0.001$; **See Figure 32**) and *HLA-DPA1* in cluster 4 ($R = -0.27$; $p < 0.001$; **See Figure 32**).

Finally, among myeloid populations, monocytes positively correlated to the expression of three MHC class II genes, *LGMM*, *CTSL*, and *HLA-DRB5* (Respectively: $R = 0.19$, 0.32 , 0.19 ; $p < 0.01$; **See Figure 31**), and one alternative APM gene *CD1D* ($R = 0.25$; $p < 0.01$; **See Figure 31**). Within cluster analysis of correlations with monocytes cluster 1 revealed a further correlation with *HLA-DRB5* ($R = 0.73$; $p < 0.05$; **See Figure 32**), cluster two additional correlation for *CD1D* and *CTSL* (Respectively: $R = 0.44$, 0.32 ; $p < 0.05$; **See Figure 32**), and lastly, a single correlation in cluster 3 with *ERAP2* ($R = 0.34$; $p < 0.05$; **See Figure 32**).

4.4.3.3 B lineage cell populations

Separately, B cells significantly negative correlated to one MHC class I gene, *CALR* ($R = -0.17$; $p < 0.01$; **See Figure 31**) and positively correlated to one alternative APM gene, *CD1D* ($R = 0.35$; $p < 0.001$; **See Figure 31**). A further three significant correlations between B cells and MHC class II was detected for *CD74*, *SPPL2A* and *HLA-DQA1* (Respectively: $R = 0.31$, -0.17 , 0.18 ; $p < 0.05$; **See Figure 31**). For B cells additional correlation were observed within individual clusters, including *CITA* and *CSDE1* in cluster 1 (Respectively: $R = -0.80$, 0.72 ; $p < 0.05$; **See Figure 32**), *CD1D*, *CD74*, *HLA-DQA1*, *HLA-DPA1*, *MR1* and *PSMB8* in cluster 2 (Respectively: $R = 0.64$, 0.37 , 0.34 , 0.31 , 0.31 , -0.27 ; p

<0.05; **See Figure 32**). Lastly, for B cell correlations, *SPPL2A* negatively correlated to B cell abundance in cluster 3 ($R = -0.31$; $p < 0.05$; **See Figure 32**).

Conversely to B cells, plasma cells only found positive correlation with APM gene expression for eight genes. Only one MHC class I gene, *TAPBPL*, correlated to plasma cell absolute scores ($R = 0.19$; $p < 0.01$; **See Figure 31**), also only one alternative APM gene, *CD1D* found a positive correlation with plasma cells scores ($R = 0.27$; $p < 0.001$; **See Figure 31**). Six MHC class II gene found positive correlation including *CD74*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DRA* and *HLA-DPA1* (Respectively: $R = 0.31, 0.16, 0.16, 0.24, 0.27$; $p < 0.05$; **See Figure 31**). No significant correlations were observed between APM gene regulators and plasma cell abundance in the total sample correlation analysis. However, within the cluster analysis *CSDE1*, *RFX5* and *CITA* found correlation with plasma cell abundance in cluster 1 (Respectively: $R = -0.80, 0.68, 0.73$; $p < 0.05$; **See Figure 32**), with a further correlation in cluster 1 found in *CD74* ($R = 0.70$; $p < 0.05$; **See Figure 32**). Within cluster 2 two further correlations with *CD1D* and *CD74* (Respectively: $R = 0.41, 0.35$; $p < 0.05$; **See Figure 32**) and an additional seven correlations in cluster 4 including *SPPL2A*, *HLA-DRA*, *TAPBPL*, *HLA-DPA1*, *CSDE1*, *CD74* and *RFX5* (Respectively: $R = 0.44, 0.34, 0.32, 0.32, 0.27, 0.26, 0.25$; $p < 0.05$; **See Figure 32**).

4.4.4 The impact of APM gene expression on immune effector/Treg cell populations in the OAC TIME.

Moving forward from correlation analysis, I progressed to statistically evaluate the difference in means/ medians for immune cell subpopulations CIBERSORT absolute scores by APM genes, which both possess a significant association to OS and a correlation between mRNA expression and immune cell subpopulation absolute scores from CIBERSORT analysis. This analysis would use the mRNA expression quantiles to dichotomise the CIBERSORT absolute scores. Starting with effector/Treg subpopulations, I firstly observed whether any statistically significant mean alterations between the low/high MHC class I mRNA expression groups as determined by quantile cut points.

Using quantile expression groups for MHC class I genes to dichotomise the CIBERSORT results for effector/Treg cell populations did find significant difference in mean scores in three of the ten correlations assessed. Foremost, I observed the association between HLAs loading complex assembly and alternative antigen presenting genes, finding the

lower quantile of *CALR* expression exhibited significantly greater CD8+ T cell abundance compared to the upper quantile (Upper: 0.09266 vs Lower: 0.1555; $p = 0.0386$; **See Figure 33 & Table 19**). Conversely, the lower quantile of *HLA-E* expression exhibited significantly lower CD8+ T cell abundance compared to the upper quantile (Upper: 0.1738 vs Lower: 0.1064; $p = 0.0433$; **See Figure 33 & Table 19**). Additionally, the lower quantile of *CD1D* expression related to a significant reduction in CD4+ T cell abundance (Upper: 1.093 vs Lower: 0.7891; $p = 0.0027$; **See Figure 33 & Table 19**).

Moving forward to the peptide generation genes, I found the lower quantile of *PSMB10* expression exhibited significantly lower Treg abundance compared to the upper quantile (Upper: 0.2008 vs Lower: 0.1438; $p = 0.0135$; **See Figure 33**). Conversely, the lower quantile of *PSMB8* expression exhibited a greater abundance of CD4+ T cells (Upper: 0.7762 vs Lower: 0.9827; $p = 0.001$; **See Figure 33 & Table 19**). Observing the *ERAP* genes I identified the lower quantile of *ERAP1* expression corresponded to a reduced presence of CD4 and NK cells in my OAC cohort (Respectively: Upper: 0.9234, 0.1844 vs Lower: 0.7459, 0.1309; $p = 0.0335, 0.0241$; **See Figure 33 & Table 19**). The final association observed among effector cells demonstrates the lower quantile of *ERAP2* expression in OAC is associated to reduced CD4+ T cell infiltration (Upper: 1.069 vs Lower: 0.8232; $p = 0.0163$; **See Figure 33 & Table 19**).

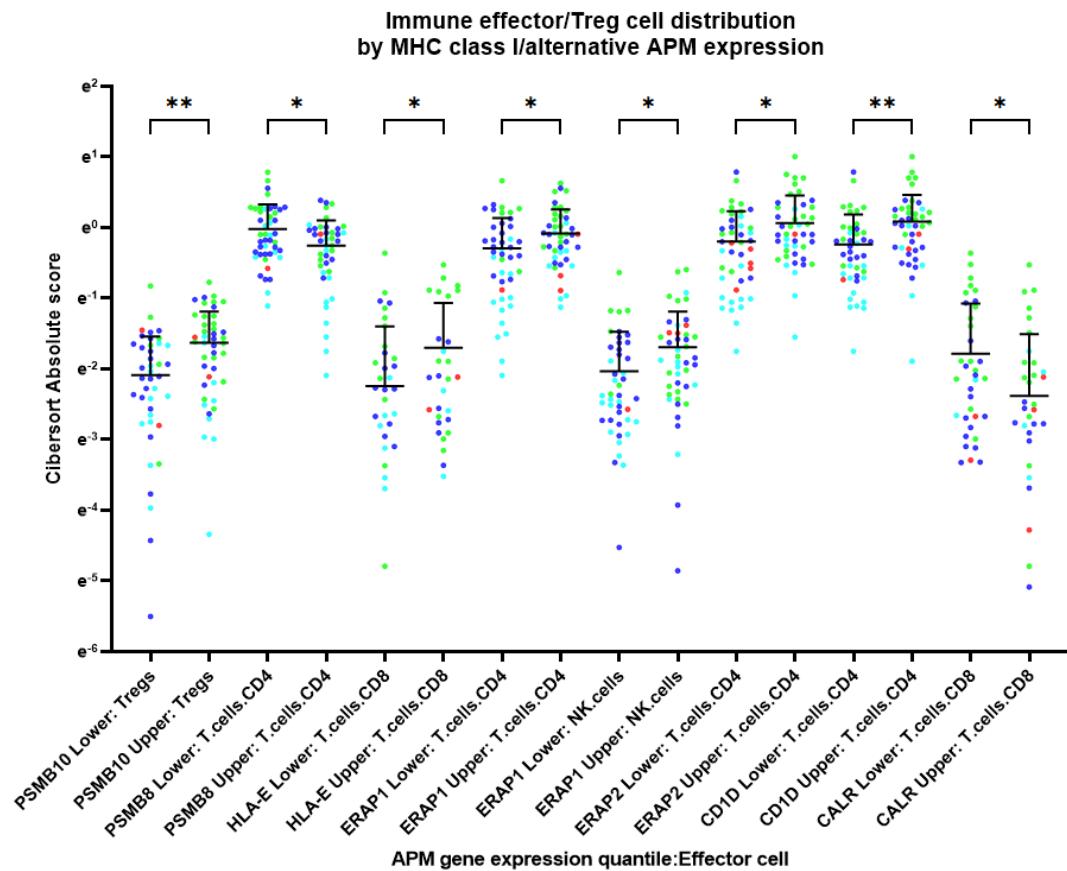


Figure 33 Boxplot of effector/Treg cell CIBERSORT absolute scores by quantile expression for MHC class I and alternative APM genes. Colours represent cluster of origin for the sample (Red: Cluster 1; Green: Cluster 2; Cyan: Cluster 3; Blue: Cluster 4). Statistical tests Mann-Whitney U test with FDR correction, p values * <0.05 , ** <0.01 , *** <0.001 , NS = non-significant.

Moving Beyond MHC class I, I assessed whether any statistically significant mean alterations between the low/high MHC class II mRNA expression groups as determined by quantile cut points. Consistently, the upper quantile of MHC class II expression in these five gene corresponded to increased Treg abundance including *CD74* (Upper: 0.2101 vs Lower: 0.1210; $p = 0.0190$; **See Figure 34 & Table 19**), the lysozyme *CTSS* (Upper: 0.2045 vs Lower: 0.1274; $p = 0.0228$; **See Figure 34 & Table 19**), and MHC class II HLAs, *HLA-DRB1* (Upper: 0.2042 vs Lower: 0.1143 ; $p = 0.0220$), *HLA-DRA* (Upper: 0.1973 vs Lower: 0.1159; $p = 0.0365$), *HLA-DRB5* (Upper: 0.1940 vs Lower: 0.1299; $p = 0.0347$), *HLA-DQA1* (Upper: 0.1177 vs Lower: 1.019; $p = 0.0166$) and *HLA-DPA1* (Upper: 0.1997 vs Lower: 0.1045; $p = 0.0098$) **See Figure 34 & Table 19**).

Increased CD4+ T cell abundance was identified among four HLA loading complex assembly genes of the MHC class II system, namely *CD74* (Upper: 1.014 vs Lower: 0.7823; $p = 0.0274$; **See Figure 34 & Table 19**), *HLA-DPA1* (Upper: 1.110 vs Lower: 0.7712; $p = 0.0018$; **See Figure 34 & Table 19**), *HLA-DQA1* (Upper: 1.019 vs Lower: 0.7241; $p = 0.0034$;

See Figure 34 & Table 19), and *HLA-DRA* (Upper: 1.032 vs Lower: 0.7833; $p = 0.0174$; See Figure 34 & Table 19).

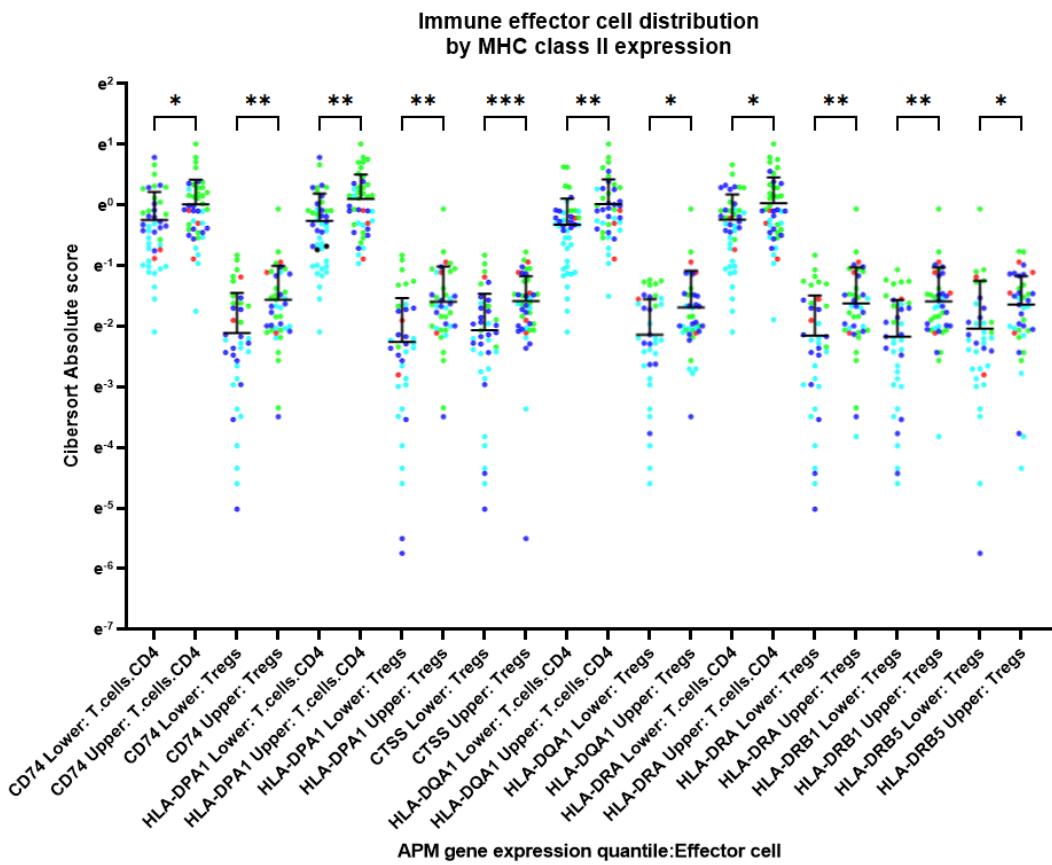


Figure 34 Boxplot of effector cell CIBERSORT absolute scores by quantile expression for MHC class II genes. Colours represent cluster of origin for the sample (Red: Cluster 1; Green: Cluster 2; Cyan: Cluster 3; Blue: Cluster 4). Statistical tests Mann-Whitney U test with FDR correction, p values* <0.05 , ** <0.01 , *** <0.001 , NS = non-significant.

Finally, for effector cells I explored the impact on *CITA* and *CSDE1* using quantile cut point dichotomisation on Treg and CD8+ T cell abundance scores, respectively. Here I found significant differences in mean scores for each interaction, with the upper quantile of *CSDE1* mRNA expression relating to lower CD8+ T cell abundance (Upper: 0.08498 vs Lower: 0.1597; $p = 0.0217$ See Figure 35 & Table 19), and the upper quantile of *CITA* expression relating to greater Treg abundance (Upper: 0.2071 vs Lower: 0.1328; $p = 0.0111$ See Figure 35 & Table 19). Additionally, the upper quantile of *CITA* expression corresponded to greater CD4+ T cell abundance within my OAC cohort (Upper: 1.066 vs Lower: 0.7869; $p = 0.0064$; See Figure 35 & Table 19). The final effector T cell APM expression regulator was observed in *RFX5*, where the lower quantile of expression

presented a reduction in Treg populations (Upper: 0.1914 vs Lower: 0.1472; $p = 0.0433$ See Figure 35 & Table 19).

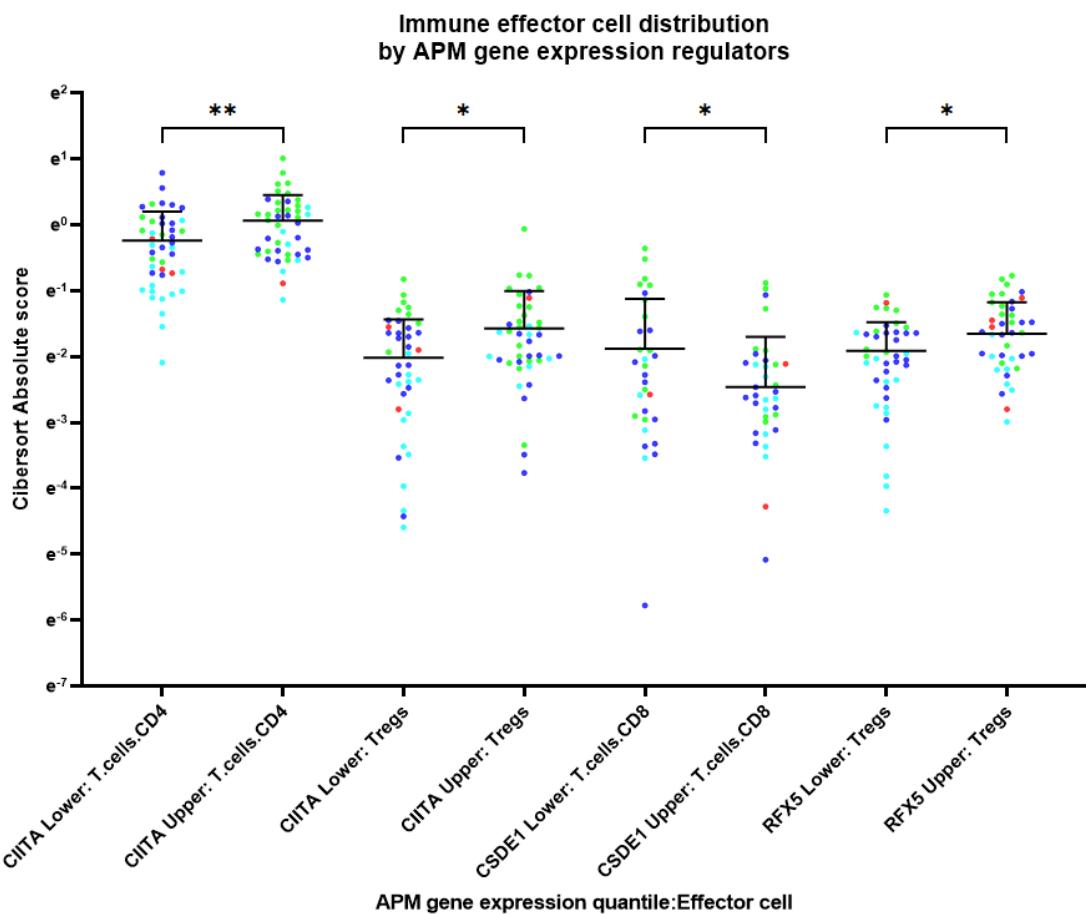


Figure 35 Boxplot of effector cell CIBERSORT absolute scores by quantile expression for APM regulator genes. Colours represent cluster of origin for the sample (Red: Cluster 1; Green: Cluster 2; Cyan: Cluster 3; Blue: Cluster 4). Statistical tests Mann-Whitney U test with FDR correction, p values * <0.05 , ** <0.01 , *** <0.001 , NS = non-significant.

4.4.5 The impact of APM gene expression on myeloid cell populations in the OAC TIME.

Next, I explored the impact of APM gene expression on myeloid population cells including macrophages, DCs, granulocytes and monocyte subpopulations dichotomised by quantile cut points, mirroring the above CIBERSORTx analysis.

Among HLA loading complex assembly genes, both *HLA-E* and *TAPBPL* demonstrated a corresponding increase in myeloid populations, specifically, the upper quantile of *HLA-E* expression related to increased monocyte populations (Upper: 0.1218 vs Lower: 0.05796; $p = 0.0437$; See Figure 36 & Table 19), while, the upper quantile of *TAPBPL* associated to

greater macrophage abundance (Upper: 0.6645 vs Lower: 0.4917; $p = 0.0479$; **See Figure 36 & Table 19**). Two peptide generation genes within the MHC class I gene candidates associated to altered myeloid populations; the upper quantile of *PSMB10* expression corresponded to lesser dendritic cell abundance (Upper: 0.1130 vs Lower: 0.1684; $p = 0.0244$; **See Figure 36 & Table 19**), conversely, the upper quantile of *ERAP1* expression related to greater granulocyte populations in OAC (Upper: 0.6066 vs Lower: 0.4048; $p = 0.0061$; **See Figure 36 & Table 19**).

Exploring myeloid population abundance differences due to quantile expression of alternative APM gene *CD1D*, of which the upper quantile of expression related to increased macrophage populations (Upper: 0.7612 vs Lower: 0.1520 ; $p = 0.0022$; **See Figure 36 & Table 19**), an increased monocyte abundance (Upper: 0.1433 vs Lower: 0.07333; $p = 0.0381$; **See Figure 36 & Table 19**), and lastly an increased granulocyte abundance (Upper: 0.6459 vs Lower: 0.4181; $p = 0.03$; **See Figure 36 & Table 19**). Another alternative APM gene, *MR1*, also found association to macrophage abundance with the OAC cohort with the upper quantile of expression resulting in greater macrophage populations (Upper: 0.7109 vs Lower: 0.5358; $p = 0.0455$; **See Figure 36 & Table 19**).

Observing the impact of MHC class II expression on myeloid cell CIBERSORT scores yielded nine significant differences in myeloid subpopulations. Observing the HLA loading complex assembly genes of the MHC class II system, the upper quantile of *CD74* expression group there was a notable increase in macrophage (Upper: 0.7401 vs Lower: 0.5040; $p = 0.0345$; **See Figure 37 & Table 19**) and granulocyte populations (Upper: 0.6757 vs Lower: 0.4550; $p = 0.0127$; **See Figure 37 & Table 19**).

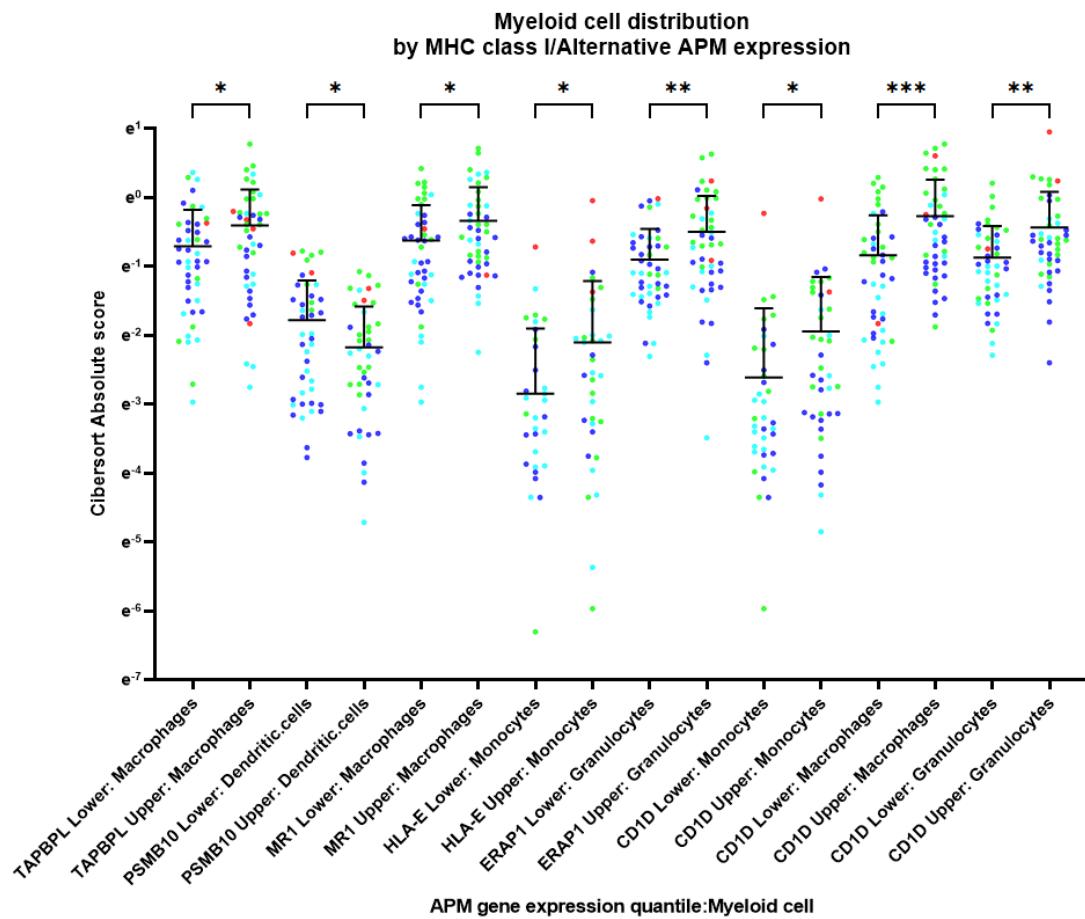


Figure 36 Boxplot of myeloid cell CIBERSORT absolute scores by quantile expression for MHC class I genes. Colours represent cluster of origin for the sample (Red: Cluster 1; Green: Cluster 2; Cyan: Cluster 3; Blue: Cluster 4). Statistical tests Mann-Whitney U test with FDR correction, p values * <0.05 , ** <0.01 , *** <0.001 , NS = non-significant.

The upper quantile of four MHC class II HLA genes corresponded to increased macrophage abundance, including *HLA-DPA1* (Upper: 0.7567 vs Lower: 0.4706; $p = 0.0043$; **See Figure 37 & Table 19**), *HLA-DQA1* (Upper: 0.7537 vs Lower: 0.4878; $p = 0.0073$; **See Figure 37 & Table 19**), *HLA-DRA* (Upper: 0.7202 vs Lower: 0.5175; $p = 0.0354$; **See Figure 36**), and *HLA-DRB5* (Upper: 0.6992 vs Lower: 0.4833; $p = 0.0251$; **See Figure 37 & Table 19**).

Finally, MHC class II peptide generation gene expression demonstrated association with altered myeloid populations, firstly, the upper quantile of *CTSL* expression group persisted a similar increase in macrophage score (Upper: 0.7968 vs Lower: 0.4802; $p = 0.0163$; **See Figure 37 & Table 19**); The upper quantile of *LGMD* expression group persisted a similar increase in macrophage score (Upper: 0.7412 vs Lower: 0.5395; $p = 0.0375$; **See Figure 37 & Table 19**).

Lastly, the upper quantile of *CTSS* expression group persisted a similar increase in granulocyte abundance (Upper: 0.6260 vs Lower: 0.4208; $p = 0.0156$; **See Figure 37 & Table 19**).

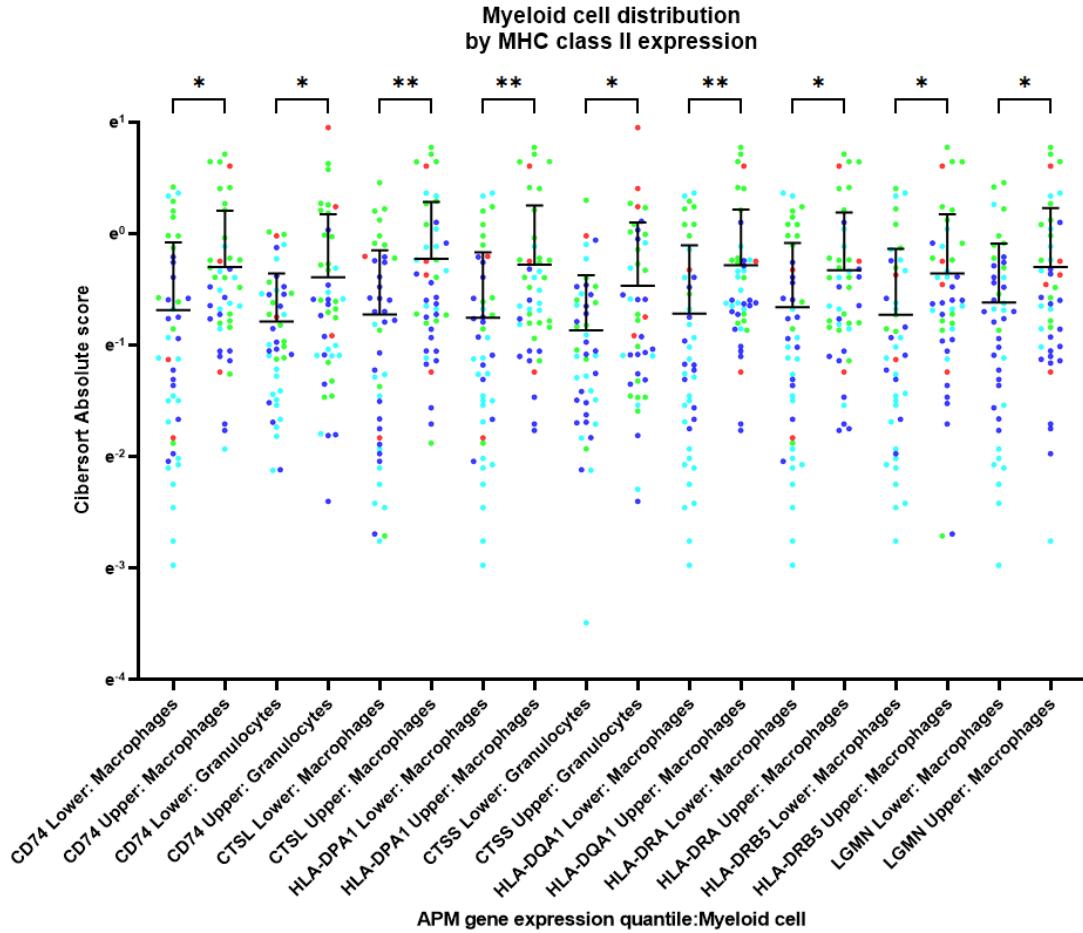


Figure 37 Boxplot of myeloid cell CIBERSORT absolute scores by quantile expression for MHC class II genes. Colours represent cluster of origin for the sample (Red: Cluster 1; Green: Cluster 2; Cyan: Cluster 3; Blue: Cluster 4). Statistical tests Mann-Whitney U test with FDR correction, p values $* < 0.05$, $** < 0.01$, $*** < 0.001$, NS = non-significant.

4.4.6 The impact of APM gene expression on B lineage cell populations in the OAC TIME

Lastly, I explored the impact of APM gene expression of B lineage cell population within my OAC cohort using quantile mRNA expression for APM genes. Assessing the MHC class I/alternative APM expression comparing quantile expression of the genes identified only two genes, namely *CD1D*, produced a significant reduction in B cell abundance within my OAC cohort (Upper: 0.6486 vs Lower: 0.3797; $p = 0.0015$; **See Figure 38 & Table 19**). The second association observed displayed the upper quantile of *CALR* expression possessed greater B cell populations in OAC (Upper: 0.4489 vs Lower: 0.6470; $p = 0.0491$; **See Figure 38 & Table 19**).

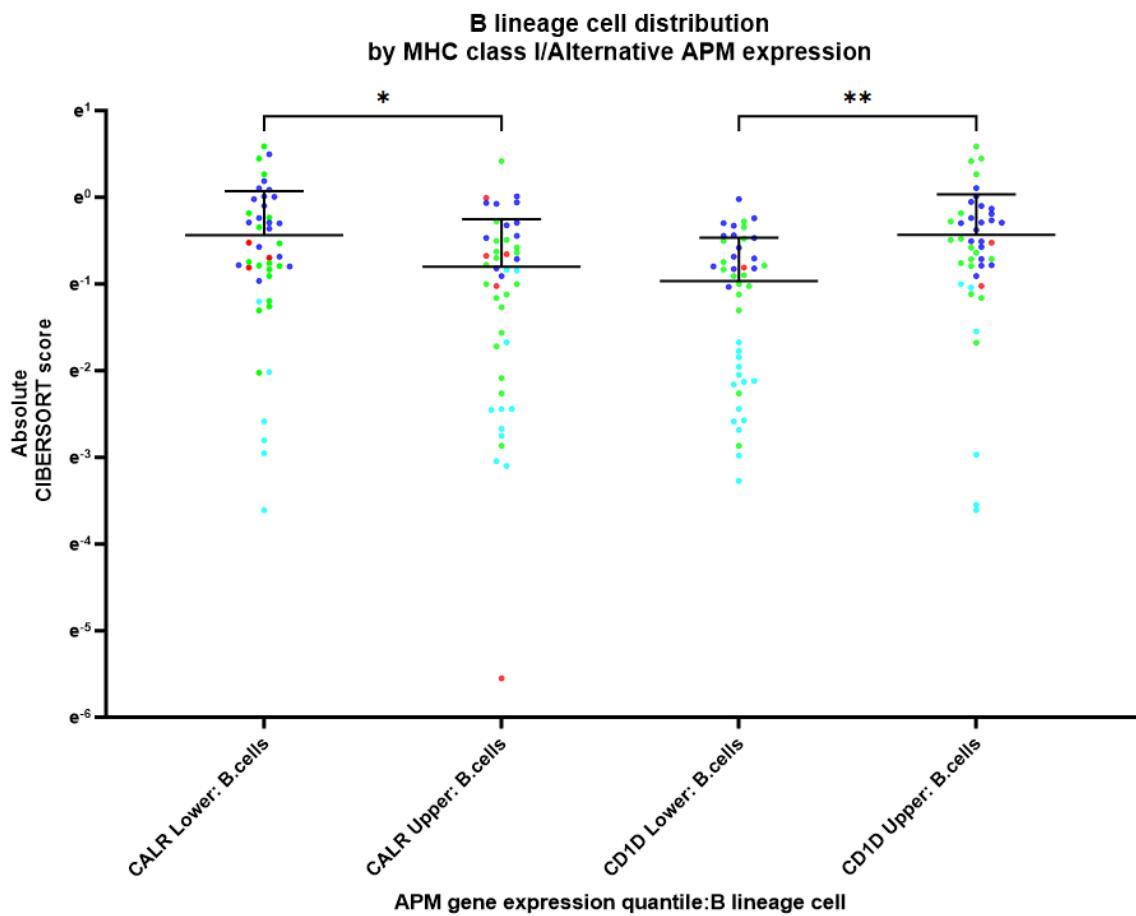


Figure 38 Boxplot of B lineage cells CIBERSORT absolute scores by quantile expression for MHC class I genes. Colours represent cluster of origin for the sample (Red: Cluster 1; Green: Cluster 2; Cyan: Cluster 3; Blue: Cluster 4). Statistical tests Mann-Whitney U test with FDR correction, p values* <0.05 , ** <0.01 , *** <0.001 , NS = non-significant.

The impact of MHC class II expression on B lineage cell distribution using the quantile cut points for MHC class II genes. Using this approach, I discovered five MHC class II genes which expression resulted significantly different B lineage cell abundance. The HLA loading complex assembly genes associated to altered B lineage cell populations, such as, the upper quantile of *HLA-DQA1* expression contained greater B cell abundance compared to the lower quantile of expression within my OAC cohort (Upper: 0.5555 vs Lower: 0.3072; $p = 0.0130$; **See Figure 39 & Table 19**). The upper quantile of *HLA-DRB1* expression within my OAC cohort scored higher for plasma cell abundance (Upper: 0.3470 vs Lower: 0.1647; $p = 0.0320$; **See Figure 39 & Table 19**). Interestingly, the upper quantile of *HLA-DRB5* expression corresponded to both increased B cell (Upper: 0.4738 vs Lower: 0.3166; $p = 0.0320$; **See Figure 39 & Table 19**) and plasma cell (Upper: 0.2686 vs Lower: 0.1238; $p = 0.0276$; **See Figure 39**) concentrations within my OAC tumours.

Finally, a peptide generation gene of the MHC class II system, *SPPL2A*, demonstrated an association with B cell populations, with the upper quantiles of expression corresponding to lower B cell abundance (Upper: 0.4228 vs Lower: 0.6135; $p = 0.0115$; **See Figure 39 & Table 19**).

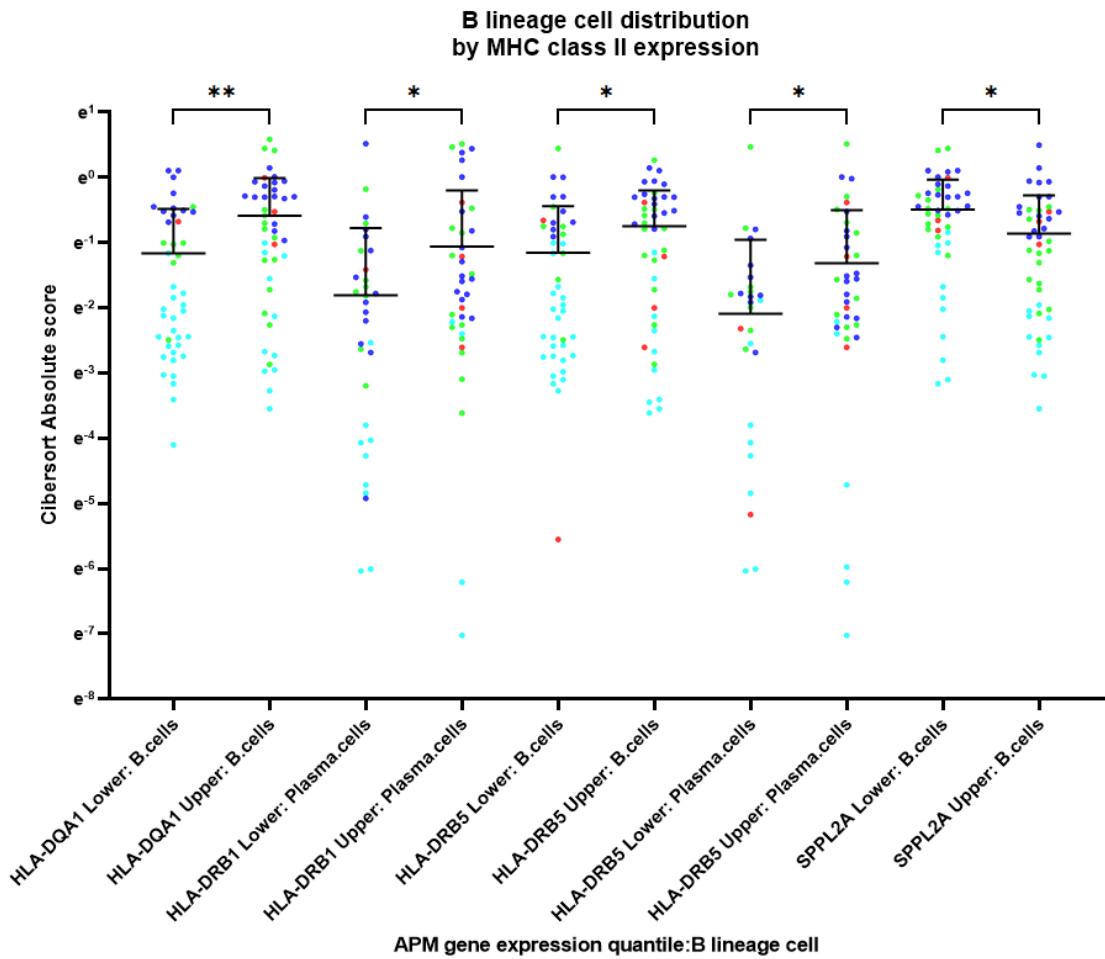


Figure 39 Boxplot of B lineage cells CIBERSORT absolute scores by quantile expression for MHC class II genes. Colours represent cluster of origin for the sample (Red: Cluster 1; Green: Cluster 2; Cyan: Cluster 3; Blue: Cluster 4). Statistical tests Mann-Whitney U test with FDR correction, p values $* < 0.05$, $** < 0.01$, $*** < 0.001$, NS = non-significant.

Table 19 Summary of significant immune cell content difference between upper and lower APM gene expression quantiles.

GENE	IMMUNE CELL TYPE	UPPER-LOWER QUANTILE IMMUNE CONTENT Δ	P VALUE
CTSS	TREG	0.077	0.0008
	GRANULOCYTES	0.21	0.016
CD1D	MACROPHAGES	0.33	0.0009
	GRANULOCYTES	0.23	0.0038
	B CELL	0.27	0.0015
	CD4+ T CELL	0.30	0.0027
	MONOCYTES	0.070	0.038
PSMB10	TREG	0.072	0.001
	DENDRITIC CELL	-0.055	0.024
HLA-DPA1	TREG	0.099	0.0012
	CD4+ T CELL	0.34	0.0018
	MACROPHAGES	0.29	0.0043
HLA-DRB1	TREG	0.090	0.0013
	PLASMA CELL	0.18	0.032
CD74	TREG	0.089	0.0031
	GRANULOCYTES	0.22	0.013
	MACROPHAGES	0.24	0.017
	CD4+ T CELL	0.23	0.027
HLA-DQA1	CD4+ T CELL	0.29	0.0034
	B CELL	0.24	0.0034
	MACROPHAGES	0.29	0.0073
	TREG	0.067	0.017
ERAP1	GRANULOCYTES	0.20	0.0061
	CD4+ T CELL	0.18	0.034
	NK CELL	0.054	0.024
CIITA	CD4+ T CELL	0.28	0.0064
	MACROPHAGES	0.22	0.019
	TREG	0.074	0.0113
HLA-DRA	TREG	0.081	0.0066
	CD4+ T CELL	0.249	0.017
	MACROPHAGES	0.20	0.035
PSMB8	CD4+ T CELL	-0.21	0.010
SPPL2A	B CELL	-0.19	0.012
ERAP2	CD4+ T CELL	0.25	0.016
HLA-DRB5	MACROPHAGES	0.22	0.025
	PLASMA CELL	0.15	0.028
	B CELL	0.16	0.032
	TREG	0.064	0.035
CALR	CD8+ T CELL	-0.076	0.026
	B CELL	-0.20	0.049
LGMN	MACROPHAGES	0.20	0.038
CSDE1	CD8+ T CELL	-0.068	0.042
HLA-E	CD8+ T CELL	0.076	0.043
	MONOCYTES	0.064	0.044
RFX5	TREG	0.044	0.043
TAPBPL	MACROPHAGES	0.17	0.048

4.5 Discussion

The analysis in this chapter, allows for broad investigation of the relationship of our prognostically significant APM genes and the TIME using the results of deconvolution themselves in context with the known literature on the TIME in cancers. A key outcome of this chapter was the potential of altered immunity due to the expression of several prognostically significant APM genes which were identified in **Chapter 3**, forming novel knowledge on the immunity of OAC tumours.

4.5.1 The TIME of OAC is diverse and is broadly defined into four phenotypes.

The first significant finding of this Chapter's analysis was the observation of four distinct immunophenotypes. These immunophenotypes can be broadly classified by their constituent immune abundances. Cluster 1 is a small cluster ($n = 9/176$, 5.1% of cases) and can be defined as a monocyte-high immunophenotype; exploring the literature surrounding the function of monocytes provides insight into the variable functions within cancer, finding both pro-tumoral and anti-tumoral roles of monocytes. Examples of pro-tumoral roles of monocytes include the suppression of T cell functions in colorectal tumours (Mouse model) and pancreatic cancer, and the recruitment of Tregs in pancreatic ductal adenocarcinoma (Mouse model). Anti-tumoral roles of monocytes include tumour cytotoxicity in ovarian cancer and antigen presentation in B16 melanoma (Mouse model) (424-427). Unfortunately, no significant survival differences could be determined in this cluster compared to other clusters, despite this, cluster 1 possessed the lowest hazard ratio for overall survival, this trend could suggest the presence of a high monocyte immunophenotype could be a positive for overall survival. However, without a greater sample size this finding cannot be validated. Interestingly, the value of this first immunophenotype has been prior publication supporting the presence of monocytes in OAC is a positive predictor of immunochemotherapy outcomes (428).

The second cluster could be described as a classical 'immune hot' immunophenotype, possessing high effector cell content compared to the other clusters; this immunophenotype represents the greatest opportunity for immunotherapy applications such as checkpoint blockade. Furthermore, this immunophenotype displays greater potential T cell derived cytotoxicity as predicted by my analysis (429-431). Interestingly, this may be reflected in OAC with the attempt to create a 'cancer-immune set point', which describes the immune profile threshold of intrinsic factors such as immune cell

abundance, cytokine secretion and genetic composition and extrinsic factors including gut microbiota and the presence of infection, which is required to pass to predict sensitivity to immune checkpoint inhibition (432). However, results also point to an unfortunate outcome in determining patient care pathways with only 55/176 (31.3%) patients (cluster 2) may be potentially able to benefit from checkpoint blockade therapy. The third cluster forms an opposing immunophenotype to cluster 2, this immunophenotype can be characterised describe as immune cold, presenting a lack of TILs compared to the other three clusters. Immune cold tumours present a significant issue in producing an effective immunotherapy to drive anti-tumoral immunity. Interestingly, this immunophenotype did present with the greater hazard ratio for overall survival, despite not being a significant finding, this could therefore represent a problematic cluster of patients forming approximately a quarter of all patients ($n = 45/176$, 25.6%), which are unlikely to benefit from ICB immunotherapy and possess a survival disadvantage (433).

However, the gross categorisation of tumours as immune cold is currently being challenged; where intermediate immunophenotypes may yield alternative approaches to immunotherapy application with the literature suggesting new definition such as immune exclude and immunosuppressed which may need these alternative approaches (434, 435). Thereby, the identification of this cluster within my OAC cohort supports the argument of immune cold tumours are highly frequent in OAC dissuading a single approach should be taken in the immunotherapy of OAC, instead observing immunological feature prior to treatment to determine the appropriate treatment pathway.

The final immunophenotype is characterised as a high B cell type; the presence of B cells in oesophageal adenocarcinoma and gastric carcinomas with their positive role in survival has been prior published, suggesting the presence of CD20+ B cells associates with survival outcomes and the recruitment of other effector cells including T cells and NK cells (383, 436, 437). The identification of this immunophenotype which makes up the majority of OAC patients ($n = 67/176$, 38.1%) demonstrates immunotherapy approaches focusing on the interaction of B lineage cells in the TIME of OAC may present an inviting opportunity for future immunotherapies. For example, the presence of B cell has correlated to suppression of T cell responses via upregulating PD-L1 expression in the TIME, as demonstrated in breast and pancreatic cancers; targeting these B cell enriched

patients with PD-1/L1 inhibitors may yield greater T cell responses in these cases (438, 439).

Overall, this analysis has outlined the diversity of immunophenotypes present in OAC, understanding the immunophenotypes may guide future efforts to stratify patients to apply the most appropriate treatment pathway. Unfortunately, this analysis highlights checkpoint blockade therapy may not be applicable in all cases; thereby, using this approach provides justification selecting immune hot patients for checkpoint blockade.

4.5.2 Effecter cell populations are increased in cases with high expression of MHC class I in OAC

Observing effector cell populations, several subpopulations appeared altered due to the expression of MHC class I genes. This result was expected, as the MHC class I system is responsible to eliciting responses from effector cells, directly interacting with CD8+ T cells.

Most of the MHC class I genes which influence effector cell populations were peptide generation genes. Firstly, the increased expression of *PSMB8* in OAC tumours appears to result in lesser CD4+ T cells. This appears to oppose prior literature in multiple cutaneous malignant melanoma showing the expression of *PSMB8* correlated with M1 macrophages, CD8 T cells, CD4+ T cells, follicular helper T cells, $\gamma\delta$ T cells, regulatory (Tregs) T cells, and activated NK cells. However, within this study *PSMB8* did negatively correlate to M0 macrophages, resting mast cells, and CD4 resting memory T cells (440). This could suggest the impact on reducing CD4 resting memory T cells is greater than potentially increase in other CD4+ T cells in OAC.

The greater expression of *PSMB10* in my analysis resulted in greater presence of Tregs, exploring the literature did not yield any prior publication demonstrating an interaction between *PSMB10* expression and Treg populations. This could suggest Treg abundance is not directly impacted by the expression of *PSMB10*, instead the increase in their presence may be due to other immune cells as *PSMB10* also alters other cells, such as dendritic cells demonstrated in the literature (441).

The latter peptide generation genes to demonstrate altered effector cell abundance were *ERAP1* and *ERAP2*. Greater *ERAP1* expression corresponded to increased CD4 and NK cell populations within my OAC cohort; for CD4+ T cell increases this could be explained by

the literature which details in ankylosing spondylitis the silencing of *ERAP1* results in suppressed CD4+ Th17 cell expansion by preventing the binding of HLA-B27 free heavy chains to KIR3DL2, suggesting this mechanism may be present in cancers such as OAC (442, 443). The interaction between *ERAP1* and NK cells as shown in my analysis has been prior demonstrated in cancer via a similar mechanism, in a transfected cell model of acute lymphoblastic lymphoma which demonstrated the inhibition *ERAP1* renders HLA-B*51:01 molecules less eligible for KIR3DL1 binding (444, 445). Thereby, using the literature in combination with my analysis it can be suggested *ERAP1* in OAC alters *HLA-B* allotype binding to KIRs, which results in altered CD4+ T cell and NK cell recruitment and expansion. *ERAP2* expression correlating with CD4+ T cell density has prior been demonstrated in squamous cell lung carcinoma, yet no mechanism is yet determined (282).

Next, I observed the associations between effector cell populations and HLA loading complex assembly genes and alternative APM genes. Firstly, greater *HLA-E* expression within my cohort resulted in higher CD8+ T cell abundance; the impact of *HLA-E* expression on CD8 T cells in OAC may relate to alternatively spliced isoforms of the antioxidant enzyme peroxiredoxin 5 peptides bound to *HLA-E* capable of being recognised by *HLA-E*-specific CD8 T cells; this interaction may further relate to findings that Peroxiredoxin 5 overexpression has been identified in gastric adenocarcinomas (367, 368).

Higher *CALR* expression within my OAC cohort corresponded to decreased CD8+ T cells, this opposes the literature which reports the loss of *CALR* expression negatively affects immunosurveillance, with the expression of *CALR* correlating with TILs in three cancer types including colorectal, breast and ovarian (446, 447). Unfortunately, the literature could not help to explain my analysis results in this case, further exploration of this novel result may be warranted in future investigations.

Finally, *CD1D* expression in my OAC cohort related to increased CD4+ T cells, this can be explained via publication demonstrating CD4+ CD1d-restricted NKT cells exist within immune microenvironments, thereby, *CD1D* may be directly interacting with these specific NKT cells in the tumour immune microenvironment of OAC (448, 449). Unfortunately, this specific subset of NKT has not been identified in oesophageal

adenocarcinoma to date, thereby, future analysis using flow-cytometry methodology could pinpoint this interaction in OAC (not attempted within my study).

Overall, my analysis presented here demonstrates the importance of MHC class I in eliciting effector cell responses in OAC, especially, peptide generation genes of the MHC class I system could specifically alter interactions with CD4+ T cells and NK cells. Exploiting these genes to produce future immunotherapies for OAC may yield an effective therapy for those possessing the upper quantile of *ERAP1/2* expression in the population (n=44/174). Assessing these results highlights the importance of intact and well-expressed MHC class I molecules to produce effective T cell responses in OAC, in particular, the expression of peptide generation genes may produce altered immunity by regulating the peptide epitope as already prior literature has found for the *PSMB8/9/10* and *ERAP1/2* genes (450-454).

4.5.3 Effecter cell populations are greater in cases with characteristically high expression of MHC class II in OAC

Next within my study, I observed altered effector cell populations, due to the expression of MHC class II genes. This result was expected, as the MHC class II system is solely present on professional APCs which engage in T cell education, directly interacting with CD4+ T cells.

For the specific CD4+ T cell density associations four HLA-loading complex assembly genes found a correspondence between greater expression and higher CD4+ T cell abundance, while six MHC class II genes were associated to increased Treg populations, including five HLA-loading complex assembly genes and a single peptide generation gene.

Firstly, *CD74* demonstrated a positive association with CD4+ T cell and Treg abundance; this gene's primary function is to function as a chaperone which regulates antigen presentation, though a secondary function is found as cell surface receptor for the cytokine macrophage migration inhibitory factor (MIF). The literature does demonstrate this relationship (*CD74*:CD4+ T cell) persists in breast cancer where the expression of *CD74* presents as a positive prognostic factor in basal-like breast cancer attributed to greater densities of TILs including CD4+ T cells (298). Interestingly, *CD74* in the literature has prior exhibited an ability to perform dendritic cell cross-priming of CD8+ T cells in viral infections, yet no increase in CD8+ T cell populations was attributed to *CD74* expression

within my study, this could suggest cross-presentation within dendritic cells is not a frequent occurrence within OAC (455). The increase of Treg populations attributed to the expression of *CD74* has prior been approached in the recent literature; here *CD74* is thought to support the accumulation and functions of Tregs. Specifically, *CD74* was located to the surface Tregs in non-small cell lung cancer and was overexpression in tumour-associated Treg as opposed to blood circulating Tregs, this could suggest *CD74* expression may be a marker of increased tumour-associated Tregs, but must be further assessed in future research of OAC to prove the relationship (456).

The expression of *HLA-DPA1* corresponded to an increase in CD4+ T cell density within my OAC cohort, this can be explained via *HLA-DPA1/CD4* mediating cell crosstalk between CD4 + T conv cells and pDC cells demonstrated in head and neck squamous cell carcinomas, suggesting the results observed in OAC may be mediated by dendritic cells in the TIME (457). *HLA-DPA1* also corresponded to Treg populations, unfortunately, this exact relationship did not find explanation in the literature, however, this association may be simply driven by co-recruitment of activated effector cells to inflammatory microenvironments observed in OAC (458). The greater expression of *HLA-DQA1* appeared to result in greater CD4+ T cells and Tregs within OAC. This could suggest expression of this gene within professional APCs corresponds to the activation and recruitment of CD4+ T cells within OAC patients to the tumour site as demonstrated in breast cancer. Whereas, extracellular *HLA-DQA1* may be a component in inducing Tregs as shown in celiac disease, but not yet observed in cancer (459, 460).

Similarly the higher expression of *HLA-DRA* was also associated to increased abundance of CD4+ T cells and Tregs within of OAC microenvironment. The correspondence with CD4+ T cells has prior been identified in non-small cell lung cancer and renal cell carcinomas and potentially is produce via two mechanisms. Namely, the recruitment and activation role of HLA-DR proteins bound to APCs on CD4+ T cells, and/or the expression of HLA-DR on CTLs (291, 461-464). Mirroring the CD4+ T cell relationship, the *HLA-DRA* and Treg association may also relate to expression of *HLA-DRA* on Tregs as shown in cervical squamous cell carcinoma; however, in a disease similar to OAC, gastric cancer, a publication suggests HLA-DR expression on pDCs, myeloid DCs (mDCs), macrophages, and B cells may relate to *ICOS-L* expression by these cells the activation of Tregs (ICOS+ Treg induction) (465, 466).

The final HLA-loading complex assembly genes to find association to effector cells were *HLA-DRB1* and *HLA-DRB5* with Treg populations, the relationship observed here may relate to prior discussed expression of these genes in Tregs as shown colorectal cancer liver metastases and ovarian cancer (467).

Lastly, the single peptide generation gene to associate with increased Tregs was identified in *CTSS*, where similarly to some HLA molecules, *CTSS* has been shown to be expressed in Tregs of bladder cancer, and relates to increased CD8+ T cell proliferation and apoptosis of cancer cells (468). Unfortunately, no relationship between *CTSS* and CD8+ T cell abundance was observed within my analysis, thus weakening the above suggestion in OAC.

Overall, the direct interaction of MHC class II antigen presenting molecules and CD4+ T cells appears to be an occurrence in OAC, and the expression of these molecules may impact activity of the CD4+ T cell subset. However, Tregs present a more complex finding with the literature supporting several of the HLA-loading complex assembly genes may be expressed by tumour associated Tregs or Treg activation may be linked to interaction with DCs expression of these genes. Which of these outcomes is most likely can be determined when exploring the associations between HLA-loading complex assembly genes and myeloid populations later in this discussion.

4.5.4 Myeloid cell populations are altered in cases due to MHC class I expression in OAC.

Following my analysis of effector cells and their association to APM genes, I explored the alteration of myeloid cell populations, due to the expression of APM genes. Starting with MHC class I genes, I observed altered myeloid cell subset abundances linked to specific MHC class I genes including two involved in each function of HLA-loading complex assembly, peptide generation and alternative antigen presentation. This result was less expected, as the MHC class I system is present on professional APCs, but presentation between cancer cells and myeloid cells does not occur in a conventional manner.

Firstly, greater expression of HLA-loading complex assembly genes *TAPBPL* and *HLA-E* lead to increased abundance of macrophages and monocytes, respectively. The *TAPBPL* relationship may be explained by its expression in macrophages and the expression-cell abundance correlation has prior been identified in breast cancer, with the low expression of *TAPBPL* relating to increased risk (260, 469). Increased *HLA-E* expression in OAC related to greater abundance of monocytes, an explanation ground in cancer was not found in cancer studies. However, in juvenile idiopathic arthritis, a disease with a highly inflammatory microenvironment akin to cancer, the upregulation of *HLA-E* expression was noted in B cell and monocyte populations with an interesting down-regulation of *HLA-E* among T cell populations (470). This does propose an interesting question about the cell of origin or mechanism the relation between *HLA-E* and monocytes employs in OAC, which could be explored in future studies.

The increased expression of peptide generation genes *PSMB10* and *ERAP1* corresponded to increased abundance of dendritic cells and granulocytes, respectively. The former relationship has been prior elucidated in dendritic cells as the expression of immunoproteasome components appears integral to the maturation of these cells, Furthermore, the lack of *PSMB10* expression in myeloid populations results in deficient MHC class I expression within DCs and correlates to poor survival outcomes (471-473). The later relationship between higher *ERAP1* expression and increased presence of granulocytes could not be explained by the literature in a cancer setting, despite this, lymphocytes stimulated by *ERAP1/2* expressing choriocarcinoma cells in a preeclampsia setting resulted in the increase of granulocyte-macrophage colony-stimulating factor (GM-CSF). GM-CSF functions to enhance neutrophils and macrophages activity and has prior been investigated in human gastrointestinal infections and cancer, suggesting activation and migration of myeloid cells such as neutrophils to inflammation sites may be regulated by this cytokine (474-476).

Finally, observing alternative APM genes found greater *MR1* expression related to increased macrophage populations; within the literature this relationship provides a complex contradiction, firstly MAIT cells which *MR1* directly interacts with appears to lead to phenotype shift of liver macrophages and within the same study *MR1* was shown to be expressed in Kupffer cell macrophages (477). On the other hand, a study of the human intestinal mucosa found a lack of expression of *MR1* in macrophages (478). Thereby, at this stage the exact mechanistic relationship between *MR1* expression and

macrophage populations cannot be explained, made especially difficult by the limitations on cell type evading MAIT cell in CIBERSORTx.

CD1D expression positively corresponded to increased populations of monocytes, macrophages and granulocytes within the OAC microenvironment, notably, this is an expected result in monocytes and macrophages as *CD1D* is a marker for these cell types (479). However, the relationship with granulocytes is more interesting and may relate to *CD1D*-restricted NKT cells encompassed by the granulocyte immune cell type, demonstrating a potential NKT cells may be reacting with monocytes and macrophages present in OAC tumours (480). Additionally, a prior publication demonstrates the *CD1D*-restricted NKT cells may produce granulocyte–macrophage colony-stimulating factor to recruit neutrophils (479).

Holistically this analysis demonstrates that although few direct conventional interactions occur between MHC class I genes and myeloid population, the expression of MHC class I genes can influence myeloid populations, however, difficulty arises when attempting to deconvolute the mechanisms behind such associations as several MHC class I and especially alternative APM genes can form markers of myeloid populations.

4.5.5 Myeloid cell populations are altered in cases with high MHC class II expression in OAC.

Following the associations between myeloid cells and MHC class I expression, I moved forward to assessed myeloid relationships to MHC class II expression, this analysis presents with added complexity as myeloid cells express MHC class II system genes as part of their primary function to directly interact with and educate CD4+ T cells. Thereby, in discussing these results particular attention must be paid in determining the likelihood that an association observed is a product of a function of a particular MHC class II gene or whether the association is likely due to the gene being a marker of a myeloid cell population.

Starting with the HLA-loading complex assembly genes of the MHC class II system, the higher expression of *CD74* corresponded to increased macrophage and granulocyte populations, *CD74* itself is expressed by macrophages, but is not expressed on the surface of neutrophils and mast cells (481, 482). Therefore, the expression of *CD74* associating to high macrophage populations may be indicative of macrophages constituent expression of the gene. However, an alternative suggestion can be made which encompasses the

relation between *CD74* expression and both macrophages and granulocyte increased populations. *CD74*, apart from functioning as a chaperone for HLAs of the MHC class II system also functions as a receptor for macrophage migration inhibitory factor (MIF) with studies in gastrointestinal tract diseases demonstrating MIF binding to *CD74* can result in the upregulation of cytokines including IL-1, IL-6, IL-8, and TNF- α (483). The increased IL-8 secretion due to MIF binding *CD74* would relate to increased neutrophil populations with IL-8 role as a key chemoattractant for this cell type; this relationship has also been observed in breast cancer, however, this mechanism functions through *CD74* binding of *TIMP1* which can be associated to granulopoiesis and neutrophilia in mouse models (484-487).

Greater expression of *HLA-DRA* and *HLA-DRB5* both correlated to increased macrophage populations within my OAC cohort. Interestingly, this opposes findings in OSCC data where lower expression of HLA-DR positively correlated with high-density of M2 TAMs promotes malignant behaviour of OSCC cells, however, this is attributed to phenotypic switching from M1 to M2 macrophage phenotype with the publication not exploring overall macrophage density (488). Thereby, the expression of *HLA-DR* genes including *HLA-DRA* and *DRB5* may relate to the phenotype of macrophages, with the corresponding relationship observed within my OAC cohort occurring due to macrophages expressing *HLA-DR* as observed in lung cancer (489).

The expression of the final two HLA-loading complex assembly genes, *HLA-DPA1* and *HLA-DQA1* corresponded to increased macrophages in the OAC TIME; once again these genes are constitutively expressed by macrophages and exploration of these genes in the literature did not find analysis which successfully deconvolutes the role of these genes from its use as a biomarker (490, 491).

The remaining MHC class II genes associated to myeloid populations are functionally grouped into peptide generation genes, namely, *CTSL*, *CTSS* and *LGMN*. Both increased expression *CTSL* and *LGMN* corresponded to increased macrophage populations in the OAC cohort samples, the expression of *CTSL* has prior been reported to modulate the polarisation of macrophages driving them towards an M2 phenotype in breast cancer (492). However, the expression of *CTSL* may be indicative of macrophages being stimulated by IL-4 in the TIME of OAC leading to expression of *CTSL*, this suggestion is supported by studies of gastric carcinomas where IL-4 is often overexpressed (493, 494).

Lastly, increased *CTSS* expression in OAC corresponded to increased granulocyte populations, this may be due to *CTSS* playing a role in truncating chemokines *CXCL1/2/3/5* by N-terminal processing creating the active form of these chemokines which function to recruit neutrophils (495).

Overall, this analysis in combination with the literature provides several potential mechanisms by which myeloid populations may be altered by the expression of MHC class II genes, however, a number of these genes are constitutively expressed by myeloid cell subpopulations thus are not informative.

4.5.6 B lineage cell populations in the OAC TIME are altered by the expression of *CALR* and *CD1D* in OAC

Observing the B cell associations with MHC class I expression found only two genes corresponded with increased B cell populations, namely, the chaperone gene *CALR* and the alternative APM gene *CD1D*. The interaction with *CALR* has prior been reported in renal cell carcinoma where the expression of *CALR* positively correlated with B cell infiltration plus immune modulators PD-1 and LAG3, while the immune cell correlation was observed with my OAC samples, the correspondence to PD-1/LAG3 expression was not identified. *CD1D* expression is noted in B cell populations, thereby using this approach could not disentangle the probability *CD1D* plays a functional role in B cell recruitment or whether this gene just forms a marker for the presence of B cells as *CD1D* plays a major role in B cell development (496). Overall, little mechanistic understanding can be gathered from the literature on the role of these two genes, excluding the knowledge of *CD1D* marking the presence of B cells.

4.5.7 B lineage cell populations in the OAC TIME are increased by the expression of MHC class II genes

My final observations investigated the association between MHC class II gene expression and the abundance of B lineage cells, finding four HLA-loading complex assembly genes and a single peptide generation gene of the MHC class II system found a corresponding relationship. B cell populations were elevated in cases with high *HLA-DQA1*, *HLA-DRB5* and *SPPL2A*; *HLA-DQA1* have demonstrated this impact in prior studies of colorectal cancer which its expression correlated with immune infiltrate in right side colorectal cancer, but not left side (497). Greater *HLA-DRB1/5* expression increased plasma cell

abundance in OAC, however, this may relate to HLA-DR being expressed on CD19+CD24-CD38hi plasma cells in the study of immunoglobulin G4 (IgG4)-related disease, gastric cancer, and melanoma (498-500). As gastric cancer is genomically similar to OAC and shares similarities in their TIME this provides confidence to this finding within my OAC cohort (501). Increased *HLA-DRB5* expression in the OAC cohort also corresponded to increased B cell populations, HLA-DR proteins are expressed on B cells, making this finding likely due to B cell expression of *HLA-DRB5*; the expression of HLA-DR in OAC epithelium has been prior reported to associate poor survival forming an independent prognostic marker (196).

Finally, the greater expression of *SPPL2A* corresponded to increased B cell populations within my OAC cohort, this finding may be due to *SPPL2A* role in clearance of CD74 n-terminal fragments (NTF), mouse models have demonstrated lack of *SPPL2A* expression in the TIME can result in the accumulation of CD74 NTF leading to the arrest of B cell maturation in the translational stage (502-504). Therefore, in OAC the expression of *SPPL2A* may result in adequate maturation of B cells compared to cases with a lack of *SPPL2A* expression. This analysis in combination with the literature demonstrates that B lineage cell abundance in the OAC microenvironment may be altered by the expression of MHC class II genes, however, with B lineage cells being capable of expressing MHC class II genes, the relationships observed may be displaying MHC class II expression as a marker of B lineage cells. Despite, these interesting results can be observed, especially in the case of *SPPL2A* which relates to a known mechanism which impacts to function of B cells in tumours.

4.5.8 Effector cell populations are altered in cases due to APM gene expression regulators in OAC.

Lastly, for immune effector cells I observed the impact of APM gene expression regulator genes including *CITA*, *RFX5* and *CSDE1*, the impact of these genes determines the expression of MHC class I and II genes within cells, in turn impacting their ability to elicit immune responses to cancer. Firstly, I observed *CITA* which primarily regulates MHC class II expression with some evidence suggesting a secondary role in the regulation of MHC class I expression. The expression of this regulator associated to the presence of CD4+ T cells and Tregs within my OAC cohort, this is explained in the literature as resulting from greater expression of MHC class II molecules in APCs promoting CD4+ T cell

priming, with greater CD4+ T cell priming an expected result would be co-recruitment of Tregs (458, 505).

RFX5 is a component of the RFX complex which acts as form an enhanceosome for MHC class II expression, in OAC the increased expression of the *RFX5* gene appears to correspond to increased Treg abundance, interestingly, a similar report has prior been published in gastric cancers with *RFX5* expression correlating to increased TILs and positively correlated to FOXP3+ biomarker (248).

Finally, greater *CSDE1* expression in OAC corresponded to reduced CD8+ T cell abundance in the OAC microenvironment, this is attributed to *CSDE1*'s role as a negative regulator of MHC class I expression. Functionally reducing STAT1 signalling in the JAK/STAT pathway which functions downstream to promote MHC class I expression. This has currently been observed in melanoma which demonstrated *CSDE1* is capable of stabilising TPTCP, a tyrosine kinase which dephosphorylates pSTAT1 prevent its translocation to the nucleus, thus inhibiting pSTAT1 binding to the upstream promoter of the HLA-locus. Thereby, I propose the expression of *CSDE1* in OAC tumour cells functions to reduce the immunogenicity of cells, this will be further explored in chapter 5 and 6 (321).

Overall, this demonstrates that APM gene expression regulators may impact the MHC class I and class II system holistically in their respective expression resulting in multiple MHC genes being altered in cancer, thus leading to altered immunity. Further exploration of *CSDE1* may yield interesting results using immunohistochemistry in tumour tissue to discovery expression localisation and knockdown models to assess the genes impact on expression of HLAs within the MHC class I pathway (Chapter 5 & 6).

4.5.9 Limitations

Deconvolution analysis of bulk-RNA from the TCGA and OCCAMS datasets have identified several prognostically significant genes from the datamining analysis impact the immune cell subpopulations in OAC. Furthermore, deconvolution is still limited to making broad evaluations as exact cell location cannot determined from the RNA 'soup', the exact cell location can be highly significant as TILs presented in the intratumorally space of OAC possess a significant impact on survival whereas TILs located in the peritumoral space do not (282). A further limitation of this analysis is the inability to determine cellular function specifically, for example no distinction is made in CIBERSORT between CD8 T cells at a resting or activate state. The final limitation of this analysis is found in immune

cell markers, with our analysis we used APM expression groups as determined by our datamining to assess differences in immune cell subpopulations, however, a number of APM genes form markers for immune cell subpopulations. For example, *CD1A* and *CD1D* are markers for dendritic cells, thereby in our analysis a high presence of dendritic cells in sample samples with high *CD1A* and/or high *CD1D* expression is most likely due to the APM gene being a marker of the cell type, rather than the APM gene impacting the TIME immune subpopulation (365, 366).

4.5.10 Conclusions

In conclusion this analysis has suggested the expression of MHC class I/II, alternative APM and APM gene regulation regulators in OAC is important in immune cell distribution and whilst elucidating the diversity found in the immunophenotypes of OAC tumours.

However, these association must be thoroughly examined to account for MHC class II and alternative APM genes which may form markers of individual immune subset cell populations. Out of the twenty-seven candidates which possessed an association to OS discovered in chapter 3 twenty-one APM genes could be associated to altered immune cell subsets in the TIME of OAC.

In particular, effector cell abundance was associated to the expression of six MHC class I genes, one alternative APM gene, seven MHC class II genes and three APM gene expression regulators; myeloid cell abundance related to the expression of four MHC class I genes, two alternative APM genes and eight MHC class II genes, with no association to APM gene expression regulators; finally, B lineage cell populations associated to the expression of one MHC class I gene, one alternative APM gene and three MHC class II genes.

Using the literature in combination with these findings allows for further filtering of APM gene candidates, namely, HLA-A/B/C/E, ERAP2, TAP1 and HLA Class II for later investigation via immunohistochemistry to validate the findings.

These candidates must be selected using a strict criterion, including available antibody markers, the gene's impact on survival outcomes, the gene's association to altered immunity and finally the literature surrounding the role of the gene in tumoral immunity.

Chapter 5 Clinical and immunohistochemical validation of prognostic APM genes

5.1 Introduction

The results gathered from the datamining section and the deconvolution section combined suggest APM gene expression in OAC does possess a significant impact on both survival and the immune cell proportions in OAC. However, the analysis thus far conducted has used bulk-RNA of OAC tumours, which measures the average expression level for each gene across the sample, due to this the data may mask exact cell types, and importantly cannot be used to determine the special positioning on immune cells in tumours. Understanding the limitation of bulk-RNA analysis the project aims to validate the findings of the bulk-RNA seq analysis whilst exploring further data gathered by immunohistochemistry (IHC). IHC is a methodology by which protein markers are stained for using antibodies which specifically target them, then a secondary antibody may be applied which is conjugated to a horseradish peroxidase (HRP) or an immunofluorescent tag. IHC is typically used on sections of tissue cut and mounted to slides using a slide-sectioning machine to produce thin sections which may be observed under microscope.

5.1.1 Immunohistochemistry for scientific and clinical applications

As prior described IHC is a versatile method which is used for the detection of specific molecules (primarily proteins) in tissues, though fixed cells may also be assessed using this method. Primarily, two factors of information can be gathered from IHC, firstly, the presence/density of a given protein within a sample and the subcellular location of the proteins stained. In the clinic IHC has been employed as an additional tool beyond simple H&E histology to determine the presence of specific proteins which are indicative of prognosis and may guide treatment, foremost of these proteins is HER2 (human epidermal growth factor receptor 2) in breast cancer which forms a dominant driver of cell proliferation and survival. HER2+ IHC staining of biopsy samples has been used to recommend trastuzumab in patient care, resulting in 50% reduction in the risk of recurrence (506-508). Since the use of IHC to recommend anti-HER2 treatment for HER2+ breast cancer, this has expanded to include other cancers overexpressing this protein including gastric with some evidence this treatment may be applicable to OAC (509, 510). However, a recent trial opposed earlier suggestion trastuzumab may be useful in the

treatment of HER2-overexpressing OAC, finding no significance improvement in patient outcomes, yet the application of trastuzumab did not lead to increased toxicity so could provide benefit in future combining treatment studies with other HER2 targeting agents in OAC (511). An additional IHC prognostic marker related to OAC is COX2 which is a rate-limiting enzyme for the conversion of arachidonic acid to prostaglandins and performing a multitude of functions including, those relevant here, such as immune evasion, angiogenesis, and proliferation. COX2 is noted as overexpressed in OAC with therapeutic intervention using inhibitors in cases of Barrett's oesophagus shown to be protective against the development of OAC (512-515). In addition to these two proteins, systematic review of IHC analysis in OAC have also indicated the importance of protein markers CD3, CD8, EGFR, p53, LgR5, Ki67 and VEGF have clinical values. Yet a number of these markers has only been successfully applied within research failing to yield significance in clinical trials (516).

5.1.2 Immunohistochemistry investigation of OAC immunity and antigen presentation machinery

As described above staining for immune cell markers using IHC is an important scientific and prognostic tool leveraged in both research and the clinic, this is especially the case in cancer where a substantial number of publications are available in multiple cancer types and prognostic features are derived from tissue samples, this includes OAC.

Firstly, a publication by Noble et al. in OAC correlated the presence of tumour infiltrating lymphocytes with improved survival; by staining with CD3+, CD4+, CD8+ and FOXP3+ antibodies in human OAC tissue three independently prognostic factors were identified between surgery-only treatment ($p = 0.015$), completeness of resection ($p = 0.001$), increased CD8+ TILs (Tumour infiltrating lymphocytes) ($p < 0.0001$) and reduced pathological N stage ($p < 0.0001$) (362). This directly shows the presence of CD8+ T cells as a key factor in cancer specific and disease-free survival in OAC patients, this finding was further supported in the Noble et al. study finding higher levels of TILs in patients associated to favourable responses to neoadjuvant therapy (517).

Multivariate analysis further found the increased levels of CD4+ and CD8+ TILs in patient samples were associated with significant local tumour regression and lymph node downstaging. Overall, this study established the link between the presence of TILs and survival in OAC, a finding previously reported in other cancers.

In the same year, another study in OAC identified High intratumoural but not peritumoral inflammatory host response was associated to better prognosis, with the intratumoural inflammation being tied to high counts of intratumoural FoxP3+, CD3+, CD8+ TILs (210). Furthermore, the combination of TILs CD3+/CD8+/FoxP3+ was further divided into three prognostic groups, triple high (all T cell subsets high infiltrate)/mixed/triple low (all T cell subsets low infiltrate) (210).

In the progression of BE to OAC, multiplex immune cell marker staining for CD3 (T cells), CD8 (cytotoxic T cells), CD163 (macrophages) and FoxP3 (Tregs) found within the stroma, a notable significant increase of these cell types from BE to low grade dysplasia and, to high grade dysplasia (79). Additionally, significant decline in CD3+CD8+ cytotoxic T cells in OAC samples compared to BE with high grade dysplasia (79). IHC has also been used to explore reoccurrence in OAC. With the presence of CD8+ T cells, despite being an overall positive prognostic indicator, may also be a potential predictor of OAC recurrence, with a higher mean densities of CD8+ T cells being detected in patients which reoccur; however, these findings were not statistically significant within this study (518).

A recent publication used IHC in OAC to explore adaptive immune and immune checkpoint landscape of neoadjuvant treated OAC. Within this project tissue microarrays from 329 OAC cases were subjected to IHC staining for adaptive immune cell markers CD3, CD4, CD8 and CD45RO and immune checkpoint biomarkers (ICOS, IDO-1, PD-L1, PD-1); this investigation produced a number of findings, firstly, OAC tumours could be broadly divided into immune hot and immune cold tumours, depicted as expressing high level of CD45RO, ICOS, CD3, CD4, CD8, PD-1 and PD-L1 in contrast to the immune cold cases (519). Furthermore, high expression activated T cell markers CD45RO/ICOS also displayed a significant survival advantage.

The study would move forward to use multiplex IHC to explore the spatial expression of T cell activation in OAC tumours, this methodology found in immune hot cases a substantially higher count of cells co-expressed CD45RO/ICOS in the tumour stroma in comparison to the immune cold cases (519). This demonstrates the significant power multiplex IHC staining possesses in the identification of specific cells co-expressing cell markers to a localised area of a tumour.

External to OAC, studies in pancreatic ductal adenocarcinoma (PDAC), a cancer type with similarities to OAC, have characterised the immune microenvironments as immune escape, immune rich and immune exhausted by the immune cells present in IHC. In

immune escape TIMEs PDACs display highly immunosuppressive, with high counts of Tregs and M2-polarised macrophages, whilst showing low numbers of effector T cells, these patients typically have poorer prognosis compared to the other two immune phenotypes (520). Conversely, the immune rich phenotype is characterised by high counts of effector CD4+ and CD8+ T cells and M1 macrophages, there is also the frequent presence of tertiary lymphoid tissues in these cases; reduced numbers of immunosuppressive immune cell populations including Tregs and M2 macrophages are also typical in immune rich TIMEs. Lastly, a mixed TIME phenotype called immune exhausted can arise, these share similar feature to their immune rich phenotypes, however, there is a distinct lack of tertiary lymphoid tissues in immune exhausted TIMEs with T cell exhibiting cell exhaustion features such as PD1, CTLA4, TIM3, and LAG3 (520, 521). This forms a model which may be in future explored in our IHC analysis of OAC to characterise the impact of APM defects on the TIME immune phenotype.

IHC based studies have also explored antigen presentation machinery in OAC and other cancer types; one specific study used IHC methodology to identify the loss of TAP2 expression in OAC due to the expression of MIR125a-5p. the same study reaffirmed prior literature in recurrence linking the increased presence of CD8A T cells in OAC to recurrence (164). Another study explored locally advanced OAC tumours using IHC, identifying the distribution of TILs, as well as the presence of PD-L1, present in 21.2% of cases, in OAC as an immune evasion mechanism. Specifically, patients with a high count of T cell infiltration (CD3 and CD8 staining) in the tumour centre displayed a significant survival advantage of 41.4 months compared to 16.3 months in T cell poor tumours ($p = 0.025$) (522). However, T cell infiltration into the invasive zone of the tumours was not correlated with survival; a notable loss of MHC class I protein expression on tumour cells was present in 32%, relating to the downregulation of APM protein expression in OAC (522).

Overall, IHC is a powerful tool which has been used to explore the immune microenvironments and antigen processing machinery of OAC and other cancers. Using prior literature on the use of IHC we can identify the immune phenotype characteristics and survival outcomes based upon them, then move further to prioritise and associate our APM gene candidates which possess maximal prognostic value with clinical survival data and impact on the TIME.

5.1.3 Limitation of immunohistochemistry

Despite the robust nature of the IHC methodology there are several well-known pitfalls when attempting to apply this methodology in the clinic and within research projects.

Firstly, immunohistochemistry does not directly show the target of interest, instead the detected emission either a fluorescent tag or a coloured substrate is visible, therefore, extra care must be applied to the staining to reduce the possibility detection is a result of off-target binding of antibodies (often referred to as background staining) (523, 524).

Secondly, the lack of observable staining in IHC may not reflect the complete absence of the target molecule. This may form out of multiple causes including insufficient affinity of the antibody to the target protein, poor tissue pre-processing or IHC technique, masking of the epitope, poor permeabilization of tissue and lack of signal amplification (solved by amplification techniques such as Avidin–Biotin Complexes) (523). Therefore, using IHC methodology a claim of complete target molecule absence should be avoided, instead relying on more sensitive methodologies such as western blotting.

The third issue forms from the assumption a tissue should not produce a signal if no target is present, however, in practice this often IHC antibodies can bind to other targets with lower affinity presenting staining which may appear on target but is not.

Lastly, IHC staining may fall victim to poor interpretation as the positive detection of a target relies heavily on additional information including knowledge of the targets expression and subcellular location (523). This could refer to issues in ubiquitous proteins expressed in the majority of cells such as signalling molecules which may appear diffuse across a tissue rather than cell localised or the potential of being misled by known additional information such as a protein appearing to stain in the nuclei rather than its known membrane location. However, IHC alone cannot provide enough data to discard this result as non-specific as the protein may have translocated to the nucleus.

5.1.4 Hypothesis and chapter aim and objectives.

The prior datamining and deconvolution investigation of the landscape of genomic defects in APM genes and their association to altered immune population identified the expression of several APM genes was prognostically significant that possessed significantly different immune cell subpopulations in the OAC cohorts. However, these findings lack validation. To validate my findings from chapter 3 and 4 immunohistochemistry (IHC) analysis will be conducted providing confidence between the association of TIME immune cell subpopulation proportions of and APM gene expression. Additionally, I will gather information of spatial localisation of APM protein expression in OAC tissue.

Hypothesis 3: 'Expression of Antigen Presentation Machinery proteins in primary oesophageal adenocarcinoma tissue impacts survival and the T lymphocyte abundance.'

Objective 3a: Immune and APM protein staining data on a Southampton OAC TMA series, will be analysed to measure the level OAC T cell infiltrate in tumour cores, determining the variance in marker expression and to validate my observations from Aims 1 & 2.

Objective 3b: APM staining scores and immune density data will be collated to identify significant correlations between APM protein expression and T cell density.

Objective 3c: APM staining scores and immune density data will be analysed with maximally selected rank statistics optimal cut-offs to measure the impact of APM protein expression on patient outcomes (constructing Univariate and Multivariate models for Overall survival, disease-free survival, and cancer-specific survival and important clinical co-variates).

5.2 Methodology

5.2.1 TMA samples and clinical data

Tissue microarray samples for IHC analysis were secured from the Southampton General Hospital tissue bank, this consists of the OES TMA 2 tumour blocks containing 185 patients, cored in triplicate. These are labelled:

- **OES Tumour 1**
- **OES Tumour 2**
- **OES Tumour 3**

Tissue microarrays (TMAs) were constructed using triplicate, randomly selected, paraffin-embedded 1-mm tumour cores and sectioned into 4- μ m sections.

5.2.2 Staining

The IHC TMA staining was conducted by the Research histology group at the Southampton general hospital (ResearchHistology@uhs.nhs.uk) using a DAKO auto-Stainer.

Within my experimental design, I selected immune and APM stains; firstly, the Immune cell panel to cover common T cell subsets which includes CD3+ (All T cells), CD8+ (Cytotoxic T cells), CD4+ (T helper cells) and FoxP3+ (T regulatory cells). Secondly, antigen processing machinery markers HLA-A/B/C, HLA-Class II TAP1, ERAP2 and HLA-E. Optimisation of antibody dilution was optimised by dilution series surrounding manufacturer recommendation in positive control tissues (i.e., tonsil) and negative controls (without primary antibody). The details of the antibodies, dilutions and positive controls are in **Table 20**. Stains were confirmed by a histologist in the research histology group at University Hospital Southampton General.

Table 20: Table of antibodies used in IHC staining.

ANTIBODY TARGET	CLONE	SUPPLIER	CATALOG NUMBER	SPECIES/HOST ISOTYPE	POSITIVE CONTROL	CHOSEN DILUTION
HLA-Class I (ABC)	EMR8-5	Abcam	ab70328	Mouse	Human Tonsil	1:2000
HLA-Class II (HLA DR/DP/DQ)	CR3/43	Abcam	ab7856	Mouse	Human Tonsil	1:200
HLA-E	MEM-E/02	Abcam	Ab2216	Mouse	Human Tonsil	1:100
CSDE1/NRU	EPR17414	Abcam	ab201688	Rabbit	Human breast cancer, Human pancreatic cancer, Human lung cancer	1:500
TAP1	Polyclonal	Proteintech	11114-1-AP	Rabbit	Human Tonsil	1:200
CD3	F7.2.38	DAKO	M7254	Mouse	Human Tonsil	1:50
CD8	C8/144B	DAKO	M7103	Mouse	Human Tonsil	1:100
CD4	4B12	DAKO	M7310	Mouse	Human Tonsil	1:80
FOXP3	236A/E7	Abcam	ab20034	Mouse	Human Tonsil	1:100

5.2.3 Imaging

Whole slide imaging will be conducted at the Biomedical imaging unit at Southampton General Hospital using Zeiss axio scan Z1 High Throughput Slide Scanner in brightfield scanning. Scanned slide images in .CZI format will be automatically uploaded to the Underwood slide scanning database for ease of remote access and long-term storage.

5.2.4 Automated analysis

Analysis was performed in an automated fashion using QuPath (0.4.4), this aimed to provided consistent and standardised analysis results across TMA block and stains. Images were converted from .CZI format to Big TIFF format with pyramids in Zeiss Zen (blue edition), these Big Tiffs were loaded in QuPath (0.4.4) using bio formats and labelled as H-DAB images. Standardised analysis achieved by recording my workflow in QuPath using the “Create workflow” command to produce groovy scripts which were run for each TMA image analysis. The scripts broadly covered three functions broken down below. Firstly, the TMA was de-arrayed to identify each core in a grid array, each core was visually assessed to discount cores which exhibit core loss or histological structures not consistent with the OAC histological structure. Following this, a tissue detection script

(available at <https://github.com/wp1g19/OAC-Thesis-Antigen-processing-machinery-and-the-immune-microenvironment>) was processed to select only areas with tissue within a core, discounting areas with no tissue.

Next, the stain vectors were estimated using the estimate stain vectors function in QuPath software for each image before running a stain quantification script, stain quantification were conducted via two methods placed in consistent scripts. Firstly, a percentage positive score was obtained by using the positive pixel count function (deprecated) which outputs the percentage of positive DAB-stained pixels. Secondly, a H-score was generated using scripts adapted from Ram, et al. 2021 study of H-scoring methodology to produce a H-score value of 0-300 (525). H-scoring functions by labelling cells as high (H), medium (M), low (L) and negatively (N) stained cells, then places them in an equation which produces a ratio of the weighted sum of the number of positive cells to the total number of detected cells.

$$H - score = ((0 \times N) + (1 \times L) + (2 \times M) + (3 \times H))$$

Equation 1: H scoring equation for scoring diffuse IHC stains.

Immune cell density was calculated for CD3+, CD4+, CD8+ and FOXP3+ cells with the fast cell counts function in QuPath (0.4.4) using a script (available at <https://github.com/wp1g19/OAC-Thesis-Antigen-processing-machinery-and-the-immune-microenvironment>) for each stain to produce a cell count per mm². Results were compared to staining intensity and coverage determinations by an expert histologist in the research histology group at university hospital Southampton General Hospital.

5.2.4.1 Super-pixel segmentation and object classification

Lastly, super-pixel segmentation analysis was conducted to firstly segment out structures using the DoG (Difference of Gaussians) super-pixel segmentation function in a script (available at <https://github.com/wp1g19/OAC-Thesis-Antigen-processing-machinery-and-the-immune-microenvironment>). This functions by calculating the difference between two smoothed versions of the same image then applies two Gaussian kernels with differing standard deviations, by subtracting of these kernels and the original input image, edges can be detected to produce segments (526).

Following segmentation an object classifier was trained on a subsection of CSDE1-stained TMA cores to identify tumour, stroma and immune populations based on histopathologist report that CSDE1 stained immune and cancer populations with greater intensity than stroma cell populations, the object classifier was trained to distinguish between tumour, immune and stromal cell populations using cell size, shape, and intensity of CSDE1 stain. Then the object classifier was applied to the entirety of CSDE1 stained TMA images.

5.2.5 Statistical analysis

TMA stain quantified scores and immune cell counts (Percentage positivity and H-scores) were z-scored scaled prior to passing data to pheatmap() function (pheatmap package) in R version 4.0.2, clustered using Euclidean distance with ward.d2 linkage (411). Optimal number of clusters was determined using the clValid() R package using hierarchical and kmeans tests (527). To assess the most accurate method of scoring diffuse stains between percentage positivity and H-scoring I performed a Bland-Altman analysis in GraphPad prism 9 which represents which scoring system is over/under predicting the value of protein expression. Correlation analysis was conducted to assess the correlation of immune cell abundance and APM stain scores using Pearson's correlation analysis with the RcmdrMisc() R package rcorr() function, which was visualised in ggplot() correlation heatmaps (528, 529). The prior deconvolution analysis was validated using Pearson correlation analysis with p-value adjustment (using the RcmdrMisc() R package rcorr() function) for CD4+, CD8+ and FOXP3+ T cells in TMA samples which overlapped with OCCAMS data, then visualised in ggplot() generated correlation heatmaps (528, 529). Finally, survival analysis was processed for overall survival using the workflow presented in chapter 3 with CoxPH univariate and multivariate OS, CSS and DFS with the finalfit() package in R (209). Multivariate model included immune and APM stain scores with age, sex and pTNM staging; this model was selected due to available clinical data and using prior IHC publications.

5.2.6 Power and sample size calculations

I used power and sample size calculator (Available at:

<https://biostat.app.vumc.org/wiki/Main/PowerSampleSize>) to assess the power of discovery in the TMA cohort using a prior OAC staining from the Noble et al. paper comparing the high/low survival expression groups between the studies (517, 530).

5.3 Results

5.3.1 TMA quality control and stain scoring optimisation.

Firstly, I explored the sample size available from the stain TMA images to determine the quality of the TMAs being affected by core loss from entirely missing cores and cores which did not contain OAC histology as described in the methodology. Performing this analysis revealed 175/185 patient samples possessed at least one recorded TMA stain measurement. However, exploring the individual stains, the HLA-ABC+ stain possessed the greatest sample size of 158/185 patients and the CSDE1 stain possess the lowest sample size of 114/185 patients (See Figure 40). The sample size which successfully measured all IHC stains (i.e., no missing data) was only 69/185 samples, with the sample size of patients with all immune stains (CD3+, CD4+, CD8+ and FOXP3+) reaching 126/185 samples and patients recording all APM stains reaching only 88/185. Only 174/185 of patient samples possessed complete clinical data (survival status, follow-up time, pTNM staging, age and gender) (See Figure 40). Using this analysis, I determined further analysis should avoid omitting all samples with missing staining to maintain power of discovery of 20% incidences.

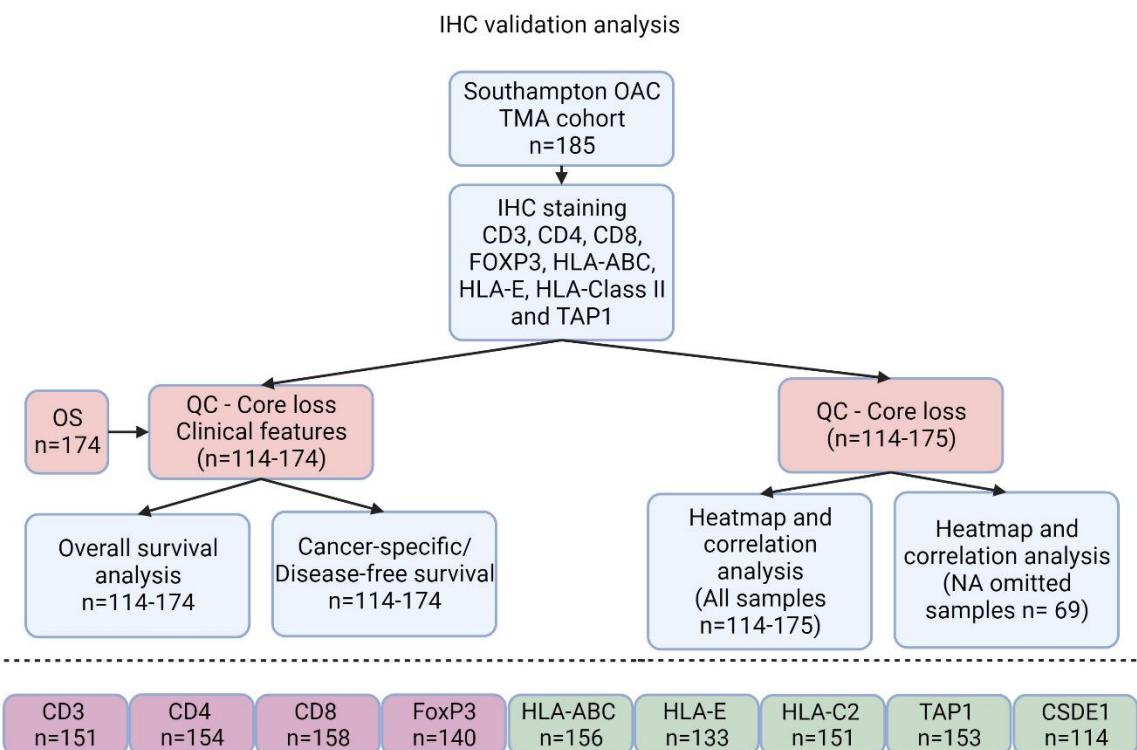


Figure 40 Consort diagram of TMA sample size available post IHC staining, including quality control, clinical data and individual stains available.

After determining the sample size due to core loss, I developed an automated detection and scoring of APM stains for my TMA images. Firstly, I explored the staining present for HLA-ABC+, HLA-E+ and TAP1+ which all appeared diffuse and with some HLA-Class II+ positive stains appearing diffuse (**See Figure 41**). Using this information and the current literature surrounding image analysis, I determined percentage positive and H scoring as the most effective manner to measure APM stains and proceeded to optimise the H scoring by visualising the staining intensity calls per cell in each APM stain and confirming these results with a histologist (**See Figure 41**) (525).

H-scoring was determined to be the most effective way to quantify IHC staining using Bland-Altman analysis of percentage positivity and H scores for HLA-A to determine whether percentage positivity under/overestimates staining quantification, finding that percentage positivity presents with a bias to underestimating HLA-A stains opposed to H scoring, though percentage positive and H scores are mostly agreement (bias = -2.38, **Figure 42**).

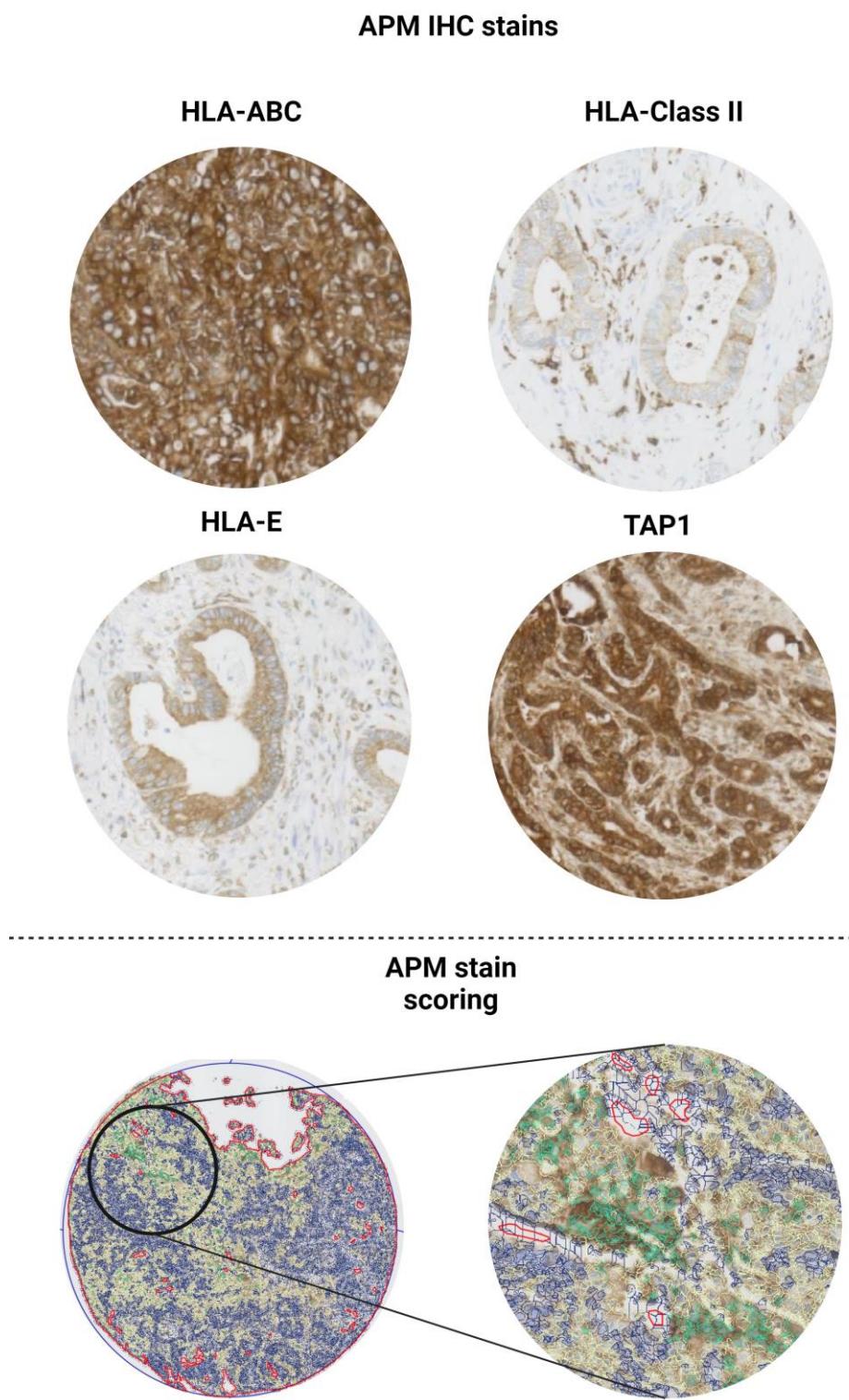


Figure 41 IHC APM staining and H scoring. Top: IHC staining of APM proteins in OAC TMA (HLA-ABC+, HLA-Class II+, HLA-E+ and TAP1+). Bottom: APM stain scoring using H-scoring (Blue: Negative, Cream: Low stain, Green: Medium stain, Purple: High stain).

Bland-Altman of percentage positivity and H scoring for HLA-A stains

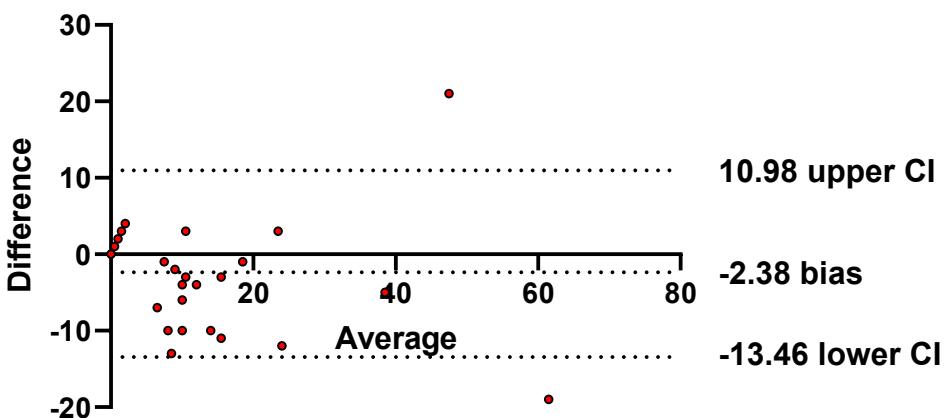
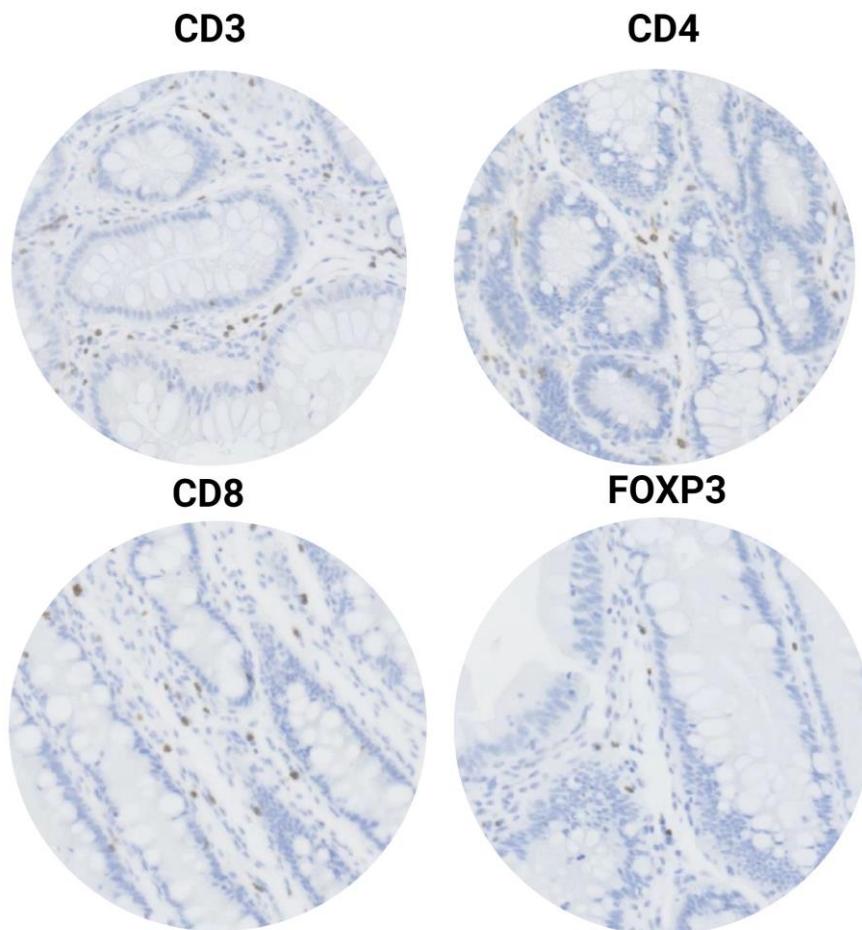


Figure 42 Bland-Altman plot comparing percentage positivity to H scores.
A negative bias here represents underestimation of HLA-A stain quantification.

Additionally, immune stains including CD3+, CD4+, CD8+ and FOXP3+ were analysed to determine in staining appropriately matched the known staining pattern confirming the appropriate cell membrane staining for CD3+, CD4+ and CD8+ markers and nuclei staining of FOXP3+ cells. The density of these cells was compared to a histologists report finding concordance, beyond this automated detection of these markers was conducted to calculate the density of positively stained cells (See Figure 43).

Lastly, I conducted DoG (Difference of Gaussian) super-pixel segmentation on OAC TMA stained with CSDE1 which identified OAC tumour core structure, this allowed for an object classifier to be trained based on knowledge CSDE1 stained tumour cells and immune cells with greater intensity than stromal populations (confirmed by a histologist, See Figure 44).

Immune IHC stains



Automated Immune cell density detection/scoring

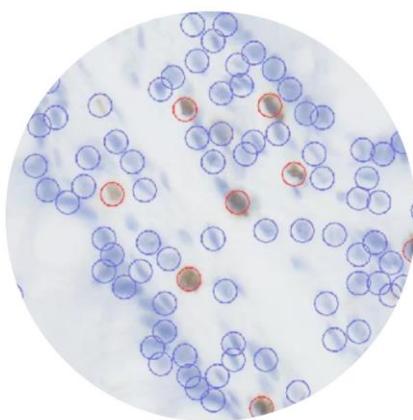


Figure 43 IHC immune cell staining and detection. Top: IHC staining of immune cells in OAC TMAs (CD3+, CD4+, CD8+ and FOXP3+). Bottom: automated immune cell detection (CD3+ Stain; Blue: Negative, Red: immune cell).

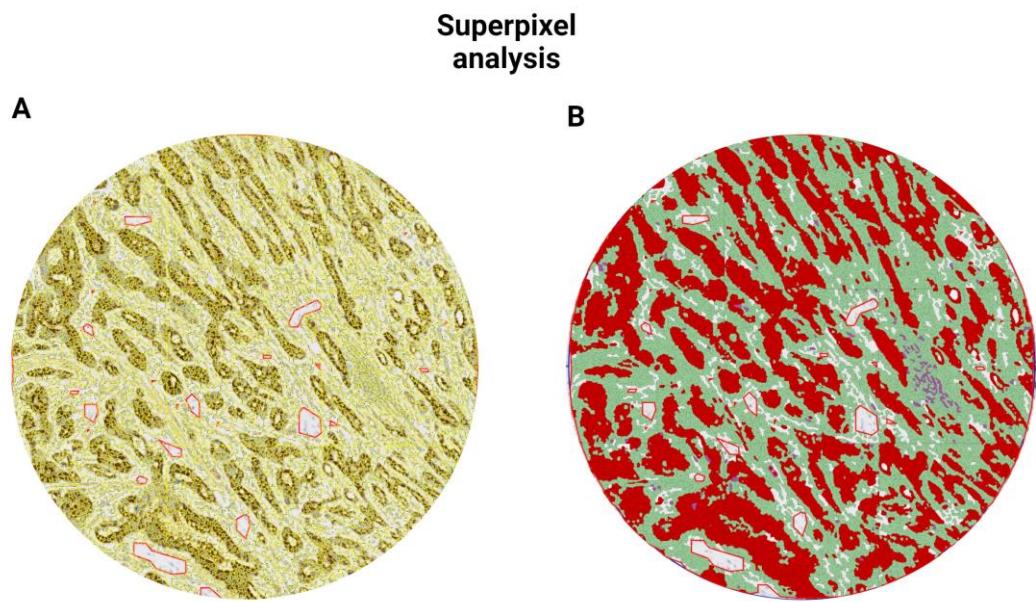


Figure 44 Super-pixel structure prediction. A: super-pixel DoG (Difference of Gaussian) segmentation of histological OAC tumour structures. B: Classification of OAC structures determined by neural networks (Red: Tumour, Green: Stroma, Purple: Immune cells).

5.3.2 Exploratory data analysis reveals high heterogeneity in APM protein expression in OAC.

After quality control the TMAs and confirming the accuracy of staining and stain scoring, I moved to assess the heterogeneity of APM protein expression detected via IHC. I found five distinct clusters of APM protein expression via percentage stain coverage, the first cluster had a lack of TAP1+ staining with low HLA-E and HLA-ABC staining (number of patients? = 16/88) with low T cell density across all T cell subsets, a second cluster was the largest cluster describing a high HLA-Class II+ cell population with greater HLA-ABC positivity than cluster 1 (12/88) and contained greater T cell density than cluster 1; cluster 3 was the largest cluster mostly driven by high HLA-ABC and TAP1 protein expression (31/88), this cluster also appears to present two distinct subclusters, a HLA-ABC positive and negative subclusters of which the HLA-ABC positive subcluster containing greater CD3+ and CD8+ T cell density; cluster 4 expressed high HLA-E protein compared to all other clusters (9/88) and possesses the highest CD3+ T cell density (**See Figure 45**). Finally, cluster 5 contained high CSDE1 expression with a subcluster lacking HLA-ABC percentage score (20/88), two distinct subclusters also appear in the cluster and HLA-ABC high and low subcluster, with the HLA-ABC high subcluster possessing greater CD3+ T cell density (**See Figure 45**).

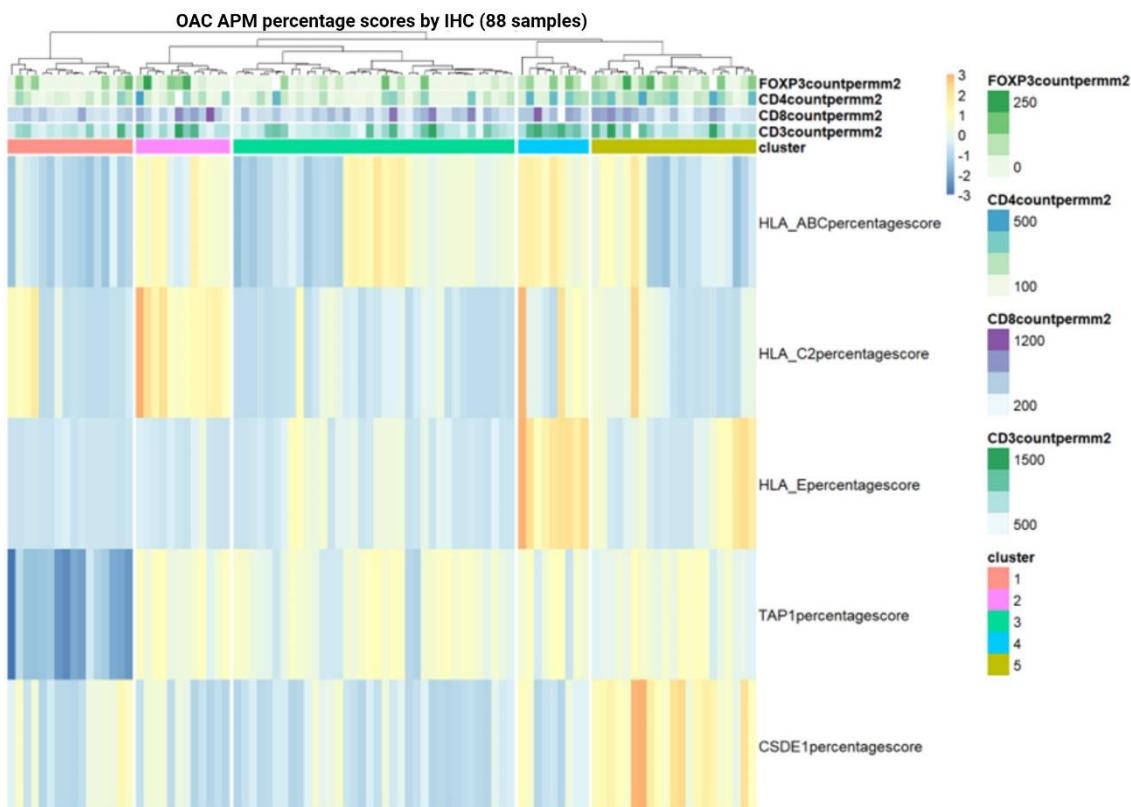


Figure 45 Z-scored heatmap of APM percentage positive staining (n = 88), clustered using Euclidean distance with Ward.D2 linkage, five distinct clusters of APM expression among OAC patients.

Interestingly, despite five clusters also being derived from H scoring of APM proteins in the OAC TMA cohort, these clusters do not match the clusters derived from percentage scores. The first cluster was the smallest and represented high HLA-ABC, HLA-E, HLA-Class II and TAP1 (3/88) and possesses the highest CD3/8+ T cell density among the clusters (See Figure 46). The second cluster was driven by HLA-ABC and TAP1 protein expression with some cases possessing high HLA-Class II (8/88) with a large proportion of patient samples also containing high CD3/4/FOXP3+ T cell density (See Figure 46). A third cluster described high CSDE1 with lesser HLA-ABC/E/Class II expression (7/88) appearing to possess lower T cell density than cluster 1/2 (See Figure 46). Interestingly, clustering analysis did identify a small cluster with high CSDE1 protein expression (H score) did possess both low HLA-ABC protein expression and low T cell infiltrate (See Appendix A). Cluster 4 is the second largest cluster, possessing low TAP1 protein expression compared to the other clusters (27/88), interestingly this cluster appears with the lowest T cell density across all subsets within the clusters. Lastly, cluster 5 had the greatest sample size and presents with high heterogeneity of APM protein expression matched with high

heterogeneity of T cell infiltration, however, notably a subset forms with high HLA-ABC and CSDE1 expression (43/88) (**See Figure 46**).

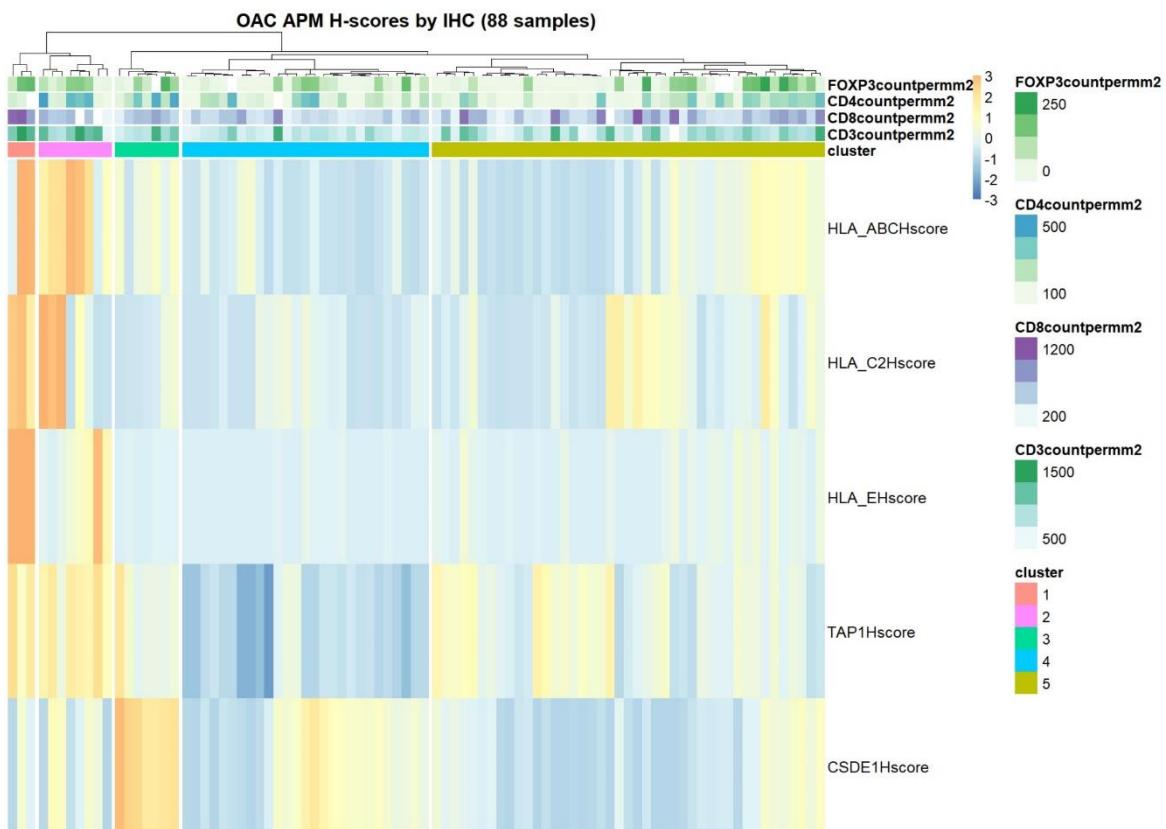


Figure 46 Z-scored heatmap of APM H scores (n = 88), clustered using Euclidean distance with Ward.D2 linkage, five distinct clusters of APM expression among OAC patients.

5.3.3 Exploratory data analysis reveals high heterogeneity in immune cell densities in OAC.

Observing the immune populations within the TMA cohort identified four distinct clusters of T cell immunophenotypes in OAC. The first cluster was the largest (64/126) presenting as an immune desert subset of patients lacking T cell infiltration, cluster 2 as the second largest cluster (31/126) presents with high FOXP3+ Tregs, lacking in CD8+ T cells suggesting this cluster represents as an immune suppressed group of tumours (**See Figure 47**). The third (13/126) and fourth clusters (22/126) are immune hot clusters with high CD3+ and CD8+ T cell infiltrate, though only cluster 4 possessed high CD4+ T cell infiltrate (**See Figure 47**).

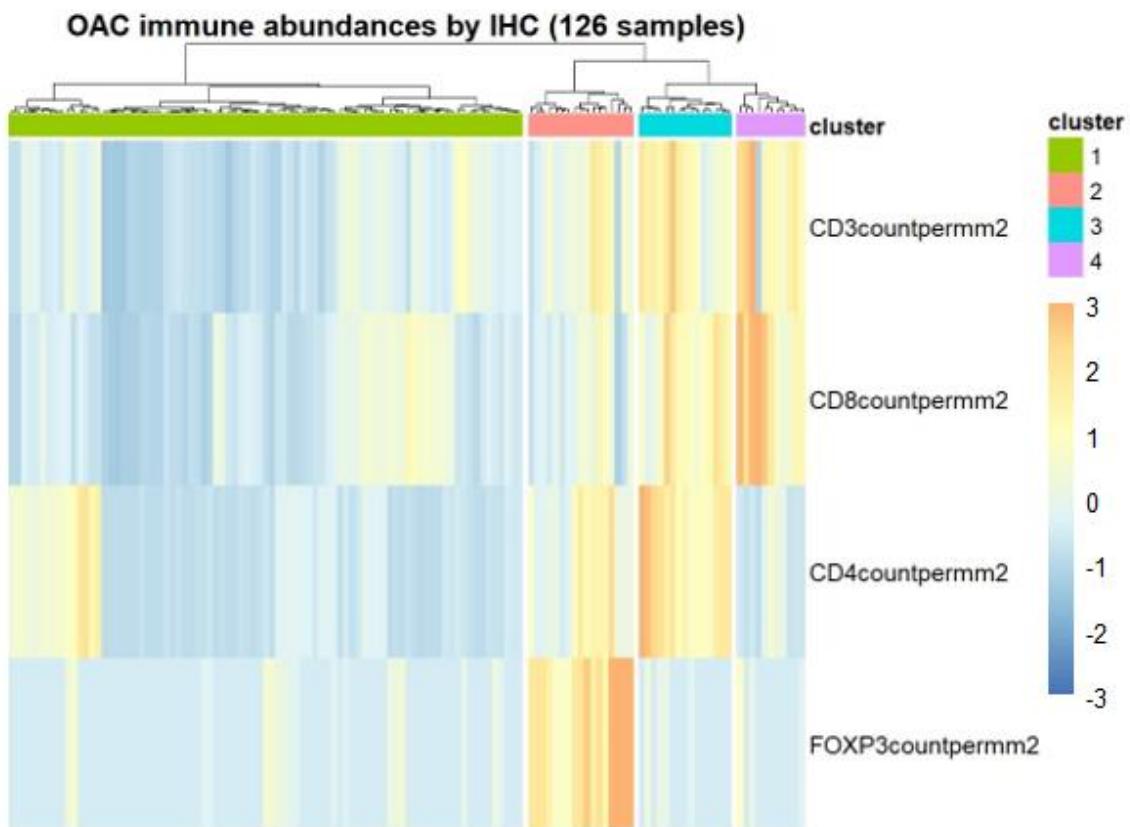


Figure 47 Z-scored heatmap of immune cell density (n = 126), clustered using Euclidean distance with Ward.D2 linkage, five distinct clusters of immune populations.

5.3.4 Correlation analysis identifies relationship between APM protein expression and immune composition in OAC.

Following the identification of APM protein expression and T cell immunophenotype clusters I proceeded to analyse the correlation between immune cell densities and APM protein expression. Using this approach, I found CD3+ T cell density positively correlated with TAP1, HLA-E, HLA-Class II, and HLA-ABC H scores (Respectively: $R = 0.28-0.38$; $p < 0.05$) and percentage positivity (Respectively: $R = 0.20-0.36$; $p < 0.05$; **see Figure 48**). CD8+ T cell density found positive correlation with TAP1, HLA-E, HLA-Class II, and HLA-ABC H scores (Respectively: $R = 0.20-0.36$; $p < 0.05$; **see Figure 48**) plus HLA-Class II and HLA-ABC percentage positivity (Respectively: $R = 0.37, 0.38$; $p < 0.05$; **see Figure 48**). The density of CD4+ T cell positively correlated to CSDE1 percentage positivity and H score (Respectively: 0.44, 0.32; $p < 0.05$; **see Figure 48**), HLA-E percentage ($R = 0.25$; $p < 0.05$; **see Figure 48**) and HLA-ABC H score ($R = 0.35$; $p < 0.05$; **see Figure 48**). Finally, Treg density (FOXP3+) positively correlated to CSDE1 percentage positivity and H score (Respectively: 0.42, 0.37; $p < 0.05$; **see Figure 48**), HLA-Class II percentage positivity and H scores (Respectively: 0.3, 0.25; $p < 0.05$; **see Figure 48**) and HLA-ABC H score ($R = 0.36$; $p < 0.05$; **see Figure 48**).

Further investigation of correlations among my APM protein expression H scores found correlation, firstly, between HLA-ABC percentage positivity scores and all other APM markers (R values: CSDE1 H score = -0.2, TAP1% = 0.49, TAP1 H score = 0.43, HLA-E % staining = 0.32, HLA-E H score = 0.37, HLA-Class II% = 0.43, HLA-Class II H score = 0.44; $p < 0.05$; **see Figure 48**). HLA-Class II percentage positivity positively correlated to TAP1 H scores ($R = 0.22$; $p < 0.05$; **see Figure 48**) HLA-E percentage positivity and H score (Respectively: $R = 0.26, 0.27$; $p < 0.05$; **see Figure 48**). HLA-E positive coverage correlated to TAP1, HLA-ABC and HLA-Class II percentage positivity (Respectively: $R = 0.18, 0.32, 0.26$; $p < 0.05$; **see Figure 48**) and TAP1, HLA-ABC, HLA-Class II H scores (Respectively: $R = 0.18-0.34$; $p < 0.05$; **see Figure 48**). TAP1 percentage positivity positively correlated to HLA-ABC and HLA-E percentage positivity (Respectively: $R = 0.43, 0.18$; $p < 0.05$; **see Figure 48**), as well as HLA-ABC, HLA-Class II, and HLA-E H scores (Respectively: $R = 0.21-0.27$; $p < 0.05$; **see Figure 48**).

Following the percentage positivity analysis, I explored the correlation between APM H scores and other APM markers. Firstly, HLA-ABC H scores positively correlated with 6

measures of APM staining scores including CSDE1, TAP1, HLA-E and HLA-Class II percentage positivity (Respectively: $R = 0.23-0.47$; $p < 0.05$; **see Figure 48**) and TAP1, HLA-E and HLA-Class II H scores (Respectively: $R = 0.41-0.54$; $p < 0.05$; **see Figure 48**). HLA-E H scores correlated to percentage positivity scores from TAP1, HLA-ABC and HLA-Class II (Respectively: $R = 0.18-0.32$; $p < 0.05$; **see Figure 48**) and H scores from TAP1, HLA-ABC, HLA-Class II (Respectively: $R = 0.36-0.44$; $p < 0.05$; **see Figure 48**). HLA-Class II H scores positively correlated to percentage positivity of HLA-ABC, TAP1 and HLA-E (Respectively: $R = 0.21-0.44$; $p < 0.05$; **see Figure 48**) and H scores of HLA-ABC, TAP1 and HLA-E (Respectively: $R = 0.34-0.54$; $p < 0.05$; **see Figure 48**). Lastly, TAP1 H score found positive correlation with HLA-ABC, HLA-Class II, and HLA-E percentage positivity (Respectively: $R = 0.18-0.49$; $p < 0.05$; **see Figure 48**) and H scores (Respectively: $R = 0.34-0.44$; $p < 0.05$; **see Figure 48**).

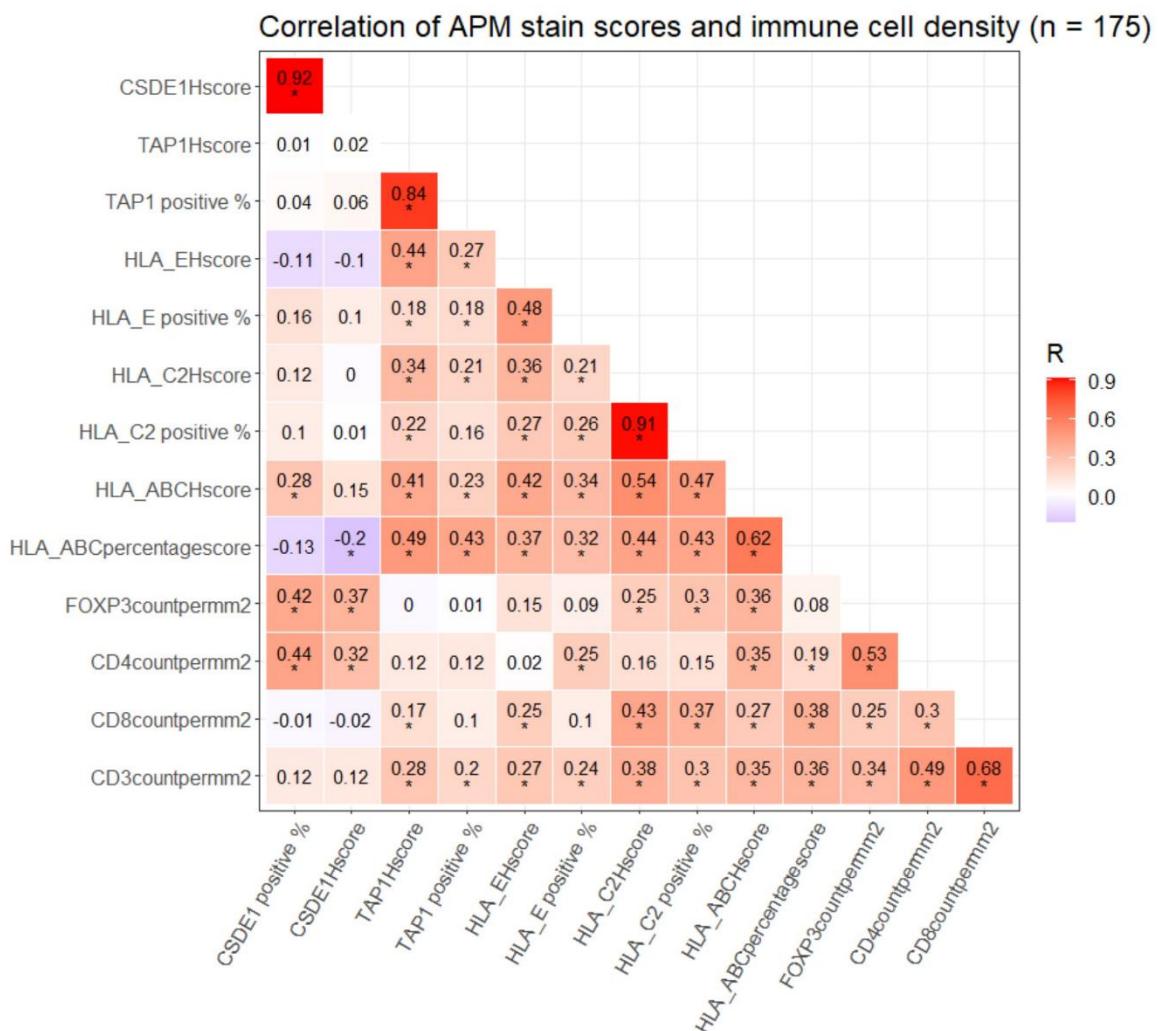


Figure 48 Correlation heatmap of APM staining scores and immune density. p values determined by Pearson's correlation test, $* < 0.05$.

5.3.5 Quantile APM protein expression analysis demonstrates altered immune composition in OAC.

Beyond correlation analysis I sought to explore the association of APM protein expression clusters identified in earlier analysis (**See Figure 45 & 46**) and T cell density. Due to the low sample size number of the APM high clusters, these clusters were combined to reflect APM high vs low protein expression in the TMA cohort. Using this approach, I identified high APM protein expression using percentage positivity as a measure possessed greater mean CD3+ T cell density within my OAC TMA cohort (High: 863.5 vs Low: 595.1, $p < 0.0001$, **See Figure 49**) and greater mean CD8+ T cell density (APM Mean, High: 524.8 vs Low: 362.5, $p = 0.015$, **See Figure 49**). This result was reflected in analysis of H score APM high vs low protein expression with APM high possessing greater mean CD3+ T cell density within my OAC TMA cohort (High: 1192 vs Low: 621.8, $p < 0.0001$, **See Figure 49**) and greater mean CD8+ T cell density (APM Mean, High: 622.3 vs Low: 403.5, $p = 0.0281$, **See Figure 49**).

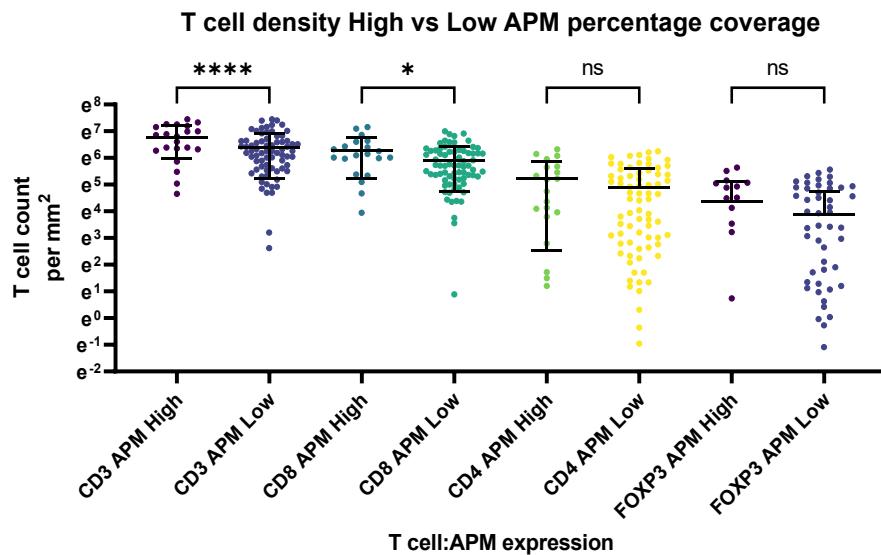
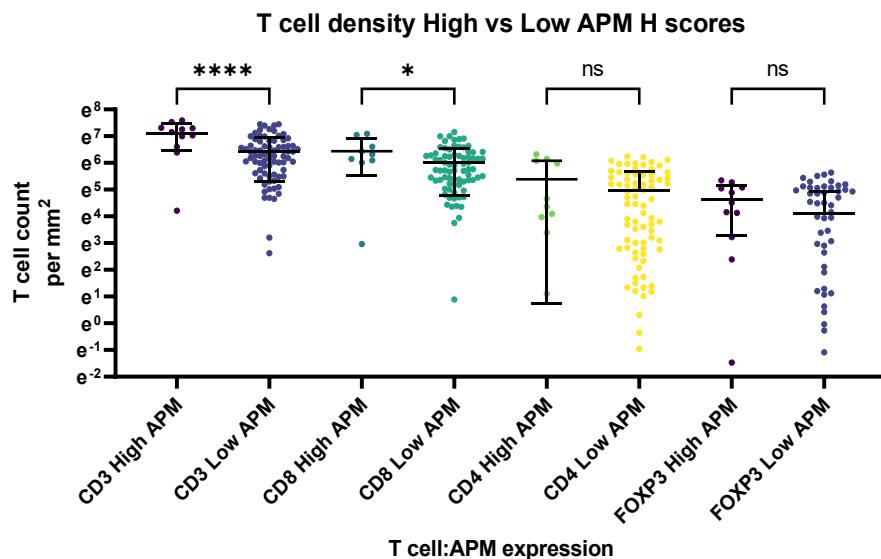
A**B**

Figure 49 Boxplots of T cell density between APM protein expression groups (High vs Low). A: T cell density compare between high/low percentage positivity APM protein expression groups. B: T cell density compare between high/low H score APM protein expression groups. Statistical test: Mann-Whitney U test, p values* <0.05 , ** <0.01 , *** <0.001 , NS = non-significant.

After observing APM expression clusters associated to altered T cell density within the OAC TMA cohort, I proceeded to explore the association between quantile IHC protein staining scores (percentage positivity and H scores) for each protein quantified and T cell density.

Starting with MHC class I HLAs, the upper quantile of percentage positivity of HLA-ABC staining possessed greater mean CD3+ and CD8+ T cell density (Respectively: Upper: 854.9 vs Lower: 414.7, $p < 0.0001$; Upper: 519.7 vs Lower: 234.5, $p < 0.0001$, **See Figure 50**). This result was replicated in the H score quantile analysis HLA-ABC staining possessed greater mean CD3+ T cell density (Upper: 906.4 vs Lower: 526.4, $p < 0.0001$, **See Figure 51**). Next the upper quantile of HLA-E percentage positivity associated with greater CD3+, CD4+ and CD8+ T cell density (Respectively: Upper: 761.9 vs Lower: 433.7, $p < 0.0001$; Upper: 220.3 vs Lower: 68.26, $p = 0.0138$; Upper: 417.1 vs Lower: 275.9, $p = 0.0211$, **See Figure 50**), these results were mirrored in the HLA-E H score upper quantile for greater CD3+, CD4+ and CD8+ T cell density (Respectively: Upper: 806.9 vs Lower: 413.8, $p < 0.0001$; Upper: 189.5 vs Lower: 61.38, $p = 0.0375$; Upper: 502.6 vs Lower: 274.8, $p < 0.001$, **See Figure 51**).

Next, I explored the association between the TAPASIN translocon protein TAP1 and T cell density finding the upper quantile TAP1 of both percentage positivity and H score protein quantification associated with greater CD3+ T cell density (Respectively: Upper: 670.6 vs Lower: 510.8, $p = 0.0160$; Upper: 717.1 vs Lower: 434.5, $p < 0.0001$, **See Figure 50 & 51**). Following the MHC class I proteins I observed the association between HLA-Class II protein expression quantiles and T cell density finding the upper quantile of percentage positivity and H score corresponded to greater CD3+, CD4+ and CD8+ T cell density (Percentage positivity Respectively: Upper: 801.3 vs Lower: 492.6, $p < 0.0001$; Upper: 516.9 vs Lower: 271.5, $p < 0.0001$; H score Respectively: Upper: 880.4 vs Lower: 438.4, $p < 0.0001$; Upper: 569.8 vs Lower: 228.5, $p < 0.0001$, **See Figure 50 & 51**). Lastly, I investigated the association between protein expression of the APM gene expression regulator CSDE1 and T cell density, observing the upper quantile of CSDE1 protein percentage positivity scores corresponded to lesser CD3+, CD4+ and CD8+ T cell density (Respectively: Upper: 534.8 vs Lower: 748.4, $p < 0.01$; Upper: 50.46 vs Lower: 261.7, $p < 0.01$; Upper: 271.8 vs Lower: 421.0, $p = 0.0414$, **See Figure 50**). However, the lower quantile of CSDE1 H scores only reflected the percentage positivity results with lesser CD3+ and CD8+ T cell density (Respectively: Upper: 763.1 vs Lower: 569.0, $p = 0.0102$; Upper: 313.3 vs Lower: 594.8, $p < 0.01$, **See Figure 51**); conversely, CD4+ T cell density was greater in the upper quantile of CSDE1 H scores (Upper: 219.9 vs Lower: 56.79, $p = 0.0308$, **See Figure 51**).

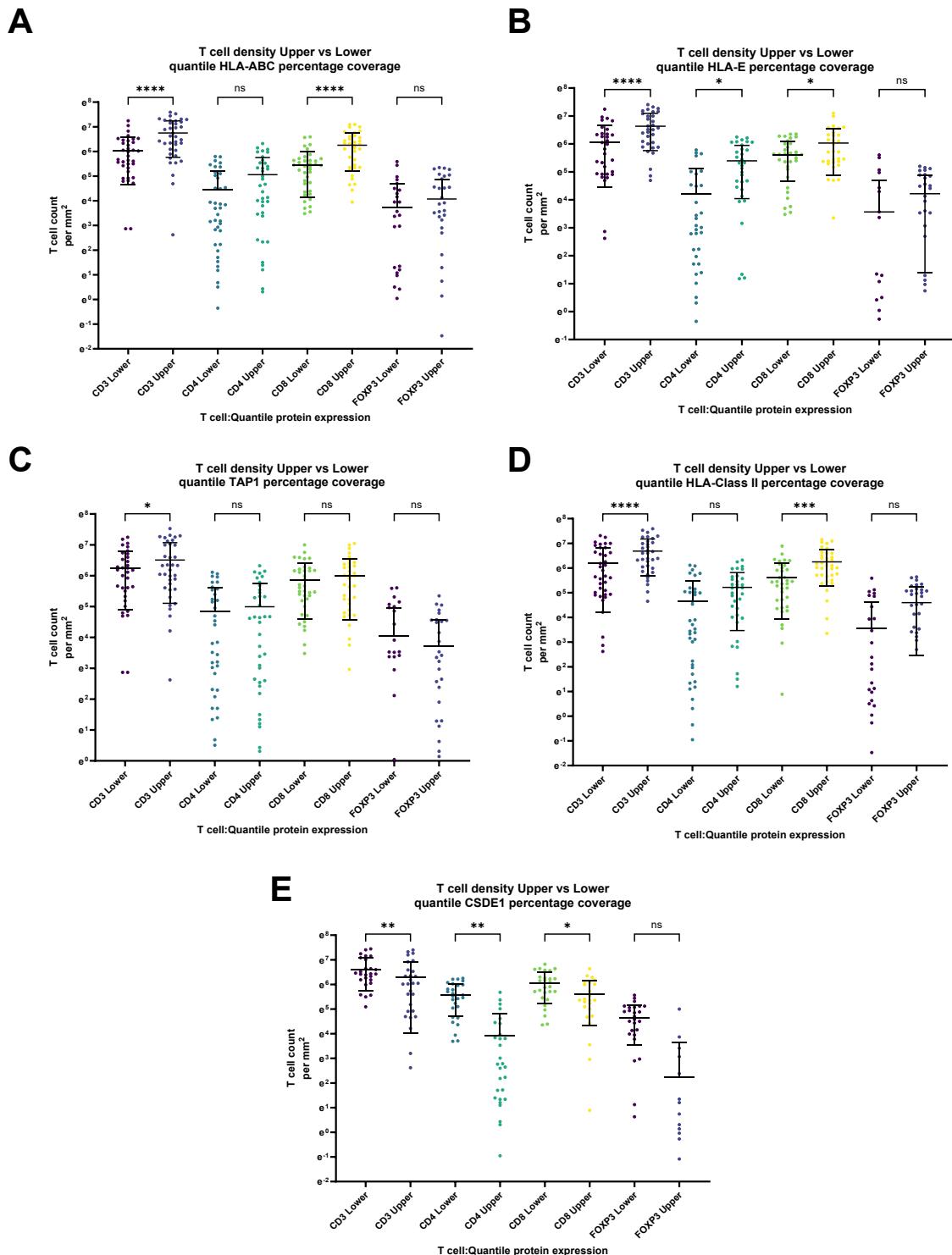


Figure 50 Boxplots of T cell density between quantile APM protein expression groups by percentage positivity (Upper vs Lower). A: Upper vs Lower quantile of HLA-ABC percentage positivity. B: Upper vs Lower quantile of HLA-E percentage positivity. C: Upper vs Lower quantile of TAP1 percentage positivity. D: Upper vs Lower quantile of HLA-Class II percentage positivity. E: Upper vs Lower quantile of CSDE1 percentage positivity. Statistical test: Mann-Whitney U test, p values* <0.05 , ** <0.01 , * <0.001 , **** <0.0001 , NS = non-significant.**

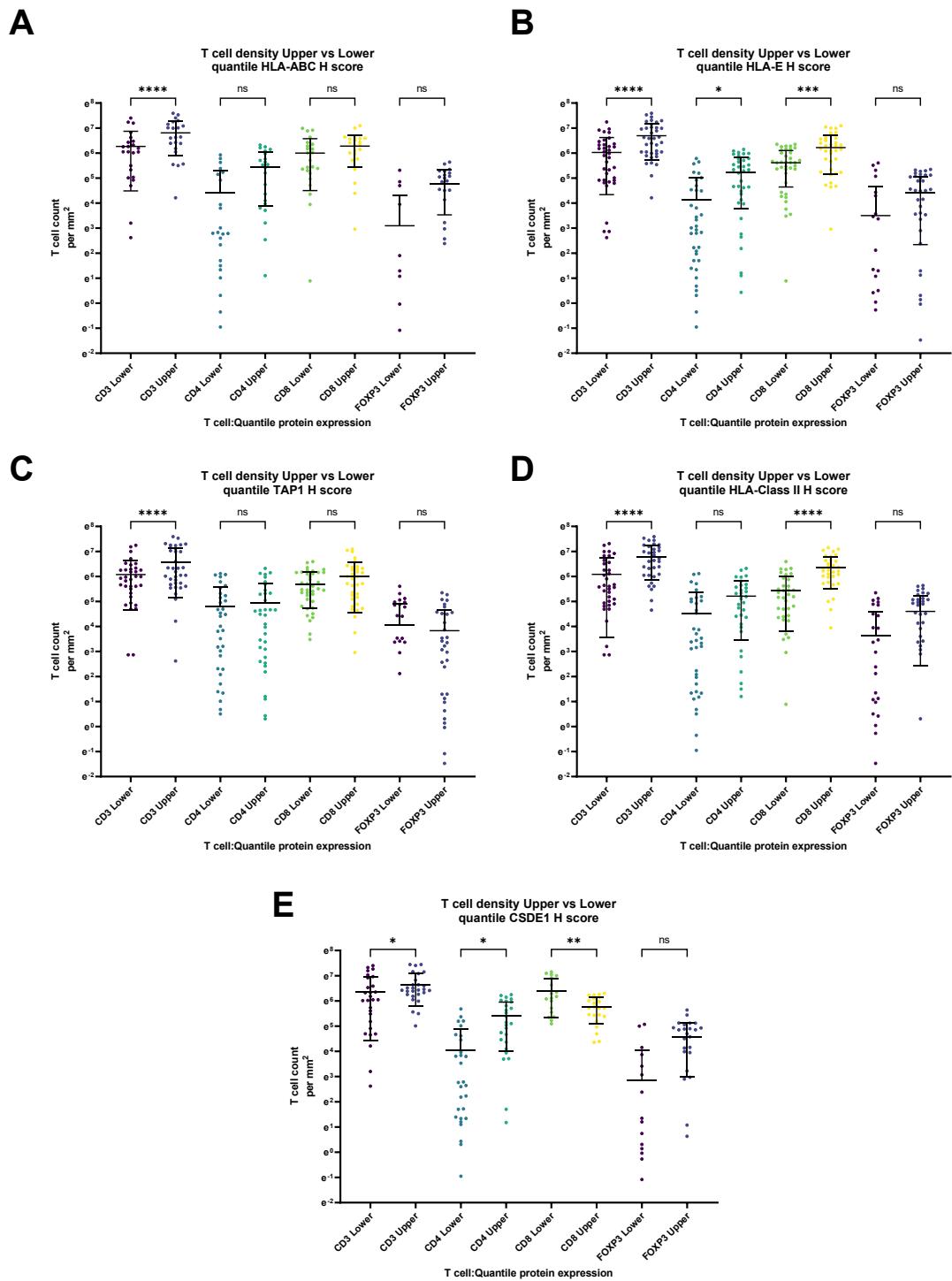


Figure 51 Boxplots of T cell density between quantile APM protein expression groups by H score (Upper vs Lower). A: Upper vs Lower quantile of HLA-ABC H-score. B: Upper vs Lower quantile of HLA-E H-score. C: Upper vs Lower quantile of TAP1 H-score. D: Upper vs Lower quantile of HLA-Class II H-score. E: Upper vs Lower quantile of CSDE1 H-score. Statistical test: Mann-Whitney U test, p values* <0.05 , ** <0.01 , *** <0.001 , **** <0.0001 , NS = non-significant.

Lastly, I sought to confirm the histologists report of CSDE1 staining tumour cells with specificity in OAC, thereby I conducted a correlation analysis between CSDE1 H scores and predicted tumour content via super pixel segmentation. Using this approach, I observed a positive correlation between CSDE1 H scores and tumour percentage in OAC ($R^2 = 0.69$, $p < 0.0001$. See Figure 52).

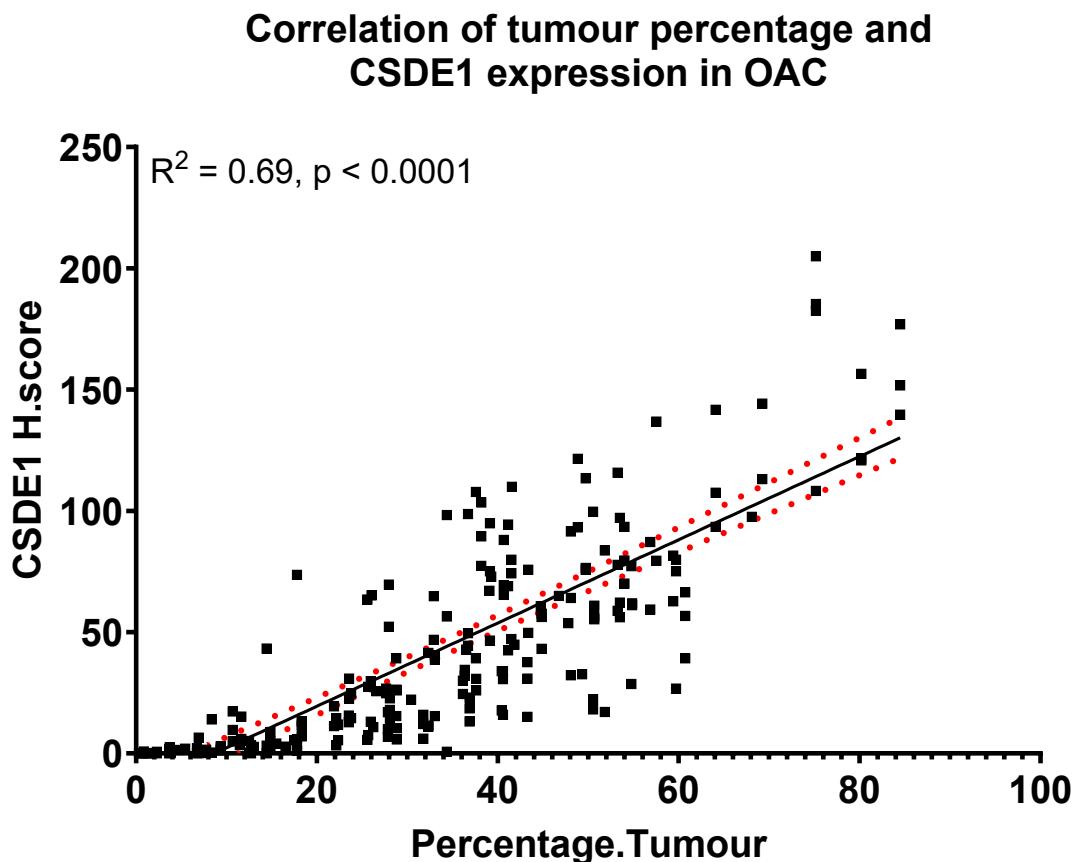


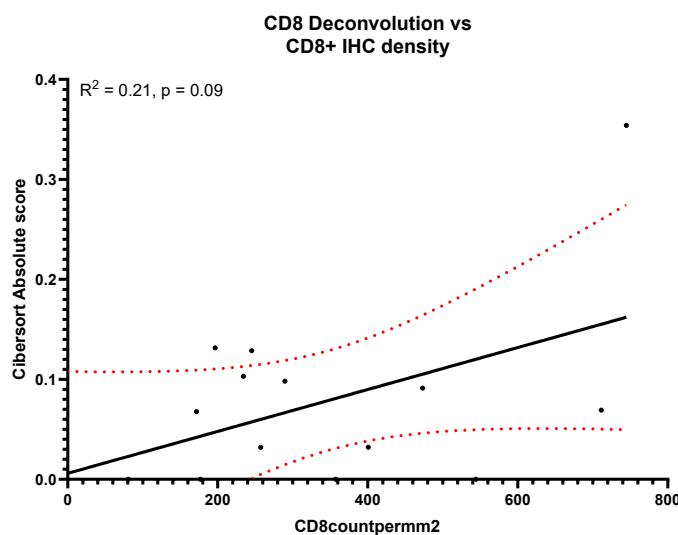
Figure 52 Scatter plot with linear regression comparing predicted tumour percentage and CSDE1 H scores in OAC TMA. Dotted red lines represent 95% confidence interval.

5.3.6 Validation of deconvolution by immunohistochemistry

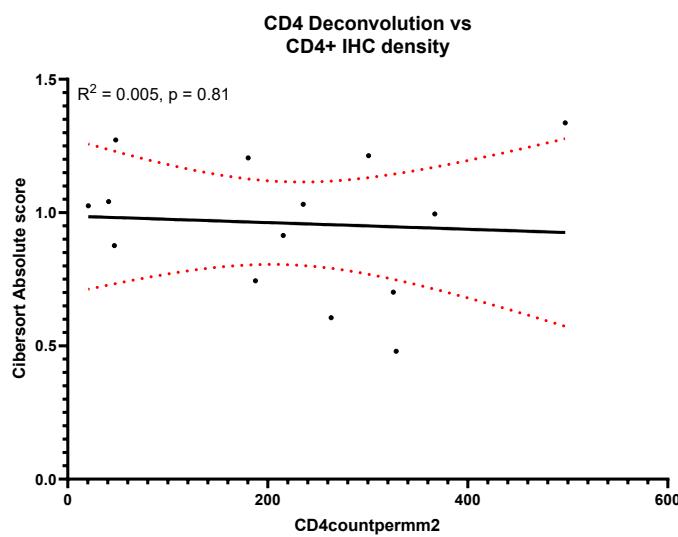
Moving forward, I proceeded to attempt to validate my deconvolution analysis presented in Chapter 4. To achieve this, I compared the T cell density of CD4+, CD8+ and FOXP3+ T cells from my IHC staining to the corresponding cell absolute abundances in my deconvolution analysis for samples which overlap between the TMA cohort and the RNA-seq cohort (CD4/8: N = 14; Tregs: N = 13). I could not establish a strong linear relationship between these variables with the closest correlation to approach significance being CD8+ T cell populations ($R^2 = 0.21$, $p = 0.09$, **See Figure 53**); CD4+ and FOXP3+ T cell correlation displayed no appreciable correlation between IHC staining and deconvolution scores (Respectively: $R^2 = 0.005$, $p = 0.81$; $R^2 = 0.027$, $p = 0.58$, **See Figure 53**). However, there is a high potential the sample size is too small to determine any significant correlation to this analysis.

To address this shortfall, I explored the using TIL percentage infiltration derived from TCGA biospecimen data (paired sample size of 14) produced from histopathologist analysis. Here I did observe a significant relationship between TILs percentage infiltration and my deconvolution of TILs, although the model possesses a low R^2 suggesting the model is weak (combine absolute values of CD4, CD8 and Tregs; ($R^2 = 0.34$, $p = 0.017$, **See Figure 54**).

A



B



C

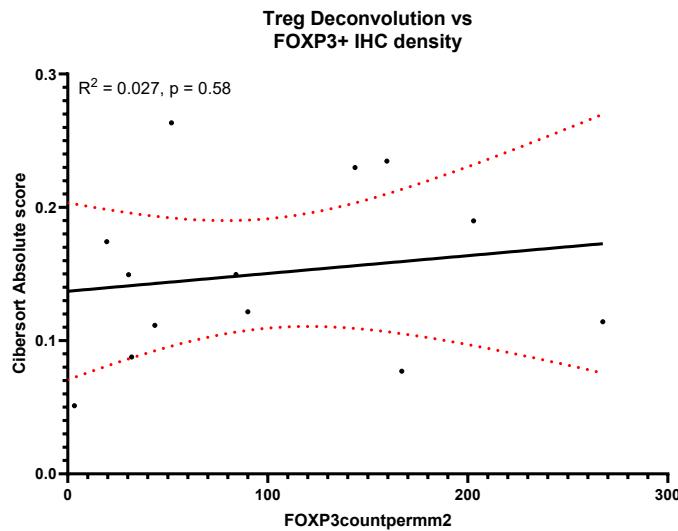


Figure 53 Scatter plot with linear regression comparing IHC T cell density with CIBERSORT Absolute score. A: IHC CD8+ T cell density vs RNA-seq CIBERSORT Absolute score. B: IHC CD4+ T cell density vs RNA-seq CIBERSORT Absolute score. C: IHC FOXP3+ T cell density vs RNA-seq CIBERSORT Absolute score. Dotted red lines represent 95% confidence interval.

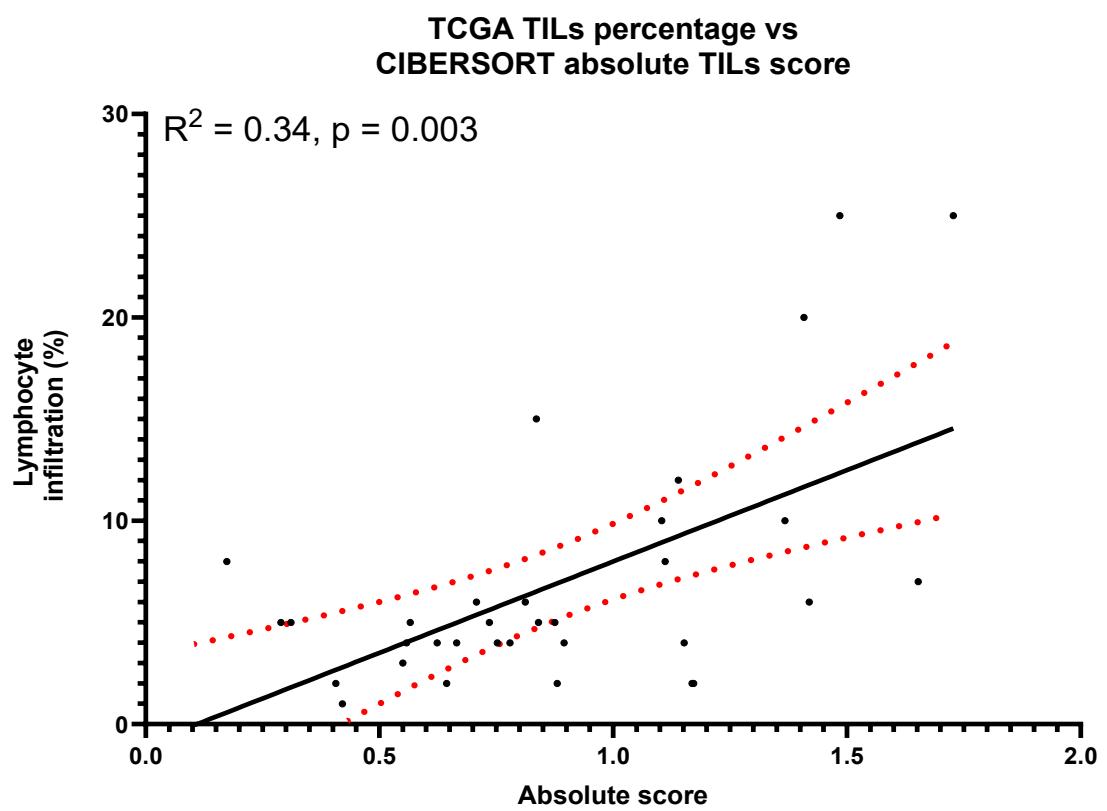


Figure 54 Scatter plot with linear regression comparing TCGA-ESCA OAC lymphocyte percentage with TILs CIBERSORT Absolute score. Dotted red lines represent 95% confidence interval.

5.3.7 Univariate survival analysis demonstrates significant impact of APM protein expression and immune composition on patient outcomes.

Beyond identifying a link between APM protein expression and the density of T cell subsets, I sought to assess the association between APM protein expression, T cell density and overall survival using optimal cut points. Among my assessed T cell markers, CD8+ T cell density associated to altered survival with low CD8+ T cell density displaying increased risk in OS (HR: 1.84, p = 0.04, **See Figure 55**); additionally, lower FOXP3+ T cells also corresponded to increased risk in OS (HR: 1.64, p = 0.05, **See Figure 55**). However, neither CD3 nor CD4 T cell density associated to altered OS (**See Figure 55**). Investigating APM protein expression their respective OS outcomes yielded four APM protein measures possessed altered OS; firstly, lower HLA-E H scores corresponded to increased risk of death (HR: 1.58, p = 0.05, **See Figure 55**), with HLA-E percentage positivity approaching significance (HR: 1.51, p = 0.06, **See Figure 55**). Interestingly, lower HLA-Class II protein expression using both percentage positivity and H score measure results in shorter OS (Respectively: HR: 1.65, p = 0.02; HR: 1.83, p = 0.02, **See Figure 55**). Lastly, high CSDE1 H scores corresponded to shorter OS in the OAC TMA cohort (HR: 1.69, p = 0.04, **See Figure 55**).

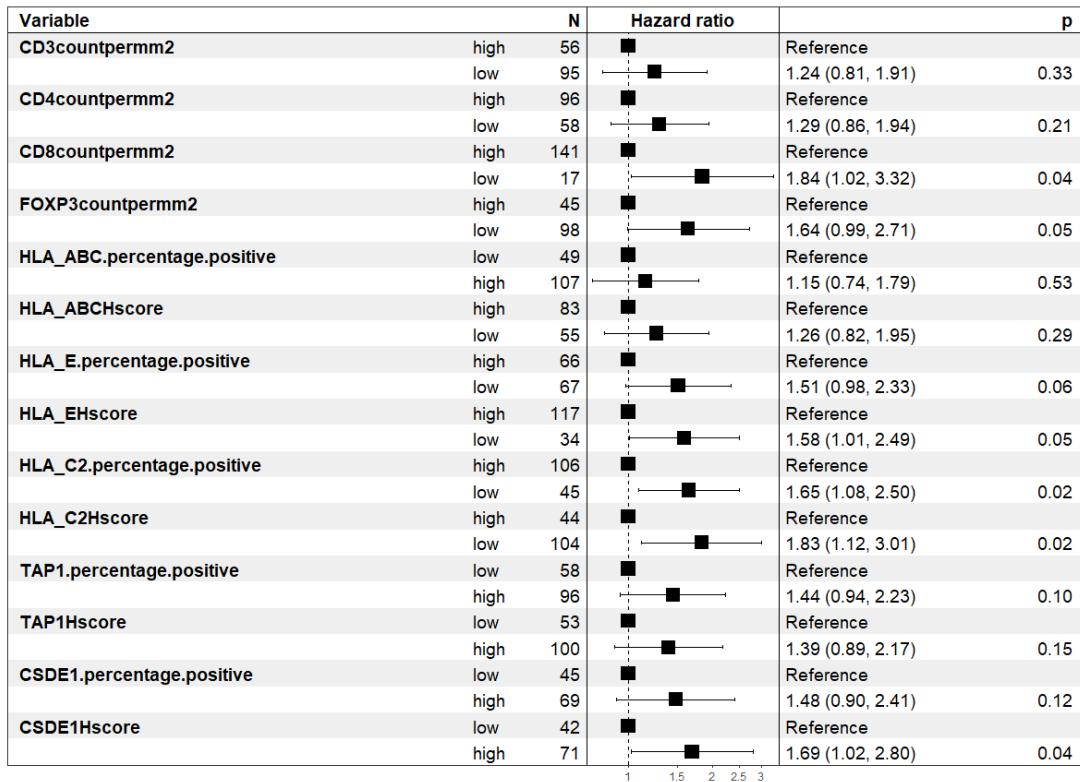


Figure 55 Forest plot of CoxPH univariate OS for TMA IHC APM protein expression and T cell density.

5.3.8 Multivariate survival analysis identifies independent prognostic APM protein expression and immune composition in OAC.

Following univariate OS analysis where I observed the APM proteins HLA-E, HLA-Class II, the APM gene expression regulator CSDE1 and T cells (CD8+ and FOXP3+) associated with altered survival, I proceeded to investigate whether these findings withstood a multivariate survival analysis using a simple clinical model. Among my assessed APM proteins, I observed only HLA-Class II protein scores produced an independent marker of survival, with lower HLA-Class II resulting in poorer OS both H scores and percentage positivity (Respectively: HR: 2.27, p = 0.022; HR: 2.24, p = 0.01, **See Table 20 & 21**).

Table 21 CoxPH multivariate OS for TMA IHC APM protein expression (H score).

VARIABLE	LEVEL	N (%)	HR (UNIVARIABLE)	HR (MULTIVARIABLE)
AGE	MEAN (SD)	70.1 (10.6)	0.97 (0.96-0.99, P=0.001)	0.99 (0.95-1.03, P=0.601)
SEX	F	19 (10.9)	-	-
	M	155 (89.1)	0.96 (0.52-1.75, P=0.885)	0.63 (0.27-1.48, P=0.288)
PT	0-1	35 (20.1)	-	-
	2-4	139 (79.9)	3.07 (1.68-5.60, P<0.001)	1.92 (0.70-5.27, P=0.207)
PN	0	76 (43.7)	-	-
	1-3	98 (56.3)	3.03 (1.99-4.60, P<0.001)	2.22 (1.11-4.43, P=0.023)
PM	0	164 (97.0)	-	-
	1	5 (3.0)	1.78 (0.65-4.85, P=0.262)	1.35 (0.14-13.21, P=0.798)
HLA-ABC H SCORE	HIGH	83 (60.1)	-	-
	LOW	55 (39.9)	1.26 (0.82-1.95, P=0.293)	1.30 (0.66-2.57, P=0.444)
HLA-Class II H SCORE	HIGH	44 (29.7)	-	-
	Low	104 (70.3)	1.83 (1.11-3.03, P=0.017)	2.27 (1.12-4.55, P=0.022)
HLA-E H SCORE	HIGH	117 (77.5)	-	-
	LOW	34 (22.5)	1.58 (1.01-2.49, P=0.047)	1.89 (0.81-4.39, P=0.140)
TAP1 H SCORE	LOW	53 (34.6)	-	-
	HIGH	100 (65.4)	1.39 (0.89-1.38, P=0.15)	1.67 (0.83-3.33, P=0.072)
CSDE1 H SCORE	LOW	42 (37.2)	-	-
	HIGH	71 (62.8)	1.69 (1.02-2.78, P=0.04)	1.28 (0.65-2.50, P=0.220)

Number in data frame = 174, Number in model = 88, Missing = 86, Number of events = 54, Concordance = 0.756 (SE = 0.037), R-squared = 0.403 (Max possible = 0.992), Likelihood ratio test = 45.390 (df = 19, p = 0.001)

Table 22 CoxPH multivariate OS for TMA IHC APM protein expression (percentage positivity).

VARIABLE	LEVEL	N (%)	HR (UNIVARIABLE)	HR (MULTIVARIABLE)
AGE	MEAN (SD)	70.1 (10.6)	0.97 (0.96-0.99, P=0.001)	0.98 (0.95-1.01, P=0.228)
SEX	F	19 (10.9)	-	-
	M	155 (89.1)	0.96 (0.52-1.75, P=0.885)	0.90 (0.35-2.32, P=0.826)
PT	0-1	35 (20.1)	-	-
	2-4	139 (79.9)	3.07 (1.68-5.60, P<0.001)	2.12 (0.73-6.16, P=0.165)
PN	0	76 (43.7)	-	-
	1-3	98 (56.3)	3.03 (1.99-4.60, P<0.001)	1.78 (0.81-3.95, P=0.153)
PM	0	164 (97.0)	-	-
	1	5 (3.0)	1.78 (0.65-4.85, P=0.262)	1.46 (0.17-12.59, P=0.729)
HLA_ABC.PERCENTAGE.POSITIVE	LOW	49 (31.4)	-	-
	HIGH	107 (68.6)	1.15 (0.74-1.79, P=0.526)	0.90 (0.45-1.78, P=0.764)
HLA_C2.PERCENTAGE.POSITIVE	HIGH	106 (70.2)	-	-
	LOW	45 (29.8)	1.65 (1.08-2.50, P=0.019)	2.24 (1.21-4.14, P=0.010)
HLA_E.PERCENTAGE.POSITIVE	HIGH	66 (49.6)	-	-
	LOW	67 (50.4)	1.51 (0.98-2.33, P=0.064)	1.27 (0.69-2.32, P=0.440)
TAP1.PERCENTAGE.POSITIVE	LOW	58 (37.7)	-	-
	HIGH	96 (62.3)	1.44 (0.94-2.23, P=0.096)	1.27 (0.61-1.27, P=0.694)
CSDE1.PERCENTAGE.POSITIVE	LOW	45 (39.5)	-	-
	HIGH	69 (60.5)	1.48 (0.90-2.41, P=0.120)	1.12 (0.61-2.08, P=0.409)

Number in data frame = 174, Number in model = 94, Missing = 80, Number of events = 57, Concordance = 0.732 (SE = 0.033), R-squared = 0.338 (Max possible = 0.992), Likelihood ratio test = 38.792 (df = 19, p = 0.005)

Moving forward I assessed whether T cell density of CD8+ and FOXP3+ T cells withstood multivariate analysis, unfortunately, neither T cell subset found significance in multivariate analysis (See Table 23).

Table 23 CoxPH multivariate OS for TMA IHC T cell density.

VARIABLE	LEVELS	N (%)	HR (UNIVARIABLE)	HR (MULTIVARIABLE)
AGE	MEAN (SD)	70.1 (10.6)	0.97 (0.96-0.99, P=0.001)	0.97 (0.95-1.00, P=0.025)
SEX	FEMALE	19 (10.9)	-	-
	MALE	155 (89.1)	0.96 (0.52-1.75, P=0.885)	0.72 (0.35-1.48, P=0.375)
PT	0-1	35 (20.1)	-	-
	2-4	139 (79.9)	3.07 (1.68-5.60, P<0.001)	1.62 (0.70-3.76, P=0.264)
PN	0	76 (43.7)	-	-
	1-3	98 (56.3)	3.03 (1.99-4.60, P<0.001)	3.78 (2.03-7.01, P<0.001)
PM	0	164 (97.0)	-	-
	1	5 (3.0)	1.78 (0.65-4.85, P=0.262)	1.92 (0.45-8.19, P=0.378)
CD3 COUNT PER MM ²	HIGH	56 (37.1)	-	-
	LOW	95 (62.9)	1.24 (0.81-1.91, P=0.327)	1.59 (0.90-2.82, P=0.110)
CD4 COUNT PER MM ²	HIGH	96 (62.3)	-	-
	LOW	58 (37.7)	1.29 (0.86-1.94, P=0.214)	1.01 (0.59-1.74, P=0.963)
CD8 COUNT PER MM ²	HIGH	141 (89.2)	-	-
	LOW	17 (10.8)	1.84 (1.02-3.32, P=0.042)	1.15 (0.56-2.34, P=0.704)
FOXP3 COUNT PER MM ²	HIGH	45 (31.5)	-	-
	LOW	98 (68.5)	1.64 (0.99-2.71, P=0.050)	1.50 (0.79-2.87, P=0.215)

Number in data frame = 174, Number in model = 130, Missing = 44, Number of events = 81, Concordance = 0.741 (SE = 0.025), R-squared = 0.361 (Max possible = 0.995), Likelihood ratio test = 58.122 (df = 18, p = 0.000)

Lastly, I assessed whether high HLA-ABC H score combined with high CD8+ T cell density would result in improved survival as this could represent intact antigen presentation with localised cytotoxic T cells capable of recognising antigens presented on these surface molecules impacts survival and by extension immunity. Here I observed the presence of high HLA-ABC, but low CD8+ T cell density resulted in poorer OS compared to high HLA-ABC and high CD8+ T cell density in multivariate analysis (HR: 2.88, p = 0.002, **See Table 24**).

Table 24 CoxPH multivariate OS for TMA IHC HLA-ABC H score:CD8 T cell density.

VARIABLE	LEVEL	N (%)	HR (UNIVARIABLE)	HR (MULTIVARIABLE)
AGE	MEAN (SD)	70.5 (10.4)	0.97 (0.95-0.99, P<0.001)	0.98 (0.96-1.00, P=0.027)
SEX	F	17 (11.8)	-	-
		127 (88.2)	1.06 (0.55-2.06, P=0.857)	0.96 (0.49-1.90, P=0.917)
PT	T0-1	30 (20.8)	-	-
		114 (79.2)	3.57 (1.83-6.95, P<0.001)	2.46 (1.16-5.20, P=0.019)
PN	0	62 (43.1)	-	-
	N1-3	82 (56.9)	3.64 (2.25-5.90, P<0.001)	2.91 (1.73-4.88, P<0.001)
PM	-	2 (1.4)	-	-
	0	137 (95.1)	0.61 (0.15-2.50, P=0.494)	2.40 (0.55-10.49, P=0.244)
	1	3 (2.1)	0.67 (0.09-4.77, P=0.690)	3.56 (0.45-28.04, P=0.228)
	X	2 (1.4)	0.57 (0.05-6.28, P=0.645)	2.85 (0.25-33.00, P=0.403)
HLA-ABC : CD8	HIGH:HIGH	26 (18.1)	-	-
	HIGH:LOW	66 (45.8)	2.15 (1.12-4.15, P=0.022)	2.88 (1.47-5.66, P=0.002)
	LOW:LOW	52 (36.1)	1.49 (0.74-2.98, P=0.264)	1.69 (0.83-3.44, P=0.147)

Number in data frame = 144, Number in model = 144, Missing = 0, Number of events = 89, Concordance = 0.736 (SE = 0.026), R-squared = 0.320 (Max possible = 0.996), Likelihood ratio test = 55.461 (df = 9, p = 0.000)

5.4 Discussion

5.4.1 APM gene expression in the OAC TMAs presents with high heterogeneity and clusters in five distinct groups.

The motivation for this investigation is to validate my findings from the RNA-seq datamining and deconvolution analysis presented in **Chapters 3 and 4** at the protein expression level in tumour tissue. This approach allows for high confidence of the APM and immune landscape of OAC, whilst affirming APM-T cell relationships prior observed. Furthermore, these findings can be assessed directly to prior literature using IHC methodology of APM genes and T cell distributions. Lastly, I pursued to confirm the presence of a negative regulatory role of CSDE1 protein on the expression of MHC class I HLAs.

Within my analysis I identified distinct clusters of APM protein expression within my OAC TMA cohort. Specifically, five clusters appear for both H scoring and percentage positivity analysis, these clusters reflected the high heterogeneity of OAC APM proteins and mostly overlap allowing them to be broadly described as 5 clusters with similar characterising features.

The first cluster is described as a TAP1 low cluster, demonstrated by the percentage positivity and H score values with relatively low T cell density, which does correspond with known literature suggesting TAP1 is downregulated in oesophageal cancer among other cancers (531). Despite this, the wider literature could not explain specific downregulation of this protein's expression in OAC but could be explained by the presence of miR-125a-5p bound to untranslated regions of TAP2 transcripts in OAC, as TAP1 is dependent on TAP2 protein for stability. Unfortunately, no antibody for miR-125a-5p is available to access this in IHC methodology although future studies could assess whether miR-125a-5p binding to TAP2 is present in OAC via immunoprecipitation analysis (532). However, an alternative explanation may be these cases are tumours with low inflammation as TAP1 expression corresponds to interferon response gene sets in gene-set enrichment analysis as this pathway acts to promote TAP1/2 expression (531).

The second cluster is described as possessing high HLA class II protein expression, which would represent that these tumours have high infiltration of professional APCs such as dendritic cell or macrophages. Interestingly, elevated dendritic cells has been prior noted

in OAC (533). However, as HLA-Class II is expressed by multiple subsets of APCs it is unknown whether these results reflect increased dendritic cells, specifically with deconvolution data presented in chapter 4 not demonstrating increased dendritic cell abundance due to HLA-Class II gene expression rather noting an increase in macrophage and granulocyte populations.

Next a cluster presented with high HLA-A/E protein (**Figure 45**). Firstly, a heterogenous cluster in percentage positivity analysis forming the largest cluster within my OAC TMA cohort, possessing relatively higher HLA-ABC percentage positivity, this parallels a publication reporting HLA-ABC positivity in 77.8% of OAC samples (355). Secondly, a H score cluster reflected high HLA-E and HLA-ABC protein expression, this cluster represented a smaller proportion of patient samples compared to prior published analysis using percentage positivity (positivity), suggesting using this measure could be overestimating which samples are truly reflecting high HLA-ABC/E protein expression (525). Interestingly the proportion of high HLA-E patients within my TMA cohort did not reflect the proportion reported in gastric cancer, which could highlight differences between APM protein expression between OAC and gastric cancer (534).

An additional cluster observed high TAP1/HLA-ABC protein expression and low CSDE1 protein expression, this could reflect these patients have intact MHC class I presentation making them a prime target for immunotherapies based in the adaptive immune system, such as checkpoint blockade (535, 536). However, to date, checkpoint blockade therapy application is not stratified by patient MHC class I HLA expression (determined by biopsy) (9).

The final cluster observed displayed high CSDE1 protein expression, but did not completely lack HLA-ABC protein expression by H scores as expected from the literature, this could reflect CSDE1 expression is specific to a subset of cells (cancer cells), with HLA-ABC expression being unaffected in stromal cell populations; an alternative explanation may be the presence of HLA-A protein expression is observed due to high interferon signalling in these tumours recovering HLA-A expression, though this cannot be determined as these TMA were not stained for IFN- γ . However, using percentage positivity indicated high CSDE1 and low HLA-ABC protein expression, this negative relationship between these proteins was expected as CSDE1 acts as a negative regulator

of MHC class I HLA expression by stabilising the tyrosine kinase TCPTP which degrades phosphorylated STAT1 preventing pSTAT1 translocation to the nucleus, which is key for promoting MHC class I HLA expression (461). Suggesting CSDE1 as a key MHC regulator in OAC for the first time.

5.4.2 T cell density reflects known immunophenotypes from deconvolution analysis.

Prior analysis of T cell distributions in OAC have observed the majority of OAC cases possess T cell infiltrate, yet the number of cases which possess prominent levels of T cell infiltrate is low (362). More recent T cell analysis in other cancers have sought to instead classify tumours into immunophenotypes including immune desert, immune suppressed and immune inflamed (537).

In my analyses in this chapter, using T cell density clustering I observed known immunophenotypes of cancer, first of which was the largest cluster forming the immune excluded phenotype lacking T cells of any type, suggesting this is cluster containing mixed immunophenotypes. The high proportion of tumour cores with low T cell density presented here reflects prior analysis in OAC suggesting only a low proportion of patients possess high T cell infiltration, though does not label these cases an immune excluded (538). The analysis presented here expands upon this earlier analysis as these cases can be labelled as immune excluded rather than immune suppressed as they lack the typical increased FOXP3+ Treg populations of immune suppressed cases which are present in the second immune cluster. Although a positive correlation between FOXP3+ T cells and CD3/4/8+ is demonstrated in my later correlation analysis, the lower abundance of CD8+ T cells in this cluster does support the suggestion these cases are immune suppressed. The last two immune clusters belong to the immune hot phenotype with high density of effector T cells (CD3/4/8+) and present as the most inviting target for checkpoint blockade therapy (539).

5.4.3 APM protein expression correlates with T cell density in OAC.

Correlation analysis of APM protein expression demonstrated positive correspondence to T cell density within my analysis. This included expected positive correlations such as HLA-ABC and CD3/4/8+ T cells as HLA-ABC is responsible for eliciting responses from CD8+ T cells, and CD4+ T cell is often co-recruited with other T cells. This is further supported

with the upper quantile of HLA-ABC protein expression corresponding to increased CD8+ T cell density. Interestingly, HLA-E H scores and not percentage positivity found positive correlation to CD8+ T cell density, supported by evidence the upper quantile of HLA-E expression possesses increased CD8+ T cells, suggesting a potential for specific cells expressing HLA-E may act to elicit HLA-E restricted T cells in the OAC TIME (540). TAP1 may be important in the activation and expansion of T cells in OAC with H scores correlating to high CD3/8+ T cell, yet the quantile analysis only demonstrated increased CD3+ T cells in the upper quantile. TAP1 has prior been demonstrated to correlate with immune cell infiltration in a pan-cancer study and relates to the potential of TAP1 expression to affect the immune epitope of cancer (531, 541).

Strikingly, the expected correlation between HLA-Class II protein expression and CD4+ T cell density was not observed within my OAC cohort (462), this results is surprising due to HLA-Class II roles in presenting antigens to CD4+ T helper cells. Instead HLA-Class II correlated to CD3/8+ T cells and Tregs with quantile analysis only displaying increased CD3/8+ T cells in the upper quantiles of HLA-Class II protein expression. However, it is difficult to suggest a potential mechanism this may occur through as HLA-Class II is expressed by multiple APC subtypes. However, one suggestion could form as high HLA-Class II represents tumours with increased IFN- γ known to induce HLA class II expression and is a marker of CD8+ T cell activation (462). Interestingly, CSDE1, a known negative regulator of MHC class I HLA expression did reflect its role in OAC with a negative correlation between CSDE1 H scores and HLA-ABC percentage positivity, however this also presents a complication as a positive correlation is observed between CSDE1 percentage positivity and HLA-ABC H scores (321). Overall, this reflects that CSDE1 expression may be localised to a subset of cells, which was noted in the histologists report CSDE1 stained tumour and immune cell populations in cancer, this finding requires further investigation which will be explored in single cell RNA-seq data analysis **(Presented in chapter 6).**

Furthermore, the upper quantile of CSDE1 protein expression in OAC possessed significantly lower CD3/8+ T cell populations but increased CD4+ T cell populations. Overall, this suggests CSDE1 expression may be localised to tumour cells in OAC and acts to reduce HLA-ABC protein levels in these tumour cells, resulting in lower tumour cell

immunogenicity and CD8+ T cell activation. This suggestion is further supported by my super pixel segmentation analysis which identified a relationship between predicted tumour content and CSDE1 protein expression within my TMA cohort (shown in **Figure 52**).

5.4.4 APM protein expression associates to altered survival outcomes in OAC.

In **Chapter 3** I observed the expression of my APM gene candidates associated to altered OS in OAC. To confirm these genes effected survival outcomes I analysed them at the protein level in my TMA cohort. Interestingly, only lower HLA-E and HLA-Class II protein possessed significantly reduced OS. Conversely to gastric cancer, HLA-E appeared as a positive indicator of univariate OS, where in gastric cancer HLA-E is noted for producing NK inhibition (542). Instead these results reflect more the observation in OSCC with HLA-E stain positivity corresponded to increased OS, as gastric cancer is considered more similar to OAC than OSCC is to OAC. This raises an important question surrounding the role of each APM protein in OAC, and whether prior assumptions of similarity in determining their respective roles is inaccurate, in relation to tumour immunity (543). This could suggest that although cancer cell gene expression in OAC is more akin to gastric cancers in prognostic gene expression programs, there may exist immunological differences in APM gene expression which play differing roles from gastric cancer, sharing features with OSCC instead. Although, this finding must be explored more directly applying the methodological approach I used for my IHC analysis in a future study of OSCC and gastric cancers to highlight the differences.

5.4.5 T cell density associates to altered univariate survival outcomes in OAC

T cell density has prior been published to associate to altered survival in OAC and gastric cancer, with greater intratumoural CD8+ T cells corresponding to improved OS (210, 362). Unfortunately, I did not observe this within the OS analysis of the OAC TMA cohort, this could be explained by a substantial majority of tumours in this cohort not expressing prominent levels of HLA-ABC, resulting in reduced anti-tumoral activity. To confirm this, I analysed the interaction between high/low HLA-ABC and CD8 T cell density finding high HLA-ABC with low CD8+ T cell density produced prognostically independent shortened OS compared to high HLA-ABC and high CD8+ T cell density. This suggests that neither HLA-ABC nor CD8+ T cell density alone impact OS, instead the presence of high CD8+ T cells in OAC tumours expressing high HLA-ABC results in improved OS due to anti-tumoral activity

of cytotoxic T cells. This finding is novel compared to prior studies which indicate CD8+ T cell density in OAC tumour cores is not an independent prognostic factor and future studies investigating T cell behaviour in OAC should stratify their results by HLA-ABC to elucidate the potential for anti-tumoral responses in OAC. Furthermore, there may be potential for this finding to possess translational value in stratifying patients which are likely to have high efficacy for checkpoint blockade therapy, by selecting patients with high CD8+ T cell density and high HLA-ABC protein expression by biopsy as well as staining for PD1/PD-L1.

5.4.6 Limitations and future work

One of the significant issues when performing this analysis was core loss over staining, with the final stain, CSDE1, experiencing significant core loss due to each section of tissue cut expending the remaining TMA. Thereby, future research effort exploring this cohort must first create new TMA from available tissue to ensure high sample size; additionally, an external cohort could be employed as further validation of these findings. A further limitation in performing TMA analysis forms from the small sample area stained, using a full section for analysis instead of tumour cores could explore further details of the heterogeneity of APM protein and immune populations. For example, whole tissue sections could be used to define tumours as either immune excluded or immune suppressed, with immune excluded tumour exhibiting denser immune populations in the tumour periphery, lacking immune populations in the tumour core and immune suppressed lacking immune cells in both the core and periphery. Additionally, the staining performed in my analysis was conducted on different sliced sections, which can reduce how comparable the staining is between sections, to improve upon this and confirm the accuracy of my results presented here future efforts could explore the use of multiplex IHC methods to overlap the staining of APM and immune markers. Additionally, the addition of cancer cell markers and immune activity markers into a multiplex IHC panel could explore the activity of T cells in proximity to cancer cells is dependent on the level of different APM protein expressions.

5.4.7 Conclusion

In conclusion, the results presented here confirm APM gene/protein expression is related to survival outcomes and the abundance of T cell subpopulations in OAC as uncovered in

Chapter 3 and 4. Furthermore, an important finding demonstrates CD8+ T cell alone is insufficient for immune rejection of OAC tumours, requiring high HLA-ABC expression to elicit these responses. Additionally, the expression of CSDE1 at both the mRNA and protein negatively corresponds to MHC class I HLA expression and should be explored further to confirm this relationship in cellular models of OAC (**Chapter 6**). Overall, my results suggest the future efforts in either producing new immuno-therapeutics or stratifying OAC patients for existing immune checkpoint blockade therapy should consider APM expression for maximizing the efficacy of patient responses.

Chapter 6 Cellular modelling of MHC class I expression regulation by CSDE1 in OAC

6.1 Introduction

6.1.1 Known immune evasion strategies used by cancer and current knowledge in OAC.

Immune evasion mechanisms are strategies often employed by cancer to evade immune detection and subsequent anti-tumoral immune responses. The immune system is equipped to monitor the microenvironment for cancer cells to detect these cells and eliminate them, however, cancer cells may develop somatically driven mechanisms to prevent recognition or promote immune cell anergy. Understanding these evasion strategies and their pathogenesis is key for the development of effective immunotherapies and for stratifying patients which are likely to respond to existing immunotherapeutic strategies. Broadly, tumour immune evasion strategies can be broken down into seven distinct categories including altering gene expression in cancer cells and adapting the tumour immune microenvironment by recruiting specific cell types and secreting immunosuppressive cytokines (544).

The seven distinct immune evasion strategies are described as follows; firstly, multiple cancers are known to downregulate the expression of MHC class I molecules to evade detection by CD8+ T cells in the immune microenvironment by hindering the presentation of tumour-associated antigen to these cytotoxic T cells (See section 6.1.2). Examples of this mechanism have been reported in OAC with miR125a-5p binding 3' untranslated region of the *TAP2* mRNA and miR148a-3p binding 3' untranslated regions of *HLA-A*, *HLA-B*, and *HLA-C* mRNAs to negatively regulate their respective expression (164).

Characterisation of alternative mechanisms of downregulation of MHC class I molecules has not yet been fully elucidated before the current study.

An additional mechanism of cancer immune evasion is found in tumour antigen heterogeneity, this describes the heterogenous antigen epitope presented by cancer cells and presents a significant challenge in producing effective anti-cancer engineered T cells. Specifically, the heterogeneity of tumour-associated antigens among cancer cells within a microenvironment means the recognition of a single tumour antigen usually does not

result in complete anti-tumoral immunity by CD8+ T cells against all cancer cells and results in clonal selection (and survival) of cancer cells which do not present the target antigen (545). Although, tumour antigen heterogeneity is not well characterised in OAC, recent analysis of OAC epitopes of seven patient samples only revealed one putative antigen from a single patient, demonstrating among OAC patients tumour antigens differ presenting a translational issue with developing immunotherapies targeted to a single antigen as the expression of a target antigen may not be ubiquitous among the patient population (546).

Immune checkpoints are critical in inhibiting immune responses towards host cells to prevent autoimmune disease. However, cancer cells often upregulate immune checkpoint molecules such as PD-L1 to promote CD8+ T cell anergy. Prior studies reported PD-L1/2 was upregulated in 40% of gastroesophageal cancers, yet the survival outcomes presented a complex picture, with increased PD-L1 expression resulting in significantly reduced OS and DFS in univariate analysis, but multivariate analysis found PD-L1 expression was an independent predictor for improved DFS, suggesting lack of immunosuppressive signalling may be protective against recurrence (101, 547-549). Despite the complex findings in PD-L1 expression in OAC, the current therapy of prior chemo-radiotherapy treated resected OAC tumour patients has recently changed in the UK to include nivolumab to target the PD-1/PD-L1 axis (550).

Cancer cells also seek to suppress immune cell responses by secreting immuno-suppressive factors such as TGF- β and IL-10, with TGF- β capable of suppressing cytotoxic expression programs and T cell proliferation and IL-10 suppressing CD4+ T cells and promoting Tregs in the tumour immune microenvironment (551, 552). Interestingly, dysfunctional TGF- β signalling by loss of TGF- β being noted in OAC cells (553, 554). Tumour also act to suppress immune response by recruiting immune regulatory cells or by polarising macrophages to an M2 phenotype to enact immune suppression. Specifically, tumours recruit Tregs and myeloid-derived suppressor cells (MDSCs) to suppress the immune response and inhibit T cell expansion in the tumour immune microenvironment. Interestingly, increased recruitment of Tregs in OAC has been associated to negative prognostic impact (CSS), however, the specific recruitment factors were not disclosed, whereas in OSSC the expression of Eomesodermin (Eomes) was found to promote progression by recruiting Tregs (555, 556).

The presence of MDSCs in oesophagogastric junction adenocarcinoma has been reported to promote progression, positively correlating with advanced staging, low grade, lymph node metastasis, and HER2- status (557). The polarisation of macrophages to the M2 phenotype was noted in OAC with nodal positivity and associated to worse overall survival, these M2 macrophages act to suppress immune responses by secreting factors by secreting immunosuppressive factors such as IL-10, arginase and TGF- β (401, 558, 559).

T cells also require co-stimulatory signals to activate their responses towards cancer cells, such as CD80 which interacts with CD28 on T cells and acts to promote activation of responses (560). However, cancer can downregulate CD80 expression to evade T cell responses as evidenced in OAC where significant downregulation has been reported which inversely correlates with TGF- β and IL-10 expression (561).

Lastly, stromal cells components (such as cancer-associated fibroblasts) can contribute to tumour immune evasion by secretion of immunosuppressive factors, building dense extracellular matrices and supporting tumoral vascularisation (562). Notably, increased rigidity of the extracellular matrix was reported in OAC and increased collagen in Barrett's oesophagus, the increase rigidity and density of the extracellular matrix in other cancers is noted to lead to immune exclusion preventing T cells and other immune effectors from reaching the tumour core (563-565).

6.1.2 The regulatory signalling pathways of MHC class I expression.

The expression of MHC class I genes is present in most cells in OAC tissue and has become a focus point in exploring the immunogenicity of cancer cells as antigens presented on HLAs of the MHC class I system are responsible for eliciting anti-tumoral immune responses from CD8+ cytotoxic T lymphocytes. Notably, a known immune escape mechanism in cancer is to reduce the expression of MHC class I genes to abrogate responses to tumour antigens by CD8+ T cells (566). These cancer cells can leverage regulatory elements for MHC class I expression which are present in normal healthy cells, which will be explored below.

Firstly, transcriptional regulation of MHC class I is enacted by multiple trans-activators, these include CIITA, which is a key transcription factor associated to MHC class II expression, however, studies indicate CIITA is also a key element in promoting MHC class I expression in cells via binding to the MHC class I promoters and thus enhancing the transcription of these genes (567).

RFX5-family genes also perform a similar role to *CIITA* by forming the RFX5-enhancersome consisting of the protein subunits *RFX5*, *RFXAP*, and *RFXANK*. The RFX5-enhancersome can bind to both CIITA and NLRC5 proteins to form the MHC class I enhanceosome which then associates to the SXYY module promoting downstream MHC class I expression (568).

The transcription factors NF- κ B, IRF1 (Interferon Regulatory Factor 1), and STAT1 (Signal Transducer and Activator of Transcription 1) act to positively regulate the expression of MHC class I genes. Specifically, STAT1 becomes phosphorylated (pSTAT1) within the JAK/STAT signalling pathway after IFN- γ binds the interferon-1 receptor on the cell surface, following phosphorylation, pSTAT1 dimers enter the nucleus and binds to the γ -activated sequence (GAS) which promotes the expression of *IRF1*, this is described as the type II IFN- γ response (569).

IRF1 after transcriptional promotion by pSTAT1 dimers binding GAS can then bind to the interferon response element (ISRE) which acts to promote MHC class I expression (570). However, the ISRE element can also be bound by a STAT1/STAT2-IRF9 complex called interferon-stimulated gene factor 3 (ISGF3) which forms downstream of the type I/III IFN- γ receptor which can result in promotion of MHC class I expression (**See Figure 56**) (571).

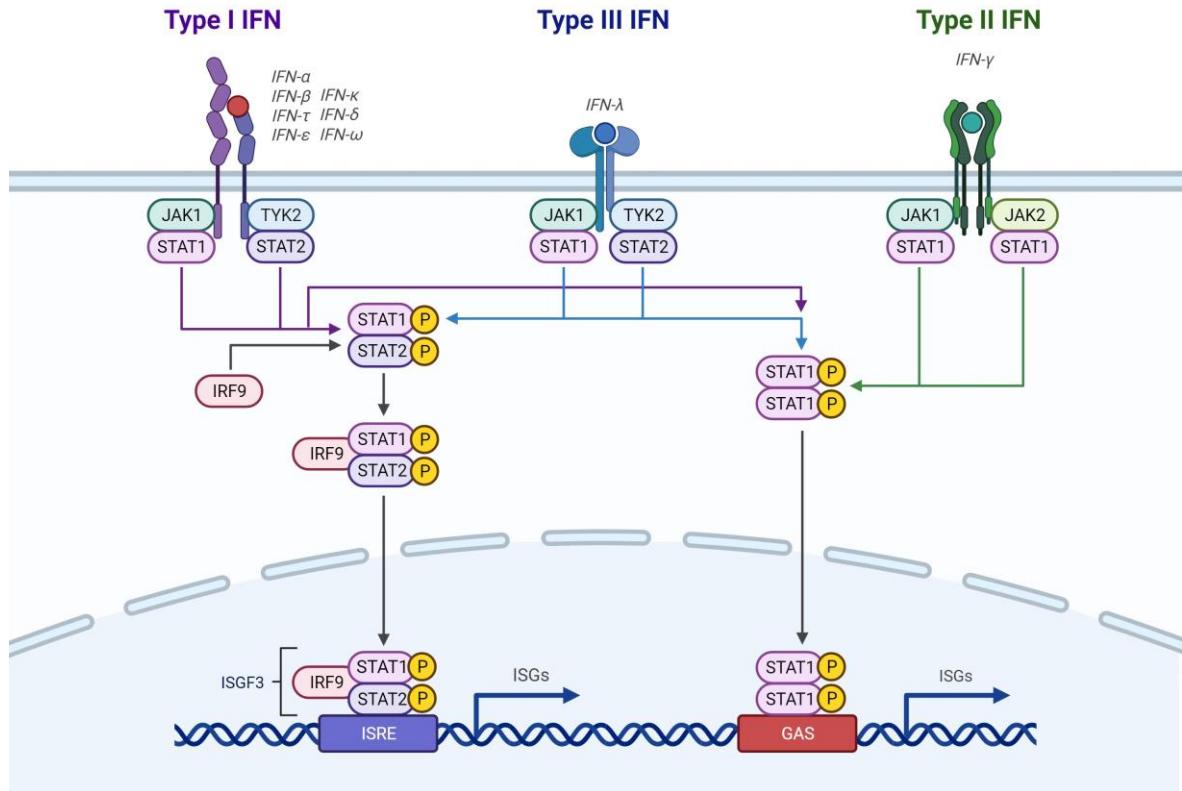


Figure 56 Type I, II and III Interferon signalling pathways for their respective receptors.

Interferons bind the type I, II and III receptors which propagate downstream signalling. In Type I responses interferons excluding IFN- γ bind the type I IFN receptor which results in the phosphorylation of STAT1/2 forming a dimer, this dimer binds IRF9 forming ISGF3 which translocates to the nucleus binding to ISRE to promote interferon-stimulated genes (ISGs). The formation of ISGF3 can also result from IFN- γ . Type II IFN receptor are bound by IFN- γ which results in a pSTAT1 dimer which translocate to the nucleus, then binds to GAS which leads to promotion of IRF1 (Created by Biorender).

Lastly, NF- κ B also acts to promote MHC class I expression by binding two NF- κ B binding sites in the enhancer A region, a requirement for induction of MHC class I expression (572). Notably, the trans-activators and transcription factors act synergistically and simultaneously to promote MHC class I expression (See Figure 57).

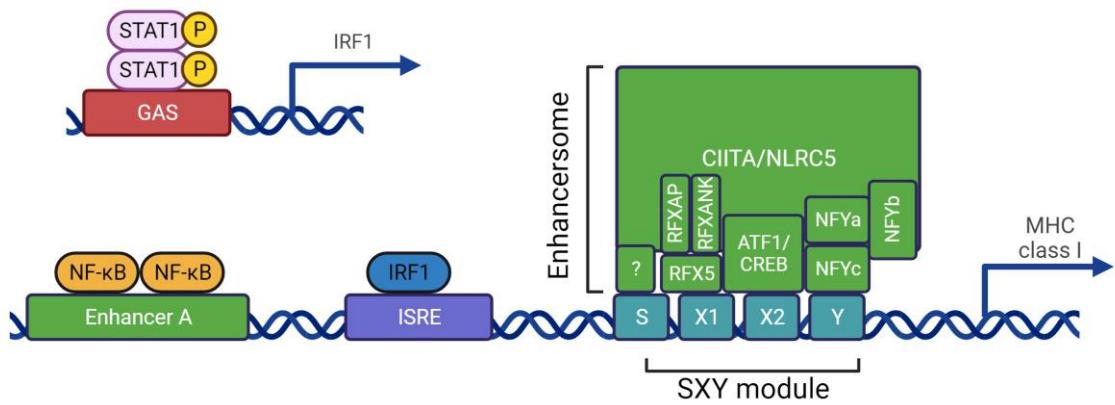


Figure 57 The MHC class I promotor region.

Translocated pSTAT1 binds GAS to promote the expression of IRF1. NLRC5/CIITA translocates into the nucleus forming the MHC class I enhanceosome with the RFX5-complex, ATF1/CREB and the NFY-complex which acts to bind the SXY module to promote the transcription of MHC class I. NLRC5/CIITA then can recruit chromatin modifiers and transcriptional elongation and initiation factors. IRF1 binds to the ISRE element within the MHC class I promotor region. NF-κB binds to Enhancer A to induce the expression of MHC class I genes (Created with Biorender).

6.1.3 The novel role of *CSDE1* in antigen presentation machinery gene expression regulation

CSDE1 is a gene coding for Cold Shock Domain Protein E1 and is implicated in several cellular processes including RNA stabilization, translational reprogramming, and protein homeostasis (573, 574). Among these roles, recent publications have indicated *CSDE1* can function as a negative regulator of MHC class I expression in cancer. The first study to uncover this role studied the effect of *CSDE1* expression on JAK/STAT signalling to produce reduced downstream promotion of MHC class I expression; this analysis was founded on the initial observation of a significant correlation between *CSDE1* mRNA expression and the expression of MHC class I genes (*HLA-A/B/C*, *TAPBP* and *TAP1/2*) within the TCGA-SKCM (skin cutaneous melanoma) dataset ($n = 471$) (575). Following this observation *CSDE1* knockout cell models of melanoma and breast cancer were generated, these cell models demonstrated increased MHC class I mRNA expression and gene ontology analysis revealed increased JAK/STAT signalling pathways. Further investigation observed *CSDE1* knockout cells possessed greater pSTAT1, suggesting *CSDE1* participated in inhibiting JAK/STAT signalling; interestingly, *CSDE1* knockout cell models observed reduced *PTPN2* (protein tyrosine phosphatase non-receptor type 2) mRNA stability uncovering the role of *CSDE1* in stabilising *PTPN2*.

The importance of *PTPN2*, also known in its protein form as TCPTP (T cell protein tyrosine phosphatase), was highlighted in this study by its function to dephosphorylate pSTAT1 within the JAK/STAT signalling pathways downstream of the IFN receptors (575). The dephosphorylation of pSTAT1 inhibits translocation of STAT1 dimers into the nucleus, preventing binding to either GAS or ISRE thus restricts the promotion of *IRF1* and MHC class I gene expression (See Figure 58) (575). Lastly, this publication identified *CSDE1* is regulated in tumorigenic cells via the *SMYD3*-mediated H3K4 trimethylation which results in transcription activation, though TCGA analysis only revealed a minor correlation between *SMYD3* and *CSDE1* expression suggesting only nuclear *SMYD3* may strongly correlate to *CSDE1* expression (575).

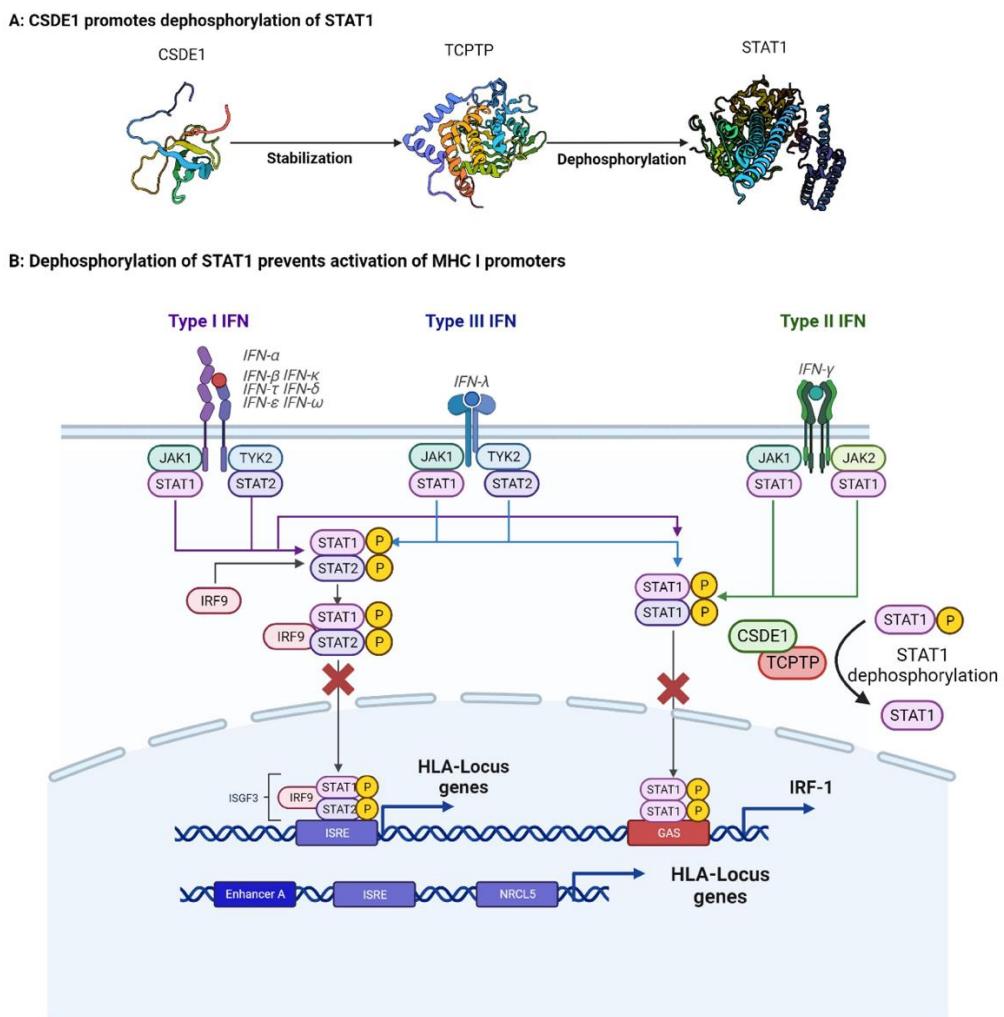


Figure 58: CSDE1 as a negative regulator of MHC class I expression via the JAK/STAT pathway. (A) CSDE1 stabilises TCPTP protein in the cytosolic compartment (Panel A reproduced, with permission from, Galassi and Galluzzi, 2023)(576, 577). (B) Stabilisation of TCPTP results in pSTAT1 dephosphorylation inhibiting pSTAT1 translocation to the nucleus preventing binding of STAT1/2 homodimers/heterodimers to promoters of IRF1 and MHC class I HLA genes.

6.1.4 Biological functions of CSDE1

CSDE1 (Cold Shock Domain Containing E1) is a gene which possesses cellular multiple functions. Unfortunately, the 3D structure of *CSDE1* has not been fully characterised, however, it is known like other cold shock binding proteins, *CSDE1* possess a five-stranded all-antiparallel β -barrel structure, with two cold shock domains, RNA-binding domains and low complexity regions (578).

The first known role of *CSDE1* protein is found in RNA Binding, where it functions to regulate mRNA stability, translation, and localisation, processes crucial for gene expression regulation. The mRNA binding role of *CSDE1* has been explored in *Drosophila* models, where it binds *msl2* and *roX* mRNA which plays a key role in the control of X-chromosome dosage compensation (579). An additional role of *CSDE1* is in stress response, specifically as a regulator of oxidative stress, where *CSDE1* and *STRAP* (Signal-transducing adaptor protein) proteins interact to convey sensitivity to oxidative stress (580).

A further role describes *CSDE1* protective role from DNA damage within the NER (nucleotide excision repair) and DSB (double strand break) pathway and then promote cell growth by modulating the expression of *RPA2*, *CHOP*, *PERK* and *GRP78* (581). *CSDE1* has also been implicated in the regulation of embryonic development, with loss of *CSDE1* accelerating neural differentiation and potentiates neurogenesis, with ectopic expression of *CSDE1* impairing neural differentiation. Specifically, *CSDE1* binds the fatty acid binding protein 7 (*FABP7*) and vimentin (*VIM*) mRNAs, and transcripts involved in neuron projection development which regulates their stability and translation (582).

Lastly, aberrant expression or dysregulation of *CSDE1* has been prior published to be critical to maintain invasive phenotype of colorectal cancer via positively regulating c-MYC and associates with epithelial-to-mesenchymal Transition (583). Specifically, this article identified *CSDE1* expression was greater in colorectal tumour samples compared to normal tissue and in metastatic origin cell lines compared to primary cell lines (583). Additionally, downregulation of *CSDE1* reduced cell viability and migration during restraint of epithelial-to-mesenchymal transition experiments, increasing sensitivity to apoptosis

(583). Lastly, they observed high *CSDE1* expression associated with poor prognosis, positively correlating to c-MYC expression in colorectal cancer tissue and cell lines (583).

6.1.5 **CSDE1 As a target for therapy in cancer**

Interestingly, a small number of recent publications identified *CSDE1* as a potential target for immunotherapy after oncolytic therapy derived by rhinoviruses as, *CSDE1* expression is required for translation of rhinoviruses (146). This article found *CSDE1* possessed a selective pressure to somatically mutate in cancer treated with an oncolytic rhinovirus to produce a C-T mutation to produce a proline to serine change at amino acid 5 (p.P5S); this mutation was found to produce a neo-epitope recognized by non-tolerized T cells, which suggests a potential to form a trap-ambush type of therapy using an oncolytic virus to produce the *CSDE1*^{C-T} escape variant with vaccination to prevent tumour escape (See **Figure 59**) (146). This investigation was followed up by exploring the use of T cells primed with *CSDE1*^{P5S} peptides, finding an optimal priming regimen possessing increased IFN- γ upon challenge. Furthermore, the research group observed responses to cells presenting *CSDE1*^{P5S} neo-epitopes could be improved with the use of the CD200AR-L checkpoint inhibitor (584).

Overall, this presents *CSDE1* as a key factor in MHC class I expression and as a targetable feature in cancer cells. Yet this mechanism has only been directly described in melanoma and breast cancer cell models, with data mining analysis indicating the mechanism may be observed in stomach adenocarcinoma, breast cancer or bladder urothelial carcinoma, colon adenocarcinoma (TCGA-COAD) and liver hepatocellular carcinoma (TCGA-LIHC) (575).

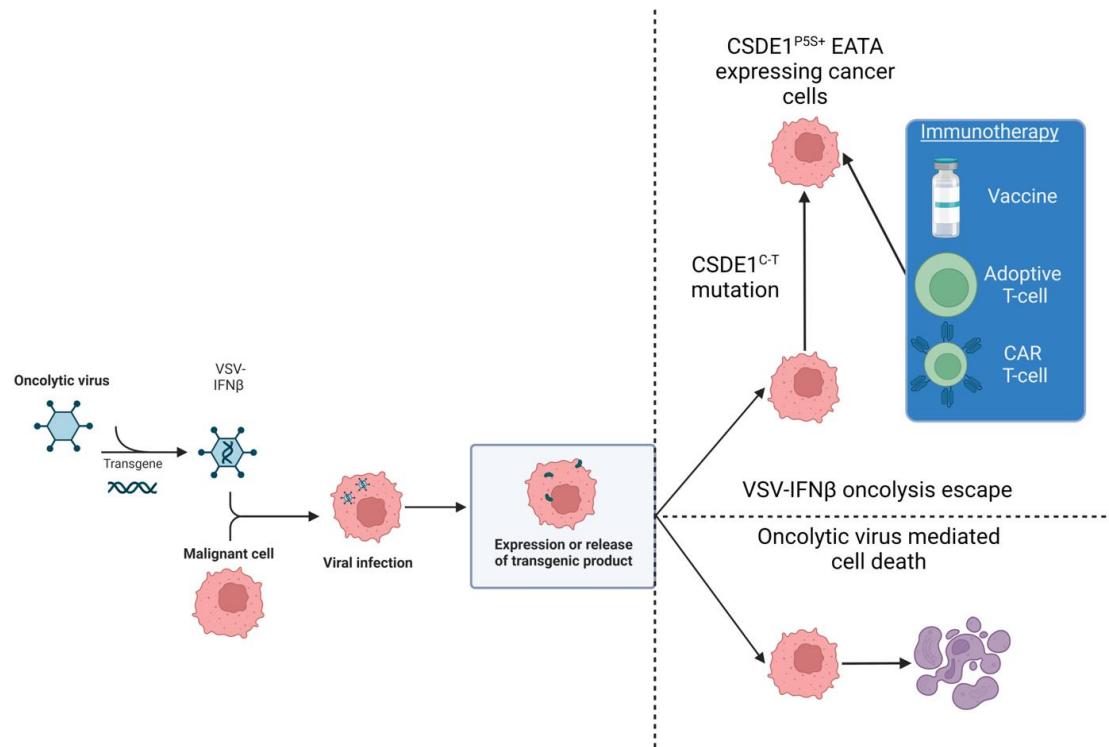


Figure 59 Oncolytic trap and ambush strategy. Targeting initial treatment failure with a highly targeted, escape-selective oncolytic virus then applying immunotherapeutic therapy to target escape variants. Conceptual framework from Kottke et al, 2021 (585).

6.1.6 Manipulating gene expression in vitro experiments

Gene expression manipulation describes the control or alteration of the genes ability to produce downstream proteins, this methodology is important in the field of functional biology to determine the roles of genes in cellular functions and possesses clinical value in the manipulation of genes to produce clinically favourable outcomes. Gene expression can be manipulated using multiple differing methods depending on the application.

Broadly, gene manipulation covers mRNA targeting molecules which acts to reduce the expression of specific mRNA from target genes, knockout methods which aim to prevent gene expression by modifying the genomic DNA of a host cell; overexpression methods which aim to increase a genes expression by knocking in a target gene or transfecting in mRNA for a target gene and finally, epigenetic modification which aims to alter the epigenetic regulation of a genes expression; these methods will be expanded on below (586, 587).

Firstly, siRNA standing for small interfering RNA describes short, double-stranded RNA molecules that can specifically target complementary sequences of genes to inhibit their expression (588). Specifically, siRNA function by guiding the protein complex RNA-induced silencing complex (RISC) to a specific messenger RNA, the siRNA binds a complementary sequence of the target mRNA, which is followed by the RISC complex cleaving the mRNA or suppresses the mRNA's translation (588). Interestingly, siRNA has already been employed in previous publication to investigate *CSDE1*'s role in the regulation of c-MYC and epithelial-to-mesenchymal transition and the role of micro-RNA in cancers (583, 589).

Short hairpin RNA (shRNA) functions similarly to siRNA, but possess a hairpin-like structure, with a stem-loop, double-stranded RNA region as a precursor to forming siRNA (590). The key difference in shRNA is the biogenesis, where typically shRNA is expressed from viral or plasmid packages, then cellular machinery forms a mature shRNA (590, 591). Mature shRNA then binds complementary sequences of mRNA directing RISC to cleave the respective mRNA or prevent translation. The study prior described to investigate the role of *CSDE1* in MHC class I expression employed shRNA to discover *CSDE1* is a negative regulator of MHC class I expression in melanoma (592).

Gene knockout refers to techniques which aim to deactivate or delete a target gene so it cannot be expressed. Generally, this involves using a targeting method such as CRISPR (clustered regularly interspaced short palindromic repeats) or zinc-finger motifs packaged in a delivery vector which is complementary to a genomic sequence, which then allows for cleavage of the DNA by an enzymes such as CAS9 in the case of CRISPR to form a double stranded break, which introduces mutation of the gene via insertion or deletion rendering the gene non-functional (593, 594). Usually when performing a gene knockout a selection marker is simultaneously introduced such as an antibiotic resistance gene, allowing for the selection of knockout clones (593). These methods can possess translational value such as attempting to knockout the *HLA-A* gene via the CRISPR-CAS9 system to reduce the incidence of alloreactive immune responses triggered by incompatible HLA molecules after organ transplantation (595).

Overexpression techniques differ from those prior described as their purpose is to increase the expression of a target gene. The vector to achieve overexpression involves plasmid type vectors which functionally achieves overexpression of a target gene by either gene cloning where a gene is cloned into a plasmid, controlled by a strong promoter, which is transfected into a host cell which then expresses the target gene. Or by an inducible system which incorporates an inducible promoter, allowing for control of a target gene's expression (596). An additional vector system involves using transposons, which are DNA sequences that can relocate within the genome, allowing for an engineered transposon system which carries a target gene which can be inserted into the host genome resulting in gene expression (597). Lastly, a modified CRISPR-CAS9 system can be implemented to activate constitutive gene expression by fusing a transcriptional activator to the CAS9 protein then targeting the gene of interest to induce expression (598). An interesting use case of overexpression in antigen presentation research, involved overexpressing *HLA-C* in colorectal cancer cells. Where the overexpression of *HLA-C* in these cells resulted in down-regulation of genes that were highly enriched in cancer-related signalling pathways such as JAK/STAT, ErbB, and Hedgehog signalling pathways (154).

An alternative method to affect the expression of a target gene would be achieved by epigenetic modification, for example the use of 5-Azacytidine, which was trialled in myelodysplastic syndromes (599). Azacytidine functions by covalently binding DNA methyltransferase resulting in DNA hypomethylation and prevents DNA synthesis, however, this mechanism is untargeted leading to a requirement of targeted epigenetic modifying tools (600). Targeted epigenetic modification have thus far been developed using zinc-finger nucleases (ZFNs), transcriptional-activator like effectors (TALEs) and CRISPR which are bound to a methyltransferase that act to methylate a target gene inhibiting transcription (601). Using methylation inhibitors such as 5-aza-2'-deoxycytidine has already been implemented to investigate MHC class I HLA expression in healthy donor cells, which notes increased *HLA-A* expression after treatment of the cells with a DNA methyltransferase inhibitor (602). Furthermore, this approach observed *HLA-A* allelic lineage-specific methylation patterns located to the *HLA-A* promoter region, suggesting *HLA-A* allelic lineage-specific variation may be partially driven by epigenetic modification (603).

6.1.7 Cell models of OAC

Cell models are critically important tools in the study of cancer, allowing researcher to perform functional studies genes and proteins involved in cancer development and progression. These frequently make use of genetic manipulation technologies, such as gene knockout/down or overexpression experiments designed to explore the roles of specific genes as examples in the above investigation of CSDE1 functional roles in melanoma and breast cancer using CRISPR-CAS9 to knockout *CSDE1* (145, 589).

Eight cell lines have commonly been used for the investigation of OAC (See table 25); firstly, OE33 is one of the most common OAC cell lines in use for the investigation of OAC was established from a lower oesophagus adenocarcinoma of a 73-year-old female patient in pathological stage IIA and poor differentiation (604). Notably, OE33 expresses *HLA-A/B/C* genes constitutively, which allows OE33 to be used as a cell model of MHC class I expression in OAC to explore the impact of genomic manipulation of *HLA-A/B/C* expression. An example of OE33 use in research can be found in the observation of FLOT (Fluorouracil, Leucovorin, Oxaliplatin and Docetaxel) directly upregulated PD-L1 on the surface of OE33 cells, which could be targeted with checkpoint blockade to improve lymphocyte killing of FLOT treated OE33 cells; these pre-clinical studies provided justification to explore the use of checkpoint blockade therapy in prior chemotherapy resected OAC tumours (605, 606).

FLO-1 is another OAC cell line used for functional genomics; FLO-1 was established from a primary distal OAC of a 68-year-old male (607). Interestingly, FLO-1 has prior been exploited to investigate AURKA (Aurora kinase A) role in regulating JAK2–STAT3 activity in OAC finding JAK2 mediates AURKA-induced phosphorylation of STAT3 (608). This process was found to promote the expression of cytokines, growth factors, and pro-survival genes that act within multiple cellular processes including survival, cell cycle and invasion (609). Despite the practicality of using cell lines in cancer studies, it is important to understand there are significant limitations in employing cell lines to investigate functional genomics in cancer. These limitations include genetic drift where cell lines can undergo genetic changes over time due to continuous passaging, the lack of tissue complexity meaning these models not accurately represent the interaction present in the tumour microenvironment, and the inability to reproduce *in vivo* complexity among others (610).

Overall, to perform functional genomic research to investigate the influence of *CSDE1* expression on MHC class I expression sought by this chapter, the commercially available cell lines must be assessed to determine if the cell model possesses the correct genomic and expression profile to do this robustly.

Table 25 Commonly used OAC cell lines in literature.

CELL LINE	AGE	SEX	ETHNICITY	HISTOLOGY	LITERATURE IN ANTIGEN PRESENTATION INVESTIGATION
ESO26	56	MALE	CAUCASIAN	ADENOCARCINOMA - GASTROESOPHAGEAL JUNCTION	YES (611)
ESO51	74	MALE	CAUCASIAN	DISTAL OESOPHAGEAL ADENOCARCINOMA	NO
FLO-1	68	MALE	CAUCASIAN	DISTAL OESOPHAGEAL ADENOCARCINOMA	YES (532)
KYAE-1	60	MALE	ASIAN	DISTAL OESOPHAGEAL ADENOCARCINOMA	NO
OACM5.1 C	47	FEMALE	CAUCASIAN	BARRETT'S ADENOCARCINOMA, ADENOCARCINOMA OF DISTAL OESOPHAGUS	NO
OE19	72	MALE	CAUCASIAN	ADENOCARCINOMA - GASTRIC CARDIA/OESOPHAGEAL GASTRIC JUNCTION	YES (532)
OE33	73	FEMALE	CAUCASIAN	OESOPHAGEAL ADENOCARCINOMA	YES (605)
SK-GT-4	89	MALE	CAUCASIAN	WELL-DIFFERENTIATED OESOPHAGEAL ADENOCARCINOMA	YES (605)
MFD-1	55	MALE	CAUCASIAN	OESOPHAGEAL ADENOCARCINOMA	NO

6.1.8 Hypothesis and chapter aim, and objectives.

From my prior analysis I found evidence suggesting *CSDE1* expression may function as a negative regulator of MHC class I HLA expression in OAC, however, this analysis requires further supporting evidence to substantiate this claim. To address this, I aimed to assess the impact of modulating *CSDE1* expression on the mRNA and protein level of MHC class I HLAs in an OAC cell model and explore single cell RNA-seq to confirm the distribution of *CSDE1* expression in OAC and compare expression between cancer cells and normal comparative cells.

Hypothesis 3: *'Expression of CSDE1 in oesophageal adenocarcinoma cells down-regulates MHC class I HLA expression.'*

Objective 4a: Knockdown/overexpression of *CSDE1* will be performed via siRNA and validated at the mRNA and protein level in OAC cell lines.

Objective 4b: MHC class I mRNA and protein expression levels will be measured in knockdown/overexpression *CSDE1* cell models, determining the effect of *CSDE1* expression on MHC class I expression.

Objective 4c: phosphorylation status of STAT1 protein levels will be measured in OAC cell lines to determine the effect of altered *CSDE1* expression on the activation of the JAK/STAT signalling pathway.

Objective 4d: *CSDE1* mRNA expression will be repartitioned into single cell populations from single-cell RNA sequencing data generated from primary human tumour tissue to determine whether *CSDE1* is over expressed in OAC cancer cells in comparison to normal cells.

6.2 Methodology

6.2.1 Cell expression profiling

Cell mRNA raw counts data was sourced from Sanger cell model passports for OE33, OE19 and FLO-1 (<https://cellmodelpassports.sanger.ac.uk>), mRNA raw counts for MFD-1 cells were downloaded from the Underwood Research file store. mRNA raw counts were then, filtered to remove lowly and highly expressed transcripts and normalised using TMM (allowing for within and between sample normalisation) to explore the cell lines for CSDE1 and HLA-A/B/C expression values (612).

6.2.2 General principles of cell culture

Cell culture is a methodology to allow for cellular expansion *in vitro* using nutrient medium. This technique allows for manipulation of cells if required and provides cellular read-outs and DNA, RNA, and proteins for downstream analysis. For my experiments routine cell culture was performed within a laminar flow hood with the tissue culture reagents and nutrient media being stored within sterile containers at 4°C. To avoid cold shock of cells, media and reagents were placed in a water bath at 37°C prior to use. All cell cultures were grown at 5% CO₂ in a humidified environment of an incubator (37°C). Cell lines were validated by STR-PCR in the Underwood laboratory.

6.2.2.1 Thawing cell stocks

Cell stocks stored in vials were removed from liquid nitrogen were placed at room temperature for 5-10 mins until completed thawed. Thawed cell stock was transferred to 15 ml tubes (BD Falcon), where 5 ml of cell media (DMEM/RPMI containing 10 % w/v heat inactivated foetal Bovine serum (FBS), 2.0 mM L-glutamine, 50 IU/ml of penicillin (100 U/ml) and 50 µg/ml of streptomycin) was added before centrifugation at 800 g for 5 minutes to produce a cell pellet. After discarding the supernatant, the remaining cell pellet was re-suspended in complete growth medium and transferred to a cell culture flask (25 cm² volume; Corning®) containing fresh cell culture media.

6.2.2.2 Cell culture

Cell lines were cultured within a humidified cell incubator at 37 °C with 5 % CO² for OE33, FLO1 and MFD1 cells. For OE33 cells complete growth medium of Roswell Park Memorial Institute 1640 medium (RPMI) supplemented with 10 % w/v heat inactivated foetal Bovine serum (FBS), 2.0 mM L-glutamine, 50 IU/ml of penicillin (100 U/ml) and 50 µg/ml of streptomycin. MFD-1 and FLO-1 cells were cultured in Dulbecco's modified eagle medium (DMEM) supplemented with 10 % w/v heat inactivated foetal calf serum (FCS), 2.0 mM glutamine, and 50 IU/ml of penicillin and 50 µg/ml streptomycin.

Cell cultures were grown as adherent monolayers in sterile cell culture flasks (25cm², 75 cm² or 175 cm² volume; Corning®) in a humidified incubator (5 % CO₂) at 37 °C. Media was changed every 3 - 4 days or when 80-90 % confluence was achieved.

Confluent cells were passaged at a 1:3 to 1:5 ratio of depending on the growth rate of the cells. To achieve passage. The existing medium was discarded, then the cell layer within the cell culture flask was washed with PBS heated to 37 °C removing the remaining media. Pre-warmed trypsin/TrypLE™ Express Enzyme (0.05 % trypsin (w/v) / 5 mM EDTA, Cat#25200056, ThermoFisher/TrypLE™ Express Enzyme (1X), no phenol red, Cat#12604013, ThermoFisher) was added to coat all cells, then incubated for 5 mins at 37 °C. FBS containing media was then added to detached cells to inactivate the trypsin/TripLE by inhibiting enzyme reaction. Finally, the free-floating cells in media suspension were divided into new tissue culture flasks and topped up with complete cell culture media.

6.2.2.3 Cell count

To perform cell counts, 0.25 ml of cells solution was added to a 0.5 ml Eppendorf, with 0.25 ml of Trypan Blue Solution (0.04 % Trypan Blue, Cat# 15250061, ThermoFisher). 10 µL of Trypan Blue + cell solution was pipetted into a C-chip haemocytometer (Neubauer Improved C-Chip Disposable Haemocytometer (2 channel), Cat#DHC-N01-50, NanoEnTek). The haemocytometer was then placed under a microscope at 10x magnification (Nikon Diaphot) and focused on the grids. Next five 1 mm² area grid squares were then counted; the average cell count of the five grid squares was then calculated, then multiplied by the dilution factor to give the cell count per mL.

6.2.3 CSDE1 knockdown by siRNA transfection

CSDE1 siRNA knockdown of FLO-1 cells was achieved via lipofection. This required plating cells at $0.25\text{--}1 \times 10^6$ in 6 well plates until 60 % confluent overnight at 37°C in 5 % CO2. Cells were then transfected overnight in Opti-MEM® Medium using Lipofectamine® RNAiMAX Reagent (Lipofectamine™ RNAiMAX Transfection Reagent, Cat#13778075, Invitrogen) with a concentration of CSDE1 siRNA smart pool (siGENOME Human CSDE1 (7812) siRNA – SMART, Cat#: M-015834-01-0010, Horizon Discovery/Dharmacon) pool of 60 pmol in 250 μL Opti-MEM® Medium (Opti-MEM™ I Reduced Serum Medium, Cat# 31985062, Gibco). Following overnight incubation cell media was replaced with complete RPMI media and incubated for 3 days before harvesting the cells.

6.2.4 CSDE1 overexpression by nucleofection

CSDE1 overexpression of OE33 was attempted during my analysis. This was conducted using a nucleofection kit (SE Cell Line 4D-Nucleofector™ X Kit S, Cat# V4XC-1032, Lonza). This required adding 2×10^5 to each 20 μL Nucleocuvette with 0.4 μg of pmaxGFP™ Vector and 20 μL SE Cell Line 4D-Nucleofector™ X Solution then placed into a Nucleofector unit where 7 different Nucleofector® programs plus 2 controls were conducted. Post nucleofection the cells were incubated in 12 well plates in complete media at 37 °C in 5 % CO2 overnight, before imaging until a fluorescent microscope. Imaged cells were analysed to determine which program produced the most successful nucleofection marked by GFP. Beyond this stage the most optimal program would have been selected for nucleofection with 0.2–1 μg of CSDE1 plasmid [CSDE1 (NM_001130523) Human Tagged ORF Clone, Cat#RC226183, Origene]. Unfortunately, after imaging it was determined cells had been infected and my remaining time in the laboratory did not allow for optimisation nor experiment conditions (see results **section 6.4.5**).

6.2.5 RNA isolation and purification

Extraction, isolation, and purification of RNA was conducted with use of RNeasy Mini Kit (RNeasy Mini Kit, Cat#74104, Qiagen). The protocol for this procedure made use of highly denaturing buffer containing guanidine-thiocyanate on lysed and homogenized samples to inactivate RNases, then selection of RNA occurs using the binding properties of a silica-based membrane using spin columns with ethanol.

The RNeasy Mini spin column allows binding of total RNA to the silica membrane with unbound elements being washed away, which isolates RNA from other cellular lysates. The RNA was extracted from cell solution which were centrifuged 300 x g for 5 mins with the supernatant discarded. Cells were first disrupted with 350 µL RLT cell pellet. Cell lysates were then homogenized using a sonicator for 30 s (3 x 10 seconds sonication with 15 second rests). The homogenized lysates were then transferred to the gDNA spin column and centrifuged for 30s at 8,000 g. The gDNA column was then discarded and the supernatant flow-through was collected, next 350 µL 70 % ethanol was added to the collected solution and thoroughly mixed. Next 700 µL of the lysate solution was transferred into a RNeasy spin column then centrifuged at \geq 8000 x g for 15s allowing binding of RNA to the silica column. Following this step, the flow through was discarded, 700 µL of RW1 buffer was used to wash the column, with further centrifugation at \geq 8000 x g for 15s. A further wash using 500 µL RPE buffer was then conducted twice, an additional centrifugation at \geq 8000 x g for 15s with the flow-through being discarded. Finally, the spin column was transferred into a 1.5 ml collection tube and eluting the RNA with 30 µL of RNase-free water. This collection tube was closed and centrifuged at 8,000 x g for 1 minute then stored at -80°C.

6.2.6 Nucleic acid quantification

To quantify purified mRNA and derived cDNA methods used a Nanodrop 1000 Spectrophotometer (ThermoFisher Scientific, UK). This required using 1 µL of nucleic acid-free water which was pipetted onto the nanodrop pedestal as calibration of the spectrophotometer. Next 1 µL of cDNA or mRNA in solution was read with the pedestal being cleaned thoroughly between readings. The DNA and RNA content for each sample were recorded in ng/µL, with the purity of the mRNA/cDNA being determined by 260/230nm and 260/280 ratios. DNA samples were satisfactory when the 260/280nm absorbance ratios; expected purity measurements for qualified further analysis were required to be within a 260/280 ratio of 1.8-2.0 for RNA and cDNA and between 2.0-2.2 for 260/230 ratio. Quantified cDNA stored at -20°C, for future analysis.

6.2.6.1 Reverse transcriptase polymer chain reaction (PCR) for cDNA synthesis

To provide ample quantity of nucleic acid (RNA) for downstream analysis copy DNA (cDNA) was produced via reverse transcription of 1 µg of RNA (High-Capacity cDNA Reverse Transcription Kit Cat#4368814, Applied Biosystems). The principles of reverse transcription involve firstly deoxy-thymine nucleotides with a short sequence [Oligo (dT)] which contains a complementary primer to the poly-A tail, and which also provides a free 3'-OH end for extension by reverse transcription. To achieve reverse transcription, a master mix of 2 µL reverse transcription buffer, 0.8 dNTP Mix, 2 µL random primers, 1 µL MultiScribe Reverse Transcriptase and 4.2 µL of nucleic acid-free water was combined for a total of 10 µL of master mix per sample. Next the required concentration of RNA was calculated to provide 40 ng/µL in 10 µL of nucleic acid-free water for all samples to allow a total of 20 µL for each samples PCR reaction (master mix + sample RNA). The solution was then placed in a thermocycler to the following program: 25°C - 10 minutes, 37°C – 120 minutes, 85°C – 5 minutes and 4°C – stop to generate complementary DNA (cDNA). cDNA was then stored at -20°C, for future analysis.

6.2.7 Quantitative Real-Time PCR (qRT-PCR) by TaqMan

qRT-PCR is a method used to quantify specific sequences of cDNA allowing the researcher to extrapolate the quantity gene expression as it is amplified by PCR in real time. The premise of TaqMan is the use of a primer and probe which are specific to the gene under study. Specifically, an oligonucleotide probe is fluorescently labelled as a reporter on the 5' end and a quencher molecule on the 3' end. The importance of the quencher dye is to reduce the fluorescence emitted by the reporter dye whilst the probe is intact, upon polymerisation the gene probe anneals downstream of the primer site allowing extension by the Taq DNA polymerase, cleaving the reporter dye. The cleavage of the reporter dye allows for detection of the reporter dye since as it is removed from the proximity to the quencher dye. This signal is further amplified by the cleavage of the quencher dye by exonucleases during the PCR cycle. These two cleavage events also allow primer extension to continue, thereby, fluorescence emitted in TaqMan qPCR is directly proportional to the number of copies of cDNA produced per cycle. The PCR reactions here are quantified when the overall fluorescence exceeds a set threshold to produce a Ct value which represents the number of PCR cycles needed to reach the required fluorescence threshold.

Therefore, a higher Ct value would represent lower expression gene under study. To establish the levels of *CSDE1* cDNA, *HLA-A*, *HLA-B* and *HLA-C* cDNA or control gene *GAPDH* cDNA in cells, qRT-PCR was performed with the QuantStudio Flex Real-Time PCR System (QuantStudio™ 7 Flex Real-Time PCR System, 384-well, desktop, Cat# 4485701, Applied Biosystems) with measurements being conducted using the standard curve method. To estimate gene expression of the genes of interest the log of cDNA concentration was plotted against Ct values for known quantities of cDNA derived from a dilution series from normal fibroblasts and the equation of this standard curve used to calculate the amount of *CSDE1*, *HLA-A*, *HLA-B*, *HLA-C* or *GAPDH* cDNA present in cells.

Commercially available primers for *CSDE1* (Hs00918650_m1, Cat#4331182, ThermoFisher Scientific), *HLA-A* (Hs01058806_g1, Cat#4331182, ThermoFisher Scientific), *HLA-B* (Hs07292706_g1, Cat#4351372, ThermoFisher Scientific), *HLA-C* (Hs00740298_g1, Cat#4331182, ThermoFisher Scientific) and *GAPDH* (Hs01060665_g1, Cat#4448489, ThermoFisher Scientific) were used to measure cDNA levels for each gene of interest in knockdown cells. 1 μ L cDNA was suspended and made up to 10 μ L with 3.5 μ L nuclease free water, 0.5 μ L of TaqMan Assay (20X) and 5.0 μ L of TaqMan Fast Advanced Master Mix (TaqMan™ Fast Advanced Master Mix for qPCR, Cat#4444557, Applied Biosystems). To ensure that the solutions were mixed thoroughly and then to remove bubbles, samples were vortexed. Following this the samples were loaded in triplicate into a 384 well then centrifuged for 300 x g for 1 min. The PCR plate was then sealed and loaded into the PCR machine and ran through 50°C (2 minutes), 95°C (20 seconds), then 40 cycles of 95°C (1 second), and 60°C (20 seconds). To effectively calculate the level of gene expression the output Ct of each gene of interest was applied to the standard curve of known RNA quantity using the equation below (y = Ct value, m = slope, x = log(quantity), b = y-intercept (613)).

$$\text{Standard curve: } y = mx + b$$

$$\text{Quantification: } x = \frac{y}{m} - \frac{b}{m}$$

Equation 2: Standard curve and mRNA quantification equations for TaqMan analysis.

6.2.8 Western blotting

6.2.8.1 Cell lysis

Cell pellets ($\sim 3.6 \times 10^6$ cells) were re-suspended in 1ml of PBS in an Eppendorf tube, and centrifuged at 8000 x g for 4 mins at 4°C. The supernatant was removed off the cell pellet and ice, reducing the potential for protein degradation. Next 20 μ L of Radio-immunoprecipitation assay (RIPA) buffer and 5 μ L of protease inhibitor and 5 μ L of phosphatase inhibitor (Halt phosphatase inhibitor cocktail, Cat#78420, ThermoFisher Scientific) was added to each cell pellet, then briefly vortexed, following being placed on ice for 30 minutes. Lysed cell mixture was then centrifuged at 8000 x g for 5 minutes at 4°C. Finally, supernatant was moved to a new Eppendorf and stored at -20°C.

6.2.8.2 Estimation of protein concentration

To allow for optimal protein loading I first determined the protein concentration achieved after cell lysis via a Bradford assay (BCA) using Bio-Rad Protein Assay Dye Reagent Concentrate (Bio-Rad Protein Assay Dye Reagent Concentrate, Cat#5000006, Bio-Rad). This assay functions by producing a colour change of Coomassie brilliant blue G-250 dye which occurs when binding amino acid residues present in protein. This colour change can then be read by a spectrophotometer set for absorbance (595nm). Bovine serum albumin (BSA) protein standards were produced using a dilution series to create set protein concentrations of 0, 125, 250, 500, 750, 1500, 2000, 4000 and 5000 μ g/ml. To achieve the analysis firstly 250 μ L of the dye concentrate was prepared at a dilution of 1:5 with deionized water then pipetted into a sterile 96-well plate, following this 1ul of protein lysate for each sample was pipetted into wells in triplicate, as well as a triplicate of 1:10 dilution of protein lysate in deionized water and the protein standards in triplicate.

The plate was then incubated for 15 minutes at 37 °C, then placed into a Varioskan Flash plate reader (Varioskan Flash, Cat#5250030, ThermoFisher Scientific) and read at 590 nm absorbance. The absorbance measurements for the protein standards were then used to produce a standard curve, with the unknown sample protein concentrations being quantified using this standard curve in Microsoft Excel using the equation below (X= unknown protein quantity, Y= absorbance, m= gradient and c= y-intercept).

$$X = (Y - c)/m$$

Equation 3: Protein quantification equation for BCA assay.

6.2.8.3 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins are separated by electrophoresis using a 1.5 mm thick pre-moulded polyacrylamide gel (NuPAGE 10 %, Bis-Tris 1.5 mm Mini Protein Gel, Cat#NP0306BOX, Invitrogen. The gel was placed in the electrophoresis tank and the chamber filled with running buffer (NuPAGE Tris-Acetate SDS Running Buffer, Cat#LA0041, Invitrogen). The first step in the SDS-PAGE process was to prepare the protein samples for loading by adding 7.5 μ L of sample buffer (NuPAGE™ LDS Sample Buffer (4X), Cat#NP0007, Invitrogen), 3 μ L of reducing agent (NuPAGE™ Sample Reducing Agent (10X), Cat#NP0004, Invitrogen) and adding nucleic acid-free water to control the concentration up to a total of 30 μ L for loading into the wells. These prepared samples were heated it at 70 °C for 10 minutes, vortexed, then placed on ice. Protein samples were loaded into wells of the polyacrylamide gel together with a protein weight ladder (PageRuler™ Pre-stained Protein Ladder, 10 to 180 kDa, Cat#26616, Thermo Scientific). Lastly, the electrophoresis was conducted for 35 minutes at 200 V constant. After running the gel was then removed and submerged in transfer buffer (NuPAGE™ Transfer Buffer (20X), Cat#NP0006, Invitrogen).

6.2.8.4 Immunoblotting

After running the SDS-PAGE gel, the separated protein was then transferred to a nitrocellulose blotting membrane (Nitrocellulose Membranes, 0.2 μ m, Cat#88013, Thermo Scientific), sandwiched between 2 filter paper and sponges then compressed, in a Mini Blot Module (Mini Blot Module, Cat#B1000, Invitrogen), and immersed in transfer buffer up to the fill line of the Mini Gel Tank (Mini Gel Tank, Cat#A25977, Invitrogen). Proteins were transferred onto the membrane at 10V, 160 mA for 60 minutes. After protein transfer, the membrane was covered in Ponceau S (Ponceau S, 0.1 % v/v soln. in 5% acetic acid, Cat# J63139.AP, ThermoFisher Scientific Chemicals) to confirm successful transfer. Following successful transfer, the membrane was placed inside a 50 ml BD Falcon tube with 5 ml of 5 % BSA to block non-specific proteins, and gently agitated for 90 minutes on a roller (180 rpm) at room temperature.

After blocking the membrane was washed 3 times with 5 ml of PBS/Tween at 5 mins and gently agitated for 90 minutes on a roller at room temperature. The membrane was then

incubated in a Falcon tube with 5 ml of diluted primary antibody in 5% BSA (See table 26) overnight on a roller (180 rpm) at 4°C.

After primary antibody incubation the stained membrane was washed 3 more times, in PBS/Tween for 5 minutes, then incubated with a secondary antibody for 1 hour on a roller at room temperature. A final wash was conducted 3 in PBS/Tween for 5 minutes at room temperature with bound antibody being detected via chemiluminescence (SuperSignal™ West Pico/Femto, Cat#34577/34094, ThermoFisher Scientific), then imaged on a ImageQuant™ 800 Western blot imaging system (ImageQuant™ 800 Western blot imaging system, Cat# 29399481, Amersham). GAPDH was probed as a loading control and subsequent re-probing of membrane for other target proteins involved stripping off the bound antibody with 5ml of stripping buffer (Restore Western Blot Stripping Buffer, Cat# 21059X4, ThermoFisher Scientific), 15 minutes at room temperature after a single wash for 5 minutes with TBS (Tris-buffered saline) at room temperature on a roller. After stripping the membrane was washed three times with 5 ml of PBS/Tween for 5 minutes at room temperature, followed by primary antibody incubation for the next protein target. Quantification of target protein was conducted in Image J, to produce a score of intensity of target protein normalised to the housekeeping intensity (GAPDH, used after laboratory testing confirming this protein expression is unchanged by non-targeting siRNA control) (614). Antibodies used in this method are detailed in **Table 26**.

Table 26 Antibodies used in immunoblotting of target proteins in western blot analysis.

ANTIBODY TARGET	CLONE	SUPPLIER	CATALOG NUMBER	SPECIES/HOST ISOTYPE	DILUTION
HLA-CLASS I (ABC)	EMR8-5	ABCBAM	AB70328	MOUSE	1:5000
CSDE1/NRU	EPR17414	ABCBAM	AB201688	RABBIT	1:1000
STAT1 ALPHA PHOSPHO STAT1	EPYR2154 EPR3146	ABCBAM	AB92506 AB109461	RABBIT	1:5000
GAPDH	6C5	ABCBAM	AB8245	MOUSE	1:5000
ANTI-RABBIT	-	ABCBAM	AB6721-1	GOAT	1:5000
ANTI-MOUSE	-	ABCBAM	AB97023	GOAT	1:5000

6.2.9 Single-cell RNA-seq analysis

Single-cell RNA-seq data was obtained from previous analysis conducted (Provided by the Matthew Rose-Zerilli/Underwood laboratory group) (615). This process involved disaggregating patient tissues into single cell suspensions using a modified protocol for fibroblast characterisation (616). Disaggregated tissue was processed through modified DropSeq protocol v3.1, allowing single cells to be co-encapsulated with a barcoded bead within a nanolitre droplet (617). These droplets were then broken, and the bead encapsulated yield were reverse transcribed followed by exonuclease and PCR steps. 500 pg of cDNA were used for a Nextera XT library prep kit (Nextera XT DNA Library Preparation Kit, CAT# FC-131-1024, Illumina) per sample before sequencing on a NextSeq 500. The raw reads were then demultiplexed and converted to Fastq files through bcl2fastq (Illumina). Converted Fastq file then underwent Dropseq Core Computational Protocol v2.1 pipeline, aligned to the HG38 genome using STAR 2.6.0a. Downstream analysis was conducted using Seurat in R (Version 4.4.0) (618). Data visualisation of single-cell RNA expression was conducted using UMAPs, Violin plots, AddModuleScores (scoring gene sets) and Dot plots functions in Seurat.

6.2.10 Statistical analyses

Statistical analysis compared CSDE1 knockdown to controls in mRNA and protein expression was conducted in GraphPad prism 10 using one-sample T-tests for protein comparisons, as only one repeat of non-targeting and wildtype was available for comparison and T-tests for mRNA quantity comparisons (619). For T tests, significance was determined at a p value <0.05 after Bonferroni multiple testing correction; Bonferroni was selected due to greater power in testing a small number of comparisons compared to the Tukey method (620). Pearson correlation analysis was employed in GraphPad prism 10 the correlation of APM protein expression, CSDE1 and STAT1 signalling level. Lastly, paired T-tests was performed in R for single-cell RNA-seq analysis of CSDE1 expression of intestinal metaplasia of precancerous lesions and comparative analysis of OAC cancer cells versus normal healthy cells.

6.3 Results

6.3.1 Cell model selection and characteristics

To determine which available cell lines would be most appropriate for knockdown and overexpression experiments for *CSDE1*. The mRNA count data was downloaded (from Cell Model Passports) for accessible OAC cell lines (**See Table 25**) then normalised using TMM in EdgeR (621). Exploring the mRNA expression level of *CSDE1*, other canonical APM gene expression regulators and MHC class I genes revealed interesting observations for each of the cell lines. Firstly, the expression of *CSDE1* was found to be the highest in OACM5-1, ESO26, ESO51 and FLO-1 cells with the greatest *HLA-A/B/E* found in ESO26 cells (**See Figure 60**). The lowest expression of MHC class I molecules was located in cluster 1 which included OE19, OACM5-1 and SK-GT-4 cells; the low MHC class I expression of OE19 could relate to OE19 cells also possessing the lowest expression of *IRF1*, *NLRCS5* and the highest expression of *PTPN2* (TCPTP), exploring the available data for OE19 reveals a copy number loss event over the *IRF1* gene locus (**See Figure 60**). Cluster 2 represents cell containing high *TAPBP*, *IRF1* and *TAP2* expression; among cluster 2, FLO-1 cells demonstrated low MHC class I expression with high *CSDE1* expression (**See Figure 60**). These finding suggests FLO-1, OACM5-1 and SK-GT-4 cells contain lower MHC class I expression, excluding *TAP2/TAPBP* with intact APM regulators and high *CSDE1* expression making FLO-1 and OACM5-1 cells suitable candidate models for *CSDE1* siRNA knockdown.

Conversely, the greatest MHC class I expression was located to cluster 3 which included OE33, ESO26 and ESO51 cell lines. OE33 possessed the greatest *B2M*, and *HLA-C/E* expression accompanied by high *NLRCS5* and *IRF1* expression among OAC cell lines and highly expressed *CSDE1* (**See Figure 60**). ESO26 possessed the greatest MHC class I expression within cluster 3, but also contained high *CSDE1* expression (**See Figure 60**). MFD-1 possessed the greatest *HLA-A/B*, *TAP2* and *TAPBP* coupled with high *IRF1* expression (**See Figure 60**). These results suggest OE33, and MFD-1 cell models are immediately available and would form effective models for *CSDE1* overexpression, whilst FLO-1 cells are ample for knockdown experiments.

Lastly, I explored the expression of the *HMGA2* (High Mobility Group AT-Hook 2) gene, which is thought to be regulated by *SMYD3* similarly to *CSDE1*.

Here I observed a similar pattern of *HMGA2* expression in the OAC cell lines compared to *CSDE1* with the greatest expression being in OACM5-1 cells and the lowest expression in OE19 and ESO26 cells.

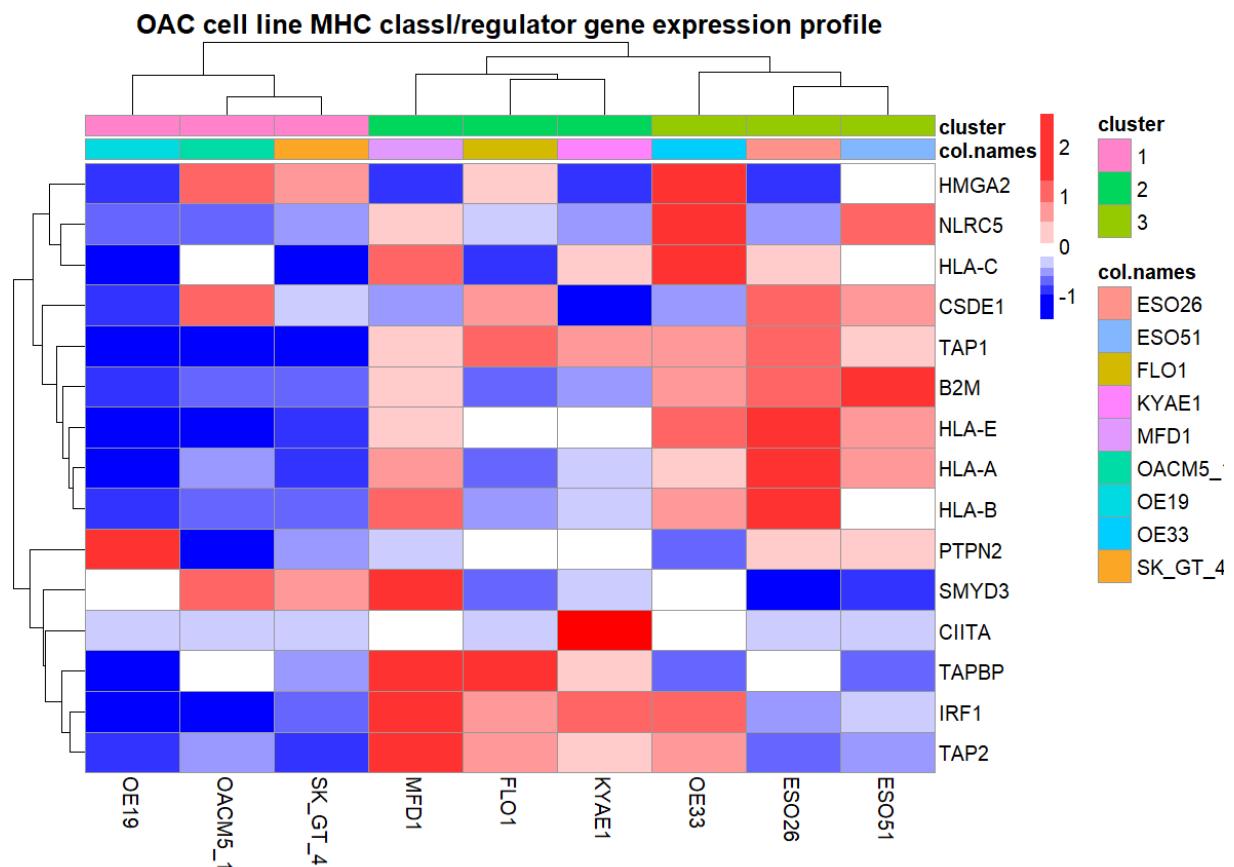


Figure 60 Heatmap of TMM normalised expression of *CSDE1*, MHC class I gene expression regulators, MHC class I genes and HMGA2 for available OAC cell lines. Z-scored, Minkowski distance, ward.D2 linkage.

6.3.2 Validation of CSDE1 knockdown in OAC cell model

After knocking down *CSDE1* expression in FLO-1 cells, I decided to validate the success of the knockdown at the protein level using a western blot (3 repeats). Using this approach, I observed a reduction of *CSDE1* protein expression in *CSDE1* knockdown cells. Specifically, a 55% reduction in *CSDE1* protein was noted between knockdown and non-targeting control (NC), a 55% reduction between knockdown and wildtype (WT) and a 61% reduction between knockdown and lipofectamine treated FLO-1 cells (See Figure 61 & 62). This result determines that *CSDE1* was successfully knocked down in FLO-1 cells using 60 pmol of *CSDE1* siRNA, allowing for further investigation of the impact of *CSDE1* knockdown on MHC class I HLA expression and STAT1 signalling in this OAC cell line. Observing the lipofectamine and non-targeting siRNA control (NC) at the mRNA expression level found no significant increase in *CSDE1*, *HLA-B* and *HLA-C* expression, however, NC treatment did exhibit an increase in HLA-A mRNA expression in FLO-1 cells compared to WT (See Figure 63).

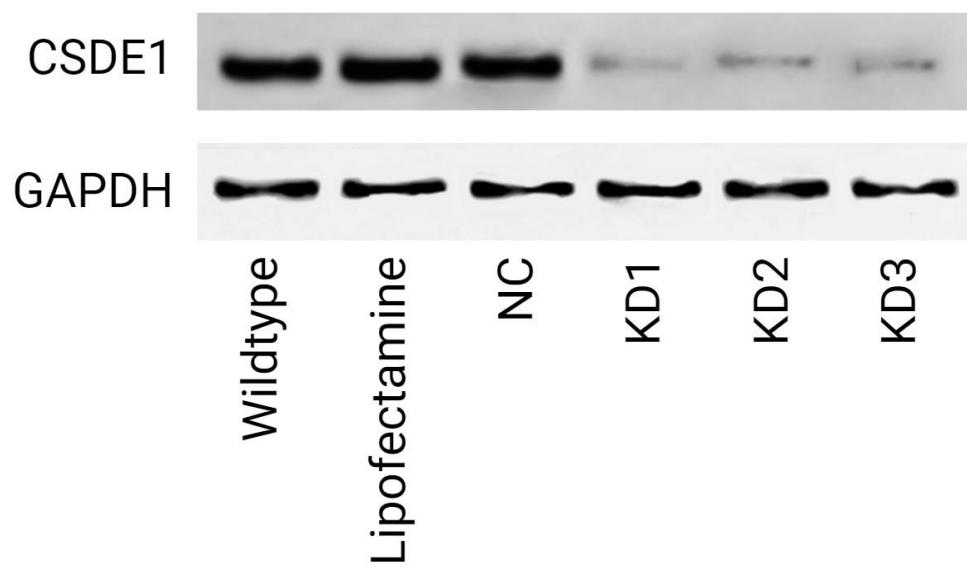


Figure 61 Western blot of CSDE1 and GAPDH expression following CSDE1 knockdown in FLO-1 cells. NC: Non-targeting control, KD: Knockdown CSDE1 60 pmol siRNA. Full images available in appendix 2A and E.

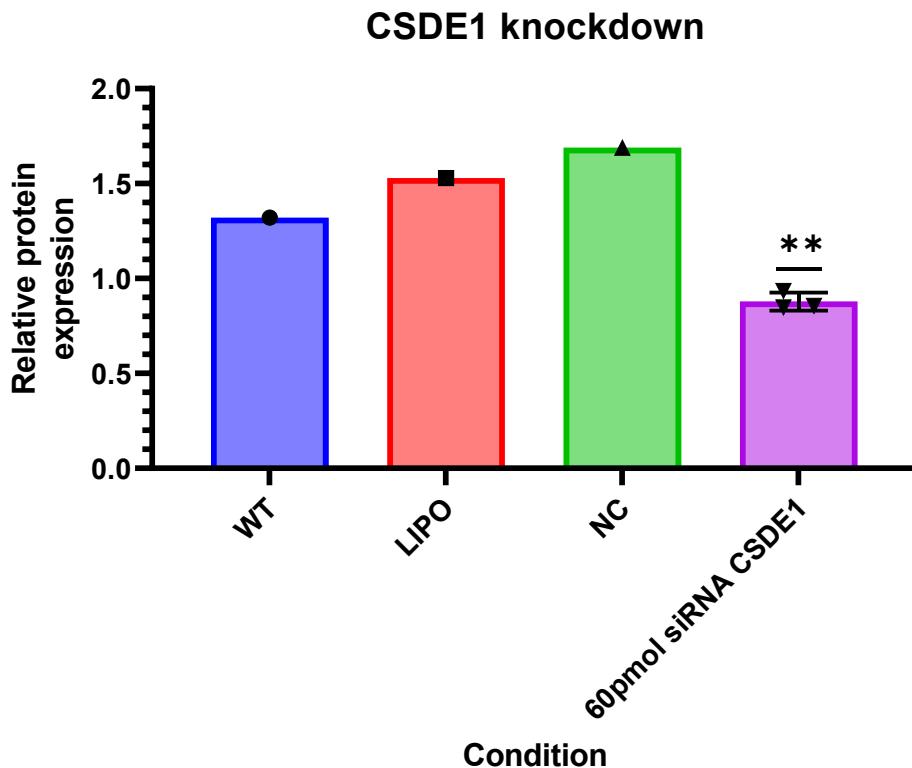


Figure 62 Protein quantification of CSDE1 normalised to GAPDH following siRNA knockdown in FLO-1 cells. NC: Non-targeting control. Error bar: Standard deviation. Statistical test: one-sample T-test comparing knockdown to NC, p values* <0.05 , ** <0.01 , *** <0.001 , **** <0.0001 .

6.3.3 APM gene expression in OAC cancer cells is downregulated by the expression of CSDE1.

Following successful knockdown of *CSDE1* expression in FLO-1 OAC cells measured at the protein level, I aimed to confirm *CSDE1* knockdown at the mRNA level and identify an association between lower *CSDE1* expression and greater MHC class I HLA (*HLA-A/B/C*) expression in FLO-1 OAC cells. Here I observed a significant reduction in *CSDE1* mRNA in FLO-1 cells with a 94% reduction in *CSDE1* mRNA in knockdown conditions (both 60 and 25 pmol of *CSDE1* siRNA) compared to NC, a 94% reduction compared to WT and lastly, a 93% reduction compared to lipofectamine treated FLO-1 cells ($p<0.0001$, **See Figure 63**); no significant difference was observed between *CSDE1* mRNA expression between 60 and 25 pmol siRNA treatment.

Following confirmation of *CSDE1* knockdown at the mRNA level, I moved to investigate the impact of the knockdown on MHC class I HLA gene expression, here I present *CSDE1* knockdown (60pmol siRNA) resulted in a significant increase of in *HLA-A* and *HLA-B* compared to WT, 34% and 68% respectively ($p<0.05$, **See Figure 63**).

However, no significant increase in was noted between *CSDE1* knockdown and NC for *HLA-A/B*, however, this may be potentially due to the NC siRNA control possessing significant increase of 18% in *HLA-A* and 62% increase *HLA-B* mRNA expression compared to the WT suggesting the NC may result in increased HLA expression. Lastly, although *HLA-C* mRNA expression was not significantly higher in knockdown *CSDE1* FLO-1 cells there is depicted an increase of 12% in quantity of *HLA-C* mRNA in 60pmol *CSDE1* siRNA condition over NC and an increase of 14% over WT (See Figure 63).

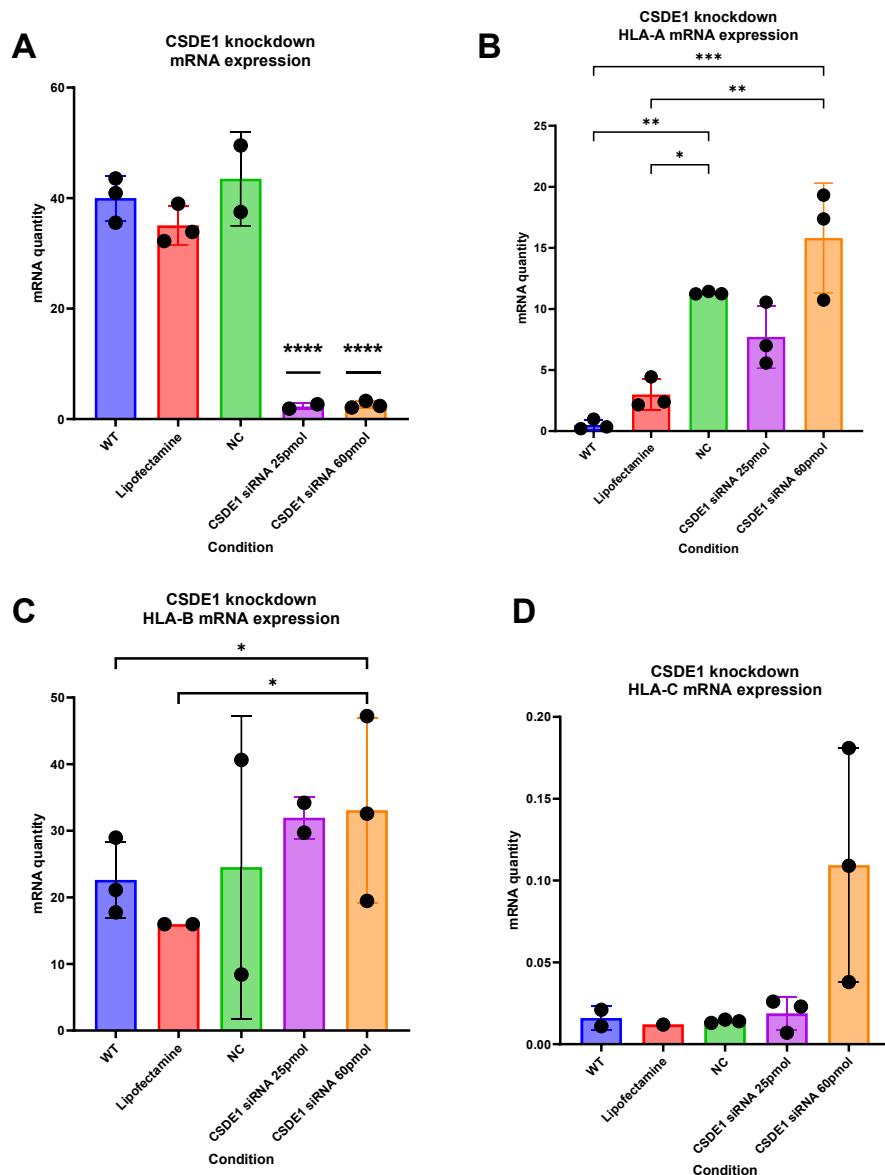


Figure 63 mRNA quantification of CSDE1, HLA-A, HLA-B and HLA-C expression following CSDE1 knockdown in FLO-1 cells. A: CSDE1 gene expression following *CSDE1* knockdown. B: *HLA-A* gene expression following *CSDE1* knockdown. C: *HLA-B* gene expression following *CSDE1* knockdown. D: *HLA-C* gene expression following *CSDE1* knockdown. NC: Non-targeting control. Error bar: Standard deviation. Statistical test: Mann-Whitney U test, p values* <0.05 , ** <0.01 , *** <0.001 , **** <0.0001 . Percentage differences compared to wildtype available in appendix 3A-D.

6.3.4 APM protein expression in OAC cancer cells may be downregulated by the expression of CSDE1 as a modulator of the STAT1/JAK pathway.

After qualifying a successful knockdown of *CSDE1* in FLO-1 cells and observing increased *HLA-A/B* expression in these knockdown cells I decided to explore if these results are reflected at the protein level using 60pmol *CSDE1* siRNA. Unfortunately, I did not observe a significant increase in *HLA-ABC* level in *CSDE1* knockdown though a minor increase of 17% was noted in *HLA-ABC* compared to NC, the lack of significance here may be due to the small number of repeats (See Figure 64 & 65). However, *CSDE1* knockdown did demonstrate significant increase in pSTAT1/STAT1 ratio of 46%, suggesting STAT1 signalling is greater with reduced *CSDE1* ($p<0.05$, See Figure 64 & 65).

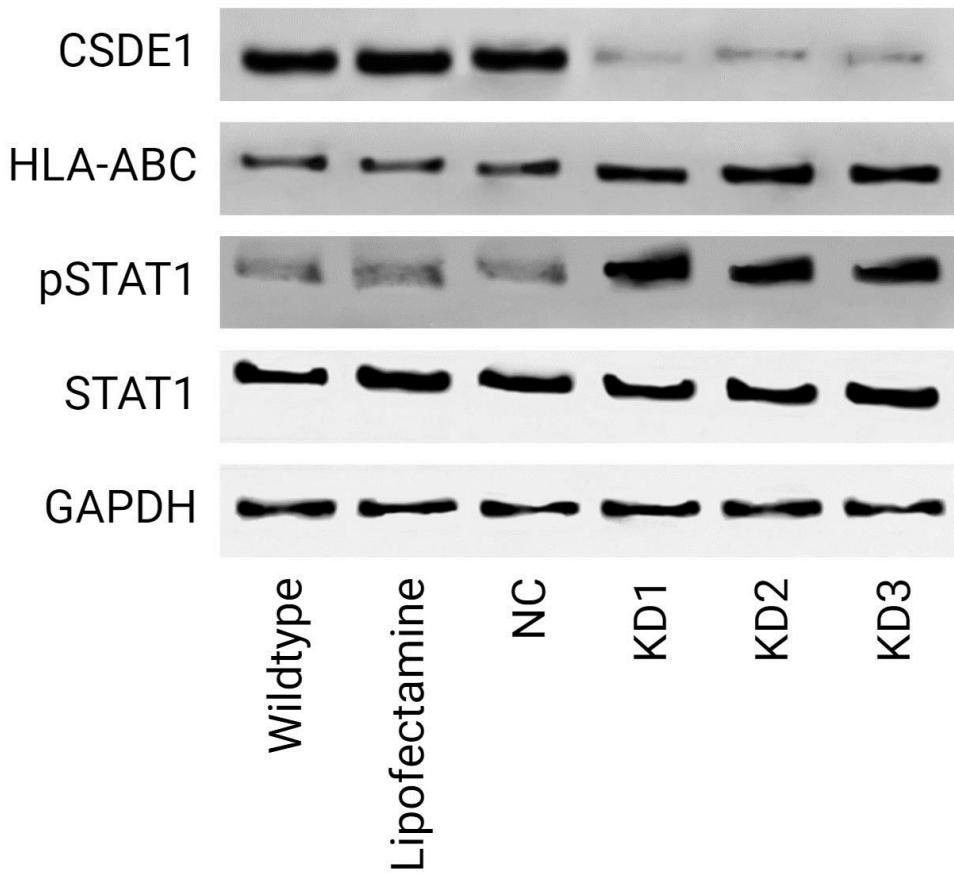


Figure 64 Western blot of CSDE1, HLA-ABC, pSTAT1, STAT1 and GAPDH expression following CSDE1 knockdown in FLO-1 cells. NC: Non-targeting control, KD: Knockdown CSDE1 60 pmol siRNA. Full images for each stain available in appendix 2A-E.

CSDE1 knockdown impact on protein expression

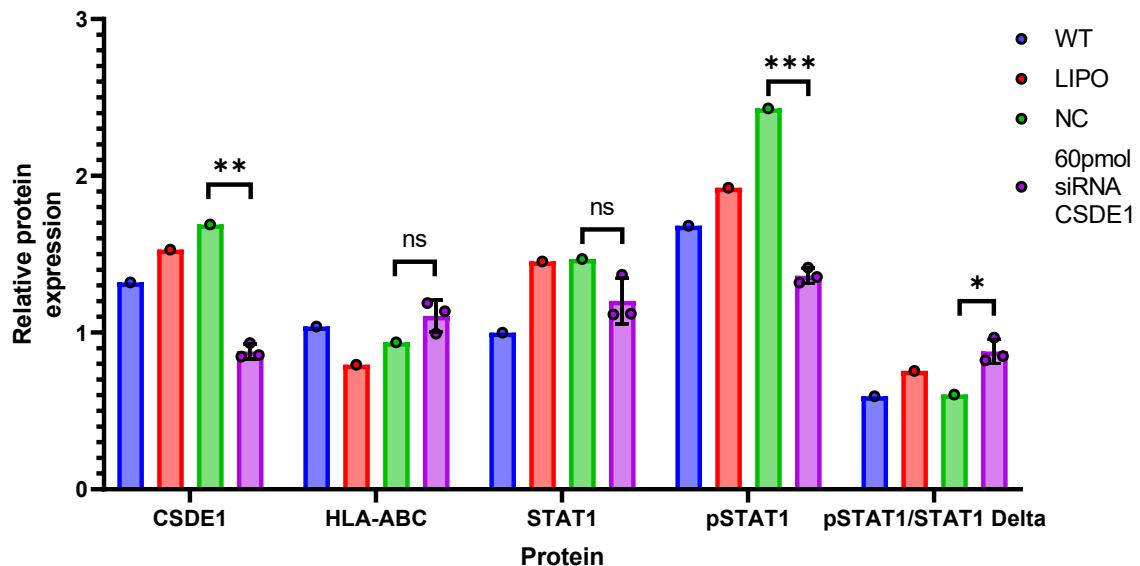


Figure 65 Protein quantification of CSDE1, HLA-ABC, pSTAT1, STAT1 and GAPDH expression following CSDE1 knockdown in FLO-1 cells. NC: Non-targeting control. Error bar: Standard deviation. Statistical test: one-sample T-test comparing knockdown to NC, p values* <0.05 , ** <0.01 , *** <0.001 , **** <0.0001 , NS = non-significant.

Lastly, I moved to explore the correlation of CSDE1, pSTAT1/STAT1 ratio and HLA-ABC at the protein level in all cell conditions. Here I observed a trend between STAT1 signalling and HLA-ABC protein expression ($R^2 = 0.07$, $p = 0.61$, **See Figure 66**); and the negative correlation between CSDE1 protein expression and STAT1 signalling trended towards significance ($R^2 = 0.59$, $p = 0.076$, **See Figure 66**). Additionally, CSDE1 protein expression demonstrated a trend towards a negatively correlation to HLA-ABC protein expression ($R^2 = 0.66$, $p = 0.096$, **See Figure 66**).

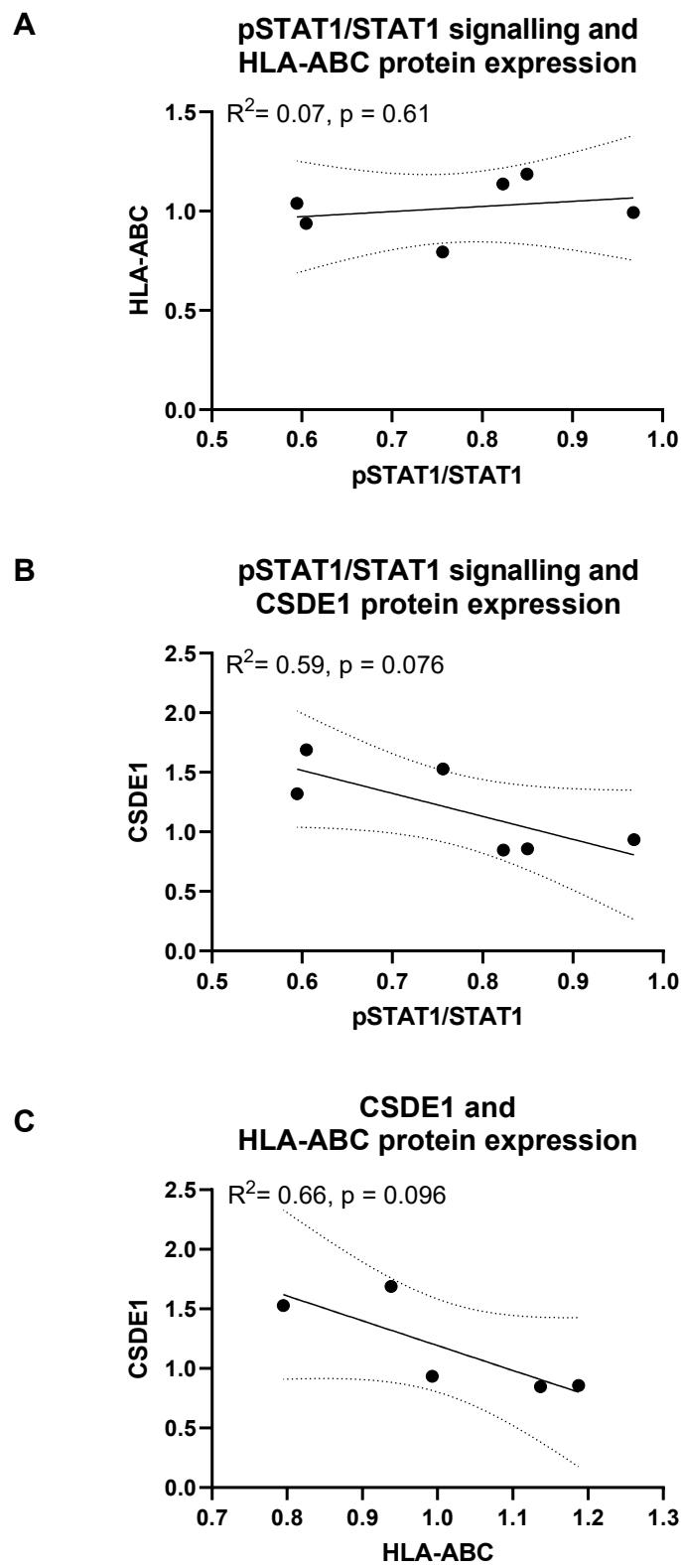


Figure 66 Correlation analysis of protein expression (CSDE1, HLA-ABC and pSTAT1/STAT1) in FLO-1 OAC cells. A: pSTAT1/STAT1 signalling and CSDE1 protein expression. B: CSDE1 and HLA-ABC protein expression. C: pSTAT1/STAT1 signalling and HLA-ABC protein expression. Dotted line: 95% confidence interval. Statistical test: Pearson's.

6.3.5 Attempted *CSDE1* overexpression in OE33 cells.

After performing the *CSDE1* knockdown experimental work above, I attempted to overexpress *CSDE1* using plasmid nucleofection in OE33 cells. Unfortunately, after imaging the optimisation stage using GFP as a marker of successful nucleofection, the cell morphology indicated excessive amounts of cell death, shortly after the nucleofection it was determined the cell culture were infected and dying and subsequently disposed of (See Figure 67).

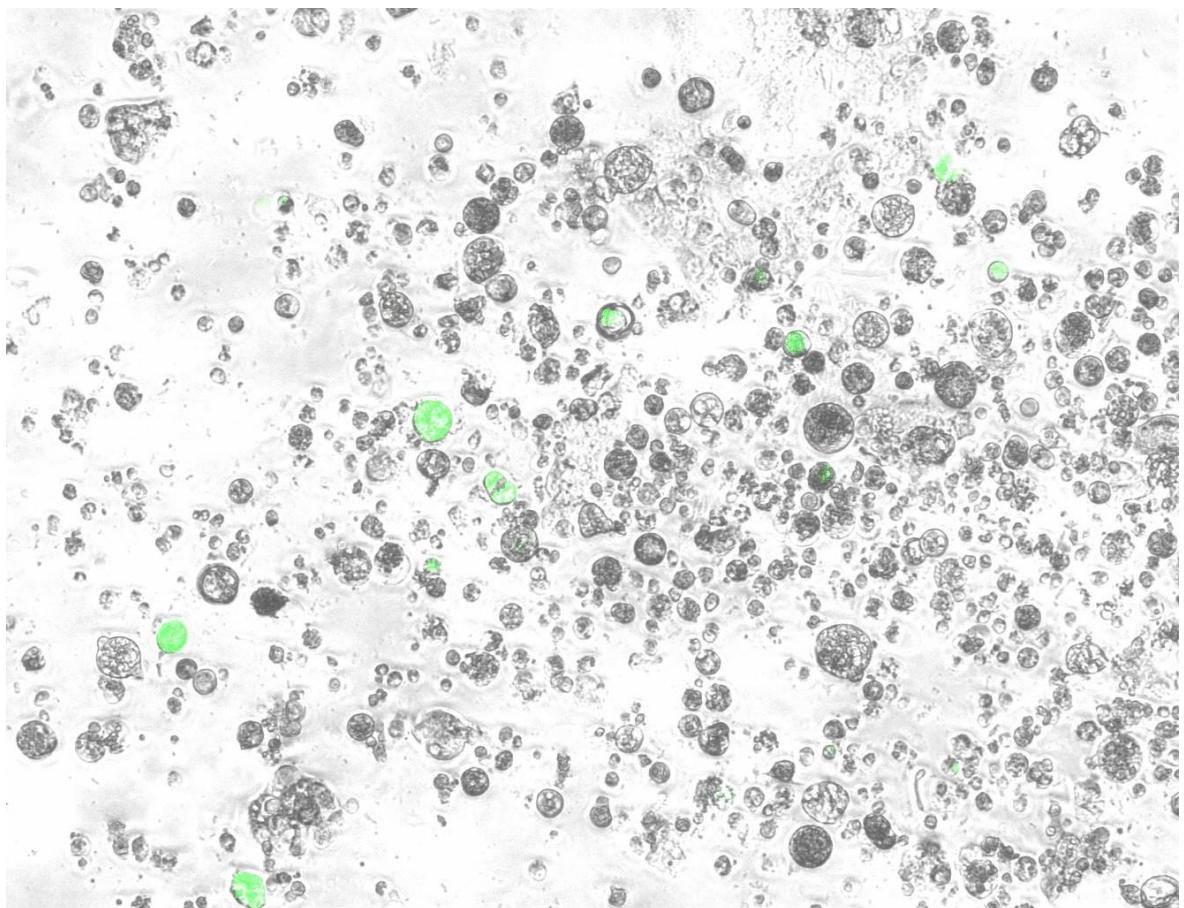


Figure 67 fluorescence microscopy of GFP nucleofected cells. Green fluorescence identifying successfully transfected cells. Large cells with poor morphology demonstrating cell stress/death. x40 magnification (Nikon Diaphot).

6.3.6 The expression of *CSDE1* is high in single cancer cells from human OAC tumours.

From my prior analysis presented in **chapter 5**, results suggested *CSDE1* expression in OAC may be localised to cancer cells as a potential mechanism of tumour immune evasion by downregulating MHC class I expression. However, to qualify this finding I sought to interrogate single-cell RNA sequencing data to assess the expression of *CSDE1* in OAC cancer cells within the tumour microenvironment, identified by canonical markers used in a prior study of OAC, compared to comparative normal healthy cells (615). To first validate this, I visualised the expression pattern of *CSDE1* expression finding a subset of cancer cells and T cells (Tregs) from both gastric cancer and OAC appeared to express *CSDE1* in higher quantity as compared to normal gastric cardia cells (**See Figure 68**).

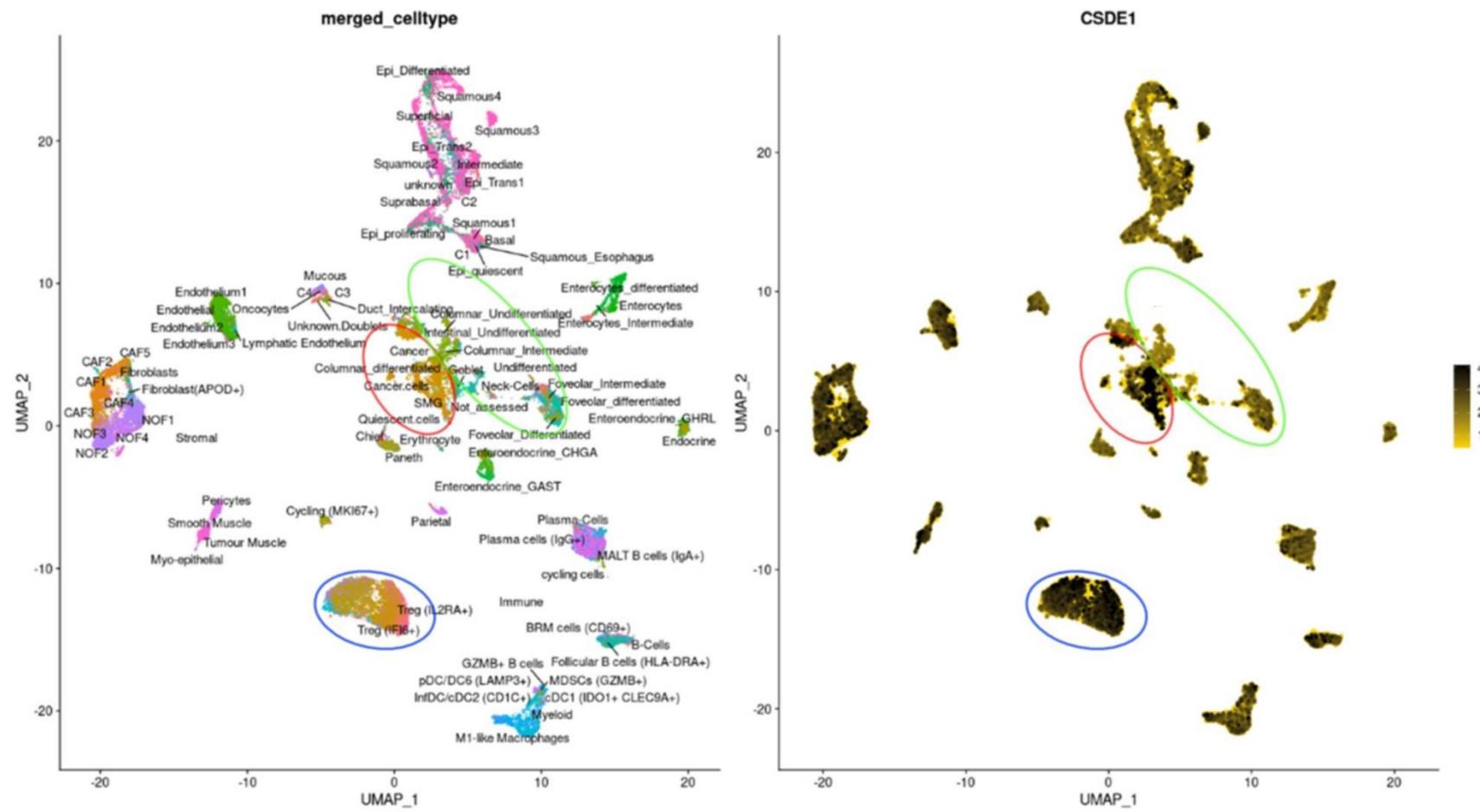


Figure 68 UMAP of single cell CSDE1 gene expression in OAC and gastric tumours. Left: Cell type labelled UMAP. Right: *CSDE1* gene expression at the single cell level. Single-cell gene expression level indicated by the colour bar. Red circle: Cancer cells. Green Circle: Normal healthy cells. Blue circle: T cells.

After identifying *CSDE1* expression in gastric and OAC cancer cells appear greater than normal gastric cardia cells, I sought to assess the *HLA-A* and *CSDE1* mRNA co-expression among gastric and OAC tumours. Here I identified co-expression of *HLA-A* and *CSDE1* in OAC and gastric cancer cells, though a small proportion of these cells possessed high *CSDE1* with low *HLA-A* mRNA expression or low *CSDE1* with high *HLA-A* mRNA (See Figure 69). Interestingly, T cells appeared to co-express *CSDE1* and *HLA-A* highly, while differentiating to differentiated squamous cells mostly possess a high *CSDE1* accompanied with low *HLA-A* mRNA expression (See Figure 69).

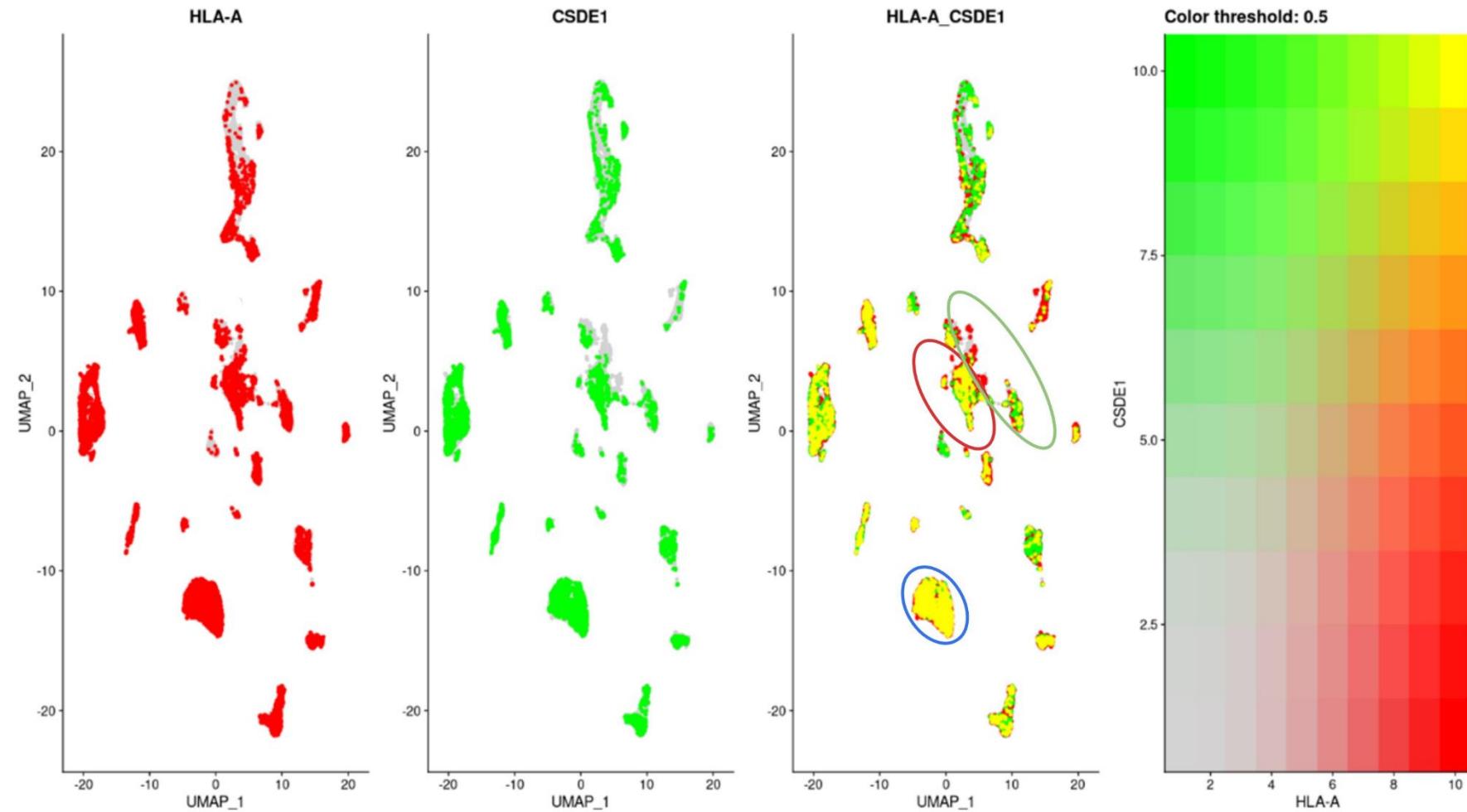


Figure 69 UMAP of single cell *CSDE1* and *HLA-A* gene co-expression in OAC and gastric cancer tumours. Green: *CSDE1* high and *HLA-A* low expression. Yellow: *CSDE1* HLA-A co-expression. Red: *HLA-A* high and *CSDE1* low expression. Red circle: Cancer cells. Green Circle: Normal healthy cells. Blue circle: T cells.

Additional analysis observing the co-expression of *CSDE1* and *HLA-A* in only OAC cancer cells revealed a high proportion of OAC cancer cells possessed high *HLA-A* and low *CSDE1* mRNA expression which cluster together, with the remaining cells mostly exhibiting co-expression both genes excluding a small number of cells with high *CSDE1* and low *HLA-A* mRNA expression (**See Figure 70**).

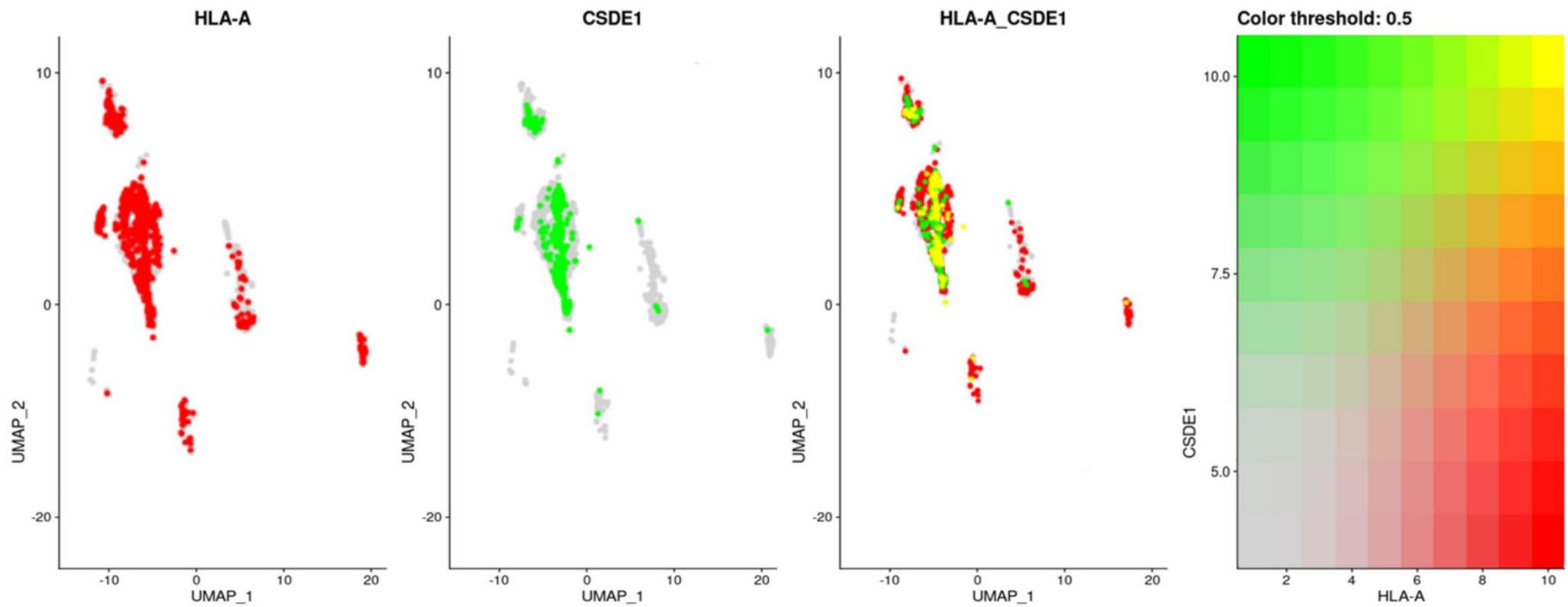


Figure 70 UMAP of single cell *CSDE1* and *HLA-A* gene co-expression in OAC cells. Green: *CSDE1* high and *HLA-A* low expression. Yellow: *CSDE1* *HLA-A* co-expression. Red: *HLA-A* high *CSDE1* low expression.

Following the analysis demonstrating OAC cancer cells possess a high ratio of *CSDE1*: *HLA-A* mRNA ratio I decided to compare the cells with comparative normal healthy cells and the distribution of ratio across cells within tumour samples. Here I observed a distinct shift towards high *CSDE1*: *HLA-A* mRNA ratio, furthermore, the *HLA-A*:*CSDE1* ratio in OAC cells was significantly greater than differentiated foveolar cells (**See Figure 71**). Interestingly, OAC cells appear to possess a higher proportion of cells with extremely high *CSDE1*: *HLA-A* mRNA ratio (>10) which are in greater abundance compared to comparative normal healthy cells and gastric cancer cells (**See Figure 71**).

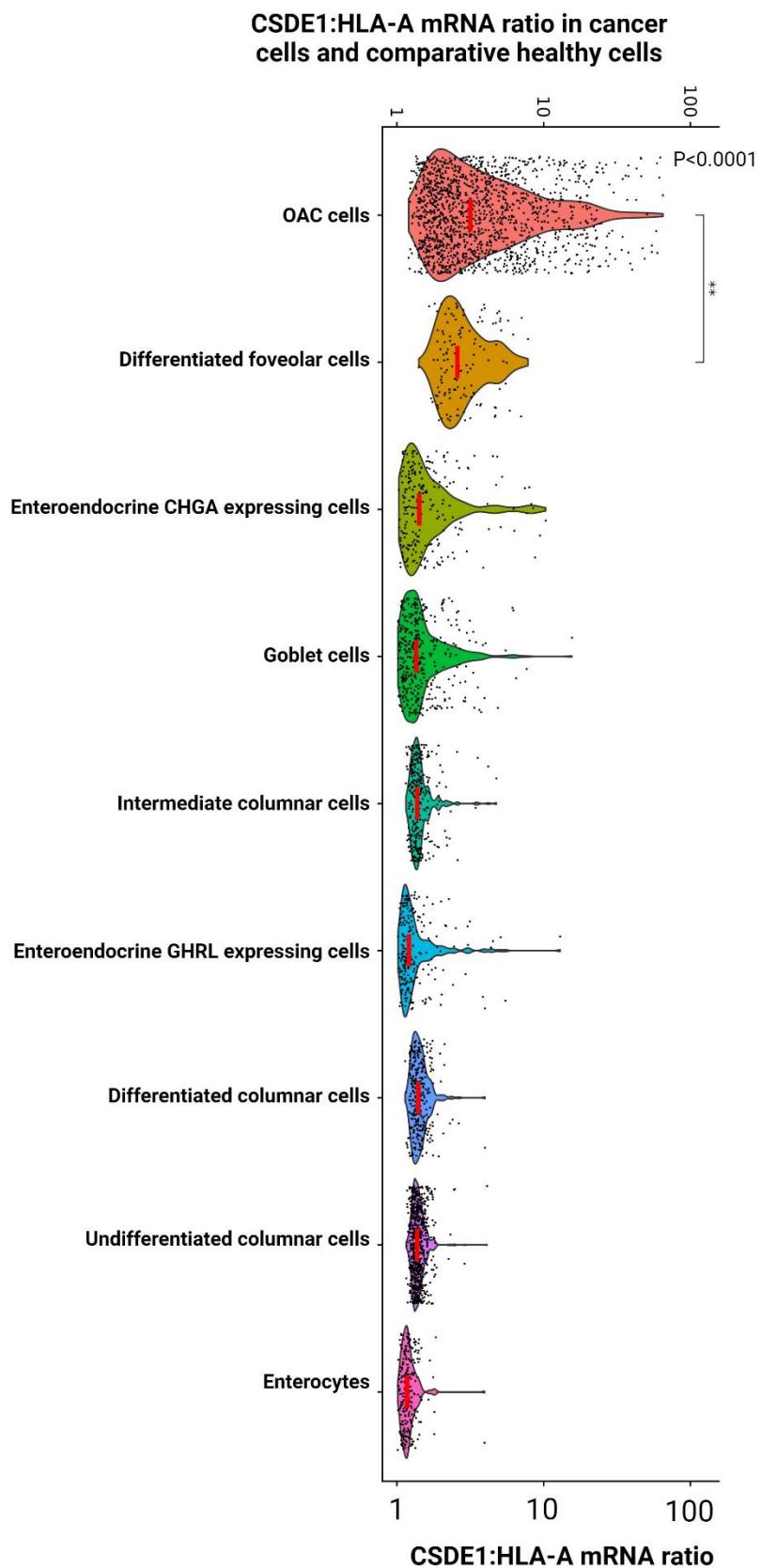


Figure 71 Violin plot of *CSDE1:HLA-A* mRNA ratio in OAC cancer cells with comparative normal healthy gastric cardia cells. Red line represents median ratio. Statistics: Wilcoxon signed-rank test, Kruskal–Wallis test. *CSDE1* expression in OAC cancer cells compared to comparative healthy cells available in appendix D.

Lastly, I explored the expression of *CSDE1*, MHC class I HLAs, interferon response gene sets (Alpha and Gamma) and regulators of the interferon signalling pathway among the low *HLA-A*:*CSDE1* mRNA ratio group (High *HLA-A*, Low *CSDE1* mRNA expression) and the high *HLA-A*:*CSDE1* mRNA ratio group (Low *HLA-A*, High *CSDE1* mRNA expression). Here I observed under low *HLA-A*:*CSDE1* mRNA ratio *CSDE1* expression was highly expressed in average expression and the percentage of cells (~75% of cells). This was paired with low expression of MHC class I HLAs (*HLA-A/B/C*), positive regulators of interferon signalling (*IRF1/9*), negative regulators of interferon signalling (*PTPN2*) and low expression of interferon response gene sets (Alpha and Gamma) (See Figure 72). The exact opposite expression profiles were observed in the high *HLA-A*:*CSDE1* mRNA ratio group with low *CSDE1* average expression with *CSDE1* being expressed by a smaller percentage of cells (<25% of cells) coupled with high expression of MHC class I HLAs (*HLA-A/B/C*), positive regulators of interferon signalling (*IRF1/9*), negative regulators of interferon signalling (*PTPN2*) and interferon response gene sets (Alpha and Gamma) (See Figure 72).

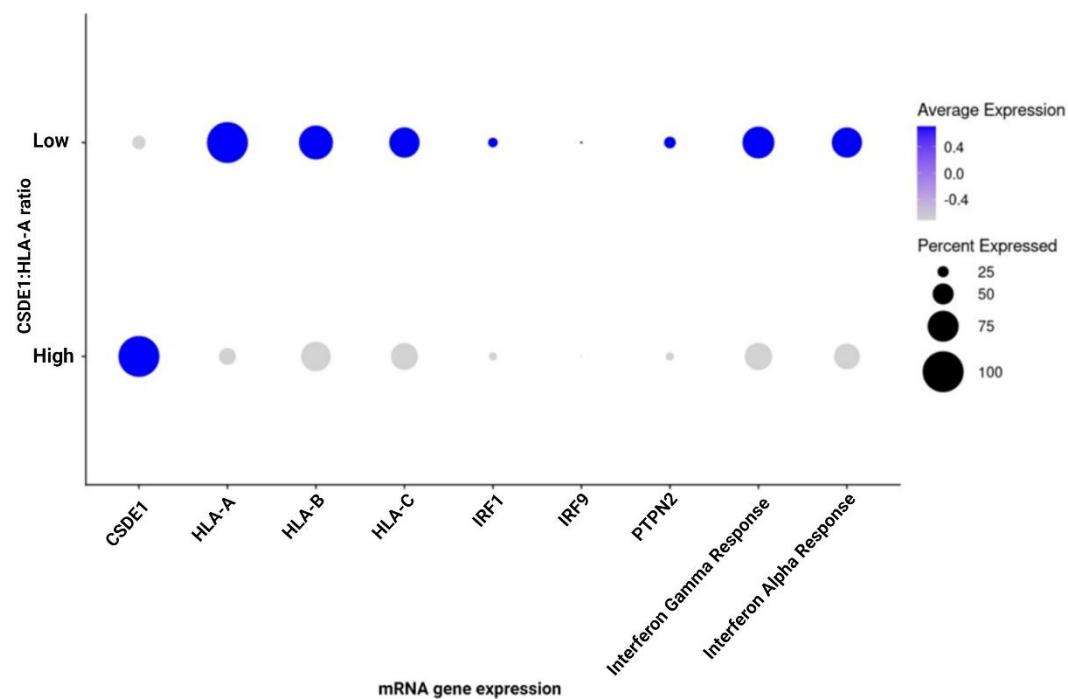


Figure 72 Dot plot of *CSDE1*, MHC class I, *PTPN2* and interferon signalling response genes for high/low *CSDE1*:*HLA-A* mRNA ratio. Circle size represents the precentage of cells expressing the gene and the blue colour scale indicates the level of scaled expression in each population of cells.

Following my analysis of *CSDE1*: *HLA-A* ratio among cancer cells, I sought to elucidate the heterogeneity of *CSDE1* and interferon signalling downstream gene expression with OAC cancer cells using clustering of single cell RNA-seq data; this approach would also allow me to further assess the relationship between *CSDE1* gene expression level and relevant interferon signalling downstream gene expression.

Here I observed 4 distinct clusters of OAC cells regarding the expression of these markers. The first observed cluster (1061/4304 cells) is driven by the high *CSDE1* expression, notably this cluster lack *IRF9* and *IFN- α/γ* response gene expression but retains the expression of *HLA-A/B/C* (See Figure 73). The second cluster (1459/4304 cells) possesses a lack of *CSDE1* and *IRF1* expression with a notable reduction in the expression of MHC class I HLAs and *IFN- α/γ* response gene expression (See Figure 73). The third observed cluster (650/4304 cells) is driven by high *IRF1* expression; interestingly, the expression of *CSDE1* is lower in the third cluster compared to the first cluster, yet no observable increase in the proportion of cells expressing MHC class I HLAs and *IFN- α/γ* response genes was identified (See Figure 73). The final cluster (1134/4304 cells) have two distinct subclusters; firstly, a high *IFN- α/γ* response gene expression subcluster of which some cases possess high *IRF1* expression with lower *CSDE1* expression compared to cluster 1. The second subcluster contains lower *IFN- α/γ* response gene expression, lacks *IRF1* expression, but retains *HLA-A/B/C* expression (See Figure 73). Interestingly, the heatmap clustering analysis found cancer cells from the same patient of origin did not predict cluster assignment, suggesting shared transcriptional states in OAC cancer cells (See Figure 74).

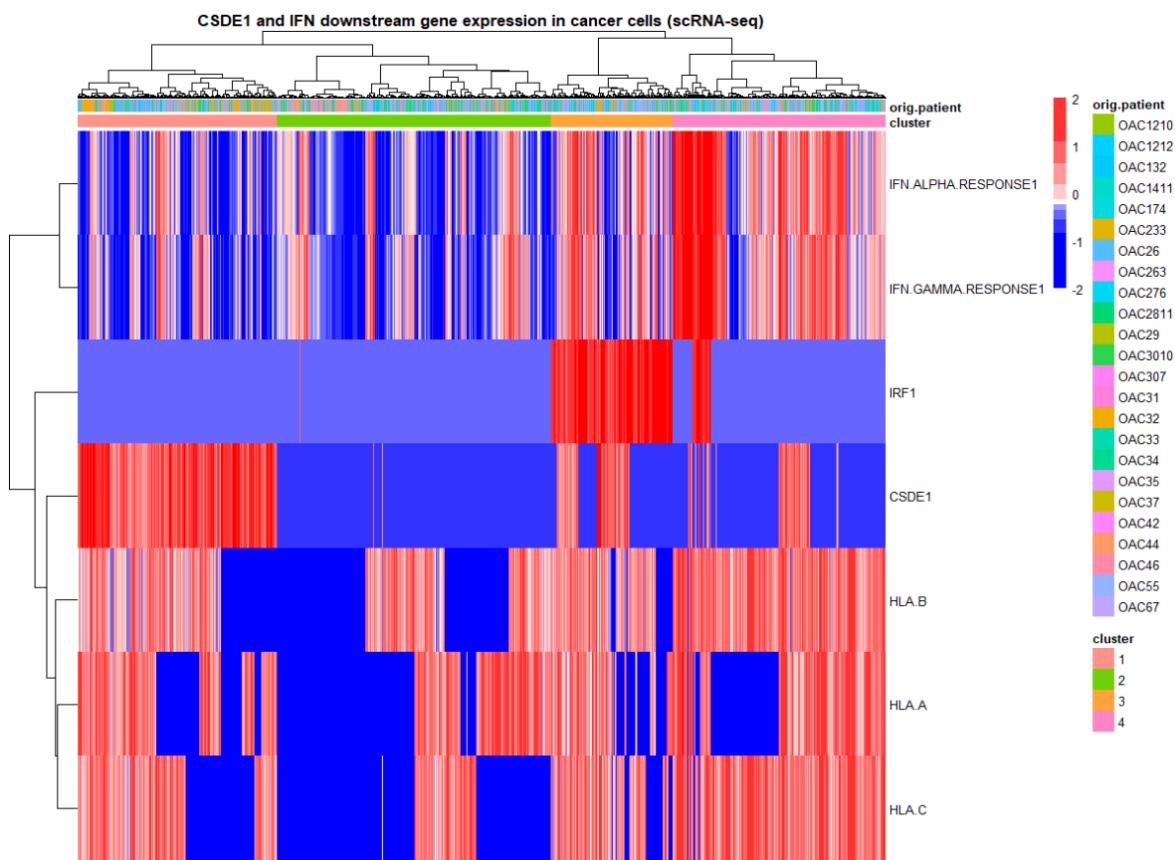


Figure 73 Heatmap of single cell RNA-seq TPM normalised expression of *CSDE1*, MHC class I genes and interferon response genes. Z-scored, maximum distance, Ward.D2 linkage with top annotation of patient of origin and cluster.

Observing expression of *CSDE1* and APM at the patient level using scRNA-seq analysis of OAC cancer cells revealed 4 distinct clusters among patients. The first cluster (6/24 patients) demonstrated low *HLA-C* expression, yet two subclusters (3 per subcluster) arose characterised by either high *CSDE1* and low IFN-Gamma/Alpha response gene or low *CSDE1* and high IFN-Gamma/Alpha response gene expression (See Figure 74). Cluster 2 (6/24 patients) was driven by low *CSDE1* and high expression of IFN-Gamma/Alpha response and MHC class I HLA gene expression; the third cluster (5/24 patients) represented high *HLA-B/C* gene expression and retained moderate expression of *CSDE1*, but with a notable lower *HLA-A* gene expression (See Figure 74). The final cluster (7/24 patients) represented the potential of *CSDE1* to inhibit the promotion of gene downstream of interferon signalling, characteristically expression high amounts of *HLA-A* and *CSDE1* mRNA yet possessed little expression of IFN-Gamma/Alpha response genes and *HLA-B* (See Figure 74).

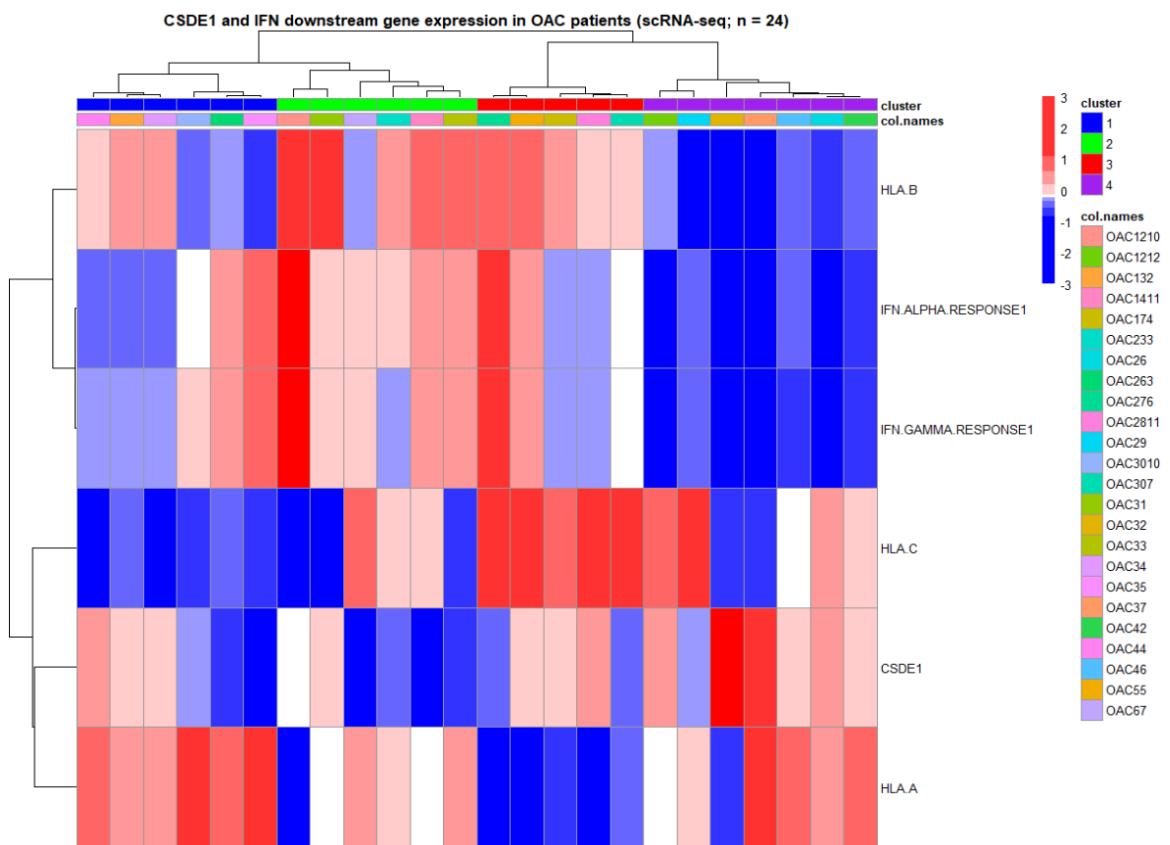


Figure 74 Heatmap of single cell RNA-seq (patient level) TPM normalised expression of CSDE1, MHC class I genes and interferon response genes. Z-scored, correlation distance, Ward.D2 linkage with top annotation of patient of origin and cluster.

Moving forward, I assessed the expression of *CSDE1* in precancerous lesions from gastric and oesophageal adenocarcinoma (Barrett's oesophagus) over intestinal metaplasia and dysplasia. Using a single cell RNA-seq approach I observed the proportion of cells expression higher levels of *CSDE1* mRNA increased with the extensity of intestinal metaplasia and dysplasia (See Figure 75). Interestingly, the expression of *CSDE1* was significantly greater in cases with higher grade dysplasia and intestinal metaplasia compared to cases with lower grade dysplasia and intestinal metaplasia (See Figure 75).

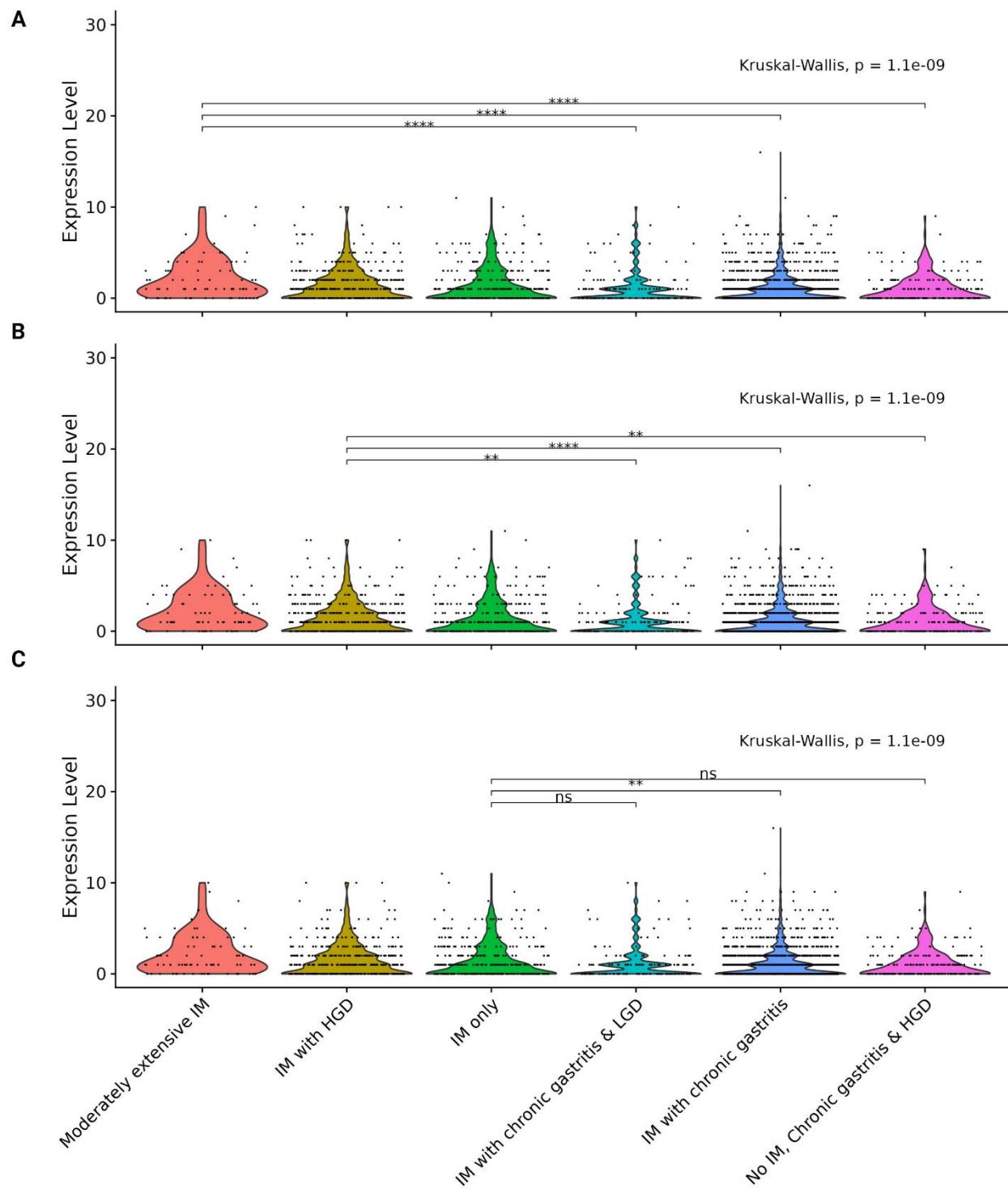


Figure 75 Violin plot of CSDE1 mRNA expression in precancerous lesions (Barrett's oesophagus). A: Pairwise statistical comparison of moderately extensive intestinal metaplasia (IM). B: Pairwise statistical comparison of intestinal metaplasia with high grade dysplasia (HGD). C: Pairwise statistical comparison of intestinal metaplasia.

6.4 Discussion

6.4.1 CSDE1 expression is a key regulator of MHC class I expression by modulating STAT1 signalling in FLO-1 OAC cells.

The premise of this investigation was to explore the impact of *CSDE1* on MHC class I expression via modulation of the JAK/STAT1 signalling pathway. The rationale of this experiment arises from prior publication demonstrating *CSDE1* was capable of stabilising TCPTP which acts to dephosphorylate STAT1 preventing MHC class I promotion in melanoma cells. This study moved forward to suggest *CSDE1* and its respective effect on MHC class I expression could be present in other cancers including gastric cancer, thereby I explored this mechanism in the context of OAC.

Firstly, I hypothesised the expression of *CSDE1* would perturb STAT1 signalling in OAC, finding *CSDE1* expression in OAC, negatively correlated with STAT1 signalling (pSTAT1/STAT1 ratio) and *CSDE1* knockdown increased STAT1 signalling in FLO-1 cells. This finding reflects the research which inspired this investigation, which reported *CSDE1* expression was a key negative regulator of STAT1 signalling via the stabilisation of TCPTP, suggesting this mechanism is active in OAC cells (145).

Interestingly, a significant increase in *HLA-A/B* and an observed non-significant increase in *HLA-C* mRNA expression was observed in *CSDE1* knockdown OAC cells, further supporting the mechanism of *CSDE1* acting to hinder the expression of MHC class I HLAs by modulating the JAK/STAT signalling pathway as prior observed in melanoma (145).

Unfortunately, the pSTAT1/STAT ratio did not correlate to MHC class I HLA protein expression, nor did any significant increase in MHC class I HLA protein level occur in *CSDE1* knockdown OAC cells, however, this may relate to the low number of repeats, as well as, the long half-life and endosomal recycling of MHC class I HLAs (622, 623). Conversely, the level of *CSDE1* protein did negatively correlate to HLA-ABC protein expression, which could suggest an alternative mechanism by which *CSDE1* may hinder MHC class I HLA protein expression. Unfortunately, at this stage it is unknown whether *CSDE1* impact MHC class I HLA mRNA stability, though *CSDE1* has prior demonstrated association to protein homeostasis and is capable of post-transcriptional repression through translational inhibition by binding to Pumilio homolog 1 (PUM1) (624).

To eliminate this possibility, future investigation should explore the role of *CSDE1* in post-transcriptional/translational repression of MHC class I HLAs.

I did not have time to explore the impact of *TCPTP* expression on *STAT1* signalling, with the original study suggesting *CSDE1* is capable of binding *PTPN2* mRNA for stability, this leaves future work to explore directly *TCPTP* expression on *STAT1* signalling. Also not explored in this investigation is the prior published role of the histone lysine methyltransferase, *SMYD3*, which is responsible for the upregulation of *CSDE1* expression which has also been published to be linked to the expression of High Mobility Group AT-Hook 2 (*HMGA2*). This is involved in the maintenance of cancer stem cell properties and regulates oral squamous cell carcinoma development and progression (625, 626). This further complicates the potential of *CSDE1* to impact survival outcomes, with further testing to determine if the survival differences prior observed are due to perturbed *STAT1* signalling resulting in lower MHC class I expression or by H3K4me3-mediated *HMGA2* transcription (627).

6.4.2 Localisation of *CSDE1* mRNA expression in patient tumour samples

From my prior analysis presented in **chapter 5**, *CSDE1* protein expression appeared to be localised to cancer cells and T cells according to an expert histologists assessment and Superpixel classification of tumour histology (**Figure 44**). To confirm the localisation of *CSDE1* expression to OAC cancer cells and T cells an analysis of single cell OAC tumours was performed and found *CSDE1* expression is greater among OAC cancer cells compared to comparative normal healthy cells. Additionally, *CSDE1* expression was greater among Tregs within tumour microenvironments, providing further useful information on top of my previous work. Furthermore, the mRNA expression ratio of *CSDE1*: *HLA-A* in cancer cells appeared to be shifted towards a higher ratio compared to normal comparative healthy cells. These findings suggest *CSDE1* expression may be somatically overexpressed in OAC cancer cells, potentially by epigenetic modification of the *CSDE1* locus mediated by *SMYD3* as prior published, yet this mechanism remains unproven in OAC at this stage (321). Interestingly, *CSDE1* and *HLA-A* appeared to be co-expressed in a large proportion of OAC and gastric cancer cells, but a distinct proportion of cancer cells did possess exclusive expression of *HLA-A* or *CSDE1* suggesting the *CSDE1*-mediated mechanism to downregulate the expression of MHC class I may be overcome by interferon signalling as

demonstrated in melanoma cell lines which overexpress *CSDE1* (321). This suggestion is supported by evidence in the high *HLA-A: CSDE1* mRNA expression group possessing greater expression of interferon response gene sets whilst possessing similar expression of positive regulators of interferon signalling (IRF1/9). However, it is important to note the significant limitation in using single-cell datasets, as they do not always sample all transcripts in a cell and may produce false negatives in read counts.

Intratumoural heterogeneity was well represented using heatmap clustering, where I uncovered 4 distinct clusters of cells expressing *CSDE1*, *IRF1*, MHC class I HLAs and interferon response genes. Interestingly, these clusters represented a high marker gene expression heterogeneity. The first cluster was driven by high *CSDE1* expression, which also possessed a lack of *IRF1* expression suggesting the high *CSDE1* expression could result in reduced promotion of *IRF1* expression in these cells as suggested in the melanoma study of *CSDE1*. Yet, the expression of MHC class I HLAs was retained in these cells. However, as noted in the prior study the mechanism of *CSDE1* negative regulation of interferon downstream promotion of gene expression may be overcome with high interferon microenvironments which are present in OAC (321). The second cluster lacked *CSDE1*, *IRF1* and interferon response gene expression with a observable lower proportion of cells highly expressing MHC class I HLAs; this cluster does not support the proposed mechanism discussed in the literature of *CSDE1* as a master regulator of MHC class I expression, however, an alternative explanation could be made these cells are not under extensive interferon signalling producing this cellular phenotype (321). A third cluster further supports the suggestion the lack of *CSDE1* expression in OAC tumour cells may result in greater JAK/STAT1 signalling to allow for increased *IRF1* promotion by possessing high *IRF1* expression, MHC class I HLA and interferon response gene expression and lower *CSDE1* gene expression compared to the first cluster forming the opposite effect observed in the first cluster (321, 628). The final cluster presented with two interesting subclusters, the first small subcluster had a notable increase in interferon response gene expression with a number of cells possessing high *IRF1* expression, these cells could be undergoing high interferon signalling to result in a promotion of this gene expression profile (628, 629). The second subcluster was observed to possess lower interferon response gene set expression compared to the first subcluster and lacked *IRF1* expression yet retained MHC class I HLA expression excepting several cells lacking *HLA-A* expression. Cells belonging to this subcluster would require further investigation to determine the factors such as the

level of interferons in the tumour microenvironment which may influence their expression profile of genes downstream of interferon signalling (628).

Additional analysis was conducted to explore the expression of *CSDE1* in premalignancy i.e., Barrett's oesophagus; here I observed an increased proportion of cells expressing higher levels of *CSDE1* with increasing intestinal metaplasia and dysplasia. This result suggests that *CSDE1* overexpression could be somatically acquired over dysplasia in premalignant conditions as an immune evasion mechanism. However, this suggestion would require further investigation to fully elucidate the stage of pathogenesis where *CSDE1* overexpression is acquired to produce inhibition of the JAK/STAT1 pathway downstream of interferon signalling which is responsible for the promotion of the MHC class I HLA locus. Furthermore, *CSDE1* has demonstrated other roles in cancer such as the induction of oncogene-induced senescence to produce senescence-associated secretory phenotype cells which help form the tumour microenvironment niche, thereby this raises a new question on the role of *CSDE1* in tumorigenesis of OAC.

Overall, these findings support the hypothesis of *CSDE1* overexpression may be somatically acquired to evade anti-tumoral immunity during the pathogenesis of OAC tumours. However, this leaves several new research questions including if the mechanism OAC exploits to overexpress *CSDE1* involves *SMYD3*-mediated epigenetic modification of the *CSDE1* locus or by an alternative mechanism, and the exact stage of pathogenesis of OAC which exploits *CSDE1* gene expression.

6.4.3 Limitations/future work

My initial plans in this investigation were to employ multiple OAC cell lines (OE33, FLO-1 and MFD-1) to overexpress and knockdown *CSDE1* gene expression and monitor the impact on the expression of MHC class I HLAs using multiple biological replicates, which unfortunately was unachievable due to candidature time constraints. As alluded to in my discussion, a significant limitation of this investigation is the sparse number of repeat experiments, future investigations should aim to validate these results using a larger number of repeats and using multiple OAC cell lines. Additionally, due to infection of cell culture, the attempt to overexpress *CSDE1* via nucleofection of OE33 cells failed, this approach should be repeated to successfully over-express *CSDE1* in OE33 cells to demonstrate an opposing effect on STAT1 signalling and MHC class I HLA expression compared to *CSDE1* knockdown.

A further limitation is presented in the long half-life of MHC class I HLA proteins which may mask the potential of *CSDE1* expression hindering the expression of MHC class I HLA protein expression. To prevent this factor impacting results a knockout *CSDE1* OAC cell line should be explored which has shown success in pancreatic cancer cell lines (630). Lastly, the role of *CSDE1* in post-transcriptional/translational modification should be ruled out for MHC class I expression using immunoprecipitation experiments. This study also cannot confirm the mechanism exploited by OAC to overexpress *CSDE1* to achieve immune evasion, with future work this could be explored to investigate the role of *SMYD3* in epigenetic modification of the *CSDE1* locus as prior published in melanoma (321). Lastly, although *CSDE1* overexpression in OAC appears somatic due to comparisons with normal healthy cells of the tumour microenvironment further research should be conducted to explore the stage of pathogenesis *CSDE1* overexpression occurs at, this could focus on comparing the level of *CSDE1* expression between healthy tissue, Barrett's oesophagus and over OAC transformation and progression.

6.4.4 Conclusion

Overall, my results in combination with the known literature suggests the expression of *CSDE1* reduces the expression of MHC class I HLAs in OAC, mirroring prior publications in melanoma. However, further investigation with a larger sample size and multiple cell models with the addition of investigating the relationship between *CSDE1* and TCPTP must be conducted to fully support the mechanisms presence in OAC. Furthermore, the role of *CSDE1* in repressing MHC class I HLA expression by post-transcriptional and post-translational modifications external to this mechanism must be ruled out. Lastly, recent publication suggests the expression of *HMGA2* resulting in poorer survival outcomes in oral squamous cell carcinoma is partly mediated by *SMYD3*, akin to *CSDE1* with brief analysis of OAC cell lines suggested *CSDE1* and *HMGA2* are both regulated by *SMYD3* warranting further investigation to deconvolute the impacts of *SMYD3*-mediated gene expression.

Chapter 7 General Discussion

Prior to my research, antigen presentation machinery (APM) had been published to associate to survival outcomes and immune composition in multiple cancer types including oesophageal squamous cell carcinoma, as well as, being noted as important for high immunotherapy efficacy, yet the role of APM defects in OAC patient outcomes and immunity was not well reported (631-635). Thereby, the ultimate focus of this research was to elucidate the landscape of APM genomic defects in OAC and assess the relationship between these defects on the immune cell populations of OAC and patient outcomes.

The first aim was to characterise the genetic and transcriptomic landscape of antigen processing machinery. Here I used a datamining approach in large publicly available OAC datasets containing mutation, copy-number, epigenetic and RNA sequencing data to assess the frequency of somatic aberration among an expertly curated candidate list of APM genes to determine the value of APM genomic defects on patient outcomes.

A further research question aimed to investigate the association between genomic defects in APM genes, APM gene expression regulators and immune compositions in OAC to determine whether recurrent APM genomic defects may alter the tumour immune microenvironment. This involved the employment of digital cytometry methods to predict the immune composition of OAC patient samples, which in turn were dichotomised by prevalent APM genomic defects to statistically identify significantly altered tumoral immunity due to the respective defect in APM. Furthermore, using this approach allowed for digital cytometric immunophenotyping of OAC RNA sequencing data which had not been performed prior to my work.

A follow-up research question focused on investigating altered APM protein expression in OAC tissue. To reach this goal, immunohistochemistry of a large TMA cohort sourced from Southampton General Hospital was stained for prognostically significant APM proteins, an APM regulator (*CSDE1*) and T cell markers followed by digital pathology analysis. Combined this analysis allowed for validation of prior findings and spatial assessment of candidate APM protein expression in OAC tissue.

My final goal was to assess the role of RNA binding protein, *CSDE1* in the regulation of MHC class I expression in OAC. Achieving this goal, I employed cellular models and assays to knockdown *CSDE1* and measure MHC class I gene and protein expression in comparison to isogenic control samples.

In my studies I identified several major headlining findings. While I demonstrated within the landscape of APM genomic defects, mutation presented as infrequent and copy-number differences on Chromosome 6 were incapable of being deconvoluted to a single gene effect. I did identify patterns of APM gene expression that possessed an association to altered survival outcomes using multivariate analysis. A highly unexpected result was found with Tapasin components such as *TAPBPL* which in univariate analysis demonstrates low expression associated to shorter CSS, yet high expression corresponded to greater risk of recurrence. This presents an interesting contradiction in the role of *TAPBPL* in survival outcomes to be discussed further here. Additionally, low *TAPBP* expression in univariate and multivariate analysis associated to increased risk of recurrence, further complicating the potential role of TAPASIN components in OAC survival outcomes.

Follow-up analysis also demonstrated these prognostically significant APM gene expression profiles possess associations with altered immune composition including *HLA-A/B/E* and *CSDE1*, suggesting the differing survival outcomes may be mediated by APM gene expression driven altered immunity. Interestingly, CD8+ T cells were observed in higher abundance with greater *HLA-E* expression, yet increased *CALR* and *CSDE1* corresponded lower CD8+ T cell abundance. These results could indicate HLA-E restricted CD8+ T cells may be present in OAC and play a significant role in anti-tumoral immunity. These results were confirmed by protein level analysis of my TMA cohort using IHC and digital pathology methodology which provided validated insight into the specific roles of each APM gene candidate in determining the potential function of the tumour immune microenvironment of OAC. Additionally, IHC methodology allowed for observation of APM protein staining across OAC tissue, finding HLA-ABC stained across the entire tissue, HLA-E/Class II stained either stained specifically to cancer cells or diffusely across the tissue and *TAP1* staining was diffuse across OAC tissue. Complete loss of APM staining was not observed, yet only a small subset of OAC tumours possessed relatively high HLA-

ABC/E/Class II expression with most patients exhibiting relatively lower protein expression of these markers. Additionally, approximately 20% of patients presented with low TAP1 protein expression which corresponded to lower CD3+ T cell abundance, which was not observed in mRNA analysis this could potentially relate to TAP1 protein being representative of TAPAPSIN dimers (TAP1/2) rather than mRNA of *TAP1* alone.

My *in vitro* investigations determined *CSDE1* expression in OAC acts as a negative regulator of MHC class I expression demonstrating knockdown of *CSDE1* produced greater *HLA-A/B/C* expression within an OAC cell line; furthermore, *CSDE1* expression was found to be highly localised to OAC cancer cell and Treg cell populations within the tumour microenvironment, with cancer cells displaying greater *CSDE1* gene expression compared to comparative normal healthy cells. These results provide confirmatory evidence that the mechanism by which *CSDE1* downregulates MHC class I HLA molecules is conserved between melanoma (as previously published by others) and OAC (321). Yet my results also provide greater granularity on the distribution of *CSDE1* expression within the individual cell types of OAC tumours and suggests *CSDE1* overexpression may yield lower tumour immunogenicity to produce effective immune evasion from CD8+ T cell responses.

7.1 Genomic disruption of APM in OAC and its impact on immune response and clinical prognosis.

In **Chapter 3**, I sought to elucidate the landscape of genomic defects within APM of OAC and ascertain the clinical value of APM genomic defects frequently observed among a large multi-omics OAC cohorts. Here I produced new findings which contributes to the current understanding of antigen presentation in OAC. First of which I explored the mutation incidence of APM genes and associated regulators, here I observed a low frequency of mutation among my APM gene candidates with the highest incidence of mutation occurring in *CD1A/C/D* (in only 2-3% of the cohort), this presents an interesting argument that APM defects are not frequently targets for somatic mutation in OAC. However, this finding may be in contradiction to other studies perspectives including a study of 551 OAC tumours which revealed *B2M* as a recurrently mutated driver gene, this could potentially relate to either the reduced sample size or potentially sampling bias available for my study or the method employed to determine recurrent mutation in the 551 study, which values the frequency of multi-hit mutations (242).

Following this analysis, I explored the copy-number incidence among my APM gene candidates, finding approximately 7.6% of OAC tumours possessed multiple deleted/copied APM genes.

Exploring the copy-number segments found high genomic copy number segment complexity over chromosome 6 proximal to the HLA locus. The complexity of the OAC genome has prior been published and there are reports of *VEGFA* being frequently amplified (636). Interestingly, *VEGFA* is proximal to the HLA locus on chromosome 6 with copy number segments often overlapping *VEGFA* and APM genes of the HLA-locus suggesting APM amplification may be driven by *VEGFA* amplification (637). Unfortunately, the impact of copy-number status on individual APM genes on survival outcomes could not be conducted due to difficulties deconvoluting the co-copied/deleted genes.

Within **Chapter 3** and **4**, I explored the association between APM gene expression and altered survival outcomes and immune composition. Specifically, the expression of 12/18 MHC class I genes were associated to altered survival, for example, low expression of *HLA-A* was associated to shorter OS, CSS, DFS in univariate survival analysis, but did not withstand multivariate testing. Yet, the low expression of *TAPBP* resulted in shorter DFS in both univariate and multivariate survival analysis; interestingly, low *TAPBP* has prior been associated recurrence in triple-negative breast cancer and non-small cell lung cancer

(147, 638). This association potentially relates to poor immune control of tumours, evidenced by the downregulation of *TAPBP* being observed in multiple cancer cell lines including small-cell lung carcinoma, pancreatic carcinoma, colon carcinoma, head and neck squamous cell carcinoma, renal cell carcinoma cell lines and prostate cancer. This associates to low MHC class I cell surface expression, suggesting low *TAPBP* expression could result in poor immune recognition by CD8+ T cells (639). Additionally, restoring *TAPBP* expression in a lung carcinoma model produced greater immune infiltrate (CD4+, CD8+, CD11C+) and was associated to improved disease-free survival (640). These results confirmed MHC class I expression in OAC may impact survival outcomes, but required further analysis to determine whether these genes may elicit altered survival via modulating the immune composition and respective immune response. To explore this question, I used digital cytometry to deconvolute the immune signature of OAC and dichotomised patients by prognostically significant MHC class I gene expression. Using this approach, I observed the expression of 8/18 MHC class I genes associated with altered immunity, for example, the upper quantile of *HLA-E* expression possessed greater abundance of CD8+ T cell and monocytes compared to the lower quantile. These results suggest that the landscape of MHC class I APM component expression may be a predictor of survival outcomes in OAC via modulation of CD8+ T cell responses to presented tumour associated antigens.

However, this poses further questions on the abundance of these presented antigens due to the expression of MHC class I machinery in OAC which could be explored using immunopeptidomics focused on prognostically significant MHC class I gene expression presented in this study. This new knowledge also differentiates the impact of expression of specific MHC class I genes between OAC and other cancers, such as *CANX* expression in OAC was not prognostic, but it is prognostic in colorectal carcinomas; these findings may potentially be harnessed in future efforts in OAC prognostication and may also be explored to explain potential differences in response to immunotherapies between OAC and closely related cancers (641). Specifically, a pan-cancer analysis exploring the subtypes of cancer (i.e., adenocarcinomas and squamous cell carcinomas) to highlight differences in survival and immune composition associations may potentially explore the impact of APM genes in modulating the immune niches of cancer subtypes.

I identified 12/20 MHC class II candidate genes were also associated to altered survival outcomes, this included *HLA-DRA* of which lower expression associated to shorter OS, CSS and DFS in univariate analysis with shorter OS notwithstanding multivariate analysis. These results suggest the expression of MHC class II components within the OAC microenvironment by professional APCs may alter the immune microenvironment composition, by eliciting responses from CD4+ T cells for beneficial anti-tumoral immunity. Additionally, MHC class II expression represents a marker for increase infiltration of professional APCs, which have prior been reported as a positive prognostic indicator in OAC (642). Additionally, 9/20 MHC class II candidate genes expression associated to altered immunity, for instance, the upper quantile of *HLA-DPA1* expression contained greater abundance of Tregs, CD4+ T cells and macrophages compared to the lower quantile.

These results demonstrate MHC class II gene expression in OAC is a key regulator of immune composition, however, a complication forms with MHC class II expression, as the cell origin for expression cannot be determined using digital cytometry methods. Specifically, MHC class II expression can form a marker of professional APC such as macrophages, which in turn may affect immune composition by secreting recruitment factors or directly interacting with CD4+ T cells. Interestingly, the literature provides evidence OAC cells within highly inflammatory microenvironments may induce MHC class II expression, yet the impact of induced MHC class II expression on the surface OAC cells induced by extreme interferon signalling on the tumour immune microenvironment as not yet been elucidated (642). However, assessment of cancer cell specific MHC class II expression in lung adenocarcinoma correlated with response to anti-PD1 therapy and increased CD4+ T cell infiltration, suggesting this could be an interesting avenue for future research (251).

Lastly, I observed 4/5 APM gene expression regulators were associated to altered survival outcomes, including *CSDE1* of which high expression corresponded to shorter OS and corresponded with a significant decrease in CD8+ T cell infiltration. Overall, these results suggest APM gene expression regulators may impact APM expression in OAC cancer cells and elicit altered survival by modulating the immune composition, yet currently the relationship between APM gene candidate, immune composition and immune checkpoint gene expression is unknown.

In **Chapter 5**, I validated my findings from **Chapter 3 & 4** for HLA-ABC, HLA-E, HLA-Class II and TAP1 demonstrating low HLA-E, HLA-Class II and high *CSDE1* protein expression associated to shorter OS in univariate analysis with low HLA-Class II expression associating to shorter OS in multivariate analysis. Furthermore, the protein expression of these APM markers demonstrated altered immune composition, for example, low HLA-ABC protein expression possessed significantly lower CD3/8+ T cell abundance. These results validate my prior findings and suggests APM machinery expression is a key modulator of survival and the tumour immune microenvironment.

These findings expand upon prior published work from others which did not explore as large of a APM gene candidate list nor assessed these APM candidates for impact on

tumour immune composition, highlighting the landscape of APM gene expression as a key factor in survival outcomes potentially enacted by altering the cellular interactions of immune recognition within the tumour immune microenvironment. There may be translational value to the findings presented throughout this study, such as assaying APM protein expression in OAC biopsies to predict efficacy to immunotherapy as suggested in prior studies (643-646). This translational value for immunotherapy is evidence by investigation of melanoma APM expression demonstrating greater MHC class I and II expression corresponded to improved responses to checkpoint blockade therapies. Unfortunately, the value of APM expression in immunotherapy responses other cancers including OAC and gastric cancer have been elusive, yet more recent trials data may be assessed in future to address this knowledge gap (647).

There were several limitations within my analyses. Firstly, in combining the available clinical data I found significant lack of detail and missing clinical data leading to a restriction in the clinical model used in multivariate survival analysis and the sample size due to exclusion of samples lacking basic clinical data, a future prospective study may allow for higher quality metadata to include a greater quantity of patients within the models. Higher quality clinical data would be of great value as it would allow for the addition of further clinically accepted co-variates for survival outcomes in multivariate models such as tumour regression grading and histological tumour grading, which are known factors in survival outcomes (648, 649). Additionally, assessing bulk-RNA seq data did not allow for partitioning the expression into single cell populations, which could mask the impact of APM cancer cell expression, therefore future analysis using single-cell RNA-seq data could explore APM expression in OAC cancer cells compared to comparative normal healthy cells as demonstrated in a recent study (428).

Whilst some single-cell RNA-seq data was available in my laboratory for this type of investigation, due to a small sample size and lack of cases treated with immunotherapy (n=1) this detailed analysis could not be conducted. Using a single-cell approach could also explore the distribution of MHC class II expression in OAC cancer cells and assess the role of APM expression on other tumour cell populations such as cancer associated fibroblasts and the sub-populations known as antigen-presenting fibroblasts could be of particular interest (650). Furthermore, this study could not explore the influence of HLA-allotypes in OAC to determine if patients expressing specific polymorphisms/ allotypes of

HLA molecules can possess a significant survival advantage and/or altered immunity.

HLA-allotypes have been prior reported to affect the immunopeptidome, response to immunotherapy and survival outcomes in cancers including OAC (546, 651, 652). Lastly, by using IHC staining protocols, they produced restrictions in which markers which could be stained, for example, HLA-ABC was staining using a single antibody, preventing separation of impact of individual HLA-A/B/C proteins on survival and relationships with immune composition in OAC tumour tissue.

Overall, in addressing the landscape of antigen presentation genomic defects in OAC and association to survival outcomes and altered microenvironments, I have uncovered evidence to support my initial hypothesis, with the expression of key APM components corresponding to altered survival outcomes and immune composition, which in future could provide great translational value in the field of cancer immunotherapy.

7.2 Potential mechanism of dysregulated APM/ MHC in OAC – therapeutic axis?

Recent work identified gene knockout of *CSDE1* produced greater MHC class I expression in a melanoma cell model. The research paper proposed *CSDE1* enacts a role of a negative regulator of MHC class I expression via stabilisation of TCPTP, which functions as a tyrosine kinase able to dephosphorylate STAT1 (321). The importance of phosphorylated STAT1 (pSTAT1), is found within the JAK/STAT signalling pathway downstream of interferon receptors; *CSDE1* stabilised TCPTP dephosphorylates homodimers of pSTAT1 which prevents translocation into the nucleus (321). pSTAT1 homodimers are functionally important for the promotion of MHC class I HLA locus as they bind to GAS which promotes the expression of *IRF-1/9*. *IRF-1/9* binds the interferon response element to promote the expression of MHC class I HLA locus (653).

My initial bioinformatic analysis on *CSDE1* observed greater expression of *CSDE1* transcripts corresponded to shorter survival outcomes and altered immunity with lower CD8+ T cell abundance. Further investigation confirmed this relationship at the protein level in immunohistochemistry analysis of OAC tissue microarrays; strikingly, *CSDE1* protein expression correlated with the tumour cell abundance in OAC tissue suggesting *CSDE1* expression is localised to OAC cancer cells. To validate this finding single cell RNA-seq analysis of OAC and gastric tumours was conducted, finding *CSDE1* gene expression was notably greater in OAC and gastric cancer cells in comparison comparative normal cells in the OAC microenvironment. Furthermore, *CSDE1* expression was observed to be greater in Treg populations in comparison to effector T cells (CD4+/CD8+) and other cells of the OAC microenvironment (**Figure 68**); the reason this may occur in OAC is not currently clear. However, TCPTP plays significant roles in inhibiting pathogenic loss of Foxp3 driven by IL-6, thereby the increased expression of *CSDE1* in Tregs may play a role in sustaining the immunosuppressive function of Tregs in OAC, yet this hypothesis requires further investigation (654). Exploring the current literature, to date no analysis have confirmed *CSDE1* to be upregulated in OAC, thus these findings provide potentially the first insight into the expression of *CSDE1* within the OAC microenvironment.

After observing the localisation of *CSDE1* to cancer cells in the OAC microenvironment my later research efforts focused on measuring the impact of *CSDE1* expression on MHC class I HLA expression in a cell model and in human tumours profiled by single cell RNA-seq.

Firstly, knocking down *CSDE1* mRNA expression (siRNA) produced an increase in MHC class I HLA expression in a OAC cell model (FLO-1) including *HLA-A/B/C* at the mRNA and protein level with greater STAT1 signalling. This suggests *CSDE1* enacts a negative role in STAT1 signalling in OAC as prior identified in melanoma, involving *CSDE1* stabilising TCPTP to elicit dephosphorylation producing downregulation of MHC class I HLA expression (321).

Further evidence to support the prevalence of this mechanism was observed in single cell RNA-seq analysis, which identified high *CSDE1* and low *HLA-A* gene expression in OAC cancer cells in comparison to normal healthy cells in the OAC microenvironment. This indicates *CSDE1* plays a negative regulatory role in MHC class I expression in OAC cancer cells. Furthermore, *CSDE1* upregulation was noted during dysplasia of Barrett's oesophagus. This analysis suggests *CSDE1* somatic overexpression occurs during the pathogenesis of Barrett's oesophagus dysplasia towards OAC, however, the mechanism driving the overexpression from early pre-cancerous lesions through to OAC is unknown and requires further investigation.

Interestingly, mouse melanoma models suggest *CSDE1* may provide a therapeutic axis for the treatment of OAC with recent publications proposing a trap-ambush therapeutic method (585). This details the use of rhinovirus based oncolytic virus which produce a selective pressure towards *CSDE1^{C-T}* (Cytosine to Thymine) point mutations forming a proline to serine amino acid change at $\alpha\alpha 5$, as *CSDE1* is required for rhinovirus infection. The mutations produced from this selective pressure form escape-associated tumour antigens which can be recognised by CD8+ T cells. By providing vaccination specific to these escape-associated tumour antigens, cancer cells which mutate *CSDE1* to evade the oncolytic virus are thus targeted by primed T cells forming a trap-ambush therapeutic method (585).

Further investigation within a mouse melanoma model also demonstrated using immune checkpoint blockade with the CD200 activation receptor ligand (CD200AR-L) peptide was capable of enhancing the response to cancer cells harbouring *CSDE1* escape-associated tumour antigens (584). Subsequently, the analysis I present in my investigation of the role of *CSDE1* in OAC immunogenicity support this therapeutic approach could be effective in the treatment of OAC, due to the specific overexpression of *CSDE1* in OAC cancer cells in comparison to normal healthy cells of the OAC microenvironment. This would suggest selective pressure formed from an engineered Rhinovirus or Rhabdovirus derived oncolytic virus target Type I IFN-defective cells, such as *CSDE1* overexpressing OAC cells, would actively select for *CSDE1*^{C-T} mutation in OAC, which could further be treated by a vaccine specific for *CSDE1*^{C-T} antigens.

Throughout my investigation of the role of *CSDE1* in OAC immunogenicity several limitations arose which future experimentation should strive to address. Namely, due to time constraints and difficulty with OAC cell cultures I could not perform further cell model experiments. Firstly, the experiment plans included a *CSDE1* overexpression experiment, in which *CSDE1* plasmids were to be used to over express *CSDE1* to demonstrated opposing effects to the knockdown experiment perform, i.e., MHC class I HLA expression to fall with increased *CSDE1* expression level. This approach is standard in functional genomics to provide an insight into the specific function of a gene of interest; in this instance, demonstrating *CSDE1* as a negative regulator of MHC class I expression would be achieved by overexpressing *CSDE1*, whereas knockdown could demonstrate targeting *CSDE1* by a targeted therapy could result in increased MHC class I expression.

Secondly, to provide further confidence in my results, I had planned to conduct a series of *CSDE1* knockdown/overexpression experiments on different cell backgrounds providing further repeats to rule out results due to chance and the biological heterogeneity association with this mechanism in OAC.

Additionally, throughout my experiments I did not measure the impact of *CSDE1* expression on the level of TCPTP protein as conducted in prior publication in a melanoma model; performing this analysis would provide further confidence in the

proposed mechanism of action for *CSDE1* driven downregulation of MHC class I HLA expression via inhibition of JAK/STAT1 signalling.

Provided additional time, my investigations could have explored the impact of *CSDE1* knockout (e.g., with CRISPR) which could provide further insight than the conducted knockdown experiments conducted, due the limitations of siRNA knockdown methods such as off-target effects, incomplete knockdowns, and the impact of protein half-life (591).

A further limitation of the current analysis is its inability to measure immune responses due the expression level of *CSDE1* in OAC cancer cells, this potentially could be achieved either by *in-vitro* immune cell – cancer co-culture experiments in which *CSDE1* knockdown, wildtype and overexpressing OAC cancer cells could be exhibited to primed matching T cells; here immune activity could be assessed using flow cytometry for T cell activation markers. An alternative approach could use humanised mouse xenograft models to explore the impact of *CSDE1* expression on immune recognition of tumours measuring survival and T cell activation using either immunohistochemistry or flow cytometry. Lastly, prior publication observed the impact of *CSDE1* overexpression on reduced JAK/STAT1 signalling could be recapitulated under high IFN- γ stimulation elevated in gastrointestinal cancers (321, 655). As the OAC microenvironment is thought to be highly inflammatory this analysis should be conducted within OAC cell models, as these results could provide further insight into the impact of *CSDE1* expression in OAC tumours with highly inflammatory microenvironments (656).

7.3 Conclusion

In conclusion, this study has employed a bioinformatic data mining approach to elucidate the landscape of APM genomic defects, identifying the significance of APM gene/protein expression in producing altered survival and shaping the immune composition of the OAC tumour immune microenvironment. Furthermore, using a cell modelling and single cell RNA-seq approach, this project has demonstrated *CSDE1* may be a significantly overexpressed gene in OAC and corresponds to reduced expression of MHC class I genes. Thereby, these findings propose the immune composition of OAC is determined by APM gene expression in OAC, *CSDE1* is somatically overexpressed in OAC and may form an actionable target for a trap-ambush therapeutic approach.

7.4 Future directions

The further experimental work required to address the findings from the current project have been discussed at length in the individual chapter and general discussion sections. However, in addition to these possibilities, the results presented in this thesis open many avenues for further research in other areas including:

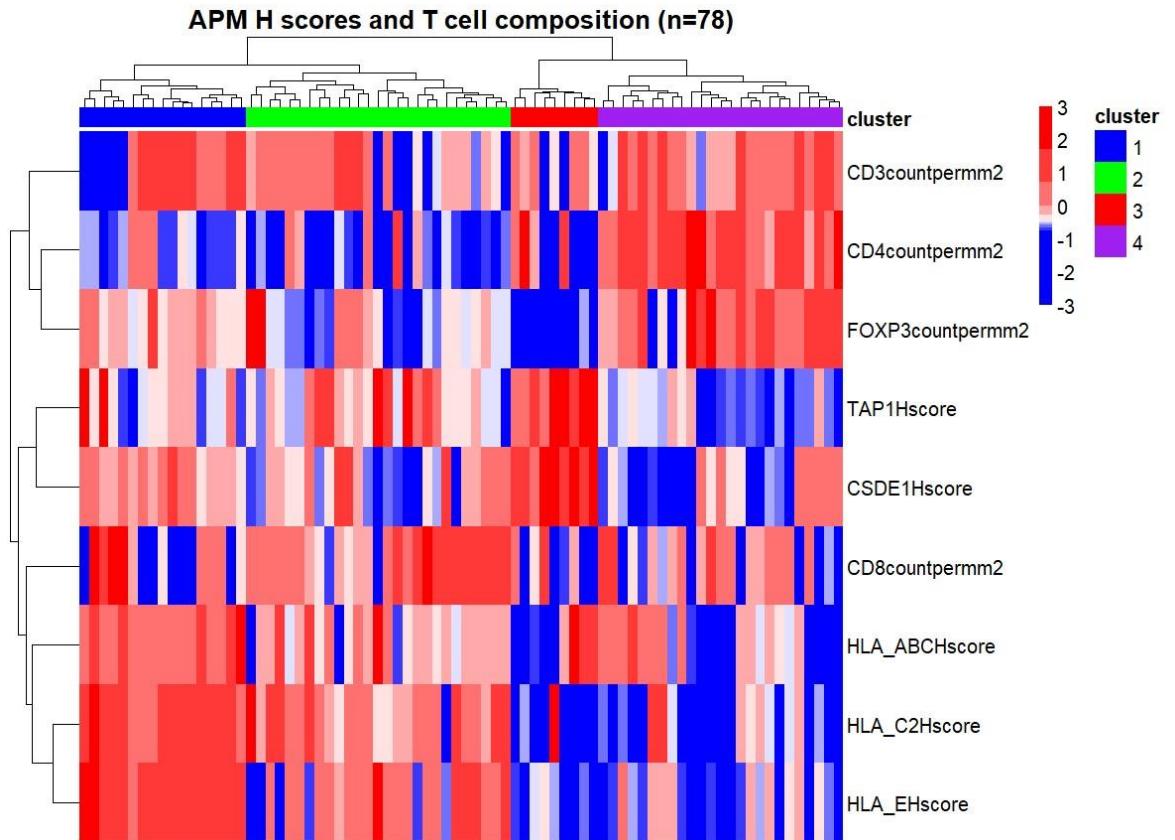
- Analysis of the effect of siRNA knockdown and overexpression of *CSDE1* in multiple OAC cell lines. Future analysis should repeat the experiment presented in **Chapter 6** and explore the impact of *CSDE1* overexpression. In addition to this investigation, providing insight into the role of *CSDE1* expression on the function of TCPTP to elicit perturbed MHC class I HLA expression should be conducted.
- Interestingly, evidence has arisen demonstrating OAC cells may express MHC class II molecules in the presence of high IFN- γ stimulation with this phenomenon being induced in the OE33 cell line (657). Yet no current analysis has attempted to observe the distribution of MHC class II expression in OAC cancer cells from tumours. This could be achieved using scRNA-seq approaches to determine the expression level of MHC class II genes within OAC cancer cells in low comparatively low inflammatory microenvironments compared to high inflammatory microenvironments.

- Determining the impact of APM gene expression and APM gene expression regulators in response to immunotherapy. A very recent publication employed the use of scRNA-seq in the investigation of a trial exploring response to immunotherapy in OAC (LUD2015-005; <https://ega-archive.org/datasets/EGAD00001009401>), with *ERAP2*, *HLA-DQA2*, *HLA-DOA*, *HLA-DMB*, *HLA-DPB1* and *HLA-DRA* being overexpressed in patients receiving checkpoint blockade therapy (658). Future investigation could access this data to explore the potential of prognostically significant APM gene expression, APM gene regulator expression and altered composition of the tumour immune microenvironment in predicting immunotherapy response.
- Determining the mechanism of *CSDE1* upregulation in OAC cancer cells. From prior publication, the upregulation of *CSDE1* gene expression was tied to the role of *SMYD3* in mediating H3K4 trimethylation of *CSDE1* locus, however, this mechanism was only explored in melanoma (321). Therefore, replicating the methods observed within the publication could help determine whether the same mechanism to upregulate *CSDE1* expression is present in OAC, or an alternative mechanism may be implicated (321).
- The effect of *CSDE1* on the tumour immune microenvironment has yet to be fully elucidated. Specifically, observed differences in immune composition due to *CSDE1* gene expression in bulk-RNA sequencing analysis was noted in my investigation. Yet, the impact of high *CSDE1* expression within OAC cancer cells on immune composition has not been explored by the analysis presented here. Therefore, further investigation focused on determining differences in immune composition due to OAC cancer cell expression of *CSDE1* using a combined scRNA-seq approach and multiplex immunohistochemistry may provide further insight into this associated impact.
- Within **Chapter 6**, high *CSDE1* expression was observed both OAC/Gastric cancer cells and Tregs. Observation of the literature did not identify publications elucidating the role of Treg specific expression *CSDE1*. Therefore, future analysis could explore further data analysis to provide further insight into the distribution

of *CSDE1* expression in T cell populations and may move forward to perform functional genomic investigation of the role of *CSDE1* in Treg populations.

- Lastly, the role of RNA-binding proteins in immunogenicity of oesophageal adenocarcinomas and precancerous conditions such as Barrett's oesophagus is currently unknown. Publications have demonstrated RNA-binding proteins such as Insulin-like growth factor 2 mRNA-binding proteins can play a negative regulation role in IFN β and IFN γ -stimulated genes within mouse melanoma models, while *CSDE1* has demonstrated a crucial role in producing a senescence-associated secretory cell phenotype which may alter immunosurveillance via secreted chemo/cytokines in the melanoma tumour immune microenvironment (659, 660). This propagates further research questions on the role of RNA-binding proteins in the regulation of tumour cell immunogenicity and immunosurveillance within the OAC microenvironment which could be explored using data mining analysis and cell modelling.

Appendix A APM H scores and Immune composition heatmap

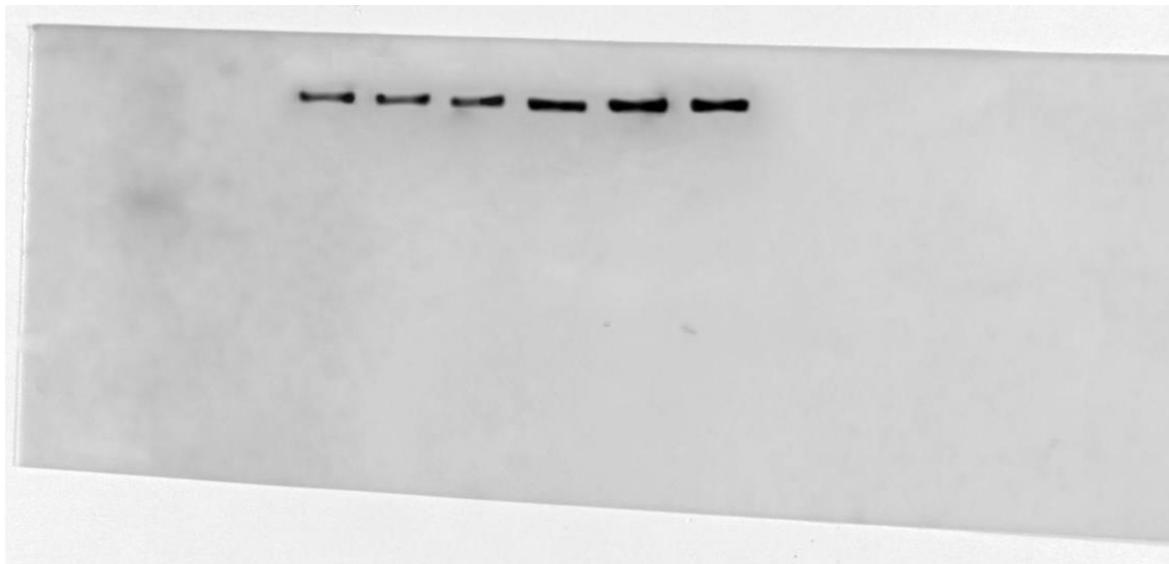


Supplementary Figure 1: Z-scored heatmap of APM H scores and immune cell density (n = 78), clustered using Canberra distance with Ward.D2 linkage, four distinct clusters of immune populations.

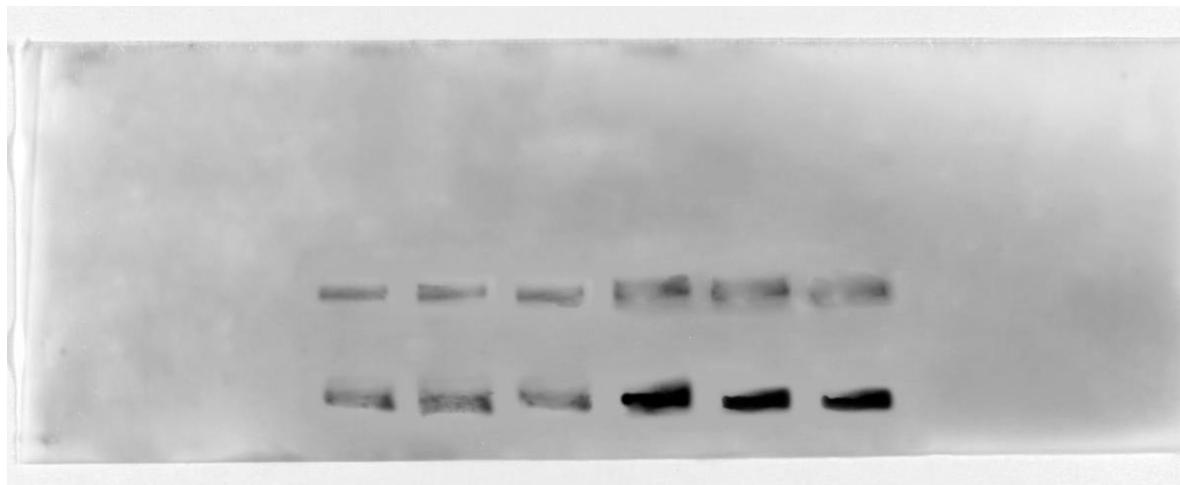
Western blots of CSDE1 knockdown experiment



Supplementary Figure 2A: Full image of Western blot of CSDE1, expression following CSDE1 knockdown in FLO-1 cells. Left to right - NC: Non-targeting control, KD: Knockdown CSDE1 60 pmol siRNA.



Supplementary Figure 2B: Full image of Western blot of HLA-ABC, expression following CSDE1 knockdown in FLO-1 cells. Left to right - NC: Non-targeting control, KD: Knockdown CSDE1 60 pmol siRNA.



Supplementary Figure 2C: Full image of Western blot of pSTAT1, expression following CSDE1 knockdown in FLO-1 cells. Left to right - NC: Non-targeting control, KD: Knockdown CSDE1 60 pmol siRNA.

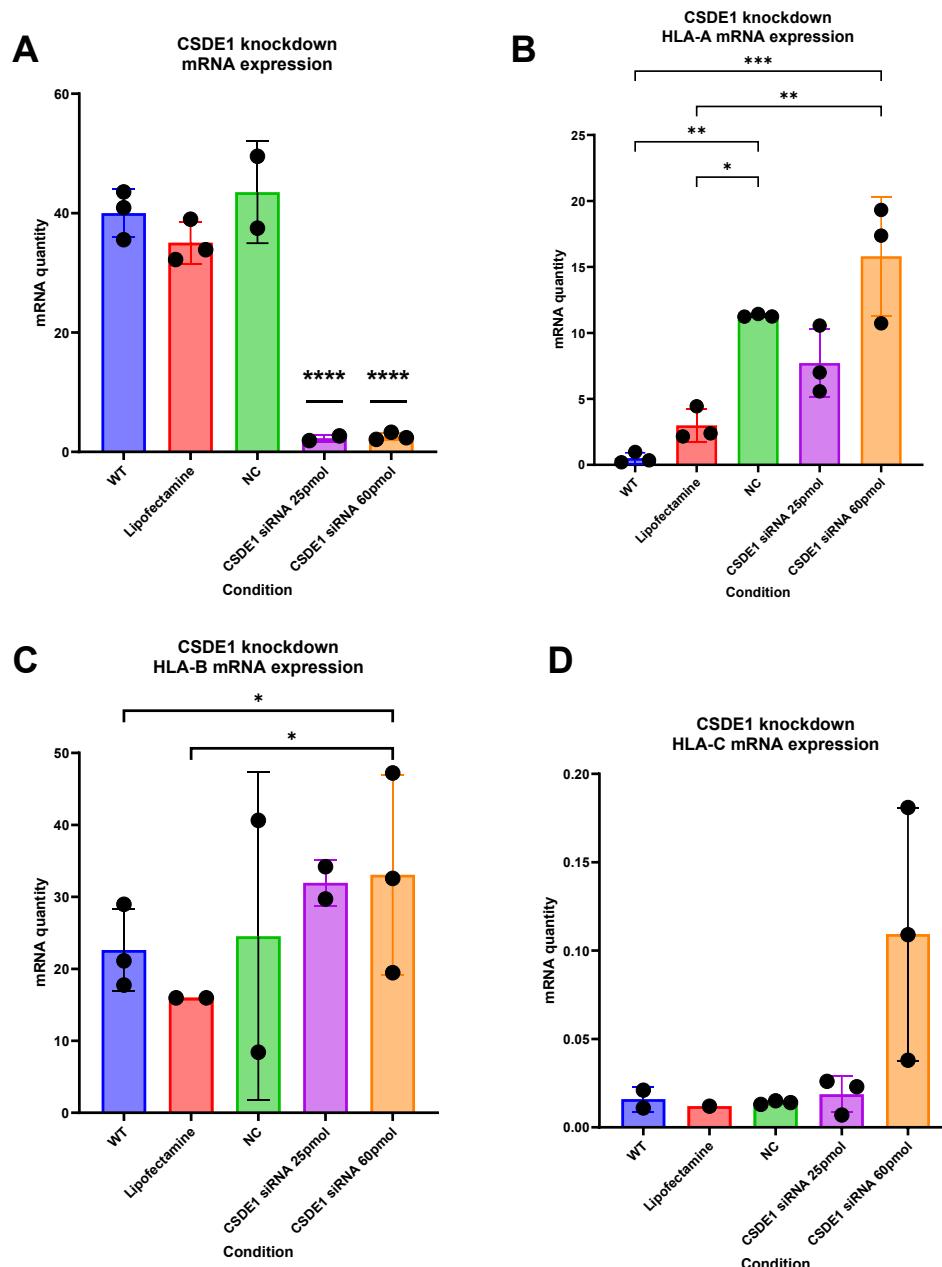


Supplementary Figure 2D: Full image of Western blot of STAT1, expression following CSDE1 knockdown in FLO-1 cells. Left to right - NC: Non-targeting control, KD: Knockdown CSDE1 60 pmol siRNA.



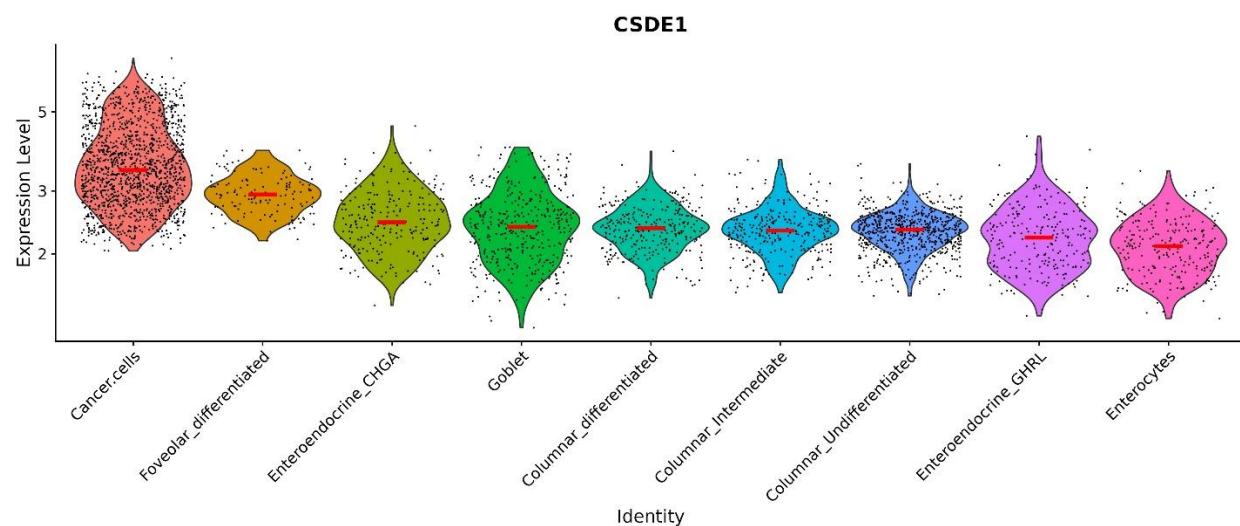
Supplementary Figure 2E: Full image of Western blot of GAPDH, expression following CSDE1 knockdown in FLO-1 cells. Left to right - NC: Non-targeting control, KD: Knockdown CSDE1 60 pmol siRNA.

Appendix B Percentage difference qPCR of CSDE1 knockdown experiment



Supplementary Figure 3: Percentage difference of mRNA quantification of CSDE1, HLA-A, HLA-B and HLA-C expression following CSDE1 knockdown in FLO-1 cells comparing to wildtype. A: CSDE1 gene expression following CSDE1 knockdown. B: HLA-A gene expression following CSDE1 knockdown. C: HLA-B gene expression following CSDE1 knockdown. D: HLA-C gene expression following CSDE1 knockdown. NC: Non-targeting control. Error bar: Standard deviation.

Appendix C CSDE1 expression in OAC cancer cells compared to healthy comparative cells.



Supplementary Figure 4: Violin plot of *CSDE1* mRNA expression in OAC cancer cells with comparative normal healthy gastric cardia cells. Red line represents median ratio.

Glossary of Terms

Term	Definition
Adenocarcinoma	Malignant tumour arising from glandular epithelial structures.
Antigen presentation machinery	Refers to a system of interacting proteins which function to process, package, and present peptide antigens on the cell surface for immune recognition.
Antigen presenting cells	A group of immune cells that engulf, process and present exogenous peptides on the cell surface via the MHC class II system to activate immune responses towards non-self/healthy antigens.
Barrett's Oesophagus	A premalignant condition arising from gastro-oesophageal reflux disease which involves metaplasia of the distal oesophagus resulting in replacement of the squamous epithelium to a columnar epithelium structure with goblet cells.
Complementary DNA (cDNA)	DNA synthesised from single stranded RNA template catalysed by the reverse transcriptase enzyme.
Cytotoxic T cells	A subset of thymic derived lymphocytes that possess the CD8 T cell receptor and can specifically target cells expressing non-self/healthy antigens on the cell surface for immune destruction.
Differential gene expression analysis	Statistical analysis used to quantify normalised gene expression fold-changes between a control and experimental group.
Gastro-oesophageal reflux disease (GORD)	A condition which involves the reflux of stomach acid into the gastro-oesophageal junction, causing irritation to the oesophagus squamous cell lining.
Human Leukocyte Antigen (HLA)	A group of cell surface proteins involved in presenting antigens to proximal T lymphocytes for immune recognition.

Glossary of Terms

Immune checkpoint inhibitors	An immunotherapeutic agent which blocks the action of immune inhibitory mechanisms often upregulated to evade immune destruction of cancer cells.
Major histocompatibility Complex class I	An endogenous system presents in almost all cells presenting antigens derived from cytosolic peptides to CD8+ T cells on cell surface HLA molecules.
Major histocompatibility Complex class II	An exogenous system present professional antigen presenting cells which presents antigens engulfed by professional antigen presenting cells to CD4+ T cells on cell surface HLA molecules.
Neoadjuvant therapy	A treatment method involving the application of chemo/chemoradiotherapy prior surgery.
Neoantigen	A novel peptide produced in cancer cells due to mutation of DNA which may be presented on cell surface MHC HLA molecules.
RNA-binding proteins	A varied class of proteins that interact with RNA to influence the regulation of gene expression, RNA processing, RNA transportation, localisation of RNA and RNA stability.
Tumour immune microenvironment	Describes the complex dynamic interactions of cellular and acellular factors which dictate the immune response or suppression in tumours.
T regulatory cells (Tregs)	A subpopulation of T helper cells expressing CD4 and FOXP3 proteins that typically function to suppress immune functions to enforce self-tolerance.

List of References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424.
2. CRUK. Oesophageal cancer incidence statistics 2020 CRUK; 2020 [Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophagealcancer/incidence#heading-Zero> . .
3. Kim J, Bowlby R, Mungall AJ, Robertson AG, Odze RD, Cherniack AD, et al. Integrated genomic characterization of oesophageal carcinoma. *Nature.* 2017;541(7636):169-75.
4. Lubin JH, Cook MB, Pandeya N, Vaughan TL, Abnet CC, Giffen C, et al. The importance of exposure rate on odds ratios by cigarette smoking and alcohol consumption for esophageal adenocarcinoma and squamous cell carcinoma in the Barrett's Esophagus and Esophageal Adenocarcinoma Consortium. *Cancer Epidemiol.* 2012;36(3):306-16.
5. Jeffrey CT, Christiana D, Charuhas D, Wei-Ting H, Seth J, Alexander H, et al. Gene signature of antigen processing and presentation machinery predicts response to checkpoint blockade in non-small cell lung cancer (NSCLC) and melanoma. *Journal for ImmunoTherapy of Cancer.* 2020;8(2):e000974.
6. Fang Y, Wang L, Wan C, Sun Y, Van der Jeught K, Zhou Z, et al. MAL2 drives immune evasion in breast cancer by suppressing tumor antigen presentation. *J Clin Invest.* 2021;131(1).
7. Ijsselsteijn ME, Petitprez F, Lacroix L, Ruano D, van der Breggen R, Julie C, et al. Revisiting immune escape in colorectal cancer in the era of immunotherapy. *British Journal of Cancer.* 2019;120(8):815-8.
8. Zhao Q, Yu J, Meng X. A good start of immunotherapy in esophageal cancer. *Cancer Med.* 2019;8(10):4519-26.
9. Hazini A, Fisher K, Seymour L. Deregulation of HLA-I in cancer and its central importance for immunotherapy. *J Immunother Cancer.* 2021;9(8).
10. Hur C, Miller M, Kong CY, Dowling EC, Nattinger KJ, Dunn M, et al. Trends in esophageal adenocarcinoma incidence and mortality. *Cancer.* 2013;119(6):1149-58.
11. Xie SH, Lagergren J. A global assessment of the male predominance in esophageal adenocarcinoma. *Oncotarget.* 2016;7(25):38876-83.
12. Arnold M, Soerjomataram I, Ferlay J, Forman D. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut.* 2015;64(3):381-7.
13. Arnold M, Colquhoun A, Cook MB, Ferlay J, Forman D, Soerjomataram I. Obesity and the Incidence of Upper Gastrointestinal Cancers: An Ecological Approach to Examine Differences across Age and Sex. *Cancer Epidemiol Biomarkers Prev.* 2016;25(1):90-7.
14. NHS. Digital N. Statistics on Obesity, Physical Activity and Diet 2019 NHS; 2019 [Available from: <https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-obesity-physical-activity-and-diet/statistics-on-obesity-physical-activity-and-diet-england-2019> . .
15. Lofdahl HE, Lane A, Lu Y, Lagergren P, Harvey RF, Blazeby JM, et al. Increased population prevalence of reflux and obesity in the United Kingdom compared with Sweden: a potential explanation for the difference in incidence of esophageal adenocarcinoma. *Eur J Gastroenterol Hepatol.* 2011;23(2):128-32.

List of References

16. Xie SH, Lagergren J. The Male Predominance in Esophageal Adenocarcinoma. *Clin Gastroenterol Hepatol*. 2016;14(3):338-47 e1.
17. Coleman HG, Xie SH, Lagergren J. The Epidemiology of Esophageal Adenocarcinoma. *Gastroenterology*. 2018;154(2):390-405.
18. Offman J, Pesola F, Sasieni P. Trends and projections in adenocarcinoma and squamous cell carcinoma of the oesophagus in England from 1971 to 2037. *Br J Cancer*. 2018;118(10):1391-8.
19. Conolly A, Davies B. Health survey for England 2017: adult and child overweight and obesity. *NHS Digital*. 2018;1:1-30.
20. Windsor-Shellard B. Adult smoking habits in the UK: 2015. *Office of National Statistics*: London, UK. 2017.
21. Peters SA, Huxley RR, Woodward M. Do smoking habits differ between women and men in contemporary Western populations? Evidence from half a million people in the UK Biobank study. *BMJ open*. 2014;4(12):e005663.
22. Souza RF. Reflux esophagitis and its role in the pathogenesis of Barrett's metaplasia. *J Gastroenterol*. 2017;52(7):767-76.
23. Mari L, Milano F, Parikh K, Straub D, Everts V, Hoeben KK, et al. A pSMAD/CDX2 complex is essential for the intestinalization of epithelial metaplasia. *Cell reports*. 2014;7(4):1197-210.
24. Kong J, Crissey MA, Funakoshi S, Kreindler JL, Lynch JP. Ectopic Cdx2 expression in murine esophagus models an intermediate stage in the emergence of Barrett's esophagus. *PLoS One*. 2011;6(4):e18280.
25. Coad RA, Woodman AC, Warner PJ, Barr H, Wright NA, Shepherd NA. On the histogenesis of Barrett's oesophagus and its associated squamous islands: a three-dimensional study of their morphological relationship with native oesophageal gland ducts. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*. 2005;206(4):388-94.
26. Leedham SJ, Preston SL, McDonald SA, 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.
27. Smith LP, Yamato JA, Galipeau PC, Paulson TG, Li X, Sanchez CA, et al. Within-patient phylogenetic reconstruction reveals early events in Barrett's Esophagus. *Evolutionary Applications*. 2021;14(2):399-415.
28. Martinez P, Mallo D, Paulson TG, Li X, Sanchez CA, Reid BJ, et al. Evolution of Barrett's esophagus through space and time at single-crypt and whole-biopsy levels. *Nature communications*. 2018;9(1):794.
29. Cook MB, Kamangar F, Whiteman DC, Freedman ND, Gammon MD, Bernstein L, et al. Cigarette smoking and adenocarcinomas of the esophagus and esophagogastric junction: a pooled analysis from the international BEACON consortium. *Journal of the National Cancer Institute*. 2010;102(17):1344-53.
30. Nowicki-Osuch K, Zhuang L, Jammula S, Bleaney CW, Mahbubani KT, Devonshire G, et al. Molecular phenotyping reveals the identity of Barrett's esophagus and its malignant transition. *Science*. 2021;373(6556):760-7.
31. Yang X, Chen X, Zhuang M, Yuan Z, Nie S, Lu M, et al. Smoking and alcohol drinking in relation to the risk of esophageal squamous cell carcinoma: a population-based case-control study in China. *Scientific reports*. 2017;7(1):17249.

32. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med.* 2003;348(17):1625-38.
33. Corley DA, Kubo A, Zhao W. Abdominal obesity and the risk of esophageal and gastric cardia carcinomas. *Cancer Epidemiology Biomarkers & Prevention.* 2008;17(2):352-8.
34. Steffen A, Huerta JM, Weiderpass E, Bueno-de-Mesquita HB, May AM, Siersema PD, et al. General and abdominal obesity and risk of esophageal and gastric adenocarcinoma in the European Prospective Investigation into Cancer and Nutrition. *International journal of cancer.* 2015;137(3):646-57.
35. Thrift AP, Shaheen NJ, Gammon MD, Bernstein L, Reid BJ, Onstad L, et al. Obesity and risk of esophageal adenocarcinoma and Barrett's esophagus: a Mendelian randomization study. *JNCI: Journal of the National Cancer Institute.* 2014;106(11).
36. Wu Y-W, Tseng P-H, Lee Y-C, Wang S-Y, Chiu H-M, Tu C-H, et al. Association of Esophageal Inflammation, Obesity and Gastroesophageal Reflux Disease: From FDG PET/CT Perspective. *PLOS ONE.* 2014;9(3):e92001.
37. El-Serag HB, Hashmi A, Garcia J, Richardson P, Alsarraj A, Fitzgerald S, et al. Visceral abdominal obesity measured by CT scan is associated with an increased risk of Barrett's oesophagus: a case-control study. *Gut.* 2014;63(2):220-9.
38. Corley DA, Kubo A, Levin TR, Block G, Habel L, Zhao W, et al. Abdominal obesity and body mass index as risk factors for Barrett's esophagus. *Gastroenterology.* 2007;133(1):34-41.
39. Nasr AO, Dillon MF, Conlon S, Downey P, Chen G, Ireland A, et al. Acid suppression increases rates of Barrett's esophagus and esophageal injury in the presence of duodenal reflux. *Surgery.* 2012;151(3):382-90.
40. Alsalahi O, Dobrian AD. Proton pump inhibitors: the culprit for Barrett's esophagus? *Frontiers in oncology.* 2015;4:373.
41. Lagergren J, Bergström R, Adami H-O, Nyrén O. Association between medications that relax the lower esophageal sphincter and risk for esophageal adenocarcinoma. *Annals of internal medicine.* 2000;133(3):165-75.
42. Rajendra S, Ackroyd R, Karim N, Mohan C, Ho JJ, Kutty M. Loss of human leucocyte antigen class I and gain of class II expression are early events in carcinogenesis: clues from a study of Barrett's oesophagus. *Journal of clinical pathology.* 2006;59(9):952-7.
43. Consortium EAG, 2 WTCCC, Su Z, Gay LJ, Strange A, Palles C, et al. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. *Nature genetics.* 2012;44(10):1131-6.
44. Dura P, van Veen EM, Salomon J, te Morsche RH, Roelofs HM, Kristinsson JO, et al. Barrett associated MHC and FOXF1 variants also increase esophageal carcinoma risk. *International journal of cancer.* 2013;133(7):1751-5.
45. Lam C, Liu W, Bel R, Chan K, Miller L, Brown M, et al. Polymorphisms of the FOXF1 and MHC locus genes in individuals undergoing esophageal acid reflux assessments. *Diseases of the Esophagus.* 2017;30(2).
46. Walker RC, Underwood TJ. Oesophageal cancer. *Surgery (Oxford).* 2017;35(11):627-34.
47. Cavallin F, Scarpa M, Cagol M, Alfieri R, Ruol A, Chiarion Sileni V, et al. Time to diagnosis in esophageal cancer: a cohort study. *Acta oncologica.* 2018;57(9):1179-84.

List of References

48. Grotenhuis BA, Van Hagen P, Wijnhoven BP, Spaander MC, Tilanus HW, Van Lanschot JJ. Delay in diagnostic workup and treatment of esophageal cancer. *Journal of Gastrointestinal Surgery*. 2010;14:476-83.
49. Zhang Y. Epidemiology of esophageal cancer. *World journal of gastroenterology: WJG*. 2013;19(34):5598.
50. Cowie A, Noble F, Underwood T. Strategies to improve outcomes in esophageal adenocarcinoma. *Expert review of anticancer therapy*. 2014;14(6):677-87.
51. Groene O, Chadwick G, Riley S, Hardwick RH, Crosby T, Greenaway K, et al. Re-organisation of oesophago-gastric cancer services in England and Wales: a follow-up assessment of progress and remaining challenges. *BMC research notes*. 2014;7(1):1-7.
52. Rice TW, Ishwaran H, Ferguson MK, Blackstone EH, Goldstraw P. Cancer of the esophagus and esophagogastric junction: an eighth edition staging primer. *Journal of Thoracic Oncology*. 2017;12(1):36-42.
53. Rice TW. Esophageal Cancer Staging. *Korean J Thorac Cardiovasc Surg*. 2015;48(3):157-63.
54. Talsma K, van Hagen P, Grotenhuis BA, Steyerberg EW, Tilanus HW, van Lanschot JJ, et al. Comparison of the 6th and 7th Editions of the UICC-AJCC TNM Classification for Esophageal Cancer. *Ann Surg Oncol*. 2012;19(7):2142-8.
55. William HA, Jane MB, Griffin SM, David C, Janusz AJ, Rachel W. Guidelines for the management of oesophageal and gastric cancer. *Gut*. 2011;60(11):1449.
56. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, et al. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med*. 2006;355(1):11-20.
57. 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;27(30):5062-7.
58. Dasari S, Tchounwou PB. Cisplatin in cancer therapy: molecular mechanisms of action. *Eur J Pharmacol*. 2014;740:364-78.
59. Longley DB, Harkin DP, Johnston PG. 5-Fluorouracil: mechanisms of action and clinical strategies. *Nature Reviews Cancer*. 2003;3(5):330-8.
60. Bonadonna G, Gianni L, Santoro A, Bonfante V, Bidoli P, Casali P, et al. Drugs ten years later: Epirubicin. *Annals of Oncology*. 1993;4(5):359-69.
61. Mayor S. More patients with oesophageal and gastric cancers are getting curative treatment plans. *BMJ*. 2016;354:i4923.
62. Howlader N NA, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2018, National Cancer Institute 2021 [Available from: https://seer.cancer.gov/archive/csr/1975_2018/].
63. Rubenstein JH, Shaheen NJ. Epidemiology, diagnosis, and management of esophageal adenocarcinoma. *Gastroenterology*. 2015;149(2):302-17. e1.
64. Chandrasoma PT. Histologic definition of gastro-esophageal reflux disease. *Curr Opin Gastroenterol*. 2013;29(4):460-7.

65. Takubo K, Vieth M, Aida J, Sawabe M, Kumagai Y, Hoshihara Y, et al. Differences in the Definitions Used for Esophageal and Gastric Diseases in Different Countries. *Digestion*. 2009;80(4):248-57.
66. Foo J, Gunnarsson EB, Leder K, Storey K. Spread of premalignant mutant clones and cancer initiation in multilayered tissue. *The Annals of Applied Probability*. 2023;33(1):299-343.
67. Dislich B, Kröll D, Langer R. Surgical pathology of adenocarcinomas arising around or within the gastroesophageal junction. *Updates in Surgery*. 2023;75(2):395-402.
68. Wang DH. The Esophageal Squamous Epithelial Cell-Still a Reasonable Candidate for the Barrett's Esophagus Cell of Origin? *Cell Mol Gastroenterol Hepatol*. 2017;4(1):157-60.
69. Jiang M, Li H, Zhang Y, Yang Y, Lu R, Liu K, et al. Transitional basal cells at the squamous–columnar junction generate Barrett's oesophagus. *Nature*. 2017;550(7677):529-33.
70. Que J, Garman KS, Souza RF, Spechler SJ. Pathogenesis and Cells of Origin of Barrett's Esophagus. *Gastroenterology*. 2019;157(2):349-64.e1.
71. Wang DH, Souza RF. Transcommitment: Paving the Way to Barrett's Metaplasia. *Adv Exp Med Biol*. 2016;908:183-212.
72. 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.
73. Derk S, Nason KS, Liao X, Stachler MD, Liu KX, Liu JB, et al. Epithelial PD-L2 Expression Marks Barrett's Esophagus and Esophageal Adenocarcinoma. *Cancer Immunol Res*. 2015;3(10):1123-9.
74. George J, Saito M, Tsuta K, Iwakawa R, Shiraishi K, Scheel AH, et al. Genomic Amplification of CD274 (PD-L1) in Small-Cell Lung Cancer. *Clin Cancer Res*. 2017;23(5):1220-6.
75. Straub M, Drecoll E, Pfarr N, Weichert W, Langer R, Hapfelmeier A, et al. CD274/PD-L1 gene amplification and PD-L1 protein expression are common events in squamous cell carcinoma of the oral cavity. *Oncotarget*. 2016;7(11):12024-34.
76. Howitt BE, Sun HH, Roemer MG, Kelley A, Chapuy B, Aviki E, et al. Genetic Basis for PD-L1 Expression in Squamous Cell Carcinomas of the Cervix and Vulva. *JAMA Oncol*. 2016;2(4):518-22.
77. Budczies J, Denkert C, Győrffy B, Schirmacher P, Stenzinger A. Chromosome 9p copy number gains involving PD-L1 are associated with a specific proliferation and immune-modulating gene expression program active across major cancer types. *BMC Med Genomics*. 2017;10(1):74.
78. Barrett MT, Anderson KS, Lenkiewicz E, Andreozzi M, Cunliffe HE, Klassen CL, et al. Genomic amplification of 9p24.1 targeting JAK2, PD-L1, and PD-L2 is enriched in high-risk triple negative breast cancer. *Oncotarget*. 2015;6(28):26483-93.
79. Lagisetty KH, McEwen DP, Nancarrow DJ, Schiebel JG, Ferrer-Torres D, Ray D, et al. Immune determinants of Barrett's progression to esophageal adenocarcinoma. *JCI Insight*. 2021;6(1).
80. Noorani A, Li X, Goddard M, Crawte J, Alexandrov LB, Secrier M, et al. Genomic evidence supports a clonal diaspora model for metastases of esophageal adenocarcinoma. *Nat Genet*. 2020;52(1):74-83.
81. Raufi AG, Almhanna K. Immune checkpoint inhibitors for esophageal cancer: are we moving in the right direction? *Ann Transl Med*. 2019;7(Suppl 3):S102.
82. Fuchs CS, Doi T, Jang RW, Muro K, Satoh T, Machado M, et al. Safety and Efficacy of Pembrolizumab Monotherapy in Patients With Previously Treated Advanced Gastric and

List of References

- Gastroesophageal Junction Cancer: Phase 2 Clinical KEYNOTE-059 Trial. *JAMA Oncol.* 2018;4(5):e180013.
83. Hamid O, Robert C, Daud A, Hodi FS, Hwu W-J, Kefford R, et al. Safety and Tumor Responses with Lambrolizumab (Anti-PD-1) in Melanoma. *New England Journal of Medicine.* 2013;369(2):134-44.
84. Robert C, Ribas A, Wolchok JD, Hodi FS, Hamid O, Kefford R, et al. Anti-programmed-death-receptor-1 treatment with pembrolizumab in ipilimumab-refractory advanced melanoma: a randomised dose-comparison cohort of a phase 1 trial. *Lancet.* 2014;384(9948):1109-17.
85. Hamanishi J, Mandai M, Ikeda T, Minami M, Kawaguchi A, Murayama T, et al. Safety and Antitumor Activity of Anti-PD-1 Antibody, Nivolumab, in Patients With Platinum-Resistant Ovarian Cancer. *J Clin Oncol.* 2015;33(34):4015-22.
86. Goode EF, Smyth EC. Immunotherapy for Gastroesophageal Cancer. *J Clin Med.* 2016;5(10).
87. Janjigian YY, Bendell J, Calvo E, Kim JW, Ascierto PA, Sharma P, et al. CheckMate-032 Study: Efficacy and Safety of Nivolumab and Nivolumab Plus Ipilimumab in Patients With Metastatic Esophagogastric Cancer. *J Clin Oncol.* 2018;36(28):2836-44.
88. Hodi FS, Chesney J, Pavlick AC, Robert C, Grossmann KF, McDermott DF, et al. Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. *Lancet Oncol.* 2016;17(11):1558-68.
89. Tsui E, Madu A, Belinsky I, Yannuzzi LA, Freund KB, Modi YS. Combination Ipilimumab and Nivolumab for Metastatic Melanoma Associated With Ciliochoroidal Effusion and Exudative Retinal Detachment. *JAMA Ophthalmol.* 2017;135(12):1455-7.
90. Callahan MK, Kluger H, Postow MA, Segal NH, Lesokhin A, Atkins MB, et al. Nivolumab Plus Ipilimumab in Patients With Advanced Melanoma: Updated Survival, Response, and Safety Data in a Phase I Dose-Escalation Study. *J Clin Oncol.* 2018;36(4):391-8.
91. Wolchok JD, Chiarion-Sileni V, Gonzalez R, Rutkowski P, Grob J-J, Cowey CL, et al. Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *New England Journal of Medicine.* 2017;377(14):1345-56.
92. Lussier DM, Johnson JL, Hingorani P, Blattman JN. Combination immunotherapy with α -CTLA-4 and α -PD-L1 antibody blockade prevents immune escape and leads to complete control of metastatic osteosarcoma. *J Immunother Cancer.* 2015;3:21.
93. Overman MJ, Lonardi S, Wong KYM, Lenz HJ, Gelsomino F, Aglietta M, et al. Durable Clinical Benefit With Nivolumab Plus Ipilimumab in DNA Mismatch Repair-Deficient/Microsatellite Instability-High Metastatic Colorectal Cancer. *J Clin Oncol.* 2018;36(8):773-9.
94. Omuro A, Vlahovic G, Lim M, Sahebjam S, Baehring J, Cloughesy T, et al. Nivolumab with or without ipilimumab in patients with recurrent glioblastoma: results from exploratory phase I cohorts of CheckMate 143. *Neuro Oncol.* 2018;20(5):674-86.
95. Hammers HJ, Plimack ER, Infante JR, Rini BI, McDermott DF, Lewis LD, et al. Safety and Efficacy of Nivolumab in Combination With Ipilimumab in Metastatic Renal Cell Carcinoma: The CheckMate 016 Study. *J Clin Oncol.* 2017;35(34):3851-8.
96. Mathur D, Root AR, Bugaj-Gaweda B, Bisulco S, Tan X, Fang W, et al. A Novel GUCY2C-CD3 T-Cell Engaging Bispecific Construct (PF-07062119) for the Treatment of Gastrointestinal Cancers. *Clin Cancer Res.* 2020;26(9):2188-202.

97. Birbe R, Palazzo JP, Walters R, Weinberg D, Schulz S, Waldman SA. Guanylyl cyclase C is a marker of intestinal metaplasia, dysplasia, and adenocarcinoma of the gastrointestinal tract. *Hum Pathol.* 2005;36(2):170-9.
98. Danaee H, Kalebic T, Wyant T, Fassan M, Mescoli C, Gao F, et al. Consistent expression of guanylyl cyclase-C in primary and metastatic gastrointestinal cancers. *PLoS One.* 2017;12(12):e0189953.
99. Kudo T, Hamamoto Y, Kato K, Ura T, Kojima T, Tsushima T, et al. Nivolumab treatment for oesophageal squamous-cell carcinoma: an open-label, multicentre, phase 2 trial. *Lancet Oncol.* 2017;18(5):631-9.
100. Kang YK, Boku N, Satoh T, Ryu MH, Chao Y, Kato K, et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTON-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet.* 2017;390(10111):2461-71.
101. Muro K, Chung HC, Shankaran V, Geva R, Catenacci D, Gupta S, et al. Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial. *The Lancet Oncology.* 2016;17(6):717-26.
102. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med.* 2015;372(26):2509-20.
103. Doi T, Piha-Paul SA, Jalal SI, Saraf S, Lunceford J, Koshiji M, et al. Safety and Antitumor Activity of the Anti-Programmed Death-1 Antibody Pembrolizumab in Patients With Advanced Esophageal Carcinoma. *J Clin Oncol.* 2018;36(1):61-7.
104. Shitara K, Ozguroglu M, Bang YJ, Di Bartolomeo M, Mandala M, Ryu MH, et al. Pembrolizumab versus paclitaxel for previously treated, advanced gastric or gastro-oesophageal junction cancer (KEYNOTE-061): a randomised, open-label, controlled, phase 3 trial. *Lancet.* 2018;392(10142):123-33.
105. Shah MA, Kojima T, Hochhauser D, Enzinger P, Raimbourg J, Hollebecque A, et al. Efficacy and Safety of Pembrolizumab for Heavily Pretreated Patients With Advanced, Metastatic Adenocarcinoma or Squamous Cell Carcinoma of the Esophagus: The Phase 2 KEYNOTE-180 Study. *JAMA Oncol.* 2019;5(4):546-50.
106. Takimoto R, Kamigaki T, Gotoda T, Takahashi T, Okada S, Ibe H, et al. Esophageal cancer responsive to the combination of immune cell therapy and low-dose nivolumab: two case reports. *J Med Case Rep.* 2021;15(1):191.
107. Wagener-Ryczek S, Schoemmel M, Kraemer M, Bruns C, Schroeder W, Zander T, et al. Immune profile and immunosurveillance in treatment-naive and neoadjuvantly treated esophageal adenocarcinoma. *Cancer Immunol Immunother.* 2020;69(4):523-33.
108. Hanahan D, Weinberg Robert A. Hallmarks of Cancer: The Next Generation. *Cell.* 2011;144(5):646-74.
109. Tsukamoto H, Fujieda K, Senju S, Ikeda T, Oshiumi H, Nishimura Y. Immune-suppressive effects of interleukin-6 on T-cell-mediated anti-tumor immunity. *Cancer Sci.* 2018;109(3):523-30.
110. Rock KL, Reits E, Neefjes J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol.* 2016;37(11):724-37.
111. Munhoz RR, Postow MA. Recent advances in understanding antitumor immunity. *F1000Res.* 2016;5:2545.

List of References

112. Chen L, Flies DB. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat Rev Immunol.* 2013;13(4):227-42.
113. Wieczorek M, Abualrous ET, Sticht J, Alvaro-Benito M, Stolzenberg S, Noe F, et al. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Front Immunol.* 2017;8:292.
114. Castro A, Ozturk K, Pyke RM, Xian S, Zanetti M, Carter H. Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes. *BMC Med Genomics.* 2019;12(Suppl 6):107.
115. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell.* 2017;171(6):1259-71 e11.
116. Wang S, He Z, Wang X, Li H, Liu XS. Antigen presentation and tumor immunogenicity in cancer immunotherapy response prediction. *Elife.* 2019;8.
117. Minias P, Bateson ZW, Whittingham LA, Johnson JA, Oyler-McCance S, Dunn PO. Contrasting evolutionary histories of MHC class I and class II loci in grouse--effects of selection and gene conversion. *Heredity (Edinb).* 2016;116(5):466-76.
118. Radwan J, Babik W, Kaufman J, Lenz TL, Winternitz J. Advances in the Evolutionary Understanding of MHC Polymorphism. *Trends Genet.* 2020;36(4):298-311.
119. Bockerstett KA, Wong CF, Koehm S, Ford EL, DiPaolo RJ. Molecular Characterization of Gastric Epithelial Cells Using Flow Cytometry. *Int J Mol Sci.* 2018;19(4).
120. Dobrzanski MJ. Expanding roles for CD4 T cells and their subpopulations in tumor immunity and therapy. *Front Oncol.* 2013;3:63.
121. Joffre OP, Segura E, Savina A, Amigorena S. Cross-presentation by dendritic cells. *Nature Reviews Immunology.* 2012;12(8):557-69.
122. Shen L, Sigal LJ, Boes M, Rock KL. Important role of cathepsin S in generating peptides for TAP-independent MHC class I crosspresentation in vivo. *Immunity.* 2004;21(2):155-65.
123. Yi Y, Zhou Z, Shu S, Fang Y, Twitty C, Hilton TL, et al. Autophagy-assisted antigen cross-presentation: Autophagosome as the argo of shared tumor-specific antigens and DAMPs. *Oncoimmunology.* 2012;1(6):976-8.
124. Sanchez-Paulete AR, Teijeira A, Cueto FJ, Garasa S, Perez-Gracia JL, Sanchez-Arraez A, et al. Antigen cross-presentation and T-cell cross-priming in cancer immunology and immunotherapy. *Ann Oncol.* 2017;28(suppl_12):xii74.
125. Muntjewerff EM, Bianchi F, Maassen S, van den Bogaart G. Measuring the molecular mechanisms of tumor antigen cross-presentation in dendritic cells and macrophages. *The Journal of Immunology.* 2019;202(1 Supplement):177.7.
126. Nguyen LT, Elford AR, Murakami K, Garza KM, Schoenberger SP, Odermatt B, et al. Tumor growth enhances cross-presentation leading to limited T cell activation without tolerance. *J Exp Med.* 2002;195(4):423-35.
127. Dickgreber N, Stoitzner P, Bai Y, Price KM, Farrand KJ, Manning K, et al. Targeting antigen to MHC class II molecules promotes efficient cross-presentation and enhances immunotherapy. *J Immunol.* 2009;182(3):1260-9.
128. Orange DE, Jegathesan M, Blachere NE, Frank MO, Scher HI, Albert ML, et al. Effective antigen cross-presentation by prostate cancer patients' dendritic cells: implications for prostate cancer immunotherapy. *Prostate Cancer Prostatic Dis.* 2004;7(1):63-72.

129. Wculek SK, Amores-Iniesta J, Conde-Garrosa R, Khouili SC, Melero I, Sancho D. Effective cancer immunotherapy by natural mouse conventional type-1 dendritic cells bearing dead tumor antigen. *J Immunother Cancer*. 2019;7(1):100.
130. van den Elsen PJ, Holling TM, Kuipers HF, van der Stoep N. Transcriptional regulation of antigen presentation. *Current opinion in immunology*. 2004;16(1):67-75.
131. Gobin SJ, Peijnenburg A, Keijsers V, van den Elsen PJ. Site α is crucial for two routes of IFNy-induced MHC class I transactivation: the ISRE-mediated route and a novel pathway involving CIITA. *Immunity*. 1997;6(5):601-11.
132. Gobin SJ, van Zutphen M, Westerheide SD, Boss JM, van den Elsen PJ. The MHC-specific enhanceosome and its role in MHC class I and β 2-microglobulin gene transactivation. *The Journal of Immunology*. 2001;167(9):5175-84.
133. Willman CL, Sever CE, Pallavicini MG, Harada H, Tanaka N, Slovak ML, et al. Deletion of IRF-1, mapping to chromosome 5q31. 1, in human leukemia and preleukemic myelodysplasia. *Science*. 1993;259(5097):968-71.
134. Kerim S, Mecucci C, Cuneo A, Vandenbergh E, Louwagie A, Stul M, et al. 5q-anomaly in lymphoid disorders. *Leukemia*. 1990;4(1):12-5.
135. Le Beau MM, Albain K, Larson R, Vardiman J, Davis E, Blough R, et al. Clinical and cytogenetic correlations in 63 patients with therapy-related myelodysplastic syndromes and acute nonlymphocytic leukemia: further evidence for characteristic abnormalities of chromosomes no. 5 and 7. *Journal of Clinical Oncology*. 1986;4(3):325-45.
136. TAMURA G, SAKATA K, NISHIZUKA S, MAESAWA C, SUZUKI Y, TERASHIMA M, et al. Allelotyping of adenoma and differentiated adenocarcinoma of the stomach. *The Journal of pathology*. 1996;180(4):371-7.
137. Cavalli LR, Riggins RB, Wang A, Clarke R, Haddad BR. Frequent loss of heterozygosity at the interferon regulatory factor-1 gene locus in breast cancer. *Breast cancer research and treatment*. 2010;121:227-31.
138. Ogasawara S, Tamura G, Maesawa C, Suzuki Y, Ishida K, Satoh N, et al. Common deleted region on the long arm of chromosome 5 in esophageal carcinoma. *Gastroenterology*. 1996;110(1):52-7.
139. Wang Y, Liu D-P, Chen P-P, Koeffler HP, Tong X-J, Xie D. Involvement of IFN regulatory factor (IRF)-1 and IRF-2 in the formation and progression of human esophageal cancers. *Cancer research*. 2007;67(6):2535-43.
140. Gobin S, Biesta P, De Steenwinkel J, Datema G, Van den Elsen P. CREB provides for an alternative pathway of HLA-G transactivation. *J Biol Chem*. 2002;277:39525-31.
141. Meissner TB, Li A, Biswas A, Lee K-H, Liu Y-J, Bayir E, et al. NLR family member NLRC5 is a transcriptional regulator of MHC class I genes. *Proceedings of the National Academy of Sciences*. 2010;107(31):13794-9.
142. Masternak K, Muhlethaler-Mottet A, Villard J, Zufferey M, Steimle V, Reith W. CIITA is a transcriptional coactivator that is recruited to MHC class II promoters by multiple synergistic interactions with an enhanceosome complex. *Genes & development*. 2000;14(9):1156-66.
143. Yoshihama S, Roszik J, Downs I, Meissner TB, Vijayan S, Chapuy B, et al. NLRC5/MHC class I transactivator is a target for immune evasion in cancer. *Proc Natl Acad Sci U S A*. 2016;113(21):5999-6004.

List of References

144. Singer D, Devaiah B. CIITA and Its Dual Roles in MHC Gene Transcription. *Frontiers in Immunology*. 2013;4.
145. Lv J, Zhou Y, Zhou N, Wang Z, Chen J, Chen H, et al. Epigenetic modification of CSDE1 locus dictates immune recognition of nascent tumorigenic cells. *Science Translational Medicine*. 2015(681):eabq6024.
146. Kottke T, Tonne J, Evgin L, Driscoll CB, van Vloten J, Jennings VA, et al. Oncolytic virotherapy induced CSDE1 neo-antigenesis restricts VSV replication but can be targeted by immunotherapy. *Nature Communications*. 2021;12(1):1930.
147. Pedersen MH, Hood BL, Beck HC, Conrads TP, Ditzel HJ, Leth-Larsen R. Downregulation of antigen presentation-associated pathway proteins is linked to poor outcome in triple-negative breast cancer patient tumors. *Oncoimmunology*. 2017;6(5):e1305531.
148. Tanaka K, Tsuchikawa T, Miyamoto M, Maki T, Ichinokawa M, Kubota KC, et al. Down-regulation of Human Leukocyte Antigen class I heavy chain in tumors is associated with a poor prognosis in advanced esophageal cancer patients. *Int J Oncol*. 2012;40(4):965-74.
149. Sokol L, Koelzer VH, Rau TT, Karamitopoulou E, Zlobec I, Lugli A. Loss of tapasin correlates with diminished CD8(+) T-cell immunity and prognosis in colorectal cancer. *J Transl Med*. 2015;13:279.
150. Porter KM, Hermann C, Traherne JA, Boyle LH. TAPBPR isoforms exhibit altered association with MHC class I. *Immunology*. 2014;142(2):289-99.
151. Hu SP, Zhou GB, Luan JA, Chen YP, Xiao DW, Deng YJ, et al. Polymorphisms of HLA-A and HLA-B genes in genetic susceptibility to esophageal carcinoma in Chaoshan Han Chinese. *Dis Esophagus*. 2010;23(1):46-52.
152. Mimura K, Shiraishi K, Mueller A, Izawa S, Kua LF, So J, et al. The MAPK pathway is a predominant regulator of HLA-A expression in esophageal and gastric cancer. *J Immunol*. 2013;191(12):6261-72.
153. Shen YQ, Zhang JQ, Miao FQ, Zhang JM, Jiang Q, Chen H, et al. Relationship between the downregulation of HLA class I antigen and clinicopathological significance in gastric cancer. *World J Gastroenterol*. 2005;11(23):3628-31.
154. Lim EB, Oh H-S, Kim KC, Kim M-H, Kim YJ, Kim BJ, et al. Identification and functional validation of HLA-C as a potential gene involved in colorectal cancer in the Korean population. *BMC Genomics*. 2022;23(1):261.
155. Ishigami S, Arigami T, Okumura H, Uchikado Y, Kita Y, Kurahara H, et al. Human leukocyte antigen (HLA)-E and HLA-F expression in gastric cancer. *Anticancer Res*. 2015;35(4):2279-85.
156. Yie SM, Yang H, Ye SR, Li K, Dong DD, Lin XM. Expression of human leukocyte antigen G (HLA-G) correlates with poor prognosis in gastric carcinoma. *Ann Surg Oncol*. 2007;14(10):2721-9.
157. Sade-Feldman M, Jiao YJ, Chen JH, Rooney MS, Barzily-Rokni M, Eliane J-P, et al. Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nature Communications*. 2017;8(1):1136.
158. Shimoda T, Wada R, Kure S, Ishino K, Kudo M, Ohashi R, et al. Expression of protein disulfide isomerase A3 and its clinicopathological association in gastric cancer. *Oncol Rep*. 2019;41(4):2265-72.
159. Peng RQ, Chen YB, Ding Y, Zhang R, Zhang X, Yu XJ, et al. Expression of calreticulin is associated with infiltration of T-cells in stage IIIB colon cancer. *World J Gastroenterol*. 2010;16(19):2428-34.

160. Ogino T, Shigyo H, Ishii H, Katayama A, Miyokawa N, Harabuchi Y, et al. HLA class I antigen down-regulation in primary laryngeal squamous cell carcinoma lesions as a poor prognostic marker. *Cancer Res.* 2006;66(18):9281-9.
161. Cifaldi L, Lo Monaco E, Forloni M, Giorda E, Lorenzi S, Petrini S, et al. Natural killer cells efficiently reject lymphoma silenced for the endoplasmic reticulum aminopeptidase associated with antigen processing. *Cancer Res.* 2011;71(5):1597-606.
162. James E, Bailey I, Sugiyarto G, Elliott T. Induction of protective antitumor immunity through attenuation of ERAAP function. *J Immunol.* 2013;190(11):5839-46.
163. Fruci D, Giacomini P, Nicotra MR, Forloni M, Fraioli R, Saveanu L, et al. Altered expression of endoplasmic reticulum aminopeptidases ERAP1 and ERAP2 in transformed non-lymphoid human tissues. *J Cell Physiol.* 2008;216(3):742-9.
164. Mari L, Hoefnagel SJM, Zito D, van de Meent M, van Endert P, Calpe S, et al. microRNA 125a Regulates MHC-I Expression on Esophageal Adenocarcinoma Cells, Associated With Suppression of Antitumor Immune Response and Poor Outcomes of Patients. *Gastroenterology.* 2018;155(3):784-98.
165. Kalaora S, Lee JS, Barnea E, Levy R, Greenberg P, Alon M, et al. Immunoproteasome expression is associated with better prognosis and response to checkpoint therapies in melanoma. *Nature Communications.* 2020;11(1):896.
166. Tertipis N, Haeggblom L, Nordfors C, Grün N, Näsman A, Vlastos A, et al. Correlation of LMP10 Expression and Clinical Outcome in Human Papillomavirus (HPV) Positive and HPV-Negative Tonsillar and Base of Tongue Cancer. *PLOS ONE.* 2014;9(4):e95624.
167. God JM, Cameron C, Figueroa J, Amria S, Hossain A, Kempkes B, et al. Elevation of c-MYC disrupts HLA class II-mediated immune recognition of human B cell tumors. *J Immunol.* 2015;194(4):1434-45.
168. Meissner M, Whiteside TL, Kaufmann R, Seliger B. CIITA versus IFN-gamma induced MHC class II expression in head and neck cancer cells. *Arch Dermatol Res.* 2009;301(2):189-93.
169. Seguin-Estevez Q, De Palma R, Krawczyk M, Leimgruber E, Villard J, Picard C, et al. The transcription factor RFX protects MHC class II genes against epigenetic silencing by DNA methylation. *J Immunol.* 2009;183(4):2545-53.
170. Leite FA, Lira RC, Fedatto PF, Antonini SR, Martinelli CE, Jr., de Castro M, et al. Low expression of HLA-DRA, HLA-DPA1, and HLA-DPB1 is associated with poor prognosis in pediatric adrenocortical tumors (ACT). *Pediatr Blood Cancer.* 2014;61(11):1940-8.
171. Kohno T, Kunitoh H, Shimada Y, Shiraishi K, Ishii Y, Goto K, et al. Individuals susceptible to lung adenocarcinoma defined by combined HLA-DQA1 and TERT genotypes. *Carcinogenesis.* 2010;31(5):834-41.
172. Magnusson PKE, Enroth H, Eriksson I, Held M, Nyren O, Engstrand L, et al. Gastric cancer and human leukocyte antigen: distinct DQ and DR alleles are associated with development of gastric cancer and infection by *Helicobacter pylori*. *Cancer Res.* 2001;61(6):2684-9.
173. Fernandez-Cuesta L, Plenker D, Osada H, Sun R, Menon R, Leenders F, et al. CD74-NRG1 fusions in lung adenocarcinoma. *Cancer Discov.* 2014;4(4):415-22.
174. Liu Z, Chu S, Yao S, Li Y, Fan S, Sun X, et al. CD74 interacts with CD44 and enhances tumorigenesis and metastasis via RHOA-mediated cofilin phosphorylation in human breast cancer cells. *Oncotarget.* 2016;7(42):68303-13.

List of References

175. Yang Y, Lim SK, Choong LY, Lee H, Chen Y, Chong PK, et al. Cathepsin S mediates gastric cancer cell migration and invasion via a putative network of metastasis-associated proteins. *J Proteome Res.* 2010;9(9):4767-78.
176. Gautam J, Banskota S, Lee H, Lee YJ, Jeon YH, Kim JA, et al. Down-regulation of cathepsin S and matrix metalloproteinase-9 via Src, a non-receptor tyrosine kinase, suppresses triple-negative breast cancer growth and metastasis. *Exp Mol Med.* 2018;50(9):1-14.
177. Sudhan DR, Rabaglino MB, Wood CE, Siemann DW. Cathepsin L in tumor angiogenesis and its therapeutic intervention by the small molecule inhibitor KGP94. *Clin Exp Metastasis.* 2016;33(5):461-73.
178. Schneppenheim J, Loock AC, Hutt S, Schweizer M, Lullmann-Rauch R, Oberg HH, et al. The Influence of MHC Class II on B Cell Defects Induced by Invariant Chain/CD74 N-Terminal Fragments. *J Immunol.* 2017;199(1):172-85.
179. Zhang Y, Wu YY, Jiang JN, Liu XS, Ji FJ, Fang XD. MiRNA-3978 regulates peritoneal gastric cancer metastasis by targeting legumain. *Oncotarget.* 2016;7(50):83223-30.
180. Rausch MP, Irvine KR, Antony PA, Restifo NP, Cresswell P, Hastings KT. GILT accelerates autoimmunity to the melanoma antigen tyrosinase-related protein 1. *J Immunol.* 2010;185(5):2828-35.
181. Phipps-Yonas H, Cui H, Sebastian N, Brunhoeber PS, Haddock E, Deymier MJ, et al. Low GILT Expression is Associated with Poor Patient Survival in Diffuse Large B-Cell Lymphoma. *Front Immunol.* 2013;4:425.
182. Xiang YJ, Guo MM, Zhou CJ, Liu L, Han B, Kong LY, et al. Absence of gamma-interferon-inducible lysosomal thiol reductase (GILT) is associated with poor disease-free survival in breast cancer patients. *PLoS One.* 2014;9(10):e109449.
183. 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;45(5):478-86.
184. Frankell AM, Jammula S, Li X, Contino G, Killcoyne S, Abbas S, et al. The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat Genet.* 2019;51(3):506-16.
185. Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P, Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene.* 2007;26(15):2157-65.
186. Gomez del Pulgar T, Benitah SA, Valeron PF, Espina C, Lacal JC. Rho GTPase expression in tumourigenesis: evidence for a significant link. *Bioessays.* 2005;27(6):602-13.
187. Kissil JL, Walmsley MJ, Hanlon L, Haigis KM, Bender Kim CF, Sweet-Cordero A, et al. Requirement for Rac1 in a K-ras induced lung cancer in the mouse. *Cancer Res.* 2007;67(17):8089-94.
188. Pan Y, Bi F, Liu N, Xue Y, Yao X, Zheng Y, et al. Expression of seven main Rho family members in gastric carcinoma. *Biochem Biophys Res Commun.* 2004;315(3):686-91.
189. Sander EE, van Delft S, ten Klooster JP, Reid T, van der Kammen RA, Michiels F, et al. Matrix-dependent Tiam1/Rac signaling in epithelial cells promotes either cell-cell adhesion or cell migration and is regulated by phosphatidylinositol 3-kinase. *J Cell Biol.* 1998;143(5):1385-98.
190. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, et al. A landscape of driver mutations in melanoma. *Cell.* 2012;150(2):251-63.

191. Zhao M, Mishra L, Deng CX. The role of TGF-beta/SMAD4 signaling in cancer. *Int J Biol Sci.* 2018;14(2):111-23.
192. Schallenberg S, Bork J, Essakly A, Alakus H, Buettner R, Hillmer AM, et al. Loss of the SWI/SNF-ATPase subunit members SMARCF1 (ARID1A), SMARCA2 (BRM), SMARCA4 (BRG1) and SMARCB1 (INI1) in oesophageal adenocarcinoma. *BMC Cancer.* 2020;20(1):12.
193. 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.
194. Song Y, Li L, Ou Y, Gao Z, Li E, Li X, et al. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature.* 2014;509(7498):91-5.
195. 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;48(10):1131-41.
196. Dunne MR, Michielsen AJ, O'Sullivan KE, Cathcart MC, Feighery R, Doyle B, et al. HLA-DR expression in tumor epithelium is an independent prognostic indicator in esophageal adenocarcinoma patients. *Cancer Immunol Immunother.* 2017;66(7):841-50.
197. Su Z, Gay LJ, Strange A, Palles C, Band G, Whiteman DC, et al. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. *Nat Genet.* 2012;44(10):1131-6.
198. Dura P, van Veen EM, Salomon J, te Morsche RH, Roelofs HM, Kristinsson JO, et al. Barrett associated MHC and FOXF1 variants also increase esophageal carcinoma risk. *Int J Cancer.* 2013;133(7):1751-5.
199. Zhang X, Lin A, Zhang JG, Bao WG, Xu DP, Ruan YY, et al. Alteration of HLA-F and HLA I antigen expression in the tumor is associated with survival in patients with esophageal squamous cell carcinoma. *Int J Cancer.* 2013;132(1):82-9.
200. Zheng J, Xu C, Chu D, Zhang X, Li J, Ji G, et al. Human leukocyte antigen G is associated with esophageal squamous cell carcinoma progression and poor prognosis. *Immunol Lett.* 2014;161(1):13-9.
201. Chen Q, Luo G, Zhang X. MiR-148a modulates HLA-G expression and influences tumor apoptosis in esophageal squamous cell carcinoma. *Exp Ther Med.* 2017;14(5):4448-52.
202. Hu J, Li L, Pang L, Chen Y, Yang L, Liu C, et al. HLA-DRB1*1501 and HLA-DQB1*0301 alleles are positively associated with HPV16 infection-related Kazakh esophageal squamous cell carcinoma in Xinjiang China. *Cancer Immunol Immunother.* 2012;61(11):2135-41.
203. Feng B, Awuti I, Deng Y, Li D, Niyazi M, Aniwar J, et al. Human papillomavirus promotes esophageal squamous cell carcinoma by regulating DNA methylation and expression of HLA-DQB1. *Asia Pac J Clin Oncol.* 2014;10(1):66-74.
204. Shen FF, Pan Y, Li JZ, Zhao F, Yang HJ, Li JK, et al. High expression of HLA-DQA1 predicts poor outcome in patients with esophageal squamous cell carcinoma in Northern China. *Medicine (Baltimore).* 2019;98(8):e14454.
205. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Research.* 2015;44(8):e71-e.
206. Hothorn T, Lausen B. Maximally selected rank statistics in R. *R News.* 2002;2(1):3-5.

List of References

207. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR genomics and bioinformatics*. 2020;2(3):lqaa078.
208. Hothorn T, Hothorn MT, Suggests T. Package ‘maxstat’. Published online. 2017.
209. Harrison E, Drake T, Ots R. Package “finalfit.”. Retrieved February. 2020;29:2020.
210. Stein AV, Dislich B, Blank A, Guldener L, Kröll D, Seiler CA, et al. High intratumoural but not peritumoural inflammatory host response is associated with better prognosis in primary resected oesophageal adenocarcinomas. *Pathology*. 2017;49(1):30-7.
211. Dhatchinamoorthy K, Colbert JD, Rock KL. Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation. *Front Immunol*. 2021;12:636568.
212. Zhang K, Wang H. [Cancer Genome Atlas Pan-cancer Analysis Project]. *Zhongguo Fei Ai Za Zhi*. 2015;18(4):219-23.
213. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, et al. The International Cancer Genome Consortium Data Portal. *Nat Biotechnol*. 2019;37(4):367-9.
214. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014;505(7484):495-501.
215. Mayakonda A. maf-tools: Summarize, Analyze and Visualize MAF files. 2019.
216. Consortium G. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318-30.
217. Collado-Torres L, Nellore A, Kammers K, Ellis SE, Taub MA, Hansen KD, et al. Reproducible RNA-seq analysis using recount2. *Nature biotechnology*. 2017;35(4):319-21.
218. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15(12):550.
219. Blighe K, S Rana, and M Lewis. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling 2018 [Available from: <https://github.com/kevinblighe/EnhancedVolcano>].
220. Li Y, Ge D, Lu C. The SMART App: an interactive web application for comprehensive DNA methylation analysis and visualization. *Epigenetics & Chromatin*. 2019;12(1):71.
221. Morgan E, Soerjomataram I, Rungay H, Coleman HG, Thrift AP, Vignat J, et al. The Global Landscape of Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma Incidence and Mortality in 2020 and Projections to 2040: New Estimates From GLOBOCAN 2020. *Gastroenterology*. 2022;163(3):649-58.e2.
222. Grotenhuis BA, van Hagen P, Wijnhoven BP, Spaander MC, Tilanus HW, van Lanschot JJ. Delay in diagnostic workup and treatment of esophageal cancer. *J Gastrointest Surg*. 2010;14(3):476-83.
223. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011;474(7353):609-15.
224. George J, Lim JS, Jang SJ, Cun Y, Ozretić L, Kong G, et al. Comprehensive genomic profiles of small cell lung cancer. *Nature*. 2015;524(7563):47-53.
225. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, et al. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep*. 2016;15(4):857-65.
226. Imaielski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 2012;150(6):1107-20.

227. Kanodia D, Nagata Y, Garg M, Lee DH, Sato A, Yoshida K, et al. Genomic landscape of liposarcoma. *Oncotarget*. 2015;6(40):42429-44.
228. Kumar A, Coleman I, Morrissey C, Zhang X, True LD, Gulati R, et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nat Med*. 2016;22(4):369-78.
229. Liu J, Lee W, Jiang Z, Chen Z, Jhunjhunwala S, Haverty PM, et al. Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. *Genome Res*. 2012;22(12):2315-27.
230. Liu Y, Easton J, Shao Y, Maciaszek J, Wang Z, Wilkinson MR, et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat Genet*. 2017;49(8):1211-8.
231. McMillan EA, Ryu MJ, Diep CH, Mendiratta S, Clemenceau JR, Vaden RM, et al. Chemistry-First Approach for Nomination of Personalized Treatment in Lung Cancer. *Cell*. 2018;173(4):864-78.e29.
232. van de Wetering M, Frances HE, Francis JM, Bounova G, Iorio F, Pronk A, et al. Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell*. 2015;161(4):933-45.
233. Yamamoto H, Perez-Piteira J, Yoshida T, Terada M, Itoh F, Imai K, et al. Gastric cancers of the microsatellite mutator phenotype display characteristic genetic and clinical features. *Gastroenterology*. 1999;116(6):1348-57.
234. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23(6):703-13.
235. Bailey A, Dalchau N, Carter R, Emmott S, Phillips A, Werner JM, et al. Selector function of MHC I molecules is determined by protein plasticity. *Sci Rep*. 2015;5:14928.
236. Frankell AM, Jammula S, Li X, Contino G, Killcoyne S, Abbas S, et al. The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat Genet*. 2019;51(3):506-16.
237. Zeng WZD, Glicksberg BS, Li Y, Chen B. Selecting precise reference normal tissue samples for cancer research using a deep learning approach. *BMC Med Genomics*. 2019;12(Suppl 1):21.
238. Middha S, Yaeger R, Shia J, Stadler ZK, King S, Guercio S, et al. Majority of B2M-Mutant and -Deficient Colorectal Carcinomas Achieve Clinical Benefit From Immune Checkpoint Inhibitor Therapy and Are Microsatellite Instability-High. *JCO Precis Oncol*. 2019;3.
239. Van Hateren A, James E, Bailey A, Phillips A, Dalchau N, Elliott T. The cell biology of major histocompatibility complex class I assembly: towards a molecular understanding. *Tissue Antigens*. 2010;76(4):259-75.
240. Castro A, Ozturk K, Pyke RM, Xian S, Zanetti M, Carter H. Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes. *BMC Medical Genomics*. 2019;12(6):107.
241. Janakiram NB, Rao CV. The role of inflammation in colon cancer. *Adv Exp Med Biol*. 2014;816:25-52.
242. Frankell AM, Jammula S, Li X, Contino G, Killcoyne S, Abbas S, et al. The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nature Genetics*. 2019;51(3):506-16.

List of References

243. Paulson TG, Maley CC, Li X, Li H, Sanchez CA, Chao DL, et al. Chromosomal instability and copy number alterations in Barrett's esophagus and esophageal adenocarcinoma. *Clin Cancer Res.* 2009;15(10):3305-14.
244. Lorenzi S, Forloni M, Cifaldi L, Antonucci C, Citti A, Boldrini R, et al. IRF1 and NF- κ B restore MHC class I-restricted tumor antigen processing and presentation to cytotoxic T cells in aggressive neuroblastoma. 2012.
245. Shao L, Hou W, Scharping NE, Vendetti FP, Srivastava R, Roy CN, et al. IRF1 Inhibits Antitumor Immunity through the Upregulation of PD-L1 in the Tumor Cell. *Cancer Immunology Research.* 2019;7(8):1258-66.
246. Guo L, Liu D. Identification of RFX5 as prognostic biomarker and associated with immune infiltration in stomach adenocarcinoma. *European Journal of Medical Research.* 2022;27(1):164.
247. Ding G, Xu X, Li D, Chen Y, Wang W, Ping D, et al. Fisetin inhibits proliferation of pancreatic adenocarcinoma by inducing DNA damage via RFXAP/KDM4A-dependent histone H3K36 demethylation. *Cell Death & Disease.* 2020;11(10):893.
248. Guo L, Liu D. Identification of RFX5 as prognostic biomarker and associated with immune infiltration in stomach adenocarcinoma. *Eur J Med Res.* 2022;27(1):164.
249. Scholl T, Mahanta SK, Strominger JL. Specific complex formation between the type II bare lymphocyte syndrome-associated transactivators CIITA and RFX5. *Proceedings of the National Academy of Sciences.* 1997;94(12):6330-4.
250. Cressman DE, Chin K-C, Taxman DJ, Ting JP-Y. A defect in the nuclear translocation of CIITA causes a form of type II bare lymphocyte syndrome. *Immunity.* 1999;10(2):163-71.
251. Johnson AM, Bullock BL, Neuwelt AJ, Poczobutt JM, Kaspar RE, Li HY, et al. Cancer Cell-Intrinsic Expression of MHC Class II Regulates the Immune Microenvironment and Response to Anti-PD-1 Therapy in Lung Adenocarcinoma. *J Immunol.* 2020;204(8):2295-307.
252. Ghasemi F, Tessier TM, Gameiro SF, Maciver AH, Cecchini MJ, Mymryk JS. High MHC-II expression in Epstein-Barr virus-associated gastric cancers suggests that tumor cells serve an important role in antigen presentation. *Scientific Reports.* 2020;10(1):14786.
253. van der Meer D, Barthorpe S, Yang W, Lightfoot H, Hall C, Gilbert J, et al. Cell Model Passports—a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Research.* 2019;47(D1):D923-D9.
254. Najafimehr H, Hajizadeh N, Nazemalhosseini-Mojarad E, Pourhoseingholi MA, Abdollahpour-Alitappeh M, Ashtari S, et al. The role of Human leukocyte antigen class I on patient survival in Gastrointestinal cancers: a systematic review and meta-analysis. *Scientific Reports.* 2020;10(1):728.
255. Shionoya Y, Kanaseki T, Miyamoto S, Tokita S, Hongo A, Kikuchi Y, et al. Loss of tapasin in human lung and colon cancer cells and escape from tumor-associated antigen-specific CTL recognition. *Oncoimmunology.* 2017;6(2):e1274476.
256. Sokol L, Koelzer VH, Rau TT, Karamitopoulou E, Zlobec I, Lugli A. Loss of tapasin correlates with diminished CD8+ T-cell immunity and prognosis in colorectal cancer. *Journal of Translational Medicine.* 2015;13(1):279.
257. Thuring C, Follin E, Geironson L, Freyhult E, Junghans V, Harndahl M, et al. HLA class I is most tightly linked to levels of tapasin compared with other antigen-processing proteins in glioblastoma. *British Journal of Cancer.* 2015;113(6):952-62.

258. Wang KW, Wang MD, Li ZX, Hu BS, Wu JJ, Yuan ZD, et al. An antigen processing and presentation signature for prognostic evaluation and immunotherapy selection in advanced gastric cancer. *Front Immunol.* 2022;13:992060.
259. Li L, Li L, Liu M, Li Y, Sun Q. Novel immune-related prognostic model and nomogram for breast cancer based on ssGSEA. *Front Genet.* 2022;13:957675.
260. Lin Y, Cui C, Su M, Silburt LK, Liu H, Zhao J, et al. Identification of TAPBPL as a novel negative regulator of T-cell function. *EMBO Mol Med.* 2021;13(5):e13404.
261. Wemeau M, Kepp O, Tesniere A, Panaretakis T, Flament C, De Botton S, et al. Calreticulin exposure on malignant blasts predicts a cellular anticancer immune response in patients with acute myeloid leukemia. *Cell death & disease.* 2010;1(12):e104-e.
262. Fucikova J, Truxova I, Hensler M, Becht E, Kasikova L, Moserova I, et al. Calreticulin exposure by malignant blasts correlates with robust anticancer immunity and improved clinical outcome in AML patients. *Blood, The Journal of the American Society of Hematology.* 2016;128(26):3113-24.
263. Schardt JA, Weber D, Eyholzer M, Mueller BU, Pabst T. Activation of the unfolded protein response is associated with favorable prognosis in acute myeloid leukemia. *Clinical cancer research.* 2009;15(11):3834-41.
264. Chao MP, Jaiswal S, Weissman-Tsukamoto R, Alizadeh AA, Gentles AJ, Volkmer J, et al. Calreticulin is the dominant pro-phagocytic signal on multiple human cancers and is counterbalanced by CD47. *Science translational medicine.* 2010;2(63):63ra94-63ra94.
265. Peng R-Q, Chen Y-B, Ding Y, Zhang R, Zhang X, Yu X-J, et al. Expression of calreticulin is associated with infiltration of T-cells in stage IIIB colon cancer. *World journal of gastroenterology: WJG.* 2010;16(19):2428.
266. Chen C-N, Chang C-C, Su T-E, Hsu W-M, Jeng Y-M, Ho M-C, et al. Identification of calreticulin as a prognosis marker and angiogenic regulator in human gastric cancer. *Annals of surgical oncology.* 2009;16:524-33.
267. Muth C, Semrau S, Rühle P-f, Frey B, Strnad A, Buslei R, et al. Primary glioblastoma multiforme tumors and recurrence. *Strahlentherapie und Onkologie.* 2016;192(3):146.
268. Hsu W-M, Hsieh F-J, Jeng Y-M, Kuo M-L, Chen C-N, Lai D-M, et al. Calreticulin expression in neuroblastoma—a novel independent prognostic factor. *Annals of Oncology.* 2005;16(2):314-21.
269. Garg AD, Elsen S, Krysko DV, Vandenabeele P, de Witte P, Agostinis P. Resistance to anticancer vaccination effect is controlled by a cancer cell-autonomous phenotype that disrupts immunogenic phagocytic removal. *Oncotarget.* 2015;6(29):26841.
270. Liu R, Gong J, Chen J, Li Q, Song C, Zhang J, et al. Calreticulin as a potential diagnostic biomarker for lung cancer. *Cancer Immunology, Immunotherapy.* 2012;61:855-64.
271. Fucikova J, Becht E, Iribarren K, Goc J, Remark R, Damotte D, et al. Calreticulin expression in human Non-Small cell lung cancers correlates with increased accumulation of antitumor immune cells and favorable prognosis. *Cancer research.* 2016;76(7):1746-56.
272. Zhang X-H, Zhang Y, Xie W-P, Sun D-S, Zhang Y-K, Hao Y-K, et al. Expression and significance of calreticulin in human osteosarcoma. *Cancer Biomarkers.* 2017;18(4):405-11.
273. Kasikova L, Hensler M, Truxova I, Skapa P, Laco J, Belicova L, et al. Calreticulin exposure correlates with robust adaptive antitumor immunity and favorable prognosis in ovarian carcinoma patients. *Journal for immunotherapy of cancer.* 2019;7(1):1-12.

List of References

274. Sheng W, Chen C, Dong M, Wang G, Zhou J, Song H, et al. Calreticulin promotes EGF-induced EMT in pancreatic cancer cells via Integrin/EGFR-ERK/MAPK signaling pathway. *Cell death & disease*. 2017;8(10):e3147-e.
275. Matsukuma S, Yoshimura K, Ueno T, Oga A, Inoue M, Watanabe Y, et al. Calreticulin is highly expressed in pancreatic cancer stem-like cells. *Cancer Science*. 2016;107(11):1599-609.
276. Cathro HP, Smolkin ME, Theodorescu D, Jo VY, Ferrone S, Frierson HF. Relationship between HLA class I antigen processing machinery component expression and the clinicopathologic characteristics of bladder carcinomas. *Cancer Immunology, Immunotherapy*. 2010;59:465-72.
277. Li Y, Liu X, Chen H, Xie P, Ma R, He J, et al. Bioinformatics analysis for the role of CALR in human cancers. *PLOS ONE*. 2021;16(12):e0261254.
278. Wang L, Chen J, Zuo Q, Wu C, Yu T, Zheng P, et al. Calreticulin enhances gastric cancer metastasis by dimethylating H3K9 in the E-cadherin promoter region mediating by G9a. *Oncogenesis*. 2022;11(1):29.
279. Sun D, Li C, Zhang F. MicroRNA-206 suppresses growth and metastasis of breast cancer stem cells via blocking EVI-1-mediated CALR expression. *PLOS ONE*. 2022;17(9):e0274919.
280. Margarido A, Whitton B, Arwert E, Cheeseman M, Kalusa A, Davison G, et al. Investigating the effects of ERAP1 inhibition on antigen presentation and the tumour immune response. *European Journal of Cancer*. 2022;174:S126.
281. Leishman A, Paes W, Coulson R, Pinggera M, Sette J, Anderton K, et al. Abstract 2065: A small molecule approach to drive novel neoantigen generation: First-in-class inhibitors of ERAP1 generate novel neoantigens driving anti-tumor effects. *Cancer Research*. 2022;82(12_Supplement):2065-.
282. Yang Z, Tian H, Bie F, Xu J, Zhou Z, Yang J, et al. ERAP2 Is Associated With Immune Infiltration and Predicts Favorable Prognosis in SqCLC. *Front Immunol*. 2021;12:788985.
283. Temponeras I, Stamatakis G, Samiotaki M, Georgiadis D, Pratsinis H, Panayotou G, et al. ERAP2 Inhibition Induces Cell-Surface Presentation by MOLT-4 Leukemia Cancer Cells of Many Novel and Potentially Antigenic Peptides. *Int J Mol Sci*. 2022;23(3).
284. Andrés AM, Dennis MY, Kretzschmar WW, Cannons JL, Lee-Lin S-Q, Hurle B, et al. Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. *PLoS genetics*. 2010;6(10):e1001157.
285. Cabrera C, Jimenez P, Cabrera T, Esparza C, Ruiz-Cabello F, Garrido F. Total loss of MHC class I in colorectal tumors can be explained by two molecular pathways: β 2-microglobulin inactivation in MSI-positive tumors and LMP7/TAP2 downregulation in MSI-negative tumors. *Tissue antigens*. 2003;61(3):211-9.
286. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419.
287. Ahmed I, Ismail N. M1 and M2 Macrophages Polarization via mTORC1 Influences Innate Immunity and Outcome of *Ehrlichia* Infection. *J Cell Immunol*. 2020;2(3):108-15.
288. Baumgart M, Moos V, Schuhbauer D, Müller B. Differential expression of major histocompatibility complex class II genes on murine macrophages associated with T cell cytokine profile and protective/suppressive effects. *Proc Natl Acad Sci U S A*. 1998;95(12):6936-40.
289. Villadangos JA, Cardoso Ma, Steptoe RJ, van Berkel D, Pooley J, Carbone FR, et al. MHC Class II Expression Is Regulated in Dendritic Cells Independently of Invariant Chain Degradation. *Immunity*. 2001;14(6):739-49.

290. Senosain M-F, Zou Y, Novitskaya T, Vasiukov G, Balar AB, Rowe DJ, et al. HLA-DR cancer cells expression correlates with T cell infiltration and is enriched in lung adenocarcinoma with indolent behavior. *Scientific Reports*. 2021;11(1):14424.
291. Mei J, Jiang G, Chen Y, Xu Y, Wan Y, Chen R, et al. HLA class II molecule HLA-DRA identifies immuno-hot tumors and predicts the therapeutic response to anti-PD-1 immunotherapy in NSCLC. *BMC Cancer*. 2022;22(1):738.
292. Chen YY, Chang WA, Lin ES, Chen YJ, Kuo PL. Expressions of HLA Class II Genes in Cutaneous Melanoma Were Associated with Clinical Outcome: Bioinformatics Approaches and Systematic Analysis of Public Microarray and RNA-Seq Datasets. *Diagnostics (Basel)*. 2019;9(2).
293. Gong X, Karchin R. Pan-cancer HLA Gene-mediated Tumor Immunogenicity and Immune Evasion. *bioRxiv*. 2021:2021.05.17.444511.
294. Shen F-F, Pan Y, Li J-Z, Zhao F, Yang H-J, Li J-K, et al. High expression of HLA-DQA1 predicts poor outcome in patients with esophageal squamous cell carcinoma in Northern China. *Medicine*. 2019;98(8).
295. Silva GBRFd, Silva TGA, Duarte RA, Neto NL, Carrara HHA, Donadi EA, et al. Expression of the Classical and Nonclassical HLA Molecules in Breast Cancer. *International Journal of Breast Cancer*. 2013;2013:250435.
296. Tariq F, Khan W, Ahmad W, Riaz SK, Khan M, Sherwani S, et al. Effect of MHC Linked 7-Gene Signature on Delayed Hepatocellular Carcinoma Recurrence. *J Pers Med*. 2021;11(11).
297. Morris KT, Nofchissey RA, Pinchuk IV, Beswick EJ. Chronic Macrophage Migration Inhibitory Factor Exposure Induces Mesenchymal Epithelial Transition and Promotes Gastric and Colon Cancers. *PLOS ONE*. 2014;9(6):e98656.
298. Wang ZQ, Milne K, Webb JR, Watson PH. CD74 and intratumoral immune response in breast cancer. *Oncotarget*. 2017;8(8):12664-74.
299. Xu S, Li X, Tang L, Liu Z, Yang K, Cheng Q. CD74 Correlated With Malignancies and Immune Microenvironment in Gliomas. *Front Mol Biosci*. 2021;8:706949.
300. Gradtke A-C, Mentrup T, Lehmann CHK, Cabrera-Cabrera F, Desel C, Okakpu D, et al. Deficiency of the Intramembrane Protease SPPL2a Alters Antimycobacterial Cytokine Responses of Dendritic Cells. *The Journal of Immunology*. 2021;206(1):164-80.
301. Mentrup T, Cabrera-Cabrera F, Fluhrer R, Schröder B. Physiological functions of SPP/SPPL intramembrane proteases. *Cellular and Molecular Life Sciences*. 2020;77(15):2959-79.
302. Talamillo A, Grande L, Ruiz-Ontañon P, Velasquez C, Mollinedo P, Torices S, et al. ODZ1 allows glioblastoma to sustain invasiveness through a Myc-dependent transcriptional upregulation of RhoA. *Oncogene*. 2017;36(12):1733-44.
303. Yixuan Y, Kiat LS, Yee CL, Huiyin L, Yunhao C, Kuan CP, et al. Cathepsin S Mediates Gastric Cancer Cell Migration and Invasion via a Putative Network of Metastasis-Associated Proteins. *Journal of Proteome Research*. 2010;9(9):4767-78.
304. Lin H-H, Chen S-J, Shen M-R, Huang Y-T, Hsieh H-P, Lin S-Y, et al. Lysosomal cysteine protease cathepsin S is involved in cancer cell motility by regulating store-operated Ca²⁺ entry. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2019;1866(12):118517.
305. Willumsen N, Bager CL, Leeming DJ, Bay-Jensen A-C, Karsdal MA. Nidogen-1 degraded by cathepsin S can be quantified in serum and is associated with non-small cell lung cancer. *Neoplasia*. 2017;19(4):271-8.

List of References

306. Zhang L, Wang H, Xu J. Cathepsin S as a cancer target. *Neoplasma*. 2015;62(1):16-26.
307. Wilkinson RDA, Young A, Burden RE, Williams R, Scott CJ. A bioavailable cathepsin S nitrile inhibitor abrogates tumor development. *Molecular Cancer*. 2016;15(1):29.
308. Zhang D-y, Liu Y, Chen Z, Zheng J, Zhao Z. CTSS is associated with immunity and affects prognosis in pan-cancer2023.
309. Wang Y, Zhang S, Wang H, Cui Y, Wang Z, Cheng X, et al. High Level of Legumain Was Correlated With Worse Prognosis and Peritoneal Metastasis in Gastric Cancer Patients. *Front Oncol*. 2020;10:966.
310. Wang H, Chen B, Lin Y, Zhou Y, Li X. Legumain Promotes Gastric Cancer Progression Through Tumor-associated Macrophages In vitro and In vivo. *Int J Biol Sci*. 2020;16(1):172-80.
311. Zhen Y, Chunlei G, Wenzhi S, Shuangtao Z, Na L, Rongrong W, et al. Clinicopathologic significance of legumain overexpression in cancer: a systematic review and meta-analysis. *Scientific Reports*. 2015;5(1):16599.
312. Crowther MD, Dolton G, Legut M, Caillaud ME, Lloyd A, Attaf M, et al. Genome-wide CRISPR–Cas9 screening reveals ubiquitous T cell cancer targeting via the monomorphic MHC class I-related protein MR1. *Nature Immunology*. 2020;21(2):178-85.
313. Adam P, Vitor HT, Khalid A, Shan EAR, Tom L, Ayse A, et al. Immune surveillance in clinical regression of pre-invasive squamous cell lung cancer. *bioRxiv*. 2019:833004.
314. Swets M, Seneby L, Boot A, van Wezel T, Gelderblom H, van de Velde CJH, et al. Promoter methylation and mRNA expression of HLA-G in relation to HLA-G protein expression in colorectal cancer. *Human Immunology*. 2016;77(9):764-72.
315. Naji A, Le Rond S, Durrbach A, Krawice-Radanne I, Creput C, Daouya M, et al. CD3+CD4low and CD3+CD8low are induced by HLA-G: novel human peripheral blood suppressor T-cell subsets involved in transplant acceptance. *Blood*. 2007;110(12):3936-48.
316. Jung G, Hernández-Illán E, Lozano JJ, Sidorova J, Muñoz J, Okada Y, et al. Epigenome-Wide DNA Methylation Profiling of Normal Mucosa Reveals HLA-F Hypermethylation as a Biomarker Candidate for Serrated Polyposis Syndrome. *The Journal of Molecular Diagnostics*. 2022;24(6):674-86.
317. Li C, Zheng Y, Pu K, Zhao D, Wang Y, Guan Q, et al. A four-DNA methylation signature as a novel prognostic biomarker for survival of patients with gastric cancer. *Cancer Cell Int*. 2020;20:88.
318. Londhe P, Zhu B, Abraham J, Keller C, Davie J. CIITA is silenced by epigenetic mechanisms that prevent the recruitment of transactivating factors in rhabdomyosarcoma cells. *Int J Cancer*. 2012;131(4):E437-48.
319. Morimoto Y, Toyota M, Satoh A, Murai M, Mita H, Suzuki H, et al. Inactivation of class II transactivator by DNA methylation and histone deacetylation associated with absence of HLA-DR induction by interferon- γ in haematopoietic tumour cells. *British Journal of Cancer*. 2004;90(4):844-52.
320. Nakamura MC. CIITA: a master regulator of adaptive immunity shows its innate side in the bone. *J Bone Miner Res*. 2014;29(2):287-9.
321. Lv J, Zhou Y, Zhou N, Wang Z, Chen J, Chen H, et al. Epigenetic modification of CSDE1 locus dictates immune recognition of nascent tumorigenic cells. *Sci Transl Med*. 2023;15(681):eabq6024.

322. Yang Z, Liu F, Li Z, Liu N, Yao X, Zhou Y, et al. Histone lysine methyltransferase SMYD3 promotes oral squamous cell carcinoma tumorigenesis via H3K4me3-mediated HMGA2 transcription. *Clinical Epigenetics*. 2023;15(1):92.
323. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol*. 2018;1711:243-59.
324. Tsoucas D, Dong R, Chen H, Zhu Q, Guo G, Yuan G-C. Accurate estimation of cell-type composition from gene expression data. *Nature Communications*. 2019;10(1):2975.
325. Hao Y, Yan M, Heath BR, Lei YL, Xie Y. Fast and robust deconvolution of tumor infiltrating lymphocyte from expression profiles using least trimmed squares. *PLoS Comput Biol*. 2019;15(5):e1006976.
326. Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D, et al. Package 'mass'. *Cran r*. 2013;538:113-20.
327. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods*. 2015;12(5):453-7.
328. Avila Cobos F, Vandesompele J, Mestdagh P, De Preter K. Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics*. 2018;34(11):1969-79.
329. Li Z, Wu H. TOAST: improving reference-free cell composition estimation by cross-cell type differential analysis. *Genome Biology*. 2019;20(1):190.
330. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nature Biotechnology*. 2019;37(7):773-82.
331. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nature Medicine*. 2015;21(8):938-45.
332. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics*. 2019;35(14):i436-i45.
333. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology*. 2017;18(1):220.
334. Marderstein AR, Uppal M, Verma A, Bhinder B, Tayyebi Z, Mezey J, et al. Demographic and genetic factors influence the abundance of infiltrating immune cells in human tissues. *Nature Communications*. 2020;11(1):2213.
335. Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P, De Preter K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nature Communications*. 2020;11(1):5650.
336. Rusk N. Expanded CIBERSORTx. *Nature Methods*. 2019;16(7):577-.
337. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife*. 2017;6:e26476.
338. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology*. 2016;17(1):218.

List of References

339. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Medicine*. 2019;11(1):34.
340. Li B, Severson E, Pignon J-C, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biology*. 2016;17(1):174.
341. Hou Y, Chen Z, Wang L, Deng Y, Liu G, Zhou Y, et al. Characterization of Immune-Related Genes and Immune Infiltration Features in Epilepsy by Multi-Transcriptome Data. *J Inflamm Res*. 2022;15:2855-76.
342. Tilsed CM, Principe N, Kidman J, Chin WL, Orozco Morales ML, Zemek RM, et al. CD4(+) T cells drive an inflammatory, TNF- α /IFN-rich tumor microenvironment responsive to chemotherapy. *Cell Rep*. 2022;41(13):111874.
343. Wang X, Park J, Susztak K, Zhang NR, Li M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nature Communications*. 2019;10(1):380.
344. Kalatskaya I, Giovannoni G, Leist T, Cerra J, Boschert U, Rolfe PA. Revealing the immune cell subtype reconstitution profile in patients from the CLARITY study using deconvolution algorithms after cladribine tablets treatment. *Scientific Reports*. 2023;13(1):8067.
345. Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, Carré C, et al. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Reports*. 2019;26(6):1627-40.e7.
346. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-74.
347. Ebbing EA, van der Zalm AP, Steins A, Creemers A, Hermsen S, Rentenaar R, et al. Stromal-derived interleukin 6 drives epithelial-to-mesenchymal transition and therapy resistance in esophageal adenocarcinoma. *Proc Natl Acad Sci U S A*. 2019;116(6):2237-42.
348. Vinocha A, Grover RK, Deepak R. Clinical significance of interleukin-6 in diagnosis of lung, oral, esophageal, and gall bladder carcinomas. *J Cancer Res Ther*. 2018;14(Supplement):S758-s60.
349. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Communications*. 2013;4(1):2612.
350. Huai Q, Guo W, Han L, Kong D, Zhao L, Song P, et al. Identification of prognostic genes and tumor-infiltrating immune cells in the tumor microenvironment of esophageal squamous cell carcinoma and esophageal adenocarcinoma. *Transl Cancer Res*. 2021;10(4):1787-803.
351. Ron Y, De Baetselier P, Gordon J, Feldman M, Segal S. Defective induction of antigen-reactive proliferating T cells in B cell-deprived mice. *Eur J Immunol*. 1981;11(12):964-8.
352. Bouaziz JD, Yanaba K, Venturi GM, Wang Y, Tisch RM, Poe JC, et al. Therapeutic B cell depletion impairs adaptive and autoreactive CD4+ T cell activation in mice. *Proc Natl Acad Sci U S A*. 2007;104(52):20878-83.
353. Ron Y, Sprent J. T cell priming in vivo: a major role for B cells in presenting antigen to T cells in lymph nodes. *J Immunol*. 1987;138(9):2848-56.
354. Harris DP, Haynes L, Sayles PC, Duso DK, Eaton SM, Lepak NM, et al. Reciprocal regulation of polarized cytokine production by effector B and T cells. *Nat Immunol*. 2000;1(6):475-82.

355. Schlößer HA, Thelen M, Lechner A, Wennhold K, Garcia-Marquez MA, Rothschild SI, et al. B cells in esophago-gastric adenocarcinoma are highly differentiated, organize in tertiary lymphoid structures and produce tumor-specific antibodies. *Oncoimmunology*. 2019;8(1):e1512458.
356. Harty JT, Tvinneim AR, White DW. CD8+ T cell effector mechanisms in resistance to infection. *Annu Rev Immunol*. 2000;18:275-308.
357. Groscurth P, Filgueira L. Killing mechanisms of cytotoxic T lymphocytes. *Physiology*. 1998;13(1):17-21.
358. Buchbinder E, Hodi FS. Cytotoxic T lymphocyte antigen-4 and immune checkpoint blockade. *J Clin Invest*. 2015;125(9):3377-83.
359. Zayac A, Almhanna K. Esophageal, gastric cancer and immunotherapy: small steps in the right direction? *Transl Gastroenterol Hepatol*. 2020;5:9.
360. Noble A, Giorgini A, Leggat JA. Cytokine-induced IL-10-secreting CD8 T cells represent a phenotypically distinct suppressor T-cell lineage. *Blood*. 2006;107(11):4475-83.
361. Yu Y, Ma X, Gong R, Zhu J, Wei L, Yao J. Recent advances in CD8(+) regulatory T cell research. *Oncol Lett*. 2018;15(6):8187-94.
362. Noble F, Mellows T, McCormick Matthews LH, Bateman AC, Harris S, Underwood TJ, et al. Tumour infiltrating lymphocytes correlate with improved survival in patients with oesophageal adenocarcinoma. *Cancer Immunol Immunother*. 2016;65(6):651-62.
363. Szabo SJ, Sullivan BM, Peng SL, Glimcher LH. Molecular Mechanisms Regulating Th1 Immune Responses. *Annual Review of Immunology*. 2003;21(1):713-58.
364. Harrington LE, Hatton RD, Mangan PR, Turner H, Murphy TL, Murphy KM, et al. Interleukin 17-producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. *Nature Immunology*. 2005;6(11):1123-32.
365. Park H, Li Z, Yang XO, Chang SH, Nurieva R, Wang YH, et al. A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. *Nat Immunol*. 2005;6(11):1133-41.
366. Kelchtermans H, Schurgers E, Geboes L, Mitera T, Van Damme J, Van Snick J, et al. Effector mechanisms of interleukin-17 in collagen-induced arthritis in the absence of interferon-gamma and counteraction by interferon-gamma. *Arthritis Res Ther*. 2009;11(4):R122.
367. Wang KS, Frank DA, Ritz J. Interleukin-2 enhances the response of natural killer cells to interleukin-12 through up-regulation of the interleukin-12 receptor and STAT4. *Blood*. 2000;95(10):3183-90.
368. Desmedt M, Rottiers P, Dooms H, Fiers W, Grootenhuis J. Macrophages induce cellular immunity by activating Th1 cell responses and suppressing Th2 cell responses. *J Immunol*. 1998;160(11):5300-8.
369. Raphael I, Nalawade S, Eagar TN, Forsthäuser TG. T cell subsets and their signature cytokines in autoimmune and inflammatory diseases. *Cytokine*. 2015;74(1):5-17.
370. Zhu J. T helper 2 (Th2) cell differentiation, type 2 innate lymphoid cell (ILC2) development and regulation of interleukin-4 (IL-4) and IL-13 production. *Cytokine*. 2015;75(1):14-24.
371. Yang XO, Pappu BP, Nurieva R, Akimzhanov A, Kang HS, Chung Y, et al. T helper 17 lineage differentiation is programmed by orphan nuclear receptors ROR alpha and ROR gamma. *Immunity*. 2008;28(1):29-39.

List of References

372. Sandquist I, Kolls J. Update on regulation and effector functions of Th17 cells. *F1000Res.* 2018;7:205.
373. Nurieva R, Yang XO, Martinez G, Zhang Y, Panopoulos AD, Ma L, et al. Essential autocrine regulation by IL-21 in the generation of inflammatory T cells. *Nature.* 2007;448(7152):480-3.
374. Liu D, Zhang R, Wu J, Pu Y, Yin X, Cheng Y, et al. Interleukin-17A promotes esophageal adenocarcinoma cell invasiveness through ROS-dependent, NF-κB-mediated MMP-2/9 activation. *Oncol Rep.* 2017;37(3):1779-85.
375. Liu J, Luo Y, Wang J, Xi C, Chen Y, Yang G, et al. Key molecules involved in the Th17/Treg balance are associated with the pathogenesis of reflux esophagitis and Barrett's esophagus. *Esophagus.* 2021;18(2):388-97.
376. Sakaguchi S, Sakaguchi N, Asano M, Itoh M, Toda M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J Immunol.* 1995;155(3):1151-64.
377. Li MO, Wan YY, Sanjabi S, Robertson AK, Flavell RA. Transforming growth factor-beta regulation of immune responses. *Annu Rev Immunol.* 2006;24:99-146.
378. Zingg U, Montani M, Frey DM, Dirnhofer S, Esterman AJ, Went P, et al. Tumour-infiltrating lymphocytes and survival in patients with adenocarcinoma of the oesophagus. *Eur J Surg Oncol.* 2010;36(7):670-7.
379. Böttcher JP, Bonavita E, Chakravarty P, Blees H, Cabeza-Cabrerizo M, Sammicheli S, et al. NK Cells Stimulate Recruitment of cDC1 into the Tumor Microenvironment Promoting Cancer Immune Control. *Cell.* 2018;172(5):1022-37.e14.
380. Adrain C, Murphy BM, Martin SJ. Molecular ordering of the caspase activation cascade initiated by the cytotoxic T lymphocyte/natural killer (CTL/NK) protease granzyme B. *J Biol Chem.* 2005;280(6):4663-73.
381. Phillips JH, Le AM, Lanier LL. Natural killer cells activated in a human mixed lymphocyte response culture identified by expression of Leu-11 and class II histocompatibility antigens. *J Exp Med.* 1984;159(4):993-1008.
382. Madjd Z, Spendlove I, Pinder SE, Ellis IO, Durrant LG. Total loss of MHC class I is an independent indicator of good prognosis in breast cancer. *Int J Cancer.* 2005;117(2):248-55.
383. Svensson MC, Warfvinge CF, Fristedt R, Hedner C, Borg D, Eberhard J, et al. The integrative clinical impact of tumor-infiltrating T lymphocytes and NK cells in relation to B lymphocyte and plasma cell density in esophageal and gastric adenocarcinoma. *Oncotarget.* 2017;8(42):72108-26.
384. Izawa S, Kono K, Mimura K, Kawaguchi Y, Watanabe M, Maruyama T, et al. H2O2 production within tumor microenvironment inversely correlated with infiltration of CD56dim NK cells in gastric and esophageal cancer: possible mechanisms of NK cell dysfunction. *Cancer Immunology, Immunotherapy.* 2011;60(12):1801-10.
385. Schmall A, Al-Tamari HM, Herold S, Kampschulte M, Weigert A, Wietelmann A, et al. Macrophage and cancer cell cross-talk via CCR2 and CX3CR1 is a fundamental mechanism driving lung cancer. *Am J Respir Crit Care Med.* 2015;191(4):437-47.
386. Qian B-Z, Li J, Zhang H, Kitamura T, Zhang J, Campion LR, et al. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature.* 2011;475(7355):222-5.
387. Olingy CE, Dinh HQ, Hedrick CC. Monocyte heterogeneity and functions in cancer. *J Leukoc Biol.* 2019;106(2):309-22.

388. Yeap WH, Wong KL, Shimasaki N, Teo EC, Quek JK, Yong HX, et al. CD16 is indispensable for antibody-dependent cellular cytotoxicity by human monocytes. *Sci Rep.* 2016;6:34310.
389. Hanna RN, Cekic C, Sag D, Tacke R, Thomas GD, Nowyhed H, et al. Patrolling monocytes control tumor metastasis to the lung. *Science.* 2015;350(6263):985-90.
390. Headley MB, Bins A, Nip A, Roberts EW, Looney MR, Gerard A, et al. Visualization of immediate immune responses to pioneer metastatic cells in the lung. *Nature.* 2016;531(7595):513-7.
391. Lavin Y, Kobayashi S, Leader A, Amir E-aD, Elefant N, Bigenwald C, et al. Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. *Cell.* 2017;169(4):750-65.e17.
392. Romano E, Kusio-Kobialka M, Foukas PG, Baumgaertner P, Meyer C, Ballabeni P, et al. Ipilimumab-dependent cell-mediated cytotoxicity of regulatory T cells ex vivo by nonclassical monocytes in melanoma patients. *Proceedings of the National Academy of Sciences.* 2015;112(19):6140-5.
393. Gordon IO, Freedman RS. Defective antitumor function of monocyte-derived macrophages from epithelial ovarian cancer patients. *Clin Cancer Res.* 2006;12(5):1515-24.
394. Griffith TS, Wiley SR, Kubin MZ, Sedger LM, Maliszewski CR, Fanger NA. Monocyte-mediated tumoricidal activity via the tumor necrosis factor-related cytokine, TRAIL. *J Exp Med.* 1999;189(8):1343-54.
395. Mantovani A, Sozzani S, Locati M, Allavena P, Sica A. Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends Immunol.* 2002;23(11):549-55.
396. Yao Y, Xu XH, Jin L. Macrophage Polarization in Physiological and Pathological Pregnancy. *Front Immunol.* 2019;10:792.
397. Mantovani A, Sica A, Sozzani S, Allavena P, Vecchi A, Locati M. The chemokine system in diverse forms of macrophage activation and polarization. *Trends Immunol.* 2004;25(12):677-86.
398. Flanagan RS, Heit B, Heinrichs DE. Antimicrobial Mechanisms of Macrophages and the Immune Evasion Strategies of *Staphylococcus aureus*. *Pathogens.* 2015;4(4):826-68.
399. Stein M, Keshav S, Harris N, Gordon S. Interleukin 4 potently enhances murine macrophage mannose receptor activity: a marker of alternative immunologic macrophage activation. *J Exp Med.* 1992;176(1):287-92.
400. Gordon S, Martinez FO. Alternative activation of macrophages: mechanism and functions. *Immunity.* 2010;32(5):593-604.
401. Cao W, Peters JH, Nieman D, Sharma M, Watson T, Yu J. Macrophage subtype predicts lymph node metastasis in oesophageal adenocarcinoma and promotes cancer cell invasion in vitro. *Br J Cancer.* 2015;113(5):738-46.
402. Wculek SK, Cueto FJ, Mujal AM, Melero I, Krummel MF, Sancho D. Dendritic cells in cancer immunology and immunotherapy. *Nature Reviews Immunology.* 2020;20(1):7-24.
403. Binnewies M, Mujal AM, Pollack JL, Combes AJ, Hardison EA, Barry KC, et al. Unleashing Type-2 Dendritic Cells to Drive Protective Antitumor CD4(+) T Cell Immunity. *Cell.* 2019;177(3):556-71.e16.

List of References

404. Bobryshev YV, Tran D, Killingsworth MC, Buckland M, Lord RV. Dendritic cell-associated immune inflammation of cardiac mucosa: a possible factor in the formation of Barrett's esophagus. *J Gastrointest Surg.* 2009;13(3):442-50.
405. Almand B, Resser JR, Lindman B, Nadaf S, Clark JI, Kwon ED, et al. Clinical significance of defective dendritic cell differentiation in cancer. *Clin Cancer Res.* 2000;6(5):1755-66.
406. Patente TA, Pinho MP, Oliveira AA, Evangelista GCM, Bergami-Santos PC, Barbuto JAM. Human Dendritic Cells: Their Heterogeneity and Clinical Application Potential in Cancer Immunotherapy. *Front Immunol.* 2018;9:3176.
407. Norrby K. Mast cells and angiogenesis. *Apmis.* 2002;110(5):355-71.
408. Nakae S, Suto H, Kakurai M, Sedgwick JD, Tsai M, Galli SJ. Mast cells enhance T cell activation: Importance of mast cell-derived TNF. *Proceedings of the National Academy of Sciences.* 2005;102(18):6467-72.
409. Maciel TT, Moura IC, Hermine O. The role of mast cells in cancers. *F1000Prime Rep.* 2015;7:09.
410. Dos Santos Cunha AC, Simon AG, Zander T, Buettner R, Bruns CJ, Schroeder W, et al. Dissecting the inflammatory tumor microenvironment of esophageal adenocarcinoma: mast cells and natural killer cells are favorable prognostic factors and associated with less extensive disease. *J Cancer Res Clin Oncol.* 2023;149(10):6917-29.
411. Cai SW, Yang SZ, Gao J, Pan K, Chen JY, Wang YL, et al. Prognostic significance of mast cell count following curative resection for pancreatic ductal adenocarcinoma. *Surgery.* 2011;149(4):576-84.
412. Weller PF, Spencer LA. Functions of tissue-resident eosinophils. *Nat Rev Immunol.* 2017;17(12):746-60.
413. Carretero R, Sektioglu IM, Garbi N, Salgado OC, Beckhove P, Hämmерling GJ. Eosinophils orchestrate cancer rejection by normalizing tumor vessels and enhancing infiltration of CD8(+) T cells. *Nat Immunol.* 2015;16(6):609-17.
414. dos Santos Cunha AC, Simon AG, Zander T, Buettner R, Bruns CJ, Schroeder W, et al. Dissecting the inflammatory tumor microenvironment of esophageal adenocarcinoma: mast cells and natural killer cells are favorable prognostic factors and associated with less extensive disease. *Journal of Cancer Research and Clinical Oncology.* 2023;149(10):6917-29.
415. Rosales C. Neutrophil: A Cell with Many Roles in Inflammation or Several Cell Types? *Front Physiol.* 2018;9:113.
416. Fridlender ZG, Sun J, Kim S, Kapoor V, Cheng G, Ling L, et al. Polarization of tumor-associated neutrophil phenotype by TGF-beta: "N1" versus "N2" TAN. *Cancer Cell.* 2009;16(3):183-94.
417. Wislez M, Rabbe N, Marchal J, Milleron B, Crestani B, Mayaud C, et al. Hepatocyte growth factor production by neutrophils infiltrating bronchioloalveolar subtype pulmonary adenocarcinoma: role in tumor progression and death. *Cancer Res.* 2003;63(6):1405-12.
418. Aoyagi Y, Oda T, Kinoshita T, Nakahashi C, Hasebe T, Ohkohchi N, et al. Overexpression of TGF-beta by infiltrated granulocytes correlates with the expression of collagen mRNA in pancreatic cancer. *Br J Cancer.* 2004;91(7):1316-26.
419. Kolde R, Kolde MR. Package 'pheatmap'. R package. 2015;1(7):790.

420. WAKIYAMA H, MASUDA T, MOTOMURA Y, HU Q, TOBO T, EGUCHI H, et al. Cytolytic Activity (CYT) Score Is a Prognostic Biomarker Reflecting Host Immune Status in Hepatocellular Carcinoma (HCC). *Anticancer Research*. 2018;38(12):6631-8.
421. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res*. 2020;48(W1):W509-w14.
422. Kassambara A, Kassambara MA. Package 'ggcorrplot'. R package version 01. *2019*;3(3):908.
423. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Cancer Systems Biology: Methods and Protocols*. 2018:243-59.
424. Jung K, Heishi T, Khan OF, Kowalski PS, Incio J, Rahbari NN, et al. Ly6C lo monocytes drive immunosuppression and confer resistance to anti-VEGFR2 cancer therapy. *The Journal of clinical investigation*. 2017;127(8):3039-51.
425. Sanford DE, Belt BA, Panni RZ, Mayer A, Deshpande AD, Carpenter D, et al. Inflammatory monocyte mobilization decreases patient survival in pancreatic cancer: a role for targeting the CCL2/CCR2 axis. *Clinical cancer research*. 2013;19(13):3404-15.
426. Gordon IO, Freedman RS. Defective antitumor function of monocyte-derived macrophages from epithelial ovarian cancer patients. *Clinical cancer research*. 2006;12(5):1515-24.
427. Sheng J, Chen Q, Soncin I, Ng SL, Karjalainen K, Ruedl C. A discrete subset of monocyte-derived cells among typical conventional type 2 dendritic cells can efficiently cross-present. *Cell reports*. 2017;21(5):1203-14.
428. Carroll TM, Chadwick JA, Owen RP, White MJ, Kaplinsky J, Peneva I, et al. Tumor monocyte content predicts immunochemotherapy outcomes in esophageal adenocarcinoma. *Cancer Cell*. 2023;41(7):1222-41.e7.
429. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*. 2014;515(7528):568-71.
430. Taube JM. Unleashing the immune system: PD-1 and PD-Ls in the pre-treatment tumor microenvironment and correlation with response to PD-1/PD-L1 blockade. *Oncimmunology*. 2014;3(11):e963413.
431. Taube JM, Galon J, Sholl LM, Rodig SJ, Cottrell TR, Giraldo NA, et al. Implications of the tumor immune microenvironment for staging and therapeutics. *Modern Pathology*. 2018;31(2):214-34.
432. Power R, Lowery MA, Reynolds JV, Dunne MR. The Cancer-Immune Set Point in Oesophageal Cancer. *Front Oncol*. 2020;10:891.
433. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nature Reviews Drug Discovery*. 2019;18(3):197-218.
434. Tiwari A, Oravecz T, Dillon LA, Italiano A, Audoly L, Fridman WH, et al. Towards a consensus definition of immune exclusion in cancer. *Front Immunol*. 2023;14:1084887.
435. Bonaventura P, Shekarian T, Alcacer V, Valladeau-Guilemond J, Valsesia-Wittmann S, Amigorena S, et al. Cold Tumors: A Therapeutic Challenge for Immunotherapy. *Front Immunol*. 2019;10:168.
436. Hennequin A, Derangère V, Boidot R, Apetoh L, Vincent J, Orry D, et al. Tumor infiltration by Tbet+ effector T cells and CD20+ B cells is associated with survival in gastric cancer patients. *Oncolimmunology*. 2016;5(2):e1054598.

List of References

437. Ni Z, Xing D, Zhang T, Ding N, Xiang D, Zhao Z, et al. Tumor-infiltrating B cell is associated with the control of progression of gastric cancer. *Immunologic Research*. 2021;69(1):43-52.
438. Guan H, Wan Y, Lan J, Wang Q, Wang Z, Li Y, et al. PD-L1 is a critical mediator of regulatory B cells and T cells in invasive breast cancer. *Scientific Reports*. 2016;6(1):35651.
439. Tong DN, Guan J, Sun JH, Zhao CY, Chen SG, Zhang ZY, et al. Characterization of B cell-mediated PD-1/PD-L1 interaction in pancreatic cancer patients. *Clin Exp Pharmacol Physiol*. 2020;47(8):1342-9.
440. Yan L, Yu Z, Wang H, Qu C, Wang Y, Yao H, et al. Bioinformatics analysis identifies PSMB8 as a key gene in the cutaneous malignant melanoma tumor microenvironment. *Ann Transl Med*. 2022;10(24):1354.
441. Eksteen B, Miles A, Curbishley SM, Tselepis C, Grant AJ, Walker LS, et al. Epithelial inflammation is associated with CCL28 production and the recruitment of regulatory T cells expressing CCR10. *The Journal of Immunology*. 2006;177(1):593-603.
442. Chen L, Ridley A, Hammitzsch A, Al-Mossawi MH, Bunting H, Georgiadis D, et al. Silencing or inhibition of endoplasmic reticulum aminopeptidase 1 (ERAP1) suppresses free heavy chain expression and Th17 responses in ankylosing spondylitis. *Annals of the Rheumatic Diseases*. 2016;75(5):916-23.
443. Rastall DPW, Alyaquob FS, O'Connell P, Pepelyayeva Y, Peters D, Godbehere-Roosa S, et al. Mice expressing human ERAP1 variants associated with ankylosing spondylitis have altered T-cell repertoires and NK cell functions, as well as increased in utero and perinatal mortality. *International Immunology*. 2017;29(6):277-89.
444. D'Amico S, D'Alicandro V, Compagnone M, Tempora P, Guida G, Romania P, et al. ERAP1 Controls the Interaction of the Inhibitory Receptor KIR3DL1 With HLA-B51:01 by Affecting Natural Killer Cell Function. *Front Immunol*. 2021;12:778103.
445. Cifaldi L, Romania P, Falco M, Lorenzi S, Meazza R, Petrini S, et al. ERAP1 Regulates Natural Killer Cell Function by Controlling the Engagement of Inhibitory Receptors. *Cancer Research*. 2015;75(5):824-34.
446. Stoll G, Iribarren K, Michels J, Leary A, Zitvogel L, Cremer I, et al. Calreticulin expression: Interaction with the immune infiltrate and impact on survival in patients with ovarian and non-small cell lung cancer. *OncolImmunology*. 2016;5(7):e1177692.
447. Kasikova L, Hensler M, Truxova I, Skapa P, Laco J, Belicova L, et al. Calreticulin exposure correlates with robust adaptive antitumor immunity and favorable prognosis in ovarian carcinoma patients. *Journal for ImmunoTherapy of Cancer*. 2019;7(1):312.
448. Gumperz JE, Miyake S, Yamamura T, Brenner MB. Functionally Distinct Subsets of CD1d-restricted Natural Killer T Cells Revealed by CD1d Tetramer Staining. *Journal of Experimental Medicine*. 2002;195(5):625-36.
449. Burks J, Olkhanud PB, Berzofsky JA. The role of NKT cells in gastrointestinal cancers. *Oncoimmunology*. 2022;11(1):2009666.
450. Tripathi SC, Peters HL, Taguchi A, Katayama H, Wang H, Momin A, et al. Immunoproteasome deficiency is a feature of non-small cell lung cancer with a mesenchymal phenotype and is associated with a poor outcome. *Proc Natl Acad Sci U S A*. 2016;113(11):E1555-64.
451. Kalaora S, Lee JS, Barnea E, Levy R, Greenberg P, Alon M, et al. Immunoproteasome expression is associated with better prognosis and response to checkpoint therapies in melanoma. *Nat Commun*. 2020;11(1):896.

452. Martín-Esteban A, Rodriguez JC, Peske D, Lopez de Castro JA, Shastri N, Sadegh-Nasseri S. The ER Aminopeptidases, ERAP1 and ERAP2, synergize to self-modulate their respective activities. *Front Immunol.* 2022;13:1066483.
453. Schmidt K, Leisegang M, Kloetzel P-M. ERAP2 supports TCR recognition of three immunotherapy targeted tumor epitopes. *Molecular Immunology.* 2023;154:61-8.
454. Chang S-C, Momburg F, Bhutani N, Goldberg AL. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a “molecular ruler” mechanism. *Proceedings of the National Academy of Sciences.* 2005;102(47):17107-12.
455. Basha G, Omilusik K, Chavez-Steenbock A, Reinicke AT, Lack N, Choi KB, et al. A CD74-dependent MHC class I endolysosomal cross-presentation pathway. *Nature Immunology.* 2012;13(3):237-45.
456. Bonnin E, Denizeau J, Maurin M, Saez JJ, Jouve M, Bonté P-E, et al. CD74 supports Treg accumulation and function in tumors. 2023.
457. Peng C, Ye H, li Z, Duan X, Yang W, Yi Z. Multi-omics characterization of a scoring system to quantify hypoxia patterns in patients with head and neck squamous cell carcinoma. *Journal of Translational Medicine.* 2023;21(1):15.
458. Huehn J, Siegmund K, Lehmann JC, Siewert C, Haubold U, Feuerer M, et al. Developmental stage, phenotype, and migration distinguish naive- and effector/memory-like CD4+ regulatory T cells. *J Exp Med.* 2004;199(3):303-13.
459. Wu G, Xiao G, Yan Y, Guo C, Hu N, Shen S. Bioinformatics analysis of the clinical significance of HLA class II in breast cancer. *Medicine (Baltimore).* 2022;101(40):e31071.
460. Hudec M, Juříčková I, Riegerová K, Ovsepian SV, Černá M, O'Leary VB. Enhanced Extracellular Transfer of HLA-DQ Activates CD3(+) Lymphocytes towards Compromised Treg Induction in Celiac Disease. *Int J Mol Sci.* 2022;23(11).
461. Sabbatino F, Liguori L, Polcaro G, Salvato I, Caramori G, Salzano FA, et al. Role of Human Leukocyte Antigen System as A Predictive Biomarker for Checkpoint-Based Immunotherapy in Cancer Patients. *Int J Mol Sci.* 2020;21(19).
462. Seliger B, Kloor M, Ferrone S. HLA class II antigen-processing pathway in tumors: Molecular defects and clinical relevance. *Oncoimmunology.* 2017;6(2):e1171447.
463. Miura Y, Anami T, Yatsuda J, Motoshima T, Oka S, Suyama K, et al. HLA-DR and CD74 Expression and the Immune Microenvironment in Renal Cell Carcinoma. *Anticancer Res.* 2021;41(6):2841-8.
464. MIURA Y, ANAMI T, YATSUDA J, MOTOSHIMA T, OKA S, SUYAMA K, et al. HLA-DR and CD74 Expression and the Immune Microenvironment in Renal Cell Carcinoma. *Anticancer Research.* 2021;41(6):2841-8.
465. Yang H, Ye S, Goswami S, Li T, Wu J, Cao C, et al. Highly immunosuppressive HLADRhi regulatory T cells are associated with unfavorable outcomes in cervical squamous cell carcinoma. *International Journal of Cancer.* 2020;146(7):1993-2006.
466. Nagase H, Takeoka T, Urakawa S, Morimoto-Okazawa A, Kawashima A, Iwahori K, et al. ICOS+ Foxp3+ TILs in gastric cancer are prognostic markers and effector regulatory T cells associated with *Helicobacter pylori*. *International Journal of Cancer.* 2017;140(3):686-95.
467. Kovacsovics-Bankowski M, Chisholm L, Vercellini J, Tucker CG, Montler R, Haley D, et al. Detailed characterization of tumor infiltrating lymphocytes in two distinct human solid malignancies show phenotypic similarities. *J Immunother Cancer.* 2014;2(1):38.

List of References

468. Yan X, Wu C, Chen T, Santos MM, Liu C-L, Yang C, et al. Cathepsin S inhibition changes regulatory T-cell activity in regulating bladder cancer and immune cell proliferation and apoptosis. *Molecular Immunology*. 2017;82:66-74.
469. Zhang J, Pan S, Han C, Jin H, Sun Q, Du J, et al. Combination of Immune-Related Network and Molecular Typing Analysis Defines a Three-Gene Signature for Predicting Prognosis of Triple-Negative Breast Cancer. *Biomolecules*. 2022;12(11).
470. Prigione I, Penco F, Martini A, Gattorno M, Pistoia V, Morandi F. HLA-G and HLA-E in patients with juvenile idiopathic arthritis. *Rheumatology*. 2011;50(5):966-72.
471. Rouette A, Trofimov A, Haberl D, Boucher G, Lavallée V-P, D'Angelo G, et al. Expression of immunoproteasome genes is regulated by cell-intrinsic and –extrinsic factors in human cancers. *Scientific Reports*. 2016;6(1):34019.
472. Macagno A, Gilliet M, Sallusto F, Lanzavecchia A, Nestle FO, Groettrup M. Dendritic cells up-regulate immunoproteasomes and the proteasome regulator PA28 during maturation. *Eur J Immunol*. 1999;29(12):4037-42.
473. Kincaid EZ, Che JW, York I, Escobar H, Reyes-Vargas E, Delgado JC, et al. Mice completely lacking immunoproteasomes show major changes in antigen presentation. *Nature Immunology*. 2012;13(2):129-35.
474. Seamon K, Kurlak LO, Warthan M, Stratikos E, Strauss JF, 3rd, Mistry HD, et al. The Differential Expression of ERAP1/ERAP2 and Immune Cell Activation in Pre-eclampsia. *Front Immunol*. 2020;11:396.
475. Egea L, Hirata Y, Kagnoff MF. GM-CSF: a role in immune and inflammatory reactions in the intestine. *Expert Rev Gastroenterol Hepatol*. 2010;4(6):723-31.
476. Khajah M, Millen B, Cara DC, Waterhouse C, McCafferty DM. Granulocyte-macrophage colony-stimulating factor (GM-CSF): a chemoattractive agent for murine leukocytes in vivo. *J Leukoc Biol*. 2011;89(6):945-53.
477. Mabire M, Hegde P, Hammoutene A, Wan J, Caér C, Sayegh RA, et al. MAIT cell inhibition promotes liver fibrosis regression via macrophage phenotype reprogramming. *Nature Communications*. 2023;14(1):1830.
478. Gozalbo-López B, Gómez del Moral M, Campos-Martín Y, Setién F, Martín P, Bellas C, et al. The MHC-related protein 1 (MR1) is expressed by a subpopulation of CD38+, IgA+ cells in the human intestinal mucosa. *Histol Histopathol*. 2009;24(11):1439-49.
479. Gillessen S, Naumov YN, Nieuwenhuis EES, Exley MA, Lee FS, Mach N, et al. CD1d-restricted T cells regulate dendritic cell function and antitumor immunity in a granulocyte–macrophage colony-stimulating factor-dependent fashion. *Proceedings of the National Academy of Sciences*. 2003;100(15):8874-9.
480. Lynch L, O'Shea D, Winter DC, Geoghegan J, Doherty DG, O'Farrelly C. Invariant NKT cells and CD1d+ cells amass in human omentum and are depleted in patients with cancer and obesity. *European Journal of Immunology*. 2009;39(7):1893-901.
481. Schröder B. The multifaceted roles of the invariant chain CD74 — More than just a chaperone. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. 2016;1863(6, Part A):1269-81.
482. Rönnberg E, Boey DZH, Ravindran A, Säfholm J, Orre AC, Al-Ameri M, et al. Immunoprofiling Reveals Novel Mast Cell Receptors and the Continuous Nature of Human Lung Mast Cell Heterogeneity. *Front Immunol*. 2021;12:804812.

483. Calandra T, Bernhagen J, Metz CN, Spiegel LA, Bacher M, Donnelly T, et al. MIF as a glucocorticoid-induced modulator of cytokine production. *Nature*. 1995;377(6544):68-71.
484. Ina K, Kusugami K, Yamaguchi T, Imada A, Hosokawa T, Ohsuga M, et al. Mucosal interleukin-8 is involved in neutrophil migration and binding to extracellular matrix in inflammatory bowel disease. *American Journal of Gastroenterology (Springer Nature)*. 1997;92(8).
485. Mazzucchelli L, Hauser C, Zgraggen K, Wagner H, Hess M, Laissue JA, et al. Expression of interleukin-8 gene in inflammatory bowel disease is related to the histological grade of active inflammation. *The American journal of pathology*. 1994;144(5):997.
486. Høeberg M, Noer JB, Vistesen MV, Bartels A, Bech EM, Nygård SB, et al. The invariant chain CD74 protein is a cell surface binding partner of TIMP-1 in breast cancer cells. *Mol Oncol*. 2023;17(8):1595-612.
487. Kobuch J, Cui H, Grünwald B, Saftig P, Knolle PA, Krüger A. TIMP-1 signaling via CD63 triggers granulopoiesis and neutrophilia in mice. *Haematologica*. 2015;100(8):1005.
488. Li J, Xie Y, Liang W, Zhang H, Wang X, Liu J, et al. HLA-DR regulates macrophage phenotypic transformation and affects malignant behavior in esophageal squamous cell carcinoma2020.
489. Singhal S, Stadanlick J, Annunziata MJ, Rao AS, Bhojnagarwala PS, O'Brien S, et al. Human tumor-associated monocytes/macrophages and their regulation of T cell responses in early-stage lung cancer. *Sci Transl Med*. 2019;11(479).
490. Yang J, Wang F, Chen B. HLA-DPA1 gene is a potential predictor with prognostic values in multiple myeloma. *BMC Cancer*. 2020;20(1):915.
491. Xu L, Yan M, Long J, Yang H, Li W. Identification of macrophage correlated biomarkers to predict the prognosis in patients with intrahepatic cholangiocarcinoma. *Frontiers in Oncology*. 2022;12:967982.
492. Oelschlaegel D, Weiss Sadan T, Salpeter S, Krug S, Blum G, Schmitz W, et al. Cathepsin Inhibition Modulates Metabolism and Polarization of Tumor-Associated Macrophages. *Cancers (Basel)*. 2020;12(9).
493. Dykes SS, Fasanya HO, Siemann DW. Cathepsin L secretion by host and neoplastic cells potentiates invasion. *Oncotarget*. 2019;10(53):5560-8.
494. Cárdenas DM, Sánchez AC, Rosas DA, Rivero E, Paparoni MD, Cruz MA, et al. Preliminary analysis of single-nucleotide polymorphisms in IL-10, IL-4, and IL-4R α genes and profile of circulating cytokines in patients with gastric Cancer. *BMC Gastroenterology*. 2018;18(1):184.
495. Repnik U, Starr AE, Overall CM, Turk B. Cysteine Cathepsins Activate ELR Chemokines and Inactivate Non-ELR Chemokines *^{sup}</sup>. *Journal of Biological Chemistry*. 2015;290(22):13800-11.
496. Palmer VL, Nganga VK, Rothermund ME, Perry GA, Swanson PC. Cd1d regulates B cell development but not B cell accumulation and IL10 production in mice with pathologic CD5+ B cell expansion. *BMC Immunology*. 2015;16(1):66.
497. Hu Y, Ding J, Wu C, Gao H, Ge M, Shao Q, et al. Differential Expression and Prognostic Correlation of Immune Related Factors Between Right and Left Side Colorectal Cancer. *Front Oncol*. 2022;12:845765.
498. Lin W, Zhang P, Chen H, Chen Y, Yang H, Zheng W, et al. Circulating plasmablasts/plasma cells: a potential biomarker for IgG4-related disease. *Arthritis Research & Therapy*. 2017;19(1):25.

List of References

499. Carlo-Stella C, Guidetti A, Di Nicola M, Lavazza C, Cleris L, Sia D, et al. IFN- γ Enhances the Antimyeloma Activity of the Fully Human Anti-Human Leukocyte Antigen-DR Monoclonal Antibody 1D09C3. *Cancer Research*. 2007;67(7):3269-75.
500. Yamakoshi Y, Tanaka H, Sakimura C, Deguchi S, Mori T, Tamura T, et al. Immunological potential of tertiary lymphoid structures surrounding the primary tumor in gastric cancer. *Int J Oncol*. 2020;57(1):171-82.
501. Quante M, Wang TC, Bass AJ. Adenocarcinoma of the oesophagus: is it gastric cancer? *Gut*. 2023;72(6):1027-9.
502. Schneppenheim J, Dressel R, Hüttl S, Lüllmann-Rauch R, Engelke M, Dittmann K, et al. The intramembrane protease SPPL2a promotes B cell development and controls endosomal traffic by cleavage of the invariant chain. *Journal of Experimental Medicine*. 2013;210(1):41-58.
503. Hüttl S, Kläsener K, Schweizer M, Schneppenheim J, Oberg H-H, Kabelitz D, et al. Processing of CD74 by the Intramembrane Protease SPPL2a Is Critical for B Cell Receptor Signaling in Transitional B Cells. *The Journal of Immunology*. 2015;195(4):1548-63.
504. Schneppenheim J, Dressel R, Hüttl S, Lüllmann-Rauch R, Engelke M, Dittmann K, et al. The intramembrane protease SPPL2a promotes B cell development and controls endosomal traffic by cleavage of the invariant chain. *Journal of Experimental Medicine*. 2012;210(1):41-58.
505. Hos BJ, Tondini E, Camps MGM, Rademaker W, van den Bulk J, Ruano D, et al. Cancer-specific T helper shared and neo-epitopes uncovered by expression of the MHC class II master regulator CIITA. *Cell Rep*. 2022;41(2):111485.
506. Joensuu H, Bono P, Kataja V, Alanko T, Kokko R, Asola R, et al. Fluorouracil, epirubicin, and cyclophosphamide with either docetaxel or vinorelbine, with or without trastuzumab, as adjuvant treatments of breast cancer: final results of the FinHer Trial. *Journal of Clinical Oncology*. 2009;27(34):5685-92.
507. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *New England Journal of Medicine*. 2005;353(16):1659-72.
508. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer Jr CE, Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *New England journal of medicine*. 2005;353(16):1673-84.
509. Phillips B, Tubbs R, Rice T, Rybicki L, Plesec T, Rodriguez C, et al. Clinicopathologic features and treatment outcomes of patients with human epidermal growth factor receptor 2-positive adenocarcinoma of the esophagus and gastroesophageal junction. *Diseases of the Esophagus*. 2013;26(3):299-304.
510. Bang Y-J, Van Cutsem E, Feyereislova A, Chung HC, Shen L, Sawaki A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *The Lancet*. 2010;376(9742):687-97.
511. Safran HP, Winter K, Ilson DH, Wigle D, DiPetrillo T, Haddock MG, et al. Trastuzumab with trimodality treatment for oesophageal adenocarcinoma with HER2 overexpression (NRG Oncology/RTOG 1010): a multicentre, randomised, phase 3 trial. *The Lancet Oncology*. 2022;23(2):259-69.
512. Prins M, Verhage R, Ten Kate F, van Hillegersberg R. Cyclooxygenase isoenzyme-2 and vascular endothelial growth factor are associated with poor prognosis in esophageal adenocarcinoma. *Journal of Gastrointestinal Surgery*. 2012;16:956-66.

513. Buskens CJ, Van Rees BP, Sivula A, Reitsma JB, Haglund C, Bosma PJ, et al. Prognostic significance of elevated cyclooxygenase 2 expression in patients with adenocarcinoma of the esophagus. *Gastroenterology*. 2002;122(7):1800-7.
514. Tuynman JB, Buskens CJ, Kemper K, Ten Kate FJ, Offerhaus GJA, Richel DJ, et al. Neoadjuvant selective COX-2 inhibition down-regulates important oncogenic pathways in patients with esophageal adenocarcinoma. *Annals of surgery*. 2005;242(6):840.
515. Corley DA, Kerlikowske K, Verma R, Buffler P. Protective association of aspirin/NSAIDs and esophageal cancer: a systematic review and meta-analysis. *Gastroenterology*. 2003;124(1):47-56.
516. McCormick Matthews LH, Noble F, Tod J, Jaynes E, Harris S, Primrose JN, et al. Systematic review and meta-analysis of immunohistochemical prognostic biomarkers in resected oesophageal adenocarcinoma. *British Journal of Cancer*. 2015;113(1):107-18.
517. Noble F, Mellows T, McCormick Matthews LH, Bateman AC, Harris S, Underwood TJ, et al. Tumour infiltrating lymphocytes correlate with improved survival in patients with oesophageal adenocarcinoma. *Cancer Immunology, Immunotherapy*. 2016;65:651-62.
518. Babar L, Kosovec JE, Jahangiri V, Chowdhury N, Zheng P, Omstead AN, et al. Prognostic immune markers for recurrence and survival in locally advanced esophageal adenocarcinoma. *Oncotarget*. 2019;10(44):4546-55.
519. Humphries MP, Craig SG, Kacprzyk R, Fisher NC, Bingham V, McQuaid S, et al. The adaptive immune and immune checkpoint landscape of neoadjuvant treated esophageal adenocarcinoma using digital pathology quantitation. *BMC Cancer*. 2020;20(1):500.
520. Karamitopoulou E. Tumour microenvironment of pancreatic cancer: immune landscape is dictated by molecular and histopathological features. *Br J Cancer*. 2019;121(1):5-14.
521. Jenkins E, Whitehead T, Fellermeyer M, Davis SJ, Sharma S. The current state and future of T-cell exhaustion research. *Oxford Open Immunology*. 2023;4(1):iqad006.
522. Schoemmel M, Loeser H, Kraemer M, Wagener-Ryczek S, Hillmer A, Bruns C, et al. Distribution of tumor-infiltrating-T-lymphocytes and possible tumor-escape mechanisms avoiding immune cell attack in locally advanced adenocarcinomas of the esophagus. *Clin Transl Oncol*. 2021;23(8):1601-10.
523. Fritschy J-M. Is my antibody-staining specific? How to deal with pitfalls of immunohistochemistry. *European Journal of Neuroscience*. 2008;28(12):2365-70.
524. Gown AM. Diagnostic Immunohistochemistry: What Can Go Wrong and How to Prevent It. *Archives of Pathology & Laboratory Medicine*. 2016;140(9):893-8.
525. Ram S, Vizcarra P, Whalen P, Deng S, Painter CL, Jackson-Fisher A, et al. Pixelwise H-score: A novel digital image analysis-based metric to quantify membrane biomarker expression from immunohistochemistry images. *PLoS One*. 2021;16(9):e0245638.
526. Misra S, Wu Y. Chapter 10 - Machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales with feature extraction and feature ranking. In: Misra S, Li H, He J, editors. *Machine Learning for Subsurface Characterization*: Gulf Professional Publishing; 2020. p. 289-314.
527. Brock G, Pihur V, Datta S, Datta S, Brock MG, Biobase S. Package 'clValid'. 2021.
528. Wickham H, Chang W, Wickham MH. Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics Version. 2016;2(1):1-189.
529. Fox J, Muenchen R, Putler D, Fox MJ. Package 'RcmdrMisc'. 2023.

List of References

530. Dupont WD, Plummer Jr WD. Power and sample size calculations: a review and computer program. *Controlled clinical trials*. 1990;11(2):116-28.
531. Tu Z, Li K, Ji Q, Huang Y, Lv S, Li J, et al. Pan-cancer analysis: predictive role of TAP1 in cancer prognosis and response to immunotherapy. *BMC Cancer*. 2023;23(1):133.
532. Mari L, Hoefnagel SJ, Zito D, van de Meent M, van Endert P, Calpe S, et al. microRNA 125a regulates MHC-I expression on esophageal adenocarcinoma cells, associated with suppression of antitumor immune response and poor outcomes of patients. *Gastroenterology*. 2018;155(3):784-98.
533. Bobryshev YV, Tran D, Killingsworth MC, Buckland M, Lord RVN. Dendritic Cells in Barrett's Esophagus and Esophageal Adenocarcinoma. *Journal of Gastrointestinal Surgery*. 2009;13(1):44-53.
534. Morinaga T, Iwatsuki M, Yamashita K, Matsumoto C, Harada K, Kurashige J, et al. Evaluation of HLA-E Expression Combined with Natural Killer Cell Status as a Prognostic Factor for Advanced Gastric Cancer. *Annals of Surgical Oncology*. 2022;29(8):4951-60.
535. Lhotakova K, Grzelak A, Polakova I, Vackova J, Smahel M. Establishment and characterization of a mouse tumor cell line with irreversible downregulation of MHC class I molecules. *Oncology Reports*. 2019;42(6):2826-35.
536. Rangan L, Galaine J, Boidot R, Hamieh M, Dosset M, Francoual J, et al. Identification of a novel PD-L1 positive solid tumor transplantable in HLA-A* 0201/DRB1* 0101 transgenic mice. *Oncotarget*. 2017;8(30):48959.
537. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature Medicine*. 2018;24(5):541-50.
538. Noble F, Mellows T, McCormick Matthews LH, Bateman AC, Harris S, Underwood TJ, et al. Tumour infiltrating lymphocytes correlate with improved survival in patients with oesophageal adenocarcinoma. *Cancer Immunology, Immunotherapy*. 2016;65(6):651-62.
539. Pérez-Romero K, Rodríguez RM, Amedei A, Barceló-Coblijn G, Lopez DH. Immune Landscape in Tumor Microenvironment: Implications for Biomarker Development and Immunotherapy. *International Journal of Molecular Sciences*. 2020;21(15):5521.
540. Oliveira CC, van Veelen PA, Querido B, de Ru A, Sluijter M, Laban S, et al. The nonpolymorphic MHC Qa-1b mediates CD8+ T cell surveillance of antigen-processing defects. *Journal of Experimental Medicine*. 2010;207(1):207-21.
541. Sandberg JK, Chambers BJ, Van Kaer L, Kärre K, Ljunggren HG. TAP1-deficient mice select a CD8+ T cell repertoire that displays both diversity and peptide specificity. *Eur J Immunol*. 1996;26(2):288-93.
542. Morinaga T, Iwatsuki M, Yamashita K, Matsumoto C, Harada K, Kurashige J, et al. Evaluation of HLA-E Expression Combined with Natural Killer Cell Status as a Prognostic Factor for Advanced Gastric Cancer. *Ann Surg Oncol*. 2022;29(8):4951-60.
543. Xu YF, Du XF, Li ZY, Fang ZP, Zhang FB. Lesion human leukocyte antigen-E is associated with favourable prognosis for patients with oesophageal squamous cell carcinoma. *J Int Med Res*. 2021;49(10):3000605211047278.
544. Kim SK, Cho SW. The Evasion Mechanisms of Cancer Immunity and Drug Intervention in the Tumor Microenvironment. *Front Pharmacol*. 2022;13:868695.

545. Albelda SM. Tumor Antigen Heterogeneity: The “Elephant in the Room” of Adoptive T-cell Therapy for Solid Tumors. *Cancer Immunology Research*. 2020;8(1):2-.
546. Nicholas B, Bailey A, McCann KJ, Wood O, Walker RC, Parker R, et al. Identification of neoantigens in oesophageal adenocarcinoma. *Immunology*. 2023;168(3):420-31.
547. Kollmann D, Ignatova D, Jedamzik J, Chang YT, Jomrich G, Baierl A, et al. PD-L1 expression is an independent predictor of favorable outcome in patients with localized esophageal adenocarcinoma. *Oncoimmunology*. 2018;7(6):e1435226.
548. Thompson ED, Zahurak M, Murphy A, Cornish T, Cuka N, Abdelfatah E, et al. Patterns of PD-L1 expression and CD8 T cell infiltration in gastric adenocarcinomas and associated immune stroma. *Gut*. 2017;66(5):794-801.
549. Derks S, Nason KS, Liao X, Stachler MD, Liu KX, Liu JB, et al. Epithelial PD-L2 Expression Marks Barrett's Esophagus and Esophageal Adenocarcinoma. *Cancer Immunology Research*. 2015;3(10):1123-9.
550. Fontana E, Smyth EC. Combination immune checkpoint blockade in advanced untreated gastroesophageal adenocarcinoma: Seeking biomarkers for durable benefit. *Cancer Cell*. 2022;40(6):606-8.
551. Yang L, Pang Y, Moses HL. TGF-beta and immune cells: an important regulatory axis in the tumor microenvironment and progression. *Trends Immunol*. 2010;31(6):220-7.
552. Brooks DG, Walsh KB, Elsaesser H, Oldstone MBA. IL-10 directly suppresses CD4 but not CD8 T cell effector and memory responses following acute viral infection. *Proceedings of the National Academy of Sciences*. 2010;107(7):3018-23.
553. Croft W, Evans RPT, Pearce H, Elshafie M, Griffiths EA, Moss P. The single cell transcriptional landscape of esophageal adenocarcinoma and its modulation by neoadjuvant chemotherapy. *Molecular Cancer*. 2022;21(1):200.
554. Mendelson J, Song S, Li Y, Maru DM, Mishra B, Davila M, et al. Dysfunctional transforming growth factor- β signaling with constitutively active Notch signaling in Barrett's esophageal adenocarcinoma. *Cancer*. 2011;117(16):3691-702.
555. Vacchelli E, Semeraro M, Enot DP, Chaba K, Poirier Colame V, Dartigues P, et al. Negative prognostic impact of regulatory T cell infiltration in surgically resected esophageal cancer post-radiochemotherapy. *Oncotarget*. 2015;6(25):20840-50.
556. Lian J, Liu S, Yue Y, Yang Q, Zhang Z, Yang S, et al. Eomes promotes esophageal carcinoma progression by recruiting Treg cells through the CCL20-CCR6 pathway. *Cancer Science*. 2021;112(1):144-54.
557. Wang Y, Sun H, Zhu N, Wu X, Sui Z, Gong L, et al. Myeloid-Derived Suppressor Cells in Immune Microenvironment Promote Progression of Esophagogastric Junction Adenocarcinoma. *Front Oncol*. 2021;11:640080.
558. Chen L, Shi Y, Zhu X, Guo W, Zhang M, Che Y, et al. IL-10 secreted by cancer-associated macrophages regulates proliferation and invasion in gastric cancer cells via c-Met/STAT3 signaling. *Oncol Rep*. 2019;42(2):595-604.
559. Yang Z, Ming XF. Functions of arginase isoforms in macrophage inflammatory responses: impact on cardiovascular diseases and metabolic disorders. *Front Immunol*. 2014;5:533.
560. Chambers CA, Allison JP. Co-stimulation in T cell responses. *Current Opinion in Immunology*. 1997;9(3):396-404.

List of References

561. Yang W, Zhang Y, Yu J, Li S. The low expression of CD80 correlated with the vascular endothelial growth factor in esophageal cancer tissue. *European Journal of Surgical Oncology (EJSO)*. 2010;36(5):501-6.
562. Sewell-Loftin MK, Bayer SVH, Crist E, Hughes T, Joison SM, Longmore GD, et al. Cancer-associated fibroblasts support vascular growth through mechanical force. *Sci Rep*. 2017;7(1):12574.
563. Cruz-Acuña R, Kariuki SW, Sugiura K, Karaikos S, Plaster EM, Loebel C, et al. Engineered hydrogel reveals contribution of matrix mechanics to esophageal adenocarcinoma and identifies matrix-activated therapeutic targets. *The Journal of Clinical Investigation*. 2023.
564. Mehravar S, Banerjee B, Chatrath H, Amirsolaimani B, Patel K, Patel C, et al. Label-free multi-photon imaging of dysplasia in Barrett's esophagus. *Biomed Opt Express*. 2016;7(1):148-57.
565. Gordon-Weeks A, Yuzhalin AE. Cancer Extracellular Matrix Proteins Regulate Tumour Immunity. *Cancers (Basel)*. 2020;12(11).
566. Cornel AM, Mimpel IL, Nierkens S. MHC Class I Downregulation in Cancer: Underlying Mechanisms and Potential Targets for Cancer Immunotherapy. *Cancers (Basel)*. 2020;12(7).
567. Martin BK, Chin KC, Olsen JC, Skinner CA, Dey A, Ozato K, et al. Induction of MHC class I expression by the MHC class II transactivator CIITA. *Immunity*. 1997;6(5):591-600.
568. Meissner TB, Liu YJ, Lee KH, Li A, Biswas A, van Eggermond MC, et al. NLRC5 cooperates with the RFX transcription factor complex to induce MHC class I gene expression. *J Immunol*. 2012;188(10):4951-8.
569. Zhou F. Molecular mechanisms of IFN-gamma to up-regulate MHC class I antigen processing and presentation. *Int Rev Immunol*. 2009;28(3-4):239-60.
570. Vijayan S, Meissner TB, Lee K-H, Liu Y-J, Downs I, Sidiq T, et al. Role of NLRC5 and IRF1 in the induction of MHC class I. *The Journal of Immunology*. 2019;202(1_Supplement):64.20-64.20.
571. Ivashkiv LB, Donlin LT. Regulation of type I interferon responses. *Nature Reviews Immunology*. 2014;14(1):36-49.
572. Jongsma MLM, Guarda G, Spaapen RM. The regulatory network behind MHC class I expression. *Molecular Immunology*. 2019;113:16-21.
573. Guo A-X, Cui J-J, Wang L-Y, Yin J-Y. The role of CSDE1 in translational reprogramming and human diseases. *Cell Communication and Signaling*. 2020;18(1):14.
574. Moore KS, Yagci N, van Alphen F, Paolini NA, Horos R, Held NM, et al. Csde1 binds transcripts involved in protein homeostasis and controls their expression in an erythroid cell line. *Scientific Reports*. 2018;8(1):2628.
575. Lv J, Zhou Y, Zhou N, Wang Z, Chen J, Chen H, et al. Epigenetic modification of CSDE1 locus dictates immune recognition of nascent tumorigenic cells. *Science translational medicine*. 2023;15:ebq6024.
576. Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S. Protein Data Bank (PDB): the single global macromolecular structure archive. *Protein crystallography: methods and protocols*. 2017:627-41.
577. Galassi C, Galluzzi L. Epigenetic escape of immunosurveillance by malignant cell precursors. *The EMBO Journal*. 2023;42(10):e114050.

578. Heinemann U, Roske Y. Cold-Shock Domains-Abundance, Structure, Properties, and Nucleic-Acid Binding. *Cancers (Basel)*. 2021;13(2).
579. Mihailovic M, Wurth L, Zambelli F, Abaza I, Militi C, Mancuso FM, et al. Widespread generation of alternative UTRs contributes to sex-specific RNA binding by UNR. *RNA*. 2012;18(1):53-64.
580. Turner DJ, Turner M. RNA Binding Proteins As Regulators of Oxidative Stress Identified by a Targeted CRISPR-Cas9 Single Guide RNA Library. *Crispr j*. 2021;4(3):427-37.
581. Cui J, Guo A-X, Wang L-Y, Yin J-Y. Abstract 1289: CSDE1 mediated translation reprogramming under stress. *Cancer Research*. 2020;80(16_Supplement):1289-.
582. Ju Lee H, Bartsch D, Xiao C, Guerrero S, Ahuja G, Schindler C, et al. A post-transcriptional program coordinated by CSDE1 prevents intrinsic neural differentiation of human embryonic stem cells. *Nature Communications*. 2017;8(1):1456.
583. Martinez-Useros J, Garcia-Carbonero N, Li W, Fernandez-Aceñero MJ, Cristobal I, Rincon R, et al. UNR/CSDE1 Expression Is Critical to Maintain Invasive Phenotype of Colorectal Cancer through Regulation of c-MYC and Epithelial-to-Mesenchymal Transition. *J Clin Med*. 2019;8(4).
584. Webb MJ, Kottke T, Kendall BL, Swanson J, Uzendu C, Tonne J, et al. Trap and ambush therapy using sequential primary and tumor escape-selective oncolytic viruses. *Molecular Therapy - Oncolytics*. 2023;29:129-42.
585. Kottke T, Tonne J, Evgin L, Driscoll CB, van Vloten J, Jennings VA, et al. Oncolytic virotherapy induced CSDE1 neo-antigenesis restricts VSV replication but can be targeted by immunotherapy. *Nat Commun*. 2021;12(1):1930.
586. Dana H, Chalbatani GM, Mahmoodzadeh H, Karimloo R, Rezaiean O, Moradzadeh A, et al. Molecular Mechanisms and Biological Functions of siRNA. *Int J Biomed Sci*. 2017;13(2):48-57.
587. Prelich G. Gene overexpression: uses, mechanisms, and interpretation. *Genetics*. 2012;190(3):841-54.
588. Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. Rational siRNA design for RNA interference. *Nature Biotechnology*. 2004;22(3):326-30.
589. Kim Y, Ko JY, Lee S-B, Oh S, Park JW, Kang H-G, et al. Reduced miR-371b-5p expression drives tumor progression via CSDE1/RAC1 regulation in triple-negative breast cancer. *Oncogene*. 2022;41(22):3151-61.
590. Moore CB, Guthrie EH, Huang MT, Taxman DJ. Short hairpin RNA (shRNA): design, delivery, and assessment of gene knockdown. *Methods Mol Biol*. 2010;629:141-58.
591. Rao DD, Vorhies JS, Senzer N, Nemunaitis J. siRNA vs. shRNA: Similarities and differences. *Advanced Drug Delivery Reviews*. 2009;61(9):746-59.
592. Lv J, Zhou Y, Zhou N, Wang Z, Chen J, Chen H, et al. Epigenetic modification of *< i>CSDE1</i>* locus dictates immune recognition of nascent tumorigenic cells. *Science Translational Medicine*. 2023;15(681):eabq6024.
593. Barrangou R, Doudna JA. Applications of CRISPR technologies in research and beyond. *Nature Biotechnology*. 2016;34(9):933-41.
594. Gaj T, Guo J, Kato Y, Sirk SJ, Barbas CF, 3rd. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat Methods*. 2012;9(8):805-7.

List of References

595. Amiri F, Ranjbar M, Pirouzfar M, Nourigorji M, Dianatpour M. HLA-A gene knockout using CRISPR/Cas9 system toward overcoming transplantation concerns. *Egyptian Journal of Medical Human Genetics*. 2021;22(1):37.
596. Prelich G. Gene Overexpression: Uses, Mechanisms, and Interpretation. *Genetics*. 2012;190(3):841-54.
597. Cain AK, Barquist L, Goodman AL, Paulsen IT, Parkhill J, van Opijnen T. A decade of advances in transposon-insertion sequencing. *Nature Reviews Genetics*. 2020;21(9):526-40.
598. Osborn MJ, Lees CJ, McElroy AN, Merkel SC, Eide CR, Mathews W, et al. CRISPR/Cas9-based cellular engineering for targeted gene overexpression. *International journal of molecular sciences*. 2018;19(4):946.
599. Silverman LR, Demakos EP, Peterson BL, Kornblith AB, Holland JC, Odchimir-Reissig R, et al. Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. *J Clin Oncol*. 2002;20(10):2429-40.
600. Issa JP, Kantarjian H. Azacitidine. *Nat Rev Drug Discov*. 2005;Suppl:S6-7.
601. Day JJ. New approaches to manipulating the epigenome. *Dialogues Clin Neurosci*. 2014;16(3):345-57.
602. Serrano A, Tanzarella S, Lionello I, Mendez R, Traversari C, Ruiz-Cabello F, et al. Expression of HLA class I antigens and restoration of antigen-specific CTL response in melanoma cells following 5-aza-2'-deoxycytidine treatment. *Int J Cancer*. 2001;94(2):243-51.
603. Ramsuran V, Kulkarni S, O'Huigin C, Yuki Y, Augusto DG, Gao X, et al. Epigenetic regulation of differential HLA-A allelic expression levels. *Hum Mol Genet*. 2015;24(15):4268-75.
604. Rockett JC, Larkin K, Darnton SJ, Morris AG, Matthews HR. Five newly established oesophageal carcinoma cell lines: phenotypic and immunological characterization. *Br J Cancer*. 1997;75(2):258-63.
605. Davern M, O'Brien RM, McGrath J, Donlon NE, Melo AM, Buckley CE, et al. PD-1 blockade enhances chemotherapy toxicity in oesophageal adenocarcinoma. *Scientific Reports*. 2022;12(1):3259.
606. Davern M, Donlon NE, Sheppard A, O'Connell F, Hayes C, Bhardwaj A, et al. Chemotherapy regimens induce inhibitory immune checkpoint protein expression on stem-like and senescent-like oesophageal adenocarcinoma cells. *Translational Oncology*. 2021;14(6):101062.
607. Contino G, Eldridge MD, Secrier M, Bower L, Fels Elliott R, Weaver J, et al. Whole-genome sequencing of nine esophageal adenocarcinoma cell lines. *F1000Res*. 2016;5:1336.
608. Katsha A, Arras J, Soutto M, Belkhiri A, El-Rifai W. AURKA regulates JAK2-STAT3 activity in human gastric and esophageal cancers. *Mol Oncol*. 2014;8(8):1419-28.
609. Katsha A, Arras J, Soutto M, Belkhiri A, El-Rifai W. AURKA regulates JAK2-STAT3 activity in human gastric and esophageal cancers. *Molecular Oncology*. 2014;8(8):1419-28.
610. Hoos A, Cordon-Cardo C. Tissue Microarray Profiling of Cancer Specimens and Cell Lines: Opportunities and Limitations. *Laboratory Investigation*. 2001;81(10):1331-8.
611. Hoppe S, Meder L, Gebauer F, Ullrich RT, Zander T, Hillmer AM, et al. Trophoblast Cell Surface Antigen 2 (TROP2) as a Predictive Bio-Marker for the Therapeutic Efficacy of Sacituzumab Govitecan in Adenocarcinoma of the Esophagus. *Cancers*. 2022;14(19):4789.

612. van der Meer D, Bartherope S, Yang W, Lightfoot H, Hall C, Gilbert J, et al. Cell Model Passports—a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic acids research*. 2019;47(D1):D923-D9.
613. Dhanasekaran S, Doherty TM, Kenneth J. Comparison of different standards for real-time PCR-based absolute quantification. *Journal of Immunological Methods*. 2010;354(1):34-9.
614. Abràmoff MD, Magalhães PJ, Ram SJ. Image processing with ImageJ. *Biophotonics international*. 2004;11(7):36-42.
615. Walker R, Breininger S, Sharpe B, Harrington J, Reddin I, Tse C, et al. Targetable fibroblast phenotypes and EMT malignant cell states cooperate to promote tumor progression in esophageal adenocarcinoma. *Research Square*; 2023.
616. Waise S, Parker R, Rose-Zerilli MJJ, Layfield DM, Wood O, West J, et al. An Optimized Method to Isolate Human Fibroblasts from Tissue for ex vivo Analysis. *Bio Protoc*. 2019;9(23):e3440.
617. Macosko Evan Z, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 2015;161(5):1202-14.
618. Dückting H, Seurat G. *Seurat*: Taschen; 2000.
619. Anderson MJ, Sundaram N, Satish N, Patwary MMA, Willke TL, Dubey P, editors. Graphpad: Optimized graph primitives for parallel and distributed platforms. 2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS); 2016: IEEE.
620. Lee S, Lee DK. What is the proper way to apply the multiple comparison test? *Korean J Anesthesiol*. 2018;71(5):353-60.
621. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics*. 2010;26(1):139-40.
622. Micheletti F, Bazzaro M, Canella A, Marastoni M, Traniello S, Gavioli R. The lifespan of major histocompatibility complex class I/peptide complexes determines the efficiency of cytotoxic T-lymphocyte responses. *Immunology*. 1999;96(3):411-5.
623. Weigert R, Yeung AC, Li J, Donaldson JG. Rab22a Regulates the Recycling of Membrane Proteins Internalized Independently of Clathrin. *Molecular Biology of the Cell*. 2004;15(8):3758-70.
624. Ju Lee H, Bartsch D, Xiao C, Guerrero S, Ahuja G, Schindler C, et al. A post-transcriptional program coordinated by CSDE1 prevents intrinsic neural differentiation of human embryonic stem cells. *Nat Commun*. 2017;8(1):1456.
625. Chien C-S, Wang M-L, Chu P-Y, Chang Y-L, Liu W-H, Yu C-C, et al. Lin28B/Let-7 regulates expression of Oct4 and Sox2 and reprograms oral squamous cell carcinoma cells to a stem-like state. *Cancer research*. 2015;75(12):2553-65.
626. Li Z, Wu X, Li J, Yu S, Ke X, Yan T, et al. HMGA2-Snai2 axis regulates tumorigenicity and stemness of head and neck squamous cell carcinoma. *Experimental Cell Research*. 2022;418(1):113271.
627. Yang Z, Liu F, Li Z, Liu N, Yao X, Zhou Y, et al. Histone lysine methyltransferase SMYD3 promotes oral squamous cell carcinoma tumorigenesis via H3K4me3-mediated HMGA2 transcription. *Clin Epigenetics*. 2023;15(1):92.
628. Alsamman K, El-Masry OS. Interferon regulatory factor 1 inactivation in human cancer. *Biosci Rep*. 2018;38(3).

List of References

629. Platanias LC. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nature Reviews Immunology*. 2005;5(5):375-86.
630. Liu H, Li X, Dun MD, Faulkner S, Jiang CC, Hondermarck H. Cold Shock Domain Containing E1 (CSDE1) Protein is Overexpressed and Can be Targeted to Inhibit Invasiveness in Pancreatic Cancer Cells. *PROTEOMICS*. 2020;20(10):1900331.
631. Jhunjhunwala S, Hammer C, Delamarre L. Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. *Nature Reviews Cancer*. 2021;21(5):298-312.
632. Hasim A, MA H, Sheyhidin I, ZHANG L-w, Abudula A. Association of defective HLA-I expression with antigen processing machinery and their association with clinicopathological characteristics in Kazak patients with esophageal cancer. *Chinese Medical Journal*. 2011;124(3):341-6.
633. Liu Q, Hao C, Su P, Shi J. Down-regulation of HLA class I antigen-processing machinery components in esophageal squamous cell carcinomas: Association with disease progression. *Scandinavian Journal of Gastroenterology*. 2009;44(8):960-9.
634. Mari L, Hoefnagel SJM, Zito D, van de Meent M, van Endert P, Calpe S, et al. microRNA 125a Regulates MHC-I Expression on Esophageal Adenocarcinoma Cells, Associated With Suppression of Antitumor Immune Response and Poor Outcomes of Patients. *Gastroenterology*. 2018;155(3):784-98.
635. Garrido F, Aptsiauri N, Doorduijn EM, Garcia Lora AM, van Hall T. The urgent need to recover MHC class I in cancers for effective immunotherapy. *Curr Opin Immunol*. 2016;39:44-51.
636. Zhang X, Wang Y, Meng L. Comparative genomic analysis of esophageal squamous cell carcinoma and adenocarcinoma: New opportunities towards molecularly targeted therapy. *Acta Pharmaceutica Sinica B*. 2022;12(3):1054-67.
637. Griffiths EA, Pritchard SA, McGrath SM, Valentine HR, Price PM, Welch IM, et al. Increasing expression of hypoxia-inducible proteins in the Barrett's metaplasia–dysplasia–adenocarcinoma sequence. *British Journal of Cancer*. 2007;96(9):1377-83.
638. Peters BA, Pass HI, Burk RD, Xue X, Goparaju C, Sollecito CC, et al. The lung microbiome, peripheral gene expression, and recurrence-free survival after resection of stage II non-small cell lung cancer. *Genome Medicine*. 2022;14(1):121.
639. Seliger B, Schreiber K, Delp K, Meissner M, Hammers S, Reichert T, et al. Downregulation of the constitutive tapasin expression in human tumor cells of distinct origin and its transcriptional upregulation by cytokines. *Tissue Antigens*. 2001;57(1):39-45.
640. Lou Y, Basha G, Seipp RP, Cai B, Chen SS, Moise AR, et al. Combining the antigen processing components TAP and Tapasin elicits enhanced tumor-free survival. *Clin Cancer Res*. 2008;14(5):1494-501.
641. Ryan D, Carberry S, Murphy Á C, Lindner AU, Fay J, Hector S, et al. Calnexin, an ER stress-induced protein, is a prognostic marker and potential therapeutic target in colorectal cancer. *J Transl Med*. 2016;14(1):196.
642. Dunne MR, Michielsen AJ, O'Sullivan KE, Cathcart MC, Feighery R, Doyle B, et al. HLA-DR expression in tumor epithelium is an independent prognostic indicator in esophageal adenocarcinoma patients. *Cancer Immunology, Immunotherapy*. 2017;66:841-50.
643. Mpakali A, Stratikos E. The Role of Antigen Processing and Presentation in Cancer and the Efficacy of Immune Checkpoint Inhibitor Immunotherapy. *Cancers (Basel)*. 2021;13(1).

644. Puccini A, Battaglin F, Iaia ML, Lenz H-J, Salem ME. Overcoming resistance to anti-PD1 and anti-PD-L1 treatment in gastrointestinal malignancies. *Journal for Immunotherapy of Cancer*. 2020;8(1).
645. Li Y, Liu J, Gao L, Liu Y, Meng F, Li X, et al. Targeting the tumor microenvironment to overcome immune checkpoint blockade therapy resistance. *Immunology letters*. 2020;220:88-96.
646. Trojaniello C, Vitale MG, Scarpato L, Esposito A, Ascierto PA. Melanoma immunotherapy: strategies to overcome pharmacological resistance. *Expert Review of Anticancer Therapy*. 2020;20(4):289-304.
647. Schaafsma E, Fugle CM, Wang X, Cheng C. Pan-cancer association of HLA gene expression with cancer prognosis and immunotherapy efficacy. *British Journal of Cancer*. 2021;125(3):422-32.
648. Mandard AM, Dalibard F, Mandard JC, Marnay J, Henry-Amar M, Petiot JF, et al. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer*. 1994;73(11):2680-6.
649. Rice TW, Blackstone EH, Rybicki LA, Adelstein DJ, Murthy SC, DeCamp MM, et al. Refining esophageal cancer staging. *The Journal of Thoracic and Cardiovascular Surgery*. 2003;125(5):1103-13.
650. Dart A. Presenting fibroblasts. *Nature Reviews Cancer*. 2022;22(4):193-.
651. Naranbhai V, Viard M, Dean M, Groha S, Braun DA, Labaki C, et al. HLA-A*03 and response to immune checkpoint blockade in cancer: an epidemiological biomarker study. *Lancet Oncol*. 2022;23(1):172-84.
652. Filip I, Wang A, Kravets O, Orenbuch R, Zhao J, Perea-Chamblee TE, et al. Pervasiveness of HLA allele-specific expression loss across tumor types. *Genome Medicine*. 2023;15(1):8.
653. Antonczyk A, Krist B, Sajek M, Michalska A, Piaszyk-Borychowska A, Plens-Galaska M, et al. Direct Inhibition of IRF-Dependent Transcriptional Regulatory Mechanisms Associated With Disease. *Front Immunol*. 2019;10:1176.
654. Svensson MND, Doody KM, Schmiedel BJ, Bhattacharyya S, Panwar B, Wiede F, et al. Reduced expression of phosphatase PTPN2 promotes pathogenic conversion of Tregs in autoimmunity. *The Journal of Clinical Investigation*. 2019;129(3):1193-210.
655. Kelly RJ. Immunotherapy for Esophageal and Gastric Cancer. *American Society of Clinical Oncology Educational Book*. 2017(37):292-300.
656. O'Sullivan KE, Phelan JJ, O'Hanlon C, Lysaght J, O'Sullivan JN, Reynolds JV. The role of inflammation in cancer of the esophagus. *Expert Rev Gastroenterol Hepatol*. 2014;8(7):749-60.
657. Davern M, Donlon NE, Sheppard AS, Majcher KD, Connell FO, Heeran AB, et al. FLOT and CROSS chemotherapy regimens alter the frequency of CD27(+) and CD69(+) T cells in oesophagogastric adenocarcinomas: implications for combination with immunotherapy. *J Cancer Res Clin Oncol*. 2023;149(7):3753-74.
658. Carroll TM, Chadwick JA, Owen RP, White MJ, Kaplinsky J, Peneva I, et al. Tumor monocyte content predicts immunochemotherapy outcomes in esophageal adenocarcinoma. *Cancer Cell*. 2023;41(7):1222-41.e7.
659. Erratum: IGF2BP family of RNA-binding proteins regulate innate and adaptive immune responses in cancer cells and tumor microenvironment. *Front Immunol*. 2023;14:1297519.

List of References

660. Avolio R, Inglés-Ferrández M, Ciocia A, Coll O, Bonnin S, Guitart T, et al. Coordinated post-transcriptional control of oncogene-induced senescence by UNR/CSDE1. *Cell Reports*. 2022;38(2):110211.