

Predicting and Understanding Outcomes on First-line Immunochemotherapy for Oesophageal Adenocarcinoma: Deep Molecular Profiling of the LUD2015-005 Trial

---

Thomas M. Carroll

A thesis submitted for the degree of
Doctor of Philosophy

Magdalen College
Nuffield Department of Clinical Medicine
University of Oxford

Trinity Term 2021

DEDICATION

TACTAGTGCGTGCTGCTACGATACATGACATACTAGTGCGCATGATATGATAGCTCAGCTACTG

AGATACATCTACTGACTATGATCACGCTATCACATCATCTACACGACTATCTGTCTACACATGA

GTATCGTCGCAGCACTCATAGACTGTATAGCTCGCGCAGCAGTATGATGTCGTATATGCTATGC

TGACAGACGCACGACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTGAGCA

ABSTRACT

Oesophageal adenocarcinoma (OAC) is a significant cause of cancer mortality that is rapidly rising in prevalence in many parts of the world. OAC is often inoperable at the time of diagnosis, and for these patients, the median survival on current first-line palliative treatment options is less than one year. Therefore, these patients represent a sizable population in need of additional treatment options. Recent phase III studies have shown that the addition of immune checkpoint inhibitors (ICI) to standard of care chemotherapy provides survival benefits for previously untreated oesophageal cancers. Selection of patients ideally suited for immunochemotherapy would be best informed by predictive biomarkers of patient outcome, but as this is a novel therapy in this first-line setting, these markers have not yet been fully investigated.

In this thesis, I undertake to deeply profile predictive biomarkers and mechanisms of response and resistance for immunochemotherapy in OAC using a comprehensive analysis of the LUD2015-005 clinical trial. LUD2015-005 is a unique phase I/II study of immunochemotherapy; the trial design includes an initial four-week window before immunochemotherapy is initiated where only ICI is given in order to disambiguate the effects attributable to the ICI agents.

Patient outcomes on the LUD2015-005 study showed reasonable agreement with reported outcomes from the immunochemotherapy arms in the recently reported phase III studies, with 41% of patients in this trial remaining alive and without reported progression at 12 months. The adverse event profile was generally acceptable, and interestingly, both immune-related and chemotherapy-related adverse events appeared to predict better outcomes.

To understand why some patients had better outcomes than others, tumours were assessed with whole genome sequencing and bulk and single-cell RNA-sequencing. The coding total mutational burden was shown to be strongly predictive of better outcome in this cohort, similar to what has been shown in other ICI settings. However, while T cell inflammation signatures are predictive of outcome on ICI in other cancers, they were not in the LUD2015-005 cohort.

Instead, an innate immune signature specific to infiltrating monocytes was a strong predictor of better outcomes.

While T cell signatures did not have predictive power at the pre-treatment timepoint, sequential sampling revealed that patients with strong on-therapy upregulation of a novel T cell inflammation signature (INCITE) showed better radiological responses during the ICI-only phase of treatment. Development of durable responses was also seen during the immunochemotherapy period even in patients who did not show strong T cell responses during the initial ICI-only window.

A droplet-based single-cell RNA-sequencing dataset is also presented in this thesis, the first reported large-scale effort to profile OAC with this methodology. This dataset was used to demonstrate a novel workflow for cryopreservation of small endoscopic biopsies, and the transcriptional profiles of all identified cell types were leveraged in a deconvolution approach to reveal the potential contribution of tumour cell composition to immunochemotherapy outcome. These data were also applied to investigate how tumour cells evolve transcriptionally in response to ICI at single-cell resolution and the implications for ICI resistance mechanisms.

In summary, this thesis deeply profiles a first-line immunochemotherapy cohort and identifies a number of predictive biomarkers and potential mechanisms of response and resistance before, during, and after treatment. If validated, these findings could be used to help improve management and outcomes of patients with inoperable OAC.

DECLARATION AND FOREWORD

I declare that the material presented in this thesis is my own work, except where otherwise noted. It must be emphasized that the multi-centre clinical trial described in this work was necessarily a highly collaborative project, particularly regarding clinical management and sample collection. However, much of the generation of transcriptomic sequencing data in this thesis, and nearly all data analysis, is my own work. Where my colleagues have made key contributions, I have clearly identified their contribution using a footnote in the relevant section. I also declare that this work has not been previously submitted to fulfil the requirements of any other degree, diploma, or qualification, and that it confirms to the thesis requirements of the University of Oxford and the Nuffield Department of Medicine (word count exclusive of bibliography, table of contents, appendices, figures and their legends, and tables is approximately 49,875 words). These analyses were generated using a cut-off date of 16 October 2020 for clinical outcome data. In the event of future publications associated with this thesis, which may use more recent clinical outcome data, the publications should be assumed to contain the more up-to-date findings. For those interested in discussing or examining the findings presented in this thesis further, I will maintain my contact information and links to publications associated with this thesis, along with relevant supporting code and data repository information, on my GitHub page (https://github.com/t-carroll).

ACKNOWLEDGEMENTS

I would first like to extend my heartfelt gratitude to all those who participated in the clinical research underpinning this thesis, particularly the patients who enrolled in the LUD2015-005 clinical trial and their families. Their willingness during the most difficult of circumstances to go above and beyond what was required for their care made this work possible. I hope they would be pleased with how their contributions were put to use in this thesis, which represents only a fraction of the scientific research they enabled.

This work would also not have been possible without the contributions of the physicians, clinical staff, and administrators who oversaw the LUD2015-005 trial and TGU Biobank. These include Naomi McGregor, Sam Dobbie, and Lucy Griffith (OCTO), Luminita Puta and the TGU Research Nurses, Stewart Norris-Bulpitt and the Sample Handling Lab (Early Phase Clinical Trials Unit), Katy McCann (Southampton General Hospital), Simon Lord (Department of Oncology), Ludwig Cancer Research (including Kristen Aufiero Ramirez and Aileen Ryan) who provided oversight, data management, and financial support, and particularly Mark Middleton, who served as Chief Investigator for the LUD2015-005 study and imparted valuable clinical insight in his role as my second supervisor. This research was conducted with support from AstraZeneca UK Limited, who provided durvalumab and tremelimumab for the LUD2015-005 trial.

From our LUD2015-005 working group at the Oxford site, I am indebted to Richard Owen and Michael White, who helped design this trial and performed much of the initial sample collection and analysis, as well as to Andrew Roth, who first arranged our working group and whose computational efforts informed many of the analyses I took on in this thesis. I am also thankful for the many others who contributed to this working group over the years, including Benedikt Kessler, Georgina Berridge, Roman Fischer, Joseph Chadwick, Joseph Kaplinsky, Benjamin Schuster-Böckler, Robert Amess, Hantao Lou, Carlos Ruiz Puig, Jaeho Chang, Phil Xie, Ayo Omiyale, and Chan Phetsouphanh.

I am grateful to the Oxford Biomedical Data Science training programme for easing my transition to bioinformatics, and Angie Green, Lorne Lonie, David Ahern, and Neil Ashley for technical guidance.

I would like to extend special thanks to my examiners, Francesca Buffa and Christian Ottensmeier, for their time and effort in reviewing what turned out to be a long and wide-ranging thesis.

Much of my growth over the past few years would not have been possible without the Rhodes Trust, whose funding allowed me to take up this opportunity, and whose community allowed me to meet and befriend some truly inspiring people. I will always be indebted to my selection committee for their investment in my development.

I am extremely grateful to my primary supervisor Xin Lu for her mentorship and guidance over the years. Xin supported me as I took time away to transition from my pure wet lab background towards bioinformatics, and I was never left wanting for resources and feedback when I came up with a new idea I wanted to chase. These can be rare opportunities during a DPhil, for which I am very thankful.

I am also glad to have had the chance to work with all the members of the Lu Lab, who could always be counted on for help and insight when I needed it. I am particularly grateful to have crossed paths with Brittany-Amber Jacobs and Joseph Chadwick. They were an ever-present source of support throughout my DPhil, including providing useful comments on this thesis. Joseph was my right-hand man in many of the projects described herein, while Brittany helped keep me focused, organized, and practical in my work. More than that, their friendship inside and outside of the lab made my time in Oxford much brighter.

Finally, I cannot begin to express my gratitude to my sister Katie and my parents Daniel and Susan, who have been my support system throughout my life, but especially during the past few years. I could not have crossed this finish line without your support.

TABLE OF CONTENTS

DEDICATION ii
ABSTRACT iii
DECLARATION AND FOREWORD v
ACKNOWLEDGEMENTS vi
TABLE OF CONTENTS viii
Chapter 1: Introduction 1
1.1 Anatomy and development of the oesophagus 1
1.2 Cancers of the oesophagus 3
1.3 Standard management of oesophageal adenocarcinoma and clinical outcomes 5
1.4 Cancer immunotherapy: Background and function 7
1.5 Immunochemotherapy for oesophageal adenocarcinoma 10
1.6 Predictive biomarkers of ICI outcome and their potential utility in OAC 12
1.7 The LUD2015-005 clinical trial and aims of this thesis 17
Chapter 2: Materials and Methods 19
2.1 Study design and treatment 19
2.2 Patient eligibility 19
2.3 Assessment of patient outcomes 19
2.4 Study procedures and sample collection 20
2.5 Supplemental sample collection 21
2.6 Single-cell RNA-sequencing 22
2.6.1 Tissue dissociation 22
2.6.2 Encapsulation, library preparation, and sequencing 23
2.7 Bulk tissue RNA-sequencing 23
2.8 Bioinformatics 24
2.8.1 Repository data access 24
2.8.2 Single-cell RNA-sequencing 24
2.8.3 Bulk RNA-sequencing 25
2.8.4 Deconvolution analysis 27
2.8.5 Survival analysis 28
2.9 Figure design 28
Chapter 3: Clinical and Safety Outcomes of the LUD2015-005 Trial 29
3.1 Introduction 29
3.2 LUD2015-005 inoperable OAC cohort: Study design and sampling protocol 30
3.3 Clinical characteristics 32

3.4 Survival outcomes from the LUD2015-005 trial....33
3.4.1 Cohort summary metrics....33
3.4.2 Propensity-matched comparison with a chemotherapy-treated patient registry....34
3.5 Response and clinical benefit outcomes from the LUD2015-005 trial....36
3.6 Assessing potential confounders of biomarker discovery....39
3.7 Adverse events and their association with patient outcomes....41
3.8 Summary and brief discussion....44
Chapter 4: Genomic and Transcriptomic Workflows to Evaluate Previously Reported and Novel Predictive Biomarkers....48
4.1 Introduction....48
4.2 Genomic biomarkers....49
4.2.1 Tumour mutational burden....49
4.2.2 HLA class I genotype....51
4.3 Evaluation of an RNA extraction and library preparation workflow for transcriptional biomarker analysis....53
4.3.1 Background and RNA-sequencing workflow....53
4.3.2 RNA extraction and quality control metrics....54
4.3.3 RNA-sequencing and assessment of alignment quality....57
4.4 Overview of bulk transcriptome profiles....59
4.4.1 PCA analysis....59
4.4.2 Characterisation of OAC in comparison with normal GI tissues....60
4.5 Finding novel predictive markers of outcomes using a *de novo* differential expression approach....68
4.5.1 Pre-treatment samples....68
4.5.1.1 Clinical benefit....68
4.5.1.2 irRECIST response....70
4.5.2 On-treatment samples (ICI-only)....74
4.5.3 Post-treatment samples (PostTx)....77
4.6 Profiling changes in transcriptional profiles to elucidate mechanisms of ICI activity....82
4.6.1 Identification of an ICI-induced transcriptional response signature (INCITE) and correlation with outcome....82
4.6.1.1 Definition and utility of the INCITE signature....82
4.6.1.2 Genes that predict INCITE-reactive and INCITE-quiescent response subtypes....86
4.6.2 Differences in ICI-induced transcriptome changes between CB and NCB patients....88
4.6.3 Assessing the effect of a priming αCTLA-4 dose on ICI-induced expression changes....90
4.6.4 Effect of ICI on normal GI control tissues....93
4.7 Comparison of results with previously reported findings....94

4.7.1 Meta-analysis of DEGs from PD-1/PD-L1-targeted ICI cohorts shows limited overlap of high-confidence predictive markers between OAC and other settings 94
4.7.2 Comparison of previously defined transcriptional markers of ICI response with LUD2015-005 predictive markers 100
4.7.3 Differing prevalence of EBV-infection and MSI is a likely driver of differences in predictive biomarkers for ICI response between oesophageal and gastric adenocarcinoma 107
4.7.4 ICI response of virus-associated tumours 109
4.8 Summary and brief discussion 111
Chapter 5: Single-cell RNA-sequencing to Profile OAC Cell Composition and Advanced Applications for Understanding Immunochemotherapy Outcomes 119
5.1 Introduction 119
5.2 Evaluation of suitability of slow-frozen whole endoscopic biopsies for scRNA-seq 121
5.2.1 Background 121
5.2.2 Performance of fresh and slow-frozen biopsies for dissociation and scRNA-seq 122
5.3 A single-cell atlas of OAC, OSCC, Barrett's, and normal GI control tissues 126
5.3.1 Overview of broad cellular compartments 126
5.3.2 Addressing cryopreservation-induced batch effect 128
5.3.2.1 Characterisation of expression variation between fresh and slow-frozen biopsies 128
5.3.2.2 Minimizing cryopreservation batch effect using a mutual nearest-neighbour approach 130
5.3.3 Identification of cell types represented in the LUD2015-005 scRNA-seq cell atlas 133
5.3.3.1 Epithelial compartment 133
5.3.3.2 Stromal compartment 145
5.3.3.3 Lymphocyte immune cell compartment 152
5.3.3.4 Phagocyte immune cell compartment 162
5.4 Deconvolution to understand and predict immunochemotherapy outcomes in OAC 168
5.4.1 Background and validation 168
5.4.2 Deconvolution of the LUD2015-005 inoperable OAC cohort 172
5.4.2.1 Deconvolution suggests, but cannot confirm, varying tumour purity of OAC biopsies 172
5.4.2.2 PreTx monocyte-like cells predict improved outcome on immunochemotherapy 173
5.4.2.3 RERGL^{+} vascular SMCs mark early resistance at the ICI-only timepoint 182
5.4.2.4 Specific T cell subtypes drive INCITE reactivity during the ICI-only window 184
5.4.2.5 Cell types associated with outcome at the PostTx timepoint 185
5.5 Further scRNA-seq approaches to understand immunochemotherapy outcomes 187

5.5.1 Predicting outcome using cell type-specific expression....187
5.5.2 Expression map of targeted checkpoint molecules....190
5.5.3 On-treatment tumour evolution....192
5.5.3.1 OAC-HDHI (No Clinical Benefit)....193
5.5.3.2 OAC-HENB (No Clinical Benefit)....194
5.5.3.3 OAC-HGFI (No Clinical Benefit)....194
5.5.3.4 OAC-PAHE (NA Clinical Benefit)....195
5.5.3.5 OAC-ACMO (Clinical Benefit)....197
5.5.3.6 OAC-JCNP (Clinical Benefit)....198
5.5.3.7 OAC-LPPN (Clinical Benefit)....199
5.5.3.8 OAC-GDBD (Clinical Benefit)....200
5.5.3.9 Operable OAC patients....201
5.5.3.10: Summary of ICI-induced transcriptional evolution....203
5.5.3.11: Genomic amplification may underlie EGFR overexpression in response to ICI in multiple patients....204
5.5.3.12 Implications of PostTx Barrett's and PostTx OAC phenotypes at the primary lesion site....207
5.6 Summary and brief discussion....210
Chapter 6: Conclusions....222
6.1 ICI shows promising activity in some OAC patients....222
6.2 Combination ICI may be a more effective strategy than targeting PD-1 alone....223
6.3 Highly OAC-specific CT antigens could be additional immunotherapeutic targets....224
6.4 Both ICI-related and chemotherapy-related adverse events are potentially useful on-treatment markers of immunochemotherapy outcome....225
6.5 T cell inflammation is not a predictor of immunochemotherapy response in OAC, but increases do correlate with early responses to ICI-only....226
6.6 Monocyte infiltration and coding TMB are markers of patients particularly likely to benefit from ICI addition....228
6.7 OAC predictive biomarkers show more overlap with NSCLC than gastric cancer....229
6.8 Pathogen-associated gastro-oesophageal cancers are prime targets for ICI....231
6.9 ICI resistance mechanisms are varied, but potentially overcome by immunochemotherapy in some patients....233
6.10 Future directions and perspectives....235
6.10.1 Next steps for further study....235
6.10.2 Towards an optimized deconvolution workflow for biomarker cohort studies....237
6.10.3 Towards a window strategy for immunochemotherapy treatment....239
6.10.4 Towards a post-treatment biopsy workflow to guide clinical management following immunochemotherapy....240

6.10.5 Closing comments 241
REFERENCES 242
APPENDICES 261
Appendix A: LUD2015-005 Eligibility Criteria 261
Appendix B: Details of Supplementary Operable OAC Patients 264
Appendix C: Significant Genes from Key Differential Expression Tests 266
Appendix C-1: OAC vs. Normal Gastrointestinal Tissues (PreTx) 267
Appendix C-2: Clinical Benefit vs. No Clinical Benefit (PreTx OAC) 290
Appendix C-3: Responder vs. Non-Responder (PreTx OAC) 292
Appendix C-4: Clinical Benefit vs. No Clinical Benefit (ICI-only OAC) 301
Appendix C-5: Clinical Benefit vs. No Clinical Benefit (PostTx OAC) 305
Appendix C-6: ICI-only vs. PreTx (All patients, OAC) 316

Chapter 1: Introduction

1.1 Anatomy and development of the oesophagus

The oesophagus is a muscular tube which functions to transfer material from the mouth to the stomach. It is unique among internal organs as the only structure to span three separate body cavities[3]. Like all organs of the gastrointestinal (GI) tract, the oesophagus is formed of four layers: the mucosa (innermost), submucosa, muscularis propria, and the outermost layer, which in the oesophagus is adventitia rather than the serosa found in other GI organs[4–6]. The mucosa is lined with a non-keratinized stratified squamous epithelium, formed of basal, suprabasal, and superficial layers, which directly encounters the oesophageal lumen. At the junction with the stomach, this squamous epithelium of the oesophagus transitions to the simple columnar epithelium of the stomach at a region known as the squamo-columnar junction (SCJ). The inner layers of the oesophageal mucosa are the lamina propria, a region of connective tissue underlying the epithelium, and the muscularis mucosa, a thin layer of smooth muscle which forms the innermost layer of the mucosa.

The oesophageal submucosa is a region of connective tissue interspersed with immune cells, vascular and lymphatic endothelium, and nerve cells[4,6]. It also contains submucosal glands, which produce mucous secretions that help clear the oesophagus and protect it from acidic gastric fluid[7,8]. Under the submucosa is a layer of muscle known as the muscularis propria. This region is composed of skeletal muscle in the proximal portion of the oesophagus, but transitions to smooth muscle at the distal end, which helps couple voluntary swallowing with involuntary peristaltic contractions to efficiently move food boluses to the stomach[9]. The outermost layer is the adventitia, a layer of connective tissue which attaches the oesophagus to nearby anatomical structures. A summary of the structure of the oesophagus is provided in Figure 1.1.

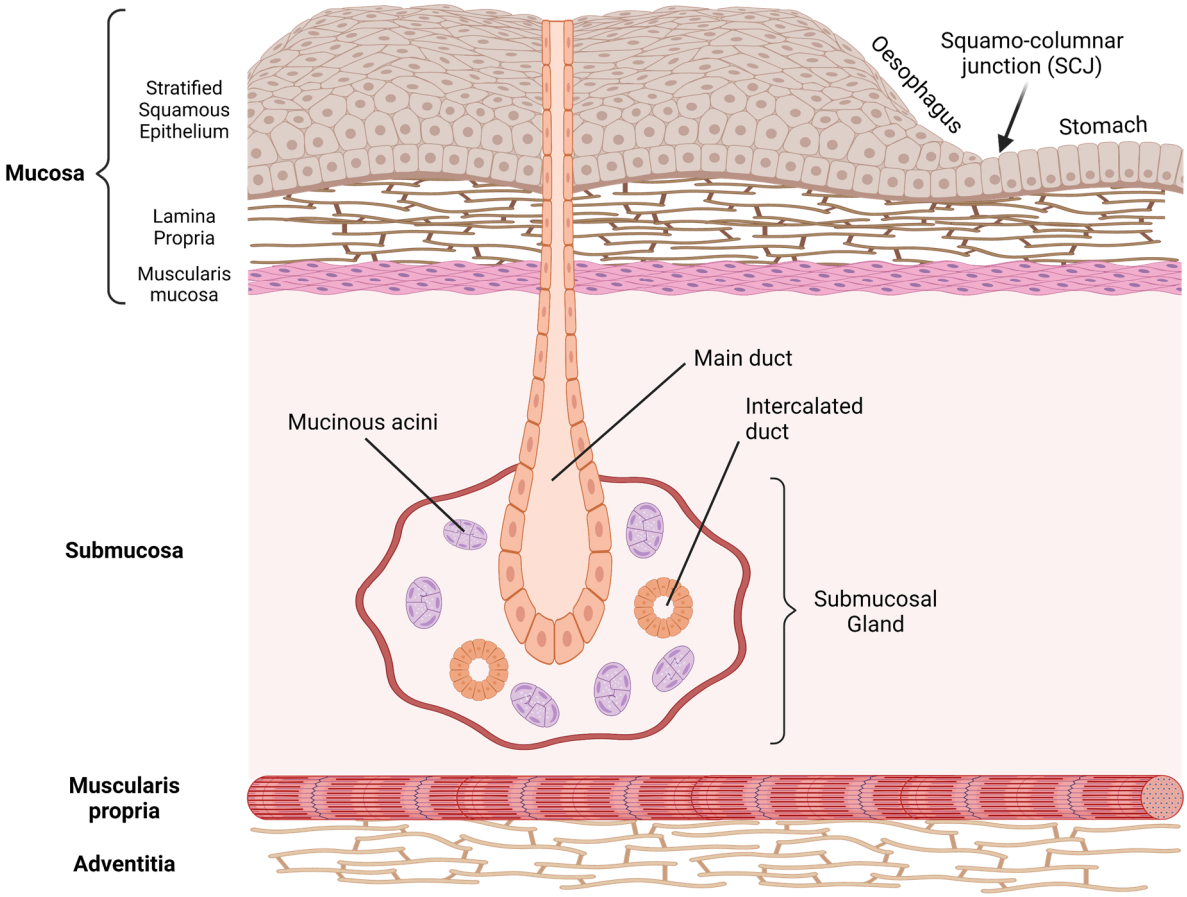


Figure 1.1: Anatomical structure of the oesophagus and junction with the stomach. The lumen of the oesophagus is at the top of the figure. Cells and structures shown are not necessarily to scale.

During embryonic development, the oesophagus arises from the foregut, a derivative of the endodermal layer. The same region of the foregut eventually develops into the trachea, lungs, and oesophagus, including both the epithelium and mucous glands[10]. Ciliated columnar epithelium derived from the foregut covers the oesophagus during much of embryonic development, but is replaced by stratified squamous epithelium before birth. The processes of oesophageal separation from the respiratory system and its subsequent squamous differentiation are driven by *SOX2*, the key squamous cell fate regulator *TP63*, and coordinated expression of a number of other transcription factors[11,12]. Additionally, while the stratified epithelium of the oesophagus differs in type from the pseudostratified columnar epithelium of the lung, these two epithelial layers share many features in adulthood that are likely related to their shared developmental origin[13]. These include the presence of basal cells expressing TP63, KRT5, and SOX2 in both organs, as well as a differentiation program regulated by Notch

signalling and common repair pathways. In summary, while the oesophagus has unique epithelial features among endoderm-derived tissues, the oesophagus also preserves some shared molecular features with developmentally associated organs.

1.2 Cancers of the oesophagus

Cancers of the oesophagus are now the eighth most prevalent form of cancer, and the sixth leading cause of cancer mortality worldwide[14]. The major histological type of oesophageal cancer globally is oesophageal squamous cell carcinoma (OSCC). However, the incidence of oesophageal adenocarcinoma (OAC) has risen rapidly in the Western world over the past few decades, and OAC is now the predominant form of oesophageal cancer in many Western countries, including the United Kingdom[15–20]. While OSCC arises from the squamous epithelium lining the oesophagus, OAC is thought to arise in many cases from a pre-malignant lesion known as Barrett's oesophagus (BO), a region of metaplastic epithelium that occurs in some people in the distal portion of the oesophagus[21,22]. The primary risk factors for OSCC include tobacco use and alcohol consumption. For OAC, tobacco use is also a major risk factor, as well as obesity, diet, and gastro-oesophageal reflux disease (GORD)[19].

In addition to the epidemiological risk factors, the likelihood of developing OAC is greatly increased following the detection of BO, befitting its status as a premalignant lesion[23]. BO involves the replacement of normal squamous epithelium with regions of metaplastic columnar epithelium, which sometimes, but not always, involves the presence of goblet cells and other features of intestinal metaplasia[24]. The development of BO is itself associated with GORD and bile reflux, and some postulate that this metaplasia may form in the oesophagus as a protective adaptation to repeated epithelial damage from stomach acid or bile[19,25]. The progression of BO to OAC is thought to occur as a gradual process involving increasing levels of dysplasia and genomic aberrations[22]. The annual risk for the general patient with non-dysplastic BO to progress to OAC is about 0.5%, but this becomes much higher once dysplasia develops, with risk estimates that BO with high-grade dysplasia will progress to OAC ranging from 16 to 59%[19]. Not all OAC patients have a known history of BO, suggesting there could be additional developmental pathways by which OAC can arise; however, modelling suggests that the low rate

of BO diagnosis drives this finding, and that BO is likely present as the putative source of OAC in most, if not all, patients[26].

There are many competing hypotheses attempting to explain the origin of Barrett's metaplasia, and therefore, the ultimate origin of OACs that arise from this condition[27,28]. The simplest could be upward migration of columnar epithelium from the stomach, but this does not convincingly explain the intestinal metaplasia and diverse cell lineages seen in BO. Transdifferentiation of normal oesophageal epithelium following GORD or bile injury is an alternative possibility. Another hypothesis identifies a population of transitional basal cells at the gastro-oesophageal junction (GOJ) that is capable or reproducing the various lineages of metaplasia seen in BO, including goblet cells. Finally, oesophageal submucosal glands have been proposed as a possible origin of BO[29,30]. As these glands can help regenerate the oesophagus following injury, it is possible an abnormal healing response from stem-like populations in the glands could regenerate the area of injury with columnar metaplasia rather than neosquamous epithelium[28,29]. Interestingly, BO seems to be associated with the proliferation of these glands, and particularly of their ducts, which have increased prevalence in some BO models[28,29,31]. A recent scRNA-seq report from our group showed transcriptional overlap between BO epithelium and cells from oesophageal submucosal glands, lending further support to the oesophageal submucosal gland hypothesis[32]. Nevertheless, consensus has not yet been reached on the exact origin and pathogenesis of this pre-malignant lesion for OAC.

While OSCC is most commonly found in the middle third of the oesophagus, OAC is more restricted to the distal third closest to the stomach[21,33]. Particularly distal OACs often involve the gastro-oesophageal junction (GOJ), which can be difficult to distinguish from gastric carcinomas involving the GOJ. The Siewert classification system was developed to classify these tumours of the GOJ, which break down into type I (adenocarcinoma of the distal oesophagus), II (true carcinoma of the gastric cardia), and III (subcardial gastric cancer infiltrating the distal oesophagus)[34]. These classifications are sometimes used to help inform which surgical intervention is most appropriate[35].

While different surgical interventions may be appropriate for different Siewert types, there is likely some degree of molecular overlap between OAC, GOJ adenocarcinoma subtypes, and gastric cancer. Indeed, a review of The Cancer Genome Atlas (TCGA) data for gastro-oesophageal cancers showed that OAC is molecularly similar to gastric cancers with chromosomal instability (CIN), one of the four major molecular subtypes of gastric cancer[33]. However, there are some key differences between OAC and other forms of gastric cancer. Other molecular subtypes of gastric cancer involve microsatellite instability (MSI) and infection with Epstein-Barr Virus (EBV), which are almost never seen in OAC without GOJ involvement[33]. Additionally, increased risk for gastric cancer is associated with infection with *Helicobacter pylori* (*H. pylori*), a gram-negative bacterium that infects the gastric lining in more than half of the world's population[36]. However, *H. pylori* is generally thought to be inversely associated with risk for OAC, perhaps through reducing the risk for GORD[37]. Therefore, while there is some molecular and anatomical overlap between OAC and CIN gastric cancer, there are also many phenotypic and aetiological differences that differentiate OAC from the majority of gastric cancers. As oesophageal cancers also share key molecular features with lung cancer[38,39], the molecular profile of oesophageal cancers may be best studied in the broader context of cancers from foregut-derived tissues, rather than presuming all gastro-oesophageal cancers to have the highest degree of molecular similarity due to their anatomical proximity.

1.3 Standard management of oesophageal adenocarcinoma and clinical outcomes

Given the rapid increase in OAC incidence in recent decades, optimizing treatment options for this disease is a matter of considerable importance. According to the current guidelines from the European Society of Medical Oncology (ESMO)[40], the treatment of choice for localised OAC is surgical resection, often by oesophagectomy unless the cancers are particularly superficial. For OACs with locally advanced spread, surgery is still employed, although the addition of neoadjuvant or perioperative treatment is also recommended to improve the chances of complete tumour resection. Treatment strategies in this setting usually involve chemotherapy with a fluoropyrimidine and platinum agent, with or without radiotherapy.

Unfortunately, a significant percentage of patients with OAC have disease that is considered inoperable at the time of diagnosis, meaning that surgery with curative intent is not feasible. The bulk of these inoperable patients have metastatic disease, which is the case for approximately 40% of all OAC patients at the time of diagnosis, but some non-metastatic patients with certain patterns of locoregional tumour spread or poor health may also not be candidates for surgery[41–43]. For these patients with inoperable disease, palliative therapy is given with the goal of extending survival and reducing symptomatic burden. Fluoropyrimidine/oxaliplatin chemotherapy combinations such as capecitabine/oxaliplatin (CapOx) are usually employed as first-line palliative therapy[40]. The evidence for the efficacy of this regimen in OAC is limited, but it is given in this setting on the basis of guidelines developed for gastric cancer[44]. Radiotherapy can also be considered depending on the clinical situation; for instance, brachytherapy can be used to relieve dysphagia and other symptoms associated with the primary oesophageal lesion.

Outcomes for patients with inoperable gastro-oesophageal cancers treated with palliative chemotherapy were summarized in a recent Cochrane review[45]. While first-line palliative CapOx likely helps extend lifespan in some patients with inoperable OAC, survival outcomes are still limited, ranging from a median overall survival (OS) of 6.4 months in a 2006 study to 10.1 months in a more recent cohort[46,47] (Table 1.1). Certain CapOx-based triplet therapy combinations may increase survival outcomes further, although the median survival on these therapies still tends to be less than one year[40,48,49]. All in all, prognosis for inoperable OAC patients remains poor on conventional chemotherapy, and therefore additional effective treatment options are sorely needed to prolong life expectancy for this expanding patient population.

Cohort	Treatment	Cohort size	Median OS (months)	Median PFS (months)	1 year OS (%)	Response rate (%)	Publication date
Jatoi et al.	Oxaliplatin, capecitabine	43	6.4	4	NR	35	2006
Dijksterhuis et al.	Oxaliplatin, capecitabine	87	10.1	6.9	NR	NR	2019
REAL-2							2008
ECF	Epirubicin, cisplatin, 5-FU	263	9.9	6.2	38	41	
ECX	Epirubicin, cisplatin, capecitabine	250	9.9	6.7	41	46	
EOF	Epirubicin, oxaliplatin, 5-FU	245	9.3	6.5	40	42	
EOX	Epirubicin, oxaliplatin, capecitabine	244	11.2	7.0	47	48	
REAL-3							2013
EOC	Epirubicin, oxaliplatin, capecitabine	238	11.3	7.4	46	42	

Table 1.1: Outcomes of patients with inoperable gastro-oesophageal cancer in studies of platinum/fluoropyrimidine-based regimens. NR: Not reported; 5-FU: 5-Fluorouracil. Response refers to tumour shrinkage following initiation of therapy as defined by RECIST criteria (see Section 1.6).

1.4 Cancer immunotherapy: Background and function

Immunotherapy agents are a class of drug that act by modulating the immune system. When applied to cancer, these agents modulate immune cells to promote their anti-tumour activity, which differentiates this approach from the direct action of chemotherapy or targeted agents on cancer cells. The first use of an immunotherapeutic strategy to combat cancer came from William Coley in 1893, who built on previous reports that some tumours would spontaneously regress following naturally-acquired pathogenic infection[50]. Coley injected patients with a mixture of pathogens ("Coley's toxins") which generated immune reactions that induced regressions in some patients, some of which may have been complete or near-complete[51].

Nevertheless, Coley's work was met with scepticism at the time, and further *in vivo* usage of cancer immunotherapeutics was limited until the 1970s, when intravesicular injection of Bacillus Calmette-Guerin (BCG), an attenuated vaccine against tuberculosis, was shown to induce immune-mediated regressions in bladder cancer[52,53]. Following regulatory approval, BCG became a standard treatment for certain forms of bladder cancer, where it remains in use to this day[54]. After the discovery of this impactful application of BCG, a revival of interest was seen in developing further immunotherapies to treat cancer. Within the past three decades, a rapidly expanding variety of immunotherapeutic approaches were developed which are now available in the clinic. These include cytokine treatment (e.g. IL-2), cancer vaccines (sipuleucel-T), adoptive cell transfer and Chimeric Antigen Receptor (CAR) T cells (tisagenlecleucel), oncolytic viruses (e.g. T-VEC), and immune checkpoint inhibitors, which are the most widely used class of immunotherapeutic agent at present[55].

An immune checkpoint inhibitor (ICI) is an antibody-based immunotherapy which works by blocking the functionality of co-inhibitory ("checkpoint") molecules expressed on T cells. Upon recognition of a particular antigen, T cells rely on a network of co-inhibitory and co-stimulatory signals, delivered by antigen-presenting cells and other components of the immune system, to determine the appropriate response. Activation of co-inhibitory receptors will encourage T cell anergy and tolerance of a particular antigen, while activation of co-stimulatory receptors will encourage T cell activation and subsequent immune responses against that antigen. Cancer cells express a variety of cancer-testis and neoantigens that should generate anti-tumoural adaptive immune responses, but they can evade the immune system by modulating these co-inhibitory pathways to promote immune tolerance. One such example is programmed death-ligand 1 (PD-L1), which triggers the co-inhibitory receptor programmed death-1 (PD-1) on T cells. PD-L1 has moderate expression in certain non-malignant cell types under normal conditions, which helps to maintain immune homeostasis, but can also be aberrantly overexpressed by malignant cells in order to induce immunosuppression[56]. ICIs work by blocking the interaction between co-inhibitory receptors and their ligands. This both interrupts tumour-specific immune evasion mechanisms and reduces general co-inhibitory signalling from non-malignant cells, which

combine to create a more pro-inflammatory state that can generate anti-tumoural immune responses. The mechanisms of commonly utilized ICIs are summarized in Figure 1.2[57,58].

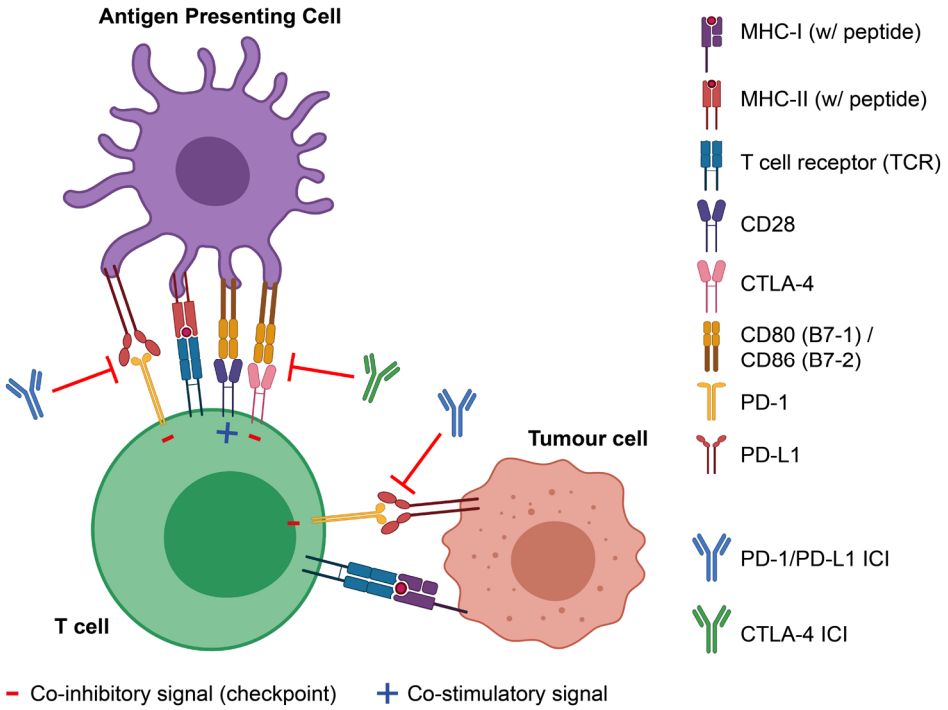


Figure 1.2: Mechanism of common ICI agents. PD-1/PD-L1 ICI blocks triggering of PD-1 co-inhibitory molecules on T cells by ligands (e.g. PD-L1) expressed on tumour or antigen-presenting cells (APCs). CTLA-4 ICI prevents triggering of CTLA-4 co-inhibition by CD80/CD86 expressed by APCs, but still allows CD80/CD86 to trigger co-stimulatory CD28 signalling. MHC: Major histocompatibility complex.

ICIs have now shown important survival benefits for patients in a number of cancer settings, and as such the research leading to their development was awarded the 2018 Nobel Prize in Physiology or Medicine[59]. ICIs targeted against the co-inhibitory receptor CTLA-4 were the first to be developed for use in the clinic, with the first indication for inoperable melanoma receiving regulatory approval in 2011[60,61]. Agents targeting the PD-1/PD-L1 axis soon followed, which demonstrated improved outcomes over CTLA-4 agents in melanoma[55,62]. The combination of CTLA-4 and PD-1/PD-L1 ICI agents showed even further survival benefits: Indeed, within a period of just two decades, the life expectancy of a patient diagnosed with inoperable melanoma went from less than one year on previous chemotherapeutic approaches to over 5 years on some modern ICI combinations[62,63]. While not all cancer types show equal responsiveness to ICI as does melanoma, ICI agents have shown survival benefits and are now used in the clinic in a

variety of other settings, including bladder, kidney, and lung cancer[55,64]. In summary, ICI agents are a recently implemented class of immunotherapy drug that have the potential to significantly improve previously poor long-term prognoses in cancer types where they have activity.

1.5 Immunochemotherapy for oesophageal adenocarcinoma

Given the success of ICI agents in other settings and the currently poor prognosis of inoperable OAC, interest has developed into whether addition of ICI could benefit patients with this disease. While ICI agents may not work well in every cancer type, one piece of supporting evidence for their potential efficacy in oesophageal tumours is that more highly mutated tumour types tend to show better responses to ICI[65]. While oesophageal cancers are not as mutated as highly ICI-sensitive cancers such as melanoma, they do show a relatively high mutational burden which is on par with other cancer types where ICI is used in the clinic (Figure 1.3).

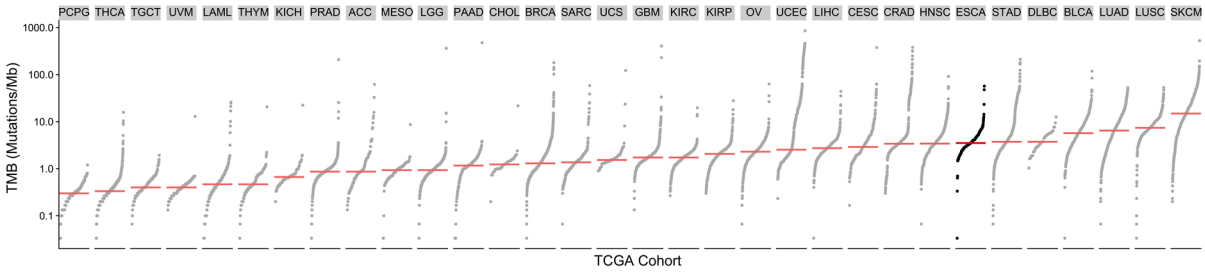


Figure 1.3: Mutational burdens of common cancer types. Total coding mutational burden (TMB) counts from the TCGA PanCancer[66] study were downloaded using cBioPortal[67,68], and converted to a rate per Megabase (Mutations/Mb) by dividing by the size of the human coding space used by cBioPortal. This rate is displayed on a log10-transformed scale, with each dot representing a separate tumour sample grouped by TCGA cancer type cohorts. The red line represents the median tumour burden for that cancer type, and cancer types are arranged according to this value, with the most highly mutated cohort (SKCM, for melanoma) on the right. Cohort names are given across the top using TCGA PanCancer codes, with the exception of colorectal adenocarcinoma (CRAD instead of COADREAD). Oesophageal cancers (ESCA), which includes both squamous cell and adenocarcinomas, are highlighted with black dots.

In oesophageal cancers, ICIs had been approved for use in certain cancers that had progressed following one or more lines of previous therapy[69,70], but no indications for use had been approved in the first-line (previously untreated) setting until recently. However, in the earlier part of this year (2021), the first regulatory approvals were issued for the use of ICI in first-line treatments for inoperable oesophageal cancers[71,72]. This followed reports from phase III trials studying the addition of ICIs targeting PD-1 to standard chemotherapy, which generally showed survival benefits over giving chemotherapy alone[73–76]. In the inoperable setting, two trials

(CheckMate 649 and KEYNOTE-590) showed significant improvements in OS for immunochemotherapy, with the average patient experiencing about two months of additional survival with the addition of ICI (Table 1.2). One trial (ATTRACTION-4) did not show a significant difference in OS for the addition of αPD-1, although there was a still highly significant increase in survival without tumour progression. The median OS under the chemotherapy-only arm of the ATTRACTION-4 was longer than other trials in the inoperable setting, a potential outlier perhaps caused by differences in patient selection or chemotherapy strategy. The longer OS on chemotherapy alone in this study could have masked detection of OS benefit with ICI addition. Alternatively, the cohort was an entirely Asian population, while the other studies had a primarily Caucasian ethnic composition, suggesting the genetic background could also play a role in the differential outcomes. Regardless, the reported PFS outcomes from ATTRACTION-4 do agree with CheckMate 649 and KEYNOTE-590 in finding that addition of ICI provides a benefit.

Subgroup analyses from these trials may show some suggestion that OAC could experience less benefit from the addition of ICI than OSCC, but no conclusive claims can be made on the basis of the currently published data[77,78]. Indeed, in a cohort largely comprised of operable OAC patients, the CheckMate 577 study showed that adjuvant αPD-1 following surgery significantly improved survival without recurrent disease when compared with adjuvant placebo, with nearly a year of additional survival without recurrence in the αPD-1 arm on average. Therefore, ICI likely has activity in some OAC patients, although future studies would be needed to clarify the effect size for inoperable OACs in comparison with other gastro-oesophageal cancers. In summary, the latest phase III data provide evidence of survival benefits for the addition of ICI to conventional first-line chemotherapy in oesophageal cancers, and with recent regulatory approvals, immunochemotherapy may soon become a standard first-line option for inoperable OAC, at least for particular subsets of patients[78].

Trial	αPD-1 + Chemo Median OS (months)	Chemo only Median OS (months)	HR (CI)	p-value	Chemotherapy agents	Setting	Notes
CheckMate 649	13.8	11.6	0.80 (99.3% CI: 0.68-0.94)	0.0002	CapOx or FOLFOX	Inoperable. 70% gastric; 16% GOJAC; 13% OAC	
ATTRACTION-4	17.5	17.2	0.90 (95% CI: 0.75-1.08)	0.257	CapOx or S-1/Oxaliplatin	Inoperable or recurrent gastric or GOJ AC.	Significantly higher PFS for ICI addition (HR 0.68, p=0.0007). Exclusively Asian cohort.
KEYNOTE-590	12.4	9.8	0.73 (95% CI: 0.62-0.86)	< 0.0001	Cisplatin + 5-FU	Inoperable OSCC or AC of oesophagus/ GOJ. 73% OSCC	

Trial	αPD-1 Median DSS (months)	Placebo Median DSS (months)	HR (CI)	p-value	Setting	Notes
CheckMate 577	22.4	11.0	0.69 (96.4% CI: 0.56-0.86)	< 0.001	Operable OAC/OSCC/GOJAC, for adjuvant treatment. 71% AC, 29% SCC.	αPD-1 given following neoadjuvant chemoradiotherapy and surgery. Tumours with complete pathological response to chemoradiotherapy were excluded from enrolment.

Table 1.2: Phase III Studies of PD-1 ICI in first-line oesophageal cancers. Survival outcomes (OS: overall survival; PFS: progression-free survival; DSS: Disease-free survival) are shown for the intent-to-treat population where given. Hazard ratio (HR), reported confidence interval (CI), and p-values are displayed as reported for the specific survival outcome. Cohorts are split into inoperable (top) and operable (bottom) groups. 5-FU: 5-fluoruracil; FOLFOX: leucovorin, 5-FU, and oxaliplatin; AC: adenocarcinoma; SCC: squamous cell carcinoma.

1.6 Predictive biomarkers of ICI outcome and their potential utility in OAC

While the addition of ICI to first-line chemotherapy is an exciting development that generally extends patient life expectancy in this setting, the potential survival benefit must be weighed against the potential increased risk for adverse events, which is a possible risk of adding any antineoplastic agent. While ICI-related adverse events are often tolerable and treatable, a potentially increased rate of serious adverse events from combination therapy may be of particular concern for frail or elderly patients with less functional reserve[79]. Indeed, the CheckMate 649 trial reported a somewhat higher rate of serious treatment-related adverse events for their immunochemotherapy arm than for chemotherapy alone (22% vs 12%, respectively), although no statistical test was conducted to determine whether that increase was statistically significant[73]. In countries where patients bear some of the expenses associated

with treatment, ICI is also associated with greater out-of-pocket costs than chemotherapy, and the addition of ICI to standard chemotherapy could further increase detrimental financial strain on some patients[80–82]. Most patients with cancer are willing to accept the possible drawbacks associated with receiving additional therapy if it is likely to provide even a small survival benefit[83]. However, for some patients, possible increases in risk for serious adverse events or financial burden may be dissuasive, depending on how much they could expect to benefit.

To help OAC patients and their physicians navigate decisions about the addition of ICI to standard-of-care chemotherapy, biomarkers that predict the benefit of ICI addition for a given individual would be particularly useful. For instance, for patients with biomarkers suggestive of better outcomes, the addition of ICI would be particularly likely to outweigh any potential of receiving additional therapy. However, for patients not predicted to receive benefit from ICI, employing conventional chemotherapy alone may be more suitable. Biomarkers indicative of benefit or resistance to immunochemotherapy could also allow the research community to develop further mechanistic understanding as to why some tumours respond better than others to this novel treatment approach. If mechanisms of response and resistance to this treatment can be identified, it could allow for additional interventions to be developed that could further improve outcomes for patients with inoperable OAC. Therefore, as discussed in a recent review[78], characterisation of predictive biomarkers for first-line immunochemotherapy in this setting is now of critical importance following the recent regulatory approvals of these treatments, both to assist patient selection and to further understand the reasons behind differing clinical outcomes.

As immunochemotherapy has only recently been approved as first-line treatment for inoperable OAC, studies specifically focused on characterising predictive biomarkers have yet to be conducted in this setting. However, many ICI biomarker studies have been conducted in other cancer types, whose findings could potentially have utility in this setting. For the purpose of clinical research, including biomarker detection, outcomes are usually summarized using survival or response metrics. While increasing survival is a major goal of treatment and a rational metric to use for the discovery of predictive biomarkers, confounding factors other than

the effectiveness of a given therapy can influence patient survival, particularly age, the performance status, and the subsequent treatment strategy taken after the conclusion of the therapy being studied[84]. As a more direct readout of therapeutic effectiveness in cancer treatment, radiological response metrics are often used, which are based on the change in tumour size in response to therapy. These on-treatment changes in tumour size are classified into response categories using criteria such as RECIST[85]. RECIST criteria identify tumours that respond to therapy, which can either be partial (>30% shrinkage) or complete (complete disappearance) responses, as well as non-responding tumours that either show progressive (>20% growth) or stable (between -30% and 20% change) disease.

Since the development of RECIST, increased recognition has developed of an on-treatment phenomenon known as pseudo-progression that can occur in some ICI-treated patients, where the radiologically-assessed tumour size temporarily increases soon after therapy onset not due to actual disease progression, but as a result of beneficial intratumoural immune cell influx[86,87]. To better categorize response patterns from immune-mediated therapies, a number of modified response criteria have been generated. These include irRECIST criteria, which uses the same thresholds as RECIST, but requires confirmation of progression using repeated scans at set intervals to help distinguish between true and pseudo-progression[88]. In addition to their clinical use in determining whether a given therapy should continue, these radiological response criteria can be used as categorical variables to identify biomarkers predictive of ICI efficacy.

In addition to response and OS, combinations of the two approaches may be valuable for certain analyses, such as the use of progression-free survival (PFS; length of survival without progression as defined by response criteria) in place of absolute OS. Related landmark endpoints such as “clinical benefit” have also been defined, which in the context of ICI often categorize patients into binary groups based on whether or not they attain PFS of a certain duration[89]. Biomarker studies usually use response or survival outcomes, or some form of clinical benefit metric, as the basis for statistical tests to identify biomarkers predictive of patient outcomes.

A recent large-scale meta-analysis of genomic and transcriptomic biomarkers of ICI outcome revealed that tumour mutational burden (TMB) and T cell-attractant chemokines such as *CXCL9* were among the most robust predictors of ICI response and survival across cancer types[90]. Tumour mutational burden is a well-studied ICI biomarker. The mechanistic rationale for the link between TMB and ICI outcomes is thought to be that a higher number of mutations generates increased immunogenic neoantigens recognizable by T cells, thereby allowing for a stronger anti-tumoural immune response following the release of T cell inhibition by ICI[91,92]. The cross-cancer link of TMB with ICI outcome is such that some ICI agents have been granted approval for use in solid tumours of any origin, provided they have a particularly high TMB (≥10 mutations per megabase) on a companion diagnostic assay, a rare tissue-agnostic regulatory approval suggestive of broad predictive power of this biomarker across multiple settings[92]. However, subsequent analyses of trial data have failed to show benefit for ICI over conventional chemotherapy in high-TMB populations, perhaps because high-TMB tumours may respond well to chemotherapy as well as ICI; this complicates the potential use of TMB to select patients for ICI regimens in the clinic[93].

In their status as predictive biomarkers, T cell chemokines such as *CXCL9* fall into the broader "hot" and "cold" tumour hypothesis that is also commonly used to try to understand responsiveness to ICI. Hot tumours are ones with high levels of T cell infiltration and inflammatory cytokines, while in cold tumours, T cells are either absent or excluded, along with the concomitant inflammatory signalling and cytotoxic activity[94,95]. Hot tumour types such as melanoma and non-small lung cell cancer (NSCLC) typically respond much better to ICI than cold tumours (e.g. pancreatic, prostate), with one potential explanation being that there is less to activate in cold tumours by using ICI due to the lack of T cell infiltration[96]. Certain signatures of the T cell inflammation associated with "hot" tumours are under investigation for use in the clinic to assist in predicting ICI outcomes[97].

In addition to these biomarkers, for therapies targeting PD-1 or PD-L1, a commonly assessed predictive biomarker is PD-L1 expression. The rationale for this is that tumours with high expression of PD-L1 may have greater inhibition of intratumoural immune escape with therapy

targeting this axis, as opposed to tumours that evade the immune system through other mechanisms that do not rely on PD-L1 overexpression. PD-L1^{high} patients receive a stronger benefit from ICI in some settings, although in some cancers there can still be benefit for giving ICI rather than conventional therapy to both PD-L1^{high} and PD-L1^{low} patients[90,98,99]. Similarly to TMB, PD-L1 expression can be assessed in the clinic by a companion diagnostic assay to determine suitability for ICI therapy. Assessment of PD-L1 expression in this clinical context is performed using a histological assay which reports the fraction of cells positive for PD-L1 (a score based on the percentage of PD-L1 expressing tumour cells, immune cells, or their combined positivity [CPS]). PD-L1 expression scores are often used to select patients for clinical trial enrolment, stratify patients for analysis of outcomes, and now in some settings as an indication for ICI therapies in the clinic[64,100].

Of these three major predictive biomarkers for ICI, the recent phase III studies of the addition of ICI to first-line OAC only reported on PD-L1 expression status, and the findings were mixed. In CheckMate 649 and KEYNOTE 590, the lowest hazard ratio for the benefit of immunochemotherapy over standard chemotherapy was in patients with the highest PD-L1 CPS, but the analysis on the full cohort still showed a survival benefit for immunochemotherapy regardless of CPS, and no test was conducted to determine whether any significant differences in effect size between patients with high CPS and low CPS[73,75]. In CheckMate 577, patients with both high and low tumour-specific expression of PD-L1 showed similar benefit from ICI over placebo in the adjuvant setting[76]. Therefore, while tumours with high PD-L1 expression may respond somewhat better to immunochemotherapy, further investigation is needed to determine whether PD-L1 expression thresholds are appropriate to use for patient selection in this setting, or whether all PD-L1 expression levels are associated with some benefit upon ICI addition.

To date, no analysis of the commonly assessed TMB and T cell inflammation predictive biomarkers has been reported from the phase III studies in this setting. Retrospective biomarkers studies have been conducted on pre-treated gastro-oesophageal cancer cohorts, which have generally shown a trend towards improved OS with increasing TMB, although the

links were either non-significant[101,102] or did not maintain significance following multivariate regression[103]. In any case, no broad biomarker study has yet been conducted in the setting of first-line immunochemotherapy for OAC, and it is therefore unclear whether findings from other settings will hold here. This is particularly true considering that the bulk of ICI biomarker research derives from cohorts in other cancer types where ICI is given by itself, rather than concurrently with chemotherapy. In summary, there is a strong need to validate whether the utility of commonly assessed ICI biomarkers from other settings will hold for immunochemotherapy in OAC, and also to conduct a broader *de novo* biomarker study to determine whether there are other biomarkers that could have stronger predictive power in this unique treatment setting.

1.7 The LUD2015-005 clinical trial and aims of this thesis

In this thesis, I endeavour to apply advanced molecular profiling techniques to identify predictive biomarkers from a cohort of inoperable OAC patients treated with first-line immunochemotherapy approach as a part of the LUD2015-005 clinical trial (NCT02735239/EudraCT 2015-005298-19)[104]. The LUD2015-005 trial is a Phase I/II study of first-line αPD-L1 or αPD-L1/αCTLA-4 ICI in addition to standard-of-care chemotherapy for patients with oesophageal cancer. Under the unique treatment design in this trial, patients are initially treated with a brief ICI-only window prior to the immunochemotherapy phase, allowing for the effects of ICI to be disambiguated from those of the combination therapy. Comprehensive molecular profiling of the samples afforded by this study design provides a particularly valuable opportunity not only to evaluate which biomarkers will be most predictive of outcome in this setting, but also to shed light on mechanisms of resistance and response to ICI and immunochemotherapy.

First, clinical findings from the LUD2015-005 study are summarized, including response and survival outcomes and an assessment of whether adverse events are linked with patient outcome. Next, results from molecular profiling using bulk sequencing methods are shown, including both whole genome sequencing and RNA-sequencing. A comparison with previous ICI-treated cohorts is then conducted, to determine how much overlap there is between ICI

biomarkers from these studies and biomarkers that predict outcome in this immunochemotherapy setting. To conclude, results from the LUD2015-005 oesophageal pathology cell atlas are presented, which contains approximately 65,000 single-cell transcriptomes from immunochemotherapy-treated OAC, normal GI tissues, and other oesophageal disease states. This comprehensive dataset is then applied in various ways to further understand immunochemotherapy outcomes of the LUD2015-005 cohort, including a deconvolution workflow to clarify the role of specific cell types in immunochemotherapy response, and an assessment of the transcriptional evolution of tumour cells over the course of therapy. A number of hypotheses are generated from the clinical and molecular profiling conducted in this study. If these findings can be validated in other cohorts, they may form the basis for future predictive companion diagnostics and methods to monitor on-treatment response and resistance, and could also one day contribute to the development of further interventions to help improve immunochemotherapy outcomes.

Chapter 2: Materials and Methods

2.1 Study design and treatment

The LUD2015-005 clinical trial (NCT02735239; REC reference: 16/SC/0015) is a phase I/II study of durvalumab as first-line therapy for patients with oesophageal or gastro-oesophageal cancer. Following a screening period to confirm eligibility, patients with metastatic or inoperable locally advanced OAC were enrolled into one of three treatment arms for inoperable patients based on a dose escalation strategy, which proceeded to the next stage following establishment of acceptable safety profile at each step. Patients in the first stage received 750 mg of intravenous durvalumab administered biweekly (Q2W) as the ICI agent, those in the second received the same regimen of durvalumab plus a single 37.5 mg priming dose of tremelimumab, and those in the third received durvalumab plus a single 75 mg priming dose of tremelimumab. After 4 weeks of treatment with these ICI agents alone, patients in both cohorts received a maximum of six cycles of chemotherapy with oxaliplatin and capecitabine, in addition to continued Q2W durvalumab. In the absence of a reason to discontinue treatment earlier (such as confirmed disease progression, withdrawal of consent, or unacceptable toxicity), durvalumab infusions continued until the end of the final chemotherapy cycle.

2.2 Patient eligibility

Adult patients with a histologically confirmed diagnosis of locally advanced or metastatic oesophageal or gastro-oesophageal cancer were eligible for enrolment into the inoperable arms of the LUD2015-005 trial. Patients with previous systemic anti-cancer therapy for this same advanced disease were excluded from this study, as were patients who had been treated with ICI in any previous setting. Inclusion criteria included an Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1, and an anticipated lifespan of greater than 4 months. A full list of eligibility criteria can be found in Appendix A.

2.3 Assessment of patient outcomes

The primary outcomes of this phase I/II study involved safety: the number of subjects reporting adverse events, number of subjects experiencing a dose-limiting toxicity, and any changes from treatment onset in laboratory evaluations. Secondary outcome measures were OS, PFS, and

tumour response. Response was assessed using irRECIST criteria[88] from CT scans conducted during the screening period and every 6 weeks thereafter during treatment. Overall tumour assessment criteria from irRECIST were used in this work, which involves the evaluation of target lesions selected for quantitative measurement based on pre-defined criteria as well as the status of new and non-target lesions. irRECIST response categories define complete responses (irCR; complete disappearance of lesions), partial responses (irPR; >30% reduction of measured target lesions in absence of progressing new or non-target lesions), progressive disease (irPR; confirmed >20% increase in measured target lesions from baseline or nadir, or confirmed progression of new or non-target lesions), and stable disease (irSD), an intermediate category that does not meet criteria for irPR or irPD. One patient with an overall response of irNN, which occurs when there is no measurable target lesion at baseline and an absence of on-treatment irPD or irCR for non-target lesions, was considered to be irSD for the purposes of this work. Three patients who did not have an on-treatment CT scan for assessment of irRECIST criteria but were deemed to have clear on-treatment clinical progression were grouped with irPD. irCR and irPR were classified as responders, and all other response categories were classified as non-responders. For this analysis, OS and PFS were calculated including data from the post-study follow-up period (data cut-off 16 October 2020). Median follow-up times for these variables were calculated using the reverse Kaplan-Meier method. A post-hoc exploratory outcome metric was also defined, which was termed “clinical benefit”. A patient was deemed to have attained clinical benefit if a centralized review process determined they had 12 months of survival from the initiation of treatment without confirmed disease progression. One patient who received an alternative therapy within 12 months of treatment onset was excluded from clinical benefit analyses.

2.4 Study procedures and sample collection

Endoscopic biopsies were collected for research analysis prior to the onset of treatment, during the fourth week of treatment with ICI alone, and at the end of the study treatment protocol (approximately 5-6 months after treatment onset). At each endoscopy, up to five 2 mm biopsy pairs were taken from the site of the tumour, while 1-2 pairs of normal control tissue were also taken from the descending duodenum (D2), gastric cardia (at least 2 cm distal to the gastro-

oesophageal junction (GOJ) or distal extent of lesion if crossing the GOJ), and normal oesophagus (at least 2 cm proximal to the GOJ or proximal extent of lesion, whichever is more proximal). Each biopsy pair was split into two cryovials, one of which was immediately snap-frozen using dry ice or liquid nitrogen, while the other was filled with 1 mL of fetal bovine serum (FBS) or heat-inactivated human serum (HS; Sigma H3667) with 10% DMSO and placed at -80°C in a controlled-rate freezing container for slow freezing. Snap and slow-frozen biopsies collected at local enrolment sites were stored at -80°C, and were shipped on dry ice to the central site for long-term storage at -80°C and subsequent analysis.

2.5 Supplemental sample collection

Earlier stage oesophageal cancer biopsies were collected from the operable arms of the LUD2015-005 trial, which enrolled patients deemed suitable for surgery with curative intent. These operable OACs were treated in the neoadjuvant setting with a similar regimen of durvalumab, to which chemotherapy or chemoradiotherapy was added following an ICI-only window. The full characterisation of these operable disease cohorts is beyond the scope of this thesis; however, the relevant treatment and outcome details of the three patients from these cohorts that were profiled for single-cell RNA-sequencing can be found in Appendix B. Following a similar protocol as the metastatic cohorts, 2 mm endoscopic biopsies were obtained from tumour and normal oesophagus prior to treatment and after 4 weeks of durvalumab alone, and were processed for snap and slow-freezing as above. At the surgery marking the end of the study protocol for these operable OAC patients, core biopsies of the tumour site and normal oesophagus were also obtained from resected material.

Additional 2mm endoscopic biopsies of BO and paired normal oesophagus (at least 2cm proximal to extent of BO) were collected from patients undergoing routine monitoring of known BO by the Translational Gastroenterology Unit (TGU) Biobank (John Radcliffe Hospital; REC reference: 11/YH/0020), and released to our team for analysis. Core biopsies from LUD2015-005 surgical specimens and the supplemental biopsies from BO patients were collected fresh in a 1.5 mL microcentrifuge tube containing DMEM/F-12 (Gibco, 11330032) with 10% HS, and were kept on ice until processing. These supplemental biopsies were then either dissociated

fresh for single-cell RNA-sequencing, or were slow-frozen in HS with 10% DMSO in a controlled-rate freezing jar, as above.

2.6 Single-cell RNA-sequencing

2.6.1 Tissue dissociation

Prior to dissociation, slow-frozen biopsies were thawed by agitation of the cryovial in a 37°C water bath. For both fresh and frozen material, biopsies were first transferred using a p1000 pipette tip into a microcentrifuge tube containing PBS, and gentle pipetting of the biopsy was performed to wash off excess blood, mucus, or trace DMSO (for slow-frozen biopsies). The biopsies were then transferred using a p1000 tip into a 15 mL Falcon tube containing 5 mL of fresh or frozen tissue dissociation medium (Table 2.1) pre-warmed to 37°C. Biopsies were dissociated under constant rotation in an incubator at 37°C and 5% CO_2 until fragments reduced significantly in size, approximately 60-90 minutes for fresh tissue, or 30-60 minutes for frozen tissue. The dissociated solution was then passed through a pre-wetted 100 µM CellTrics filter (Sysmex, 04-004-2328), and any tissue fragments remaining on the filter were washed with 10 mL quench buffer to maximize yield. The resulting filtrate was spun down at 300 g for 5 minutes, and supernatant was discarded. For fresh material, red blood cell lysis was performed using red blood cell lysis solution (Miltenyi, 130-094-183), and dead cells were depleted using a magnetic-bead based approach (Miltenyi Dead Cell Removal Kit, 130-090-101). For frozen material, red blood cell contamination of dissociated pellets was not visible, and dead cell removal did not improve viability in optimization tests. Therefore, these purification steps were not performed on frozen material.

Cell pellets at this stage were then resuspended in pre-chilled cell resuspension buffer, and were passed through a 70 µM tip strainer (Flowmi Cell, Sigma-Aldrich, BAH136800070). Filtered cell suspensions were then counted using an automated cell counter (Bio-Rad, TC20), diluted to a concentration of about 1 million total cells/mL in cell resuspension buffer, and kept on ice until encapsulation.

Reagent	Composition
Fresh tissue dissociation medium	DMEM/F12 (Gibco, 11330032) with 10% HS, 0.24 mg/mL Animal Free Collagenase/Dispase Blend II (Sigma-Aldrich, SCR140), and 0.1 mg/mL DNAse I (Sigma-Aldrich, D5025)
Frozen tissue dissociation medium	DMEM with 10% FBS, 2.5 mg/mL Collagenase D (Sigma-Aldrich, 11088866001), 0.5 mg/mL Liberase DL (Sigma-Aldrich, 5466202001), 0.2 mg/mL DNAse I (Sigma-Aldrich, D5025)
Quench buffer	Phosphate-buffered saline (PBS) with 6% FBS
Cell resuspension buffer	PBS with 0.04% BSA

Table 2.1: Composition of dissociation reagents.

2.6.2 Encapsulation, library preparation, and sequencing

Single-cell RNA-sequencing (scRNA-seq) was conducted using a 5' scRNA-seq gene expression workflow (Chromium Single Cell Immune Profiling, Solution v1.1, 10x Genomics). Encapsulation of cells was performed using the Chromium Controller, and then GEM-RT, cDNA amplification, and construction of final libraries was conducted following manufacturer's instructions. Size profiles of amplified cDNA and final sequencing-ready libraries was verified by on-chip electrophoresis (Agilent 2100 Bioanalyzer system) using the High Sensitivity DNA Kit (Agilent, 5067-4626), and concentration of final libraries was assessed by Qubit (BR DNA Assay, ThermoFisher, Q32853). Libraries were sequenced on an Illumina NovaSeq 6000 or NextSeq 550 (26 cycles read 1, 8 cycles i7 index, 98 cycles read 2), targeting a minimum of 20,000 reads per cell.

2.7 Bulk tissue RNA-sequencing

Total RNA was extracted from whole endoscopic biopsies using the mirVana miRNA Isolation Kit (AM1560, ThermoFisher). Briefly, snap-frozen biopsies were transferred directly into pre-chilled lysis/binding buffer in a 1.5 mL microcentrifuge tube, and the tissue was homogenized on ice using a disposable RNAse-free pestle (ThermoFisher, 13236679). Acid-Phenol:Chloroform RNA extraction was then performed following manufacturer's instructions. Genomic DNA was digested from the eluted RNA using a 30 minute incubation with TURBO DNase (TURBO DNA-free Kit, ThermoFisher, AM1907). RNA concentration was then calculated using the Qubit RNA BR Assay (ThermoFisher, Q10211) and A260/230 and A260/280 ratios

were calculated by spectrophotometer (NanoDrop One, ThermoFisher). To assess the RNA integrity number (RIN) for each sample, 1 ng of total RNA in 1 uL sample volume was loaded onto the RNA Pico Chip (Agilent, 5067-1513), and on-chip electrophoresis and RIN calculation was performed by the Agilent 2100 Bioanalyzer system (Total RNA Assay). RNA was then plated and stored at -80°C.

Prior to library preparation, thawed RNA was purified and concentrated using AMPure XP (BeckmanCoulter, A63880) at a 2.8x ratio. Bead-based rRNA depletion and total RNA-sequencing library preparation with dual sample indexing was then conducted using the TruSeq Stranded Total RNA Library Prep Human/Mouse/Rat Kit (Illumina, 20020596). Libraries were sequenced on the Illumina HiSeq (75 cycles read 1, 8 cycles i7 index, 8 cycles i5 index, 75 cycles read 2) to a targeted depth of 50 million reads per library.

2.8 Bioinformatics

2.8.1 Repository data access

For deconvolution, public bulk RNA-sequencing data for normal oesophagus, BO, and OAC were retrieved from accession PRJEB11797 (Maag et al. cohort, European Nucleotide Archive), and pre-processed RNA-seq counts for OAC and OSCC were downloaded from the Genomic Data Commons, and clinical metadata was obtained using RTCGA (v1.16.0). Data access for the publicly available ICI-treated cohorts are summarized in Table 2.2 (see Bulk RNA-sequencing).

2.8.2 Single-cell RNA-sequencing

Processing of scRNA-seq data was conducted using the Cell Ranger pipeline (v3.1.0) from 10x Genomics. Briefly, raw BCL files were converted to demultiplexed FASTQ files using cellranger mkfastq with --use-bases-mask=Y26n*,I8,Y98n*. A custom reference sequence was then prepared using the concatenation of the GRCh38 human genome (ENSEMBL100) and supplementary contigs from the Genomic Data Commons alignment reference[a], namely the decoy contig hs38.d1 and a collection of viral genomes commonly found in human cancer samples. The full GENCODE v34 GTF was used as a gene annotation file, filtering out readthrough transcripts and annotations to PAR regions in chromosome Y as per Cell Ranger

[a] As obtained from https://gdc.cancer.gov/about-data/gdc-data-processing/gdc-reference-files.

recommendations. A custom Cell Ranger reference set was then prepared from these sequence and annotation files using cellranger mkref with default arguments. Finally, the identification of true cells and construction of a UMI counts table for each sample was performed using cellranger count, and these per-sample counts were combined into a single matrix for downstream analysis using cellranger aggr without normalization.

Downstream analysis was conducted using Seurat (v3.9.9.9038). To begin, low-quality cells with less than 200 genes detected or more than 25% mitochondrial reads were removed. For initial clustering of the full dataset, we normalized using SCTransform (v0.3.2.9002), performed UMAP dimensionality reduction (umap-learn v0.5.0), and identified clusters with the default Louvain algorithm. Broad cell type partitions were then identified and subsetted into separate Seurat objects. To characterize smaller cell subsets, the above workflow was repeated on each of these partitioned objects. To minimize batch effect between fresh and frozen samples for these smaller objects, we conducted SCTransform normalization (sctransform v0.3.2.9002) on fresh and frozen samples separately, and then integrated these two datasets using FastMNN (batchelor v1.2.4), a mutual nearest neighbour batch correction method, before proceeding to dimensionality reduction. Marker genes for each identified cluster were calculated using the FindAllMarkers function, and these markers were used to manually assign cell type and subtype labels. For detection of cell type-specific markers associated with patient outcome, a pseudobulk differential expression method was designed to minimize patient-specific genes and prioritize genes associated with outcome in multiple patients. Briefly, SCTransform-normalized counts from all cells in a given biopsy were summed, resulting in a pseudobulk sample. Differential expression was then conducted on these pseudobulk samples using DESeq2 as discuss below, with clinical benefit as the design formula. Figures visualizing the scRNA-seq findings were generated using a combination of Seurat and ggplot2 (v3.3.3).

2.8.3 Bulk RNA-sequencing

Following adapter trimming with TrimGalore (v0.6.4), reads were aligned to the same custom reference genome as for scRNA-seq processing using STAR (v2.7.5), and gene-level counts were prepared from the output by featureCounts (v2.0.1). Evaluation of sequencing and mapping

quality metrics was conducted using FastQC (v0.11.8), MultiQC (v1.9), and Picard Tools (GATK v4.1.7.0) CollectRnaSeqMetrics and CollectMultipleMetrics modules, including assessment of rRNA depletion efficacy by quantifying reads mapped to rRNA regions (both nuclear and mitochondrial rRNA, as well as rRNA pseudogenes).

Differential expression was conducted using DESeq2 (v1.30.1), using a significance threshold of FDR < 0.1 to identify differentially expressed genes. Default arguments were used, with the exception of the use of the "local" fit type to model dispersion estimates. Wald tests were employed unless otherwise specified. Moderated gene-level log-fold changes were calculated with the "ashr" method, and were used as the test statistic for gene set enrichment analysis, which was conducted using fgsea (v1.17.1) on a database of pathways from Reactome and MSigDB (Hallmark and Gene Ontology-Biological Process). Log-normalized counts for heatmap visualization were generated using the varianceStabilizingTransformation function.

A summary of the DESeq2 design formulas and (where applicable) descriptions of batch variables used in static analyses of LUD2015-005 data and publicly available ICI cohorts are summarized in Table 2.2. To assess changes between timepoints, a design formula of ~patient+timepoint was used. Following DESeq2 guidelines, a design formula of ~group+group:ind.n+group:timepoint was used to assess changes across timepoint that were different between groups (such as clinical benefit or treatment group), where ind.n is an integer uniquely identifying individuals within groups. All-zero columns of the model matrix were manually removed in this analysis format.

Cohort	Data Accession	NCT Identifier	Design formula	Batch description
LUD2015-005, OAC (Clinical Benefit)	Present study	NCT02735239	~Clinical_benefit	n/a
LUD2015-005, OAC (Response)	Present study	NCT02735239	~Response_binary	n/a
Samsung Medical Center, Gastric Cancer	PRJEB25780	NCT02589496	~Response_binary	n/a
Yonsei Cancer Centre, NSCLC	GSE126044	NA	~batch+Clinical_benefit	FFPE vs. fresh
Samsung Medical Center, NSCLC	GSE135222	NA	~batch+Clinical_benefit	FFPE/reversely stranded library vs. fresh/unstranded library
UCLA/VIC, Melanoma	GSE78220	NA	~batch+Response_binary	reversely stranded vs. unstranded library
BMS038, Melanoma	GSE91061	NCT01621490 (Part 1)	~Clinical_benefit	n/a
Melanoma Institute Australia, Melanoma	PRJEB23709	NA	~Clinical_benefit	n/a
IMvigor210, Urothelial Carcinoma (UC)	IMvigor210CoreBiologies/ EGAS00001002556	NCT02951767/ NCT02108652	~Response_binary	n/a

Table 2.2: Summary of clinical bulk RNA-sequencing cohorts assessed for genes predictive of ICI outcome. Cohort descriptions and accession information are provided for each study. The DESeq2 design formulas used in each analysis are shown, and an explanation of batch effects that were controlled for (if applicable) are described. NSCLC = non-small cell lung cancer; UC = urothelial carcinoma.

2.8.4 Deconvolution analysis

Normalized counts in a linear scale were used for deconvolution, which were obtained for scRNA-seq data using the vst function from the sctransform package and for bulk RNA-seq data using the median of ratios method embedded in the estimateSizeFactors function from DESeq2. The combination of these normalization methods has been shown to provide accurate results across a variety of deconvolution approaches[105]. The scRNA-seq data was further subsetted to the 3000 most variable genes (as calculated by FindVariableFeatures from Seurat) prior to deconvolution, with exclusion of potentially confounding gene types (mitochondrial, ribosomal, sex-specific, and TCR/BCR variable region genes). Cell proportions were then estimated from these normalized inputs using the BayesPrism deconvolution algorithm (v2.0) using cell type and subtype labels identified in section 5.3 from scRNA-seq data, with OAC epithelial clusters specified as the tumour type for the deconvolution call. Normalized and scaled cell type proportion estimates were used to calculate UMAP projections and perform hierarchical clustering of samples based on their cell composition (using the hclust method with “ward.D” linkage).

2.8.5 Survival analysis

The link between variables and survival outcomes were assessed using the survivalAnalysis package (v0.2.0) in R. analyze_survival was used to conduct univariate Cox regressions and to generate Kaplan-Meier plots, while analyze_multivariate was used for multivariate Cox regressions and forest plots. Unless otherwise noted, continuous variables were scaled and centered prior to regression. For propensity matching analysis, patient treatment, demographic, and clinical outcome data were extracted from the Surveillance, Epidemiology, and End Results Program (SEER), a US-based registry of disease outcomes, using SEER*STAT v8.3.6. For comparison with LUD2015-005 inoperable OAC patients, only patients with stage III or IV OAC treated with chemotherapy were included. Propensity matching between LUD2015-005 and this SEER subset was performed using MatchIt v4.2.0 with optimal matching and a 1:4 ratio.

2.9 Figure design

Figures in this thesis were largely generated in R using ggplot2 graphics (v3.3.3). Schematic and mechanistic diagrams, including Figures 1.1, 1.2, 4.3, 5.2, and 5.62, were prepared using BioRender design tools (BioRender.com).

Chapter 3: Clinical and Safety Outcomes of the LUD2015-005 Trial

3.1 Introduction

As discussed in the general introduction, the prognosis for patients with inoperable OAC is poor under conventional palliative therapy, with expected survival of less than one year on commonly used first-line chemotherapy regimens such as CapOx (Table 1.1). Therefore, there is a strong need for additional treatment options to improve the prognosis for patients in this setting. Recent phase III studies have shown that the addition of ICI agents to standard first-line chemotherapy regimens does improve patient outcomes in gastro-oesophageal cancers[73–76]. While an exciting development, patients and their physicians would need to weigh the expected benefit against potential drawbacks to determine the utility of this combination treatment for their individual case. This decision-making would be best informed by a robust set of predictive factors that could help identify patients most likely to derive additional benefit from the addition of ICI. To date, no study specifically focused on finding biomarkers predictive of immunochemotherapy outcome has yet been conducted in this setting.

In addition to predictive factors, dynamic on-treatment markers that are associated with outcome are also of interest to patients and their physicians, both for prognostic information and to identify patients for whom the current therapy is not working, who may benefit from additional interventions. The change in tumour size, assessed using immunotherapy-specific radiological response criteria such as irRECIST[88], is one such dynamic marker. However, other treatment-emergent markers that are able to predict resistance before it is detected radiologically could be of additional utility, to mark patients for additional screening or potentially earlier intervention with second-line therapy. One such dynamic marker could be the appearance of on-therapy immune-related adverse events, which have been correlated with improved outcome on immunotherapy in other settings, including ICI alone for gastro-oesophageal cancers that have progressed on prior therapies[106–110]. However, the utility of this on-treatment marker of ICI response has not yet been assessed for first-line immunochemotherapy combinations in OAC.

To date, only αPD-1-based immunochemotherapy has regulatory approval for the first-line setting in inoperable OAC. While these agents bring exciting and statistically significant benefit for a disease with otherwise poor prognosis, the effect size appears smaller than for some other highly ICI-sensitive cancers, such as melanoma. This raises the question of whether combination ICI could potentially further improve outcomes, as has been shown in other settings[62]. Clinical and molecular profiling of an inoperable OAC cohort treated with combination ICI could shed light on whether a combination ICI strategy could help further improve prognosis in this setting, in addition to evaluating potential predictive biomarkers

This work seeks to address these key research questions using a detailed study of clinical and molecular data from the LUD2015-005 clinical trial[111]. The LUD2015-005 trial is a phase I/II study of αPD-L1 or αPD-L1/αCTLA-4 ICI in addition to standard of care chemotherapy for patients with oesophageal cancers. In this introductory results chapter, I first describe the structure of the LUD2015-005 trial and the clinical outcomes for patients with inoperable OAC on this immunochemotherapy treatment. As a basis for the further biomarker studies in later chapters, I also explore whether any patient demographic or clinical presentation characteristics are potential confounders of clinical outcome in this cohort. Finally, aspects of the safety profile of this regimen and the potential utility of using adverse events as an on-treatment marker of immunochemotherapy outcome in this setting are also assessed.

3.2 LUD2015-005 inoperable OAC cohort: Study design and sampling protocol[b]

The LUD2015-005 clinical trial enrolled two major cohorts: an operable cohort receiving neoadjuvant immunochemotherapy prior to surgery, and an inoperable cohort receiving palliative immunochemotherapy[111]. The inoperable cohort is the focus of this thesis, while the operable cohort will be studied in future work from our group.

Within the inoperable cohort of the LUD2015-005 clinical trial, 38 patients with previously untreated inoperable oesophageal or gastro-oesophageal cancers were enrolled. Patients were

[b]Design and implementation of this trial involved key contributions from many individuals, including Mark Middleton, Xin Lu, Richard Owen, Michael White, Simon Lord, Ioannis Karydis, and many clinical and research collaborators (see Acknowledgements). The Oncology Clinical Trials Office at the University of Oxford (OCTO) and Ludwig Cancer Research (LCR) provided key oversight of study procedures and data management and processing. AstraZeneca UK Limited provided the investigational drugs for this trial.

allocated into three stages according to a dose-escalation strategy to ensure the safety of the addition of the ICI agents used in this study to chemotherapy in this setting (see Methods). In the first stage, patients received αPD-L1 agent (durvalumab) alone as ICI. Following establishment of an acceptable safety profile, patients in the last two stages received αPD-L1 in combination with a single priming dose of the αCTLA-4 agent tremelimumab at one of two dosing levels (see Methods). To allow for the characterisation of response to ICI without the potentially confounding effect of chemotherapy, all patients began treatment with a four-week window where only αPD-L1 or αPD-L1/αCTLA-4 ICI was given (ICI-only). Following this 4-week ICI-only period, patients then received a standard of care chemotherapy regimen (CapOx) alongside continued αPD-L1. Endoscopic biopsies of tumour and control healthy GI tissues were taken prior to treatment onset, at the end of the 4-week ICI-only period, and at the end of the immunochemotherapy phase to profile potential predictive biomarkers. CT scans were taken throughout the treatment protocol to monitor response to therapy (see Figure 3.1 and Methods).

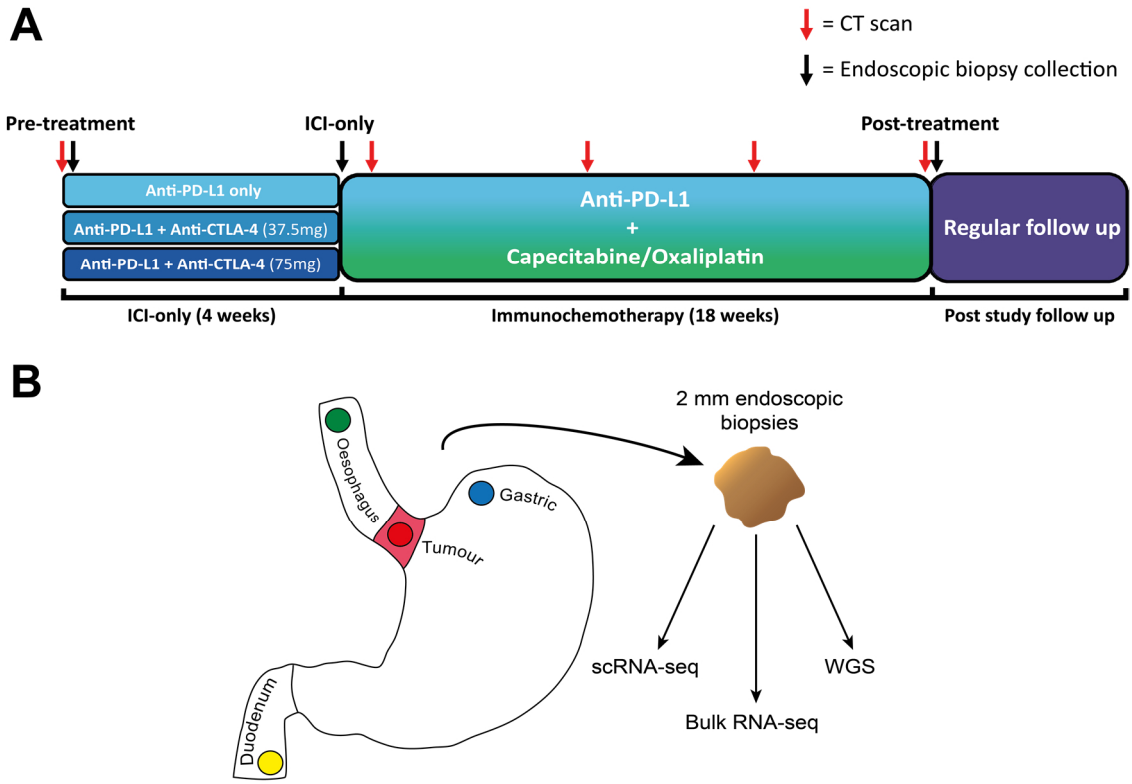


Figure 3.1: Study design and sampling protocol for the LUD2015-005 inoperable cancer cohorts. A) Timeline showing the relative chronology of different treatment, sampling, and CT response assessment events during the LUD2015-005 trial. B) Location of endoscopic sampling sites for biomarker analyses. Oesophageal and gastric biopsies were taken at least 2cm away from the tumour and gastro-oesophageal junction. This figure was provided courtesy of Joseph Chadwick.

3.3 Clinical characteristics[c]

Of the 38 patients with inoperable oesophageal cancer in the LUD2015-005 trial, the vast majority had OAC (n = 35), while a small number had OSCC (n = 3). Given the significant molecular differences between adenocarcinomas and squamous cell carcinomas[38,112], only the 35 OAC patients were selected for detailed analysis in this work, to provide a more homogenous cohort for biomarker discovery. The characteristics of the 35 patients comprising this inoperable OAC cohort are given in Table 3.1. Following the non-randomised dose-escalation format of this study, the first 11 patients with inoperable OAC were treated with αPD-L1 alone for the ICI component of the trial, followed by 5 patients receiving αPD-L1 plus a single lower dose of αCTLA-4, and finally 19 patients receiving αPD-L1 plus a single higher dose of αCTLA-4. αPD-L1-only patients, enrolled during the phase I component of the trial, all had an ECOG performance status of 0 (fully active), while αPD-L1/αCTLA-4 patients were more evenly split between ECOG grades 0 and 1 (ambulatory but restricted in physically strenuous activity). Otherwise, demographic and clinical characteristics of treatment subgroups were similar, each having predominately male patients with stage IV cancer.

[c]Collection and formatting of these data for this analysis was conducted by OCTO and LCR.

ICI Agent	αPD-L1 (n=11)	αPD-L1+ 37.5mg αCTLA-4 (n=5)	αPD-L1+ 75mg αCTLA-4 (n=19)	Total (n=35)
Age at consent- years				
Median (range)	59 (23-75)	55 (29-67)	59 (48-78)	59 (23-78)
Sex- no. (%)				
Male	10 (90.9%)	4 (80%)	15 (78.9%)	29 (82.9%)
Female	1 (9.1%)	1 (20%)	4 (21.1%)	6 (17.1%)
Ethnicity- no. (%)				
White, British	9 (81.8%)	5 (100%)	19 (100%)	33 (94.3%)
White, Any other	2 (18.2%)	0	0	2 (5.7%)
Stage at study entry- no. (%)				
IIIA	0	0	1 (5.3%)	1 (2.9%)
IIIB	1 (9.1%)	0	0	1 (2.9%)
IIIC	0	0	1 (5.3%)	1 (2.9%)
IV	10 (90.9%)	5 (100%)	17 (89.5%)	32 (91.4%)
ECOG Status- no. (%)				
0	11 (100%)	3 (60%)	11 (57.9%)	25 (71.4%)
1	0	2 (40%)	8 (42.1%)	10 (28.6%)

Table 3.1: Clinical characteristics of the LUD2015-005 inoperable OAC cohort. The characteristics at enrolment of the 35 patients with inoperable OAC from the LUD2015-005 study are presented, categorized by the ICI agent used. ECOG refers to the patient performance status: 0 is fully active, while 1 is ambulatory but restricted in physically strenuous activity.

3.4 Survival outcomes from the LUD2015-005 trial

3.4.1 Cohort summary metrics

To evaluate patient outcomes, survival outcomes were first assessed (Figure 3.2). Overall, patients in the inoperable OAC cohort had a median OS of 11.9 months (median OS follow-up of 27.4 months), and a median PFS of 9.3 months (median PFS follow-up of 10.5 months). 10 patients remained alive at the time of the data cut-off used for analysis, including 3 patients with ongoing OS longer than three years. Treatment arm-specific survival outcomes are summarized along with response outcomes later in this chapter (Table 3.2). This trial was not designed or powered to test for significant differences between the survival outcomes of these treatment arms, particularly given the variation in performance status between groups. Therefore, while the group receiving αPD-L1 with the highest dose of αCTLA-4 had the longest OS (15.6 months) and PFS (11.5 months) of the three treatment arms (Table 3.2), no statistical test for significant differences in survival outcomes between treatment arms was conducted in this thesis, and these findings must be interpreted with caution.

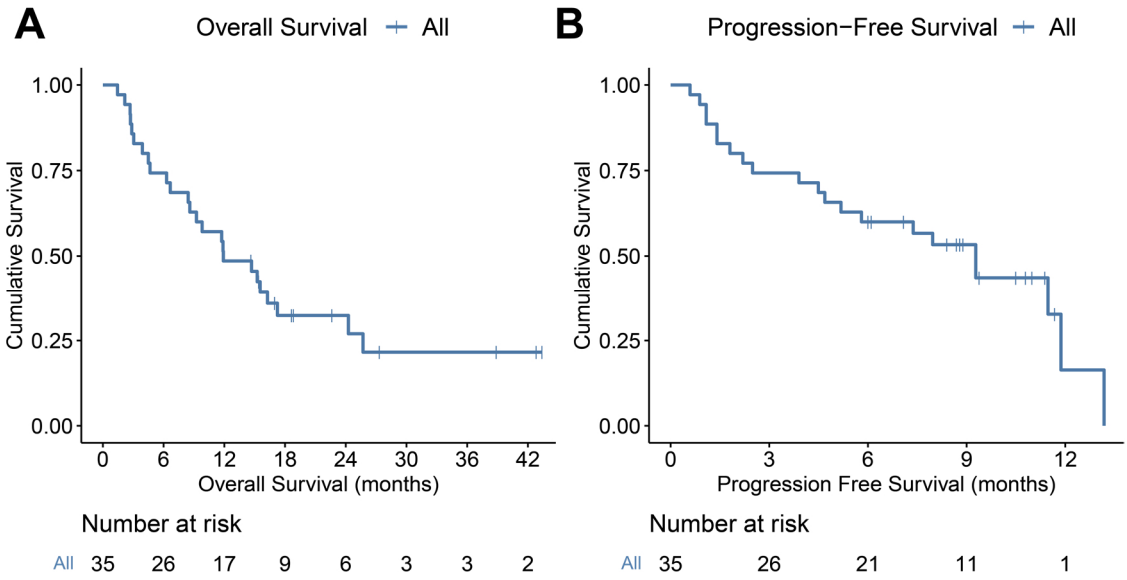


Figure 3.2: Survival outcomes of the LUD2015-005 inoperable OAC cohort. Kaplan-Meier curves for A) OS and B) PFS for all inoperable OAC patients in the LUD2015-005 trial. Dashes represent right-censored patient outcomes: either ongoing survival at the time of last follow-up, or if alternative therapy was initiated prior to progression (for PFS). Risk table at bottom shows number of patients at each timepoint with ongoing survival.

3.4.2 Propensity-matched comparison with a chemotherapy-treated patient registry

The LUD2015-005 trial is a phase I/II study, and therefore did not include a chemotherapy-only comparator. However, the median OS and PFS appeared to be slightly higher than what has been reported in the literature for similar cohorts treated with CapOx (Table 1.1), suggesting the addition of ICI may have had some impact. To further explore this, a retrospective analysis was conducted comparing outcomes from this study to those from OAC patients reported in SEER, a US-based registry of outcomes from clinical practice. To attain better comparability with the LUD2015-005 cohort, only white OAC patients treated with chemotherapy, but not radiotherapy, since 2010 were included from SEER. To further reduce potential inter-cohort bias, propensity-matching was conducted on the basis of age, sex, and stage. Survival outcomes were then compared between these propensity-matched cohorts (Figure 3.3). In this analysis, OS was relatively similar between the two cohorts at 1 year (LUD2015-005 48.6% [95% CI: 34.5%-68.3%]; SEER 41.9% [95% CI: 33.7-52.1%]). However, at later timepoints, more difference between the survival curves became visible, with the 3-year OS estimate for the ICI-cohort falling outside of the 95% CI from the SEER cohort (LUD2015-005 21.6% [95% CI: 10.2%-45.8%]; SEER 6.84% [95% CI: 2.64%-17.7%]). Caution must be exercised in interpreting

this analysis, as patients from trial cohorts like the LUD2015-005 study are more carefully selected for enrolment characteristics that might result in improved survival than in registries like SEER, no information was available as to the exact chemotherapy regimen used for SEER patients, and there were only a handful of patients had ongoing follow-up in both cohorts at the 3-year timepoint. Additionally, the overall difference between the two survival curves was not significant, although the limited size of the LUD2015-005 cohort limits the statistical power to detect moderate differences in this analysis. Nevertheless, these data suggest that the potential survival differences caused by the addition of this ICI strategy to standard of care chemotherapy in this setting may be more visible at long-term endpoints.

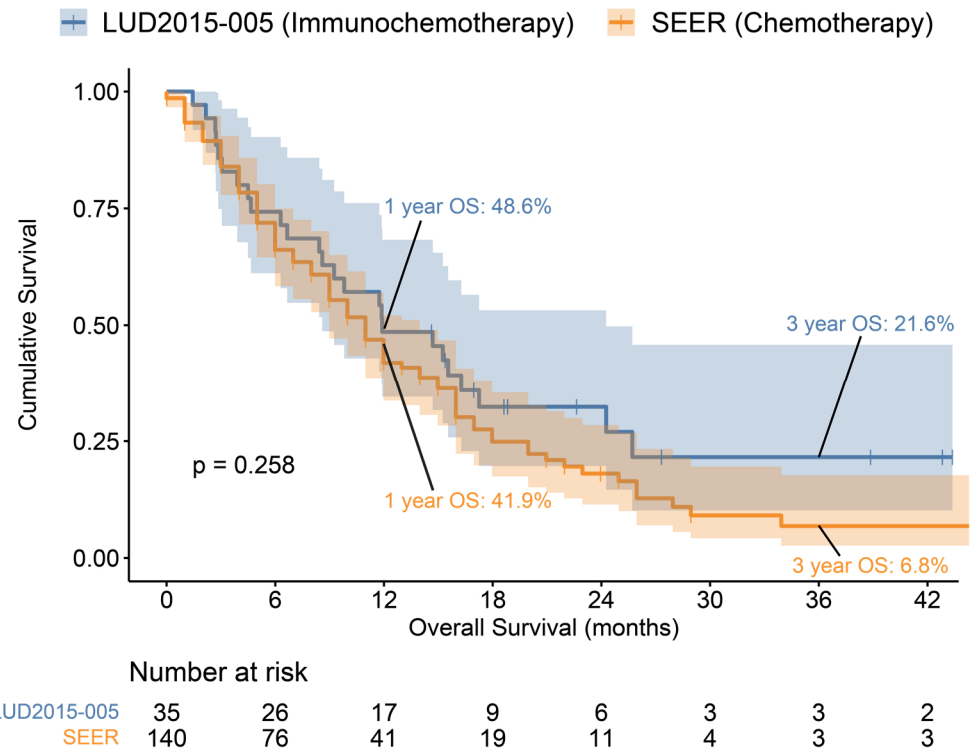


Figure 3.3: Propensity-matched cohort analysis between SEER and LUD2015-005 inoperable OAC patients. White SEER patients with stage III or IV OAC with reported chemotherapy treatment since 2010 were selected for comparison. Propensity matching was conducted with optimal matching using a 1:4 ratio (see Methods), although similar outcomes were seen with matching ratios of 3-10. The shaded region shows the 95% CI for each curve. Displayed p value is calculated by the log-rank test, testing the overall difference between the two curves. Listed OS percentages at fixed timepoints were extracted from the generated survival curve model using the survfit function from the survival package.

3.5 Response and clinical benefit outcomes from the LUD2015-005 trial

In addition to survival, response outcomes, which summarize the growth or shrinkage of the tumour during treatment, can give important information as to the direct impact of therapy on the tumour. Briefly, changes to tumour size were classified using irRECIST response criteria[88] into irCR, irPR, irSD, or irPD categories (see Methods). The best overall response (BOR) attained by each patient is summarized in Figure 3.4. Of the 35 patients in the inoperable OAC cohort, 17 (49%) achieved an irRECIST response during the study period. Of these, 15 were partial responses (irPR), and 2 were complete (irCR; complete disappearance of all lesions). Three patients (1 irCR, 2 irSD) had no measurable target lesions at the pre-treatment CT scan, and therefore are not shown in target lesion response plots (Figures 3.4 and 3.5). Three responses had already occurred by the first CT scan, which occurred just after the conclusion of the 4-week ICI-only window and only about one week into immunochemotherapy, suggesting a response primarily mediated by the ICI component of treatment. The most common timepoint to first detect a response was during the 3rd cycle of chemotherapy (n = 11), while a small number were not detected until the 5th or 6th cycles of chemotherapy (n = 1 and n = 2 respectively).

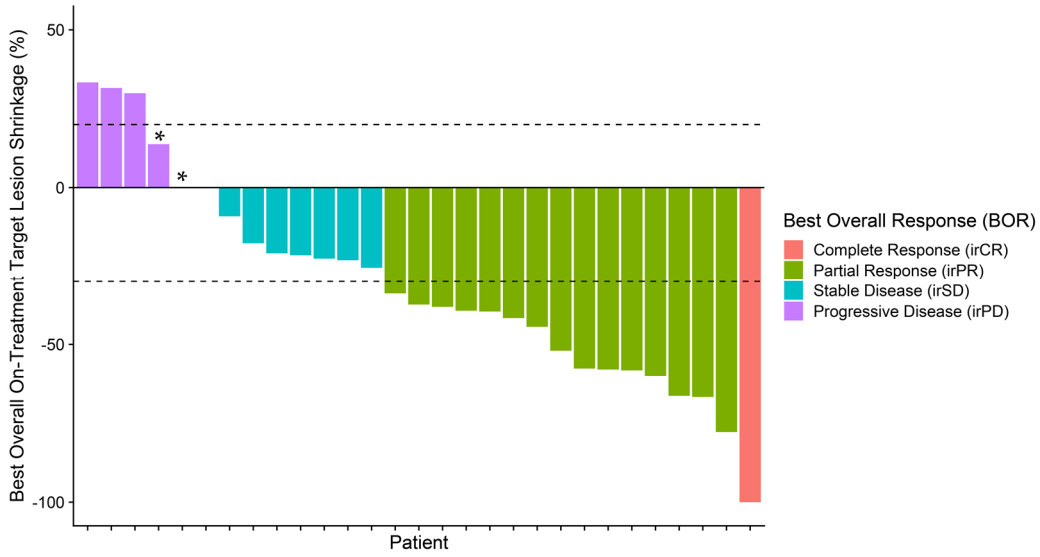


Figure 3.4: Best overall responses (BOR) the LUD2015-005 inoperable OAC cohort. The maximal shrinkage of target lesions (sum of diameters; not including new or non-measurable lesions) attained by each patient at any CT scan conducted during the study period is shown. Three patients (1 irCR, 2 SD) are not shown due to absence of measurable target lesions. Three additional patients who passed away from clear clinical progression prior to any on-treatment CT scan (clinical PD) are also not shown. Asterisks mark patients with irSD of target lesions but unequivocal progression due to new lesions (overall irRECIST response of irPD).

While irRECIST response can be a useful categorical variable for biomarker discovery, some challenges arise especially in its use as a summary metric for this cohort. First, some patients displayed vastly different responses during the different phases of therapy. Of the patients in this cohort that initially had unconfirmed progressions during ICI-only, some resolved to stable disease or even partial response by the end of the immunochemotherapy period (Figure 3.5). Using BOR allows for the summarization of these varied outcomes with a single metric, but this too showed limitations in classifying some patient outcomes. For instance, the two patients achieving irCR in this cohort showed strongly divergent long-term outcomes, with one progressing at 9.3 months and passing away within one year of treatment onset, while the other has survived for over 3.5 years. Similarly, four of ten patients with a BOR of irSD achieved quite good clinical outcomes (more than 1 year of survival without reported progression), despite never recording an irRECIST response.

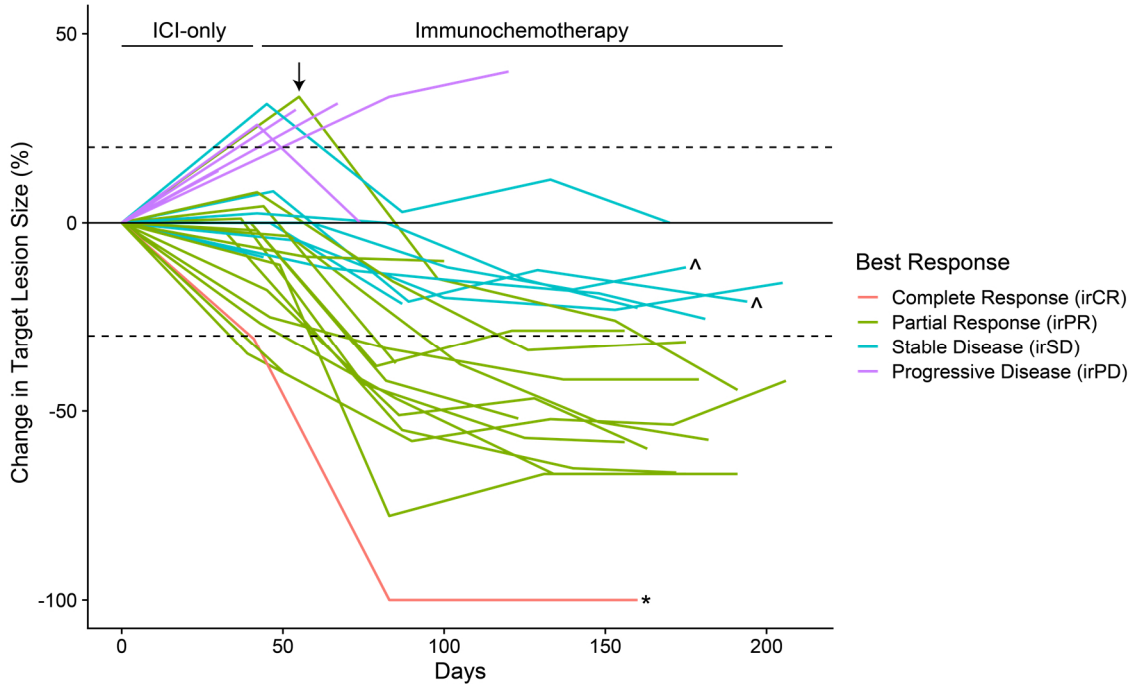


Figure 3.5: Spider plot of target lesion size over the course of treatment. The change in target lesion size (sum of diameters) over time is displayed for each patient for all CT scans during the study period. Patients are highlighted to show the difficulty in summarizing outcome using response: ↓ marks a patient whose lesions grew during the ICI-only period, but eventually attained irPR during immunochemotherapy; * marks a patient with irCR who had average PFS (progressed at 9.3 months); ^ marks 2 patients with PFS > 12 months who did not attain an irRECIST response (two additional cases were non-assessable for target lesion size).

Given the occasional disconnect between irRECIST response and long-term survival outcome in this cohort, and following the practice of other ICI biomarker studies[89,113], an additional clinical benefit classification system was defined retrospectively. Patients were deemed to have attained clinical benefit (CB) in this cohort if they attained 12 months of survival without confirmed irRECIST progression, and to have attained no clinical benefit (NCB) if they progressed prior to this endpoint. The 12 month landmark endpoint is roughly double the reported median PFS for platinum/fluoropyrimidine-based chemotherapy in this setting[47], and therefore patients attaining CB are likely to be enriched for those who benefit the most from the addition of ICI to standard of care chemotherapy. Importantly, this metric is agnostic of whether patients survive with an irRECIST response or long-term maintenance of stable disease, which occurs in a number of patients in this cohort. As expected from the incorporation of a survival landmark in the definition of this metric, CB is more closely linked with OS than irRECIST in this cohort (Figure 3.6). 14/34 assessable patients (41%) in the inoperable OAC cohort met the definition of CB, with one patient excluded as they went on to alternative therapy prior to the 12-month threshold. In summary, CB is a retrospectively assigned binary classification system that combines aspects of survival and irRECIST response, prioritizing long-term disease control regardless of whether or not the threshold for an irRECIST response is met. A summary of key clinical outcome metrics for the full inoperable OAC cohort of the LUD2015-005 trial, as well as for each specific treatment subgroup, is provided in Table 3.2.

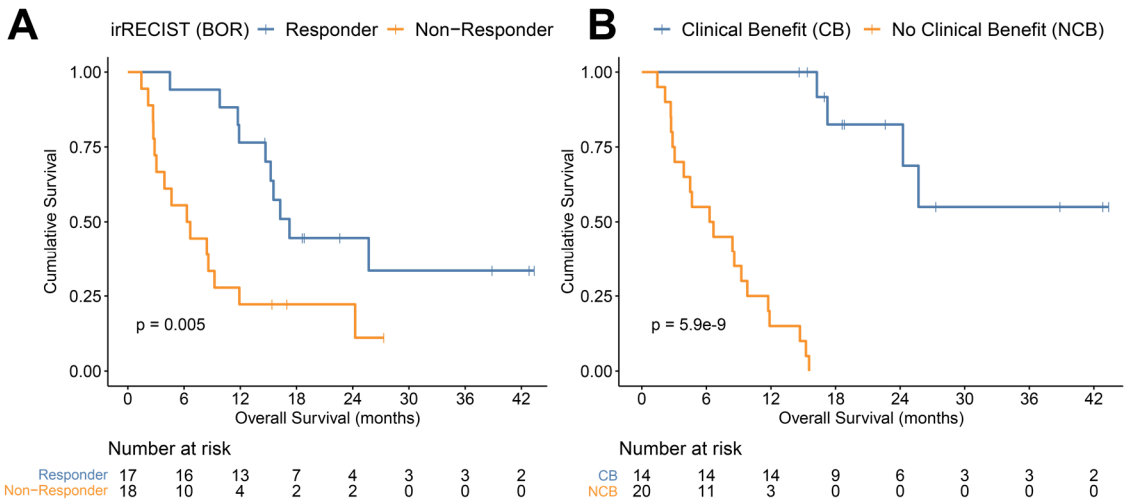


Figure 3.6: Comparison of binary outcome classification systems and their link with long-term overall survival. Kaplan-Meier curves are shown with groupings by A) irRECIST BOR and B) clinical benefit. Displayed p-values were calculated using the log-rank method. One patient not assessable for CB was excluded from B.

ICI Agent	αPD-L1 only (n=11)	αPD-L1+ 37.5mg αCTLA-4 (n=5)	αPD-L1+ 75mg αCTLA-4 (n=19)	Total (n=35)
irRECIST BOR				
Response (irCR/irPR)	5 (45.5%)	3 (60%)	9 (47.4%)	17 (48.6%)
Non-Response (irSD/irPD*)	6 (54.5%)	2 (40%)	10 (52.6%)	18 (51.4%)
Clinical Benefit (>12 months PFS)				
Yes (CB)	5 (45.5%)	0	9 (47.4%)	14 (40%)
No (NCB)	6 (54.5%)	5 (100%)	9 (47.4%)	20 (57.1%)
Not assessable	0	0	1 (5.3%)	1 (2.9%)
Overall survival (months)				
Median (95% CI)	11.8 (6.7-NA)	8.6 (4.5-NA)	15.6 (11.9-NA)	11.9 (8.6-24.3)
Progression-free survival (months)				
Median (95% CI)	9.3 (2.5-NA)	5.2 (4.5-NA)	11.5 (9.3-NA)	9.3 (5.2-NA)

Table 3.2: Summary of clinical outcomes of the LUD2015-005 inoperable OAC cohort. irRECIST BOR and CB outcomes and survival summary statistics are presented for each ICI treatment subgroup. For survival outcomes, NA means that the upper end of the confidence interval was not yet reached due to ongoing patient survival. *irSD represents BOR of irSD or irNN, and irPD represents a BOR of irPD or clinical PD (clear clinical progression with no on-study CT scans; see Methods).

3.6 Assessing potential confounders of biomarker discovery

The primary aim of using LUD2015-005 trial data in this thesis is for discovery of potential biomarkers that could predict outcomes on immunochemotherapy. To evaluate whether variation in clinical presentation, demographics, and treatment strategy could potentially confound biomarker analyses, these variables were investigated to assess whether they could

have a strong link with survival outcomes that could interfere with predictive biomarker signal. The sex, stage, age, and performance status of the patient were included in this analysis, as variables that have previously been linked with either general or ICI-specific prognosis[114–117]. As the addition of αCTLA-4 impacts survival outcomes in other settings[62], the dose of αCTLA-4 was also assessed. With the exception of age, no variable in this analysis was significantly linked with survival outcomes either in univariate or multivariate approaches (Table 3.3). Older age was associated with improved PFS in a univariate approach; however, this effect did not reach significance in the multivariate approach or in the OS analysis, which suggests the effect may not be robustly linked with outcomes in this cohort. Therefore, while the small size of the assessed cohort limited the power of this analysis, these characteristics were judged to at least not have a strong impact on survival signal for this cohort, and were thus deemed likely to not be major potential confounders for the biomarker discovery analysis conducted later in this work.

Variable	OS				PFS			
	Univariable		Multivariable		Univariable		Multivariable	
	HR (95% CI)	p.value	HR (95% CI)	p.value	HR (95% CI)	p.value	HR (95% CI)	p.value
Tremelimumab dose (reference: 0 mg, n=11)								
37.5 mg (n=5)	2.84 (0.85-9.47)	0.089	1.84 (0.49-6.88)	0.366	2.67 (0.77-9.28)	0.121	1.41 (0.34-5.93)	0.637
75 mg (n=19)	1.00 (0.39-2.54)	0.993	0.63 (0.20-2.06)	0.448	0.87 (0.31-2.45)	0.788	0.59 (0.16-2.15)	0.421
Sex (reference: Female, n=6)								
Male (n=29)	0.47 (0.19-1.18)	0.108	0.61 (0.22-1.69)	0.339	0.74 (0.24-2.29)	0.599	0.97 (0.29-3.19)	0.954
Stage (reference: III, n=3)								
IV (n=32)	1.51 (0.35-6.44)	0.578	1.74 (0.37-8.16)	0.484	1.94 (0.26-14.69)	0.521	2.45 (0.30-20.09)	0.404
ECOG (reference: 0, n=25)								
1 (n=10)	1.71 (0.74-3.96)	0.211	2.21 (0.72-6.79)	0.167	1.41 (0.56-3.57)	0.466	2.13 (0.66-6.89)	0.206
Age (continuous)	0.14 (0.02-1.12)	0.064	0.18 (0.02-1.15)	0.117	**0.10 (0.01-0.97)**	**0.047**	0.16 (0.02-1.67)	0.125

Table 3.3: Impact of ICI treatment subgroup, demographic, and clinical presentation characteristics on survival. Each listed variable was included in a univariable Cox proportional hazards model to predict OS and PFS, as well as a multivariable model to predict the same using all listed variables as predictors. Continuous variables were scaled prior to regression. Hazard ratio (HR), 95% confidence interval (95% CI), and unadjusted p-value for the given regression are listed for each variable, with significant entries ($p < 0.05$) in bold.

3.7 Adverse events and their association with patient outcomes[d]

As a component of this phase I/II study, treatment-emergent adverse events (AEs) were recorded and graded using Common Terminology Criteria for Adverse Events (CTCAE; v4.03)[118]. Study investigators classified each AE as immune-related (irAE) or not (non-irAE), and assessed the probability of relationship between the AE and each of the ICI and chemotherapy agents. AEs determined to be unrelated to all treatment agents were excluded from this analysis. Generally, treatment was tolerated acceptably in the inoperable OAC cohort, with most reported AEs being mild or moderate (CTCAE Grade 1-2, 92%). 94% of patients (33/35) reported at least one AE, with the most common categories being nausea/vomiting (71.4%), neuropathies[e] (65.7%), and fatigue/weakness (62.9%). Higher-grade AEs (CTCAE Grade 3-5) were reported in 57% (20/35) of patients, the most frequent of which were elevated lipase (11.4%), neutropenia (8.6%), and diarrhoea (8.6%). While no statistical test was conducted due to the limited size of some arms in this cohort, the incidence of higher-grade AEs did not appear to be increased by the addition of αCTLA-4 (Table 3.4).

ICI Agent	αPD-L1 only (n=11)	αPD-L1/low-dose αCTLA-4 (n=5)	αPD-L1/high-dose αCTLA-4 (n=19)	Total (n = 35)
All AEs	10/11 (91%)	5/5 (100%)	18/19 (95%)	33/35 (94%)
Higher-grade AEs	8/11 (73%)	2/5 (40%)	10/19 (53%)	20/35 (57%)
All irAEs	6/11 (55%)	4/5 (80%)	13/19 (68%)	23/35 (57%)
Higher-grade irAEs	3/11 (27%)	1/5 (20%)	5/19 (26%)	9/35 (26%)

Table 3.4: Summary overview of treatment-emergent adverse events in the inoperable OAC cohort. Rates of occurrence for each of the AE classifications, excluding AEs deemed by investigators to have no relation to all treatment components. Rates are shown separately for all three treatment subgroups, as well as combined across the full cohort (Total). Higher-grade AEs were CTCAE grades 3, 4, and 5.

[d]Data on AEs were collected from study clinical staff by OCTO and processed by LCR for this analysis.

[e]The neuropathies category was collapsed from AE codes for neuropathy (peripheral or otherwise unspecified), "pins and needles", paraesthesia, and hot/cold hypersensitivity/dysaesthesia, with neuropathy being the most frequent (31 of 41 events). Similar results are seen when limiting to just the neuropathy codes (peripheral and otherwise unspecified).

irAEs were reported in 63% (22/35) of the inoperable OAC cohort, and 26% reported higher-grade irAEs (9/35). The prevalence of higher-grade irAEs also did not appear to be higher in the combination ICI arms of this cohort (Table 3.4). Interestingly, across the inoperable OAC cohort, patients with any reported irAE were found to have significantly better OS (Fig 3.6A). To assess whether this effect was independent of the total dosage received by each patient, the analysis was repeated on only the patients who received the full number of durvalumab infusions (total dosage of 8250mg). A similar proportion of patients in this full treatment subset experienced irAEs to the full cohort (12/17, 71%), and irAEs were still significantly associated with improved OS (Fig 3.6B). In summary, development of irAEs could indeed serve as a useful on-treatment predictive marker in this setting, and did not appear to be confounded by cumulative dosage of ICI.

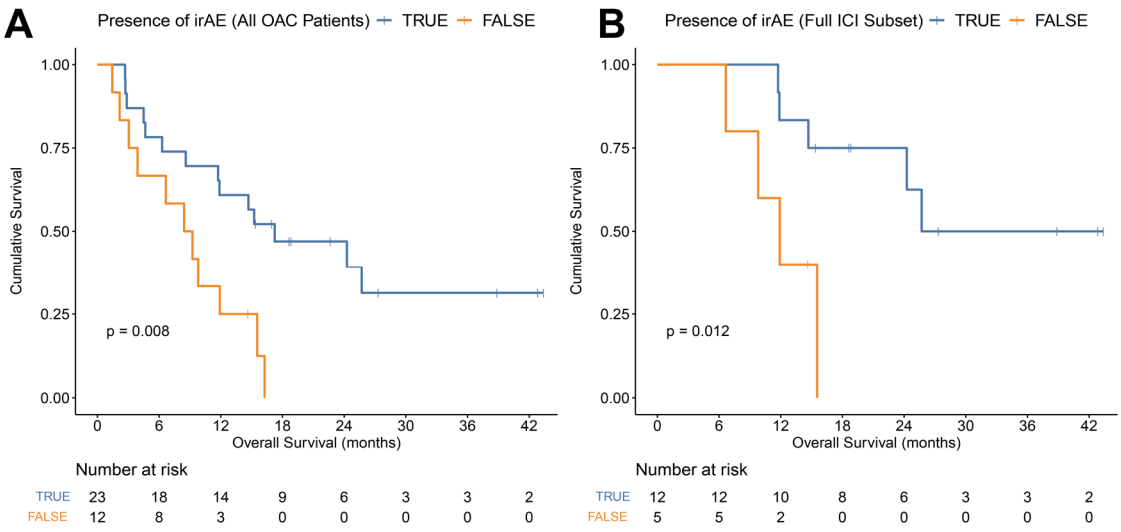


Figure 3.7: Development of irAEs is a marker of improved overall survival on immunochemotherapy. A) Kaplan-Meier curve showing OS for the full inoperable OAC cohort with grouping based on whether the patient experienced a treatment-related irAE. B) Same as in A, except only patients receiving the full 8250mg dosage of durvalumab were included (full ICI subset). All p values were calculated by the log-rank test.

As a control, non-irAEs were also interrogated for any link with OS. As nearly all patients reported at least one non-irAE, individual AE categories were assessed instead. Of the most frequent non-irAE categories, nausea/vomiting and fatigue/weakness were not linked with OS across the inoperable OAC cohort ($p = 0.547$ and $p = 0.865$, respectively). However, the development of neuropathies did show an association with survival outcomes. 41 AEs in the

neuropathy category were reported in 23 patients, the vast majority of which were mild (36/41 CTCAE Grade 1, 4/41 Grade 2, 1/41 Grade 3). All neuropathies were investigator-assessed to be probably or likely linked to oxaliplatin, and indeed no events in this category occurred prior to oxaliplatin initiation. Median OS for those who reported neuropathy was 24.3 months, compared to just 4.2 months OS for those who did not. However, the magnitude of this survival difference was likely driven by the cumulative neurotoxicity which is known to accumulate with repeated oxaliplatin treatment[119], which would likely not affect patients who did not survive long enough to receive multiple doses to the same extent. Indeed, the prevalence of neuropathy was much higher in those surviving long enough to receive the full six doses of oxaliplatin (19/21, 91%) compared with those that did not (4/14, 29%).

To determine whether there was a link with this AE and OS that was independent of the duration of oxaliplatin treatment, restriction to the full-treatment subset could not be employed, as nearly all patients receiving six doses of oxaliplatin experienced neuropathy. Therefore, the timing of neuropathy onset after the first dose was employed instead, which allowed the same analysis to be applied to all oxaliplatin-treated patients (n = 32). As the second dose of oxaliplatin was given as early as 17 days following the first (range 17-69 days), a window of 16 days following initiation of oxaliplatin was applied across all patients to mark early-onset neuropathies associated with the first oxaliplatin dose. Interestingly, oxaliplatin-treated patients who experienced neuropathy in this window following the first dose had significantly better OS than those that did not (Figure 3.7A). Furthermore, the median number of days after the first oxaliplatin dose until neuropathy was first reported was significantly less in patients who would go on to attain clinical benefit (CB: 4.5 days; NCB: not reached), although there was no significant link with irRECIST response (Figure 3.7B-C). Therefore, the early development of oxaliplatin-related neuropathies is associated with better survival outcomes for this OAC cohort, even after controlling for potential bias due to a cumulative toxicity effect. This unexpected finding suggests that, in addition to irAEs, some non-irAEs may also have utility in assessing patient outcomes on immunochemotherapy.

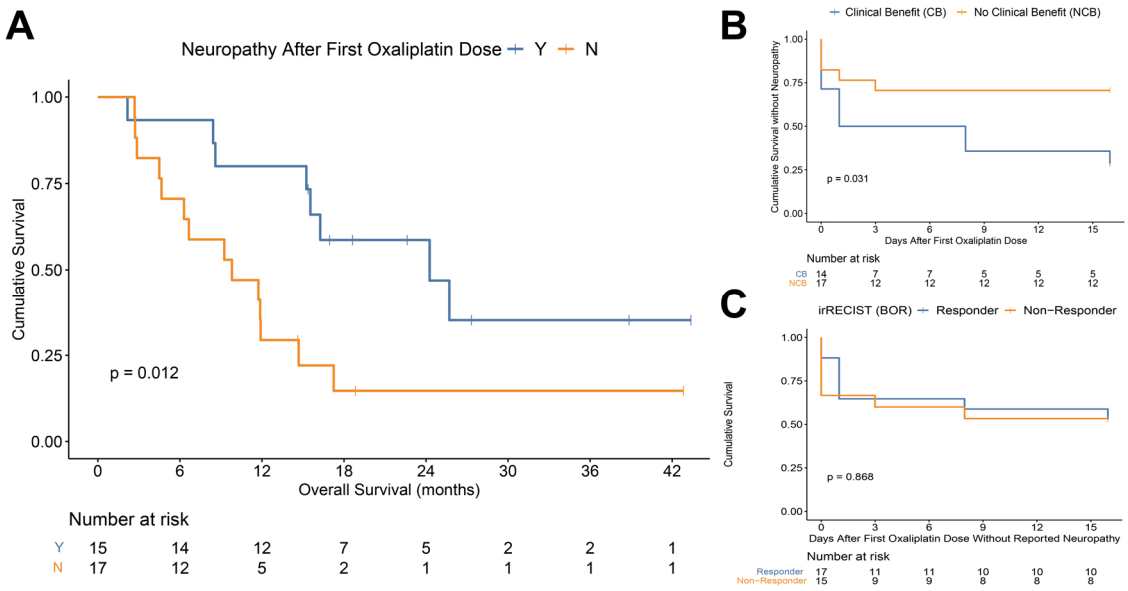


Figure 3.8: Development of neuropathy after the first oxaliplatin dose is associated with better overall survival, but not irRECIST response, on immunochemotherapy. A) Kaplan-Meier curve showing OS (from ICI treatment onset) for patients that do or do not develop neuropathies within 16 days of their first oxaliplatin dose. Neuropathies were defined as the presence of neuropathy (peripheral or otherwise unspecified), "pins and needles", paraesthesia, and hot/cold hypersensitivity/dysaesthesia (similar results also seen when limiting classification to only peripheral or otherwise unspecified neuropathy). Three patients that did not receive any oxaliplatin doses were not included in this analysis. The right facet shows "neuropathy-free survival" during the 16 days after the first oxaliplatin dose; as no mortality was reported in this window, this plot directly shows the risk of developing neuropathy over this period for each group. All patients were included who received oxaliplatin and were assessable for clinical benefit (B) and irRECIST BOR (C). Neuropathy-free survival was censored at 16 days if no neuropathy was reported by that time to avoid bias from the timing of the second oxaliplatin dose. For A-C, p values were calculated by the log-rank test.

3.8 Summary and brief discussion

In this chapter, I summarize clinical outcomes from a cohort of inoperable OAC patients treated with first-line immunochemotherapy as a part of the LUD2015-005 trial. While this analysis cohort is comparatively small (n=35), the median OS of 11.9 months is similar to outcomes from the immunochemotherapy arm of a recently reported phase III trial in inoperable cancers of the oesophagus and gastro-oesophageal junction, which reported 12.4 months median OS[75]. The LUD2015-005 treatment arm receiving αPD-L1 in combination with the highest dose of αCTLA-4 showed the longest survival outcomes (median OS: 15.6; median PFS: 11.5 months) in this cohort. However, no statistical comparison between the different treatment arms was conducted, as this was not a part of the prospective study design, and the differences in performance status between arms is likely particularly confounding, given that the variable is

often strongly linked with patient survival outcomes[117]. Nevertheless, especially as no signal was seen in this analysis that this dosage of αPD-L1/αCTLA-4 combination ICI resulted in increased incidence of higher-grade AEs, analysis in a larger cohort is perhaps warranted to confirm whether combination ICI does improve outcomes.

A propensity-matched analysis with a public registry of clinical outcomes (SEER) suggested that the benefit of adding this ICI approach to standard of care chemotherapy could be most visible at long-term endpoints in this setting, while less effect was seen over the first year of treatment. While caution needs to be exercised with interpreting propensity-matched analyses from this small cohort, this is reminiscent of the "tail-of-the-curve" survival benefit that has been shown in many ICI studies, where survival of ICI-treated patients tends to plateau at late timepoints more so than standard of care comparators. It is thought that this long-term survival plateau is due to the generation of robust and durable anti-tumour adaptive immunity in a proportion of patients, which may be more likely under ICI than conventional therapies[120]. Analysis of long-term outcomes from the recently reported phase III trial data for immunochemotherapy in OAC would be better suited to confirming the extent of any tail-of-the-curve benefit that may be occurring in this setting.

For any cohort used for biomarker discovery, it is important to ensure that biomarker analysis is not confounded by particular patient enrolment or treatment characteristics, or at least to be aware of potential confounders that need to be controlled for in downstream analysis. In this cohort, an analysis of treatment subgroups, patient demographics, and clinical presentation characteristics determined that none of these factors had a strong impact on clinical outcomes, suggesting that it would be reasonable to treat this entire cohort as a single group for the purpose of biomarker discovery. Such an approach does not allow for differentiation between αPD-L1 and αPD-L1/αCTLA-4 predictive biomarkers, but the gain in statistical power could greatly assist in identifying general biomarkers of outcome on αPD-L1-based immunochemotherapy.

irRECIST tumour responses of this cohort were characterized, including some of the potential limitations in using these response categories for biomarker discovery in this small cohort. To

address some of these limitations, a clinical benefit variable was introduced based on whether patients reached a PFS landmark of 12 months. While this metric is cohort-specific and defined post-hoc, it allows for the use of categorical analytical approaches to discover biomarkers that could predict longer disease control and survival in this setting. This could arguably have more utility than markers that predict pure response, which can be disconnected from long-term disease control and survival. This is perhaps especially true for unique trial designs such as the LUD2015-005 study, where a staggered immunotherapy and chemotherapy onset is employed and patients can show different responses during each treatment phase.

In addition to radiological measures of response, irAEs have been queried as a marker of immunotherapy response in other settings[106–109]. In this cohort, patients with irAEs had better OS, largely agreeing with previous findings and suggesting that mounting of a stronger anti-tumour immune response as a result of ICI can be associated with stronger immune-related reactions. These could be mediated by autoimmunity, or could perhaps be thought of as markers of an immune system more primed to generate general immune responses following ICI. In this way, absence of on-treatment irAEs could potentially be used to identify patients with poor systemic immune responses on immunochemotherapy in this setting, for whom treatment modification or alternative interventions may need to be considered.

Most of these AE studies have been conducted in ICI monotherapy, and the relationship between chemotherapy AEs and outcome in immunochemotherapy settings has not yet been well characterized. Interestingly, a link between increased survival and early development of neuropathies, which are AEs seen particularly with platinum-based chemotherapy agents[119,121], was also noted in this OAC cohort. The potential link between oxaliplatin-induced nerve toxicity and immunochemotherapy outcome does not have an immediately obvious mechanistic explanation. It could be possible that increased bioavailability of oxaliplatin in some patients results in both increased toxicity and a stronger anti-tumour response. This could be mediated through stronger direct anti-tumour activity of the platinum agent, or perhaps also through synergy with ICI, by generating more DNA damage and subsequent neoepitopes that could be recognized by responding T cells. Another possibility is that the propensity to develop

neuropathy following oxaliplatin treatment could be related to the immune status of each patient. Increased systemic neutrophil- and monocyte-to-lymphocyte ratios have been associated with development of taxane-induced neuropathy[122], and increased infiltration and activity of these innate immune cell types also seem to play a role in oxaliplatin-induced neuropathy[123,124]. Additionally, chemotherapy-induced neuropathy can be reduced in laboratory models by blocking inflammatory cytokines such as TNFα and monocyte chemoattractant protein-1 (MCP-1/*CCL2*)[125]. It is therefore possible that early development of neuropathy could be a marker of a certain systemic immune state which is primed for better outcomes on immunochemotherapy. Regardless of the mechanism, the presence of this AE was significantly linked with outcome in this study, and could warrant further validation in a larger cohort. As more data from phase III trials continue to become available from this setting, chemotherapy AEs such as oxaliplatin-induced neuropathy should be assessed alongside irAEs, with particular attention to whether AEs correlate with a particular pre-treatment systemic immune phenotype.

In summary, this chapter validated the use of irAEs, non-irAEs, and radiological response metrics as on-treatment dynamic markers of survival outcomes in this setting. Additionally, no pre-treatment demographic or clinical characteristic was found to be strongly predictive of outcome, suggesting it would be appropriate to consider the entire LUD2015-005 inoperable OAC cohort as a single entity for the purpose of identifying predictive biomarkers for immunochemotherapy.

Chapter 4: Genomic and Transcriptomic Workflows to Evaluate Previously Reported and Novel Predictive Biomarkers

4.1 Introduction

Much work has been done to characterize biomarkers that predict response to ICI in other cancer types. However, as first-line immunochemotherapy has only recently been approved for use in OAC, little is known about how effective these biomarkers will be in this treatment setting. In this chapter, I describe the generation of genomic and transcriptomic datasets from the LUD2015-005 trial cohort, and investigate whether previously reported predictive biomarkers derived from other cancer types are also predictive in this setting. Furthermore, I present my own *de novo* approach to identify predictive transcriptional biomarkers from this dataset and from ICI cohorts across a range of other cancer types. Finally, I investigate how on-treatment changes in the transcriptome correlate with response to the ICI-only and immunochemotherapy phases of treatment in the LUD2015-005 trial.

As discussed in the general introduction, TMB is one of the more well-studied genomic biomarkers in the context of ICI outcome[90]. Some ICI agents have been granted regulatory approval for any hypermutated solid tumour with TMB above a certain threshold. However, the benefit of ICI over chemotherapy in hypermutated populations is in question, and previous studies have raised doubts as to the significance of its predictive power in gastro-oesophageal cancer[60,90,110]. Therefore, clarifying whether TMB is predictive of outcomes on first-line immunochemotherapy in OAC is a key question to address.

Another genomic factor reported to link with ICI outcome is the genotype at HLA Class I (HLA-I) loci. Chowell et al. assessed the HLA status of a cohort of more than 1500 ICI-treated patients with advanced cancer[126]. They found that, among melanoma patients, the B44 supertype was associated with improved survival, while the B62 supertype was associated with inferior survival. The authors also showed inferior survival for patients with two copies of the same allele (homozygosity) at one or more HLA-I loci across all cancers. A separate report from NSCLC showed a similar link between HLA-I homozygosity and inferior survival, although in this cohort, it was the A02 supertype linked with improved outcome rather than B44[127]. This

suggests that HLA-I homozygosity could predict ICI responses in multiple settings, while specific allele supertypes linked with outcome may be more tumour-specific.

The key transcriptomic biomarkers predictive of positive outcome on αPD-L1 ICI include PD-L1 expression and T cell inflammation signatures, as summarized in the general introduction. Phase III studies in this setting have assessed PD-L1 using immunohistochemistry, with mixed findings for the association between PD-L1 expression and outcomes[73,75,76]. Following the sample collection protocols for the LUD2015-005 inoperable OAC cohort, no formalin-fixed, paraffin-embedded tumour material, the collection format usually required for PD-L1 companion diagnostic immunohistochemistry assays[128], was available from the pre-treatment timepoint for this work. However, RNA-sequencing assays likely also have utility to study associations between PD-L1 and ICI outcome[129], and RNA-sequencing data were available for the vast majority of patients in this cohort. Beyond PD-L1, no assessment of T cell inflammation or other transcriptional biomarkers predictive of positive outcome has yet been conducted in this setting, nor of transcriptional signatures of ICI resistance, which have not yet been shown to have the same potential cross-cancer applicability but have also been reported within cohorts[130]. Additionally, the mechanisms of how immunochemotherapy modulates the intratumoural microenvironment to affect response is not yet clear. As immunochemotherapy protocols enter the clinic for implementation in the first-line setting for OAC, evaluation of potential predictive biomarkers and characterisation of mechanisms of ICI activity remain important unmet needs. In this chapter, these questions will be addressed using bulk sequencing methods, which will allow for both the assessment of previously reported predictive biomarkers, and the establishment of a *de novo* workflow to find biomarkers most predictive of outcome in this setting.

4.2 Genomic biomarkers

4.2.1 Tumour mutational burden

In order to assess the utility of TMB and other genomic biomarkers to predict outcome on immunochemotherapy in OAC, tumour and paired-normal tissue pre-treatment biopsies were

prepared by our group for whole genome sequencing (WGS)[f]. WGS was conducted on all patients in the inoperable OAC cohort with primary tumour material available (33 of 35). To allow for a more direct comparison with TMB estimates from whole-exome sequencing and targeted-exome sequencing companion diagnostic assays, TMB quantification was restricted to the protein coding space (see Methods)[g].

A median TMB of 4.1 coding mutations per megabase (muts/Mb) were detected across the inoperable OAC cohort (IQR 1.2-5.4 muts/Mb). Patients attaining CB had a significantly higher TMB than those that did not (median TMB of 5.3 vs. 2.7 muts/Mb, p=0.007; Figure 4.1D). One patient was deemed to be MSI-high[h]; this patient had a particularly high TMB of 16.7 muts/Mb and attained CB. Both stratifying patients by their median TMB and a Cox regression with the TMB values themselves showed that patients with higher TMB had significantly improved OS (Figure 4.1A,C). This significant link between TMB and OS was preserved even after excluding the lone MSI-high patient (hazard ratio [95% CI]: 0.62 [0.40-0.95]). Interestingly, a similar analysis on inoperable OAC patients from the TCGA, as a representative cohort of cancers treated with conventional therapy, was not able to show any predictive power for TMB (Figure 4.1B), suggesting this biomarker may be specifically predictive of immunochemotherapy outcomes in this setting. In summary, these analyses show that a higher TMB is robustly associated with better clinical outcomes in this OAC cohort, and that this well-studied biomarker could help inform the selection of inoperable OAC patients who are likely to show greater benefit from the addition of ICI to conventional chemotherapy.

[f] DNA extraction was conducted by Jaeho Chang, Phil Xie, Michael White, and Richard Owen. WGS library preparation and sequencing was performed by the Oxford Genomics Centre at the Wellcome Centre for Human Genetics.

[g] WGS processing and variant calling was conducted by Benjamin Schuster-Böckler, Joseph Kaplinsky, and Robert Amess.

[h] MSI-high identification performed by Benjamin Schuster-Böckler on the basis of mutation signature profile, which is outside the scope of this thesis.

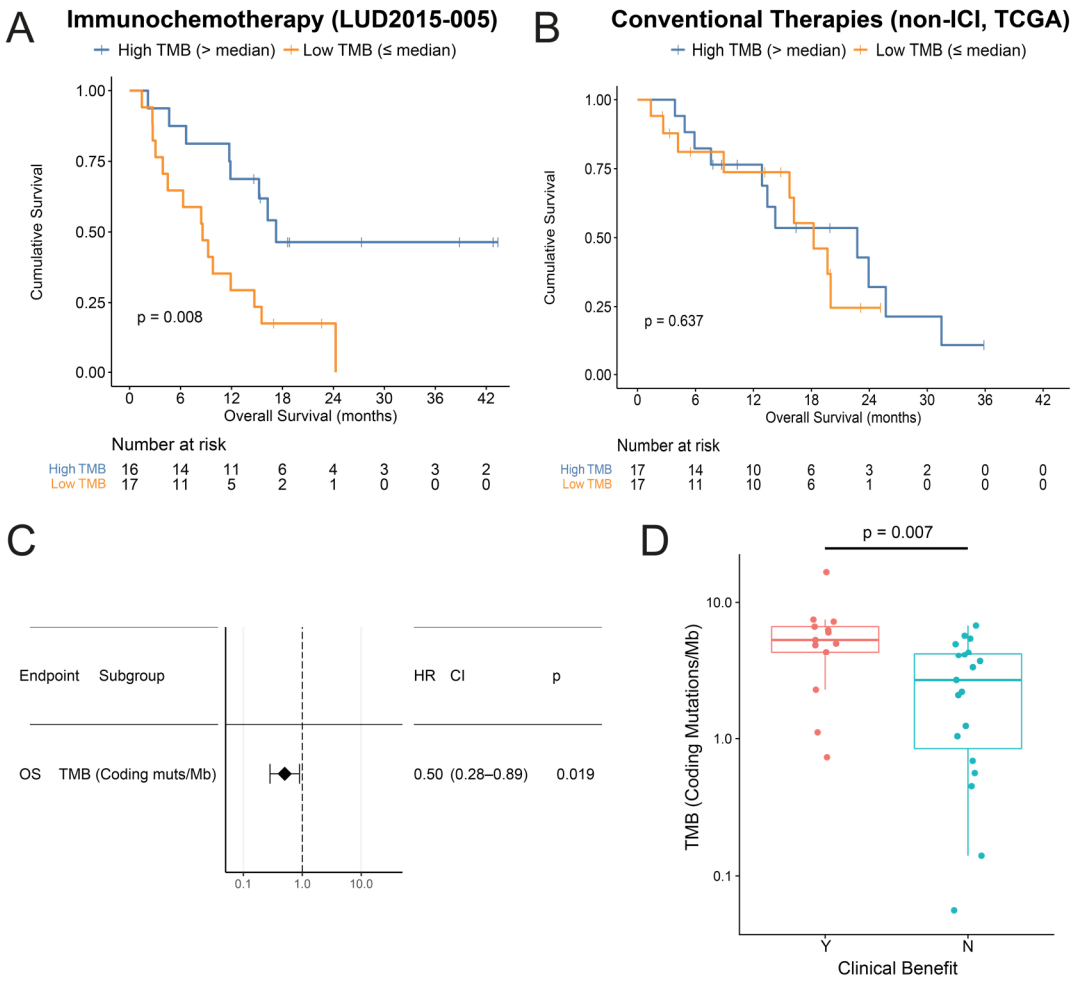


Figure 4.1: Pre-treatment coding tumour mutational burden is a robust predictor of immunochemotherapy outcome in OAC. Top: Patients for the A) LUD2015-005 inoperable OAC cohort and B) Stage III and IV OAC tumours from the TCGA database were stratified based on their median coding TMB rate, expressed as the rate of coding mutations per megabase (Mb). For TCGA, mutation counts and clinical data were downloaded from cBioportal, and subset to only the stage III and IV tumours to maintain better comparability with the present cohort. C) Coding TMB rates per Mb for the LUD2015-005 inoperable OAC cohort were input into a univariable Cox regression with OS. Coding TMB rates were scaled and centered prior to regression. D) Coding TMB rate (Coding TMB) for each sample in the LUD2015-005 inoperable OAC cohort, grouped by clinical benefit status. P value between groups was calculated by a Mann-Whitney U-test.

4.2.2 HLA class I genotype[i]

The supertype analysis reported by Chowell et al. was reproduced on the LUD2015-005 inoperable OAC cohort, grouping related HLA class I alleles into their supertypes[131] in order to assess their link with outcome. While A02 alleles had a trend towards predicting improved OS, no supertype was significantly linked with outcome in this cohort (Table 4.1). However, patients who were homozygous for at least one classical Class I HLA loci did have inferior OS in this

[i] 4-digit HLA typing from WGS data was conducted by Joseph Kaplinsky.

cohort (Figure 4.2). Therefore, while the allele diversity at class I loci may be linked with clinical outcome, a significant link between specific class I supertypes and outcome could not be shown in this cohort, perhaps due to the comparatively smaller size of this cohort compared to those assessed in previous reports.

De novo assessment of genomic predictive biomarkers is beyond the scope of this thesis, and will be assessed in future work from other collaborators working on the LUD2015-005 study. Nevertheless, it appears that previously reported TMB and HLA biomarkers are useful predictive factors of ICI outcome in first-line immunochemotherapy for OAC. However, some patients with low TMB and HLA-I heterozygosity still achieve good long-term survival outcomes, suggesting other classes of biomarkers could also have important utility.

HLA Supertype	**Frequency (%)**	**Univariable - OS**		
		HR (95% CI)	p	FDR
A01	12/35 (34.3%)	1.75 (0.78-3.92)	0.173	0.480
A01A24	4/35 (11.4%)	0.81 (0.19-3.47)	0.780	0.903
A02	19/35 (54.3%)	0.46 (0.20-1.01)	0.054	0.480
A03	16/35 (45.7%)	1.15 (0.52-2.54)	0.732	0.903
A24	5/35 (14.3%)	0.68 (0.20-2.30)	0.534	0.903
B07	17/35 (48.6%)	0.91 (0.41-2.00)	0.813	0.903
B08	10/35 (28.6%)	1.94 (0.85-4.44)	0.117	0.480
B27	6/35 (17.1%)	1.15 (0.42-3.11)	0.786	0.903
B44	18/35 (51.4%)	0.59 (0.27-1.30)	0.192	0.480
B62	6/35 (17.1%)	0.96 (0.33-2.79)	0.934	0.934

Table 4.1: HLA supertype and link with overall survival. HLA-A and -B alleles were grouped into allele supertype groups[131], and Cox regressions were performed with OS. HLA-C alleles do not have reported supertype groupings. All results are shown for the presence vs. absence of the given supertype, with an HR above 1 representing worse OS for the presence of that supertype. p-value adjustment was conducted using the Benjamini-Hochberg procedure (FDR). The prevalence of each supertype in the inoperable OAC cohort is also reported.

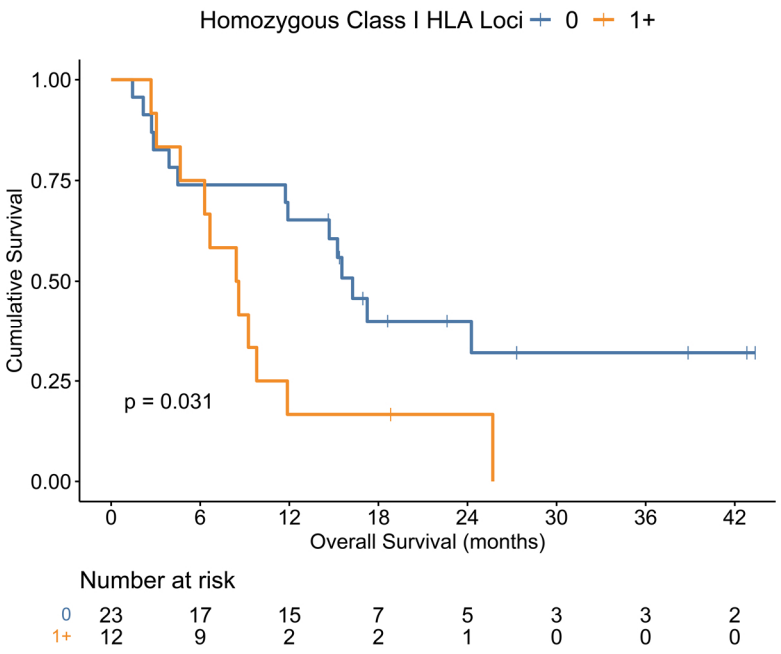


Figure 4.2: HLA Class I homozygosity and link with overall survival. The count of homozygous classical HLA class I loci (A, B, and C) for each patient was determined, and the link with OS is presented using a Kaplan-Meier plot. A locus was considered homozygous if both alleles had the same protein sequence (4-digit HLA allele designation). Patients with 1-3 homozygous class I alleles were grouped for analysis. P-value was calculated by the log-rank method.

4.3 Evaluation of an RNA extraction and library preparation workflow for transcriptional biomarker analysis

4.3.1 Background and RNA-sequencing workflow

In order to profile additional predictive biomarkers in this setting, transcriptome profiling using RNA-sequencing (RNA-seq) was selected, which allows for assessment of previously reported markers such as PD-L1 and inflammatory gene expression signatures, but also allows for *de novo* characterisation of genes and signatures that are most predictive for this setting.

Compared to DNA profiling, transcriptome profiling has an additional layer of difficulty due to the much more rapid rate of RNA degradation, which can have a major adverse impact on gene expression profiles. This adverse impact is of particular concern for RNA extraction from tumour tissue biopsies, where necrosis and associated degradation mechanisms are often prevalent[132]. To assess the quality of RNA extractions, the total yield and RNA integrity number (RIN), a measure of RNA degradation based on the electrophoretic profile which ranges from 1 (completely degraded) to 10 (completely intact)[133], are often used. A recent biobanking study of

flash-frozen endoscopic OAC biopsies reported a median RIN of 5.5 from extracted RNA, showing that increased time between the collection of the biopsy and freezing was a major driver of decreasing RNA quality[134]. This suggests that RNA degradation during standard sampling and storage of OAC endoscopic biopsies is a likely source of potentially confounding variation, and therefore careful consideration of this factor is warranted when generating RNA-seq datasets for biomarker analysis in this setting.

With this in mind, an RNA-seq workflow for GI tract biopsies was designed with an aim to maximize RNA quality, while still allowing for reproducible implementation by biobankers working across the multiple enrolment sites of the LUD2015-005 trial[j]. Given the reported link between time-to-freezing and RNA degradation, this workflow prioritized immediate flash-freezing, generally in-theatre, prior to long-term storage at -80°C. For RNA extraction, as working with a mortar and pestle on small biopsies can be challenging, a single-use polypropylene pestle was used for rapid disruption and homogenization of biopsies directly in pre-chilled lysis buffer, in order to minimize degradation associated with tissue thawing. Finally, a ribodepletion library preparation method was employed for RNA-seq, which tends to perform better on degraded RNA samples than poly-A mRNA selection methods[135].

4.3.2 RNA extraction and quality control metrics

From all patients in the inoperable LUD2015-005 cohorts with biopsies available from the primary tumour, RNA was extracted from tumour samples taken pre-treatment (PreTx), on-treatment at the end of the four-week ICI-only window (ICI-only), and at the end of the study following the immunochemotherapy period (PostTx). For some patients, RNA was also extracted from biopsies of paired normal control tissues- oesophagus, gastric cardia, and descending duodenum. All in all, RNA was extracted from a total of 145 flash-frozen biopsies[k] (Figure 4.3).

[j] The biopsy preservation and RNA-seq extraction workflows for this study were originally developed by Richard Owen and Michael White, with later updates and optimization from myself.

[k] Richard Owen and Michael White extracted RNA from earlier batches, while RNA extractions from later batches were split between Joseph Chadwick and myself.

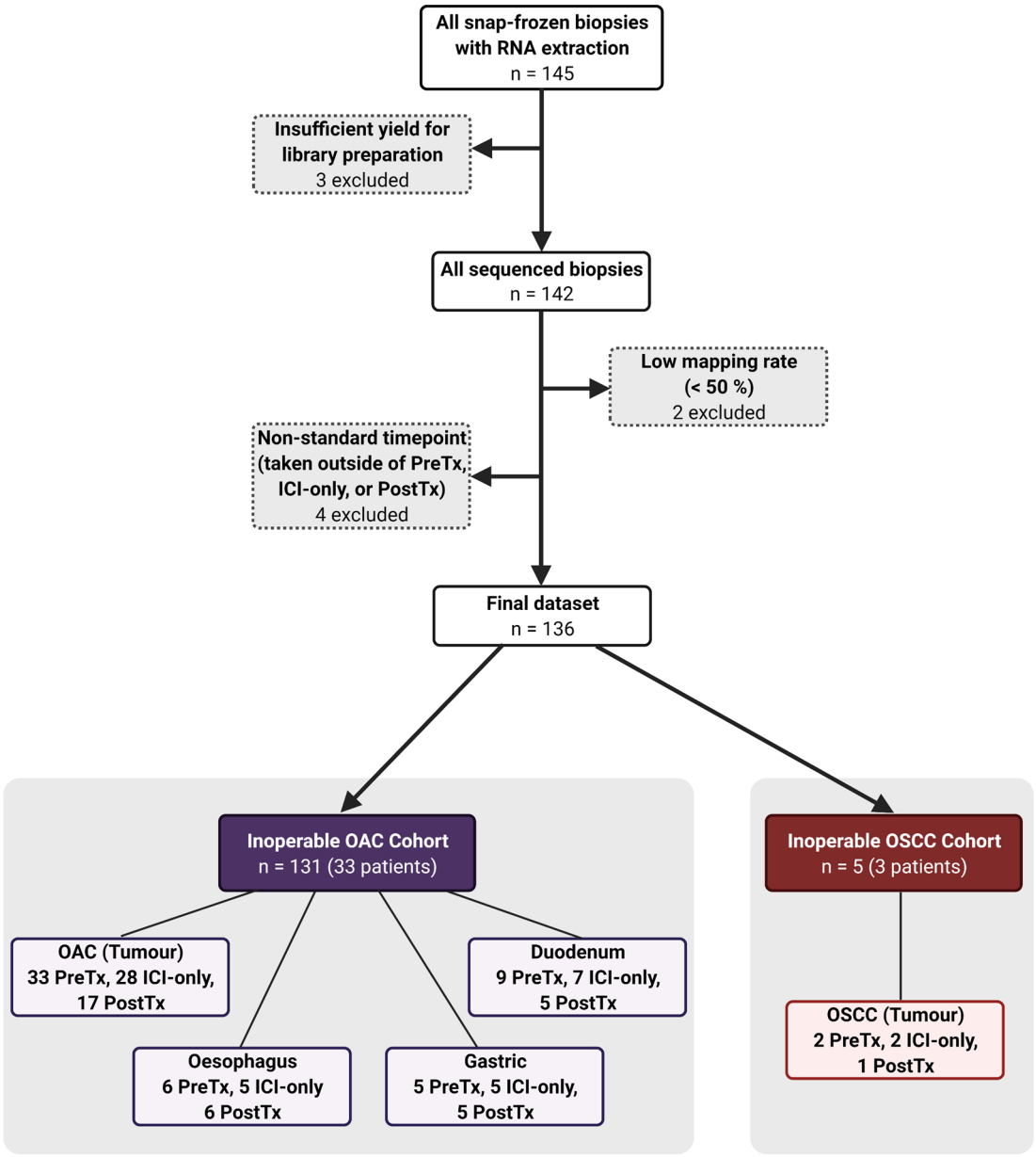


Figure 4.3: Summary of RNA-seq dataset construction for the LUD2015-005 trial. Workflow showing processing and filtering of RNA samples taken from inoperable cancer patients on the LUD2015-005 study. All patients with available biopsies of the primary lesion were included (36 of 38). For biopsies excluded due to low yield or low mapping rate, a replacement biopsy from the same patient, tissue, and timepoint was included in the final dataset.

To evaluate the quality of preserved RNA extracted by the workflow described above, the RIN and yield of each extraction was assessed. The quality of RNA extracted for tumour tissues was generally good, with a median RIN of 6.7 (range 1.3-9.7) and median RNA yield of 5.6µg (range 0.14-21.6µg), a slight improvement over previous reports[134]. However, a significant proportion (31/89, 34.8%) of tumour samples still showed substantial RNA degradation (RIN < 5).

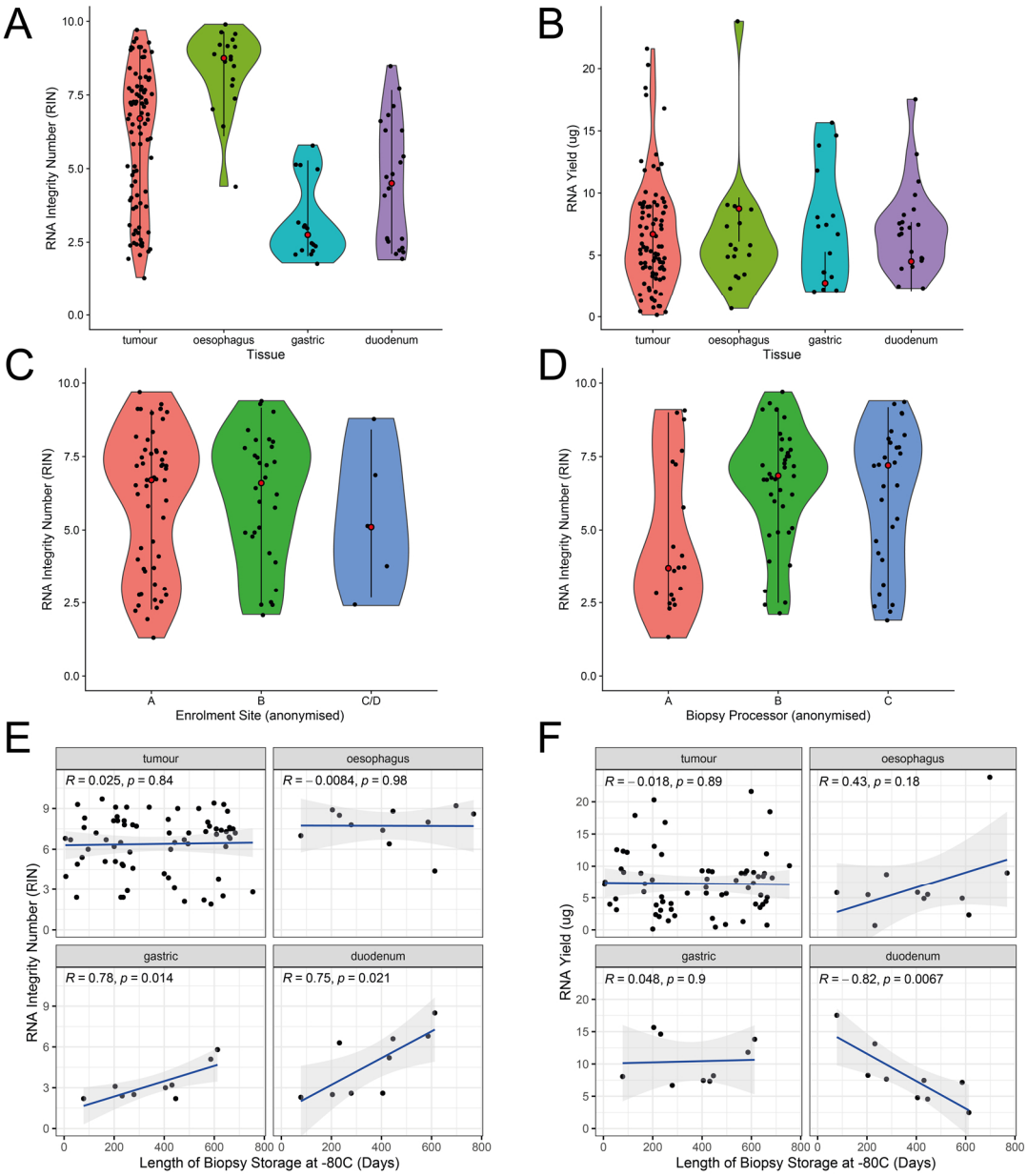


Figure 4.4: RNA QC metrics for all extracted biopsies. Top: The RNA integrity number (RIN) (A) and total yield in micrograms (B) is plotted for each tissue type. Each black dot represents an individual sample, while the violin shows the density of points along the y-axis. For each tissue type, the median (red dot) and a vertical line showing a 90% confidence interval are provided. Middle: Same as above, except RINs of only the tumour biopsies are assessed, grouped by the enrolment site (C) and the individual processing the biopsy for RNA extraction (D), to check for possible batch effects. Bottom: The effect of long-term storage at -80°C on RIN (E) and yield (F) are assessed. The trend line (blue) shows the linear regression, and correlation coefficients and p-values (Pearson method) are shown. For B and F, 3 biopsies with yield below limit of detection were excluded.

There were significant differences between the quality of RNA extracted from the various normal control tissues, with oesophageal biopsies showing higher RINs. However, no differences were noted in RNA yield between the different tissue types (Figure 4.4A-B). RNA quality was consistent for biopsies collected at different enrolment sites, and despite some

variance between RNA extractors, confidence intervals for RNA quality remained overlapping (Figure 4.4C-D). These findings suggest the developed workflow was reproducible and likely did not result in major batch effects on the basis of these factors. Biopsies were also stored at -80°C for a wide range of durations prior to RNA extraction (median 410.5 days, range 5-768 days), raising a potential concern that biopsies stored for longer periods could show increased RNA degradation. However, increasing storage length did not correlate with decreasing RIN for any tissue type assessed. With respect to RNA yield, only duodenal biopsies showed any significant decrease with longer storage (Figure 4.6E-F). In summary, while degraded RNA was still a concern for some samples, this RNA extraction workflow was generally successful, delivering moderate improvements in RNA quality over previous reports that were reproducibly obtained across multiple enrolment sites and biopsy processors involved in this trial.

4.3.3 RNA-sequencing and assessment of alignment quality

142 flash-frozen biopsies had sufficient yield and were taken forward for RNA-seq[l]. Alignment to the human genome was performed using STAR[136,137], a splice-aware mapping and alignment tool frequently used for large-scale cohort studies such as The Cancer Genome Atlas (TCGA)[138]. STAR showed good mapping rates across most samples, with a median unique alignment rate of 90.2% (IQR 87.9-91.3%). 2 tumour samples with particularly low mapping rates by STAR (<50%) were excluded (Figure 4.3), leaving 140 biopsies in the final dataset.

Despite variation in RIN, most samples showed uniform gene body coverage (Figure 4.5A). However, samples with lower RINs tended to have a much lower percentage of reads assigned to mRNA coding sequences and untranslated regions (UTRs; Figure 4.5B-C), the typical regions assessed during differential gene expression. While 19/19 (100%) of tumour samples with relatively intact RNA (RIN > 8) had at least 50% of mapped reads falling into these mRNA regions, this was only true for 4/28 (14.3%) of tumour samples with degraded RNA (RIN < 5). Therefore, samples with intact RNA yielded preferable results in this cohort, with a higher yield of reads that fell into transcriptomic regions informative for differential gene expression.

[l] RNA-seq library preparation was conducted by the Oxford Genomics Centre (OGC) at the Wellcome Centre for Human Genetics, Joseph Chadwick, and myself. OGC and the Imperial College London Genomics Facility performed sequencing. Robert Amess provided automation of RNA-seq processing and QC workflows.

However, given the robust preservation of even coverage across the gene body despite variation in RIN, and the fact that downstream analysis tools typically normalize for the total number of transcriptomic counts per sample, all non-excluded samples at this stage were considered to be of sufficient quality for further analysis. Therefore, all 140 remaining RNA-seq samples were taken forward for biomarker study, 131 of which were relevant to the inoperable OAC cohort (Figure 4.3).

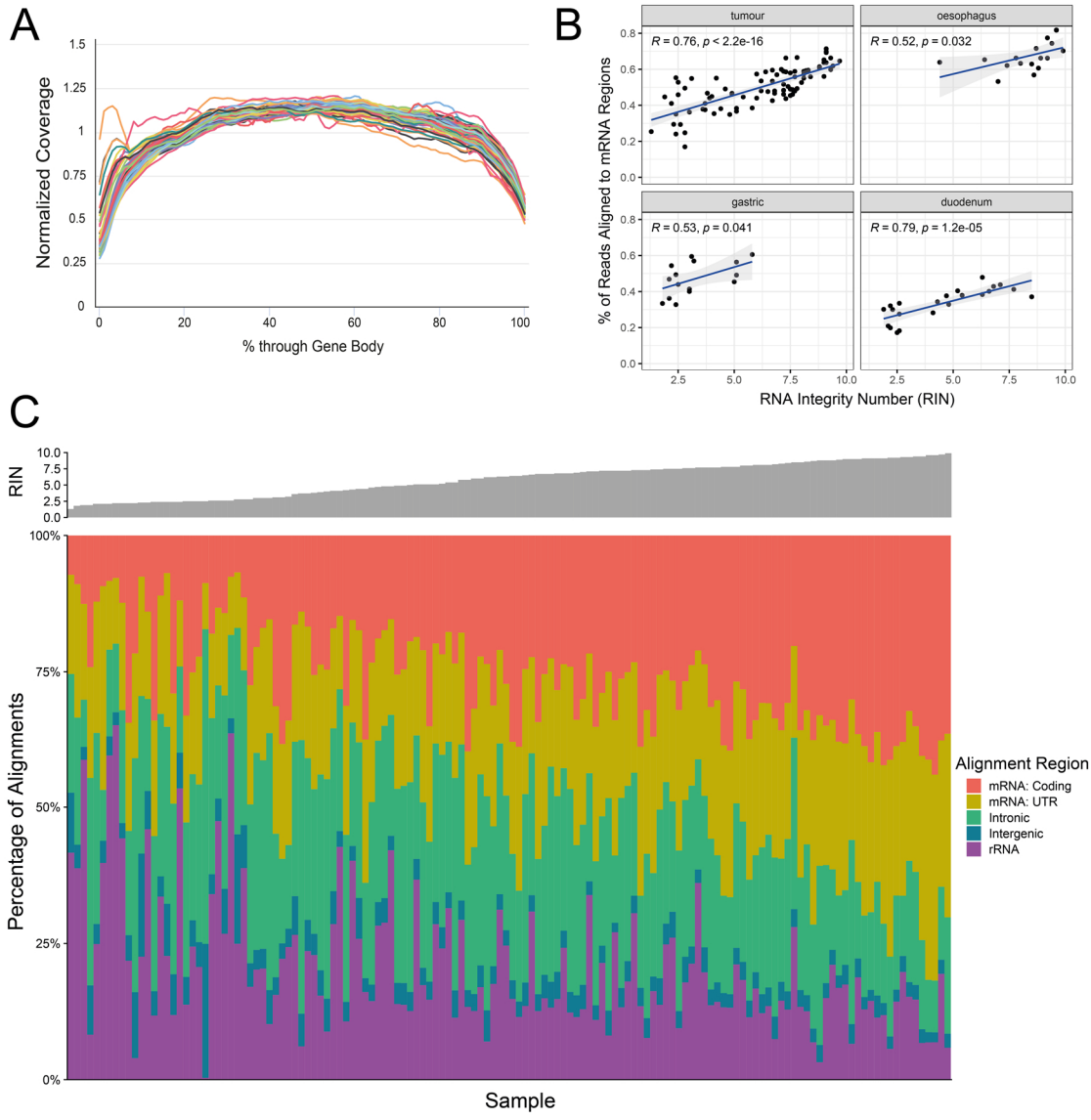


Figure 4.5: Dependence of RNA integrity on alignment QC metrics. A) The average sequencing coverage across all gene bodies (scaled according to the length of each individual gene prior to averaging) is shown, with each line representing a single sample. B) RIN is plotted against the percentage of mapped reads that align to mRNA regions (coding and UTR), as assessed by Picard using the input gene annotation file. Pearson correlation coefficients and their associated p-values are shown. C) All non-excluded samples were ordered by their RIN, which is shown on the top subplot. The stacked bar chart shows the percentage of mapped alignments for each sample that are assigned to each of the listed genomic regions, as determined by Picard.

4.4 Overview of bulk transcriptome profiles

4.4.1 PCA analysis

Prior to beginning ICI biomarker analysis, broad overview analyses were first conducted to summarize the structure of the dataset and identify features that differentiated OAC from normal GI tissues. As a first step, PCA was conducted on all samples from the inoperable OAC cohort. At all three timepoints, healthy control duodenal, gastric, and oesophageal biopsies were resolved in discrete groupings by PCA. Contrastingly, OAC samples fell in a spread between the oesophageal and gastric tissues, but were not clearly resolved, perhaps speaking to the greater variability inherent in these malignant tissues (Figure 4.6A). Interestingly, more overlap was noted between OAC tumour biopsies and oesophageal samples at the PostTx timepoint than PreTx or ICI-only. This PostTx cluster of patients overlapping the normal squamous oesophagus grouping appeared to attain a marginally higher rate of clinical benefit (Figure 4.6B). This could perhaps reflect a successful localized response to therapy which returns the phenotype at the site of the primary lesion towards that of healthy squamous oesophagus, and that this finding post-therapy could be suggestive of improved long-term outcome. However, the marginally higher rate of clinical benefit in this group was not statistically significant ($p = 0.15$). Finally, differences in tissue type, rather than RIN, seemed to drive the clustering at all three timepoints (Figure 4.6C), supporting the decision to keep in the full set of samples regardless of input RNA quality.

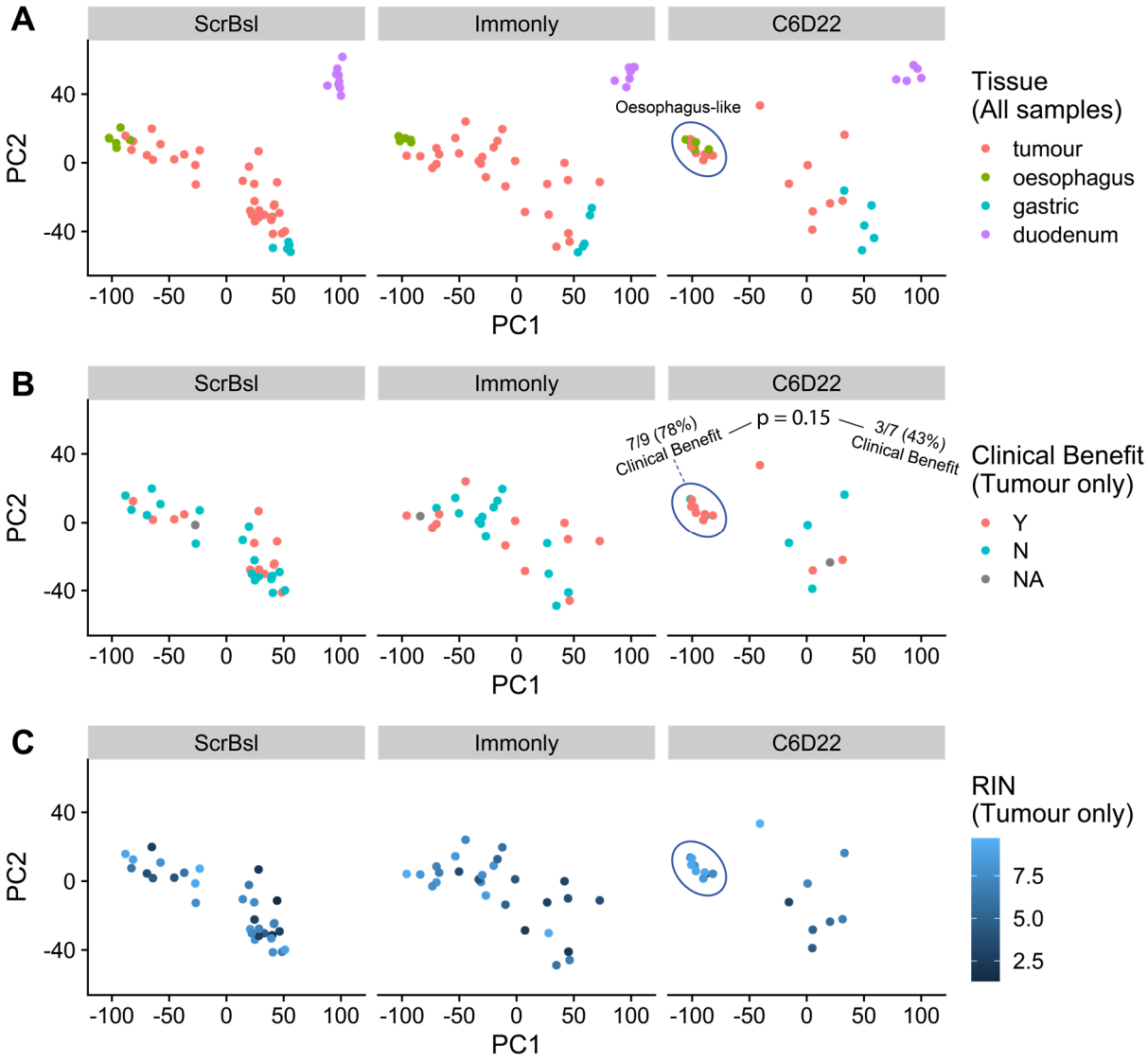


Figure 4.6: Principal component overview of the inoperable OAC RNA-seq dataset. Principle component analysis (PCA) was conducted on variance stabilizing transformation (VST)-normalized counts across all samples in the inoperable OAC cohort, and the first two PCs are plotted. Points are separated into separate facets based on the sampling timepoint. A) Samples are coloured by their tissue of origin. A grouping of oesophageal samples and transcriptionally similar tumour site biopsies is highlighted at PostTx (right). B) Only tumour samples are shown, with the same loadings as in (A). Points are coloured by their clinical benefit status, and the rate of benefit is highlighted for the groupings at PostTx (right) p-value was calculated using a two-proportion z-test. C) Same as (B), except points are coloured according to their RIN.

4.4.2 Characterisation of OAC in comparison with normal GI tissues

The collection of tumour and paired-normal tissues in the same dataset provides an opportunity to define the transcriptional profile of OAC in comparison with non-malignant GI tissues. In addition to gaining insight about the general characteristics that differentiate OAC, this approach also allows for tissue-specific quantification of immune infiltrates and checkpoint

molecules, which could be important to evaluate the potential efficacy of this and other immunotherapeutic treatment strategies.

To assess transcriptional signatures specific to OAC, differentially expressed genes (DEGs) were calculated between OAC and normal control tissues (oesophagus, gastric, and duodenum). To avoid potential bias caused by any tissue-specific impact of therapy, only PreTx biopsies were included in this comparison. Using a false discovery rate (FDR[139]) adjusted p-value threshold of 0.1 (the default FDR threshold for DESeq2 analysis), 7183 DEGs were upregulated in OAC compared to normal control tissues, and 6449 DEGs were downregulated in OAC (Appendix C-1). Downregulated genes returned by this approach were generally dominated by marker genes of one normal tissue type rather than tumour-specific downregulation (data not shown); therefore, upregulated genes were chosen as the focus for this analysis. Inspection of the top DEGs (Table 4.2) revealed four categories of highly upregulated genes in OAC: 1) Cancer-Testis (CT) antigens (*MAGEA4*, *PRAME*), 2) markers of myeloid cells (*FPR1* and *FCGR3A*/CD16) and their chemoattractants (CXCR2 ligands: *CXCL1* and *CXCL8*[140,141]), 3) stromal markers (matrix metallopeptidases and endothelial genes), and 4) potential epithelial-specific drivers of OAC, such as homeobox proteins which have previously been reported as potential drivers of Barrett's oesophagus (BO) and are also enriched in Barrett's-related OAC[142,143]. These data suggest significant dysregulation in both the malignant epithelium and the tumour microenvironment in OAC.

Gene	Description	log2FC	FDR	Category
HOXB7	Homeobox B7	4.98	2.06E-39	Homeobox
SERPINE1	Serpin family E member 1	4.99	6.03E-38	Other
MAGEA4	MAGE family member A4	22.13	6.55E-38	CT Antigen
HOXB6	Homeobox B6	5.19	2.47E-36	Homeobox
MMP9	Matrix metallopeptidase 9	5.14	1.88E-35	Stromal (MP)
CXCL8	C-X-C motif chemokine ligand 8	6.49	1.25E-34	Myeloid (chemokine)
MIR4435-2HG	MIR4435-2 host gene	2.78	5.18E-34	Other (lncRNA)
HOXB9	Homeobox B9	7.20	9.26E-34	Homeobox
MMP12	Matrix metallopeptidase 12	6.86	1.96E-33	Stromal (MP)
CXCL1	C-X-C motif chemokine ligand 1	5.34	1.76E-32	Myeloid (chemokine)
SPP1	Secreted phosphoprotein 1	6.59	1.66E-31	Other
SOCS3	Suppressor of cytokine signalling 3	4.72	3.25E-31	Other
TUBB3	Tubulin beta 3 class III	4.44	3.90E-31	Other
FCGR3A	Fc fragment of IgG receptor IIIa	4.36	4.15E-28	Myeloid (marker)
ANGPT2	Angiopoietin 2	3.90	5.51E-28	Stromal (endothelial)
HOXA13	Homeobox A13	7.53	5.68E-28	Homeobox
TIMP1	TIMP metallopeptidase inhibitor 1	3.84	1.62E-27	Other
SALL4	Spalt like transcription factor 4	5.27	6.45E-27	Other
FKBP10	FKBP prolyl isomerase 10	4.40	3.83E-26	Other
MTHFD1L	Methylenetetrahydrofolate dehydrogenase 1 like	2.15	9.15E-26	Other
PRRX2	Paired related homeobox 2	4.26	1.06E-25	Homeobox
OSR2	Odd-skipped related transcription factor 2	4.74	1.85E-25	Other
LUCAT1	Lung cancer associated transcript 1	5.46	2.90E-25	Other (lncRNA)
FPR1	Formyl peptide receptor 1	5.47	1.11E-24	Myeloid (marker)
PTGS2	Prostaglandin-endoperoxide synthase 2	5.30	1.76E-24	Other
ADAMTS4	ADAM MP with thrombospondin type 1 motif 4	3.71	1.95E-24	Stromal (MP)
PRAME	PRAME nuclear receptor transcriptional regulator	6.87	1.95E-24	CT Antigen
ESM1	Endothelial cell specific molecule 1	5.04	2.60E-24	Stromal (endothelial)
LAMC2	Laminin subunit gamma 2	3.42	3.58E-24	Stromal (ECM)
SAA1	Serum amyloid A1	6.55	4.56E-24	Other

Table 4.2: Top 30 genes upregulated in oesophageal cancer over normal GI tissues. Differential expression was conducted using DESeq2 between all pre-treatment OAC biopsies and all pre-treatment normal tissue biopsies (oesophagus, gastric, and duodenum). Moderated fold changes (log2 scale, log2FC) and adjusted p-value (FDR) are presented as calculated by DESeq2 for each of the 30 most significantly upregulated genes. Gene categories were assigned manually. MP = metallopeptidase; ECM = extracellular matrix component.

Given the immune infiltrate differences suggested in the top DEGs, a full panel of immune cell type markers were assessed to profile the differences in the immune cell microenvironment between OAC and normal GI tissues (Figure 4.7A). Overall, while levels of T cell infiltration (*CD2,*

CD3E) were not higher in OAC, *CD4* and *FOXP3* were increased compared with normal tissues. Contrastingly, infiltration of cytotoxic T (*CD8A*, *CD8B*) and NK cells (*NCAM1*/CD56, *GNLY*) was marginally lower in OAC than in normal control tissues, particularly when compared with gastric and duodenal biopsies. Commonly used markers for cytolytic activity of these cell types (*GZMA*, *PRF1*) showed a similar trend. Therefore, increased T cell infiltration relative to normal GI tissues does not appear to be a feature of OAC, and the T cell infiltrate that is present skews more towards helper and regulatory T cells than cytolytic effectors. Markers of B cell and plasma cell infiltrate were low in normal squamous oesophagus, but were highly present in gastric, duodenal, and OAC biopsies. Comparing antibody isotypes between OAC and gastric and duodenal tissues, the IgG/IgA ratio appeared marginally higher in OAC compared to gastric and duodenal biopsies.

In contrast with T and B cells, general myeloid infiltration (*ITGAM*/CD11b, *ITGAX*/CD11c) appeared highly enriched in OAC, as a key feature differentiating this tumour type from all normal GI tissues. Within this group, monocyte (*CD14*, *FCGR3A*/CD16a) and neutrophil (*FCGR3B*/CD16b) markers were particularly enriched in tumour tissues. The pan-macrophage marker *CD68* was distributed more evenly between OAC and normal tissues, although M2 markers (*CD163*, *MRC1*/CD206) were more highly expressed in OAC, gastric, and duodenum than in normal oesophagus. Interestingly, normal oesophagus had highly enriched expression of a number of conventional dendritic cell markers, including *XCR1* (cDC1), *CD1C* (cDC2), and *CD1A* and *CD207* (Langerhans-like DCs), whereas plasmacytoid DC markers (*IL3RA*/CD123) were more specific to OAC.

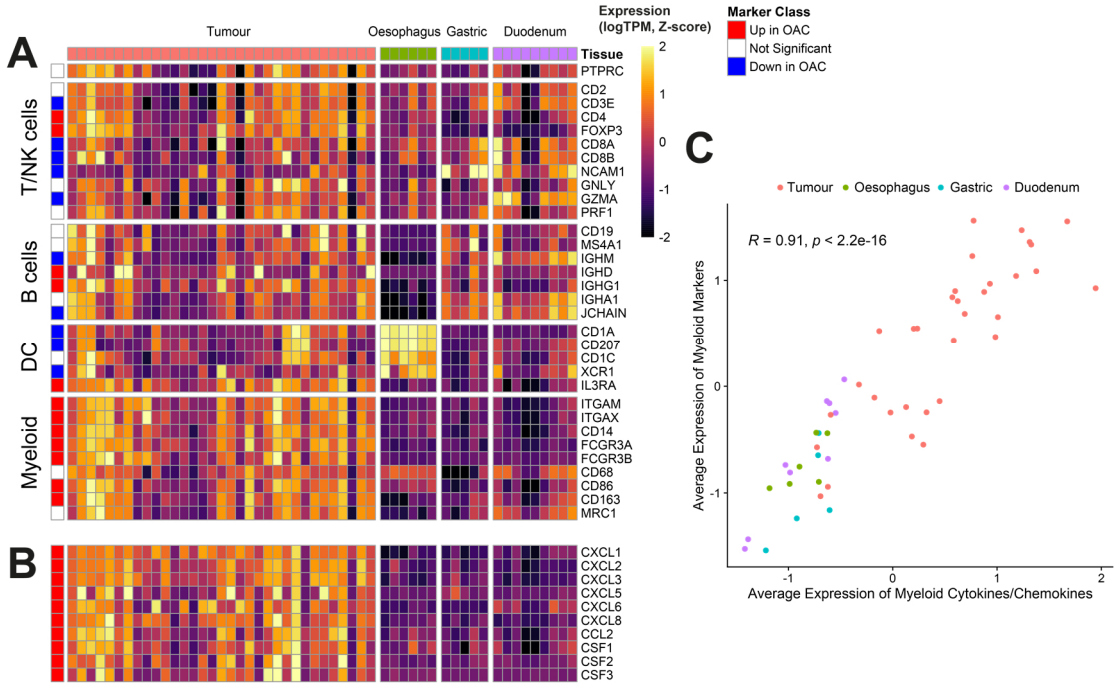


Figure 4.7: Myeloid infiltration is a key feature of the OAC immune landscape compared to normal GI tissues. A) Z-score normalized logTPM (log10 TPM, with 0.1 offset) expression of each gene is shown, separated into blocks for markers of general immune infiltration (*PTPRC*/CD45, top), T/NK cells, B cells, dendritic cells (DC), and other myeloid cells. Genes are marked as significantly up or down in OAC if they met a significance threshold (FDR < 0.1). B) Same as (A), but for a panel of myeloid-targeted cytokines and chemokines. C) The mean of logTPM Z-scores for the myeloid markers (from A) and myeloid cytokines and chemokines (from B) are plotted for each sample. The Pearson correlation coefficient and p value for this trend are shown.

To further explore the reason for increased myeloid infiltration in OAC, the expression of myeloid-targeted cytokines and chemokines were assessed, namely CXCR2 ligands, *CCL2*, and members of the colony-stimulating factor (CSF) family, which are thought to be key mediators of myeloid infiltration and maintenance in tumours[144]. These genes were also highly enriched in OAC (Figure 4.7B). Interestingly, there was a high degree of correlation between these myeloid-targeted molecules and the presence of the myeloid markers themselves (Figure 4.7C), suggesting a likely connection between the presence of myeloid cells in the tissue and the expression of these cytokines and chemokines. In summary, the immune microenvironment in OAC appears to be dominated by myeloid infiltration, particularly with cells expressing monocyte and neutrophil markers, which is likely downstream of increased myeloid cytokines and chemokines secreted by the tumour or its microenvironment.

While CT antigens are not specifically targeted by αPD-L1/αCTLA-4 ICI, their expression in tumours outside of immune-privileged environments can generate T cell mediated anti-

tumoural responses, and therefore they are potentially relevant to ICI. Additionally, as MAGE-A3 and NY-ESO-1 CT antigens have been targeted for immunotherapeutic interventions in oesophageal cancer[145,146], characterisation of additional CT antigens relevant in this setting could assist the development of novel immunotherapies for future use in the clinic. Towards this end, all putative CT antigens annotated in the CTdatabase[147] were queried within this PreTx analysis.

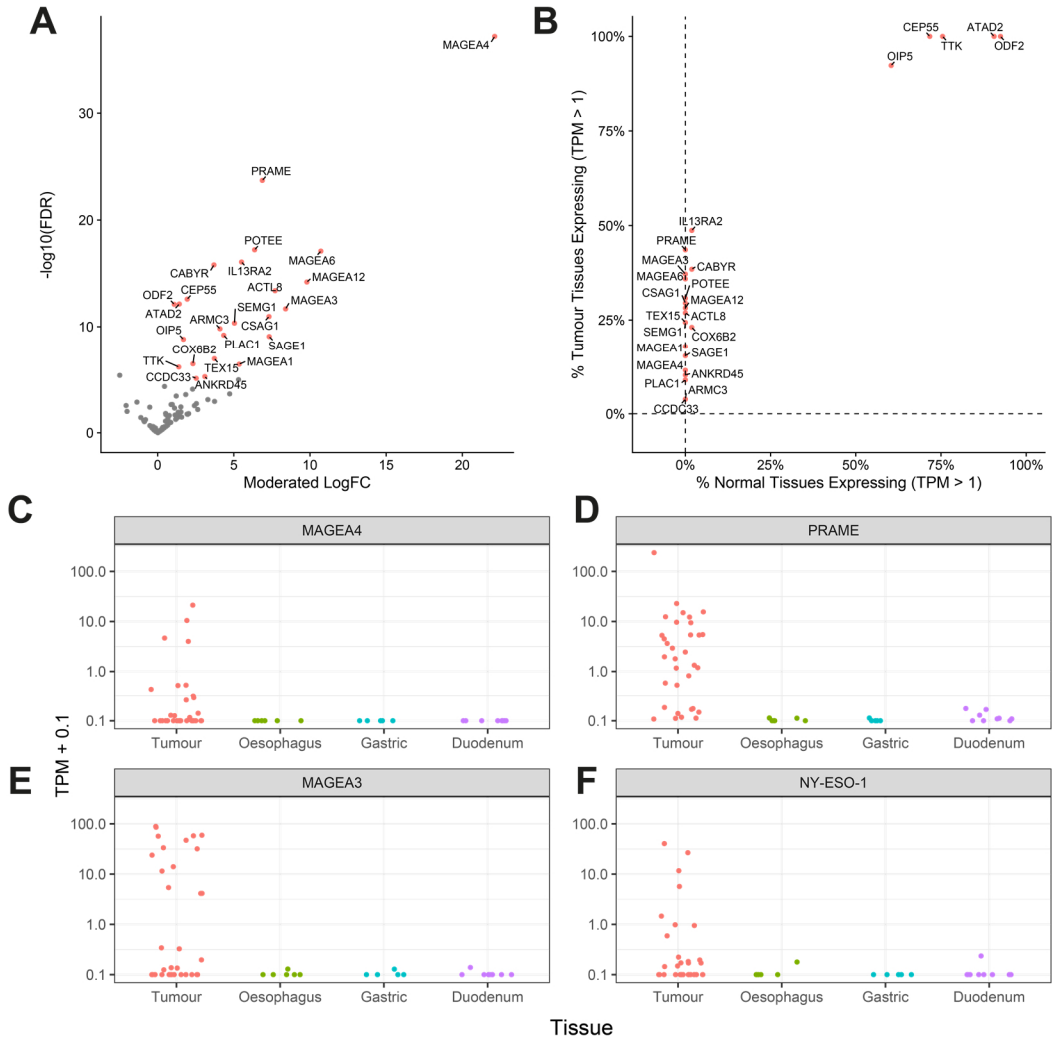


Figure 4.8: Pre-treatment expression of known cancer-testis antigens in OAC. A) Differential expression results between OAC and all normal tissues (as per Table 4.2) were filtered for CT antigens from the CTdatabase. The moderated fold change ('ashr' method) and significance (-log10 of the FDR) are shown for each gene. 24 genes with FDR < 1e-5 and higher expression in OAC are highlighted in red. B) For the 24 genes highlighted in A, the percentage of tumour and normal samples with at least 1 TPM expression are shown. C-F) TPM counts (0.1 offset) are plotted for each tissue type to show specificity of selected genes. Y axis is on a log10 scale. NY-ESO-1 TPM (F) are calculated as the sum of *CTAG1A* and *CTAG1B* TPM, including multi-mapping read counts (not default).

Of all CT antigens from the CTdatabase, 58 of 107 with detectable expression in PreTx samples were significantly upregulated in OAC (FDR > 0.1), and 24 of these were highly upregulated (FDR > 1e-5). Of these, *MAGEA4*, *PRAME*, *POTEE*, and *MAGEA6* were the most specifically expressed in OAC compared with normal control tissues, although *MAGEA3* also showed strong OAC specificity (Figure 4.8A). Some highly upregulated CT antigens did not have strong expression (TPM > 1) in any normal biopsies, while others were less specific (Figure 4.8B). For most CT antigens, some low-level expression below this 1 TPM threshold was occasionally detected in normal tissue, however for *MAGEA4*, expression in normal tissues was completely undetectable in the normal biopsies profiled (Figure 4.8C-D).

NY-ESO-1 was not detected by the standard processing pipeline used in this analysis, as this protein is encoded by two separate, nearly entirely homologous genes (*CTAG1A* and *CTAG1B*). Reads from NY-ESO-1 will therefore necessarily have multi-mapping alignments, which are excluded by default from quantification. However, a variant of this pipeline allowing for quantification of multi-mapping alignments showed robust expression of NY-ESO-1 in some OAC patients, although at a somewhat lower frequency than MAGE-A3 (Figure 4.8E-F). This analysis confirms significantly enriched expression of NY-ESO-1 and MAGE-A3 in OAC. However, it should be noted that other CT antigens such as *MAGEA4* and *PRAME*, which have not yet been targeted in the clinic in this cancer type, are more significantly enriched in OAC compared to normal GI tissues.

Finally, as the expression of the targeted checkpoint molecules are frequently used as predictive markers for outcome on ICI, the ligands and receptors targeted in this study (PD-L1 and CTLA-4 and their binding partners) were assessed across tissue types, alongside a panel of other potentially targetable coinhibitory and costimulatory T cell receptors and ligands (Figure 4.9A-B). Of all 19 checkpoint genes in this panel, *CTLA4* was most significantly enriched in OAC versus normal control tissues. Interestingly, the full CTLA-4 axis, including CTLA-4 ligands (*CD80, CD86*) in addition to CTLA-4 itself, was more significantly enriched in OAC than PD-L1 (*CD274*) or its receptor (PD-1/*PDCD1*). Indeed, only a fraction of patients in this cohort had increased expression of PD-1 or PD-L1 in OAC compared to the highest level seen in normal

oesophageal biopsies, whereas CTLA-4 was much more commonly upregulated compared to normal oesophagus (Figure 4.9C). Of the checkpoint molecules not targeted in this trial, the co-inhibitory molecule *CD276* (B7-H3) and co-stimulatory molecule *TNFRSF9* (4-1BB) had the most significant enrichment in OAC compared to normal tissues. Overall, PreTx inoperable OAC shows enriched expression of a number of potentially targetable T cell checkpoint molecules. While CTLA-4 and its ligands appear more frequently upregulated in OAC compared with normal tissues, upregulation of the PD-1/PD-L1 axis occurs only in a smaller subset of OACs. This could be an important consideration in this setting, where currently only single-agent ICI targeting the PD-1/PD-L1 axis is approved for first-line use.

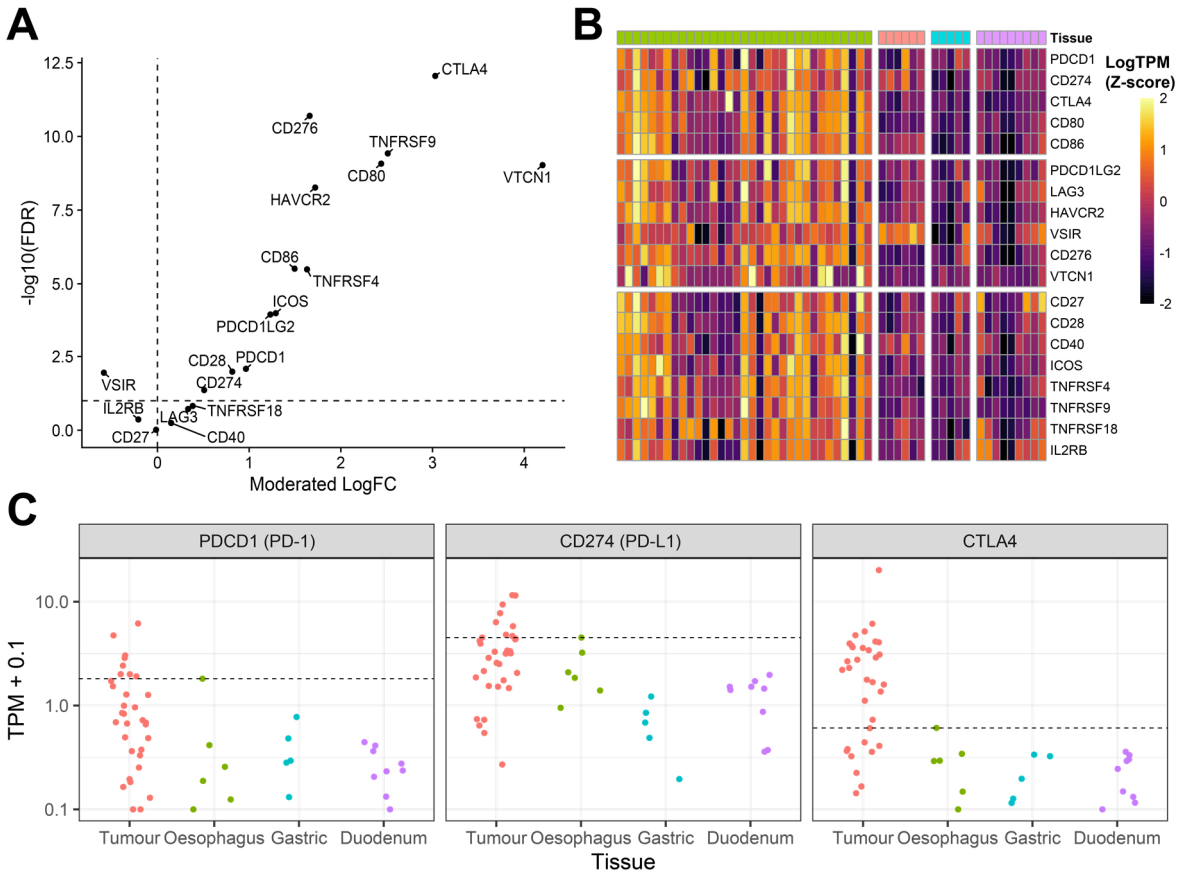


Figure 4.9: Pre-treatment expression of known co-inhibitory and co-stimulatory molecules in OAC. A) The moderated fold change ('ashr' method) and significance (-log10 of the FDR) are shown for the enrichment of known co-inhibitory and co-stimulatory molecules in OAC compared to normal tissues. The CTLA-4 axis (*CTLA4*, *CD80*, *CD86*) are more significantly enriched than the PD-L1 axis (*PDCD1*, *CD274*). Dashed lines show a log2 fold change of zero, and an FDR significance threshold of 0.1. B) Z-scores of TPM expression (log10 base) are shown for the same genes, organized into PD-L1/CTLA-4 pathway genes and other co-inhibitory and co-stimulatory molecules not targeted by this therapy. C) To give a sense of absolute expression of PD-1, PD-L1, and CTLA-4, TPM expression levels are displayed (0.1 offset, y axis plotted on a log10 scale). Dashed lines are placed at the highest expression level from a normal tissue sample.

4.5 Finding novel predictive markers of outcomes using a *de novo* differential expression approach

4.5.1 Pre-treatment samples

4.5.1.1 Clinical benefit

Having summarized the transcriptomic features that distinguish OAC from normal tissues, biomarker analysis was next conducted to determine which features were predictive of outcome in the setting of first-line immunochemotherapy. First, a *de novo* analysis was employed to find genes and pathways that were most predictive of clinical benefit (CB), beginning with pre-treatment (PreTx) OAC biopsies. PreTx biopsies were available from 33 patients, 32 of which were assessable for clinical benefit. 58 DEGs were detected in this analysis, 42 of which were higher in patients with CB, and 16 higher in patients with NCB (Figure 4.10A; Appendix C-2). Gene set enrichment analysis (GSEA) of these results revealed that a number of immune-related pathways were associated with CB (Figure 4.10B). Interestingly, these pathways associated with CB were enriched for innate immune gene sets, rather than the T cell-mediated inflammation typically reported to mark ICI response.

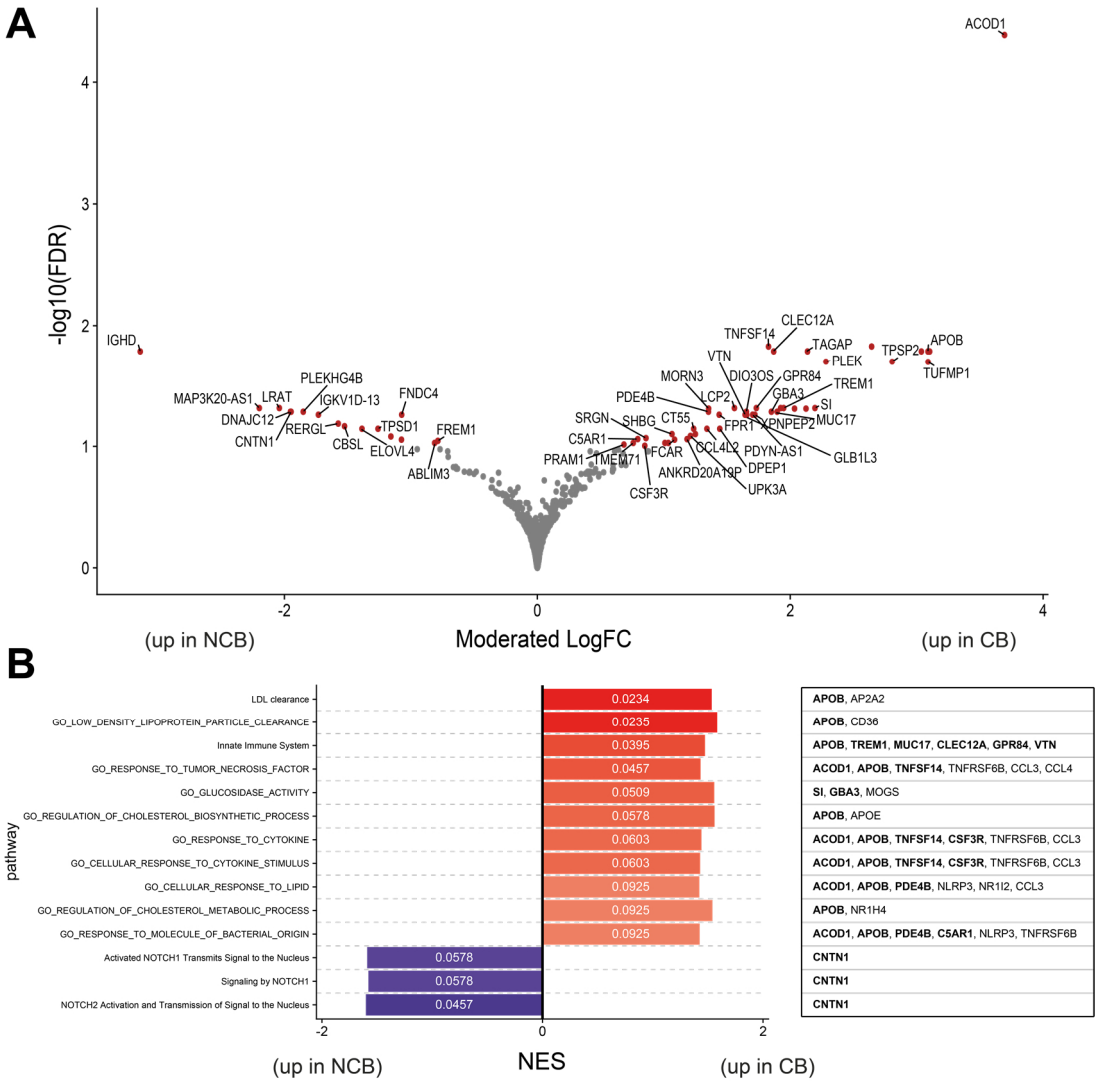


Figure 4.10: Genes and pathways associated with Clinical Benefit (CB) or No Clinical Benefit (NCB) at the pre-treatment timepoint. A) Differential expression analysis between CB and NCB biopsies, showing the moderated log2 fold change (ashr method, "Moderated LogFC") and FDR (-log10 transformed). Significant DEGs (FDR < 0.1) are highlighted in red and labelled. To assist with overplotting, some non-coding genes without informative gene symbols are omitted from labelling. B) Significantly enriched (FDR < 0.1) pathways predictive of CB/NCB. The length of the bar shows the normalized enrichment score (NES), while the colour denotes the FDR, which is also displayed on each bar. Right: leading edge genes for each pathway; individual significant DEGs are bolded.

The specific cell types contributing to this innate immune enrichment were not clear. As none of the general markers assessed in Figure 4.7A were among these DEGs, this did not appear to be driven by general myeloid infiltration. A number of innate chemotactic genes (e.g. *FPR1*, *C5AR1, CSF3R*) were noted among CB genes, but these have key roles in both neutrophils and other myeloid cell types, such as monocytes[148–150]. However, many of these upregulated innate immunity genes in CB fell into pathways for responses to cytokines, TNF, and sensing of bacterial molecules, suggestive of an inflammatory phenotype similar to that induced in myeloid

cells following detection of pathogens. Indeed, the top DEG associated with CB, ACOD1 (Immunoresponsive Gene 1), is a gene predominantly upregulated in myeloid cells that have been activated by inflammatory signals, including infection, cytokines, and pathogen- or damage-associated molecular patterns (DAMPS/PAMPS)[151]. Additionally, two other top hits associated with CB, TNFSF14 and LCP2, have been shown to be strongly upregulated in human monocytes following LPS challenge[152]. Therefore, while cell type specificity was not attainable from this bulk RNA-seq analysis, it appears that a particular inflammatory myeloid phenotype is driving the association of innate immunity pathways with CB.

Less genes and pathways were significantly upregulated in the NCB group, perhaps suggesting a weaker signal in this group or more varied mechanisms of resistance in this setting. Among the top 15 significant pathways identified by GSEA, only three were enriched in NCB compared to CB (Figure 4.10B), all of which were NOTCH signalling pathways driven by a single leading-edge gene (*CNTN1*). However, manual inspection of the DEGs associated with NCB did show some immune-related genes. The top DEG associated with NCB was *IGHD*, encoding the heavy chain of the IgD immunoglobulin isotype associated with naïve B cells. No other isotype-specific constant region genes (i.e. IgA, IgG, IgM, or IgE-specific markers) were significantly associated with CB or NCB. The immunoglobulin variable region gene *IGKV1D-13*, one of the most important tolerance-associated B cell genes in the context of renal transplants[153,154], was also significantly associated with NCB. In summary, while PreTx transcriptomic signatures associated with NCB were more difficult to categorize, the presence of genes associated with antigen-inexperienced or tolerogenic B cells and the absence of inflammatory myeloid markers were predictive of poor outcome in this analysis.

4.5.1.2 irRECIST response

To further assess whether these results were robust predictors of outcome in this setting, differential gene expression was also assessed according to the best overall irRECIST response (BOR) of each patient, another commonly used categorical classification system in biomarker studies. While irRECIST classifications are not necessarily linked with long-term survival, they do give complementary additional information on how the size of the tumour responds, and are

less open to potential bias, as it relies on a pre-determined set of criteria. Therefore, differential gene expression was also assessed according to BOR, and the findings were compared to the predictive markers extracted from the CB analysis. Within the inoperable OAC cohort, DEGs were calculated between the 16 patients with available material who attained an irRECIST response (responders, R [irCR/irPR]) and the 17 patients who did not (non-responders, NR [irSD/irPD]).

142 genes were upregulated in R at the PreTx timepoint, and 282 were upregulated in NR (Appendix C-3). Assessment of the top DEGs and significant pathways again showed suggested enrichment of innate immune activity in the responding patients (Figure 4.11A-B), showing good concordance with the CB analysis. Indeed, a number of innate immune genes, particularly those related to the myeloid compartment (*FPR1, C5AR1*), were significantly upregulated in both R and CB (Figure 4.11C). That these genes are predictive of both tumour shrinkage and long-term disease control show that they are robust predictors of positive outcome with immunochemotherapy for this cohort, and are not an artefact of the post-hoc threshold set for CB. Contrastingly, R also showed enrichment for squamous keratinocyte pathways and related genes (*DSG1, KRT14, KRT17, FLG*), which were not enriched with CB (Figure 4.11, 4.12). Further characterisation of these squamous genes showed that, while they were associated with higher rates of response, they were not good predictors of OS, and in multivariable analyses even predicted worse survival within BOR groups (Figure 4.12B). This was in contrast to innate immune markers, which showed more robust prediction of both survival and response outcomes on immunochemotherapy in this cohort.

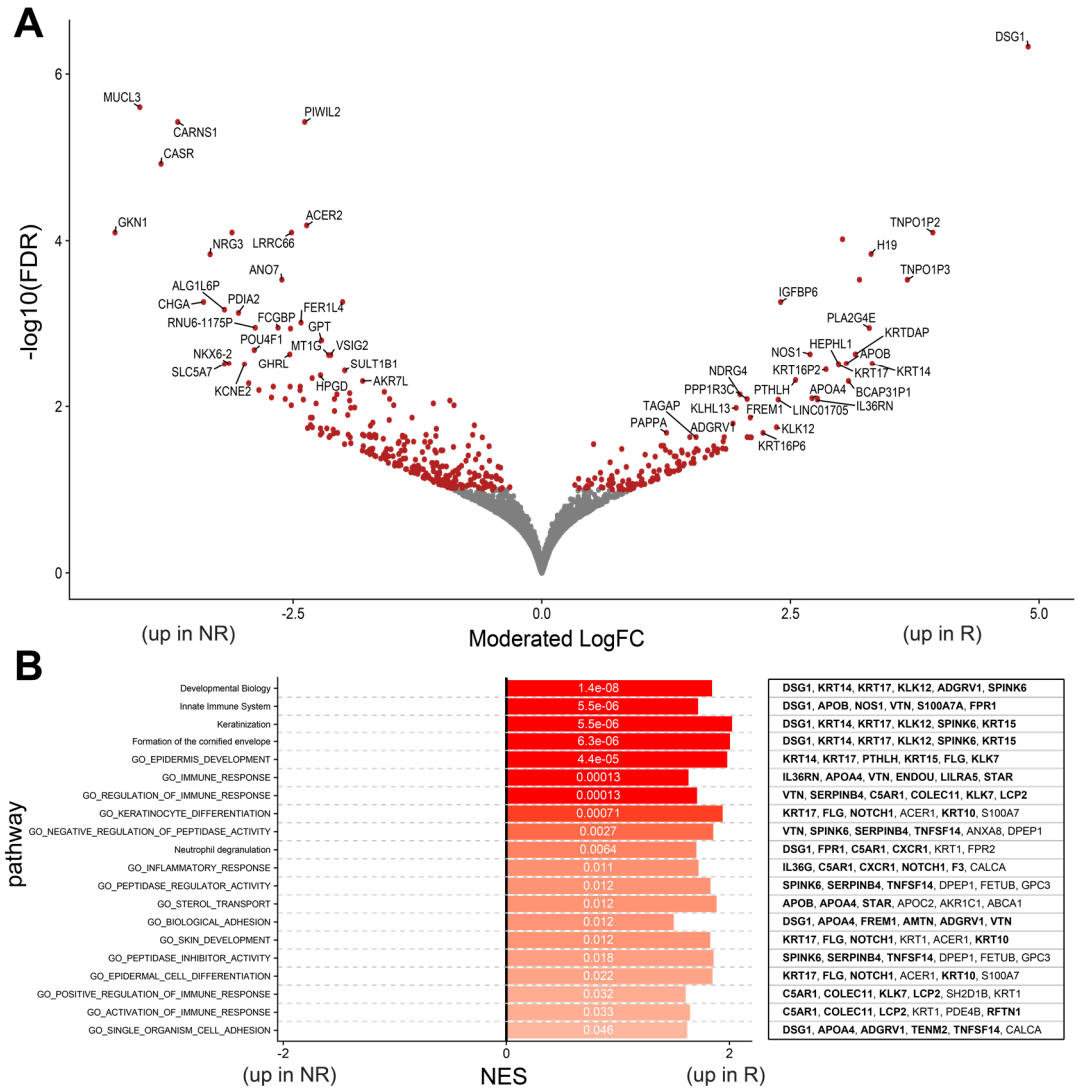
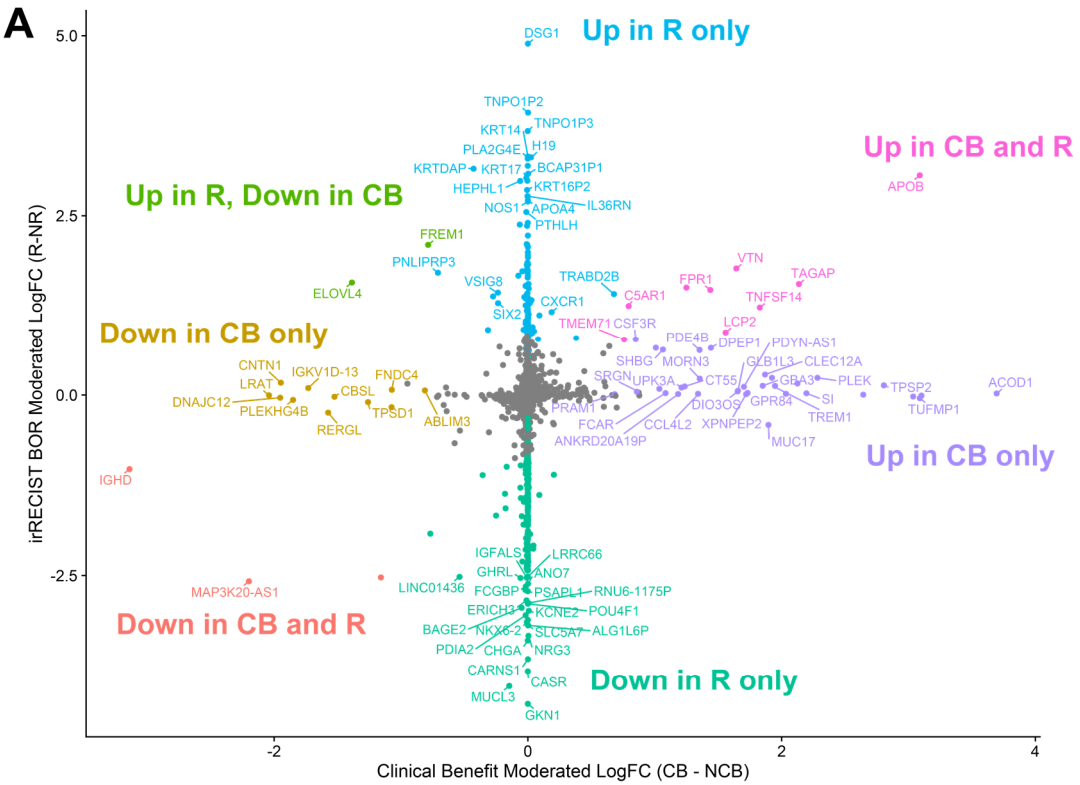


Figure 4.11: Differential expression by BOR shows innate immunity and squamous differentiation genes are associated with irRECIST response to therapy. A) Differential expression analysis between irRECIST R and NR biopsies, showing the moderated log fold change and FDR (-log10 transformed). Significant DEGs (FDR < 0.1) are highlighted in red, and the top 30 DEGs upregulated in R and NR are labelled. B) Significantly enriched (FDR < 0.1) pathways predictive of R/NR. The length of the bar shows the normalized enrichment score (NES), while the colour denotes the FDR, which is also displayed on each bar. Right: leading edge genes for each pathway; individual significant DEGs are bolded.



B

Category	Example Gene	Univariate (OS)		Multivariate (OS; w/ BOR)	
		HR (95% CI)	p.value	HR (95% CI)	p.value
Keratinization (BOR only)	DSG1	1.08 (0.73-1.60)	0.695	**1.65 (1.01-2.72)**	**0.047**
Innate Immunity (BOR + CB)	FPR1	**0.65 (0.43-0.99)**	**0.044**	0.73 (0.47-1.12)	0.15
Innate Immunity (CB only)	ACOD1	**0.54 (0.30-0.99)**	**0.046**	0.61 (0.34-1.09)	0.096

Figure 4.12: Comparison of predictive biomarkers returned by clinical benefit and irRECIST BOR groupings. A) Comparison of moderated log fold changes between BOR analysis (y-axis) and clinical benefit analysis (x-axis). Coloured genes are significant (FDR < 0.1) in one or more analysis, as labelled. D) Cox regression results for example genes significant in BOR, CB, and BOR & CB analyses, showing some BOR markers are not good predictors of survival. First results column shows univariable Cox regression with OS for each gene, and the second results column shows the gene-specific results for a multivariable analysis which also includes whether or not the patient achieved an irRECIST response. Significant results (p.value < 0.05) are bolded.

Similar to the NCB group, there were not many pathways significantly enriched in NR, again suggesting increased heterogeneity among patients with poor outcome. However, a number of genes associated with gastric (*GKN1, MUCL3*) and neuroendocrine cells (*CHGA, GHRL, NRG3*) were among the top DEGs associated with NR, although these were not associated with NCB (Fig 4.12A). *IGHD* was enriched in both NR and NCB, suggesting this marker of antigen-naïve B cells could be a robust predictor of poor outcome on immunotherapy, as judged by both tumour shrinkage and long-term disease control.

In summary, this comparison shows limited overlap between DEGs that predict BOR and CB classifications in this cohort. These were mainly confined to a subset of innate immune genes which predicted positive outcome in both systems, and *IGHD* which consistently predicted negative outcome. These results highlight the potential limitations of using BOR classifications to find predictive markers associated with survival, and provides evidence that the top hits returned by the CB system may be more suitable for predicting long-term outcome in this setting. Therefore, CB classification will be used for biomarker analysis at the other timepoints assessed in this section.

4.5.2 On-treatment samples (ICI-only)

Determining predictive markers at an early on-treatment timepoint could help identify patients likely to progress in advance of radiographic detection of progression, and could help show early mechanisms of resistance that could be targeted by supplemental therapy. 28 patients (12 CB, 15 NCB, 1 not assessable) had OAC biopsies collected at the end ICI-only window. ICI-only biopsies were taken approximately 4 weeks after treatment onset (median 27 days, IQR 26-30 days, range 18-63 days), all prior to the onset of the chemotherapy. Comparing CB and NCB transcriptional profiles revealed 158 DEGs, 80 upregulated in CB and 78 in NCB (Figure 4.13A; Appendix C-4).

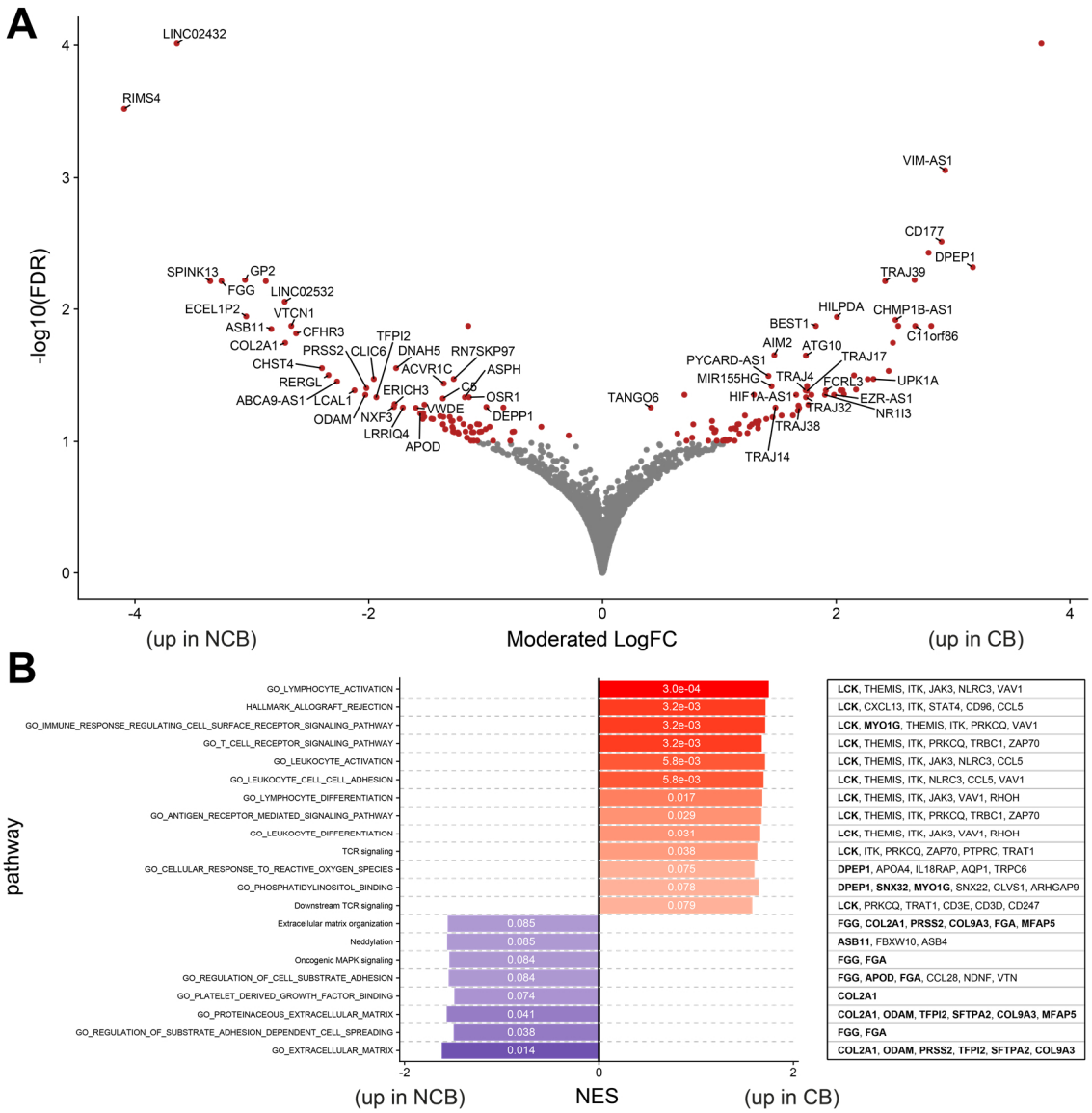


Figure 4.13: Differential expression at ICI-only timepoint shows enrichment of T cell signalling in CB patients, and stromal genes in NCB patients. A) Differential expression analysis between CB and NCB biopsies at the ICI-only timepoint. Significant DEGs (FDR < 0.1) are highlighted in red, and the top 80 genes with informative gene symbols are labelled. B) Significantly enriched (FDR < 0.1) pathways predictive of CB/NCB. The length of the bar shows the normalized enrichment score (NES), while the colour denotes the FDR, which is also displayed on each bar. Right: leading edge genes for each pathway; individual significant DEGs are bolded.

In contrast with PreTx biopsies, many genes associated with T cell response were among the DEGs significantly associated with CB after 4 weeks of ICI (Figure 4.13A; Appendix C-4). These included 11 TCRα joining region components (TRAJ), suggesting enriched TCR expression in CB across multiple T cell clonotypes, or general increased infiltration of T cells at this timepoint. *Lck*, which phosphorylates the T cell receptor (TCR) to initiate the TCR signalling pathway upon antigen binding[155,156], was also significantly enriched in CB. DEGs associated with NCB at this

timepoint included genes associated with complement (C5), clotting (*FGG*, *FGA*, *TFPI2*), and the extracellular matrix (*COL2A1*, *PRSS2*, *COL9A3*). These suggest that on-treatment enrichment of tumour-associated stroma and the complement/clotting cascade could play a role in resistance to the ICI component of therapy.

Comparing DEGs from this timepoint to those from the PreTx analysis, *DPEP1*, a cell-surface protein that can function as an adhesion receptor mediating neutrophil tissue infiltration[157], was the only gene associated with CB at both timepoints. Similarly, only *RERGL*, a less well-characterized member of the RAS GTPase superfamily[158], was associated with NCB at both PreTx and ICI-only. In summary, there was limited overlap in potentially useful predictive markers between these two timepoints, suggesting significant transcriptional changes to the tumour and its microenvironment during the ICI-only phase of treatment in this study.

Category		ICI-only		
		Down in CB	NS	Up in CB
PreTx	Up in CB	0	42	**1 *(DPEP1)***
	NS	77	27368	79
	Down in CB	**1 *(RERGL)***	15	0

Table 4.3: Overlap between DEGs between PreTx and ICI-only timepoints. DEGs (FDR < 0.1) at each timepoint are compared to detect genes predictive of CB status at either timepoint. Only *DPEP1* and *RERGL* are predictive of CB and NCB (respectively) at both timepoints. NS = not significant.

4.5.3 Post-treatment samples (PostTx)

Post-treatment (PostTx) biopsies are not routinely taken to evaluate response to a course of therapy in advanced OAC; however, sampling at the conclusion of therapy could potentially be useful in identifying patients likely to maintain long-term disease control, as well as patients likely to progress imminently that may benefit from additional surveillance or therapy. At the conclusion of the immunochemotherapy protocol used in this trial, 17 patients from the inoperable OAC cohort remained on study and had biopsies available for molecular profiling. These biopsies were taken approximately 5-6 months after the patient's pre-treatment biopsy (median 5.3 months, IQR 5.2-5.9 months, range 5.0-6.9 months). DEGs were assessed at this timepoint between the patients who would go on to attain CB versus those that would not (CB [n = 10] vs. NCB [n = 6]; 1 un-assessable). 119 DEGs were significantly higher in CB patients at this timepoint, and 342 were higher in NCB (Figure 4.14A; Appendix C-5). GSEA was dominated by enrichment of translation, ribosomal, and rRNA processing pathways among CB biopsies (Figure 4.14B).

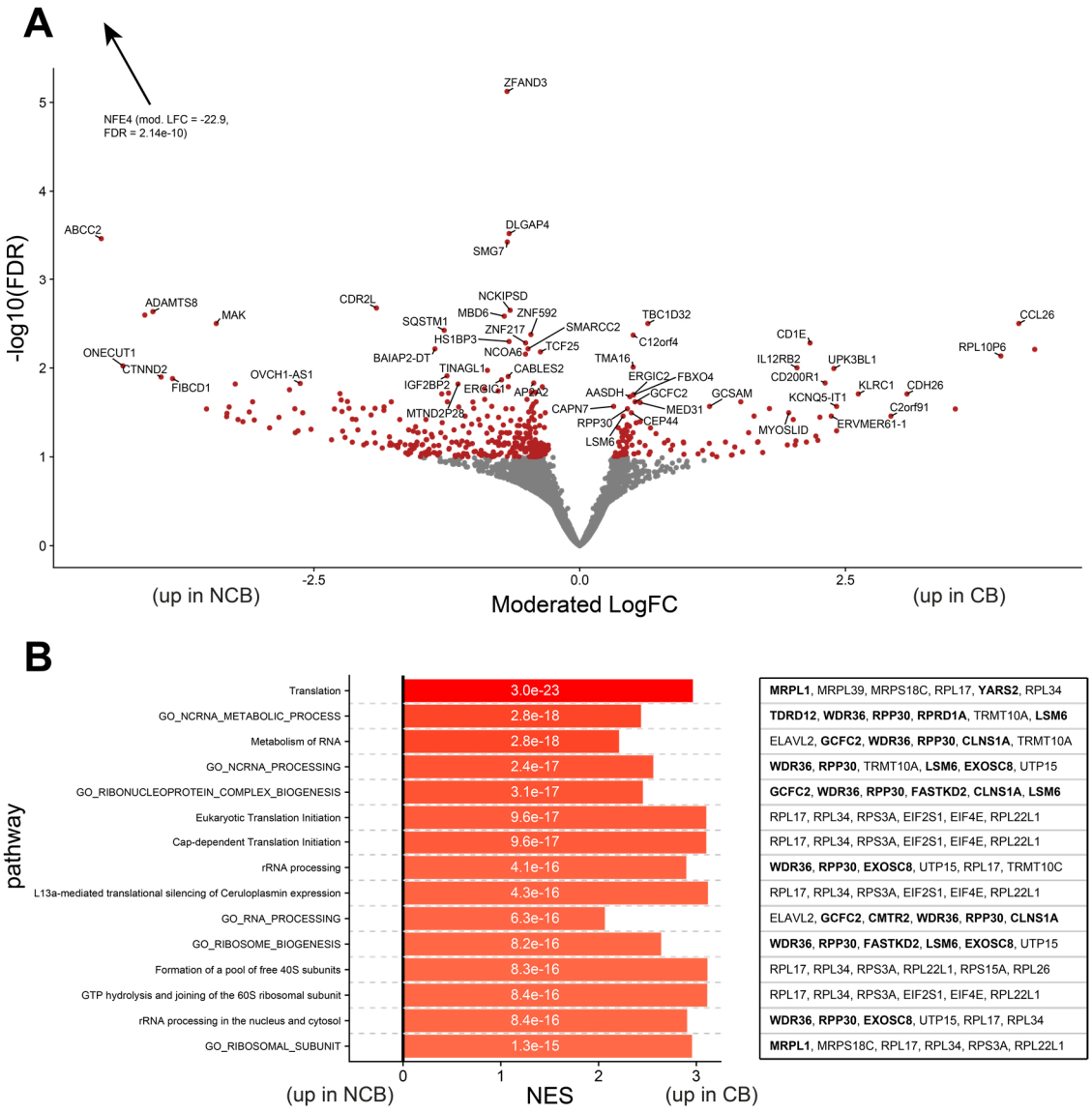


Figure 4.14: Detection of markers predictive of clinical benefit at post-treatment (PostTx) biopsy. A) Differential expression results between CB and NCB, showing moderated log fold change (moderated logFC) and FDR. Significant DEGs are highlighted in red; and the top 60 DEGs (30 up in CB, 30 up in NCB) are labelled. The top DEG (*NFE4*) has a much more extreme FDR and fold change than other DEGs and is cropped from the plot to maintain legibility, but the FDR and moderated logFC are given on the plot. B) Top 15 significantly enriched pathways (FDR < 1e-16), all enriched in CB. The length of the bar shows the normalized enrichment score (NES), while the colour denotes the FDR, which is also displayed on each bar. Right: leading edge genes for each pathway; individual significant DEGs are bolded.

As GSEA was dominated by a single cluster of related pathways, DEGs were manually assessed to find additional pathways potentially linked with response or resistance to therapy at this timepoint. Given the PCA analysis in 4.4.1, where oesophageal-like tumour site biopsies have a non-significantly higher rate of CB than biopsies that have more broad transcriptional similarities to PreTx tumours, one hypothesis could be that the maintenance of a malignant epithelial markers at site of the primary lesion could predict poorer outcome. Indeed, NCB

biopsies do maintain significantly higher transcriptional levels of certain OAC-specific CT antigens (*PRAME*, *POTEE*) in the post-treatment biopsy (Table 4.4). However, no significant difference was seen in markers of squamous epithelium (TP63, KRT14) in this analysis (FDR = 0.44 and 0.73, respectively). Therefore, while residual markers of malignant epithelium can predict NCB, the return of markers of normal squamous oesophagus is not significantly linked with CB.

In addition to these epithelial differences, various immune cell markers were significantly upregulated in CB patients compared to NCB at PostTx. Among T and NK cell markers, *ITGAE* (CD103) and *KLRC1*, whose co-expression has been shown in CD8+ tissue-resident memory T cells (T_{RM}) in lung cancer[159], are both upregulated in CB (Figure 4.15; Table 4.4). ICI-only appeared to upregulate these genes in many patients, but their upregulation did not differentiate CB from NCB until the PostTx timepoint (Figure 4.15). Members of the CD1 family (*CD1C*, *CD1E*), which participate in lipid antigen presentation to T cells[160], were also upregulated in CB. These immune markers suggest that the maintenance of infiltrating T_{RM} cells at the end of immunochemotherapy, as well as certain specialized antigen presentation pathways, could help predict long-term outcome at this timepoint. Among NCB DEGs at this timepoint were two ABC transporter genes, which promote multidrug resistance through efflux of chemotherapy substrates: *ABCC2* (Multidrug resistance protein 2/MDR2) and *ABCG2* (Breast cancer resistance protein/BCRP) (Table 4.4). While both have been reported to predict poor prognosis in oesophageal cancer[161], *ABCC2* is of particular interest due its known role in limiting intracellular oxaliplatin concentrations and subsequently promoting resistance to oxaliplatin-containing chemotherapy regimens in gastrointestinal cancers[162,163].

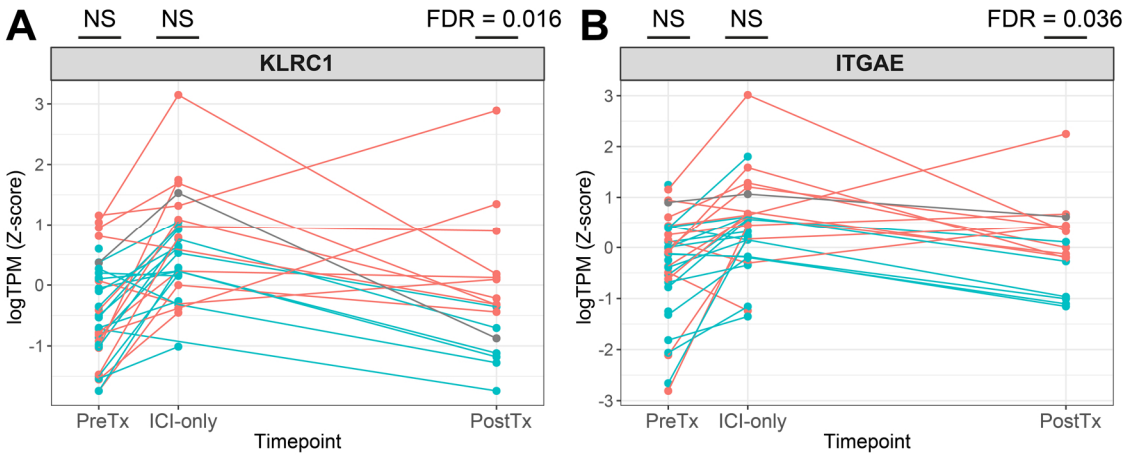


Figure 4.15: Change in selected immune markers over the course of immunochemotherapy. Z-scaled TPMs (log10 scale, 0.1 offset) are shown for A) *KLRC1* and B) *ITGAE*, with lines connecting results for tumour biopsies from the same patient over the course of therapy. Above the plot, results for differential expression between CB and NCB at each timepoint are shown. NS: not significant (FDR > 0.1). X-axis is scaled to represent the average time between the different sampling timepoints. Red colour denotes CB patients, and blue NCB.

To further explore the utility of these markers in predicting survival after conclusion of therapy, the link between the expression of these markers and OS after the PostTx biopsy were assessed, alongside more traditional radiological measures of tumour response (Table 4.4). For radiological measures, there did appear to be a trend connecting greater burden of residual tumour after therapy with reduced OS after the PostTx timepoint, but neither this measure nor the percentage change in tumour size from the pre-treatment measurement reached significance. Contrastingly, many of the transcriptional markers profiled above were more robust predictors of OS after the PostTx timepoint, particularly for the immune markers, ribosomal and RNA translation genes, and *ABCC2*.

In summary, molecular profiling of PostTx biopsies returned a number of genes that were associated with CB, suggesting that biopsies at this timepoint could help predict a disease-control outcome that was still 6-7 months from the time of sampling. Therefore, tumour biopsies following a course of immunochemotherapy could supply valuable information that could potentially be used alongside radiological measures to provide improved predictions of long-term outcomes. These data suggest PostTx sampling should be considered to help inform decisions about patient care following first-line immunochemotherapy in this setting.

Predictive Variable	DE (CB vs. NCB)		OS (after PostTx)	
	log2FC	FDR	HR (95% CI)	p
Radiological Response				
Target lesion size (SumD)	-	-	1.69 (0.95-3.00)	0.072
% Change from baseline	-	-	0.96 (0.47-1.96)	0.908
Tumour Markers/CT Antigens				
POTEE	**-3.06**	**0.038**	1.47 (0.89-2.42)	0.135
PRAME	**-2.92**	**0.047**	**1.94 (1.05-3.57)**	**0.034**
Squamous Epithelium Markers				
TP63	0.14	0.482	**0.47 (0.22-0.97)**	**0.04**
KRT14	0.06	0.761	0.64 (0.32-1.28)	0.206
T/NK Cell Markers				
KLRC1	**2.62**	**0.020**	**0.21 (0.06-0.70)**	**0.011**
ITGAE (CD103)	**0.57**	**0.040**	**0.24 (0.08-0.73)**	**0.012**
CD1 (Lipid Antigen Presentation)				
CD1E	**2.17**	**0.005**	**0.25 (0.09-0.70)**	**0.008**
CD1C	**0.83**	**0.065**	**0.36 (0.17-0.80)**	**0.012**
Translation/rRNA processing				
GCFC2	**0.52**	**0.024**	**0.39 (0.17-0.92)**	**0.031**
WDR36	**0.46**	**0.044**	**0.34 (0.15-0.77)**	**0.009**
MRPL1	**0.44**	**0.068**	**0.46 (0.23-0.93)**	**0.032**
ABC Transporters (Multidrug Resistance)				
ABCC2 (MRP2)	**-4.50**	**3.4E-04**	**1.82 (1.00-3.32)**	**0.049**
ABCG2 (BCRP/MXR)	**-1.52**	**0.053**	2.03 (0.98-4.21)	0.057

Table 4.4: Potential utility of selected PostTx transcriptional markers in predicting outcome after conclusion of therapy. Top: The link between radiological measures of tumour size, including the absolute size of target lesions at the PostTx timepoint (sum of diameters) and the change in this measurement from the pre-treatment baseline, and survival outcomes are shown, using a univariable Cox regression. OS for this section was calculated from the PostTx CT scan, rather than treatment onset. Target lesion burden was equivalent to total measurable tumour burden (TMTB) for all patients at this timepoint. Below: moderated log-fold change and FDR from differential expression (DE) analysis are shown from the comparison between CB and NCB at the PostTx timepoint. Cox regression results between expression of the given gene (log10 TPM) with OS (from date of PostTx biopsy) are shown. Significant results (DE FDR < 0.1; Cox p value < 0.05) are bolded. All variables were scaled and centered prior to Cox regression.

4.6 Profiling changes in transcriptional profiles to elucidate mechanisms of ICI activity

4.6.1 Identification of an ICI-induced transcriptional response signature (INCITE) and correlation with outcome

4.6.1.1 Definition and utility of the INCITE signature

In addition to assessing static markers of outcome at each timepoint, the sequential sampling protocol implemented in the LUD2015-005 study allowed for the study of transcriptional evolution in response to therapy. Given the unique study design of the LUD2015-005 trial, one important question to address was whether the four-week ICI-only window at the beginning of treatment was sufficient to induce strong T cell infiltration and activity prior to initiation of chemotherapy. For patients with paired biopsies available at PreTx and ICI-only timepoints (n = 28), transcriptomic changes induced during the ICI-only window were assessed, controlling for patient-specific effects. Across all 28 patients, 547 DEGs were significantly upregulated at the ICI-only timepoint compared to PreTx, and 146 were significantly downregulated (Appendix C-6). The top 70 significant DEGs were all upregulated at ICI-only, and were dominated by cytotoxic and immune response genes (Figure 4.16A), including T cell chemokines (*CXCL9, CXCL10, CXCL11*), T/NK cell markers (*CD2, CD3D/E*, *CD8A/B*, Killer cell lectin-like receptor (KLR) family, *NKG7*, *TRGC2*), cytotoxic effector molecules (*GZMA, GZMH, GZMK, PRF1, FASLG*), and markers of $CD8^+$ T_{RM} cells (*ITGAE*/CD103, *ZNF683*/Hobit). Interestingly, some immunosuppressive and immunoregulatory components (*PDCD1*/PD-1, *IDO1*, *CD244*/SLAMF4) were also among the top genes upregulated by therapy. This could represent resistance to the induced inflammation; however, these molecules are often upregulated following functional TCR engagement[164], and so could simply be a further readout of successful immune activation. While not among the top 70 DEGs, interferon gamma (*IFNG*), a key regulator of anti-tumoural inflammatory responses, was significantly upregulated during ICI-only (p.adj = 0.010). Nearly all patients showed upregulation of these top 70 genes during the ICI-only period, regardless of whether they attained CB (Figure 4.16B). These data provide evidence that intratumoural cytotoxic activity is increased in most OAC patients during the first four weeks of ICI therapy.

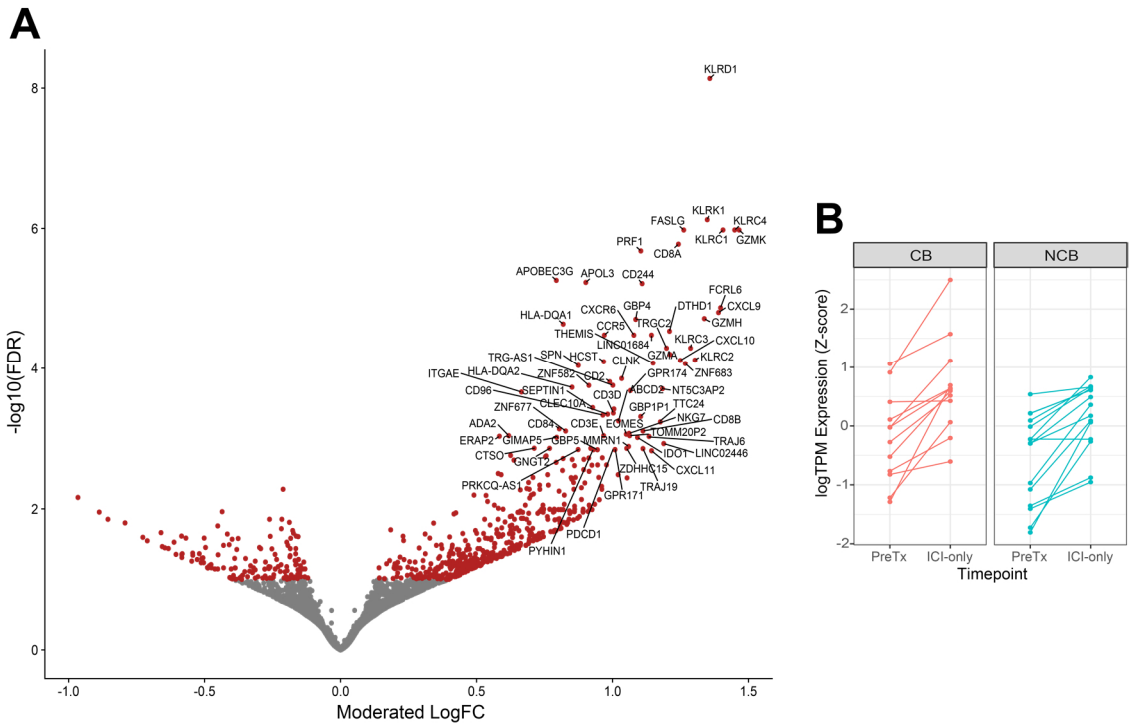


Figure 4.16: Robust cytotoxic activity is induced by 4 weeks of ICI across all patients. A) Differential expression results between ICI-only and PreTx biopsies are shown, regressing out patient-specific effects (design = ~patient + timepoint). Moderated log-fold changes and FDR are shown. The top 70 DEGs, many of which are related to cytotoxic cell activity, are labelled. B) Sum of z-score normalized log10 TPM expression values (0.1 offset) for the 70 top DEGs labelled in A. Patients are grouped by clinical benefit (left) and no clinical benefit (right). PreTx and ICI-only values for each patient are connected by a single line.

To further compare these immune-related transcriptional changes with patient outcome, these top 70 upregulated DEGs were used to define as a gene set termed INCITE (ImmuNe Checkpoint Inhibitor-induced T cell Enrichment). The full list of INCITE genes are provided in bold in Appendix C-6. While nearly all patients experienced upregulation of INCITE genes to some extent during the ICI-only period (Figure 4.16B), there was a wide range in the magnitude of upregulation of these genes, which did not appear to be linked with overall CB (Figure 4.17A). However, patients with strongest INCITE upregulation during the ICI-only window did appear to be more likely to demonstrate an early irRECIST response, particularly for those in the top third of INCITE upregulation (10/28, "INCITE-reactive"). This group included all 3 patients who responded at the CT scan marking the end of the ICI-only period (C1D8). Radiological responses at this timepoint are likely to be primarily driven by ICI, as chemotherapy will only have given for approximately one week at this scan. Four additional INCITE-reactive patients demonstrated response at the following scan during the middle of the immunochemotherapy period (C3D8).

There were three INCITE-reactive patients who did not attain an irRECIST response: one was non-assessable for target lesion size throughout (irNN/irSD) but nonetheless attained clinical benefit, another had stable disease throughout (no change at C1D8) but nonetheless attained clinical benefit, and the third showed mild tumour shrinkage at C1D8 (~9%) but died as a result of adverse effects from treatment, specifically the chemotherapy component, before tumour shrinkage could be assessed at C3D8. Regardless, the enrichment for early responders in this group suggested a possible link between INCITE reactivity and ICI-induced anti-tumour activity. Indeed, the extent of upregulation in INCITE genes was significantly correlated with tumour shrinkage during the ICI-only window (Figure 4.17B-C). Interestingly, the lone MSI-high tumour showed the greatest extent of both INCITE gene upregulation and tumour shrinkage (~40%) during this window. Therefore, strong upregulation of this ICI-response gene set after the first few weeks of treatment is a good transcriptional marker of patients who are able to generate strong anti-tumour T cell-mediated responses in response to ICI.

However, a number of INCITE-quiescent patients responded, all at the C3D8 timepoint or later. This suggests that the chemotherapy component likely plays an important role in generating an effective anti-tumour response in some patients, and that the lack of a strong T cell response during the first few weeks of ICI monotherapy does not necessarily preclude eventual response to immunochemotherapy. Indeed, overall CB in this cohort was not associated with INCITE upregulation, and no survival differences were noted between INCITE-reactive patients and INCITE-quiescent patients who nevertheless responded during the immunochemotherapy period, with both groups showing improved survival over INCITE-quiescent non-responders (Figure 4.17D-E). In summary, greater INCITE reactivity appears to be a feature of OAC patients who have the strongest responses to the ICI-only phase of this treatment protocol. However, this metric has limited utility in predicting all OAC patients who will respond or maintain long-term disease control on the full course of immunochemotherapy. These data suggest that some fraction of patients may be able to derive a response from ICI alone in this setting, while others may need the chemotherapy component or its combination with ICI to generate a response.

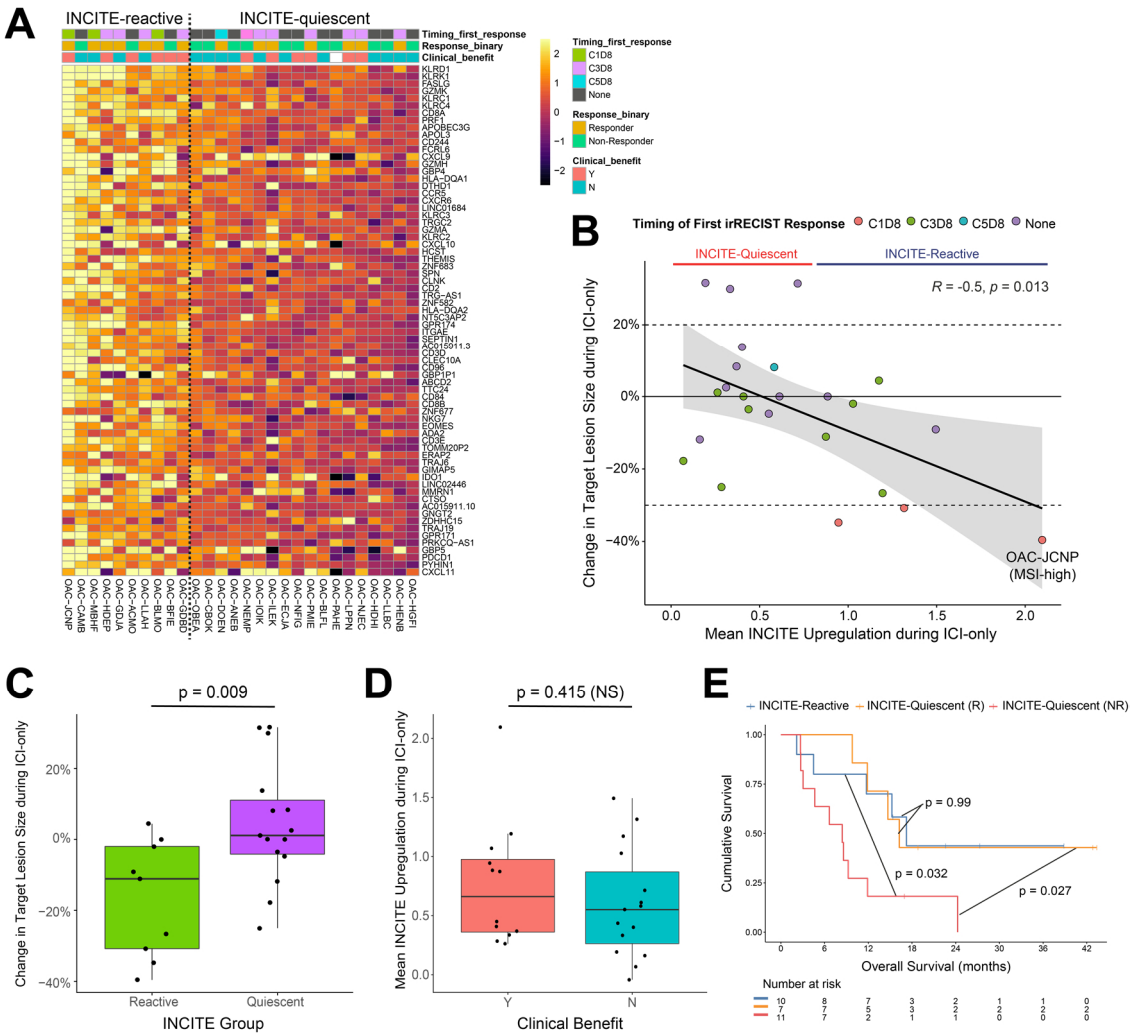


Figure 4.17: Stronger induction of INCITE genes during the ICI-only period is linked with early tumour shrinkage, but not with overall outcome on immunochemotherapy. Variance-stabilizing transformation (VST) normalized expression was used rather than logTPM in section, as this transformation accounts for the dependence of variance on mean expression, and therefore returns more homoscedastic values that are more appropriate for summarizing differences in gene expression across a wide range of expression levels. The VST transformation used herein (blind = F) also uses the same design formula used in 4.15 to estimate gene dispersion, thereby taking patient-specific differences into consideration. A) Changes in INCITE genes were calculated as the difference between VST expression values at ICI-only and PreTx for each patient. Patients (columns) were ordered according to the average magnitude in these VST expression changes (scaled without centering so that each gene contributes similarly while preserving the direction of the ICI-only – PreTx changes). Samples with the highest level of increased expression were deemed "INCITE-reactive". Hierarchical clustering was performed on rows (ward.D linkage). B) The mean of VST changes (scaled without centering) of INCITE genes (used to order A) was compared to the percentage change of the tumour at the C1D8 scan from pre-treatment baseline (negative % change means tumour shrinkage). Points are coloured by the timing of first reported irRECIST response, and the MSI-high patient (OAC-JCNP) is highlighted. C) Tumour shrinkage measures from (B) are summarized as a boxplot, grouped by the INCITE classification introduced in (A). A p-value calculated by the Welch's t-test shows the significance of the increased tumour shrinkage experienced during the ICI-only phase by INCITE-reactive patients. D) The average of z-scores for the VST change of INCITE genes is compared by clinical benefit (p value by Welch's t-test). E) OS of INCITE-reactive patients was compared to INCITE-quiescent patients with R and NR. Pair-wise log-rank test p values (unadjusted) are shown for the difference between each group, showing decreased survival for INCITE-quiescent non-responders versus INCITE-reactive and INCITE-quiescent responding patients.

4.6.1.2 Genes that predict INCITE-reactive and INCITE-quiescent response subtypes

As INCITE-reactive patients show significantly stronger early radiological responses to ICI-only therapy, predictive markers of INCITE reactivity may be of interest to identify patients that would be likely to have good outcome on ICI monotherapy. Similarly, among INCITE-quiescent patients, there may be transcriptional differences between patients that will go on to respond to immunochemotherapy and those that will not. Therefore, a likelihood ratio test (LRT) was conducted to find DEGs between INCITE-reactive patients, INCITE-quiescent responders, and INCITE-quiescent non-responders at the PreTx timepoint. 341 significant DEGs (FDR < 0.1) were detected by this LRT method.

Similar to the DEGs differentiating overall irRECIST responders from non-responders in 4.5.1.2, epithelial markers differentiated INCITE-reactive Pre-Tx biopsies from the INCITE-quiescent non-responders (Figure 4.18A). Markers of squamous epithelium, particularly suprabasal keratins (*KRT1*/*KRT10*), marked INCITE-reactive patients at baseline. The presence of these squamous keratins raised a concern about potential contaminating normal squamous oesophagus cells that could be confounding results. However, markers like KRT10 were actually expressed at much higher levels in the few INCITE-reactive OACs driving the enrichment of this gene than in normal oesophageal biopsies (Figure 4.18B), suggesting this could represent a poorly differentiated or adenosquamous subtype of OAC tumours that express aberrant levels of this suprabasal keratin. These INCITE-reactive tumours were also enriched for the gene encoding the tumour-specific antigen CA-125 (*MUC16*) and inflammatory mediators, including *NOS1* and the interferon-inducible genes *AIM2* and *ILF3*[165,166], all of which could help drive the inflammatory response to ICI seen in these patients.

Contrastingly to the squamous keratins seen in the INCITE-reactive group, INCITE-quiescent non-responders were enriched for gastric tissue markers (*GKN1*/Gastrokine 1; *CBLIF*/Gastric Intrinsic Factor; *LIPF*/Lipase F, gastric-type; gastric neuroendocrine markers *CHGA*/Chromogranin A and *GHRL*/Ghrelin). Some markers of metaplastic epithelial cell types seen in BO (*LEFTY1*; *SPINK4*[32]) were also enriched in this group. *IGHD* was also found to be a marker of the INCITE-quiescent non-responder cluster, suggesting a link between gastric or Barrett's-like differentiation, an

antigen-naïve B cell phenotype, and poor outcomes. On the other hand, IgM and IgG1 heavy chain genes, along with a variety of other immunoglobulin genes in the variable and constant regions, were more associated with INCITE-quiescent responders (Figure 4.18A). Therefore, while all these B cell markers appeared to predict muted INCITE responses during the ICI-only window, the presence of certain B cell antibody isotypes may help differentiate between INCITE-quiescent patients likely to obtain responses during the immunochemotherapy phase (IgG1/IgM) and those likely to exhibit continued resistance (IgD).

Interestingly, the innate markers identified in the PreTx comparisons as predictive of overall clinical benefit were not significant in this analysis, suggesting the inflammatory genes predictive of INCITE reactivity followed a different pattern from the innate inflammation genes predictive of long-term outcome on immunochemotherapy. For instance, the interferon-stimulated gene *ILF3* was specifically enriched in INCITE-reactive patients at the PreTx timepoint, whereas *ACOD1*, the DEG most associated with CB in section 4.5.1.1, appeared to mark clinical benefit patients regardless of their INCITE group. (Figure 4.18C-D). These findings suggest that the innate immune signature predictive of CB is likely contributed to at least in part by the chemotherapy component of treatment, as individual innate immunity genes could not predict INCITE reactivity in this analysis. The power of this three-way analysis is limited by the small numbers in each subgroup, but gives further insight into predictive markers that could help differentiate patients responding to ICI alone or immunochemotherapy, as well as those that are more likely to resist both phases.

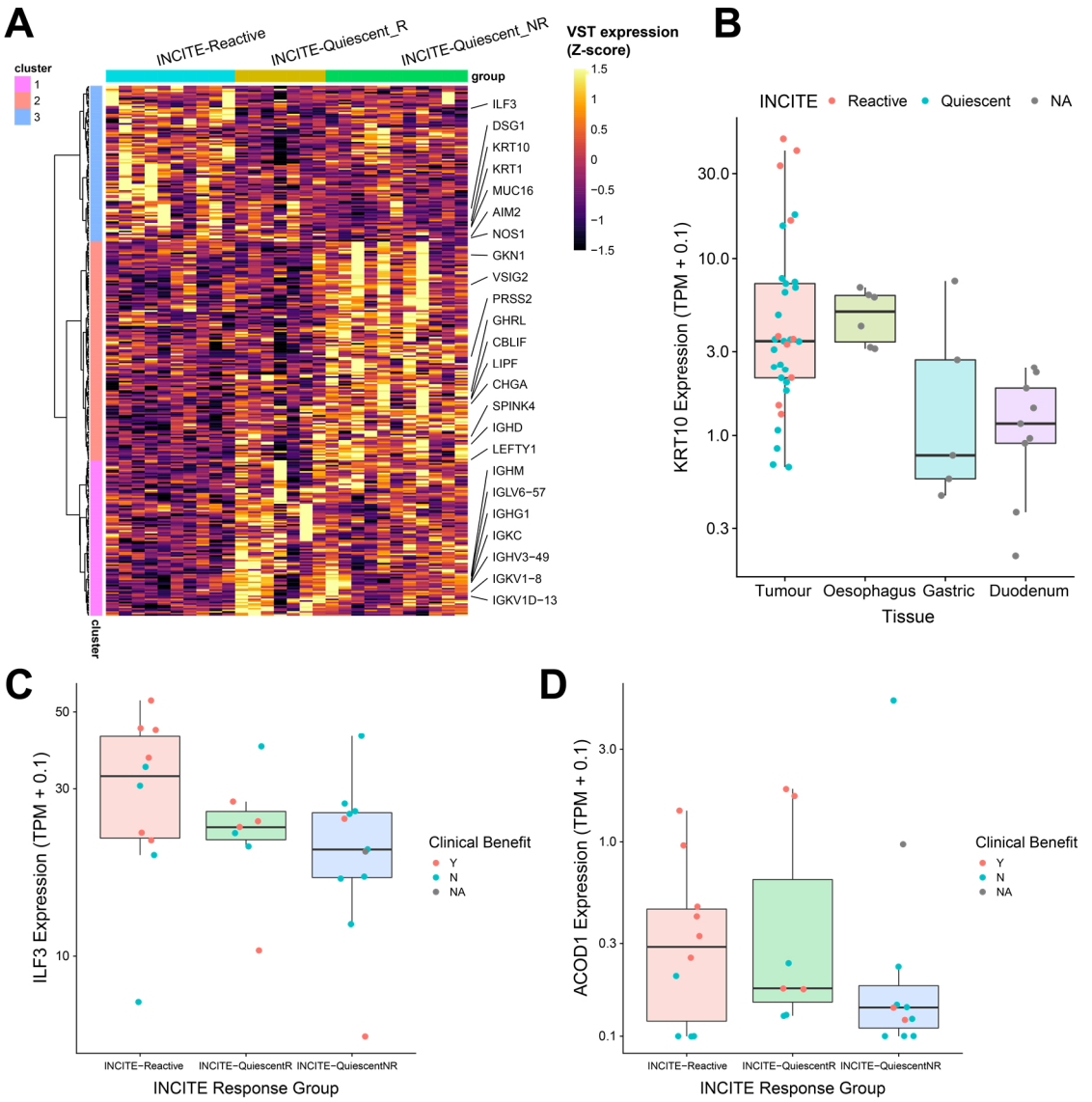


Figure 4.18: Differentially expressed genes according to INCITE/Response groupings. A) Genes with an FDR < 0.1 on a likelihood ratio test between INCITE-reactive, INCITE-quiescent responders (R), and INCITE-quiescent non-responders (NR) are shown. Z-scores of normalized expression (VST) are used to colour each cell, and hierarchical clustering was used to order rows, and the 3 major branches of this hierarchical clustering were used to define gene clusters. B) PreTx expression of KRT10 (TPM, 0.1 offset) are plotted according to different tissue types at baseline. Points are coloured according to whether or not the tumour showed INCITE reactivity. Normal biopsies and biopsies without paired PreTx/ICI-only biopsies are coloured grey. Bottom: TPM expression of C) *ILF3* and D) *ACOD1*, grouped by their INCITE reactivity/response groups. Points are coloured by whether the patient attained long-term disease control (clinical benefit).

4.6.2 Differences in ICI-induced transcriptome changes between CB and NCB patients

INCITE reactivity performs well in marking early response to the ICI component of treatment, but it does not robustly link with overall CB, perhaps due to differences between tumours likely to respond to ICI-only and those likely to respond to immunochemotherapy. To more directly test for ICI-induced expression changes that were linked with overall clinical benefit, a

statistical method was employed to specifically test for changes during the ICI-only period that were different between CB and NCB patients (see Methods). However, this test returned no statistically significant DEGs, suggesting there were indeed few treatment-induced changes that consistently differentiated CB and NCB. As a less stringent method, treatment-induced DEGs were calculated for CB and NCB patients as separate cohorts, and the results were compared (Figure 4.19A). As expected from the analysis in section 4.6.1, both CB and NCB patients were able to significantly upregulate many INCITE genes, including *CXCL9* and *CXCL10*, markers of cytotoxic cell infiltration, and some cytotoxic effector molecules (*FASLG*, *PRF1*). Investigating the genes that were significantly upregulated by ICI in NCB but not CB, a diverse pool of immune cell type markers were noted, including general markers of antigen presenting cells (MHC-II complex), the cDC2 marker *CLEC10A*, and *CD163*, a marker of M2-like immunosuppressive macrophages. Contrastingly, genes significantly upregulated in CB but not NCB were more restricted to T and B cell molecules (*IGHV4-31*, *CD8B*, *TRAJ3*, *TRBC1*). This provides some suggestion that CB patients may respond to ICI with focused cytotoxic and adaptive immune infiltration, while NCB patients have infiltration of immunosuppressive myeloid cells and other immune cell types alongside cytotoxic effectors, which may contribute to ICI resistance. However, while these genes met significance thresholds in one group but not the other, it must be emphasized that the differences in these trends were not statistically significant by the tests employed in this section (Figure 4.19B-C). Therefore, no conclusive claims can be made about ICI-induced expression changes and their link with CB in this cohort, with the exception of noting that many INCITE genes are upregulated to similar extents in both CB and NCB patients. Given the apparent variability in response subtypes, analyses of larger immunochemotherapy cohorts would likely be needed to attain sufficient power to detect transcriptional changes which more consistently distinguish patients that attain CB from those that do not.

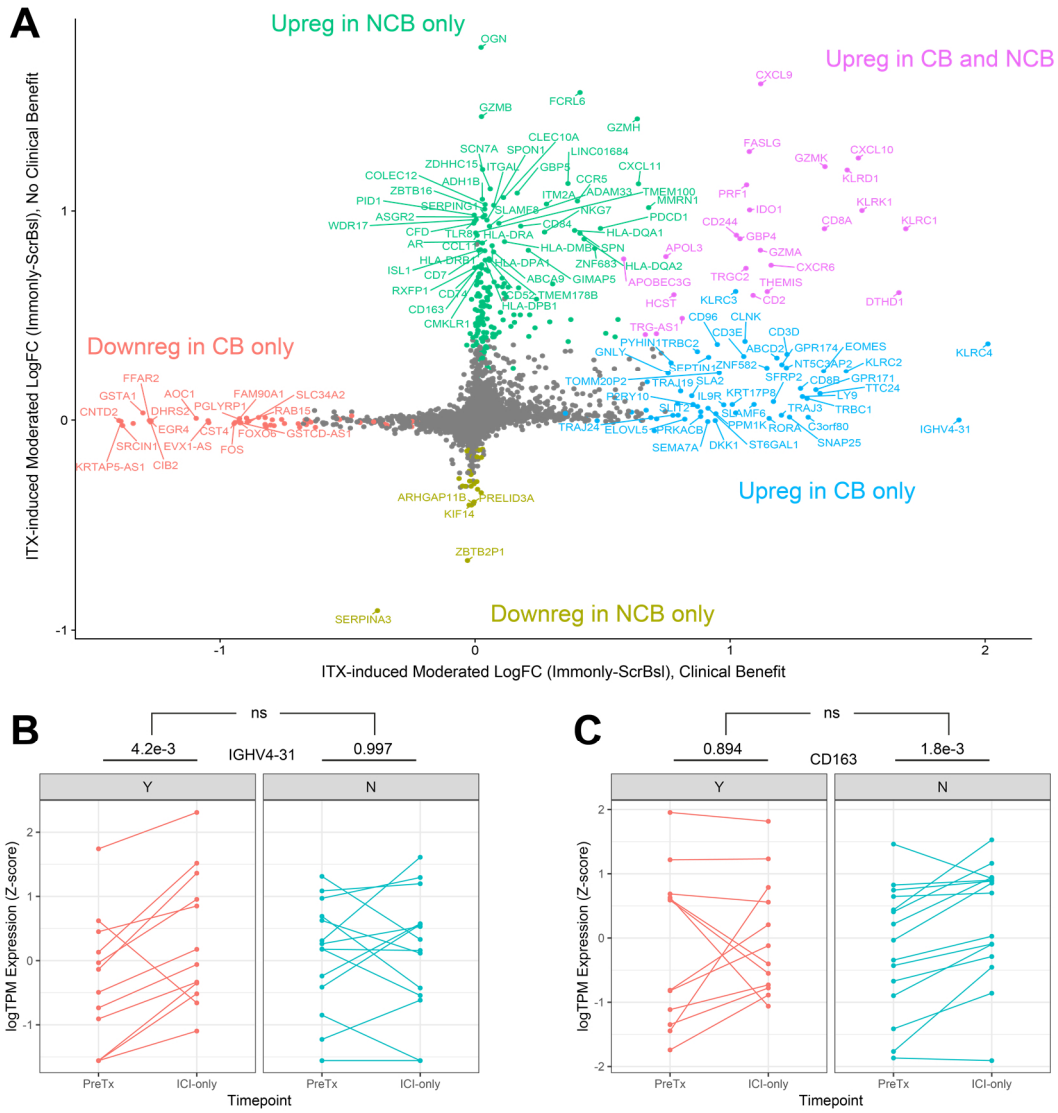


Figure 4.19: Comparison of genes meeting significance thresholds for ICI-induced differential expression for CB vs NCB. A) Differential expression analysis for the difference between ICI-only and PreTx biopsies was conducted separately for CB and NCB biopsies. Moderated LFCs are compared for CB (x axis) and NCB (y axis). Genes are coloured based on which dataset(s) they meet a 0.1 FDR significance threshold for, as labelled. Bottom: Z-scaled logTPM expression data are shown for B) *IGHV4-31* and C) *CD163*, presented separately for CB and NCB. Measurements for each patient across timepoints are connected by a line. The FDR testing for upregulation of each gene at ICI-only is presented for CB and NCB. "ns" (top) signifies that the difference between CB and NCB trends is not statistically significant for both genes (FDR > 0.99).

4.6.3 Assessing the effect of a priming αCTLA-4 dose on ICI-induced expression changes

In this study, a single priming dose of αCTLA-4 ICI was given for a subset of patients. One potential point of interest for clinicians and researchers is whether giving this single priming dose had a significant impact on transcriptional changes during the ICI-only window. To assess this, treatment-induced DEGs were calculated separately for patients receiving only αPD-L1 therapy and for those receiving combined αPD-L1/αCTLA-4 therapy, who had biopsies available

at both timepoints (n = 9 and n = 19, respectively). Comparing results from these analyses, both groups were able to significantly induce some key immune response genes in the INCITE gene set, including *CXCL9*, *CXCL10*, *CD8A*, *CD2, KLRC1*, *KLRC4*, and *KLRD1* (Figure 4.20).

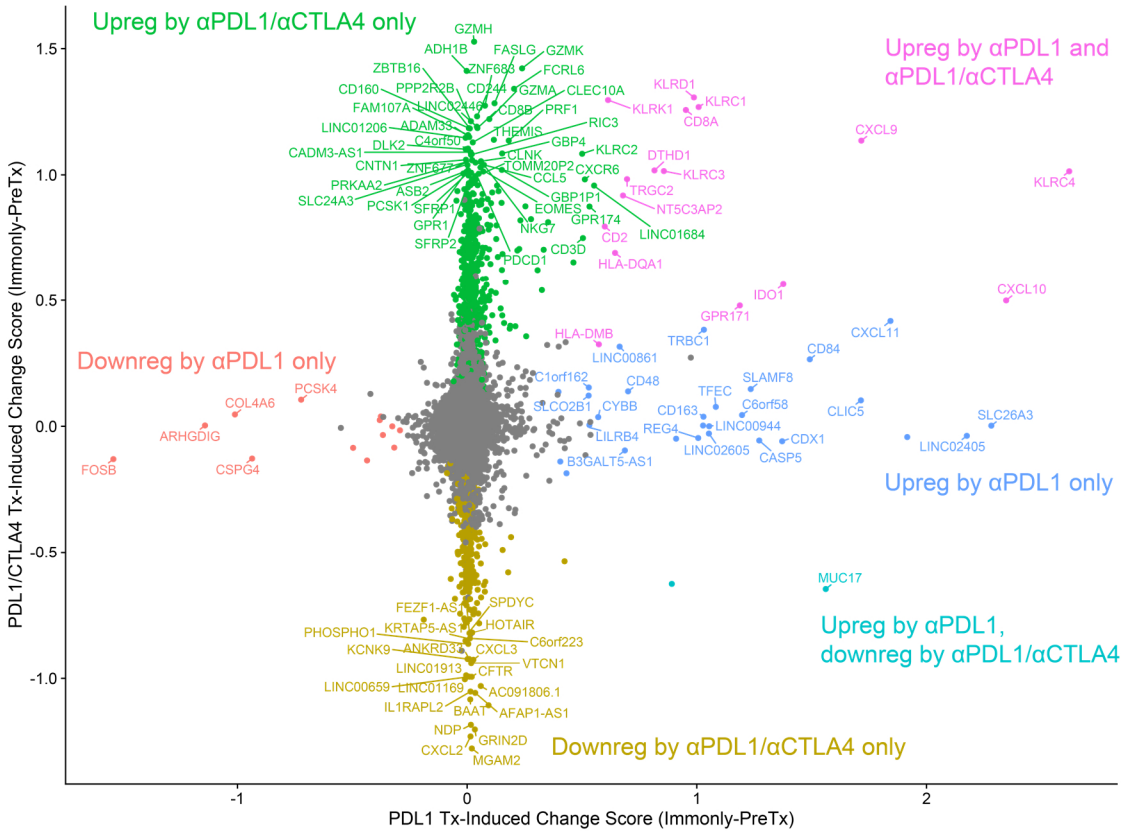


Figure 4.20: Comparison of genes meeting significance thresholds for ICI-induced differential expression for αPD-L1- and αPD-L1/αCTLA-4-treated patients. Moderated fold changes (log2 scale) calculated separately for αPD-L1 monotherapy (x axis) and αPD-L1/αCTLA-4 (y axis) are shown. Genes are coloured according to whether they meet significance thresholds (FDR < 0.1) in one or more treatment groups, as labelled.

However, significant upregulation in other T cell related genes, particularly cytotoxic effector molecules such as granzymes, *PRF1*, and *FASLG*, were significantly upregulated by αPD-L1/αCTLA-4 therapy, but did not reach significance for upregulation by αPD-L1 monotherapy (Figure 4.20). Genes significantly upregulated only by αPD-L1 monotherapy included the M2-marker *CD163*. Interestingly, myeloid chemo-attractants *CXCL2* and *CXCL3* were significantly downregulated by αPD-L1/αCTLA-4, but did not reach significance for αPD-L1 alone. None of these treatment-specific patterns were significantly different from each other according to rigorous statistical testing, as in section 4.6.2. However, nine genes were found to have

significantly different treatment-induced trends between these two treatment groups (Table 4.5): especially the IL-22 regulated epithelial barrier protein *MUC17*[167,168], which was significantly upregulated after αPD-L1 monotherapy, but significantly downregulated after αPD-L1/αCTLA-4 (Figure 4.20). However, no clear patterns that further differentiated the immune activity of αPD-L1 and αPD-L1/αCTLA-4 could be determined on the basis of these genes. In summary, these findings provide some suggestion that the addition of αCTLA-4 to αPD-L1 was associated with stronger cytolytic immune responses during ICI-only in this setting, and perhaps also with reduced accumulation of immunosuppressive macrophage infiltration. However, none of these differences reached significance in this analysis, so these trends cannot be conclusively demonstrated on the basis of this cohort alone. Similar to 4.6.2, larger cohort sizes are likely needed to increase the power to detect further statistical differences by these approaches. Sampling after repeated dosing of αPD-L1/αCTLA-4 rather than a single priming shot may also help clarify how the transcriptional impact of this combination ICI differs from that of αPD-L1 monotherapy.

Gene	Description	ICI-induced DE (αPD-L1 group)		ICI-induced DE (αPD-L1/αCTLA-4 group)		Difference between groups	
		mLFC	FDR	mLFC	FDR	mLFC	FDR
DDX41	DEAD-box helicase 41	**-0.38**	**0.006**	0.04	0.714	**0.40**	**0.033**
MUC17	Mucin 17, cell surface associated	**1.56**	**0.022**	**-0.65**	**0.039**	**-3.06**	**0.033**
AC120498.4		0.09	0.412	**-1.11**	**0.010**	**-2.25**	**0.067**
AC020656.2		**0.89**	**0.061**	**-0.63**	**0.055**	**-2.08**	**0.067**
LINC02863	Long intergenic non-protein coding RNA 2863	0.21	0.252	-0.34	0.115	**-1.45**	**0.087**
AP000943.3		0.18	0.279	**-0.58**	**0.063**	**-1.56**	**0.087**
NR1H4	Nuclear receptor subfamily 1 group H member 4	0.21	0.199	-0.33	0.125	**-1.45**	**0.0999**
AC066616.2		0.42	NA	**-0.54**	**0.057**	**-1.25**	**0.0999**
CDC37	Cell division cycle 37, HSP90 cochaperone	-0.19	0.232	0.05	0.599	**0.36**	**0.0999**

Table 4.5: Genes with significant differences in ICI-induced gene expression changes between αPD-L1 and αPD-L1/αCTLA-4 groups. Left: Gene name and description (where available) for all genes that show changes between PreTx and ICI-only that are significantly different (FDR < 0.1) between αPD-L1 and αPD-L1/αCTLA-4 groups. Middle: Differential expression (DE) between ICI-only and PreTx timepoints, calculated separately for αPD-L1 and αPD-L1/αCTLA-4 groups. Moderated log2 fold change (mLFC) and FDR are reported. Right: Same as middle, except testing the difference between the ICI-induced trends in αPD-L1 and αPD-L1/αCTLA-4 groups. All significant values (FDR < 0.1) are bolded.

4.6.4 Effect of ICI on normal GI control tissues

As immunochemotherapy was associated with treatment-emergent AEs, which included gastrointestinal AEs in some patients, the effect of ICI on paired normal tissues may also be of interest. Treatment-induced changes were therefore assessed in the normal control tissues. Paired PreTx and ICI-only biopsies were available for 5 patients in the normal oesophagus and gastric tissues, and for 6 patients with duodenal tissue, a smaller sample size than for tumour that likely results in comparatively reduced power to call DEGs. A limited number of ICI-induced significant DEGs were detected in oesophagus and gastric, while none were detected for duodenal tissue (Figure 4.21). Comparing oesophageal and gastric DEGs with findings from the tumour, no shared ICI-downregulated genes were detected. However, the INCITE genes *CXCL10* and *CXCL11* were found to be significantly upregulated by ICI in tumour, gastric, and oesophageal tissues. Additionally, *CXCL9*, *IDO1*, *HLA-DQB1*, and *WARS1* were significantly upregulated in one normal tissue site in addition to tumour. While the reduced sample size for normal tissues in this dataset may result in relative underestimation of the number of DEGs in normal tissues compared to in OAC, these data still show that a restricted subset of INCITE genes, particularly CXCR3-ligand T cell chemoattractants, can be induced by ICI in adjacent normal GI tissues in addition to OAC. However, no significant evidence for this induction was detected in the duodenal control site, suggesting these effects may be more specific to the tumour and the immediately adjacent normal GI tissues.

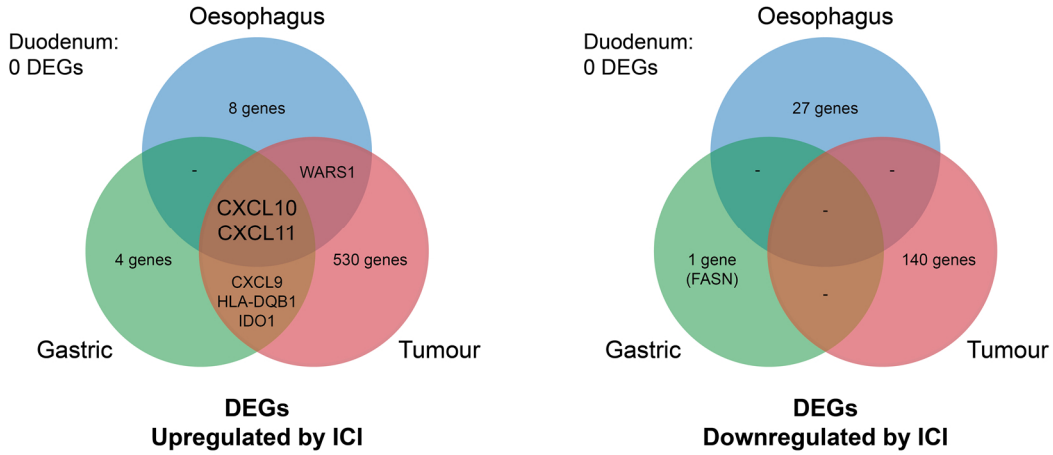


Figure 4.21: Comparison of transcriptional changes induced in the tumour during the ICI-only period with transcriptional changes induced in normal GI tissues. All treatment-induced DEGs were calculated at each tissue site separately, accounting for patient-specific differences. Significant DEGs (FDR < 0.1) were extracted from each comparison and compared. These Venn diagrams show DEGs that are upregulated (left) or downregulated (right), and how they overlap between OAC, Gastric, and Oesophagus. No significant ICI-induced DEGs were detected in duodenum, which is excluded from the Venn diagram.

4.7 Comparison of results with previously reported findings

4.7.1 Meta-analysis of DEGs from PD-1/PD-L1-targeted ICI cohorts shows limited overlap of high-confidence predictive markers between OAC and other settings

In the differential expression analyses in sections 4.5 and 4.6, T cell inflammation genes were found to be strongly activated by ICI, but did not have any link with CB or BOR at the PreTx timepoint. In fact, it was innate immune markers rather than T cell-related pathways that predicted positive outcome in this OAC immunochemotherapy cohort. This is in apparent contrast with reports and predictive signatures derived from other settings[90]. Therefore, I set out to characterize whether these innate immunity markers were OAC-specific or could apply to other cancer types, and to more directly assess the utility of previously reported predictive transcriptional signatures in OAC. Towards this end, a literature search was conducted for ICI cohorts with openly accessible bulk transcriptome data and outcome metrics. Raw RNA-seq data and clinical outcome information were obtained from seven additional advanced cancer cohorts from four different cancer types, all treated with ICI agents targeting the PD-1/PD-L1 axis (Table 4.6).

Cohort	Short Name	Tumour Type	ICI Agents	Chemo?	# Assessed	Line of therapy
LUD2015-005, OAC (Clinical Benefit)	OAC_LUD	Oesophageal/GOJ adenocarcinoma	αPD-L1 +/- αCTLA-4	Yes	32	1st
LUD2015-005, OAC (Response)	OAC_LUD	Oesophageal/GOJ adenocarcinoma	αPD-L1 +/- αCTLA-4	Yes	33	1st
Samsung Medical Center, Gastric Cancer	Gastric_Kim	Gastric/GOJ cancer	αPD-1	No	45	2+ (prior platinum-based chemotherapy)
Yonsei Cancer Centre, NSCLC	NSCLC_Cho	NSCLC	αPD-1	No	16	NR
Samsung Medical Center, NSCLC	NSCLC_Kim	NSCLC	αPD-1 or αPD-L1	No	27	NR
UCLA/VIC, Melanoma	Melanoma_Hugo	Melanoma	αPD-1	No	26	Mixed. 11/26 with prior MAPKi. Other lines NR
BMS038, Melanoma	Melanoma_Riaz	Melanoma	αPD-1	No	51	Mixed. 26 with priot αCTLA-4. Other previous lines NR
Melanoma Institute Australia, Melanoma	Melanoma_Gide	Melanoma	αPD-1 +/- αCTLA-4	No	73	Mixed. ~17% with prior BRAFi (18/105 in full cohort)
IMvigor210, Urothelial Carcinoma (UC)	UC_Mariathasan	Urothelial carcinoma	αPD-L1	No	298	Mixed. ~75% with prior platinum-based chemotherapy

Table 4.6: Characteristics of ICI cohorts assessed for predictive transcriptional biomarkers. "Chemo?" column denotes whether concurrent chemotherapy was given to patients during the ICI study period. "# assessed" refers to the number of patients with both available PreTx transcriptomes and assessable outcome information (response or clinical benefit, as reported by authors). NSCLC = non-small cell lung cancer; UC = urothelial carcinoma.

RNA-seq data from these sources were then processed for downstream analysis according to the same pipeline used for the inoperable OAC cohort. As exact survival information was not available for all cohorts, DEGs were calculated according to author-reported outcome classifications. While the terminology used varied, patient outcomes in these cohorts were generally classified according to their RECIST1.1 or irRECIST BOR (R vs. NR), or according to a CB-style metrics, which required maintenance of PFS for a certain duration. For the Melanoma_Riaz cohort, the authors reported both response and CB-style classifications; of these, the CB metric was selected due to the small proportion of responders in this cohort. DEGs were calculated for each cohort using the same significance threshold (p.adj < 0.1), and the results were integrated with the findings from the inoperable OAC cohort.

As an unbiased meta-analysis approach to detect genes predictive of outcome in multiple settings, the number of times each gene was significantly associated with outcome (FDR < 0.1)

was summed across all cohorts (integration score: +1 per cohort for a DEG upregulated in the positive outcome classification, -1 for a DEG upregulated in the negative outcome classification). Both response and clinical benefit PreTx analyses were included from the inoperable OAC cohort. To avoid biasing the findings towards this OAC cohort by double-counting genes shared with benefit and response, results for each classification strategy were assigned only half of the standard weight. Using this strategy, 101 genes were found to predict positive outcome in 3 or more cohorts (integration score >= +2.5), while 157 genes predicted negative outcome in 3 or more cohorts (integration score <= - 2.5).

Pathway analysis and manual categorization revealed that well over half of the genes that predicted positive outcome in 3 or more cohorts were related to T cell inflammation pathways which are classically used to predict ICI response (Figure 4.22), showing good agreement with previous reports and validating the efficacy of this meta-analysis approach. These included antigen presentation (especially classical and non-classical MHC class I and associated genes), *IFNG* and interferon-stimulated genes, NK and T cell markers, cytotoxic effector molecules, T cell coinhibitory and costimulatory proteins, and T cell chemoattractants (Figures 4.22, 4.24A). While most genes in these categories tended to predict positive outcome in at least one cohort from each of the four other cancer types studied, none were significant markers of clinical benefit or response in this OAC cohort. However, other markers were shared between OAC and multiple other cohorts, including CCR5 ligands (*CCL4L2*), markers of immune activation that have reported functions in multiple immune cell types (*LCP2, GPR84*, *TAGAP*), and the myeloid marker *CLEC12A*. These genes showed particular overlap with the NSCLC cohorts. Among the uncategorized genes predictive of outcome in multiple cohorts, *MUC16*, the gene encoding the tumour marker CA-125, was found to predict better response in OAC, gastric cancer, and urothelial carcinoma. Interestingly, the most consistent predictor of positive outcome was the unclassified gene vitronectin (*VTN*), which was upregulated in positive outcome for 5 of 8 cohorts assessed, including the LUD2015-005 OAC cohort, where it significantly predicted both CB and R.

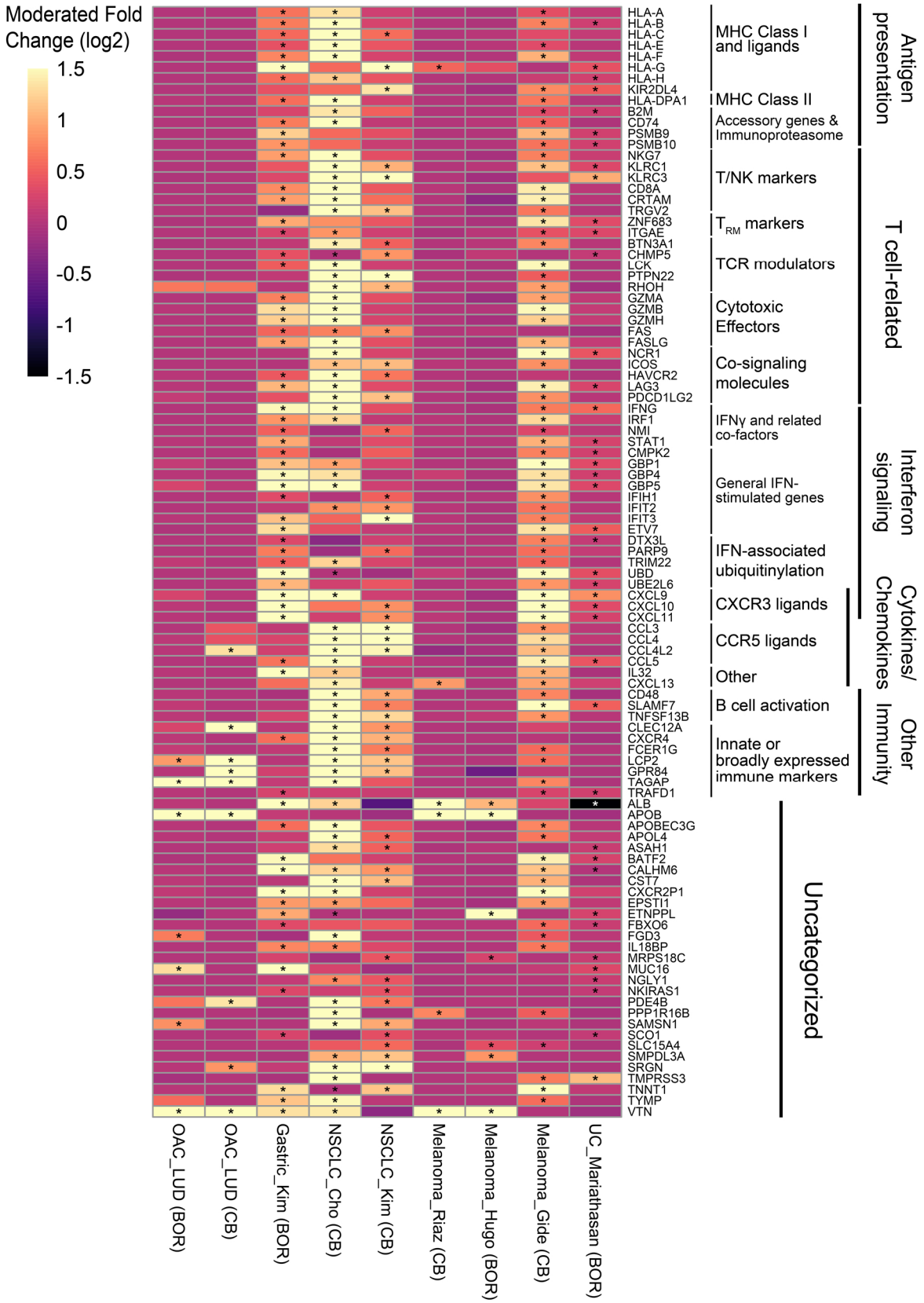


Figure 4.22: Genes predictive of positive outcome in multiple ICI cohorts. All genes upregulated in responders or patients attaining clinical benefit in 3 more cohorts are included in this plot, and categorized based on manual annotation and gene set enrichment (see Figure 4.24A). Each cell is coloured by the moderated log fold change for each dataset. Significant genes (FDR < 0.1) are marked by an asterisk- coloured black if the gene is upregulated in positive outcome, or white if upregulated in negative outcome.

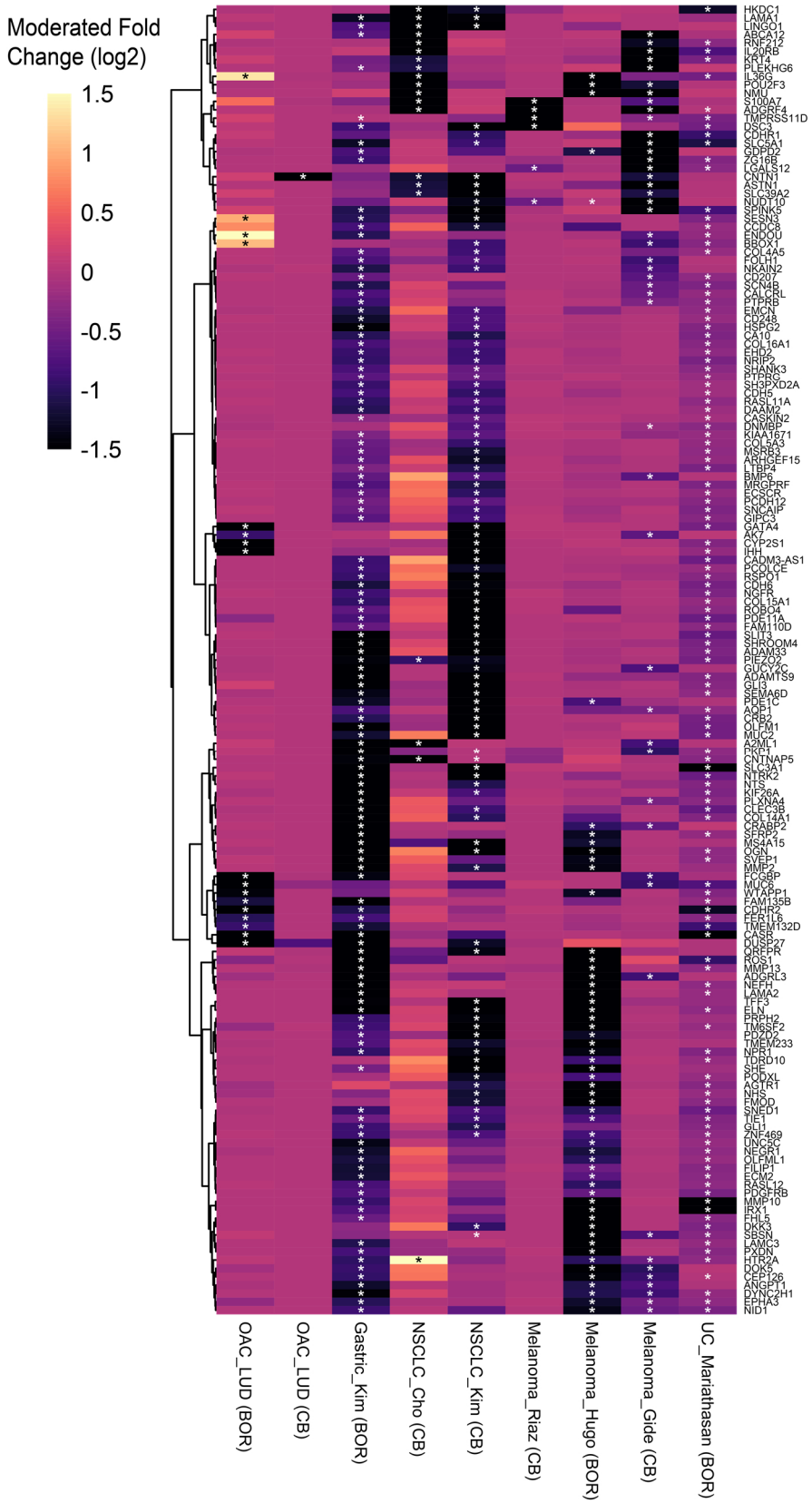


Figure 4.23: Genes predictive of negative outcome in multiple ICI cohorts. All genes upregulated non-responders/non-clinical benefit in 3 more cohorts are included in this plot. Each cell is coloured by the moderated log fold change, and rows are ordered by hierarchical clustering (ward.D linkage) rather than manual categorization. Significant genes (FDR < 0.1) are marked by an asterisk- white if the gene is upregulated in negative outcome, or black if upregulated in positive outcome.

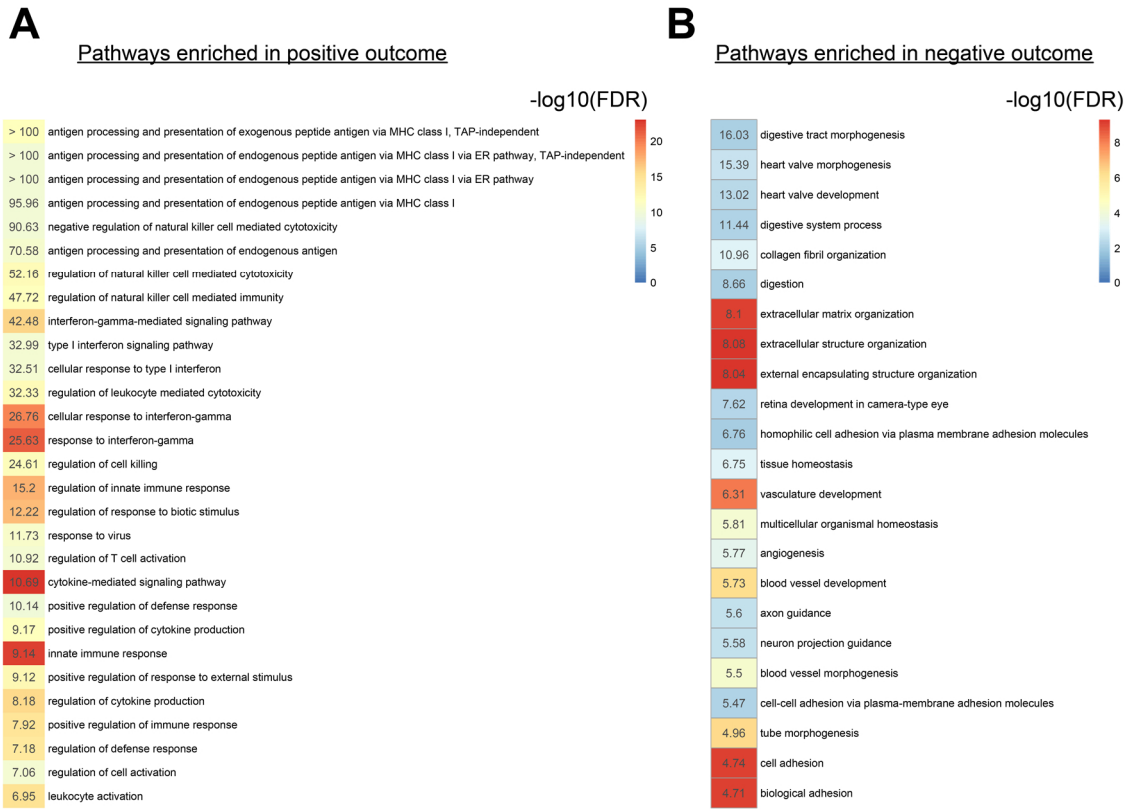


Figure 4.24: Pathways analysis of genes predictive of ICI outcome in multiple cohorts. The genes significantly upregulated in positive outcome (A) or negative outcome (B) in 3 or more ICI cohorts were subjected to pathway analysis using the PANTHER overrepresentation test using the GO biological processes database. Cells for each pathway are labelled according to the fold enrichment over background (all genes in the human reference). Cells are coloured according to the significance of the enrichment (-log10 of the FDR).

Genes associated with negative outcome in three or more cohorts were more heterogeneous and less susceptible to broad classification. Gene ontology analysis (Figure 4.24B) showed less significant pathway scores than for positive outcome genes, but still revealed enrichment for genes in stromal pathways, particularly involving the extracellular matrix (*MMP2*, *MMP13*, *COL4A5*, *ELN*) and development of vasculature (*ECSCR*, *TIE1*, *ANGPT1*). Genes involved in digestion (*TFF3*, *MUC2*, *MUC6*) and digestive tract morphogenesis (Wnt pathway: *GATA4*, *SFRP2*; Hedgehog pathway: *IHH*, *GLI1*, *GLI3*), were also commonly associated with poor outcome, even in non-GI tract cancers (Figure 4.23).

In summary, genes associated with T cell activity, antigen presentation, and interferon responses are indeed strongly predictive of outcome on ICI in multiple cancers, but do not predict outcome in this OAC immunochemotherapy cohort. However, markers of other classes of immunity, including innate and general immune activation markers, are shared between OAC

and other cohorts, particularly NSCLC. Genes associated with negative outcome are less easily categorized into discrete categories, but appear to be enriched for pathways associated with stroma and organ morphogenesis.

4.7.2 Comparison of previously defined transcriptional markers of ICI response with LUD2015-005 predictive markers

The fact that the pan-cancer markers of positive outcome returned in 4.7.1 have little overlap with markers predictive of outcome in this OAC immunochemotherapy cohort suggests that commonly used transcriptional signatures reported in the literature to predict ICI response may not have much utility in this OAC cohort. To assess this, ICI-predictive transcriptional signatures were mined from the literature[130,169–174]. These broadly fell into two categories: T cell inflammation-related gene sets used to predict response to ICI, and a heterogeneous group of signatures that has been reported to predict resistance to ICI in melanoma (IPRES) (Table 4.7). Interestingly, many of the T cell inflammation predictive signatures shared genes with the INCITE gene set identified in this work, and INCITE scores showed high correlation with scores assessed by these other inflammatory predictors in the OAC patients (Figure 4.25). Therefore, the INCITE gene set was also included for comparison. Expression of targeted checkpoint molecules and stand-alone T cell markers, which have also been used to predict ICI response, were considered as well. The performance of these markers and gene sets were compared to the genes predicting response and clinical benefit in the LUD2015-005 OAC cohort, as well as an innate immunity gene set derived from the clinical benefit gene list (Table 4.7).

Inflammatory / T cell signatures predictive of ICI response		
Gene set	Genes	Reference
BMS_4Gene_Inflammatory	CD274, CD8A, LAG3, STAT1	[169]
BMS_18Gene_Tcell_Inflamed	CXCR6, TIGIT, CD27, PDCD1LG2, CD274, CD8A, LAG3, NKG7, CCL5, CMKLR1, PSMB10, CXCL9, IDO1, HLA-DQA1, CD276, STAT1, HLA-DRB1, HLA-E	[170]
Cytolytic_Activity_Signature	GZMA, PRF1	[171]
Gajewski_13Gene_Inflammatory	CCL2, CCL3, CCL4, CD8A, CXCL10, CXCL9, GZMK, HLA-DMA, HLA-DMB, HLA-DOA, HLA-DOB, ICOS, IRF1	[172]
Merck_6Gene_Interferon_Gamma	CXCL10, CXCL9, HLA-DRA, IDO1, IFNG, STAT1	[173]
Merck_10Gene_Interferon_Gamma	CCR5, CXCL10, CXCL11, CXCL9, GZMA, HLA-DRA, IDO1, IFNG, PRF1, STAT1	[174]
INCITE	70 genes (see Section 4.6.1/Appendix C)	This work

Markers predictive of positive ICI outcome from this work		
Gene set	Genes	Reference
OAC_CB_Markers	43 genes (see Section 4.5.1.1/Appendix C)	This work
OAC_CB_InnateImmune	ACOD1, TNFSF14, APOB, CLEC12A, GPR84, LCP2, TREM1, MUC17, PDE4B, FPR1, VTN, C5AR1, FCAR, CSF3R	This work
OAC_R_Markers	142 genes (see Section 4.5.1.2/Appendix C)	This work

Signatures predictive of ICI resistance		
Gene set	Genes	Reference
IPRES signatures	See reference (26 gene sets)	[130]

Table 4.7: Definition of pre-defined ICI response gene sets. For each gene set, the name, member genes, and the source reference citation are provided. For larger gene sets defined in this work, the size of the gene set is provided along with the location where the full set can be found. The OAC_CB_InnateImmune gene set is derived from OAC_CB_Markers (genes significantly predictive of CB in 4.5.1.1) that fall into pathways related to innate immunity and general inflammatory response pathways, namely: Innate Immune System (Reactome) and Gene Ontology pathways for RESPONSE_TO_TUMOR_NECROSIS_FACTOR, RESPONSE_TO_CYTOKINE, CELLULAR_RESPONSE_TO_CYTOKINE_STIMULUS, and RESPONSE_TO_MOLECULE_OF_BACTERIAL_ORIGIN.

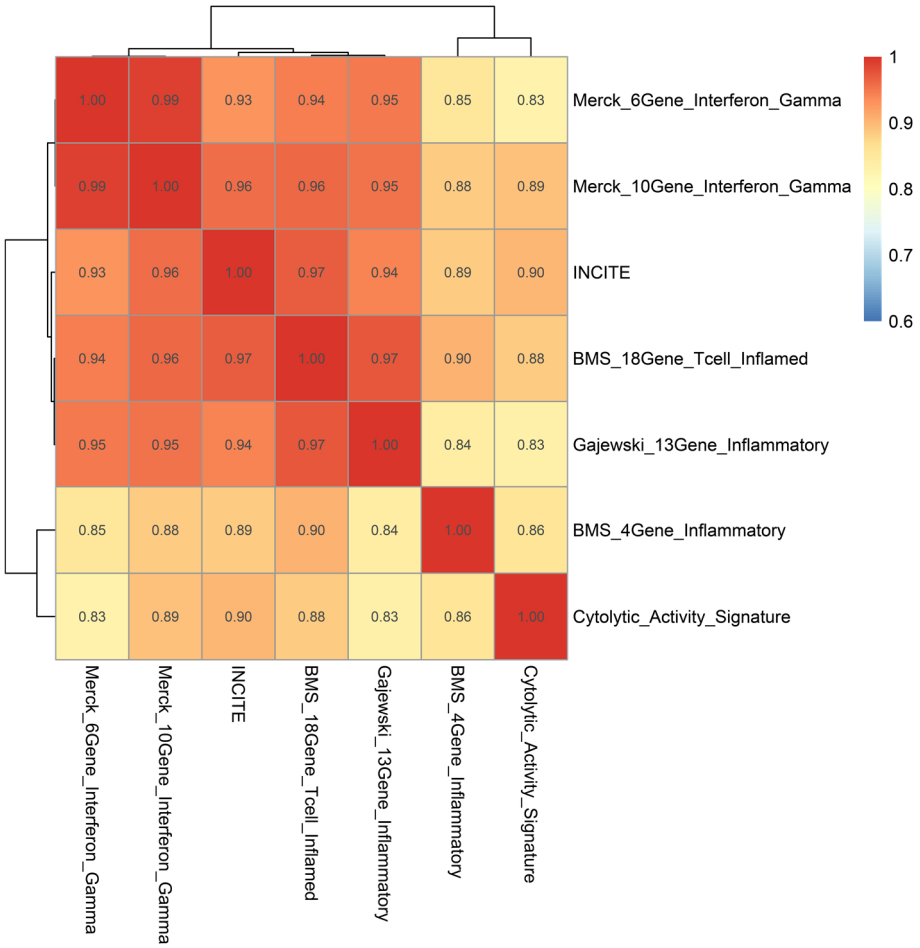


Figure 4.25: Gene set scores show high correlation between all T cell inflammation signatures and INCITE. For each gene set, the Z-score normalized logTPM expression for each gene was summed, resulting in a gene set score for each sample. These scores were calculated for all pre-treatment OAC biopsies in the LUD2015-005 cohort (n = 33), and pair-wise Pearson correlation coefficients were calculated between all T cell inflammation signatures, including INCITE. Correlation coefficients are labelled on each cell, and were used as the clustering distance to perform hierarchical clustering of rows and columns (ward.D linkage).

To assess the performance of these pre-defined markers and gene sets in OAC and other cancers, a score was calculated for each gene or gene set, as the sum of scaled logTPMs. A simple t-test was then used to assess the difference in scores between patients attaining positive and negative outcomes within each cohort. The results (Figure 4.26) show that T cell inflammation signatures predict better outcome in roughly half of the cohorts assessed, with INCITE performing similarly to the previously reported T cell inflammation signatures. Notably, these inflammatory signatures predicted CB in the Melanoma_Gide cohort, but did not predict positive outcome in the other 2 melanoma cohorts. This is indicative of the difficulty in finding robust predictive signatures from clinical cohorts, even within the same tumour type. Of stand-alone T cell markers and checkpoint molecules, PD-1/PD-L1 and cytotoxic T cell markers (*CD8A/CD8B*)

showed the strongest predictive power across cohorts, with a similar pattern to the T cell inflammation signatures.

None of these T cell inflammation markers or gene sets were predictive of positive outcome in OAC. Of the standalone markers, *CD274* (PD-L1) did have higher expression in OAC responders by this TPM-based method, although it was still not associated with CB. However, it should be noted that, while there was also a trend towards increased expression of PD-L1 in responders in the DESeq2 differential expression test in 4.5.1.2, it did not reach FDR-controlled significance (p.value = 0.017; FDR = 0.295). In summary, while PD-L1 expression may have some level of predictive power, T cell inflammation signatures are not predictive of immunochemotherapy outcome in this OAC cohort.

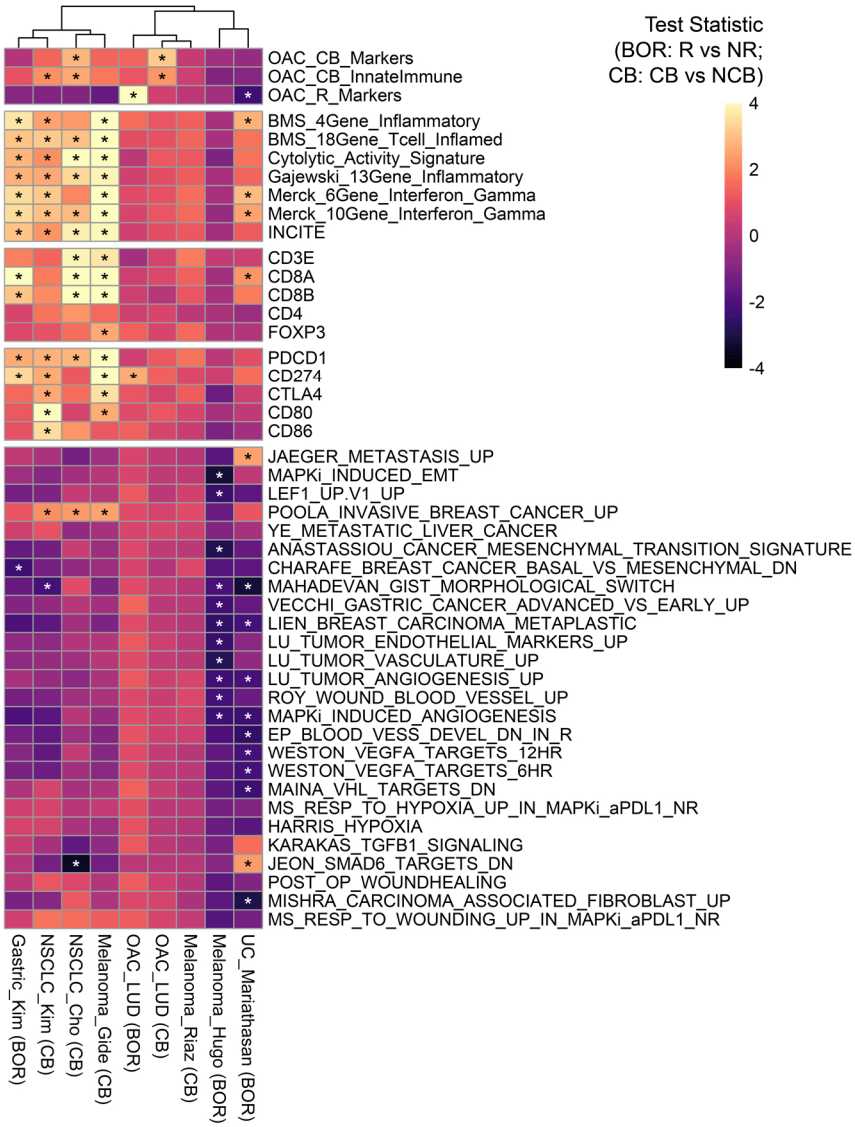


Figure 4.26: Evaluation of pre-defined ICI predictive signatures across PD-1/PD-L1 immunotherapy cohorts. For each gene set listed in Table 4.7, a gene set score was calculated as the sum of logTPM z-scores for all member genes. Scores were also calculated for stand-alone T cell markers and checkpoint molecules as listed in the figure using a similar approach (z-score logTPM of the individual marker). For each cohort, these scores were compared between R and NR (if BOR was used for outcome) or CB and NCB (if clinical benefit was used) using a Welch's t-test. The test statistic returned is used to colour cells in this plot, while significant values (unadjusted p-values < 0.05) are marked with an asterisk (black if gene is upregulated in positive outcome, white if upregulated in negative outcome). Columns are ordered by hierarchical clustering (ward.D linkage).

IPRES signature scores calculated by this method did predict poorer outcome in the melanoma cohort from which they were first reported (Melanoma_Hugo). However, while IPRES signatures did also predict poor outcome in urothelial carcinoma, they did not predict poor outcome the other two melanoma cohorts, again highlighting the inter-cohort variability for many of these predictive signatures. IPRES signatures did not significantly predict worse outcome for the OAC cohort. In fact, while IPRES was not linked with absolute irRECIST

response, the PreTx tumours having the highest IPRES expression appeared to be enriched for INCITE-quiescent responders (Figure 4.27A). This enrichment for high IPRES scores in INCITE-quiescent responders was even more visible at the ICI-only timepoint, where higher IPRES signalling could also be seen in INCITE-quiescent non-responders (Figure 4.27B). Interestingly, the enrichment in IPRES expression for INCITE-quiescent patients at ICI-only was mirrored by their inferior radiological responses during the ICI-only window (Figure 4.27C). Therefore, while PreTx IPRES expression should not be used to predict poor outcome for OAC on immunochemotherapy, its expression at ICI-only may be linked with resistance to ICI-only, as measured by both transcriptional (INCITE quiescence) and radiological methods. However, some INCITE-quiescent patients went on to respond during the immunochemotherapy phase of treatment despite high expression of these putative immunotherapy resistance pathways at the ICI-only timepoint. This suggests these signatures could be used to identify tumours resisting ICI-only, which may in some cases benefit from subsequent immunochemotherapy.

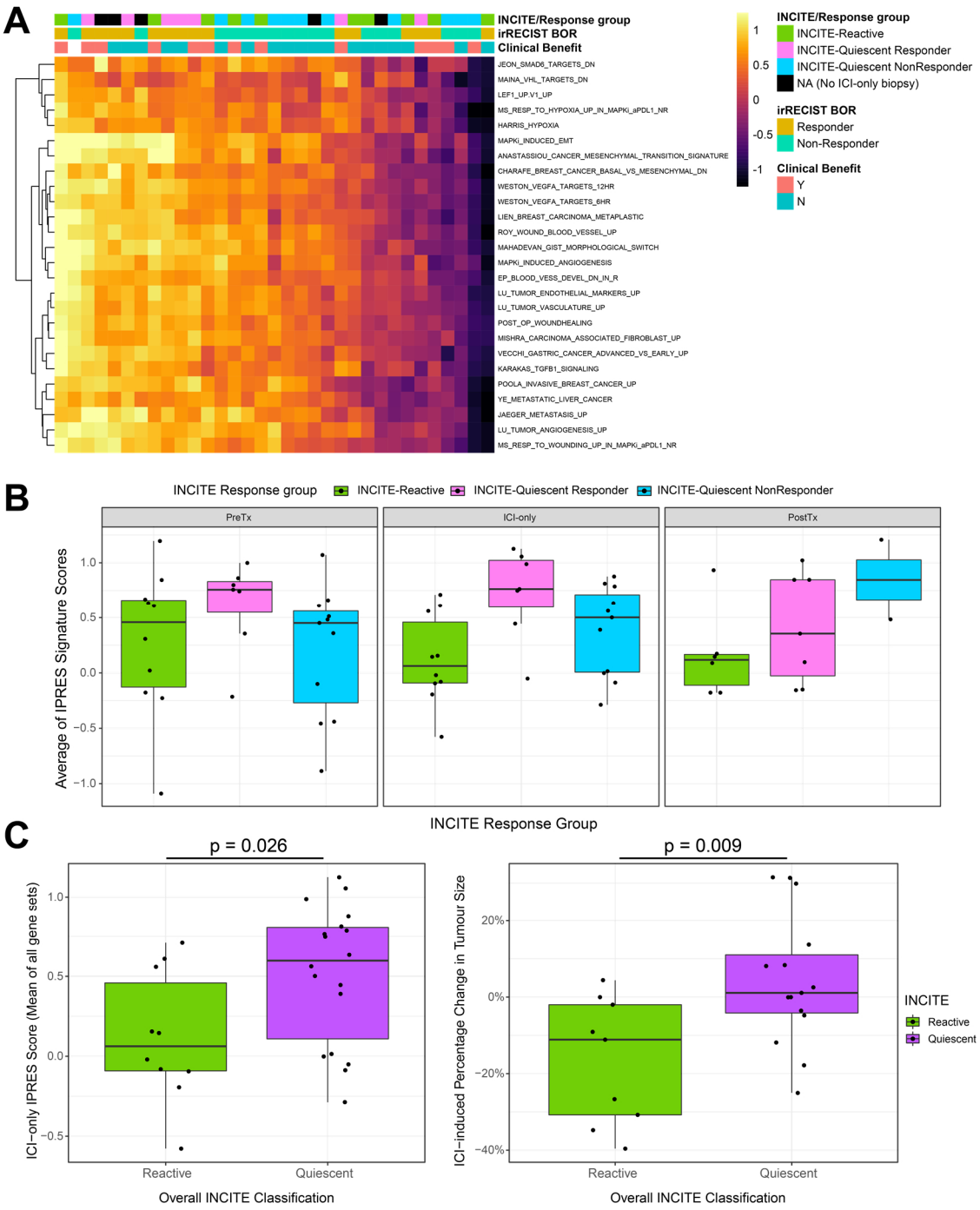


Figure 4.27: IPRES signatures and their link to response types in the LUD2015-005 OAC cohort. A) Signature scores, calculated as the sum of logTPM z-scores, are shown for each of the IPRES gene sets (rows) and PreTx OAC biopsies (columns). Columns are ordered by the average of all IPRES gene set scores for that biopsy. Rows are ordered according to hierarchical clustering (ward.D linkage). Annotations are provided for clinical benefit, irRECIST BOR, and INCITE groupings (based on Figure 4.18, with quiescent patients split into whether or not they responded during the immunochemotherapy component). B) IPRES signature averages were calculated as the average of scores for IPRES component gene sets shown in (A). These scores are compared for OAC biopsies across all 3 treatment timepoints (facets from left to right), and are grouped by the INCITE/Response categorization developed in (A). C) IPRES scores (left) and the percentage change of target lesion size during the ICI-only phase of treatment (right) are shown, as grouped by the overall INCITE classification from Figure 4.18. p values were calculated using a Welch's t-test.

The DEGs identified in this thesis to be predictive of positive outcome on immunochemotherapy for OAC showed limited utility in ICI cohorts in other tumour types. Genes predicting response in the LUD2015-005 OAC cohort did not positively predict outcome in any cohort. However, genes predicting CB were predictive of positive outcome in NSCLC, particularly for the subset of innate immunity genes (Figure 4.26). In summary, while the predictive power of signatures derived from this study appears limited in other cancer types, these signatures may have some predictive power in NSCLC. Additionally, previously reported signatures predictive of ICI outcome do not have strong predictive power at the PreTx timepoint in this OAC immunochemotherapy cohort, although on-treatment expression of IPRES signatures may help identify tumours resisting the ICI component of therapy.

4.7.3 Differing prevalence of EBV-infection and MSI is a likely driver of differences in predictive biomarkers for ICI response between oesophageal and gastric adenocarcinoma

In the above integrated analysis, there was little overlap between markers predictive of positive outcome in the Gastric_Kim cohort and OAC cohort, a somewhat surprising finding given the previously reported molecular similarity between OAC and some subtypes of gastric cancer[112]. In the αPD-1-treated gastric cancer cohort described by Kim et al., T cell inflammation signatures at PreTx were strongly predictive of ICI response[175,176]. This was particularly driven by expression of *CXCR3* ligands (*CXCL9, CXCL10,* and *CXCL11*), the gamma-interferon-induced family of T cell chemoattractants, which made up 3 of the 4 top DEGs enriched in responders in my reanalysis. However, these T cell chemoattractants and related T cell inflammation signatures showed little predictive power in the LUD2015-005 OAC immunotherapy cohort (sections 4.7.1-4.7.2).

To explore the reasons for the differences seen between OAC and gastric cancer cohorts, a refined reanalysis of the Kim et al. gastric cancer cohort was conducted based on the molecular subtype of each tumour. Integrated analysis of TCGA data has divided gastric cancers into four molecular subtypes[112]: 1) EBV infection-associated (EBV), 2) microsatellite instability (MSI), 3) genomic stability (GS), and 4) chromosomal instability (CIN), which was determined to have the

greatest similarity to OAC. Conversely, MSI-high and EBV+ tumours are thought to be rare or absent in OAC.

Subtype analysis of the Kim et al. cohort showed that many inflammatory markers highly predictive of response in this cohort, such as *IFNG* and *CXCL9*, had the highest expression in EBV-associated cancers, suggesting virus-associated cancers were associated with this hyper-inflamed pre-treatment phenotype (Figure 4.28). Most, but not all, of the MSI gastric cancers also showed similarly high expression of *IFNG* and *CXCL9* to EBV+ gastric cancers. These inflamed EBV and MSI subtypes of gastric cancer showed the highest rates of response to ICI monotherapy in the Kim et al. publication (100% and 85.7% response rate, respectively, compared to 24.6% in the overall cohort). Contrastingly, CIN and GS cancers tended to have much lower expression of T cell inflammation genes, and much lower ICI response rates. Comparing expression across cohorts, OAC tumours showed much lower absolute expression of these inflammatory genes than in the EBV and MSI-high gastric cancers, similar to or even marginally lower than the inflammatory gene expression seen in the related CIN gastric cancer subtype.

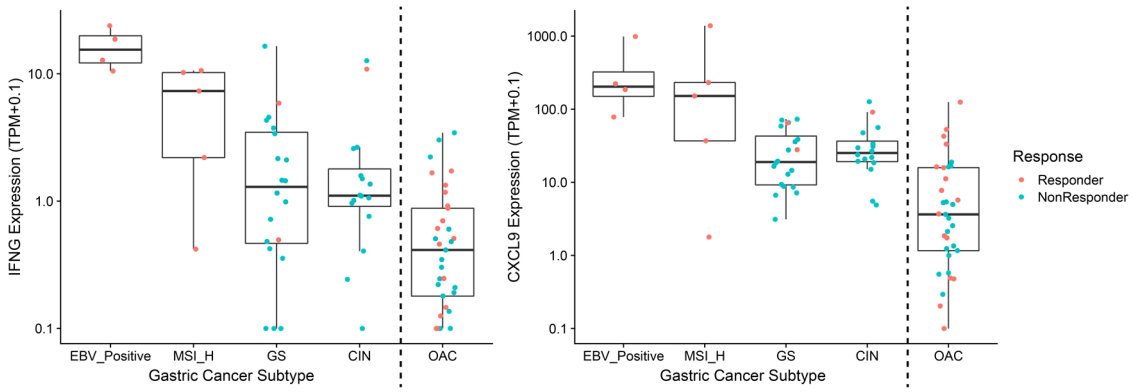


Figure 4.28: Comparison of inflammatory gene expression between OAC and various molecular subtypes of gastric cancer. Absolute expression, in the form of TPMs (0.1 offset) are shown for *IFNG* (left) and *CXCL9* (right), as exemplars of inflammatory gene signalling. Each point represents a single tumour sample. Molecular subtypes, as reported by Kim et al., are used to group the gastric cancer samples, while OAC TPMs are provided to the right of the dashed line as a separate group for comparison. TPMs are plotted on a log scale.

These data suggest another potential reason why T cell inflammation signatures are not predictive of outcome in the LUD2015-005 inoperable OAC cohort, even though they are strongly linked with response in this gastric cancer cohort: gastric cancers are much more likely

to be associated with MSI and EBV infection, which are much more likely to display this inflammatory pre-treatment profile and respond to ICI. In the LUD2015-005 cohort, only 1 out of 35 OACs was determined to be MSI-high, and there were no known EBV-associated tumours (see section 4.7.4). Additionally, no OAC showed the hyper-inflammatory pre-treatment phenotype characteristic of particularly the EBV-associated cancers in the Kim et al. cohort (*IFNG* > 10 TPM). This suggests that OACs do behave similarly to CIN gastric cancers in terms of their comparatively limited pre-treatment expression of interferon-related ICI biomarkers, and also that inflammatory MSI/EBV gastric cancer subtypes may return different predictive biomarkers in ICI studies than these OAC and CIN gastric cancers.

4.7.4 ICI response of virus-associated tumours

As noted in the previous section, Kim et al. reported a 100% response rate (6/6) for EBV-positive gastric cancers treated with αPD-1, suggesting high immunogenicity of some virally-driven tumours. While EBV positivity was absent in oesophageal cancers reported in the TCGA[112], the high response rate reported by Kim et al. warranted further examination of the LUD2015-005 dataset to search for evidence of rare EBV-associated OACs or other common cancer-associated viruses that could be linked with ICI response. Kim et al. determined EBV positivity using histological methods, which is the standard method used in the clinic for gastric cancers, but this was unavailable for the LUD2015-005 OAC cohort. The deposited sequencing data from the Gastric_Kim cohort was therefore analysed to assess whether EBV-driven tumours could be robustly detected using bioinformatics approaches. Indeed, all four EBV+ tumours with available sequencing data in this gastric cancer cohort had an EBV alignment rate that was orders of magnitude higher than all other samples (60-250 unique alignments per million (UAPM) to the EBV genome) (Fig 4.29). Contrastingly, patients with EBV-negative tumours either had zero detected EBV alignments or a minimal background level (<1 UAPM), which likely represented either low-level ambient contamination, false alignments, or presence of latent EBV in small amounts (such as in infiltrating lymphocytes). A threshold of 10 unique alignments per million (UAPM) to the EBV chromosome was chosen to distinguish histologically-confirmed EBV-positive tumours from all other samples.

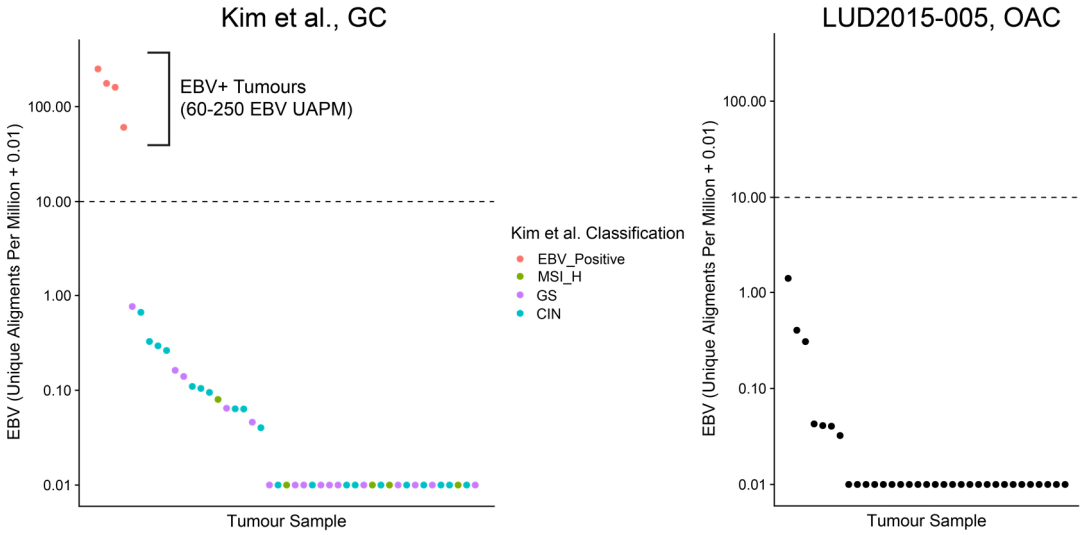


Figure 4.29: Transcriptional detection of EBV+ gastro-oesophageal cancers. For each sample, the number of reads uniquely aligning to the EBV chromosome per million unique alignments (Unique Alignments Per Million; UAPM) was calculated. Pre-treatment tumour samples were then ordered by decreasing EBV UAPM, and plotted for the Gastric_Kim cohort (left) and the inoperable OAC cohort (right). A pseudocount of 0.01 was added to UAPM values (such that 0.01 represents an absence of detectable EBV alignments), and these values are displayed on a log scale.

In the LUD2015-005 inoperable OAC cohort, no sample was determined to be convincingly EBV-positive by this threshold, with the maximum baseline tumour sample scoring only 1.4 UAPM to EBV. Therefore, no OACs in this cohort appeared to be associated with EBV infection, agreeing with the findings from the TCGA dataset. A similar workflow was implemented to check for other transcriptionally active cancer-associated viruses, which showed that none met this 10 UAPM threshold in OAC samples. However, while OSCC patients have been excluded thus far from analyses in this thesis, assessment of the two OSCC patients with PreTx RNA-seq available revealed that one patient was positive for Human Papilloma Virus (HPV, strain 16, 14 UAPM), despite completely undetectable HPV expression for all other LUD2015-005 pre-treatment samples. This patient experienced a strong outcome on immunochemotherapy, showing a partial irRECIST response (27% reduction during ICI-only, maximal reduction of 50%) which was accompanied by a complete loss of detectable HPV in bulk RNA-seq data at the ICI-only timepoint. This patient attained clinical benefit, and at the time of last follow-up, had continued survival of over 19 months. Interestingly, this patient also had the highest pre-treatment expression of key inflammatory genes such as *IFNG* and *CXCL9* out of all LUD2015-005 tumour biopsies, similar to what is seen in EBV-positive gastric cancers (Figure 4.30). In summary, while no virus-associated OACs were found in this cohort, a patient with an HPV-16-associated

OSCC was detected. This patient had the highest levels of inflammatory gene expression of any pre-treatment biopsy in this cohort, and also attained good outcome on immunochemotherapy. While this is just a single case, in combination with the reanalysis of the Gastric_Kim cohort, these data provide further evidence that virus-associated gastro-oesophageal cancers have an inflammatory pre-treatment phenotype that is associated with particular responsiveness to ICI treatment regimens.

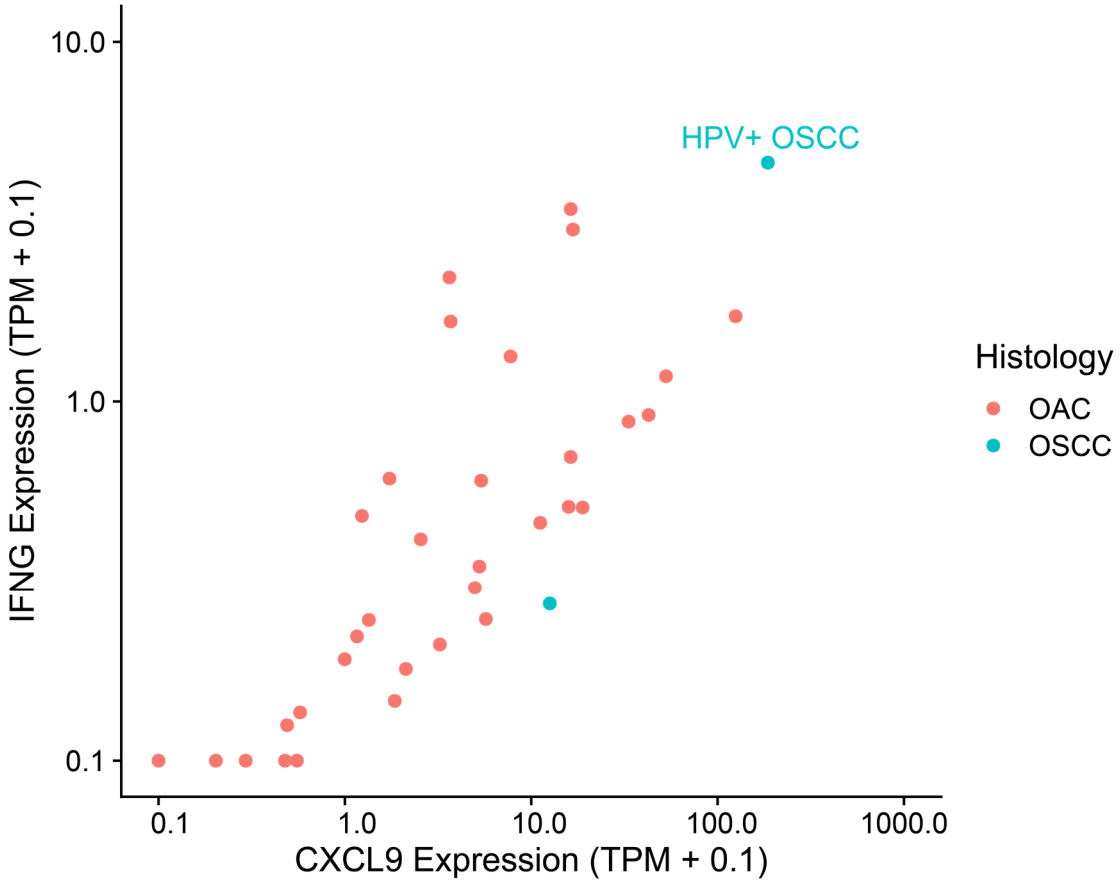


Figure 4.30: Inflammatory pre-treatment phenotype of an HPV-associated OSCC. Absolute expression, in the form of TPMs (0.1 offset) on a log scale, are plotted for *IFNG* and *CXCL9*. Two pre-treatment OSCC biopsies (blue) are shown in comparison to the 33 samples in the inoperable OAC cohort (red). The HPV+ OSCC patient is labelled.

4.8 Summary and brief discussion

In this chapter, I assess a variety of genomic and transcriptomic features and their link with outcome for this OAC immunotherapy cohort. For genomic biomarkers, this study is the first to specifically address the suitability of the widely used TMB biomarker for first-line immunochemotherapy in OAC. While a previous report from a cohort of previously-treated

gastro-oesophageal cancers suggested that TMB may not be a robust predictor of outcome on ICI in that setting[110], this work provides evidence that higher TMB is in fact a strong predictor in OAC of positive outcome with first-line immunochemotherapy. Additionally, analysis of TCGA data did not show the same association between TMB and outcome, suggesting this association is likely immunochemotherapy-specific. This well-studied ICI biomarker could therefore be particularly useful to help select patients with greater odds of deriving benefit from the addition of ICI to conventional chemotherapy in this setting. Similar to what has been attempted for PD-L1 in recent phase III trials in this setting, future randomized studies would be needed to assess whether TMB is indeed useful to select patients for ICI addition, and whether any TMB threshold is particularly suitable for this purpose.

Previous studies have demonstrated particularly high response rates for MSI-high and virus-associated gastric cancers to ICI targeting the PD-1/PD-L1 axis[175]. While there was only one MSI-high tumour and one virus-associated tumour (an HPV-16+ OSCC) in this cohort, both showed strong tumour shrinkage during the ICI-only component of therapy and maintained disease control for longer than 12 months. In combination with the data from Kim et al, these results provide further evidence that MSI-high and virus-associated gastro-oesophageal cancers are strong candidates for response to ICI-containing treatment protocols, including first-line immunochemotherapy for OAC. On the basis of the responses shown by Kim et al. and the regression seen in our study in the ICI-only window, it appears that ICI alone can also generate strong responses in these tumour subtypes even in the absence of chemotherapy. Further studies would be needed to determine whether ICI monotherapy could be a suitable first-line treatment option for these groups of patients. Nevertheless, given these results, regular screening for HPV (OSCC), EBV (gastric cancer), and MSI status (mainly in gastric cancer, but also in some oesophageal cancers) should be strongly considered to assist with selection of patients for first-line immunochemotherapy.

In this study, transcriptomic profiles of normal tissues were obtained in addition to OAC. These data are a valuable resource to assess tumour-specific expression, which could be used to inform novel immunotherapeutic treatment approaches. For instance, the highly OAC-specific

expression of *MAGEA4* would support the use of immunotherapies targeting this CT antigen, which are currently in clinical development in other cancers[177]. Additionally, while *PRAME* has low-level background expression in normal tissues, it is also highly enriched in OAC, and the fraction of patients who maintain its expression at the PostTx timepoint have particularly worse long-term outcomes. Therefore, targeted therapies against *PRAME*[178] may be useful in generating anti-tumoural responses in patients who otherwise resist ICI.

These transcriptomic data also provide support for therapies targeting CTLA-4 in first-line OAC, in addition to PD-1/PD-L1, which is currently the only axis targeted by ICIs approved for this setting. Expression of CTLA-4 and its ligands had much greater OAC-specific expression than PD-1/PD-L1 in this cohort, which was unselected on the basis of checkpoint molecule expression. While clinical data were not formally assessed in this thesis to determine whether addition of αCTLA-4 impacted survival outcomes, the addition of this drug did show a trend towards increased induction of cytolytic effector molecules during the ICI-only window. This finding suggests that a larger cohort designed to test this comparison may show a therapeutic benefit for the ICI combination over αPD-L1 alone. Other highly enriched checkpoint molecules (B7-H3, 4-1BB) could also be considered as potential therapeutic targets in this setting.

In PreTx differential expression analyses, while T cell markers were not associated with outcome in this cohort, a number of immune markers were. First, naïve and tolerance-associated B cell genes appeared to be linked with poor response in multiple PreTx analyses in this study. Of these, *IGHD* was a particularly robust marker of poor outcome, showing associations with NCB, NR, and INCITE-quiescent non-responders, who have poor responses during both the ICI-only and immunochemotherapy phase. It is possible the presence of intratumoural naïve IgD^+ B cells may represent a tumour type that is particularly resistant to generation of adaptive immune responses. In any case, this gene is the most robust PreTx marker of poor outcome to immunochemotherapy in this cohort. PreTx genes most predictive of positive outcome in this cohort appeared to be related to inflammatory myeloid responses. While myeloid infiltration is a general feature of OAC, it appears this specific form of inflammatory myeloid infiltration is a key driver of overall outcome on immunochemotherapy.

However, bulk RNA-seq analysis is limited in its ability to specifically define the cell type identity or polarization state responsible for this predictive power. Characterisation at the single-cell level would be better placed to more fully characterize this inflammatory myeloid phenotype. In conclusion, B cell and myeloid markers were more suitable as PreTx markers in this cohort, while potential role of T cell markers was restricted more to study of transcriptional responses during the ICI-only window and static analysis of the ICI-only window.

While PD-L1 expression was not histologically determined in this study, its PreTx RNA expression level did appear to show some link with response to immunochemotherapy, although this did not survive p-value correction against all genes in the standard differential expression workflow. Additionally, no link with PD-L1 expression and CB was apparent. To date, PD-L1 is the only expression-based biomarker that has been used to stratify phase III trial data to identify patient groups particularly likely to benefit from addition of ICI to chemotherapy. While the analysis in this thesis cannot directly assess that question as the LUD2015-005 trial did not have a chemotherapy-only cohort, the moderate or absent link between PD-L1 expression and positive immunochemotherapy outcomes suggests it may not be the ideal biomarker to inform patient selection, at least for the addition of this ICI regimen. Future studies should consider developing other PreTx markers beyond PD-L1 for patient stratification and selection in this setting.

While not significant in the present OAC cohort, the predictive biomarkers and genes returned by the pan-cancer ICI meta-analysis in this chapter confirm the importance of T cell inflammation genes in predicting outcome across cancer types, showing good agreement with other meta-analysis approaches[90]. This analysis suggests that, were T cell inflammation strongly predictive of outcome in this OAC immunochemotherapy cohort, it would be readily detectable; however, this finding was not seen in this cohort. There are four potential explanations for this. First, the inherent variability associated with biomarker studies in clinical cohorts. Only one of three melanoma cohorts showed strong predictive power for T cell inflammation signatures, and these cohorts also showed poor agreement for other signatures such as IPRES. Due to confounding enrolment characteristics, differences in reported outcome classifications, and

inherent variability of tumours, a robust finding may not be detectable to the same extent in each cohort. Therefore, one must consider the possibility that T cell inflammation is in fact a strong predictor in OAC that did not reach significance in this cohort due to limited power and confounding variation.

Second, this could be a phenomenon specific to a specific tissue type. The innate immune markers returned in OAC show the greatest predictive power in NSCLC, which may be of particularly interest given the features shared between these two cancer types and between the developmentally-linked tissues from which they arise[13,38]. Additionally, the HLA supertype A02 has been suggested to predict improved survival on ICI in NSCLC[127], and this supertype had a non-significant trend towards predicting OS in this setting as well (HR 0.46 [95% CI: 0.20-1.01]). This could suggest a conserved immune phenotype shared between these related tissue types with implications for ICI therapy that apply to both tissues. However, T cell inflammation signatures were even better predictors in NSCLC than this OAC-derived innate immune signature, so other factors must be at play.

Third, all cohorts assessed in this meta-analysis either had some form of prior therapy in at least some patients, or did not report whether there were previous lines of treatment. For instance, all patients in the Gastric_Kim ICI cohort had previously progressed on an oxaliplatin-containing chemotherapy regimen, which has been shown to have immunogenic properties[179]. This could explain why ICI responders in other cohorts have high expression of T cell inflammation genes at PreTx, while in our first-line cohort, these genes are induced by immunotherapy (INCITE) in ICI-only responders: tumours likely to respond to therapy in an immunogenic way have already done so, whereas in the first-line setting, the same stimulus has not yet been applied. This would explain why OAC tumours in this cohort appear to have even less inflammatory gene expression than the related CIN gastric cancer subtype from this cohort. Consideration of prior therapy subgroups during analysis, as well as additional validation of markers in first-line cohorts, should be considered to validate whether these T cell inflammation signatures are as predictive in the absence of prior immunogenic chemotherapy.

Finally, these findings may speak to differences between predictive biomarkers for ICI-only compared to those predictive of immunochemotherapy outcome. This OAC cohort was the only ICI group in the meta-analysis that was treated with concurrent chemotherapy, and many findings from this chapter suggest that signatures related to the chemotherapy component are likely to be major contributors to the overall predictive signal. For example, the expression of a multi-drug resistance protein (*ABCC2*/MRP2) implicated in resistance to oxaliplatin-based chemotherapy was one of the top genes predictive of poor outcome at the PostTx timepoint. Therefore, genes predicting overall clinical benefit or response in this cohort are likely to be a mixture of predictive factors for ICI and chemotherapy, or their combination, which may differ from predictive factors for ICI alone.

Using the unique study design of the LUD2015-005 trial, this question was able to be partially addressed by assessing the change during the four-week ICI-only window, prior to the addition of chemotherapy. While nearly all patients showed upregulation of T cell infiltration and inflammatory activity during this period, as summarized by the INCITE gene signature, this upregulation was much stronger in some patients than others. Patients with highest upregulation of INCITE genes had greater tumour shrinkage during the ICI-only window, showing that this transcriptional response was associated with radiologically measurable anti-tumour activity, and could be used as a marker of response to the ICI component of therapy. INCITE-quiescent patients, on the other hand, had high expression of putative PD-1/PD-L1 ICI resistance signatures such as IPRES at the ICI-only timepoint, as well as reduced tumour shrinkage during the ICI-only window. Importantly, many IPRES-high INCITE-quiescent patients generated late responses during the immunochemotherapy phase of this trial, suggesting chemotherapy may help overcome ICI resistance in some patients. These findings support the notion that cohorts from this setting contain some patients that may be responding primarily to the ICI component of therapy, and others responding to immunochemotherapy, which may have different radiological and transcriptional biomarker profiles.

One important finding from the on-treatment analyses in this thesis is that nearly all patients develop some level of INCITE induction during the ICI-only window in the LUD2015-005 trial,

even though phase III studies suggest that only a minority of patients likely derive clinical benefit from the addition of ICI agents[73–75]. This suggests that ICI resistance mechanisms are able to overcome the induction of T cell inflammation in most advanced OAC tumours. With this in mind, careful attention should be paid to the transcriptomic signatures associated with NCB at the ICI-only timepoint, including increased expression of extracellular matrix components and clotting factors, which likely represent early on-treatment resistance mechanisms. Overcoming these ICI resistance factors could be critical in expanding the number of OAC patients who develop strong clinical responses to the T cell inflammation which is successfully induced by ICI-only in this setting.

In addition to T cell inflammation markers, the integrative meta-analysis in this chapter provided evidence for additional markers that predict positive ICI outcome in multiple cohorts. For instance, *MUC16*, encoding the key tumour-specific antigen CA-125, predicted ICI response in gastric cancer, OAC, and bladder cancer. *MUC16* has been explored as a potential antigenic target for immunotherapeutic intervention in ovarian cancer models[180], but also showed enriched expression in adenocarcinomas throughout the gastrointestinal tract, where it has been associated with poorer survival on standard therapy[181]. Therefore, *MUC16*/CA-125 is potentially an important marker of gastrointestinal adenocarcinomas that could particularly benefit from the addition of ICI to conventional chemotherapy. Additionally, vitronectin (*VTN*) was returned as the gene that most consistently predicted positive outcome in this analysis (5 out of 8 cohorts assessed), surprisingly scoring even higher than the classical predictive genes in T cell infiltration, interferon-gamma, and antigen presentation pathways. The role of this gene in immunotherapy response has not yet been characterized. *VTN* is a multifunctional extracellular matrix (ECM) protein produced by multiple cell types, which can promote ECM remodelling in the context of tumour invasion and metastasis[182]. However, it also has a role in modulating complement and clotting pathways[183], which were associated with on-treatment ICI resistance in this cohort. *VTN* is also a binding target for extracellular integrin receptors and other cell-adhesion molecules, and can therefore function as a mediator of intratumoural lymphocyte infiltration[184]. In summary, *VTN*, *MUC16*, and the other unclassified markers returned by this analysis represent an exciting opportunity to develop new pan-cancer ICI

biomarkers and also to understand additional contributory ICI response mechanisms, which are not well captured by the current focus of the field on T cell inflammation signatures. Further exploration of the biology behind these markers and validation of their predictive utility in ICI should be strongly considered.

Finally, the integrative analysis in this chapter emphasized the importance of considering molecular subtype information when attempting to find predictive ICI markers with utility across gastro-oesophageal cancers. To date, immunochemotherapy studies have generally grouped gastro-oesophageal cancers together. However, the present analysis highlighted that interferon-related signatures associated with MSI and EBV-positive tumours are likely to drive the overall predictive biomarkers identified in gastric cancer ICI cohorts, particularly given the high response rate of these subtypes. However, these markers are not likely to hold for OAC and CIN/GS gastric cancer subtypes, where hyper-inflammatory phenotypes are rare. Given these findings and the molecular similarity between OAC and CIN gastric cancer, it may be more appropriate to group OAC and CIN gastric cancer for the purpose of biomarker study, and treat hyperinflammatory gastric cancer subtypes separately. As only one CIN gastric cancer patient responded to second-line ICI treatment in the Kim et al. cohort, it was not possible to directly test whether OAC and CIN gastric cancer shared predictive biomarkers. Future studies in first-line gastric cancer setting are needed to clarify whether these groups do indeed show similar profiles in ICI biomarker studies. Nevertheless, these data highlight a need to consider molecular subtype when designing and assessing predictive biomarker analyses in gastro-oesophageal cancer, to evaluate whether identified biomarkers have utility in all subtypes.

Chapter 5: Single-cell RNA-sequencing to Profile OAC Cell Composition and Advanced Applications for Understanding Immunochemotherapy Outcomes

5.1 Introduction

Single-cell RNA-sequencing (scRNA-seq) is a useful tool to understand the composition of tumours and identify tumour subtypes that could be best targeted by various agents. Indeed, with the advent of microfluidic approaches allowing for easier and cheaper generation of large scRNA-seq datasets, many common tumours have been profiled with this method. For OAC, however, relatively little data is available. scRNA-seq from OAC has been reported previously, but only for a limited number of single cells (150)[185]. Additionally, only the malignant epithelium was profiled, so no analysis of the TME can be conducted. An abstract has reported the generation of a larger dataset from OAC, but the full results and dataset have not been reported yet[186]. Therefore, at the time of writing, no large-scale scRNA-seq resource exists that can be used to further the understanding of OAC, despite its prevalence in many parts of the world.

In this chapter, I describe the generation of a combined single-cell atlas of OAC, paired-normal GI tissues, and other oesophageal pathologies, namely the pre-malignant condition Barrett's oesophagus and OSCC[m]. This dataset contains over 65,000 single-cell transcriptomes derived from nearly 60 samples (Table 5.1), greatly expanding the scRNA-seq data available to researchers to profile these conditions. The vast majority of these cells are derived from ICI-treated patients over the course of treatment on the LUD2015-005 study, and therefore this resource is also a valuable tool to further understand predictive biomarkers and mechanisms of response to the treatments used in this trial. First, a new method for cryopreservation of endoscopic biopsies for scRNA-seq will be discussed, followed by an introduction to the cell types that were identified in this dataset. Finally, I will show how this dataset can be leveraged to gain further insight into immunochemotherapy for OAC, including an assessment of tumour

[m] The generation of this dataset was made possible through collaboration with Joseph Chadwick. For most batches, Joseph and I worked in tandem to process the biopsies into cell suspensions, conduct scRNA-seq, and prepare libraries for sequencing.

evolution during the ICI-only window and evaluating how cell composition links with outcome in the LUD2015-005 inoperable OAC cohort.

Patient	Source	Timepoints	Tissues	Biopsy method	Preparation	Treatment
OAC-ACMO	LUD2015-005*	PreTx; ICI-only; PostTx	OAC; Gastric; Oesophagus	Endoscopic	Slow-frozen	αPD-L1/CTLA-4 + CapOx
OAC-GDBD	LUD2015-005*	PreTx; ICI-only; PostTx	OAC; Gastric; Oesophagus	Endoscopic	Slow-frozen (all); fresh replicate (PostTx)	αPD-L1/CTLA-4 + CapOx
OAC-HDHI	LUD2015-005*	PreTx; ICI-only	OAC	Endoscopic	Slow-frozen	αPD-L1/CTLA-4 + CapOx
OAC-HENB	LUD2015-005*	PreTx; ICI-only	OAC; Gastric; Oesophagus	Endoscopic	Slow-frozen	αPD-L1/CTLA-4 + CapOx
OAC-HGFI	LUD2015-005*	PreTx; ICI-only	OAC; Gastric; Oesophagus	Endoscopic	Slow-frozen	αPD-L1/CTLA-4 + CapOx
OAC-JCNP	LUD2015-005*	PreTx; ICI-only; PostTx	OAC; Gastric; Oesophagus	Endoscopic	Slow-frozen	αPD-L1/CTLA-4 + CapOx
OAC-LPPN	LUD2015-005*	PreTx; ICI-only; PostTx	OAC	Endoscopic	Slow-frozen	αPD-L1/CTLA-4 + CapOx
OAC-PAHE	LUD2015-005*	PreTx; ICI-only; PostTx	OAC	Endoscopic	Slow-frozen	αPD-L1/CTLA-4 + CapOx
OAC.Op-IKIJ	LUD2015-005	PreTx; ICI-only; PostTx	OAC (Operable); Oesophagus (PostTx)	Endoscopic; Core Biopsy (PostTx)	Slow-frozen; PostTx (fresh)	αPD-L1 + PC + RadioTx
OAC.Op-JJHD	LUD2015-005	PreTx; ICI-only; PostTx	OAC (Operable); Oesophagus (PostTx)	Endoscopic; Core Biopsy (PostTx)	Fresh	αPD-L1 + PC + RadioTx
OAC.Op-ODHL	LUD2015-005	PreTx; ICI-only; PostTx	OAC (Operable): Oesophagus (PostTx)	Endoscopic; Core Biopsy (PostTx)	Slow-frozen; PostTx (fresh)	αPD-L1 + FLOT
OSCC-FLJG	LUD2015-005	PreTx; ICI-only	OSCC	Endoscopic	Slow-frozen	αPD-L1 + CapOx
OSCC-JHIE	LUD2015-005	PreTx; ICI-only	OSCC	Endoscopic	Slow-frozen	αPD-L1/CTLA-4 + CapOx
OSCC-LAFI	LUD2015-005	ICI-only	OSCC	Endoscopic	Slow-frozen	αPD-L1/CTLA-4 + CapOx
BARR-CMAG	TGU	NA	Oesophagus; BO	Endoscopic	Fresh and slow-frozen	NA
BARR-IMPO	TGU	NA	BO	Endoscopic	Slow-frozen	NA
BARR-INPN	TGU	NA	Oesophagus; BO	Endoscopic	Fresh and slow-frozen	NA
BARR-MEPJ	TGU	NA	BO	Endoscopic	Slow-frozen	NA

Table 5.1: Sample summary for the LUD2015-005 Oesophageal Pathology Cell Atlas. Details are provided for each deidentified patient ID in this dataset. TGU: Translational Gastroenterology Unit, the biobank providing supplemental biopsies from BO patients. *Patient was a part of the analysis cohort of inoperable OAC described in Chapters 3-4. OAC.Op: Operable OAC, taken from other arms of the LUD2015-005 study. Treatment abbreviations: RadioTx = Radiotherapy; FLOT = 5-fluorouracil, leucovorin, oxaliplatin, and docetaxel; PC: paclitaxel and carboplatin; CapOx: Capecitabine and oxaliplatin.

5.2 Evaluation of suitability of slow-frozen whole endoscopic biopsies for scRNA-seq

5.2.1 Background

One of the key considerations in scRNA-seq experimental design is minimizing potential batch effects, meaning differences in transcriptional profiles due to technical factors which can confound the meaningful biological variation. A well-known source of significant batch effect is the use of different scRNA-seq technologies[187,188]. However, even for intra-institutional datasets created using a single technology, technical variation can still be introduced by other factors, including processing samples in different batches[189].

In an ideal case, this source of technical variation could be mitigated by processing all samples in a single batch, or by designing projects with balanced batches such that technical and biological variation can be disambiguated. With cell line work, it is usually feasible to execute these pre-planned multi-sample batches. However, material from clinical trials often arrives from a single patient at a given time, making it difficult to create multi-patient batches using standard workflows, which require the timely use of fresh biopsies.

Cryopreserved material would allow for multiple stored samples to be processed at once, thereby reducing potential batch effect, cost, and operator time. Two such approaches have been developed for scRNA-seq from tissue samples: first, cells can be dissociated from tissue biopsies upon collection and cryopreserved in suspension for later use, either by standard controlled-rate freezing ("slow-frozen") or cold storage in a methanol fixative[190]. Alternatively, nuclei (rather than intact cells) can be isolated from previously snap-frozen tissue[191]. However, both methods have limitations. Cryopreservation of cell suspensions with current methods requires independent dissociation of tissue biopsies prior to cryopreservation, which could be a source of batch effect due to varying enzyme concentration or dissociation time. This is not necessarily the case for nuclei isolation from snap-frozen material; however, nuclear RNA makes up only a subset of the whole cell transcriptome. Importantly, neither methanol fixation nor nuclei isolation allow for isolation of live cells after freezing, which limits the potential utility of material stored in these ways in other assays.

5.2.2 Performance of fresh and slow-frozen biopsies for dissociation and scRNA-seq

For the LUD2015-005 study, up to half of the biopsies at each collection were cryopreserved using controlled-rate freezing in serum with 10% DMSO prior to long-term storage at -80°C (see Methods), with the goal of preserving live cells for later use. It was decided to explore whether it would be possible to use these slow-frozen whole biopsies for scRNA-seq, and whether data from both fresh and frozen material could be integrated to create an expanded scRNA-seq resource. A similar optimized dissociation protocol was therefore developed for both fresh and slow-frozen whole biopsies (see Methods).

In total, 59 samples from 18 individuals were processed for scRNA-seq using a 5'-based method with microfluidic droplet encapsulation (10X Genomics), often with multiple tissues and treatment timepoints profiled for the same patient (Table 5.1). 13 samples were prepared from fresh biopsies, and 46 from slow-frozen biopsies.

Viability of input cell suspensions was significantly lower when prepared from frozen material than from fresh (Figure 5.1A), often significantly lower than the 70% viability threshold recommended by the manufacturer[n]. This was likely in part due to the extended length of storage at -80°C (longest storage: 3.3 years, OSCC-FLJG) and the absence of a dead cell depletion strategy for the cryopreserved dissociation protocol (see Methods). However, it was still possible to generate successful cDNA and final libraries from both preparation methods.

After sequencing and pre-processing, approximately 175,000 cell barcodes from these 59 samples were called as true cells. However, a significant proportion of cells showed signs of death or membrane compromise, namely a high percentage of mitochondrial UMIs and low number of unique features detected. As an initial quality control (QC) filter, these low-quality cells were removed by selecting only cells with at least 200 unique features and less than 25% mitochondrial UMIs for analysis. In general, a much lower fraction of cells passed this QC filter from slow-frozen material compared with from fresh material (Figure 5.1B). The percentage of cells passing this filter showed a significant correlation with the viability of the input cell

[n] Manufacturer communication: "Tumor Dissociation for Single Cell RNA Sequencing (Demonstrated Protocol)". *https://support.10xgenomics.com/single-cell-gene-expression/sample-prep/doc/demonstrated-protocol-tumor-dissociation-for-single-cell-rna-sequencing*. Accessed 23 June 2021.

suspension (Figure 5.1C), suggesting the higher percentage of low-quality cell barcodes likely corresponded to the higher fraction of cells called as non-viable by trypan blue staining. In addition to this QC threshold, some cell barcodes were removed if they belonged to clusters whose marker gene profile was suspected to be that of doublet, ambient, or non-viable cells (Figure 5.2). Despite the increased number of cell barcodes removed by these filters in slow-frozen tissue, the remaining high quality-cells had similar levels of unique features and UMI counts detected per cell between methods (Figure 5.3). This suggests that the quality of information captured by cell barcodes that pass QC thresholding is not significantly different between conditions, and that high-quality cells derived from slow-frozen tissues can provide transcriptome profiles of similar quality to those from fresh tissues.

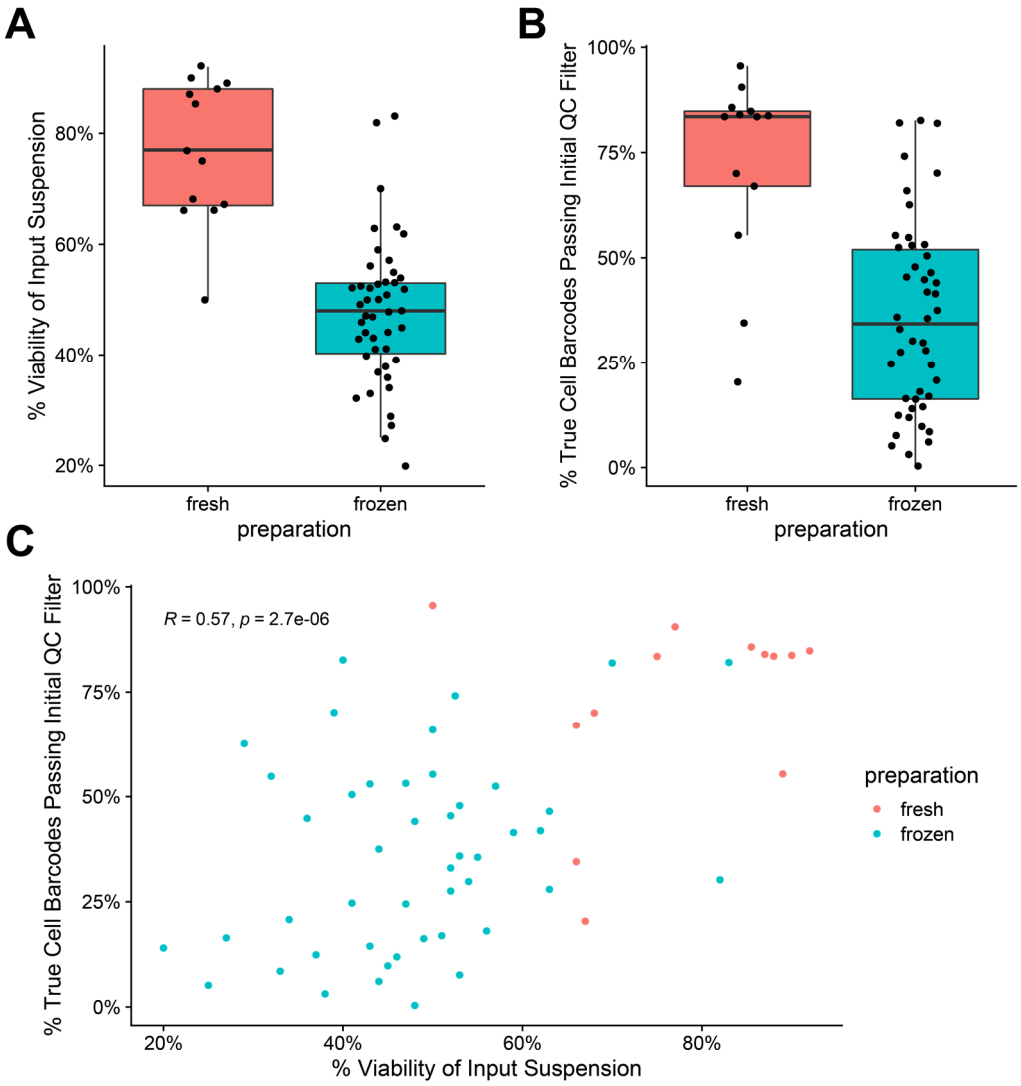


Figure 5.1: Comparison of viability and scRNA-seq performance of cell suspensions produced by dissociation of freshly-isolated and slow-frozen tissue biopsies. A) Viability of input cell suspensions immediately prior to loading onto the Chromium Controller for scRNA-seq, as measured by trypan blue exclusion. B) The percentage of true cell barcodes (as called by Cell Ranger) that had 200 unique features and less than 25% mitochondrial UMIs, thresholds which were used as the initial QC filter before downstream analysis. C) The relationship between viability and the percentage of cell barcodes passing the initial QC filter. Pearson correlation coefficient and p-value are shown.

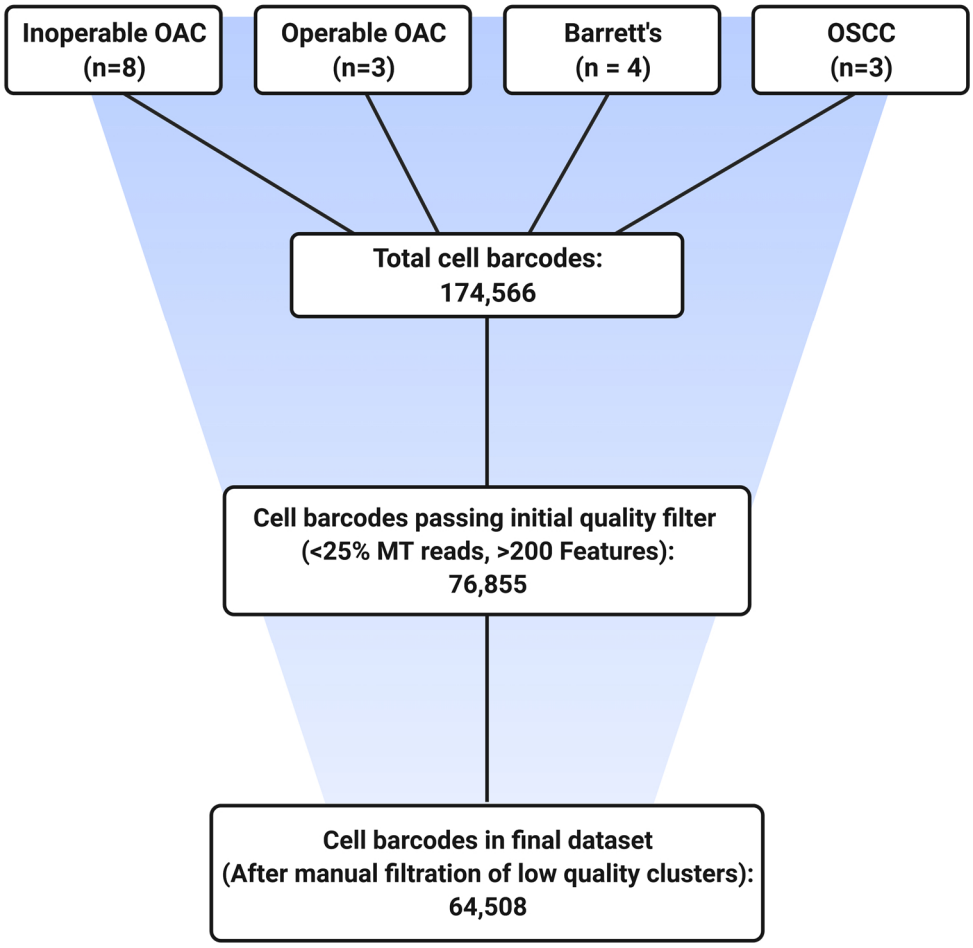


Figure 5.2: Cell barcode filtering summary for the LUD2015-005 cell atlas. Diagram showing the number of cell barcodes remaining after each filtration step. Total cell barcodes (top) refers to the number of true cell barcodes called by Cell Ranger in the filtered feature-barcode matrix.

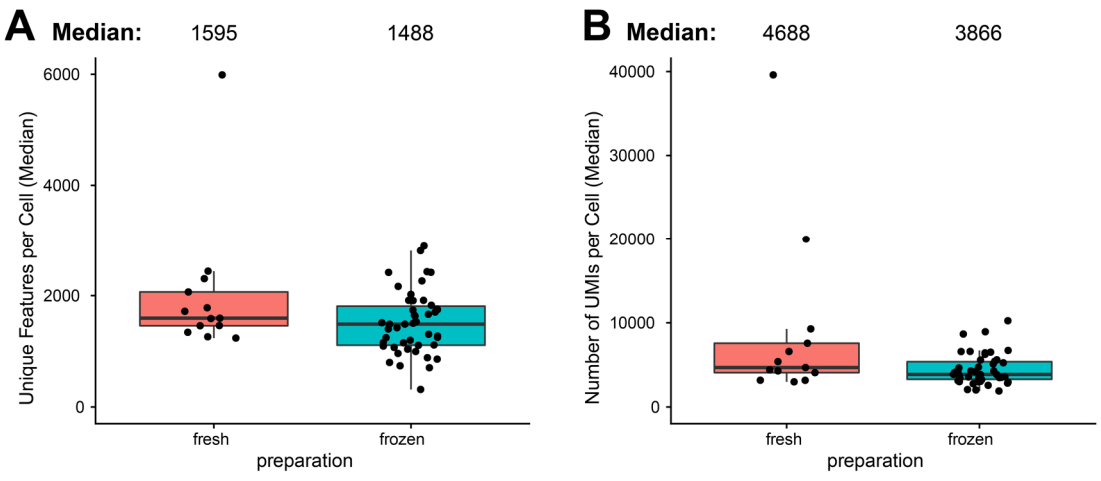


Figure 5.3: Comparison of transcriptome quality for high-quality cell barcodes that survive QC filtering. A) A summary of the unique features per cell, with each dot representing the median value for each sample. The median of these sample-wise medians for each preparation method is displayed along the top of each plot. B) Same as in (A), but comparing the median number of UMIs detected per cell for each sample.

5.3 A single-cell atlas of OAC, OSCC, Barrett's, and normal GI control tissues

5.3.1 Overview of broad cellular compartments

Following QC, all cell barcodes in the final dataset were subjected to dimensionality reduction and clustering. Cells appeared to fall into four broad compartments (Figure 5.4A): 1) Epithelium (malignant and non-malignant); 2) Stroma (endothelium, fibroblasts, and mural cells); 3) Lymphocytes (T, B, Plasma, and NK cells), and 4) Phagocytes (monocytes, macrophages, neutrophils, dendritic cells, and mast cells). While some potential batch effect was seen between the different tissue preservation methods (Figure 5.4B), the general clustering appeared to be driven mainly by the biological compartment across the full dataset, with clear expression of expected compartment markers across preparation methods. For instance, Vimentin^{-}E-cadherin^{+} epithelial cells, Vimentin^{+}CD45^{-} stromal cells, Vimentin^{+}CD45$^{+/mid}$CD11b^{-} lymphocytes, and Vimentin^{+}CD45^{+}CD11b^{+} phagocytes from the myeloid lineage resolved into discrete groupings regardless of preparation method (Figure 5.4C-F).

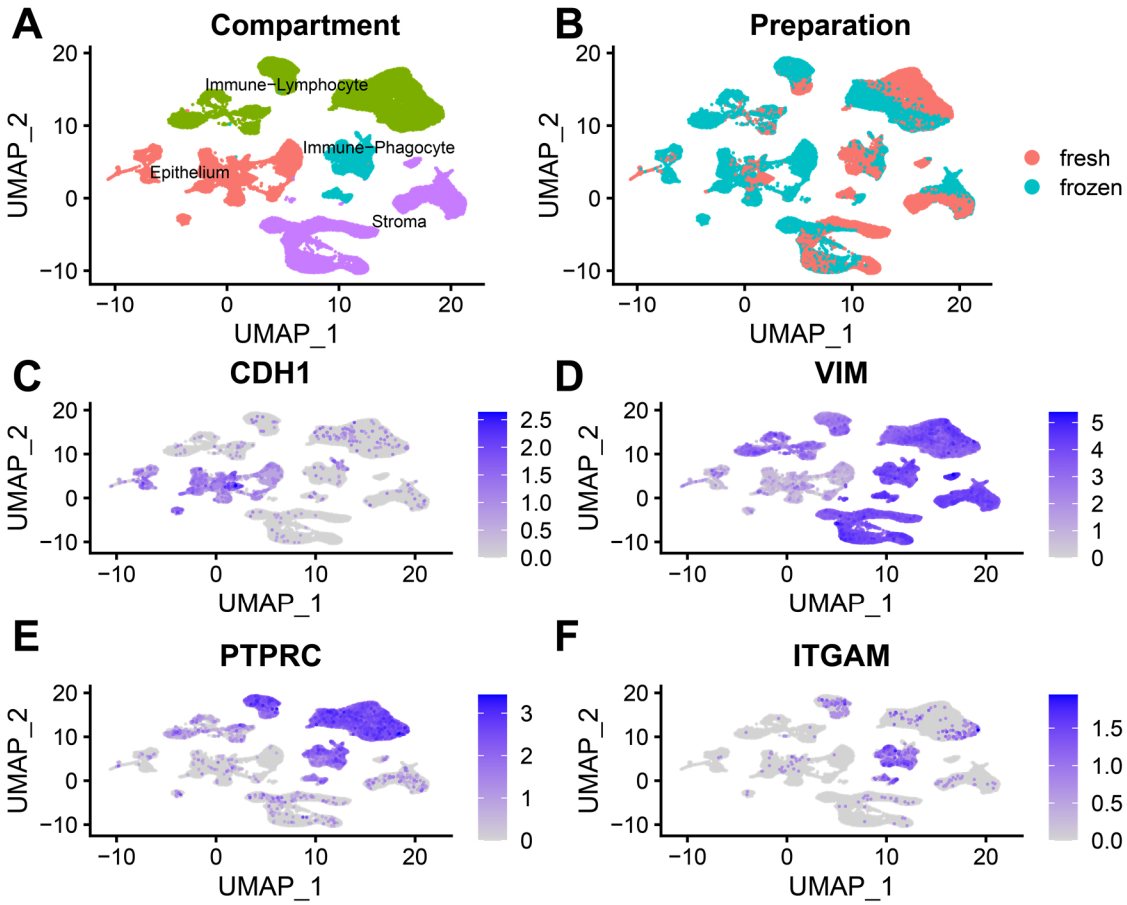


Figure 5.4: Overview of broad cellular compartments in the LUD2015-005 scRNA-seq atlas. UMAP dimensionality reductions of all cells surviving final QC and filtering. A) Points are coloured by the identified cellular compartment. B) Points are coloured by the tissue preparation method, where frozen refers to slow-frozen biopsies, and fresh refers to freshly collected and dissociated material. Bottom: Plots showing markers of compartment identity, where points are coloured by expression of C) *CDH1*/E-Cadherin, D) *VIM*/Vimentin, E) *PTPRC*/CD45, and F) *ITGAM*/CD11b.

5.3.2 Addressing cryopreservation-induced batch effect

5.3.2.1 Characterisation of expression variation between fresh and slow-frozen biopsies

The effect of a freeze-thaw cycle prior to dissociation on gene expression is unknown, but it is possible it could induce a stress response that could confound transcriptome profiles. Indeed, Figure 5.4 suggests that there may be some effect on expression profiles, which may cause confounding signal particularly within compartments. To assess what might be driving this confounding signal, DEGs were identified within compartments between cells from freshly-isolated and slow-frozen biopsies were identified. These DEGs were then filtered to select those that showed the same direction of change in at least three of four compartments, to identify genes that consistently drive technical variation between the two preparation methods across cell types. As shown in Figure 5.5A, almost all of the significant genes in this analysis were consistently upregulated in freshly-isolated cells compared with cells from slow-frozen biopsies. These include cell cycle (*CDKN1A*) and apoptosis regulators (*MCL1*, *BCL3*), as well as inflammatory stress response genes (*FOSL2*, *REL*, *JUND*, *IER2*). The variation in these genes between preparation methods appeared to be strongest in the phagocyte and stromal compartments (Figure 5.5A-B). It was unclear whether this meant that fresh dissociation induced stronger overexpression of these genes, or that slow-freezing reduced their expression below physiological levels. However, it did not appear that the freezing and thawing process induced increases in major stress responses that would confound biological signal; if anything, stress response genes showed reduced expression by this method. Indeed, a previously reported 17-gene scRNA-seq signature of dissociation-induced stress response[190] showed decreased levels in slow-frozen cells from all compartments, particularly the stroma (Figure 5.5C). Therefore, while there is likely to be technical variation between freshly isolated and slow-frozen biopsies that will need to be accounted for, using slow-frozen material may actually reduce confounding stress responses associated with standard enzymatic dissociation, at least in cells that survive QC.

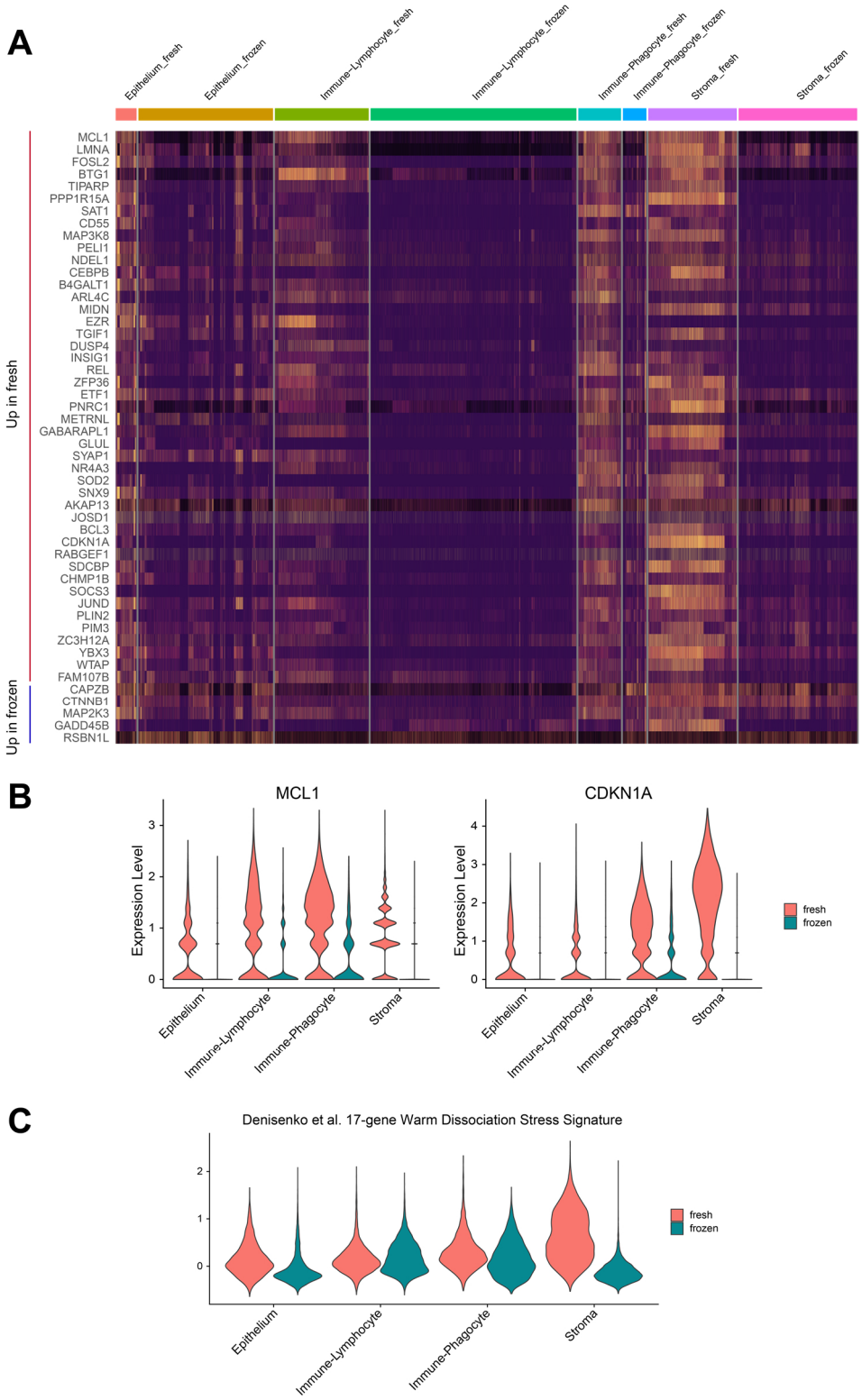


Figure 5.5: Technical variation in gene expression caused by differences in dissociation method. A) Top markers returned by FindConservedMarkers (Seurat) for differential expression across preparation methods seen in multiple compartments. For genes where fold changes agreed for at least three of four compartments, the top 50 genes (with lowest maximum p-value) were selected. The rows are sorted by the average upregulation in fresh compared to frozen, with the directionality of the conserved changes marked. Expression is given as scaled normalized counts, with yellow representing higher expression. B) Log-normalized expression of two example genes upregulated in fresh tissue. C) Gene module score of 17 genes reported by Denisenko et al. as a dissociation stress signature upregulated by dissociation at 37C, as used in the LUD2015-005 dataset.

5.3.2.2 Minimizing cryopreservation batch effect using a mutual nearest-neighbour approach

While the difference between cellular compartments outweighed the effect of the tissue preparation method across the full dataset, some clustering differences likely driven by this technical variation were still apparent. This became much more visible and problematic when conducting separate analysis within cell compartments to identify cell types and subtypes, as the technical variation in gene expression caused by the preparation method began to outweigh the biological variation between cell subtypes. As an example, when clustering was performed on just mast cells (*TPSAB1*$^{+}$*KIT*$^{+}$ cell clusters), separate groupings appeared upon dimensionality reduction for each tissue preparation method (Figure 5.6). This batch-driven grouping outweighed apparent biological signal, such as the presence of a subset of mast cells expressing a variant mast cell tryptase *(TPSD1*). *TPSD1*^{hi} cells were present in both conditions but clustered separately. This variation appeared to be driven by some of the key batch-responsive genes profiled in 5.3.2.1, including *CDKN1A* and *REL*, as well as some other stress response (*FOSB*) and mitochondrial genes (*MT-RNR2*). This suggested the clustering within this cell type was likely driven by technical variation related to cellular responses to the different dissociation methods, rather than markers of biologically discrete mast cell subsets.

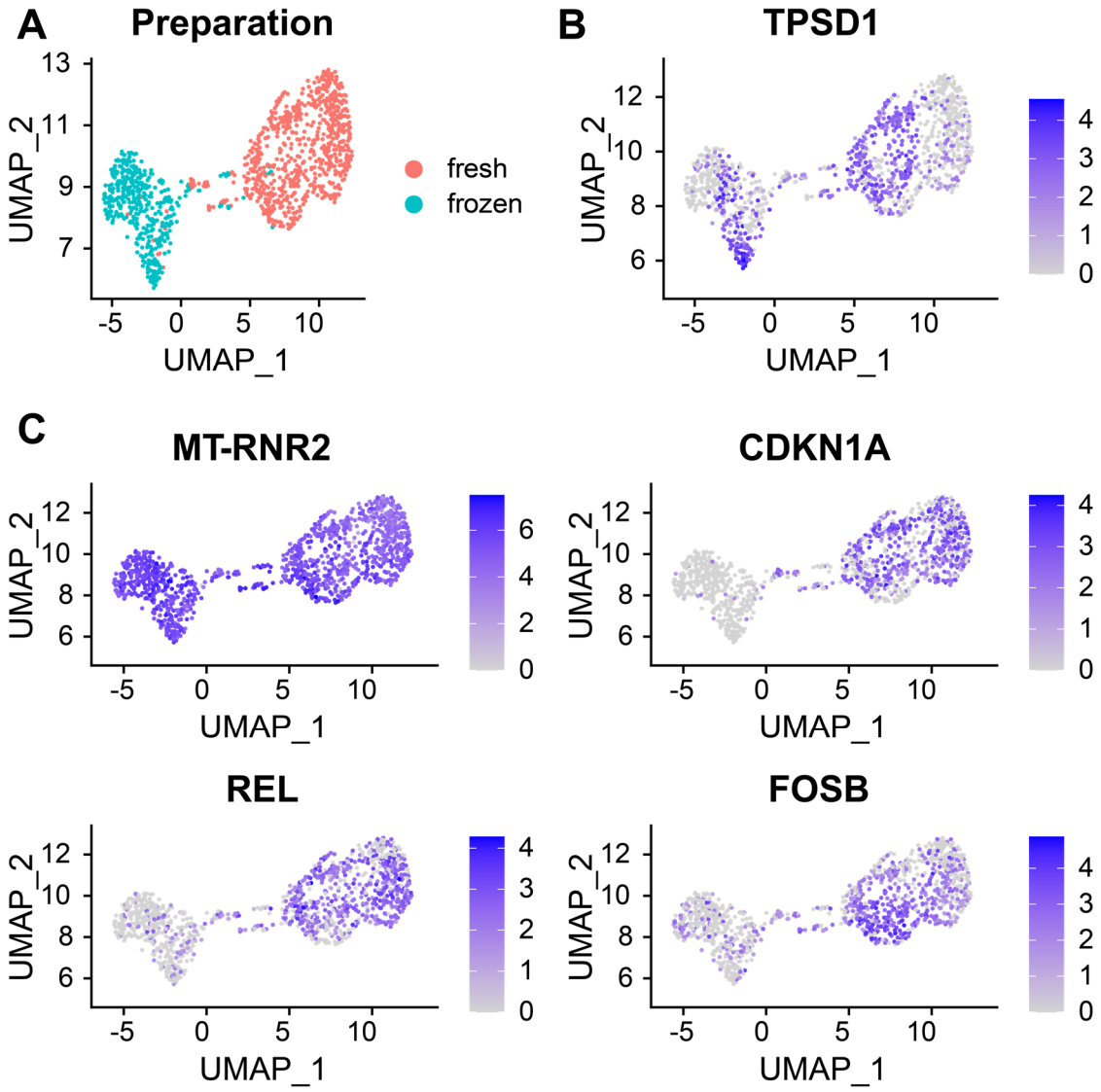


Figure 5.6: Mast cell dimensionality reduction without batch effect correction. A) UMAP visualization following standard SCTransform processing, with cells coloured by preparation method. B) Log-normalized expression of presumptive marker of biological variation in mast cells. C) Log-normalized expression of four genes that appear to be correlated with technical variation between fresh and slow-frozen biopsies.

To reduce this confounding effect, a mutual nearest-neighbour (MNN) approach was employed to correct for technical variation between fresh and slow-frozen batches. This method appeared to successfully reduce the effect of technical variation, and after application, mast cells were resolved into *TPSD1*^{hi} and *TPSD1*^{lo} cell subsets regardless of tissue preparation method (Figure 5.7). *TPSD1* expression has been shown to increase in response to *in vitro* culture with a combination of mast cell activation cytokines[192], and therefore it appears likely that clustering driven by *TPSD1* variation better represents true biological variation between mast cell

subtypes, which is more clearly visible following removal of the technical variation between slow-frozen and freshly-isolated tissues. Therefore, this MNN approach was deemed effective in minimizing batch effects due to preparation method on this dataset, and was taken forward for the identification of cell types and subtypes within compartments in the following section (5.3.3). MNN batch effect reduction was conducted on all compartments and sub-compartments in this analysis, with the exception of the gastric/neuroendocrine (almost entirely cells from slow-frozen material) and OAC/BO sub-compartments (highly divergent malignant cells with patient-specific clustering means MNNs may not be appropriately identified).

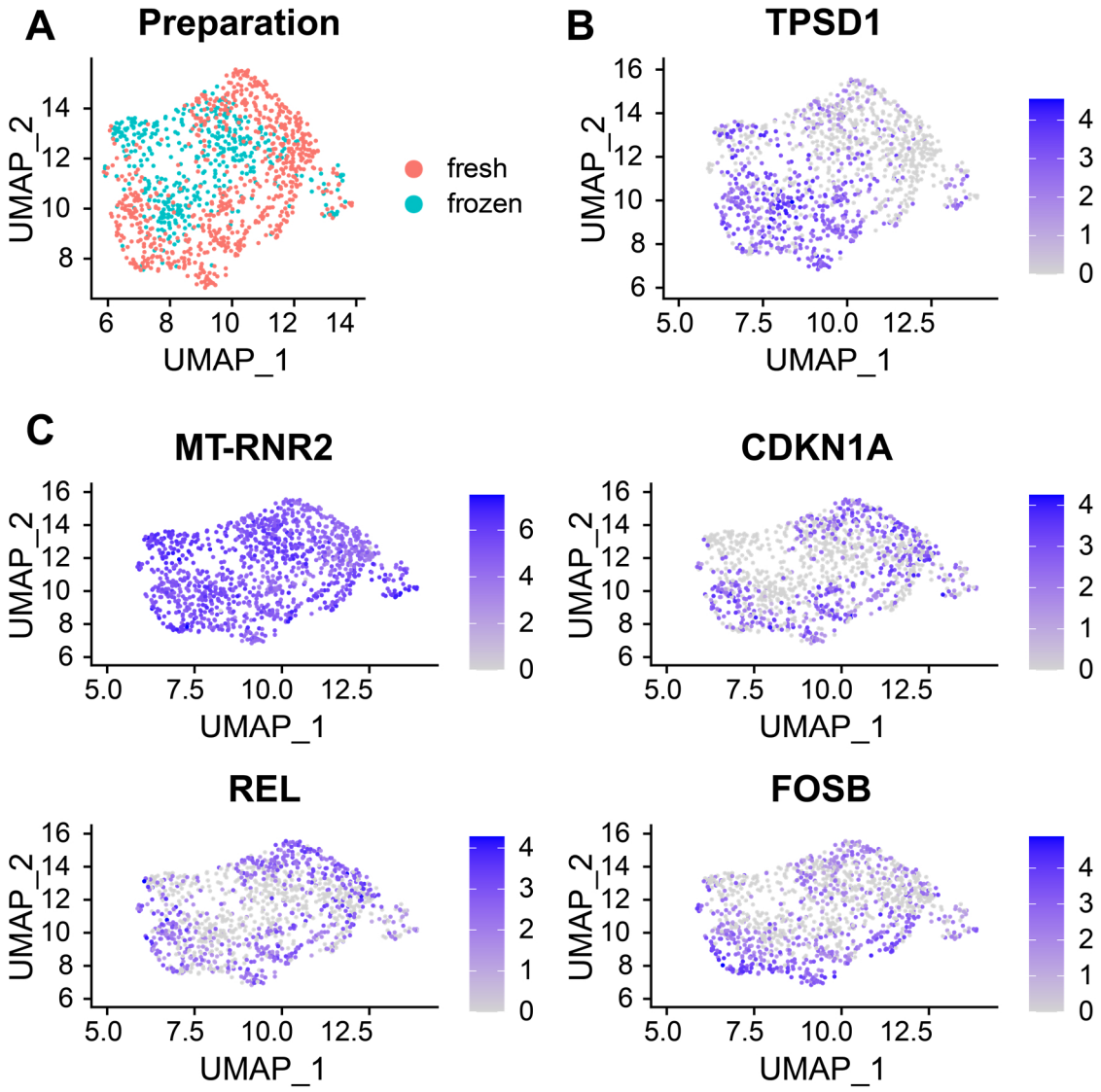


Figure 5.7: Mast cell dimensionality reduction following batch effect correction with FastMNN. A) As in 5.6, UMAP visualization with cells coloured by preparation method, but following MNN-based batch correction. B) Log-normalized expression of presumptive marker of biological variation in mast cells. C) Log-normalized expression of four genes correlated with technical variation between preparation methods.

5.3.3 Identification of cell types represented in the LUD2015-005 scRNA-seq cell atlas

5.3.3.1 Epithelial compartment

The information contained in the LUD2015-005 scRNA-seq atlas could be a useful resource to those studying oesophageal pathologies, but the biological importance of the cell types present may be opaque to the general researcher due to the highly-specialized nature of both scRNA-seq analysis and cell type identification. Therefore, as a primer to the utility of this dataset, the major cell types were identified and profiled on the basis of marker expression. First, epithelial cells were characterized, which broadly broke into squamous and non-squamous sub-compartments. Both sub-compartments had contribution from biopsies of malignant (OAC and OSCC) and non-malignant tissues. Due to the heterogeneity apparent in the non-squamous sub-compartment, cells in this category were further split into further subgroups for cell type identification, one containing gastric and neuroendocrine cells, and the other dominated by other non-squamous epithelial cells from BO and OAC, which were presumably metaplastic or malignant epithelium.

Neuroendocrine cells expressing a range of neuropeptide hormones and neural markers were contributed mainly from BO, OAC, and gastric biopsies, with only a single cell from oesophagus and none from OSCC, suggesting the presence of this cell type is a general feature of non-squamous metaplasia or malignancy in the oesophagus. Four major neuroendocrine subtypes were identified (Figure 5.8). The largest neuroendocrine subtype had high expression of ghrelin (*GHRL*), a neuropeptide known to be secreted by the stomach to modulate hypothalamic control of food intake[193]. This subtype had a large contribution from gastric tissue, but also OAC. An $HDC^{+}CHGA^{bright}$ subtype was more specific to gastric tissue. Interestingly, both BO and OAC tissue contributed a subtype expressing the neuroendocrine progenitor marker *NEUROG3* and dopamine metabolism genes (*DDC*, *GCH1*), which were rarer in the normal gastric tissue. A final subtype expressed weaker levels of neuroendocrine markers (*CHGA*), but also markers of normal gastric chief cells (*PGA5*), and could thus represent a doublet or intermediate cell type.

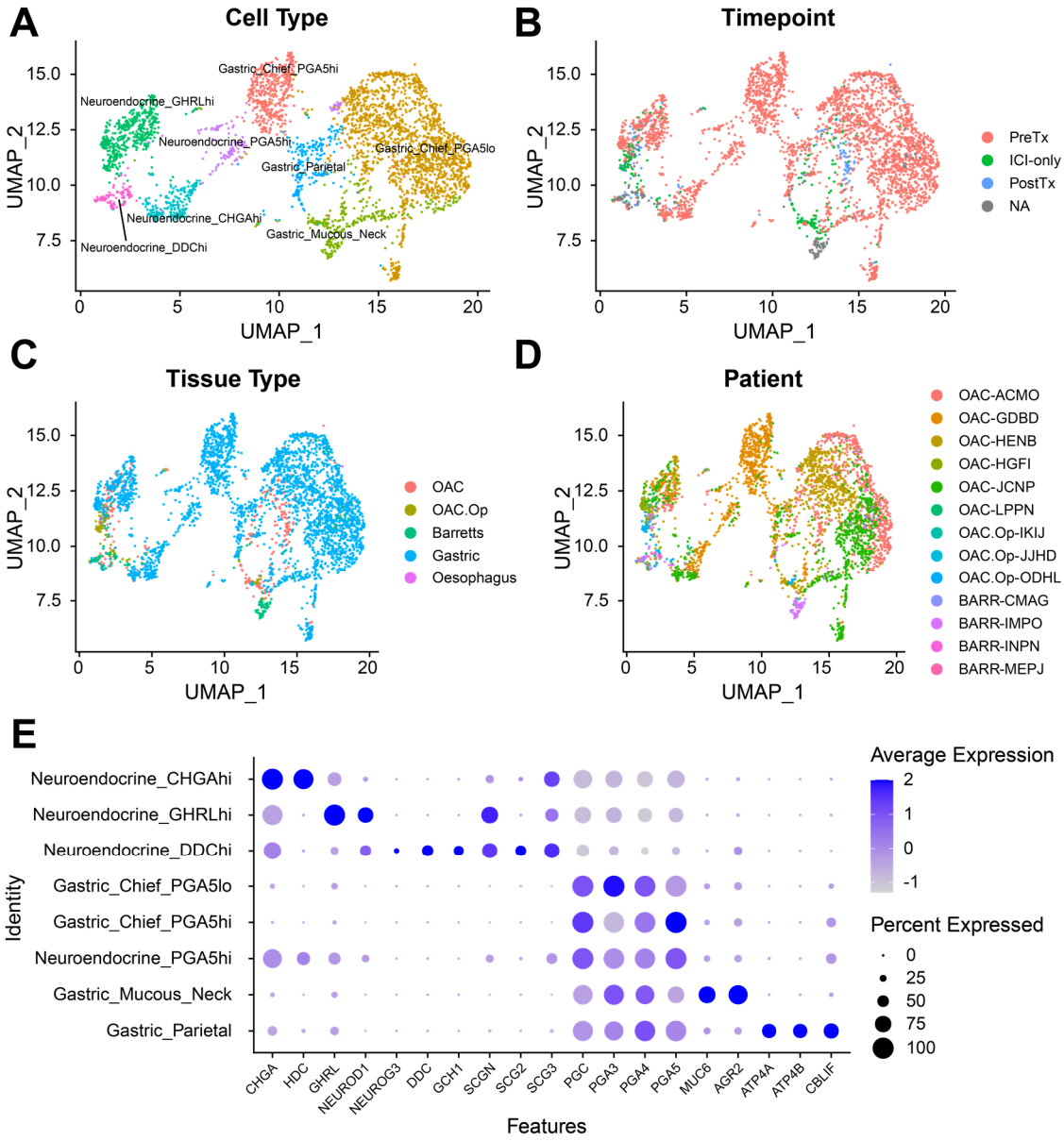


Figure 5.8: Cell map of the gastric and neuroendocrine sub-compartment. Top and middle: UMAP visualizations coloured by the identified cell type (A), LUD2015-005 treatment timepoint (B), tissue type (C), and patient of origin (D). For B, dots without annotated treatment (NA) are from BO patients outside of the LUD2015-005 trial. Bottom (E): Dot plot of selected markers that differentiate identified cell types. The size of the dot represents the percent of cells within each column expressing the gene, while the colour is the average strength of expression. Rows are ordered by hierarchical clustering.

Gastric-like cells in this dataset (Figure 5.8) were dominated by cells expressing chief cell markers, particularly pepsinogens (*PGC*, *PGA3*, *PGA4*, *PGA5*). These were classified into two major subpopulations based on their expression of *PGA5*, with the *PGA5*^{hi} cluster contributed mainly by a single patient. Smaller groups of mucous-neck cells (*AGR2*, *MUC6*) and acid-producing parietal cells (*ATP4A*, *ATP4B,* gastric intrinsic factor/*CBLIF*) were also present. Another major population in gastric biopsies should be foveolar cells (also called surface

mucous cells), expressing gastrokines (*GKN1*/*GKN2*) and *TFF1*/*MUC5AC*[32,194], but these were not found in this dataset. As normal gastric biopsies (all slow-frozen) were profiled from five separate patients with good epithelial yield, it is unlikely a major cell type was missed due to low cell numbers. It is possible this cell type does not survive cryopreservation or is not isolated well by the dissociation protocol used. Regardless, while the remaining gastric cell types were mainly contributed from gastric biopsies, OAC and BO biopsies also contributed to each of the mucous-neck, parietal, and *PGA5*^{low} chief cell types, suggesting that gastric metaplasia is present in these conditions which involves many of the gastric epithelial cell types.

Cells from BO and OAC biopsies that did not cluster with neuroendocrine or gastric epithelial cells were classified as Barrett's metaplasia and malignant OAC epithelium (BO/OAC sub-compartment). Among this sub-compartment, most cells from OAC patients tended to group into separate patient-specific clusters, suggesting these were malignant epithelium that exhibited significant transcriptional differences between individual patients (Figure 5.9). BO clusters were identified as the non-squamous cells from BO; these BO cells showed more overlap than OAC, with several clusters shared between multiple patients, although there were some patient-specific clusters. Some cells from OAC showed significant transcriptional overlap with BO clusters, away from the major patient-specific OAC clusters for those patients. This perhaps represented BO-like cells intermixed in OAC biopsies, which were particularly prevalent at the PostTx timepoint. Interestingly, OAC cells from earlier-stage operable tumours clustered closer with BO than cells from the metastatic tumours in the inoperable OAC cohort (Figure 5.9C), suggesting increased transcriptional evolution away from the putative pre-malignant lesion in the later-stage tumours.

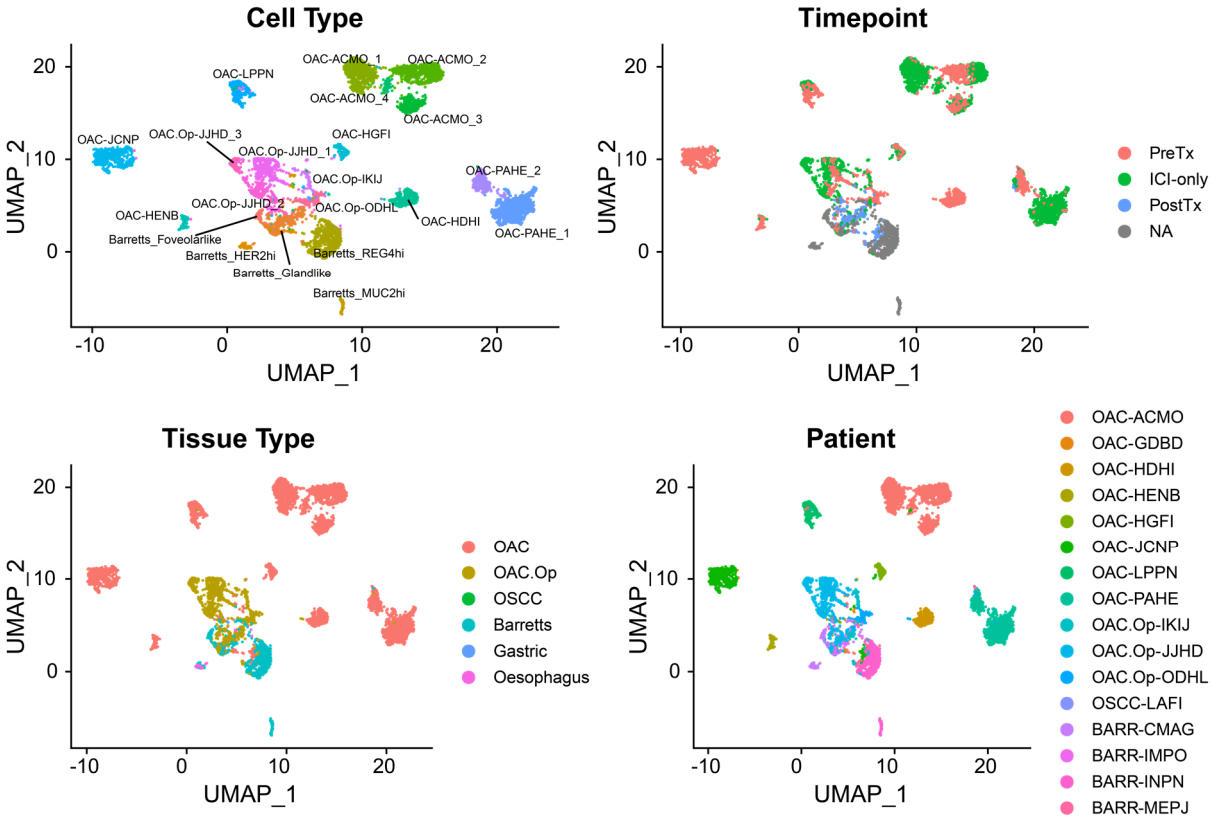


Figure 5.9: Cell map of the OAC/BO sub-compartment. UMAP visualizations coloured by the identified cell type (A), LUD2015-005 treatment timepoint (B), tissue type (C), and patient of origin (D). For B, dots without annotated treatment (NA) are from Barrett's patients outside of the LUD2015-005 trial.

A previous scRNA-seq study on BO from our group identified two major subtypes of non-neuroendocrine BO epithelium. The primary subtype in that dataset was a *LEFTY1*$^{+}$*OLFM4*$^{+}$ "Barrett's type", which was found to have transcriptional similarity to cells from submucosal glands in the normal oesophagus. The other subpopulation was goblet cells, including *ITLN1*$^{+}$*SPINK4*$^{+}$ early goblet cells and *MUC2*$^{+}$ differentiated goblet cells, which had some overlap. In the present dataset, five major clusters of BO epithelium were identified. The most prevalent cluster contained a number of *LEFTY1*$^{+}$*OLFM4*$^{+}$ Barrett's type cells, as well as a small number of *ITLN1*$^{+}$*SPINK4*$^{+}$ early goblet cells within the same cluster (Figure 5.10). However, many of the cells in this grouping were negative for all four of these markers. Instead, this BO cluster was best characterized by high expression of *REG4*, which is a marker of early secretory cells found deep within colonic crypts[195]. A smaller *MUC2*^{hi} cluster was also found in a single patient, likely representing more differentiated goblet cells (Figure 5.10). In addition to this these *MUC2*^{hi} and *REG4*^{hi} clusters, two clusters contributed by a single patient were identified: a cluster of BO cells expressing foveolar-like markers (*GKN1*/*TFF1*/*MUC5AC*), likely representing the gastric-type

metaplasia seen in some BO patients[196], and a HER2^{hi} cluster (Figure 5.11A). This HER2^{hi} patient-specific subtype was discrete from the gland-like and foveolar-like BO cells from the same patient (BARR-CMAG). Interestingly, an scRNA-seq-based copy number profiling method (inferCNV) suggested this overexpression of HER2 was downstream of a high-level copy number gain in these cells, to a similar extent seen in HER2-overexpressing OACs (OAC-ACMO and OAC-HDHI; Figure 5.11B-C). This putative amplification perhaps represented an early malignant or pre-malignant genomic change in a subset of cells in this BO biopsy. However, this cluster also had some contribution from the oesophagus biopsy of the same patient (17 oesophagus, 87 BO). It was unclear whether this was due to cross-contamination during the collection of biopsies, or perhaps represented some patient-specific submucosal gland phenotype that was not restricted to the region of visible BO.

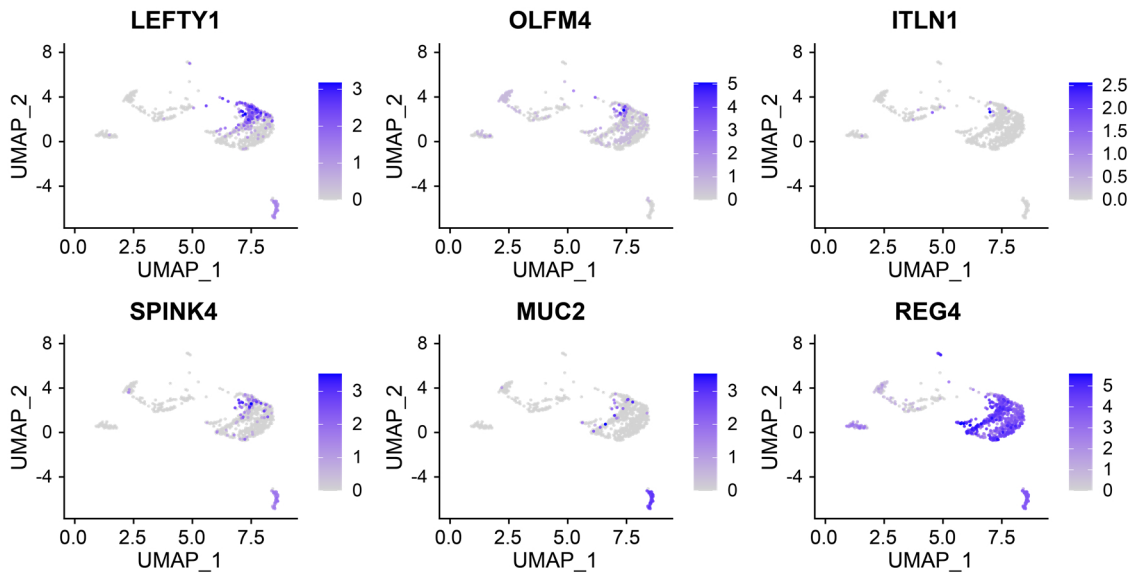


Figure 5.10: Assessing markers of Barrett's-type and goblet cell markers in non-squamous epithelium from Barrett's oesophagus. Dimensionality reduction from Figure 5.9, but limited to cells in BO clusters from a BO biopsy. Subplots are coloured by log-normalized expression of Barrett's-type (*LEFTY1*, *OLFM4*), early goblet (*ITLN1*, *SPINK4*) and differentiated goblet (*MUC2*) as described in Owen et al[32]. Expression of intestinal early secretory marker *REG4* is shown for comparison.

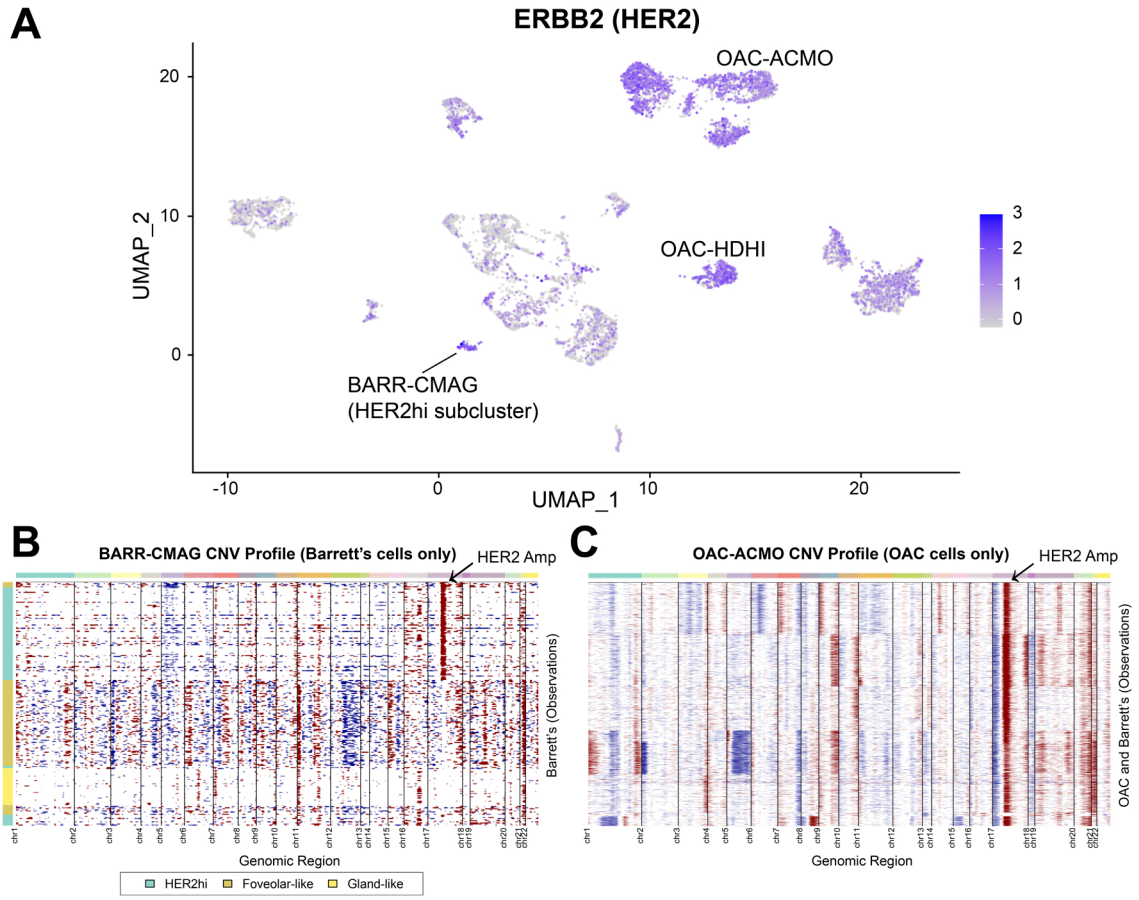


Figure 5.11: HER2^{hi} Barrett's subtype from BARR-CMAG shares features of HER2-amplified cancers. A) Log-normalized expression of *ERBB2* (HER2) in BO/OAC sub-compartment. HER2-overexpressing OAC and BO clusters are labelled. Bottom: Estimated copy-number variation (CNV) profiles of B) BO cells from BARR-CMAG and C) OAC and BO cells from OAC-ACMO. CNV levels are estimated on a scale of 0 (loss of both copies, deep blue) to >2 (3+ copy number gain, deep red) using normal cells as reference (not shown). For both B and C, the location of a 3+ copy HER2 gain is indicated with an arrow.

The remaining BO subcluster in this dataset had major contributions from two separate patients, and had a secretory phenotype (*SCGB2A1*$^{+}$), despite absent goblet or Barrett's-type markers. This raised the question of whether this could represent an another BO epithelial phenotype with resemblance to secretory submucosal glands, in addition to the previously described Barrett's-type. There was low yield of non-squamous cells from normal oesophagus in this dataset, suggestive of near absence of submucosal glands. However, the Human Cell Atlas (HCA) has published data from the normal oesophagus which included submucosal gland cells that were partitioned into duct-like and mucous-secreting subsets[197]. Interestingly, a number of markers were highly specific to both HCA submucosal glands (particularly in the duct-like cells) and this secretory BO subcluster, including *SCGB2A1*, *BPIFB1*, and *FOLR1*, suggesting close

resemblance between the two cell types (Figure 5.12; Figure 5.13). Additionally, these gland-like Barrett's cells showed reduced CNV events to other Barrett's clusters (Figure 5.11B), perhaps indicating a closer connection to a non-metaplastic cell of origin. However, gland-like Barrett's subtype was distinguishable from normal submucosal gland cells by a number of markers, including the gastric mucous neck marker *MUC6* (Figure 5.12; Figure 5.13). This gland-like Barrett's cell type could therefore represent another Barrett's phenotype with transcriptional similarity to oesophageal submucosal glands, which also shows some metaplastic features, including expression of gastric secretory markers which are not known to be present in the normal oesophageal submucosa.

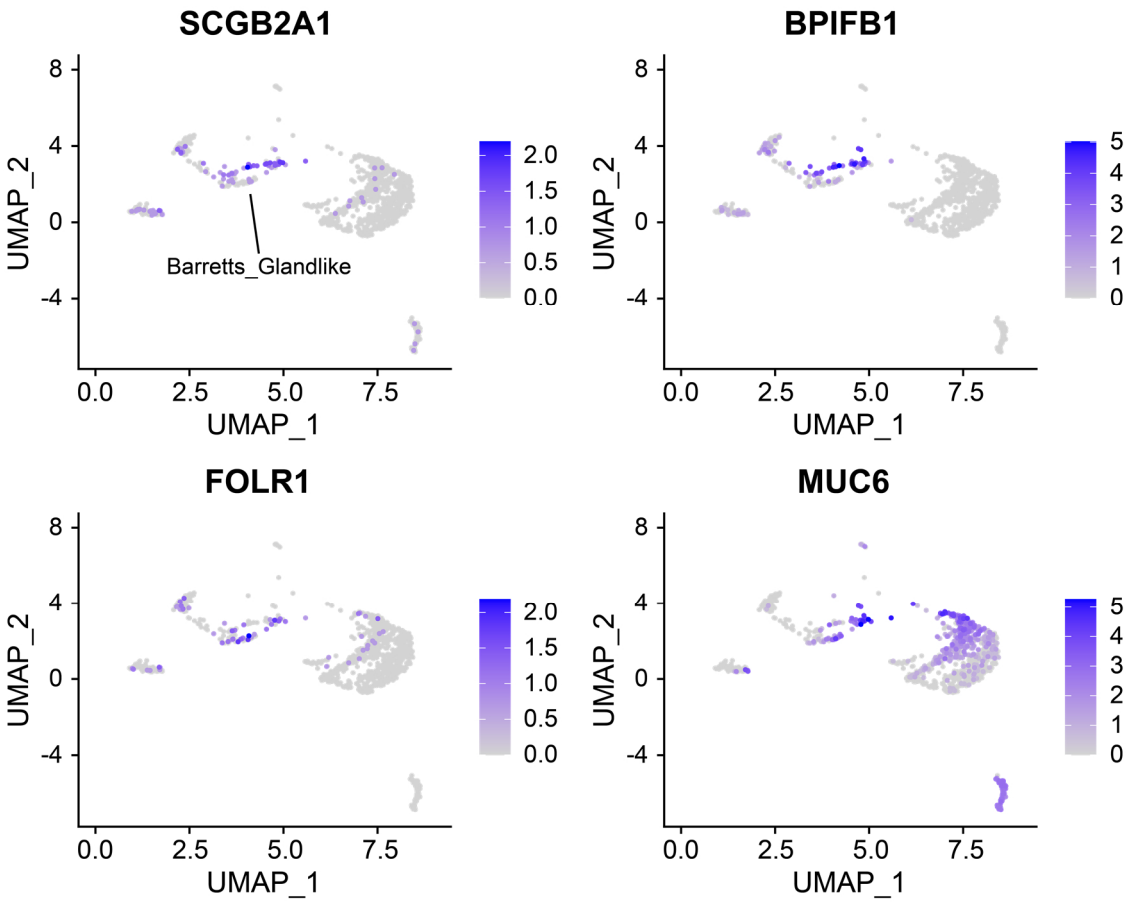


Figure 5.12: Selected markers of a submucosal gland-like Barrett's subtype. Dimensionality reduction from Figure 5.9, but limited to cells in BO clusters from a BO biopsy. Subplots are coloured by log-normalized expression of the indicated gene, with the gland-like grouping identified.

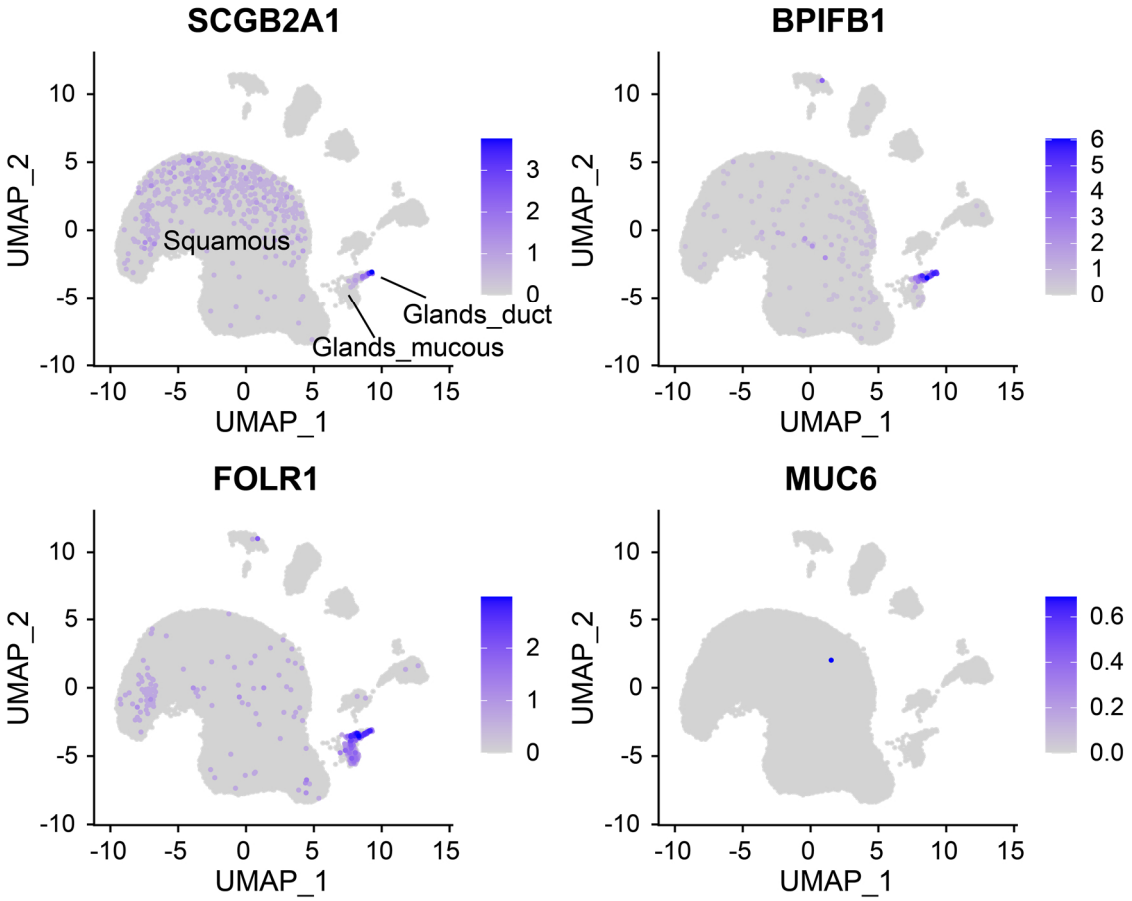


Figure 5.13: Expression of Barrett's gland-like subcluster markers in the Human Cell Atlas Dataset. Expression of markers identified in Figure 5.12 overlaid on the Human Cell Atlas dataset for normal squamous oesophagus[197]. The count data, dimensionality reduction, and cluster labels from the original publication were used to generate these plots. Labels show the approximate location of squamous and submucosal gland clusters as identified by the authors (using the July 2020 annotation update).

All OAC patients had at least one identifiable patient-specific cluster of malignant epithelium, with the exception of OAC-GDBD, where insufficient non-squamous cells survived QC to be resolved as a discrete cluster (n=18). Markers of OAC clusters were diverse, and will be expanded on further in the discussion of tumour evolution in section 5.5.3. The distribution of selected BO and patient-specific OAC markers is shown in Figure 5.14.

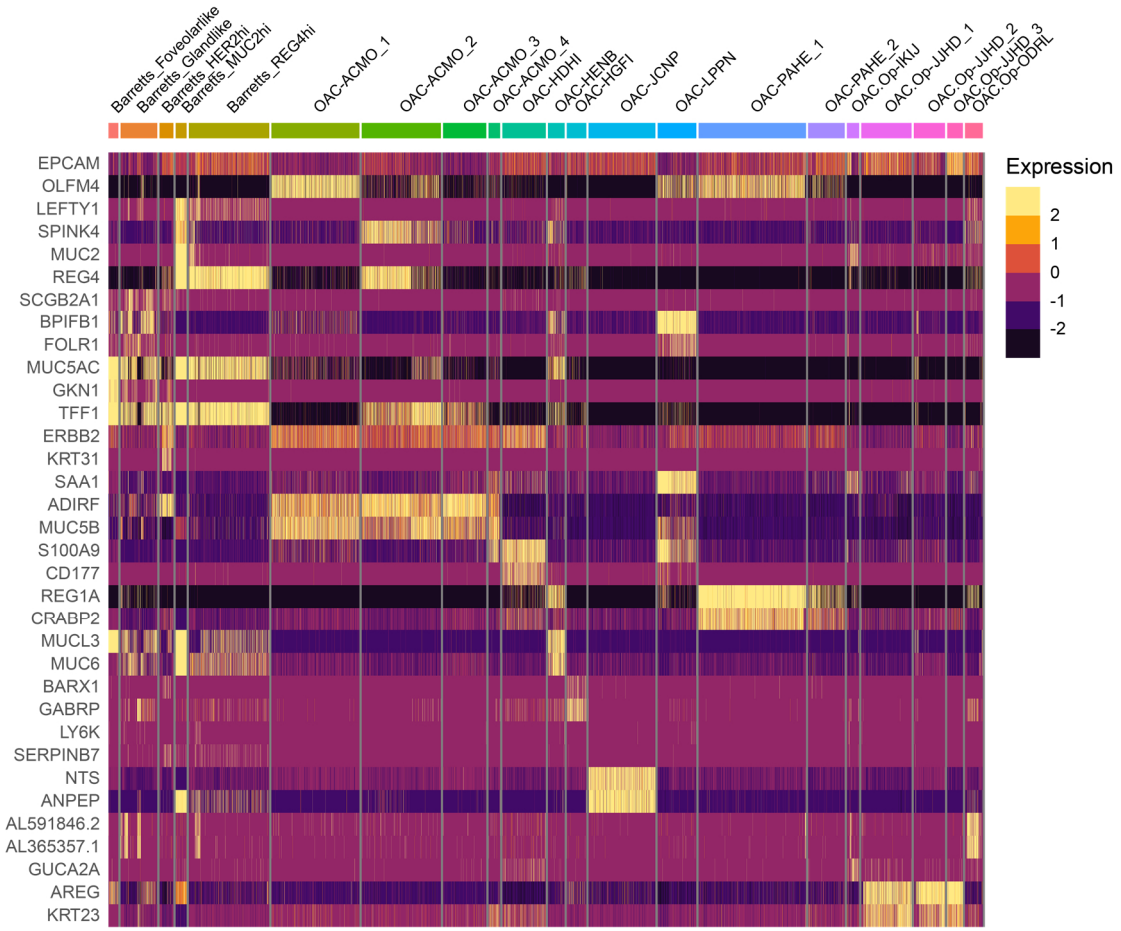


Figure 5.14: Selected markers of Barrett's and OAC subclusters. Scaled log-normalized expression values are shown, with each column representing a single cell grouped within identified cell types. The top 2 genes specific to each patient-specific OAC cluster and manually selected BO subcluster markers are used.

Within the squamous sub-compartment, normal squamous epithelium shared between multiple patients was identified, as well as distinct patient-specific OSCC clusters, similar to what was seen in the non-squamous compartment. While some cells identified as being in normal squamous clusters were derived from Barrett's, OAC, and OSCC biopsies, the vast majority of these clusters were contributed by normal oesophageal biopsies, suggesting the contamination of the profiled metaplastic and dysplastic oesophageal biopsies with normal squamous oesophagus is relatively low (Figure 5.15).

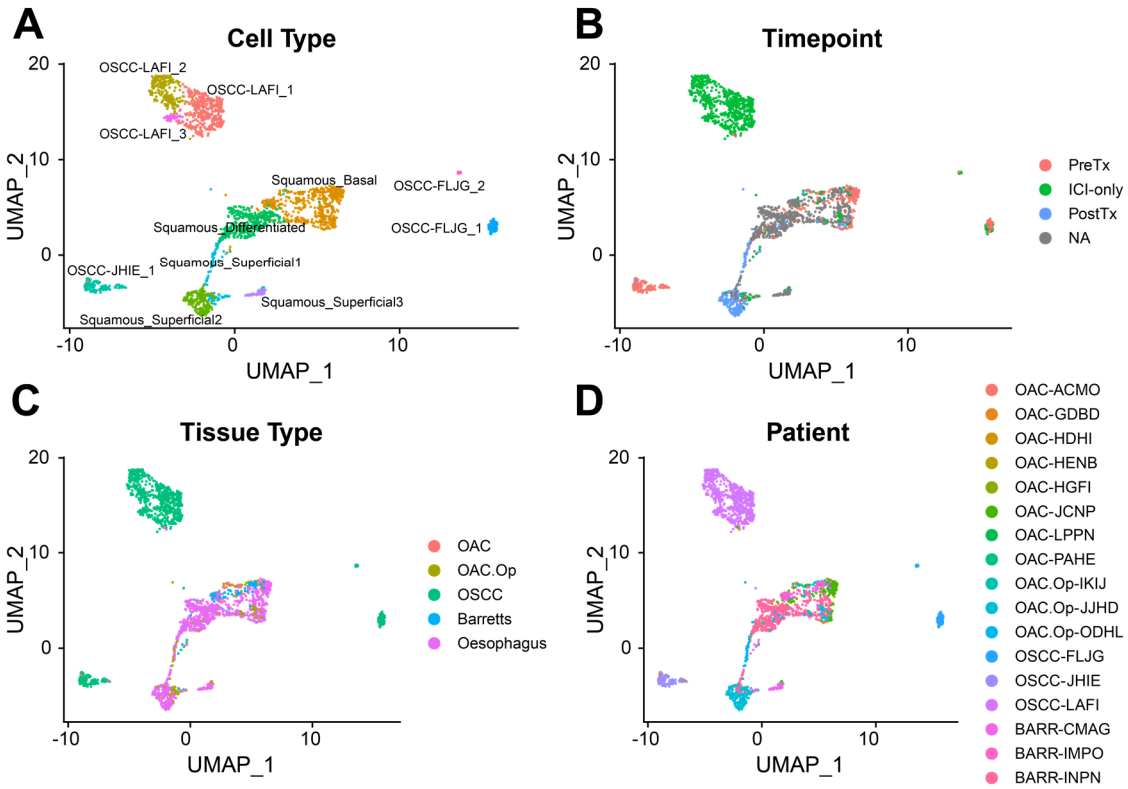


Figure 5.15: Cell map of the squamous sub-compartment. UMAP visualizations coloured by the identified cell type (A), LUD2015-005 treatment timepoint (B), tissue type (C), and patient of origin (D). For B, dots without annotated treatment (NA) are from Barrett's patients outside of the LUD2015-005 trial.

The HCA scRNA-seq analysis of the normal oesophagus identified basal, suprabasal, stratified, and superficial ("upper") clusters among 80,000 squamous epithelial cells[197]. In the present dataset, basal (*KRT15*^{hi}*TP63*^{hi}) and superficial (*KRT78*^{hi}*SPRR3*^{hi}*IVL*^{hi}) normal squamous epithelial clusters were clearly identifiable, but suprabasal and stratified cells were classified as "differentiated", as they could not be robustly separated in this analysis due to the smaller number of cells (Figure 5.16A). Nevertheless, a transition between keratins and associated differentiation markers could still be seen in basal through superficial which showed strong agreement with the classifications assigned in the HCA dataset (Figure 5.16B).

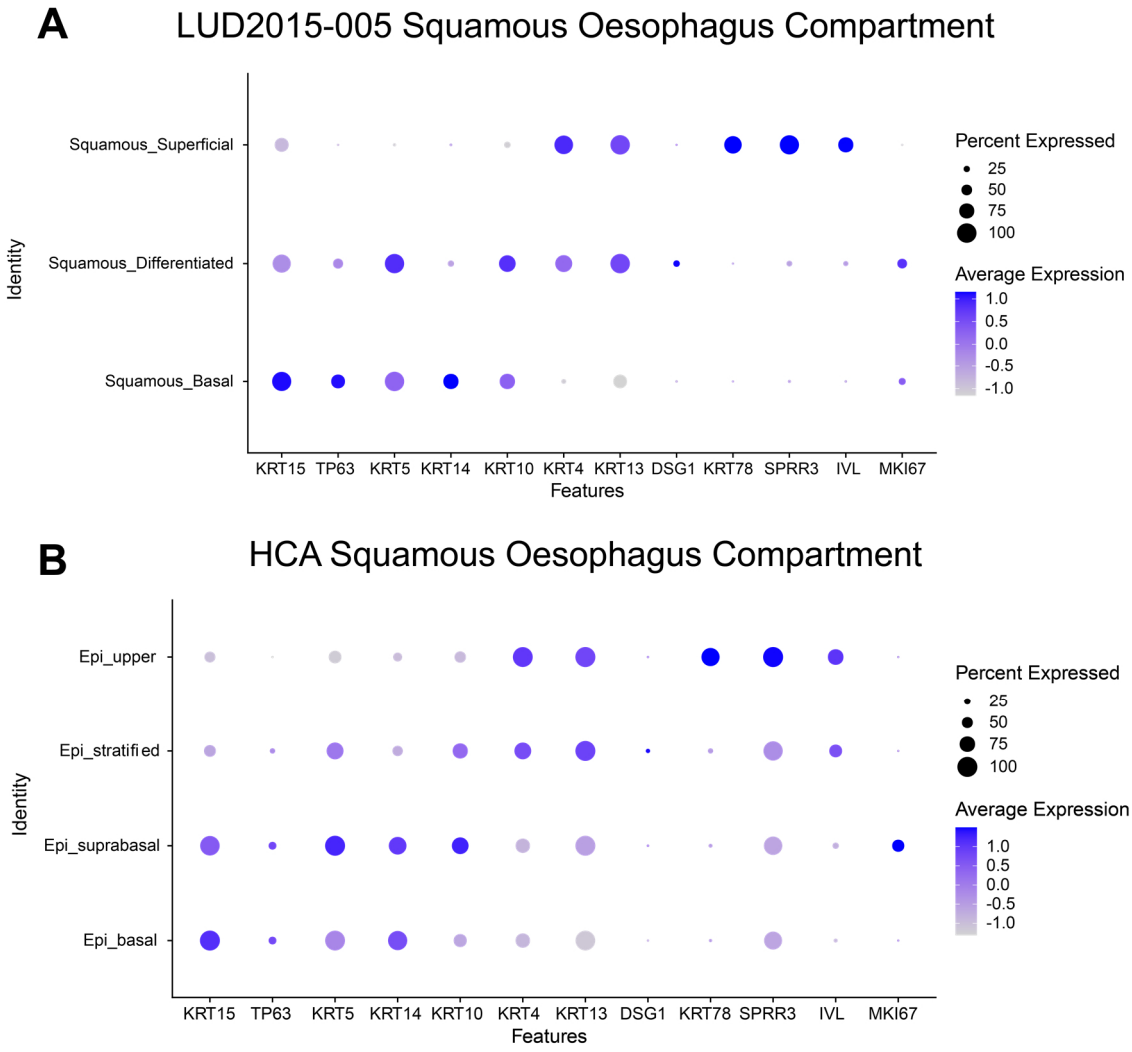


Figure 5.16: Comparison of squamous differentiation markers distribution in HCA and LUD2015-005 datasets. Expression of genes marking different stages of squamous differentiation are shown for the normal squamous sub-compartments of the A) LUD2015-005 and B) HCA dataset. For B, cell type labels and normalized expression values were utilized as published (July 2020 annotation update). Rows are ordered manually to show the transition from basal (bottom) to terminal differentiation (top).

Compared to normal squamous epithelium, OSCC-specific clusters showed higher levels of columnar markers in some patients (*EPCAM*, *KRT8/KRT18*), and expression of several CT antigens (MAGE-A family and *PRAME*) which were not detected in normal squamous clusters (Figure 5.17). Due to the lack of these tumour markers, cells from OAC and OSCC that clustered with normal squamous clusters were deemed more likely to represent contaminating normal squamous epithelium than subtypes of malignant epithelium.



Figure 5.17: Selected markers of OSCC in the LUD2015-005 atlas. Manually selected DEGs associated with OSCC in comparison with normal squamous oesophagus, as well as patient-specific markers of OSCC, are shown alongside a condensed version of squamous differentiation markers from Figure 5.16. Each column is a cell, grouped within their annotated cell type. Cells are coloured by expression (scaled log-normalized counts).

The HPV-positive OSCC (OSCC-JHIE) was included in this scRNA-seq dataset. To confirm whether this was a true HPV-associated cancer, the raw data was queried to assess whether HPV-16 expression could be detected in their patient-specific OSCC cluster. Indeed, UMIs uniquely aligned to HPV-16 were found at high levels in the malignant epithelium for this patient, but not for any other patient (Figure 5.18A). For OSCC-JHIE, HPV-16 was strongly expressed just within the malignant epithelium cluster, and found only at background levels in other sub-compartments, consistent with an HPV-associated malignancy (Figure 5.18B). Therefore, it appears this patient's OSCC was properly associated with HPV-16, and the finding was not caused by contamination of ambient RNA or some other technical artefact.

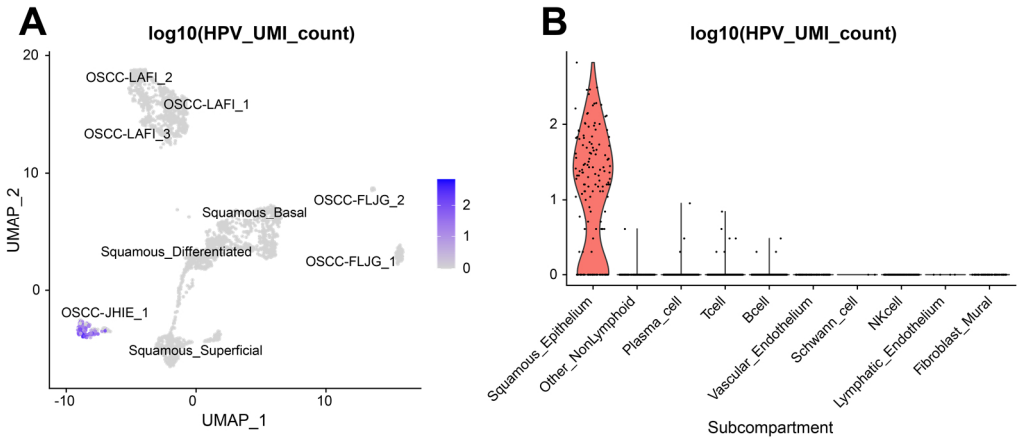


Figure 5.18: OSCC-specific HPV-16 expression in OSCC-JHIE. A) Squamous sub-compartment UMAP (Figure 5.15) coloured by the number of UMIs uniquely aligning to the HPV-16 genome per cell. Values are on a log10 scale (2 = 100 UMIs aligned to HPV-16). HPV-16 expression is only detected in the malignant cluster for OSCC-JHIE. B) Violin plot of UMI count from A, now shown across all sub-compartments. Only the PreTx biopsy from OSCC-JHIE is displayed for this plot. Sub-compartments are arranged by decreasing mean value.

In summary, an overview of the epithelial compartment identified many of the expected cell types present in normal oesophagus and gastric tissue, and the diversity of BO, OAC, and OSCC clusters was assessed, showing greater patient-specific heterogeneity in the epithelium from these oesophageal pathologies. This compartment therefore represents a valuable resource for further understanding of both normal epithelial differentiation and metaplastic and malignant conditions.

5.3.3.2 Stromal compartment

Next, stromal cell types were identified. The stromal compartment broadly broke down into *LYVE1*$^+$ lymphatic endothelium, CD31/*PECAM1*$^+$ vascular endothelium, and a diverse fibroblast/mesenchymal mural cell sub-compartment (*MMP2*$^+$ or *ACTA2*$^+$) (Figure 5.19). A small discrete cluster of *CDH19*$^+$*S100B*$^+$ cells was also identified, with a few cells contributed by multiple patients. These are likely a rare peripheral nerve sheath cell type, presumably of the Schwann cell lineage[198], which is likely associated with the enteric nervous system[199]. Lymphatic endothelium was fairly homogenous, and was contributed mainly by PostTx OAC and oesophageal biopsies from the three patients with operable OAC.

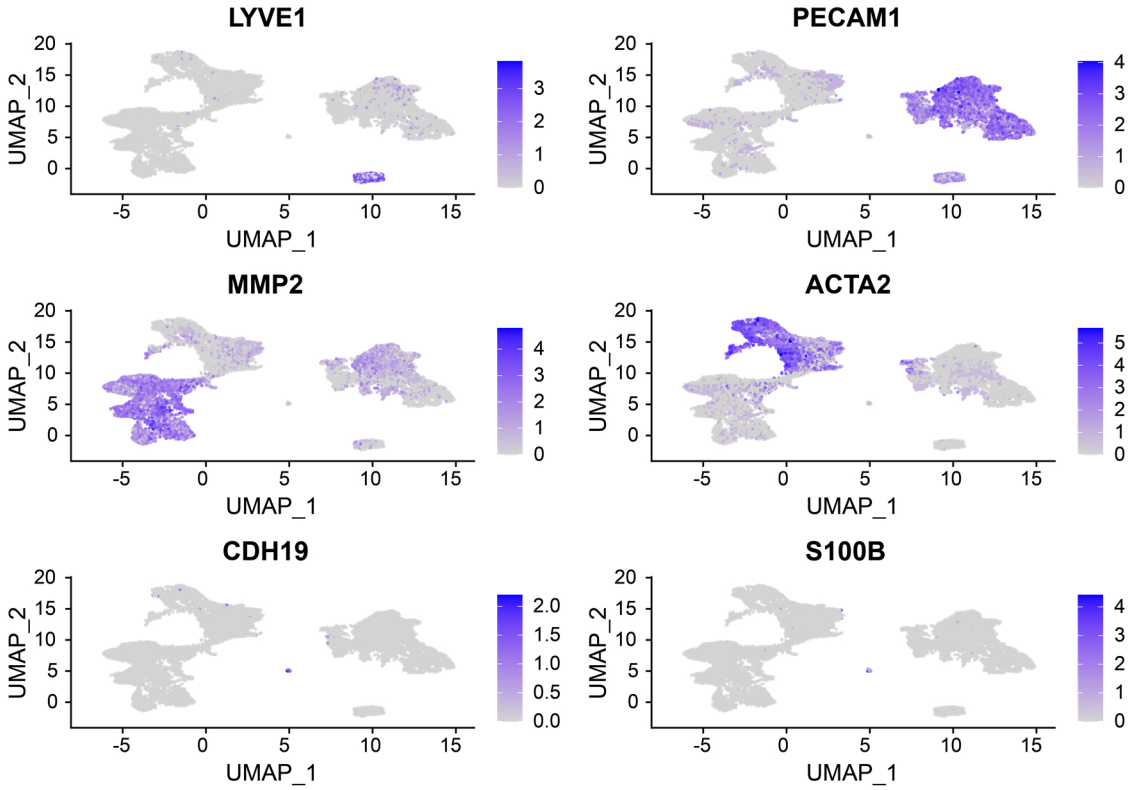


Figure 5.19: Identification of stromal sub-compartments. UMAP dimensionality reduction plot of the stromal compartment. Subplots are coloured by log-normalized expression of the indicated gene, showing lymphatic endothelium (*LYVE1*), vascular endothelium (*PECAM1*/CD31), fibroblasts (*MMP2*), myofibroblasts and mesenchymal mural cells (*ACTA2*), and Schwann cells (*CDH19*, *S100B*).

Vascular endothelium and the fibroblast/mural cell sub-compartments appeared heterogeneous and were broken down into further cell types. Among vascular endothelium, strong differences were seen between tumour-associated endothelial cells (TAEC) and endothelium from normal tissue (Figure 5.20), which appeared particularly driven by expression of basal lamina-related collagens (*COL4A1*/*COL4A2*) and previously reported markers of tumour-specific neovasculature (i.e. Endoglin/*ENG*, *MCAM*) in TAECs[200,201]. Normal vasculature was split into arterial cells (*GJA4*$^{+}$*GJA5*$^{+}$*EFNB2*$^{+}$), capillaries (*CD36*$^{+}$ or an *EDNRB*^{hi} subtype, which shares some markers with specialized aerocyte capillaries in the lung[202]), and post-capillary venules (PCV; *ACKR1*$^{+}$). While most PCV cells showed some markers of activation (MHC class II, P-selectin), post-treatment PCV cells from oesophagus and operable OACs expressed even more inflammatory markers playing a role in immune cell extravasation into tissue (E-selectin, *ICAM1*, *CSF3*; Figure 5.20E). TAEC clusters also showed a number of subtypes (Figure 5.20A,E), including a *ANGPT2*$^{+}$*PGF*$^{+}$*CXCR4*$^{+}$ tip-cell phenotype[203] and a particularly

proliferative (*MKI67*$^{+}$) subcluster, suggestive of actively developing neovasculature in oesophageal tumours.

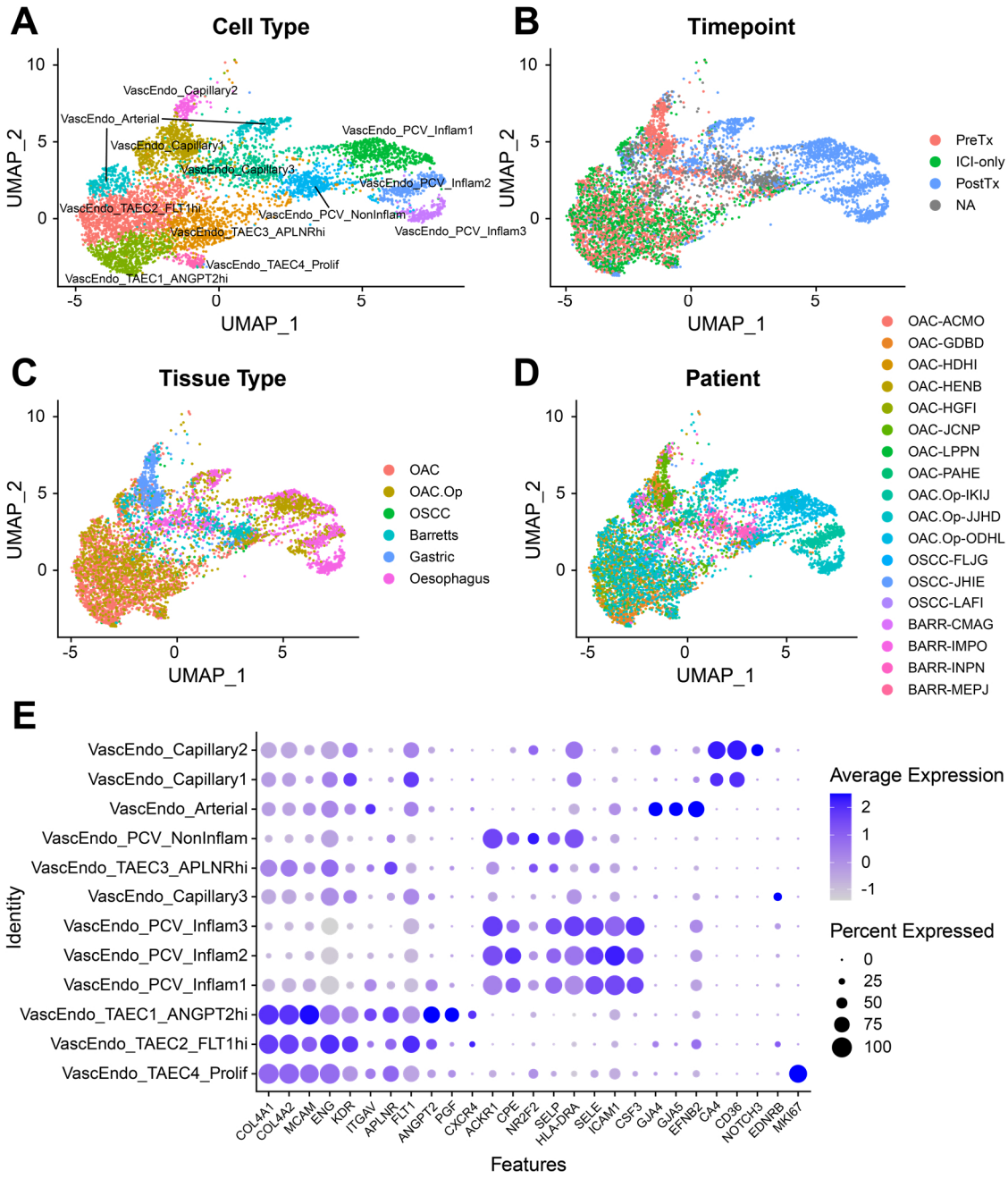


Figure 5.20: Cell map of the vascular endothelium sub-compartment. Top and middle: UMAP visualizations coloured by the identified cell type (A), LUD2015-005 treatment timepoint (B), tissue type (C), and patient of origin (D). For B, dots without annotated treatment (NA) are from Barrett's patients outside of the LUD2015-005 trial. Bottom (E): Dot plot of selected markers that differentiate identified cell types. The size of the dot represents the percent of cells within each column expressing the gene, while the colour is the average strength of expression. Rows are ordered by hierarchical clustering.

Among non-endothelial cells, *MMP2*$^{+}$ fibroblasts and a variety of *ACTA2*$^{+}$ cells were noted. While *ACTA2* is a commonly used marker of myofibroblasts, it is also expressed in mural cells of mesenchymal origin, including pericytes and smooth muscle cells[204]. Among *MMP2*$^{+}$ fibroblasts, a large population specific to tumour biopsies were seen, which were annotated as tumour-associated fibroblasts (TAFs; Figure 5.21). CHI3L1+ and CST1+ TAF subsets were identified, which have been linked with immunosuppressive and senescent fibroblast phenotypes, respectively[205,206]. The major fibroblast subtype in BO and gastric biopsies belonged to a separate POSTN^{hi} grouping.

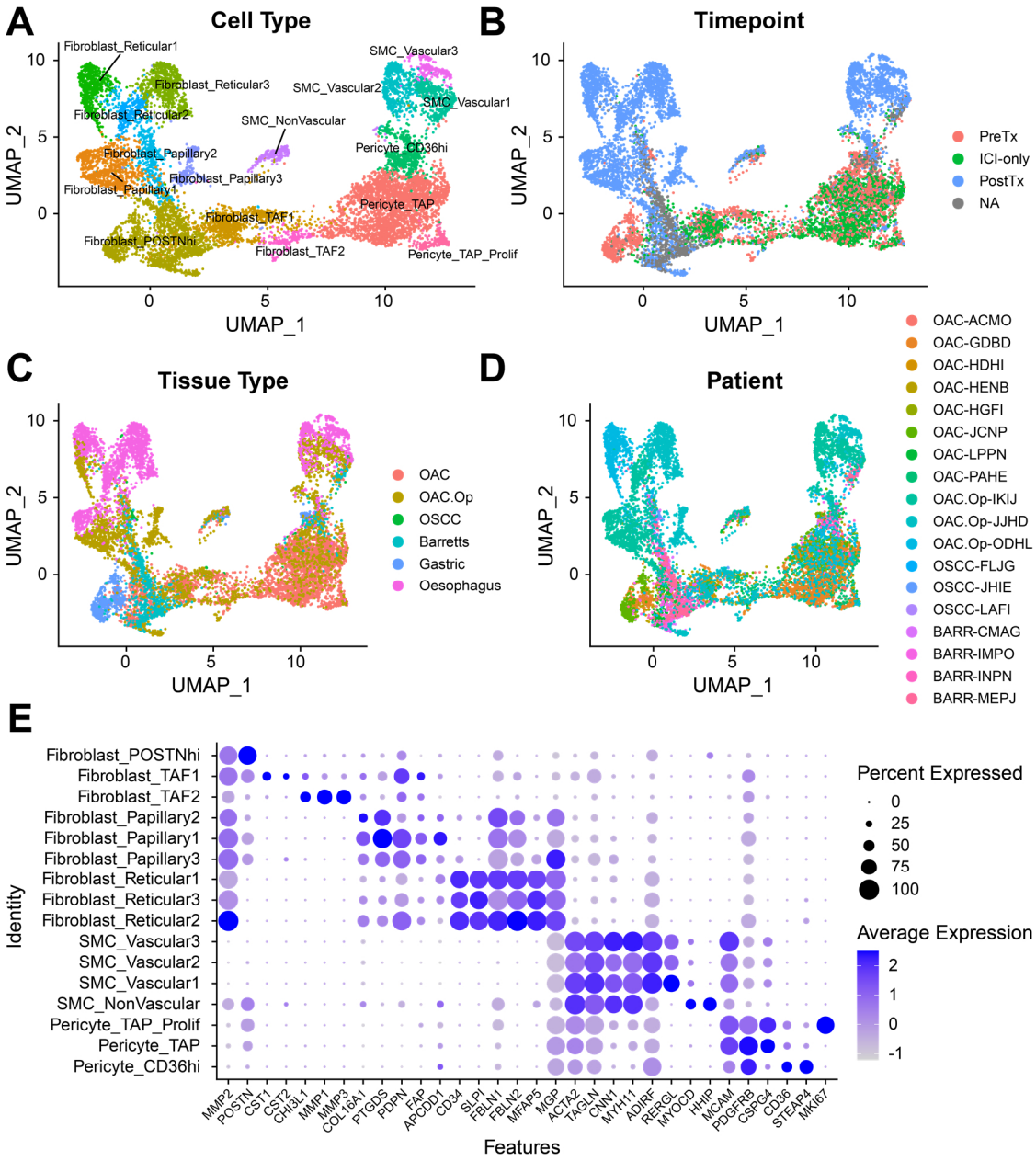


Figure 5.21: Cell map of the fibroblast and mural cell sub-compartment. Top and middle: UMAP visualizations coloured by the identified cell type (A), LUD2015-005 treatment timepoint (B), tissue type (C), and patient of origin (D). For B, dots without annotated treatment (NA) are from Barrett's patients outside of the LUD2015-005 trial. Bottom (E): Dot plot of selected markers that differentiate identified cell types. The size of the dot represents the percent of cells within each column expressing the gene, while the colour is the average strength of expression. Rows are ordered by hierarchical clustering.

Finally, two fibroblast subtypes were detected that resembled previously described human dermal fibroblast populations, with *CD34*$^{+}$*SLPI*$^{+}$*MFAP5*$^{+}$ resembling fibroblasts from the reticular layer of the dermis, and *COL16A1*$^{+}$*FAP*$^{+}$*PTGDS*$^{+}$ clusters resembling fibroblasts from the more superficial papillary layer[207–209]. These may therefore represent normal oesophageal resident populations with resemblance to populations found in the skin. While the structure of

the skin differs in some ways from that of the oesophagus, a similar *CD34*$^{+}$ fibroblast population has been demonstrated in reticular structures in both the dermal layer of the skin and the submucosa of all human GI tissues[210]. These reticular fibroblasts were mainly contributed by core biopsies from surgically resected material (PostTx operable OAC patients), supporting the notion that this cluster likely represents the previously characterized submucosal reticular population, which may be easier to sample with deeper core biopsies. For the remaining *COL16A1*$^{+}$ fibroblasts, while conventional terminology does not explicitly define a papillary layer of the oesophagus as in the skin, the lamina propria forms papillary protrusions into oesophageal epithelium reminiscent of the papillary dermis[211], and this cluster may therefore represent fibroblasts from this papillae-forming layer of the oesophageal mucosa. Consistent with this hypothesis, fibroblasts with this papillary fibroblast-like phenotype were contributed by superficial endoscopic biopsies, especially in Barrett's oesophagus (Figure 5.21C).

Among MMP2^{-} cells, a large tumour-specific population expressing *ACTA2* was present (Figure 5.21E). These could be annotated as myofibroblasts, but they also expressed high level of pericyte markers (*PDGFRB*, *MCAM*, NG2/*CSPG4*). There is significant overlap between markers used to call pericytes and myofibroblasts in the literature, which is further confounded by the fact that vessel-associated pericytes can detach and differentiate into ACTA2-expressing myofibroblasts in response to PDGF-BB, especially in the context of tumours[212]. This process is usually accompanied by decrease in pericyte markers, which were preserved in the major MMP2^{-} TA populations, supporting the classification of these cells as pericytes. However, a minor cluster of *CD36*$^{+}$*STEAP4*$^{+}$ cells found in multiple tissue types was identified, where pericyte markers such as *MCAM* and NG2 were decreased despite preserved *PDGFRB* expression, suggestive of cells that might be transitioning towards myofibroblasts. Indeed, Shami et al. showed that human pericytes preserving *MCAM* expression ("muscular pericytes") were restricted to the perivascular space, while *MCAM*$^{-}$CD36^{+} "fibroblastic pericytes" could be found detached from vessels in the interstitial space[213]. Therefore, these *PDGFRB* clusters were annotated as tumour-associated pericytes (TAP) and *CD36*^{hi} pericytes, although the *CD36*^{hi} population could potentially represent pericytes in the process of differentiating towards a

myofibroblast-like cell state. Other fibroblast cells, particularly a subset of CST1^{hi} TAFs, also expressed *ACTA2* and could be considered to have a myofibroblast-like phenotype (Figure 5.22).

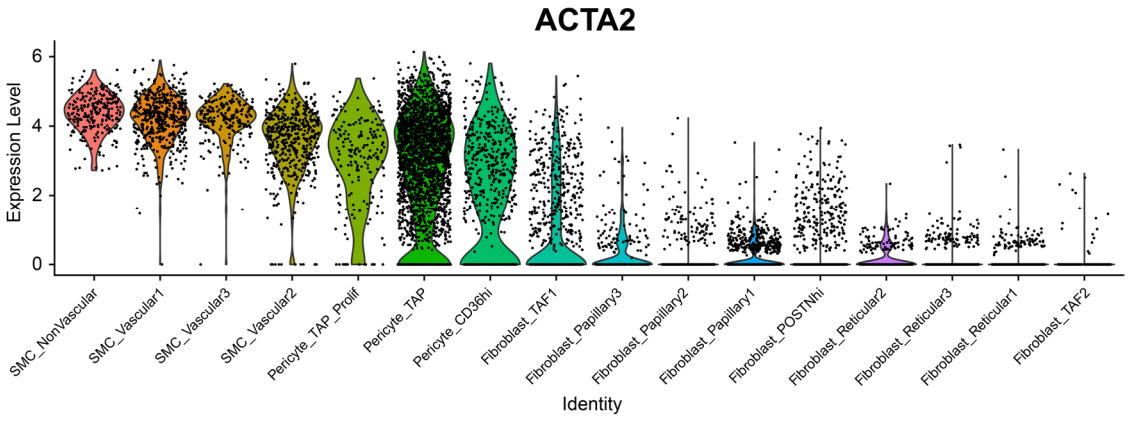


Figure 5.22: Distribution of *ACTA2* across fibroblast and mural cell sub-compartment. Log-normalized expression of *ACTA2* is shown by cell subtype, which each point representing a single cell. Cell subtypes are arranged by decreasing mean expression.

Cells in NG2^{lo}ACTA2^{hi} clusters expressed classical smooth muscle cell (SMC) markers, including *CNN1*, *TAGLN*, and *MYH11*[204] (Figure 5.21E), but were absent for skeletal muscle markers (*MYOD1*, *MYOG*; data not shown), consistent with SMC identity. These SMCs separated into two major clusters. Based on previously reported literature[214,215], these clusters were identified as vascular (*ADIRF*^{hi}) and non-vascular (*HHIP*^{hi}) SMC subsets, with the vascular subset clustering closer to pericytes. Similar to reticular fibroblasts, vascular SMCs were found mainly in PostTx core biopsies of resected material, suggesting increased prevalence in deeper layers of the tissue. However, one subtype (SMC_Vascular1) was also found in endoscopic biopsies from multiple tissue types and treatment timepoints, suggesting this cell type can also be found in more superficial biopsies (Figure 5.21A-B). It is possible these vascular SMC subtypes represent different populations from the lamina propria and deeper submucosal connective tissue. Non-vascular tissue SMCs were contributed from multiple timepoints, primarily from superficial biopsies of OAC and gastric tissue. As these SMCs were found in superficial biopsies, they likely represent the smooth muscle of the muscularis mucosa. In conclusion, fibroblast and mural cells show high heterogeneity in the LUD2015-005 dataset, with both groups having distinct TA subsets, as was shown for vascular endothelium.

5.3.3.3 Lymphocyte immune cell compartment

Lymphocytes were assessed next, which were classified into sub-compartments for plasma cells (*CD37*$^{-}$*CD38*$^{+}$)[216], non-plasma B cells (*CD37*$^{+}$*CD38*$^{-}$), T cells, and natural killer (NK) cells, which clustered closely with T cells. Within the plasma cell sub-compartment, cells clustered mainly on the basis of their antibody heavy chain isotype, with IgA, IgG, and IgM-secreting subgroups identified (Figure 5.23). IgE and IgD expression was rare in this dataset. Additional cluster differences were driven by the light chain subtype (kappa or lambda). No major differences were seen according to treatment timepoint; however, the antibody isotype was highly dependent on the tissue type (Figure 5.23; Table 5.2). As suggested in Figure 4.7A, normal oesophagus had little infiltration of IgA+ or IgG+ plasma cells, while gastric tissue had higher infiltration particularly of IgA+ plasma cells. Interestingly, this effect was even more extreme for BO. While the magnitude is likely an artefact of the comparatively smaller number of cells profiled, 28.4% of all cells surviving QC from BO were IgA+ plasma cells. In OAC, there was a more even balance of IgA and IgG plasma cells, which were both increased over normal oesophagus, while the 3 OSCC patients profiled in this dataset were dominated by IgG infiltration.

Isotype	OAC (inoperable)	OAC (operable)	OSCC	Barrett's	Gastric	Oesophagus
IgA	3.71%	3.17%	0.95%	28.40%	6.55%	0.40%
IgG	3.60%	1.92%	8.81%	1.15%	0.37%	0.11%
IgM	0.76%	0.14%	0.74%	0.25%	0.25%	0%

Table 5.2: Prevalence of plasma cell antibody isotypes by tissue type. Percentages shown are the percentage of all total cells that fall into IgA^{+}, IgG^{+} or IgM^{+} plasma cell clusters (as in Figure 5.17A), including all samples of the given tissue type. Proliferative plasma cells were excluded from this analysis.

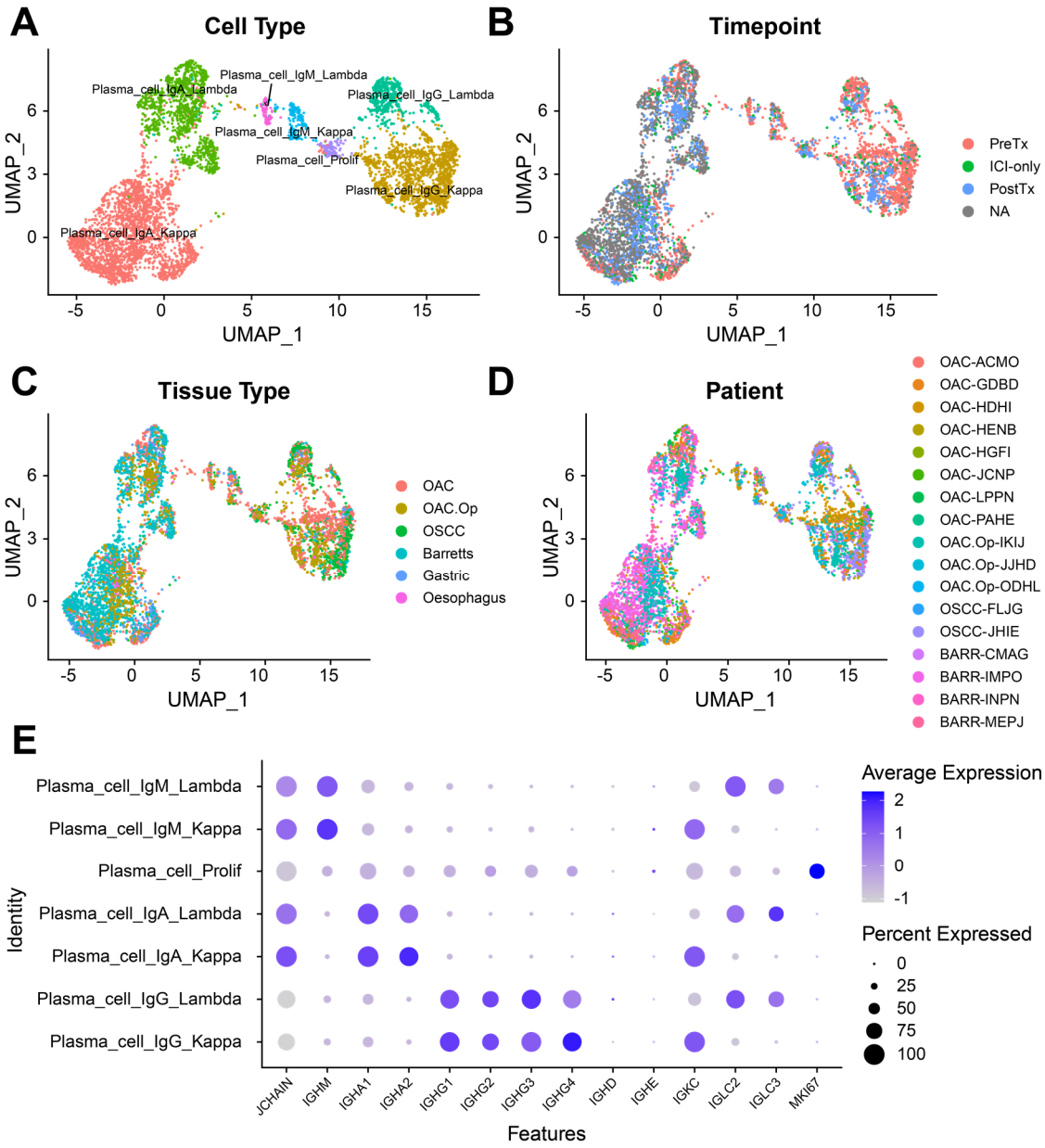


Figure 5.23: Cell map of the plasma cell sub-compartment. Top and middle: UMAP visualizations coloured by the identified cell type (A), LUD2015-005 treatment timepoint (B), tissue type (C), and patient of origin (D). For B, dots without annotated treatment (NA) are from BO patients outside of the LUD2015-005 trial. Bottom (E): Dot plot of selected markers that differentiate identified cell types. The size of the dot represents the percent of cells within each column expressing the gene, while the colour is the average strength of expression. Rows are ordered by hierarchical clustering.

Clustering of non-plasma B cells was less driven by antibody isotype, presumably due to their lower level of immunoglobulin expression compared with plasma cells. Three major non-plasma B cell subtypes were identified: Naïve B cells (CD27-), mature B cells (CD27+), and germinal center (GC) B cells (Figure 5.24). Two major subtypes of GC B cells were identified: *CD83/CXCR4*$^{+}$ GC B cells, and *AICDA*^{hi} GC cells, which also express *RGS13* and were highly

proliferative. All of these molecules have been reported in B cells as key mediators of GC formation and function[217–220]. Naïve B cells expressed IgD and IgM, representing an antigen-naïve phenotype, while more mature B cells appeared to have isotype class-switching characteristic of previous B cell receptor activation. GC-like B cells had reduced immunoglobulin expression across isotypes, which is a known feature of GC B cells[221]. Within GC cells, as *AICDA* expression and proliferation are key mediators of class switch recombination in B cells[220,222], the *AICDA*^{hi} subtype likely represents GC B cells in the process of class switching. Finally, mature and CD83/CXCR4^{+} GC B cells were further subdivided into TNF^{hi} and TNF^{lo} subtypes.

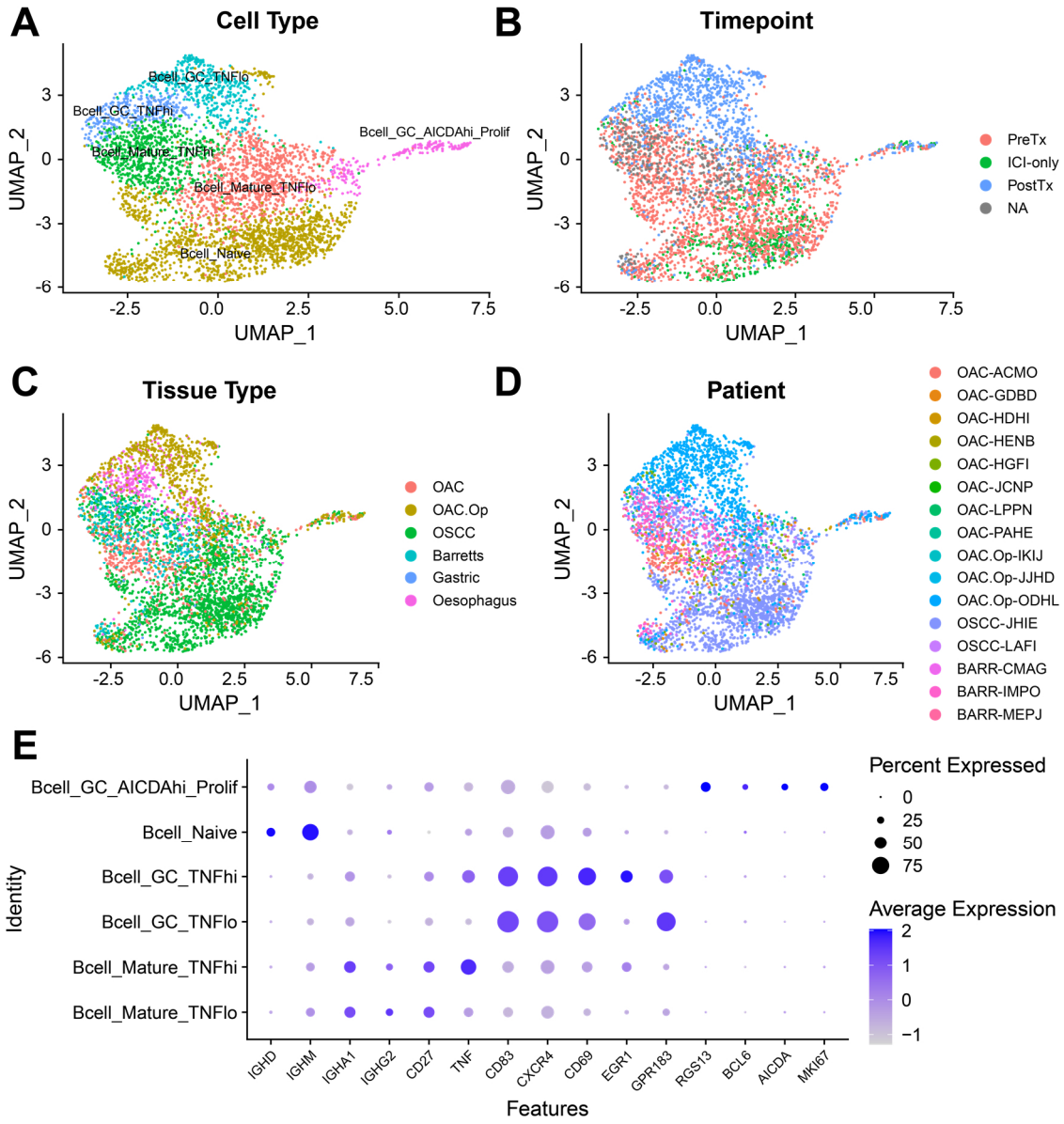


Figure 5.24: Cell map of the B cell sub-compartment. Top and middle: UMAP visualizations coloured by the identified cell type (A), LUD2015-005 treatment timepoint (B), tissue type (C), and patient of origin (D). For B, dots without annotated treatment (NA) are from Barrett's patients outside of the LUD2015-005 trial. Bottom (E): Dot plot of selected markers that differentiate identified cell types. The size of the dot represents the percent of cells within each column expressing the gene, while the colour is the average strength of expression. Rows are ordered by hierarchical clustering.

Interestingly, most GC B cells were found at the PostTx timepoint, potentially suggestive of the formation of germinal centers and associated complex lymphoid structures in response to immunochemotherapy. However, the vast majority of the *CD83/CXCR4*$^{+}$ GC subtype (674) were contributed by PostTx core biopsies from a single patient with operable OAC (OAC.Op-ODHL, the lone patient receiving FLOT), both in the tumour and an adjacent oesophageal biopsy. Therefore, this could also just be a patient-specific finding, or could just represent an increased

propensity to sample these GC structures in deeper core biopsies. However, a smaller number of cells in this GC subtype from were also contributed by treatment-naïve superficial biopsies from the HPV-associated OSCC patient and a Barrett's oesophagus patient. The *AICDA*^{hi} GC subtype had more balanced contributions from multiple patients which did not show a clear timepoint-specific pattern (Figure 5.24A-B). In summary, among oesophageal pathologies, GC-like B cells can be found in a variety of disease states and treatment timepoints.

NK cells were relatively few in number among lymphoid cells, and therefore more difficult to break into discrete subclusters. However, CD16^{hi}, *GZMK*^{hi}, L-Selectin^{hi}, and *KRT86*^{hi} (which may represent an exhausted NK cell phenotype[223]) subsets could still be identified (Figure 5.25). Additionally, a discrete cluster of cells in this sub-compartment were noted which expressed innate lymphoid cell (ILC) markers, particularly resembling a group 3 "lymphoid tissue inducer" ILC (*IL7R*, *KIT*, *RORC*, *CCR6*). These lymphoid tissue inducers play a role in the development of Peyer's patches and other lymphoid structures within the intestines[224,225], and are therefore a potentially important contributor to the local immune system despite their relative rarity. Assessing distribution of the NK cell subtypes, CD16^{hi} and *KRT86*^{hi} subsets appeared to be particularly common at the ICI-only timepoint (Figure 5.25B).

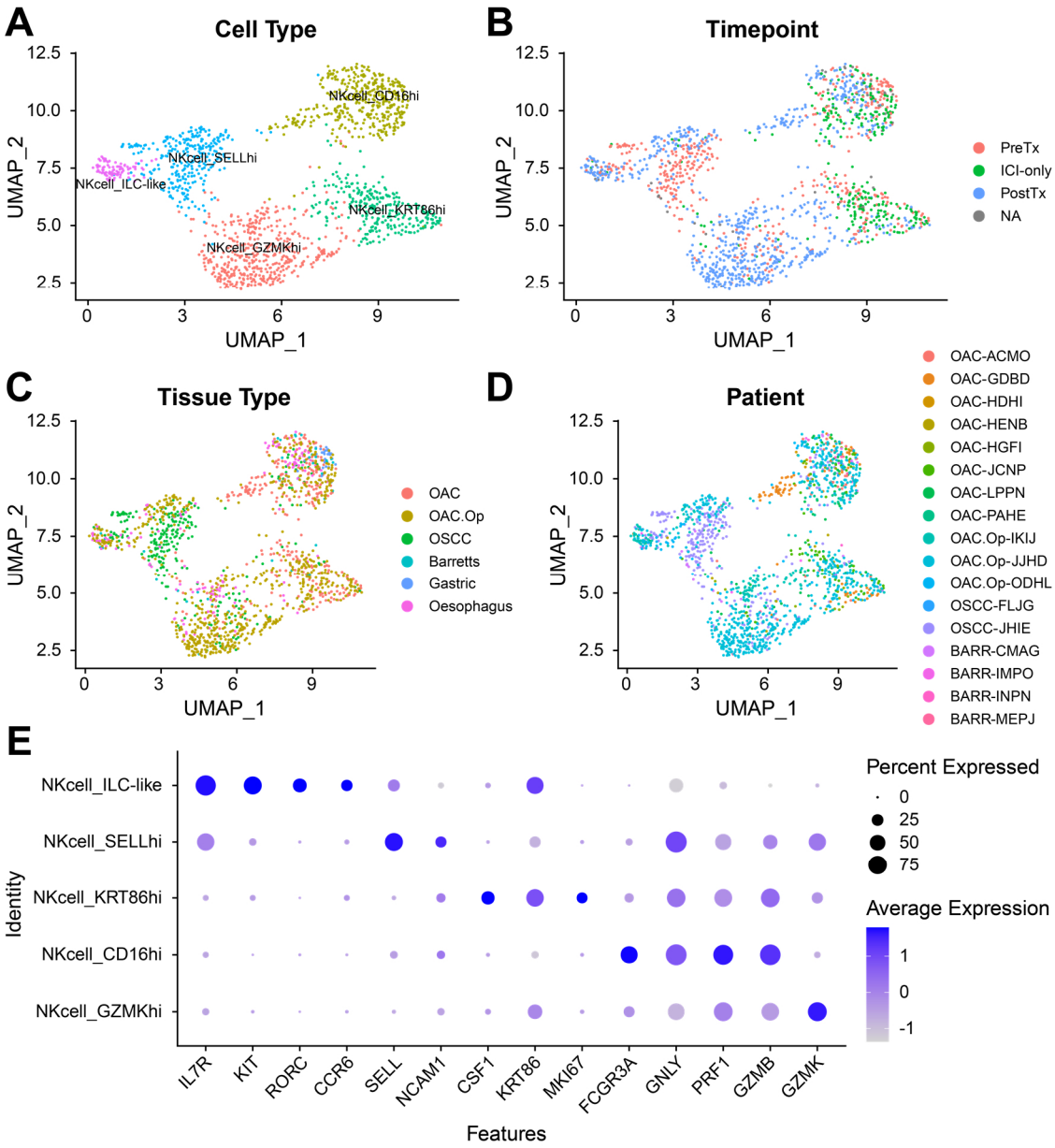


Figure 5.25: Cell map of the natural killer cell sub-compartment. Top and middle: UMAP visualizations coloured by the identified cell type (A), LUD2015-005 treatment timepoint (B), tissue type (C), and patient of origin (D). For B, dots without annotated treatment (NA) are from Barrett's patients outside of the LUD2015-005 trial. Bottom (E): Dot plot of selected markers that differentiate identified cell types. The size of the dot represents the percent of cells within each column expressing the gene, while the colour is the average strength of expression. Rows are ordered by hierarchical clustering.

The T cell sub-compartment was classified into a number of CD4 and CD8 subtypes, as well as a single proliferative (*MKI67*^{hi}) cluster that contained both CD4 and CD8 T cells (Figure 5.26). CD8^{+} T cells were broadly split into subgroups based on whether or not they expressed tissue resident memory (T_{RM}) markers. Among T_{RM} CD8^{+} cells, one cluster ($T_{RM}1$) appeared to be particularly specific to OAC biopsies. This cluster was best defined by expression of *KRT86*, but also expressed T_{RM} markers *ITGAE* (CD103) and *ENTPD1* (CD39), had high markers of cytolytic

activity (*GZMB, PRF1*) without elevated *IFNG* and *TNF*, and expressed other T cell exhaustion markers (TIM-3/*HAVCR2*, PD-1/*PDCD1*). This combination of markers has been reported to identify a tumour-reactive T_{RM} subtype in lung cancer and other solid tumours, which maintains proliferative and cytotoxic capacity despite elevated expression of exhaustion markers[226,227]. While full TCR clonotype analysis is beyond the scope of this thesis, some large T cell clones expanded by immunotherapy were detected within this cluster (Figure 5.27), providing evidence that this T_{RM} subtype is a likely effector of ICI-induced anti-tumoural T cell responses in at least some OAC patients. Other T_{RM} subsets included a HOBIT/*ZNF683*^{hi} subset ($T_{RM}2$), which was the primary CD8 T cell found in gastric biopsies and had little cytolytic or exhaustive gene expression, and a *CD69*^{hi} subset ($T_{RM}3$), which expressed high levels of *TNF* and *IFNG* and could be found in multiple tissue types.

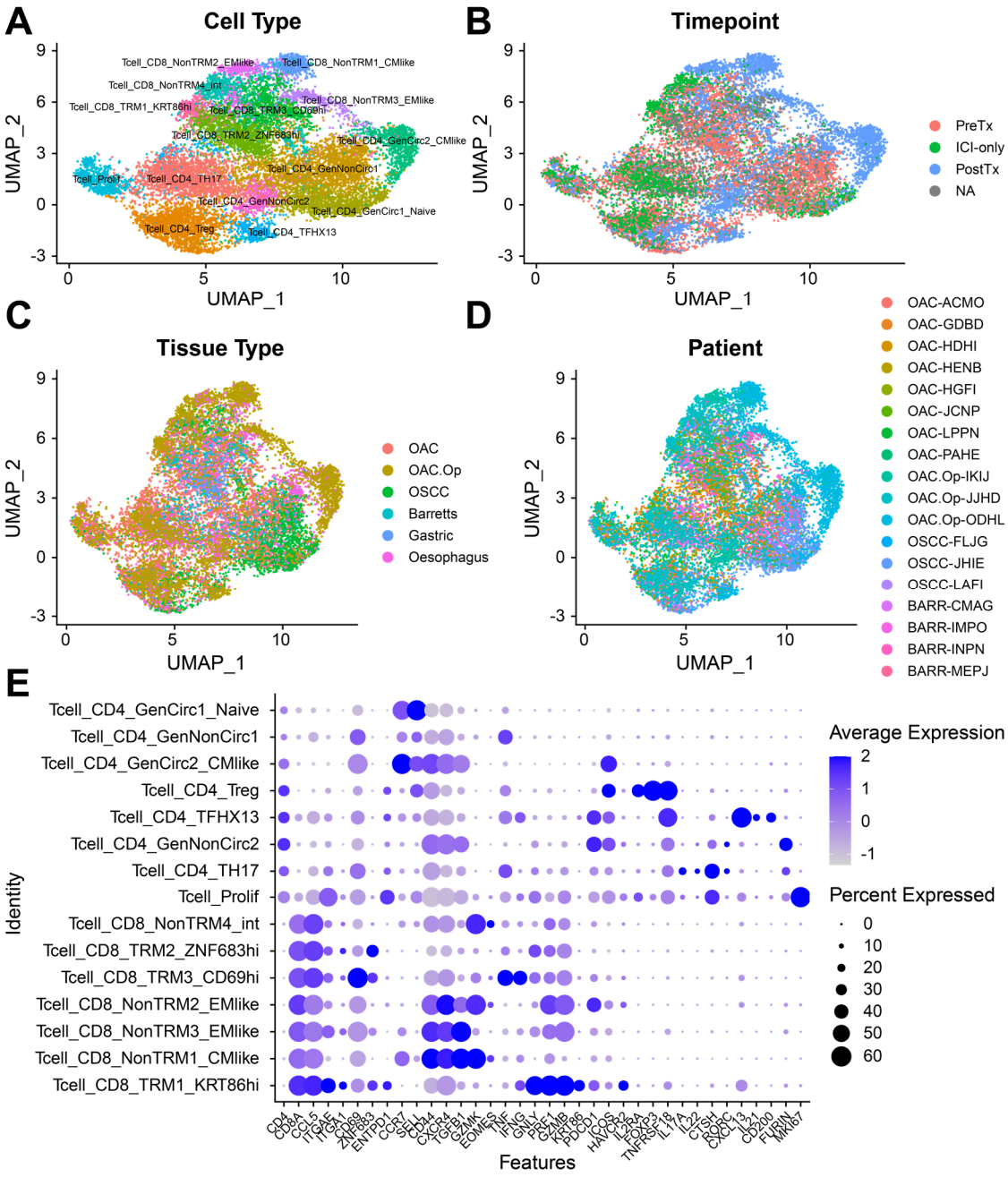


Figure 5.26: Cell map of the T cell sub-compartment. Top and middle: UMAP visualizations coloured by the identified cell type (A), LUD2015-005 treatment timepoint (B), tissue type (C), and patient of origin (D). For B, dots without annotated treatment (NA) are from Barrett's patients outside of the LUD2015-005 trial. Bottom (E): Dot plot of selected markers that differentiate identified cell types. The size of the dot represents the percent of cells within each column expressing the gene, while the colour is the average strength of expression. Rows are ordered by hierarchical clustering.

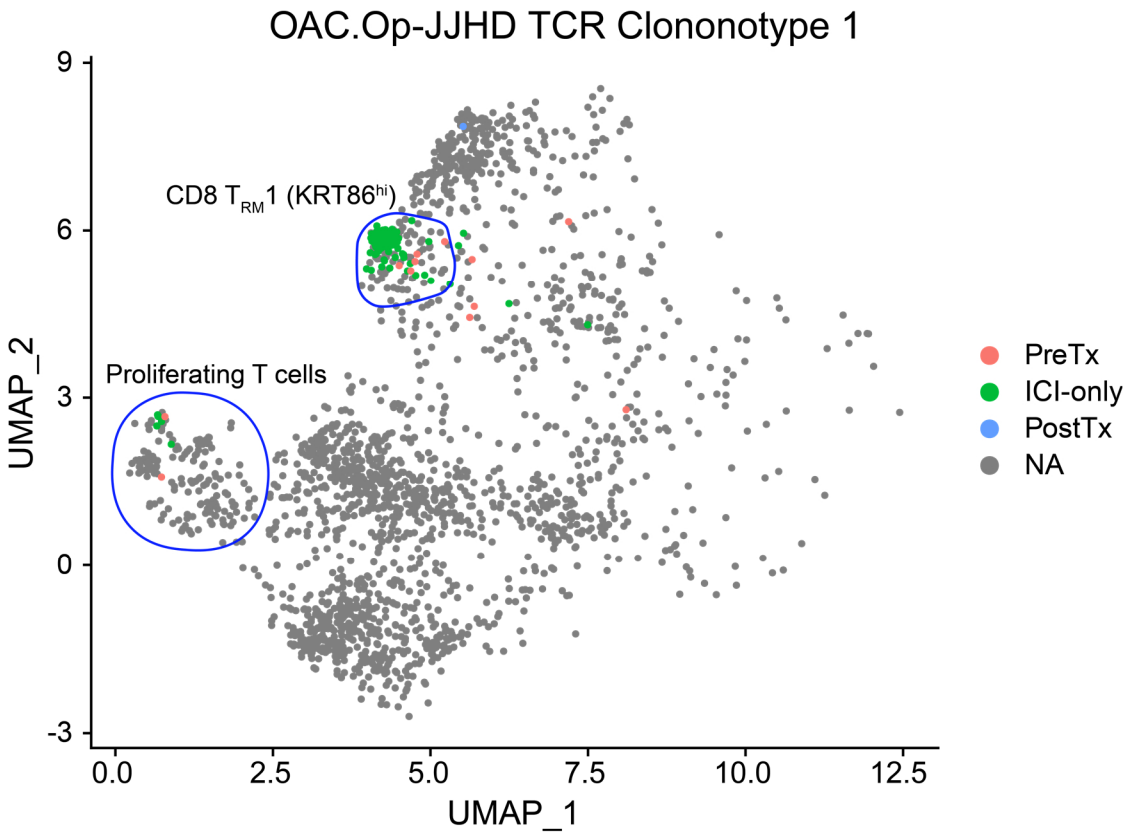


Figure 5.27: Distribution of the largest TCR clonotype from patient OAC.Op-JJHD. Paired TCR-sequencing was conducted using the 10x V(D)J assay, and TCR clonotypes were identified and matched up with gene expression profiles using Cell Ranger. Only T cells for patient OAC.Op-JJHD are shown, using the dimensionality reduction from Figure 5.26. Cells from this largest clonotype are assigned a colour based on whether they were found at PreTx, ICI-only, or PostTx, while all other clonotypes (from all timepoints) are plotted in grey (NA). The approximate location of the $T_{RM}1$ and proliferating T cell cluster are shown.

NonT_{RM} CD8 T cells generally expressed markers of antigen experience and memory formation (*CD44*) and *CXCR4*, an important mediator of central memory homing to bone marrow[228,229], and therefore likely represent a memory phenotype more primed for circulation to other lymphoid structures rather than tissue residence. Within these nonT_{RM} clusters, central memory-like (CM-like; *CD44*^{hi} CD62L/*SELL*^{hi}) and effector memory-like (EM-like; *CD44*^{hi} CD62L/*SELL*^{lo}) groups were identified (Figure 5.26). The majority of the EM-like and CM-like clusters were contributed by PostTx core biopsies of operable OAC, but there were also smaller contributions from endoscopic biopsies of inoperable OAC in each subtype. A final nonT_{RM} cluster with even contributions from multiple OAC and OSCC patients (NonT_{rm}4) did not have increased expression of *CD44*, but did have high expression of *EOMES*, which enables self-renewal and central memory differentiation of CD8$^+$ T cells after antigen experience[230]. This cluster may

therefore represent a group of cells transitioning from tissue residence towards a CM-like phenotype following antigen exposure.

Among $CD4^+$ T cells, specialized Treg (*IL2RA*/CD25, *FOXP3*) and Th17 (*IL17A*, *IL22*, *RORC*) cells formed discrete clusters (Figure 5.26). Another major $CD4^+$ cluster was identified as a specific T-follicular helper cell (T_{FH}) subtype known as T_{FH}X13, based on high marker concordance (*CXCR5*^{lo}CXCL13^{hi}CD200^{hi}PD-1^{hi}ICOS ^{hi}IL21^{hi}) with a previously reported cell type of the same name, which was described to mediate intratumoural B cell infiltration (likely via *CXCL13*) and potentially also tertiary lymphoid structures (TLS) formation in breast cancer[231]. Finally, a number of general $CD4^+$ clusters without identifiable specialization markers were seen. These were classified into circulatory or non-circulatory based on their expression of homing receptors to lymph nodes and other lymphoid structures, such as *CCR7* and *SELL*[232]. Circulatory CD4s were *CCR7*$^+$*SELL*$^+$, and could either be *CD44*^{hi} (CM-like) or *CD44*^{lo} (naïve). Similar to $CD8^+$ T cells, the CM-like phenotype was primarily (but not entirely) contributed by a single PostTx biopsy, while the naïve grouping had contributions from multiple patients. Non-circulatory general CD4 T cells were classified into two subtypes, the more prevalent of which (NonCirc1) was *TNF*^{hi}, with the minor subpopulation (NonCirc2) being *FURIN*^{hi}.

A direct comparison of PreTx proportions of T cell subtypes between tissues showed that gastric and oesophageal T cell infiltration patterns were dominated by *ZNF683*^{hi} T_{RM} and Th17 cells, respectively (Figure 5.28A). While OAC had comparatively less of these cell types, they show increased proportions of the KRT86^{hi} cytolytic-exhausted T_{RM}1 phenotype, T_{FH}X13, and proliferative T cells, suggestive of a multi-faceted ongoing anti-tumoural response even before therapy. However, Tregs were also increased relative to normal controls, indicative of concomitant enrichment in immunosuppression. T cell subtypes of metastatic OAC biopsies were relatively consistent over treatment, but did show some differences in certain subtypes. During the ICI-only window, the frequency of the *CD69*^{hi} T_{RM}3 subtype expressing high levels of *TNF* and *IFNG* increased slightly, but there were no major differences in other cell types, even in proliferative T cells. However, by the end of immunochemotherapy, there were significant differences in T cell proportions: Namely, the proportion of Tregs and T_{FH}X13 decreased, while

$T_{RM}2$, $T_{RM}3$, and the EOMES^{hi} NonTRM4 CD8^{+} subtypes were increased (Figure 5.28B). Therefore, the overall phenotype of the T cell compartment shifted towards CD8 subtypes during immunochemotherapy at the expense of CD4 subtypes, either due to the effect of treatment or as a function of the reduced tumour content at this timepoint.

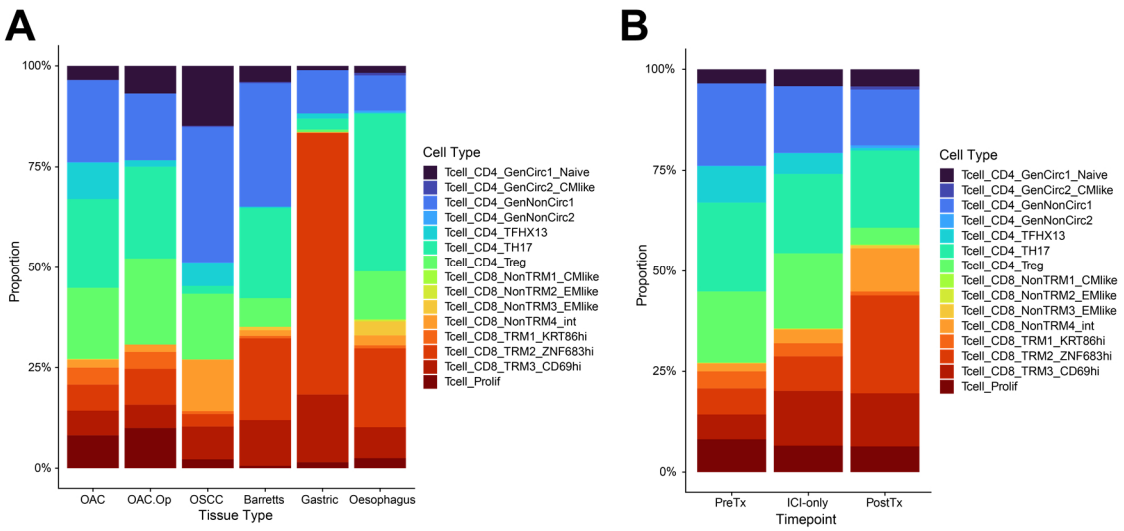


Figure 5.28: Differences in T cell states between tissue types and treatment timepoints. A) Proportion of all pre-treatment T cells that fall into indicated cell states for all tissue types, with all samples from a given tissue grouped as a single bar. B) As in A, except showing T cell states for only inoperable OAC biopsies, and how the proportions change over treatment timepoints. All samples for the given timepoint are grouped as a single bar.

5.3.3.4 Phagocyte immune cell compartment

The phagocyte compartment contained two major two sub-compartments: one of monocytes, macrophages, neutrophils, and dendritic cells, and another of mast cells, which clustered separately. The vast majority of phagocytes expressed markers of the myeloid lineage (CD11b/CD11c, Figure 5.4), which the notable exception of plasmacytoid dendritic cells, which can be derived from lymphoid progenitors[233]. For the mast cell sub-compartment, mast cells were previously characterized in section 5.3.2 as a fairly homogenous cell type with major subsets based on the presence or absence of a specific mast cell tryptase (*TPSD1*). Contrastingly, the other sub-compartment appeared much more heterogeneous. A number of dendritic cell (DC) subtypes were identified, including both major conventional dendritic cell (cDC) subsets (cDC1: *XCR1*^{hi}*CLEC9A*^{hi}; cDC2: *CD1C*^{hi}*CLEC10A*^{hi}) and a *IL3RA*/CD123^{hi} plasmacytoid dendritic cell subset[234]. cDC2s could be split into two subsets: a Langerhans-like (*CD1A*, *CD207*) CD207^{hi}

subset, and a CD207^{lo} subset that shared some features with immunosuppressive macrophages (*CD163*, *IL10*). Normal tissues, particularly the oesophagus, primarily contributed the CD207^{hi} cDC2 subset, while the CD207^{lo} subset were mainly found in PostTx OAC and oesophageal biopsies (Figure 5.29; Figure 5.30). A separate DC cluster expressing DC-LAMP (*LAMP3*), a mature DC subtype previously reported in multiple cancer types[235,236], was also detected. Agreeing with these reports, this mature DC subtype expressed high levels of key molecules involved in T cell co-stimulation (*CD80*, *CD86*), but also molecules that promote immunosuppression and tumour progression (PD-L1/*CD274*, *IDO1*, Fascin/*FSCN1*), suggesting a complex role in modulating anti-tumour responses in this setting. A neutrophil cluster was also identified (CD16b/*FCGR3B*, *PROK2*, *HCAR2*). This cell type showed a particularly low number of unique features (median 270, range 204-600), which is a known feature of granulocyte populations using this scRNA-seq workflow[o].

The remaining cells in this sub-compartment were identified as monocytes and macrophages. While exact discrimination between monocytes and macrophages phenotypes in tissue is complicated, two clusters which appeared to correspond to classical *(CD14*$^{+}$, *FCGR3A*/CD16^{-}, *SPN*/CD43^{-}) and non-classical (*CD14*^{lo}, *FCGR3A*/CD16^{+}, *SPN*/CD43^{+}) monocytes[237] were identifiable (Figure 5.29). Of these, the classical subset was the predominant population, which also displayed expression of previously reported markers of inflammatory monocyte-like cells found in Crohn's disease and ulcerative colitis (*FPR1*, *S100A9*, *S100A12*)[238].

Other clusters in this sub-compartment appeared to express markers of more differentiated macrophage states. Macrophages were traditionally classified using a dichotomy between inflammatory M1 macrophages and anti-inflammatory M2 macrophages, however this system is now thought to be oversimplified, particularly in cancer where tumour-associated macrophages (TAMs) are thought to exist in a complex continuum of states between M1 and M2[239]. For the

[o] Manufacturer communication: "Can I process neutrophils (or other granulocytes) using 10x Single Cell applications?". *https://kb.10xgenomics.com/hc/en-us/articles/360004024032-Can-I-process-neutrophils-or-other-granulocytes-using-10x-Single-Cell-applications-*. Accessed 23 June 2021.

sake of simplicity, M1 and M2 language will still be used to describe phenotypes approaching the M1 and M2 ends of this continuum.

At the M2 end, cells expressing high levels of immunosuppressive macrophage markers (*CD163*, *MRC1*/CD206, *CD209*, *IL10*) were noted. These were mainly found in PostTx core biopsies of tumour and oesophagus from patients with operable OAC (Figure 5.29, Figure 5.30), and either represented a tissue-resident phenotype found in deeper layers of the oesophagus, or perhaps an M2 phenotype enriched by the radiotherapy given to these patients, which has been previously suggested[240]. A separate group of macrophages with more moderate expression of *CD163* and CD206 macrophages were seen, which were derived primarily from endoscopic biopsies of OAC but not normal tissue. Due to their tumour-specificity and phenotype in between M1 and M2-like states, these were deemed to be the TAM population. Three clusters of these TAMs were identified, all of which expressed high levels of the C1Q complex and *APOE*. Finally, on the inflammatory end of the spectrum, more M1-polarized macrophages were identified; these were also fairly tumour-specific, but were termed M1-like on the basis of their inflammatory cytokine profile and high expression of M1-associated markers such as Activin A *(INHBA*)[241], which differentiated them from the intermediate TAM phenotype. Several M1-like macrophage subtypes were identified. The predominant subtypes at the PreTx timepoint in inoperable OAC were *CXCL1*^{hi} (chemoattractant) and *GBP5*^{hi} (IFN-stimulated[242]) phenotypes. Both of these highly expressed markers of M1-like inflammatory cytokine production (*IL1A*, *TNF*, *IL6*) and reactive oxygen species (ROS) generation (*SOD2*). Two additional M1-like INHBA^{hi} subclusters (*VEGFA*^{hi}/angiogenic and *HILPDA*^{hi}/hypoxia-lipid response[243]) that had more moderate inflammatory cytokine expression were mainly contributed by a single patient (OAC-GDBD) at the PostTx timepoint. There was a notable overlap in marker expression between classical monocyte-like cells and the *VEGFA*^{hi} M1-like subtype, and a more limited set of markers (*CD300E*, *TREM1*) were shared between these monocytes and other M1 subtypes, suggesting some level of overlap between monocytic and M1-like macrophage states.

To further assess this overlap, this sub-compartment was assessed for transcriptomic signatures of cell states identified in a recent cross-species scRNA-seq study of infiltrating

myeloid cells in NSCLC, which identified classical monocytes, non-classical monocytes, mature monocyte-derived macrophages, and tissue-resident alveolar macrophages[244]. Classical and non-classical monocyte gene signatures from the NSCLC dataset were specific to the same cell types identified in the present work, providing supporting evidence that these monocyte identifications were correctly made (Figure 5.29A-C). While cells expressing alveolar macrophage signatures were not found in this work (data not shown), likely due to the organ specificity of that macrophage subtype, most M2 and TAM macrophages expressed high levels of mature monocyte-derived macrophage markers from the NSCLC study (Figure 5.29D). However, for M1-like macrophages, one pole expressed mature macrophage markers, while the other pole closest to monocytes appeared to have less markers of maturity, with increased expression of monocyte-like signatures. Therefore, while these cells all exhibit some markers of M1-like inflammatory state that separate them from resting monocyte populations, they may represent different stages of differentiation between M1-polarized monocytes[245] or immature M1-like macrophages and fully differentiated M1-like macrophages. To better represent the likely presence of various states of maturity in the M1-like grouping, these cell types were annotated as differentiating M1-like macrophages ("Macrophage_DiffM1").

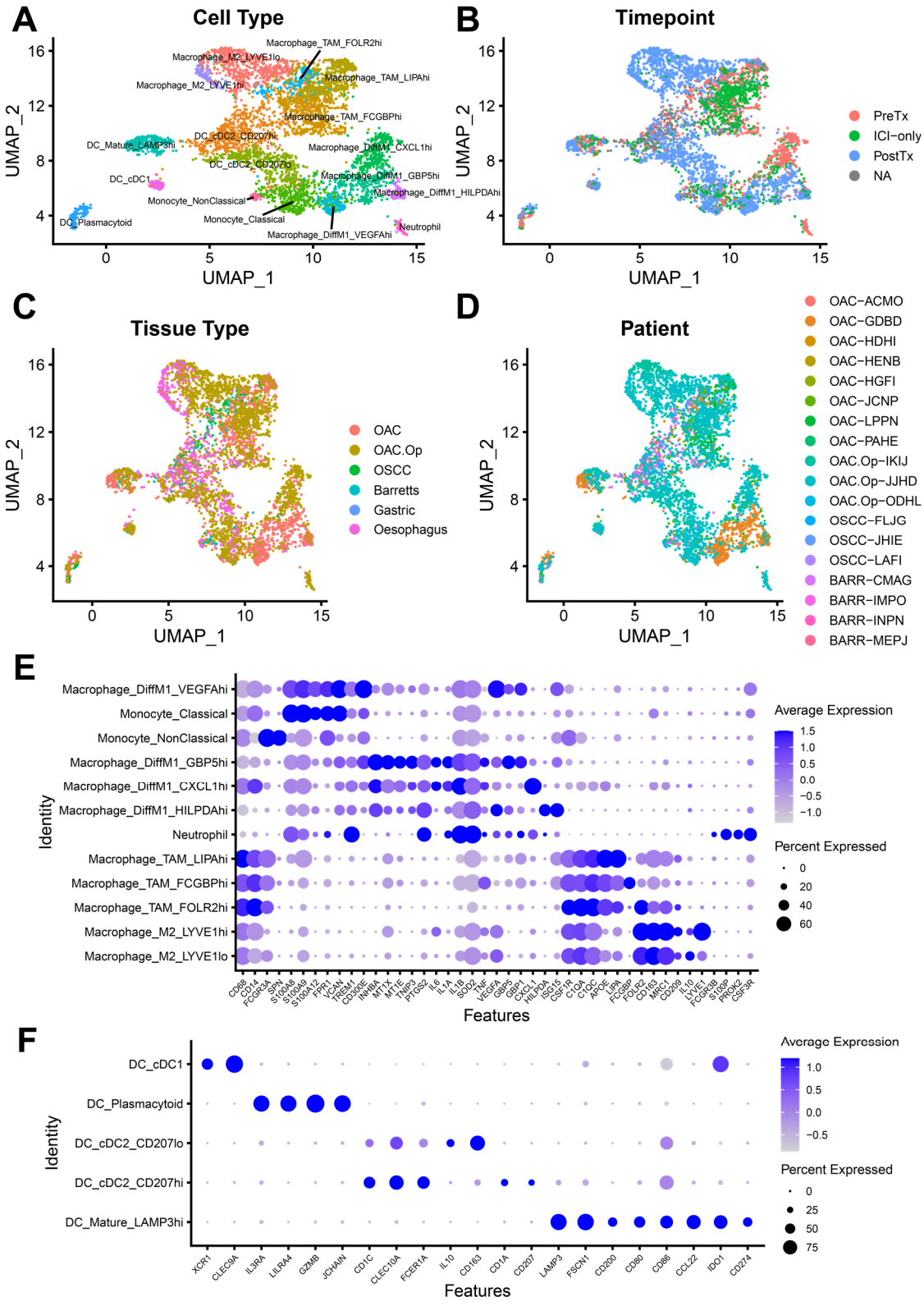


Figure 5.29: Cell map of the phagocytic immune cell sub-compartment. Top and middle: UMAP visualizations of phagocytic immune cells (excluding mast cells) coloured by the identified cell type (A), LUD2015-005 treatment timepoint (B), tissue type (C), and patient (D). For B, dots without colour (NA) do not have a timepoint, as they do not come from the LUD2015-005 trial. Bottom (E): Dot plot of selected markers that differentiate identified cell types. The size of the dot represents the percent of cells within each column expressing the gene, while the colour is the average expression. Rows are ordered by hierarchical clustering.

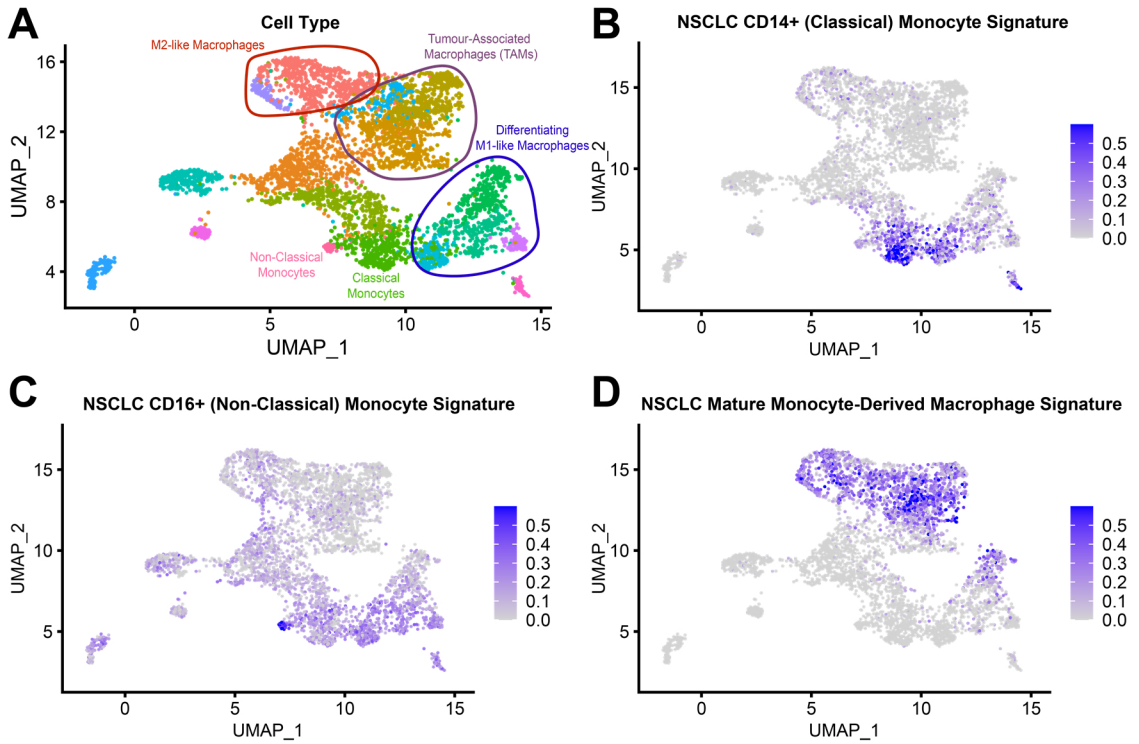


Figure 5.30: Expression of reported monocyte and mature macrophage markers in the LUD2015-005 dataset. A) Dimensionality reduction from 5.29A, grouped by cell type. Monocyte subtypes and broad macrophage states are identified. B-D): Expression of cross-species markers of classical monocytes (group III), non-classical monocytes (group IV), and mature monocyte-derived macrophages (group II) from Figure 1C of Casanova-Acebes, et al[244]. Gene signatures were calculated using AddModuleScore from Seurat, and the range of plotted values were truncated to a range between 0 and 0.6.

Comparing PreTx distribution across tissue types, the predominant cell types within this sub-compartment in oesophagus were cDCs (Figure 5.31A), agreeing with results seen from bulk RNA (Figure 4.7). Contrastingly, inoperable OAC had much fewer cDCs, but was dominated by infiltration of monocytes and macrophages, particularly TAMs. Zooming in on these monocytes and macrophages, the enrichment of TAMs seen at PreTx appeared to be strongly reduced during the ICI-only and immunochemotherapy periods, with the predominant subtype present at PostTx being M1-like macrophages (Figure 5.31B). Even after removing patient OAC-GDBD, who contributed the majority of some M1-like subclusters, a reduction in TAMs by therapy was still seen, which was already visible after the ICI-only window (Figure 5.31C). This suggests this immunochemotherapy protocol is able to induce a more inflammatory phenotype in myeloid cells. In summary, phagocytic cell types show a high level of heterogeneity in this dataset, and phagocytic states appear to be particularly sensitive to change during immunochemotherapy.

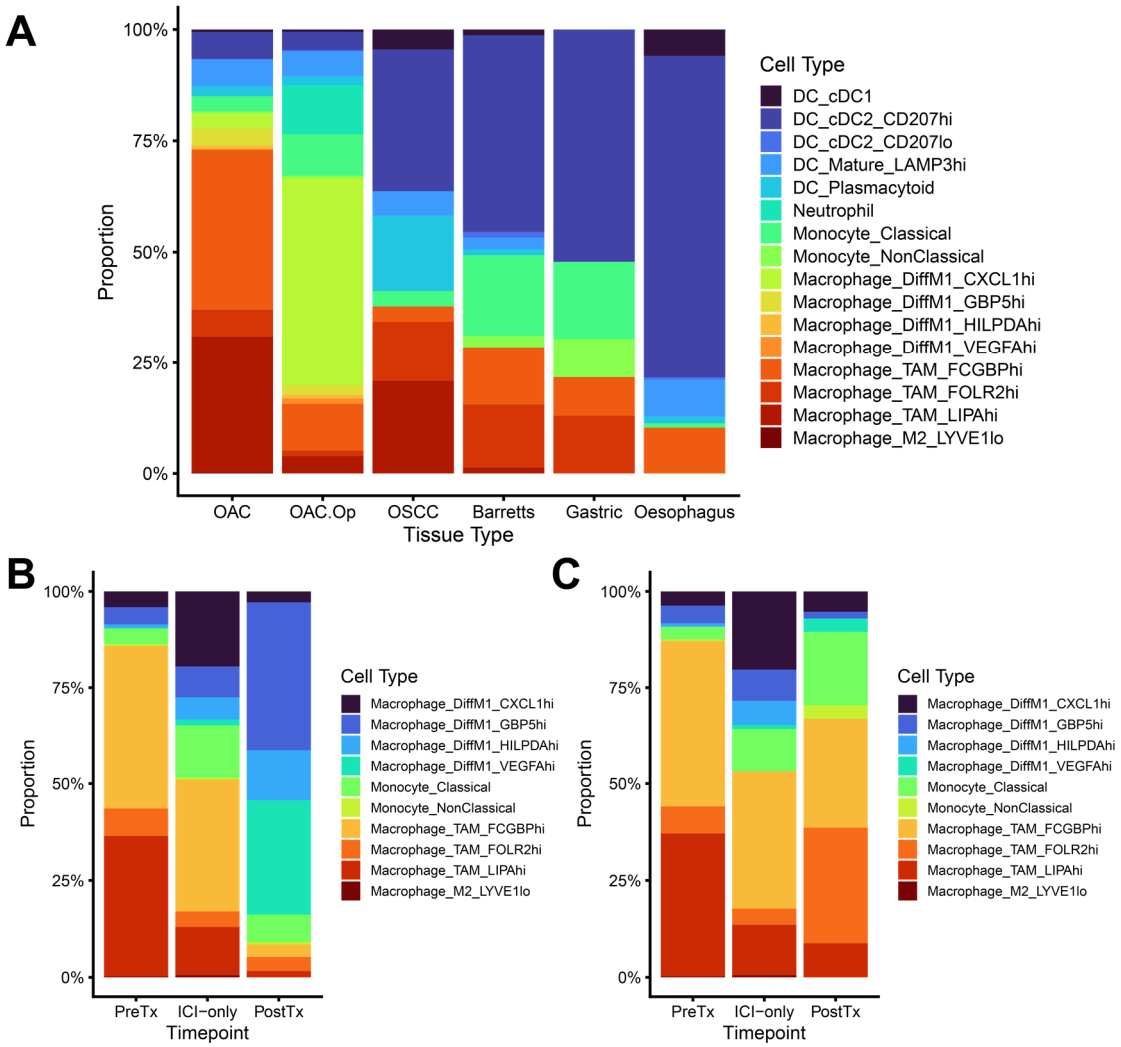


Figure 5.31: Differences in myeloid cell states between tissue types and treatment timepoints. A) Proportion of all pre-treatment phagocytic immune cells (without mast cells) that fall into indicated cell states for all tissue types, with all samples from a given tissue grouped as a single bar. B) As in A, except showing only macrophage and monocyte states for just the inoperable OAC biopsies, and how the proportions change over treatment timepoints. All samples for the given timepoint are grouped as a single bar. C) Same as B, but patient OAC-GDBD, who contributes the majority of some M1-like clusters, is excluded.

5.4 Deconvolution to understand and predict immunochemotherapy outcomes in OAC

5.4.1 Background and validation

The LUD2015-005 oesophageal pathology cell atlas is a powerful resource for understanding the profiled disease states and their response to treatment. However, the statistical power of this dataset to predict outcome on immunochemotherapy for the full inoperable OAC cohort is limited by the reduced subset of patients that were profiled using scRNA-seq (n = 8). Fortunately, the detailed transcriptional profiles of each cell type defined in the previous

sections can be leveraged to gain additional information from the bulk transcriptomes described in Chapter 4, which were available from 33 of 35 patients in the inoperable OAC cohort. Using a process known as deconvolution, gene expression profiles from scRNA-seq data can be used to estimate the contribution of each cell type to a given bulk transcriptome, thus increasing the statistical power to detect cell types that may correspond with patient outcomes. A deconvolution workflow using the BayesPrism algorithm[246] was therefore implemented to clarify the role of specific cell types in predicting outcome in the inoperable OAC cohort. Transcriptome profiles for 121 of 122 cell subtypes defined in section 5.3 were passed to BayesPrism to estimate the proportion of the 45 broad cell types they comprised. Schwann cells were excluded from the deconvolution workflow due to their rarity and high degree of dissimilarity from other cell types.

As there was insufficient material suitable for histological validation of deconvolution findings in the inoperable OAC cohort, deconvolution of a test cohort was first performed to ensure the workflow returned sensible results. For this *in silico* validation analysis, an RNA-seq dataset published by Maag et al. was assessed, where 19 histologically-confirmed biopsies of normal squamous oesophagus, 19 of BO, 8 of BO with low-grade dysplasia (LGD), and 17 of OAC were profiled[247].

Deconvolution results were assessed to determine whether cell composition estimates agreed with the expected composition of these tissue types from the analysis of the LUD2015-005 scRNA-seq dataset in section 5.3. First, the ability to correctly identify the appropriate epithelial cell type was evaluated. Indeed, deconvolution estimated that normal squamous cells made up the vast majority of normal oesophagus biopsies, but a much smaller fraction of BO and OAC samples (Table 5.3; Figure 5.32). Similarly, the BO cell type identified by scRNA-seq was found to be highly enriched in BO tissue (with or without LGD), but was present at a lower level in OAC and at a minimal level in normal oesophagus. The estimates for OAC cell composition were less specific to OAC biopsies, perhaps due to the greater heterogeneity seen in the malignant OAC clusters and the close relation particularly of operable OAC to BO. Nevertheless, on average, the highest OAC proportions were estimated in OAC tissue by a fair margin.

Interestingly, the OAC signature was higher in BO with LGD compared to non-dysplastic BO, perhaps recognizing that LGD represents a more advanced stage along the progression of this pre-malignant lesion towards OAC.

Estimated Proportion		Tissue Type			
		Oesophagus (n = 19)	BO (n = 19)	BO with LGD (n = 8)	OAC (n = 17)
Cell Type	Squamous	**0.910**	0.174	0.065	0.052
	BO	1.26E-04	**0.383**	**0.467**	0.079
	OAC	0.016	0.219	0.321	**0.551**

Table 5.3: BayesPrism deconvolution correctly identifies epithelial type in closely related oesophageal pathologies. The mean proportion of reads assigned to each of the indicated cell type signature (rows) for each of the tissue types (columns) in the Maag et al. dataset are indicated.

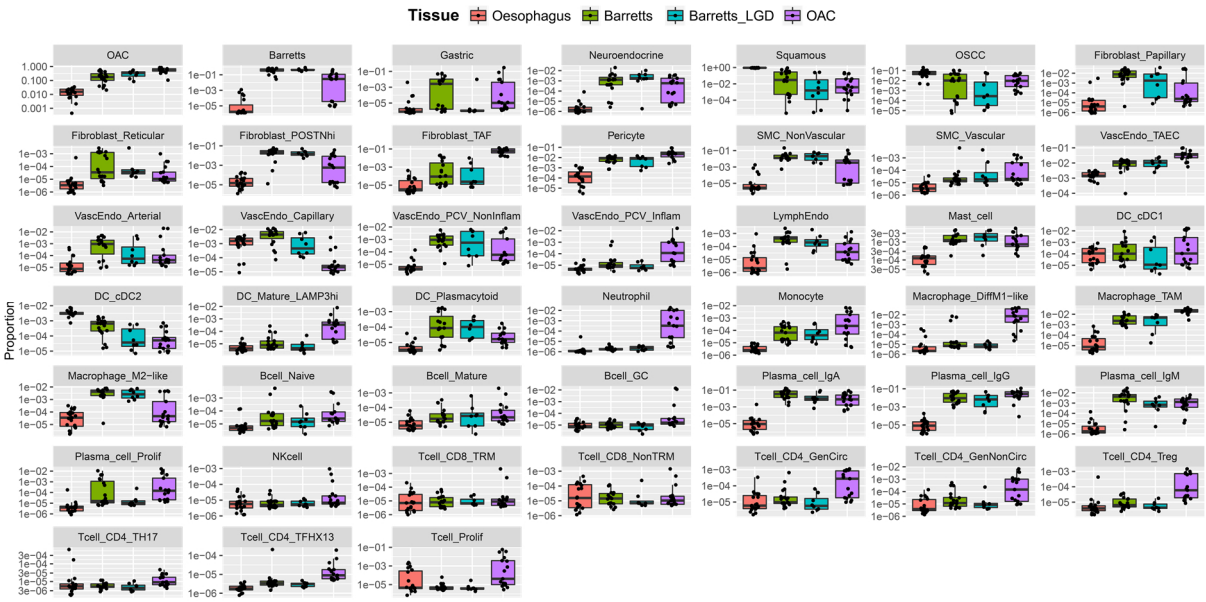


Figure 5.32: Deconvolution results show changes in cell type composition in metaplastic and malignant conditions of the oesophagus. For each subplot, the proportion of a sample contributed by reads from a specific cell type are shown. Within each cell type subplot, each point represents a single sample, and are grouped according to the tissue type of origin in the Maag et al. dataset. Proportion values are shown on a log10-transformed y-axis to assist with visualization.

Additional features of the microenvironment also appeared to be correctly detected by deconvolution. For instance, plasma cell contribution in the Maag dataset was greatly increased in BO and OAC over normal oesophagus (Figure 5.32), as was seen in the LUD2015-005 scRNA-seq dataset (Table 5.3). Deconvolution of this cohort also found that IgA enrichment was strongest in BO, while OAC had higher levels of IgG plasma cells than BO, as shown in scRNA-seq data. Treg, $T_{FH}X13$, and proliferative T cells estimates also showed the expected enrichment in OAC. For phagocytic immune cells, enrichment of the cDC2 cell type characteristic of normal oesophagus was seen in oesophageal biopsies, while tumour showed increased rates of TA macrophages, M1-like macrophages, and monocytes, as shown in section 5.3. Interestingly, deconvolution detected robustly increased levels of neutrophils in OAC, which were found only sporadically in scRNA-seq data for OAC, likely due to issues with detecting granulocytes using droplet-based methods. This shows deconvolution applications may be able to overcome some technical variation associated with scRNA-seq. This analysis of the Maag dataset also suggested that the transition from oesophagus through Barrett's to OAC was accompanied by increases in tumour-specific stromal cell types from scRNA-seq analysis, including TA fibroblasts, TA endothelium, and pericytes. Indeed, the overall cell composition profile was found to be a good tool for separating oesophagus from Barrett's and Barrett's from OAC (Figure 5.33). In summary, this deconvolution analysis identified changes across multiple compartments during the progression from normal oesophagus to OAC, which showed good agreement with findings from the LUD2015-005 scRNA-seq atlas. This suggests this deconvolution workflow returned appropriate biological findings from bulk RNA-seq data collected from oesophageal pathologies. This workflow was therefore taken forward for analysis of the LUD2015-005 cohort.

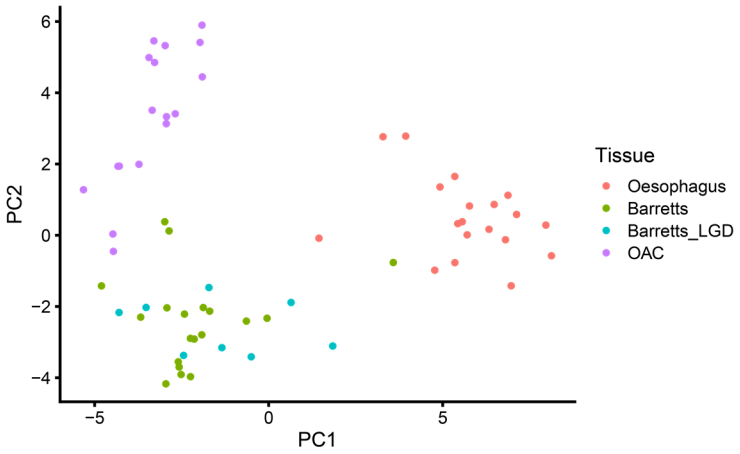


Figure 5.33: Deconvolution estimates of cell composition differentiate normal oesophagus, Barrett's oesophagus, and OAC. Dimensionality reduction using principal component analysis (PCA) of cell composition estimates for the Maag et al. dataset, using proportions from all 45 cell types. Proportions were log10-transformed and scaled prior to dimensionality reduction. Each point is coloured by the tissue type of origin.

5.4.2 Deconvolution of the LUD2015-005 inoperable OAC cohort

5.4.2.1 Deconvolution suggests, but cannot confirm, varying tumour purity of OAC biopsies

Having established the ability of deconvolution to distinguish between oesophageal phenotypes in bulk RNA-seq, the workflow was next applied to the LUD2015-005 inoperable OAC cohort. One potential confounder revealed by deconvolution was that some samples had low estimated tumour purity, particularly in those with high squamous signature (Figure 5.34). These samples with low OAC estimates were particularly noticeable at PostTx, where a return to normal squamous phenotype in some patients has already been suggested (Section 4.4.1), but could also be seen in some PreTx and ICI-only samples. However, the lone patient with a high-squamous, low-OAC PreTx bulk RNA-seq estimate and available PreTx scRNA-seq (OAC-JCNP, separate biopsies) showed high tumour purity (64% OAC) and no cells identified as squamous epithelium in scRNA-seq data, in stark contrast to the deconvolution findings. While inter-biopsy variation may be a confounder, these data disagreed as to whether the high squamous estimates truly represented low-purity samples with contaminating normal oesophagus. An alternative explanation, such as OAC cells expressing some squamous genes, could also perhaps explain these deconvolution results. As the available data could not prove either hypothesis, all samples were taken forward for analysis, regardless of deconvolution-assessed purity.

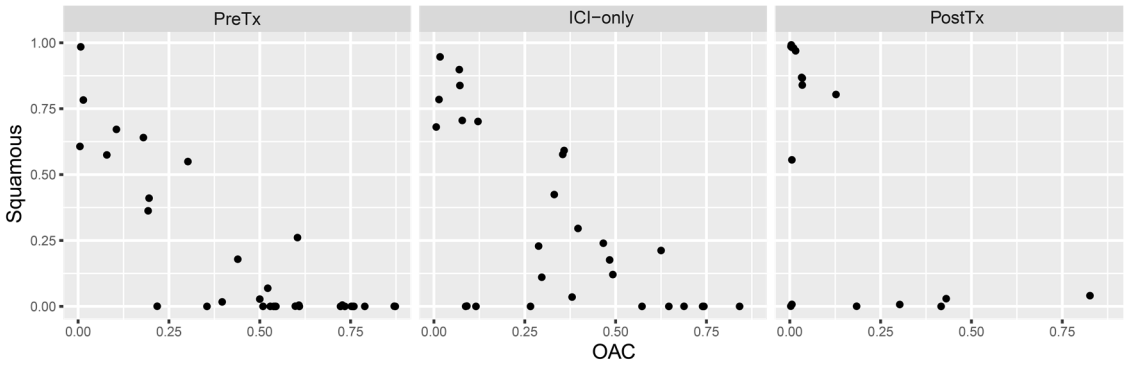


Figure 5.34: Deconvolution-assessed tumour purity. Estimated proportions of OAC and squamous cell types in bulk RNA-seq tumour samples as assessed by BayesPrism, faceted according to timepoint. A squamous-high sample is labelled with the OAC and squamous percentage detected in scRNA-seq data from a separate biopsy.

5.4.2.2 PreTx monocyte-like cells predict improved outcome on immunochemotherapy

First, deconvolution results from PreTx OAC biopsies were assessed to look for any links between cell composition and patient outcome. An overview of the estimated PreTx TME was first conducted. In the phagocyte cell compartment, hierarchical clustering showed one patient grouping with particularly high levels of monocyte infiltration that attained a high rate of CB (8/12, Figure 5.35). No such enrichment for CB was noted when assessing any of the other cell compartments (data not shown).

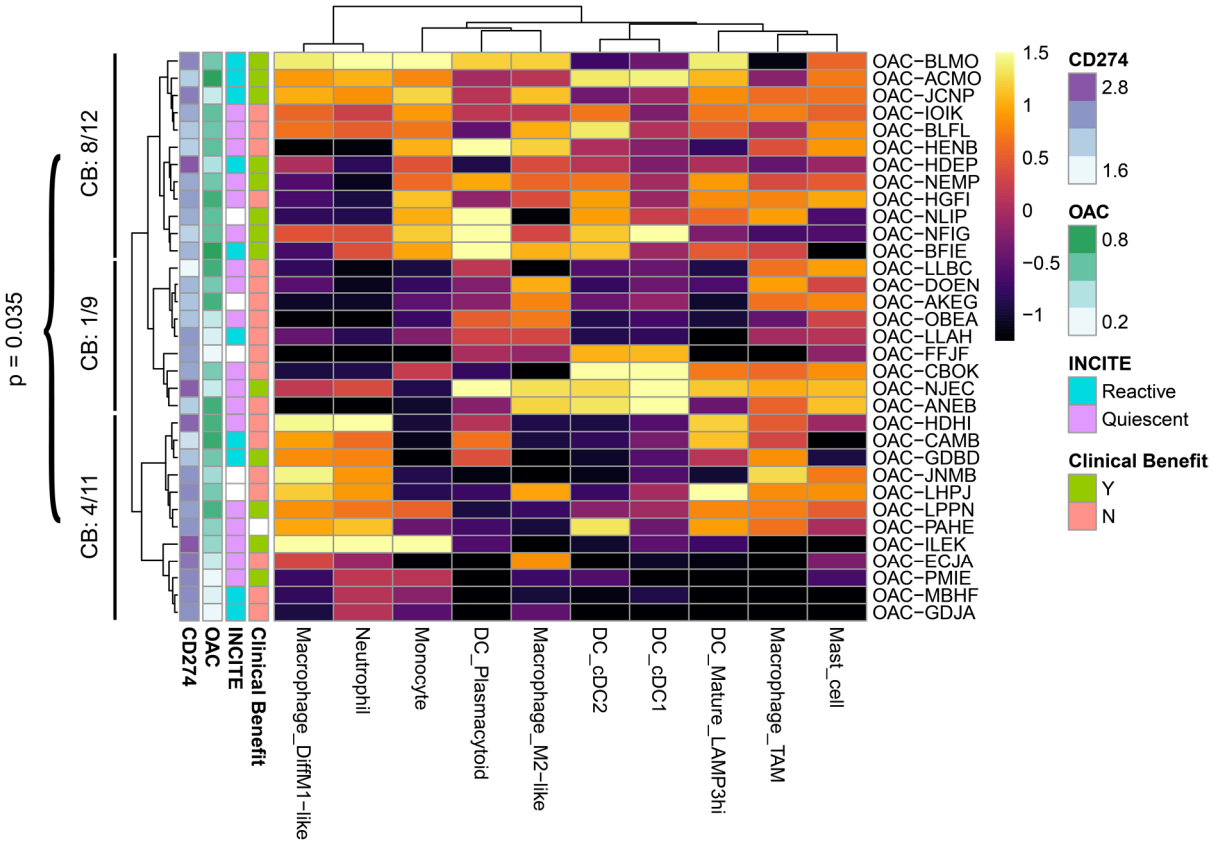


Figure 5.35: PreTx TME composition phenotypes associate with clinical benefit in the LUD2015-005 inoperable OAC cohort. Log10-transformed cell proportion estimates for all cell types in the phagocytic immune cell compartment were scaled by row, and are used to colour the heatmap cells. Hierarchical clustering was conducted using ward.D linkage. Estimated tumour purity (as deconvolution-assessed OAC proportion), INCITE grouping, and PD-L1 expression are shown as a row annotation. The rates of clinical benefit are shown for the 3 major hierarchical cluster branches. Displayed p-value for the difference between these rates was calculated using a 3-sample test for equality of proportions.

To clarify whether monocytes or any other individual cell type was associated with patient survival outcomes on immunochemotherapy, the estimated proportions of all 45 cell types were directly compared with patient survival. This analysis showed that monocytes were the only cell type significantly linked with OS following p-value correction, with patients having an increased level of monocytes showing greater OS (Table 5.4). The presence of Tregs also had a trend towards improved OS, as well as differentiating M1-like macrophages and neutrophils, suggesting a broader inflammatory myeloid phenotype could contribute towards improved prognosis, although none of these cell types reached significance following p-value adjustment. Similarly, vascular SMCs showed the strongest trend towards reduced OS, although the adjusted p-value was not significant.

Cell Type	Hazard Ratio (95% CI)	p.value	FDR
Monocyte	**0.40 (0.23-0.69)**	**0.001**	**0.047**
Tcell_CD4_Treg	0.61 (0.38-0.97)	0.038	0.665
Macrophage_DiffM1-like	0.62 (0.38-1.01)	0.054	0.665
Neutrophil	0.65 (0.41-1.02)	0.059	0.665
SMC_Vascular	1.39 (0.92-2.09)	0.122	0.853

Table 5.4: PreTx cell types with greatest association with OS. A univariate Cox regression with estimated cell type proportions (log10-transformed, scaled, and centered prior to regression) and OS was conducted. Hazard ratio with 95% CI and unadjusted p-value from the Cox regression are shown for the results with lowest p-values (p-value < 0.15). P-values for all 45 cell types were adjusted for FDR. Results shown in bold meet the FDR significance threshold (FDR < 0.1).

In addition to the Cox regression, using the cohort median of the monocyte-like cell proportions to classify patients into binary groups effectively stratified patient outcomes, with the monocyte-high group showing much improved OS compared with monocyte-low patients (median OS 24.3 vs. 7.6 months; Figure 5.36A). This difference remained significant even when removing potentially low-purity samples with an estimated OAC proportion less than 25% (log-rank p-value 0.01, data not shown). As a representation of a cohort treated with non-ICI therapies, similar deconvolution was also conducted on advanced OAC patients from the TCGA, which showed no significant difference between monocyte-high and monocyte-low tumours (Figure 5.36B). This suggests the predictive power of PreTx monocyte infiltration may be immunochemotherapy-specific, similar to TMB.

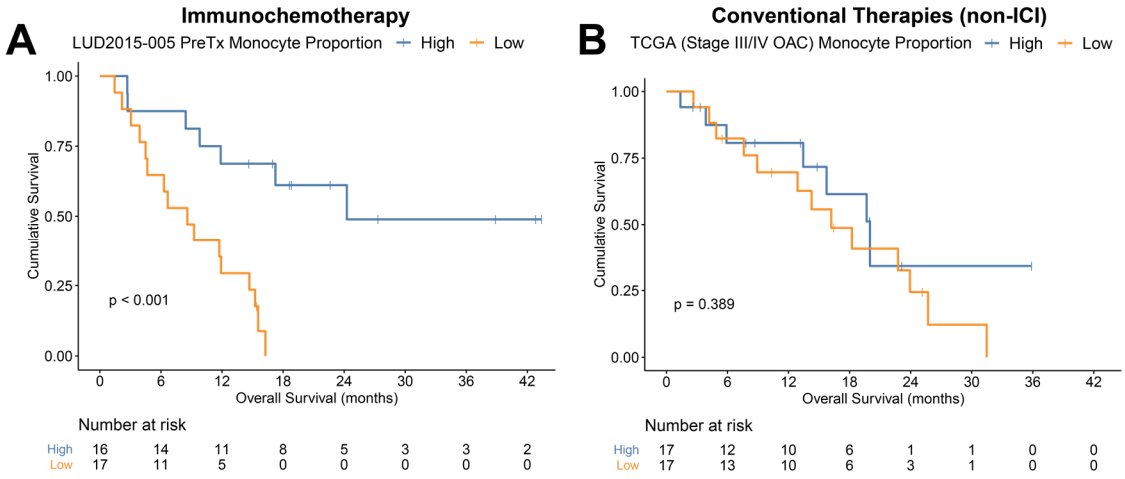


Figure 5.36: PreTx monocyte-like proportion is predictive of OS on immunochemotherapy cohort, but not under conventional therapies (TCGA). Estimated monocyte-like cell proportions were split by the cohort-specific median, to define monocyte-high and monocyte-low groups for the A) LUD2015-005 and B) TCGA cohorts. For TCGA, only stage III and IV OAC tumours are included. Displayed p-value for the difference between survival curves is calculated by the log-rank method.

Next, the relationship between monocyte proportion and TMB was assessed, to determine whether these strong predictive biomarkers in this cohort were highly correlated or independent predictors of improved patient outcome. Interestingly, there was only a non-significant trend for correlation between monocyte proportion and TMB, and in a multivariate model including both markers, both monocyte proportion and TMB remained significant predictors of greater OS in this cohort (Figure 5.37). In summary, monocyte proportions showed additive predictive power to TMB, such that having both high monocyte proportions and high TMB predicted significantly better outcomes on immunochemotherapy in this cohort.

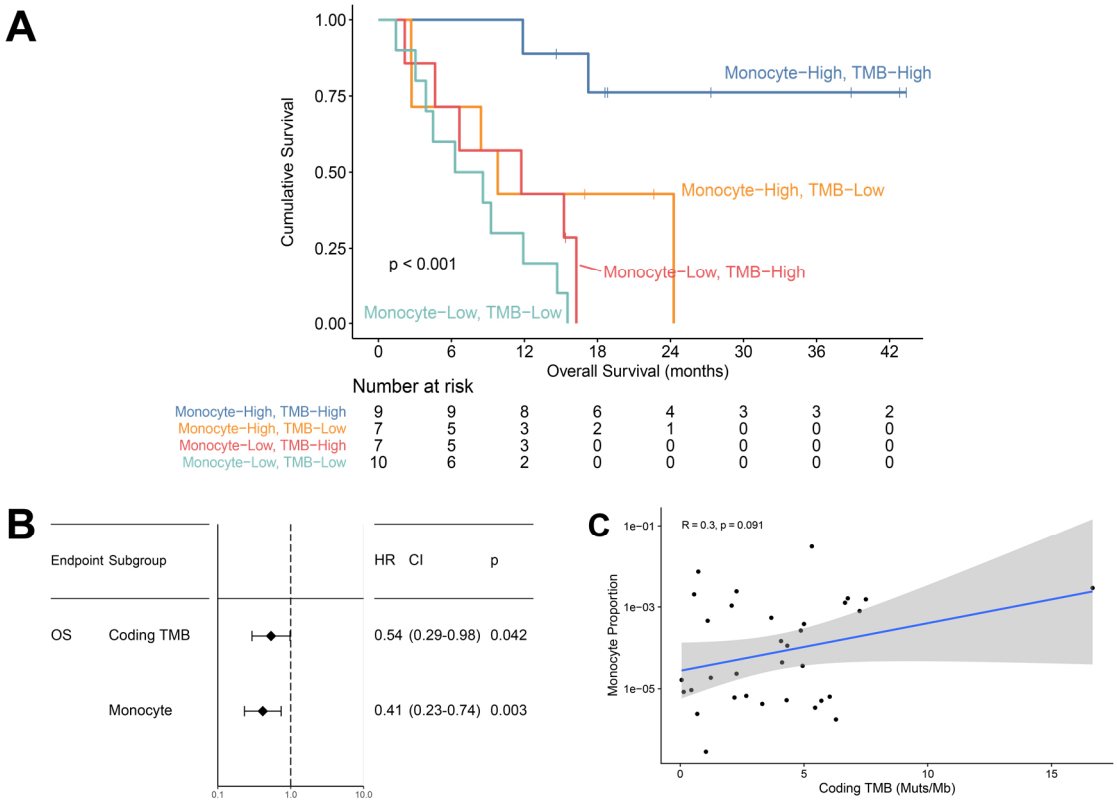


Figure 5.37: Monocyte-like proportion and TMB independently predict improved OS. A) Estimated monocyte-like cell proportion and coding TMB (muts/Mb) were split by the cohort-specific median, to define four groups based on these two binary variables. The displayed p-value is calculated by the log-rank test, testing the overall difference of the four survival curves. B) Cox regression conducted directly on estimated monocyte proportion (log10-transformed) and coding mutational burden (muts/Mb). Both values were scaled and centered prior to regression. C) Monocyte proportion (log10 transformed) and coding TMB (muts/Mb) are compared. P-value and correlation coefficient calculated using the Pearson method.

To compare the deconvolution findings with the results from traditional bulk RNA-sequencing, the expression context of the innate immune-related DEGs significantly associated with CB in section 4.5.1.1 was evaluated in the scRNA-seq data. Indeed, many bulk DEGs (*CLEC12A*, *FPR1*, *C5AR1*, *FCAR*) had higher average expression in monocyte cell clusters than any other cell type, although many of these marker genes were shared with M1-like macrophages and neutrophils as well (Figure 5.38). This provided further evidence that the inflammatory myeloid DEGs associated with CB in section 4.5.1.1 was likely linked most closely with the presence of these monocyte-like cells, although other inflammatory myeloid cell types with transcriptional similarity may have some contribution.

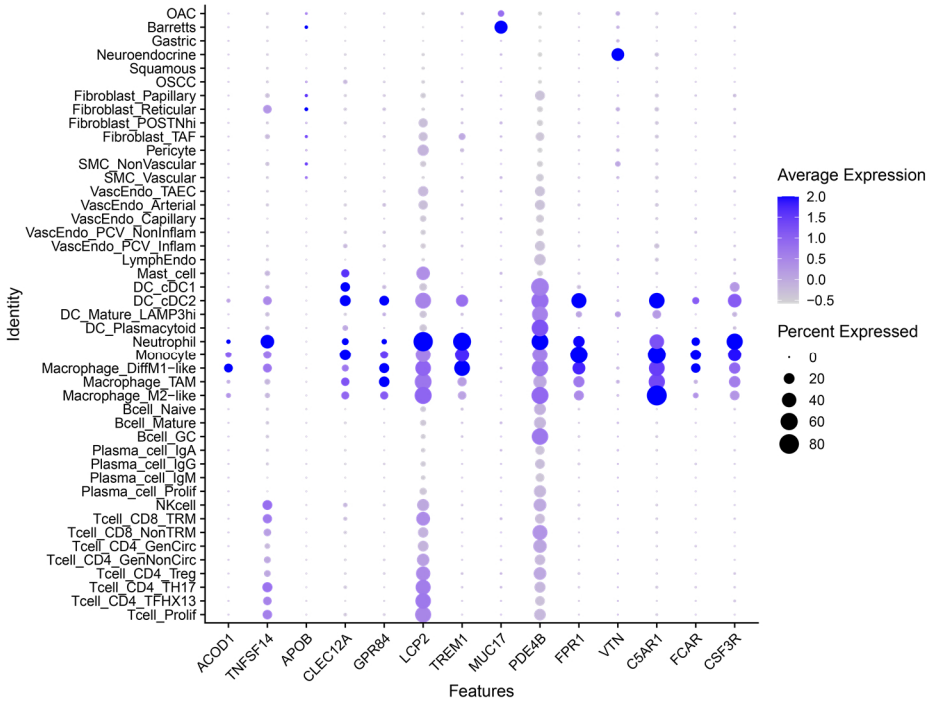


Figure 5.38: Cell type context of innate immune markers predictive of CB in bulk PreTx differential expression. Significant DEGs from 4.5.1.1 which fell into innate immune-related gene sets were assessed for their expression in scRNA-seq data. Dots for each cell type are coloured according to their average scaled, log-normalized expression, while the size of the dot is the percentage of cells in that cell type with any detectable expression.

With the exception of moderate *IL1B* expression, monocyte-like cells did not express as high levels of cytotoxic, antigen-presenting, or checkpoint molecules as other myeloid cell types (see section 5.3.3.4), and the rationale for why monocytes had stronger links with patient outcome were therefore unclear. One hypothesis could be that immature monocyte-like cells could serve as a pool that could differentiate into more inflammatory M1-like macrophages or perhaps also

antigen presenting monocyte-derived dendritic cells in response to treatment. Indeed, trajectory analysis suggested that monocyte-like cells represented an intermediate cell state that could differentiate towards a number of more mature myeloid states, including a differentiation trajectory towards M1-like macrophages and another through cDCs towards LAMP3^{hi} mature DCs (Figure 5.39). In agreement with this notion, patients with greater monocyte infiltration at PreTx tended to have higher amounts of mature LAMP3^{hi} DCs following the ICI-only window, although neither LAMP3^{hi} DCs nor M1-like macrophages showed significant differences following p-value adjustment against all cell types (Figure 5.40A). Lymphatic endothelium and non-TRM CD8^{+} T cells had similar trends towards higher levels in this group, likely due to the propensity of the mature LAMP3^{hi} DC cell type to secrete a number of chemokines that bind to lymphatic endothelium receptors to promote lymphangiogenesis, including the *CCL22*/*ACKR2* axis[248] (Figure 5.40B-C). In summary, monocyte-high cells are associated with higher levels of mature pro-inflammatory myeloid cells, lymphatic endothelium, and non-TRM CD8^{+} T cells following therapy, although these do not reach significance following p-value adjustment. Monocytes may help drive this pro-inflammatory state at ICI-only by differentiating into the pro-inflammatory myeloid cells directly, or otherwise promoting their influx or expansion.

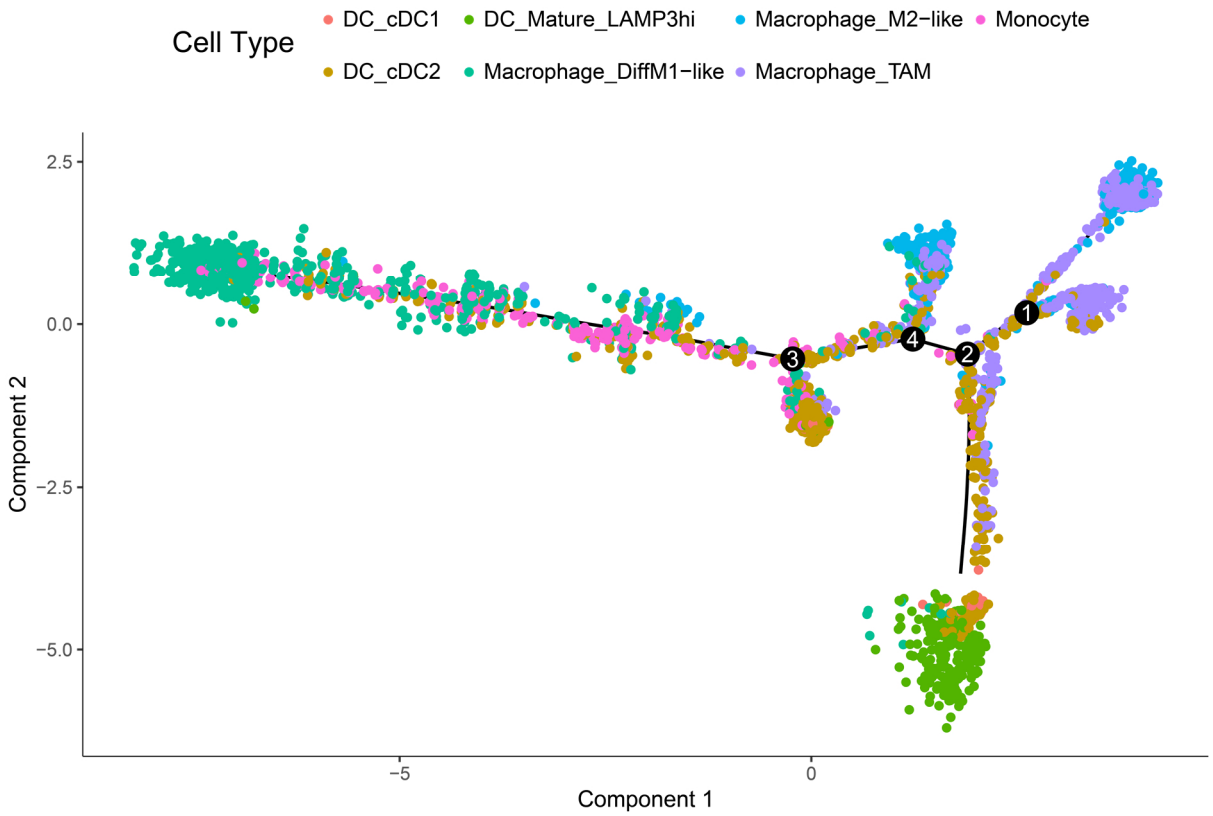


Figure 5.39: Trajectory analysis of monocytes, macrophages, and dendritic cells. Further dimensionality reduction was conducted on these cell types from the LUD2015-005 dataset using the DDRTree method of Monocle 2, regressing out the effect of tissue preparation (fresh vs. frozen). The reconstructed trajectory is overlaid on top of the dimensionality reduction, and trajectory branch points are labelled with a number. Cells are coloured by the cell type identity assigned during the analysis of the full dataset.

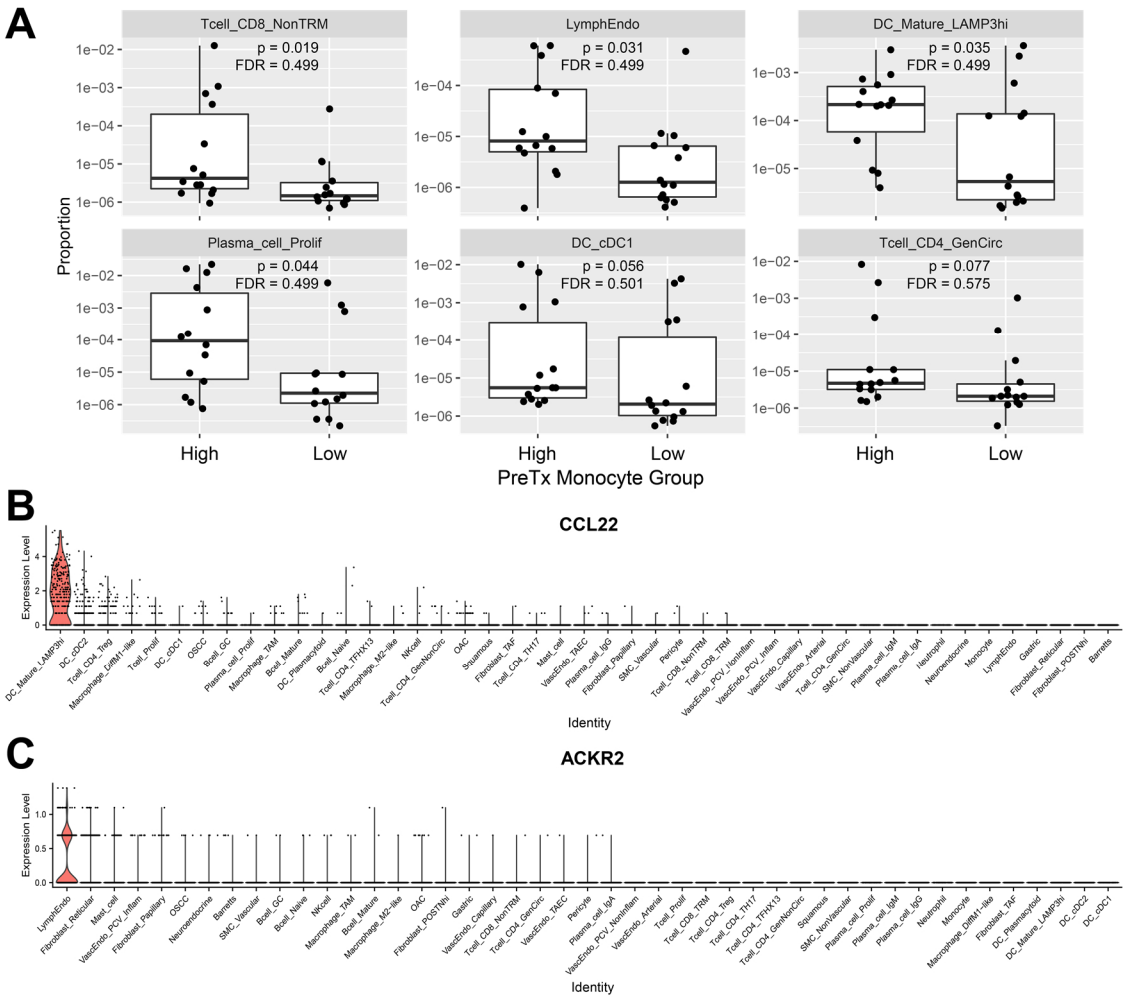


Figure 5.40: ICI-only cell composition phenotypes associated with PreTx monocyte infiltration. A) All samples with paired PreTx and ICI-only samples available were split by their PreTx monocyte-like median proportion into high and low groups, and then ICI-only proportion of all cell types was tested between groups. Proportions and unadjusted p-values from the Mann-Whitney U-test are shown for the top 8 significant cell types at this ICI-only timepoint, as well as FDR-adjusted p-values (all equal and not significant). Bottom: Log-normalized expression of B) *CCL22* and C) *ACKR2*, which are examples of a LAMP3^{hi} DC-expressed chemokine and its corresponding lymphatic endothelium-expressed receptor, respectively. Cell types are sorted based on average expression of each gene.

At PreTx, one patient (OAC-ILEK) stood out as an outlier in regard to the monocyte composition of their tumour, having an estimated monocyte-like cell proportion of 3.05%, which was far higher than all other patients (maximum 0.73%). This patient also had an excellent outcome on the LUD2015-005 trial, attaining partial response, clinical benefit, and ongoing OS at 43.6 months. Therefore, this patient was examined in depth to determine whether there were any additional factors that might explain the extreme level of monocyte infiltration and positive outcome in this patient.

As monocytes generally have a key role in pathogen defence, and the monocyte cells in this dataset express a number of key pathogen-activated signalling and/or chemotactic receptors (*ACOD1*, *TREM1*, *FPR1*), it is possible the local microbiome could play a role in monocyte infiltration into tumours. Therefore, the local microbiome of the PreTx tumour biopsy from OAC-ILEK was examined. While no strong evidence for viral infection was detected, this patient appeared to have an ongoing *Helicobacter pylori (H. pylori*) infection, which is associated with gastric oncogenesis through an inflammatory mechanism[249]. 26% of reads not mapped to the human genome from this patient's PreTx gastric sample were assigned to *H. pylori*. The tumour sample from this patient also had a relatively high level of reads assigned to *Helicobacter* (1%). These were both far higher than any other sample from this dataset (next highest: 0.06%), suggesting high likelihood of an active *H. pylori* infection present in the normal gastric biopsy site, if not the OAC tumour itself.

As *H. pylori* is more strongly linked with gastric cancers[37,249], the Kim et al. gastric cancer ICI cohort was examined to assess whether any further evidence could be found linking *H. pylori* with ICI outcome. While no sample in this RNA-sequencing data showed as high of a level of *H. pylori* alignments as the gastric sample of OAC-ILEK, the three patients with the highest level of detectable *Helicobacter* in normal gastric biopsies (but not tumour biopsies) also achieved responses to their second-line ICI regimen (Figure 5.41). In summary, it is possible that the high levels of monocyte infiltration seen in OAC-ILEK could be downstream of an associated *H. pylori* infection. More broadly, while the sample size is small, tumours that arise following *H. pylori* infection in the stomach could potentially have greater susceptibility to ICI response in gastro-oesophageal cancers, which may in some cases be associated with increased intratumoural monocyte infiltration.

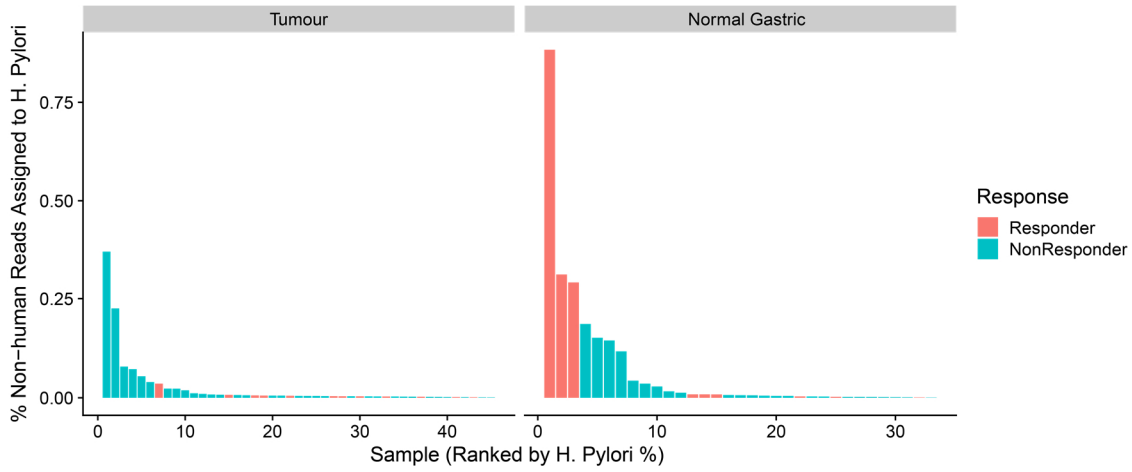


Figure 5.41: H. pylori prevalence in biopsies from the Kim et al. gastric cancer dataset. Percentage of unaligned reads assigned to the Helicobacter genus are shown for each sample, ordered by decreasing magnitude. Plots are separated into separate facets for gastric tumour (left) and normal gastric (right) biopsies.

5.4.2.3 RERGL⁺ vascular SMCs mark early resistance at the ICI-only timepoint

At the ICI-only timepoint, all cell types were again tested in univariate Cox regression with OS. At this timepoint, only vascular SMCs were significantly associated with OS following p-value adjustment, strongly predicting poor outcome at this timepoint (Table 5.5). There was also a trend towards Vascular SMCs predicting worse outcome in PreTx (section 5.4.2.2), suggesting this cell type could be linked with pre- and on-therapy markers of resistance, although the effect was stronger after four weeks of ICI (Figure 5.42). This pattern was highly similar to what was shown for the gene *RERGL* in the bulk RNA-sequencing, as the only gene predictive of poor outcome at both of these timepoints (section 4.5.2). Indeed, *RERGL* was highly specific to vascular SMCs in this dataset (Figure 5.21E; Figure 5.43). Therefore, the vascular SMC cell type appeared to predict treatment resistance in this immunochemotherapy cohort, and *RERGL* expression may be a useful marker of this cell type in bulk sequencing data when deconvolution is not available.

Cell Type	Hazard Ratio (95% CI)	p.value	FDR
SMC_Vascular	**2.02 (1.30-3.15)**	**0.002**	**0.080**
SMC_NonVascular	1.64 (1.03-2.61)	0.038	0.589
Macrophage_TAM	1.72 (0.97-3.06)	0.063	0.589
Gastric	1.49 (0.97-2.28)	0.067	0.589
Fibroblast_Reticular	1.48 (0.96-2.28)	0.076	0.589
Barretts	1.48 (0.92-2.37)	0.106	0.589
Tcell_CD4_Treg	0.68 (0.42-1.11)	0.119	0.589
Mast_cell	1.56 (0.89-2.72)	0.120	0.589
VascEndo_Arterial	1.37 (0.91-2.04)	0.128	0.589

Table 5.5: ICI-only cell types with greatest association with OS. A univariate Cox regression with estimated cell type proportions (log10-transformed, scaled, and centered prior to regression) and OS was conducted. Hazard ratio with 95% CI and unadjusted p-value from the Cox regression are shown for the results with lowest p-values (p-value < 0.15). P-values for all 45 cell types were adjusted for FDR. Results shown in bold meet the FDR significance threshold (FDR < 0.1).

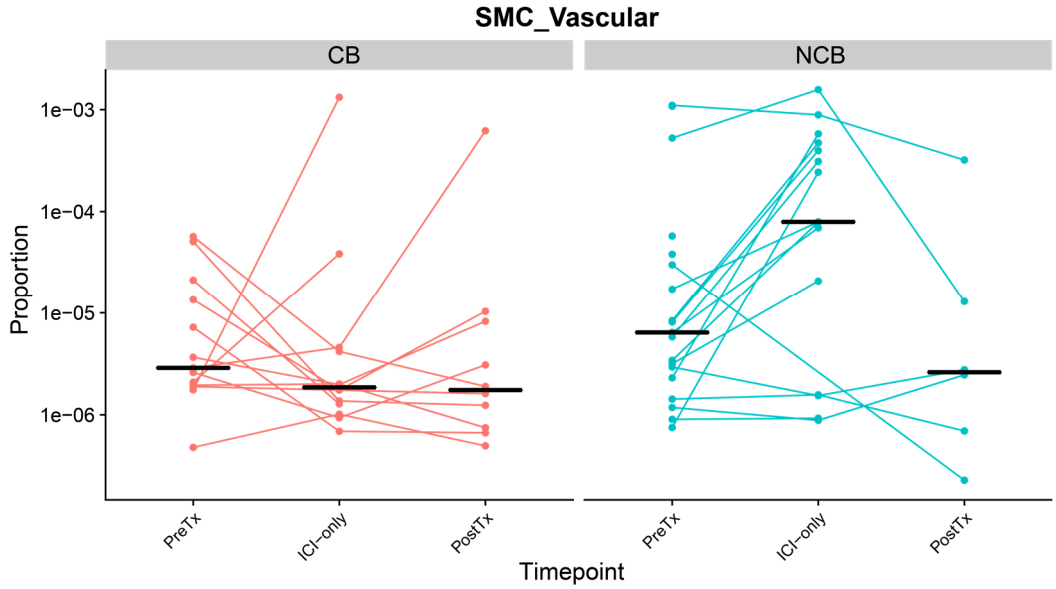


Figure 5.42: Estimated vascular SMC proportion during the course of immunochemotherapy. Deconvolution-estimated vascular SMC proportion for each sample is shown across all three treatment timepoints, with values from the same patient connected by a line. Patients are split into facets according to their outcome (CB or NCB). The median value for each timepoint within each facet is shown by a black line. The y-axis is log10-transformed to assist visualization.

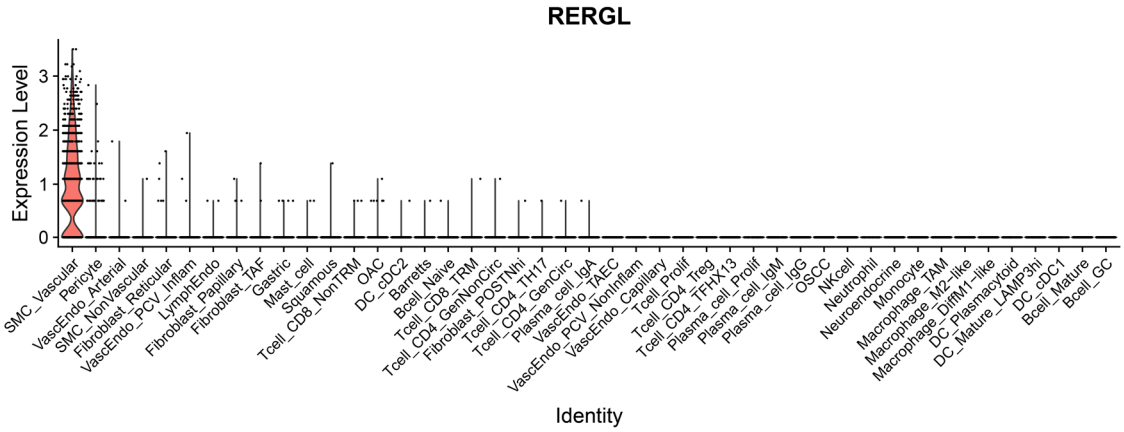


Figure 5.43: RERGL is a specific marker of the vascular SMC cell type. Log-normalized expression of *RERGL* is shown across all 45 cell types profiled for deconvolution, sorted by average expression of this gene.

5.4.2.4 Specific T cell subtypes drive INCITE reactivity during the ICI-only window

In section 4.6.1.1, a transcriptional signature (INCITE) involving apparent cytotoxic cell infiltration and activation was found to be induced to some extent in nearly all patients following four weeks of ICI-only. To clarify which cell types were primarily responsible for increases in this signature during ICI-only, ICI-induced changes in cell composition were assessed using deconvolution, and results were compared based on INCITE reactivity (Figure 5.44). This analysis revealed that non-T_{RM} CD8 T cells had significantly higher increases during ICI-only in INCITE-reactive patients than in INCITE-quiescent, although there was also a similar trend for T_{RM} CD8 T cells that did not meet the adjusted p-value threshold. Specialized CD4 T cell subsets (T_{FH}X13 and Th17) also had apparent increases following therapy in INCITE-reactive patients which did not reach statistical significance. This suggests that INCITE reactivity during ICI-only was driven primarily by accumulation of nonT_{RM} CD8 T cells, although other lymphocyte subtypes likely contribute. Particularly given the non-significant trend towards increases in cell types associated with complex lymphoid structures in INCITE-reactive patients, including T_{FH}X13 cells, lymphatic endothelium, germinal B cells, and plasma cells (Figure 5.44), this increase in non-resident CD8 T cells could represent the development of T cells with a circulatory or memory-like phenotype in intratumoral lymphoid structures. Regardless, increases in INCITE reactivity were primarily driven by this nonT_{RM} CD8 T cell subset, although a variety of adaptive immune cells appear to contribute to the increases in inflammation apparent in INCITE-reactive patients following ICI-only.

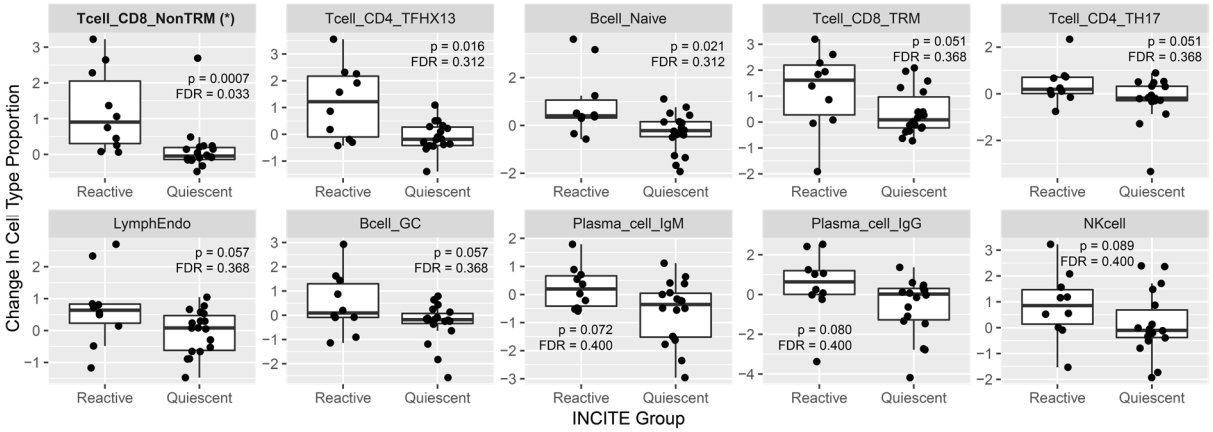


Figure 5.44: Cell type proportion changes during the ICI-only window associated with INCITE reactivity. Cell type changes were calculated as the log10-transformed proportion at ICI-only minus the log10-transformed proportion at PreTx. Samples were grouped by INCITE classification, and a non-parametric test (Mann-Whitney U-test) was calculated for difference between the groups for all cell types. The cell types with lowest p-values (p value < 0.10) are shown. P-value and FDR-adjusted p-values are shown; panels with significant results (FDR < 0.1) are titled in bold and annotated with an asterisk.

5.4.2.5 Cell types associated with outcome at the PostTx timepoint

At the PostTx timepoint, no individual cell types were associated with OS (neither from treatment onset nor from the time of the PostTx sample), likely due to the stringency of p-value adjustment and the smaller sample size at this timepoint (n = 17). However, some stromal and myeloid cell types showed a trend towards predictive power in this analysis, with reticular fibroblasts and cDC2 trending towards worse and greater OS respectively (Table 5.6). Additionally, there was a trend towards residual OAC at this timepoint predicting worse OS, while a return towards a squamous epithelial signature trended towards predicting greater OS. Indeed, while the OAC proportion was strongly reduced at the end of immunochemotherapy in most patients, it remained high in a small number of patients, particularly those with NCB (Figure 5.45).

Cell Type	Hazard Ratio (95% CI)	p.value	FDR
DC_cDC2	0.34 (0.16-0.76)	0.008	0.196
Fibroblast_Reticular	2.98 (1.32-6.73)	0.009	0.196
Monocyte	1.92 (1.03-3.59)	0.041	0.606
Pericyte	2.29 (0.93-5.62)	0.070	0.606
VascEndo_TAEC	1.98 (0.92-4.26)	0.081	0.606
Squamous	0.59 (0.32-1.1)	0.100	0.606
OAC	2.04 (0.86-4.82)	0.105	0.606
DC_cDC1	0.47 (0.19-1.18)	0.108	0.606
Plasma_cell_IgA	1.82 (0.82-4.04)	0.139	0.697

Table 5.6: PostTx cell types with greatest association with OS. A univariate Cox regression with estimated cell type proportions (log10-transformed, scaled, and centered prior to regression) and OS was conducted. OS used here is calculated from treatment onset, although similar results are seen when calculating from the date of PostTx biopsy (data not shown). Hazard ratio with 95% CI and unadjusted p-value from the Cox regression are shown for the results with lowest p-values (p-value < 0.15). P-values for all 45 cell types were adjusted for FDR. Results shown in bold meet the FDR significance threshold (FDR < 0.1).

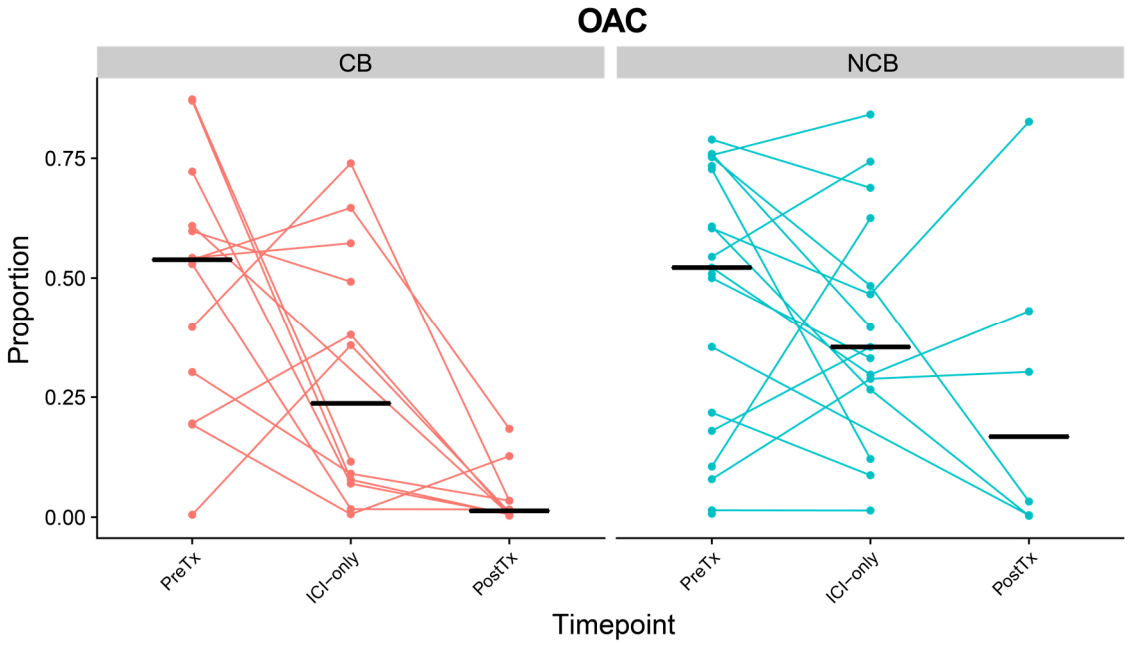


Figure 5.45: Estimated OAC cell proportion during the course of immunochemotherapy. Deconvolution-estimated OAC cell proportion for each sample is shown across all three treatment timepoints, with values from the same patient connected by a line. Patients are split into facets according to their outcome (CB or NCB). The median value for each timepoint within each facet is shown by a black line.

5.5 Further scRNA-seq approaches to understand immunochemotherapy outcomes

5.5.1 Predicting outcome using cell type-specific expression

Deconvolution analysis of cell composition showed a significant link with PreTx myeloid subtypes and patient outcome, but not epithelial or T cell subtypes, which are both key cell types of interest for anti-tumoural ICI response. One possibility is that these cell types could modulate efficacy of ICI through expression of different markers within cell types, rather than through differing distributions of specific cell subtypes. To assess this, T cell and OAC cells were queried at the single-cell level in patients with inoperable OAC from the LUD2015-005 scRNA-seq atlas, to find cell-specific expression of PreTx markers that are linked with CB.

A pseudobulk-based differential expression approach was employed to prioritize markers shared by multiple patients over those only found in cells from one patient (see Methods). This method returned a number of markers that were significantly different between OAC cells from CB and NCB patients (Figure 5.46A). The top gene associated with CB in this analysis was *IGFBP2* (FDR = 3.2e-10), which had high average expression in OAC cells from all four patients with CB but very low expression in all three with NCB (Figure 5.46A,C). Tumour-specific expression of this gene has been previously linked with an ICI-sensitive TME in melanoma[250]. NCB OAC cells, on the other hand, were enriched for a number of markers, including some cytokines (*IL33*, *CCL15*) and a decoy receptor for IL1β (*IL1R2*). None of these markers were called as significant DEGs in traditional bulk sequencing, probably due to the differing expression patterns seen for these genes in different compartments (Figure 5.46C).

A similar analysis conducted on all T cells also returned a number of DEGs between CB and NCB patients. CB T cells expressed higher levels of *ITM2C* and *SMPD3*, which are relatively uncharacterized in the context of T cells (Figure 5.46B). NCB T cells, however, had higher levels of a variety of interferon-stimulated genes (ISGs), including a number with key roles in antiviral responses (*OAS1*, *MX1*). These NCB patients were three of the four patients with the lowest levels of INCITE reactivity (Figure 4.17A), and those with the highest ISG levels (OAC-HDHI and OAC-HGFI) had particularly poor OS on ICI (4.7 and 2.7 months, respectively). Therefore, high

PreTx expression of ISGs in T cells was associated with poor inflammatory responses and poor clinical outcomes on immunochemotherapy. Additionally, two TCR variable region genes (*TRGV2*, *TRAV22*) were enriched in NCB T cells, suggesting the presence of specific TCR clonotypes at PreTx could also be associated with outcome (Figure 5.46B). Similar to the OAC analysis, marker genes from this analysis were expressed in a different pattern in the T cell compartment compared with other compartments, which could mask the ability to detect this finding using bulk data (Figure 5.46D). Therefore, this scRNA-seq differential expression analysis enabled the detection of genes whose cell-type specific expression patterns may influence outcome on immunochemotherapy in OAC.

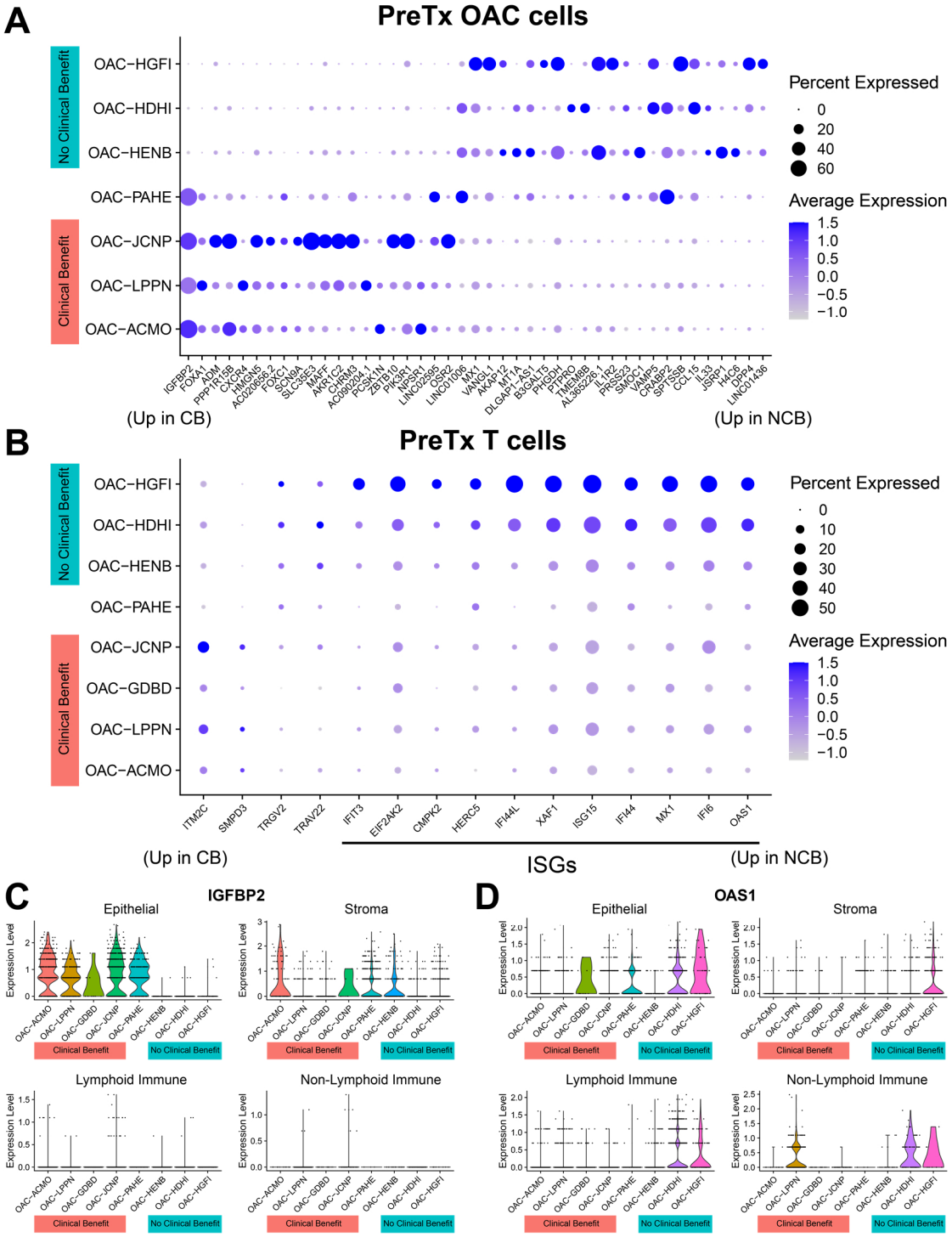


Figure 5.46: Markers of clinical outcome expressed specifically in OAC and T cells. A) Pseudobulk samples of OAC were compared to find DEGs between CB and NCB using DESeq2 (see Methods). All DEGs with FDR < 0.1 are plotted, with the size of the dot representing percentage expressed within each patient, and the colour the average expression (scaled log-normalized counts). DEGs are sorted according to FDR and the sign of change, such that genes most significantly associated with CB are at the left edge, and those most associated with NCB are at the right edge. B) As in A, but for T cell-specific DEGs. Bottom: Log-normalized expression of selected markers C) *IGFBP2* and D) *OAS1*, grouped by patient and separated into separate facets for each compartment.

5.5.2 Expression map of targeted checkpoint molecules

The direct activity of ICI agents targeting the PD-1/PD-L1 and CTLA-4 axes has not yet been fully characterized in oesophageal cancers, particularly regarding which cell types these agents might act on. Therefore, the cell type expression profiles of checkpoint molecules relevant to these axes were characterized. As expected, checkpoint receptors *PDCD1* (PD-1) and *CTLA4* were specifically expressed in T cells, with *CTLA4* showing more restricted expression to Tregs and T_{FH}X13 CD4 cell types, and *PDCD1* expressed more broadly across CD4 and CD8 T cell subtypes (Figure 5.47A). Both receptors were strongly expressed in proliferative T cell subsets as well. PD-1 and CTLA-4 ligands were most strongly expressed in myeloid cells, particularly mature $LAMP3^{hi}$ DCs. Contrastingly, these ligands were only rarely expressed in oesophageal cancers, and only a small number of OAC cells had detectable expression of PD-L1 (Figure 5.47B). This suggests that immunosuppressive modulation of these axes in this setting may be more commonly mediated through the TME rather than direct expression of coinhibitory ligands on OAC cells.

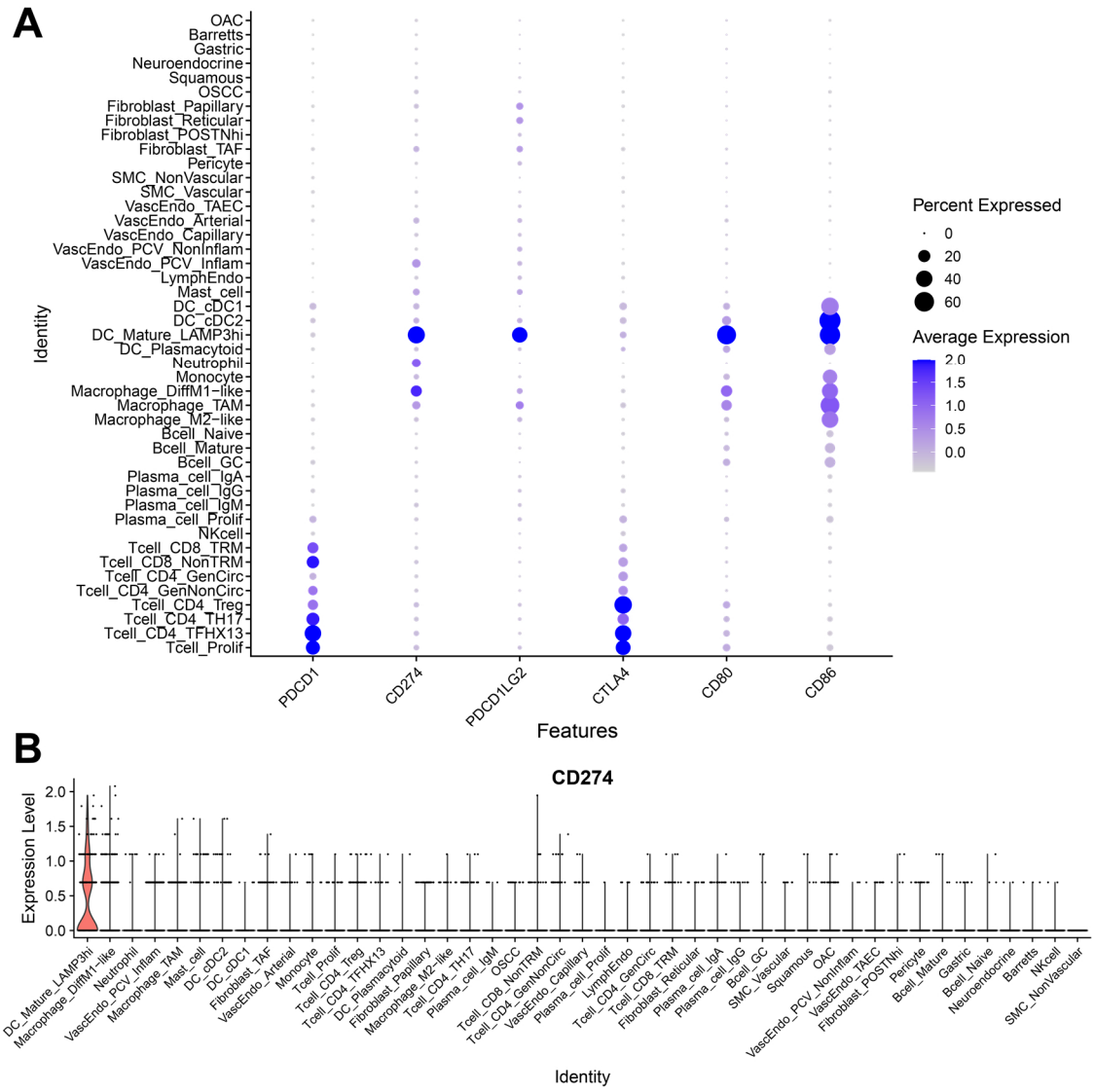


Figure 5.47: Cell-type expression context of checkpoint molecules in PD-1/PD-L1 and CTLA-4 axes. A) Expression of checkpoint molecules in axes modulated in the LUD2015-005 trial are shown. Dots for each cell type are coloured according to their average scaled, log-normalized expression, while the size of the dot is the percentage of cells in that cell type with any detectable expression. B) Log-normalized expression of *CD274* (PD-L1) is shown for each cell type, sorted according to the average expression for that cell type.

5.5.3 On-treatment tumour evolution

All tumours are likely to show some evolution under the selective pressure of treatment, such that subclusters remaining at ICI-only and PostTx may show more resistant characteristics than PreTx clusters which are effectively cleared by therapy. With the LUD2015-005 atlas, changes in gene expression within tumour cells can be directly assessed in response to treatment, which could be masked in analyses of bulk RNA-sequencing data from the same patients. This application of the scRNA-seq dataset therefore allows for further understanding of how oesophageal cancers respond to ICI-only and immunochemotherapy, and potentially identifying which changes correspond with greater odds of resistance. In this section, the transcriptional heterogeneity of each OAC patient's tumour subclusters will be examined, and changes over the course of therapy will be assessed. To identify changes that may result in higher likelihood of resistance to therapy, patterns of on-treatment transcriptional evolution will then be summarized and compared with patient outcomes.

5.5.3.1 OAC-HDHI (No Clinical Benefit)

This patient only had a single major group of cells at PreTx, and a small group of transcriptionally distinct ICI-only cells that clustered far apart during the analysis of the full Barrett's/OAC sub-compartment (see Figure 5.48). The ICI-only cells were difficult to resolve into a discrete cluster in this sub-analysis due to their small numbers. Therefore, marker expression profiling was performed by timepoint rather than by subcluster, as little intra-timepoint heterogeneity was detected. At ICI-only, while no change was seen in HLA gene expression, proteasome components (C10/*PSMB3*, p58/*PSMD3*) were strongly underexpressed, which could have knock-on effects on antigen presentation. ICI-only OAC cells also showed strong upregulation of a number of genes associated with ICI-resistance. This included genes related to EMT (*FN1*, *BMP2*, *TGFB1*, and *NNMT*, which have been shown to promote EMT in gastric cancer[251]), which is a major component of the IPRES ICI-resistance phenotype. Gamma-fibrinogen (*FGG*), which was associated with poor outcome at the ICI-only timepoint (section 4.5.2), was also upregulated in the ICI-only cluster, along with *EGFR*. Finally, CT-antigens were upregulated in cancer cells at this timepoint (*MAGEA3*; *GAGE2A*). This suggests complex resistance to ICI in this patient, including modulation of EMT and the complement-clotting cascade, which allows for the outgrowth of subclusters expressing putatively immunogenic CT antigens even in the context of a T cell stimulating therapy.

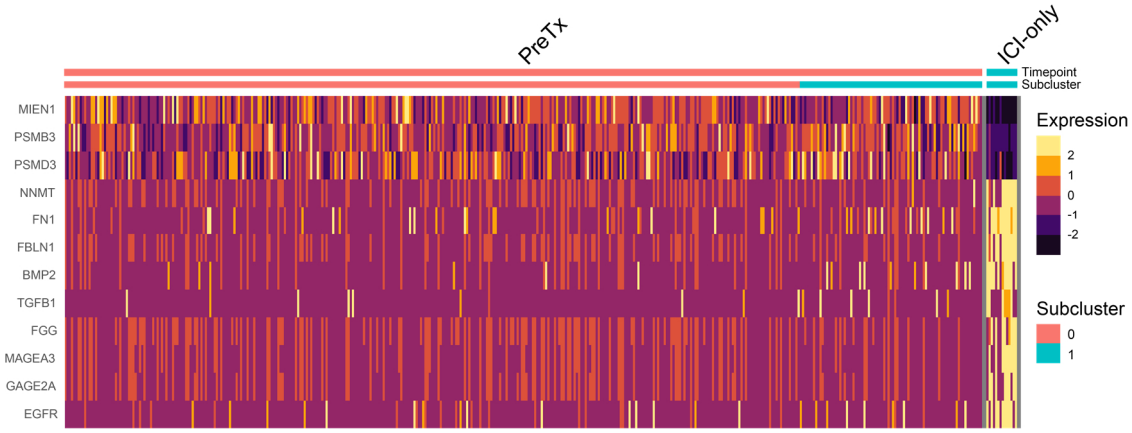


Figure 5.48: Heatmap of OAC-HDHI subcluster markers. Barrett's/OAC sub-compartment cells are sorted according to subcluster and timepoint. For this patient, timepoint was for primary sort key and marker evaluation, as identified clusters did not recapitulate the significant difference seen between PreTx and ICI-only cells in the full dataset for this patient, likely due to the small number of cells in this subset analysis. Scaled log-normalized expression values (from SCT workflow) are plotted for selected marker genes differentiating PreTx and ICI-only cells.

5.5.3.2 OAC-HENB (No Clinical Benefit)

OAC-HENB had two major subclusters (Figure 5.49): subcluster 1 was the primary subtype present at PreTx, while subcluster 0 predominated at ICI-only, although it was present at a smaller but not-insubstantial frequency at PreTx. The ICI-only-predominant subcluster was distinguished by downregulation of both class I and II HLA genes, as well as downregulation of interferon-stimulated genes (*IFI27*). Subcluster 0 had comparatively higher levels of early-goblet marker *SPINK4* and CT antigens (*CSAG1*, *LAGE3*). *APOBEC2* and RNA polymerase II components and regulators (*ELOC*, *POLR2E*) were also upregulated in this subcluster, suggesting modulation of both mutational and transcriptional processes in this subcluster expanding during ICI. In summary, we again see the outgrowth of a subcluster with potentially reduced antigen presentation and higher levels of CT antigens during the ICI-only window in a patient with poor outcome.

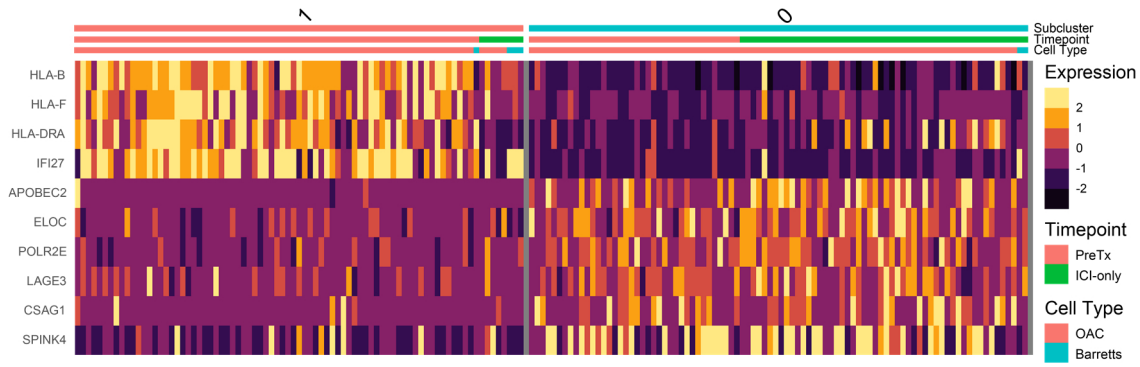


Figure 5.49: Heatmap of OAC-HENB subcluster markers. Barrett's/OAC sub-compartment cells are sorted according to subcluster, timepoint and identified cell type (OAC or Barrett's). Scaled log-normalized expression values (from SCT workflow) are plotted for selected marker genes differentiating identified subclusters.

5.5.3.3 OAC-HGFI (No Clinical Benefit)

This patient showed the highest level of overlap in tumour subclusters between timepoints. Two major subclusters were identified (0 and 1), which were both found at similar levels at PreTx and ICI-only (Figure 5.50); while subcluster 1 was found at a marginally higher proportion at ICI-only, 0 was still the predominant subtype at both timepoints. The subcluster with marginally greater levels at ICI-only was characterized by increased expression of apoptosis-resistance genes (*BIRC3*) and reduced expression of proteasome components

(C10/*PSMB3*; PA28α/*PSME1*). Overall, there was very little clear gene expression change during the ICI-only window in this patient, perhaps suggestive of a weaker anti-tumour response induced by ICI with a subsequently weaker selective pressure for transcriptional change.

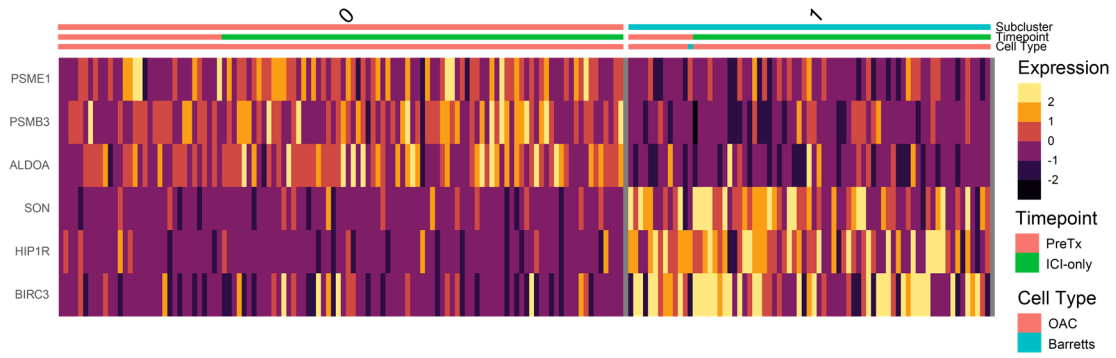


Figure 5.50: Heatmap of OAC-HGFI subcluster markers. Barrett's/OAC sub-compartment cells are sorted according to subcluster, timepoint and identified cell type (OAC or Barrett's). Scaled log-normalized expression values (from SCT workflow) are plotted for selected marker genes differentiating identified subclusters.

5.5.3.4 OAC-PAHE (NA Clinical Benefit)

OAC-PAHE was censored for PFS before clinical benefit could be assessed, but their BOR (irSD) and OS (11.9 months) were suggestive of a middling outcome in this cohort. Significant subclusters of cells that did not cluster with Barrett's (presumably malignant epithelium) were present for all three timepoints for this patient (Figure 5.51). The major PreTx phenotype was subcluster 1, which appeared to be closely related to a small ICI-only subcluster (7). In addition to 7, a number of other ICI-only subclusters were seen (0, 2, 3, 4, 6) that had little or no contribution from PreTx cells. Finally, a single PostTx subcluster was seen (5), which was separated from both PreTx and ICI-only subclusters.

This predominant PreTx subcluster (1) expressed *IGFBP2*, which largely disappeared in ICI-only subclusters, with the exception of the closely related subcluster. This small ICI-only subcluster was differentiated by increased inflammatory gene expression (*IFI27*, *CCL20*), although also showed increased expression of the CT antigen *CSAG1*. The major ICI-only subclusters (0, 2, 3, 4, 6) appeared more related, all expressing higher levels of *NUPR1*, *REG1A*, and *HLA-DRA* than clusters 1 and 7. Interestingly, the PostTx cluster shared some key characteristics with the predominant PreTx cluster which were largely absent during ICI-only, including expression of

KLK6, *CTSE*, and *CCK*. This is suggestive of a reversion towards the original PreTx tumour subtype at the end of treatment to some extent; however some unique genes did distinguish the PostTx subcluster, such as the Barrett's marker *REG4* and dermokine (*DMKN*), an inflammatory epidermal gene with a putative role in EMT[252]. This PostTx cluster also had high levels of the multi-drug resistance gene *ABCC2*, which was present to some extent at PreTx, but exaggerated in these cells. This appeared to correspond to the resurgence of the CT antigen PRAME at PostTx, which had been reduced in multiple ICI-only subclusters. Therefore, scRNA-seq support MDR-mediated resistance to the chemotherapy component of treatment, which appeared to arise in a subcluster related to the original PreTx cells despite their relative disappearance during the ICI-only phase.

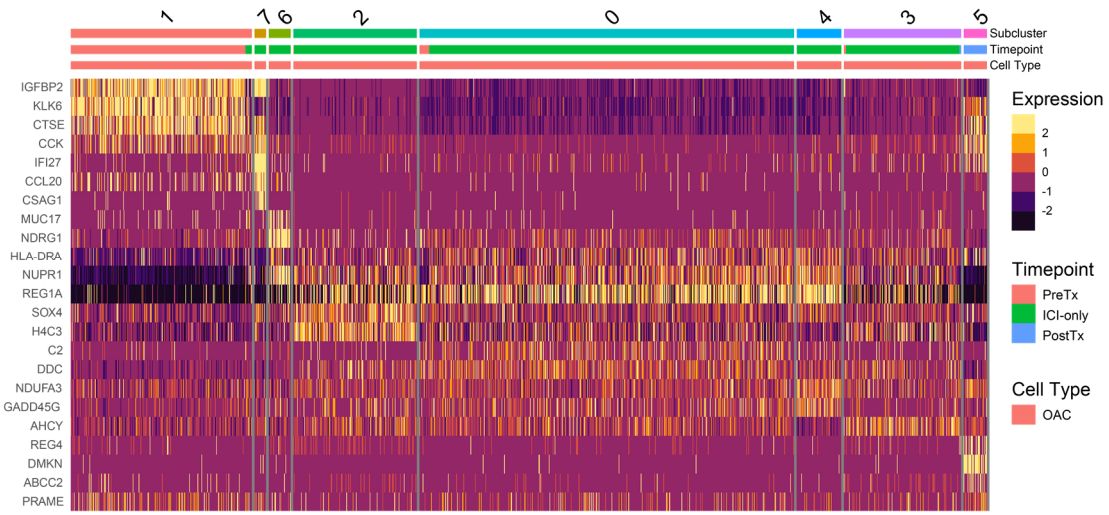


Figure 5.51: Heatmap of OAC-PAHE subcluster markers. Barrett's/OAC sub-compartment cells are sorted according to subcluster, timepoint and identified cell type (OAC or Barrett's). Scaled log-normalized expression values (from SCT workflow) are plotted for selected marker genes differentiating identified subclusters.

5.5.3.5 OAC-ACMO (Clinical Benefit)

Two major PreTx subclusters were identified (3 and 6), which appeared to evolve into related subclusters after ICI-only (1 and 2, respectively) (Figure 5.52). The transition from 3 to 1 was accompanied by increases in early goblet marker *SPINK4*, complement pathway (*CFD*), and interferon-induced genes (*IFITM1*), while subcluster 6 downregulated expression of *IGFBP2* while upregulating *PSCA* relative to 2. Three other subclusters appeared to completely unique to the ICI-only biopsy (0, 4, and 5). Related subclusters 0 and 4 expressed the *PRSS2* gene associated with poor outcome at ICI-only in bulk sequencing, and were differentiated by higher expression of *OLFM4* in 0. Subcluster 5 had some markers of EMT (*NNMT*), but also high upregulation of inflammatory response genes, including *IFITM1*, MHC class II, and T cell chemoattractants *CXCL10*). In summary, there was strong evolution in response to therapy, with some subclusters upregulating genes associated with resistance, and others upregulating genes that would presumably assist with T cell response at ICI-only. At PostTx, only a small discrete cluster of epithelial cells remained, which clustered with Barrett's oesophagus rather than other cancer cells from this patient.

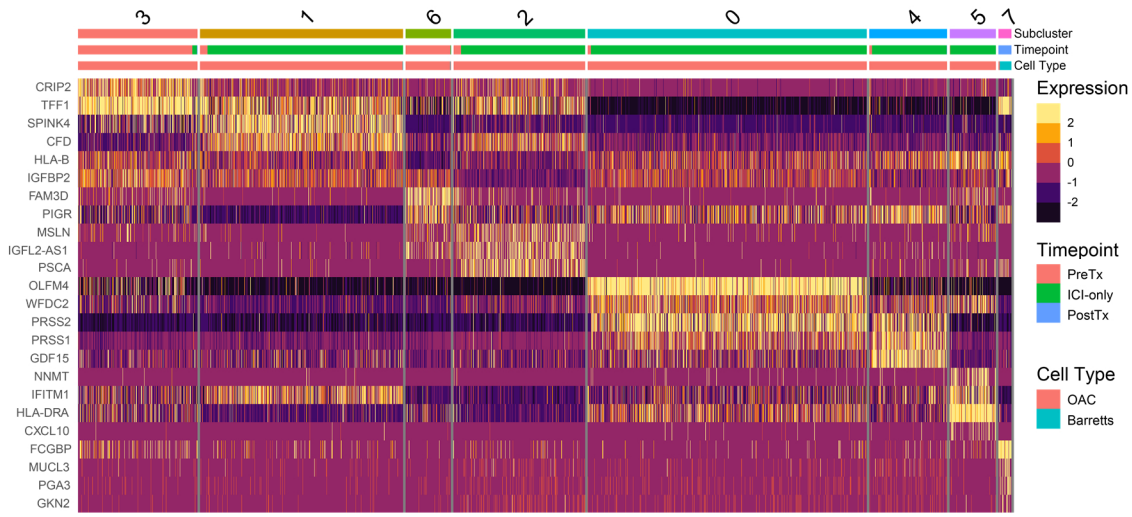


Figure 5.52: Heatmap of OAC-ACMO subcluster markers. Barrett's/OAC sub-compartment cells are sorted according to subcluster, timepoint and identified cell type (OAC or Barrett's). Scaled log-normalized expression values (from SCT workflow) are plotted for selected marker genes differentiating identified subclusters.

5.5.3.6 OAC-JCNP (Clinical Benefit)

This MSI-high patient had the strongest radiological (target lesion shrinkage) and transcriptional (INCITE reactivity) response to the ICI-window. Their tumour showed a number of large PreTx subclusters (0, 1, 2, and 3), and a small number of cells from ICI-only that clustered separately, close to a group of PostTx cells that were identified as Barrett's in the overall analysis (subcluster 4; Figure 5.53). The PreTx clusters had strong variation in a number of marker genes, indicative of high tumour heterogeneity prior to therapy. However, all PreTx subclusters expressed some common markers, including the CT-antigen PRAME, *NTS*, and *ANPEP*, which disappeared during ICI-only. The ICI-only cells shared expression of *SPINK1* with PreTx subcluster 0, but were differentiated from that cluster by increased expression of MHC class I and II antigen presentation genes and interferon-stimulated genes (*IFITM1*, *IFI27*). The ICI-only cells had also upregulated some markers that were shared with the Barrett's-like cells found at PostTx (*TFF1*, *PIGR*, *MSLN*). In summary, this patient showed decrease of all PreTx subclusters, which was replaced by a small subcluster with some relation to the SPINK1^{hi} subpopulation that had strongly increased antigen presentation and interferon-response signatures. The remaining PostTx cells were indistinguishable from Barrett's oesophagus in the overall analysis.

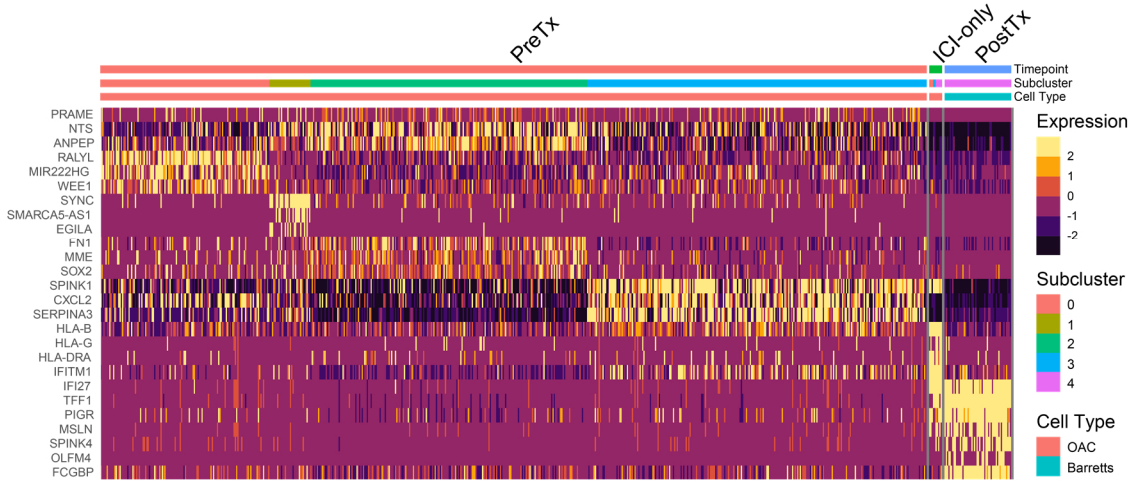


Figure 5.53: Heatmap of OAC-JCNP subcluster markers. Barrett's/OAC sub-compartment cells are sorted according to timepoint, subcluster, and identified cell type (OAC or Barrett's). Primary sort key is timepoint for this patient to assist with visualization of differences at ICI-only timepoint. Scaled log-normalized expression values (from SCT workflow) are plotted for selected marker genes differentiating identified subclusters.

5.5.3.7 OAC-LPPN (Clinical Benefit)

Three major PreTx clusters were detected in this CB patient (0, 2, and 3), all of which were fairly specific to PreTx, although 0 had a minor contribution from ICI-only (Figure 5.54). Contrastingly, subcluster 1 was almost entirely contributed by ICI-only cells, suggesting evolution or significant expansion of this subtype during ICI-only. The only epithelium in this sub-compartment present at PostTx (subcluster 4) was identified as Barrett's oesophagus. Of the PreTx clusters, subcluster 3 (*CLDN6*^{hi}, HLA^{lo}) appeared to be discrete, while 2 and 0 has shared overlap in a number of markers (*PSCA*, *PLAAT4*). The ICI-only specific subcluster (1) was characterized by high expression of MHC class I and II genes, interferon-induced genes (*IFITM1*), and epithelial inflammatory genes (*S100A9*). However, these inflammatory changes were accompanied by increased expression of immunosuppressive *IDO1*, a Fas decoy receptor (*TNFRSF6B*), and a subset of EMT-relevant genes (*NNMT*, although not *TGFB1* or clear mesenchymal genes), suggesting potential resistance mechanisms. However, these were absent at the PostTx timepoint, where the *FCGBP*^{hi}*REG4*^{hi} cells were identified as Barrett's oesophagus in the overall analysis.

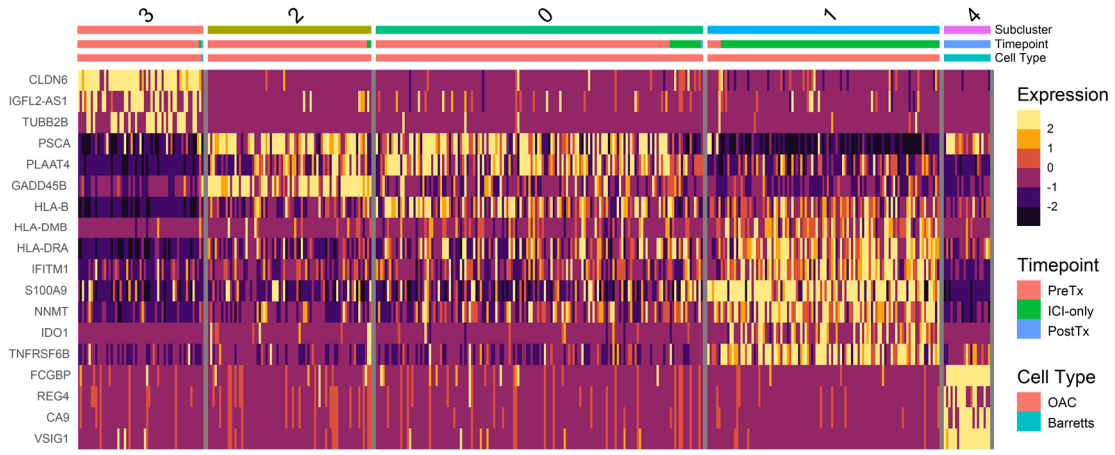


Figure 5.54: Heatmap of OAC-LPPN subcluster markers. Barrett's/OAC sub-compartment cells are sorted according to subcluster, timepoint and identified cell type (OAC or Barrett's). Scaled log-normalized expression values (from SCT workflow) are plotted for selected marker genes differentiating identified subclusters.

5.5.3.8 OAC-GDBD (Clinical Benefit)

This patient had poor yield of non-squamous epithelium in the LUD2015-005 cell atlas, with only 18 cells spread across the three treatment timepoints assessed. However, a separate set of samples from freshly-isolated biopsies were profiled for this patient during the optimization phase prior to creation of the cell atlas, using a different scRNA-seq methodology (see Methods). While these samples were not included in the final cell atlas for consistency due to the different technology used, PreTx and ICI-only biopsies using this methodology with greater non-squamous epithelial yield were available, and were therefore used as substitutes to assess tumour evolution within this patient.

Most tumour subclusters were shared across timepoints for this patient (Figure 5.55). Interestingly, one subcluster 3 expressing a CT antigen (*MAGEC2*) largely disappeared during ICI, although two clusters expressing *MAGEA3* did not appear to be selected against. Subcluster 4 was only found at the end of the ICI-only period. This subcluster expressed higher levels of interferon-induced genes (*IFITM1*; *IFI44L*), MHC class I & II genes (*HLA-B*; *HLA-DRA*), and high levels of *EGFR*, a feature that was shared with a subset of subcluster 1 (particularly the ICI-only cells, although some were present at PreTx).

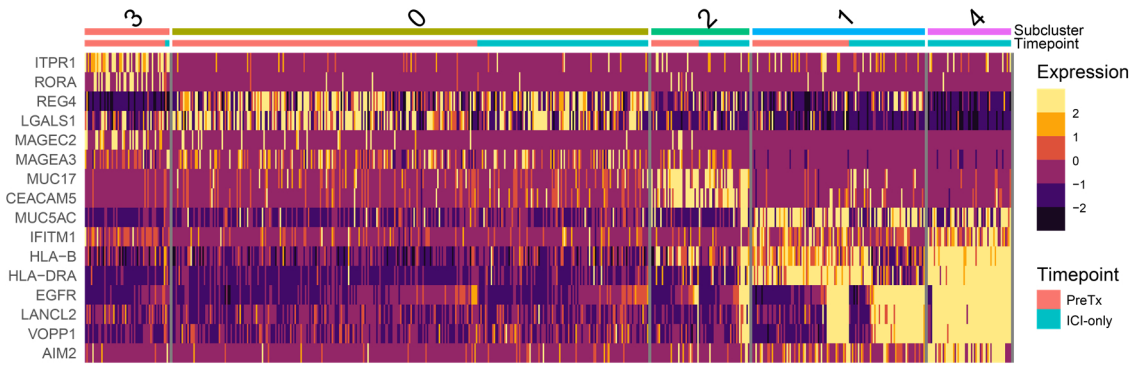


Figure 5.55: Heatmap of OAC-GDBD subcluster markers. Barrett's/OAC sub-compartment cells from a supplemental set of optimization biopsies for this patient (not included in LUD2015-005 scRNA-seq atlas) are sorted according to subcluster and timepoint, as well as by *EGFR* expression level to more clearly visualize expression patterns shown in cells overexpressing *EGFR*. Scaled log-normalized expression values (from SCT workflow) are plotted for selected marker genes differentiating identified subclusters.

Interestingly, the genes immediately downstream of *EGFR* (*LANCL2*, *VOPP1*) were also expressed very highly in subcluster 1, in exactly the same cells expressing high *EGFR* (Figure 5.55). This was perhaps suggestive of an amplification around the *EGFR* locus, with resultant

overexpression of the genes on the amplified segment (see section 5.5.3.11). This *EGFR*-overexpressing subcluster appeared following four weeks of ICI in this patient, and appeared to arise in cells that also have strongest upregulation of interferon-induced gene expression. No PostTx biopsies were available using this methodology to assess the evolution of this cluster during immunochemotherapy.

5.5.3.9 Operable OAC patients

Supplementary OAC biopsies included in this cohort were not used to derive predictive biomarkers linked with particular clinical outcomes, as they received different combination therapies to the inoperable OAC cohort and their long-term outcome information was not yet as mature as for the inoperable patients in this study. However, as all three patients received an αPD-L1-based treatment regimen, treatment-induced changes are still of interest particularly during the ICI-only window. These will be briefly summarized here.

OAC.Op-IKIJ had few assessable PreTx cells, making this analysis less informative for this patient. Of their ICI-only cells, two clusters were able to be differentiated, in part by differing expression of proteasome components (Figure 5.56). At PostTx, cells were generally identified as Barrett's from the overall analysis, including markers seen in PostTx Barrett's cells from inoperable OAC patients (*CTSE*, *MUC5AC*, *TFF1*, *MSLN*, *FCGBP*). MHC class II expression was also high in these PostTx cells.

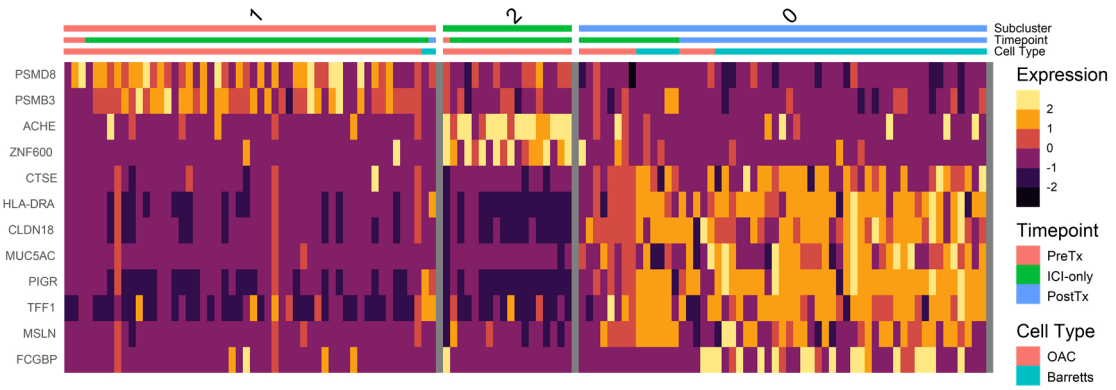


Figure 5.56: Heatmap of OAC.Op-IKIJ subcluster markers. Barrett's/OAC sub-compartment cells are sorted according to subcluster, timepoint and identified cell type (OAC or Barrett's). Scaled log-normalized expression values (from SCT workflow) are plotted for selected marker genes differentiating identified subclusters.

For OAC.Op-JJHD, clusters were identifiable from all three treatment timepoints (Figure 5.57). The predominant PreTx biopsy, expressing *IGFBP1* and high levels of MHC class I, was only found rarely following ICI-only. A number of subclusters appeared after the ICI-only window, including one with high immunoproteasome component expression (*PSMD10*), and another subset with interferon-induced gene expression (*IFITM1*). Interestingly, one of these *IFITM1*^{hi} ICI-only subclusters had very high expression of *EGFR* and surrounding genes, both upstream (*SEC61G*) and downstream *(LANCL2*, *VOPP1*). This was similar to the *EGFR*^{hi}*IFITM1*^{hi} subcluster that arose in patient OAC-GDBD during ICI-only, again suggestive of a possible *EGFR* amplification arising after four weeks of ICI. The predominant PostTx subcluster was identifiable as Barrett's, and a few Barrett's-like cells in this subcluster were also contributed by PreTx and ICI-only biopsies.

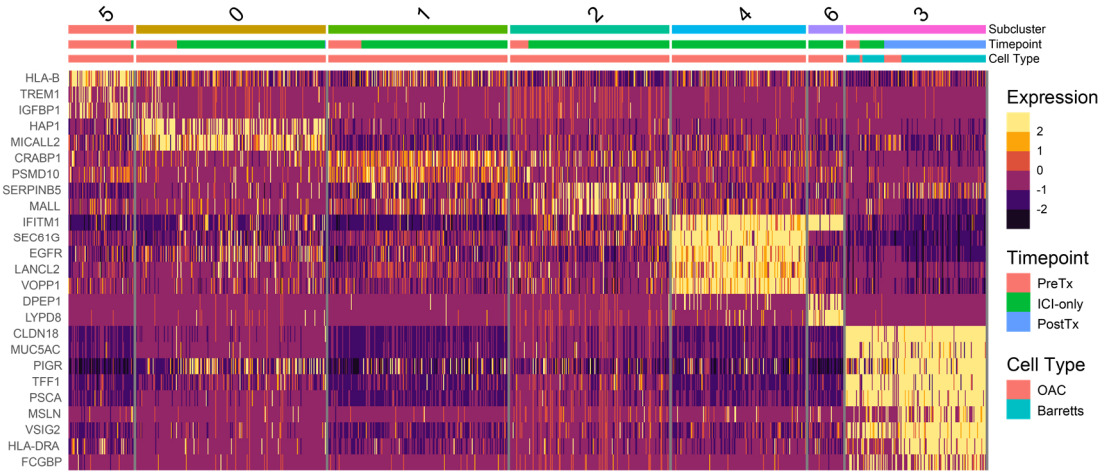


Figure 5.57: Heatmap of OAC.Op-JJHD subcluster markers. Barrett's/OAC sub-compartment cells are sorted according to subcluster, timepoint and identified cell type (OAC or Barrett's). Scaled log-normalized expression values (from SCT workflow) are plotted for selected marker genes differentiating identified subclusters.

Patient OAC.Op-ODHL had three major PreTx subclusters, one of which (1) was shared with cells from the PostTx timepoint (Figure 5.58). This subcluster had higher expressions of interferon-stimulated genes, including *IFITM1*. Interestingly, it was the ICI-only biopsy in this patient, rather than the PostTx biopsy, that clustered with Barrett's oesophagus in the full analysis; however, the close clustering of operable OAC with Barrett's makes this distinction more uncertain than for inoperable OAC. The ICI-only cluster did express a number of genes

seen in PostTx Barrett's in other patients (*CTSE*, *TFF1*, *MUC5AC*, *FCGBP*), suggestive of a related phenotype.

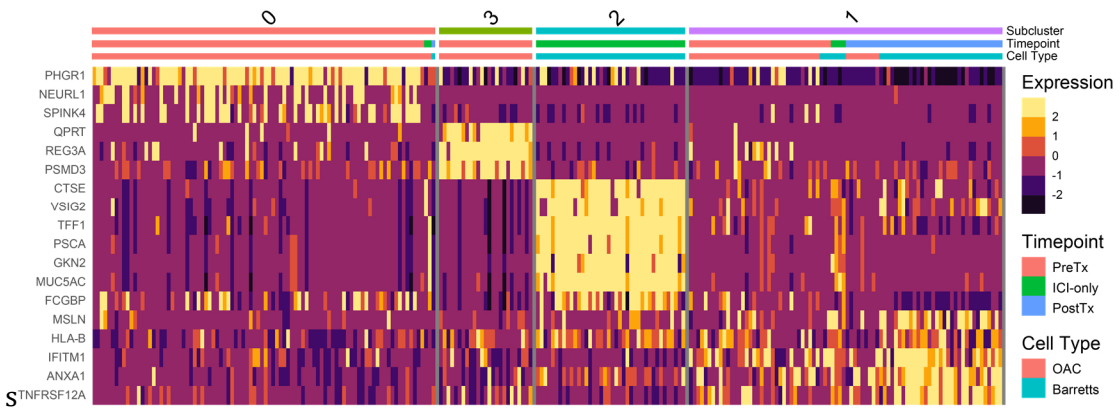


Figure 5.58: Heatmap of OAC.Op-ODHL subcluster markers. Barrett's/OAC sub-compartment cells are sorted according to subcluster, timepoint and identified cell type (OAC or Barrett's). Scaled log-normalized expression values (from SCT workflow) are plotted for selected marker genes differentiating identified subclusters.

5.5.3.10: Summary of ICI-induced transcriptional evolution

In the previous sections, a number of diverse transcriptional changes were seen during the ICI-only period, many of which were patient-specific. However, a number of common changes to tumour epithelium were noted across multiple patients, which are summarized here. All three patients with NCB showed decreased expression of one or more antigen presentation machinery components, such as MHC-I genes (*HLA-B*) and proteasome components (*PSMB3*, *PSMD3*, *PSME1*). While none of the proteasome genes profiled were components of the *IFNG*-induced immunoproteasome associated with specialized immune functions, increased expression of both *PSMB3* and *PSME1* have been linked with increased MHC-I antigen presentation[253,254], suggesting their downregulation could also impair this pathway. CT antigens were also seen to increase in prevalence following ICI-only in NCB patients, as well as in a small subcluster of cells from OAC-PAHE (not assessable for CB). In contrast, patients with CB all upregulated inflammatory response genes, particularly MHC-II expression (*HLA-DRA*) and ISGs (*IFITM1*) which have been previously linked with tumour epithelium in inflammatory environments[255–257]. Finally, three patients overexpressed *EGFR* in at least one subcluster enriched for ICI-only cells, and in two patients, adjacent genes on the locus immediately surrounding *EGFR* on chromosome 7 were overexpressed to a similar extent in the same cells.

However, this phenomenon did not show a clear link with outcome. A summary of these shared transcriptional responses to ICI are shown in Table 5.7.

Outcome	Patient	Proteasome/ HLA Underexp.	CT Antigen Overexp.	Inflammatory (MHC class II, ISG Overexp.)	EGFR Overexp.	EGFR locus Overexp.
NCB	OAC-HDHI	✓	✓		✓	
	OAC-HENB	✓	✓			
	OAC-HGFI	✓				
NA	OAC-PAHE		✓	✓		
CB	OAC-ACMO			✓		
	OAC-JCNP			✓		
	OAC-LPPN			✓		
	OAC-GDBD			✓	✓	✓
NA	OAC.Op-JJHD	✓		✓	✓	✓

Table 5.7: Summary of transcriptional evolution induced in OAC cells after four weeks of ICI alone. The patient-specific changes highlighted in 5.5.3.1-9 were manually assessed for common patterns of transcriptional response in tumour cells during the ICI-only window. Overexpression (Overexp.) and underexpression (Underexp.) refer to ICI-only cells in comparison with PreTx cells for that patient. EGFR locus refers to genes immediately adjacent to *EGFR* on chromosome 7.

5.5.3.11: Genomic amplification may underlie EGFR overexpression in response to ICI in multiple patients

In OAC-GDBD and OAC.Op-JJHD, both *EGFR* and adjacent genes were strongly overexpressed in a pattern that was specific to or enriched in cells at the ICI-only timepoint. This shared pattern raised the question of whether the pressure of ICI treatment resulted in the amplification of the segment of chromosome 7 containing *EGFR* or selected for cells containing this alteration.

To see whether genomic alterations might underlie the pattern of overexpression seen in these patients, the copy number landscape of the Barrett's/OAC sub-compartment was estimated using the inferCNV method employed in section 5.3.3. Indeed, this tool predicted a gain of three or more copies (the maximum predictable CNV gain using this method) at the locus surrounding *EGFR* for both OAC-GDBD and OAC.Op-JJHD; however, no amplification involving *EGFR* was

detected for OAC-HDHI, where *EGFR* was overexpressed at ICI-only but not neighbouring genes. For OAC-GDBD, the *EGFR* amplification was specific to the cells showing *EGFR* overexpression (Figure 5.59A,C). For OAC.Op-JJHD, a maximally predictable *EGFR* gain (3+ copies) was already detected in nearly all PreTx and ICI-only cells, but not PostTx Barrett's (Figure 5.59B,D). As a result, it could not be determined whether there was further amplification of the *EGFR* locus at ICI-only over the baseline value that drove the *EGFR* overexpression seen in subcluster 4 from this patient.

However, chromosomal rearrangements involving the *EGFR* locus on chromosomal 7 were detected in scRNA-seq data from OAC.Op-JJHD at ICI-only that were not present at PreTx, which were identifiable through fusion transcripts. This included a possible intrachromosomal inversion involving *EGFR*, as well as an apparent translocation just downstream of *EGFR* that involved a segment of chromosome 13 which was amplified specifically in subcluster 4 (Figure 5.59, Figure 5.60). Therefore, while a further copy number amplification of chromosome 7 during ICI-only cannot be confirmed or ruled out, the *EGFR* overexpression in subcluster 4 could also be caused by chromosomal rearrangements specific to the ICI-only timepoint that cause transcriptional activation of a pre-existing *EGFR* amplification. In summary, two patients in this cohort developed *EGFR* overexpression specific to an ICI-treated subcluster of cells, which may involve underlying genomic changes to this locus that are induced or selected for during ICI-only.

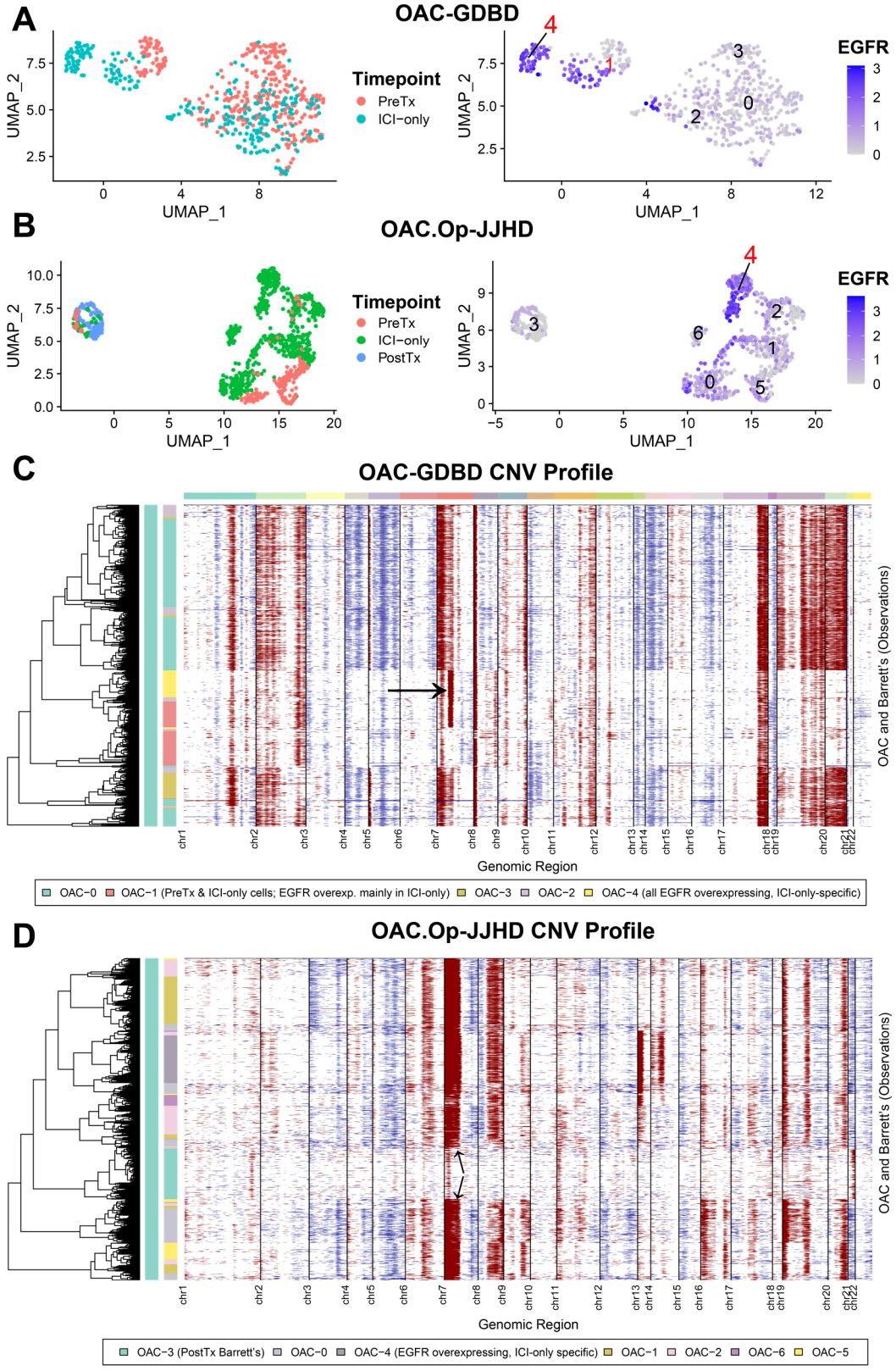


Figure 5.59: Amplifications as a possible mechanism for ICI-only subclonal overexpression of EGFR. A) and B): UMAP of non-squamous epithelium for the identified patient coloured by treatment timepoint (left) and log-normalized *EGFR* expression (right). Identified clusters are labelled, with red clusters denoting *EGFR* overexpressing clusters. C) and D): CNV levels, estimated on a scale of 0 (loss of both copies, deep blue) to >2 (3+ copy number gain, deep red) using normal cells (not shown) as reference, are shown for the indicated patients. The region of 3+ copy number gain on chromosome 7 including EGFR is indicated with an arrow.

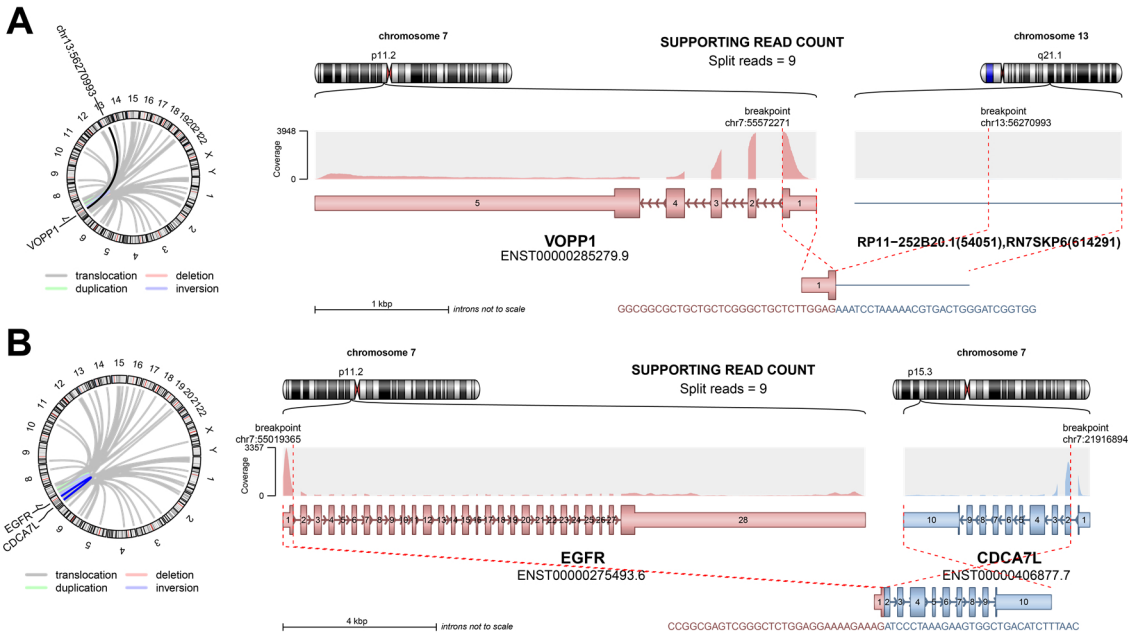


Figure 5.60: Genomic rearrangements involving the *EGFR* locus in OAC.Op-JJHD tumour cells at ICI-only. Gene fusion predictions returned by Arriba. For each subplot, the left facet displays a map summarizing the predicted rearrangement, while the right facet shows the details of the fusion sequence, predicted fusion protein, and read coverage. 'Split reads' refers to the number of reads spanning the junction of the indicated fusion transcript. ICI-only-specific rearrangements involving A) *VOPP1* and B) *EGFR* are shown, which both lie in p11.2 on chromosome 7. Both rearrangements are predicted with 'medium' confidence by Arriba.

5.5.3.12 Implications of PostTx Barrett's and PostTx OAC phenotypes at the primary lesion site

Following the course of immunochemotherapy administered in this study, many patients had residual non-squamous epithelium cells detectable in PostTx biopsies. For most patients, these cells showed a phenotype that was more similar to BO than malignant cells from PreTx and ICI-only timepoints for that patient. This could represent a successful clearance of most tumour cells by therapy, leaving behind true BO cells from the pre-malignant lesion which are not as effectively cleared by this treatment. Alternatively, these non-squamous cells could be residual malignant cells, which adopt a more Barrett's-like phenotype under the selective pressure of therapy. As some BO and OAC epithelial cells had very similar transcriptional profiles, the question of whether these cells were truly BO or residual malignant cells could not be confidently answered by the transcriptional profile alone.

Therefore, CNV profiles were examined as in 5.5.3.11, to determine the extent of shared CNV events between PostTx Barrett's-like cells and malignant cells from other timepoints that might imply some evolutionary connection or ancestral origin. For patients with PostTx cells

clustering with BO, these PostTx cells showed very little copy number aberrations compared with malignant epithelial clusters contributed from other timepoints, with overall levels of copy number aberration more similar to the non-malignant cells. An example of the PostTx BO CNV phenotype is shown above in subcluster 3 of OAC.Op-JJHD, which had less overall CNV and did not share the otherwise clonal amplifications in chromosomes 7 and 8 (Figure 5.59D). Other patients with PostTx BO showed similar CNV patterns (data not shown).

However, for OAC-PAHE, who was the only inoperable OAC patient with a PostTx cluster identified as an OAC cell type rather than BO, the PostTx subcluster (5) had similar numbers of CNV events to other OAC subclusters, consistent with retained malignant identity in these cells (Figure 5.61). Additionally, this PostTx subcluster showed many shared copy number events with the predominant PreTx sub-cluster (1), providing further evidence for their ancestral relationship, in combination with the shared marker expression discussed in 5.5.3.4. In summary, PostTx cells identified as OAC in this dataset were similar to malignant clusters from other timepoints in their CNV profiles, while the Barrett's-like cells present after therapy had reduced CNV events, and did not appear to be linked with patient-specific PreTx and ICI-only CNV subclusters. This suggests that the PostTx BO phenotype seen in most patients following immunochemotherapy is representative of true BO, rather than Barrett's-like OAC cells that retain a malignant character.

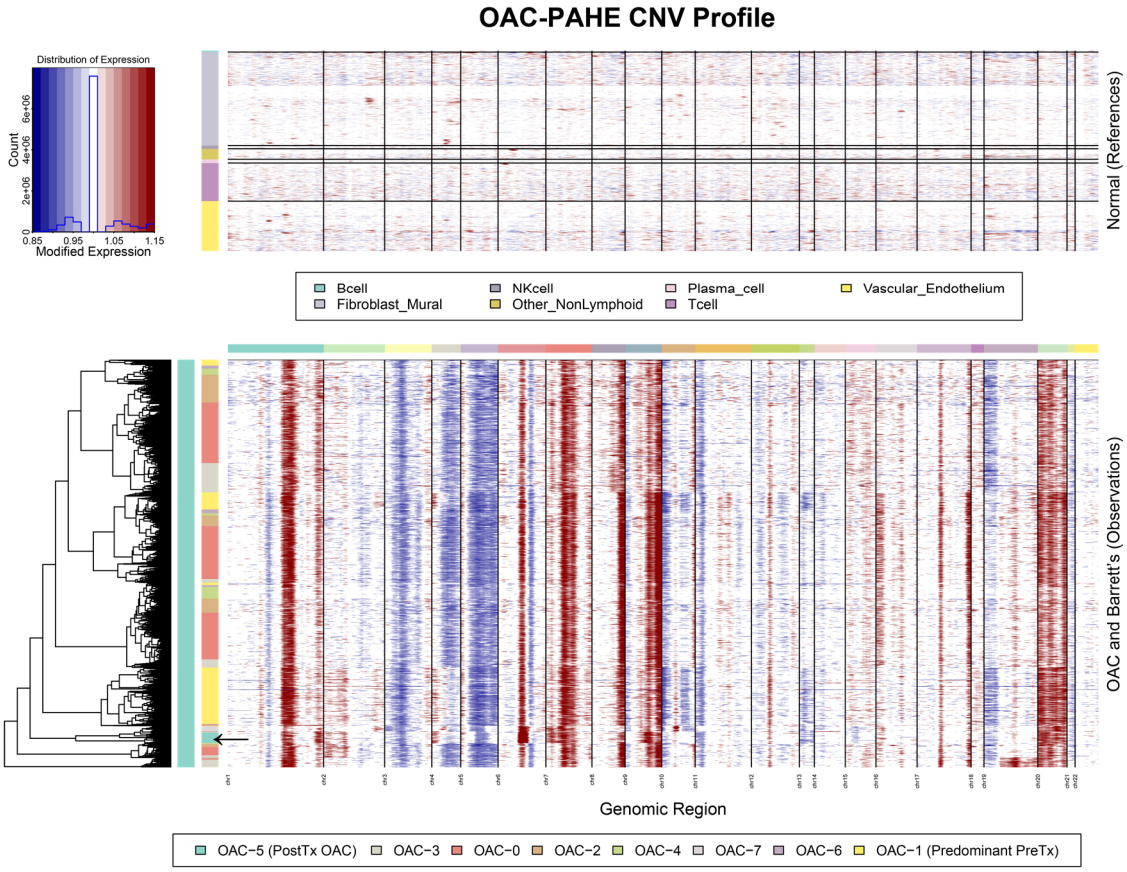


Figure 5.61: scRNA-seq copy number profile of a patient maintaining a PostTx OAC phenotype. Estimated copy-number variation (CNV) profiles of OAC and BO cells (bottom) from patient OAC-PAHE. CNV levels are estimated on a scale of 0 (loss of both copies, deep blue) to >2 (3+ copy number gain, deep red) using normal cells (top) as reference. The PostTx OAC subcluster (3) is highlighted with an arrow.

Of the inoperable OAC patients, the lone patient with a PostTx subcluster identified as OAC following immunochemotherapy had inferior OS (11.9 months) compared to the three patients exhibiting the PostTx BO phenotype (ongoing OS of 15-27 months; all without reported progression at the 12-month cut-off). Distinguishing between these phenotypes may therefore be useful to help predict late resistance and long-term outcomes among patients that remain alive and on-treatment following a full course of immunochemotherapy. Consequently, genes were examined to identify expression markers that could differentiate these two PostTx phenotypes, which would be easier to use in-clinic than broad genomic methods such as copy number profiling. Including operable patients to increase power, the marker *FCGBP* was found to be highly expressed in PostTx BO cells from 5 of 5 patients with the phenotype, but was not at all expressed in PostTx OAC cells from two patients (Figure 5.62A). The expression of *FCGBP* in PostTx BO is similar to that seen in BO from untreated patients without cancer (Figure 5.62B),

and has highest expression in this cell type, although it is also expressed in some myeloid cell types (data not shown). In summary, the expression of *FCGBP* in non-squamous epithelium strongly differentiated non-malignant BO from residual malignant cells at the PostTx timepoint in this study. This marker may therefore be able to inform assessments of immunochemotherapy efficacy using PostTx sampling.

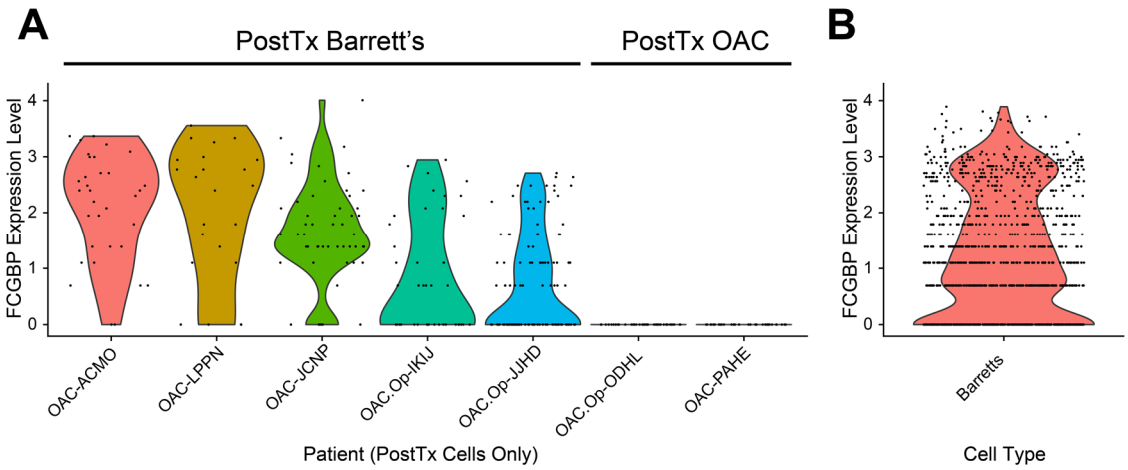


Figure 5.62: *FCGBP* differentiates malignant and non-malignant non-squamous epithelium following immunochemotherapy. A) Log-normalized expression of *FCGBP* in PostTx cells in the BO/OAC sub-compartment, grouped by patient. Five samples with a PostTx BO phenotype, and two with a PostTx OAC phenotype, are indicated. B) As a reference, the log-normalized expression of *FCGBP* in BO cells from BO biopsies taken from patients without cancer is shown. The same y-axis scale is used in A and B.

5.6 Summary and brief discussion

In this chapter, I outline the LUD2015-005 oesophageal pathology atlas, a scRNA-seq dataset that can yield further insight into the transcriptional profiles of cell types present in a number of normal and pathological upper GI tissue types. One major contribution of this chapter is the validation of a new method to preserve small endoscopic biopsies from the gastrointestinal tract for scRNA-seq using controlled-rate freezing in serum with DMSO. A similar method has recently been applied successfully by our collaborators in head and neck squamous cell carcinoma[226], suggesting it could have utility in multiple settings. The slow-freezing methodology will already be familiar to most laboratories working with cell culture, and would be easily implementable without needing additional specialized buffers or materials. Cells derived from slow-frozen material that survived QC filtering had similar quality transcriptomes to those of cells isolated from fresh material, and even expressed lower signatures of

dissociation-related stress. The lower stress signatures could be down to slight differences in the enzyme mix used between fresh and slow-frozen dissociation methods in this study, but it is also possible hypothermia associated with cryopreservation induced some protection against stress responses induced by subsequent dissociation. Indeed, dissociation at hypothermic temperatures has previously been shown to minimize dissociation-induced stress responses in scRNA-seq data[258]. In any case, slow-freezing provided a usable method to preserve whole biopsies for later isolation of viable cells even after prolonged storage at -80°C, which were suitable for use with scRNA-seq and potentially other live-cell assays. Additionally, data from fresh and slow-frozen biopsies were able to be combined using standard scRNA-seq dataset integration methods, including MNN-based approaches.

The downside of the slow-frozen method in this application was the reduction in cell viability compared with freshly-isolated biopsies, which appeared to result in lower yields of cell barcodes passing QC thresholds. While low-quality barcodes could be filtered out and did not substantially affect the final analysis, the high amount of low-quality barcodes reduces the amount of usable cells per sample and increases the effective cost per cell. Therefore, those implementing the slow-frozen workflow in the future should focus on improving post-dissociation viability, which is likely to ameliorate this disadvantage. Use of specialized commercial cryopreservation media rather than serum, as well as storage in liquid nitrogen rather than -80°C, could help maintain higher viability during long-term storage. An optimized dead cell depletion step, which was not included in this slow-frozen workflow, would likely significantly improve viability of isolated cell suspensions. While magnetic bead-based dead cell depletion anecdotally did not improve viability in our hands, optimization of magnetic bead, cell sorting, or density-based[259] dead cell depletion methods would significantly improve the suitability of this method for general use.

46 cell types were defined in this atlas, many of which could be further resolved into subtypes. In addition to malignant epithelium, a number of non-malignant cell types showed tumour-specific subtypes, including endothelium, fibroblast, pericyte, and macrophage subsets,

highlighting significant changes that develop throughout the tumour microenvironment in oesophageal cancers.

Interestingly, a number of cell types were identified that are likely associated with tertiary lymphoid structures (TLS). Antigen-specific adaptive immune cells are generated in primary (e.g. thymus, bone marrow) and secondary (e.g. lymph nodes, Peyer's patches) lymphoid structures, but also in TLS, which are smaller ectopic lymphoid structures that can develop during inflammation induced by a number of disease states, including cancer[260]. It is generally thought that TLS are sites where key antigen-specific anti-tumoural responses can be generated. Their presence has been associated with better outcome on ICI, as well as better prognosis even for patients not treated with ICI agents[261–263].

Because of this, the function of TLS in tumours is a well-studied area, and a summary of what is known about the functionality of these structures follows[260,261,264,265]. TLS resemble secondary lymphoid organs to some extent, and are generally organized into B and T cell zones. The mature B cell zone is further subdivided into "light" and "dark" zones, although this light and dark polarity is not always seen in some cancers[266]. Upon appropriate BCR-ligand recognition and secondary signalling, B cells enter the dark zone, where they proliferate and undergo somatic hypermutation (SHM). These B cells with modified BCRs then enter the light zone, where follicular dendritic cells (FDCs) and T_{FH} cells select cells with the appropriate affinity and provide them with survival and co-stimulatory signals. Following this process of affinity maturation, mature antigen-specific memory B cells or plasma cells are generated. In the T cell zone of TLS, mature dendritic cells expressing DC-LAMP (*LAMP3*) play a key role in activating T cells towards more mature states through a variety of co-stimulatory signals. TLS are associated with a specific form of vascular endothelium known as high endothelial venules (HEV), which are vascular endothelial vessels specialized for immune cell extravasation which express high levels of immune cell adhesion proteins. HEVs are thought to be the source of naïve B cells and naïve or CM-like T cells for maturation in the TLS. Lymphatic endothelial cells are also associated with TLS, and may facilitate trafficking of matured cells to the draining lymph node and systemic immune system.

Even in the absence of histological TLS detection, a number of the cell subtypes profiled in the LUD2015-005 atlas have marker profiles thought to be highly specific to TLS, to the extent that the likely TLS structure in OAC can be reconstructed (Figure 5.62). For example, the proliferative GC cell subset expressing *AICDA*, *BCL6,* and *KI67* almost certainly represents the proliferative B cell subset undergoing SHM, likely in the dark zone of GCs, sharing many markers with previous literature[266,267]. The other major GC subset expresses genes more enriched in the light zone (*CD83*, *CD69*, *GPR183, EGR1*[268]), but also the dark zone marker *CXCR4*. As this subset is largely contributed by a single patient, this could be a more immature TLS without light and dark polarity, as seen in melanoma[266]. Alternatively, it could be that non-proliferative GC B cells in light and dark zones are closely related subtypes[269], which are not easily separated by general clustering of shallow scRNA-seq. Besides these GC B cells, a *CXCL13*-producing T_{FH} subtype previously identified in breast cancer to be linked with TLS functionality[231] was also identified in this OAC dataset. In the T cell zone, mature DC-LAMP/*LAMP3*^{hi} DCs are the key DC population responsible for T cell activation, and some have even stated that these DCs are specific to TLS[270]. The key endothelial cell types are also present, including lymphatic endothelium, and an inflammatory *ICAM1*$^{+}$ post-capillary venule subset, which likely represents HEV[260]. Therefore, while the follicular DC cell type associated with TLS could not be identified, likely due to its relative rarity, most of the key cell types present in TLS were readily identifiable.

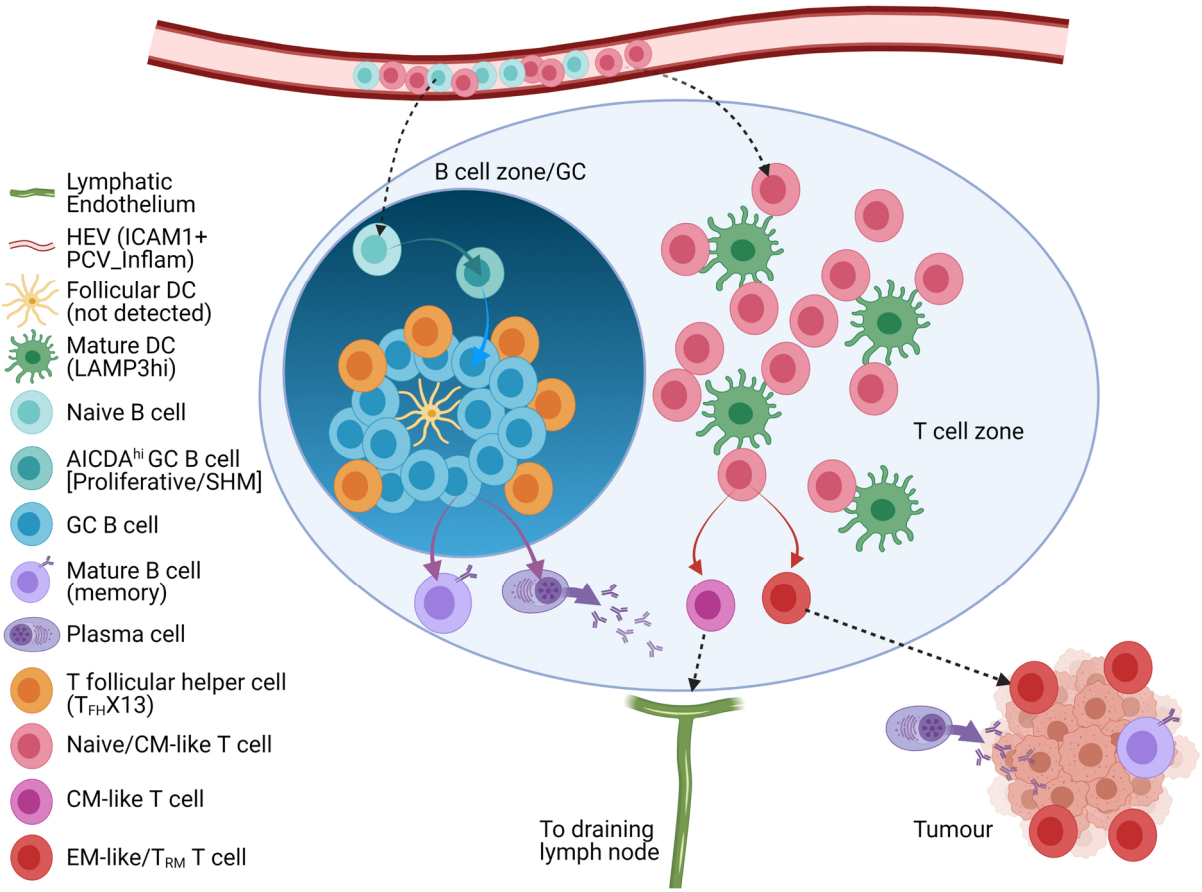


Figure 5.62: Proposed structure of tertiary lymphoid structures (TLS) in OAC. All cell types are represented in scRNA-seq data with the exception of follicular DCs, but future histological work would be needed to confirm their presence in TLS. HEV: High endothelial venule; PCV_Inflam: Inflamed post-capillary venule; SHM: somatic hypermutation.

Interestingly, many of these cell types, particularly CXCR4/CD83^{+} GC B cells, HEV and lymphatic endothelium, came from the PostTx timepoint. This could suggest that immunochemotherapy plays an important role in triggering the maturation of TLS. However, these PostTx cells were mainly contributed from the three operable OAC patients, which unfortunately means that the different stage of cancer, deeper biopsy method unique to these samples (core biopsies), and therapies used in addition to ICI (see Table 5.1) could also play a confounding role. In any case, it seems likely that TLS are present in OAC, and could be of particular interest to study the development of anti-tumoural immunity at the site of the lesion. Future studies would be needed to histologically confirm the presence of these structures, determine how they are modulated by immunochemotherapy, and examine how they are formed, particularly with regard to whether potential LTIs such as the ILC-like NK cell subtype may play an initiating role.

Core biopsies from surgical resections may increase the odds of TLS detection in this setting, and should be considered for studies investigating TLS in OAC when available.

This dataset also sheds further light on the heterogeneity of metaplastic Barrett's epithelial cells. In this thesis, a *REG4*$^{+}$ cluster was identified as the major cell type present in BO, which includes the early-goblet and Barrett's-type populations identified by Owen et al[32]. *REG4* has previously been shown to be expressed more highly in BO than normal[271,272], and as the gene is a marker of secretory cells deep within colonic crypts. Along with the MUC2^{hi} differentiated goblet cell cluster, these findings suggest a shift towards an intestinal transcriptional programme. These clusters are likely consistent with the markers of intestinal metaplasia often seen in BO, while the foveolar-like BO cluster could represent the gastric-type metaplasia seen in some patients[24]; however, the exact implication of foveolar-like cells in Barrett's oesophagus is difficult to work out due to the low yield of normal foveolar cells from gastric biopsies in this dataset.

Finally, in this work the presence of a gland-like BO subtype was identified, which shares some markers (*SCGB2A1*, *BPIFB1*, *FOLR1*) otherwise unique to submucosal glands[197]. In one patient, this subtype was shown to have less copy number aberrations than other BO subtypes from the same individual. As increasing levels of genome-wide copy number variation have been reported to mark Barrett's with increased odds of progressing to cancer, this low-CNV cluster may represent BO closer to the cell of origin, with less CNV changes associated with metaplasia. It is possible this cell type could also represent a non-metaplastic submucosal gland, but the expression of *MUC6* appears to be aberrant. Therefore, these gland-like cells likely represent another cell type with high transcriptional similarity to submucosal glands, in addition to the Barrett's-type previously identified by our group[32]. This provides further support to the hypothesis that BO develops from submucosal gland cells. Integrated analysis of the present work, that of Owen et al., and the HCA squamous oesophagus dataset would likely shed further light on the origins and heterogeneity of BO.

This study highlights the transcriptional heterogeneity present even with the same OAC biopsy, and identified some common patterns in how OAC epithelium evolves transcriptionally in

response to ICI. In this study, on-treatment downregulation of HLA-I and proteasome components is identified in NCB patients. Impaired HLA-I antigen presentation has been associated with acquired resistance to ICI in NSCLC, which included alterations to both the HLA-I complex (*B2M*) and proteasome components[273]. While it is unclear whether the changes in our study are mediated through genetic or transcriptional mechanisms, this provides further support to loss of antigen presentation as a common mechanism for ICI resistance in some patients. Notably, the patients with downregulation of antigen presentation components in the present study also had some of the lowest INCITE upregulation on-treatment, suggesting that this evasion mechanism is associated with muted induction of inflammation and T cell infiltration.

In patients with positive outcome, inflammatory response from tumour cells were noted, particularly HLA-II expression (*HLA-DRA*) and interferon-stimulated genes (*IFITM1*). HLA-II is not normally expressed on epithelium, but can be induced under inflammatory conditions including exposure to IFNG[274]. In OAC treated with standard care, this is associated with poor prognosis[256], but in melanoma the finding is associated with good outcome on ICI[255], agreeing with the present results from the ICI-only window in OAC. The upregulation of these inflammatory response genes are likely related to the upregulation of IFNG expression and inflammatory stimuli during the ICI-only window. Induction of these genes tended to be subclonal, suggesting that some OAC cells experience more inflammatory stimuli than others. It is possible these inflamed subclusters are on the outer edge of the tumour or closer to responding immune structures, and therefore may be under more inflammatory pressure than other subclusters.

This raises the question of whether these subclusters are under increased selective pressure to evolve in ways that resist ICI-related cytotoxicity. Indeed, *EGFR* overexpression possibly linked with genomic alterations was detected in two patients at the ICI-only clusters, and always in *IFITM1*^{hi} subclusters. In NSCLC, EGFR-mutated tumours have been shown to be particularly resistant to ICI[275]. It is possible EGFR amplification and subsequent overexpression, similarly to activating EGFR mutations, initiates downstream signalling pathways which helps to resist the

inflammatory environment induced during ICI. While this EGFR finding was not associated with poor outcome in this study, it could provide a mechanism for late resistance in tumours that initially generate a strong inflammatory response. If EGFR amplifications or other overexpression mechanisms can be demonstrated in progressing tumours, it could have clinical implications for potential second-line treatment options in this setting, as tumours that develop wild-type EGFR amplifications as a resistance mechanism may be sensitive to certain tyrosine kinase inhibition (TKI) agents[276].

The disappearance of tumour cells upon pathological assessment (pathological response) is often used particularly in the setting of neoadjuvant chemotherapy, where better pathological responses are linked with improved long-term patient outcomes[277]. While the sample size is limiting in the present study, one patient with residual tumour detectable by scRNA-seq at the PostTx had worse outcomes than other patients that survived to the PostTx timepoint, suggesting a similar concept could have utility in this setting. However, assessing pathological response following immunochemotherapy may be difficult, as most patients without detectable OAC by scRNA-seq had BO cells present. These BO cells may be difficult to distinguish from residual cancer, particularly if they have some degree of dysplasia. In this study, *FCGBP* is highlighted as a marker of general BO cells, including BO found at PostTx in OAC patients, which could help pathologists differentiate between BO and residual OAC in assessing pathological response. Indeed, antibodies targeting the GCTM-5 epitope on FCGBP are already in use for pathological study[278], and could easily be added in to a pathological response workflow in the setting of OAC.

In this chapter, deconvolution emerged as a powerful tool for leveraging the cell type profiles defined in the LUD2015-005 scRNA-seq atlas to gain additional insight from bulk transcriptomes of many oesophageal tissue types. First, deconvolution was used to analyse a previously published dataset from Maag et al.[247], which showed that normal oesophagus, BO, and OAC could be separated by both their epithelial and microenvironment composition. Deconvolution of the bulk transcriptomes from the LUD2015-005 study revealed that the innate immune signature associated with improved outcome was particularly specific to monocytes,

although DEGs associated with CB could be found in monocytes, differentiating M1-like macrophages, and neutrophils. The monocyte signature at PreTx showed a strong immunochemotherapy-specific association with improved OS, and appeared to have additive power when used with other useful predictors in this cohort, such as coding TMB. The combination of this previously reported and novel biomarker provides an exciting method to predict survival outcomes on immunochemotherapy in this setting.

The question of why monocytes predict better outcome in this study is a key consideration. A 50-gene signature derived from a subtype of monocytes has previously been linked with better TCGA outcome in OSCC patients[279]. However, in this thesis, monocyte infiltration did not correlate with TCGA outcomes, only showing a link with outcomes for patients receiving immunochemotherapy. This suggests key differences between these two reports, which could include how the monocyte populations were defined, methodological differences between gene signature workflows and the present deconvolution approach, or differences in the underlying disease type, and therefore the question of why monocytes associate specifically with immunochemotherapy outcome in OAC is still left open. Monocytes themselves may participate in direct tumour killing, both through phagocytosis and secretion of inflammatory and tumouricidal molecules[280]. However, in this dataset, M1-macrophages and neutrophils appeared to secrete more anti-tumour effector molecules than did monocytes, rendering this direct explanation unlikely. One alternative explanation is that monocytes could differentiate into more inflammatory macrophage states under the stimulus of ICI, which may not be as feasible for myeloid cells that are already terminally differentiated. Indeed, immunochemotherapy did appear to skew the myeloid compartment towards M1 macrophages and away from the major TAM population having a more intermediate phenotype, and patients with high monocytes at PreTx had a non-significant trend towards increased levels of M1 macrophages at ICI-only.

Alternatively, the PreTx monocyte proportion also had an interesting association with increased levels of key cell types associated with TLS at the ICI-only timepoint, namely mature LAMP3^{hi} DCs and lymphatic endothelium. Monocyte-derived cells have the ability to differentiate directly into LAMP3^{hi} DCs, which has been shown under inflammatory stimuli *in vitro*[281]. However,

monocytes have also been proposed to serve a role as lymphoid tissue inducers (LTIs) in the context of initiating TLS, independent of differentiation into DCs. The development of secondary lymphoid organs are known to be initiated by the group 3 ILC-like LTIs, such as the ones identified in the LUD2015-005 dataset. However, these cell types have not yet been shown to trigger formation of TLS, and in fact it has been proposed that inflammatory myeloid cells, including monocytes, may take over this initiating LTI role for TLS[282]. Therefore, these monocyte-like cells may play a key role in triggering formation of TLS and antigen-specific anti-tumour responses in response to ICI-only, even if they do not differentiate directly into the mature DC type that is involved in TLS-mediated T cell activation. This monocyte-TLS link was not associated with INCITE reactivity during the ICI-only window, suggesting a link with outcome even in patients who respond to immunochemotherapy rather than ICI alone. This agrees with the notion discussed above that the presence of TLS is generally a positive prognostic marker, even outside the context of ICI.

Deconvolution also highlighted the role of vascular SMCs in marking ICI resistance. The role of the gene *RERGL*, a relatively uncharacterized gene that was the only DEG predictive of NCB at both PreTx and ICI-only, was clarified as well, suggesting that its link with predicting outcome is likely a function of its status as a highly specific marker of this vascular SMC cell type. Vascular SMCs are not frequently discussed in the context of ICI, and the presence of this cell type could simply mark a tumour with more invasion into the underlying stroma and vasculature, which may have a poor prognosis on any treatment. However, no other vessel-associated cell types have a link with outcome, suggesting an alternative mechanism could be at play. A hint may come from atherosclerosis, in which vascular SMCs are known to mediate the concept of medial immunoprivilege, whereby immune cells are excluded from the middle layer of the arterial wall (tunica media) despite the overall inflammatory nature of the condition. This exclusion of immune cells by vascular SMCs is thought to occur due to a combination of mechanical (i.e. vascular SMCs form a barrier that limits immune cell extravasation into this layer) and biochemical (i.e. secretion of immunosuppressive molecules from vascular SMCs) reasons[283]. It is possible this cell type could serve a similar function in this setting, creating a barrier that limits the ability of responding immune cells to move from the vasculature and infiltrate into

the tumour site, thereby allowing OAC to resist ICI. Further study of how vascular SMCs contribute to resistance is warranted, in this setting and perhaps also ICI generally; for this purpose, *RERGL* and the other markers profiled in section 5.3.3.2 may serve as a useful tools to help differentiate vascular SMCs from non-vascular SMCs present in the oesophagus and other GI tissues.

Finally, cell-type specific biomarker analysis was shown to be another potentially exciting application of the LUD2015-005 cell atlas, although comparatively limited by the smaller sample size. Nevertheless, potentially interesting biology was revealed that would have been masked by analysis of bulk samples due to the varied expression of many DEGs across compartments. In the tumour epithelium, *IGFBP2* was shown to be a candidate gene whose tumour-specific expression was associated with CB. In a previous report from melanoma, tumour-specific *IGFBP2* expression was associated with increased PD-L1 expression, and tumours high in both *IGFBP2*/PD-L1 had an immune composition that was predicted to have higher responsiveness to ICI targeting the PD-1/PD-L1 axis[250]. In T cells, PreTx ISGs were associated with NCB patients that had particularly poor INCITE reactivity during the ICI-only window. These ISGs play a key role in the antiviral response to diseases such as HIV and COVID-19, but their sustained expression in T cells from these diseases has been linked with T cell dysfunction and death downstream of *XAF1*-mediated apoptosis[284,285]. Therefore, T cell specific ISG expression could also be a marker of dysfunctional or apoptosis-prone T cells not likely to respond to ICI, which agrees with the low INCITE reactivity seen in these patients. These findings will need to be validated in larger cohorts, but they show how scRNA-seq can potentially reveal predictive signal that may not be discoverable through bulk sequencing methods.

In summary, the LUD2015-005 oesophageal pathology cell atlas showed useful applications for the understanding of the biology of multiple oesophageal tissue types, and allowed for further insight into key predictive signatures for immunochemotherapy response in OAC. Using the deconvolution method described herein, this dataset can potentially be leveraged against any bulk transcriptome dataset from normal oesophagus, Barrett's oesophagus, OAC, OSCC, or gastric tissue to reveal how changes in cell composition are linked with a variable of interest. It

is my hope that this chapter will serve as a useful primer to the LUD2015-005 cell atlas for upper GI researchers interested in leveraging the complex information contained herein on their own research questions.

Chapter 6: Conclusions

In this thesis, a combination of clinical, bulk sequencing, and single-cell RNA-sequencing data were used to deeply profile outcomes on immunochemotherapy in OAC. The implications of these findings have been discussed in further detail in the individual chapter discussions, but the overall key findings and opportunities for future studies will be highlighted in this chapter.

6.1 ICI shows promising activity in some OAC patients

As a phase I/II study, the LUD2015-005 trial was not designed to test whether the addition of ICI to standard of care chemotherapy resulted in more benefit for patients over chemotherapy alone. However, some evidence presented in this thesis suggests that the addition of the ICI strategy implemented in the LUD2015-005 study may be beneficial to patients with inoperable OAC:

- Median survival outcomes were similar to outcomes from immunochemotherapy arms of Phase III studies which have demonstrated benefit ICI addition (Table 1.2; Table 3.2).
- While overall outcomes were not significantly different between the LUD2015-005 trial and a propensity-matched cohort of chemotherapy-treated OAC patients (SEER), outcomes at the tail of the curve were significantly better in the LUD2015-005 cohort (3-year OS: 21.6% in LUD2015-005 vs. 6.8% in SEER, Figure 3.3).
- Three patients already showed irRECIST responses (>30% tumour shrinkage) during the first four weeks where ICI-only was given (Figure 3.5).
- Almost all patients showed some induction of T cell inflammation genes (INCITE) during ICI-only. Those with greater INCITE induction during ICI-only had greater tumour shrinkage during this window (Figure 4.16, Figure 4.17).

These data are in agreement with the recent large-scale phase III trials of αPD-1 ICI, suggesting that the addition of αPD-L1/αCTLA-4 ICI strategies may also benefit oesophageal cancer patients. While the effect sizes in these trials suggest a smaller magnitude of long-term benefit in oesophageal cancer than for highly ICI-sensitive cancers such as melanoma, it appears likely that at least some patients do benefit from the addition of ICI, which may be particularly visible

with respect to long-term survival outcomes. This claim must be taken with some caution regarding OAC, as subgroup analysis for one recent phase III study has reported a hazard ratio for OAC that overlaps one for the addition of ICI to chemotherapy, in contrast to the significant hazard ratio reported for gastric cancer in the same study (CheckMate 649: OAC HR 0.82 [0.60-1.13], n=211; gastric cancer HR 0.76 [0.66-0.87], n=1100)[73]. This could suggest the activity of ICI is less in OAC than gastric cancer; however, this result is potentially explained by the smaller proportion of OAC patients in this study, as the numerical difference between hazard ratios of these two subsets is not large. Indeed, the sizable transcriptomic and radiological responses seen during the ICI-only window in some patients from the LUD2015-005 study are strong evidence that certain subgroups of OAC patients would likely benefit from receiving ICI in addition to chemotherapy. In summary, while ICI addition appears to benefit oesophageal cancer patients as a whole, the percentage of patients in whom ICI addition is likely to improve long-term survival appears to be less in OAC and other oesophageal cancers than in highly ICI-sensitive cancers such as melanoma. Therefore, targeting those who are most likely to benefit from the addition should be a particularly important consideration in this setting.

In considering where αPD-L1 ICI regimens may be a particularly useful alternative to currently approved αPD-1 agents, it is interesting to note that older patients had improved progression-free survival in this cohort in a univariate analysis (Table 3.3). Given the multiple testing conducted in that analysis and that this finding did not survive multivariate regression, age was not determined to be a strong driver of outcomes in this cohort. However, a previous meta-analysis has also suggested that older patients attain more survival benefit from αPD-L1 immunochemotherapy than younger patients, while the opposite trend is seen with αPD-1 immunochemotherapy[116]. Future studies may wish to consider whether αPD-L1 immunochemotherapy is a particularly beneficial option for elderly OAC patients.

6.2 Combination ICI may be a more effective strategy than targeting PD-1 alone

Currently, only ICI targeting PD-1 is approved for first-line immunochemotherapy in OAC. In this study, no test was conducted to test differences in survival between αPD-L1 and αPD-L1/αCTLA-4 arms, as the variation in performance status and the limited power of this phase

I/II study would make this sort of retrospective comparison inappropriate. However, some findings in this work suggest that combination with αCTLA-4, and perhaps other ICI agents, may in fact result in additional benefit:

- PD-1 and PD-L1 are only mildly enriched in OAC compared to normal GI tissues, whereas CTLA-4 is much more specifically enriched in OAC (Figure 4.9).
- Both PD-1 and CTLA-4 are strongly expressed in T cells from OAC biopsies. PD-L1 is rarely expressed on OAC cells, although both PD-1 and CTLA-4 ligands are strongly expressed in the microenvironment, particularly on LAMP3^{hi} DCs (Figure 5.47).
- There was a non-significant trend towards increased induction of cytolytic effectors by αPD-L1/αCTLA-4 during ICI-only compared with αPD-L1, while the αPD-L1 cohort showed increased induction of immunosuppressive macrophages during ICI-only (Figure 4.20).
- Other potentially actionable checkpoint molecules (e.g. *CD276*, *TNFRSF9*) are more highly enriched in OAC than PD-1/PD-L1 (Figure 4.9). These targets could also be considered for use in novel combination ICI approaches.

In summary, while αPD-1/PD-L1 shows benefit in OAC, the specificity of expression patterns suggests that combination with αCTLA-4 or ICIs targeting other checkpoint molecules could yield better responses. Additionally, in this small cohort, patients treated with the αPD-L1 and the highest level of αCTLA-4 had particularly good survival outcomes (Table 3.2), although no statistical comparison of treatment arms was conducted. Especially given that no evidence of increased higher-grade AEs was seen in this analysis with combination ICI, these data suggest that αPD-L1/αCTLA-4 or alternative combination ICI approaches to improve patient outcomes may be worth exploring further as a strategy to extend patient outcomes on immunochemotherapy.

6.3 Highly OAC-specific CT antigens could be additional immunotherapeutic targets

While commonly targeted checkpoint molecule ligands are not strongly expressed in OAC cells in this dataset, this work shows that alternative immunotherapy strategies could be beneficial:

- A number of established CT antigens show an expression pattern that is highly enriched in OAC over normal tissues (Figure 4.8). Some of these have been previously targeted in OAC (MAGE-A3, NY-ESO-1), while others could potentially be targetable (PRAME, MAGE-A4).
- At the PostTx timepoint, strong expression of CT antigens like PRAME is associated with inferior long-term outcomes, likely as a marker of residual OAC cells that were not completely cleared by therapy (Table 4.4, Figure 5.51).

These findings suggest that immunotherapies targeting OAC-specific CT antigens, such as CAR-T cells or cancer vaccines, could lead to effective anti-tumoural responses in some patients. One potentially interesting application would be to use a CT antigen such as PRAME to detect residual disease following immunochemotherapy, and then to target those residual tumour cells resisting broad ICI approaches with an immunotherapy targeted at this specific antigen.

6.4 Both ICI-related and chemotherapy-related adverse events are potentially useful on-treatment markers of immunochemotherapy outcome

Immune-related adverse events are well-studied in the context of ICIs. The findings from this thesis that irAEs are linked with improved outcome on immunochemotherapy are in agreement with a report on an ICI cohort of previously-treated gastro-oesophageal cancers[110], and suggest irAEs also have predictive power in this first-line setting. However, the present study also suggests chemotherapy-related adverse effects may be important:

- In both the full cohort and in the subset of patients receiving the full dosage of αPD-L1, the development of any treatment-related irAE was associated with better long-term survival (Figure 3.7).
- In an analysis that controlled for any possible biases due to duration of oxaliplatin therapy and cumulative toxicity, patients who had a neuropathy event following the first dose of oxaliplatin had significantly improved OS (Figure 3.8).

The link with oxaliplatin-related neuropathy and improved survival on chemotherapy has not yet been established in the literature. Using data from other cohorts, future studies could assess whether this link holds in conventional chemotherapy settings as well, or whether it is specific

to immunochemotherapy. Given the link with neuropathy and damage mediated by myeloid cells, and the positive outcome seen with certain myeloid phenotypes in this study, one hypothesis could be that the development of neuropathy is more likely in patients with a systemic immune state more primed towards a pro-inflammatory myeloid environment. Deeper profiling of the systemic immune state in addition to the intratumoural would be needed to assess this claim.

In any case, both irAEs and non-irAEs are both promising dynamic markers of outcomes on immunochemotherapy. Patients who do not develop neuropathy or irAEs within a set window from treatment onset could be considered likely candidates to develop resistance even in advance of radiological progression, and thus could be potentially targeted for increased surveillance or earlier re-intervention with second-line approaches.

6.5 T cell inflammation is not a predictor of immunochemotherapy response in OAC, but increases do correlate with early responses to ICI-only

T cell inflammation is a well-studied predictor of ICI outcome at the PreTx timepoint in other cancers, where relevant RUO/IUO products are currently in development. In this study, PreTx T cell inflammation was not found to predict outcome on immunochemotherapy for OAC patients, but the increase in these genes were likely to mark patients with good response to ICI-only. Below is a summary of the key findings relevant to T cell inflammation:

- A cross-cancer meta-analysis showed that over half of the genes that consistently predict positive outcome on αPD-1/PD-L1 ICI are relevant to T cell inflammation (antigen presentation, T cell related, interferon signalling, cytokines). Additionally, previously defined T cell inflammation signatures predict positive outcome in many, but not all, ICI cohorts across cancer types. However, these genes and signatures are not predictive of response or clinical benefit on immunochemotherapy in the LUD2015-005 inoperable OAC cohort (Figure 4.22, Figure 4.26).
- The top 70 genes induced by ICI in this cohort are nearly all T cell inflammation genes, and a signature score for these genes (INCITE) shows high correlation with T cell

inflammation signatures used to predict ICI in other settings. Greater INCITE induction was linked with ICI-only responses (Figure 4.16, Figure 4.17).

These findings suggest that the use of PreTx T cell inflammation signatures for first-line immunochemotherapy in OAC would not be useful. There are a few reasons that could explain why our findings on these signatures disagree with some cohorts from other settings:

1. Line of treatment: This cohort is a first-line setting, while many cohorts from which predictive T cell inflammation signatures are derived from cohorts of patients who have already progressed on another therapy. Patients likely have inflammatory responses to certain first-line treatments, particularly with DNA damage-inducing chemotherapies, and it could be that the strength of this inflammatory response to first-line therapy predicts ICI outcome in these cohorts, rather than the treatment-naïve state of the tumour. Supporting this notion, the ICI-responsive genes in this cohort showed good overlap with the PreTx predictive genes from other cohorts.
2. Cancer type: This is the only first-line OAC cohort that has been assessed for T cell inflammation in relation to ICI response. It could therefore be that these signatures predict response in other cancers, but not OAC. As mutational load in OAC is on par with other cancer types included in this study, this could be mediated through the unique microenvironment in OAC, which shows low inflammatory signatures in comparison with some subtypes of gastric cancer and low CD8 infiltration compared with normal GI tissues (Figure 4.7, Figure 4.28).
3. Patient variability: There are many confounding factors inherent to clinical studies that can mask predictive signal. Indeed, despite the widespread study of T cell inflammation signatures in melanoma, only 1 of 3 melanoma cohorts assessed showed strong predictive value for these signatures (Figure 4.26). Similar results were also shown for the predictive power of IPRES resistance signatures. Therefore, it is possible T cell signatures do play a role in predicting outcome in this OAC setting, but did not reach significance in this cohort.

6.6 Monocyte infiltration and coding TMB are markers of patients particularly likely to benefit from ICI addition

While T cell infiltration was not predictive of outcome, a combination of two key molecular features was more than sufficient to stratify patient outcomes in this cohort. The evidence supporting these potentially immunochemotherapy-specific predictors, monocyte infiltration and coding TMB, is summarized below:

- A PreTx inflammatory myeloid signature was found to be predictive of CB in this study (Figure 4.10, Figure 4.26). scRNA-seq revealed that this signal was particularly specific to monocyte cells, although some overlap with differentiating M1-like macrophages and neutrophils was detected (Figure 5.36, Figure 5.38). Interestingly, this signature predicted outcomes for inoperable OAC patients treated with immunochemotherapy in the LUD2015-005 study, but not a similar cohort from the TCGA as a representative of conventional treatment (Figure 5.36)
- The importance of monocytes before therapy could be mediated through an increase in $LAMP3^{hi}$ mature DCs and lymphatic endothelium following ICI. While this did not survive p-value adjustment, if validated in further studies it could suggest these monocyte cells could perhaps contribute to TLS formation or other complex immune structures in response to ICI, thus improving anti-tumour immune responses (Figure 5.40).
- Higher tumour mutational burden, a well-studied marker of ICI response across cancers, was also significantly associated with CB and improved OS for the LUD2015-005 study, but not a similar cohort from the TCGA (Figure 4.1).
- The combination of monocyte-like score and coding TMB load shows more predictive power in this setting than using either variable alone, with the high TMB/high monocyte PreTx phenotype showing the best outcomes by far in this cohort (median OS not met compared with median OS < 12 months for all other combinations; Figure 5.37).

In summary, the combination of TMB and monocyte infiltration shows strong promise as a predictive signature of immunochemotherapy outcome in this first-line OAC cohort.

Importantly, neither of these major PreTx markers was linked with OS in the TCGA, which was used as a representative cohort of conventional therapy in this setting. While some caveats exist for this sort of comparison, particularly given the lack of specificity and completeness regarding TCGA treatment information[286], it appears these two major predictive molecular biomarkers identified in this thesis are both powerfully predictive and immunochemotherapy-specific. Therefore, TMB and monocyte infiltration show great promise as tools to identify patients most likely to benefit from the addition of ICI to standard of care chemotherapy in this setting.

6.7 OAC predictive biomarkers show more overlap with NSCLC than gastric cancer

OACs are often grouped with gastric cancers for clinical trials and previous biomarker studies, likely due to their anatomical overlap around the GOJ which can be difficult to distinguish. However, findings from this study suggest that predictive markers of outcome on ICI-containing regimens are not likely to overlap with overall findings from gastric cancer cohorts, and integration with NSCLC may be more suitable:

- ICI responses in gastric cancer (at least in the second-line setting) are driven by EBV-associated and MSI cancers, which have a unique hyper-inflamed phenotype and are nearly absent in OAC (Figure 4.28, Figure 4.29).
- The innate immune gene signature associated with CB in OAC also predicted ICI outcome in two NSCLC cohorts, but not a gastric cohort (Figure 4.26). This appeared particularly driven by the *CLEC12A* gene, which was most specific to monocyte cells in scRNA-seq data (Figure 4.22, Figure 5.38).
- Positive ICI outcomes have been associated with the A02 HLA supertype in NSCLC[127], and this was the supertype with closest link to predicting greater OS on immunochemotherapy in this OAC cohort (HR: 0.46 (0.20-1.01); Table 4.1). HLA-I homozygosity was also shown to predict poor outcome for ICI in this same NSCLC cohort, as well as in the present study in OAC (Figure 4.2).

The link between oesophageal cancers and NSCLC is plausible due to the shared developmental origin of the oesophagus and the lung, which also share a number of features even in

adulthood[13,287]. With innate immune genes and possibly HLA alleles as the supportive evidence for this connection, it seems likely that any similarities are mediated through immune features shared by the two tumour types. Alveolar macrophages are a primary component of the immune microenvironment of the normal lung, playing a key role in protecting against inhaled pollutants and pathogens[288], and the myeloid phenotype of tumours of the lung are therefore likely important in determining the overall immune state and thus ICI responsiveness. OAC and NSCLC may therefore share the importance of an inflammatory myeloid phenotype related to monocytes in determining ICI response. This finding could be particularly useful for advancing the research on immunochemotherapy in OAC, as the work on ICI biomarkers in NSCLC is much more developed than for OAC. However, histological subtyping of NSCLC would be important for optimal integration with the oesophageal setting, which may not be feasible for smaller NSCLC cohorts.

Predictive biomarkers from OAC are likely to be very different from overall biomarkers for gastric cancer cohorts that are unselected on the basis of molecular subtyping. This is due to the prevalence of EBV-positive and MSI-high gastric cancers. These subtypes appear highly responsive to ICI in gastric cancer, thereby dominating biomarker analysis of gastric cancer ICI cohorts[175], and limiting the suitability of direct comparisons with OAC where these subtypes are rare or absent. The higher rate of EBV-positive and MSI-high tumours could help explain the subgroup analysis of CheckMate 649, which showed a somewhat stronger effect size in gastric cancer than OAC[73]. In combination, these findings suggest that grouping OAC with unselected gastric cancers as one large cohort may be inappropriate for clinical studies of ICI. As biomarkers driving response and potentially patient outcomes on immunochemotherapy may be different between unselected gastric cancer and OAC, future clinical studies should either analyse these tumour types in separate cohorts, or structure enrolment such that there are sufficient numbers to separately assess benefit in both OAC and gastric cancer robustly. Alternatively, future approaches to clinical and biomarker studies could attempt to integrate gastric and OAC tumours by using molecular subtyping. While MSI/EBV-associated cancers in particular may be best analysed separately in integrative analyses, CIN gastric cancers may be more appropriate to combine with OAC for analysis due to their shared molecular profiles.

Analysis of further ICI-responsive CIN gastric cancers would be needed to validate this hypothesis further. In summary, more attention should be paid to molecular subtyping before attempting to integrate OAC with gastric cancer in clinical studies involving ICI. The data also suggest that broader integration with other cancers from foregut-derived tissues such as NSCLC may yield additional useful insight, provided the data are carefully curated to exclude MSI and virus-associated tumours which may otherwise skew results.

6.8 Pathogen-associated gastro-oesophageal cancers are prime targets for ICI

Pathogen-associated cancers are particularly prevalent in the stomach, and may also occur to a lesser extent in cancers of the oesophagus or gastro-oesophageal junction. In this work, several analyses suggest particular sensitivity of these cancers to ICI:

- In a gastric cancer cohort, EBV-associated cancers had a 100% response rate to ICI, and showed a hyper-inflamed PreTx state (Kim et al.[175], Figure 4.28).
- EBV-positive cancers were not found in the LUD2015-005 inoperable cohort, but one of three OSCCs was found to be associated with HPV-16 (Figure 5.18). This OSCC also had a hyper-inflamed PreTx state, and showed strong outcome on immunochemotherapy (Figure 4.30).
- Gastro-oesophageal cancers where *H. pylori* is detected in the paired gastric biopsy may respond better to ICI, although supporting cases are rare and unconfirmed (Figure 5.41). One such inoperable OAC patient with detectable *H. pylori* showed strikingly high PreTx monocyte infiltration and has ongoing OS of over 3.5 years.

For viral cancers, the link with EBV-associated gastric cancers demonstrated by Kim et al. is particularly striking. The lone HPV-associated OSCC in the LUD2015-005 cohort showed a very similar PreTx and ICI response profile, suggesting it may be appropriate to group these virus-associated cancers into the same ICI-sensitive category, if further HPV-associated OSCCs show a similar pattern. Interestingly, a previous analysis of TCGA data did not detect any HPV-associated OSCCs[112], although non-sequencing methods typically find that a substantial proportion of OSCCs do have a link with HPV in many populations[289–291]. This could suggest sequencing methods are less sensitive in detecting HPV, or perhaps there was lower HPV

prevalence among the OSCC patients sampled by the TCGA by chance, as estimates for the prevalence of HPV-associated OSCC can vary significantly even within the same geographical region[291]. In any case, while more patients will need to be profiled, the potential similarity of PreTx phenotype and ICI outcome of this OSCC to EBV-associated gastric cancers suggests viral-associated gastro-oesophageal cancers may be particularly primed for ICI response, and that ICI alone may even be sufficient to generate strong responses in these patients. Based on these data, I propose the following:

1. EBV screening should be a routine practice for gastric cancers. EBV-associated cancers should be strongly considered for ICI-containing regimens, as these should be likely to generate responses. Oesophageal cancers are unlikely to harbour EBV, but cancers of the gastro-oesophageal junction could also be considered for EBV screening.
2. HPV is associated with some OSCCs, and should also be screened for, especially where the virus is particularly prevalent. Non-NGS methods (e.g. PCR, IHC) may be more sensitive than NGS. If larger cohorts can confirm the findings of this study for HPV-associated OSCCs, this could be a particularly useful assay to assist patient selection for the addition of ICI in OSCC. Additionally, HPV-targeted immunotherapeutic interventions (e.g. systemic or intratumoural administration of an HPV vaccine[292,293]) could then be considered as an option for patients with HPV-associated OSCCs that progress on previous lines of therapy.

The link of the lone *H. pylori*-associated OAC with high monocyte infiltration is enticing, as it suggests this monocyte feature linked with outcome could perhaps be linked with features of the microbiome, and that perhaps modulation of the microbiome could help improve immunochemotherapy outcome. For instance, if the monocyte finding is validated, it is possible modulating the microbiome surrounding the tumour with a bacteria that induces monocyte infiltration could be a potential strategy for a complementary therapy. However, much more work needs to be done to clarify whether there is a link between non-viral pathogens and immunochemotherapy response. For this purpose, targeted 16S rRNA sequencing would be

ideally suited to understand the microbiome composition in these patients, and how it might correlate with other predictors found in this study.

6.9 ICI resistance mechanisms are varied, but potentially overcome by immunochemotherapy in some patients

From phase III clinical trial data, while immunochemotherapy provides benefit over chemotherapy, most patients with inoperable OAC are still likely to eventually progress on this first-line therapy. Profiling of the resistance mechanisms leading to progression could help develop supplemental therapies that could prevent or delay progression, and could also help inform monitoring and second-line treatment approaches following immunochemotherapy. The following is a summary of findings associated with treatment resistance in this study:

- *RERGL*$^{+}$ vascular SMCs were identified as the lone cell type associated with NCB following the ICI-only period (Table 5.5). *RERGL* was also the lone DEG to predict NCB at both PreTx and ICI-only timepoints (Table 4.3, Figure 5.42). The role of vascular SMCs are relatively uncharacterized in the context of ICI, but could contribute to ICI resistance by mediating immune cell exclusion as in atherosclerosis[294].
- A lack of induction of T cell inflammatory responses during ICI-only (INCITE quiescence) was associated with less tumour shrinkage during the ICI-only phase of therapy (Figure 4.27). In scRNA-seq data, three INCITE-quiescent patients attaining NCB all showed downregulation of genes related to antigen presentation during the ICI-only window (Table 5.7), a resistance mechanism that has been seen in NSCLC[273]. A T cell-specific ISG-related phenotype showing similar features to apoptotic T cells found in severe viral illnesses was also particularly enriched in two of these quiescent patients (Figure 5.46). Therefore, tumours may limit upregulation of INCITE genes on therapy by limiting presentation of neoantigens to T cells, and certain dysfunctional PreTx T cell states may be less capable of generating inflammatory responses on-therapy.
- IPRES signatures predicting αPD-1 resistance in the Hugo et al. melanoma cohort[130] did not predict resistance at the PreTx timepoint, but were strongly enriched in INCITE-quiescent patients following ICI-only, such that IPRES-high patients also had reduced tumour shrinkage during ICI-only (Figure 4.27). Therefore IPRES signatures, which

include EMT, hypoxia, angiogenesis, and metastasis-related pathways, may also serve as on-treatment markers of ICI-resistance.

- While INCITE-quiescent/IPRES-high tumours generally showed inferior responses to ICI-only, a number of these patients developed strong responses during the immunochemotherapy phase of treatment and had durable long-term survival outcomes (Figure 3.5, Figure 4.17, Figure 4.27). These data suggest that the addition of chemotherapy to ICI can generate responses in tumours that resist ICI-only treatment.
- Two patients upregulated *EGFR* in tumour subclusters showing strong inflammatory responses to ICI, in a pattern that was consistent with amplification and/or genomic rearrangements. This pattern was not linked with poor outcome, but particularly given the resistance of *EGFR*-mutant tumours to ICI, this could represent a common resistance mechanism that occurs under particularly inflammatory environments.
- Following immunochemotherapy, multidrug resistance genes were also found to be associated with NCB and reduced OS after the end of treatment. In particular, *ABCC2* (MRP2) was upregulated in the relatively rare OAC cells surviving immunochemotherapy from one patient in scRNA-seq data (Table 4.4, Figure 5.51). This suggests chemotherapy resistance factors can also drive outcomes in this study.

If these resistance mechanisms can be validated by further study, interventions could help prevent or address the impact of these mechanisms. For INCITE-quiescent tumours, overcoming downregulation of antigen pathways and T cell dysfunction could be challenging, but it appears chemotherapy was able to overcome this resistance mechanism in some patients, which can also be said for IPRES-high tumours. For tumours with resistance to ICI due to *EGFR* overexpression following strong inflammatory responses, TKIs with activity on WT *EGFR* may provide some benefit. MRP-related chemotherapy resistance could also be targeted with agents that inhibit MRP activity, some of which are already in use in clinic for other purposes[295,296]. Finally, if vascular SMCs truly mediate ICI resistance, development of additional interventions targeting this cell type would rely on further examination of the exact mechanism through which this resistance is affected. If vascular SMCs are preventing immune cell infiltration through a physical mechanism, such as an overly contractile phenotype arranged around

tumour-associated vasculature, it is possible that PDE5 inhibitors or other agents that relax vascular SMCs could reverse this function. Indeed, PDE5 inhibition has been linked with increased cytotoxic cell infiltration, which could help explain the association between PDE5 inhibition and better post-surgery outcomes in colorectal cancer[297–299]. A summary of resistance mechanisms and possible ameliorating interventions is found in Table 6.1.

Possible resistance mechanism	Possible interventions
INCITE quiescence (impaired antigen presentation, T cell dysfunction)	Chemotherapy
IPRES overexpression	Chemotherapy
EGFR amplification/overexpression	1st or 2nd generation TKIs
Multidrug resistance protein overexpression	MRP inhibitors
Vascular SMC-mediated immune cell exclusion	PDE5 inhibitors

Table 6.1: Putative resistance mechanisms in the LUD2015-005 inoperable OAC cohort. For each mechanism, a potential therapeutic intervention that could overcome resistance is listed.

6.10 Future directions and perspectives

6.10.1 Next steps for further study

The deep profiling conducted in this thesis generated a number of potentially clinically actionable hypotheses. However, as these were generated from a relatively small cohort of 35 patients, these will need to be validated in other cohorts, particularly in light of the variability seen even within tumour types in the meta-analysis in this thesis. Additionally, as immunochemotherapy is also being trialled as a first-line option in the context of operable gastro-oesophageal cancers, the applicability of these findings outside of the inoperable setting is a key question. Further analysis of the operable arms in the LUD2015-005 study will be important towards addressing the validity and extensibility of these findings, and this will be a focus of our group moving forward.

Of the findings in this thesis that merit particular consideration for further study are the predictive factors that are currently less well characterized in the context of ICI, particularly the notion that monocyte-like cells link with good outcomes on immunochemotherapy, and vascular SMCs link with on-treatment resistance. For vascular SMCs, it remains to be

determined how this cell type mediates resistance to immunochemotherapy in this setting. This could potentially be accomplished through physical prevention of immune cell infiltration, or they could produce substances that suppress the immune system or support cancer progression. If a physical mechanism can be shown, agents that induce SMC relaxation could be a potential option to minimize this resistance mechanism, as discussed in section 6.9. Future studies to address this will need to not only model this cell type, but also the associated tumour, tumour vasculature, and responding adaptive immune cells, in order to fully investigate how this cell type may contribute to ICI resistance. Therefore, *in vivo* or complex *ex vivo* co-culture studies may be needed.

For monocyte-like cells and their role in predicting outcome at the PreTx timepoint, it is known that monocytes can perform direct anti-tumour and antigen presentation functions, but no compelling evidence for monocytes as particularly effective contributors to these pathways could be found in this study. Therefore, the capacity of monocyte-like cells to differentiate into inflammatory myeloid effector cells and to trigger TLS formation should be investigated further. For instance, if monocytes are able to differentiate into or otherwise increase the prevalence of TLS-associated mature DCs, it could explain why they are able to predict better outcome in this setting.

As the monocyte signature was one of the strongest predictive signatures of positive outcome on immunochemotherapy in this study, if this finding is validated and the mechanism is clarified further, it would represent an exciting avenue to explore for supplemental therapies that could further improve immunochemotherapy outcomes. For instance, agents that increase intratumoural monocyte infiltration could be given alongside of or as a sensitizing pre-treatment for immunochemotherapy. As the only *H. pylori*-associated tumour in this cohort had extremely high levels of estimated monocyte infiltration, it is possible that interventions that modify the local microbiome to encourage infiltration of responding monocytes could be an attractive option. Another possibility to encourage monocyte infiltration could follow from oncolytic viruses approaches such as T-VEC, which expresses GM-CSF during replication to facilitate intratumoural dendritic cell and macrophage infiltration in addition to direct oncolytic

activity in tumour cells[300]. A similar approach, perhaps with slight modifications to exchange GM-CSF for a more monocyte-specific chemoattractant, could help induce a microenvironment more favourable to response to immunochemotherapy.

In summary, validation of the monocyte and SMC findings in this thesis and development of further mechanistic understanding for their association with immunochemotherapy outcomes could be important first steps towards developing supplemental therapeutics that can further improve patient outcomes.

6.10.2 Towards an optimized deconvolution workflow for biomarker cohort studies

Deconvolution was found to be a powerful tool in this study to clarify the context of predictive biomarkers, particularly as histological information was unavailable. Deconvolution is thought to be more accurate when all cell types present in the biopsies are present[105], making an approach based on scRNA-seq such as the one in this study likely a more useful than generally-applicable workflows based off of bulk sequencing of sorted cell types. This combination of scRNA-seq on a subset of patients with bulk RNA-seq from the larger cohort allowed for a cost-effective way of assessing the contribution of specific cell types to immunochemotherapy outcome and a number of other related assays, and represents an attractive methodological approach for similar cohort studies. However, this workflow is prone to potential biases of dissociation-based workflows which may be difficult to completely remove, which include dissociation-induced stress and the possibility that different cell types may survive dissociation at different rates. For example, in addition to the known issues with detecting granulocytes by these methods, 28.4% of cells profiled from BO biopsies in scRNA-seq data from this study were identified as plasma cells. This proportion is almost certainly higher than the plasma cell prevalence in the average region of BO, which could have been caused by more efficient collection of live plasma cells during the dissociation process in these patients. Additionally, the spatial context of cells can provide key information which are not assessable by dissociation-based workflows; for instance, in this study certain immune cell types were identified that may participate in TLS, but to truly confirm presence of TLS, a particular patten of co-occurrence of these cell types must be established, which could not be done here.

To address these concerns, deconvolution workflows for cohort studies could substitute in spatial scRNA-seq methods in place of droplet-based methods that rely on single-cell dissociation. Spatial scRNA-seq cells allowed cells to be profiled in sections of tissue, which removes the potential bias of cell-type specific dissociation efficiencies as well as confounding expression signatures associated with dissociation stress. Spatial methods have some drawbacks, including lower RNA detection efficiency than droplet-based methods and resolution wider than the size of a normal cell which may make defining cell type-specific profiles difficult, although newer methods are approaching single-cell resolution (~10 μM)[301,302]. However, the use of spatial methods would also allow to verify the spatial context of cells, which could be critical in identifying complex lymphoid structures and interaction of other cell types. It would also allow for streamlined validation of major findings using immunohistochemistry or immunofluorescence on serial sections. Therefore, particularly as spatial methods continue to improve with RNA capture efficiency and resolution, deconvolution workflows based on spatial scRNA-seq provide an attractive option for cohort studies to allow for both broad hypothesis generation and specific validation of major findings.

Ribodepletion-based bulk RNA-sequencing using snap-frozen endoscopic GI biopsies using the processing methodology described in this study was found to return data of suitable quality, although RNA degradation and low mapping rates were still a concern in some samples. For future bulk RNA-seq cohort studies from the GI tract, long-term storage in liquid nitrogen rather than -80°C may further improve RNA quality, although no degradation associated with increasing length of storage at -80°C was noted in this study. Storage of biopsies in RNA preservative solutions (e.g. RNA*later*) could also potentially provide some improvement in the quality of preserved RNA, although snap-frozen workflows should be preferred if assessment of the proteome is desired[303]. In summary, the workflow described in this thesis could provide a useful methodological foundation for future cohort studies, which could use the findings and suggestions presented herein to further improve the quality and utility of molecular data used for biomarker analysis.

6.10.3 Towards a window strategy for immunochemotherapy treatment

In this study, a window design is implemented which is, to the best of my knowledge, unique among immunochemotherapy studies. This design afforded an opportunity to disambiguate the effects of ICI alone from immunochemotherapy, but also provides a model that could serve some utility for clinical practice. Most patients will likely show the best clinical outcome on combination immunochemotherapy, but there could be some patients for whom ICI alone could be beneficial, perhaps including those with low performance statuses who may not be able to tolerate combination therapy and who also have tumour subtypes particularly sensitive to ICI (e.g. virus-associated or MSI/ hypermutated). This study demonstrates that ICI sensitivity can already be determined to some extent at 4 weeks. In practice, a patient could similarly be started on ICI-only and checked for radiological and transcriptional responses after the first month. If they show strong regression and high induction of inflammatory responses, chemotherapy could be postponed or omitted, particularly in frail patients. However, for those upregulating IPRES resistance signatures and showing minimal inflammatory responses and regression, this work shows that the addition of chemotherapy at this point can still generate strong immunochemotherapy responses.

This window approach could also be beneficial for combining T cell mediated therapies with chemotherapy generally. Chemotherapies generally have off-target effects against many cell types, including T cells in some cases, and the effect can be strongest against rapidly proliferating cells[304]. Therefore, allowing ICI to stimulate T cell influx and proliferation of tumour-specific T cells for a period prior to the introduction of cytotoxic chemotherapy could potentially allow for stronger immune responses to be generated. This trial did not include patients that were treated with a simultaneous immunochemotherapy onset, so this hypothesis cannot be assessed without further studies. In summary, the window approach employed herein is an interesting potential strategy for combination therapies such as first-line immunochemotherapy in OAC, both to assess whether selected patients could continue on ICI without the chemotherapy component and to potentially allow for improved T cell responses.

6.10.4 Towards a post-treatment biopsy workflow to guide clinical management following immunochemotherapy

Even among the patients that maintained progression-free survival throughout the 5-6 months of immunochemotherapy given in the inoperable OAC cohort, long-term outcomes were widely different, with some patients progressing shortly thereafter, and others enjoying continued OS for a number of years. Differentiating between these long-term outcomes following this first-line course could help target patients at higher risk of progression for additional screening and potentially earlier intervention with alternative second-line therapies. This thesis shows that a number of factors could be linked with long-term outcome at the PostTx timepoint:

- High residual tumour content, including residual OAC cells by scRNA-seq or preserved expression of CT antigens by bulk RNA-seq, at the site of the primary oesophageal lesion is associated with poor outcome (Table 4.4, Section 5.5.3.12). *FCGBP* can be used to help distinguish residual OAC from PostTx BO, which is prevalent in most patients (Figure 5.62)
- Transcriptional expression of chemotherapy resistance markers (esp. multi-drug resistance proteins) are also associated with poor outcome following PostTx (Table 4.4).
- Certain markers of maintained cytotoxic cell infiltration (*KLRC1* and *ITGAE*/CD103), other immune infiltrate with lipid presentation machinery (*CD1C* and *CD1E*), and high signatures of translation/rRNA processing (*GCFC2*, *WDR36*, *MRPL1*) could be used as markers of patients with longer expected OS following immunochemotherapy (Table 4.4).
- Expression-based markers may perform better in this setting than radiological measures (Table 4.4).

A biopsy workflow at the PostTx workflow similar to those used in the neoadjuvant setting to assess pathological response could therefore show major improvements over radiological assessment alone to predict patient outcomes and manage care following immunochemotherapy. Biopsies could be assessed for residual OAC and immune infiltrate using the markers described above, and patients with poor prognostic factors (high residual tumour,

low T/NK infiltrate, chemotherapy resistance genes) could derive benefit from more regular screening to detect radiological progression, or even earlier targeting with second-line therapy in advance of progression. Less invasive approaches could also be developed; for instance, the Cytosponge device has recently been implemented as an alternative approach to detecting BO, which collects cells from the oesophagus using a sponge attached to a string, thereby avoiding the need for anaesthesia or invasive biopsies[305]. A similar non-invasive approach could perhaps be adapted for post-immunochemotherapy sampling, perhaps by using the Barrett's marker *FCGBP* to assist in differentiation of residual OAC from BO. Regardless of the methodology, PostTx sampling to assess pathological response to immunochemotherapy would give patients information on their prognosis following this first-line course, and could help physicians make more informed decisions about monitoring and second-line treatment.

6.10.5 Closing comments

This work has generated a number of findings and hypotheses that could potentially have clinical utility before, during, and after treatment with first-line immunochemotherapy. However, all conclusions from this study must be taken with caution until they can be replicated in an independent cohort. The validation of these and other biomarker analyses from this setting, and the further study of their implications, are now of primary importance as immunochemotherapy enters the clinic for OAC, as they could lead to improved patient selection, on-treatment monitoring, and supplemental therapies for immunochemotherapeutic regimens. The pursuit of these aims will help derive the most benefit from the advent of this novel treatment strategy, and offers a promising path forward in continuing to improve the prognosis of patients with inoperable OAC.

REFERENCES

1. Goldman, N. *et al.* Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **494**, 77–80 (2013).

2. Costa, A. *https://github.com/allanino/DNA/releases/tag/v0.4*. (2015).

3. Oezcelik, A. & DeMeester, S. R. General anatomy of the esophagus. *Thorac Surg Clin* **21**, 289–297, x (2011).

4. Kuo, B. & Urma, D. Esophagus - anatomy and development. *GI Motility online* (2006) doi:10.1038/gimo6.

5. Rao, J. N. & Wang, J.-Y. *Intestinal Architecture and Development*. *Regulation of Gastrointestinal Mucosal Growth* (Morgan & Claypool Life Sciences, 2010).

6. Zhang, X. *et al.* The microscopic anatomy of the esophagus including the individual layers, specialized tissues, and unique components and their responses to injury. *Annals of the New York Academy of Sciences* **1434**, 304–318 (2018).

7. Long, J. D. & Orlando, R. C. Esophageal submucosal glands: structure and function. *Am J Gastroenterol* **94**, 2818–2824 (1999).

8. Abdulnour-Nakhoul, S. *et al.* HCO3− secretion in the esophageal submucosal glands. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **288**, G736–G744 (2005).

9. Paterson, W. G. Esophageal peristalsis. *GI Motility online* (2006) doi:10.1038/gimo13.

10. Katz, J. M., Malik, A. & Basit, H. Embryology, Esophagus. in *StatPearls* (StatPearls Publishing, 2021).

11. Jacobs, I. J., Ku, W.-Y. & Que, J. Genetic and Cellular Mechanisms Regulating Anterior Foregut and Esophageal Development. *Dev Biol* **369**, 54–64 (2012).

12. Rosekrans, S. L., Baan, B., Muncan, V. & van den Brink, G. R. Esophageal development and epithelial homeostasis. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **309**, G216–G228 (2015).

13. Morrisey, E. E. & Rustgi, A. K. The Lung and Esophagus: Developmental and Regenerative Overlap. *Trends in Cell Biology* **28**, 738–748 (2018).

14. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68**, 394–424 (2018).

15. Devesa, S. S., Blot, W. J. & Fraumeni, J. F. Changing patterns in the incidence of esophageal and gastric carcinoma in the United States. *Cancer* **83**, 2049–2053 (1998).

16. Fitzgerald, R. C. Review article: Barrett's oesophagus and associated adenocarcinoma – a UK perspective. *Alimentary Pharmacology & Therapeutics* **20**, 45–49 (2004).

17. Lepage, C., Rachet, B., Jooste, V., Faivre, J. & Coleman, M. P. Continuing rapid increase in esophageal adenocarcinoma in England and Wales. *Am. J. Gastroenterol.* **103**, 2694–2699 (2008).

18. Zhang, Y. Epidemiology of esophageal cancer. *World J Gastroenterol* **19**, 5598–5606 (2013).

19. Pennathur, A., Gibson, M. K., Jobe, B. A. & Luketich, J. D. Oesophageal carcinoma. *The Lancet* **381**, 400–412 (2013).

20. Groulx, S. *et al.* Guideline on screening for esophageal adenocarcinoma in patients with chronic gastroesophageal reflux disease. *CMAJ* **192**, E768–E777 (2020).

21. Jain, S. & Dhingra, S. Pathology of esophageal cancer and Barrett's esophagus. *Ann Cardiothorac Surg* **6**, 99–109 (2017).

22. Conteduca, V. *et al.* Barrett's esophagus and esophageal cancer: An overview. *International Journal of Oncology* **41**, 414–424 (2012).

23. Solaymani-Dodaran, M., Logan, R. F. A., West, J., Card, T. & Coupland, C. Risk of oesophageal cancer in Barrett's oesophagus and gastro-oesophageal reflux. *Gut* **53**, 1070–1074 (2004).

24. Spechler, S. J. Barrett's Esophagus: Is the Goblet Half-Empty? *Clin Gastroenterol Hepatol* **10**, 1237–1238 (2012).

25. Jovov, B. *et al.* Claudin-18: a dominant tight junction protein in Barrett's esophagus and likely contributor to its acid resistance. *Am J Physiol Gastrointest Liver Physiol* **293**, G1106-1113 (2007).

26. Curtius, K., Rubenstein, J. H., Chak, A. & Inadomi, J. M. Computational modelling suggests that Barrett's oesophagus may be the precursor of all oesophageal adenocarcinomas. *Gut* **70**, 1435–1440 (2021).

27. Barbera, M. & Fitzgerald, R. C. Cellular origin of Barrett's metaplasia and oesophageal stem cells. *Biochemical Society Transactions* **38**, 370–373 (2010).

28. Que, J., Garman, K. S., Souza, R. F. & Spechler, S. J. Pathogenesis and Cells of Origin of Barrett's Esophagus. *Gastroenterology* **157**, 349-364.e1 (2019).

29. Garman, K. S. Origin of Barrett's Epithelium: Esophageal Submucosal Glands. *Cellular and Molecular Gastroenterology and Hepatology* **4**, 153–156 (2017).

30. Leedham, S. J. *et al.* Individual crypt genetic heterogeneity and the origin of metaplastic glandular epithelium in human Barrett's oesophagus. *Gut* **57**, 1041–1048 (2008).

31. Krüger, L. *et al.* Ductular and proliferative response of esophageal submucosal glands in a porcine model of esophageal injury and repair. *Am J Physiol Gastrointest Liver Physiol* **313**, G180–G191 (2017).

32. Owen, R. P. *et al.* Single cell RNA-seq reveals profound transcriptional similarity between Barrett's oesophagus and oesophageal submucosal glands. *Nat Commun* **9**, 4261 (2018).

33. The Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–175 (2017).

34. Rüdiger Siewert, J., Feith, M., Werner, M. & Stein, H. J. Adenocarcinoma of the Esophagogastric Junction. *Ann Surg* **232**, 353–361 (2000).

35. Zhang, S., Orita, H. & Fukunaga, T. Current surgical treatment of esophagogastric junction adenocarcinoma. *World J Gastrointest Oncol* **11**, 567–578 (2019).

36. Hooi, J. K. Y. *et al.* Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. *Gastroenterology* **153**, 420–429 (2017).

37. Xie, F.-J. *et al.* Helicobacter pylori infection and esophageal cancer risk: An updated meta-analysis. *World J Gastroenterol* **19**, 6098–6107 (2013).

38. Lin, E. W. *et al.* Comparative transcriptomes of adenocarcinomas and squamous cell carcinomas reveal molecular similarities that span classical anatomic boundaries. *PLOS Genetics* **13**, e1006938 (2017).

39. Yokota, E. *et al.* Targeted silencing of SOX2 by an artificial transcription factor showed antitumor effect in lung and esophageal squamous cell carcinoma. *Oncotarget* **8**, 103063–103076 (2017).

40. Lordick, F. *et al.* Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **27**, v50–v57 (2016).

41. Smyth, E. C. *et al.* Oesophageal Cancer. *Nat Rev Dis Primers* **3**, 17048 (2017).

42. Akutsu, Y. & Matsubara, H. Chemoradiotherapy and surgery for T4 esophageal cancer in Japan. *Surg Today* **45**, 1360–1365 (2015).

43. Makino, T. *et al.* Treatment and clinical outcome of clinical T4 esophageal cancer: A systematic review. *Ann Gastroenterol Surg* **3**, 169–180 (2018).

44. Smyth, E. C. *et al.* Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Annals of Oncology* **27**, v38–v49 (2016).

45. Janmaat, V. T. *et al.* Palliative chemotherapy and targeted therapies for esophageal and gastroesophageal junction cancer. *Cochrane Database Syst Rev* **2017**, CD004063 (2017).

46. Jatoi, A. *et al.* Oxaliplatin and capecitabine in patients with metastatic adenocarcinoma of the esophagus, gastroesophageal junction and gastric cardia: a phase II study from the North Central Cancer Treatment Group. *Ann Oncol* **17**, 29–34 (2006).

47. Dijksterhuis, W. P. M. *et al.* Association between body composition, survival, and toxicity in advanced esophagogastric cancer patients receiving palliative chemotherapy. *J Cachexia Sarcopenia Muscle* **10**, 199–206 (2019).

48. Cunningham, D. *et al.* Capecitabine and Oxaliplatin for Advanced Esophagogastric Cancer. *New England Journal of Medicine* **358**, 36–46 (2008).

49. Waddell, T. *et al.* Epirubicin, oxaliplatin, and capecitabine with or without panitumumab for patients with previously untreated advanced oesophagogastric cancer (REAL3): a randomised, open-label phase 3 trial. *Lancet Oncol* **14**, 481–489 (2013).

50. Coley, W. B. The treatment of malignant tumors by repeated inoculations of erysipelas. With a report of ten original cases. 1893. *Clin Orthop Relat Res* 3–11 (1991).

51. Nauts, H. C. & McLaren, J. R. Coley Toxins — The First Century. in *Consensus on Hyperthermia for the 1990s: Clinical Practice in Cancer Treatment* (eds. Bicher, H. I., McLaren, J. R. & Pigliucci, G. M.) 483–500 (Springer US, 1990). doi:10.1007/978-1-4684-5766-7_52.

52. Morales, A., Eidinger, D. & Bruce, A. W. Intracavitary Bacillus Calmette-Guerin in the treatment of superficial bladder tumors. *J Urol* **116**, 180–183 (1976).

53. van Puffelen, J. H. *et al.* Trained immunity as a molecular mechanism for BCG immunotherapy in bladder cancer. *Nat Rev Urol* **17**, 513–525 (2020).

54. Bree, K. K., Brooks, N. A. & Kamat, A. M. Current Therapy and Emerging Intravesical Agents to Treat Non-Muscle Invasive Bladder Cancer. *Hematol Oncol Clin North Am* **35**, 513–529 (2021).

55. Dobosz, P. & Dzieciątkowski, T. The Intriguing History of Cancer Immunotherapy. *Front Immunol* **10**, (2019).

56. Zhang, Y. & Zhang, Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell Mol Immunol* **17**, 807–821 (2020).

57. Buchbinder, E. I. & Desai, A. CTLA-4 and PD-1 Pathways. *Am J Clin Oncol* **39**, 98–106 (2016).

58. Zaravinos, A. An updated overview of HPV-associated head and neck carcinomas. *Oncotarget* **5**, 3956–3969 (2014).

59. Waldman, A. D., Fritz, J. M. & Lenardo, M. J. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat Rev Immunol* **20**, 651–668 (2020).

60. Hodi, F. S. *et al.* Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *New England Journal of Medicine* **363**, 711–723 (2010).

61. Traynor, K. Ipilimumab approved for metastatic melanoma. *American Journal of Health-System Pharmacy* **68**, 768–768 (2011).

62. Larkin, J. *et al.* Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N Engl J Med* **381**, 1535–1546 (2019).

63. McDermott, D. *et al.* Durable benefit and the potential for long-term survival with immunotherapy in advanced melanoma. *Cancer Treat Rev* **40**, 1056–1064 (2014).

64. Arora, S. *et al.* Existing and Emerging Biomarkers for Immune Checkpoint Immunotherapy in Solid Tumors. *Adv Ther* **36**, 2638–2678 (2019).

65. Yarchoan, M., Hopkins, A. & Jaffee, E. M. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *http://dx.doi.org/10.1056/NEJMc1713444* (2017) doi:10.1056/NEJMc1713444.

66. Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat Genet* **45**, 1113–1120 (2013).

67. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401–404 (2012).

68. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).

69. Pembrolizumab Approved for Esophageal Cancer - National Cancer Institute. https://www.cancer.gov/news-events/cancer-currents-blog/2019/pembrolizumab-esophageal-fda-approval (2019).

70. Pembrolizumab Secures FDA Approval in Stomach Cancer - National Cancer Institute. https://www.cancer.gov/news-events/cancer-currents-blog/2017/gastric-cancer-fda-keytruda (2017).

71. FDA Center for Drug Evaluation and Research. FDA approves nivolumab in combination with chemotherapy for metastatic gastric cancer and esophageal adenocarcinoma. *https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-nivolumab-combination-chemotherapy-metastatic-gastric-cancer-and-esophageal* (2021).

72. FDA Center for Drug Evaluation and Research. FDA approves pembrolizumab for esophageal or GEJ carcinoma. *https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-pembrolizumab-esophageal-or-gej-carcinoma* (2021).

73. Janjigian, Y. Y. *et al.* First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. *The Lancet* **0**, (2021).

74. Boku, N. *et al.* LBA7_PR Nivolumab plus chemotherapy versus chemotherapy alone in patients with previously untreated advanced or recurrent gastric/gastroesophageal junction (G/GEJ) cancer: ATTRACTION-4 (ONO-4538-37) study. *Annals of Oncology* **31**, S1192 (2020).

75. Kato, K. *et al.* LBA8_PR Pembrolizumab plus chemotherapy versus chemotherapy as first-line therapy in patients with advanced esophageal cancer: The phase 3 KEYNOTE-590 study. *Annals of Oncology* **31**, S1192–S1193 (2020).

76. Kelly, R. J. *et al.* Adjuvant Nivolumab in Resected Esophageal or Gastroesophageal Junction Cancer. *New England Journal of Medicine* **384**, 1191–1203 (2021).

77. Högner, A. & Thuss-Patience, P. Immune Checkpoint Inhibition in Oesophago-Gastric Carcinoma. *Pharmaceuticals (Basel)* **14**, 151 (2021).

78. Smyth, E. C., Gambardella, V., Cervantes, A. & Fleitas, T. Checkpoint inhibitors for gastroesophageal cancers: dissecting heterogeneity to better understand their role in first-line and adjuvant therapy. *Annals of Oncology* **32**, 590–599 (2021).

79. van Holstein, Y. *et al.* Efficacy and Adverse Events of Immunotherapy with Checkpoint Inhibitors in Older Patients with Cancer. *Drugs Aging* **36**, 927–938 (2019).

80. Vasekar, M. K., Agbese, E. & Leslie, D. The value of immunotherapy: Comparison of annual cost per patient receiving immunotherapy versus chemotherapy in patients with non-small cell lung cancer. *JCO* **38**, e19364–e19364 (2020).

81. Desai, A. & Gyawali, B. Financial toxicity of cancer treatment: Moving the discussion from acknowledgement of the problem to identifying solutions. *EClinicalMedicine* **20**, (2020).

82. Hernandez, I. *et al.* Pricing of monoclonal antibody therapies: higher if used for cancer? *Am J Manag Care* **24**, 109–112 (2018).

83. Acevedo, F. *et al.* How much survival benefit is necessary for breast cancer patients to opt for adjuvant chemotherapy? Results from a Chilean survey. *Ecancermedicalscience* **8**, 391 (2014).

84. Zietemann, V. D., Schuster, T. & Duell, T. H. Post-study therapy as a source of confounding in survival analysis of first-line studies in patients with advanced non-small-cell lung cancer. *J Thorac Dis* **3**, 88–98 (2011).

85. Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* **45**, 228–247 (2009).

86. Le Lay, J., Jarraya, H., Lebellec, L. & Penel, N. irRECIST and iRECIST: the devil is in the details. *Ann Oncol* **28**, 1676–1678 (2017).

87. Kataoka, Y. & Hirano, K. Which criteria should we use to evaluate the efficacy of immune-checkpoint inhibitors? *Ann Transl Med* **6**, (2018).

88. Bohnsack, O., Hoos, A. & Ludajic, K. 1070P - Adaptation of the Immune Related Response Criteria: Irrecist. *Annals of Oncology* **25**, iv369 (2014).

89. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).

90. Litchfield, K. *et al.* Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596-614.e14 (2021).

91. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463–1469 (2016).

92. Strickler, J. H., Hanks, B. A. & Khasraw, M. Tumor Mutational Burden as a Predictor of Immunotherapy Response: Is More Always Better? *Clin Cancer Res* **27**, 1236–1241 (2021).

93. Hellmann, M. D. *et al.* Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. *New England Journal of Medicine* **378**, 2093–2104 (2018).

94. Galon, J. & Bruni, D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat Rev Drug Discov* **18**, 197–218 (2019).

95. Puccini, A., Battaglin, F., Iaia, M. L., Lenz, H.-J. & Salem, M. E. Overcoming resistance to anti-PD1 and anti-PD-L1 treatment in gastrointestinal malignancies. *J Immunother Cancer* **8**, e000404 (2020).

96. Maleki Vareki, S. High and low mutational burden tumors versus immunologically hot and cold tumors and response to immune checkpoint inhibitors. *Journal for ImmunoTherapy of Cancer* **6**, 157 (2018).

97. Damotte, D. *et al.* The tumor inflammation signature (TIS) is associated with anti-PD-1 treatment benefit in the CERTIM pan-cancer cohort. *Journal of Translational Medicine* **17**, 357 (2019).

98. Shen, X. & Zhao, B. Efficacy of PD-1 or PD-L1 inhibitors and PD-L1 expression status in cancer: meta-analysis. *BMJ* **362**, k3529 (2018).

99. Liu, X. *et al.* Association of PD-L1 expression status with the efficacy of PD-1/PD-L1 inhibitors and overall survival in solid tumours: A systematic review and meta-analysis. *International Journal of Cancer* **147**, 116–127 (2020).

100. Davis, A. A. & Patel, V. G. The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors. *Journal for ImmunoTherapy of Cancer* **7**, 278 (2019).

101. Samstein, R. M. *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206 (2019).

102. Janjigian, Y. Y. *et al.* Genetic predictors of response to systemic therapy in esophagogastric cancer. *Cancer Discov* **8**, 49–58 (2018).

103. Greally, M. *et al.* Clinical and molecular predictors of response to immune checkpoint inhibitors in patients with advanced esophagogastric cancer. *Clin Cancer Res* **25**, 6160–6169 (2019).

104. Study of Anti-PD-L1 in Combination With Chemo(Radio)Therapy for Oesophageal Cancer - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT02735239.

105. Avila Cobos, F., Alquicira-Hernandez, J., Powell, J. E., Mestdagh, P. & De Preter, K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun* **11**, 5650 (2020).

106. Sanlorenzo, M. *et al.* Pembrolizumab Cutaneous Adverse Events and Their Association With Disease Progression. *JAMA Dermatol* **151**, 1206–1212 (2015).

107. Downey, S. G. *et al.* Prognostic factors related to clinical response in patients with metastatic melanoma treated by CTL-associated antigen-4 blockade. *Clin Cancer Res* **13**, 6681–6688 (2007).

108. Freeman-Keller, M. *et al.* Nivolumab in Resected and Unresectable Metastatic Melanoma: Characteristics of Immune-Related Adverse Events and Association with Outcomes. *Clin Cancer Res* **22**, 886–894 (2016).

109. Das, S. & Johnson, D. B. Immune-related adverse events and anti-tumor efficacy of immune checkpoint inhibitors. *J Immunother Cancer* **7**, 306 (2019).

110. Greally, M. *et al.* Clinical and Molecular Predictors of Response to Immune Checkpoint Inhibitors in Patients with Advanced Esophagogastric Cancer. *Clin. Cancer Res.* **25**, 6160–6169 (2019).

111. Ludwig Institute for Cancer Research. *Phase 1/2 Study of Anti-PD-L1 in Combination With Chemo(Radio)Therapy for Oesophageal Cancer*. https://clinicaltrials.gov/ct2/show/NCT02735239 (2021).

112. Kim, J. *et al.* Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–175 (2017).

113. Gide, T. N. *et al.* Distinct Immune Cell Populations Define Response to Anti-PD-1 Monotherapy and Anti-PD-1/Anti-CTLA-4 Combined Therapy. *Cancer Cell* **35**, 238-255.e6 (2019).

114. Katai, H. *et al.* Five-year survival analysis of surgically resected gastric cancer cases in Japan: a retrospective analysis of more than 100,000 patients from the nationwide registry of the Japanese Gastric Cancer Association (2001–2007). *Gastric Cancer* **21**, 144–154 (2018).

115. Conforti, F. *et al.* Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis. *The Lancet Oncology* **19**, 737–746 (2018).

116. Duan, J. *et al.* Immunotherapy efficacy and patients' age: A systematic review and meta-analysis. *JCO* **37**, e14168–e14168 (2019).

117. Abdel-Rahman, O. ECOG performance score 0 versus 1: impact on efficacy and safety of first-line 5-FU-based chemotherapy among patients with metastatic colorectal cancer included in five randomized trials. *Int J Colorectal Dis* **34**, 2143–2150 (2019).

118. US Department of Health and Human Services. Common Terminology Criteria for Adverse Events (CTCAE), Version 4.0.3. 80 (2010).

119. Saif, M. W. & Reardon, J. Management of oxaliplatin-induced peripheral neuropathy. *Ther Clin Risk Manag* **1**, 249–258 (2005).

120. Ribas, A. *et al.* New Challenges in Endpoints for Drug Development in Advanced Melanoma. *Clin Cancer Res* **18**, 336–341 (2012).

121. Starobova, H. & Vetter, I. Pathophysiology of Chemotherapy-Induced Peripheral Neuropathy. *Front Mol Neurosci* **10**, 174 (2017).

122. Yamanouchi, K. *et al.* The Relationship Between Peripheral Neuropathy Induced by Docetaxel and Systemic Inflammation-based Parameters in Patients with Breast Cancer. *Anticancer Research* **37**, 6947–6951 (2017).

123. Li, D. *et al.* Preventive Effects of Bee Venom Derived Phospholipase A2 on Oxaliplatin-Induced Neuropathic Pain in Mice. *Toxins* **8**, 27 (2016).

124. Lees, J. G. *et al.* Immune-mediated processes implicated in chemotherapy-induced peripheral neuropathy. *European Journal of Cancer* **73**, 22–29 (2017).

125. Al-Mazidi, S. *et al.* Blocking of cytokines signalling attenuates evoked and spontaneous neuropathic pain behaviours in the paclitaxel rat model of chemotherapy-induced neuropathy. *European Journal of Pain* **22**, 810–821 (2018).

126. Chowell, D. *et al.* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* **359**, 582–587 (2018).

127. Abed, A. *et al.* Prognostic value of HLA-I homozygosity in patients with non-small cell lung cancer treated with single agent immunotherapy. *J Immunother Cancer* **8**, e001620 (2020).

128. Roach, C. *et al.* Development of a Companion Diagnostic PD-L1 Immunohistochemistry Assay for Pembrolizumab Therapy in Non-Small-cell Lung Cancer. *Appl Immunohistochem Mol Morphol* **24**, 392–397 (2016).

129. Conroy, J. M. *et al.* Next generation sequencing of PD-L1 for predicting response to immune checkpoint inhibitors. *Journal for ImmunoTherapy of Cancer* **7**, 18 (2019).

130. Hugo, W. *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165**, 35–44 (2016).

131. Sidney, J., Peters, B., Frahm, N., Brander, C. & Sette, A. HLA class I supertypes: a revised and updated classification. *BMC Immunology* **9**, 1 (2008).

132. Opitz, L. *et al.* Impact of RNA degradation on gene expression profiling. *BMC Med Genomics* **3**, 36 (2010).

133. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* **7**, 3 (2006).

134. Stiekema, J. *et al.* Biobanking of fresh-frozen endoscopic biopsy specimens from esophageal adenocarcinoma. *Dis Esophagus* **29**, 1100–1106 (2016).

135. Zhao, S., Zhang, Y., Gamini, R., Zhang, B. & von Schack, D. Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. *Sci Rep* **8**, 4781 (2018).

136. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

137. Dobin, A. & Gingeras, T. R. Optimizing RNA-Seq Mapping with STAR. *Methods Mol Biol* **1415**, 245–262 (2016).

138. Gao, G. F. *et al.* Before and After: Comparison of Legacy and Harmonized TCGA Genomic Data Commons' Data. *Cell Syst* **9**, 24-34.e10 (2019).

139. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).

140. Ha, H., Debnath, B. & Neamati, N. Role of the CXCL8-CXCR1/2 Axis in Cancer and Inflammatory Diseases. *Theranostics* **7**, 1543–1588 (2017).

141. Vonderheide, R. H. & Bear, A. S. Tumor-Derived Myeloid Cell Chemoattractants and T Cell Exclusion in Pancreatic Cancer. *Front Immunol* **11**, 605619 (2020).

142. Pietro, M. D. & Fitzgerald, R. C. HOXB cluster genes are novel potential regulators of Barrett's oesophagus development. *Gut* **60**, A20–A21 (2011).

143. Errico, M. C., Jin, K., Sukumar, S. & Carè, A. The widening sphere of influence of HOXB7 in solid tumors. *Cancer Res* **76**, 2857–2862 (2016).

144. D'ambrosio, D., Mariani, M., Panina-Bordignon, P. & Sinigaglia, F. Chemokines and Their Receptors Guiding T Lymphocyte Recruitment in Lung Inflammation. *Am J Respir Crit Care Med* **164**, 1266–1275 (2001).

145. Morgan, R. A. *et al.* Cancer regression and neurologic toxicity following anti-MAGE-A3 TCR gene therapy. *J Immunother* **36**, 133–151 (2013).

146. Raza, A. *et al.* Unleashing the immune response to NY-ESO-1 cancer testis antigen as a potential target for cancer immunotherapy. *J Transl Med* **18**, 1–11 (2020).

147. Almeida, L. G. *et al.* CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res* **37**, D816-819 (2009).

148. Guo, R.-F. & Ward, P. A. Role of C5a in Inflammatory Responses. *Annual Review of Immunology* **23**, 821–852 (2005).

149. Chen, K. *et al.* Regulation of inflammation by members of the formyl-peptide receptor family. *J Autoimmun* **85**, 64–77 (2017).

150. Martinez, F. O. The transcriptome of human monocyte subsets begins to emerge. *Journal of Biology* **8**, 99 (2009).

151. Wu, R., Chen, F., Wang, N., Tang, D. & Kang, R. ACOD1 in immunometabolism and disease. *Cellular & Molecular Immunology* **17**, 822–833 (2020).

152. Rødland, E. K. *et al.* Expression of genes in normal human monocytes in response to Aspergillus fumigatus. *Med Mycol* **46**, 327–336 (2008).

153. Newell, K. A. *et al.* Longitudinal studies of a B cell-derived signature of tolerance in renal transplant recipients. *Am J Transplant* **15**, 2908–2920 (2015).

154. Peng, B., Ming, Y. & Yang, C. Regulatory B cells: the cutting edge of immune tolerance in kidney transplantation. *Cell Death Dis* **9**, 109 (2018).

155. Palacios, E. H. & Weiss, A. Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation. *Oncogene* **23**, 7990–8000 (2004).

156. Rossy, J., Williamson, D. J. & Gaus, K. How does the kinase Lck phosphorylate the T cell receptor? Spatial organization as a regulatory mechanism. *Front Immunol* **3**, 167 (2012).

157. Choudhury, S. R. *et al.* Dipeptidase-1 Is an Adhesion Receptor for Neutrophil Recruitment in Lungs and Liver. *Cell* **178**, 1205-1221.e17 (2019).

158. Besray Unal, E. *et al.* Systems level expression correlation of Ras GTPase regulators. *Cell Commun Signal* **16**, (2018).

159. Chen, Y. *et al.* CD8+ T Cells Form the Predominant Subset of NKG2A+ Cells in Human Lung Cancer. *Front. Immunol.* **10**, (2020).

160. Chancellor, A., Gadola, S. D. & Mansour, S. The versatility of the CD1 lipid antigen presentation pathway. *Immunology* **154**, 196–203 (2018).

161. Vrana, D. *et al.* ABC Transporters and Their Role in the Neoadjuvant Treatment of Esophageal Cancer. *Int J Mol Sci* **19**, (2018).

162. Myint, K. *et al.* Identification of MRP2 as a targetable factor limiting oxaliplatin accumulation and response in gastrointestinal cancer. *Sci Rep* **9**, 2245 (2019).

163. Biswas, R. *et al.* Transport-Mediated Oxaliplatin Resistance Associated with Endogenous Overexpression of MRP2 in Caco-2 and PANC-1 Cells. *Cancers (Basel)* **11**, (2019).

164. Simon, S. & Labarriere, N. PD-1 expression on tumor-specific T cells: Friend or foe for immunotherapy? *Oncoimmunology* **7**, e1364828 (2017).

165. Ponomareva, L. *et al.* AIM2, an IFN-inducible cytosolic DNA sensor, in the development of benign prostate hyperplasia and prostate cancer. *Mol Cancer Res* **11**, 1193–1202 (2013).

166. Watson, S. F., Bellora, N. & Macias, S. ILF3 contributes to the establishment of the antiviral type I interferon program. *Nucleic Acids Res* **48**, 116–129 (2020).

167. Resta-Lenert, S., Das, S., Batra, S. K. & Ho, S. B. Muc17 protects intestinal epithelial cells from enteroinvasive E. coli infection by promoting epithelial barrier integrity. *Am J Physiol Gastrointest Liver Physiol* **300**, G1144-1155 (2011).

168. Layunta, E., Jäverfelt, S., Dolan, B., Arike, L. & Pelaseyed, T. IL-22 promotes the formation of a MUC17 glycocalyx barrier in the postnatal small intestine during weaning. *Cell Reports* **34**, 108757 (2021).

169. Sangro, B. *et al.* Association of inflammatory biomarkers with clinical outcomes in nivolumab-treated patients with advanced hepatocellular carcinoma. *J Hepatol* **73**, 1460–1469 (2020).

170. Ott, P. A. *et al.* T-Cell–Inflamed Gene-Expression Profile, Programmed Death Ligand 1 Expression, and Tumor Mutational Burden Predict Efficacy in Patients Treated With Pembrolizumab Across 20 Cancers: KEYNOTE-028. *JCO* **37**, 318–327 (2019).

171. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).

172. Spranger, S. *et al.* Density of immunogenic antigens does not explain the presence or absence of the T-cell–inflamed tumor microenvironment in melanoma. *Proc Natl Acad Sci U S A* **113**, E7759–E7768 (2016).

173. Seiwert, T. Y. *et al.* Safety and clinical activity of pembrolizumab for treatment of recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-012): an open-label, multicentre, phase 1b trial. *The Lancet Oncology* **17**, 956–965 (2016).

174. Ayers, M. *et al.* IFN-γ–related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* **127**, 2930–2940.

175. Kim, S. T. *et al.* Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nature Medicine* **24**, 1449–1458 (2018).

176. Noh, M.-G. *et al.* Practical prediction model of the clinical response to programmed death-ligand 1 inhibitors in advanced gastric cancer. *Experimental & Molecular Medicine* **53**, 223–234 (2021).

177. T Cells Targeting MAGE-A4 Shrink Tumors | Cancer Discovery. https://cancerdiscovery.aacrjournals.org/content/10/8/OF2.long.

178. Orlando, D. *et al.* Adoptive Immunotherapy Using PRAME-Specific T Cells in Medulloblastoma. *Cancer Res* **78**, 3337–3349 (2018).

179. Srivastava, S. *et al.* Immunogenic Chemotherapy Enhances Recruitment of CAR-T Cells to Lung Tumors and Improves Antitumor Efficacy when Combined with Checkpoint Blockade. *Cancer Cell* **39**, 193-208.e10 (2021).

180. Li, T. & Wang, J. Therapeutic effect of dual CAR-T targeting PDL1 and MUC16 antigens on ovarian cancer cells in mice. *BMC Cancer* **20**, 678 (2020).

181. Streppel, M. M. *et al.* Mucin 16 (cancer antigen 125) expression in human tissues and cell lines and correlation with clinical outcome in adenocarcinomas of the pancreas, esophagus, stomach, and colon,. *Hum Pathol* **43**, 1755–1763 (2012).

182. Burgos-Panadero, R., Noguera, I., Cañete, A., Navarro, S. & Noguera, R. Vitronectin as a molecular player of the tumor microenvironment in neuroblastoma. *BMC Cancer* **19**, 479 (2019).

183. Preissner, K. T. The role of vitronectin as multifunctional regulator in the hemostatic and immune systems. *Blut* **59**, 419–431 (1989).

184. Edwards, S., Lalor, P. F., Tuncer, C. & Adams, D. H. Vitronectin in human hepatic tumours contributes to the recruitment of lymphocytes in an alpha v beta3-independent manner. *Br J Cancer* **95**, 1545–1554 (2006).

185. Wu, H. *et al.* Single-cell RNA sequencing reveals diverse intratumoral heterogeneities and gene signatures of two types of esophageal cancers. *Cancer Letters* **438**, 133–143 (2018).

186. Thakkar, P. V. *et al.* Abstract 975: Single cell RNA sequencing of esophageal adenocarcinoma before and after chemotherapy alone and with pembrolizumab identifies novel tumor immune microenvironment alterations. *Cancer Res* **80**, 975–975 (2020).

187. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

188. Tran, H. T. N. *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biology* **21**, 12 (2020).

189. Tung, P.-Y. *et al.* Batch effects and the effective design of single-cell gene expression studies. *Sci Rep* **7**, (2017).

190. Denisenko, E. *et al.* Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol* **21**, (2020).

191. Slyper, M. *et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* **26**, 792–802 (2020).

192. Maaninka, K., Lappalainen, J. & Kovanen, P. T. Human mast cells arise from a common circulating progenitor. *Journal of Allergy and Clinical Immunology* **132**, 463-469.e3 (2013).

193. Kohno, D., Gao, H.-Z., Muroya, S., Kikuyama, S. & Yada, T. Ghrelin directly interacts with neuropeptide-Y-containing neurons in the rat arcuate nucleus: Ca2+ signaling via protein kinase A and N-type channel-dependent mechanisms and cross-talk with leptin and orexin. *Diabetes* **52**, 948–956 (2003).

194. Kouznetsova, I. *et al.* Biosynthesis of gastrokine-2 in the human gastric mucosa: restricted spatial expression along the antral gland axis and differential interaction with TFF1, TFF2 and mucins. *Cell Physiol Biochem* **20**, 899–908 (2007).

195. Sasaki, N. *et al.* Reg4 deep crypt secretory cells function as epithelial niche for Lgr5 stem cells in colon. *PNAS* **113**, E5399–E5407 (2016).

196. Malhi-Chowla, N., Ringley, R. & Wolfsen, H. Gastric Metaplasia of the Proximal Esophagus Associated with Esophageal Adenocarcinoma and Barrett's Esophagus: What Is the Connection? *DDI* **18**, 183–185 (2000).

197. Madissoon, E. *et al.* scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biology* **21**, 1 (2019).

198. Kim, H.-S., Kim, J. Y., Song, C. L., Jeong, J. E. & Cho, Y. S. Directly induced human Schwann cell precursors as a valuable source of Schwann cells. *Stem Cell Res Ther* **11**, 257 (2020).

199. Kameneva, P., Kastriti, M. E. & Adameyko, I. Neuronal lineages derived from the nerve-associated Schwann cell precursors. *Cell. Mol. Life Sci.* **78**, 513–529 (2021).

200. Nassiri, F. *et al.* Endoglin (CD105): A Review of its Role in Angiogenesis and Tumor Diagnosis, Progression and Therapy. *Anticancer Research* **31**, 2283–2290 (2011).

201. Wragg, J. W. *et al.* MCAM and LAMA4 are highly enriched in tumor blood vessels of renal cell carcinoma and predict patient outcome. *Cancer Res* **76**, 2314–2326 (2016).

202. Schupp, J. C. *et al.* Integrated Single Cell Atlas of Endothelial Cells of the Human Lung. *Circulation* **0**,.

203. Goveia, J. *et al.* An Integrated Gene Expression Landscape Profiling Approach to Identify Lung Tumor Endothelial Cell Heterogeneity and Angiogenic Candidates. *Cancer Cell* **37**, 21-36.e13 (2020).

204. Kumar, A. *et al.* Specification and Diversification of Pericytes and Smooth Muscle Cells from Mesenchymoangioblasts. *Cell Rep* **19**, 1902–1916 (2017).

205. Cohen, N. *et al.* Fibroblasts drive an immunosuppressive and growth-promoting microenvironment in breast cancer via secretion of Chitinase 3-like 1. *Oncogene* **36**, 4457–4468 (2017).

206. Keppler, D., Zhang, J., Bihani, T. & Lin, A. W. Novel Expression of CST1 as Candidate Senescence Marker. *The Journals of Gerontology: Series A* **66A**, 723–731 (2011).

207. Solé-Boldo, L. *et al.* Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. *Commun Biol* **3**, 1–12 (2020).

208. Korosec, A. *et al.* Lineage Identity and Location within the Dermis Determine the Function of Papillary and Reticular Fibroblasts in Human Skin. *Journal of Investigative Dermatology* **139**, 342–351 (2019).

209. Grässel, S., Unsöld, C., Schäcke, H., Bruckner-Tuderman, L. & Bruckner, P. Collagen XVI is expressed by human dermal fibroblasts and keratinocytes and is associated with the microfibrillar apparatus in the upper papillary dermis. *Matrix Biol* **18**, 309–317 (1999).

210. Benias, P. C. *et al.* Structure and Distribution of an Unrecognized Interstitium in Human Tissues. *Sci Rep* **8**, 4947 (2018).

211. Alcolea, M. P. Oesophageal Stem Cells and Cancer. *Adv Exp Med Biol* **1041**, 187–206 (2017).

212. Hosaka, K. *et al.* Pericyte-fibroblast transition promotes tumor growth and metastasis. *Proc Natl Acad Sci U S A* **113**, E5618-5627 (2016).

213. Shami, A. N. *et al.* Single-Cell RNA Sequencing of Human, Macaque, and Mouse Testes Uncovers Conserved and Divergent Features of Mammalian Spermatogenesis. *Developmental Cell* **54**, 529-547.e12 (2020).

214. Danopoulos, S., Bhattacharya, S., Mariani, T. J. & Alam, D. A. Transcriptional Characterization of Human Lung Cells Identifies Novel Mesenchymal Lineage Markers. *Eur Respir J* **55**, 1900746 (2020).

215. Lee, S. J., Blanchett-Anderson, S., Keep, S. G., Gasche, M. B. & Wang, M. M. Tripartite factors leading to molecular divergence between human and murine smooth muscle. *PLoS One* **15**, e0227672 (2020).

216. Barrena, S. *et al.* Aberrant expression of tetraspanin molecules in B-cell chronic lymphoproliferative disorders and its correlation with normal B-cell maturation. *Leukemia* **19**, 1376–1383 (2005).

217. Shi, G.-X., Harrison, K., Wilson, G. L., Moratz, C. & Kehrl, J. H. RGS13 regulates germinal center B lymphocytes responsiveness to CXC chemokine ligand (CXCL)12 and CXCL13. *J Immunol* **169**, 2507–2515 (2002).

218. Krzyzak, L. *et al.* CD83 Modulates B Cell Activation and Germinal Center Responses. *The Journal of Immunology* **196**, 3581–3594 (2016).

219. Allen, C. D. C. *et al.* Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. *Nat Immunol* **5**, 943–952 (2004).

220. Park, S.-R. Activation-induced Cytidine Deaminase in B Cell Immunity and Cancers. *Immune Netw* **12**, 230–239 (2012).

221. Todo, K., Koga, O., Nishikawa, M. & Hikida, M. Modulation of Igβ is essential for the B cell selection in germinal center. *Sci Rep* **5**, 10303 (2015).

222. Stavnezer, J., Guikema, J. E. J. & Schrader, C. E. Mechanism and Regulation of Class Switch Recombination. *Annu Rev Immunol* **26**, 261–292 (2008).

223. Foroutan, M. *et al.* The ratio of exhausted to resident infiltrating lymphocytes is prognostic for colorectal cancer patient outcome. *bioRxiv* 2020.12.19.423379 (2020) doi:10.1101/2020.12.19.423379.

224. Allan, D. S. *et al.* An in vitro model of innate lymphoid cell function and differentiation. *Mucosal Immunol* **8**, 340–351 (2015).

225. Zhong, C. *et al.* Group 3 innate lymphoid cells continuously require the transcription factor GATA3 after commitment. *Nat Immunol* **17**, 169–178 (2016).

226. Clarke, J. *et al.* Single-cell transcriptomic analysis of tissue-resident memory T cells in human lung cancer. *Journal of Experimental Medicine* **216**, 2128–2149 (2019).

227. Duhen, T. *et al.* Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors. *Nat Commun* **9**, 2724 (2018).

228. Chaix, J. *et al.* Cutting Edge: CXCR4 Is Critical for CD8+ Memory T Cell Homeostatic Self-Renewal but Not Rechallenge Self-Renewal. *The Journal of Immunology* **193**, 1013–1016 (2014).

229. Baaten, B., Tinoco, R., Chen, A. & Bradley, L. Regulation of Antigen-Experienced T Cells: Lessons from the Quintessential Memory Marker CD44. *Front. Immunol.* **3**, (2012).

230. Banerjee, A. *et al.* The Transcription Factor Eomesodermin Enables CD8+ T Cells to Compete for the Memory Cell Niche. *J Immunol* **185**, 4988–4992 (2010).

231. Gu-Trantien, C. *et al.* CXCL13-producing TFH cells link immune suppression and adaptive memory in human breast cancer. *JCI Insight* **2**, e91487.

232. Bjorkdahl, O. *et al.* Characterization of CC-chemokine receptor 7 expression on murine T cells in lymphoid tissues. *Immunology* **110**, 170–179 (2003).

233. Patente, T. A. *et al.* Human Dendritic Cells: Their Heterogeneity and Clinical Application Potential in Cancer Immunotherapy. *Front. Immunol.* **0**, (2019).

234. Collin, M. & Bigley, V. Human dendritic cell subsets: an update. *Immunology* **154**, 3–20 (2018).

235. Liao, X. *et al.* High Expression of LAMP3 Is a Novel Biomarker of Poor Prognosis in Patients with Esophageal Squamous Cell Carcinoma. *Int J Mol Sci* **16**, 17655–17667 (2015).

236. Wang, Y. *et al.* Fascin inhibitor increases intratumoral dendritic cell activation and anti-cancer immunity. *Cell Reports* **35**, 108948 (2021).

237. Wong, K. L. *et al.* The three human monocyte subsets: implications for health and disease. *Immunol Res* **53**, 41–57 (2012).

238. Chapuy, L. & Sarfati, M. Single-Cell Protein and RNA Expression Analysis of Mononuclear Phagocytes in Intestinal Mucosa and Mesenteric Lymph Nodes of Ulcerative Colitis and Crohn's Disease Patients. *Cells* **9**, 813 (2020).

239. Aras, S. & Zaidi, M. R. TAMeless traitors: macrophages in cancer progression and metastasis. *Br J Cancer* **117**, 1583–1591 (2017).

240. Leblond, M. M. *et al.* M2 macrophages are more resistant than M1 macrophages following radiation therapy in the context of glioblastoma. *Oncotarget* **8**, 72597–72612 (2017).

241. Sierra-Filardi, E. *et al.* Activin A skews macrophage polarization by promoting a proinflammatory phenotype and inhibiting the acquisition of anti-inflammatory macrophage markers. *Blood* **117**, 5092–5101 (2011).

242. Fujiwara, Y. *et al.* Guanylate-binding protein 5 is a marker of interferon-γ-induced classically activated macrophages. *Clin Transl Immunology* **5**, e111 (2016).

243. van Dierendonck, X. A. M. H. *et al.* HILPDA Uncouples Lipid Droplet Accumulation in Adipose Tissue Macrophages from Inflammation and Metabolic Dysregulation. *Cell Reports* **30**, 1811-1822.e6 (2020).

244. Casanova-Acebes, M. *et al.* Tissue-resident macrophages provide a pro-tumorigenic niche to early NSCLC cells. *Nature* **595**, 578–584 (2021).

245. Cole, S. L. *et al.* M1-like monocytes are a major immunological determinant of severity in previously healthy adults with life-threatening influenza. *JCI Insight* **2**, e91868.

246. Chu, T. & Danko, C. G. Bayesian cell-type deconvolution and gene expression inference reveals tumor-microenvironment interactions. *bioRxiv* 2020.01.07.897900 (2020) doi:10.1101/2020.01.07.897900.

247. Maag, J. L. V. *et al.* Novel Aberrations Uncovered in Barrett's Esophagus and Esophageal Adenocarcinoma Using Whole Transcriptome Sequencing. *Mol Cancer Res* **15**, 1558–1569 (2017).

248. Farnsworth, R. H., Karnezis, T., Maciburko, S. J., Mueller, S. N. & Stacker, S. A. The Interplay Between Lymphatic Vessels and Chemokines. *Front. Immunol.* **0**, (2019).

249. Zhang, X.-Y., Zhang, P.-Y. & Aboul-Soud, M. A. M. From inflammation to gastric cancer: Role of Helicobacter pylori. *Oncol Lett* **13**, 543–548 (2017).

250. Li, T. *et al.* IGFBP2 regulates PD-L1 expression by activating the EGFR-STAT3 signaling pathway in malignant melanoma. *Cancer Lett* **477**, 19–30 (2020).

251. Liang, L., Zeng, M., Pan, H., Liu, H. & He, Y. Nicotinamide N-methyltransferase promotes epithelial-mesenchymal transition in gastric cancer cells by activating transforming growth factor-β1 expression. *Oncol Lett* **15**, 4592–4598 (2018).

252. Huang, C. *et al.* Dermokine contributes to epithelial–mesenchymal transition through increased activation of signal transducer and activator of transcription 3 in pancreatic cancer. *Cancer Science* **108**, 2130–2141 (2017).

253. Tanaka, K. The proteasome: Overview of structure and functions. *Proc Jpn Acad Ser B Phys Biol Sci* **85**, 12–36 (2009).

254. Kubiniok, P. *et al.* Global Analysis of the Mammalian MHC class I Immunopeptidome at the Organism-Wide Scale. *bioRxiv* 2020.09.28.317750 (2021) doi:10.1101/2020.09.28.317750.

255. Johnson, D. B. *et al.* Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nat Commun* **7**, 10582 (2016).

256. Dunne, M. R. *et al.* HLA-DR expression in tumor epithelium is an independent prognostic indicator in esophageal adenocarcinoma patients. *Cancer Immunol Immunother* **66**, 841–850 (2017).

257. Yan, J., Jiang, Y., Lu, J., Wu, J. & Zhang, M. Inhibiting of Proliferation, Migration, and Invasion in Lung Cancer Induced by Silencing Interferon-Induced Transmembrane Protein 1 (IFITM1). *BioMed Research International* **2019**, e9085435 (2019).

258. O'Flanagan, C. H. *et al.* Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *Genome Biology* **20**, 210 (2019).

259. Petit, V. *et al.* Optimization of tumor xenograft dissociation for the profiling of cell surface markers and nutrient transporters. *Lab Invest* **93**, 611–621 (2013).

260. Ruddle, N. H. High Endothelial Venules and Lymphatic Vessels in Tertiary Lymphoid Organs: Characteristics, Functions, and Regulation. *Front. Immunol.* **0**, (2016).

261. Dieu-Nosjean, M.-C., Goc, J., Giraldo, N. A., Sautès-Fridman, C. & Fridman, W. H. Tertiary lymphoid structures in cancer and beyond. *Trends in Immunology* **35**, 571–580 (2014).

262. Pipi, E. *et al.* Tertiary Lymphoid Structures: Autoimmunity Goes Local. *Front Immunol* **9**, 1952 (2018).

263. Petitprez, F. *et al.* B cells are associated with survival and immunotherapy response in sarcoma. *Nature* **577**, 556–560 (2020).

264. Munoz-Erazo, L., Rhodes, J. L., Marion, V. C. & Kemp, R. A. Tertiary lymphoid structures in cancer – considerations for patient prognosis. *Cell Mol Immunol* **17**, 570–575 (2020).

265. Colbeck, E. J., Ager, A., Gallimore, A. & Jones, G. W. Tertiary Lymphoid Structures in Cancer: Drivers of Antitumor Immunity, Immunosuppression, or Bystander Sentinels in Disease? *Front. Immunol.* **0**, (2017).

266. Werner, F. *et al.* A Standardized Analysis of Tertiary Lymphoid Structures in Human Melanoma: Disease Progression- and Tumor Site-Associated Changes With Germinal Center Alteration. *Front. Immunol.* **0**, (2021).

267. Zaheen, A. & Martin, A. Activation-Induced Cytidine Deaminase and Aberrant Germinal Center Selection in the Development of Humoral Autoimmunities. *Am J Pathol* **178**, 462–471 (2011).

268. Victora, G. D. *et al.* Germinal Center Dynamics Revealed by Multiphoton Microscopy Using a Photoactivatable Fluorescent Reporter. *Cell* **143**, 592–605 (2010).

269. Victora, G. D. *et al.* Identification of human germinal center light and dark zone cells and their relationship to human B-cell lymphomas. *Blood* **120**, 2240–2248 (2012).

270. Goc, J., Fridman, W.-H., Hammond, S. A., Sautès-Fridman, C. & Dieu-Nosjean, M.-C. Tertiary lymphoid structures in human lung cancers, a new driver of antitumor immune responses. *Oncoimmunology* **3**, e28976 (2014).

271. Dai, Y. *et al.* Genome-Wide Analysis of Barrett's Adenocarcinoma. A First Step Towards Identifying Patients at Risk and Developing Therapeutic Paths. *Translational Oncology* **11**, 116–124 (2018).

272. Wang, J. *et al.* Differential Gene Expression in Normal Esophagus and Barrett's Esophagus. *J Gastroenterol* **44**, 10.1007/s00535-009-0082–2 (2009).

273. Gettinger, S. *et al.* Impaired HLA Class I Antigen Processing and Presentation as a Mechanism of Acquired Resistance to Immune Checkpoint Inhibitors in Lung Cancer. *Cancer Discov* **7**, 1420–1435 (2017).

274. Todd, I., Pujol-Borrell, R., Hammond, L. J., Bottazzo, G. F. & Feldmann, M. Interferon-gamma induces HLA-DR expression by thyroid epithelium. *Clin Exp Immunol* **61**, 265–273 (1985).

275. Lee, C. K. *et al.* Checkpoint Inhibitors in Metastatic EGFR-Mutated Non-Small Cell Lung Cancer-A Meta-Analysis. *J Thorac Oncol* **12**, 403–407 (2017).

276. Nukaga, S. *et al.* Amplification of EGFR Wild-Type Alleles in Non–Small Cell Lung Cancer Cells Confers Acquired Resistance to Mutation-Selective EGFR Tyrosine Kinase Inhibitors. *Cancer Res* **77**, 2078–2089 (2017).

277. Spring, L. M. *et al.* Pathologic Complete Response after Neoadjuvant Chemotherapy and Impact on Breast Cancer Recurrence and Survival: A Comprehensive Meta-analysis. *Clin Cancer Res* **26**, 2838–2848 (2020).

278. Stamp, L. A. *et al.* The GCTM-5 Epitope Associated with the Mucin-Like Glycoprotein FCGBP Marks Progenitor Cells in Tissues of Endodermal Origin. *Stem Cells* **30**, 1999–2009 (2012).

279. Zheng, Y. *et al.* Immune suppressive landscape in the human esophageal squamous cell carcinoma microenvironment. *Nat Commun* **11**, 6268 (2020).

280. Olingy, C. E., Dinh, H. Q. & Hedrick, C. C. Monocyte heterogeneity and functions in cancer. *J Leukoc Biol* **106**, 309–322 (2019).

281. Malaguarnera, L., Marsullo, A., Zorena, K., Musumeci, G. & Di Rosa, M. Vitamin D3 regulates LAMP3 expression in monocyte derived dendritic cells. *Cell Immunol* **311**, 13–21 (2017).

282. Johansson-Percival, A. & Ganss, R. Therapeutic Induction of Tertiary Lymphoid Structures in Cancer Through Stromal Remodeling. *Front. Immunol.* **0**, (2021).

283. Tellides, G. & Pober, J. S. Inflammatory and Immune Responses in the Arterial Media. *Circulation Research* **116**, 312–322 (2015).

284. Cheng, L. *et al.* Type I interferons suppress viral replication but contribute to T cell depletion and dysfunction during chronic HIV-1 infection. *JCI Insight* **2**, 94366 (2017).

285. Zhu, L. *et al.* Single-Cell Sequencing of Peripheral Mononuclear Cells Reveals Distinct Immune Response Landscapes of COVID-19 and Influenza Patients. *Immunity* **53**, 685-696.e3 (2020).

286. Liu, J. *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400-416.e11 (2018).

287. Shu, W. *et al.* Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development* **134**, 1991–2000 (2007).

288. Hu, G. & Christman, J. W. Editorial: Alveolar Macrophages in Lung Inflammation and Resolution. *Front. Immunol.* **0**, (2019).

289. Zhang, S.-K. *et al.* Prevalence of human papillomavirus 16 in esophageal cancer among the Chinese population: a systematic review and meta-analysis. *Asian Pac J Cancer Prev* **15**, 10143–10149 (2014).

290. Benamouzig, R. *et al.* Human papillomavirus infection in esophageal squamous-cell carcinoma in western countries. *Int J Cancer* **50**, 549–552 (1992).

291. Petrick, J. L. *et al.* Prevalence of human papillomavirus among oesophageal squamous cell carcinoma cases: systematic review and meta-analysis. *Br J Cancer* **110**, 2369–2377 (2014).

292. Nichols, A. J. *et al.* Combined Systemic and Intratumoral Administration of Human Papillomavirus Vaccine to Treat Multiple Cutaneous Basaloid Squamous Cell Carcinomas. *JAMA Dermatol* **154**, 927–930 (2018).

293. Nichols, A. J. *et al.* Systemic and intratumoral 9-valent human papillomavirus vaccine treatment for squamous cell carcinoma in situ in a renal transplant recipient. *JAAD Case Rep* **6**, 289–291 (2020).

294. Tellides, G. & Pober, J. S. Inflammatory and Immune Responses in the Arterial Media. *Circulation Research* **116**, 312–322 (2015).

295. Weiss, J., Theile, D., Ketabi-Kiyanvash, N., Lindenmaier, H. & Haefeli, W. E. Inhibition of MRP1/ABCC1, MRP2/ABCC2, and MRP3/ABCC3 by nucleoside, nucleotide, and non-nucleoside reverse transcriptase inhibitors. *Drug Metab Dispos* **35**, 340–344 (2007).

296. Baiceanu, E. *et al.* 2-Indolylmethylenebenzofuranones as first effective inhibitors of ABCC2. *European Journal of Medicinal Chemistry* **122**, 408–418 (2016).

297. Serafini, P. *et al.* Phosphodiesterase-5 inhibition augments endogenous antitumor immunity by reducing myeloid-derived suppressor cell function. *The Journal of Experimental Medicine* **203**, 2691 (2006).

298. Weed, D. T. *et al.* Tadalafil reduces myeloid-derived suppressor cells and regulatory T cells and promotes tumor immunity in patients with head and neck squamous cell carcinoma. *Clin Cancer Res* **21**, 39–48 (2015).

299. Huang, W., Sundquist, J., Sundquist, K. & Ji, J. Phosphodiesterase-5 inhibitors use and risk for mortality and metastases among male patients with colorectal cancer. *Nat Commun* **11**, 3191 (2020).

300. Rehman, H., Silk, A. W., Kane, M. P. & Kaufman, H. L. Into the clinic: Talimogene laherparepvec (T-VEC), a first-in-class intratumoral oncolytic viral therapy. *J Immunother Cancer* **4**, 53 (2016).

301. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol* **39**, 313–319 (2021).

302. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat Methods* **18**, 9–14 (2021).

303. Wang, M., Ji, X., Wang, B., Li, Q. & Zhou, J. Simultaneous Evaluation of the Preservative Effect of RNAlater on Different Tissues by Biomolecular and Histological Analysis. *Biopreservation and Biobanking* **16**, 426–433 (2018).

304. Wu, Y. *et al.* Repeated cycles of 5-fluorouracil chemotherapy impaired anti-tumor functions of cytotoxic T cells in a CT26 tumor-bearing mouse model. *BMC Immunology* **17**, 29 (2016).

305. Fitzgerald, R. C. *et al.* Cytosponge-trefoil factor 3 versus usual care to identify Barrett's oesophagus in a primary care setting: a multicentre, pragmatic, randomised controlled trial. *Lancet* **396**, 333–344 (2020).

APPENDICES

Appendix A: LUD2015-005 Eligibility Criteria[p]

Synopsis of inclusion criteria

1. Histological diagnosis of oesophageal or gastro-oesophageal cancer.

   -Inoperable Cohorts (A & B): metastatic or locally advanced cancer

   -Operable Cohorts (C, C-FLOT & D): deemed suitable for surgery with curative intent
2. Recovered from prior therapy (persistent Grade 1 AEs are acceptable)
3. Anticipated lifespan greater than 4 months
4. Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1
5. Subjects with brain metastases must be asymptomatic for at least 4 weeks
6. Adequate normal organ and marrow function, with laboratory parameters for vital functions in the normal range
7. Written informed consent must be obtained prior to performing any protocol-related procedures, including protocol-specific screening evaluations
8. 18 years or older of age
9. Must have been informed of other treatment options
10. Willing and able to comply with the protocol for the duration of the study

Synopsis of exclusion criteria

1. Involvement in the planning and/or conduct of the study, or previous enrolment in the present study
2. Prior treatment in another clinical trial with an IMP within the last 4 weeks
3. Prior or concurrent systemic full-dose anti-cancer therapy
4. Mean QT interval corrected for heart rate (QTc) ≥ 470ms calculated from 3 ECGs using Fredericia's Correction

[p] This summary of eligibility criteria is not my work, and is taken from the publicly available study schema (*https://www.oncology.ox.ac.uk/clinical-trials/oncology-clinical-trials-office-octo/current-trials/lud2015-005*) with only slight modifications to match spelling and nomenclature used in this thesis.

5. Active or prior documented autoimmune disease within the past 2 years (excluding vitiligo, Grave's disease or psoriasis that has not required systemic treatment in past 2 years)
6. Active or prior documented inflammatory bowel disease
7. History of allogeneic organ transplant
8. Uncontrolled intercurrent illness
9. Known history of previous clinical diagnosis of tuberculosis
10. Prior exposure to tremelimumab/durvalumab or checkpoint inhibitors
11. History of severe allergic reactions to any unknown allergens or any components of the study drugs.
12. Known dihydropyrimidine dehydrogenase (DPD) deficiency
13. Treatment with sorivudine or chemically related analogues
14. Peripheral sensitive neuropathy with functional impairment prior to first course
15. History of sarcoidosis syndrome
16. Metastatic disease to CNS for which other therapeutic options may be available, including radiotherapy
17. History of pneumonitis or interstitial lung disease
18. Major surgical procedure within 30 days of Day 1, or if still recovering from prior surgery.
19. Women who are breast feeding or pregnant. Female subjects of childbearing potential and male subjects must who are sexually active with a female of childbearing potential must use at least one highly effective means of contraception. All subjects should be aware of the requirement to continue using highly effective means of contraception for several months after finishing study treatment.
20. Any condition that, in the clinical judgment of the treating physician, is likely to interfere with the study.
21. Subjects should not donate blood while participating in this study or for at least 90 days following the last infusion of durvalumab or 6 months following last tremelimumab or until the time specified in the prescribing information of oxaliplatin, capecitabine,

paclitaxel, carboplatin, 5-fluorouracil (5-FU), leucovorin or docetaxel, whichever is longest

22. For oxaliplatin, capecitabine, paclitaxel, carboplatin, and FLOT (5-fluorouracil (5-FU), leucovorin, oxaliplatin, and docetaxel): refer to prescribing information for additional information

Appendix B: Details of Supplementary Operable OAC Patients[q]

Three patients with operable OAC were included in the LUD2015-005 cell atlas. All patients in operable cohorts of the LUD2015-005 trial received a similar ICI regimen to the αPD-L1 ICI arm of the inoperable cohort, with 750mg of Q2W durvalumab administered intravenously. On-study αPD-L1 treatment could continue for up to 12 months for these operable patients. While these patients were treated with different agents, all followed a similar neoadjuvant treatment protocol consisting of an initial ICI-only window followed by ICI in combination with chemotherapy or chemoradiotherapy. Then, following a rest period with no treatment, surgical resection occurred, after which adjuvant αPD-L1 doses could be given. While long-term outcome data are not yet available, the following is a summary of the currently available outcome information and treatment details for the patients profiled by scRNA-seq in this thesis from this operable cohort:

OAC.Op-ODHL

- OAC.Op-ODHL had stage II cancer at the time of study enrolment.
- Following an ICI-only window, this patient received FLOT (5-fluorouracil (5-FU), leucovorin, oxaliplatin, and docetaxel) along with continued αPD-L1.
- A response of irNN was recorded at the time of surgery. No adjuvant αPD-L1 or post-surgery determinations of response were reported.

OAC.Op-JJHD

- OAC.Op-JJHD had stage III cancer at the time of study enrolment.
- Following an ICI-only window, this patient received carboplatin, paclitaxel, and radiotherapy along with continued αPD-L1.
- A response of irNN was recorded at the time of surgery. This patient received additional adjuvant αPD-L1 doses following surgery. At a scan following the last dose of on-study αPD-L1, the response was determined to be irCR.

OAC.Op-IKIJ

- OAC.Op-IKIJ had stage III cancer at the time of study enrolment.

[q] Collection and processing of outcome data for these patients was conducted by OCTO and LCR.

- Following an ICI-only window, this patient received carboplatin, paclitaxel, and radiotherapy along with continued αPD-L1, the same regimen as OAC.Op-JJHD.
- A response of irNN was recorded at the time of surgery. This patient also received adjuvant αPD-L1 doses following surgery. At a scan following the last dose of on-study αPD-L1, the response was determined to be irCR.

Appendix C: Significant Genes from Key Differential Expression Tests

For these tables, all significant (FDR < 0.1) genes are shown, with the exception of table C-1, where only the 1000 most significant differentially expressed genes are shown due to space considerations (13,632 met the FDR cut-off of 0.1 for the analysis generating C-1; see Section 4.4.2). Column names follow DESeq2 output conventions following fold change moderation, such that baseMean refers to average expression across all samples, log2FC refers to the fold change with moderation using the 'ashr' method (log2 scale), pvalue is the p-value produced by the Wald test, and padj is the FDR-corrected p-value. Short gene descriptions obtained from the mygene package are also included. For all tables, log2FoldChanges are shown for the first group listed in the section title relative to the second (i.e. for Appendix C-1, positive log2FoldChanges are upregulated in CB with respect to NCB, while negative ones are lower in CB). For Appendix C-6, INCITE genes are listed in bold.

Appendix C-1: OAC vs. Normal Gastrointestinal Tissues (PreTx)

Gene	baseMean	log2FC	pvalue	padj	description
CYP3A4	5865.43	-10.307	1.52E-53	4.33E-49	cytochrome P450 family 3 subfamily A member 4
HOXB7	116.67	4.979	1.45E-43	2.06E-39	homeobox B7
SERPINE1	489.98	4.993	6.36E-42	6.03E-38	serpin family E member 1
ASAH2	2351.29	-7.090	9.16E-42	6.51E-38	N-acylsphingosine amidohydrolase 2
MAGEA4	14.15	22.126	1.15E-41	6.55E-38	MAGE family member A4
LCT	15221.60	-9.796	2.64E-41	1.25E-37	lactase
HOXB6	257.47	5.189	6.07E-40	2.47E-36	homeobox B6
MMP9	291.67	5.141	5.28E-39	1.88E-35	matrix metallopeptidase 9
CXCL8	1276.78	6.489	3.96E-38	1.25E-34	C-X-C motif chemokine ligand 8
MIR4435-2HG	983.75	2.782	1.82E-37	5.18E-34	MIR4435-2 host gene
HOXB9	101.70	7.198	3.58E-37	9.26E-34	homeobox B9
MMP12	589.16	6.864	8.27E-37	1.96E-33	matrix metallopeptidase 12
CXCL1	913.52	5.343	8.04E-36	1.76E-32	C-X-C motif chemokine ligand 1
SPP1	985.82	6.587	8.15E-35	1.66E-31	secreted phosphoprotein 1
SOCS3	1770.08	4.723	1.72E-34	3.25E-31	suppressor of cytokine signaling 3
MEP1B	2991.87	-9.528	2.14E-34	3.80E-31	meprin A subunit beta
TUBB3	286.93	4.441	2.33E-34	3.90E-31	tubulin beta 3 class III
FOLH1	3192.41	-6.641	1.56E-32	2.47E-29	folate hydrolase 1
C6orf58	683.40	-8.460	3.48E-32	5.21E-29	chromosome 6 open reading frame 58
ADA	1926.58	-4.887	8.16E-32	1.16E-28	adenosine deaminase
FCGR3A	261.02	4.357	3.07E-31	4.15E-28	Fc fragment of IgG receptor IIIa
ANGPT2	243.85	3.901	4.27E-31	5.51E-28	angiopoietin 2
HOXA13	189.62	7.533	4.59E-31	5.68E-28	homeobox A13
DEFA5	2465.34	-13.790	1.32E-30	1.57E-27	defensin alpha 5
TIMP1	1194.34	3.842	1.42E-30	1.62E-27	TIMP metallopeptidase inhibitor 1
TMEM236	731.45	-6.137	1.67E-30	1.82E-27	transmembrane protein 236
SALL4	238.84	5.270	6.13E-30	6.45E-27	spalt like transcription factor 4
ENPP3	621.53	-6.865	6.83E-30	6.93E-27	ectonucleotide pyrophosphatase/phosphodiesterase 3
FKBP10	1509.58	4.402	3.91E-29	3.83E-26	FKBP prolyl isomerase 10
MTHFD1L	547.94	2.155	9.65E-29	9.15E-26	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1 like
PRRX2	49.60	4.263	1.15E-28	1.06E-25	paired related homeobox 2
OSR2	48.55	4.743	2.09E-28	1.85E-25	odd-skipped related transciption factor 2
LUCAT1	354.68	5.463	3.37E-28	2.90E-25	lung cancer associated transcript 1
ALDOB	20839.49	-7.903	4.06E-28	3.40E-25	aldolase, fructose-bisphosphate B
SLC2A5	1301.94	-5.567	1.05E-27	8.54E-25	solute carrier family 2 member 5
FPR1	222.46	5.474	1.40E-27	1.11E-24	formyl peptide receptor 1
PTGS2	979.63	5.301	2.29E-27	1.76E-24	prostaglandin-endoperoxide synthase 2
ADAMTS4	220.13	3.709	2.68E-27	1.95E-24	ADAM metallopeptidase with thrombospondin type 1 motif 4
PRAME	122.08	6.870	2.64E-27	1.95E-24	PRAME nuclear receptor transcriptional regulator
ACADSB	1351.33	-2.636	2.92E-27	2.08E-24	acyl-CoA dehydrogenase short/branched chain
TMPRSS15	16734.52	-10.015	3.70E-27	2.57E-24	transmembrane serine protease 15

ALDH6A1	1464.48	-2.127	3.81E-27	2.58E-24	aldehyde dehydrogenase 6 family member A1
ESM1	151.21	5.040	3.94E-27	2.60E-24	endothelial cell specific molecule 1
LAMC2	3367.73	3.421	5.54E-27	3.58E-24	laminin subunit gamma 2
SAA1	266.94	6.549	7.22E-27	4.56E-24	serum amyloid A1
MFAP2	85.67	3.590	7.86E-27	4.86E-24	microfibril associated protein 2
CRADD	229.66	-1.257	8.57E-27	5.19E-24	CASP2 and RIPK1 domain containing adaptor with death domain
MMP3	1081.79	6.797	1.64E-26	9.72E-24	matrix metallopeptidase 3
LINC02577	44.86	6.086	1.90E-26	1.10E-23	long intergenic non-protein coding RNA 2577
SERPINH1	1525.41	2.198	3.16E-26	1.80E-23	serpin family H member 1
HOXA10	163.07	5.993	3.62E-26	2.02E-23	homeobox A10
LRFN4	325.61	2.761	4.20E-26	2.29E-23	leucine rich repeat and fibronectin type III domain containing 4
ACSL6	164.09	-4.585	5.10E-26	2.73E-23	acyl-CoA synthetase long chain family member 6
HOXC10	138.77	6.129	5.50E-26	2.84E-23	homeobox C10
IFITM3	1782.72	2.535	5.46E-26	2.84E-23	interferon induced transmembrane protein 3
ICAM1	1341.85	3.545	7.12E-26	3.62E-23	intercellular adhesion molecule 1
FOSB	582.29	5.032	9.28E-26	4.63E-23	FosB proto-oncogene, AP-1 transcription factor subunit
AC010198.2	115.68	4.404	1.02E-25	4.98E-23	NA
HOXA11	45.07	7.410	1.13E-25	5.33E-23	homeobox A11
UGT2B17	1484.57	-8.621	1.12E-25	5.33E-23	UDP glucuronosyltransferase family 2 member B17
PLA2G7	140.67	3.930	1.17E-25	5.45E-23	phospholipase A2 group VII
FEZF1-AS1	104.02	7.131	1.21E-25	5.53E-23	FEZF1 antisense RNA 1
AQP9	189.52	5.804	1.28E-25	5.76E-23	aquaporin 9
SLC6A14	1401.07	5.201	1.40E-25	6.22E-23	solute carrier family 6 member 14
PLEKHG2	687.14	2.271	1.46E-25	6.37E-23	pleckstrin homology and RhoGEF domain containing G2
ECT2	1890.67	2.247	1.63E-25	7.03E-23	epithelial cell transforming 2
SAA2	252.43	5.561	1.76E-25	7.45E-23	serum amyloid A2
FAP	118.57	3.401	2.20E-25	9.19E-23	fibroblast activation protein alpha
SNAI1	45.19	3.416	2.59E-25	1.07E-22	snail family transcriptional repressor 1
POTEF	194.74	6.933	2.87E-25	1.17E-22	POTE ankyrin domain family member F
LPCAT1	929.20	2.451	3.16E-25	1.27E-22	lysophosphatidylcholine acyltransferase 1
SLC11A1	355.87	4.348	4.23E-25	1.67E-22	solute carrier family 11 member 1
GRIN2D	194.82	3.988	4.70E-25	1.83E-22	glutamate ionotropic receptor NMDA type subunit 2D
TREM1	287.46	6.018	5.63E-25	2.16E-22	triggering receptor expressed on myeloid cells 1
CEBPB	426.27	2.148	5.86E-25	2.22E-22	CCAAT enhancer binding protein beta
TEAD4	246.90	2.462	9.05E-25	3.39E-22	TEA domain transcription factor 4
COCH	204.65	5.034	1.11E-24	4.12E-22	cochlin
HOTAIR	41.69	6.711	1.41E-24	5.13E-22	HOX transcript antisense RNA
FOXN3	3057.12	-1.416	1.52E-24	5.49E-22	forkhead box N3
FCGR3B	266.54	6.200	1.60E-24	5.69E-22	Fc fragment of IgG receptor IIIb
COL4A1	7048.00	3.315	2.05E-24	7.19E-22	collagen type IV alpha 1 chain
STRIP2	254.60	2.613	2.85E-24	9.88E-22	striatin interacting protein 2
MMP7	522.68	6.955	3.82E-24	1.31E-21	matrix metallopeptidase 7
SLC26A3	6716.41	-8.651	5.32E-24	1.80E-21	solute carrier family 26 member 3
CXCL2	400.00	4.021	5.40E-24	1.81E-21	C-X-C motif chemokine ligand 2

ZG16	370.74	-8.063	6.70E-24	2.21E-21	zymogen granule protein 16
AC103702.2	60.39	5.147	8.43E-24	2.72E-21	NA
SMIM25	93.65	4.417	8.35E-24	2.72E-21	plaque enriched lncRNA in atherosclerotic and inflammatory bowel macrophage regulation
COL25A1	230.25	-6.259	8.69E-24	2.77E-21	collagen type XXV alpha 1 chain
SELL	245.28	4.531	8.76E-24	2.77E-21	selectin L
ETFDH	1318.21	-1.895	1.02E-23	3.20E-21	electron transfer flavoprotein dehydrogenase
HOXB5	61.95	4.357	1.18E-23	3.63E-21	homeobox B5
AQP10	851.52	-7.598	1.51E-23	4.58E-21	aquaporin 10
TRPM2-AS	43.41	4.955	1.51E-23	4.58E-21	TRPM2 antisense RNA
OSMR	1610.84	2.652	1.63E-23	4.89E-21	oncostatin M receptor
FAR2P1	91.67	8.262	1.76E-23	5.20E-21	fatty acyl-CoA reductase 2 pseudogene 1
SLC51A	179.00	-5.694	3.28E-23	9.62E-21	solute carrier family 51 subunit alpha
CDKN2A	263.34	3.162	3.36E-23	9.76E-21	cyclin dependent kinase inhibitor 2A
CD36	2106.51	-4.364	4.40E-23	1.26E-20	CD36 molecule
LINC01270	70.13	4.488	4.59E-23	1.31E-20	long intergenic non-protein coding RNA 1270
OSM	261.95	6.165	6.89E-23	1.94E-20	oncostatin M
TREH	526.37	-5.062	7.02E-23	1.96E-20	trehalase
ACAN	119.05	5.480	8.06E-23	2.23E-20	aggrecan
ADH4	1235.14	-7.605	9.08E-23	2.48E-20	alcohol dehydrogenase 4 (class II), pi polypeptide
SPINDOC	217.29	1.656	9.96E-23	2.70E-20	spindlin interactor and repressor of chromatin binding
LINC01234	97.99	6.604	1.17E-22	3.13E-20	long intergenic non-protein coding RNA 1234
APOBEC3A	196.61	5.115	1.32E-22	3.52E-20	apolipoprotein B mRNA editing enzyme catalytic subunit 3A
HOXC11	53.35	4.901	1.70E-22	4.48E-20	homeobox C11
COL12A1	4238.15	3.367	3.10E-22	8.05E-20	collagen type XII alpha 1 chain
S100G	447.22	-11.106	3.11E-22	8.05E-20	S100 calcium binding protein G
CD300E	115.29	5.909	5.17E-22	1.32E-19	CD300e molecule
BGN	586.14	3.126	6.74E-22	1.71E-19	biglycan
FCGR2A	716.57	3.916	6.82E-22	1.72E-19	Fc fragment of IgG receptor IIa
CGAS	343.21	1.801	7.45E-22	1.86E-19	cyclic GMP-AMP synthase
LILRB3	82.85	4.038	1.06E-21	2.61E-19	leukocyte immunoglobulin like receptor B3
ZNF695	157.44	2.750	1.33E-21	3.26E-19	zinc finger protein 695
SLC46A1	4382.29	-4.548	1.37E-21	3.33E-19	solute carrier family 46 member 1
MS4A10	944.85	-9.610	1.63E-21	3.92E-19	membrane spanning 4-domains A10
FPR2	93.24	6.096	1.67E-21	3.98E-19	formyl peptide receptor 2
AC012363.2	56.31	3.341	2.01E-21	4.77E-19	NA
HOXC8	37.90	4.849	2.22E-21	5.21E-19	homeobox C8
EPOP	139.94	3.735	2.47E-21	5.76E-19	elongin BC and polycomb repressive complex 2 associated protein
CSF3R	779.82	4.472	3.51E-21	8.13E-19	colony stimulating factor 3 receptor
SLC25A34	239.26	-3.359	3.86E-21	8.86E-19	solute carrier family 25 member 34
COLGALT1	1395.13	1.223	4.13E-21	9.38E-19	collagen beta(1-O)galactosyltransferase 1
IL1B	1924.49	5.083	5.45E-21	1.23E-18	interleukin 1 beta
IL24	53.56	7.771	6.58E-21	1.47E-18	interleukin 24
PGA5	2142.31	-9.527	6.70E-21	1.49E-18	pepsinogen A5
HOXC9	63.33	4.680	7.80E-21	1.72E-18	homeobox C9

ACOX2	305.83	-3.873	9.62E-21	2.10E-18	acyl-CoA oxidase 2
STRA6	131.02	5.039	9.69E-21	2.10E-18	signaling receptor and transporter of retinol STRA6
KHK	623.02	-3.962	1.08E-20	2.32E-18	ketohexokinase
LAMC1	4263.20	2.277	1.15E-20	2.46E-18	laminin subunit gamma 1
LILRA6	50.95	4.765	1.16E-20	2.47E-18	leukocyte immunoglobulin like receptor A6
TMED6	146.19	-4.086	1.69E-20	3.56E-18	transmembrane p24 trafficking protein 6
AC125807.2	98.20	2.322	1.77E-20	3.71E-18	NA
WNT2	26.48	6.003	1.84E-20	3.82E-18	Wnt family member 2
CD14	312.94	2.778	2.00E-20	4.12E-18	CD14 molecule
CREB5	210.49	3.097	2.44E-20	4.99E-18	cAMP responsive element binding protein 5
EGR3	116.86	3.784	2.86E-20	5.82E-18	early growth response 3
POTEE	118.29	6.364	3.01E-20	6.08E-18	POTE ankyrin domain family member E
ITGA5	1072.73	2.620	3.08E-20	6.17E-18	integrin subunit alpha 5
THY1	448.47	3.658	3.57E-20	7.09E-18	Thy-1 cell surface antigen
MAGEA6	141.23	10.713	4.09E-20	8.08E-18	MAGE family member A6
TMEM253	299.61	-4.412	4.58E-20	8.98E-18	transmembrane protein 253
STRIT1	49.41	-6.939	5.02E-20	9.77E-18	small transmembrane regulator of ion transport 1
AURKA	805.02	2.057	5.88E-20	1.14E-17	aurora kinase A
AFAP1L1	249.71	2.289	6.12E-20	1.18E-17	actin filament associated protein 1 like 1
CIDEB	1094.33	-3.438	7.62E-20	1.45E-17	cell death inducing DFFA like effector b
C17orf78	1365.90	-9.768	8.54E-20	1.62E-17	chromosome 17 open reading frame 78
FXYD5	956.95	2.632	8.60E-20	1.62E-17	FXYD domain containing ion transport regulator 5
DUXAP8	101.02	4.327	1.02E-19	1.91E-17	double homeobox A pseudogene 8
HOXA11-AS	17.53	6.209	1.17E-19	2.17E-17	HOXA11 antisense RNA
PDLIM7	428.95	1.659	1.24E-19	2.29E-17	PDZ and LIM domain 7
CST1	98.38	7.461	1.36E-19	2.49E-17	cystatin SN
FSTL3	213.83	2.666	1.67E-19	3.05E-17	follistatin like 3
SDS	55.38	4.084	1.73E-19	3.13E-17	serine dehydratase
APOA4	23621.52	-8.761	1.82E-19	3.27E-17	apolipoprotein A4
RGMB	1017.79	-2.055	1.95E-19	3.49E-17	repulsive guidance molecule BMP co-receptor b
PCK2	2276.39	-3.280	2.00E-19	3.56E-17	phosphoenolpyruvate carboxykinase 2, mitochondrial
THBS1	3064.53	2.818	2.14E-19	3.78E-17	thrombospondin 1
TNFAIP6	58.08	4.615	2.16E-19	3.80E-17	TNF alpha induced protein 6
LINC02163	38.61	7.196	2.22E-19	3.87E-17	long intergenic non-protein coding RNA 2163
HOTTIP	53.56	6.498	2.61E-19	4.52E-17	HOXA distal transcript antisense RNA
EMILIN2	313.74	2.609	2.81E-19	4.84E-17	elastin microfibril interfacer 2
SI	20709.47	-7.187	2.83E-19	4.85E-17	sucrase-isomaltase
ABCB1	534.08	-3.315	2.89E-19	4.92E-17	ATP binding cassette subfamily B member 1
GPR4	45.46	2.843	3.00E-19	5.08E-17	G protein-coupled receptor 4
PCSK2	28.73	-5.254	3.13E-19	5.27E-17	proprotein convertase subtilisin/kexin type 2
LINC01285	26.88	3.407	3.21E-19	5.37E-17	long intergenic non-protein coding RNA 1285
PGRMC2	1324.03	-1.526	3.70E-19	6.15E-17	progesterone receptor membrane component 2
MYBL2	891.06	2.468	3.86E-19	6.38E-17	MYB proto-oncogene like 2
AC005993.1	38.36	5.073	4.68E-19	7.69E-17	NA
MME	8576.43	-5.605	4.76E-19	7.78E-17	membrane metalloendopeptidase

AL117329.1	265.14	5.471	4.81E-19	7.82E-17	uncharacterized LOC284930
SLC5A4	230.28	-4.616	4.87E-19	7.88E-17	solute carrier family 5 member 4
IL13RA2	35.44	5.507	5.30E-19	8.47E-17	interleukin 13 receptor subunit alpha 2
OLFML2B	124.73	3.327	5.30E-19	8.47E-17	olfactomedin like 2B
MARVELD1	222.98	1.795	8.36E-19	1.33E-16	MARVEL domain containing 1
LINC01833	77.02	5.431	9.64E-19	1.52E-16	long intergenic non-protein coding RNA 1833
CABYR	102.99	3.693	1.00E-18	1.57E-16	calcium binding tyrosine phosphorylation regulated
RBM38	407.31	1.523	1.04E-18	1.63E-16	RNA binding motif protein 38
IL6	96.44	6.310	1.10E-18	1.71E-16	interleukin 6
SH3PXD2B	881.70	2.003	1.18E-18	1.82E-16	SH3 and PX domains 2B
VNN3	95.60	4.489	1.23E-18	1.89E-16	vanin 3
CCSER2	1918.41	-1.363	1.60E-18	2.45E-16	coiled-coil serine rich protein 2
HOXA9	60.04	4.969	1.63E-18	2.47E-16	homeobox A9
NAT8B	85.85	-5.764	2.15E-18	3.25E-16	N-acetyltransferase 8B (putative, gene/pseudogene)
ITGAX	609.22	3.237	2.18E-18	3.28E-16	integrin subunit alpha X
CDC25B	1440.62	1.971	2.37E-18	3.55E-16	cell division cycle 25B
SEMA6D	813.59	-4.011	2.47E-18	3.68E-16	semaphorin 6D
TRIP13	370.68	2.052	2.56E-18	3.80E-16	thyroid hormone receptor interactor 13
LAMC3	121.08	2.737	2.61E-18	3.84E-16	laminin subunit gamma 3
STC2	132.29	3.769	2.69E-18	3.94E-16	stanniocalcin 2
E2F1	281.15	1.694	2.71E-18	3.95E-16	E2F transcription factor 1
CTSL	934.14	2.611	2.88E-18	4.17E-16	cathepsin L
AC012531.1	57.63	4.168	2.91E-18	4.18E-16	NA
ALPI	2941.94	-7.146	2.91E-18	4.18E-16	alkaline phosphatase, intestinal
BAAT	123.96	5.417	3.36E-18	4.80E-16	bile acid-CoA:amino acid N-acyltransferase
BID	464.18	1.687	3.57E-18	5.00E-16	BH3 interacting domain death agonist
DNMT3B	296.50	2.502	3.57E-18	5.00E-16	DNA methyltransferase 3 beta
MT1G	1065.35	-4.027	3.53E-18	5.00E-16	metallothionein 1G
NFE2	42.83	4.549	3.57E-18	5.00E-16	nuclear factor, erythroid 2
MSX2	82.33	2.817	3.88E-18	5.40E-16	msh homeobox 2
CCL3	135.51	4.599	4.02E-18	5.57E-16	C-C motif chemokine ligand 3
FCGR1A	25.24	3.506	4.10E-18	5.65E-16	Fc fragment of IgG receptor Ia
EPHX2	1270.86	-2.491	5.16E-18	7.09E-16	epoxide hydrolase 2
KDM8	223.58	-2.295	6.75E-18	9.23E-16	lysine demethylase 8
SAR1B	2479.13	-1.427	6.80E-18	9.25E-16	secretion associated Ras related GTPase 1B
ADGRE2	257.72	2.736	7.23E-18	9.79E-16	adhesion G protein-coupled receptor E2
SERPINA3	1564.85	4.799	7.42E-18	1.00E-15	serpin family A member 3
SCUBE2	161.87	-3.125	7.48E-18	1.00E-15	signal peptide, CUB domain and EGF like domain containing 2
CILP2	20.38	5.333	7.74E-18	1.03E-15	cartilage intermediate layer protein 2
CLCF1	92.91	2.264	8.12E-18	1.08E-15	cardiotrophin like cytokine factor 1
AUNIP	103.60	2.283	8.17E-18	1.08E-15	aurora kinase A and ninein interacting protein
APOC3	3570.28	-8.332	8.56E-18	1.13E-15	apolipoprotein C3
BMS1P1	763.64	-2.028	8.66E-18	1.13E-15	BMS1 pseudogene 1
APOC1	149.07	3.336	9.20E-18	1.20E-15	apolipoprotein C1
OAT	5369.94	-2.591	9.30E-18	1.21E-15	ornithine aminotransferase

CHIA	426.34	-10.043	9.63E-18	1.24E-15	chitinase acidic
MST1	479.11	-3.020	9.91E-18	1.28E-15	macrophage stimulating 1
BTD	1329.68	-2.208	9.96E-18	1.28E-15	biotinidase
ABHD6	504.79	-2.112	1.00E-17	1.28E-15	abhydrolase domain containing 6, acylglycerol lipase
ACE2	3509.95	-4.583	1.03E-17	1.31E-15	angiotensin converting enzyme 2
AC007128.2	43.05	4.809	1.14E-17	1.44E-15	NA
MEFV	107.29	3.851	1.24E-17	1.56E-15	MEFV innate immuity regulator, pyrin
PHYHIPL	189.05	-7.000	1.27E-17	1.59E-15	phytanoyl-CoA 2-hydroxylase interacting protein like
CT83	46.48	6.538	1.36E-17	1.70E-15	cancer/testis antigen 83
ZNF540	66.94	-1.820	1.41E-17	1.76E-15	zinc finger protein 540
CLCA1	2040.88	-7.428	1.53E-17	1.89E-15	chloride channel accessory 1
SH2D5	35.03	3.889	1.65E-17	2.03E-15	SH2 domain containing 5
ADCY3	667.16	1.376	1.78E-17	2.19E-15	adenylate cyclase 3
INHBA	343.68	4.156	1.84E-17	2.25E-15	inhibin subunit beta A
AC068491.3	28.36	2.995	2.05E-17	2.49E-15	NA
HOXC-AS3	11.28	5.207	2.16E-17	2.61E-15	HOXC cluster antisense RNA 3
C1orf112	392.76	1.331	2.24E-17	2.70E-15	chromosome 1 open reading frame 112
MCAM	1039.41	2.561	2.33E-17	2.79E-15	melanoma cell adhesion molecule
C1orf216	129.69	1.889	2.58E-17	3.07E-15	chromosome 1 open reading frame 216
HTR4	79.18	-5.062	2.58E-17	3.07E-15	5-hydroxytryptamine receptor 4
KIF14	658.28	1.856	2.75E-17	3.26E-15	kinesin family member 14
MYO1B	2786.51	1.831	3.36E-17	3.96E-15	myosin IB
GPRIN1	78.40	2.120	3.43E-17	4.03E-15	G protein regulated inducer of neurite outgrowth 1
CYP27B1	74.12	2.346	3.51E-17	4.11E-15	cytochrome P450 family 27 subfamily B member 1
MMUT	806.46	-1.015	3.59E-17	4.18E-15	methylmalonyl-CoA mutase
CCN2	488.77	2.411	3.86E-17	4.49E-15	cellular communication network factor 2
LIPF	31797.49	-7.254	3.93E-17	4.55E-15	lipase F, gastric type
LINC02086	60.54	4.570	4.00E-17	4.60E-15	long intergenic non-protein coding RNA 2086
C8G	141.94	-3.661	5.26E-17	6.04E-15	complement C8 gamma chain
CRYL1	958.54	-2.206	5.51E-17	6.29E-15	crystallin lambda 1
MAGEA12	81.66	9.797	5.53E-17	6.30E-15	MAGE family member A12
FOXD1	26.56	3.637	5.56E-17	6.30E-15	forkhead box D1
BUB1	929.98	1.930	5.75E-17	6.49E-15	BUB1 mitotic checkpoint serine/threonine kinase
AL360169.3	467.72	-7.224	5.95E-17	6.68E-15	NA
UNC5D	184.48	-5.471	5.98E-17	6.70E-15	unc-5 netrin receptor D
CXCL3	343.98	3.912	6.01E-17	6.70E-15	C-X-C motif chemokine ligand 3
DHRS11	1328.94	-2.864	6.03E-17	6.70E-15	dehydrogenase/reductase 11
ACO2	3065.87	-1.531	6.88E-17	7.61E-15	aconitase 2
TWNK	256.98	1.303	7.45E-17	8.21E-15	twinkle mtDNA helicase
AQP11	221.63	-3.988	7.68E-17	8.40E-15	aquaporin 11
COL4A2	5133.33	2.476	7.67E-17	8.40E-15	collagen type IV alpha 2 chain
AC098614.2	174.41	-3.411	8.02E-17	8.74E-15	NA
EGR2	46.68	3.040	8.12E-17	8.81E-15	early growth response 2
AC051619.2	29.64	-4.860	8.17E-17	8.83E-15	NA

HEBP1	639.74	-2.587	9.74E-17	1.05E-14	heme binding protein 1
DGKI	43.48	1.906	9.96E-17	1.07E-14	diacylglycerol kinase iota
ZC3H6	550.82	-1.254	1.04E-16	1.12E-14	zinc finger CCCH-type containing 6
IER5L	187.81	1.790	1.08E-16	1.15E-14	immediate early response 5 like
KLHDC1	81.79	-1.724	1.08E-16	1.15E-14	kelch domain containing 1
AC026369.2	11.72	4.365	1.11E-16	1.17E-14	NA
CLSPN	443.89	1.644	1.11E-16	1.17E-14	claspin
UBE2C	650.49	2.276	1.21E-16	1.27E-14	ubiquitin conjugating enzyme E2 C
B4GALNT1	67.87	4.333	1.51E-16	1.58E-14	beta-1,4-N-acetyl-galactosaminyltransferase 1
IL11	52.25	4.126	1.56E-16	1.63E-14	interleukin 11
CXCR1	99.04	4.756	1.62E-16	1.68E-14	C-X-C motif chemokine receptor 1
FCAR	53.44	5.733	1.83E-16	1.89E-14	Fc fragment of IgA receptor
LILRA5	54.29	4.468	1.83E-16	1.89E-14	leukocyte immunoglobulin like receptor A5
SLC2A12	829.78	-3.042	1.86E-16	1.90E-14	solute carrier family 2 member 12
TTC36	32.81	-2.907	1.88E-16	1.92E-14	tetratricopeptide repeat domain 36
CYP4F27P	23.89	6.633	1.92E-16	1.96E-14	cytochrome P450 family 4 subfamily F member 27, pseudogene
SEMA6B	140.05	2.591	1.97E-16	2.00E-14	semaphorin 6B
RIPK2	343.60	1.688	2.08E-16	2.11E-14	receptor interacting serine/threonine kinase 2
CIP2A	501.12	1.686	2.22E-16	2.24E-14	cellular inhibitor of PP2A
MMP11	253.16	3.443	2.34E-16	2.35E-14	matrix metallopeptidase 11
PLEKHD1	41.90	3.682	2.41E-16	2.41E-14	pleckstrin homology and coiled-coil domain containing D1
TREM2	34.69	3.431	2.52E-16	2.52E-14	triggering receptor expressed on myeloid cells 2
FAR2P4	68.55	7.870	2.95E-16	2.93E-14	uncharacterized LOC440910
MIR3945HG	35.85	5.594	2.99E-16	2.96E-14	MIR3945 host gene
TMEM158	40.91	3.110	3.29E-16	3.25E-14	transmembrane protein 158
STIL	775.63	1.422	3.34E-16	3.29E-14	STIL centriolar assembly protein
MTTP	4511.14	-6.763	3.45E-16	3.38E-14	microsomal triglyceride transfer protein
HOXB13	53.85	5.206	3.78E-16	3.70E-14	homeobox B13
PSMC3IP	107.93	1.477	3.88E-16	3.78E-14	PSMC3 interacting protein
AK3	1776.85	-1.280	4.15E-16	4.02E-14	adenylate kinase 3
MIR196A1	11.80	5.224	4.14E-16	4.02E-14	microRNA 196a-1
LIMK1	615.94	1.375	4.18E-16	4.03E-14	LIM domain kinase 1
ACTL8	174.46	7.707	4.28E-16	4.11E-14	actin like 8
CD177	112.39	4.543	4.35E-16	4.17E-14	CD177 molecule
MEP1A	2860.40	-5.921	4.45E-16	4.24E-14	meprin A subunit alpha
AC009501.3	60.29	4.975	5.04E-16	4.79E-14	NA
COL1A1	6851.64	2.772	5.14E-16	4.86E-14	collagen type I alpha 1 chain
GBA3	842.67	-6.262	5.14E-16	4.86E-14	glucosylceramidase beta 3 (gene/pseudogene)
GABRD	10.78	3.626	5.46E-16	5.14E-14	gamma-aminobutyric acid type A receptor subunit delta
CCK	128.35	-5.309	5.67E-16	5.32E-14	cholecystokinin
MGAM	5040.67	-4.818	6.12E-16	5.73E-14	maltase-glucoamylase
C11orf54	897.44	-1.191	6.18E-16	5.76E-14	chromosome 11 open reading frame 54
PI4K2B	1001.57	-1.402	6.31E-16	5.86E-14	phosphatidylinositol 4-kinase type 2 beta
FANCI	1345.45	1.421	6.57E-16	6.08E-14	FA complementation group I
KIF3C	137.24	2.053	6.78E-16	6.26E-14	kinesin family member 3C

DUXAP9	25.90	5.138	7.10E-16	6.53E-14	double homeobox A pseudogene 9
UGT2B7	511.36	-5.496	7.23E-16	6.64E-14	UDP glucuronosyltransferase family 2 member B7
CDC6	929.67	2.501	7.50E-16	6.86E-14	cell division cycle 6
ANK1	177.07	3.205	7.57E-16	6.90E-14	ankyrin 1
NT5DC4	19.09	5.214	7.60E-16	6.91E-14	5'-nucleotidase domain containing 4
ANK3	2546.82	-1.685	7.79E-16	7.05E-14	ankyrin 3
DYSF	657.22	2.614	9.00E-16	8.12E-14	dysferlin
MYOT	13.18	-5.030	9.22E-16	8.30E-14	myotilin
RNFT2	205.47	2.762	9.40E-16	8.43E-14	ring finger protein, transmembrane 2
BCAT1	576.70	3.536	9.54E-16	8.51E-14	branched chain amino acid transaminase 1
TKFC	1351.42	-2.198	9.54E-16	8.51E-14	triokinase and FMN cyclase
FOSL1	186.09	3.161	1.03E-15	9.12E-14	FOS like 1, AP-1 transcription factor subunit
CYP2R1	446.58	-0.964	1.04E-15	9.18E-14	cytochrome P450 family 2 subfamily R member 1
SLC5A9	782.62	-5.259	1.04E-15	9.22E-14	solute carrier family 5 member 9
HOXC13	25.97	4.932	1.06E-15	9.29E-14	homeobox C13
PDGFB	186.26	1.844	1.07E-15	9.36E-14	platelet derived growth factor subunit B
SPAG5	931.74	1.522	1.08E-15	9.48E-14	sperm associated antigen 5
ADAM12	116.28	2.768	1.10E-15	9.62E-14	ADAM metallopeptidase domain 12
ZFPM2-AS1	129.17	4.262	1.12E-15	9.70E-14	ZFPM2 antisense RNA 1
MAOA	2911.31	-2.043	1.16E-15	1.01E-13	monoamine oxidase A
RARRES1	269.63	3.960	1.20E-15	1.04E-13	retinoic acid receptor responder 1
TGIF2	428.89	1.584	1.23E-15	1.06E-13	TGFB induced factor homeobox 2
ACADS	741.90	-1.973	1.27E-15	1.09E-13	acyl-CoA dehydrogenase short chain
NPY6R	196.35	-7.330	1.30E-15	1.12E-13	neuropeptide Y receptor Y6 (pseudogene)
HTRA1	550.66	2.337	1.31E-15	1.12E-13	HtrA serine peptidase 1
S100A7A	333.91	8.761	1.32E-15	1.12E-13	S100 calcium binding protein A7A
FSTL1	1832.82	2.287	1.36E-15	1.15E-13	follistatin like 1
PBLD	2068.11	-3.403	1.36E-15	1.15E-13	phenazine biosynthesis like protein domain containing
CLEC4E	29.14	4.747	1.46E-15	1.23E-13	C-type lectin domain family 4 member E
PGA3	3735.08	-8.570	1.57E-15	1.32E-13	pepsinogen A3
KLHDC7B	35.80	3.097	1.67E-15	1.40E-13	kelch domain containing 7B
MMP25	132.02	3.036	1.68E-15	1.41E-13	matrix metallopeptidase 25
CCL18	105.29	4.482	1.72E-15	1.44E-13	C-C motif chemokine ligand 18
PGF	127.23	2.527	1.78E-15	1.48E-13	placental growth factor
MCEMP1	12.40	5.031	1.79E-15	1.49E-13	mast cell expressed membrane protein 1
CDCA5	389.14	1.947	1.86E-15	1.54E-13	cell division cycle associated 5
MUC5B	6481.92	5.179	1.87E-15	1.54E-13	mucin 5B, oligomeric mucus/gel-forming
LEAP2	100.17	-3.117	1.90E-15	1.56E-13	liver enriched antimicrobial peptide 2
PRKG2	349.45	-3.470	1.92E-15	1.57E-13	protein kinase cGMP-dependent 2
ASPA	264.71	-4.209	1.93E-15	1.58E-13	aspartoacylase
C5AR1	218.89	3.078	2.10E-15	1.71E-13	complement C5a receptor 1
DMBX1	17.78	4.656	2.14E-15	1.74E-13	diencephalon/mesencephalon homeobox 1
PPM1A	2020.77	-1.093	2.16E-15	1.75E-13	protein phosphatase, Mg2+/Mn2+ dependent 1A
CENPO	574.68	1.282	2.29E-15	1.85E-13	centromere protein O
RRP12	756.93	1.131	2.33E-15	1.87E-13	ribosomal RNA processing 12 homolog

IL19	49.81	4.592	2.38E-15	1.91E-13	interleukin 19
ANO10	1074.54	-1.625	2.46E-15	1.97E-13	anoctamin 10
COL5A2	2249.15	2.447	2.55E-15	2.04E-13	collagen type V alpha 2 chain
SAA4	30.03	5.040	2.63E-15	2.10E-13	serum amyloid A4, constitutive
GPR3	9.74	4.075	2.64E-15	2.10E-13	G protein-coupled receptor 3
AGO2	4004.71	0.935	2.79E-15	2.21E-13	argonaute RISC catalytic component 2
DCAF11	2014.72	-1.056	2.81E-15	2.21E-13	DDB1 and CUL4 associated factor 11
NAT2	119.38	-5.520	2.80E-15	2.21E-13	N-acetyltransferase 2
HADHB	2981.90	-1.433	2.95E-15	2.32E-13	hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit beta
SCML4	181.51	-3.063	3.01E-15	2.36E-13	Scm polycomb group protein like 4
IDNK	153.78	-1.637	3.06E-15	2.39E-13	IDNK gluconokinase
SPDL1	395.01	1.228	3.18E-15	2.48E-13	spindle apparatus coiled-coil protein 1
CEP55	617.41	1.939	3.26E-15	2.53E-13	centrosomal protein 55
CTHRC1	117.55	3.597	3.33E-15	2.57E-13	collagen triple helix repeat containing 1
SLC4A7	5654.49	-3.214	3.33E-15	2.57E-13	solute carrier family 4 member 7
TIMELESS	770.62	1.220	3.42E-15	2.64E-13	timeless circadian regulator
RPS14P4	24.38	2.561	3.51E-15	2.70E-13	ribosomal protein S14 pseudogene 4
C16orf89	227.56	-5.883	3.53E-15	2.71E-13	chromosome 16 open reading frame 89
FN1	8852.81	3.707	3.73E-15	2.85E-13	fibronectin 1
MTARC2	629.86	-2.334	3.94E-15	3.01E-13	mitochondrial amidoxime reducing component 2
GLT1D1	50.00	3.789	4.02E-15	3.06E-13	glycosyltransferase 1 domain containing 1
FOXJ1	130.76	4.634	4.11E-15	3.12E-13	forkhead box J1
KIF18A	380.47	1.874	4.17E-15	3.16E-13	kinesin family member 18A
MTHFD2	1237.11	1.462	4.26E-15	3.21E-13	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase
C3	1958.43	3.707	4.35E-15	3.27E-13	complement C3
DAB1	336.88	-4.892	4.34E-15	3.27E-13	DAB adaptor protein 1
TMEM132E	44.66	-4.132	4.37E-15	3.27E-13	transmembrane protein 132E
SELENOP	6172.68	-2.757	4.40E-15	3.28E-13	selenoprotein P
LGALS1	640.85	2.233	4.57E-15	3.40E-13	galectin 1
C4BPA	81.44	6.043	4.69E-15	3.48E-13	complement component 4 binding protein alpha
MNDA	201.83	3.374	4.70E-15	3.48E-13	myeloid cell nuclear differentiation antigen
NNMT	216.47	3.072	4.87E-15	3.60E-13	nicotinamide N-methyltransferase
RELT	317.14	1.900	4.96E-15	3.65E-13	RELT TNF receptor
RBP2	4418.04	-7.344	5.12E-15	3.76E-13	retinol binding protein 2
ZER1	1126.28	-1.200	5.13E-15	3.76E-13	zyg-11 related cell cycle regulator
ITGAM	207.92	2.764	5.38E-15	3.93E-13	integrin subunit alpha M
MAFTRR	23.01	-3.072	5.45E-15	3.96E-13	MAF transcriptional regulator RNA
MAML3	666.01	-1.210	5.45E-15	3.96E-13	mastermind like transcriptional coactivator 3
ULBP1	27.18	3.697	5.51E-15	4.00E-13	UL16 binding protein 1
AC243654.3	54.74	-8.498	5.59E-15	4.05E-13	NA
GINS4	256.77	1.714	5.75E-15	4.15E-13	GINS complex subunit 4
ACO1	1557.06	-1.380	5.79E-15	4.17E-13	aconitase 1
VWA1	1069.89	1.917	5.86E-15	4.21E-13	von Willebrand factor A domain containing 1
NCF2	454.91	2.953	5.90E-15	4.23E-13	neutrophil cytosolic factor 2
FEZF1	31.12	6.646	5.94E-15	4.24E-13	FEZ family zinc finger 1

NMNAT2	122.12	4.107	5.96E-15	4.24E-13	nicotinamide nucleotide adenylyltransferase 2
PGM2L1	708.76	2.296	5.95E-15	4.24E-13	phosphoglucomutase 2 like 1
FCN1	98.00	3.839	6.07E-15	4.30E-13	ficolin 1
FOXM1	998.42	1.953	6.30E-15	4.46E-13	forkhead box M1
SLC28A1	425.05	-5.941	6.57E-15	4.64E-13	solute carrier family 28 member 1
CASC19	1849.66	3.172	6.83E-15	4.81E-13	prostate cancer associated transcript 2
PLBD2	681.28	1.059	6.93E-15	4.87E-13	phospholipase B domain containing 2
PNO1	409.35	1.417	6.96E-15	4.88E-13	partner of NOB1 homolog
C3orf85	834.36	-7.168	7.14E-15	4.99E-13	chromosome 3 open reading frame 85
AP005262.1	27.46	6.241	7.28E-15	5.06E-13	NA
RGS19	158.01	1.511	7.28E-15	5.06E-13	regulator of G protein signaling 19
ACADM	1670.00	-1.854	7.44E-15	5.16E-13	acyl-CoA dehydrogenase medium chain
TGFB2	129.14	2.561	7.70E-15	5.33E-13	transforming growth factor beta 2
ACAA1	1487.94	-1.567	8.04E-15	5.55E-13	acetyl-CoA acyltransferase 1
WDR66	125.17	1.957	8.73E-15	6.01E-13	cilia and flagella associated protein 251
MMP13	11.13	4.661	8.76E-15	6.02E-13	matrix metallopeptidase 13
HAPLN3	135.19	2.084	9.01E-15	6.17E-13	hyaluronan and proteoglycan link protein 3
TMEM25	443.25	-2.519	9.09E-15	6.21E-13	transmembrane protein 25
HACL1	774.90	-2.015	9.56E-15	6.52E-13	2-hydroxyacyl-CoA lyase 1
CPA2	286.61	-6.253	9.97E-15	6.78E-13	carboxypeptidase A2
CEP170	880.51	1.419	1.01E-14	6.84E-13	centrosomal protein 170
GPR176	106.93	2.907	1.01E-14	6.84E-13	G protein-coupled receptor 176
OLR1	103.43	3.783	1.02E-14	6.87E-13	oxidized low density lipoprotein receptor 1
AC106865.1	13.59	5.184	1.06E-14	7.13E-13	toll like receptor 2 pseudogene
GAL3ST2	22.60	4.815	1.08E-14	7.26E-13	galactose-3-O-sulfotransferase 2
ATAD2	1908.85	1.421	1.09E-14	7.26E-13	ATPase family AAA domain containing 2
DSCC1	179.89	1.889	1.08E-14	7.26E-13	DNA replication and sister chromatid cohesion 1
CYBRD1	4243.77	-3.159	1.09E-14	7.29E-13	cytochrome b reductase 1
LINC01550	43.41	-2.099	1.10E-14	7.33E-13	long intergenic non-protein coding RNA 1550
IGF2BP1	374.37	5.729	1.13E-14	7.49E-13	insulin like growth factor 2 mRNA binding protein 1
SUCLG2	1692.53	-1.251	1.15E-14	7.65E-13	succinate-CoA ligase GDP-forming subunit beta
ODF2	1063.71	1.090	1.18E-14	7.81E-13	outer dense fiber of sperm tails 2
DNAH17-AS1	12.80	4.014	1.20E-14	7.91E-13	DNAH17 antisense RNA 1
LINC01060	8.98	5.005	1.20E-14	7.91E-13	long intergenic non-protein coding RNA 1060
RTCA-AS1	138.59	-1.472	1.21E-14	7.92E-13	RTCA antisense RNA 1
BRCA1	1149.88	1.634	1.26E-14	8.25E-13	BRCA1 DNA repair associated
L3MBTL4	248.37	-2.413	1.31E-14	8.57E-13	L3MBTL histone methyl-lysine binding protein 4
SLC38A7	390.42	0.758	1.33E-14	8.70E-13	solute carrier family 38 member 7
PTPRU	724.72	2.695	1.36E-14	8.84E-13	protein tyrosine phosphatase receptor type U
GGT4P	30.44	-5.493	1.37E-14	8.89E-13	gamma-glutamyltransferase 4 pseudogene
CTLA4	46.92	3.030	1.37E-14	8.90E-13	cytotoxic T-lymphocyte associated protein 4
GHRL	257.30	-4.208	1.40E-14	9.05E-13	ghrelin and obestatin prepropeptide
LILRB4	123.71	3.313	1.45E-14	9.37E-13	leukocyte immunoglobulin like receptor B4
BLM	459.27	1.759	1.48E-14	9.50E-13	BLM RecQ like helicase
FABP2	2187.75	-7.227	1.52E-14	9.77E-13	fatty acid binding protein 2

MCM10	389.71	1.915	1.53E-14	9.77E-13	minichromosome maintenance 10 replication initiation factor
SLC35G1	1090.02	-3.245	1.53E-14	9.77E-13	solute carrier family 35 member G1
TNFRSF11B	141.13	3.835	1.53E-14	9.77E-13	TNF receptor superfamily member 11b
AC073842.1	9.67	-2.533	1.57E-14	9.98E-13	NA
YTHDF1	1166.82	0.711	1.58E-14	1.00E-12	YTH N6-methyladenosine RNA binding protein 1
PLEK	1080.54	3.742	1.75E-14	1.11E-12	pleckstrin
LINC01595	19.92	-6.873	1.98E-14	1.25E-12	long intergenic non-protein coding RNA 1595
DHDH	102.72	-3.121	2.16E-14	1.36E-12	dihydrodiol dehydrogenase
CCN1	227.36	2.627	2.30E-14	1.44E-12	cellular communication network factor 1
COL8A1	275.01	3.270	2.36E-14	1.48E-12	collagen type VIII alpha 1 chain
HOXD9	13.80	2.645	2.38E-14	1.49E-12	homeobox D9
RPL5P30	12.82	-3.304	2.43E-14	1.52E-12	ribosomal protein L5 pseudogene 30
TUBAL3	169.19	-4.739	2.54E-14	1.58E-12	tubulin alpha like 3
C11orf95	306.05	1.308	2.56E-14	1.60E-12	zinc finger translocation associated
KIF4A	620.74	1.709	2.62E-14	1.63E-12	kinesin family member 4A
ADAMTS12	257.93	3.046	2.66E-14	1.65E-12	ADAM metallopeptidase with thrombospondin type 1 motif 12
HAPLN4	239.97	-5.147	2.78E-14	1.72E-12	hyaluronan and proteoglycan link protein 4
POTEKP	64.01	5.871	2.87E-14	1.77E-12	POTE ankyrin domain family member K, pseudogene
SCP2	2960.04	-1.293	2.88E-14	1.77E-12	sterol carrier protein 2
TPX2	1931.64	1.988	3.06E-14	1.88E-12	TPX2 microtubule nucleation factor
ORC6	229.46	1.391	3.09E-14	1.89E-12	origin recognition complex subunit 6
CPNE8	480.19	2.075	3.18E-14	1.94E-12	copine 8
BBOX1-AS1	95.27	3.025	3.19E-14	1.95E-12	BBOX1 antisense RNA 1
CD3EAP	249.38	1.650	3.30E-14	2.01E-12	RNA polymerase I subunit G
TESK2	338.41	-1.604	3.38E-14	2.06E-12	testis associated actin remodelling kinase 2
MAGEA3	184.85	8.399	3.41E-14	2.07E-12	MAGE family member A3
PPARA	2253.32	-1.452	3.43E-14	2.08E-12	peroxisome proliferator activated receptor alpha
SSX2IP	1521.82	-2.394	3.48E-14	2.10E-12	SSX family member 2 interacting protein
SLC2A3	670.74	3.235	3.62E-14	2.18E-12	solute carrier family 2 member 3
LCAL1	72.18	5.110	3.64E-14	2.19E-12	lung cancer associated lncRNA 1
AC090015.1	56.63	4.255	3.70E-14	2.22E-12	long intergenic non-protein coding RNA 2882
CALU	2576.47	1.384	3.76E-14	2.25E-12	calumenin
EGR1	741.11	2.461	3.77E-14	2.25E-12	early growth response 1
SLC23A1	161.99	-4.084	3.77E-14	2.25E-12	solute carrier family 23 member 1
GSEC	36.35	2.893	3.90E-14	2.32E-12	G-quadruplex forming sequence containing lncRNA
CYB5A	1488.65	-1.839	4.02E-14	2.39E-12	cytochrome b5 type A
SLC43A3	816.82	1.970	4.06E-14	2.41E-12	solute carrier family 43 member 3
MT1F	172.29	-2.822	4.31E-14	2.55E-12	metallothionein 1F
DEFA6	1164.03	-8.858	4.36E-14	2.57E-12	defensin alpha 6
NCAPH	466.37	1.712	4.39E-14	2.59E-12	non-SMC condensin I complex subunit H
SCRN1	1712.88	1.986	4.41E-14	2.59E-12	secernin 1
NFKB2	1015.32	1.293	4.59E-14	2.69E-12	nuclear factor kappa B subunit 2
SEPSECS	674.30	-1.373	4.61E-14	2.70E-12	Sep (O-phosphoserine) tRNA:Sec (selenocysteine) tRNA synthase

OTX1	66.28	2.468	4.82E-14	2.81E-12	orthodenticle homeobox 1
ALKBH7	293.62	-1.266	4.96E-14	2.89E-12	alkB homolog 7
CENPL	262.60	1.335	5.18E-14	3.01E-12	centromere protein L
APLN	148.49	3.491	5.20E-14	3.02E-12	apelin
STEAP1B	43.39	3.461	5.65E-14	3.27E-12	STEAP family member 1B
UBE2T	197.63	1.760	5.65E-14	3.27E-12	ubiquitin conjugating enzyme E2 T
IL17C	9.10	5.127	5.70E-14	3.29E-12	interleukin 17C
UGT1A1	568.73	-3.377	5.82E-14	3.35E-12	UDP glucuronosyltransferase family 1 member A1
CCL20	510.89	3.701	5.90E-14	3.39E-12	C-C motif chemokine ligand 20
PFDN2	180.37	1.123	5.95E-14	3.41E-12	prefoldin subunit 2
ZNF493	903.53	-1.091	5.96E-14	3.41E-12	zinc finger protein 493
PLXNA1	1990.99	1.264	6.05E-14	3.45E-12	plexin A1
FABP1	3099.06	-6.271	6.07E-14	3.46E-12	fatty acid binding protein 1
AMPD1	54.49	-3.607	6.34E-14	3.61E-12	adenosine monophosphate deaminase 1
CCNG1	2136.98	-1.499	6.47E-14	3.68E-12	cyclin G1
AGAP13P	36.51	-2.139	7.23E-14	4.10E-12	ArfGAP with GTPase domain, ankyrin repeat and PH domain 13, pseudogene
GULOP	20.77	-2.749	7.40E-14	4.18E-12	gulonolactone (L-) oxidase, pseudogene
IBSP	12.83	5.792	7.49E-14	4.23E-12	integrin binding sialoprotein
SLC30A10	1273.85	-5.582	7.90E-14	4.45E-12	solute carrier family 30 member 10
CELSR3	492.74	2.676	8.72E-14	4.90E-12	cadherin EGF LAG seven-pass G-type receptor 3
MINDY4	299.51	1.350	8.85E-14	4.97E-12	MINDY lysine 48 deubiquitinase 4
PMEPA1	925.48	2.362	9.68E-14	5.42E-12	prostate transmembrane protein, androgen induced 1
LYRM7	497.48	-0.901	9.83E-14	5.49E-12	LYR motif containing 7
CDC45	252.98	1.458	1.02E-13	5.67E-12	cell division cycle 45
RUVBL1	651.79	1.450	1.04E-13	5.81E-12	RuvB like AAA ATPase 1
REEP6	571.53	-3.074	1.09E-13	6.04E-12	receptor accessory protein 6
ENTPD5	2547.65	-1.971	1.09E-13	6.05E-12	ectonucleoside triphosphate diphosphohydrolase 5 (inactive)
RBKS	175.40	-2.051	1.11E-13	6.14E-12	ribokinase
AC008627.1	9.72	-5.379	1.14E-13	6.28E-12	NA
TGFBI	1542.49	2.329	1.14E-13	6.28E-12	transforming growth factor beta induced
CD160	32.79	-3.778	1.16E-13	6.36E-12	CD160 molecule
CKAP5	3814.60	1.080	1.16E-13	6.37E-12	cytoskeleton associated protein 5
GRAMD2B	624.04	-1.517	1.20E-13	6.59E-12	GRAM domain containing 2B
PLAU	884.88	2.463	1.22E-13	6.65E-12	plasminogen activator, urokinase
BAMBI	197.49	3.027	1.29E-13	7.05E-12	BMP and activin membrane bound inhibitor
BATF	77.71	2.740	1.33E-13	7.25E-12	basic leucine zipper ATF-like transcription factor
CASC9	81.78	3.500	1.35E-13	7.33E-12	cancer susceptibility 9
MT1H	209.40	-3.754	1.40E-13	7.57E-12	metallothionein 1H
LRP8	619.05	2.258	1.43E-13	7.76E-12	LDL receptor related protein 8
GLP2R	58.80	-4.198	1.44E-13	7.79E-12	glucagon like peptide 2 receptor
MS4A8	896.55	-4.977	1.46E-13	7.88E-12	membrane spanning 4-domains A8
CRAT	1116.36	-1.613	1.47E-13	7.91E-12	carnitine O-acetyltransferase
ASAH2B	370.23	-2.190	1.53E-13	8.24E-12	N-acylsphingosine amidohydrolase 2B
GAD1	100.41	3.912	1.54E-13	8.28E-12	glutamate decarboxylase 1

APOA1	9756.51	-6.022	1.56E-13	8.38E-12	apolipoprotein A1
PRKN	104.06	-1.968	1.62E-13	8.65E-12	parkin RBR E3 ubiquitin protein ligase
SLC25A51	448.43	-1.089	1.63E-13	8.71E-12	solute carrier family 25 member 51
NR4A3	88.38	2.766	1.66E-13	8.85E-12	nuclear receptor subfamily 4 group A member 3
KIF23	797.26	1.535	1.67E-13	8.88E-12	kinesin family member 23
MIA2	3646.23	-1.600	1.68E-13	8.93E-12	MIA SH3 domain ER export factor 2
SLC5A1	7048.12	-4.085	1.82E-13	9.66E-12	solute carrier family 5 member 1
AQP7	180.15	-3.455	1.85E-13	9.75E-12	aquaporin 7
SNHG15	404.15	1.309	1.85E-13	9.75E-12	small nucleolar RNA host gene 15
C14orf28	102.95	-1.219	1.92E-13	1.01E-11	chromosome 14 open reading frame 28
SLC30A4	621.97	-2.155	1.92E-13	1.01E-11	solute carrier family 30 member 4
ZIC5	24.87	5.947	1.93E-13	1.01E-11	Zic family member 5
SLC52A2	720.19	1.208	1.97E-13	1.03E-11	solute carrier family 52 member 2
PRSS51	92.22	4.492	2.00E-13	1.04E-11	serine protease 51
WDR62	333.79	1.617	2.04E-13	1.06E-11	WD repeat domain 62
GDF15	342.03	2.929	2.04E-13	1.06E-11	growth differentiation factor 15
CENPI	344.74	1.518	2.06E-13	1.07E-11	centromere protein I
EPB41L3	1842.58	-2.852	2.07E-13	1.08E-11	erythrocyte membrane protein band 4.1 like 3
CSAG1	66.96	7.308	2.09E-13	1.08E-11	chondrosarcoma associated gene 1
LOXL2	511.35	2.320	2.09E-13	1.08E-11	lysyl oxidase like 2
PROK2	54.34	5.044	2.13E-13	1.10E-11	prokineticin 2
ASPDH	18.06	-6.051	2.16E-13	1.11E-11	aspartate dehydrogenase domain containing
MTMR10	2117.06	-1.412	2.18E-13	1.12E-11	myotubularin related protein 10
CDKL1	333.89	-1.823	2.19E-13	1.12E-11	cyclin dependent kinase like 1
TBX15	21.69	4.080	2.22E-13	1.14E-11	T-box transcription factor 15
PRRX1	161.21	2.922	2.32E-13	1.18E-11	paired related homeobox 1
AC073365.1	23.58	5.616	2.36E-13	1.20E-11	NA
ACOX1	7879.23	-1.456	2.36E-13	1.20E-11	acyl-CoA oxidase 1
AC015917.2	22.09	-3.589	2.36E-13	1.20E-11	NA
CLDN15	988.06	-2.919	2.37E-13	1.20E-11	claudin 15
GRB14	112.20	2.478	2.41E-13	1.22E-11	growth factor receptor bound protein 14
ISYNA1	160.20	1.951	2.44E-13	1.23E-11	inositol-3-phosphate synthase 1
CCDC25	951.34	-1.321	2.55E-13	1.29E-11	coiled-coil domain containing 25
EPHB4	1800.62	1.506	2.61E-13	1.32E-11	EPH receptor B4
FAM89A	109.59	2.181	2.67E-13	1.34E-11	family with sequence similarity 89 member A
MIIP	308.57	1.019	2.73E-13	1.37E-11	migration and invasion inhibitory protein
HROB	124.45	1.546	2.77E-13	1.39E-11	homologous recombination factor with OB-fold
BOP1	661.87	1.498	2.79E-13	1.40E-11	BOP1 ribosomal biogenesis factor
C1QTNF1	253.78	2.874	2.81E-13	1.40E-11	C1q and TNF related 1
IFITM1	484.62	2.000	2.87E-13	1.43E-11	interferon induced transmembrane protein 1
APOE	544.18	3.047	2.91E-13	1.45E-11	apolipoprotein E
GCNT4	1664.26	-3.040	2.92E-13	1.45E-11	glucosaminyl (N-acetyl) transferase 4
HOXC12	28.53	6.081	2.94E-13	1.46E-11	homeobox C12
PCK1	1803.53	-5.425	2.94E-13	1.46E-11	phosphoenolpyruvate carboxykinase 1
MTFR2	97.99	1.760	2.99E-13	1.48E-11	mitochondrial fission regulator 2
COL10A1	46.70	4.797	3.02E-13	1.49E-11	collagen type X alpha 1 chain

TLR2	433.92	2.061	3.06E-13	1.51E-11	toll like receptor 2
TAPT1	526.45	-0.708	3.06E-13	1.51E-11	transmembrane anterior posterior transformation 1
DECR1	1119.92	-1.033	3.17E-13	1.55E-11	2,4-dienoyl-CoA reductase 1
PTGES	191.92	2.844	3.17E-13	1.55E-11	prostaglandin E synthase
BCL2A1	137.04	3.443	3.21E-13	1.57E-11	BCL2 related protein A1
SPARC	4435.63	2.404	3.21E-13	1.57E-11	secreted protein acidic and cysteine rich
HEYL	107.97	2.470	3.21E-13	1.57E-11	hes related family bHLH transcription factor with YRPW motif like
MUC16	214.55	4.004	3.22E-13	1.57E-11	mucin 16, cell surface associated
TDRD3	798.19	-0.810	3.29E-13	1.60E-11	tudor domain containing 3
NLRP12	19.57	3.211	3.30E-13	1.60E-11	NLR family pyrin domain containing 12
AC061992.2	13.78	3.988	3.40E-13	1.65E-11	SOCS3 divergent transcript
CSF3	19.69	4.689	3.67E-13	1.77E-11	colony stimulating factor 3
HSD17B14	36.65	1.779	3.66E-13	1.77E-11	hydroxysteroid 17-beta dehydrogenase 14
LINC00355	77.49	6.753	3.72E-13	1.79E-11	long intergenic non-protein coding RNA 355
SLC36A1	1722.94	-2.226	3.85E-13	1.85E-11	solute carrier family 36 member 1
TMEM161B-AS1	1996.64	-1.426	3.98E-13	1.91E-11	TMEM161B divergent transcript
SNX10	160.80	2.430	4.08E-13	1.96E-11	sorting nexin 10
DHRS4	203.82	-1.241	4.12E-13	1.97E-11	dehydrogenase/reductase 4
CD276	710.71	1.660	4.21E-13	2.01E-11	CD276 molecule
HAVCR1	89.17	5.334	4.31E-13	2.06E-11	hepatitis A virus cellular receptor 1
CD200R1L-AS1	13.74	4.590	4.40E-13	2.10E-11	CD200R1L antisense RNA 1
LINC02428	20.91	5.043	4.43E-13	2.11E-11	long intergenic non-protein coding RNA 2428
FBXO41	275.19	1.764	4.53E-13	2.15E-11	F-box protein 41
NPFFR1	33.04	3.346	4.55E-13	2.16E-11	neuropeptide FF receptor 1
URB2	629.32	1.218	4.90E-13	2.32E-11	URB2 ribosome biogenesis homolog
CCDC86	276.55	1.156	5.00E-13	2.36E-11	coiled-coil domain containing 86
FBXL17	569.41	-0.956	5.16E-13	2.44E-11	F-box and leucine rich repeat protein 17
PANX1	514.21	1.054	5.20E-13	2.45E-11	pannexin 1
MAPK15	40.27	2.329	5.28E-13	2.48E-11	mitogen-activated protein kinase 15
MAOB	656.93	-3.096	5.29E-13	2.48E-11	monoamine oxidase B
LITAF	1697.74	1.388	5.39E-13	2.52E-11	lipopolysaccharide induced TNF factor
AC037198.1	24.73	2.432	5.40E-13	2.52E-11	THBS1 intronic transcript 1
TCOF1	1209.32	0.909	5.52E-13	2.58E-11	treacle ribosome biogenesis factor 1
AGRN	3381.66	1.392	5.68E-13	2.65E-11	agrin
PDGFRB	965.94	2.212	5.68E-13	2.65E-11	platelet derived growth factor receptor beta
ZNF280A	12.36	5.452	5.76E-13	2.68E-11	zinc finger protein 280A
PEBP1	4459.78	-1.536	5.78E-13	2.68E-11	phosphatidylethanolamine binding protein 1
KPNA2	1972.91	1.805	5.96E-13	2.76E-11	karyopherin subunit alpha 2
LY6E	2030.91	2.162	6.04E-13	2.79E-11	lymphocyte antigen 6 family member E
CYTOR	437.64	1.843	6.12E-13	2.82E-11	cytoskeleton regulator RNA
EIF4E3	615.81	-1.424	6.17E-13	2.84E-11	eukaryotic translation initiation factor 4E family member 3
ANLN	2162.55	1.909	6.42E-13	2.96E-11	anillin actin binding protein
CCL3L1	52.52	4.583	6.61E-13	3.04E-11	C-C motif chemokine ligand 3 like 3
SMPD1	390.75	-1.152	6.63E-13	3.04E-11	sphingomyelin phosphodiesterase 1

DYRK3	287.00	1.340	6.76E-13	3.09E-11	dual specificity tyrosine phosphorylation regulated kinase 3
ABCG2	1150.58	-3.333	6.90E-13	3.15E-11	ATP binding cassette subfamily G member 2 (Junior blood group)
GPR55	22.66	-2.015	6.93E-13	3.16E-11	G protein-coupled receptor 55
HELLPAR	642.84	2.064	6.99E-13	3.18E-11	HELLP associated long non-coding RNA
RAD51	135.90	1.424	6.99E-13	3.18E-11	RAD51 recombinase
MYADM	988.60	2.280	7.02E-13	3.19E-11	myeloid associated differentiation marker
MTFR1L	799.08	-1.114	7.07E-13	3.20E-11	mitochondrial fission regulator 1 like
AC011365.2	86.21	-4.889	7.10E-13	3.21E-11	SLIT3 antisense RNA 2
PRC1	1311.58	1.557	7.10E-13	3.21E-11	protein regulator of cytokinesis 1
AL589743.1	57.89	3.679	7.36E-13	3.32E-11	double homeobox A pseudogene 10
CDC14A	458.68	-1.638	7.39E-13	3.33E-11	cell division cycle 14A
RGS13	35.20	-3.298	7.43E-13	3.34E-11	regulator of G protein signaling 13
RNVU1-31	59.67	2.792	7.65E-13	3.44E-11	RNA, variant U1 small nuclear 31
TNFRSF6B	147.98	2.613	7.71E-13	3.46E-11	TNF receptor superfamily member 6b
ABCD3	2711.23	-1.012	7.73E-13	3.46E-11	ATP binding cassette subfamily D member 3
KRT222	12.94	-3.608	8.04E-13	3.59E-11	keratin 222
AP005060.1	42.34	-2.449	8.17E-13	3.65E-11	NA
ETV4	471.77	2.452	8.28E-13	3.69E-11	ETS variant transcription factor 4
CD244	43.37	-1.781	8.49E-13	3.78E-11	CD244 molecule
TOM1L1	1890.17	-1.520	8.62E-13	3.83E-11	target of myb1 like 1 membrane trafficking protein
GIMD1	102.25	-7.777	8.64E-13	3.83E-11	GIMAP family P-loop NTPase domain containing 1
EXO1	335.76	1.726	8.68E-13	3.84E-11	exonuclease 1
DLX4	20.93	3.301	8.69E-13	3.84E-11	distal-less homeobox 4
PEX11A	336.12	-1.253	8.72E-13	3.85E-11	peroxisomal biogenesis factor 11 alpha
SHCBP1	308.48	1.562	9.26E-13	4.08E-11	SHC binding and spindle associated 1
DUSP10	222.73	1.975	9.28E-13	4.09E-11	dual specificity phosphatase 10
SLC2A7	25.62	-6.461	9.31E-13	4.09E-11	solute carrier family 2 member 7
IKBIP	170.96	1.428	9.33E-13	4.09E-11	IKBKB interacting protein
BIRC7	9.08	4.290	9.40E-13	4.11E-11	baculoviral IAP repeat containing 7
RGS16	95.75	1.990	9.39E-13	4.11E-11	regulator of G protein signaling 16
MMP19	137.77	3.019	9.46E-13	4.13E-11	matrix metallopeptidase 19
FZD2	40.53	2.262	9.58E-13	4.18E-11	frizzled class receptor 2
TRIM50	45.04	-4.230	9.59E-13	4.18E-11	tripartite motif containing 50
MEX3A	149.80	2.734	9.63E-13	4.19E-11	mex-3 RNA binding family member A
SLC39A1	1813.68	1.328	9.82E-13	4.26E-11	solute carrier family 39 member 1
TOP2A	6328.74	2.279	9.83E-13	4.26E-11	DNA topoisomerase II alpha
AP5M1	1472.56	-0.939	1.02E-12	4.40E-11	adaptor related protein complex 5 subunit mu 1
ABAT	1034.76	-2.136	1.02E-12	4.40E-11	4-aminobutyrate aminotransferase
SMDT1	175.88	-1.001	1.03E-12	4.43E-11	single-pass membrane protein with aspartate rich tail 1
NCOA4	8618.90	-1.553	1.04E-12	4.49E-11	nuclear receptor coactivator 4
FANCB	174.37	1.825	1.05E-12	4.52E-11	FA complementation group B
SEMG1	10.78	5.041	1.05E-12	4.53E-11	semenogelin 1
FOXP3	25.22	2.112	1.07E-12	4.59E-11	forkhead box P3
TNNI3	14.74	3.758	1.08E-12	4.64E-11	troponin I3, cardiac type

ADAM8	719.95	2.300	1.09E-12	4.68E-11	ADAM metallopeptidase domain 8
FBP1	1577.14	-3.152	1.11E-12	4.72E-11	fructose-bisphosphatase 1
S100A7	6941.11	6.838	1.12E-12	4.79E-11	S100 calcium binding protein A7
CADM2	34.17	-4.210	1.13E-12	4.83E-11	cell adhesion molecule 2
AC007406.3	9.49	3.082	1.14E-12	4.83E-11	NA
TSHZ1	628.20	-1.356	1.14E-12	4.83E-11	teashirt zinc finger homeobox 1
AC019069.1	17.63	2.636	1.14E-12	4.84E-11	HK2 divergent transcript
HJURP	522.62	1.618	1.15E-12	4.85E-11	Holliday junction recognition protein
CCKAR	65.85	-6.349	1.15E-12	4.87E-11	cholecystokinin A receptor
TRIM36	501.91	-2.927	1.19E-12	5.03E-11	tripartite motif containing 36
WNT9B	37.62	-3.556	1.21E-12	5.08E-11	Wnt family member 9B
ACTBL2	5.27	4.695	1.23E-12	5.19E-11	actin beta like 2
ODC1	1815.89	1.735	1.25E-12	5.23E-11	ornithine decarboxylase 1
HYI	348.30	-1.549	1.26E-12	5.27E-11	hydroxypyruvate isomerase (putative)
CCR9	21.01	-4.217	1.28E-12	5.34E-11	C-C motif chemokine receptor 9
AC105046.1	12.85	4.124	1.32E-12	5.50E-11	NA
AC090164.5	31.24	3.708	1.36E-12	5.66E-11	long intergenic non-protein coding RNA 2348
LMX1B	50.86	4.282	1.36E-12	5.66E-11	LIM homeobox transcription factor 1 beta
THEMIS2	359.16	1.841	1.36E-12	5.66E-11	thymocyte selection associated family member 2
ANKRD9	1039.00	-1.674	1.37E-12	5.70E-11	ankyrin repeat domain 9
CCN6	32.01	5.839	1.38E-12	5.72E-11	cellular communication network factor 6
TMEM65	1260.34	-1.975	1.40E-12	5.79E-11	transmembrane protein 65
SPC25	110.42	1.685	1.44E-12	5.95E-11	SPC25 component of NDC80 kinetochore complex
RRP9	255.24	1.187	1.45E-12	5.99E-11	ribosomal RNA processing 9, U3 small nucleolar RNA binding protein
KLHL34	16.89	-4.824	1.46E-12	6.03E-11	kelch like family member 34
TLR8	48.44	2.706	1.47E-12	6.04E-11	toll like receptor 8
P4HA3	16.82	2.771	1.57E-12	6.45E-11	prolyl 4-hydroxylase subunit alpha 3
U62317.1	10.23	3.127	1.57E-12	6.45E-11	KLHDC7B divergent transcript
UBD	11.78	4.526	1.57E-12	6.45E-11	ubiquitin D
RAD51AP1	268.34	1.705	1.58E-12	6.46E-11	RAD51 associated protein 1
P2RY6	93.23	2.470	1.61E-12	6.58E-11	pyrimidinergic receptor P2Y6
NES	534.73	1.847	1.63E-12	6.67E-11	nestin
SIRPB2	50.95	2.151	1.64E-12	6.70E-11	signal regulatory protein beta 2
GALNT18	183.70	1.860	1.65E-12	6.71E-11	polypeptide N-acetylgalactosaminyltransferase 18
BST2	721.98	2.122	1.65E-12	6.71E-11	bone marrow stromal cell antigen 2
SLC13A2	1022.66	-5.911	1.65E-12	6.72E-11	solute carrier family 13 member 2
MPP1	710.61	-2.665	1.69E-12	6.86E-11	membrane palmitoylated protein 1
SAT2	531.03	-1.684	1.69E-12	6.86E-11	spermidine/spermine N1-acetyltransferase family member 2
CHP2	885.39	-5.218	1.72E-12	6.96E-11	calcineurin like EF-hand protein 2
TPH1	126.32	-3.059	1.74E-12	7.04E-11	tryptophan hydroxylase 1
AC005050.3	17.22	5.831	1.76E-12	7.09E-11	NA
ARHGAP11A	1020.03	1.534	1.79E-12	7.22E-11	Rho GTPase activating protein 11A
TBC1D16	1566.87	1.211	1.80E-12	7.23E-11	TBC1 domain family member 16
XPNPEP2	1372.08	-5.790	1.80E-12	7.23E-11	X-prolyl aminopeptidase 2

GJC1	246.63	2.208	1.83E-12	7.35E-11	gap junction protein gamma 1
RGS1	206.09	2.537	1.84E-12	7.35E-11	regulator of G protein signaling 1
HOXA10-AS	14.83	4.695	1.85E-12	7.39E-11	HOXA10 antisense RNA
CCL4L2	78.63	4.239	1.86E-12	7.40E-11	C-C motif chemokine ligand 4 like 2
SLC39A10	599.21	1.410	1.85E-12	7.40E-11	solute carrier family 39 member 10
CCL2	149.33	2.638	1.89E-12	7.50E-11	C-C motif chemokine ligand 2
TBX18	114.09	3.851	1.88E-12	7.50E-11	T-box transcription factor 18
NOP2	959.58	1.346	1.91E-12	7.60E-11	NOP2 nucleolar protein
SLC9B2	246.52	1.860	1.96E-12	7.76E-11	solute carrier family 9 member B2
AC034199.1	9.89	5.045	1.99E-12	7.87E-11	uncharacterized LOC100128059
SLC22A5	860.89	-1.774	2.03E-12	8.04E-11	solute carrier family 22 member 5
LRRC19	405.79	-5.408	2.08E-12	8.21E-11	leucine rich repeat containing 19
SLC16A10	632.44	-3.223	2.09E-12	8.26E-11	solute carrier family 16 member 10
SLFN11	727.52	2.023	2.10E-12	8.27E-11	schlafen family member 11
DNAH2	151.37	3.136	2.11E-12	8.29E-11	dynein axonemal heavy chain 2
NUF2	311.66	1.715	2.12E-12	8.33E-11	NUF2 component of NDC80 kinetochore complex
SOWAHA	377.93	-3.303	2.13E-12	8.35E-11	sosondowah ankyrin repeat domain family member A
NKILA	28.33	2.591	2.19E-12	8.58E-11	NF-kappaB interacting lncRNA
PRR11	809.03	1.667	2.23E-12	8.74E-11	proline rich 11
FAM72A	34.67	1.801	2.25E-12	8.81E-11	family with sequence similarity 72 member A
NCAPG2	690.73	1.067	2.31E-12	9.01E-11	non-SMC condensin II complex subunit G2
CES2	8841.84	-2.530	2.35E-12	9.16E-11	carboxylesterase 2
RETSAT	2731.62	-1.416	2.38E-12	9.26E-11	retinol saturase
TWIST1	15.31	3.006	2.43E-12	9.45E-11	twist family bHLH transcription factor 1
AC234582.1	46.74	2.335	2.45E-12	9.49E-11	THBS3 antisense RNA 1
STK31	270.38	2.671	2.53E-12	9.81E-11	serine/threonine kinase 31
MELTF	513.94	2.722	2.57E-12	9.94E-11	melanotransferrin
LINC02688	102.79	-4.270	2.62E-12	1.01E-10	long intergenic non-protein coding RNA 2688
ACADVL	8025.60	-1.084	2.64E-12	1.02E-10	acyl-CoA dehydrogenase very long chain
ADHFE1	142.36	-2.301	2.65E-12	1.02E-10	alcohol dehydrogenase iron containing 1
GPAT3	1089.15	-2.107	2.70E-12	1.04E-10	glycerol-3-phosphate acyltransferase 3
IL4I1	50.38	2.710	2.70E-12	1.04E-10	interleukin 4 induced 1
LINC01050	6.28	4.600	2.70E-12	1.04E-10	long intergenic non-protein coding RNA 1050
GAST	46.83	-5.897	2.70E-12	1.04E-10	gastrin
PLAUR	1038.13	2.610	2.72E-12	1.04E-10	plasminogen activator, urokinase receptor
Z98257.1	13.86	4.488	2.78E-12	1.06E-10	NA
AL049734.1	6.80	4.966	2.83E-12	1.08E-10	NLRP7 pseudogene 1
MTND5P32	10.73	5.155	2.85E-12	1.09E-10	MT-ND5 pseudogene 32
AC093898.1	22.07	-4.596	2.92E-12	1.11E-10	NA
ARID5A	247.97	1.621	2.99E-12	1.14E-10	AT-rich interaction domain 5A
SEMA4F	169.76	1.455	3.02E-12	1.15E-10	ssemaphorin 4F
MMP1	2225.12	3.698	3.07E-12	1.16E-10	matrix metallopeptidase 1
IL2RA	80.25	2.597	3.14E-12	1.19E-10	interleukin 2 receptor subunit alpha
CYP4A11	12.21	-4.675	3.19E-12	1.21E-10	cytochrome P450 family 4 subfamily A member 11
SLC51B	107.58	-4.877	3.22E-12	1.22E-10	solute carrier family 51 subunit beta

CYBA	1329.81	1.384	3.34E-12	1.26E-10	cytochrome b-245 alpha chain
CDKN2B-AS1	251.00	1.682	3.38E-12	1.27E-10	CDKN2B antisense RNA 1
AL359397.2	60.19	-2.021	3.42E-12	1.29E-10	NA
UPRT	309.48	-1.108	3.44E-12	1.29E-10	uracil phosphoribosyltransferase homolog
SOS2	2063.66	-1.084	3.50E-12	1.31E-10	SOS Ras/Rho guanine nucleotide exchange factor 2
BUB1B	686.75	1.508	3.51E-12	1.32E-10	BUB1 mitotic checkpoint serine/threonine kinase B
NUP62	1233.54	0.582	3.65E-12	1.37E-10	nucleoporin 62
ACAT1	1094.90	-1.614	3.66E-12	1.37E-10	acetyl-CoA acetyltransferase 1
AL109615.3	18.02	2.873	3.66E-12	1.37E-10	stem cell inhibitory RNA transcript
MAF	2443.40	-2.381	3.81E-12	1.42E-10	MAF bZIP transcription factor
DPY19L1	698.92	1.477	3.83E-12	1.43E-10	dpy-19 like C-mannosyltransferase 1
AMIGO2	170.39	2.147	3.84E-12	1.43E-10	adhesion molecule with Ig like domain 2
ARAP3	450.35	1.349	3.86E-12	1.43E-10	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3
CHN1	151.36	2.013	3.89E-12	1.44E-10	chimerin 1
FITM1	13.20	-1.396	3.91E-12	1.45E-10	fat storage inducing transmembrane protein 1
PLEKHO2	400.39	1.488	3.92E-12	1.45E-10	pleckstrin homology domain containing O2
RAET1K	7.44	3.634	3.93E-12	1.45E-10	retinoic acid early transcript 1K pseudogene
SERPINB9	662.95	2.226	3.97E-12	1.46E-10	serpin family B member 9
ASB11	33.19	-5.576	3.98E-12	1.46E-10	ankyrin repeat and SOCS box containing 11
CENPA	137.70	1.665	3.99E-12	1.47E-10	centromere protein A
CDHR2	5309.87	-4.143	4.04E-12	1.48E-10	cadherin related family member 2
CENPF	3518.05	1.563	4.04E-12	1.48E-10	centromere protein F
WDR20	792.12	-0.720	4.09E-12	1.50E-10	WD repeat domain 20
ARMC3	12.41	4.087	4.19E-12	1.53E-10	armadillo repeat containing 3
MIR3936HG	325.43	-1.787	4.19E-12	1.53E-10	MIR3936 host gene
XRCC2	272.33	1.489	4.20E-12	1.53E-10	X-ray repair cross complementing 2
ZADH2	875.81	-0.863	4.48E-12	1.63E-10	zinc binding alcohol dehydrogenase domain containing 2
SULF1	932.83	2.606	4.57E-12	1.66E-10	sulfatase 1
DBP	241.72	-1.393	4.65E-12	1.69E-10	D-box binding PAR bZIP transcription factor
LINC02754	68.69	-3.550	4.69E-12	1.70E-10	long intergenic non-protein coding RNA 2754
PANK1	515.69	-1.543	4.69E-12	1.70E-10	pantothenate kinase 1
AC063952.4	7.35	4.553	4.86E-12	1.76E-10	NA
SYNJ2BP	2028.78	-1.135	5.08E-12	1.84E-10	synaptojanin 2 binding protein
BYSL	225.44	1.371	5.12E-12	1.85E-10	bystin like
PVT1	1585.51	1.707	5.18E-12	1.87E-10	Pvt1 oncogene
EME1	114.61	1.690	5.37E-12	1.93E-10	essential meiotic structure-specific endonuclease 1
MOGAT2	1216.94	-4.881	5.39E-12	1.94E-10	monoacylglycerol O-acyltransferase 2
RFC3	453.11	1.733	5.40E-12	1.94E-10	replication factor C subunit 3
IQCJ-SCHIP1	24.17	1.977	5.45E-12	1.96E-10	IQCJ-SCHIP1 readthrough
ZNF429	339.74	-1.380	5.51E-12	1.98E-10	zinc finger protein 429
NRGN	26.33	2.053	5.82E-12	2.09E-10	neurogranin
SLC6A4	313.40	-4.341	5.86E-12	2.10E-10	solute carrier family 6 member 4
CNOT6L	1525.99	-0.874	5.93E-12	2.12E-10	CCR4-NOT transcription complex subunit 6 like
RNF125	491.13	-2.142	5.94E-12	2.12E-10	ring finger protein 125

TEX9	345.59	-1.464	5.94E-12	2.12E-10	testis expressed 9
ACAD11	943.47	-1.728	5.96E-12	2.12E-10	acyl-CoA dehydrogenase family member 11
BRCA2	1240.79	1.835	5.98E-12	2.12E-10	BRCA2 DNA repair associated
SLC2A2	795.48	-7.062	6.07E-12	2.15E-10	solute carrier family 2 member 2
MTBP	300.30	1.368	6.12E-12	2.17E-10	MDM2 binding protein
TRIM3	502.43	-1.790	6.16E-12	2.18E-10	tripartite motif containing 3
LHFPL5	15.84	2.321	6.17E-12	2.18E-10	LHFPL tetraspan subfamily member 5
TMEM220	245.92	-2.299	6.32E-12	2.23E-10	transmembrane protein 220
RUNDC3B	61.34	-2.646	6.34E-12	2.24E-10	RUN domain containing 3B
PHKB	2506.78	-0.758	6.58E-12	2.32E-10	phosphorylase kinase regulatory subunit beta
TTC38	1044.03	-1.659	6.58E-12	2.32E-10	tetratricopeptide repeat domain 38
AGPAT4	387.59	1.873	6.60E-12	2.32E-10	1-acylglycerol-3-phosphate O-acyltransferase 4
LRRC28	529.63	-1.147	6.61E-12	2.32E-10	leucine rich repeat containing 28
DNAH17	71.79	2.945	6.66E-12	2.34E-10	dynein axonemal heavy chain 17
EVX1	12.25	5.393	6.68E-12	2.34E-10	even-skipped homeobox 1
FAM172A	620.39	-0.888	6.84E-12	2.39E-10	family with sequence similarity 172 member A
TEX41	44.47	3.466	6.86E-12	2.40E-10	testis expressed 41
AC010789.1	29.17	5.505	6.95E-12	2.42E-10	NA
EBF2	16.91	3.048	6.95E-12	2.42E-10	EBF transcription factor 2
AC009159.3	13.53	-2.095	6.98E-12	2.43E-10	NA
PGC	46479.12	-4.935	7.05E-12	2.45E-10	progastricsin
TMEM231P1	88.34	-3.401	7.31E-12	2.54E-10	TMEM231 pseudogene 1
TKTL1	121.04	6.817	7.33E-12	2.54E-10	transketolase like 1
ZNF57	279.97	-0.995	7.34E-12	2.54E-10	zinc finger protein 57
CCL4	115.95	3.209	7.46E-12	2.58E-10	C-C motif chemokine ligand 4
DBF4B	242.71	1.287	7.52E-12	2.60E-10	DBF4 zinc finger B
IQGAP3	2110.32	1.674	7.55E-12	2.61E-10	IQ motif containing GTPase activating protein 3
NOP56	1747.21	1.038	7.61E-12	2.62E-10	NOP56 ribonucleoprotein
DEPDC7	258.84	-2.523	7.70E-12	2.65E-10	DEP domain containing 7
SGO1	252.65	1.481	7.71E-12	2.65E-10	shugoshin 1
IL3RA	64.80	1.995	7.74E-12	2.66E-10	interleukin 3 receptor subunit alpha
AL157756.1	14.01	-1.462	7.83E-12	2.69E-10	uncharacterized LOC101927702
CDHR5	5249.07	-4.289	7.89E-12	2.70E-10	cadherin related family member 5
HAGH	464.44	-0.932	8.14E-12	2.79E-10	hydroxyacylglutathione hydrolase
AC010336.2	8.83	-3.616	8.34E-12	2.85E-10	NA
PPRC1	1067.73	0.887	8.59E-12	2.93E-10	PPARG related coactivator 1
PLA2G12A	778.02	-0.771	8.66E-12	2.95E-10	phospholipase A2 group XIIA
SOX6	873.29	-2.013	8.71E-12	2.96E-10	SRY-box transcription factor 6
C5orf34	101.36	1.686	8.74E-12	2.97E-10	chromosome 5 open reading frame 34
GNAT3	16.26	-5.864	8.73E-12	2.97E-10	G protein subunit alpha transducin 3
KLHDC2	1105.86	-0.909	8.73E-12	2.97E-10	kelch domain containing 2
KCNAB2	369.16	1.729	9.24E-12	3.13E-10	potassium voltage-gated channel subfamily A regulatory beta subunit 2
DDX21	3284.41	1.157	9.31E-12	3.15E-10	DExD-box helicase 21
SPOCD1	40.40	2.333	9.34E-12	3.16E-10	SPOC domain containing 1
SULT1B1	3964.77	-3.350	9.39E-12	3.17E-10	sulfotransferase family 1B member 1
AFF2	43.72	3.042	9.58E-12	3.23E-10	AF4/FMR2 family member 2

DDX39A	1006.40	0.797	9.59E-12	3.23E-10	DExD-box helicase 39A
CRBN	731.45	-0.857	9.62E-12	3.24E-10	cereblon
CCDC150	263.41	1.778	9.68E-12	3.25E-10	coiled-coil domain containing 150
TMEM116	292.17	-1.199	9.81E-12	3.29E-10	transmembrane protein 116
UBXN10	103.39	2.818	9.88E-12	3.31E-10	UBX domain protein 10
IGHG4	1360.96	4.723	1.00E-11	3.35E-10	immunoglobulin heavy constant gamma 4 (G4m marker)
ST18	62.47	-2.637	1.00E-11	3.35E-10	ST18 C2H2C-type zinc finger transcription factor
ZYX	1129.54	1.041	1.01E-11	3.38E-10	zyxin
C1orf115	1342.14	-2.741	1.03E-11	3.43E-10	chromosome 1 open reading frame 115
SULT1A2	214.68	-3.681	1.06E-11	3.55E-10	sulfotransferase family 1A member 2
LINC01303	17.32	3.422	1.07E-11	3.55E-10	long intergenic non-protein coding RNA 1303
NOLC1	1740.07	1.083	1.07E-11	3.55E-10	nucleolar and coiled-body phosphoprotein 1
AP006216.3	36.91	-7.690	1.08E-11	3.57E-10	NA
TNFRSF10C	125.15	3.200	1.08E-11	3.59E-10	TNF receptor superfamily member 10c
C8orf31	72.02	2.839	1.09E-11	3.60E-10	long intergenic non-protein coding RNA 2904
FP325330.3	17.78	5.582	1.10E-11	3.63E-10	NA
ILDR2	124.09	2.805	1.10E-11	3.63E-10	immunoglobulin like domain containing receptor 2
PCOLCE	168.15	1.970	1.11E-11	3.67E-10	procollagen C-endopeptidase enhancer
NID2	404.22	2.209	1.12E-11	3.70E-10	nidogen 2
TOP1MT	334.06	1.182	1.12E-11	3.71E-10	DNA topoisomerase I mitochondrial
TNFRSF9	73.95	2.512	1.15E-11	3.78E-10	TNF receptor superfamily member 9
LINC02177	94.92	-5.257	1.16E-11	3.83E-10	long intergenic non-protein coding RNA 2177
NOX4	32.15	2.234	1.20E-11	3.93E-10	NADPH oxidase 4
KIF18B	473.38	1.520	1.20E-11	3.94E-10	kinesin family member 18B
FOXK1	2551.92	0.874	1.23E-11	4.02E-10	forkhead box K1
CD93	1314.99	1.883	1.23E-11	4.02E-10	CD93 molecule
MS4A18	68.59	-7.684	1.25E-11	4.08E-10	membrane spanning 4-domains A18
ZBTB16	126.63	-2.406	1.26E-11	4.10E-10	zinc finger and BTB domain containing 16
C1orf61	61.50	3.198	1.29E-11	4.19E-10	MIR9-1 host gene
ADPRM	89.25	-0.888	1.29E-11	4.20E-10	ADP-ribose/CDP-alcohol diphosphatase, manganese dependent
ALPP	39.92	6.477	1.30E-11	4.24E-10	alkaline phosphatase, placental
NEB	1072.27	2.664	1.31E-11	4.24E-10	nebulin
AL096816.1	21.04	-3.074	1.31E-11	4.25E-10	NA
DUSP2	82.64	1.978	1.34E-11	4.33E-10	dual specificity phosphatase 2
MSANTD3	288.63	1.118	1.35E-11	4.37E-10	Myb/SANT DNA binding domain containing 3
CYP2C9	1719.14	-3.404	1.38E-11	4.45E-10	cytochrome P450 family 2 subfamily C member 9
CCR8	9.61	2.735	1.39E-11	4.48E-10	C-C motif chemokine receptor 8
GSTK1	1997.11	-1.120	1.40E-11	4.50E-10	glutathione S-transferase kappa 1
NR1I3	20.54	-4.383	1.40E-11	4.50E-10	nuclear receptor subfamily 1 group I member 3
TM6SF2	628.42	-4.392	1.40E-11	4.50E-10	transmembrane 6 superfamily member 2
LILRB2	133.21	2.837	1.43E-11	4.60E-10	leukocyte immunoglobulin like receptor B2
HEATR1	2315.47	0.989	1.43E-11	4.60E-10	HEAT repeat containing 1
OR7E117P	4.29	-5.167	1.44E-11	4.62E-10	olfactory receptor family 7 subfamily E member 117 pseudogene
TOB1-AS1	69.97	-1.571	1.46E-11	4.67E-10	TOB1 antisense RNA 1

AC246817.2	32.53	-2.677	1.47E-11	4.72E-10	uncharacterized LOC101927815
QSOX2	561.16	1.314	1.50E-11	4.78E-10	quiescin sulfhydryl oxidase 2
ADGRG4	186.07	-7.026	1.52E-11	4.85E-10	adhesion G protein-coupled receptor G4
DCSTAMP	4.52	4.437	1.54E-11	4.92E-10	dendrocyte expressed seven transmembrane protein
AC015909.1	11.59	4.678	1.56E-11	4.97E-10	NA
GALNT2	1623.69	0.756	1.57E-11	5.00E-10	polypeptide N-acetylgalactosaminyltransferase 2
CPNE1	1605.19	0.902	1.58E-11	5.03E-10	copine 1
MPND	191.54	-1.014	1.60E-11	5.07E-10	MPN domain containing
PDE4B	683.82	2.809	1.60E-11	5.07E-10	phosphodiesterase 4B
ZNF473	362.94	0.868	1.61E-11	5.11E-10	zinc finger protein 473
HOXC13-AS	7.51	4.793	1.62E-11	5.15E-10	HOXC13 antisense RNA
B3GAT1	109.00	-3.897	1.63E-11	5.16E-10	beta-1,3-glucuronyltransferase 1
NEMP1	974.59	1.478	1.65E-11	5.22E-10	nuclear envelope integral membrane protein 1
FAM72D	17.33	2.269	1.66E-11	5.25E-10	family with sequence similarity 72 member D
TYMS	406.74	1.324	1.71E-11	5.38E-10	thymidylate synthetase
CYP4A22-AS1	10.75	2.744	1.73E-11	5.45E-10	CYP4A22 antisense RNA 1
SIGLEC14	32.73	2.897	1.76E-11	5.55E-10	sialic acid binding Ig like lectin 14
VIPR1	734.86	-2.279	1.77E-11	5.58E-10	vasoactive intestinal peptide receptor 1
SLC2A6	156.84	2.339	1.80E-11	5.64E-10	solute carrier family 2 member 6
CCR4	41.45	2.141	1.82E-11	5.70E-10	C-C motif chemokine receptor 4
TMEM120A	420.39	-0.994	1.85E-11	5.80E-10	transmembrane protein 120A
DGAT1	2531.35	-2.130	1.89E-11	5.90E-10	diacylglycerol O-acyltransferase 1
AC022031.2	27.76	5.940	1.90E-11	5.93E-10	NA
TMEM232	35.57	-1.676	1.90E-11	5.93E-10	transmembrane protein 232
GPA33	1538.98	-4.681	1.93E-11	6.03E-10	glycoprotein A33
FAM13A	2293.85	-2.466	1.95E-11	6.07E-10	family with sequence similarity 13 member A
LRP2	115.81	4.912	1.95E-11	6.08E-10	LDL receptor related protein 2
ANAPC1P4	32.05	1.975	1.97E-11	6.13E-10	ANAPC1 pseudogene 4
ZC3H12A	1287.78	1.432	1.98E-11	6.15E-10	zinc finger CCCH-type containing 12A
FAM72B	41.70	1.821	2.00E-11	6.20E-10	family with sequence similarity 72 member B
MSLN	549.03	3.287	2.01E-11	6.21E-10	mesothelin
BPHL	464.37	-1.359	2.02E-11	6.23E-10	biphenyl hydrolase like
PLAC1	10.07	4.337	2.02E-11	6.23E-10	placenta enriched 1
DSN1	476.31	0.980	2.03E-11	6.28E-10	DSN1 component of MIS12 kinetochore complex
RBL1	640.25	0.992	2.11E-11	6.50E-10	RB transcriptional corepressor like 1
RNVU1-22	10.87	2.416	2.15E-11	6.62E-10	RNA, variant U1 small nuclear 22
ACTN1	2747.00	1.375	2.16E-11	6.66E-10	actinin alpha 1
COL1A2	6203.05	2.473	2.25E-11	6.91E-10	collagen type I alpha 2 chain
DKC1	1125.96	1.069	2.29E-11	7.02E-10	dyskerin pseudouridine synthase 1
GOLGA8N	31.62	-1.492	2.31E-11	7.10E-10	golgin A8 family member N
NR3C1	1978.19	-1.128	2.35E-11	7.22E-10	nuclear receptor subfamily 3 group C member 1
DOCK4	697.17	1.459	2.43E-11	7.43E-10	dedicator of cytokinesis 4
NCAM1	134.76	-3.136	2.44E-11	7.46E-10	neural cell adhesion molecule 1
AL033380.1	11.72	-4.416	2.46E-11	7.52E-10	NA
SCG2	137.31	-3.578	2.56E-11	7.80E-10	secretogranin II

ADAMTS20	21.70	5.897	2.63E-11	8.02E-10	ADAM metallopeptidase with thrombospondin type 1 motif 20
MFSD8	584.95	-0.608	2.66E-11	8.11E-10	major facilitator superfamily domain containing 8
CLN6	699.12	1.003	2.68E-11	8.14E-10	CLN6 transmembrane ER protein
SAGE1	34.16	7.323	2.74E-11	8.32E-10	sarcoma antigen 1
CD80	65.92	2.440	2.77E-11	8.41E-10	CD80 molecule
FEZ1	99.73	2.329	2.79E-11	8.47E-10	fasciculation and elongation protein zeta 1
AC012073.1	52.48	1.215	2.80E-11	8.49E-10	NA
BDH2	622.91	-1.812	2.81E-11	8.51E-10	3-hydroxybutyrate dehydrogenase 2
PXMP2	365.18	-1.585	2.90E-11	8.76E-10	peroxisomal membrane protein 2
ZNF114	31.44	3.374	2.92E-11	8.81E-10	zinc finger protein 114
CCDC28B	31.38	1.701	3.01E-11	9.09E-10	coiled-coil domain containing 28B
ERCC6L	204.87	1.770	3.03E-11	9.13E-10	ERCC excision repair 6 like, spindle assembly checkpoint helicase
PLEKHA4	257.99	1.531	3.08E-11	9.27E-10	pleckstrin homology domain containing A4
LINC00346	141.01	2.295	3.08E-11	9.27E-10	p53 regulated carcinoma associated Stat3 activating long intergenic non-protein coding transcript
PEPD	1251.35	-1.672	3.10E-11	9.32E-10	peptidase D
VTCN1	112.58	4.200	3.14E-11	9.43E-10	V-set domain containing T cell activation inhibitor 1
AL356599.1	58.93	-1.426	3.18E-11	9.53E-10	FBXO30 divergent transcript
SLC30A1	1605.25	-1.415	3.31E-11	9.91E-10	solute carrier family 30 member 1
CCNE1	529.35	2.800	3.34E-11	9.97E-10	cyclin E1
RHPN1	303.55	1.849	3.35E-11	1.00E-09	rhophilin Rho GTPase binding protein 1
FEN1	552.43	1.394	3.35E-11	1.00E-09	flap structure-specific endonuclease 1
LINC00476	181.96	-1.130	3.35E-11	1.00E-09	long intergenic non-protein coding RNA 476
AC093001.1	23.24	5.074	3.37E-11	1.00E-09	NA
NDC80	440.08	1.542	3.38E-11	1.00E-09	NDC80 kinetochore complex component
SNRPD1	968.70	1.431	3.38E-11	1.00E-09	small nuclear ribonucleoprotein D1 polypeptide
ANGPTL2	232.72	2.102	3.39E-11	1.01E-09	angiopoietin like 2
MMP10	216.63	3.957	3.40E-11	1.01E-09	matrix metallopeptidase 10
ENTR1	660.35	0.854	3.41E-11	1.01E-09	endosome associated trafficking regulator 1
ABCA1	2431.48	1.268	3.42E-11	1.01E-09	ATP binding cassette subfamily A member 1
C3P1	38.89	5.136	3.45E-11	1.02E-09	complement component 3 precursor pseudogene
UBE2S	424.00	1.406	3.52E-11	1.04E-09	ubiquitin conjugating enzyme E2 S
DPF1	12.79	3.414	3.60E-11	1.06E-09	double PHD fingers 1
RFC4	338.77	1.242	3.61E-11	1.06E-09	replication factor C subunit 4
DOK3	184.29	1.513	3.76E-11	1.11E-09	docking protein 3
ANKRD20A21P	81.57	-3.435	3.85E-11	1.13E-09	ankyrin repeat domain 20 family member A21, pseudogene
ZNF469	230.46	2.678	3.88E-11	1.14E-09	zinc finger protein 469
TREML2	34.91	3.453	3.90E-11	1.14E-09	triggering receptor expressed on myeloid cells like 2
MSN	3223.79	1.542	3.95E-11	1.16E-09	moesin
SIGLEC9	44.09	2.276	3.95E-11	1.16E-09	sialic acid binding Ig like lectin 9
RHPN1-AS1	8.29	2.837	4.12E-11	1.21E-09	RHPN1 antisense RNA 1 (head to head)
PKD2L1	9.23	3.396	4.15E-11	1.21E-09	polycystin 2 like 1, transient receptor potential cation channel
TMEM161B	935.62	-1.014	4.17E-11	1.22E-09	transmembrane protein 161B

AC026401.3	89.14	0.982	4.19E-11	1.22E-09	NA
KNSTRN	430.03	1.165	4.20E-11	1.22E-09	kinetochore localized astrin (SPAG5) binding protein
C10orf143	41.11	-1.194	4.23E-11	1.23E-09	chromosome 10 open reading frame 143
CHEK1	440.63	1.285	4.23E-11	1.23E-09	checkpoint kinase 1
PKMYT1	268.26	1.411	4.26E-11	1.24E-09	protein kinase, membrane associated tyrosine/threonine 1
CCR1	183.80	2.692	4.34E-11	1.26E-09	C-C motif chemokine receptor 1
AL355612.1	6.52	-5.352	4.38E-11	1.27E-09	NA
CTSO	411.64	-1.860	4.39E-11	1.27E-09	cathepsin O
RPLP0P2	19.62	3.652	4.41E-11	1.28E-09	ribosomal protein lateral stalk subunit P0 pseudogene 2
SKA3	280.29	1.523	4.44E-11	1.28E-09	spindle and kinetochore associated complex subunit 3
ASAP1	1128.00	1.122	4.45E-11	1.29E-09	ArfGAP with SH3 domain, ankyrin repeat and PH domain 1
HK3	109.17	2.900	4.46E-11	1.29E-09	hexokinase 3
KIF2C	570.24	1.625	4.50E-11	1.30E-09	kinesin family member 2C
ISG20L2	889.31	0.677	4.59E-11	1.32E-09	interferon stimulated exonuclease gene 20 like 2
SLC15A1	3381.53	-4.013	4.65E-11	1.34E-09	solute carrier family 15 member 1
SRCIN1	283.25	2.365	4.68E-11	1.34E-09	SRC kinase signaling inhibitor 1
GUCA2B	88.58	-4.822	4.74E-11	1.36E-09	guanylate cyclase activator 2B
STX12	1723.26	-1.065	4.78E-11	1.37E-09	syntaxin 12
AL391056.1	48.11	3.693	4.79E-11	1.37E-09	NA
BTNL2	15.41	-5.188	4.86E-11	1.39E-09	butyrophilin like 2
SLC66A1L	52.96	3.237	4.89E-11	1.40E-09	solute carrier family 66 member 1 like
AC108463.3	20.12	2.349	4.94E-11	1.41E-09	NA
GNGT1	29.08	4.836	5.01E-11	1.43E-09	G protein subunit gamma transducin 1
RILPL2	191.25	1.199	5.09E-11	1.45E-09	Rab interacting lysosomal protein like 2
SH2B3	504.83	1.614	5.11E-11	1.45E-09	SH2B adaptor protein 3
LRRC17	54.89	-2.843	5.13E-11	1.46E-09	leucine rich repeat containing 17

Appendix C-2: Clinical Benefit vs. No Clinical Benefit (PreTx OAC)

Gene	baseMean	log2FC	pvalue	padj	description
ACOD1	11.12	3.695	1.56E-09	4.11E-05	aconitate decarboxylase 1
AC010491.2	17.08	2.644	1.59E-06	1.48E-02	NA
TNFSF14	91.76	1.828	1.69E-06	1.48E-02	TNF superfamily member 14
AC080079.2	7.73	3.037	4.70E-06	1.63E-02	NA
AC123912.2	3.77	3.103	5.58E-06	1.63E-02	NA
APOB	165.69	3.089	3.99E-06	1.63E-02	apolipoprotein B
CLEC12A	36.33	1.869	3.98E-06	1.63E-02	C-type lectin domain family 12 member A
IGHD	64.32	-3.140	5.20E-06	1.63E-02	immunoglobulin heavy constant delta
TAGAP	493.06	2.136	2.96E-06	1.63E-02	T cell activation RhoGTPase activating protein
PLEK	1414.20	2.282	7.90E-06	1.99E-02	pleckstrin
TPSP2	12.27	2.806	8.32E-06	1.99E-02	tryptase pseudogene 2
TUFMP1	21.11	3.090	9.11E-06	2.00E-02	Tu translation elongation factor, mitochondrial pseudogene 1
AL137792.1	4.61	1.924	2.53E-05	4.84E-02	NA
GPR84	41.85	1.732	2.84E-05	4.84E-02	G protein-coupled receptor 84
LCP2	1216.28	1.559	2.60E-05	4.84E-02	lymphocyte cytosolic protein 2
LRAT	36.75	-2.040	3.40E-05	4.84E-02	lecithin retinol acyltransferase
MAP3K20-AS1	67.53	-2.199	3.49E-05	4.84E-02	MAP3K20 antisense RNA 1
SI	318.18	2.195	3.47E-05	4.84E-02	sucrase-isomaltase
TREM1	407.55	1.948	3.21E-05	4.84E-02	triggering receptor expressed on myeloid cells 1
AC107953.2	5.39	2.125	4.09E-05	4.89E-02	uncharacterized LOC101927513
AL354811.1	3.89	2.034	4.00E-05	4.89E-02	NA
MORN3	30.66	1.355	3.87E-05	4.89E-02	MORN repeat containing 3
CNTN1	141.41	-1.947	5.26E-05	5.19E-02	contactin 1
DIO3OS	48.16	1.653	5.62E-05	5.19E-02	DIO3 opposite strand upstream RNA
DNAJC12	152.12	-1.952	4.76E-05	5.19E-02	DnaJ heat shock protein family (Hsp40) member C12
GBA3	24.89	1.851	5.41E-05	5.19E-02	glucosylceramidase beta 3 (gene/pseudogene)
MUC17	4853.10	1.896	5.30E-05	5.19E-02	mucin 17, cell surface associated
PDE4B	929.70	1.353	5.72E-05	5.19E-02	phosphodiesterase 4B
PLEKHG4B	170.96	-1.851	5.22E-05	5.19E-02	pleckstrin homology and RhoGEF domain containing G4B
FNDC4	213.95	-1.073	7.03E-05	5.49E-02	fibronectin type III domain containing 4
FPR1	339.91	1.438	7.30E-05	5.49E-02	formyl peptide receptor 1
IGKV1D-13	14.35	-1.731	7.14E-05	5.49E-02	immunoglobulin kappa variable 1D-13
PDYN-AS1	3.77	1.703	7.30E-05	5.49E-02	PDYN antisense RNA 1
VTN	60.46	1.643	7.09E-05	5.49E-02	vitronectin
XPNPEP2	20.00	1.721	6.58E-05	5.49E-02	X-prolyl aminopeptidase 2
GLB1L3	30.21	1.656	7.54E-05	5.51E-02	galactosidase beta 1 like 3
RERGL	25.57	-1.574	9.14E-05	6.50E-02	RERG like
CBSL	7.47	-1.525	9.86E-05	6.83E-02	cystathionine beta-synthase like
CCL4L2	121.67	1.341	1.17E-04	7.16E-02	C-C motif chemokine ligand 4 like 2
CT55	8.10	1.236	1.17E-04	7.16E-02	cancer/testis antigen 55
DPEP1	35.99	1.443	1.10E-04	7.16E-02	dipeptidase 1

ELOVL4	147.20	-1.387	1.09E-04	7.16E-02	ELOVL fatty acid elongase 4
TPSD1	9.31	-1.260	1.12E-04	7.16E-02	tryptase delta 1
AC114291.1	13.35	1.250	1.36E-04	7.92E-02	uncharacterized LOC101928857
SHBG	4.88	1.065	1.35E-04	7.92E-02	sex hormone binding globulin
UPK3A	9.11	1.211	1.43E-04	8.20E-02	uroplakin 3A
AC007663.2	16.04	-1.158	1.48E-04	8.29E-02	long intergenic non-protein coding RNA 2891
SRGN	1236.09	0.859	1.56E-04	8.55E-02	serglycin
C5AR1	316.21	0.794	1.62E-04	8.71E-02	complement C5a receptor 1
ANKRD20A19P	4.75	1.184	1.66E-04	8.72E-02	ankyrin repeat domain 20 family member A19, pseudogene
AP001099.1	6.04	-1.075	1.74E-04	8.81E-02	NA
FCAR	75.18	1.085	1.73E-04	8.81E-02	Fc fragment of IgA receptor
FREM1	41.44	-0.786	1.81E-04	9.01E-02	FRAS1 related extracellular matrix 1
ABLIM3	1116.95	-0.812	2.03E-04	9.37E-02	actin binding LIM protein family member 3
AC005253.1	5.99	1.009	2.01E-04	9.37E-02	NA
AL139254.3	4.84	1.036	2.01E-04	9.37E-02	NA
TMEM71	107.93	0.759	2.03E-04	9.37E-02	transmembrane protein 71
PRAM1	46.77	0.685	2.13E-04	9.66E-02	PML-RARA regulated adaptor molecule 1
CSF3R	1198.03	0.850	2.22E-04	9.89E-02	colony stimulating factor 3 receptor

Appendix C-3: Responder vs. Non-Responder (PreTx OAC)

Gene	baseMean	log2FC	pvalue	padj	description
DSG1	1559.33	4.890	1.71E-11	4.69E-07	desmoglein 1
MUCL3	2118.07	-4.041	1.82E-10	2.50E-06	mucin like 3
CARNS1	178.28	-3.660	4.31E-10	3.76E-06	carnosine synthase 1
PIWIL2	66.80	-2.383	5.47E-10	3.76E-06	piwi like RNA-mediated gene silencing 2
CASR	20.27	-3.828	2.17E-09	1.20E-05	calcium sensing receptor
ACER2	1149.10	-2.364	1.43E-08	6.55E-05	alkaline ceramidase 2
AC084026.2	15.81	-3.115	2.67E-08	7.99E-05	NA
GKN1	1611.96	-4.291	2.91E-08	7.99E-05	gastrokine 1
LRRC66	231.73	-2.516	2.62E-08	7.99E-05	leucine rich repeat containing 66
TNPO1P2	14.30	3.932	2.86E-08	7.99E-05	transportin 1 pseudogene 2
AC016717.2	40.49	3.024	3.85E-08	9.62E-05	NA
H19	1246.12	3.311	6.31E-08	1.44E-04	H19 imprinted maternally expressed transcript
NRG3	11.08	-3.335	6.89E-08	1.46E-04	neuregulin 3
AC103718.1	602.82	3.194	1.71E-07	2.94E-04	NA
ANO7	424.52	-2.613	1.71E-07	2.94E-04	anoctamin 7
TNPO1P3	45.91	3.676	1.59E-07	2.94E-04	transportin 1 pseudogene 3
AL158206.1	559.80	-2.002	3.72E-07	5.45E-04	NA
CHGA	85.21	-3.401	3.70E-07	5.45E-04	chromogranin A
IGFBP6	263.07	2.402	3.77E-07	5.45E-04	insulin like growth factor binding protein 6
ALG1L6P	6.19	-3.190	4.90E-07	6.73E-04	ALG1 like 6, pseudogene
PDIA2	379.32	-3.050	5.62E-07	7.36E-04	protein disulfide isomerase family A member 2
FER1L4	4768.72	-2.421	7.70E-07	9.63E-04	fer-1 like family member 4 (pseudogene)
FCGBP	12173.30	-2.652	9.62E-07	1.11E-03	Fc fragment of IgG binding protein
RNU6-1175P	10.19	-2.881	9.70E-07	1.11E-03	RNA, U6 small nuclear 1175, pseudogene
PLA2G4E	488.40	3.293	1.02E-06	1.12E-03	phospholipase A2 group IVE
AC007663.2	15.19	-2.526	1.07E-06	1.14E-03	long intergenic non-protein coding RNA 2891
GPT	248.16	-2.215	1.55E-06	1.58E-03	glutamic--pyruvic transaminase
POU4F1	8.90	-2.891	2.10E-06	2.06E-03	POU class 4 homeobox 1
GHRL	53.70	-2.534	2.62E-06	2.32E-03	ghrelin and obestatin prepropeptide
KRTDAP	59.56	3.153	2.54E-06	2.32E-03	keratinocyte differentiation associated protein
NOS1	40.57	2.697	2.55E-06	2.32E-03	nitric oxide synthase 1
MT1G	176.44	-2.142	2.84E-06	2.36E-03	metallothionein 1G
VSIG2	1994.31	-2.122	2.76E-06	2.36E-03	V-set and immunoglobulin domain containing 2
APOB	283.21	3.061	3.86E-06	3.02E-03	apolipoprotein B
KRT14	34120.73	3.322	3.82E-06	3.02E-03	keratin 14
NKX6-2	16.18	-3.146	3.96E-06	3.02E-03	NK6 homeobox 2
HEPHL1	1605.87	2.983	4.09E-06	3.04E-03	hephaestin like 1
KCNE2	35.49	-2.989	4.27E-06	3.07E-03	potassium voltage-gated channel subfamily E regulatory subunit 2
SLC5A7	5.62	-3.191	4.36E-06	3.07E-03	solute carrier family 5 member 7
KRT17	19573.48	2.986	4.57E-06	3.14E-03	keratin 17
KRT16P2	138.30	2.858	5.34E-06	3.58E-03	keratin 16 pseudogene 2
SULT1B1	978.20	-1.981	5.65E-06	3.70E-03	sulfotransferase family 1B member 1
HPGD	3198.51	-2.225	6.61E-06	4.23E-03	15-hydroxyprostaglandin dehydrogenase
AC099521.3	105.92	-2.309	7.34E-06	4.59E-03	NA
PTHLH	210.42	2.551	7.90E-06	4.83E-03	parathyroid hormone like hormone
AKR7L	313.22	-1.801	8.49E-06	4.96E-03	aldo-keto reductase family 7 like (gene/pseudogene)
BCAP31P1	3.40	3.084	8.34E-06	4.96E-03	B cell receptor associated protein 31 pseudogene 1
BAGE2	11.22	-2.947	9.21E-06	5.27E-03	BAGE family member 2
AC108751.5	25.62	-2.700	1.04E-05	5.80E-03	NA
AP001994.3	92.31	-2.429	1.08E-05	5.80E-03	NA

IGFALS	27.13	-2.530	1.07E-05	5.80E-03	insulin like growth factor binding protein acid labile subunit
ERICH3	27.02	-2.844	1.21E-05	6.38E-03	glutamate rich 3
LSAMP-AS1	8.85	-2.430	1.24E-05	6.41E-03	LSAMP antisense RNA 1
AL132780.1	25.11	-1.583	1.32E-05	6.71E-03	PRMT5 divergent transcript
CFAP65	22.24	-1.931	1.39E-05	6.96E-03	cilia and flagella associated protein 65
NDRG4	380.14	1.993	1.49E-05	7.18E-03	NDRG family member 4
WDR17	34.12	-2.060	1.47E-05	7.18E-03	WD repeat domain 17
PSAPL1	316.33	-2.718	1.65E-05	7.83E-03	prosaposin like 1
APOA4	16.00	2.717	1.72E-05	8.00E-03	apolipoprotein A4
AL683887.1	4.37	2.771	1.76E-05	8.08E-03	NA
AL645929.2	10.35	-1.530	1.90E-05	8.17E-03	HLA complex P5B
CCDC89	92.34	-2.085	1.90E-05	8.17E-03	coiled-coil domain containing 89
MAP3K20-AS1	66.73	-2.579	1.85E-05	8.17E-03	MAP3K20 antisense RNA 1
PPP1R3C	860.87	2.065	1.87E-05	8.17E-03	protein phosphatase 1 regulatory subunit 3C
AC092490.1	19.85	-2.398	2.03E-05	8.31E-03	NA
IL36RN	684.54	2.772	2.01E-05	8.31E-03	interleukin 36 receptor antagonist
LINC01705	43.85	2.377	2.03E-05	8.31E-03	long intergenic non-protein coding RNA 1705
AK7	83.25	-0.924	2.14E-05	8.54E-03	adenylate kinase 7
CAPN8	4627.05	-1.932	2.14E-05	8.54E-03	calpain 8
AC009237.3	424.10	-1.090	2.35E-05	9.23E-03	NA
AC009093.3	13.25	-2.269	2.39E-05	9.26E-03	NA
CYP2S1	1836.54	-1.489	2.54E-05	9.70E-03	cytochrome P450 family 2 subfamily S member 1
GALE	1744.35	-0.883	2.62E-05	9.73E-03	UDP-galactose-4-epimerase
LINC01436	49.74	-2.519	2.61E-05	9.73E-03	long intergenic non-protein coding RNA 1436
RPL21P7	7.14	-1.881	2.82E-05	1.03E-02	ribosomal protein L21 pseudogene 7
AC091544.4	28.80	-1.920	2.91E-05	1.04E-02	NA
KLHL13	203.19	1.953	2.93E-05	1.04E-02	kelch like family member 13
AL355499.1	25.51	-2.084	3.24E-05	1.14E-02	NA
SMPD3	967.06	-1.951	3.65E-05	1.27E-02	sphingomyelin phosphodiesterase 3
FREM1	95.89	2.096	3.97E-05	1.37E-02	FRAS1 related extracellular matrix 1
SPATA6L	336.26	-1.304	4.08E-05	1.37E-02	spermatogenesis associated 6 like
TPRG1LP1	17.08	-2.325	4.07E-05	1.37E-02	tumor protein p63 regulated 1 like pseudogene 1
SULT1C2P1	58.16	-2.308	4.14E-05	1.37E-02	sulfotransferase family 1C member 2 pseudogene 1
AC011379.1	200.85	-1.534	4.46E-05	1.46E-02	NA
AC103923.1	44.46	-1.699	4.85E-05	1.57E-02	NA
ADGRV1	298.91	1.922	5.03E-05	1.61E-02	adhesion G protein-coupled receptor V1
ZSCAN4	30.52	-2.324	5.32E-05	1.68E-02	zinc finger and SCAN domain containing 4
CBLB	1790.48	-0.678	5.42E-05	1.69E-02	Cbl proto-oncogene B
AL590096.1	4.83	-1.722	5.54E-05	1.71E-02	NA
DMBT1L1	13.87	-2.307	5.87E-05	1.78E-02	deleted in malignant brain tumors 1 like 1 (pseudogene)
KLK12	704.13	2.360	5.91E-05	1.78E-02	kallikrein related peptidase 12
AC114930.1	28.47	-2.355	6.12E-05	1.83E-02	NA
AL391834.3	2.72	-2.161	6.30E-05	1.84E-02	NA
NKX2-1	7.93	-2.225	6.24E-05	1.84E-02	NK2 homeobox 1
SPATA6	361.75	-1.060	6.44E-05	1.86E-02	spermatogenesis associated 6
FAM177B	272.86	-2.097	6.65E-05	1.90E-02	family with sequence similarity 177 member B
PAPPA	710.94	1.254	7.34E-05	2.08E-02	pappalysin 1
GATA4	1304.89	-2.053	7.58E-05	2.08E-02	GATA binding protein 4
KRT16P6	609.78	2.225	7.52E-05	2.08E-02	keratin 16 pseudogene 6
PHGR1	64.09	-2.211	7.58E-05	2.08E-02	proline, histidine and glycine rich 1
LINC00365	153.43	-2.093	7.79E-05	2.12E-02	long intergenic non-protein coding RNA 365
AP003555.2	81.52	-1.946	8.34E-05	2.19E-02	NA
JAKMIP3	64.89	-1.733	8.36E-05	2.19E-02	Janus kinase and microtubule interacting protein 3

LINC00930	13.21	-1.801	8.35E-05	2.19E-02	long intergenic non-protein coding RNA 930
MGAT4C	6.87	-2.226	8.26E-05	2.19E-02	MGAT4 family member C
AC007389.1	28.45	-1.682	8.55E-05	2.20E-02	uncharacterized LOC101927533
ANKRD36C	5046.75	-1.163	8.64E-05	2.20E-02	ankyrin repeat domain 36C
LLGL2	5118.49	-0.695	8.59E-05	2.20E-02	LLGL scribble cell polarity complex component 2
MLPH	4059.84	-1.668	8.86E-05	2.23E-02	melanophilin
AC096751.2	46.25	-1.785	9.08E-05	2.25E-02	NA
ANO1	2823.32	-1.401	9.02E-05	2.25E-02	anoctamin 1
ANGPTL3	2.62	-2.129	9.34E-05	2.27E-02	angiopoietin like 3
SAMD13	33.79	-1.065	9.33E-05	2.27E-02	sterile alpha motif domain containing 13
IHH	471.85	-1.733	9.61E-05	2.32E-02	Indian hedgehog signaling molecule
AC018521.2	28.40	-1.375	9.81E-05	2.33E-02	NA
OSBPL7	1530.63	-1.400	9.82E-05	2.33E-02	oxysterol binding protein like 7
AL356361.2	3.29	-2.138	1.01E-04	2.34E-02	NA
MYO1A	979.13	-1.716	1.01E-04	2.34E-02	myosin IA
TAGAP	492.81	1.551	1.00E-04	2.34E-02	T cell activation RhoGTPase activating protein
AC007834.2	20.85	1.834	1.04E-04	2.34E-02	NA
EVA1A	33.33	1.489	1.03E-04	2.34E-02	eva-1 homolog A, regulator of programmed cell death
OR7E28P	3.36	2.065	1.04E-04	2.34E-02	olfactory receptor family 7 subfamily E member 28 pseudogene
AMTN	57.05	2.093	1.08E-04	2.36E-02	amelotin
SYT14P1	27.57	2.106	1.08E-04	2.36E-02	synaptotagmin 14 pseudogene 1
TSEN2	697.16	-0.598	1.07E-04	2.36E-02	tRNA splicing endonuclease subunit 2
WSCD2	36.26	-1.927	1.06E-04	2.36E-02	WSC domain containing 2
SNRPEP10	29.10	-1.669	1.10E-04	2.37E-02	SNRPE pseudogene 10
AC008870.5	87.06	-1.594	1.14E-04	2.46E-02	NA
CYSTM1	3382.28	-1.349	1.25E-04	2.64E-02	cysteine rich transmembrane module containing 1
PCYT1B	18.84	1.824	1.25E-04	2.64E-02	phosphate cytidylyltransferase 1B, choline
FOXH1	15.78	-1.289	1.27E-04	2.66E-02	forkhead box H1
AC093323.2	28.02	-1.120	1.29E-04	2.67E-02	NA
WTAPP1	64.32	-1.435	1.29E-04	2.67E-02	Wilms tumor 1 associated protein pseudogene 1
AC009133.1	385.41	-1.177	1.30E-04	2.67E-02	NA
DUSP27	72.35	-1.920	1.37E-04	2.78E-02	serine/threonine/tyrosine interacting like 2
GPER1	39.12	-1.443	1.37E-04	2.78E-02	G protein-coupled estrogen receptor 1
ATP4B	15.02	-1.983	1.42E-04	2.85E-02	ATPase H+/K+ transporting subunit beta
LIG4	549.24	0.521	1.43E-04	2.85E-02	DNA ligase 4
AATK	482.21	-1.385	1.48E-04	2.93E-02	apoptosis associated tyrosine kinase
PIPOX	65.64	1.200	1.53E-04	2.98E-02	pipecolic acid and sarcosine oxidase
RAB40B	512.33	-0.738	1.53E-04	2.98E-02	RAB40B, member RAS oncogene family
TNFSF14	91.68	1.225	1.54E-04	2.98E-02	TNF superfamily member 14
ADGRG2	67.40	-1.795	1.58E-04	3.01E-02	adhesion G protein-coupled receptor G2
DLX5	24.53	1.729	1.58E-04	3.01E-02	distal-less homeobox 5
GCC2	3390.21	-0.530	1.59E-04	3.02E-02	GRIP and coiled-coil domain containing 2
AC106782.5	9.37	-1.145	1.63E-04	3.07E-02	NA
MIR29B2CHG	1799.81	-0.888	1.65E-04	3.08E-02	MIR29B2 and MIR29C host gene
YWHAEP3	5.23	1.819	1.73E-04	3.21E-02	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein epsilon pseudogene 3
AL590627.1	42.84	-1.700	1.75E-04	3.23E-02	uncharacterized LOC651337
AL137143.1	21.05	-1.712	1.78E-04	3.25E-02	Metazoan signal recognition particle RNA
AL161431.1	88.95	1.850	1.82E-04	3.25E-02	NA
LILRA5	86.99	1.553	1.79E-04	3.25E-02	leukocyte immunoglobulin like receptor A5
NOTCH1	3163.55	0.804	1.82E-04	3.25E-02	notch receptor 1
TEKT5	12.94	-1.559	1.82E-04	3.25E-02	tektin 5

CDHR2	750.91	-1.517	1.86E-04	3.28E-02	cadherin related family member 2
VTN	64.24	1.768	1.86E-04	3.28E-02	vitronectin
HSPE1P5	8.18	1.742	1.91E-04	3.33E-02	heat shock protein family E (Hsp10) member 1 pseudogene 5
UNC5B-AS1	23.12	-1.731	1.91E-04	3.33E-02	UNC5B antisense RNA 1
C5AR1	319.14	1.244	1.93E-04	3.33E-02	complement C5a receptor 1
AC092338.3	28.87	-0.996	2.04E-04	3.38E-02	NA
AKR7A3	730.89	-1.377	1.98E-04	3.38E-02	aldo-keto reductase family 7 member A3
ELOVL4	75.06	1.572	2.06E-04	3.38E-02	ELOVL fatty acid elongase 4
LINC02481	69.11	-1.012	2.07E-04	3.38E-02	long intergenic non-protein coding RNA 2481
MT1F	60.55	-1.075	2.01E-04	3.38E-02	metallothionein 1F
OBP2B	5.40	1.667	2.02E-04	3.38E-02	odorant binding protein 2B
S100A7A	479.45	1.669	2.03E-04	3.38E-02	S100 calcium binding protein A7A
SPINK6	5.62	1.735	2.00E-04	3.38E-02	serine peptidase inhibitor Kazal type 6
TUBB2B	107.44	1.405	2.05E-04	3.38E-02	tubulin beta 2B class IIb
AL592166.1	15.38	-1.413	2.10E-04	3.41E-02	NA
SULT1C2P2	10.05	-1.804	2.11E-04	3.41E-02	sulfotransferase family, cytosolic, 1C, member 2 pseudogene 2
CBLIF	80.01	-1.710	2.23E-04	3.55E-02	cobalamin binding intrinsic factor
CYP17A1	10.71	-1.761	2.23E-04	3.55E-02	cytochrome P450 family 17 subfamily A member 1
PARM1	600.76	-1.201	2.21E-04	3.55E-02	prostate androgen-regulated mucin-like protein 1
LINC00242	8.71	1.289	2.29E-04	3.57E-02	long intergenic non-protein coding RNA 242
PM20D1	10.17	-1.573	2.27E-04	3.57E-02	peptidase M20 domain containing 1
STAR	6.30	1.533	2.27E-04	3.57E-02	steroidogenic acute regulatory protein
ENDOU	419.29	1.765	2.30E-04	3.57E-02	endonuclease, poly(U) specific
AC127521.1	18.85	-1.063	2.34E-04	3.60E-02	NA
TRABD2B	45.92	1.412	2.36E-04	3.60E-02	TraB domain containing 2B
XXYLT1-AS2	18.25	-1.570	2.35E-04	3.60E-02	XXYLT1 antisense RNA 2
OXCT1-AS1	3.33	-1.437	2.47E-04	3.75E-02	OXCT1 antisense RNA 1
LINC01300	2.65	-1.727	2.50E-04	3.76E-02	long intergenic non-protein coding RNA 1300
TMEM94	1658.14	-0.635	2.50E-04	3.76E-02	transmembrane protein 94
CASTOR1	139.16	1.039	2.57E-04	3.80E-02	cytosolic arginine sensor for mTORC1 subunit 1
FPR1	346.23	1.469	2.56E-04	3.80E-02	formyl peptide receptor 1
PNLIPRP3	14.00	1.709	2.57E-04	3.80E-02	pancreatic lipase related protein 3
CTSE	11553.03	-1.590	2.61E-04	3.82E-02	cathepsin E
LGALS7B	278.12	1.708	2.61E-04	3.82E-02	galectin 7B
AL355388.1	28.38	-0.740	2.68E-04	3.90E-02	ARHGEF2 antisense RNA 2
LIPF	493.39	-1.679	2.71E-04	3.92E-02	lipase F, gastric type
OR7E22P	11.15	-1.496	2.78E-04	3.98E-02	olfactory receptor family 7 subfamily E member 22 pseudogene
RFLNA	2029.56	-1.541	2.76E-04	3.98E-02	refilin A
AC099329.2	12.58	1.332	2.80E-04	3.98E-02	NA
MUC6	1482.10	-1.670	2.81E-04	3.98E-02	mucin 6, oligomeric mucus/gel-forming
NDUFA4L2	642.71	1.349	2.84E-04	4.00E-02	NDUFA4 mitochondrial complex associated like 2
PTPRD	139.16	1.140	2.87E-04	4.02E-02	protein tyrosine phosphatase receptor type D
XIST	80.47	1.642	2.96E-04	4.13E-02	X inactive specific transcript
AC084734.1	4.12	-1.618	3.06E-04	4.23E-02	NA
CLIC4P1	7.00	-1.370	3.05E-04	4.23E-02	chloride intracellular channel 4 pseudogene 1
AC016735.1	43.32	-1.394	3.11E-04	4.26E-02	NA
ANKRD35	316.25	1.525	3.11E-04	4.26E-02	ankyrin repeat domain 35
MIR657	3.94	-1.569	3.16E-04	4.30E-02	microRNA 657
CRYZP1	7.87	-1.180	3.21E-04	4.34E-02	crystallin zeta pseudogene 1
PLEKHA6	4033.44	-0.665	3.22E-04	4.34E-02	pleckstrin homology domain containing A6
SYTL2	6462.02	-1.271	3.27E-04	4.38E-02	synaptotagmin like 2
AL117344.2	53.78	-0.526	3.39E-04	4.51E-02	NA

NAPEPLD	853.80	-0.762	3.38E-04	4.51E-02	N-acyl phosphatidylethanolamine phospholipase D
STK32A	41.19	-1.428	3.42E-04	4.52E-02	serine/threonine kinase 32A
AC036176.3	44.27	1.379	3.50E-04	4.61E-02	NA
SLC34A3	47.55	1.524	3.59E-04	4.69E-02	solute carrier family 34 member 3
MYO5A	1437.16	0.687	3.61E-04	4.70E-02	myosin VA
ABCC2	355.89	1.365	3.82E-04	4.79E-02	ATP binding cassette subfamily C member 2
AC009501.2	14.12	-1.476	3.74E-04	4.79E-02	NA
AC092447.7	2.77	-1.543	3.72E-04	4.79E-02	NA
AC114291.1	18.84	1.502	3.79E-04	4.79E-02	uncharacterized LOC101928857
AL512598.1	9.07	-1.096	3.82E-04	4.79E-02	NA
CELF3	4.18	-1.429	3.78E-04	4.79E-02	CUGBP Elav-like family member 3
GIPR	261.34	-1.105	3.81E-04	4.79E-02	gastric inhibitory polypeptide receptor
HPDL	188.98	-1.056	3.83E-04	4.79E-02	4-hydroxyphenylpyruvate dioxygenase like
TENM2	538.85	1.531	3.82E-04	4.79E-02	teneurin transmembrane protein 2
VSIG8	15.66	1.432	3.95E-04	4.91E-02	V-set and immunoglobulin domain containing 8
HTR1B	81.36	-1.417	4.09E-04	5.07E-02	5-hydroxytryptamine receptor 1B
VILL	2078.82	-1.061	4.13E-04	5.09E-02	villin like
AC009090.3	5.50	-1.243	4.31E-04	5.15E-02	NA
AC087857.1	2.96	-1.474	4.20E-04	5.15E-02	NA
CAPN9	427.89	-1.480	4.25E-04	5.15E-02	calpain 9
HHEX	304.45	1.075	4.23E-04	5.15E-02	hematopoietically expressed homeobox
KRT16P5	12.87	1.384	4.30E-04	5.15E-02	keratin 16 pseudogene 5
SERPINB4	1766.99	1.423	4.30E-04	5.15E-02	serpin family B member 4
SMIM32	21.11	-1.437	4.27E-04	5.15E-02	small integral membrane protein 32
MST1R	1718.02	-0.808	4.35E-04	5.18E-02	macrophage stimulating 1 receptor
LIPN	14.71	1.261	4.43E-04	5.25E-02	lipase family member N
NEIL2	414.50	-0.998	4.46E-04	5.25E-02	nei like DNA glycosylase 2
PSMD10P2	38.38	-1.284	4.46E-04	5.25E-02	proteasome 26S subunit, non-ATPase, 10 pseudogene 2
PRR15L	1979.11	-0.959	4.52E-04	5.29E-02	proline rich 15 like
NECTIN2	2880.55	-0.438	4.56E-04	5.31E-02	nectin cell adhesion molecule 2
GK	967.42	0.850	4.62E-04	5.35E-02	glycerol kinase
MAST1	19.84	-1.092	4.63E-04	5.35E-02	microtubule associated serine/threonine kinase 1
SLC45A3	746.53	-1.024	4.69E-04	5.39E-02	solute carrier family 45 member 3
ARL14	785.70	-1.232	4.74E-04	5.43E-02	ADP ribosylation factor like GTPase 14
RRNAD1	450.36	-0.415	4.78E-04	5.46E-02	methyltransferase like 25B
MT1DP	2.51	-1.303	4.81E-04	5.47E-02	metallothionein 1D, pseudogene
ARHGAP12	3634.77	-0.601	4.83E-04	5.47E-02	Rho GTPase activating protein 12
AP000967.1	13.66	-0.834	4.91E-04	5.53E-02	Y RNA
AC008397.1	101.43	-1.361	5.07E-04	5.68E-02	uncharacterized LOC729966
LINC02806	12.61	-1.368	5.08E-04	5.68E-02	long intergenic non-protein coding RNA 2806
B3GALT1	5.38	-1.255	5.10E-04	5.68E-02	beta-1,3-galactosyltransferase 1
EHMT2-AS1	71.48	-1.193	5.14E-04	5.70E-02	EHMT2 and SLC44A4 antisense RNA 1
LINC02457	9.47	1.384	5.21E-04	5.75E-02	long intergenic non-protein coding RNA 2457
AL121987.1	11.94	-1.233	5.33E-04	5.84E-02	NA
COLEC11	46.56	1.181	5.35E-04	5.84E-02	collectin subfamily member 11
IL36G	207.32	1.370	5.35E-04	5.84E-02	interleukin 36 gamma
SLC9A1	2941.04	-0.868	5.46E-04	5.93E-02	solute carrier family 9 member A1
MUC16	392.30	1.289	5.53E-04	5.98E-02	mucin 16, cell surface associated
AC025884.1	5.19	1.254	5.57E-04	6.01E-02	NA
AC130456.2	107.79	-1.090	5.60E-04	6.01E-02	NA
FGF19	9.92	1.327	5.65E-04	6.04E-02	fibroblast growth factor 19
AC005329.1	12.32	1.128	5.68E-04	6.06E-02	NA
SESN3	2006.74	1.004	5.72E-04	6.07E-02	sestrin 3
F3	2188.79	0.758	5.77E-04	6.10E-02	coagulation factor III, tissue factor

CLDN23	643.86	-1.028	5.90E-04	6.20E-02	claudin 23
LINC00342	3314.88	-0.894	5.93E-04	6.20E-02	long intergenic non-protein coding RNA 342
RPL17P10	13.91	-0.727	5.93E-04	6.20E-02	ribosomal protein L17 pseudogene 10
FBP2	26.05	-1.316	5.97E-04	6.21E-02	fructose-bisphosphatase 2
TPST2	583.72	0.590	6.04E-04	6.27E-02	tyrosylprotein sulfotransferase 2
LINC01115	14.77	-1.279	6.08E-04	6.28E-02	long intergenic non-protein coding RNA 1115
AC093583.1	11.46	-1.163	6.23E-04	6.38E-02	NA
AC130456.7	73.12	-1.056	6.20E-04	6.38E-02	NA
KRT15	4598.82	1.269	6.26E-04	6.38E-02	keratin 15
SIX2	40.38	1.285	6.27E-04	6.38E-02	SIX homeobox 2
ETNK1	4606.62	-0.559	6.30E-04	6.39E-02	ethanolamine kinase 1
SLC25A25-AS1	1522.47	-0.676	6.33E-04	6.40E-02	SLC25A25 antisense RNA 1
CYP4F29P	102.29	1.277	6.36E-04	6.41E-02	cytochrome P450 family 4 subfamily F member 29, pseudogene
AC021723.2	13.22	-1.247	6.47E-04	6.47E-02	NA
DGKD	2539.43	-0.838	6.47E-04	6.47E-02	diacylglycerol kinase delta
LINC01978	216.89	-1.030	6.53E-04	6.50E-02	long intergenic non-protein coding RNA 1978
LINC01133	928.13	-0.927	6.81E-04	6.76E-02	long intergenic non-protein coding RNA 1133
WARS2-AS1	752.00	-0.631	6.89E-04	6.82E-02	WARS2 antisense RNA 1
ALDH1A3	1640.09	0.981	6.95E-04	6.84E-02	aldehyde dehydrogenase 1 family member A3
LINC02805	8.25	-1.182	6.97E-04	6.84E-02	long intergenic non-protein coding RNA 2805
AC012236.1	70.26	-1.055	7.22E-04	6.84E-02	NA
CXCR1	148.32	1.160	7.18E-04	6.84E-02	C-X-C motif chemokine receptor 1
FAM118B	587.16	-0.458	7.06E-04	6.84E-02	family with sequence similarity 118 member B
FAM135B	14.23	-1.164	7.07E-04	6.84E-02	family with sequence similarity 135 member B
FGD3	410.73	0.703	7.21E-04	6.84E-02	FYVE, RhoGEF and PH domain containing 3
ITPRID2	7579.28	-0.565	7.14E-04	6.84E-02	ITPR interacting domain containing 2
LINC02435	3.16	-0.964	7.14E-04	6.84E-02	long intergenic non-protein coding RNA 2435
MAGIX	149.95	-0.652	7.03E-04	6.84E-02	MAGI family member, X-linked
MAP7D2	113.57	-1.139	7.16E-04	6.84E-02	MAP7 domain containing 2
MCU	2705.50	-0.543	7.20E-04	6.84E-02	mitochondrial calcium uniporter
AC009962.1	37.24	-0.941	7.33E-04	6.93E-02	NA
IL1R2	1093.42	-0.952	7.48E-04	7.04E-02	interleukin 1 receptor type 2
STEAP3	1340.96	-0.504	7.52E-04	7.06E-02	STEAP3 metalloreductase
CRACR2B	912.71	-0.870	7.63E-04	7.13E-02	calcium release activated channel regulator 2B
SLC39A14	4627.15	-0.678	7.70E-04	7.15E-02	solute carrier family 39 member 14
SULT1C2	3976.02	-1.193	7.67E-04	7.15E-02	sulfotransferase family 1C member 2
TRIM54	28.61	-1.134	7.79E-04	7.21E-02	tripartite motif containing 54
ABCC8	19.22	-1.178	7.92E-04	7.28E-02	ATP binding cassette subfamily C member 8
PLEKHG2	1208.26	0.533	7.90E-04	7.28E-02	pleckstrin homology and RhoGEF domain containing G2
L1CAM	138.80	1.114	7.98E-04	7.32E-02	L1 cell adhesion molecule
AC139495.2	30.18	0.709	8.11E-04	7.40E-02	NA
TSGA10	331.44	-0.454	8.13E-04	7.40E-02	testis specific 10
EPAS1	4280.99	0.371	8.17E-04	7.41E-02	endothelial PAS domain protein 1
PTPRQ	21.99	1.168	8.21E-04	7.42E-02	protein tyrosine phosphatase receptor type Q
MIA3	3392.68	-0.423	8.30E-04	7.48E-02	MIA SH3 domain ER export factor 3
Z97056.2	5.52	1.164	8.33E-04	7.48E-02	NA
CP	674.20	1.126	8.46E-04	7.55E-02	ceruloplasmin
DNMBP-AS1	27.79	-1.032	8.54E-04	7.55E-02	DNMBP antisense RNA 1
FAM89A	192.03	0.601	8.53E-04	7.55E-02	family with sequence similarity 89 member A
OR7E161P	5.86	-1.110	8.50E-04	7.55E-02	olfactory receptor family 7 subfamily E member 161 pseudogene
TNFAIP8L3	77.98	1.007	8.45E-04	7.55E-02	TNF alpha induced protein 8 like 3

AC009093.4	5.29	-0.988	8.65E-04	7.62E-02	NA
FAM222A	195.57	0.681	8.72E-04	7.66E-02	family with sequence similarity 222 member A
PLLP	572.59	-0.716	8.77E-04	7.67E-02	plasmolipin
AL033528.3	7.96	-0.712	8.84E-04	7.72E-02	NA
LCP2	1225.93	0.875	8.87E-04	7.72E-02	lymphocyte cytosolic protein 2
AL035071.1	178.98	0.468	8.94E-04	7.75E-02	NA
MT1H	40.53	-1.077	8.97E-04	7.75E-02	metallothionein 1H
ACP6	799.80	-0.408	9.06E-04	7.75E-02	acid phosphatase 6, lysophosphatidic
CYP3A4	11.87	-1.014	9.03E-04	7.75E-02	cytochrome P450 family 3 subfamily A member 4
KIF5C	88.98	-0.952	9.08E-04	7.75E-02	kinesin family member 5C
PDZD3	160.15	-1.060	9.04E-04	7.75E-02	PDZ domain containing 3
CMBL	1019.54	-0.879	9.13E-04	7.77E-02	carboxymethylenebutenolidase homolog
SCIN	788.89	-0.864	9.19E-04	7.80E-02	scinderin
CD2BP2-DT	42.84	-0.591	9.31E-04	7.88E-02	CD2BP2 divergent transcript
AL035413.2	25.86	-0.942	9.42E-04	7.92E-02	NA
MIR4251	5.27	-1.091	9.42E-04	7.92E-02	microRNA 4251
AC097521.1	3.36	-0.946	9.46E-04	7.93E-02	NA
AC140479.6	19.55	1.105	9.53E-04	7.94E-02	NA
SAMSN1	322.38	0.814	9.52E-04	7.94E-02	SAM domain, SH3 domain and nuclear localization signals 1
ANO1-AS1	2.75	-1.041	9.67E-04	7.96E-02	ANO1 antisense RNA 1
BBOX1	59.02	1.093	9.64E-04	7.96E-02	gamma-butyrobetaine hydroxylase 1
DEFB4A	168.29	1.000	9.66E-04	7.96E-02	defensin beta 4A
SAP30	225.72	-0.599	9.62E-04	7.96E-02	Sin3A associated protein 30
AC013726.2	4.90	-1.037	9.93E-04	8.11E-02	NA
AP003900.1	13.29	-1.038	9.94E-04	8.11E-02	NA
GALNT13	31.86	-1.087	9.94E-04	8.11E-02	polypeptide N-acetylgalactosaminyltransferase 13
PTGER2	148.84	-0.932	1.00E-03	8.15E-02	prostaglandin E receptor 2
EDA2R	39.40	0.717	1.03E-03	8.35E-02	ectodysplasin A2 receptor
FHL2	1151.54	-0.565	1.04E-03	8.36E-02	four and a half LIM domains 2
PKDREJ	15.74	-0.913	1.04E-03	8.36E-02	polycystin family receptor for egg jelly
CHRD	114.22	0.914	1.07E-03	8.55E-02	chordin
CSGALNACT 1	380.26	0.501	1.07E-03	8.55E-02	chondroitin sulfate N-acetylgalactosaminyltransferase 1
AC093849.3	11.51	-0.915	1.08E-03	8.56E-02	NA
FLG	401.54	1.001	1.07E-03	8.56E-02	filaggrin
LINC01003	54.56	-0.510	1.08E-03	8.56E-02	long intergenic non-protein coding RNA 1003
AC005796.1	2.93	-1.033	1.09E-03	8.59E-02	NA
AL139095.3	3.23	1.023	1.09E-03	8.59E-02	NA
ERN2	2089.88	-0.875	1.10E-03	8.59E-02	endoplasmic reticulum to nucleus signaling 2
HCN4	16.68	-0.994	1.10E-03	8.59E-02	hyperpolarization activated cyclic nucleotide gated potassium channel 4
LUCAT1	431.12	0.865	1.10E-03	8.59E-02	lung cancer associated transcript 1
AC011369.2	12.08	1.035	1.11E-03	8.59E-02	NA
CA9	1112.04	-1.036	1.11E-03	8.59E-02	carbonic anhydrase 9
FAM25BP	3.12	0.954	1.10E-03	8.59E-02	family with sequence similarity 25 member B, pseudogene
FER1L6	3661.40	-1.036	1.11E-03	8.59E-02	fer-1 like family member 6
FAM221A	468.54	-0.514	1.12E-03	8.64E-02	family with sequence similarity 221 member A
IGHD	51.27	-1.027	1.12E-03	8.64E-02	immunoglobulin heavy constant delta
AC137630.3	7.62	0.949	1.13E-03	8.66E-02	NA
KBTBD12	205.25	-0.985	1.14E-03	8.69E-02	kelch repeat and BTB domain containing 12
AC007683.1	54.60	-0.713	1.14E-03	8.69E-02	NA
CLMAT3	7.76	-1.000	1.14E-03	8.71E-02	colorectal liver metastasis associated transcript 3
AL391839.3	5.55	-0.820	1.16E-03	8.73E-02	NA
CAPN5	3222.02	-0.563	1.16E-03	8.73E-02	calpain 5

OIT3	56.61	-0.906	1.15E-03	8.73E-02	oncoprotein induced transcript 3
AC127526.3	2.86	-1.001	1.18E-03	8.74E-02	NA
CYP3A5	4084.72	-0.680	1.17E-03	8.74E-02	cytochrome P450 family 3 subfamily A member 5
LY6G6C	24.42	0.876	1.17E-03	8.74E-02	lymphocyte antigen 6 family member G6C
RBP1	279.92	0.903	1.17E-03	8.74E-02	retinol binding protein 1
RPL22P1	78.42	-0.580	1.18E-03	8.74E-02	ribosomal protein L22 pseudogene 1
SLC30A8	4.33	-1.010	1.18E-03	8.74E-02	solute carrier family 30 member 8
SSTR1	298.46	-0.970	1.17E-03	8.74E-02	somatostatin receptor 1
ATF7IP2	1104.19	-0.678	1.18E-03	8.74E-02	activating transcription factor 7 interacting protein 2
THSD1	242.71	0.668	1.19E-03	8.77E-02	thrombospondin type 1 domain containing 1
AL357078.2	15.88	-0.738	1.20E-03	8.81E-02	NA
HDHD2	538.08	0.333	1.20E-03	8.81E-02	haloacid dehalogenase like hydrolase domain containing 2
CPA4	257.71	0.998	1.21E-03	8.81E-02	carboxypeptidase A4
SNORA11B	11.21	-0.538	1.21E-03	8.85E-02	small nucleolar RNA, H/ACA box 11B
AC006026.3	5.53	-0.801	1.22E-03	8.86E-02	NA
MYT1	6.41	-0.976	1.22E-03	8.88E-02	myelin transcription factor 1
CAPN13	614.97	-0.880	1.24E-03	8.97E-02	calpain 13
LYPD5	356.02	0.895	1.24E-03	8.98E-02	LY6/PLAUR domain containing 5
AC025580.2	285.78	-0.755	1.26E-03	9.03E-02	NA
BCAS2P1	2.01	-0.901	1.26E-03	9.03E-02	BCAS2 pseudogene 1
KRT10	190.60	0.690	1.26E-03	9.03E-02	keratin 10
MALINC1	42.16	-0.540	1.28E-03	9.07E-02	mitosis associated long intergenic non-coding RNA 1
NIN	1992.36	0.454	1.27E-03	9.07E-02	ninein
SMIM6	169.01	-0.941	1.28E-03	9.07E-02	small integral membrane protein 6
HOXA9	116.86	0.881	1.28E-03	9.08E-02	homeobox A9
CASS4	150.26	0.795	1.29E-03	9.11E-02	Cas scaffold protein family member 4
TMEM71	108.65	0.773	1.30E-03	9.15E-02	transmembrane protein 71
GKN2	250.13	-0.859	1.31E-03	9.21E-02	gastrokine 2
AC112492.1	3.03	0.598	1.34E-03	9.33E-02	ORC1 pseudogene 1
CYMP	9.91	-0.895	1.34E-03	9.33E-02	chymosin, pseudogene
CYP2AB1P	63.13	-0.938	1.34E-03	9.33E-02	cytochrome P450 family 2 subfamily AB member 1, pseudogene
MECOM	6411.86	-0.565	1.33E-03	9.33E-02	MDS1 and EVI1 complex locus
VWA3B	46.30	-0.883	1.33E-03	9.33E-02	von Willebrand factor A domain containing 3B
AADACP1	220.67	-0.911	1.35E-03	9.36E-02	arylacetamide deacetylase pseudogene 1
ZNF514	952.90	-0.321	1.35E-03	9.36E-02	zinc finger protein 514
AC009495.1	10.52	-0.637	1.37E-03	9.36E-02	NA
AC010998.3	23.20	-0.947	1.37E-03	9.36E-02	NA
AC090181.2	184.00	-0.479	1.36E-03	9.36E-02	NA
C13orf46	164.03	-0.887	1.37E-03	9.36E-02	chromosome 13 open reading frame 46
NBPF10	330.89	0.467	1.37E-03	9.36E-02	NBPF member 10
CEP170P1	21.04	0.614	1.40E-03	9.50E-02	centrosomal protein 170 pseudogene 1
AC011379.2	97.50	-0.634	1.40E-03	9.52E-02	uncharacterized LOC101929719
KLK7	2165.99	0.925	1.41E-03	9.54E-02	kallikrein related peptidase 7
AL357972.1	5.28	-0.905	1.44E-03	9.58E-02	NA
CREB3L1	2296.22	-0.800	1.44E-03	9.58E-02	cAMP responsive element binding protein 3 like 1
GPAT2	216.28	-0.903	1.44E-03	9.58E-02	glycerol-3-phosphate acyltransferase 2, mitochondrial
NEFL	78.01	0.910	1.44E-03	9.58E-02	neurofilament light chain
RFTN1	514.51	0.439	1.44E-03	9.58E-02	raftlin, lipid raft linker 1
STAT3	8883.69	0.377	1.43E-03	9.58E-02	signal transducer and activator of transcription 3
TMEM132D	4.45	-0.914	1.44E-03	9.58E-02	transmembrane protein 132D
ZDHHC7	2445.16	-0.398	1.44E-03	9.58E-02	zinc finger DHHC-type palmitoyltransferase 7

TEX35	21.32	-0.869	1.45E-03	9.58E-02	testis expressed 35
IGKV2-24	56.87	0.886	1.46E-03	9.62E-02	immunoglobulin kappa variable 2-24
HMGN2P28	29.29	-0.596	1.48E-03	9.75E-02	high mobility group nucleosomal binding domain 2 pseudogene 28
COL11A2	154.33	0.780	1.50E-03	9.87E-02	collagen type XI alpha 2 chain
PFKFB2	2970.56	-0.414	1.50E-03	9.87E-02	6-phosphofructo-2-kinase/ fructose-2,6-biphosphatase 2
DLX1	9.71	0.867	1.52E-03	9.94E-02	distal-less homeobox 1
ALDH2	2852.59	-0.496	1.53E-03	9.96E-02	aldehyde dehydrogenase 2 family member
IGKV2-30	94.42	0.893	1.53E-03	9.96E-02	immunoglobulin kappa variable 2-30
KCNH3	36.90	0.819	1.53E-03	9.96E-02	potassium voltage-gated channel subfamily H member 3
EFEMP1	577.28	0.710	1.54E-03	9.98E-02	EGF containing fibulin extracellular matrix protein 1

Appendix C-4: Clinical Benefit vs. No Clinical Benefit (ICI-only OAC)

Gene	baseMean	log2FC	pvalue	padj	description
AC091814.1	19.50	3.756	7.25E-09	9.71E-05	NA
LINC02432	254.03	-3.641	7.21E-09	9.71E-05	long intergenic non-protein coding RNA 2432
RIMS4	21.20	-4.093	3.37E-08	3.01E-04	regulating synaptic membrane exocytosis 4
VIM-AS1	232.38	2.932	1.31E-07	8.80E-04	VIM antisense RNA 1
CD177	354.86	2.901	5.67E-07	3.04E-03	CD177 molecule
AC115618.1	21.27	2.790	8.26E-07	3.69E-03	NA
DPEP1	66.88	3.169	1.24E-06	4.74E-03	dipeptidase 1
AC025164.2	78.44	2.670	2.02E-06	6.02E-03	NA
GP2	564.21	-3.058	1.81E-06	6.02E-03	glycoprotein 2
AL096677.1	7.35	-2.881	2.77E-06	6.14E-03	NA
FGG	5.96	-3.260	2.85E-06	6.14E-03	fibrinogen gamma chain
SPINK13	9.90	-3.356	2.41E-06	6.14E-03	serine peptidase inhibitor Kazal type 13
TRAJ39	7.08	2.418	2.98E-06	6.14E-03	T cell receptor alpha joining 39
LINC02532	242.12	-2.719	4.60E-06	8.80E-03	long intergenic non-protein coding RNA 2532
ECEL1P2	7.31	-3.048	6.36E-06	1.14E-02	endothelin converting enzyme like 1 pseudogene 2
HILPDA	644.48	2.006	6.85E-06	1.15E-02	hypoxia inducible lipid droplet associated
CHMP1B-AS1	20.93	2.506	7.66E-06	1.21E-02	CHMP1B antisense RNA 1
AC010148.1	31.64	-1.149	9.81E-06	1.34E-02	NA
AC044849.1	14.83	2.813	1.15E-05	1.34E-02	NA
AL138963.4	171.88	2.531	1.02E-05	1.34E-02	NA
BEST1	298.72	1.826	1.11E-05	1.34E-02	bestrophin 1
C11orf86	57.96	2.676	1.13E-05	1.34E-02	chromosome 11 open reading frame 86
VTCN1	121.75	-2.663	1.09E-05	1.34E-02	V-set domain containing T cell activation inhibitor 1
ASB11	6.23	-2.833	1.27E-05	1.41E-02	ankyrin repeat and SOCS box containing 11
CFHR3	6.37	-2.622	1.42E-05	1.53E-02	complement factor H related 3
AL645941.3	12.52	2.485	1.81E-05	1.80E-02	NA
COL2A1	50.73	-2.715	1.81E-05	1.80E-02	collagen type II alpha 1 chain
AIM2	123.30	1.470	2.33E-05	2.23E-02	absent in melanoma 2
ATG10	481.69	1.740	2.43E-05	2.24E-02	autophagy related 10
CHST4	71.16	-2.400	3.22E-05	2.80E-02	carbohydrate sulfotransferase 4
DNAH5	431.01	-1.766	3.23E-05	2.80E-02	dynein axonemal heavy chain 5
AC244100.2	11.38	2.448	3.50E-05	2.93E-02	NA
RERGL	25.36	-2.344	3.88E-05	3.15E-02	RERG like
AL671277.1	92.72	2.152	4.02E-05	3.17E-02	NA
PYCARD-AS1	27.55	1.420	4.18E-05	3.20E-02	PYCARD antisense RNA 1
CLIC6	488.92	-1.956	4.71E-05	3.37E-02	chloride intracellular channel 6
RN7SKP97	68.90	-1.273	4.78E-05	3.37E-02	RN7SK pseudogene 97
UPK1A	148.62	2.318	4.70E-05	3.37E-02	uroplakin 1A
AL592166.1	47.80	2.270	4.92E-05	3.38E-02	NA
ABCA9-AS1	11.66	-2.271	5.26E-05	3.52E-02	ABCA9 antisense RNA 1
ACVR1C	119.56	-1.358	5.58E-05	3.65E-02	activin A receptor type 1C
AC133644.1	9.10	1.750	5.97E-05	3.81E-02	NA

MIR155HG	48.19	1.447	6.15E-05	3.83E-02	MIR155 host gene
PRSS2	95.78	-2.020	6.49E-05	3.95E-02	serine protease 2
AC005703.6	12.11	2.170	6.80E-05	4.05E-02	NA
AC018761.2	13.64	2.054	7.14E-05	4.11E-02	NA
AC120057.3	28.48	2.037	7.64E-05	4.11E-02	NA
FCRL3	119.49	1.911	7.23E-05	4.11E-02	Fc receptor like 3
LCAL1	230.52	-2.122	7.56E-05	4.11E-02	lung cancer associated lncRNA 1
TRAJ4	9.89	1.738	7.67E-05	4.11E-02	T cell receptor alpha joining 4
TRAJ17	7.82	1.749	7.88E-05	4.14E-02	T cell receptor alpha joining 17
AP003465.2	15.19	2.066	8.52E-05	4.39E-02	NA
AC004847.1	17.73	1.296	8.95E-05	4.44E-02	NA
AC135983.3	198.93	0.701	9.57E-05	4.44E-02	NA
EZR-AS1	12.77	1.981	9.18E-05	4.44E-02	EZR antisense RNA 1
HIF1A-AS1	8.18	1.657	9.77E-05	4.44E-02	HIF1A antisense RNA 1
NR1I3	3.46	1.903	9.68E-05	4.44E-02	nuclear receptor subfamily 1 group I member 3
ODAM	11.63	-2.029	9.26E-05	4.44E-02	odontogenic, ameloblast associated
U47924.2	26.80	1.788	9.39E-05	4.44E-02	NA
ASPH	12094.72	-1.178	1.09E-04	4.63E-02	aspartate beta-hydroxylase
OSR1	42.96	-1.142	1.04E-04	4.63E-02	odd-skipped related transcription factor 1
TFPI2	119.59	-1.935	1.08E-04	4.63E-02	tissue factor pathway inhibitor 2
TRAJ32	7.60	1.741	1.09E-04	4.63E-02	T cell receptor alpha joining 32
C5	406.66	-1.366	1.13E-04	4.73E-02	complement C5
ERICH3	39.93	-1.779	1.27E-04	5.22E-02	glutamate rich 3
AC012645.3	17.80	1.761	1.31E-04	5.27E-02	CORO1A antisense RNA 1
AL139275.2	3.60	-1.523	1.32E-04	5.27E-02	NA
AC007991.2	9.90	1.677	1.36E-04	5.35E-02	NA
DEPP1	548.22	-0.994	1.43E-04	5.49E-02	DEPP1 autophagy regulator
NXF3	47.61	-1.784	1.43E-04	5.49E-02	nuclear RNA export factor 3
AC090340.1	11.51	-0.848	1.49E-04	5.54E-02	NA
LRRIQ4	17.11	-1.708	1.51E-04	5.54E-02	leucine rich repeats and IQ motif containing 4
TANGO6	547.99	0.413	1.53E-04	5.54E-02	transport and golgi organization 6 homolog
TRAJ14	10.45	1.480	1.47E-04	5.54E-02	T cell receptor alpha joining 14
TRAJ38	8.46	1.683	1.57E-04	5.56E-02	T cell receptor alpha joining 38
VWDE	145.36	-1.597	1.58E-04	5.56E-02	von Willebrand factor D and EGF domains
AP002433.1	4.66	1.671	1.68E-04	5.86E-02	NA
AC020656.2	283.06	-1.533	1.78E-04	6.11E-02	NA
APOD	196.15	-1.559	1.81E-04	6.14E-02	apolipoprotein D
AC092053.3	4.96	1.629	1.90E-04	6.35E-02	NA
RPL10P6	15.05	1.532	1.93E-04	6.36E-02	ribosomal protein L10 pseudogene 6
TRAJ12	12.04	1.219	1.95E-04	6.36E-02	T cell receptor alpha joining 12
AL355482.1	10.35	-1.389	1.99E-04	6.43E-02	NA
COL9A3	120.71	-1.361	2.05E-04	6.52E-02	collagen type IX alpha 3 chain
CSTL1	16.62	-1.525	2.09E-04	6.52E-02	cystatin like 1
FGA	89.56	-1.307	2.07E-04	6.52E-02	fibrinogen alpha chain
RN7SL471P	73.12	1.456	2.13E-04	6.55E-02	RNA, 7SL, cytoplasmic 471, pseudogene
LRRIQ1	28.92	-1.464	2.21E-04	6.73E-02	leucine rich repeats and IQ motif containing 1

AC096589.1	11.64	-1.538	2.28E-04	6.75E-02	NA
AL392086.3	15.27	1.402	2.29E-04	6.75E-02	NA
HHIPL2	20.36	-1.224	2.27E-04	6.75E-02	HHIP like 2
SFTPA2	23.93	-1.451	2.32E-04	6.77E-02	surfactant protein A2
AC138627.1	21.73	-1.293	2.45E-04	6.99E-02	PSMD7 divergent transcript
BCL2L15	1262.16	-1.282	2.44E-04	6.99E-02	BCL2 like 15
CLEC2D	574.38	0.938	2.48E-04	7.00E-02	C-type lectin domain family 2 member D
AL138767.1	4.78	1.338	2.51E-04	7.02E-02	NA
CLMAT3	36.64	1.338	2.54E-04	7.03E-02	colorectal liver metastasis associated transcript 3
C12orf60	68.92	1.311	2.59E-04	7.09E-02	chromosome 12 open reading frame 60
CPB2-AS1	17.87	1.330	2.66E-04	7.21E-02	CPB2 antisense RNA 1
AL035252.6	15.35	-1.190	2.82E-04	7.39E-02	NA
CCKAR	7.62	-1.073	2.82E-04	7.39E-02	cholecystokinin A receptor
TRIM50	17.44	-1.365	2.84E-04	7.39E-02	tripartite motif containing 50
TTN	2140.66	1.166	2.77E-04	7.39E-02	titin
AC007381.1	8.05	1.329	2.95E-04	7.42E-02	MIR4432 host gene
AC092661.1	3.48	-1.366	2.98E-04	7.42E-02	NA
LINC02068	42.16	-1.056	2.93E-04	7.42E-02	long intergenic non-protein coding RNA 2068
MYO1G	383.50	0.777	3.02E-04	7.42E-02	myosin IG
PYHIN1	148.07	1.156	3.00E-04	7.42E-02	pyrin and HIN domain family member 1
TRAJ30	6.50	1.278	2.96E-04	7.42E-02	T cell receptor alpha joining 30
TRAJ36	7.13	1.259	3.16E-04	7.71E-02	T cell receptor alpha joining 36
CCDC162P	323.44	-1.128	3.28E-04	7.74E-02	coiled-coil domain containing 162, pseudogene
OVGP1	53.02	-0.523	3.24E-04	7.74E-02	oviductal glycoprotein 1
PGA5	61.70	-0.965	3.23E-04	7.74E-02	pepsinogen A5
PGC	14352.15	-1.285	3.29E-04	7.74E-02	progastricsin
SCARNA10	6.98	0.939	3.39E-04	7.89E-02	small Cajal body-specific RNA 10
MIR1282	30.63	1.331	3.50E-04	7.94E-02	microRNA 1282
TNPO1P2	12.09	-1.280	3.45E-04	7.94E-02	transportin 1 pseudogene 2
ZNF831	227.78	1.125	3.49E-04	7.94E-02	zinc finger protein 831
GPR174	285.09	1.098	3.56E-04	8.01E-02	G protein-coupled receptor 174
LINC00342	3722.12	-0.999	3.67E-04	8.06E-02	long intergenic non-protein coding RNA 342
NT5C3AP2	32.95	1.143	3.64E-04	8.06E-02	NT5C3A pseudogene 2
PLAAT5	30.98	-1.343	3.64E-04	8.06E-02	phospholipase A and acyltransferase 5
LCK	321.12	0.969	3.77E-04	8.22E-02	LCK proto-oncogene, Src family tyrosine kinase
AC188616.1	10.00	-1.276	4.01E-04	8.41E-02	uncharacterized LOC105375118
AL359764.1	5.02	-1.172	3.96E-04	8.41E-02	NA
CORIN	49.44	-1.037	4.02E-04	8.41E-02	corin, serine peptidase
HMGB1P11	3.89	-1.280	3.95E-04	8.41E-02	high mobility group box 1 pseudogene 11
PGA3	231.05	-0.757	3.92E-04	8.41E-02	pepsinogen A3
SNORA54	193.67	0.964	4.07E-04	8.45E-02	small nucleolar RNA, H/ACA box 54
AC096589.2	5.82	-1.052	4.13E-04	8.46E-02	NA
LINC01719	26.56	-0.770	4.12E-04	8.46E-02	long intergenic non-protein coding RNA 1719
AC046129.1	5.92	-1.267	4.23E-04	8.50E-02	NA
ARHGAP31-AS1	5.71	-1.115	4.28E-04	8.50E-02	ARHGAP31 antisense RNA 1

BPIFB4	7.98	-1.104	4.19E-04	8.50E-02	BPI fold containing family B member 4
TRGC1	30.79	0.958	4.28E-04	8.50E-02	TCR gamma alternate reading frame protein
SLC22A31	12.76	-1.238	4.33E-04	8.53E-02	solute carrier family 22 member 31
AC011446.2	13.26	1.241	4.42E-04	8.64E-02	NA
AL122035.1	19.45	0.640	4.50E-04	8.72E-02	NA
AL671883.2	50.25	1.172	4.56E-04	8.72E-02	NA
LINC00930	27.33	-1.231	4.53E-04	8.72E-02	long intergenic non-protein coding RNA 930
P4HA2	956.25	-0.289	4.74E-04	9.01E-02	prolyl 4-hydroxylase subunit alpha 2
AC090954.1	3.37	-1.158	5.02E-04	9.35E-02	NA
AC105020.6	21.16	0.767	4.96E-04	9.35E-02	NA
LINC02456	8.03	-1.160	4.99E-04	9.35E-02	long intergenic non-protein coding RNA 2456
AC010761.1	50.39	1.049	5.28E-04	9.69E-02	NA
TRAJ27	6.81	1.085	5.25E-04	9.69E-02	T cell receptor alpha joining 27
AC004593.2	11.84	1.131	5.54E-04	9.87E-02	chimerin 2
AC011815.2	10.00	-0.936	5.46E-04	9.87E-02	NA
AF129075.3	11.97	0.906	5.82E-04	9.87E-02	NA
AL159972.1	5.63	1.054	5.64E-04	9.87E-02	NA
AP001107.1	11.13	1.029	5.79E-04	9.87E-02	NA
GABRB2	35.66	-1.102	5.77E-04	9.87E-02	gamma-aminobutyric acid type A receptor subunit beta2
GVINP1	855.87	0.719	5.70E-04	9.87E-02	GTPase, very large interferon inducible pseudogene 1
MFAP5	54.49	-1.072	5.77E-04	9.87E-02	microfibril associated protein 5
SNX32	42.99	1.066	5.58E-04	9.87E-02	sorting nexin 32
TEX19	3.88	-1.125	5.60E-04	9.87E-02	testis expressed 19
TLL2	70.24	-0.789	5.63E-04	9.87E-02	tolloid like 2
TRAJ31	9.58	0.978	5.48E-04	9.87E-02	T cell receptor alpha joining 31

Appendix C-5: Clinical Benefit vs. No Clinical Benefit (PostTx OAC)

Gene	baseMean	log2FC	pvalue	padj	description
NFE4	16.18	-22.869	9.09E-15	2.14E-10	nuclear factor, erythroid 4
ZFAND3	1202.49	-0.684	6.40E-10	7.52E-06	zinc finger AN1-type containing 3
DLGAP4	1163.83	-0.664	3.82E-08	2.99E-04	DLG associated protein 4
ABCC2	617.89	-4.500	5.81E-08	3.41E-04	ATP binding cassette subfamily C member 2
SMG7	2481.08	-0.681	7.90E-08	3.72E-04	SMG7 nonsense mediated mRNA decay factor
CDR2L	579.17	-1.912	5.39E-07	2.11E-03	cerebellar degeneration related protein 2 like
NCKIPSD	512.95	-0.654	6.67E-07	2.24E-03	NCK interacting protein with SH3 domain
ADAMTS8	31.43	-4.015	7.90E-07	2.32E-03	ADAM metallopeptidase with thrombospondin type 1 motif 8
AP003119.1	50.90	-4.091	9.70E-07	2.53E-03	NA
MBD6	1282.66	-0.709	1.11E-06	2.61E-03	methyl-CpG binding domain protein 6
CCL26	131.36	4.131	1.61E-06	3.15E-03	C-C motif chemokine ligand 26
MAK	19.38	-3.419	1.50E-06	3.15E-03	male germ cell associated kinase
TBC1D32	361.10	0.641	1.74E-06	3.15E-03	TBC1 domain family member 32
SQSTM1	11545.47	-1.275	2.24E-06	3.76E-03	sequestosome 1
ZNF592	1686.88	-0.459	2.68E-06	4.20E-03	zinc finger protein 592
C12orf4	783.06	0.503	2.90E-06	4.26E-03	chromosome 12 open reading frame 4
HS1BP3	747.39	-0.664	3.64E-06	5.03E-03	HCLS1 binding protein 3
CD1E	53.44	2.167	4.22E-06	5.22E-03	CD1e molecule
ZNF217	2447.60	-0.509	4.05E-06	5.22E-03	zinc finger protein 217
BAIAP2-DT	350.28	-1.362	5.27E-06	6.09E-03	BAIAP2 divergent transcript
SMARCC2	2893.95	-0.486	5.44E-06	6.09E-03	SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2
AP005018.2	11.62	4.280	5.77E-06	6.17E-03	NA
TCF25	2447.58	-0.369	6.44E-06	6.58E-03	transcription factor 25
NCOA6	2371.55	-0.510	7.12E-06	6.97E-03	nuclear receptor coactivator 6
RPL10P6	79.01	3.962	7.78E-06	7.32E-03	ribosomal protein L10 pseudogene 6
ONECUT1	68.65	-4.297	1.05E-05	9.47E-03	one cut homeobox 1
TMA16	370.07	0.503	1.12E-05	9.77E-03	translation machinery associated 16 homolog
IL12RB2	156.77	2.044	1.18E-05	9.94E-03	interleukin 12 receptor subunit beta 2
UPK3BL1	688.99	2.391	1.25E-05	1.01E-02	uroplakin 3B like 1
TINAGL1	1524.70	-0.868	1.36E-05	1.06E-02	tubulointerstitial nephritis antigen like 1
IGF2BP2	967.23	-1.248	1.62E-05	1.23E-02	insulin like growth factor 2 mRNA binding protein 2
CABLES2	187.44	-0.673	1.70E-05	1.25E-02	Cdk5 and Abl enzyme substrate 2
CTNND2	53.27	-3.938	1.77E-05	1.26E-02	catenin delta 2
FIBCD1	24.17	-3.830	1.90E-05	1.32E-02	fibrinogen C domain containing 1
ERGIC1	4705.09	-0.735	2.02E-05	1.36E-02	endoplasmic reticulum-golgi intermediate compartment 1
AP2A2	1408.77	-0.432	2.28E-05	1.47E-02	adaptor related protein complex 2 subunit alpha 2
CD200R1	221.35	2.309	2.32E-05	1.47E-02	CD200 receptor 1
OVCH1-AS1	38.08	-2.629	2.41E-05	1.49E-02	OVCH1 antisense RNA 1
AC020651.2	7.79	-3.240	2.58E-05	1.51E-02	NA

MTND2P28	324.01	-1.145	2.53E-05	1.51E-02	MT-ND2 pseudogene 28
CASKIN2	710.06	-0.672	2.91E-05	1.63E-02	CASK interacting protein 2
STK25	1883.83	-0.352	2.84E-05	1.63E-02	serine/threonine kinase 25
PCGF2	339.08	-0.898	3.12E-05	1.71E-02	polycomb group ring finger 2
ADRB1	39.27	-2.730	3.29E-05	1.76E-02	adrenoceptor beta 1
PACSIN2	2558.46	-0.767	3.47E-05	1.81E-02	protein kinase C and casein kinase substrate in neurons 2
SPTAN1	14948.17	-0.415	3.70E-05	1.88E-02	spectrin alpha, non-erythrocytic 1
SUPT6H	3947.51	-0.455	3.76E-05	1.88E-02	SPT6 homolog, histone chaperone and transcription elongation factor
MT-ND4	65852.75	-1.234	3.91E-05	1.92E-02	NADH dehydrogenase subunit 4
ICAM1	1355.34	-2.255	4.05E-05	1.94E-02	intercellular adhesion molecule 1
CDH26	1399.31	3.079	4.29E-05	1.96E-02	cadherin 26
KLRC1	26.30	2.621	4.33E-05	1.96E-02	killer cell lectin like receptor C1
PXN	2905.03	-1.298	4.18E-05	1.96E-02	paxillin
ERGIC2	1443.34	0.506	4.49E-05	1.99E-02	ERGIC and golgi 2
FBXO4	216.04	0.506	4.62E-05	2.01E-02	F-box protein 4
AASDH	516.57	0.472	4.92E-05	2.10E-02	aminoadipate-semialdehyde dehydrogenase
MAST2	939.59	-0.894	5.39E-05	2.25E-02	microtubule associated serine/threonine kinase 2
SP2	396.44	-0.496	5.47E-05	2.25E-02	Sp2 transcription factor
RAMP3	348.19	-2.243	5.65E-05	2.29E-02	receptor activity modifying protein 3
RAP1GAP	1060.80	-2.185	5.98E-05	2.38E-02	RAP1 GTPase activating protein
AP000465.1	12.04	1.515	6.61E-05	2.40E-02	NA
GCFC2	440.55	0.521	6.30E-05	2.40E-02	GC-rich sequence DNA-binding factor 2
LMAN2L	404.20	-0.411	6.64E-05	2.40E-02	lectin, mannose binding 2 like
MT-ND2	17249.26	-1.245	6.62E-05	2.40E-02	NADH dehydrogenase subunit 2
PCDHA11	23.78	-3.077	6.49E-05	2.40E-02	protocadherin alpha 11
PNKD	1185.82	-1.772	6.59E-05	2.40E-02	PNKD metallo-beta-lactamase domain containing
MT-ND5	41690.28	-1.005	6.78E-05	2.42E-02	NADH dehydrogenase subunit 5
MED31	207.19	0.569	6.96E-05	2.44E-02	mediator complex subunit 31
FYCO1	2245.86	-0.436	7.35E-05	2.54E-02	FYVE and coiled-coil domain autophagy adaptor 1
ATP6V0A1	2436.32	-0.447	8.05E-05	2.70E-02	ATPase H+ transporting V0 subunit a1
CAPN7	1526.05	0.319	8.33E-05	2.70E-02	calpain 7
DNTTIP1	379.96	-0.827	8.17E-05	2.70E-02	deoxynucleotidyltransferase terminal interacting protein 1
GCSAM	70.70	1.221	8.39E-05	2.70E-02	germinal center associated signaling and motility
KCNQ5-IT1	10.84	2.417	8.33E-05	2.70E-02	NA
DLX6-AS1	57.68	-3.298	8.76E-05	2.74E-02	DLX6 antisense RNA 1
GRB10	1047.85	-1.138	8.63E-05	2.74E-02	growth factor receptor bound protein 10
MTND1P23	4.63	-1.841	8.87E-05	2.74E-02	MT-ND1 pseudogene 23
MT-TV	51.43	-2.038	9.29E-05	2.84E-02	mitochondrially encoded tRNA-Val (GUN)
CDK9	982.51	-0.454	9.42E-05	2.84E-02	cyclin dependent kinase 9
HYAL1	597.08	-1.925	9.80E-05	2.85E-02	hyaluronidase 1
SGSH	490.01	-0.996	9.67E-05	2.85E-02	N-sulfoglucosamine sulfohydrolase

TXNRD1	5792.13	-2.170	9.68E-05	2.85E-02	thioredoxin reductase 1
AC079610.1	15.16	1.785	9.98E-05	2.86E-02	NA
BMPR1B	113.11	-2.680	1.03E-04	2.87E-02	bone morphogenetic protein receptor type 1B
BRI3	759.55	-0.694	1.05E-04	2.87E-02	brain protein I3
MT-RNR1	1329997.30	-0.633	1.02E-04	2.87E-02	mitochondrially encoded 12S rRNA
RPP30	454.74	0.449	1.04E-04	2.87E-02	ribonuclease P/MRP subunit p30
AC021683.2	10.72	-2.307	1.09E-04	2.89E-02	NA
AC107373.2	4.83	3.534	1.09E-04	2.89E-02	NA
ECEL1P2	9.68	-3.510	1.08E-04	2.89E-02	endothelin converting enzyme like 1 pseudogene 2
MMP15	1308.65	-1.842	1.15E-04	3.00E-02	matrix metallopeptidase 15
PCDHGB1	79.25	-2.236	1.19E-04	3.09E-02	protocadherin gamma subfamily B, 1
AC090164.5	27.01	-3.108	1.23E-04	3.12E-02	long intergenic non-protein coding RNA 2348
RNPEPL1	1925.17	-0.480	1.24E-04	3.12E-02	arginyl aminopeptidase like 1
ACHE	338.67	-2.605	1.31E-04	3.19E-02	acetylcholinesterase (Cartwright blood group)
BAZ2A	5703.95	-0.324	1.33E-04	3.19E-02	bromodomain adjacent to zinc finger domain 2A
BCL9	602.99	-0.433	1.30E-04	3.19E-02	BCL9 transcription coactivator
CEP44	554.53	0.486	1.28E-04	3.19E-02	centrosomal protein 44
MYOSLID	22.15	1.966	1.33E-04	3.19E-02	myocardin-induced smooth muscle lncRNA, inducer of differentiation
MGAM	717.06	-3.318	1.36E-04	3.24E-02	maltase-glucoamylase
ADCY5	200.74	-1.963	1.44E-04	3.38E-02	adenylate cyclase 5
AC092745.4	26.09	1.640	1.58E-04	3.48E-02	NA
C2orf91	6.87	2.927	1.51E-04	3.48E-02	long intergenic non-protein coding RNA 2898
ERVMER61-1	4.49	2.366	1.58E-04	3.48E-02	long intergenic non-protein coding RNA 1036
LSM6	264.14	0.409	1.58E-04	3.48E-02	LSM6 homolog, U6 small nuclear RNA and mRNA degradation associated
MT-ND1	23034.70	-0.868	1.55E-04	3.48E-02	NADH dehydrogenase subunit 1
MT-ND4L	15255.65	-1.078	1.58E-04	3.48E-02	NADH dehydrogenase subunit 4L
UBR4	12751.40	-0.369	1.52E-04	3.48E-02	ubiquitin protein ligase E3 component n-recognin 4
LINC01234	19.13	-3.321	1.62E-04	3.52E-02	long intergenic non-protein coding RNA 1234
RPL10P9	88.55	2.259	1.66E-04	3.58E-02	ribosomal protein L10 pseudogene 9
AC138649.1	5.35	-2.829	1.69E-04	3.59E-02	uncharacterized LOC283683
MLXIP	2691.20	-0.763	1.70E-04	3.59E-02	MLX interacting protein
ZNF629	748.58	-0.450	1.72E-04	3.61E-02	zinc finger protein 629
AKR1C3	3494.26	-2.138	1.81E-04	3.76E-02	aldo-keto reductase family 1 member C3
AC009237.3	192.43	-0.950	1.93E-04	3.80E-02	NA
AC011447.3	96.39	-2.105	1.93E-04	3.80E-02	NA
ITGA7	277.73	-1.446	1.88E-04	3.80E-02	integrin subunit alpha 7
NME4	281.40	-0.561	1.94E-04	3.80E-02	NME/NM23 nucleoside diphosphate kinase 4
PITX2	91.32	2.009	1.94E-04	3.80E-02	paired like homeodomain 2
VAX2	14.53	-1.693	1.89E-04	3.80E-02	ventral anterior homeobox 2
WNT2	43.21	-3.214	1.92E-04	3.80E-02	Wnt family member 2
POTEE	22.47	-3.061	1.99E-04	3.85E-02	POTE ankyrin domain family member E
SOX13	799.60	-0.428	2.00E-04	3.85E-02	SRY-box transcription factor 13

AGRN	3825.28	-0.648	2.04E-04	3.91E-02	agrin
GUSB	708.69	-0.478	2.09E-04	3.96E-02	glucuronidase beta
CETN3	307.89	0.566	2.17E-04	4.02E-02	centrin 3
ITGAE	394.86	0.570	2.16E-04	4.02E-02	integrin subunit alpha E
MUC5B	10881.71	-3.168	2.14E-04	4.02E-02	mucin 5B, oligomeric mucus/gel-forming
TNFSF15	824.78	-1.293	2.25E-04	4.12E-02	TNF superfamily member 15
FZD6	1388.65	0.531	2.27E-04	4.14E-02	frizzled class receptor 6
NEU1	1341.91	-1.016	2.37E-04	4.28E-02	neuraminidase 1
LINS1	717.72	0.446	2.42E-04	4.34E-02	lines homolog 1
THRA	922.92	-0.718	2.46E-04	4.38E-02	thyroid hormone receptor alpha
MCF2L	1351.65	-0.938	2.48E-04	4.38E-02	MCF.2 cell line derived transforming sequence like
FAM222B	853.48	-0.484	2.51E-04	4.41E-02	family with sequence similarity 222 member B
WDR36	1821.39	0.456	2.54E-04	4.42E-02	WD repeat domain 36
NCOR2	3189.65	-0.521	2.56E-04	4.42E-02	nuclear receptor corepressor 2
DGAT1	1211.77	-0.587	2.60E-04	4.44E-02	diacylglycerol O-acyltransferase 1
ZNF687	1075.45	-0.378	2.60E-04	4.44E-02	zinc finger protein 687
EPHB4	2281.81	-0.569	2.65E-04	4.45E-02	EPH receptor B4
FASTKD2	1468.26	0.445	2.64E-04	4.45E-02	FAST kinase domains 2
SATB2	186.12	-1.097	2.69E-04	4.49E-02	SATB homeobox 2
AGPAT1	864.68	-0.513	2.75E-04	4.53E-02	1-acylglycerol-3-phosphate O-acyltransferase 1
BTG3	449.23	0.468	2.74E-04	4.53E-02	BTG anti-proliferation factor 3
OSGIN1	552.91	-1.543	2.80E-04	4.58E-02	oxidative stress induced growth inhibitor 1
ANKRD52	2188.87	-0.483	2.87E-04	4.65E-02	ankyrin repeat domain 52
ZNF646	1118.20	-0.419	2.89E-04	4.65E-02	zinc finger protein 646
RNF123	926.06	-0.332	2.92E-04	4.67E-02	ring finger protein 123
CCDC90B	814.01	0.362	2.97E-04	4.69E-02	coiled-coil domain containing 90B
MFSD13A	163.41	-0.840	2.97E-04	4.69E-02	major facilitator superfamily domain containing 13A
PRAME	34.30	-2.917	3.01E-04	4.71E-02	PRAME nuclear receptor transcriptional regulator
C21orf91	894.40	0.666	3.03E-04	4.72E-02	chromosome 21 open reading frame 91
CSNK1D	1766.04	-0.468	3.06E-04	4.73E-02	casein kinase 1 delta
PCDHAC1	43.27	-2.528	3.19E-04	4.89E-02	protocadherin alpha subfamily C, 1
PYURF	722.07	0.428	3.23E-04	4.93E-02	PIGY upstream open reading frame
LRP5	1799.06	-0.824	3.26E-04	4.94E-02	LDL receptor related protein 5
ILDR2	75.14	-2.087	3.36E-04	4.97E-02	immunoglobulin like domain containing receptor 2
KRT8P36	17.59	-1.609	3.32E-04	4.97E-02	keratin 8 pseudogene 36
P3H4	275.08	-1.182	3.35E-04	4.97E-02	prolyl 3-hydroxylase family member 4 (inactive)
PI4KA	2608.83	-0.464	3.32E-04	4.97E-02	phosphatidylinositol 4-kinase alpha
C6orf15	46.59	2.417	3.48E-04	5.09E-02	chromosome 6 open reading frame 15
XPO4	1721.44	0.393	3.49E-04	5.09E-02	exportin 4
SLC1A7	37.56	-2.650	3.51E-04	5.09E-02	solute carrier family 1 member 7
AC006525.1	8.91	-2.230	3.60E-04	5.13E-02	NA
AL445645.1	7.63	-1.566	3.60E-04	5.13E-02	NA
RPRD1A	1477.75	0.446	3.57E-04	5.13E-02	regulation of nuclear pre-mRNA domain containing 1A

PCDHGA4	69.77	-1.458	3.67E-04	5.20E-02	protocadherin gamma subfamily A, 4
ABCG2	336.51	-1.520	4.14E-04	5.34E-02	ATP binding cassette subfamily G member 2 (Junior blood group)
AL590064.1	179.65	-1.306	4.10E-04	5.34E-02	NA
ALB	39.70	-2.676	4.07E-04	5.34E-02	albumin
EXOSC8	490.89	0.401	3.93E-04	5.34E-02	exosome component 8
FCER1A	153.99	1.589	3.88E-04	5.34E-02	Fc fragment of IgE receptor Ia
FTH1P2	12.55	-1.525	3.89E-04	5.34E-02	ferritin heavy chain 1 pseudogene 2
HCFC2	950.41	0.396	3.99E-04	5.34E-02	host cell factor C2
MCOLN1	326.62	-1.042	4.06E-04	5.34E-02	mucolipin TRP cation channel 1
MT-ATP6	27254.33	-0.660	4.03E-04	5.34E-02	ATP synthase F0 subunit 6
MTATP6P1	2594.46	-0.731	4.08E-04	5.34E-02	MT-ATP6 pseudogene 1
NT5DC3	376.47	-0.741	4.00E-04	5.34E-02	5'-nucleotidase domain containing 3
PACS2	1348.61	-0.414	4.12E-04	5.34E-02	phosphofurin acidic cluster sorting protein 2
TLR4	1269.90	-1.934	3.95E-04	5.34E-02	toll like receptor 4
TTC39C	739.94	0.502	4.01E-04	5.34E-02	tetratricopeptide repeat domain 39C
ZFHX3	1367.04	-0.589	4.01E-04	5.34E-02	zinc finger homeobox 3
ZNF703	379.00	-0.690	3.80E-04	5.34E-02	zinc finger protein 703
CLN3	856.70	-0.504	4.27E-04	5.41E-02	CLN3 lysosomal/endosomal transmembrane protein, battenin
HNRNPA3P6	10.53	-0.675	4.28E-04	5.41E-02	heterogeneous nuclear ribonucleoprotein A3 pseudogene 6
MYORG	1176.28	-0.873	4.28E-04	5.41E-02	myogenesis regulating glycosidase (putative)
SLC4A2	1746.54	-0.562	4.25E-04	5.41E-02	solute carrier family 4 member 2
DIP2C	658.98	-0.469	4.37E-04	5.48E-02	disco interacting protein 2 homolog C
ZC2HC1A	282.43	0.718	4.38E-04	5.48E-02	zinc finger C2HC-type containing 1A
CDH6	316.18	-1.395	4.49E-04	5.59E-02	cadherin 6
VRK2	864.07	0.436	4.59E-04	5.68E-02	VRK serine/threonine kinase 2
AASDHPPT	846.17	0.438	4.63E-04	5.70E-02	aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase
OR7E108P	5.70	2.222	4.79E-04	5.84E-02	olfactory receptor family 7 subfamily E member 108 pseudogene
TMEM120B	390.29	-0.470	4.79E-04	5.84E-02	transmembrane protein 120B
ERMARD	514.14	0.403	4.89E-04	5.93E-02	ER membrane associated RNA degradation
KYAT3	938.81	0.471	4.92E-04	5.93E-02	kynurenine aminotransferase 3
EFHD1	66.75	-1.789	4.96E-04	5.96E-02	EF-hand domain family member D1
AC087175.1	5.34	-1.386	5.00E-04	5.97E-02	NA
CMTR2	1473.77	0.474	5.06E-04	6.01E-02	cap methyltransferase 2
MFSD8	613.20	0.398	5.08E-04	6.01E-02	major facilitator superfamily domain containing 8
CNNM2	441.36	-0.573	5.18E-04	6.04E-02	cyclin and CBS domain divalent metal cation transport mediator 2
H3C4	1154.23	0.480	5.15E-04	6.04E-02	H3 clustered histone 4
SNX8	657.34	-0.464	5.22E-04	6.04E-02	sorting nexin 8
SS18	2578.23	0.458	5.19E-04	6.04E-02	SS18 subunit of BAF chromatin remodeling complex
AL138828.1	12.53	-2.207	5.28E-04	6.04E-02	uncharacterized LOC644135
CAPS	299.88	-0.752	5.29E-04	6.04E-02	calcyphosine
MYO15B	2353.64	-1.266	5.29E-04	6.04E-02	myosin XVB
ATG2A	1313.79	-0.610	5.38E-04	6.08E-02	autophagy related 2A

FLCN	1185.19	-0.499	5.36E-04	6.08E-02	folliculin
SPATA2L	208.60	-0.517	5.57E-04	6.26E-02	spermatogenesis associated 2 like
SPRED2	1012.99	-0.543	5.59E-04	6.26E-02	sprouty related EVH1 domain containing 2
MTX3	1059.43	0.465	5.72E-04	6.37E-02	metaxin 3
OSR2	71.90	-1.440	5.77E-04	6.40E-02	odd-skipped related transciption factor 2
AC104667.2	12.88	-1.444	5.84E-04	6.41E-02	uncharacterized LOC105373958
PTPN23	1256.32	-0.418	5.81E-04	6.41E-02	protein tyrosine phosphatase non-receptor type 23
AC093627.7	8.56	-2.334	5.87E-04	6.42E-02	NA
CD1C	70.45	0.830	5.99E-04	6.51E-02	CD1c molecule
CERS2	2166.26	-0.824	6.02E-04	6.51E-02	ceramide synthase 2
LINC02687	4.41	2.240	6.04E-04	6.51E-02	long intergenic non-protein coding RNA 2687
OMA1	676.16	0.406	6.06E-04	6.51E-02	OMA1 zinc metallopeptidase
NNAT	7.89	2.034	6.18E-04	6.60E-02	neuronatin
FN3K	99.87	-1.107	6.23E-04	6.63E-02	fructosamine 3 kinase
ABCD4	1019.13	-0.416	6.30E-04	6.68E-02	ATP binding cassette subfamily D member 4
TAF7L	6.85	1.678	6.42E-04	6.77E-02	TATA-box binding protein associated factor 7 like
ARL3	183.65	-0.417	6.47E-04	6.79E-02	ADP ribosylation factor like GTPase 3
MC1R	127.41	-0.780	6.50E-04	6.79E-02	melanocortin 1 receptor
AC084277.1	6.59	2.133	6.77E-04	6.81E-02	NA
ANGPTL2	440.25	-1.159	6.62E-04	6.81E-02	angiopoietin like 2
GOLT1A	55.14	-1.689	6.76E-04	6.81E-02	golgi transport 1A
LINC02569	7.22	-1.060	6.73E-04	6.81E-02	long intergenic non-protein coding RNA 2569
NAP1L2	45.55	0.975	6.73E-04	6.81E-02	nucleosome assembly protein 1 like 2
PTPN20	141.19	1.371	6.77E-04	6.81E-02	protein tyrosine phosphatase non-receptor type 20
RAMP1	174.80	-1.530	6.60E-04	6.81E-02	receptor activity modifying protein 1
SELENON	1691.49	-0.447	6.65E-04	6.81E-02	selenoprotein N
ZDHHC8	441.52	-0.430	6.78E-04	6.81E-02	zinc finger DHHC-type palmitoyltransferase 8
ANGEL2	791.10	0.371	6.81E-04	6.82E-02	angel homolog 2
SLC12A9	1107.69	-0.531	6.85E-04	6.82E-02	solute carrier family 12 member 9
RHOC	2762.68	-0.859	6.88E-04	6.82E-02	ras homolog family member C
IGKV1D-39	9.20	-2.139	6.94E-04	6.83E-02	immunoglobulin kappa variable 1D-39
LRCH4	1174.81	-0.411	6.96E-04	6.83E-02	leucine rich repeats and calponin homology domain containing 4
METTL15	512.04	0.366	6.98E-04	6.83E-02	methyltransferase like 15
MRPL1	390.67	0.440	7.01E-04	6.83E-02	mitochondrial ribosomal protein L1
MELTF	222.93	-0.763	7.04E-04	6.84E-02	melanotransferrin
AC011451.3	7.77	1.674	7.10E-04	6.84E-02	NA
NEK3	348.37	-1.203	7.10E-04	6.84E-02	NIMA related kinase 3
SLC47A1	80.30	0.563	7.14E-04	6.85E-02	solute carrier family 47 member 1
C2orf76	102.41	0.441	7.20E-04	6.88E-02	chromosome 2 open reading frame 76
PABPC1L	411.52	-0.796	7.25E-04	6.88E-02	poly(A) binding protein cytoplasmic 1 like
PRRC2A	5948.73	-0.424	7.24E-04	6.88E-02	proline rich coiled-coil 2A
AC005410.2	6.64	0.797	7.50E-04	6.96E-02	NA
AL390728.2	20.66	-0.610	7.47E-04	6.96E-02	NA
ALOX15P1	30.67	0.978	7.43E-04	6.96E-02	arachidonate 15-lipoxygenase pseudogene 1

ARHGEF10	567.92	-0.705	7.66E-04	6.96E-02	Rho guanine nucleotide exchange factor 10
NRSN2	149.24	-0.595	7.64E-04	6.96E-02	neurensin 2
PEG10	394.85	-1.782	7.55E-04	6.96E-02	paternally expressed 10
RERE	2685.57	-0.364	7.63E-04	6.96E-02	arginine-glutamic acid dipeptide repeats
RPP25	280.10	-0.594	7.47E-04	6.96E-02	ribonuclease P and MRP subunit p25
SALL4	107.28	-1.923	7.66E-04	6.96E-02	spalt like transcription factor 4
TXNRD2	349.81	-0.475	7.66E-04	6.96E-02	thioredoxin reductase 2
ZNF608	868.88	-0.441	7.59E-04	6.96E-02	zinc finger protein 608
AL121832.3	64.71	-0.672	7.77E-04	7.01E-02	NA
LINC00659	18.06	-2.006	7.78E-04	7.01E-02	long intergenic non-protein coding RNA 659
ZNF492	22.47	1.219	7.86E-04	7.05E-02	zinc finger protein 492
CABLES1	736.37	-0.758	7.95E-04	7.09E-02	Cdk5 and Abl enzyme substrate 1
PYCR1	834.11	-1.010	7.96E-04	7.09E-02	pyrroline-5-carboxylate reductase 1
SPTBN1	21659.82	-0.669	8.05E-04	7.14E-02	spectrin beta, non-erythrocytic 1
ANKRD11	3603.58	-0.387	8.08E-04	7.14E-02	ankyrin repeat domain 11
ATN1	1484.97	-0.463	8.27E-04	7.24E-02	atrophin 1
MT-TW	161.55	-1.041	8.29E-04	7.24E-02	mitochondrially encoded tRNA-Trp (UGA/G)
RRNAD1	383.62	-0.338	8.23E-04	7.24E-02	methyltransferase like 25B
AL021937.3	11.61	-1.478	8.36E-04	7.26E-02	NA
GTSF1	7.75	-1.766	8.37E-04	7.26E-02	gametocyte specific factor 1
LRP3	130.52	-0.466	8.47E-04	7.30E-02	LDL receptor related protein 3
RIPOR3	142.86	-1.643	8.46E-04	7.30E-02	RIPOR family member 3
AP2B1	4950.13	-0.727	8.70E-04	7.33E-02	adaptor related protein complex 2 subunit beta 1
CILP2	4.55	-1.591	8.55E-04	7.33E-02	cartilage intermediate layer protein 2
HECW1	97.63	-1.483	8.78E-04	7.33E-02	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 1
LAMA5	4951.57	-0.542	8.75E-04	7.33E-02	laminin subunit alpha 5
MIR644A	4.84	2.026	8.70E-04	7.33E-02	microRNA 644a
PGAP6	1523.22	-0.433	8.82E-04	7.33E-02	post-glycosylphosphatidylinositol attachment to proteins 6
TMEM97	1119.98	-1.132	8.76E-04	7.33E-02	transmembrane protein 97
TNRC18	2943.88	-0.458	8.70E-04	7.33E-02	trinucleotide repeat containing 18
TSPEAR-AS1	7.68	-2.032	8.81E-04	7.33E-02	TSPEAR antisense RNA 1
ZC3H3	519.69	-0.456	8.76E-04	7.33E-02	zinc finger CCCH-type containing 3
AC011601.1	10.38	1.953	8.86E-04	7.34E-02	NA
AC021683.6	4.80	-1.550	9.24E-04	7.38E-02	NA
AC147067.2	7.69	-1.787	8.99E-04	7.38E-02	NA
AKR1C2	5286.36	-1.536	9.26E-04	7.38E-02	aldo-keto reductase family 1 member C2
COL28A1	237.43	-1.707	9.08E-04	7.38E-02	collagen type XXVIII alpha 1 chain
DAP	1521.33	-0.712	9.18E-04	7.38E-02	death associated protein
DDX42	3062.34	-0.296	9.07E-04	7.38E-02	DEAD-box helicase 42
FAR1	4913.84	0.492	9.11E-04	7.38E-02	fatty acyl-CoA reductase 1
PADI2	1287.05	-1.771	9.16E-04	7.38E-02	peptidyl arginine deiminase 2
PLOD1	1344.34	-0.541	9.06E-04	7.38E-02	procollagen-lysine,2-oxoglutarate 5-dioxygenase 1
POLR3G	185.52	0.875	8.98E-04	7.38E-02	RNA polymerase III subunit G

TANC2	2533.01	-0.435	9.22E-04	7.38E-02	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 2
FBXO31	715.53	-0.362	9.42E-04	7.43E-02	F-box protein 31
GRB14	89.41	-1.325	9.37E-04	7.43E-02	growth factor receptor bound protein 14
NAA60	761.84	-0.449	9.41E-04	7.43E-02	N-alpha-acetyltransferase 60, NatF catalytic subunit
SCARB1	775.84	-0.762	9.45E-04	7.43E-02	scavenger receptor class B member 1
CCDC9	308.11	-0.465	9.58E-04	7.46E-02	coiled-coil domain containing 9
SESN3	3975.04	0.615	9.53E-04	7.46E-02	sestrin 3
SLC4A1	18.29	-1.729	9.55E-04	7.46E-02	solute carrier family 4 member 1 (Diego blood group)
MSL3P1	5.00	-1.372	9.67E-04	7.50E-02	MSL complex subunit 3 pseudogene 1
CD63	6122.56	-0.557	9.79E-04	7.55E-02	CD63 molecule
R3HDM2	1476.88	-0.318	9.76E-04	7.55E-02	R3H domain containing 2
PLAA	1201.71	0.377	9.87E-04	7.58E-02	phospholipase A2 activating protein
ASPH	8101.18	-0.594	1.01E-03	7.73E-02	aspartate beta-hydroxylase
FAHD2A	350.86	-0.394	1.01E-03	7.73E-02	fumarylacetoacetate hydrolase domain containing 2A
CLNS1A	2121.55	0.444	1.02E-03	7.74E-02	chloride nucleotide-sensitive channel 1A
FAM50A	701.17	-0.421	1.02E-03	7.74E-02	family with sequence similarity 50 member A
LINC00391	9.78	1.459	1.02E-03	7.74E-02	long intergenic non-protein coding RNA 391
KCNMB4	46.16	-1.071	1.04E-03	7.81E-02	potassium calcium-activated channel subfamily M regulatory beta subunit 4
SLC22A18	694.09	-0.626	1.04E-03	7.81E-02	solute carrier family 22 member 18
LINC02561	6.36	-1.710	1.05E-03	7.84E-02	long intergenic non-protein coding RNA 2561
LINC00884	154.54	-1.556	1.06E-03	7.94E-02	long intergenic non-protein coding RNA 884
MRGBP	380.39	-0.499	1.07E-03	7.94E-02	MRG domain binding protein
RRN3P4	4.65	-1.857	1.07E-03	7.94E-02	RRN3 pseudogene 4
SPEN	5323.42	-0.335	1.07E-03	7.94E-02	spen family transcriptional repressor
LRFN4	235.92	-0.626	1.09E-03	8.01E-02	leucine rich repeat and fibronectin type III domain containing 4
CTDSP1	2635.70	-0.342	1.09E-03	8.02E-02	CTD small phosphatase 1
DXO	262.21	-0.401	1.10E-03	8.03E-02	decapping exoribonuclease
SPRY4-AS1	10.38	-1.228	1.10E-03	8.03E-02	SPRY4 antisense RNA 1
PLBD2	893.30	-0.526	1.12E-03	8.13E-02	phospholipase B domain containing 2
ASL	1612.09	-0.556	1.14E-03	8.13E-02	argininosuccinate lyase
CNTD2	4.93	-1.570	1.14E-03	8.13E-02	cyclin P
MAML1	1571.23	-0.351	1.13E-03	8.13E-02	mastermind like transcriptional coactivator 1
MT-TL1	184.70	-0.632	1.14E-03	8.13E-02	mitochondrially encoded tRNA-Leu (UUA/G) 1
STX3	1988.37	-0.473	1.14E-03	8.13E-02	syntaxin 3
TBC1D13	706.44	-0.293	1.12E-03	8.13E-02	TBC1 domain family member 13
TMEM184B	1388.97	-0.439	1.13E-03	8.13E-02	transmembrane protein 184B
ZNF114	26.09	-1.230	1.15E-03	8.18E-02	zinc finger protein 114
BTNL9	167.03	-1.417	1.16E-03	8.19E-02	butyrophilin like 9
LNPEP	6126.15	0.412	1.16E-03	8.19E-02	leucyl and cystinyl aminopeptidase
TRIP4	629.03	-0.353	1.16E-03	8.19E-02	thyroid hormone receptor interactor 4
LASP1	7824.92	-0.494	1.17E-03	8.21E-02	LIM and SH3 protein 1
ITGB3BP	286.84	0.426	1.18E-03	8.27E-02	integrin subunit beta 3 binding protein

ENKD1	140.68	-0.395	1.19E-03	8.32E-02	enkurin domain containing 1
TMEM74	35.69	0.857	1.20E-03	8.33E-02	transmembrane protein 74
TMUB2	847.14	-0.332	1.21E-03	8.37E-02	transmembrane and ubiquitin like domain containing 2
ZNF76	579.34	-0.379	1.21E-03	8.37E-02	zinc finger protein 76
PCBP2	5549.33	-0.304	1.22E-03	8.42E-02	poly(rC) binding protein 2
SYPL2	18.19	-1.598	1.23E-03	8.42E-02	synaptophysin like 2
TSPAN18	278.19	-0.927	1.23E-03	8.45E-02	tetraspanin 18
AC007684.2	8.71	1.156	1.24E-03	8.46E-02	NA
ALDH1A1	6441.96	-1.433	1.27E-03	8.65E-02	aldehyde dehydrogenase 1 family member A1
ADCY4	531.17	-0.981	1.28E-03	8.67E-02	adenylate cyclase 4
BCORL1	319.18	-0.506	1.30E-03	8.75E-02	BCL6 corepressor like 1
LDB3	30.17	-0.993	1.30E-03	8.75E-02	LIM domain binding 3
USP36	1118.34	-0.428	1.30E-03	8.75E-02	ubiquitin specific peptidase 36
SLC16A3	2764.94	-0.651	1.31E-03	8.77E-02	solute carrier family 16 member 3
AL021807.1	16.78	-0.812	1.32E-03	8.79E-02	NA
DLGAP1-AS5	24.22	1.533	1.32E-03	8.79E-02	DLGAP1 antisense RNA 5
EEPD1	532.19	-0.507	1.31E-03	8.79E-02	endonuclease/exonuclease/phosphatase family domain containing 1
GGH	547.44	0.663	1.33E-03	8.79E-02	gamma-glutamyl hydrolase
PRXL2B	734.28	-0.515	1.33E-03	8.79E-02	peroxiredoxin like 2B
SRMS	43.91	-1.371	1.35E-03	8.93E-02	src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristylation sites
LYPLAL1	425.08	0.436	1.36E-03	8.93E-02	lysophospholipase like 1
TDRD12	15.29	1.722	1.36E-03	8.93E-02	tudor domain containing 12
CENPC	1433.37	0.397	1.37E-03	8.94E-02	centromere protein C
ATP8A2	53.55	-1.458	1.39E-03	8.97E-02	ATPase phospholipid transporting 8A2
CDK20	113.91	-0.418	1.40E-03	8.97E-02	cyclin dependent kinase 20
KRT8P12	122.97	-0.441	1.38E-03	8.97E-02	keratin 8 pseudogene 12
PRPF39	1026.48	0.337	1.38E-03	8.97E-02	pre-mRNA processing factor 39
RAB20	467.23	-1.317	1.39E-03	8.97E-02	RAB20, member RAS oncogene family
RASSF3	1287.94	-0.658	1.41E-03	8.97E-02	Ras association domain family member 3
RPL4P5	12.77	0.552	1.41E-03	8.97E-02	ribosomal protein L4 pseudogene 5
SEPHS2	626.53	-0.601	1.41E-03	8.97E-02	selenophosphate synthetase 2
TMEM94	1411.15	-0.437	1.41E-03	8.97E-02	transmembrane protein 94
VSTM1	6.20	-1.590	1.40E-03	8.97E-02	V-set and transmembrane domain containing 1
FAM131A	395.69	-0.451	1.43E-03	9.01E-02	family with sequence similarity 131 member A
GLB1L	139.56	-0.739	1.43E-03	9.01E-02	galactosidase beta 1 like
GPX8	377.83	-0.607	1.43E-03	9.01E-02	glutathione peroxidase 8 (putative)
IGHJ4	5.96	-1.361	1.43E-03	9.01E-02	immunoglobulin heavy joining 4
VWA5B1	10.10	-1.636	1.43E-03	9.01E-02	von Willebrand factor A domain containing 5B1
ELP4	801.74	0.381	1.44E-03	9.06E-02	elongator acetyltransferase complex subunit 4
CCHCR1	508.35	-0.394	1.45E-03	9.08E-02	coiled-coil alpha-helical rod protein 1
EMBP1	76.54	-1.092	1.47E-03	9.15E-02	embigin pseudogene 1
HMBOX1	1298.47	-0.339	1.48E-03	9.15E-02	homeobox containing 1
ILRUN	3978.76	-0.402	1.47E-03	9.15E-02	inflammation and lipid regulator with UBA-like and NBR1-like domains

CCL11	141.66	-1.413	1.49E-03	9.19E-02	C-C motif chemokine ligand 11
COCH	108.70	-1.488	1.49E-03	9.19E-02	cochlin
KCNS3	338.53	-0.897	1.50E-03	9.23E-02	potassium voltage-gated channel modifier subfamily S member 3
TIFA	744.24	0.510	1.51E-03	9.25E-02	TRAF interacting protein with forkhead associated domain
NT5C3AP2	10.07	1.114	1.52E-03	9.26E-02	NT5C3A pseudogene 2
RAB5C	2392.05	-0.395	1.51E-03	9.26E-02	RAB5C, member RAS oncogene family
AL356740.1	13.88	-1.080	1.53E-03	9.26E-02	NA
KCNQ3	44.40	-1.062	1.53E-03	9.26E-02	potassium voltage-gated channel subfamily Q member 3
SMAD6	385.85	-0.946	1.53E-03	9.26E-02	SMAD family member 6
TRIM13	1710.80	0.513	1.53E-03	9.26E-02	tripartite motif containing 13
CCNK	928.90	-0.375	1.54E-03	9.26E-02	cyclin K
FAM189B	285.04	-0.421	1.54E-03	9.26E-02	family with sequence similarity 189 member B
KAT8	669.24	-0.324	1.56E-03	9.26E-02	lysine acetyltransferase 8
MAP4K2	361.28	-0.449	1.55E-03	9.26E-02	mitogen-activated protein kinase kinase kinase kinase 2
MGRN1	1226.58	-0.424	1.55E-03	9.26E-02	mahogunin ring finger 1
RUNDC1	734.97	-0.428	1.55E-03	9.26E-02	RUN domain containing 1
ATF6B	1908.51	-0.371	1.57E-03	9.31E-02	activating transcription factor 6 beta
PLRG1	1658.57	0.413	1.58E-03	9.31E-02	pleiotropic regulator 1
ZBTB33	1499.32	0.435	1.57E-03	9.31E-02	zinc finger and BTB domain containing 33
CD1B	7.17	0.980	1.59E-03	9.36E-02	CD1b molecule
ADAT1	724.58	-0.386	1.61E-03	9.38E-02	adenosine deaminase tRNA specific 1
CYP7B1	1228.04	0.652	1.61E-03	9.38E-02	cytochrome P450 family 7 subfamily B member 1
OR13A1	10.74	1.248	1.60E-03	9.38E-02	olfactory receptor family 13 subfamily A member 1
SOX11	4.98	-1.615	1.61E-03	9.38E-02	SRY-box transcription factor 11
RAMACL	26.20	-0.757	1.62E-03	9.41E-02	RNA guanine-7 methyltransferase activating subunit like (pseudogene)
GLMN	281.89	0.415	1.63E-03	9.44E-02	glomulin, FKBP associated protein
FCN3	36.35	-1.142	1.64E-03	9.48E-02	ficolin 3
NOC3L	818.41	0.391	1.64E-03	9.48E-02	NOC3 like DNA replication regulator
NPPC	8.67	1.434	1.65E-03	9.50E-02	natriuretic peptide C
LINC00968	5.81	-1.337	1.66E-03	9.51E-02	long intergenic non-protein coding RNA 968
ZNF696	163.92	-0.381	1.66E-03	9.51E-02	zinc finger protein 696
ENTPD1	1375.53	-0.791	1.67E-03	9.56E-02	ectonucleoside triphosphate diphosphohydrolase 1
LINC02471	62.41	1.428	1.67E-03	9.56E-02	long intergenic non-protein coding RNA 2471
JAZF1-AS1	4.57	-1.443	1.68E-03	9.57E-02	JAZF1 antisense RNA 1
GREB1	87.28	-0.875	1.69E-03	9.59E-02	growth regulating estrogen receptor binding 1
SLC26A6	606.30	-0.614	1.69E-03	9.59E-02	solute carrier family 26 member 6
AC011407.1	5.44	-1.172	1.70E-03	9.63E-02	NA
AL021707.3	54.00	-0.926	1.71E-03	9.66E-02	NA
PCDHB16	61.08	-0.835	1.73E-03	9.71E-02	protocadherin beta 16
SLC8B1	927.14	-0.364	1.73E-03	9.71E-02	solute carrier family 8 member B1
YARS2	388.53	0.359	1.76E-03	9.84E-02	tyrosyl-tRNA synthetase 2

TDP1	622.41	0.352	1.77E-03	9.87E-02	tyrosyl-DNA phosphodiesterase 1
MORC2	1010.88	-0.382	1.79E-03	9.87E-02	MORC family CW-type zinc finger 2
MTHFS	223.43	-0.614	1.79E-03	9.87E-02	methenyltetrahydrofolate synthetase
PLOD3	802.65	-0.807	1.78E-03	9.87E-02	procollagen-lysine,2-oxoglutarate 5-dioxygenase 3
RARA-AS1	46.48	-0.681	1.78E-03	9.87E-02	RARA antisense RNA 1
ZCCHC4	297.21	0.326	1.78E-03	9.87E-02	zinc finger CCHC-type containing 4
PPP1R12B	2620.85	-0.453	1.80E-03	9.92E-02	protein phosphatase 1 regulatory subunit 12B
ENHO	4.47	-1.499	1.81E-03	9.93E-02	energy homeostasis associated
AL670729.3	37.07	-1.168	1.82E-03	9.94E-02	NA
PTPN18	1104.37	-0.473	1.82E-03	9.94E-02	protein tyrosine phosphatase non-receptor type 18
SH3BP4	1598.96	-0.571	1.82E-03	9.94E-02	SH3 domain binding protein 4
IKBKG	180.25	-0.404	1.83E-03	9.94E-02	inhibitor of nuclear factor kappa B kinase regulatory subunit gamma
SIDT2	1082.23	-0.434	1.83E-03	9.94E-02	SID1 transmembrane family member 2
AP000640.1	18.60	-0.826	1.85E-03	9.95E-02	NA
MT-TD	11.74	-1.105	1.84E-03	9.95E-02	mitochondrially encoded tRNA-Asp (GAU/C)
SLC26A1	26.41	-1.051	1.84E-03	9.95E-02	solute carrier family 26 member 1
GABRA4	80.41	1.289	1.86E-03	9.97E-02	gamma-aminobutyric acid type A receptor subunit alpha4
RBPMS	885.07	-0.417	1.85E-03	9.97E-02	RNA binding protein, mRNA processing factor
MTUS1	4403.82	-0.442	1.86E-03	9.98E-02	microtubule associated scaffold protein 1

Appendix C-6: ICI-only vs. PreTx (All patients, OAC)

Gene	baseMean	log2FC	pvalue	padj	description
KLRD1	143.39	1.358	2.60E-13	7.28E-09	killer cell lectin like receptor D1
KLRK1	136.68	1.348	5.35E-11	7.50E-07	killer cell lectin like receptor K1
FASLG	42.89	1.262	2.25E-10	1.05E-06	Fas ligand
GZMK	37.54	1.466	1.93E-10	1.05E-06	granzyme K
KLRC1	25.58	1.406	1.60E-10	1.05E-06	killer cell lectin like receptor C1
KLRC4	33.26	1.449	1.35E-10	1.05E-06	killer cell lectin like receptor C4
CD8A	147.20	1.242	4.18E-10	1.67E-06	CD8a molecule
PRF1	151.83	1.104	5.96E-10	2.09E-06	perforin 1
APOBEC3G	260.33	0.793	1.76E-09	5.49E-06	apolipoprotein B mRNA editing enzyme catalytic subunit 3G
APOL3	581.38	0.902	2.10E-09	5.88E-06	apolipoprotein L3
CD244	53.18	1.109	2.40E-09	6.12E-06	CD244 molecule
FCRL6	28.91	1.397	5.78E-09	1.35E-05	Fc receptor like 6
CXCL9	686.21	1.390	7.34E-09	1.58E-05	C-X-C motif chemokine ligand 9
GZMH	43.00	1.337	9.66E-09	1.94E-05	granzyme H
GBP4	1928.82	1.085	1.06E-08	1.98E-05	guanylate binding protein 4
HLA-DQA1	643.72	0.819	1.32E-08	2.32E-05	major histocompatibility complex, class II, DQ alpha 1
DTHD1	81.56	1.210	1.78E-08	2.93E-05	death domain containing 1
CCR5	132.75	0.970	2.27E-08	3.33E-05	C-C motif chemokine receptor 5
CXCR6	124.10	1.078	2.38E-08	3.33E-05	C-X-C motif chemokine receptor 6
LINC01684	28.25	1.143	2.35E-08	3.33E-05	long intergenic non-protein coding RNA 1684
KLRC3	14.38	1.287	4.02E-08	5.12E-05	killer cell lectin like receptor C3
TRGC2	32.18	1.199	3.99E-08	5.12E-05	T cell receptor gamma constant 2
GZMA	113.18	1.211	5.19E-08	6.32E-05	granzyme A
KLRC2	18.95	1.303	6.52E-08	7.62E-05	killer cell lectin like receptor C2
CXCL10	341.52	1.249	6.87E-08	7.70E-05	C-X-C motif chemokine ligand 10
HCST	35.61	0.968	7.45E-08	8.04E-05	hematopoietic cell signal transducer
THEMIS	129.60	1.149	8.11E-08	8.42E-05	thymocyte selection associated
ZNF683	22.63	1.268	8.54E-08	8.55E-05	zinc finger protein 683
SPN	272.91	0.874	9.40E-08	9.09E-05	sialophorin
CLNK	36.18	1.034	1.50E-07	1.40E-04	cytokine dependent hematopoietic cell linker
CD2	237.95	0.990	1.73E-07	1.56E-04	CD2 molecule
TRG-AS1	72.55	1.001	2.00E-07	1.75E-04	T cell receptor gamma locus antisense RNA 1
ZNF582	25.43	0.913	2.07E-07	1.76E-04	zinc finger protein 582
HLA-DQA2	135.25	0.851	2.27E-07	1.87E-04	major histocompatibility complex, class II, DQ alpha 2
NT5C3AP2	19.84	1.184	2.46E-07	1.97E-04	NT5C3A pseudogene 2
GPR174	178.62	1.066	2.67E-07	2.08E-04	G protein-coupled receptor 174
ITGAE	390.28	0.665	2.89E-07	2.19E-04	integrin subunit alpha E
SEPTIN1	116.96	0.927	4.94E-07	3.65E-04	septin 1
AC015911.3	26.84	1.005	5.31E-07	3.82E-04	NA
CD3D	97.56	1.003	6.22E-07	4.36E-04	CD3d molecule
CLEC10A	60.91	0.982	6.62E-07	4.53E-04	C-type lectin domain containing 10A

CD96	226.15	0.964	6.98E-07	4.66E-04	CD96 molecule
GBP1P1	175.49	1.103	7.50E-07	4.89E-04	guanylate binding protein 1 pseudogene 1
ABCD2	55.02	1.022	8.85E-07	5.64E-04	ATP binding cassette subfamily D member 2
TTC24	10.28	1.176	9.36E-07	5.83E-04	tetratricopeptide repeat domain 24
CD84	481.35	0.804	1.20E-06	7.31E-04	CD84 molecule
CD8B	50.70	1.112	1.34E-06	7.85E-04	CD8b molecule
ZNF677	95.58	0.828	1.32E-06	7.85E-04	zinc finger protein 677
NKG7	71.23	1.061	1.49E-06	8.50E-04	natural killer cell granule protein 7
EOMES	56.56	1.051	1.55E-06	8.72E-04	eomesodermin
ADA2	611.44	0.618	1.73E-06	9.14E-04	adenosine deaminase 2
CD3E	140.47	0.967	1.73E-06	9.14E-04	CD3e molecule
TOMM20P2	23.11	1.063	1.69E-06	9.14E-04	TOMM20 pseudogene 2
ERAP2	1529.69	0.584	1.79E-06	9.32E-04	endoplasmic reticulum aminopeptidase 2
TRAJ6	5.77	1.134	1.84E-06	9.39E-04	T cell receptor alpha joining 6
GIMAP5	334.34	0.795	1.93E-06	9.67E-04	GTPase, IMAP family member 5
IDO1	246.09	1.092	1.97E-06	9.69E-04	indoleamine 2,3-dioxygenase 1
LINC02446	21.84	1.188	2.45E-06	1.19E-03	long intergenic non-protein coding RNA 2446
MMRN1	183.40	1.059	2.74E-06	1.30E-03	multimerin 1
CTSO	325.80	0.712	2.95E-06	1.38E-03	cathepsin O
AC015911.10	71.56	0.920	3.11E-06	1.38E-03	NA
GNGT2	23.06	0.768	3.09E-06	1.38E-03	G protein subunit gamma transducin 2
ZDHHC15	18.98	1.052	3.11E-06	1.38E-03	zinc finger DHHC-type palmitoyltransferase 15
TRAJ19	7.07	1.112	3.19E-06	1.40E-03	T cell receptor alpha joining 19 (non-functional)
GPR171	39.75	1.009	3.39E-06	1.44E-03	G protein-coupled receptor 171
PRKCQ-AS1	61.60	0.874	3.35E-06	1.44E-03	PRKCQ antisense RNA 1
GBP5	993.95	0.945	3.54E-06	1.45E-03	guanylate binding protein 5
PDCD1	55.04	1.009	3.57E-06	1.45E-03	programmed cell death 1
PYHIN1	96.23	0.930	3.48E-06	1.45E-03	pyrin and HIN domain family member 1
CXCL11	107.61	1.144	3.74E-06	1.50E-03	C-X-C motif chemokine ligand 11
APOL6	5247.56	0.625	4.42E-06	1.74E-03	apolipoprotein L6
CYSLTR2	39.06	0.756	4.55E-06	1.77E-03	cysteinyl leukotriene receptor 2
IL2RB	289.14	0.753	4.74E-06	1.82E-03	interleukin 2 receptor subunit beta
LINC01550	41.84	0.962	4.98E-06	1.89E-03	long intergenic non-protein coding RNA 1550
TRBC2	189.05	0.913	5.07E-06	1.90E-03	T cell receptor beta constant 2
IL9R	21.22	0.818	5.23E-06	1.93E-03	interleukin 9 receptor
ITM2A	279.09	0.894	5.45E-06	1.98E-03	integral membrane protein 2A
ITGAL	579.89	0.852	5.55E-06	2.00E-03	integrin subunit alpha L
MS4A6A	492.20	0.637	5.74E-06	2.04E-03	membrane spanning 4-domains A6A
GVINP1	603.21	0.793	6.24E-06	2.19E-03	GTPase, very large interferon inducible pseudogene 1
CCL5	369.07	0.945	6.91E-06	2.38E-03	C-C motif chemokine ligand 5
CXCR3	33.78	0.978	6.96E-06	2.38E-03	C-X-C motif chemokine receptor 3
CD3G	103.55	0.924	7.10E-06	2.40E-03	CD3g molecule
GNLY	143.36	0.860	7.30E-06	2.44E-03	granulysin
TRAC	187.37	0.894	8.42E-06	2.78E-03	T cell receptor alpha constant
NLRC3	310.36	0.816	8.74E-06	2.85E-03	NLR family CARD domain containing 3
SH2D1A	55.80	0.914	9.11E-06	2.94E-03	SH2 domain containing 1A

BTN3A3	749.19	0.579	9.88E-06	3.15E-03	butyrophilin subfamily 3 member A3
ABCA9	139.60	0.762	1.05E-05	3.27E-03	ATP binding cassette subfamily A member 9
FKBP5	786.60	0.591	1.06E-05	3.27E-03	FKBP prolyl isomerase 5
TRAJ33	5.63	1.021	1.06E-05	3.27E-03	T cell receptor alpha joining 33
GIMAP7	124.03	0.708	1.19E-05	3.58E-03	GTPase, IMAP family member 7
TIGIT	105.14	0.864	1.18E-05	3.58E-03	T cell immunoreceptor with Ig and ITIM domains
TRBC1	108.63	0.949	1.20E-05	3.58E-03	T cell receptor beta constant 1
APOBEC3H	11.15	0.848	1.22E-05	3.61E-03	apolipoprotein B mRNA editing enzyme catalytic subunit 3H
TRGV10	10.21	1.053	1.24E-05	3.64E-03	T cell receptor gamma variable 10 (non-functional)
SLIT2	76.08	0.850	1.31E-05	3.80E-03	slit guidance ligand 2
PPP1R16B	441.23	0.691	1.49E-05	4.22E-03	protein phosphatase 1 regulatory subunit 16B
Z95114.3	8.27	0.857	1.49E-05	4.22E-03	NA
LY9	80.98	0.898	1.53E-05	4.29E-03	lymphocyte antigen 9
HLA-DMB	601.78	0.732	1.67E-05	4.60E-03	major histocompatibility complex, class II, DM beta
SLA2	79.98	0.764	1.66E-05	4.60E-03	Src like adaptor 2
CD200R1	121.36	0.960	1.76E-05	4.79E-03	CD200 receptor 1
SIGLEC8	27.82	0.886	1.86E-05	5.01E-03	sialic acid binding Ig like lectin 8
EPB41L3	661.95	0.685	1.93E-05	5.15E-03	erythrocyte membrane protein band 4.1 like 3
ABCF1	1858.52	-0.211	1.99E-05	5.27E-03	ATP binding cassette subfamily F member 1
ACKR1	122.39	0.961	2.05E-05	5.27E-03	atypical chemokine receptor 1 (Duffy blood group)
CELF2	535.21	0.694	2.05E-05	5.27E-03	CUGBP Elav-like family member 2
CEP250-AS1	8.39	0.858	2.03E-05	5.27E-03	CEP250 antisense RNA 1
HLA-DQB1	1037.89	0.733	2.07E-05	5.27E-03	major histocompatibility complex, class II, DQ beta 1
NCKAP1L	705.77	0.661	2.13E-05	5.38E-03	NCK associated protein 1 like
TBC1D10C	170.08	0.803	2.24E-05	5.60E-03	TBC1 domain family member 10C
AL357054.4	10.82	0.710	2.47E-05	6.14E-03	NA
AOAH-IT1	12.73	0.823	2.60E-05	6.38E-03	AOAH intronic transcript 1
BTN3A2	1098.63	0.490	2.63E-05	6.38E-03	butyrophilin subfamily 3 member A2
RASAL3	230.99	0.701	2.64E-05	6.38E-03	RAS protein activator like 3
C1orf162	88.09	0.535	2.69E-05	6.45E-03	chromosome 1 open reading frame 162
SERPINA3	4608.82	-0.966	2.90E-05	6.89E-03	serpin family A member 3
HLA-DPA1	2446.97	0.695	2.98E-05	7.02E-03	major histocompatibility complex, class II, DP alpha 1
ZBTB16	99.05	0.886	3.15E-05	7.36E-03	zinc finger and BTB domain containing 16
TRAJ12	7.91	0.951	3.20E-05	7.42E-03	T cell receptor alpha joining 12
ABI3	225.23	0.541	3.42E-05	7.85E-03	ABI family member 3
DUX4L50	10.03	0.934	3.77E-05	8.59E-03	double homeobox 4 like 50 (pseudogene)
LINC01266	13.54	0.823	3.80E-05	8.60E-03	long intergenic non-protein coding RNA 1266
CLEC2D	434.32	0.566	4.04E-05	8.86E-03	C-type lectin domain family 2 member D
HLA-DOA	350.85	0.744	4.03E-05	8.86E-03	major histocompatibility complex, class II, DO alpha
SPOCK2	1954.36	0.679	4.01E-05	8.86E-03	SPARC (osteonectin), cwcv and kazal like domains proteoglycan 2
TRAJ7	7.35	0.914	4.03E-05	8.86E-03	T cell receptor alpha joining 7

TBX21	31.35	0.860	4.21E-05	9.16E-03	T-box transcription factor 21
PPP2R2B	63.08	0.917	4.47E-05	9.64E-03	protein phosphatase 2 regulatory subunit Bbeta
EVL	928.94	0.580	4.59E-05	9.82E-03	Enah/Vasp-like
BNC2	184.48	0.747	4.84E-05	1.01E-02	basonuclin 2
P2RY10	79.10	0.786	4.86E-05	1.01E-02	P2Y receptor family member 10
PPM1K	646.54	0.597	4.85E-05	1.01E-02	protein phosphatase, Mg2+/Mn2+ dependent 1K
PTN	193.58	0.892	4.76E-05	1.01E-02	pleiotrophin
LINC00861	75.27	0.861	4.98E-05	1.01E-02	long intergenic non-protein coding RNA 861
LINC01934	99.13	0.769	4.97E-05	1.01E-02	long intergenic non-protein coding RNA 1934
RNF150	67.70	0.739	4.97E-05	1.01E-02	ring finger protein 150
AL683807.1	17.32	0.811	5.18E-05	1.02E-02	NA
BRAFP1	3.38	0.925	5.15E-05	1.02E-02	BRAF pseudogene 1
KLRB1	49.45	0.801	5.20E-05	1.02E-02	killer cell lectin like receptor B1
SLAMF6	111.12	0.818	5.11E-05	1.02E-02	SLAM family member 6
ZNF831	150.82	0.810	5.22E-05	1.02E-02	zinc finger protein 831
CCL14	78.12	0.842	5.49E-05	1.07E-02	C-C motif chemokine ligand 14
SLAMF8	221.74	0.661	5.55E-05	1.07E-02	SLAM family member 8
AL392086.3	13.84	0.891	5.62E-05	1.08E-02	NA
GRAP2	126.84	0.732	5.77E-05	1.09E-02	GRB2 related adaptor protein 2
IFNG	18.44	0.910	5.77E-05	1.09E-02	interferon gamma
ITK	206.63	0.786	5.77E-05	1.09E-02	IL2 inducible T cell kinase
ASS1P1	18.88	0.860	5.87E-05	1.10E-02	argininosuccinate synthetase 1 pseudogene 1
CIT	1470.41	-0.435	5.90E-05	1.10E-02	citron rho-interacting serine/threonine kinase
CLEC2B	314.31	0.744	6.07E-05	1.11E-02	C-type lectin domain family 2 member B
GZMB	90.35	0.846	6.12E-05	1.11E-02	granzyme B
TPSP2	9.33	-0.887	6.06E-05	1.11E-02	tryptase pseudogene 2
UBASH3A	102.07	0.782	6.15E-05	1.11E-02	ubiquitin associated and SH3 domain containing A
ZYG11A	34.58	0.891	6.18E-05	1.11E-02	zyg-11 family member A, cell cycle regulator
LINC01358	29.74	0.745	6.25E-05	1.12E-02	long intergenic non-protein coding RNA 1358
CD6	181.14	0.658	6.38E-05	1.12E-02	CD6 molecule
IKZF1	508.92	0.670	6.39E-05	1.12E-02	IKAROS family zinc finger 1
IL21R	130.53	0.652	6.39E-05	1.12E-02	interleukin 21 receptor
SRR	73.95	0.424	6.54E-05	1.14E-02	serine racemase
CD52	137.22	0.725	6.68E-05	1.15E-02	CD52 molecule
TESPA1	68.26	0.692	6.71E-05	1.15E-02	thymocyte expressed, positive selection associated 1
CRTAM	23.61	0.843	6.79E-05	1.16E-02	cytotoxic and regulatory T cell molecule
ATP8B4	226.83	0.417	6.94E-05	1.18E-02	ATPase phospholipid transporting 8B4 (putative)
HLA-DPB1	1509.02	0.630	7.49E-05	1.27E-02	major histocompatibility complex, class II, DP beta 1
TRAV19	6.00	0.878	7.64E-05	1.28E-02	T cell receptor alpha variable 19
MAP4K1	148.66	0.624	7.70E-05	1.29E-02	mitogen-activated protein kinase kinase kinase kinase 1
ZNF471	116.64	0.760	7.77E-05	1.29E-02	zinc finger protein 471
TRAV21	4.77	0.871	8.24E-05	1.36E-02	T cell receptor alpha variable 21
PRKCQ	97.36	0.665	8.36E-05	1.36E-02	protein kinase C theta

SAMD9L	3756.08	0.672	8.35E-05	1.36E-02	sterile alpha motif domain containing 9 like
CH25H	29.64	0.804	8.51E-05	1.38E-02	cholesterol 25-hydroxylase
ADAM33	104.81	0.800	8.72E-05	1.40E-02	ADAM metallopeptidase domain 33
AL034550.3	5.28	-0.855	8.80E-05	1.40E-02	NA
FGL2	540.17	0.530	9.04E-05	1.40E-02	fibrinogen like 2
GYPC	138.61	0.579	9.07E-05	1.40E-02	glycophorin C (Gerbich blood group)
LINC02328	28.08	0.640	8.94E-05	1.40E-02	long intergenic non-protein coding RNA 2328
OGN	14.30	0.840	8.99E-05	1.40E-02	osteoglycin
SERPING1	850.15	0.684	9.11E-05	1.40E-02	serpin family G member 1
TIFAB	21.78	0.762	9.06E-05	1.40E-02	TIFA inhibitor
TRAF3IP1	426.01	-0.261	8.98E-05	1.40E-02	TRAF3 interacting protein 1
FLOT1	1814.50	-0.216	9.17E-05	1.40E-02	flotillin 1
EDARADD	75.85	0.856	9.45E-05	1.43E-02	EDAR associated death domain
SAMD3	50.43	0.788	9.41E-05	1.43E-02	sterile alpha motif domain containing 3
ANKRD44-IT1	80.33	0.605	1.02E-04	1.53E-02	ANKRD44 intronic transcript 1
TRAJ13	6.29	0.834	1.03E-04	1.54E-02	T cell receptor alpha joining 13
AL096816.1	11.05	0.820	1.05E-04	1.55E-02	NA
MTA1	1097.46	-0.188	1.05E-04	1.55E-02	metastasis associated 1
SFRP2	66.84	0.835	1.07E-04	1.57E-02	secreted frizzled related protein 2
SRCIN1	380.47	-0.793	1.08E-04	1.58E-02	SRC kinase signaling inhibitor 1
TRAJ29	6.05	0.829	1.08E-04	1.58E-02	T cell receptor alpha joining 29
AC079140.6	5.35	0.807	1.10E-04	1.59E-02	NA
LINC00426	75.11	0.693	1.09E-04	1.59E-02	long intergenic non-protein coding RNA 426
LINC01145	140.27	0.342	1.11E-04	1.59E-02	long intergenic non-protein coding RNA 1145
TRAJ11	7.03	0.832	1.16E-04	1.66E-02	T cell receptor alpha joining 11
GIMAP6	358.52	0.550	1.19E-04	1.69E-02	GTPase, IMAP family member 6
AOAH	321.51	0.630	1.20E-04	1.70E-02	acyloxyacyl hydrolase
VEGFA	6009.62	-0.452	1.24E-04	1.75E-02	vascular endothelial growth factor A
ABCA6	258.85	0.623	1.25E-04	1.75E-02	ATP binding cassette subfamily A member 6
CD160	13.81	0.808	1.29E-04	1.80E-02	CD160 molecule
KLRG1	40.46	0.648	1.34E-04	1.86E-02	killer cell lectin like receptor G1
ROBO1	1409.01	0.593	1.34E-04	1.86E-02	roundabout guidance receptor 1
SFRP1	97.93	0.810	1.38E-04	1.88E-02	secreted frizzled related protein 1
TRAJ3	5.34	0.811	1.37E-04	1.88E-02	T cell receptor alpha joining 3
ZAP70	229.27	0.707	1.39E-04	1.89E-02	zeta chain of T cell receptor associated protein kinase 70
ADORA3	31.49	0.686	1.40E-04	1.90E-02	adenosine A3 receptor
CADM3-AS1	42.74	0.807	1.41E-04	1.91E-02	CADM3 antisense RNA 1
PCED1B-AS1	67.34	0.641	1.42E-04	1.91E-02	PCED1B antisense RNA 1
POM121	1183.14	-0.218	1.43E-04	1.91E-02	POM121 transmembrane nucleoporin
SLC15A2	501.76	0.575	1.48E-04	1.97E-02	solute carrier family 15 member 2
SNTB2	1495.65	0.184	1.50E-04	1.98E-02	syntrophin beta 2
ACAP1	276.45	0.626	1.52E-04	2.00E-02	ArfGAP with coiled-coil, ankyrin repeat and PH domains 1
TRAJ40	6.60	0.789	1.53E-04	2.01E-02	T cell receptor alpha joining 40
SCN7A	35.62	0.789	1.55E-04	2.02E-02	sodium voltage-gated channel alpha subunit 7
ANTXRLP1	18.85	0.792	1.58E-04	2.03E-02	ANTXR like pseudogene 1

BRF1	912.14	-0.176	1.58E-04	2.03E-02	BRF1 RNA polymerase III transcription initiation factor subunit
PRKCH	1432.27	0.496	1.59E-04	2.03E-02	protein kinase C eta
ST6GAL2	21.73	0.693	1.58E-04	2.03E-02	ST6 beta-galactoside alpha-2,6-sialyltransferase 2
TRBV19	7.49	0.791	1.59E-04	2.03E-02	T cell receptor beta variable 19
ASB2	82.71	0.656	1.61E-04	2.04E-02	ankyrin repeat and SOCS box containing 2
TRAJ16	6.03	0.786	1.61E-04	2.04E-02	T cell receptor alpha joining 16
AC020549.2	3.55	0.773	1.68E-04	2.10E-02	NA
RIC3	94.00	0.783	1.68E-04	2.10E-02	RIC3 acetylcholine receptor chaperone
BLACAT1	281.55	-0.660	1.75E-04	2.18E-02	bladder cancer associated transcript 1
XCL2	6.60	0.743	1.75E-04	2.18E-02	X-C motif chemokine ligand 2
SRRT	1990.46	-0.173	1.79E-04	2.21E-02	serrate, RNA effector molecule
SFMBT2	435.65	0.501	1.80E-04	2.22E-02	Scm like with four mbt domains 2
AC021766.1	19.44	0.702	1.82E-04	2.23E-02	NA
FCGR2C	160.35	0.574	1.83E-04	2.23E-02	Fc fragment of IgG receptor IIc (gene/pseudogene)
CLIC5	347.09	0.731	1.85E-04	2.24E-02	chloride intracellular channel 5
FLT3LG	90.79	0.449	1.86E-04	2.24E-02	fms related receptor tyrosine kinase 3 ligand
ZNF415	86.00	0.526	1.85E-04	2.24E-02	zinc finger protein 415
RPS3AP37	7.21	0.621	1.88E-04	2.25E-02	RPS3A pseudogene 37
ZNF829	77.47	0.634	1.87E-04	2.25E-02	zinc finger protein 829
C1QB	1075.90	0.581	1.92E-04	2.27E-02	complement C1q B chain
RPS6KA5	862.99	0.311	1.92E-04	2.27E-02	ribosomal protein S6 kinase A5
HIP1R	3027.86	-0.416	1.93E-04	2.28E-02	huntingtin interacting protein 1 related
ZNF135	62.55	0.597	1.98E-04	2.32E-02	zinc finger protein 135
AC006033.2	11.95	0.696	2.02E-04	2.36E-02	NA
FAM107A	49.13	0.722	2.06E-04	2.40E-02	family with sequence similarity 107 member A
CD48	181.71	0.591	2.12E-04	2.43E-02	CD48 molecule
CDAN1	520.57	-0.164	2.12E-04	2.43E-02	codanin 1
DNMT3B	319.39	-0.585	2.12E-04	2.43E-02	DNA methyltransferase 3 beta
SPON1	369.69	0.711	2.10E-04	2.43E-02	spondin 1
ZNF430	582.96	0.230	2.14E-04	2.44E-02	zinc finger protein 430
C1orf35	260.77	-0.259	2.16E-04	2.45E-02	chromosome 1 open reading frame 35
SCIMP	70.94	0.581	2.18E-04	2.47E-02	SLP adaptor and CSK interacting membrane protein
ADH1B	106.97	0.746	2.21E-04	2.47E-02	alcohol dehydrogenase 1B (class I), beta polypeptide
ZNF142	946.58	-0.156	2.20E-04	2.47E-02	zinc finger protein 142
CADM3	7.06	0.713	2.26E-04	2.50E-02	cell adhesion molecule 3
CASP1	723.64	0.401	2.26E-04	2.50E-02	caspase 1
LINC01871	11.52	0.730	2.24E-04	2.50E-02	long intergenic non-protein coding RNA 1871
SIRPG	52.46	0.665	2.25E-04	2.50E-02	signal regulatory protein gamma
PLA2G2F	74.17	-0.727	2.30E-04	2.53E-02	phospholipase A2 group IIF
SRPX2	723.71	0.633	2.33E-04	2.55E-02	sushi repeat containing protein X-linked 2
AC098934.1	93.84	-0.569	2.43E-04	2.60E-02	tubulin alpha 5, pseudogene
AC104232.2	25.73	0.735	2.42E-04	2.60E-02	NA
CD1D	42.55	0.532	2.39E-04	2.60E-02	CD1d molecule

LINC02836	4.94	0.733	2.43E-04	2.60E-02	long intergenic non-protein coding RNA 2836
RTCA-AS1	119.34	0.328	2.43E-04	2.60E-02	RTCA antisense RNA 1
SASH3	307.27	0.554	2.39E-04	2.60E-02	SAM and SH3 domain containing 3
WDR62	439.42	-0.502	2.44E-04	2.60E-02	WD repeat domain 62
NSUN5	367.65	-0.184	2.46E-04	2.61E-02	NOP2/Sun RNA methyltransferase 5
GIMAP1	107.61	0.541	2.49E-04	2.63E-02	GTPase, IMAP family member 1
RTKN	819.89	-0.377	2.49E-04	2.63E-02	rhotekin
TMEM178B	59.56	0.730	2.51E-04	2.63E-02	transmembrane protein 178B
IL12A	55.81	0.727	2.55E-04	2.65E-02	interleukin 12A
TRDC	25.82	0.721	2.56E-04	2.65E-02	T cell receptor delta constant
ZBTB20-AS1	13.85	0.694	2.55E-04	2.65E-02	ZBTB20 antisense RNA 1
PRKAA2	231.13	0.728	2.57E-04	2.66E-02	protein kinase AMP-activated catalytic subunit alpha 2
CLCN5	1335.89	-0.399	2.62E-04	2.70E-02	chloride voltage-gated channel 5
KRT17P8	9.61	0.722	2.66E-04	2.73E-02	keratin 17 pseudogene 8
ZNF568	127.84	0.554	2.67E-04	2.73E-02	zinc finger protein 568
AL135818.1	35.42	0.509	2.69E-04	2.74E-02	NA
TRAT1	31.26	0.694	2.70E-04	2.75E-02	T cell receptor associated transmembrane adaptor 1
MFAP4	317.30	0.705	2.77E-04	2.80E-02	microfibril associated protein 4
FAM78A	188.65	0.459	2.80E-04	2.81E-02	family with sequence similarity 78 member A
NAAA	389.78	0.232	2.80E-04	2.81E-02	N-acylethanolamine acid amidase
ARMC3	37.00	-0.711	2.84E-04	2.83E-02	armadillo repeat containing 3
RSPO1	7.17	0.708	2.84E-04	2.83E-02	R-spondin 1
LINC00278	350.96	0.693	2.96E-04	2.95E-02	long intergenic non-protein coding RNA 278
CALHM6	40.70	0.620	2.99E-04	2.97E-02	calcium homeostasis modulator family member 6
SYNE4	82.27	-0.612	3.02E-04	2.97E-02	spectrin repeat containing nuclear envelope family member 4
TAF6	711.26	-0.198	3.02E-04	2.97E-02	TATA-box binding protein associated factor 6
SIT1	28.64	0.611	3.10E-04	3.04E-02	signaling threshold regulating transmembrane adaptor 1
SLC9A9	353.69	0.627	3.11E-04	3.04E-02	solute carrier family 9 member A9
ANKRD20A5P	691.55	0.692	3.14E-04	3.06E-02	ankyrin repeat domain 20 family member A5, pseudogene
COLEC12	67.77	0.662	3.19E-04	3.09E-02	collectin subfamily member 12
C4orf50	23.09	0.691	3.21E-04	3.10E-02	chromosome 4 open reading frame 50
CD7	82.50	0.568	3.23E-04	3.11E-02	CD7 molecule
HLA-DRA	5462.44	0.566	3.23E-04	3.11E-02	major histocompatibility complex, class II, DR alpha
ARHGAP39	281.44	-0.409	3.37E-04	3.22E-02	Rho GTPase activating protein 39
BOC	235.79	0.648	3.40E-04	3.25E-02	BOC cell adhesion associated, oncogene regulated
IRF1	2678.35	0.370	3.42E-04	3.25E-02	interferon regulatory factor 1
LAX1	147.41	0.649	3.43E-04	3.25E-02	lymphocyte transmembrane adaptor 1
CD40	246.43	0.411	3.46E-04	3.27E-02	CD40 molecule
ARRDC5	17.17	0.681	3.49E-04	3.29E-02	arrestin domain containing 5
CTSW	69.06	0.633	3.54E-04	3.29E-02	cathepsin W
MPEG1	796.82	0.352	3.56E-04	3.29E-02	macrophage expressed 1

MUSK	27.55	0.682	3.55E-04	3.29E-02	muscle associated receptor tyrosine kinase
RTN2	95.00	-0.430	3.57E-04	3.29E-02	reticulon 2
SLCO2B1	1104.14	0.469	3.55E-04	3.29E-02	solute carrier organic anion transporter family member 2B1
TRAJ1	3.55	0.661	3.55E-04	3.29E-02	T cell receptor alpha joining 1 (non-functional)
TTLL4	1117.71	-0.194	3.58E-04	3.29E-02	tubulin tyrosine ligase like 4
DPT	48.89	0.650	3.70E-04	3.39E-02	dermatopontin
FOSB	760.43	-0.656	3.75E-04	3.42E-02	FosB proto-oncogene, AP-1 transcription factor subunit
CCDC180	109.74	0.566	3.87E-04	3.51E-02	coiled-coil domain containing 180
IRF1-AS1	617.07	0.410	3.87E-04	3.51E-02	IRF1 antisense RNA 1
ZNF793	127.50	0.629	3.88E-04	3.51E-02	zinc finger protein 793
AC005332.5	157.67	-0.298	3.91E-04	3.53E-02	NA
RHPN1	426.55	-0.496	3.92E-04	3.53E-02	rhophilin Rho GTPase binding protein 1
RAET1K	9.88	-0.646	3.98E-04	3.57E-02	retinoic acid early transcript 1K pseudogene
ARHGAP20	65.66	0.630	4.03E-04	3.59E-02	Rho GTPase activating protein 20
TRIM69	854.18	0.289	4.03E-04	3.59E-02	tripartite motif containing 69
FGD2	415.76	0.373	4.11E-04	3.65E-02	FYVE, RhoGEF and PH domain containing 2
AC005256.1	11.94	-0.633	4.13E-04	3.65E-02	NA
C3orf80	60.66	0.622	4.14E-04	3.65E-02	chromosome 3 open reading frame 80
PHKG2	585.75	-0.165	4.16E-04	3.66E-02	phosphorylase kinase catalytic subunit gamma 2
ATG2A	1302.57	-0.185	4.28E-04	3.66E-02	autophagy related 2A
BX284613.2	30.74	0.646	4.29E-04	3.66E-02	NA
CD74	12885.63	0.512	4.18E-04	3.66E-02	CD74 molecule
CHAF1B	290.37	-0.294	4.28E-04	3.66E-02	chromatin assembly factor 1 subunit B
GAS1RR	8.72	0.640	4.25E-04	3.66E-02	GAS1 adjacent regulatory RNA
HDX	62.81	0.505	4.29E-04	3.66E-02	highly divergent homeobox
LMO7DN-IT1	9.61	0.645	4.28E-04	3.66E-02	LMO7DN intronic transcript 1
RPL7AP64	8.11	0.645	4.21E-04	3.66E-02	ribosomal protein L7a pseudogene 64
TRAJ28	4.86	0.655	4.28E-04	3.66E-02	T cell receptor alpha joining 28
RGMA	293.71	0.653	4.32E-04	3.68E-02	repulsive guidance molecule BMP co-receptor a
CDC42EP4	1801.42	-0.253	4.34E-04	3.69E-02	CDC42 effector protein 4
AC008969.1	140.95	0.474	4.40E-04	3.73E-02	uncharacterized LOC100128398
AC092535.4	9.87	0.635	4.43E-04	3.74E-02	uncharacterized LOC105374344
LAG3	61.96	0.464	4.44E-04	3.74E-02	lymphocyte activating 3
CMKLR1	247.07	0.473	4.46E-04	3.74E-02	chemerin chemokine-like receptor 1
H2BE1	3.20	0.641	4.48E-04	3.75E-02	H2B.E variant histone 1
PPIF	2318.80	-0.382	4.53E-04	3.77E-02	peptidylprolyl isomerase F
TEF	463.78	0.429	4.53E-04	3.77E-02	TEF transcription factor, PAR bZIP family member
KLHDC1	69.42	0.436	4.57E-04	3.79E-02	kelch domain containing 1
FLG2	60.33	0.463	4.62E-04	3.82E-02	filaggrin family member 2
GZMM	15.94	0.629	4.63E-04	3.82E-02	granzyme M
AC010623.1	13.54	0.640	4.71E-04	3.87E-02	NA
GCSAM	39.36	0.601	4.72E-04	3.87E-02	germinal center associated signaling and motility
CD38	254.15	0.523	4.78E-04	3.91E-02	CD38 molecule

HS3ST3B1	153.13	0.406	4.83E-04	3.94E-02	heparan sulfate-glucosamine 3-sulfotransferase 3B1
PTPRC	2125.00	0.539	5.03E-04	4.09E-02	protein tyrosine phosphatase receptor type C
MUC22	1002.10	0.405	5.09E-04	4.13E-02	mucin 22
ACOX2	63.25	0.561	5.12E-04	4.14E-02	acyl-CoA oxidase 2
CYP7B1	400.81	0.598	5.19E-04	4.19E-02	cytochrome P450 family 7 subfamily B member 1
AL591518.1	6.89	0.622	5.27E-04	4.23E-02	NA
RN7SL751P	13.38	0.609	5.28E-04	4.23E-02	RNA, 7SL, cytoplasmic 751, pseudogene
ASGR2	18.77	0.589	5.29E-04	4.23E-02	asialoglycoprotein receptor 2
MAPK10	284.18	0.532	5.32E-04	4.23E-02	mitogen-activated protein kinase 10
LINC01357	25.23	-0.544	5.34E-04	4.25E-02	long intergenic non-protein coding RNA 1357
CCNL2P1	20.34	0.553	5.36E-04	4.25E-02	cyclin L2 pseudogene 1
GHR	90.90	0.596	5.39E-04	4.25E-02	growth hormone receptor
DOCK10	938.63	0.401	5.48E-04	4.32E-02	dedicator of cytokinesis 10
CIB2	112.05	-0.519	5.53E-04	4.33E-02	calcium and integrin binding family member 2
PSMB9	649.54	0.414	5.51E-04	4.33E-02	proteasome 20S subunit beta 9
PTPN22	285.72	0.434	5.54E-04	4.33E-02	protein tyrosine phosphatase non-receptor type 22
AC100858.3	36.38	0.603	5.57E-04	4.33E-02	uncharacterized LOC105375743
TRBV27	3.82	0.614	5.56E-04	4.33E-02	T cell receptor beta variable 27
APOBEC3C	457.21	0.452	5.63E-04	4.33E-02	apolipoprotein B mRNA editing enzyme catalytic subunit 3C
CSF1R	658.30	0.312	5.61E-04	4.33E-02	colony stimulating factor 1 receptor
LCK	229.79	0.496	5.64E-04	4.33E-02	LCK proto-oncogene, Src family tyrosine kinase
LINC01169	36.24	-0.582	5.65E-04	4.33E-02	long intergenic non-protein coding RNA 1169
ZNF382	69.32	0.404	5.59E-04	4.33E-02	zinc finger protein 382
RIOX1	263.07	-0.165	5.68E-04	4.34E-02	ribosomal oxygenase 1
CCL19	18.06	0.602	5.70E-04	4.34E-02	C-C motif chemokine ligand 19
AC105046.1	16.15	-0.605	5.87E-04	4.42E-02	NA
ADCK5	383.73	-0.241	5.84E-04	4.42E-02	aarF domain containing kinase 5
GFI1	102.21	0.445	5.86E-04	4.42E-02	growth factor independent 1 transcriptional repressor
MELTF	681.08	-0.584	5.83E-04	4.42E-02	melanotransferrin
CCDC80	535.97	0.545	5.93E-04	4.46E-02	coiled-coil domain containing 80
SLC13A4	62.61	0.527	6.01E-04	4.48E-02	solute carrier family 13 member 4
SLFN12L	233.74	0.476	6.00E-04	4.48E-02	schlafen family member 12 like
YBX2P2	10.70	0.548	6.00E-04	4.48E-02	YBX2 pseudogene 2
EPPK1	5064.97	-0.450	6.03E-04	4.48E-02	epiplakin 1
PRKACB	970.83	0.436	6.05E-04	4.49E-02	protein kinase cAMP-activated catalytic subunit beta
NPAS3	20.87	0.575	6.15E-04	4.54E-02	neuronal PAS domain protein 3
SLC24A3	303.15	0.582	6.14E-04	4.54E-02	solute carrier family 24 member 3
NR4A1	769.29	-0.543	6.18E-04	4.55E-02	nuclear receptor subfamily 4 group A member 1
CEP250	1714.30	-0.185	6.27E-04	4.60E-02	centrosomal protein 250
TRAJ18	6.12	0.594	6.40E-04	4.69E-02	T cell receptor alpha joining 18
ZNF667-AS1	46.82	0.524	6.45E-04	4.71E-02	ZNF667 antisense RNA 1 (head to head)
AC016582.3	26.19	0.560	6.49E-04	4.72E-02	NA
ZNF606	348.25	0.399	6.48E-04	4.72E-02	zinc finger protein 606

AC013474.1	7.57	0.592	6.52E-04	4.73E-02	NA
PTGDR	29.31	0.585	6.56E-04	4.74E-02	prostaglandin D2 receptor
RAB39B	18.14	0.565	6.61E-04	4.75E-02	RAB39B, member RAS oncogene family
ZNF141	937.75	0.365	6.60E-04	4.75E-02	zinc finger protein 141
AATF	1129.18	-0.187	6.66E-04	4.78E-02	apoptosis antagonizing transcription factor
RORA	1144.88	0.460	6.72E-04	4.81E-02	RAR related orphan receptor A
SLC14A1	37.79	0.566	6.77E-04	4.83E-02	solute carrier family 14 member 1 (Kidd blood group)
ADPRM	88.25	0.219	6.80E-04	4.84E-02	ADP-ribose/CDP-alcohol diphosphatase, manganese dependent
AC120049.1	9.82	0.559	6.83E-04	4.85E-02	SETBP1 divergent transcript
CCL26	10.96	-0.584	6.92E-04	4.89E-02	C-C motif chemokine ligand 26
PSTPIP1	205.48	0.453	6.91E-04	4.89E-02	proline-serine-threonine phosphatase interacting protein 1
C1QA	758.03	0.460	7.01E-04	4.92E-02	complement C1q A chain
FCGR2B	79.61	0.492	7.01E-04	4.92E-02	Fc fragment of IgG receptor IIb
TRAJ5	4.98	0.545	7.02E-04	4.92E-02	T cell receptor alpha joining 5
CDC6	1213.88	-0.506	7.08E-04	4.95E-02	cell division cycle 6
BCORP1	25.17	0.563	7.11E-04	4.96E-02	BCL6 corepressor pseudogene 1
GAB3	172.47	0.360	7.21E-04	5.02E-02	GRB2 associated binding protein 3
CD27	83.75	0.529	7.36E-04	5.04E-02	CD27 molecule
CX3CL1	339.52	0.489	7.37E-04	5.04E-02	C-X3-C motif chemokine ligand 1
FMO2	4895.62	0.498	7.31E-04	5.04E-02	flavin containing dimethylaniline monoxygenase 2
PRAME	162.20	-0.510	7.35E-04	5.04E-02	PRAME nuclear receptor transcriptional regulator
RPL12P20	9.58	0.574	7.29E-04	5.04E-02	ribosomal protein L12 pseudogene 20
SCX	66.01	-0.389	7.31E-04	5.04E-02	scleraxis bHLH transcription factor
TRAJ42	3.74	0.534	7.35E-04	5.04E-02	T cell receptor alpha joining 42
SPAG5	1206.21	-0.349	7.41E-04	5.06E-02	sperm associated antigen 5
NAPSB	94.27	0.521	7.47E-04	5.09E-02	napsin B aspartic peptidase, pseudogene
MIR99AHG	430.08	0.552	7.52E-04	5.10E-02	mir-99a-let-7c cluster host gene
IL12RB1	112.79	0.492	7.56E-04	5.11E-02	interleukin 12 receptor subunit beta 1
SEMA6A	655.60	0.549	7.55E-04	5.11E-02	semaphorin 6A
AKT1	3094.78	-0.154	7.79E-04	5.22E-02	AKT serine/threonine kinase 1
C14orf132	217.36	0.504	7.77E-04	5.22E-02	chromosome 14 open reading frame 132
GBP1	2047.14	0.440	7.75E-04	5.22E-02	guanylate binding protein 1
NCAPG2	932.96	-0.235	7.79E-04	5.22E-02	non-SMC condensin II complex subunit G2
AC005993.1	62.97	-0.527	7.83E-04	5.23E-02	NA
TTC14	1833.04	0.233	7.87E-04	5.24E-02	tetratricopeptide repeat domain 14
WNT5A	2056.68	0.478	7.98E-04	5.30E-02	Wnt family member 5A
NLRC5	2944.34	0.303	8.20E-04	5.41E-02	NLR family CARD domain containing 5
PURA	1824.92	0.180	8.17E-04	5.41E-02	purine rich element binding protein A
ZMIZ2	2025.56	-0.168	8.19E-04	5.41E-02	zinc finger MIZ-type containing 2
TACR1	17.43	0.553	8.28E-04	5.45E-02	tachykinin receptor 1
SNORA11B	14.62	0.400	8.37E-04	5.50E-02	small nucleolar RNA, H/ACA box 11B
USP30-AS1	15.78	0.533	8.41E-04	5.51E-02	USP30 antisense RNA 1

CSPG4	776.28	-0.547	8.53E-04	5.51E-02	chondroitin sulfate proteoglycan 4
DDAH2	529.45	-0.314	8.47E-04	5.51E-02	dimethylarginine dimethylaminohydrolase 2
HLA-DRB1	2793.70	0.432	8.46E-04	5.51E-02	major histocompatibility complex, class II, DR beta 1
LINC00298	5.08	0.541	8.53E-04	5.51E-02	long intergenic non-protein coding RNA 298
TOGARAM2	57.63	0.413	8.52E-04	5.51E-02	TOG array regulator of axonemal microtubules 2
TSPAN5	395.55	0.496	8.52E-04	5.51E-02	tetraspanin 5
TRIM22	1869.10	0.473	8.59E-04	5.54E-02	tripartite motif containing 22
MYRIP	79.63	0.540	8.76E-04	5.62E-02	myosin VIIA and Rab interacting protein
SEPTIN6	530.03	0.254	8.74E-04	5.62E-02	septin 6
DPYD	1620.03	0.463	8.81E-04	5.64E-02	dihydropyrimidine dehydrogenase
EPHB4	2307.47	-0.276	8.88E-04	5.67E-02	EPH receptor B4
ZNF625	70.59	0.403	8.91E-04	5.68E-02	zinc finger protein 625
NSRP1	716.14	-0.135	8.95E-04	5.68E-02	nuclear speckle splicing regulatory protein 1
SLC25A46	1177.16	0.274	8.95E-04	5.68E-02	solute carrier family 25 member 46
AL512590.2	39.75	0.465	9.01E-04	5.70E-02	NA
NUDT6	88.55	0.210	9.11E-04	5.73E-02	nudix hydrolase 6
RORB	23.37	0.525	9.12E-04	5.73E-02	RAR related orphan receptor B
SSBP2	397.69	0.459	9.10E-04	5.73E-02	single stranded DNA binding protein 2
AC093001.1	157.31	-0.504	9.24E-04	5.80E-02	NA
GYS2	69.29	0.475	9.30E-04	5.80E-02	glycogen synthase 2
LINC01913	28.36	-0.513	9.30E-04	5.80E-02	long intergenic non-protein coding RNA 1913
PLXDC2	1729.54	0.386	9.31E-04	5.80E-02	plexin domain containing 2
ANKRD44	801.58	0.375	9.37E-04	5.82E-02	ankyrin repeat domain 44
SCML4	92.72	0.494	9.40E-04	5.83E-02	Scm polycomb group protein like 4
MEF2C-AS1	22.50	0.481	9.44E-04	5.83E-02	MEF2C antisense RNA 1
TFEC	200.96	0.438	9.43E-04	5.83E-02	transcription factor EC
AL359220.1	91.27	0.421	9.57E-04	5.89E-02	NA
DENND1C	137.62	0.414	9.55E-04	5.89E-02	DENN domain containing 1C
GGTA1P	85.11	0.509	9.76E-04	5.95E-02	glycoprotein alpha-galactosyltransferase 1 (inactive)
L3MBTL4	131.29	0.511	9.80E-04	5.95E-02	L3MBTL histone methyl-lysine binding protein 4
LDOC1	45.66	0.520	9.76E-04	5.95E-02	LDOC1 regulator of NFKB signaling
MS4A4E	35.61	0.458	9.80E-04	5.95E-02	membrane spanning 4-domains A4E
TRAJ24	7.00	0.518	9.72E-04	5.95E-02	T cell receptor alpha joining 24
TTTY10	34.13	0.473	9.81E-04	5.95E-02	testis-specific transcript, Y-linked 10
CFH	2015.02	0.446	9.90E-04	6.00E-02	complement factor H
POMC	6.73	0.461	9.93E-04	6.00E-02	proopiomelanocortin
MAT2B	990.13	0.200	9.95E-04	6.00E-02	methionine adenosyltransferase 2B
CIITA	1433.74	0.391	1.00E-03	6.00E-02	class II major histocompatibility complex transactivator
PSMB8-AS1	177.18	0.332	1.00E-03	6.00E-02	PSMB8 antisense RNA 1 (head to head)
RNU6-316P	13.33	0.495	1.00E-03	6.00E-02	RNA, U6 small nuclear 316, pseudogene
XRN1	3997.64	0.169	1.00E-03	6.00E-02	5'-3' exoribonuclease 1
ZSCAN18	282.06	0.445	1.01E-03	6.01E-02	zinc finger and SCAN domain containing 18
SHARPIN	720.03	-0.163	1.02E-03	6.07E-02	SHANK associated RH domain interactor
SNHG14	1566.49	0.434	1.03E-03	6.13E-02	small nucleolar RNA host gene 14

ZSCAN12P1	74.51	-0.476	1.04E-03	6.17E-02	zinc finger and SCAN domain containing 12 pseudogene 1
SNX21	465.74	-0.219	1.05E-03	6.20E-02	sorting nexin family member 21
HSF1	1460.04	-0.165	1.05E-03	6.21E-02	heat shock transcription factor 1
AC025539.1	34.83	0.517	1.06E-03	6.24E-02	NA
CD247	106.66	0.425	1.07E-03	6.28E-02	CD247 molecule
AR	90.69	0.489	1.07E-03	6.29E-02	androgen receptor
AC008434.1	14.04	0.369	1.09E-03	6.38E-02	NA
TSPAN32	37.28	0.497	1.09E-03	6.38E-02	tetraspanin 32
FTH1P22	4.68	0.514	1.10E-03	6.39E-02	ferritin heavy chain 1 pseudogene 22
KRTAP5-AS1	84.94	-0.482	1.10E-03	6.39E-02	KRTAP5-1/KRTAP5-2 antisense RNA 1
TRAJ34	3.40	0.489	1.10E-03	6.39E-02	T cell receptor alpha joining 34
AC234582.1	69.00	-0.331	1.11E-03	6.40E-02	THBS3 antisense RNA 1
ACBD6	925.09	-0.147	1.11E-03	6.40E-02	acyl-CoA binding domain containing 6
IKZF3	1040.31	0.467	1.11E-03	6.40E-02	IKAROS family zinc finger 3
TRAJ44	3.87	0.472	1.11E-03	6.40E-02	T cell receptor alpha joining 44
AC026150.2	3.53	0.501	1.12E-03	6.46E-02	NA
CNTN1	172.95	0.453	1.13E-03	6.48E-02	contactin 1
APOBEC3F	172.77	0.198	1.14E-03	6.52E-02	apolipoprotein B mRNA editing enzyme catalytic subunit 3F
TRAJ22	5.40	0.489	1.14E-03	6.53E-02	T cell receptor alpha joining 22
CDHR1	171.94	0.444	1.15E-03	6.56E-02	cadherin related family member 1
PTPN7	271.74	0.413	1.15E-03	6.57E-02	protein tyrosine phosphatase non-receptor type 7
MYO1G	339.85	0.370	1.16E-03	6.59E-02	myosin IG
AGAP1	1582.31	-0.198	1.17E-03	6.61E-02	ArfGAP with GTPase domain, ankyrin repeat and PH domain 1
MROH1	3303.90	-0.230	1.17E-03	6.61E-02	maestro heat like repeat family member 1
SMS	1326.93	-0.264	1.17E-03	6.61E-02	spermine synthase
XCR1	29.70	0.498	1.17E-03	6.61E-02	X-C motif chemokine receptor 1
ACOT9	819.34	-0.259	1.19E-03	6.67E-02	acyl-CoA thioesterase 9
RNF165	48.44	0.497	1.19E-03	6.69E-02	ring finger protein 165
AC022126.1	22.40	0.459	1.20E-03	6.70E-02	NA
DOCK2	957.37	0.402	1.20E-03	6.70E-02	dedicator of cytokinesis 2
TAPBPL	802.31	0.247	1.20E-03	6.70E-02	TAP binding protein like
AOX1	52.93	0.487	1.22E-03	6.77E-02	aldehyde oxidase 1
AC074141.1	64.33	0.357	1.22E-03	6.77E-02	NA
BTN3A1	1307.44	0.304	1.22E-03	6.77E-02	butyrophilin subfamily 3 member A1
IGIP	313.91	0.209	1.23E-03	6.77E-02	IgA inducing protein
RSL24D1P8	4.74	0.495	1.23E-03	6.77E-02	ribosomal L24 domain containing 1 pseudogene 8
IL16	391.12	0.408	1.23E-03	6.79E-02	interleukin 16
UNC5CL	746.40	-0.490	1.24E-03	6.80E-02	unc-5 family C-terminal like
AL365361.1	61.61	0.470	1.25E-03	6.86E-02	NA
FAM241B	42.16	-0.320	1.26E-03	6.86E-02	family with sequence similarity 241 member B
ZNF619	280.00	0.141	1.25E-03	6.86E-02	zinc finger protein 619
ZNF695	217.47	-0.475	1.26E-03	6.86E-02	zinc finger protein 695

AP003559.1	30.59	-0.430	1.26E-03	6.86E-02	NA
CCDC40	121.57	-0.388	1.27E-03	6.88E-02	coiled-coil domain containing 40
ERAP1	2158.29	0.196	1.27E-03	6.88E-02	endoplasmic reticulum aminopeptidase 1
DUSP8	255.67	-0.387	1.27E-03	6.88E-02	dual specificity phosphatase 8
AC130462.2	4.26	0.488	1.27E-03	6.89E-02	NA
CXCL2	589.28	-0.488	1.28E-03	6.90E-02	C-X-C motif chemokine ligand 2
LINC00892	6.45	0.442	1.28E-03	6.90E-02	long intergenic non-protein coding RNA 892
PLSCR4	338.28	0.219	1.28E-03	6.90E-02	phospholipid scramblase 4
AC004702.1	6.75	0.462	1.29E-03	6.90E-02	NA
SLC1A3	596.76	0.466	1.29E-03	6.93E-02	solute carrier family 1 member 3
AC022898.2	7.23	0.488	1.30E-03	6.93E-02	NA
Z82206.1	9.97	0.472	1.30E-03	6.93E-02	NA
FAM114A2	493.84	0.199	1.32E-03	7.02E-02	family with sequence similarity 114 member A2
GPR25	6.21	0.480	1.32E-03	7.02E-02	G protein-coupled receptor 25
PPP1R15A	2445.53	-0.264	1.32E-03	7.02E-02	protein phosphatase 1 regulatory subunit 15A
AL365475.1	9.22	0.482	1.33E-03	7.06E-02	NA
IGHJ3P	57.06	0.466	1.34E-03	7.07E-02	immunoglobulin heavy joining 3P (pseudogene)
ZNF570	156.10	0.360	1.34E-03	7.07E-02	zinc finger protein 570
BEX4	186.93	0.432	1.34E-03	7.07E-02	brain expressed X-linked 4
DHX16	1102.72	-0.131	1.36E-03	7.12E-02	DEAH-box helicase 16
GPR1	58.44	0.430	1.36E-03	7.12E-02	chemerin chemokine-like receptor 2
LONP1	1370.16	-0.138	1.36E-03	7.12E-02	lon peptidase 1, mitochondrial
AC012181.1	33.44	0.329	1.38E-03	7.12E-02	NA
AC023024.2	31.79	-0.402	1.37E-03	7.12E-02	NA
AC092718.4	105.64	-0.240	1.39E-03	7.12E-02	NA
CCDC114	20.61	-0.452	1.37E-03	7.12E-02	outer dynein arm docking complex subunit 1
CST6	105.41	0.435	1.38E-03	7.12E-02	cystatin E/M
FOXO6	18.32	-0.462	1.38E-03	7.12E-02	forkhead box O6
JAK2	1895.60	0.263	1.38E-03	7.12E-02	Janus kinase 2
KIR2DL4	9.57	0.464	1.38E-03	7.12E-02	killer cell immunoglobulin like receptor, two Ig domains and long cytoplasmic tail 4
TNFAIP8L2	36.75	0.349	1.37E-03	7.12E-02	TNF alpha induced protein 8 like 2
TNNI3K	22.63	0.405	1.38E-03	7.12E-02	TNNI3 interacting kinase
PAQR6	106.49	-0.240	1.39E-03	7.15E-02	progestin and adipoQ receptor family member 6
C21orf58	311.62	-0.279	1.40E-03	7.15E-02	chromosome 21 open reading frame 58
DDX54	1419.25	-0.157	1.44E-03	7.32E-02	DEAD-box helicase 54
GBP2	2315.60	0.278	1.44E-03	7.32E-02	guanylate binding protein 2
TONSL	787.32	-0.314	1.43E-03	7.32E-02	tonsoku like, DNA repair protein
SMG5	4092.11	-0.154	1.44E-03	7.32E-02	SMG5 nonsense mediated mRNA decay factor
LRWD1	339.78	-0.196	1.48E-03	7.52E-02	leucine rich repeats and WD repeat domain containing 1
ZNF334	150.49	0.468	1.49E-03	7.55E-02	zinc finger protein 334
AC021205.3	9.64	0.426	1.50E-03	7.58E-02	NA
ZFP28	106.43	0.346	1.51E-03	7.60E-02	ZFP28 zinc finger protein
AL022098.1	8.02	0.452	1.52E-03	7.63E-02	NA
GPR18	15.68	0.451	1.53E-03	7.65E-02	G protein-coupled receptor 18
SAMD12-AS1	77.35	-0.334	1.53E-03	7.65E-02	SAMD12 antisense RNA 1

TRAJ17	5.14	0.454	1.53E-03	7.65E-02	T cell receptor alpha joining 17
TRAJ21	5.11	0.455	1.53E-03	7.65E-02	T cell receptor alpha joining 21
AC010536.1	48.48	-0.222	1.55E-03	7.72E-02	uncharacterized LOC100129215
CHST9	45.44	0.324	1.56E-03	7.78E-02	carbohydrate sulfotransferase 9
SOX9-AS1	95.42	-0.352	1.57E-03	7.78E-02	SOX9 antisense RNA 1
AC023830.3	28.42	0.425	1.57E-03	7.79E-02	NA
KIF24	421.91	-0.204	1.58E-03	7.84E-02	kinesin family member 24
AC027097.2	39.41	0.351	1.59E-03	7.87E-02	NA
ZBED2	60.61	0.459	1.59E-03	7.87E-02	zinc finger BED-type containing 2
IPCEF1	178.82	0.313	1.61E-03	7.96E-02	interaction protein for cytohesin exchange factors 1
PASK	669.68	-0.150	1.62E-03	7.97E-02	PAS domain containing serine/threonine kinase
AC007381.1	8.88	0.430	1.63E-03	8.01E-02	MIR4432 host gene
RECQL4	674.80	-0.324	1.63E-03	8.01E-02	RecQ like helicase 4
FRMD6	2074.21	0.449	1.65E-03	8.04E-02	FERM domain containing 6
ZNF793-AS1	7.78	0.441	1.65E-03	8.04E-02	ZNF793 antisense RNA 1 (head to head)
ELK3	1510.22	0.210	1.66E-03	8.09E-02	ETS transcription factor ELK3
ATP12A	100.42	0.388	1.67E-03	8.12E-02	ATPase H+/K+ transporting non-gastric alpha2 subunit
ZNF675	450.08	0.293	1.71E-03	8.30E-02	zinc finger protein 675
AL078622.1	27.54	0.439	1.73E-03	8.38E-02	uncharacterized LOC105373085
AL645568.1	104.70	0.260	1.76E-03	8.50E-02	PRDX6 antisense RNA 1
STAG3L3	213.60	-0.158	1.76E-03	8.50E-02	tripartite motif containing 74
GIMAP2	152.05	0.349	1.76E-03	8.52E-02	GTPase, IMAP family member 2
SLC52A1	54.83	0.439	1.77E-03	8.52E-02	solute carrier family 52 member 1
GNAO1	58.64	0.376	1.79E-03	8.58E-02	G protein subunit alpha o1
TLR3	570.65	0.430	1.79E-03	8.58E-02	toll like receptor 3
ZNF208	60.82	0.403	1.79E-03	8.58E-02	zinc finger protein 208
FANCA	1171.69	-0.246	1.80E-03	8.59E-02	FA complementation group A
GAS8	642.33	-0.182	1.81E-03	8.63E-02	growth arrest specific 8
AL590764.1	23.10	0.439	1.82E-03	8.64E-02	NA
IFI16	2571.19	0.351	1.82E-03	8.64E-02	interferon gamma inducible protein 16
ZNF540	50.15	0.334	1.81E-03	8.64E-02	zinc finger protein 540
UBE2L6	1095.37	0.287	1.83E-03	8.67E-02	ubiquitin conjugating enzyme E2 L6
USP9Y	1425.14	0.415	1.83E-03	8.67E-02	ubiquitin specific peptidase 9 Y-linked
TRAJ46	3.20	0.422	1.85E-03	8.72E-02	T cell receptor alpha joining 46
WARS1	3714.04	0.284	1.85E-03	8.72E-02	tryptophanyl-tRNA synthetase 1
RAB15	572.69	-0.360	1.85E-03	8.73E-02	RAB15, member RAS oncogene family
KEL	17.61	0.435	1.87E-03	8.77E-02	Kell metallo-endopeptidase (Kell blood group)
ZNF879	89.83	0.344	1.87E-03	8.77E-02	zinc finger protein 879
PRCP	1465.43	0.245	1.87E-03	8.77E-02	prolylcarboxypeptidase
ZBTB2P1	14.26	-0.394	1.88E-03	8.80E-02	ZBTB2 pseudogene 1
RNGTTP1	4.76	0.412	1.89E-03	8.82E-02	RNA guanylyltransferase and 5'-phosphatase pseudogene 1
AC007370.2	13.70	0.419	1.90E-03	8.84E-02	NA
CNGA1	150.51	0.432	1.91E-03	8.88E-02	cyclic nucleotide gated channel subunit alpha 1
VSNL1	434.98	0.362	1.92E-03	8.95E-02	visinin like 1
TMEM51-AS1	782.94	-0.277	1.93E-03	8.97E-02	TMEM51 antisense RNA 1

LCNL1	9.71	0.407	1.94E-03	8.98E-02	lipocalin like 1
ANKRD33	13.37	-0.365	1.99E-03	9.11E-02	ankyrin repeat domain 33
KRT31	42.37	0.289	1.98E-03	9.11E-02	keratin 31
NAALAD2	16.11	0.424	1.98E-03	9.11E-02	N-acetylated alpha-linked acidic dipeptidase 2
PRKRIP1	480.38	-0.157	1.97E-03	9.11E-02	PRKR interacting protein 1
RN7SL546P	7.07	0.404	1.98E-03	9.11E-02	RNA, 7SL, cytoplasmic 546, pseudogene
STK3	1145.69	-0.201	1.97E-03	9.11E-02	serine/threonine kinase 3
ZC3H6	485.98	0.207	1.98E-03	9.11E-02	zinc finger CCCH-type containing 6
AC016629.1	10.28	-0.342	2.01E-03	9.15E-02	NA
POMT2	477.03	-0.146	2.01E-03	9.15E-02	protein O-mannosyltransferase 2
SERPINB9P1	179.32	0.355	2.01E-03	9.15E-02	serpin family B member 9 pseudogene 1
SLCO4A1	1052.39	-0.304	2.00E-03	9.15E-02	solute carrier organic anion transporter family member 4A1
AP003557.1	4.12	0.420	2.02E-03	9.15E-02	NA
BMS1P10	42.15	0.402	2.02E-03	9.15E-02	BMS1 pseudogene 10
PIK3IP1	740.22	0.300	2.02E-03	9.15E-02	phosphoinositide-3-kinase interacting protein 1
TMEM232	28.99	0.343	2.01E-03	9.15E-02	transmembrane protein 232
SCNN1B	786.85	0.324	2.03E-03	9.15E-02	sodium channel epithelial 1 subunit beta
RNASE6	76.35	0.310	2.03E-03	9.16E-02	ribonuclease A family member k6
ZNF385C	68.96	-0.394	2.03E-03	9.16E-02	zinc finger protein 385C
LINC00987	26.64	0.352	2.04E-03	9.17E-02	long intergenic non-protein coding RNA 987
ASB6	678.22	-0.131	2.05E-03	9.20E-02	ankyrin repeat and SOCS box containing 6
CD4	345.24	0.268	2.05E-03	9.20E-02	CD4 molecule
ODF2	1404.71	-0.170	2.07E-03	9.27E-02	outer dense fiber of sperm tails 2
CCL21	29.56	0.309	2.08E-03	9.27E-02	C-C motif chemokine ligand 21
IER5L	286.50	-0.247	2.08E-03	9.27E-02	immediate early response 5 like
TBCEL	404.71	0.288	2.09E-03	9.30E-02	tubulin folding cofactor E like
AL158055.1	5.49	0.399	2.10E-03	9.31E-02	NA
PAX8-AS1	2114.64	-0.278	2.11E-03	9.34E-02	PAX8 antisense RNA 1
ZWINT	807.17	-0.196	2.12E-03	9.41E-02	ZW10 interacting kinetochore protein
RAB30	424.38	0.269	2.13E-03	9.42E-02	RAB30, member RAS oncogene family
SYN2	145.05	-0.408	2.13E-03	9.42E-02	synapsin II
CD72	54.98	0.324	2.14E-03	9.45E-02	CD72 molecule
TRBJ1-6	3.40	0.392	2.15E-03	9.48E-02	T cell receptor beta joining 1-6
KDM5D	1272.97	0.373	2.17E-03	9.52E-02	lysine demethylase 5D
LINC02848	6.90	0.405	2.17E-03	9.52E-02	long intergenic non-protein coding RNA 2848
MCM3AP-AS1	245.12	-0.172	2.17E-03	9.52E-02	MCM3AP antisense RNA 1
PCSK6	955.15	-0.260	2.18E-03	9.53E-02	proprotein convertase subtilisin/kexin type 6
PPP1R10	2526.07	-0.120	2.18E-03	9.53E-02	protein phosphatase 1 regulatory subunit 10
LINC01273	18.21	-0.371	2.19E-03	9.53E-02	long intergenic non-protein coding RNA 1273
AC005838.3	54.24	0.223	2.20E-03	9.55E-02	NA
CSNK1G3	1041.27	0.143	2.19E-03	9.55E-02	casein kinase 1 gamma 3
CDK16	1894.98	-0.188	2.22E-03	9.57E-02	cyclin dependent kinase 16
CNTNAP3B	58.89	0.404	2.21E-03	9.57E-02	contactin associated protein family member 3B
CST4	5.39	-0.314	2.22E-03	9.57E-02	cystatin S
GNG4	103.53	0.391	2.20E-03	9.57E-02	G protein subunit gamma 4
KIF14	904.17	-0.286	2.23E-03	9.57E-02	kinesin family member 14

LEPR	239.29	0.287	2.22E-03	9.57E-02	leptin receptor
TMEM9	705.51	-0.208	2.22E-03	9.57E-02	transmembrane protein 9
LINC02006	29.59	0.327	2.23E-03	9.58E-02	long intergenic non-protein coding RNA 2006
TBC1D16	1995.20	-0.252	2.23E-03	9.58E-02	TBC1 domain family member 16
AC010889.2	14.97	0.404	2.27E-03	9.65E-02	uncharacterized LOC105377224
CASC2	189.23	-0.272	2.26E-03	9.65E-02	cancer susceptibility 2
DOCK8	1546.42	0.279	2.27E-03	9.65E-02	dedicator of cytokinesis 8
GIMAP4	279.70	0.300	2.26E-03	9.65E-02	GTPase, IMAP family member 4
SEC14L4	81.20	-0.403	2.27E-03	9.65E-02	SEC14 like lipid binding 4
LINC01206	56.08	0.358	2.27E-03	9.66E-02	long intergenic non-protein coding RNA 1206
ST6GALNAC6	536.06	0.193	2.29E-03	9.71E-02	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 6
TCF3	1177.87	-0.159	2.30E-03	9.73E-02	transcription factor 3
CHRM3-AS2	37.54	0.390	2.30E-03	9.73E-02	CHRM3 antisense RNA 2
HCFC2	627.55	0.161	2.31E-03	9.75E-02	host cell factor C2
AC010889.1	13.61	0.400	2.32E-03	9.76E-02	NA
MCM7	2489.65	-0.187	2.32E-03	9.76E-02	minichromosome maintenance complex component 7
TRAF3IP3	289.93	0.357	2.32E-03	9.76E-02	TRAF3 interacting protein 3
CCDC69	587.75	0.279	2.33E-03	9.76E-02	coiled-coil domain containing 69
HLA-DQB2	57.62	0.368	2.35E-03	9.84E-02	major histocompatibility complex, class II, DQ beta 2
VTN	49.27	-0.398	2.35E-03	9.84E-02	vitronectin
SEC24AP1	10.14	0.389	2.37E-03	9.89E-02	SEC24 homolog A pseudogene 1
LPAR2	388.19	-0.197	2.38E-03	9.93E-02	lysophosphatidic acid receptor 2
WASIR2	5.76	-0.386	2.38E-03	9.93E-02	WASH and IL9R antisense RNA 2
RNU7-119P	6.54	0.398	2.40E-03	9.99E-02	RNA, U7 small nuclear 119 pseudogene
AC005498.3	9.72	0.389	2.41E-03	1.00E-01	ZNF470 divergent transcript
GPX2	3222.92	-0.363	2.41E-03	1.00E-01	glutathione peroxidase 2