The Immunobiology of B Lymphocytes in Non-Small Cell Lung Cancer

Akshay Jatin Patel

A Thesis Submitted to the University of Birmingham for the degree of DOCTOR of PHILOSOPHY

Institute of Immunology and Immunotherapy

College of Medical and Dental Sciences

The University of Birmingham

Edgbaston

B15 2TT

United Kingdom

February 2022

Abstract

Lung cancer is the second most diagnosed cancer, after breast cancer, worldwide. However, it is still the leading cause of cancer-specific mortality globally, contributing to 18% of all cancer-related deaths. Non-small cell lung cancer (NSCLC) makes up 85% of lung cancers and dependent on the stage, is amenable to a wide of treatments from surgery to systemic therapy. Immune responses within the tumour microenvironment have increasingly been implicated as determining factors in tumour progression and aggressiveness, and the focus has predominated on T-cell biology. The immune response is a complex interplay between the primary tumour and microenvironment, T and B cells. The role of the B cell in tumour survival is unclear but clearly has a function as tumour infiltration is commonly reported.

Through deep phenotyping and multispectral tissue imaging techniques, we identified key differences in the effector and suppressive B cell composition between the tumour and peripheral blood compartments. IL10 positive suppressive B regulatory phenotypes were significantly more abundant in the circulation of patients who recurred post-operatively. Using a broad spectrum immunome array, we employed machine learning techniques and identified an auto-antibody signature in the serum of NSCLC patients that was highly predictive for post-operative recurrence in two independent cohorts.

In addition to the techniques described above, we utilised functional ex vivo B cell assays to interrogate the response to checkpoint blockade in advanced disease patients and how this relates to B cell dynamics. Our findings demonstrated that lack of a suppressive B cell “brake” predisposed patients to high grade immune related adverse events post-treatment. Moreover, the B cells from toxicity patients were not only functionally defective in their ability

to produce IL10 but also displayed a pan cytokine failure affecting pro-inflammatory cytokines thus suggesting B cell exhaustion in these patients. These findings significantly enhanced our understanding of the aetiopathogenesis of auto-immune toxicity secondary to checkpoint blockade with anti PD-1/PDL-1.

In summary, this study aimed to explore the role of B cell biology in NSCLC by employing deep phenotyping and functional assay techniques at the blood and tissue level in both early and advanced stage disease. Our findings are likely to be informative in biomarker development for predicting response to treatment, post-operative relapse and for therapeutic adjuvant polyepitopic vaccine strategies in high-risk patients.

Papers and abstracts submitted during thesis period

Papers

1. **Patel AJ**, Tan TM, Richter AG, Naidu B, Blackburn JM, Middleton GW. A highly predictive autoantibody-based biomarker panel for prognosis in early-stage NSCLC with potential therapeutic implications. *Br J Cancer*. 2022 Feb;126(2):238-246. Doi: 10.1038/s41416-021-01572-x. Epub 2021 Nov 2. 2.
2. **Patel AJ**, Nightingale P, Naidu B, Drayson MT, Middleton GW, Richer AG. Characterising the Impact of Pneumonia on Outcome in Non-Small Cell Lung Cancer: Identifying Preventative Strategies. *J Thorac Dis*. 2020 May;12(5):2236-2246. Doi:10.21037/jtd.2020.04.49.
3. **Patel AJ**, Richter AG, Drayson MT, Middleton GW. The Immunobiology of B Lymphocytes in Non-Small Cell Lung Cancer. *Cancer Immunology Immunotherapy*. 2020 Jan 4. Doi: 10.1007/s00262-019-02461-2.
4. **Patel AJ**, Willsmore Z, Khan N, Richter A, Naidu B, Drayson M, Papa S, Cope A, Karagiannis S, Perucha E, Middleton GW. Regulatory B cell repertoire defects predispose lung cancer patients to immune-related toxicity following checkpoint blockade. Nat Commun. 2022 Jun 7;13(1):3148. doi: 10.1038/s41467-022-30863-x.

Abstracts

1. *Identifying Risk Factors for Respiratory Illness in NSCLC: A Comparative Analysis*. **Patel AJ**, Nightingale P, Naidu B, Drayson MT, Middleton GW, Richter AG. October 2019: European Association of Cardiothoracic Surgery (EACTS) Annual Congress, Lisbon, Portugal [Oral Presentation].
2. The Impact of Respiratory Illness on Morbidity and Mortality in Non-Small Cell Lung Cancer. **Patel AJ**, Nightingale P, Naidu B, Drayson MT, Middleton GW, Richter AG. March 2019: SCTS Annual Meeting & Cardiothoracic Forum, London, UK [Oral Presentation].

Acknowledgements

I am eternally grateful to the Clinical Immunology Service and department of Cancer sciences at the University of Birmingham, and Cancer Research UK for providing me with the support, infrastructure, and funding to undertake research for this thesis. To my academic supervisors, Professor Gary Middleton, and Professor Alex Richter, thank you both for your unyielding support, mentorship, and guidance over the last 3 years. You have helped me forge the foundations of an academic career and I could not have done this work without you. To my surgical supervisor, Mr. Babu Naidu for keeping me grounded and connected to the world of surgery and along with Professor Middleton, being instrumental in securing my fellowship funding.

I am indebted to the countless lung cancer patients who donated their time, their samples, and their insight to help deliver my research. I would like to thank the thoracic research team at University Hospitals Birmingham (Amy Kerr, Hollie Bancroft, Helen Shackleford, Zara Jalal, and Salma Kadiri) and the doctors, nursing staff and phlebotomists at UHB cancer outpatients, without whom I would not have successfully completed patient recruitment.

Thank you to Dr. Naeem Khan not only for his mentorship in complex experimental methods throughout the course of my research but also for his friendship. To Dr. Esperanza Perucha for helping me set up my functional ex vivo assay and welcoming me into her lab as a collaborator for the last two years. To Dr. Sian Faustini for her unrelenting wisdom and time throughout the course of my project.

A very special thanks to the support of my parents, who throughout my entire life have always been there for me without expectation and without having to ask. Thank you for being my strength, always.

Lastly, to my incredible wife, Dr. Geera Patel. Her patience, support and understanding are what has given me the focus to persevere through the last few years. The proverb, “behind every great man is an even greater woman”, I do not think I am great, but my wife most certainly is. This project is just as much her success as it is mine.

Ethics committees

All ethical approvals for this project were sought and secured within the first 9 months of this project. Each objective had separate ethical approvals in place.

Objective 1: Deep B cell phenotyping in the surgical NSCLC population. Ethical remit allowed for the collection of blood and tumour tissue pre-operatively.

REC reference: 17/WM/0272 (MLTC), Sponsor: University of Birmingham.

Objective 2: Deep B cell phenotyping in the advanced disease NSCLC population. Ethical remit allowed for the collection of blood pre- and post-treatment (immunotherapy/chemotherapy).

REC reference: 10/H0501/39, Sponsor: University of Birmingham.

Objective 3: Characterising the proteome in NSCLC. Ethical remit was for the historical CLUB serum sample dataset in NSCLC. The ethical remit allowed for the use of these serum samples in further experimental work pertaining to discovery and translational research in NSCLC.

REC reference: 04/Q2704/34 (CLUB), Sponsor: University of Birmingham.

Objective 4: Determining the Functional in Vivo B cell response in NSCLC. Ethical remit allowed for the collection of blood pre- and post-vaccination with Pneumovax (PPSV23) at serial time points.

REC reference: 19/WM/0182 (FUNCTION), Sponsor: University of Birmingham.

Table of Contents

Abbreviations

Abbreviations	Meaning
AAb	Auto-antibody
AoA	Another other Admission
AUC	Area under Curve
BCR	B cell receptor
Breg	Regulatory B cell
BSA	Bovine Serum Albumin
CAP	Community Acquired Pneumonia
CB	Checkpoint blockade
CCB	Combination Checkpoint blockade
CIS	Clinical Immunology Service
COPD	Chronic Obstructive Pulmonary Disease
CpG	CpG oligodeoxynucleotides
CSR	Class Switch Recombination
CT	Computed Tomography
CTAGs	Cancer Testis Antigens
CTLA-4	Cytotoxic T-lymphocyte-associated protein 4
CyTOF	Cytometry by Time of Flight
DA	Differential abundance
DC(s)	Dendritic cell(s)
DFS	Disease free Survival
DMSO	Dimethyl Sulfoxide
ESMO	European Society of Medical Oncology
FCS	Fetal Calf Serum
FSC	Forward scatter
FT3	Triiodothyronine
GC	Germinal Centre
HR	Hazard Ratio
IASLC	International Association for the Study of Lung Cancer
ICFC	Intracellular Flow Cytometry
IFNγ	Interferon gamma
IgA	Immunoglobulin A
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IL10	Interleukin-10
IL17	Interleukin-17
IL2	Interleukin-2
IL6	Interleukin-6
IrAE	Immune related Adverse Event
MDSC(s)	Myeloid Derived Suppressor cell(s)

MHC	Major Histocompatibility Complex
NSCLC	Non-small cell lung cancer
OS	Overall Survival
PBMC	Peripheral Blood Mononuclear cell
PCA	Principal Component Analysis
PD1	Programmed cell death 1
PD-L1	Programmed cell death ligand 1
PET	Positron Emission Tomography
pFC	Penetrance Fold change
PPSV23	Pneumovax-23
QC	Quality Control
R&D	Research and Development
RCT	Randomised Controlled Trial
REC	Research Ethics Committee
RFE	Recursive Feature Elimination
RFU	Relative Fluorescence Unit
ROC	Receiver operating characteristic
RT	Room Temperature
SCLC	Small cell lung cancer
SHM	Somatic Hypermutation
SLOS	Secondary Lymphoid Organs
Tfh	T follicular helper cell
Tfr	T follicular regulatory cell
TILBs	Tumour infiltrating B Lymphocyte
TILs	Tumour infiltrating Lymphocyte
TLC	Toxicity linked cluster
TLRs	Toll-Like Receptors
TLS	Tertiary Lymphoid Structure
TME	Tumour Microenvironment
TNFα	Tumour Necrosis Factor alpha
Treg	Regulatory T cell
tSNE	t-Stochastic Neighbour Embedding
UHA	Unplanned Hospital Admission
UHB	University Hospitals Birmingham
UMAP	Uniform Manifold Approximation and Projection
UoB	University of Birmingham

List of Figures

List of Tables

Overarching Project Summary

This section will summarise the individual hypotheses and different patient cohorts that were investigated in this project.

Objective 1

Hypothesis 1

B cell immune phenotype is predictive of subsequent post-operative outcome in non-small cell lung cancer

The first objective aimed to better characterise B cell populations in matched resected NSCLC tumour and peripheral blood samples from a cohort of patients with NSCLC that underwent surgical resection. In the field of renal cell carcinoma, distinct immune cell compositions were identified and correlated with progression free survival. A robust human atlas of the immune cell signature in these patients revealed potential biomarkers for disease as well as targets for immunotherapy development (1). Through the use of mass cytometry and multiplex immunofluorescence (VECTRA Polaris) analyses, we were able to perform deep B cell phenotyping in this patient population in paired blood and tumour samples.

Objective 2

Hypothesis 2

B cell changes are predictive of outcome and immune related adverse events (irAE) in NSCLC patients receiving checkpoint blockade therapy +/- chemotherapy

The second objective aimed to determine if B cell phenotype was predictive of outcome and toxicity (immune-related adverse events; irAE) in patients with advanced NSCLC being treated with immune-oncotherapy (IO; checkpoint blockade agents) alone, or in combination with other immunotherapy agents (IO/IO) or chemotherapy (IO/chemo).

Unlocking the phenotypic diversity of the B cell milieu in NSCLC will enable the identification of immune cell patterns that correlate with outcome (negative or positive) which may harbour new avenues for immunotherapy. There is a changing landscape in checkpoint immunotherapy in NSCLC, whether it is monotherapy with a checkpoint inhibitor (PD1 or CTLA4 or both) or in combination with an angiogenesis inhibitor or even chemotherapy, the regime is evolving to the point that combination therapy in patients with advanced disease will become the standard of care over the course of this project.

Treatment with IO has shown promise in the treatment of advanced melanoma (2) however its application is stifled by the incidence of irAEs which lead to the discontinuation of therapy. Whilst these events are thought to be T-cell mediated, there is evidence for B cell implication (3). Moreover, loss of CTLA4 in humans has, in addition to T cell dysregulation, resulted in progressive circulating B cell loss and accumulation of autoreactive B cells (CD21lo) with subsequent deregulated lymphocyte homeostasis (4). It was therefore desirable to know if changes in B cell profile affect can predict for the development of irAE. Early B cell changes have been detected in melanoma patients on IO/IO combination therapy; low levels of circulating B cells and high numbers of CD21-Lo B cells and plasmablasts were found in patients who experienced higher rates of grade 3 or higher irAE six months after treatment

(2). Through the use of mass cytometry and various functional B cell and CD4+ T cell assays, we explored the B cell signatures in the blood of these patients pre- and post-treatment.

Objective 3

Hypothesis 3

The circulating tumour-associated auto-antibody signature in NSCLC is predictive of outcome post-surgical resection

The third objective sought to characterise the proteome in the serum of NSCLC patients who underwent surgical resection. We utilised the pre-operative serum samples from a historical cohort of NSCLC patients with early-stage disease. This cohort was clinically robust with clear stratification between survivors and non-survivors. Survivors were patients who are alive and disease-free at a minimum of 5 years post resection. Non-survivors died from disease recurrence within 12 months post resection. We employed the Immunome™ protein array precision medicine technology developed by Sengenics® to determine the expression of >1600 autoantibodies in the pre-operative serum of these patients.

Chapter 1. Introduction

1.1 Overview

The overview of this report is the investigation of the humoral immune response in the pathophysiology of non-small cell lung cancer, specifically with reference to the role B cells and their soluble antibody products play in tumorigenesis. B cell responses have been implicated in both the anti-cancer and pro-cancer responses in lung cancer, but their definitive role remains unclear. Thus, this report will focus on exploring deep B cell phenotypes in early and advanced stage non-small cell lung cancers, response to treatment with checkpoint blockade therapies and the role of functional in vivo B cell responses. This chapter will discuss the epidemiology and natural history of lung cancer and novel emerging treatments. The immunobiology of B cells in disease and malignancy will be explored, focussing on pathogenesis and potential role in non-small cell lung cancer and functional B cell immunity and infection.

1.2 Burden of Disease

1.2.1 Lung Cancer Epidemiology

Lung cancer is on the rise globally and is the commonest cause of cancer-related deaths, with tobacco smoking posing the single biggest risk factor for malignancy (5). Despite advances in our understanding of lung cancer genetics, the role of the immune system in controlling lung cancer and allied treatment options, the number of year-on-year cases is rising and survival continues to be low. In 2020, GLOBOCAN estimated 2.2 million new cases (11.6% of total cancer cases) and 1.76 million deaths (18.4% of total cancer deaths), which were higher than the 2012 reported rates (1.8 million new cases and 1.6 million deaths) (6). In the UK, there

are 48,000 new lung cancer cases and 35,100 lung cancer-related deaths each year, with very little improvement in overall lung cancer survival in the last 40 years (7).

1.2.2 Natural History and outcome

The survival of patients with lung cancer at 5 years after diagnosis is between 11 and 17% in most countries among those diagnosed during 2010 through 2014, although rates are higher in Japan (33%), Israel (27%), and the Republic of Korea (25%) (6,8). These rates are also dependent on stage, histological and molecular subtypes. There is wide variation in tumour growth rates reflecting the heterogeneity of this disease. Cancers diagnosed at a very early stage during routine medical practice can be rapidly fatal, hence the need for prompt specialist referral and evaluation (9,10). Large scale registry data has shown that in the first 6 year after diagnosis irrespective of stage, lung cancer itself is the predominant cause of death in these patients, largely due to respiratory failure (11). With time, the relative contribution of cardiovascular disease, COPD and infection increased however the absolute contribution of lung cancer remained high, particularly in non-metastatic disease.

1.3 Detection and Screening

1.3.1 Clinical Presentation and Investigation

There are usually no symptoms or signs in the early stages of lung cancer, but these can develop with disease progression and include, a persistent cough, haemoptysis, breathlessness, as well as unexplained fatigue, malaise and weight loss (12). A plain film radiograph of the chest is usually the first test used to detect any abnormalities within the chest. This is usually followed by a formal assessment of the lungs and thoracic cavity with cross-sectional imaging. Computed Tomography (CT) – Positron Emission Tomography (PET)

scan can detect lesions suspicious for malignancy within the chest cavity and determine the degree of metabolic activity within these lesions; metabolically active lesions ("hot") usually correlate with malignancy, but this is not always absolute. A definitive tissue diagnosis is usually performed with a biopsy via bronchoscopy, radiologically under CT guidance or at the time of surgery (12). Information gathered through these tests will enable the disease to be staged according to the recent iteration of the International Association for the Study of Lung Cancer Staging Project (13). This will inform which treatment strategies are employed.

1.3.2 Histological Subtypes

Primary lung cancer is broadly divided into two groups: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). The former accounts for 80-85% of all lung cancers and are grouped together owing to the way in which they behave, progress and respond to treatment. Adenocarcinoma, Squamous cell carcinoma and Large cell carcinoma are all examples of NSCLC (14). SCLC accounts for 15-20% of all lung cancer and tends to be detected at a late stage, which confers poorer outcomes. This group belongs to a broader class of tumour known as neuroendocrine tumour (NET) (14).

1.3.3 Screening

Screening with low dose computed tomography (CT) for high-risk individuals (current and former heavy smokers) can help diagnose cancer early, when successful treatment is more likely. The efficacy of annual low-dose CT screening in reducing lung cancer mortality has been confirmed in several independent, international, randomized controlled clinical trials (15–18). Low-dose CT based screening of high-risk smokers reduced lung cancer-specific mortality; consensus from a recently published position statement by European investigators recommended implementation of screening across Europe (17). Recently, the use of volume

CT screening, with growth-rate assessment as an imaging biomarker for indeterminate tests, resulted in a lung cancer mortality reduction at 10 years of follow-up of 24% in men and 33% in women compared with no screening.

In 2010, the US-based National Lung Cancer Screening Trial (NLST) was stopped 1 year early because a 20% relative reduction in lung cancer mortality had been achieved by low-dose CT compared with chest x-ray after three annual screens and 6 years of follow-up (15). Subsequently, the IASLC has made six recommendations with regard to the implementation of CT screening in future screening initiatives (19).

a) Optimisation of identification of high-risk individuals.
b) Development of radiological guidelines.
c) Development of guidelines for the clinical workup of indeterminate nodules.
d) Development of guidelines for pathology reporting.
e) Definition of criteria for surgical and therapeutic interventions of suspicious nodules identified through lung cancer CT screening programmes.
f) Development of recommendations for the integration of smoking cessation practices into future national lung cancer CT screening programmes.

The pilot UK lung cancer screening (UKLS) RCT compared low dose CT scanning screening to usual care in order to demonstrate the effectiveness of risk prediction modelling for the selection of high risk participants, to evaluate the use of volumetric analysis in nodule management and to determine cost-effectiveness based on modelling of the pilot approach (19). The UKLS trial demonstrated that it is possible to detect lung cancer at an early stage and deliver potentially curative treatment in over 80% of cases on a cost-effective basis. The

long-term NHS plan, by 2028, is to have increased the diagnosis rate of early-stage cancers to 75%. The primary aim of which is to reduce mortality from lung cancer. This must be achieved with minimal physical and psychological harm. The NHS England document *"Target screening for lung cancer with low radiation dose computed tomography – standard protocol"* has outlined an incremental roll out service to deliver targeted lung health checks over the next five years (6).

1.4 Aetiology of Lung Cancer

Figure 1.1 below illustrates the common risk factors and causal factors linked to the development of lung cancer.

Figure 1.1. A schema to illustrate the common risk factors causally linked to increased risk of lung cancer around the world. This figure has been adapted from an educational website (https://ib.bioninja.co.au).

1.4.1 Genetics of Lung Cancer

1.4.1.1 Inheritance

Most cases of lung cancer are not related to an inherited genetic mutation, but instead are due to somatic mutations occurring in certain cells within the lungs or airways. When lung cancer is related to inherited genetic changes, the risk of developing cancer follows an autosomal dominant pattern, which means one copy of the altered gene in each cell is sufficient to increase a person's chance of developing the disease (20).

1.4.1.2 Culprit Genes

The occurrence of genetic mutations in critical genes, those that control cell growth, division, DNA damage repair pathways, often result in unregulated cell proliferation. The most commonly identified types of genetic mutations in lung cancer include missense/nonsense mutations, small insertions or deletion, alternative splicing and chromosomal fusion rearrangements (21). Predominantly implicated genes include eGFR, KRAS, TP53, MET, LKB1, BRAF, PIK3CA, ALK, RET and ROS1 (21). Copy number alternations also play a significant role, both eGFR and KRAS copy number gains have been found to occur more frequently in tumours harbouring activating mutations for these oncogenes (22) which results in enhanced cellular proliferation. Following the discovery of a specific cluster of eGFR mutations in lung adenocarcinomas and the correlation with enhanced clinical response to eGFR tyrosine kinase inhibitors (TKIs) (23,24), this paved the way for the development of the “targeted therapy” paradigm in lung cancer. KRAS gene mutations are generally more prevalent in adenocarcinomas (15-25%) and tend to be found in lung cancers with wild-type eGFR protein (25,26). Contrastingly, over-expression of the MET oncogene is common in NSCLC and correlates with over-expression of eGFR (27).

Somatic mutations in the TP53 gene are one of the most frequent aberrations in human cancers and were some of the earliest described in NSCLC (28). Germline mutations in this gene are the underlying cause of Li-Fraumeni syndrome which predisposes to a wide range of early-onset cancers (29). The p53 protein is a nuclear phosphoprotein that maintains the integrity of the genome by exerting control over the G1-S and G2-M cell cycle checkpoints to prevent the replication of damaged DNA and cell division (30). If the DNA repair pathway fails, p53 triggers cellular apoptosis (30). Aberrations in this protein account for over 60% of human lung cancer cases (28–30).

The LKB1 (STK11) gene is another tumour suppressor gene and germline mutations in this gene are the causative insult leading to the development of Peutz-Jeghers syndrome (21). Inactivating mutations of LKB1 occur in 4% of NSCLC cases, with a preponderance for adenocarcinoma and co-exist with activating KRAS mutations (21,31). BRAF and PIK3CA are examples of commonly over-expressed kinases in NSCLC which are responsible for uncontrolled cellular growth, proliferation, and differentiation. ALK, RET and ROS1 are all examples of receptor tyrosine kinases which form chimeric fusion gene variants through oncogenic chromosomal translocations. Only 3-7% of NSCLC harbour ALK fusions, with the EML4-ALK fusion driver mutant kinase being reported in 2007, the presence of which correlates with eGFR TKI resistance (32,33).

Decades of research into NSCLC has led to the discovery of these culprit genes, which has transformed treatment paradigms but clinical outcomes for these patients' still remains poor, reflecting the unmet need for further insight into the biology of lung cancer and how this tracks with outcome.

1.4.2 Lung Cancer Evolution

The study of tumour evolution has expanded rapidly over the last decade, particularly through the advent of the TRACERx study in NSCLC (34,35). The TRACERx (Tracking Cancer Evolution through therapy) lung study, set up in 2014, aims to integrate new technologies to study cancer evolution in up to 840 lung cancer patients. This ranges from analysing longitudinal changes in tumour DNA, to studying how these changes can be detected in blood and expanding into the characterization of tumour immune and tissue microenvironments. By tracking the clonal and subclonal mutations in tumours, it has been shown that tumours follow distinct evolutionary patterns, either Darwinian selection of subclonal events or an accumulation of mutations following a clonal driver event (34–36). A prospective analysis of 100 NSCLC tumour samples through whole exome sequencing, determined that intratumour heterogeneity of copy-number events is associated with adverse outcome (34). Moreover, the presence of such intratumour heterogeneity can render gene-expression biomarkers vulnerable to sampling bias, whereas a panel of robust clonally expressed genes (present in all cells) can offer greater diagnostic or indeed prognostic utility (37). Early in tumour evolution, there are strong selection pressures to evade host immune surveillance through numerous immune escape methods (38) and those tumours with multiple regions of poor immune cell infiltration associate with worse outcome (39). Work from TRACERx has shed light on the dynamic and ever changing genomic landscape in NSCLC, the therapeutic utility of targeting clonal neoantigens and the use of circulating tumour DNA in the detection of minimal residual disease in the adjuvant setting (35,36). Translating this knowledge into longitudinal prospective clinical trials is the next step. Understanding the deeper role of the anti-tumour and indeed pro-tumour response will help to inform therapeutic strategies which can be incorporated into such clinical trials.

1.4.3 Environmental factors

The aetiology of lung carcinogenesis is multi-factorial. Cellular, genetic and environmental factors are all responsible for neoplastic transformation. There are numerous lifestyle and environmental factors that expose individuals to carcinogens and increase the rate at which somatic mutations occur, contributing to a person's risk of developing lung cancer. Tobacco smoking confers a significant relative risk of 20-25% and an attributable risk of 85-90% for lung cancer (40). Other risk factors include exposure to environmental tobacco smoke, air pollution, and occupational exposure to radon, asbestos, certain metals and chemicals (41). Of these, only smoking prevention and cessation program have been shown to reduce lung cancer risk (40). Smokers who quit for more than 15 years have an 80-90% reduction in their risk for lung cancer compared with persons who continue to smoke (41). In addition to those mentioned, data from a 2008 risk-scoring model determined that a prior diagnosis of pneumonia (OR 1.83) and a family history of lung cancer (OR 2.02 for early-onset (<60 years) and OR 1.18 for late-onset (>60 years)) were significant independent risk factors for developing lung cancer (42). The rationale of employing so-called “chemopreventative agents” to limit DNA damage and hence malignant progression has demonstrated success in breast (tamoxifen) and colon (celecoxib) (41). However the use of chemopreventative agents such as Vitamin A, β-Carotene, N-acetylcysteine and selenium have not shown efficacy in preventing lung cancer in high-risk individuals in large-scale clinical trials (40,41).

1.5 Treatment Options

The treatment for lung cancer tends to involve multiple modalities managed by a team of healthcare professionals. The most common treatment options include surgical resection,

chemotherapy, radiotherapy, and immunotherapy. The type or combination of treatment modality employed depends on the type of lung cancer (NSCLC or SCLC), the size and position of the cancer within the thoracic cavity, the stage of cancer and overall health status of the patient. Figure 1.2 below summarises the current available treatment pathways in lung cancer.

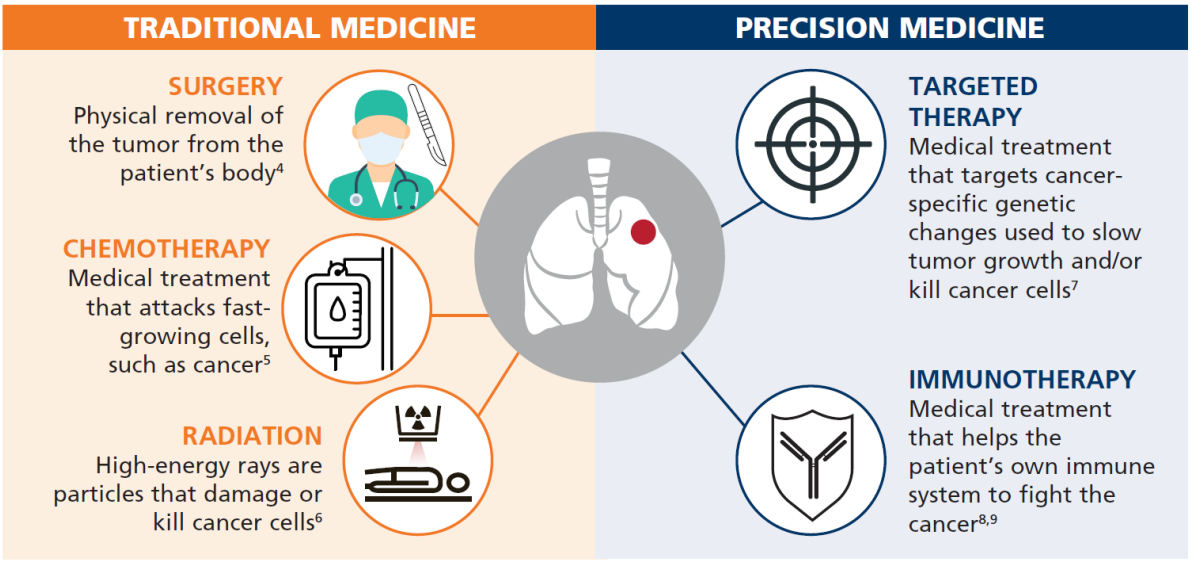


Figure 1.2. A schematic to illustrate currently available treatment options in lung cancer, both early and advanced stages. The modalities have been stratified according to traditional options (left panel) which include surgical resection and chemoradiotherapy and more novel precision medicine options (right panel) which include targeted therapy and immunotherapy. Targeted therapy involves the use of specific enzyme inhibitor molecules (TKIs) to target specific mutations in lung cancer. Immunotherapy uses the body's own immune system to target the cancer. Figure adapted from Conquer Magazine, April 2020, Volume 6, issue 2, TLG1784-2.

1.5.1 Surgery

Treatment of NSCLC is stage-specific and generally early-stage cancers (I/II) undergo complete surgical resection when not contraindicated. Lobectomy (complete removal of a single lobe) with systematic lymph node dissection is considered the gold standard operation (43). Registry data from Norway showed that for operated patients, 5-year relative survival for

stage I and II disease was 58.4% and 28.4% respectively (43). Sublobar resection (segmentectomy or wedge resection) however has the advantage of preserving lung function as well as offering lower morbidity and mortality post-operatively (44). Historical data from the pre-PET scan era demonstrated that sublobar resections associated with a significantly higher local recurrence rate compared with lobectomy (45). In the current era, the question of whether segmentectomy is superior to lobectomy for early-stage NSCLC is being addressed by the Japan Clinical Oncology Group 0802/WJOG 4607L randomised controlled trial.

1.5.1.1 Neoadjuvant and Adjuvant Therapy

Chemotherapy can be employed pre-operatively (neoadjuvant) to downstage potentially resectable tumours and clear micro-metastatic disease or post-operatively (adjuvant) to reduce the chance of distant metastasis following a potentially curative resection (46). Meta-analytical data from 13 RCTs has shown that neoadjuvant chemotherapy improves the overall survival of operable NSCLC patients including those with stage III disease (HR 0.84, $p<0.05$) (47). However it is important to note that this analysis did not include data from a large RCT from the same time involving 624 patients where no difference was found between groups in overall or disease-free survival (48). Pooled trial from the adjuvant setting has demonstrated an absolute survival benefit of 5.4% at 5 years, this is pronounced in tumours exceeding 4cm, stage II disease and after complete resection in stage IIIa disease (46,49). The use of TKIs and immunotherapeutic agents in the neoadjuvant and adjuvant setting is currently being explored through large scale phase III RCTs.

1.5.2 Chemo-radiation

Pre-immunotherapy, the standard treatment regimen for advanced metastatic NSCLC was to treat with a combination of a platinum-doublet agent in combination with Gemcitabine,

Vinorelbine or taxanes (Paclitaxel or Docetaxel) (46). Pemetrexed, a multi-targeted folate in combination with Cisplatin showed significant survival benefit (HR 0.81) in patients with non-squamous histology and became the standard of care in these patients (46). Moreover, for unresectable stage III NSCLC, concurrent chemo-radiation was the preferred treatment however survival remains poor at 15% at 5 years (46). Four year data from the PACIFIC trial (50,51) showed that stage III patients treated with Durvalumab post chemo-radiation had significantly better overall (HR 0.71) and disease free (HR 0.55) survival compared to placebo.

1.5.3 Immunotherapy and Checkpoint Blockade

The advent of immunotherapy has rapidly transformed the landscape of anti-cancer treatment not only in lung cancer but across a range of other malignancies such as melanoma and renal cell carcinoma. The success of single agent and combination checkpoint blockade, which enables antigen-experienced micro-environmental T cells, has transformed the treatment paradigm of advanced NSCLC both as a front-line option and in the platinum-refractory setting. Despite the successful application of immunotherapy, a large proportion of patients with advanced stages cancers do not experience significant improvements in overall survival, this is largely reflective of the intricate and highly regulated nature of the host immune system and its interplay with the ever-changing and the ever-adapting tumour microenvironment (TME) (52).

1.5.3.1 Manipulating the Adaptive Immune System

The adaptive immune system was characterised by the discovery of two major subtypes of lymphocytes; the B lymphocytes and T lymphocytes (53). B lymphocytes respond to antigens in circulation by secreting protective antibodies, where T cells recognise peptide antigen loaded onto cell surface MHC molecules (antigen presentation). Two broad classes of T cell

exist, CD4+, which detect MHC class II-loaded antigen and coordinate the adaptive immune system by secreting chemotactic and pro-inflammatory cytokines (54,55) and CD8+ cells which detect MHC class I loaded antigen and carry out direct cytotoxic reactions against infected or neoplastic cells (54). A subclass of CD4 T cells known as Tregs or regulatory T cells (CD4+ CD25+) suppress the immune response (55,56). The role of T cells in modern cancer immunotherapies has been extensively explored, and as we shall discuss later, in recent years, B cells have been increasingly explored with great interest not only in the context of immunotherapy but in the context of solid cancers also.

1.5.3.2 Concept of "Immune Checkpoint Blockade"

The rationale behind blocking key immune "checkpoints" was to reactivate and unleash powerful T cell responses against tumour cells. These checkpoint molecules exist as a fail-safe within the body to regulate the immune response and protect against hyperactive T cell responses (57). The programmed cell death 1 (PD1) and cytotoxic T lymphocyte antigen 4 (CTLA4) molecules are critical checkpoints which regulate cell function throughout distinct sites in the body (58). Targeting these molecules through the use of monoclonal antibodies effectively releases this brake on T cell function, enabling T cell activity against recalcitrant cancers.

1.5.3.3 CTLA-4 Axis

During T cell activation, the T cell receptor (TCR) on the surface of naïve T cells interacts with its cognate peptide-antigen loaded on MHC molecules on antigen-presenting cells (APC). Concurrently, the co-stimulatory receptor, CD28 (T cell surface) binds to its complementary B7 ligands (B7.1, B7.2) on the surface of APCs. This process results in T cell activation (57). CTLA-4 possesses sequence similarity with CD28 in the extracellular binding domain and

competes with CD28 with greater affinity and avidity for the B7 ligands on APCs. CTLA4 (inhibitory) is stored within vesicles in naïve T cells and constitutively expressed on the surface of immunosuppressive Tregs (57). Following T cell activation in the lymphoid tissues, CTLA4 is displayed on the cell surface and binds to B7.1/B7.2 on APC, preventing their binding to CD28 and rendering activated T cells anergic (57). Treg-expressed CTLA4 leads to trans-endocytosis of B7 ligands and thus interferes with the co-stimulatory ability of CD28 (57).

1.5.3.4 PD1/PD-L1 Axis

Early after activation within the peripheral tissues, PD1 transcription is upregulated on the T cell surface which continues resulting in significant surface expression of PD1. This binds to its complementary ligands PD-L1 and PD-L2 on the surface of other immune cells or tumour cells resulting in T cell exhaustion (57). PD-L1 (also known as B7-H1) and PDL2 (also known as B7-DC), are present constitutively on APCs and can be induced in non-haematopoietic tissues by pro-inflammatory cytokines (57). Cells which express PD-L1 include B cells, macrophages, dendritic cells and epithelial cells under the right inflammatory conditions. Blocking of the PD1 axis through the administration of an anti-PD1 (or anti-PDL1 or anti-PDL2) antibody prevents this inhibitory interaction and unleashes anti-tumoural T lymphocyte activity by promoting increased T cell activation and proliferation, by enhancing their effector functions and by supporting the formation of memory cells. Consequently, more T cells bind to tumour antigens presented on tumour cells by MHC molecules via their T cell receptors (TCRs). This ultimately leads to the release of cytolytic mediators, such as perforin and granzyme, causing enhanced tumour killing.

1.5.3.5 Immune-related Adverse Events (irAE)

Blocking immune checkpoints overrides the normal mechanism of immune tolerance to self-tissues, as such the use of these therapies is associated with autoimmune toxicity (59). The incidence of this relatively common (15-90%) (60,61) and any organ be targeted, with the most frequently occurring irAEs affecting the skin, colon, endocrine organs, liver, lungs and kidneys (62). Others are very infrequent but may be very serious, even lethal, such as neurological disorders and myocarditis. Toxicities affecting the gastrointestinal tract and brain are more common with anti-CTLA4 therapy, whereas patients treated with PD1/PD-L1 axis targeted therapies are at higher risk of hypothyroidism, hepatotoxicity and pneumonitis (63). Clinical management of these irAE is the same and the toxicities are graded according to the 2009 National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) severity scale (63). These range from mild (grade I) which are not typically treated, to severe and life-threatening toxicity (grade III/IV respectively) which require complete discontinuation of treatment and use of glucocorticoids, immunosuppressants and life-saving measures (62,63). This poses a huge healthcare burden and is associated with considerable detriment to patient quality of life and overall survival.

1.5.3.6 Clinical Trial Data

Numerous monoclonal antibodies targeting the PD1/PD-L1 axis have reached the clinic following large scale phase III RCTs. Antibodies targeting PD1 (Nivolumab, Pembrolizumab) and the ligand PD-L1 (Atezolizumab, Durvalumab, Avelumab) have shown durables responses in up to 20% of patients with advanced NSCLC (64–67). Five year follow-up data has recently emerged from the early immunotherapy trials in NSCLC; KEYNOTE 024 showed a 5-year overall survival rate that was almost doubled in metastatic NSCLC patients with PD-L1 tumour

proportion score (TPS) of >50% compared with those receiving standard chemotherapy (31.9% versus 16.3%) (66). Three year data from KEYNOTE 189 showed that combination chemo-immunotherapy improved overall and disease free survival compared to chemotherapy alone in metastatic non-squamous NSCLC irrespective of PD-L1 expression (68), and a similar relationship was observed at median follow-up (14.3 months) in metastatic squamous cell cancers from KEYNOTE 407 (69,70). In two phase III trials (CheckMate 017 and CheckMate 057), Nivolumab showed an improvement in overall survival and favourable safety versus docetaxel in patients with previously treated, advanced squamous and non-squamous NSCLC, respectively (67,71). Five-year pooled results showed that patients derived long-term survival benefit and durable responses with Nivolumab versus chemotherapy, regardless of histology and PD-L1 expression (71). The clinical trial space in immunotherapy for lung cancer is vast and evolving at a rapid rate.

1.5.4 Targeted Therapies

Use of agents to target specific oncogenic driver mutations (discussed in section 1.4.2) in NSCLC has changed the treatment profile in the past few years. High response rates have been achieved in eGFR mutant cancers, through the use of TKIs and monoclonal antibodies targeting the eGFR mutant receptor (46). However, resistance to therapy in these cancers develops after 6-12 months; the T790M missense mutation at exon 20 (up to 60% of patients) is a common mechanism of resistance, requiring switching to Osimertinib (third generation agent) (72). Mutations in PIK3CA, amplification of HER-2 and MET are other secondary resistance mechanisms in these patients (73,74).

The small subset of cancers with ALK and ROS1 translocations (1-5%) (75) are targeted with Crizotinib, a TKI which demonstrated a higher response rate (65% versus 20%) and disease

free survival compared to standard therapy in ALK positive tumours that progressed on first-line chemotherapy (76). Alectinib and Lorlatinib are new generation ALK inhibitors with the latter demonstrating activity against all known resistance mutations in ALK translocation positive cancers (77). In ROS1 positive cancers, Crizotinib is the preferred first-line therapy (78) with newer agents such as Ceritinib and Cabozantinib under investigation for those who progress on first-line therapy (79). Other targetable mutations include BRAF (2% of NSCLC), half of whom have the V600E mutation which can be targeted by the TKI, Vemurafenib (80). Neurotrophic Receptor Tyrosine Kinase (NTRK) gene fusion positive cancers that have progressed on chemo/immunotherapy can be targeted with Larotrectinib, an oral tropomyosin receptor kinase inhibitor (81). Although KRAS mutations are the most prevalent in NSCLC, there are no suitable agents that have undergone large scale phase III trials as yet (82).

1.6 Immunobiology of Lung Cancer

Immune responses within the tumour microenvironment (TME) are important determinants of tumour behaviour, progression and aggressiveness and are often very different to the systemic immune responses in terms of inflammatory cell infiltrate and cytokine milieu (83). Flow cytometric analyses have identified thirteen predominant immune cell phenotypes in NSCLC, with the landscape largely dominated by T cells (47% of all CD45+ cells) with a close to even split between CD4+ (26%) and CD8+ (22%) subtypes (84). B lymphocytes, macrophages and natural killer (NK) cells comprised 16%, 4.7% and 4.5% respectively (84). Three types of dendritic cell (DC) were also identified, plasmacytoid DCs, CD1c+ DCs and CD141+ DCs, collectively accounting for 2.1% of all tumour infiltrating immune cells (84).

Immune cells have been studied extensively in an attempt to understand their role in prognosticating disease. The interaction between immune cells and the tumour in the TME is influenced by the type of immune cells, their density and location (tumour stroma or nest) (85). In NSCLC, cytotoxic CD8+ T cells are generally thought to be protective in lung cancer but the literature is conflicted and owing to heterogeneity between the studies in terms of stage, histology of NSCLC and tumour regions studied (nest or stroma), this makes it challenging to further determine their role (83). Stromal CD4+, stromal CD20+ and co-localisation of stromal CD4+ and CD8+ lymphocytes have all shown an association with improved survival in NSCLC (83,86). Dieu-Nosjean et al showed that the presence of mature Lamp+ DCs within tertiary lymphoid structures (TLS) correlated with higher T and B lymphocyte infiltration and better overall survival (87). The link between tumour-associated macrophages (TAMs) and outcome is also dependent on the location of the immune cell infiltrate (88); intra-tumoural macrophages (M1) associate with better survival whereas stromal macrophages (M2) and their associated cytokines, IL-8, IL-10 and TREM-1 associate with poor survival (83,88–90). FoxP3+ Tregs also associate with poor survival (91) and attenuation of their activity *in vivo* decreased overall tumour burden (92).

Ultimately the TME is influenced by these cellular interactions between immune and tumour cells and their cytokines. This alters the balance from a pro-inflammatory to an anti-inflammatory TME and it is this that determines the natural progression of the disease.

1.7 B Lymphocytes in Health and Disease

The role of T cell-mediated immune responses in solid tumours is well established and has become the driving force for major therapeutic advances, specifically with the advent of

immune checkpoint inhibitors (93). As a consequence, our understanding of the biology of these T cells has expanded exponentially. In contrast, our knowledge of B cell biology in cancer is less well developed. Tumour-infiltrating B cells have been observed in all stages of lung cancer development, with their presence differing according to stage and histological subtype. Given that they play a role in both humoral and cellular immunity B cell parameters may be important in determining both responsiveness to and toxicity of checkpoint blockade. Manipulation of B cell biology might offer significant immune-therapeutic opportunities (94). Thus, understanding B cell biology in NSCLC is of fundamental importance in informing potential novel multi-faceted approaches to favourably reset the immune contexture of the cancer microenvironment.

1.7.1 Normal Development and Ontogeny

In the bone marrow, haematopoietic stem cells serve as the common lymphoid progenitor and continually give rise to B cells throughout life (95) (Figure 1.3). B cell development and differentiation are subsequently regulated through the B cell receptor (BCR). The pro-B cell is BCR naïve; the transition from the pro-B cell phase to the pre-B cell phase is dependent on the generation of a functional immunoglobulin μ heavy chain by rearranging the immunoglobulin heavy chain variable (V), diversity (D) and joining (J) segments and the VDJ locus is spliced to the μ constant region (Cμ) gene. The pre-B cell receptor is composed of the heavy chain paired with the surrogate light chain (95). These pre-B cells then undergo light chain V and J segment rearrangement with either κ or λ constant regions. This complex results in the formation of immunoglobulin M (IgM) expressed on immature B cells which are still resident within the bone marrow. Alternative splicing of the heavy chain VDJ locus to the δ constant region (Cδ) gene generates IgD, which is co-expressed with IgM on mature B cells in the periphery (95). Subsequent development takes place in B cell follicles within secondary

lymphoid organs (SLOS), where germinal centres (GC) develop in response to antigen encounter. Mature B cells undergo a process known as somatic hypermutation (SHM) where the enzyme, Activation Induced Cytidine Deaminase (AICD) introduces mutations in the heavy and light chain variable regions in order to further diversify the immunological repertoire. AICD also mediates immunological class switch recombination (CSR), replacing the μ constant region for another heavy chain region in order to generate IgA, IgE or IgG. The cells with the highest affinity B cell receptors (BCR) are positively selected (this requires signalling through the BCR for the cell to survive, negative selection occurs when the B cell antigen receptor binds self-antigen and therefore undergoes cell death) and undergo clonal expansion in the germinal centre. These class-switched B cells can then give rise to long-lived plasma cells or memory B cells. Mature B cells leaving the bone marrow that do not undergo SHM or CSR in the GC reside within lymph node follicles or splenic sinusoids (follicular B cells) or traffic to the marginal zone (marginal zone B cells); this is an area located at the interface of the circulation and lymphoid tissue of SLOS (95). These cells often express polyreactive BCRs and respond to conserved microbial antigens. Germinal centres, which are normally found in SLOS such as the spleen and lymph nodes, can also occur as tertiary lymphoid structures (TLS) in tumours; this has been reported in colon, breast and NSCLC (5–7). The de novo formation of ectopic lymphoid tissue can occur at the site of inflammation in potentially any organ system (98). There is evidence that the adaptive immune system can be initiated independent of SLOS in these TLS in NSCLC (87).

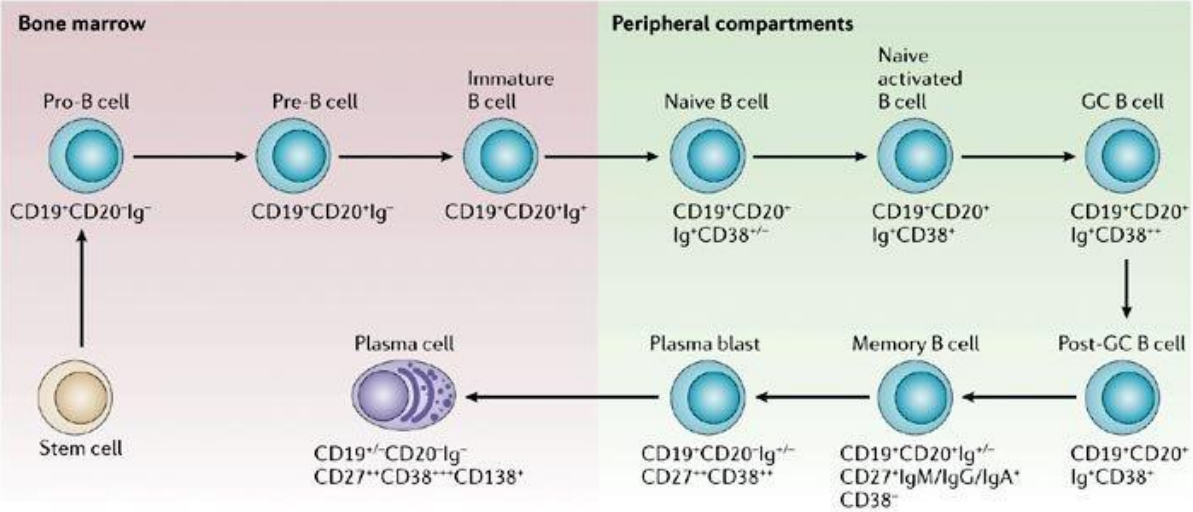


Figure 1.3. This flowchart illustrates the development stages in B cell maturation from haematopoietic stem cells in the bone marrow to terminally differentiated plasma cells which hone back to the bone marrow. The canonical B cell lineage markers are shown at each stage of maturation, with CD19 and CD20 being the classic B cell lineage markers. This figure has been taken from Edwards JCW et al, *Nat Rev Immunol* 2006; PMID 16622478.

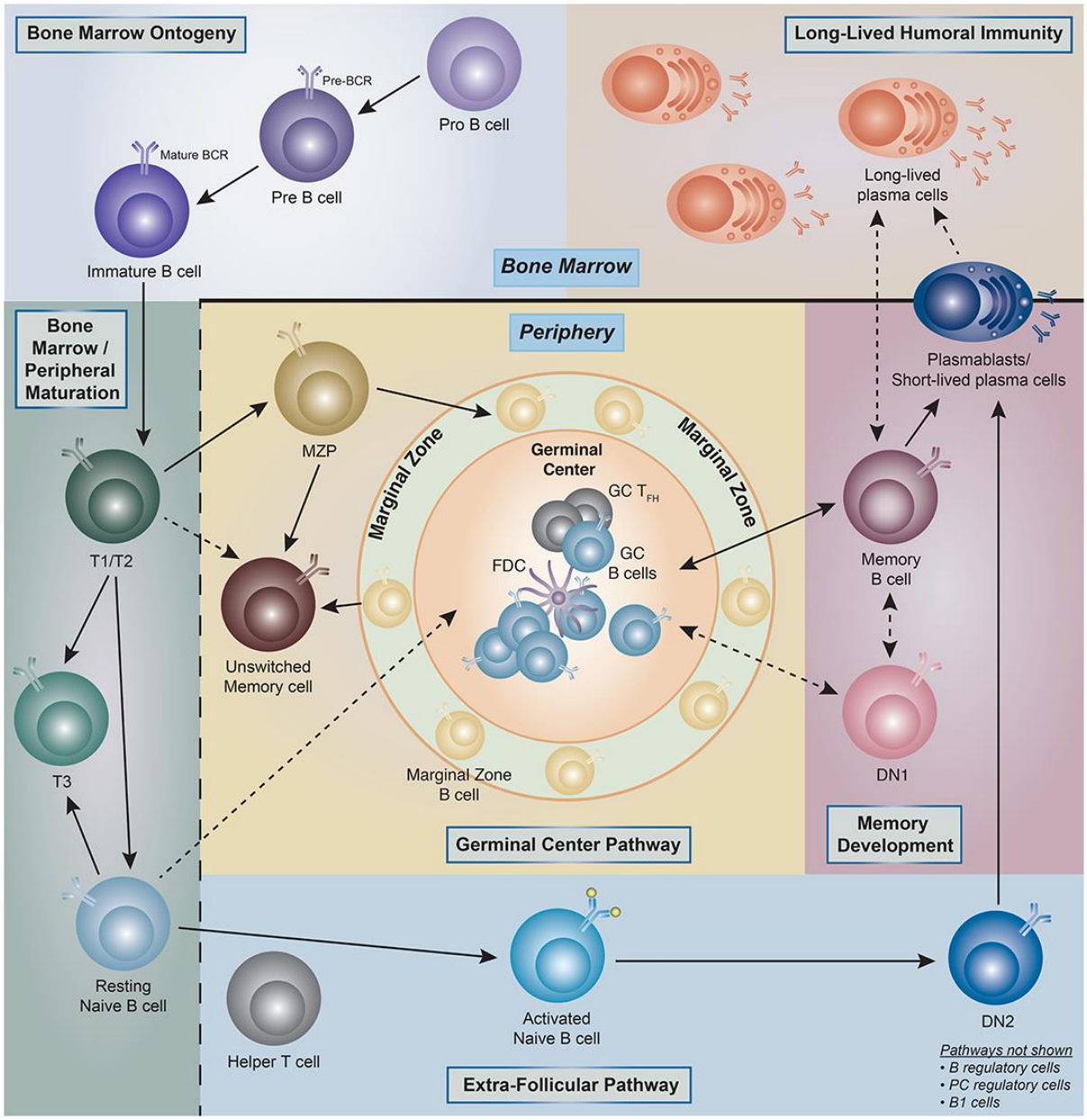


Figure 1.4. This schematic illustrates the more complex interplay in B cell ontogeny highlighting different biological compartments and in particular the extra-follicular pathway of B cell maturation towards the plasma cell phenotype. Memory B cells can differentiate through both germinal centre and extra-follicular pathways and traffic from the periphery to the bone marrow and vice-versa. This figure has been taken from Sanz I et al, *Front Immunol* 2019; PMID 31681331.

1.7.2 B cell Function

1.7.2.1 Antigen-Presentation

T cell immune responses can be activated by B cells. Antigen-specific interactions require antigen internalisation via the BCR with subsequent presentation to T cells in an MHC-restricted manner (99). Depleting host B cells from normal adult mice results in sub-optimal CD4+ T cell activation during immune responses to low-dose foreign antigen (100). Activated B cells derived from peripheral donor blood from healthy controls present antigens to CD4+ and CD8+ T cells (12). B cells and dendritic cells further provide a co-stimulatory signal which is critical to the expansion of an effector T cell population, namely the cross-linking of CD40 with CD154 (CD40 Ligand) on CD4+ T cells (99) which in turn induces the expression and thence stabilisation of CD86 (B7.2) and CD80 (B7.1) on the B cell surface (second co-stimulatory signal).

1.7.2.2 Antibody Production

B cells exist as long-lived plasma cells to produce antigen-specific antibody. The functional BCR can recognise an array of foreign antigens in the circulation and lymphoid organs, which then triggers antigen-specific antibody responses. Following antigen binding, BCR stimulation results in a signalling cascade mediated by membrane bound protein tyrosine kinases, Spleen Tyrosine Kinase (SYK), Bruton Tyrosine Kinase (BTK) and PI3K, in particular PI3Kδ which is a p110 isoform with high lymphocyte specificity (95). This in turn results in B cell differentiation into plasma and memory cells. The plasma cells will traffic to the bone marrow or reside in SLOS where they will continue to secrete antibody. BCR signalling controls this process from early B cell precursor development to terminal plasma cell differentiation, long-lived IgG plasma cells are devoid of the BCR (102).

1.7.2.3 Immunosuppression

B cells play a vital role in the development of the immune system but are also responsible for immune homeostasis. Their immunosuppressive role has been illustrated by their ability to function as regulatory cells or Bregs whereby they influence T cell differentiation and thus T cell mediated inflammatory responses through IL-10 production (103). B regulatory cells are associated with limiting excessive inflammation (104) and mice lacking IL-10 producing Bregs develop chronic inflammation (105). IL-10 producing Bregs induce the Treg phenotype by skewing T cell differentiation in mice (106) and humans (107). Other cytokines that have been implicated in B-cell specific immunosuppression are TGF-beta and IL-35. TGF-beta when produced by B cells can induce CD4+ T cell apoptosis (108) and CD8+ T cell anergy (109). Chimaeric mice lacking either p35 or EBi3 (IL-35 subunits) in B cells develop accentuated autoimmune responses and have greater protection against Salmonella induced sepsis (110).

IL-10 has been useful for Breg characterisation in both mice and humans however this can be up or down-regulated during immune activation and is not a stable inducible trait (104,111). A lineage-specific marker for Bregs equivalent to Foxp3 has not been identified. B cells are polarised to Bregs in response to microenvironmental cues: Bregs have been derived by treating human derived peripheral B cells with tumour-conditioned media; these Bregs were able to suppress the activity of human T cells *in vitro* and they exhibited low surface expression of CD20 unlike healthy control human B cells (112,113). These tumour evoked Bregs did not utilise IL-10 dependent suppression but instead primarily functioned to promote the differentiation of Tregs (CD25+) via the TGFβ signalling axis (112).

1.7.3 Role of B cells in Malignancy

1.7.3.1 Pre-Clinical Studies of the Role of B cells in Cancer

There appears to be a significant difference in the role of B cells in animal models which represent different stages in the development of cancer. In murine models of pre-malignancy, B cells appear to drive inflammation which in turn induces pre-malignancy.

1.7.3.1.1 Pre-Malignancy Models

Early evidence from the K14-HPV16 model (these RAG-1 knockout (KO) mice in which carcinogenesis is initiated by HPV lack T and B cells) has shown that the lack of an adaptive immune response results in failure to initiate leucocyte infiltration during pre-malignancy (114) and this halts progression towards carcinogenesis. Adoptive transfer of B lymphocytes or serum from HPV16 mice into HPV16/RAG1−/− mice resulted in significant infiltration of CD45+ leucocytes, macrophages and granulocytes in the dermal stroma as well as detectably higher levels of serum Ig which then enhanced malignant progression. In this model of squamous carcinogenesis, B cells were shown to play a role in activating Fcγ receptors (FcγRs) on resident and recruited myeloid cells likely via circulating immune complexes (CIC) detected in the serum of HPV16 mice and formed from IgG bound to cognate HPV16 antigens. These CICs were localised to both the epidermal and dermal components of neoplastic skin. The recruitment of these chronically activated leucocytes was dependent on FcγR expression (115), as shown by the reduced level of inflammatory cell infiltration in neoplastic skin in FcγR-/- KO mice. Subsequently it was shown in K14-HPV16 mice that premalignant dysplasia could be prevented through B cell depletion (116). Administration of anti-CD20 monoclonal antibodies depleted B cells in peripheral blood and SLOS with resultant reduced levels of

circulating IgG, immune complex deposition and trafficking of myeloid cells (CD11+), CD45+ leucocytes, mast cells and GR1+ cells to neoplastic skin (116).

In a DMBA/TPA murine model of skin carcinogenesis adoptive transfer of B cells from DMBA/TPA wild type (WT) mice into TNF-α knockout mice significantly increased papilloma development ($p<0.05$), an effect not seen when B cells from the TNF-α knockout mice were transferred to RAG2-/- mice (117). TNF-α is a potent inflammatory cytokine and tumour promotor, Bregs are known producers of TNF-α. Selective TNF-α deletion in CD19+-Cre B cells significantly reduced papilloma development compared to B cells from WT mice ($p<0.002$). TNF-α knockout mice showed increased levels of IFN- γ and CD8+ T cell skin infiltration, but also a significant reduction in the number of splenic CD19+ CD21hi, IL-10 producing Bregs compared with WT mice ($p<0.01$). Further experimental data showed that TNF-α blockade of LPS-induced B cell activation significantly reduced IL-10 production with no difference in IL-2, -4, -5, -12 or IFN-γ. The results of this study identify Bregs as contributory to squamous carcinogenesis, with their activity likely regulated by TNF-alpha with Bregs themselves acting as a cellular source of TNF-α.

The impact of B cells in established cancer appears to be very different to their role in pre-malignant disease.

1.7.3.1.2 Established Cancer Models

B cell depletion using anti-CD20 in a B16 melanoma murine model, resulted in increased tumour burden and the development of pulmonary metastasis (118). CpG (TLR 9 ligand) primed B cells caused tumour regressions in a B cell deficient melanoma model (119). CpG

treatment of B cells induced higher expression of MHC class I and II as well as CD20. It was also noted that in the mice that underwent adoptive B cell transfer, the lung tumours expressed significantly lower levels of BCL-2 (an anti-apoptotic protein) and increased levels of TRAIL (TNF-related-apoptosis-inducing-ligand, a pro-apoptotic protein), TRAIL expression is highest in germinal centre B cells and thus the upregulation in these murine lung tumours is likely B cell driven (119).

In a 4T1 breast cancer model treatment with an anti-CD20 antibody resulted in massive cancer growth and metastasis. Eradication of $CD20^{hi}$B cells, enriched for a select population of $CD20^{LO}$ Bregs which escaped CD20-directed depletion and thus significantly suppressed CD4+ and CD8+ T cell activity thus abrogating anti-tumour responses (113). Targeted delivery of CpG-ODN to CXCR5 expressing cells, reversed the phenotype of these tumour evoked Bregs (which upregulated CD20) and restored effector B cell responses. This was demonstrated by a complete abrogation of tumour metastasis in anti-CD20 treated 4T1.2 cancer bearing mice after adoptive transfer of CpG treated B cells (120). Using the same murine model, Tao and colleagues demonstrated that IL-10 inhibition significantly augmented the therapeutic efficacy of adoptive B cell transfer, as demonstrated by increased trafficking of CD8+ T cells into the tumour microenvironment as well as *in vitro* antigen-specific B cell dependent, FasL-mediated tumour cell killing (121). Activated B cells from this model produce IgG and mediate complement-dependent tumour cell lysis *in vitro* (122). Finally, the use of intra-tumoural TLR9 immune stimulation in combination with PD-1 blockade has demonstrated clinical efficacy in advanced melanoma, with increased B cell infiltration noted on post-treatment tumour biopsies (123). Phase I clinical trial data from patients with metastatic solid tumours showed a non-significant increase in TLR9-expressing naïve B cells during therapy (124).

1.7.3.2 Clinical Studies of the Role of B cells in Cancer

Across all the studies in solid human cancers, assessing the role of B cells in outcome, CD20+ tumour infiltrating lymphocyte characterisation using IHC is the most commonly reported parameter (54 separate cohorts encompassing 15 types of cancer) (125). As a marker of naïve and memory B cells, this method attempts to capture B cells across most stages of differentiation. The prognostic effect of CD20+ TIL was positive in 27 (50%), neutral in 22 (40.7%), and negative in 5 (9.3%) studies (125). This again is reflective of the heterogeneity between studies in terms of clinico-pathological factors, experimental technique, type of tissue section, counting strategies and methods of statistical analysis. On balance however, B cells associate with a positive outcome in the large majority of solid cancers, including breast (126), colorectal (127), gastric (128), ovarian (129), and hepatocellular (130) carcinomas, and cutaneous melanoma (118,131,132).

1.8 Tumour-Infiltrating B Lymphocytes in Human Lung Cancer

Numerous studies have identified TIL-Bs in tissue and examined their associated prognostic significance [See Patel AJ, 2020 PMID 31901949]. These studies largely focus on NSCLC which is where the correlation between TIL-Bs and disease-specific outcome has been shown to be strongest by comparison with other forms of Lung cancer. However, groups have investigated the immune milieu in large cell carcinoma and Small Cell Lung Carcinoma (SCLC). In the former, Eerola et al have demonstrated that TIL-Bs do correlate with better overall survival (133) whereas in SCLC, B cell infiltration is significantly reduced compared with CD8+ T cell and Macrophage infiltration and TIL-B number has not been shown to associate with survival ($p=0.634$ by log rank) (134).

Most studies have used a combination of immunofluorescence or immunohistochemistry in order to identify these cells in tissue. An anti-BCL-6 antibody and an anti-CD21 antibody were utilised in order to identify Germinal Centre B cells and mature B cells respectively. Immunohistochemistry was used to determine B cell density, location and phenotype within NSCLC tissue (97). Studies have utilised PCR-microarray and mRNA sequencing techniques in order to identify primarily humoral immunity related gene signatures in NSCLC specimens (135,136). CD20+ B cell infiltration has been shown to be positively prognostic in NSCLC by a number of different groups (38–43). Disease-free and overall survival was significantly higher in non-smokers with non-squamous NSCLC (138). Significantly improved survival has also been shown in large cell carcinomas with higher degrees of CD20+ B cell infiltration (133). Associations have been made between TLS (GCs forming as ectopic foci of follicular B cells and clusters of mature DC-Lamp^{+ve} Dendritic cells and T cells in cancer tissue, in response to antigen stimulation) in NSCLC and improved long-term survival. The presence of both types of antigen-presenting cells and mature dendritic cells in these TLS strongly predicts the outcome of patients (87,137). A low density of both follicular B cells and mature dendritic cells allows the identification of patients at high risk of poor survival. A higher prevalence of intra-tumoural GC formation was found in NSCLC stage I tumours compared with higher stage (II-IV) tumours ($p<0.02$) (97). In a recent study (142) the expression of a tumour-induced plasmablast-like B cell signature (TIPB) was significantly correlated with the expression of CD8a signatures and the density of CD8+ cells. High expression of the TIPB signature was correlated with overall survival in the melanoma TCGA dataset. Importantly, a cohort of melanoma patients treated with anti-CD20 antibodies, showed significant on-treatment down-regulation of the TIPB signature: the signature was highly correlated with tumour inflammatory score, interferon gamma and T cell effector signatures all of which significantly

decreased on anti-CD20 therapy. There was a marked depletion in in both CD4+ and CD8+ cell density at the invasive tumour-stroma margin and a reduction in the TLS area, an effect which was prolonged. In support of this data suggesting the importance of B cells in a successful anti-cancer immune response, long-term follow-up of CD20 depletion with Rituximab in patients with lymphoma, it was shown that CD20 depletion was an independent risk factor for the development of secondary solid tumour malignancy in both univariate and multivariate analyses (143).

CD138+ plasma cell infiltration has correlated with both good and poor outcomes in NSCLC depending on the study (62–64), with a predilection for a favourable prognosis in adenocarcinoma subtypes. CD138 expression has been associated with increased survival across all NSCLC stages (145,146), and IgG4+ (145) and Ig kappa C expression (146) have independently been shown to be positively prognostic. IgG4 is the least abundant subclass of IgG and is produced in response to chronic antigenic stimulus and inflammation which explains its abundance in the setting of cancer. Reports of IgG4 antibodies and IgG4+ B cells in different cancers suggest the involvement of IgG4 in tumour escape from immune surveillance through a number of potential mechanisms, including IgG4 blockade of IgG1-mediated effector functions (147). In melanoma, IgG4 subclass antibodies effectively impair anti-tumour immunity and IgG4 secreting polarised B cells increased VEGF in cultures and correlated with increased tumour cell IL-10 secretion (148). The association with IgG4 accumulation and poor outcome with links to tumour immune evasion mechanisms has been further delineated in oesophageal cancer (149). Hald and colleagues conferred no prognostic correlation for epithelial or stromal CD138 expression in NSCLC ($p=0.292$) (150). However, in this 55-patient cohort, 24 patient tumours were $CD138^{LO}$ and 29 were $CD138^{HI}$ and 5-year

overall survival rates between the two sub-groups were 35% and 54% respectively. Interrogation of distinct plasma cell subtypes in NSCLC identified a subset of CD138+ plasma cells as an independent negative prognostic in all cases of lung adenocarcinoma and stage 1 cases overall ($p<0.001$) (144). This study further determined this to not be a conventional Ig producing population, but instead a plasmablast-type population displaying Breg characteristics through IL-35 production (144).

Finally, the prognostic impact of follicular B cells was evaluated in two patient cohorts; early-stage untreated NSCLC and advanced stage NSCLC treated with neoadjuvant chemotherapy. "Foll-B-Hi" patients had significantly prolonged survival in early-stage disease, (97% DFS at 4 years compared with 62% in the "Foll-B-Lo" group), and in advanced stage disease, a benefit was demonstrated albeit not significant (56 months median DFS compared with 23 months in the "Foll-B-Lo" group). The global increase in follicular B cell density was associated with an overall increase in mature DC density. When the combined immune populations were taken into account and correlated with survival, "Foll-B-Hi/mDC-Lamp^{HI}" patients had the highest median survival, 100% of early stage patients ($p<0.04$) and 55% of advanced disease patients ($p=0.007$) were alive after a follow-up of 50 and 60 months respectively (137). "Foll-B-Lo/mDC-Lamp^{LO}" patients had the worst prognosis.

Some studies have not demonstrated a prognostic impact of B cell density on NSCLC outcomes (141,144,150–152). However, it may be that the lack of prognostic impact may relate to the high density of Bregs in such studies and it would appear to be essential that Breg density be considered separately (141,150,151). An explicit analysis of whether Breg density is negatively prognostic for outcome however has not been performed. Finally, the

prognostic impact may be dependant not only on enumeration of the appropriate B cell subsets (TIMPs, follicular) but also by enumeration of tumour-associated B cells in the appropriate compartment, as opposed to analysis of un-segmented tumoural B cell density. Most B cells are found at the invasive tumour-stroma margin, and it is here that the cancer cells are likely to polarise B cells to the immune-stimulatory TIPB phenotype.

1.8.1 Tumour-specific B cell responses

1.8.1.1 Antibody Specificity and Antigen Presentation

Tumour antigen-specific B cell responses are evidenced by the production of tumour-specific antibody and the oligo-clonality of TIL-Bs in the TLS (153). B cells cultured from TLS' have been shown to produce tumour-specific IgG and IgA antibodies (137). LAGE-1 was identified as the most immunogenic tumour antigen in NSCLC, followed by MAGE family antigens, p53 and NY-ESO-1(137). Plasma cell and Ig expression associates favourably with outcome in NSCLC (136,145,146). IgG4+ plasma cell infiltration correlates favourably with prognosis (145).

In NSCLC tissue TIL-Bs present antigen to CD4+ TILs with resultant effector responses (153). CD4+ T cells, B cells and DCs were co-cultured together with protein +/- co-stimulation with an anti-CD40/anti-CD28 antibody. Anti-HLA-DR, -DP and –DQ was used to block MHC class II antigen presentation. T cell responses were stratified according to "activated" (spontaneous presentation of antigen to CD4+ T cells), "antigen-associated" (presentation of antigen following re-stimulation by the antigen itself) and "non-responsive". Activated TIL-Bs (CD19+ CD20+ CD69+ CD27+ CD21+) and antigen associated B cells mediated an effector T-cell response (IFN- γ producing CD4+ T cells). Conversely, exhausted phenotype TIL-Bs (CD19+ CD20+ CD69+ CD27- CD21-) were associated with a regulatory T-cell phenotype (Foxp3+ CD4+

TILs) (153). Exhausted TIL-Bs were still able to antigen present but controlled host damage from chronic antigen exposure by inducing a Treg phenotype and ultimately dampening anti-tumour immunity (153).

1.8.1.2 Antibody Production

In the context of NSCLC, plasma cell and Ig expression associates favourably with outcome (136,145,146). IgG4+ plasma cell infiltration correlates favourably with prognosis (145). Moreover B cells cultured from TLS' appear to be participating in the anti-tumour response given that they have been shown to produce tumour-specific IgG and IgA antibodies (137). Meta-analysis of non-TCGA transcriptomic data from over 18000 human tumours found intra-tumoural plasma cell presence to be a significant positive prognostic factor for survival in the case of non-brain solid tumours, except for large cell lung carcinoma (154). IgG+ plasma cells often produce IgG1 with extreme specificity and as part of large hypermutating clonal populations, these cells partake in tumour cell killing via opsonisation, complement fixation and promotion of antigen presentation by dendritic cells (131,155–157). These populations often exist on a phenotypic spectrum and work in synchrony with other IgG+ effector B cells with produce Granzyme B, interferon-γ and MHC class II partaking in direct tumour cell killing and antigen-presentation as well as shifting the balance towards anti-tumour M1 macrophage production (131,155–157).

B cell subsets have been identified through flow cytometry using CD19 as a pan marker as well as differential expression of IgD together with CD38 and CD138 for plasma cells. Infiltrating plasma cells produced tumour-specific antibody. This was determined by B cell cultures from the patient tumours which were analysed by ELISA for IgG and IgA binding

against a series of 33 tumour antigens. LAGE-1 was identified as the most immunogenic tumour antigen in NSCLC, followed by MAGE family antigens, p53 and NY-ESO-1(their prognostic correlations were not described in this study) (137). There is a clear organisation of B cells into follicles intra-tumourally within TLS'. The development of which is associated with the development of antigen-specific humoral responses such as the production of tumour-antigen-specific immunoglobulin.

Plasma cell infiltration has also been shown to be an independent negative prognostic factor in papillary/acinar adenocarcinoma (144). Kurebayashi and colleagues have demonstrated that the fraction of IgG4+ or IgA+ plasma cells varied amongst the cases with increased plasma cell infiltration which may be reflective of the differing cytokine milieu in each adenocarcinoma; however the fraction of neither IgG4- nor IgA-producing plasma cells influenced prognosis (144), which contrasts the findings from Fujimoto's group where IgG4+ infiltration associated with improved outcome (145). This however starkly contrasts the finding of numerous studies implicating IgG4 in tumour immune escape, and furthermore in non-cancerous disease processes such as atopic allergy and SLE (149).

Plasma cell survival is dependent on the expression of B cell Activating Factor (BAFF) and a proliferation-inducing ligand; APRIL which can be secreted by different cell types depending on the environment (158). Increased numbers of APRIL-producing cells were observed in cases where there was higher plasma cell infiltration (144). The co-infiltration of both cell types suggests that the environment in question is conducive to prolonged plasma cell survival in the cancer stroma. Moreover, plasma cells intermingled with B cells in the adenocarcinoma tissue did not display the germinal centre reaction typically seen in the

lymphoid follicle, but instead mimicked the extra-follicular differentiation of plasma cells that is induced by chronic inflammation (144,159). The plasma cells were mainly derived from *in situ* B cell differentiation in the cancer stroma driven by the favourable environment as opposed to being trafficked in from regional lymph nodes (144). The environment therefore influences a plasma cell subtype with a particular phenotype, whether it is IL-35 production (an immunosuppressive cytokine that is CD25+ Foxp3 Treg or Breg derived) or indeed IgG4/IGKC expression, and this could be influencing the differing associations with outcome in NSCLC. These immunosuppressive B cells and plasma cells (often IgA+ CD20-) tend to exert their effector responses via IL-10, TGF-beta, PDL1 and IL-35 and convert CD4+ T cells into Tregs and promote macrophage conversion to the pro-tumoural M2 phenotype (156,160).

1.8.1.3 Immunosuppression

In NSCLC, significantly higher frequencies of peripheral Bregs (CD19+ $CD24^{hi}$ CD27+) and CD19+ IL-10+ B cells were detected compared with healthy controls (161). IL-10+ B cell infiltration has positively correlated with CD25+ Treg expression and advanced clinical stage (162). Multiple subsets exist with similarities in effector function and phenotype; in humans Br1 cells (CD19+ $CD25^{hi}$ $CD71^{hi}$) are strongly IL-10 positive. The differentiation of activated B cells into plasma cells is controlled by the expression of transcription factors, IRF4 and BLIMP-1 which are also critical for B cell suppressive functions and some T cell suppressive functions (IRF4 expression in Tregs is dependent on Foxp3) (92). B cells that co-express IRF4 and BLIMP-1 are the main source of B cell derived IL-10 *in vivo,* and plasma cells are therefore a significant contributor however not all plasma cells produce IL-10 (163). Currently, the signals that are required for differentiation into these regulatory B cells are not known.

Numerous studies have phenotypically characterised various Breg subsets in health and disease, (161,162,164–183). The findings from these studies of human Bregs in cancer underscores the ability of different Breg subsets to mediate immunosuppression in support of tumour growth through a variety of mechanisms; suppressive cytokine production (IL-10, TGF-β, IL-35), suppression of T cells and NK cells and the expansion of suppressive Tregs and myeloid-derived suppressor cells (120), expression of inhibitory ligands such as PD-L1 to dampen anti-tumour immunity (184) and STAT3 mediated promotion of angiogenesis and Treg augmentation (112,171,185). It is unclear whether Bregs enhance tumour progression directly or if an increase in the Breg population is merely reflective of the immune response being mounted against the tumour (186).

1.8.2 B cell repertoire in NSCLC

B cells can exist in a continuum of naïve cells to terminally differentiated plasma cells within the TME and more specifically within the TLS (137). Determining the ratio between these so-called “anti-tumour” TLS derived TIL-Bs and the “pro-tumour”, inhibitory Bregs is important in order to understand the biology and long-term outcome from this disease. This balance is likely influenced by the micro-environmental cues which play a role in determining B cell polarity. CXCL13 and Lymphotoxin have been identified as two factors critical to the formation and development of lymphoid follicles in the gut (187), and in lung cancer, B cells produce CXCL13 and Lymphotoxin via TLR4 signalling which acts as a positive feedback loop to support the formation and high density of TLS (188,189). CXCR5 expressing B cells stimulated by CXC13 coupled CpG-ODN can trigger the cytolytic effect of CD8+ T cells leading to the abrogation of metastasis in 4T1.2 tumour-bearing mice (113). Resveratrol, Lipoxin, Glucosides of Paeony have also inhibited Bregs through STAT3 and/or ERK inactivation leading to a reduction in IL-10 and TGF-β levels thus exerting an anti-tumour effect (190). B cell homeostasis and thus

polarity will largely be determined by the degree of inflammation within the tumour, factors such as tissue hypoxia, intra-tumoural vascularity, cytokine milieu and cellular infiltration are all factors which are likely to exert control over the pro versus anti-tumour B cell balance but as yet there is little evidence describing the Breg/B effector ratio in tumour biology, and this is likely due to the transient inducible nature of Bregs.

1.8.3 The Interplay of B cells and Checkpoint Blockade

In the first study to investigate whether B cell density impacts outcome with checkpoint blockade, B cell content was determined in 34 melanoma patients undergoing PD-1 blockade monotherapy and evaluated for response (191). Dichotomising the patients at the median of CD20 positive cells in at least one histospot there was no difference in response or survival between those with high and low B cell density. The comments made earlier about enumeration of specific B cell populations in specific microenvironmental segments also apply here. B cell depletion in the MC38 (colon carcinoma) and YUMMER1.7 (melanoma) models did not impact the efficacy of anti-PD-1 treatment. Anti-PD-1 outcomes were similar in in *mu*MT mice (mice lacking B cells) and WT mice bearing MC38 tumours (191). In human studies, immunofluorescence staining of CXCR5 and CXCL13 in combination with CD20 in melanoma revealed the formation of TLS, the presence of which was translated into an associative gene signature which was able to predict improved clinical outcomes in patients undergoing checkpoint blockade therapy (131). B cell signatures in melanoma have further been shown to be enriched in treatment responders where immune cells were localised to TLS, a strong proportion of which were class-switched memory cells (157). Work in soft-tissue sarcoma demonstrated that tumours rich in TLS and B cells demonstrated significantly

improved survival and a high response rate to PD1 blockade with Pembrolizumab in phase II trials (192).

Larger datasets in other cancers need to be interrogated to fully understand whether there is any impact of specific intra-tumoural B cell populations on the outcome of checkpoint blockade. This is particularly the case given earlier data showing that clinical benefit with ipilimumab was greater in melanoma patients with sero-positivity against NY-ESO (193): as mentioned above, the CTags appear to be potent immunogens stimulating antibody responses. Furthermore, gene-expression profiling in urothelial carcinoma and melanoma patients undergoing both anti-PD-1/PD-L1 and anti-CTLA4 therapy identified a memory B cell (MBL) signature which was significantly and reproducibly elevated in patients showing clinical benefit (194). It significantly outperformed other immune cell signatures and remained significantly associated with outcome when including tumour mutational burden, copy number aberration burden and checkpoint expression. Samples enriched for an innate PD-1 resistance scores had significantly lower levels of MBL scores. The MBL score positively correlated with BCR heavy chain expression and the expression of T cell activation genes, MHC class II and genes responsible for B cell proliferation and activation within the TME. Finally, high expression of the TIPB signature was associated with improved survival in melanoma patients treated with anti-PD-1, and plasmablast-like and naïve B cell frequencies were significantly higher in patients responding to immune checkpoint blockade (142). The use of anti-PD-1 treatment in murine models has shown to increase total IgG and OVA-specific IgG production in OVA-immunised mice (195). The enhanced humoural response in these mice is thought to be mediated by CD4+ ICOS+ T cells which are presumably of the T follicular helper (Tfh) phenotype that go on to augment terminal B cell differentiation in the germinal centre

(195). PD-1/PD-L1 interactions between Tfh and Bregs control this axis (196,197) and by blocking this checkpoint, Tfh cells are released from Breg-induced suppression. This demonstrates the importance of heterogeneity of the B cell repertoire and how checkpoint blockade can impact downstream immune responses by targeting select populations. Importantly, none of the above studies have examined B cell density as predictive of response to checkpoint blockade in NSCLC or to combination chemo/immunotherapy which has become a 1st line standard of care in this disease.

Given that PD-1 is expressed on B cells and can limit B cell responsiveness (198,199) and further, that certain auto-immune conditions are mediated via auto-antibody formation, the association of B cell sub-populations with checkpoint blockade toxicity has become a focus of investigation. Das et al demonstrated in melanoma patients, a detectable decline in circulating B cell numbers together with an increase in CD21^{LO} B cells and plasmablasts after the first cycle of combination checkpoint blockade therapy (2). These treatment induced changes in B cell numbers preceded and correlated with both the frequency and timing of immune related adverse events (irAE). Early B cell changes correlated with a higher rate of grade 3 or higher irAEs 6 months after starting treatment. Contrastingly, several groups have shown through case report series' that B cell depletion therapy using Rituximab successfully treated B-cell mediated irAEs in NSCLC (200,201), SCLC (202), Melanoma (203) and Urothelial Carcinoma (204). This recapitulates the idea of B cell enumeration and selective targeting of certain micro environmental B cell populations. Bregs which by nature are immunosuppressive and dampen down inflammation would limit anti-tumour activity in response to PD-1 blockade but effector B cell populations with robust humoural responses and T cell activation mechanisms are driving autoimmunity and irAE in response to PD-1 blockade. These intriguing analyses need to be validated, extended and applied to other

cancers amenable to checkpoint blockade and the mechanisms underlying these observations identified.

1.8.4 B cell derived Treatment Strategies

Studies in mouse models of pre-malignancy suggest that B cell-mediated inflammation may be important in promoting the progression to invasive malignancy. Given the huge promise of reversing the pre-malignant phenotype to reduce the cancer burden, there is an urgent need to understand the role of B cells in human metaplasia, dysplasia and in situ cancer and how they mediate progression through these stages in order to decide whether B cell-directed strategies may be of value in reducing the progression of pre-malignancy.

Studies examining B cells with a regulatory phenotype (Bregs) consistently suggest that Breg infiltration may enhance tumour progression. The factors that induce Bregs in human malignancy need to be defined. Specifically are there particular microbes, TLR ligands or cancer cell produced cytokines in the TME that polarise B cells to a Breg phenotype (104,205). Currently used B-cell depleting antibodies cannot distinguish between effector and regulatory B cell subsets, therefore meticulous phenotypic characterisation and study of this subset in the TME (104,205) is required in order to identify Breg specific targets that can be exploited to selectively deplete Breg populations but more fundamentally to fully understand the role of Bregs in human cancer. There are some current potential anti-Breg strategies. *In vivo* murine studies have displayed selective Breg depletion using LXA4 without affecting conventional B cell proliferation, differentiation and germinal centre formation thus promoting an anti-tumour responses (190). An alternate to Breg depletion would be repolarisation of this subpopulation into B effector cells, as has been shown with TLR9 ligands

in vitro (112,113). Adoptive transfer of CpG-pulsed B cells with effector phenotypes into patients with established cancer could be employed to shift the balance in favour of an anti-tumour B cell response within the TME.

More work is needed to understand the anti-tumour impact of antibodies against tumour associate antigens, particularly CTags which appear to be strong immunogens, and to identify new humoural immunity targets. The disappointing results of the MAGRIT trial vaccinating NSCLC patients in the adjuvant setting (206) should not be taken as suggesting that harnessing the anti-tumour antibody response should be deprioritised: mono-epitopic vaccination as cancer therapy has a long history of failure. Multi-valent vaccines, preferably against personalised B cell antigens, are one option. Building on the model of the chimaeric antigen receptor T cells (CART), highly specific B cell receptors to critical tumour antigens could be cloned into autologous B cells and transferred into patients with resultant high specificity and high affinity anti-tumour Ig production. Alternatively, antibodies could be produced ex-vivo and adoptively transferred. Given the role of B cell PD-1 expression in mediating B cell hypo-responsiveness, the role of PD-1 blockade in augmenting these strategies should be explored, as a research priority. Understanding B cell biology will help to refine the understanding behind the effects of checkpoint blockade on the immune milieu. Toxicity from these therapies is the Achilles heel of this treatment strategy. As was alluded to earlier, work in mice and humans has demonstrated that PD-L1^{hi} Bregs play a role in the suppression of humoural immunity through Tfh cell regulation, moreover these cells are resistant to classical anti-CD20 therapy (197). Firmly understanding the ontogeny of these B cells and their relationship to other B cell subsets, including other Breg phenotypes is of paramount importance if we hope

to be able to refine therapeutic strategies so as to augment anti-tumour protective immunity and dampen down autoimmune and hence toxic responses.

Finally, large scale prospective and careful B cell sub-type specific and microenvironment segment specific analyses are required in lung cancer and in other cancers to clarify the role of B cells in modulating the responsiveness to checkpoint blockade and in mediating the toxicity to these therapies. These studies will define the role of B cell-targeted strategies in augmenting the activity of, reducing resistance to and the ameliorating toxicity of this crucial class of anti-cancer agents.

1.9 Functional B cell Immunity and Infection

The role of the B cell in tumour survival clearly has a function as tumour infiltration is commonly reported. The functional capabilities of B cells in NSCLC and the degree of patient immune competency remain to be further elucidated. The incidence of most common cancers increases with age, and this occurs in association with a decline in the quality of the immune response, known as immune senescence. Defects arise in the humoral (B cell) and cellular (T cell) arms of the immune system with age which parallels the age-associated peak incidences of cancer (207). Correlations have been made between age and primary resistance to novel immunotherapeutic anti-cancer agents; a phenomenon also likely to be linked to immune senescence (208). When dominant B cell clones become non-functional (immunoparesis) as in multiple myeloma, this can lead to an increased burden of infections. It is possible tumours act in a similar, inducing non-functional clones and switching off anti-tumour responses.

One of our best tools in the box for assessing immune competence is measuring antibody response to vaccination. The current literature suggests that there are a significant proportion of cancer patients who die from non-cancer related causes such as infection. Data from Netherlands and Belgium has shown that patients with lung cancer aged above 60 are increasingly likely to die from cardiovascular disease and respiratory infections (11). Furthermore, follow-up work done by a similar group showed that 22% of all unexpected hospital admissions within a cohort of lung cancer patients were due to respiratory infection (209). Research has shown that vaccinating against Streptococcus pneumoniae in elderly lung cancer patients decreased hospitalisation from community acquired pneumonia and improved survival (210); and moreover there is an increasing drive to vaccinate against influenza and pneumococcal pneumonia in these patients. The current guidelines from the department of health mandate that any person over the age of 65 should be vaccinated against pneumococcus as a standard of care. We thus sought to determine the burden of pneumonia and vaccine uptake in our regional population of NSCLC patients at a tertiary treatment site (Appendix 1) and moreover designed an observational interventional study which involved vaccinating NSCLC patients with Pneumovax (PPSV23) and measuring the functional *in vivo* B cell response (chapter 7).

Chapter 2. Key Methods

The methodology section has been ordered such that it maps the specific experimental techniques used in each objective of this project. The structure is as follows.

i. Patient recruitment and sample processing for Objectives 1 and 2

ii. Experimental Techniques for Objective 1 – Mass cytometry and RNA sequencing

iii. Experimental Techniques for Objective 2 – Mass cytometry (discussed with Objective 1 methodology), Functional Ex vivo B and T cell assays, Intracellular Flow Cytometry and ELISA

iv. Experimental Techniques for Objective 3 – Sengenics^{TM} Immunome Protein Microarray platform

v. Data Acquisition and Analysis – mirrored to sections I to IV

The methodology has been crafted with expert input from Tan Ti-Myen (bioinformatician from Sengenics Inc), who helped oversee and confirm validity of the machine learning pipeline (Objective 3). Dr. Peter Nightingale from the institute of translational medicine, University of Birmingham helped oversee and craft the general statistical methodology. Dr. Boris Noyvert from the Department of Genomics and Cancer Sciences at the University of Birmingham helped to oversee and craft the high dimensional RNA sequencing data analysis pipelines. Technical field experts, James Youell and Thomas Adejumo gave input and advice on mass cytometry marker orientation and panel set-up.

2.1 Patient Recruitment

2.1.1 Objective 1

Deep B cell immune phenotyping in the acute surgical NSCLC cohort.

Patients were recruited at the pre-operative thoracic surgical clinic typically 1-2 weeks before the date of surgery for suspected lung cancer. Patient data was recorded which included age, demographic, lung cancer staging and histology as well as any significant co-morbidity (including smoking history). All data capture, recording and storage were performed in line with the ethical remit and the trust R&D policies.

2.1.2 Objective 2

Deep B cell immune phenotyping in the advanced disease NSCLC cohort.

Patients were recruited at the pre-treatment oncology clinic typically 3-4 weeks prior to starting definitive anti-cancer treatment (chemotherapy, immunotherapy, radiotherapy) for advanced stage lung cancer. Patient data was recorded which included age, demographic, lung cancer staging and histology as well as any significant co-morbidity (including smoking history). All data capture, recording and storage were performed in line with the ethical remit and the trust R&D policies.

2.2 Patient Samples and Processing

2.2.1 Peripheral Blood Samples

Blood was taken from pre-operative NSCLC patients in the pre-operative thoracic surgical clinic by peripheral venepuncture into 3 BD vacutainer® CPT bottles (NH: ~130IU FICOLL™ 2.0ml) [each bottle contained 8mls of blood] after taking consent as per ethical protocol guidelines and then processed in the laboratory within 2-4 hours. The bloods were typically taken in the morning between 0930am and 1230pm to reduce the effect of diurnal variation on lymphocyte count and subsets.

Blood was taken from healthy volunteers by peripheral venepuncture into 3 BD vacutainer® CPT bottles (NH: ~130IU FICOLL™ 2.0ml) within the University of Birmingham Medical School,

Clinical Immunology Service Department after taking as per university guidelines and immediately processed in the laboratory.

2.2.1.1 Isolation of Peripheral Blood Mononuclear Cells (PBMCs) [BLOOD]

Blood was taken from patients by peripheral venepuncture into 3 BD vacutainer® CPT bottles (NH: ~130IU FICOLL™ 2.0ml) [each bottle contained 8mls of blood] as part of the ethics (REC 17/WM/0272 MLTC). The vacutainer® CPT tubes prevent coagulation and were inverted several times to ensure mixing (8-10 times prior to centrifugation). The CPT tubes were transported to the laboratory and were processed within 2-4 hours of venepuncture for PBMC isolation. After collection, the tubes were stored upright at room temperature (18-25^{o}C) until centrifugation. The tubes were centrifuged directly in a horizontal rotor (swing-out head) for 30 minutes at 1500g (the relative centrifugal force [RCF]). After centrifugation, the mononuclear cell layer (whitish layer – see Figure 2.1) was collected with a 2.5 ml Pasteur pipette (Alphalabs, Eastleigh, UK) and transferred to a 50ml size sterile conical shaped Falcon tube.

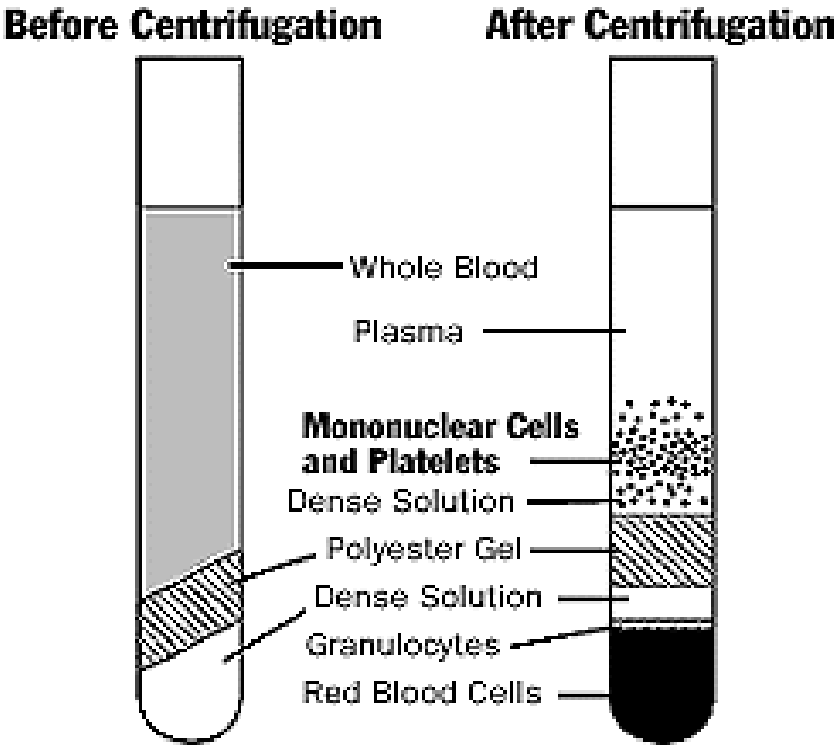


Figure 2.1. A schematic to demonstrate whole blood layering before and after centrifugation with vacutainer CPT tubes.

The 50ml Falcon tubes were then diluted with 10mls of warmed culture media [RPMI 1640 media; 10% FBS] and then mixed by inversion. The 50ml Falcon™ tubes were then centrifuged at 300g with zero acceleration and zero brake for 15 minutes at room temperature. The supernatant was aspirated as much as possible without disturbing the cell pellet. The pellet was then re-suspended and washed in a further 10mls of RPMI 1640. The tube was tapped and gently pipetted with a P1000 in order to resuspend the pellet. The cells were then centrifuged at 300g with zero acceleration and zero brake for 10 minutes at room temperature. After the second wash, 1ml of RPMI was added to the cell pellet following supernatant discard and resuspended. A cell count was thence performed (detailed below).

Following the cell count, a final wash with 10mls RPMI and centrifugation with zero acceleration and zero brake for 10 minutes at room temperature was performed.

2.2.2 NSCLC Tissue Samples

Lung resections samples collected as part of the ethics (REC) were immediately taken "fresh" to the histopathology suite at the Birmingham Heartlands Hospital, UHB NHS Trust. Following examination by the consultant histopathologist, two sections of the tumour were sampled with each sample containing tissue from the core as well as periphery to ensure as much tumour microenvironment representation as possible. The samples taken were not required for diagnostic or clinical purposes and thus did not compromise the clinicopathological analysis of the tumour. One sample was immediately placed in sterile sealed container containing Miltenyi Tumour Storage Solution and stored in a fridge at 4^{o}C [cell labelling and cytometric analysis specimen] and the second sample was immediately placed in a sterile sealed container containing formalin for fixation and stored in a fridge at 4^{o}C [VECTRA multiplex immunofluorescence analysis specimen]. All samples were transported to the laboratory for processing and analysis within 24-48 hours.

2.2.2.1 Isolation of Peripheral Blood Mononuclear Cells (pBMCs) [TUMOUR]

The lung tumour sample acquired was stored in Tumour Storage Solution (Miltenyi Biotech) for up to 72h at 4^{o}C. Within the 72h window, the tumour tissue was dissociated into a single cell suspension by combining mechanical dissociation with enzymatic degradation of the extracellular matrix, which maintains the structural integrity of the tissue. The tumour tissue was enzymatically digested using kit components and the gentleMACS^{TM} Octo dissociator was used for the mechanical dissociation steps. Tumour tissue was weighed to ensure that it fit

within a range of 0.01g to 1g in order to ensure the appropriate volume of enzyme mix and RPMI 1640 culture media (approximately 5mls TOTAL).

2.2.2.1.1 Reagent Preparation

All reagents were supplied by Miltenyi Biotech and reconstituted in a standardised way. Enzyme H was reconstituted from lyophilized powder in a volume of 3ml RPMI 1640 with subsequent aliquoting into volumes of 200µl. Enzyme R was reconstituted from lyophilized powder in a volume of 2.7ml RPMI 1640 with subsequent aliquoting into volumes of 100µl. Enzyme A was reconstituted from lyophilized powder in a volume of 1ml Buffer A with subsequent aliquoting into volumes of 25µl. Aliquots were prepared to avoid repeated freeze-thaw cycles.

2.2.2.1.2 Dissociation of Lung tumour specimen

The tumour sample was cut up into small 2-4mm pieces in a sterile fashion using a 24-blade sterile disposable scalpel (Swann-Morton®, Sheffield, England, 24 blade, REF 0511). Necrotic areas, excess fat or fibrotic tissue was discarded. The tumour pieces were placed into a gentleMACS® C Tube along with the enzyme mix (200µl enzyme H, 100µl enzyme R and 25µl enzyme A) and 4.7ml of warmed RPMI 1640. The C tube was then inverted and placed onto a gentleMACS® Octo Dissociator and covered with a heated jacket. The *37C_h_TDK_2* program was run for 1 hour to allow full dissociation into a single cell suspension. Prior to filtration, a 70µm filter was pre-wet with 2ml of warmed RPMI 1640. The tumour suspension was applied to a MACS® SmartStrainer (70µm filter) which was placed over a 50ml Falcon tube. This was then washed through with 20mls of warmed RPMI1640. The cell suspension was subsequently centrifuged at 300RCF for 7 minutes. Following this, the supernatant was

aspirated, and the cell pellet gently resuspended in 1ml of warmed RPMI 1640 using pipetting with a P1000. A cell count was subsequently performed as detailed below.

2.2.3 Cell Count and Viability

Cell count was performed using a HORIBA Cell Counter (HORIBA Inc. Kyoto, Japan) method. A 50µl aliquot was sampled from the 1ml PBMC suspension and placed under the cell counting apparatus. Cells were then re-suspended in culture media dependent on total cell count to an initial volume such that the concentration was $1x10^6$ for use in phenotyping.

2.2.4 Freezing down PBMCs

Following cell count and repeat centrifugation, the supernatant was discarded, taking care not to disturb the pellet. The pellet was then re-suspended in freezing solution (sterilised mix of 90% heat inactivated fetal calf serum and 10% DMSO) at a density of 4-10 $x10^6$/ml. The freezing solution DMSO was added in a dropwise fashion to the cell suspension. The cell/freezing solution suspension was then aliquoted into pre-labelled 1ml cryovials. The cryovials were then transferred into a Coolcell freezing container and placed in a -80°C freezer for a minimum of 4 hours. The cryovials containing patients PBMCs were transferred into the main -80°C compartment within 72 hours of freezing down and their position was recorded in a sample log.

2.2.5 Defrosting PBMCs for Phenotyping on Mass Cytometer (CyTOF)

Cryovials containing PBMCs were removed from the freezer and transported to the lab on ice. Cryovials were subsequently placed into a 37°C incubator and gently swirled until a few ice crystals remained. All contents were then transferred into a pre-labelled 50ml Falcon tube and washed with 15mls of warmed RPMI 1640. This was then spun at 300RCF for 10 minutes in a similar fashion to above. The supernatant was decanted, and the cells resuspended in

their residual volume. This was further diluted with 2mls of warmed RPMI 1640 and 1 microlitre of Benzonase (Merck 70664-3, 10KU vial, 25U/microlitre) for every $1x10^6$ cells that were frozen down (titrated according to cell number). The cells were then incubated for 1 hour at 37°C. Following this, the cells were washed again with 10mls of warmed RPMI 1640 and spun at 300g for 10 minutes. The cells were then resuspended in 1ml of RPMI 1640 in preparation for a cell count as described above. Expected cellular reconstitution rate was 80-90%.

2.3 Experimental Methods

2.3.1 Cytometry by Time of Flight (CyTOF) Phenotyping

Mass Cytometry enables the measurement of up to 50 features on a single cell. This technique was developed as an alternative to fluorescence flow cytometry, where instead of fluorophores, mass cytometry uses metal ions as the labelling reagent. Inductively-coupled plasma mass spectrometry (ICP-MS) is used for detection and quantification instead of photomultiplier tubes (211,212). This eliminates the issue of spectral overlap and theoretically allows 100 or more features to be measured simultaneously, however in reality the majority of experiments are performed less than 50 channels owing to the limitations both in isotopically pure elements with suitable valence and atomic mass, and stable chelation for certain sizes or valences of the metal ions (212). The pre-labelled cells (with antibody-metal ion tag) are fed through the device, and this enables readouts as a function of time of flight through the apparatus, hence the name CyTOF; Cytometry by Time of Flight. The conjugation of metals to antibody is discussed further below. Sample preparation involves incubations of cells with metal-containing reagents and iridium or rhodium containing DNA intercalators to allow for cell detection regardless of whether any metal-

labelled antibodies bind. This is discussed further below. Currently, mass cytometry is unable to offer information on the size or complexity of cells, i.e., there is no parameter analogous to side scatter in fluorescence flow cytometry. Groups have reported use of plasma membrane staining assays to evaluate mammalian cell size in mass cytometry. This technique employed wheat germ agglutinin staining or Osmium tetroxide (O_sO_4) which have preferential affinity for cell membranes (213). We found that the number of different antibodies that can be simultaneously employed per sample can also circumvent this issue somewhat by providing increased detail on cell phenotypes and populations (212).

We thus employed mass cytometry as our primary phenotyping method to determine the properties and phenotype of B cells circulating in the blood of NSCLC patients undergoing surgery and in those with advanced disease due to start on checkpoint blockade therapy +/- chemotherapy. Using custom-designed B cell CyTOF antibody panels (detailed below), we used mass cytometry to profile the B cell immune signature in these lung cancer patients. In order to comprehensively analyse the immune landscape in the tumour microenvironment, millions of cells from human tumours must be characterised by simultaneous analysis of dozens of markers (214). Mass cytometry enables the quantification of more than 50 readouts at the single-cell level by combining metal isotope-labelled antibodies with mass spectrometry detection (1); this can be done in a time and cost efficient manner, and therefore served as our primary phenotyping method.

2.3.1.1 Antibody Panels

Given the paucity of deep B cell immune-profiling in humans with lung cancer, I based the antibody panels (Table 2.1) on broad B cell subsets, B cell ontogeny and on work done in melanoma (2) as well as work done within the clinical immunology service lab on myeloma

and myeloid leukaemia. CyTOF panels were designed using the panel design tool via the Fluidigm website and antibody titrations at serial concentrations were carried out to validate the entire panel. This was performed on healthy control samples. Each antibody was serially titrated starting with a fixed volume of 0.5µL and depending on the signal intensity, cell capture and degree of spillover, this amount was titrated upwards or downwards. For rarer markers such as IL-10, full dose of 1µL per patient sample was used in the antibody cocktail. All antibodies were titrated and validated in this way individually on healthy control PBMCs before combining together in a cocktail for use on patient samples. Figure 2.2 below illustrates the gating of healthy controls at multiple antibody volumes to achieve the optimum antibody concentration, here CD20 is used as an example.

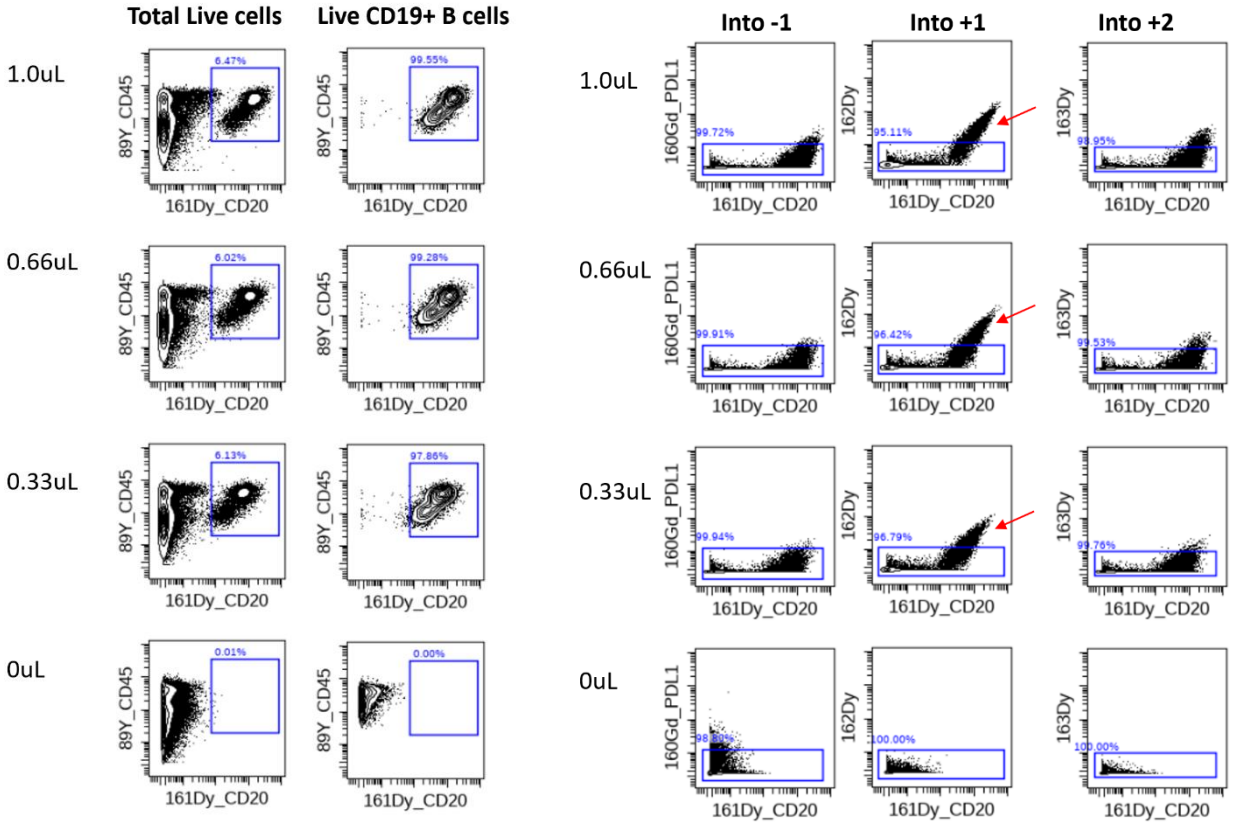


Figure 2.2. Validation of antibody titrations and CyTOF panel set-up is shown above using CD20 titrations as an example. Multiple antibody volumes are trialled, from 0µl to 1µl. Under live CD19+ B cells, staining is more intense at higher volumes however some intermediate non-specific B cell

staining is seen. This must be taken into account along with the spillover index at varying volumes. The red arrows in the third column indicate spillover into neighbouring channels (CD20 is on 161Dy here), therefore +1 indicates 162Dy, +2 indicates 163Dy and -1 indicates 160Dy. There is less spillover at lower concentrations such as 0.33µL, indicating that the optimal volume is between 0.1 and 0.3µL.

2.3.1.2 Probe Selection

The most commonly used probes for antigen detection are antibodies, which bind to the biomolecules of interest. Other probes also exist, designed for investigating more specific biological processes such as carbohydrate-binding proteins (215), small molecules such as tellurium-based oxygen sensors (216), or modified nucleotides for cell-cycle analysis (217). Ensuring that the probes are specific is crucial as cross-reactivity for other targets (secondary affinity) will compromise the integrity of the data. We chose complementary antibodies specific to the antigens of interest as marked out on the panels.

2.3.1.3 Panel Design

We considered the following aspects when designing the CyTOF panels, on the advice of Fluidigm® field application specialists (James Youell and Thomas Adejumo from Fluidigm Inc).

a) Matching antibody probes with appropriate elemental isotopes in order to achieve adequate signal intensity. Given that the majority of antibodies were procured directly from Fluidigm®, the pairings were already biologically validated. The degree of signal intensity also correlates with likelihood for single spillover into neighbouring channels.
b) Oxide formation when cells travel through the ion cloud results in the addition of 16 mass units (16O) to an element's mass. Oxide formation is primarily a problem for La, Pr, Nd, and Sm-labeled probes, and can never be completely eliminated (215,218). Standard daily tuning of the instrument uses the signal from the tuning solution, in

order to optimise machine parameters to limit the 139La + 16O signal to less than 3% of the 159Tb signal present in the tuning solution.

c) The uniformity of the distribution of ion arrival times at the detector is dependent on the mass of ions i.e., this distribution is narrower for low-mass ions (e.g., 133Cs in the tuning solution) and broader for higher-mass ions (e.g., 193Ir). As ion optics have improved, this has decreased from ~2% for the first CyTOF model to ~0.3% for CyTOF2™ and Helios™ for 159Tb (Fluidigm Handbook®). All our samples were run on the Helios™ instrument.

2.3.1.4 Custom Conjugations

Developing the panels required adhering to the principle of "one antigen per metal" whilst simultaneously trying to reduce significant spillover into neighbouring channels. Most antibodies were procured from Fluidigm® as "off the shelf" biologically validated antibodies. Six antibodies in panel 1 and 13 antibodies in panel 2 required custom conjugation with lanthanide metals in order to fit on the panel-appropriate channel. These custom antigens are specified in Table 1. The Conjugation Protocol is summarised below in the flow diagram (Figure 2.3). Broadly speaking this involves, binding the metal ions with a polymer that possesses pendant chelating groups and then the metal-loaded polymer is used to label antibodies or other specific reagents. We followed the Fluidigm® MaxPar antibody labelling guide protocol however modified the protocol in order to optimise the yielded antibody conjugates. The steps are outlined in detail below.

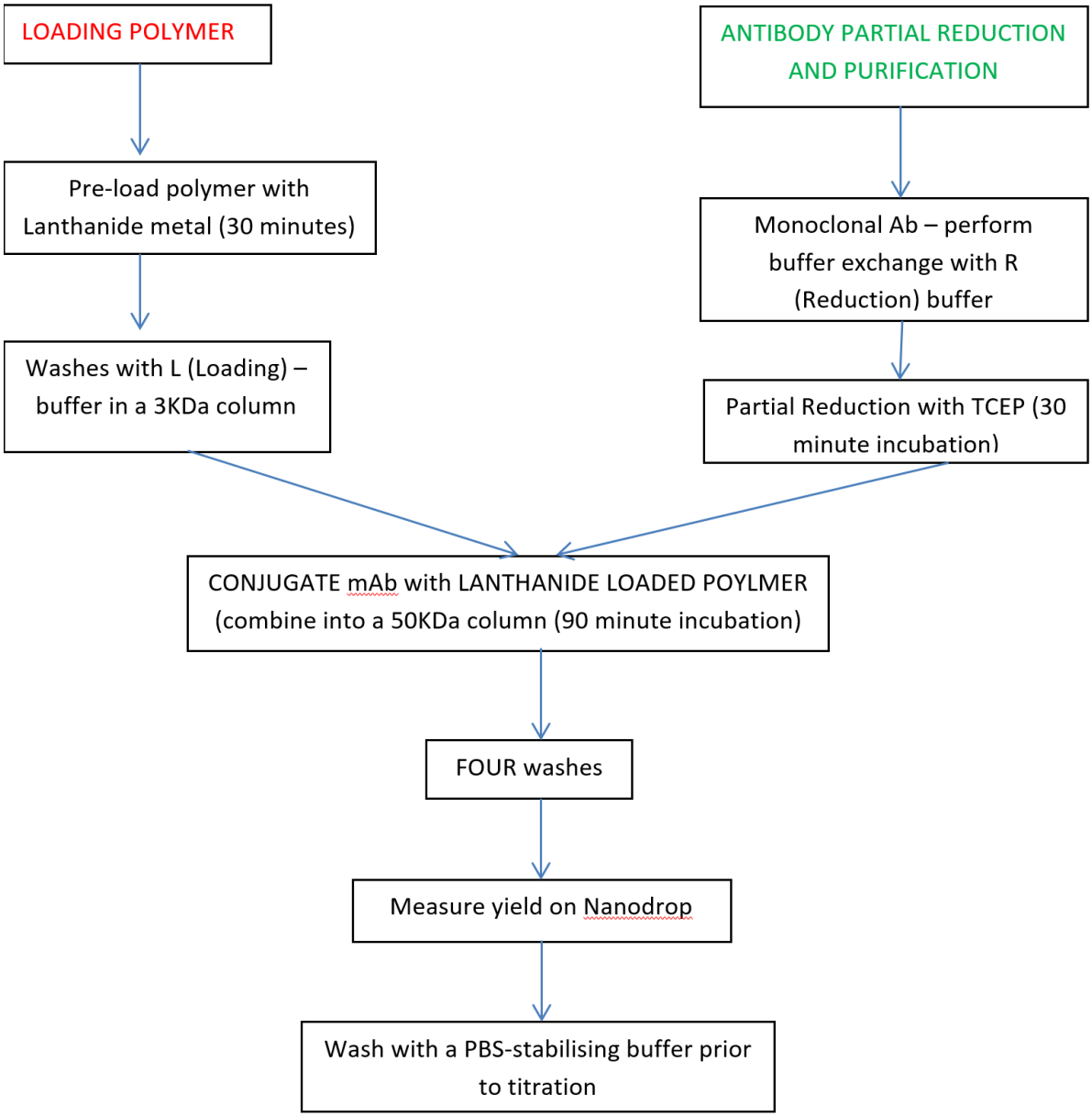


Figure 2.3. Schematic illustrating custom antibody conjugation protocol for mass cytometry.

2.3.1.5 Conjugation Protocol

The first phase of the protocol was to load polymer with the lanthanide metal of interest (Figure 2.2). The polymer was spun down in an Eppendorf tube to ensure all had collected at the bottom (10 seconds at 10,000RPM). This was then resuspended in 95µl of Loading (L) buffer and mixed thoroughly. 5µl of Lanthanide metal of interest was added to this tube to

yield a final concentration of 2.5mM in 100µl. This metal-polymer conjugate was then incubated for 40 minutes in a 37°C sterile incubator.

At this point the manufacturer's recommendation was to move to buffer exchange and partial reduction of the antibody to be conjugated to the metal. However due to the time critical nature of this step, we elected to perform this step later.

Following polymer incubation, we retrieved the metal-polymer conjugate and transferred the volume (100µl) to a 3kDa filter pre-loaded with 200µl L-buffer. This was centrifuged at 12,000g for 25 minutes and room temperature. The wash was repeated with 400µl of Conjugation (C) buffer which was added directly to the filter and centrifuged at 12,000g for 30 minutes at room temperature. During this second centrifugation we proceeded to partially reduce the antibody.

100µg of stock antibody was diluted in 400µl of Reducing ® buffer in a 50kDa filter. This was centrifuged in a separate microfuge at 12,000g for 10 minutes at room temperature. During the centrifugation step, 0.5M TCEP stock was diluted in R buffer in order to achieve a 4mM working solution (992µl of R buffer was added to 8µl of 0.5M TCEP stock). For every antibody that required partial reduction, 100µl of 4mM TCEP solution was required and added accordingly to the antibody in the 50kDa filter and incubated in a 37°C sterile incubator for no more than 30 minutes. This step was time critical as exceeding this reduction step would completely reduce the antibody by irreversibly breaking the disulphide linkages between the antibody Fc regions rendering it unusable.

Following incubation, the partially reduced antibody would be washed in the 50kDa filter with 300µl of C buffer and centrifuged at 12,000g for 10 minutes at room temperature. This step was repeated once with 400µl of C buffer for the second wash.

The purified lanthanide-loaded polymer was retrieved following centrifugation was resuspended in 60µl of C buffer. The resuspended contents were transferred from the 3kDa filter to the 50kDa filter containing the partially reduced antibody resulting in a final volume of circa 100µl. This was gently mixed by pipetting and incubated for 90 minutes. This was followed by 4 wash steps whereby the antibody-conjugation mixture was resuspended in up to 400µl of Wash (W) buffer and then centrifuged at 12,000g for 10 minutes and repeated accordingly. After the final wash, the conjugate was diluted in 80µl of W buffer, with the walls of the filter washed up and down to ensure no antibody-conjugate mixture was wasted. At this point, 1µl of this circa 100µl mixture was used to measure the conjugated antibody yield to ensure that the process had worked properly. The nanodrop device measures the absorbance at 280nm against a W buffer blank. The yield generated however were often unreliable and not at all correlatives as to how the antibody would perform on the mass cytometer. Therefore, we elected to forego this step in the majority of conjugation experiments instead relying on healthy pilot control runs on the mass cytometer as a way of determining antibody performance.

The antibody-conjugate was centrifuged at 12,000g for 10 minutes to remove the W buffer. 100µl of antibody stabilisation buffer (supplemented with 0.05% Sodium Azide) was added to the conjugate and collected in a separate tube by inverting the 50kDa filter over a new tube and spinning at 1,000g for 2 minutes. This was then stored at 4°C until ready to titrate. Titration experiments usually took place using healthy control donor PBMCs at a variety of concentrations (0.25µg/ml, 0.4µg/ml, 0.5µg/ml, 0.75µg/ml, 1µg/ml).

Base B cell Panel

Metal Tag	Antibody	Clone	Source
89Y	CD45	HI30	Fluidigm
141Pr	CD3	UCHT1	Fluidigm
142Nd	CD19	HIB19	Fluidigm
144Nd	CD38	HIT2	Fluidigm
145Nd	CD81	5A6	Fluidigm
146Nd	IgD	IA6-2	Fluidigm
147Sm	CD20	2H7	Fluidigm
148Nd	CD8a	SK1 (Custom)	Biolegend
149Sm	CD25 (IL2R)	2A3	Fluidigm
150Nd	CD138	DL-101	Fluidigm
151Eu	HLA-DR	G46-6	Fluidigm
152Sm	CD21	BL13	Fluidigm
153Eu	Ig Lambda	MHL-38 (Custom)	Biolegend
154Sm	IgG R10	R10 (Custom)	Biolegend
155Gd	CD279 (PD-1)	EH12.2H7	Fluidigm
156Gd	CD274 (PD-L1)	29E.2A3	Fluidigm
158Gd	CD10	HI10a	Fluidigm
159Tb	CD22	HIB22	Fluidigm
160Gd	Ig Kappa	MHK-49	Fluidigm
161Dy	CD5	UCHT2 (Custom)	Biolegend
162Dy	CD79B	CB3-1	Fluidigm
163Dy	BCL-6	K112-91	Fluidigm
164Dy	CD95/Fas	DX2	Fluidigm
165Ho	CD40	5C3	Fluidigm
166Er	IL-10	JES3-9D7	Fluidigm
167Er	CD27	L128	Fluidigm
168Er	Ki-67	B56	Fluidigm
169Tm	CD24	ML5	Fluidigm
170Er	TGF-Beta	TW4-6H10 (Custom)	Biolegend
171Yb	CD185 (CXCR5)	51505	Fluidigm
172Yb	IgM	MHM-88	Fluidigm
175Lu	CD28	CD28.2 (Custom)	Biolegend
176Yb	CD4	RPA-T4	Fluidigm
209Bi	CD16	3G8	Fluidigm

Breg/Tfh Panel

Metal Tag	Antibody	Clone	Source
89Y	CD45	HI30	Fluidigm
141Pr	CD3	UCHT1	Fluidigm
142Nd	CD19	HIB19	Fluidigm
143Nd	ICOS (CD278)	C398.4A	Fluidigm
144Nd	CD38	HIT2	Fluidigm
145Nd	CTLA-4 (CD152)	14D3 (Custom)	Biolegend
146Nd	IgD	IA6-2	Fluidigm
147Sm	IL-21	3A3-N2 (Custom)	Biolegend
148Nd	CD8a	SK1 (Custom)	Biolegend
149Sm	CD25 (IL2R)	2A3	Fluidigm
150Nd	CD71	OKT-9 (Custom)	Biolegend
151Eu	CD40L	24-31	Fluidigm
152Sm	CD1d	51.1 (Custom)	Biolegend
153Eu	TIGIT	MBSA43	Fluidigm
154Sm	IL-4	MP4-25D2 (Custom)	Biolegend
155Gd	CD279 (PD-1)	EH12.2H7	Fluidigm
156Gd	CD274 (PD-L1)	29E.2A3	Fluidigm
158Gd	CXCL13	53610 (Custom)	Bio-Techne
159Tb	IL-35	27537 (Custom)	Bio-Techne
160Gd	CD39	A1	Fluidigm
161Dy	CD5	UCHT2 (Custom)	Biolegend
162Dy	TIM-1	1D12 (Custom)	Biolegend
163Dy	BCL-6	K112-91	Fluidigm
164Dy	CD95/Fas	DX2	Fluidigm
165Ho	CD40	5C3	Fluidigm
166Er	IL-10	JES3-9D7	Fluidigm
167Er	CD27	L128	Fluidigm
168Er	Ki-67	B56	Fluidigm
169Tm	CD24	ML5	Fluidigm
170Er	TGF-Beta	TW4-6H10 (Custom)	Biolegend
171Yb	CD185 (CXCR5)	51505	Fluidigm
172Yb	IgM	MHM-88	Fluidigm
173Yb	CD73	AD2 (Custom)	Biolegend
174Yb	HLA-DR	L243	Fluidigm
175Lu	CD28	CD28.2 (Custom)	Biolegend
176Yb	CD4	RPA-T4	Fluidigm
209Bi	CD16	3G8	Fluidigm

Table 2.1. This illustrates the markers used in each antibody panel. The Base B cell panel has 34 markers and the Breg/Tfh panel has 37 markers. The antigen, corresponding metal tag, clone and source are shown for both panels.

2.3.1.6 Phenotyping Staining Protocol

Following cellular resuspension, the cells were filtered through a blue strainer cap FACS tube in a volume of approximately 50 microlitres (µl). The cell volume can be titrated upwards (if low volumes) with MaxPar® cell staining buffer (Fluidigm®). MaxPar® cell staining buffer (Fluidigm®) was added (2-3mls) with a 2.5 ml Pasteur pipette (Alphalabs, Eastleigh, UK), and then spun at 1500RPM for 5 minutes. The antibody cocktail was made up according to pre-determined amounts of antibody required per patient sample. Separate surface and intracellular marker cocktails were made. This was then multiplied by the total number of samples (n) requiring staining in that session to avoid working with very low antibody volumes and therefore risking reagent loss. Moreover, due to the repeated pipetting and filtering/washing steps, antibody cocktail was made up for "n+2" samples to account for any inter-step reagent loss. The supernatant was decanted, and the cell suspension transferred to regular 5ml FACS tubes; 5µl of Fc blocking agent (Biolegend®) was added and, the cells were incubated at room temperature for 10 minutes. Without washing off the Fc block, the antibody cocktail was added to the suspension (surface cocktail) and vortexed gently for a few seconds. The total staining reaction volume was 100µl; the required volume of MaxPar® cell staining buffer (Fluidigm®) is added to achieve this. This was followed by 30-minute incubation at room temperature. During the last two minutes of this incubation, a cisplatin viability stain was performed. Cisplatin was added at a concentration of 5µM to the cell suspension and incubated at room temperature for no more than 2 minutes. The samples were then washed with 4 ml of MaxPar® cell staining buffer (Fluidigm®) and centrifuged at 1500RPM for 5 minutes. The supernatant was aspirated, and the wash and centrifugation were repeated.

Following surface antibody staining, the cells were permeabilised for intracellular staining. The cells were first fixed by adding 1ml of MaxPar Fix I Buffer (Fluidigm®) to each tube and gently vortexed. This was followed by a 15-minute incubation period at room temperature. The cells were then washed with 2mls of MaxPar Perm-S Buffer (Fluidigm®) and centrifuged for 5 minutes at 1500RPM. The supernatant was discarded, and the cell pellet resuspended in residual volume; the wash step with the MaxPar Perm-S Buffer (Fluidigm®) was repeated along with repeat centrifugation. The cells were resuspended and 2μl of Heparin solution (2kU/ml stock) was added to each sample to prevent non-specific binding of charged eosinophils. This was followed by a 10-minute incubation period. The antibody cocktail (intracellular) (required volume) was then added, and tubes were gently agitated. The total volume was 100μl; the required volume of MaxPar® cell staining buffer (Fluidigm®) is added to achieve this. This was followed by a 30-minute incubation period at room temperature. The samples were washed with 4mls of MaxPar® cell staining buffer (Fluidigm®) and centrifuged at 1500RPM for 5 minutes and the supernatant discarded. During the wash, 500microlitres of cell intercalation solution (CIS) for each sample was prepared by adding Rhodium 103 (4μl) to 4mls of Fix and Perm Buffer (Fluidigm®).

Following addition of the 0.5ml of CIS, the samples were gently vortexed and incubated at room temperature for 1 hour or left overnight at 4^{o}C. For the validation work, cells were extracted from whole blood and directly stained as per the method described and left in CIS for 24h and 48h prior to running on the machine. Cells were also defrosted from the -80^{o}C freezer and stained and left in CIS for 24h and 48h prior to running on the CyTOF. A batch of cells was stained and then instead of placing in CIS, they were placed in 9 parts heat-inactivated fetal calf serum and 1-part DMSO and frozen down at -80^{o}C. This batch was

defrosted and run on the CyTOF to provide additional validation to the staining protocol variations. The most consistent staining protocols with good antibody capture and binding of metal tagged antibody to antigen were staining from fresh blood and staining from defrosted PBMCs (one defrost cycle) and running within 24hours. Owing to logistics and sample acquisition rates, all samples were frozen down at -80 degrees in a uniform manner and defrosted 24 hours prior to running on the mass cytometer.

2.3.1.7 Cell number

When considering starting cell number, the following factors were taken into account.

a) Initial cell viability: if this was low to begin with, more aliquots were used given that the defrosting process would result in circa 10-20% loss of PBMCs. Poor initial viability would result in fewer live intact singlets.
b) Rare populations: B cells were our target population of interest, and this makes up 5% of the total PBMC population. On average, we would aim to collect between 40,000-50,000 Live CD19+ B cells as this would enable reasonable interrogation of rare subsets such as Bregs. Assuming the "5%" rule was true across all samples; we would want to collect $0.8*10^6 – 1.0*10^6$ live PBMCs.
c) The staining protocol particularly intracellular staining would result in around 50-60% cell loss, therefore a starting amount of $2*10^6$ PBMCs per sample would be reasonable.
d) The maximum amount we would stain in one sample tube would be $3*10^6$ cells, as beyond this, antibodies would saturate leading to poor binding and suboptimal data.
e) Length of sample acquisition was performed at an average rate of 350 events per second, with view to collecting a minimum of 800,000 events (38 minutes). It was

logistically prohibitive to run for longer and still gather data on 4-6 samples in one session. Enriching for CD19+ B cells prior to running was a possibility however this would have not given us information about other populations of interest and their relative frequencies. Enriching also risks potentially losing cells.

2.3.2 RNA Extraction from Tumour Specimens

Tumour RNA was extracted and purified using the Qiagen RNeasy® Plus Mini Kit according to manufacturer protocol. Tumour samples were defrosted according to the protocol outlined above. This kit was chosen for its versatility, avoidance of use of toxic substances such as phenol which can affect RNA integrity and the ability for parallel processing of samples. Time efficiency was key in RNA extraction, the sooner the samples were placed on ice and kept cool, the higher the chance of improved RNA integrity. The "plus" kit further allowed for simultaneous selective removal of double stranded DNA without the need for additional DNase digestion.

Samples were first lysed and homogenised in a highly denaturing buffer, which deactivated the RNases to ensure isolation of intact RNA. The lysate was then passed through a gDNA Eliminator spin column. This column, in combination with the optimized high-salt buffer, allowed efficient removal of genomic DNA. Ethanol was added to the "flow-through" to provide appropriate binding conditions for RNA, and the sample was then applied to a RNeasy spin column, where total RNA binds to the membrane and contaminants washed away. High-quality RNA was then eluted in 30-45µl of water. The procedure is highlighted in Figure 2.3.

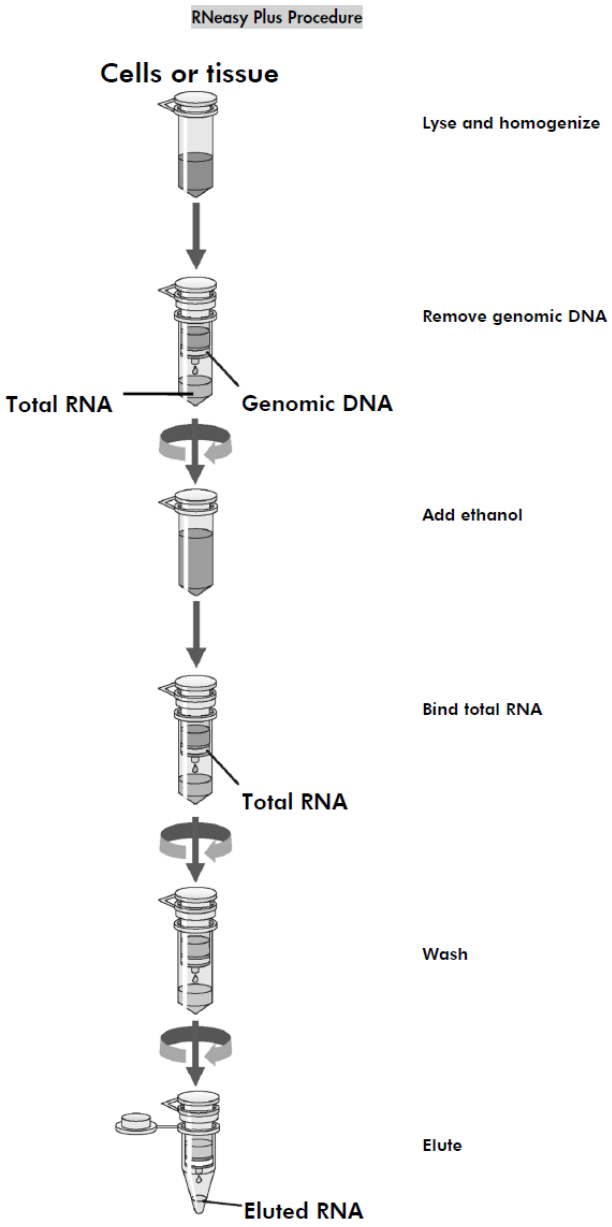


Figure 2.4. Overview of Qiagen RNeasy® Plus Mini Elution protocol

Per sample, the maximum amount of cellular starting material did not exceed $1*10^7$ cells so as to avoid overloading the gDNA Eliminator spin columns and therefore contaminating purified RNA with DNA.

2.3.2.1 Reagent Preparation

1. β-mercaptoethanol (β-ME) was added to Buffer RLT plus in order to deal with potentially high RNase concentration in the cell suspensions [10µl β-ME per 1 ml Buffer RLT Plus]. This solution was stable at room temperature for up to 30 days.
2. Buffer RPE was supplied as a concentrate, therefore 4 volumes of 96-100% ethanol was added to obtain a working solution.
3. All steps were performed at room temperature including those requiring a microcentrifuge.

2.3.2.2 Protocol

1. Reagents were appropriately prepared on the day of RNA extraction.
2. Cellular material was defrosted as per in-house protocol and collected as a cell pellet.
3. The cell pellet was resuspended in 350µl of buffer RLT plus and then transferred to a QIA shredder column for homogenisation and disruption of the cellular material. Homogenisation was carried out in order to maximise RNA yield and prevent clogging of the RNA spin easy column.
4. The QIA shredder spin columns were then placed in a 2 ml collection tube and centrifuged for 2 min at maximum speed (12,000RPM).
5. The supernatant was carefully removed by pipetting and transferred to a gDNA Eliminator spin column placed in a 2 ml collection tube. This was centrifuge for 30 seconds at 10,000RPM.

6. The column was discarded and “flow-through” preserved. This step was to remove any insoluble material that could clog the gDNA Eliminator spin column and interfere with DNA removal.
7. Centrifugation was repeated as required in order to remove any remaining liquid from the column membrane.
8. 350µl of 70% ethanol was then added to the flow-through and mixed by pipetting.
9. 700µl of the sample was then transferred to a RNeasy spin column placed in a 2 ml collection tube, centrifuged for 15 seconds at 10,000RPM and flow-through discarded. This step was carried out twice if the starting volume was greater than 700µl.
10. 700µl of Buffer RW1 was added to the RNeasy spin column and centrifuged for 15 seconds at 10,000RPM to wash the spin column membrane and flow-through discarded.
11. 500µl of Buffer RPE was added to the RNeasy spin column, centrifuged for 15 seconds at 10,000RPM to wash the spin column membrane and flow-through discarded. This step was performed *twice,* the second time however, the centrifugation was for 2 minutes so as to dry the spin column membrane, ensuring that no ethanol is carried over during RNA elution. Residual ethanol may interfere with downstream reactions.
12. The RNeasy spin column was then placed in a new 1.5ml collection tube and 30–45µl of RNase-free water was added directly to the spin column membrane. This was then centrifuged for 1 minute at 10,000RPM to elute the RNA. In cases where the expected yield was >30µg, this step was repeated using another 30–45µl of RNase-free water.
13. RNA was then frozen down at -80 degrees in two batches per sample.
14. Two separate aliquots were frozen, a 5µl for RNA quality control check and the rest for full experimental sequencing.

2.3.3 Functional *Ex vivo* B and T cell Assays

2.3.3.1 Cell isolation and culture

PBMCs were isolated as described and defrosted from -80°C as per the protocol. All subsequent steps involved in isolation and culture were performed sterile in the hood. CD4+ T cells and CD19+ B cells were subsequently isolated by magnetic cell sorting, based on CD4 microbead positive selection and subsequent CD19 negative selection (MACS, Miltenyi Biotech); purity was consistently >98%. The cellular isolation was carried out according to the manufacturer protocols (Miltenyi Biotech). Briefly, this involved resuspension of defrosted PBMC pellets in MACS running buffer, incubation with microbeads and staining antibody cocktails and then passing through appropriately sized filter columns placed within a magnet to carry magnetic cell separation. All steps were performed "cold", all reagents were kept on ice to preserve cells and ensure correct labelling with microbeads and antibodies. CD4+ T cells were collected first (positive selection) and then CD19+ B cells were collected. During the CD19+ elution step, CD4+ T cells were labelled with cell trace violet (CTV) according to manufacturer instructions. In order to ensure correct, plating concentrations, the cell number was determined using the cell counter method or manual counting using a haemocytometer to calculate volumes of culture buffer for resuspension. Cells were cultured in 10% FBS, RPMI-1640 (Sigma) with 2mM L- glutamine (Sigma), 0.1g/L sodium bicarbonate (Sigma), supplemented with 100U/mL penicillin and 0.1mg/mL Streptomycin (PAA), and 10mM Hepes (PAA), in round bottom 96-well Nunclon plates (Thermo Scientific) at 0.3 x 10^6 cells/well. To induce IL-10 expression, the TLR9 ligand CpG ODN 2006 (achieve a 1µM concentration) (Miltenyi Biotech) was used, alongside IL-2 (250µg/ml, Aldesleukin, Peprotech). All B cell only cultures lasted for 40 hours. This time point was chosen as most appropriate for the detection of IL-10, based on expression kinetics. Cells were re-stimulated with PMA (1mg/mL) and

Ionomyocin (1mg/mL, Sigma) in the presence of Brefeldin A, and GolgiStop (both 1mg/ml) for three hours at a 1:1000 concentration, and then stained for intracellular cytokine production.

2.3.3.2 T and B cell co-culture

Autologous CD4+ T and CD19+ B cells were isolated as described above. CD4+ T cells were stimulated with plate bound CD3 and CD28 (both 1mg/ml, Biolegend). B cells were stimulated to induce IL-10 expression in the same manner as for B cell only assays: the TLR9 ligand CpG ODN 2006 (achieve a 1µM concentration) (Miltenyi Biotech) was used, alongside IL-2 (250µg/ml, Aldesleukin, Peprotech) and cells cultured for 40 hours. The B cells were then added to the T cells in a 1:1, 1:2 and 1:4 ratio and cultured for 4 days. After 91 hours, cells were re-stimulated with PMA (1mg/ml) and Ionomyocin (1mg/mL, Sigma) in the presence of Brefeldin A, and GolgiStop (both 1mg/ml) for five hours, and then stained for intracellular cytokine production. Various control wells were used as well as technical replicates (at least two) per sample. The plate set-up including description of the wells is shown below in Figure 2.5.

SAMPLE 1:	B cell IL-10 Assay only (1)	B cell IL-10 Assay only (2)		ANY CD4+ T cells - IL-2 only, CTV-ve, NO activation	ANY CD4+ T cells - IL-2 only, CTV+ve, NO activation	CD4+ T cells (Sample), CTV+, FULL activation *[stimulated]*	CD4+ T cells (Sample), CTV+, FULL activation *[unstimulated]*		CD4:B co-culture **(1:1)**	CD4:B co-culture **(1:2)**	CD4:B co-culture **(1:4)**	*stimulated*
									CD4:B co-culture **(1:1)**	CD4:B co-culture **(1:2)**	CD4:B co-culture **(1:4)**	*unstimulated*
SAMPLE 2:	B cell IL-10 Assay only (1)	B cell IL-10 Assay only (2)		ANY CD4+ T cells - IL-2 only, CTV-ve, NO activation	ANY CD4+ T cells - IL-2 only, CTV+ve, NO activation	CD4+ T cells (Sample), CTV+, FULL activation *[stimulated]*	CD4+ T cells (Sample), CTV+, FULL activation *[unstimulated]*		CD4:B co-culture **(1:1)**	CD4:B co-culture **(1:2)**	CD4:B co-culture **(1:4)**	*stimulated*
									CD4:B co-culture **(1:1)**	CD4:B co-culture **(1:2)**	CD4:B co-culture **(1:4)**	*unstimulated*
	ELISA and ICFC			**Negative Control**	**Positive Control**	**MAX Proliferation**	**MAX Proliferation (unstimulated)**		Full activation for all wells here			

Figure 2.5. T and B cell co-culture set up including B cell only culture set up

i. CTV negative control: (Any CD4+ cells, healthy donor): Unstained CD4+ T cells preserved in IL-2 but no activation with CD3/CD28 (negative gate in your flow cytometry assay)

ii. CTV positive control: As above, but CTV-stained cells preserved in IL-2 during culture time but no activation with CD3/CD28. This will be the brightest (undivided) population

iii. CD4+ T cells alone: Maximum CD4+ T cell proliferation expected with FULL activation (plate bound CD3/CD28) conditions and CTV labelling. Stimulated and unstimulated fractions (with PMA/Ionomycin/Brefeldin/Monensin cocktail)

iv. CD4:CD19 co-culture wells [1:1, 1:2, and 1:4]: CD4+ T cells as per iii. Different ratios of CD19+ B cells added. Stimulated and unstimulated fractions (with PMA/Ionomycin/Brefeldin/Monensin cocktail)

2.3.3.3 CD4+ T cell Proliferation Assay

To assess the effect of IL-10 on CD4+ T cell proliferation, cell trace violet (Thermofisher Scientific) was used. CD4+ T cells isolated from PBMCs were stained with 5µM cell trace violet (Thermofisher Scientific), following manufacturer's instructions. Briefly the CD4+ T cells were resuspended in a warm 1:1000 diluted solution of CTV: DMSO in PBS and incubated for 30 minutes before being washed and resuspended in culture media. Cells were then cultured with varying concentrations of CD19+ B cells from 0.6 x 10^5 – 0.24 x 10^6/well. They were subsequently cultured in 96-well flat bottom plates, containing plate bound CD3 and CD28 (both 1mg/ml), at 0.3 x 10^6 cells per well. After 4 days, cells were harvested and acquired.

2.4 Data Acquisition

2.4.1 CyTOF

Immediately prior to running on the mass cytometer, cells were washed with 4mls of MaxPar® cell staining buffer (Fluidigm®), centrifuged as above and the supernatant discarded. Prior to acquisition on the CyTOF, the cells underwent two cycles of wash with ddH_2O and 5-minute centrifugation at 800g. After removal of supernatant, cells were left pelleted until ready to run on the CyTOF. EQ beads (Fluidigm®) were vortexed and diluted at a ratio of 9 parts ddH_2O to 1 part EQ beads. The cell pellet was resuspended in the 0.1×EQ bead solution and then the cells were filtered through cell strainer cap tubes. The cell concentration was adjusted with EQ bead solution to achieve 0.5-1.0 $x10^6$ cells/ml. Concentrations exceeding this, were likely to cause a higher incidence of ion cloud fusion and doublet acquisition on the CyTOF (Helios flow rate is fixed at 30 microlitres/minute thus 1.0 $x10^6$ cells/ml is 500 events per second which is optimal). In order to stay within a safe window, samples were run at a rate of 300-400 events per second.

2.4.1.1 CyTOF Data Pre-processing

Before informative data analysis was carried out, a number of pre-processing steps were completed in order to convert the estimated ion counts in the FCS file into a matrix of ion counts for live, intact, single cells to enable appropriate analysis (212). There are two types of file formats, IMD (integrated mass data format) and FCS (flow cytometry standard format). The latter was the primary type of file format we utilised.

2.4.1.2 Normalisation and Batch correction

The use of the normalisation beads (EQ) pre-sample loading acts as a reference to which reported signals can be normalised. This is because during the course of a CyTOF run,

instrument sensitivity may decline due to cone contamination and mass calibration drift and cause overall signal intensity decay (212). The normalisation of data can be carried out on the instrument using Fluidigm software

The occurrence of batch effect due to use of different lots of reagent, runs on different days (219), or runs on different CyTOF instruments (219) was also taken into account. All experimental design was standardised and the analysis pipeline, which is discussed later, was used to account for potential batch effect within the data. The inclusion of well-characterized healthy controls was another precautionary measure (220).

2.4.2 Next Generation Bulk RNA sequencing

RNA, once extracted from tumour specimens was frozen down at -80 degrees in two batches per sample, one for quality control assessment and the other for bulk sequencing. The first step in RNA sequencing was generation of the library preparation. This involves fragmenting the RNA and converting the RNA fragments into double stranded DNA as this is more stable than RNA. Moreover, it can be amplified and modified. Sequencing adaptors are then added in order to allow the sequencing apparatus to recognise fragments and to sequence multiple samples at the same time. PCR is employed to amplify the sequences which then undergo QC to ensure they are of adequate length and concentration. This is followed by sequencing of the library where millions of sequences are generated and known as “reads”. The outputs are stored as FASTq files (Illumina only outputs) or SRA files (if dealing with SRA archive). Pre-processing of these files prior to data analysis is discussed further below.

We used the Lexogen QuantSeq 3’ mRNA sequencing kit. This kit provides library preparation protocols that generate illumina-compatible libraries from poly-adenylated RNA within 4.5 hours. The protocol generates one fragment per transcript, resulting in accurate gene

expression values with sequences obtained closed to the 3' end of the transcript. There are two read directions, FWD (forward) and REV (reverse). We used the FWD kit. This method contains the Read 1 linker sequence in the second strand synthesis primer, hence the next generation sequence reads are generated towards the poly(A) tail and directly correspond to the mRNA sequence.

An overview schematic of the RNA sequencing process can be found in the Lexogen handbook along with detailed procedural steps (221).

2.4.2.1 Pre-processing of Data

FASTq files which are generated from each sequencing read are structured as four lines of raw data. The first line starts with an "@" and is followed by a unique ID for the sequence that follows. The second line are the bases called for the sequence fragment. The third line is always a "+" character and the fourth line contains quality scores for each base in the sequenced fragment. This raw data undergoes several quality control processes prior to analysis and the initial phases are performed through a linux terminal. Whereas most downstream steps for analysis are done in R, the pre-processing steps are very computationally intensive and beyond the remit of most personal computer processors. Quality control steps included filtering out any "garbage" reads, with further trimming of the data with Trimmomatic or Cutadapt (222,223). The high quality reads were then aligned to the genome in a process known as "mapping" using HISAT2 or STAR with subsequent quality control checks using RSeQC (224–226). The "counting" and generation of read count files were carried out with STAR (226) (HTSeq or Subread packages (227,228) can also be used to count the number of reads per gene (pre-specified by intron or exon and attribute type such

as “gene ID”) and generate read count files in .tsv and table format). The raw read count files were then imported into R for differential gene expression analysis.

2.4.3 Multiplex Immunofluorescence (VECTRA Polaris)

The NSCLC samples were fixed in 4% isotonic formaldehyde for no more than 24 hours, dehydrated and embedded in paraffin. Sections (4-µm) were cut from each paraffin-embedded tissue and stained with hematoxylin and eosin (HE) to evaluate tumour pathology.

Formalin-fixed, paraffin-embedded (FFPE) tissue sections of (4µm) were baked for 2 h at 60 °C before staining. Deparaffinization and antigen retrieval (pH9 for 20 minutes at 100°C) were performed on the Leica BondRx Automated IHC stainer. Primary antibody dilutions were optimized individually in a chromogenic DAB staining. Control tissue was stained with the Bond Polymer Refine Detection kit (DS9800) and evaluated by a pathologist for specificity. Each marker was then assessed by a single fluorescence staining, to optimise the fluorophores dilution and to generate a library for spectral separation, using the Opal Polaris 7 Colour Automation IHC Detection Kit (NEL871001KT) from Akoya Biosciences. Each marker was tested in the six different positions to evaluate the effect of the heat deactivation steps and the epitope stability and determine their sequence in the panel accordingly (Table 2.2).

Marker	Supplier	Catalogue Number	Final Dilution	Control Tissue	Reporter	Reporter Dilution	Position
BCl-6	Leica	PA0204	RTU	Tonsil	520	1:150	5
CD4	abcam	ab13545	1:200	Tonsil	480	1:150	3
CD8	Leica	NCL-L-CD8-4B11	1:400	Tonsil	780	1:10	6
CD20	Dako	M0755	1:200	Tonsil	690	1:150	4
CD138	Dako	IS642	RTU	Tonsil	620	1:100	2
IL-10	Proteintech	60269-I-Ig	1:400	Lung	570	1:150	1

Table 2.2. Markers of interest used during tissue immunofluorescence analysis, with corresponding supplier information, dilution, reporter channel and position.

Slides were serially stained with the following antibodies: IL-10 (1:400), CD138 (RTU), anti-CD4 (1:200), -CD20 (1:200), -BCL6 (RTU) and CD8 (1:400), with an incubation of 30 minutes. Secondary antibody used was OPAL POLYMER HRP MS + RB (ARH1001EA) from Akoya Biosciences, incubated for 10 minutes. TSA-conjugated fluorophores used to visualize each biomarker were Opal 480, Opal 780, Opal 690, Opal 620, Opal 570 and Opal 520, with a 10-minute incubation. Opal 780 was incubated for 60 minutes, preceded by a 10-minute incubation in TSA-DIG. Slides were mounted with ProLong Diamond Antifade Mountant (Fisher Scientific Ltd, 15205739) and stored at 4°C before imaging. Image acquisitions (20 × magnification as multispectral images) were performed using the Vectra Polaris multispectral imaging platform (Akoya Biosciences), with the entire slide image being scanned and 7-10 representative regions of interest chosen by the pathologist. DAPI was used to count number of cells per slide. Negative controls (PBS instead of primary antibody) were run simultaneously with these samples.

2.4.4 Intracellular Flow Cytometry (ICFC)

Antibodies used, and their dilutions, are shown in Table 2.3. For cytokine detection, cells were re-stimulated in the final 3 (CD19+ only) or 5 (co-culture) hours of culture with PMA (1mg/mL) and Ionomycin (1mg/mL, Sigma) in the presence of Brefeldin A, and GolgiStop (both 1mg/ml, both BD Biosciences). Cells were harvested and incubated with Live/Dead stain in a dark incubation chamber at room temperature (eBioscience™ Fixable Viability Dye eFluor™), then washed and fixed in 4% paraformaldehyde (Fixation Buffer, Biolegend) for 15 minutes in a dark incubation chamber at room temperature. Following fixation, cells were washed twice, firstly with 3ml PBS and secondly with 3ml PBS supplemented with 1% FBS. Cells were then

incubated with the appropriate volume of antibody, diluted in 0.1% Saponin in 0.5% BSA in PBS for 60 minutes in a dark incubation chamber at room temperature. Cells were acquired using a BD FACSCanto II (BD Biosciences), and analysis was conducted using MRC Cytobank software (Beckman Coulter Inc.). A general gating strategy to identify live human B cells is shown in Figure 2.5. Gating was carried out using either a fluorescence minus one control, and/or unstimulated cells.

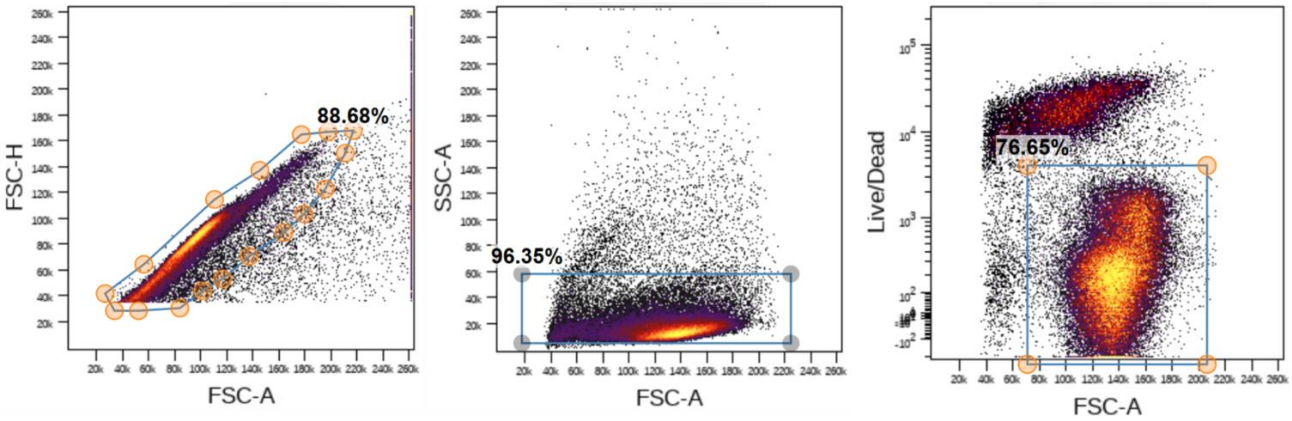


Figure 2.6. General Gating Strategy to Identify purified CD19+ B cells (lymphocytes)

Marker	Biolegend Fluorophore
Fixability Viability Dye eFluor 780	Equivalent to APC-eFluor 780 or APC Alexa Fluor 750
IL-2	Ax647
IL-6	PE/Dazzle 594
IL-10	Ax488
IL-17	PE
TNF-alpha	BV421
IFN-gamma	PE-Cy7
Cell Trace Violet	401-450 Excitation

Table 2.3. Intracellular antigen targets of interest along with corresponding fluorophores/detection channels

2.4.5 ELISA

Prior to ICFC staining, cell culture supernates were harvested (circa 80-100µl) and stored at -20°C until analysis. Sandwich ELISA was used to analyse supernatant analytes. ELISAs for IL-10 (Human Quantikine R&D systems) were conducted as per manufacturer's protocols. Quantification of concentrations was drawn from a standard curve run on each plate.

2.4.6 Proteomics Immunome Array

Objective 3 was centred on characterising the proteome in NSCLC. We utilised pre-operative serum samples from a historical NSCLC trial (CLUB) cohort and identified the circulating auto-antibody signature using Immunome microarray technology from Sengenics Inc. This technology is based on their KREX patented protein-folding technology, which enables the expression of functional proteins to create high-density arrays. It provides a simple and rapid way to immobilize and purify correctly folded proteins for the detection of autoantibodies. A schematic below illustrates this principle (Figure 2.6).

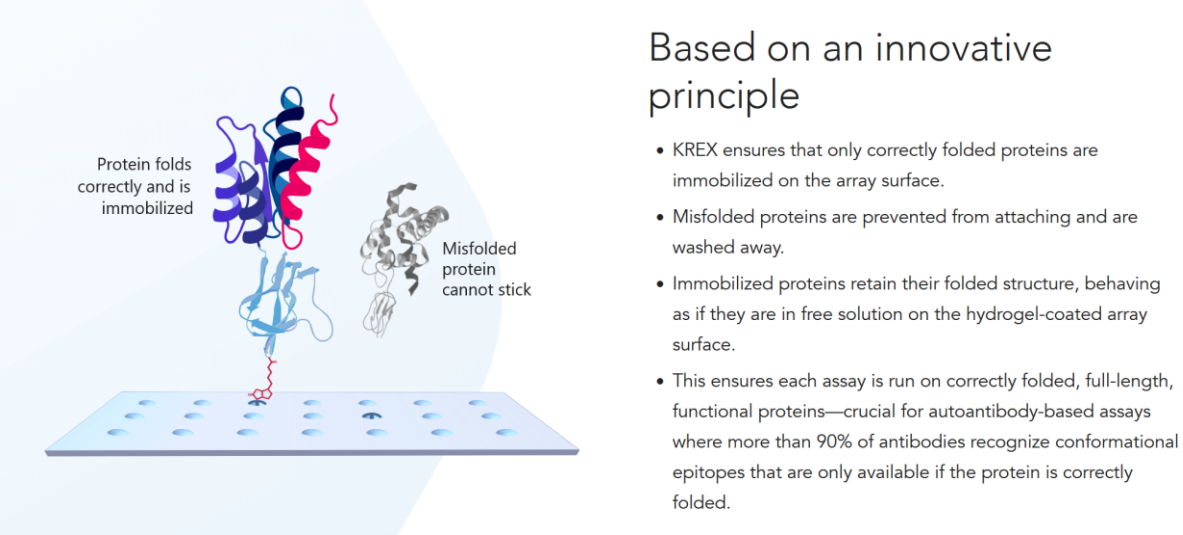


Figure 2.7. A schematic illustrating KREX binding principle from the Sengenics Inc Website (https://sengenics.com/technology/krex-protein-folding-technology/).

The process of sample application to the immunome array is illustrated below in Figure 2.7.

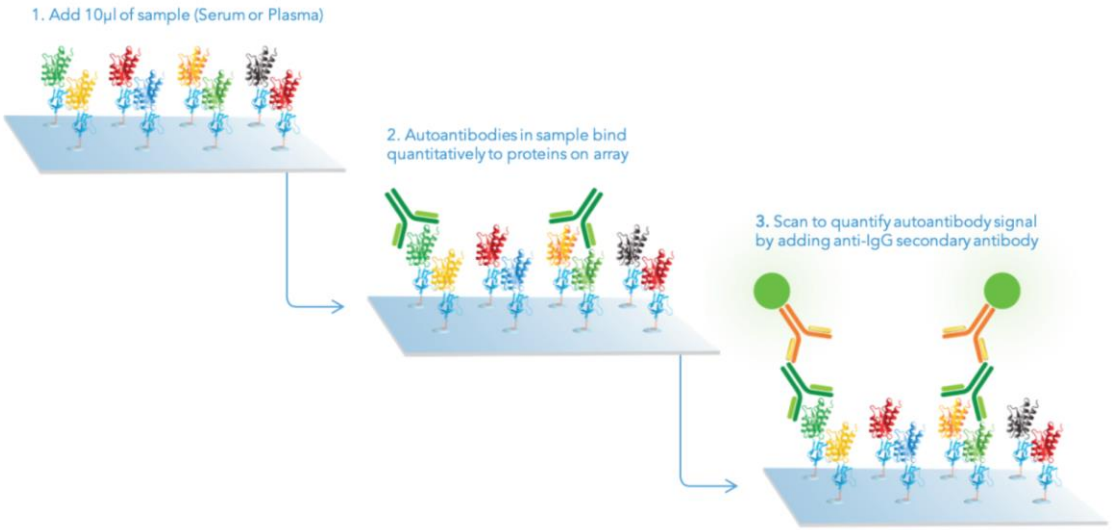


Figure 2.8. Graphic overview of the Sengenics Immunome Protein Array assay. Autoantibodies in the sample are captured by the immobilised, native recombinant human proteins. The unbound material is removed by washing and the captured autoantibodies are detected by anti-human IgG coupled to Cy-3 fluorescent dye. Available at https://sengenics.com/wp-content/uploads/2022/07/i-Ome-Protein-Array-Kit-Instruction-Manual-07-22-2022.pdf

2.4.6.1 Sample Collection

Serum samples were taken at enrolment or prior to surgery. Samples were collected from all participants in a starved state to maintain uniformity. A sample of 7 ml whole venous blood was taken into standard collection tubes and allowed to clot for 2 h. Samples were centrifuged at 3000G for 20 minutes. Serum was then carefully aspirated, divided into aliquots and stored at −80 °C (229).

2.4.6.2 Serum/Plasma dilution

Samples were placed in a shaking incubator set at 20°C to allow thawing for 30 minutes. When completely thawed, each sample was vortexed three times at full speed and spun down for 3 minutes at 13,000RPM using a micro-centrifuge. An 11.25 µL sample volume was then

pipetted into Serum Assay Buffer (SAB, sample buffer) which consists of 0.1 % v/v Triton, 0.1% w/v BSA in PBS (20°C) to a final volume of 4.5 mL and were mixed three times by pipetting. The tube was tilted during aspiration to ensure that the sera were sampled from below the lipid layer at the top. This serum/plasma dilution process was carried out in a class II biological safety cabinet.

2.4.6.3 Sample incubation

The slide was removed from the storage buffer using forceps, placed in the slide box and rack containing 200mL cold SAB and shaken on an orbital shaker at 50RPM, for 5 minutes. When the slides had completed washing, the slide was placed, array side up, in a slide hybridization chamber with individual sera, which had been diluted earlier. All slides were scanned using the barcode scanner into the relevant batch record and incubated on a horizontal shaker at 50RPM for 2 hours at 20°C.

2.4.6.4 Array Washing After Serum Binding

The protein array slide was then rinsed twice in individual "Pap jars" with 30 mL SAB, followed by 200 mL of SAB buffer in the slide-staining box for 20 minutes on the shaker at 50RPM at room temperature. All slides were transferred sequentially and in the same orientation.

2.4.6.5 Incubation with Cy3-anti IgG

After incubation, the slide was dipped in 200 mL of SAB buffer, three times for 5 minutes at 50RPM at room temperature. Excess buffer was removed by immersing the slide in 200mL of pure water for a few minutes. Slides were then dried for 2 minutes at 240G at room temperature. Slides were then stored at room temperature until scanning (preferably the same day). Hybridization signals were measured with a microarray laser scanner (Agilent Scanner) at 10µm resolution.

2.4.6.6 Data pre-processing

Scanned images were pre-processed and quality control checks were performed on the generated data using the Sengenics internal pipeline (230). Composite normalization of the data was done subsequently by using both quantile-based and intensity-based modules on the Cy3-labelled biotinylated BSA positive control probes as reported by Duarte et al (231). Autoantibody binding towards specific proteins were presented as relative fluorescent intensities (RFU) and used as inputs for downstream analysis.

2.4.6.7 Penetrance fold change analysis

The penetrance fold change (pFC) analysis compares both the frequency and strength of autoantibody signals with the intention of identifying biomarkers which are highly elevated in survivors. To achieve this, individual fold changes of survivors and non-survivors were estimated using the equation below:

$$IFC_{(protein\ A,\ sample\ X)} = RFU_{(protein\ A,\ sample\ X)} / \mu\ RFU_{(protein\ A,\ control\ group)}$$

Protein A represents each protein in the Immunome array and X represent every sample assayed in the microarray platform. The mean RFU value for each protein in the control group were used as a background threshold.

For both the survivor and non-survivor groups respectively, pFC values for each group were obtained by calculating the mean IFC of patients which passes the IFC threshold of ≥ 2. The penetrance frequencies were then calculated by estimating the number of patients (in each group) which has an IFC ≥ 2 (232). Biomarkers were further filtered based on the criteria of i) pFC of survivors ≥2, ii) % penetrance frequency of survivors ≥ 10% and iii) penetrance frequency of non-survivors ≤10%.

2.5 Quantification and Data Analysis

2.5.1 CyTOF

Files (.fcs) were processed and normalised as described and uploaded into Cytobank, populations of interest were manually gated, biaxial marker expression was performed for visualisation in Cytobank and events of interest were exported as .fcs files. CD19+ sample 'clean-up' was performed by gating on intact (103Rh+ DNA stain), non-beads (140Ce−), live (194Pt+), events followed by gating out T-cells (CD3−, 141Pr-), immature granulocytes and natural killer cells (CD16−, 209Bi) and then using a CD45+ CD19+ gate to select B cells. Pre-processing of files in order to rename files, edit panels, remove unwanted parameters and perform bead-based normalisation was also carried out using the Premessa package (233,234).

For downstream analysis, .fcs files were loaded into R (R Core Development Team, 2015). Signal intensities for each channel were arcsinh transformed with a cofactor of 5 (x_transf=asinh(x/5)). In order to facilitate differential discovery and analysis within our dataset, we employed a hybrid R-based pipeline largely based on Bioconductor packages; flowCore (235), FlowSOM (236), CATALYST (237,238), diffCYT (239).

High-resolution, unsupervised clustering and meta-clustering was performed using the FlowSOM and ConsensusClusterPlus packages respectively, the former allowed for scaling of millions of cells therefore no sub-sampling of the data was required (235,236). Visualisation of data was carried out through the CATALYST package which employs the ggplot2 R package as the graphical engine. To visualize the high-dimensional cell populations in two dimensions, the UMAP (Uniform Manifold Approximation and Projection) algorithm (240) was applied in order to represent characteristics of the annotated cell populations and identified

biomarkers. Differential cell abundance analysis was performed using generalised linear mixed models (GLMM) and of marker intensities using linear mixed models (LMM), implemented via the diffCYT package (237,239) using a false discovery rate adjustment (at 5% using the Benjamini-Hochberg method) for multiple hypothesis testing. Internally, model fitting is performed with packages lme4 and stats, and hypothesis testing with the multcomp package (237). The use of these various regression models overcomes the issues of simple statistical tests (t-test, Mann-Whitney) in assessing arbitrary experimental design and issues such as batch effects and paired experiments. As in genomic datasets, logistic regression cannot capture the over-dispersion that is present in high dimensional CyTOF data. The GLMM model offers additional variation whereby sample ID, patient pairings, and patient-patient variability can be modelled as random effects in order to model the over-dispersion proportionally (237,241).

In order to identify the main cell subsets using both B cell panels, FlowSOM was run with the parameter k ((x dim = 10 x ydim=10) = 100), defining the number of nearest neighbours, set to 100. The function will then metacluster populations into 2 through maxk (default 20) clusters (236). In order to confirm and extend our biological discovery, the clustering algorithm was modified to detect a maximum of 8 meta-clusters after assessing the initial unsupervised 20 meta-clusters for biological relevance and this was to deduce which clusters were deemed most important according to the algorithm. A more objective way of determining the optimal cluster number was by creating a delta plot as shown in the Nowicka pipeline to see at which point the clustering algorithm stabilises, this is usually beyond 8 meta-clusters. Clustering too few or too many can result in loss of biological interpretability and rare populations and incur too many insignificant populations respectively. Furthermore, selective marker clustering algorithms were run to assure us of true marker expression within

the clusters of interest. In order to further define specific B cell clusters, runs were carried out with Principal Component Analysis (PCA) pre-processing incorporating all markers on the panel (including those for T cell lineage) and then run without these markers (namely CD3, CD4, CD8) to exclude those that are not expressed on B cells and likely to add "noise" in the cluster generation process and increase the impact of the biologically relevant markers (1,238).

2.5.2 Bulk RNA sequencing differential gene expression analysis

Once raw read counts had been imported into R. We employed the DESeq2 package (242) in order to analyse the data. The typical distribution of count data is not normal; it behaves like an atypical Poisson distribution where the variance increases in a non-linear fashion as the mean count number increases, i.e., a negative binomial distribution. DESeq2 must also adjust for differences in library sizes between samples and adjusting for differences in library composition (e.g., if comparing different tissue types where different tissues have different sets of tissue specific genes). DESeq2 accounts for both of these issues. Scaling using DESeq2 accounts for composition bias (where certain genes may be more highly expressed in certain samples).

The processes that DESeq2 uses to analyse count data are as follows: estimates size factors, estimates gene-wise dispersion, fits curve to gene-wise dispersion estimates, shrinks gene-wise dispersion estimates, Generalised Linear Model (GLM) fit for each gene (the GLM model takes non-normally distributed data and fits a linear model to it, hence the "generalised" prefix) (242–245).

The outputs generated include the log2fold change expression of each gene between conditions and a p value for significance, which is adjusted using Benjamini-Hochberg

correction. We took the significant differentially expressed genes (DEGs) and mapped the Ensemble IDs to obtain gene names and Entrez IDs using annotation DBi (homosapiens) and biomaRt packages (246,247). The complex heatmap and enhanced volcano packages were used to generate heatmaps and volcano plots respectively (248,249). Gene Set Enrichment analysis (GSEA), Gene ontology pathway analysis and KEGG pathway analysis were performed using the gage, clusterProfiler and pathview packages (250–252). Broadly speaking these analyses relied on ranking all genes in the data set, identifying the rank positions of all members of the gene set in the ranked data set and then calculating an enrichment score (ES) that represents the difference between the observed rankings and that which would be expected assuming a random rank distribution.

2.5.3 Multiplex Immunofluorescence (VECTRA Polaris)

Once quantitative imagery had been acquired, regions of interest were selected using the Phenochart image viewer. Data extraction was then performed using the inForm software interface which utilised machine learning to create tissue segmentation and cell phenotyping algorithms which were then applied to the entire cohort. Cells were phenotyped by thresholding markers for positive or negative signal with the assistance of a pathologist. Differential expression analysis was performed using the Wilcoxon rank sums test. Non-parametric correlation analyses were performed for cell populations in each tissue compartment.

2.5.4 Intracellular Flow Cytometry

Data was exported as .fcs files and gated as described above. Cytokine positive cells were gated with live CD4+ or CD19+ cells as the parent population. Downstream analysis in R was carried out using the cytokine positive populations as a percentage of the parent population.

Inter-group differences were analysed for CD19+ B cells and differences between CD4: B co-culture conditions were also carried out. Statistical significance was determined using a 2-tailed non-parametric test for unpaired (Mann-Whitney U test) or paired (Wilcoxon Rank Sums test) samples and for more than 2 independent groups, Kruskal-Wallis. A p value less than 0.05 was considered significant. Multiple comparisons correction was applied using the Benjamini-Hochberg method. Graphical representation was carried using the ggpubr package in R (253).

2.5.5 ELISA

Raw data readouts from the plate reader were calculated by generating a four-parameter logistic (4-PL) curve fit. The absorbance reads and the concentrations reads from the standard curve were used in subsequent inter-group analyses. Multiple wells were used for each sample (technical replicates) with calculation of inter-well CV errors. Statistical significance and testing was carried out as for ICFC analysis, with graphical representation using the ggpubr package in R (253).

2.5.6 Proteomic Data Analysis and Biomarker Panel Generation using Machine Learning Pipelines

A schematic of the machine learning processes, and the bespoke data analysis algorithm is illustrated below in Figure 2.6. A combination of feature selection and machine learning methodologies were used to determine the optimal number of biomarkers that were able to provide the best stratification between survivors and non survivors (254). For feature selection, univariate statistical tests, random forest importance and mutual information metrics were the filter methods used to rank biomarkers. Given the degree of multi-collinearity between the biomarkers, Recursive-Feature Elimination (RFE) with Random

Forest modelling was applied to the dataset, looping across 100 unsupervised iterations using random seeds for marker reliability. The topmost stable biomarkers were used to generate biomarker panels by additively selecting the top-ranking biomarkers (top 3.75% of biomarkers, n=60) in a cumulative fashion, starting with the most stable biomarker from the RFE set (i.e. 1st, 1st + 2nd, 1st + 2nd + 3rd etc). ROC metrics were determined for each additive model and the top-performing combination taken forward as input to machine learning models. Any further addition of biomarkers did not lead to significant improvements of model performance but only further increases in computational time. To determine the biomarker panel performance, ROC, sensitivity and specificity were evaluated and the biomarker panel with the best sensitivity and specificity was deemed the optimal panel to stratify between survivors and non-survivors. For this analysis, Boosted Logistic Regression was performed under default settings using accuracy estimation methods, repeated cross-fold validation and leave-one out cross validation (LOOCV) (255).

2.5.6.1 Model Selection

To corroborate marker selection from the RFE algorithm, we used lasso regression with repeated tenfold cross-validation in the training set. This was applied using the R package glmnet. We set the elastic-net penalty, α, that bridges the gap between lasso (α = 1, the default) and ridge regression (α = 0), to 0.9 for numerical stability (256). Furthermore, we processed proteomics data using DESeq2 (v.4.0.2) software to identify differentially expressed proteins between survivors and non-survivors. A cut-off of gene-expression fold change of ≥ 2 or ≤ 0.5 and an FDR q ≤ 0.05 was applied to select the most differentially expressed proteins.

2.5.6.2 Akaike Information Criterion

We adopted a model averaging approach using the Akaike information criterion (AIC) weights (231,257) in order to estimate the in-sample prediction error and thereby the relative quality of the statistical models for a given set of data. We used an information theoretics approach to calculate the AIC for each model permutation within the top-ranking biomarkers using the glmulti and MuMIn packages in order to determine the most parsimonious model with the greatest explanatory predictive power. The AIC is a measure of how well a model fits the data relative to the other possible models given the data analysed and favours fewer parameters (258). The model with the lowest AIC is the best model approximating the outcome of interest. AIC can be expressed as:

$$AIC=-2(\text{loglikelihood})+2K,$$

K = number of model parameters and log-likelihood is a measure of model fit. In this study, as $n/K \leq 60$ for sample size n and the model with the largest value of K, we used the second-order bias correction version of the AIC (AICc):

$$AICc=-2(\text{loglikelihood})+2K+\frac{2K(K+1)}{n-K-1},$$

$$AICc=AIC+\frac{2K(K+1)}{n-K-1},$$

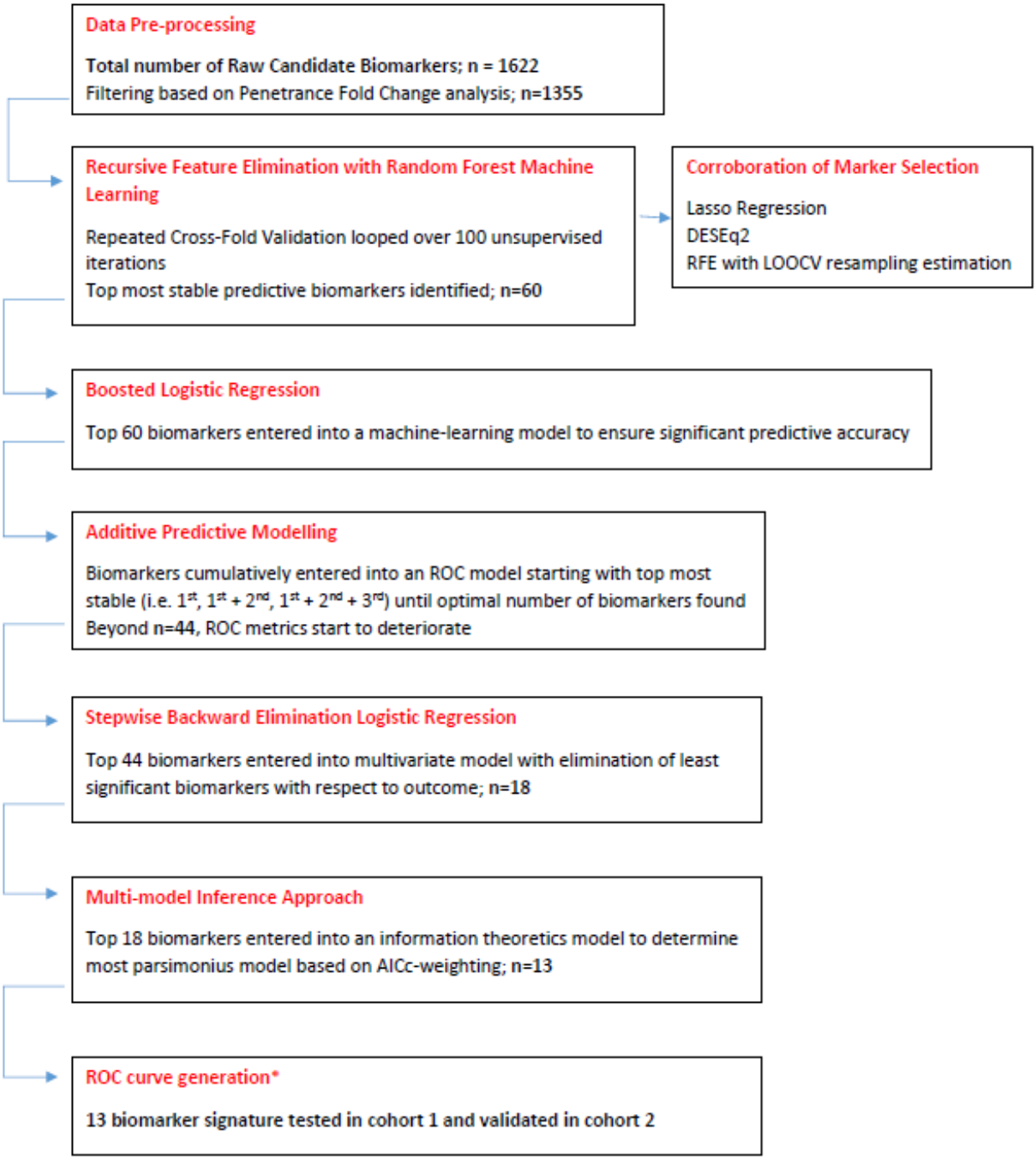
where n = sample size, K = number of model parameters and log-likelihood is a measure of model fit (257,259). From an information-theoretic perspective, the Akaike weights for a particular model can be regarded as the probability or “weight of evidence” that the model is the best model (in a Kullback-Leibler sense of minimizing the loss of information when approximating full reality by a fitted model) out of all of the models considered/fitted based on the available data set (257,258).

2.5.6.3 Pathway analysis

Biological process pathway analysis was carried out using Gene Ontology and PANTHER25. UniProt accession numbers of proteins corresponding to the biomarkers selected from RFE were uploaded to http://geneontology.org and all Homo sapiens genes in the database were used as a reference list. Fisher's exact with false discovery rate (FDR) multiple test correction was used for determining pathway significance.

2.5.6.4 Additional Statistical Analyses

All other statistical analyses were done using the RFU values of 1600+ proteins using the R platform. ROC analyses were performed using the package OptimalCutpoints (260) and plotted using ggplot2 (261). Survival analyses were performed using survminer package (262). Machine learning analyses were performed using the mlr (263), party (264), ranger (255), randomForest (265) and praznik (266) and caret (267) package. Power calculations were performed using the samplesize and sizepower packages (267–269). Data presentation in table format was implemented using the gtsummary package (270).



*All analysis steps carried out in Cohort 1, with final step (*) carried out in cohort 1 and validated in cohort 2

Figure 2.9. Data analysis algorithm for unsupervised biomarker panel selection, this was crafted with oversight and supervision from Tan Ti-Myen, Department of Bioinformatic, Sengenics Inc.

Chapter 3. Objective 1

B cell immune phenotype is predictive of subsequent post-operative outcome in non-small cell lung cancer

3.1 Surgery in NSCLC

Surgery serves an important role in the diagnosis, staging, and definitive management of non-small cell lung cancer (NSCLC). Resection is the treatment of choice for stage I and II NSCLC and an important component of the multimodality approach to stage IIIA disease (271). Standard resections include removal of the lobe involved with tumour and systematic evaluation of ipsilateral hilar and mediastinal lymph nodes. For early stage disease the evolving surgical treatment goals are aimed at decreasing morbidity and mortality through less invasive approaches including video-assisted thoracoscopic surgery and robotic approaches, and potentially decreasing the volume of lung removed for select patients with well-staged small peripheral tumours (271,272). The gold standard operation for early-stage lung cancer is lobectomy with systematic lymph node dissection.

Despite the successes of surgery and the curative potential, recurrence rates remain high. Between 30 and 55% of NSCLC patients recur following resection which is the major contributor to mortality (273,274). Complete removal of cancer needs to be performed at both the macroscopic and microscopic level. The presence of occult micro-metastatic circulating tumour cells at the time of surgery which cannot be detected by modern staging methods are likely to drive recurrence post-resection of the tumour bulk (273). Clinicopathological correlates such as those based on the TNM staging system and which include lymphovascular invasion, nodal involvement and tumour size can help to predict recurrence, but this is by no means a perfect system. Increasingly we must delve deeper into

the biology of lung cancer and use molecular biological techniques to help map the course of this aggressive disease. Immune responses within the tumour microenvironment are increasingly implicated as determining factors in tumour progression and aggressiveness (83). Immune research in NSCLC has focused predominantly around T-cell immune-biology, however, in reality the immune response is a complex interplay between the primary tumour and microenvironment, T and B cells (153). The role of the B cell in tumour survival is unclear but clearly has a function as tumour infiltration is commonly reported; B cell presence has been shown to correlate with improved survival and lower relapse rates in ovarian, cervical and NSCLC (129,139,275). The role of B cells and their antibody products is poorly defined in NSCLC and potentially could be tumour protective or tumour suppressive. Defining B cell or antibody capability offers the potential for tumour surveillance through biomarker identification and tumour treatment through monoclonal antibody production.

In order to explore the relationship between B cells and post-resection outcome in these early stage cancers, we utilised deep phenotyping techniques with a B cell specific CyTOF panel, multiplex immunofluorescence tissue imaging (VECTRA Polaris), bulk RNA sequencing, robust high dimensional display techniques, and regression models to analyse the importance of specific B cell immunophenotypes in the circulation and within the tumour microenvironment on various clinical correlates including disease-specific outcome. We explored the B cell signatures in NSCLC in publicly available genomics datasets to corroborate our findings.

3.2 Early Disease Surgical NSCLC Cohort

Our early-stage NSCLC cohort (stage I-IIIa) consisted of 56 patients; full demographic data is shown below in Figure 3.1. Median follow-up in these patients was 3 years. Median age was 74 years, with a preponderance of males (66%). Most cancers removed were

adenocarcinomas (54%), with a relatively stage III preponderance (48%). Fourteen patients (25%) in this cohort died within the follow-up period and twelve patients succumbed to disease recurrence (21%). Median overall and disease-free survival was 809 and 740 days respectively.

Characteristic	N = 56[1]
Age	74 (65, 75)
Gender	
F	19 (34%)
M	37 (66%)
Histology	
Adenocarcinoma	30 (54%)
Adenosquamous	1 (1.8%)
Large Cell Carcinoma	3 (5.4%)
Squamous Cell Carcinoma	22 (39%)
IASLC stage	
I	11 (20%)
II	18 (32%)
III	27 (48%)
Lymphovascular Invasion	35 (62%)
Mediastinal Pleural Invasion	3 (5.4%)
Visceral Pleural Invasion	22 (39%)
Adjuvant Chemotherapy	10 (18%)
Mortality	14 (25%)
Recurrence	12 (21%)
OS (days)	809 (632, 954)
DFS (days)	740 (645, 902)
[1] Median (IQR); n (%)	

Figure 3.1. A table to illustrate the overall demographic, oncological and outcome data in the cohort (n=56).

We examined the risk factors which emerged as significant independent predictors of overall and disease-free survival on multivariate testing (Figure 3.2). Squamous cell carcinoma was

associated with reduced risk of death (HR 0.17, p-0.014) and male gender was associated with an increased risk of death (HR 5.45, p=0.034). In terms of recurrence, advanced stage (III) was associated with an increased risk (HR 25.6, p=0.016) and adjuvant chemotherapy was associated with a reduced risk (HR 0.04, p=0.018).

	Overall Survival			Disease Free Survival		
Characteristic	**HR**[1]	**95% CI**[1]	**p-value**	**HR**[1]	**95% CI**[1]	**p-value**
Histology						
Adenocarcinoma	—	—		—	—	
Squamous Cell Carcinoma	0.17	0.04, 0.70	**0.014**	0.37	0.09, 1.45	0.154
Gender						
F	—	—				
M	5.45	1.13, 26.2	**0.034**			
IASLC_Stage						
I	—	—		—	—	
II	0.34	0.03, 3.71	0.373	0.79	0.11, 5.67	0.816
III	4.05	0.76, 21.6	0.101	25.6	1.84, 358	**0.016**
Adjuvant_Chemotherapy						
No				—	—	
Yes				0.04	0.00, 0.57	**0.018**

[1] HR = Hazard Ratio, CI = Confidence Interval

Figure 3.2. A backward elimination stepwise multivariate regression model to illustrate the significant independent predictors of outcome in our cohort. Overall survival/death (left hand column) and disease-free survival/recurrence (right hand panel) are both shown. Cox proportional hazards modelling was employed, $p<0.05$ was deemed significant.

3.3 Deep B cell phenotype in NSCLC

We performed large-scale mass cytometry analysis of paired peripheral blood and tumour samples from patients with early-stage NSCLC (IASLC stage I-IIIa). The following schematic illustrates the analyses undertaken.

Surrogacy (n=48)	Circulating versus Intra-tumoural B cell profile	
Histology	Circulation (Peripheral Blood) (n=52)	Intratumoural (n=48)
Stage	Circulation (Peripheral Blood) (n=52)	Intratumoural (n=48)
Lymphovascular Invasion	Circulation (Peripheral Blood) (n=52)	Intratumoural (n=48)
Recurrence	Circulation (Peripheral Blood) (n=52)	Intratumoural (n=48)

Table 3.1. A table to illustrate mass cytometric analyses undertaken.

3.3.1 Antibody Panels

Cells were stained with one antibody panel (Base B cell panel) created for this study (Table 2.1). There is a paucity of data in the literature demonstrating deep B cell phenotyping. The first panel was designed to detect the expression of B cells at various stages of maturation (activated, transitional, marginal zone, follicular, germinal centre, class-switched and plasma) as well as rarer B cell populations such B regulatory (Breg) cells. All patients were run on this first “base” B cell panel. This panel included markers for natural killer cells, T cells and granulocytes.

3.3.2 B cell Milieu in Early-Stage NSCLC

In order to generate a comprehensive picture of the B cell landscape we created two-dimensional maps of the data, using dimensionality reduction algorithms (UMAP and diffusion maps). UMAP was used in preference to t-SNE owing to faster run times, especially with large perplexity hyper-parameters and better preservation of the global cellular architecture with competitive visual quality (240). The arrangement of cells along their development path, with bifurcations along points of differentiation was achieved using diffusion maps. Distinct cellular phenotypes were determined following application of the FlowSOM clustering algorithm which employs a Self-Organising Map to order determine how

markers are behaving on cells in high dimensional space thus creating well partitioned cellular subsets (236).

3.3.2.1 Deep B cell phenotyping reveals clear differences between the peripheral circulating and tumour-infiltrating B cell milieu in NSCLC

CD19+ cells were taken forward into all further downstream analyses. In order to map cell phenotypes, FlowSOM clustering was performed, and expression of B cell clusters were visualised as a heatmap (Figure 3.3) with heterogeneity in marker level displayed at single cell level using UMAP.

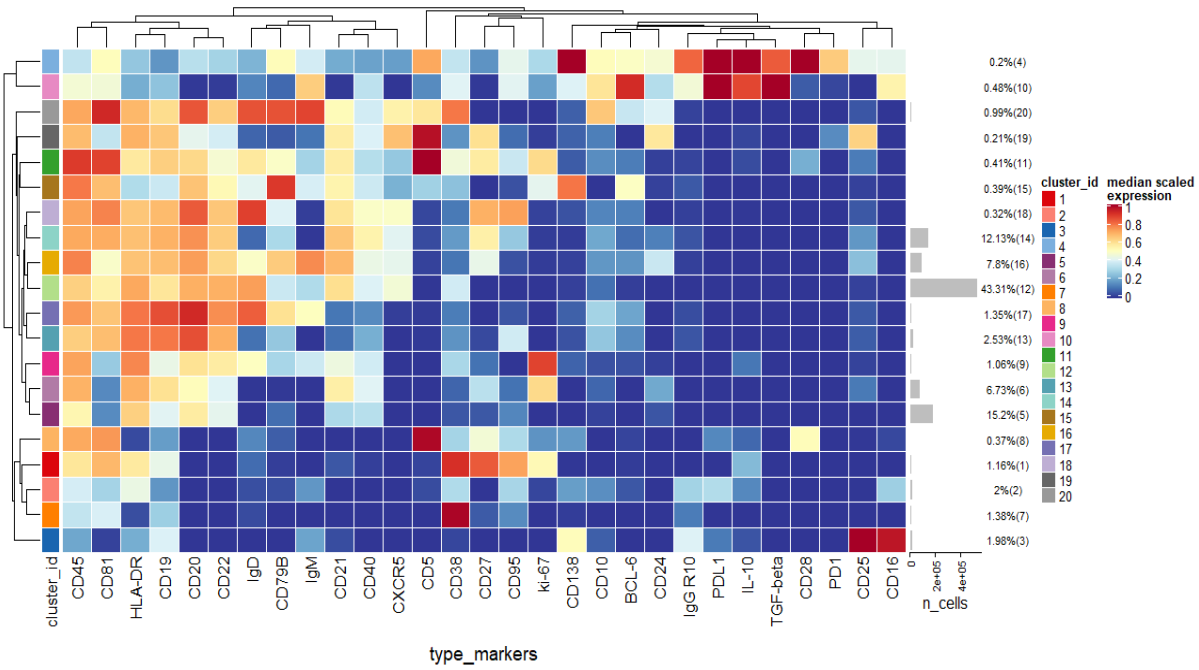


Figure 3.3. A heatmap demonstrating predominant 20 clusters in the B cell repertoire, as seen following FlowSOM clustering.

Table 3.2 below has enumerated all 20 clusters and contains their likely phenotype based on the surface marker expression in the heatmap (Figure 3.3).

Cluster Number	Population
1	ki67^{hi} IL10+ CD27+ CD38+ plasmablasts
2	immature PDL1+ IL10+ Breg
3	Ig- Plasma cells characterised by low/dim CD19 expression and CD138 positivity
4	Insignificant cluster (0.2%) cannot be identified
5	Ig- CD38^{lo} Plasma
6	ki67^{hi} CD24 CD25 CD27 B10 Breg
7	Ig- Plasma cells characterised by low/dim CD19 expression and CD138 positivity
8	immature PDL1+ IL10+ CD5^{hi} Breg
9	ki67^{hi} IL10+ CD27+ CD38+ plasmablasts
10	Insignificant cluster (0.48%) cannot be identified
11	CD5^{hi} CD10+ CD27^{hi} CD38+ Transitional
12	Follicular (CD20^{hi} CD22^{hi} IgD+ IgM^{lo})
13	Double negative memory B cells (CD27^{lo} IgD- IgM-)
14	Fully affinity matured, class-switched B cells (CD27^{hi} IgD- IgM-)
15	Ig- Plasma cells characterised by low/dim CD19 expression and CD138 positivity
16	Activated (IgM+ IgD+ CD25+ CD27+)
17	Germinal Centre
18	Early memory
19	PD1+ CD5^{hi} CD25^{hi} IL10+ Breg
20	Transitional (IgM^{hi} IgD^{hi} CD24^{hi} CD38^{hi} CD10^{hi} CD5+)

Table 3.2. A table to illustrate the likely B cell populations in the circulation and tumour microenvironment in early-stage non-small cell lung cancer patients. Clusters 4 and 10 were not identifiable and very sparsely represented in the overall population.

Unsupervised multi-dimensional scaling (principal component analysis) shows the broad differences in B cell make=up between blood and tumour samples. There is clear separation of these samples indicating likely differences in immunophenotype expression in the different environments (Figure 3.4).

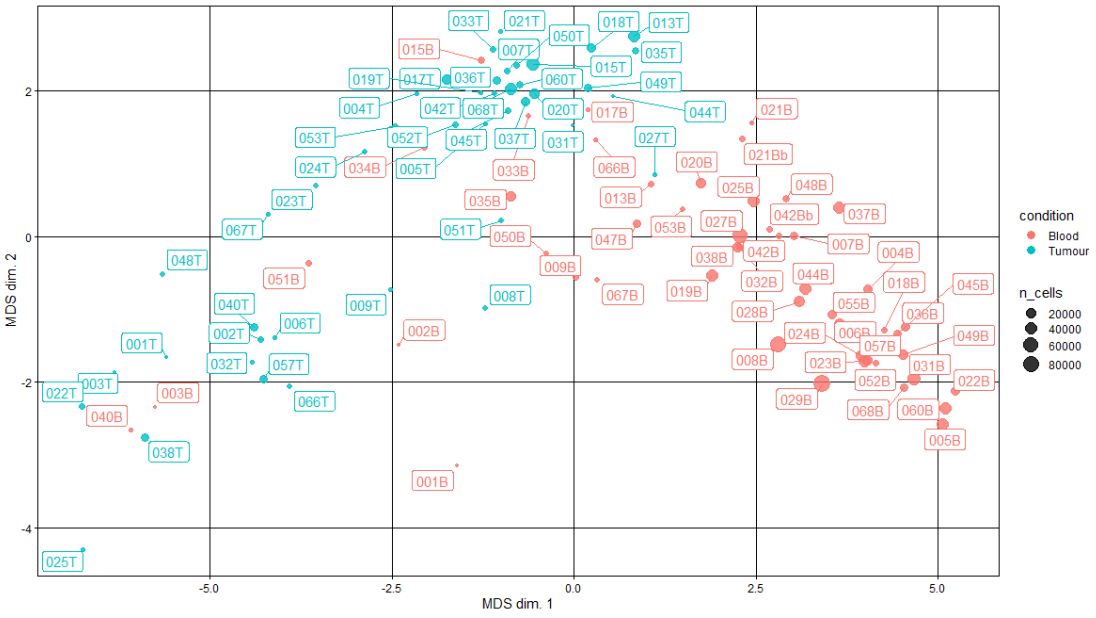


Figure 3.4. A multi-dimensional scaled plot to display samples according to type of tissue and to determine if this shows any differences based on principal component analysis.

We defined 20 distinct B cell clusters at various stages of maturation. The mostly frequently observed cluster was of the follicular B cell lineage (clusters 12, 43.41% of the total population (Figure 3.3) characterised by high expression of CD20, CD22 and IgD and with lower levels of IgM. Activated (IgM+ IgD+ CD25+ CD27+) (cluster 16) and Transitional B cell (IgM^{hi} IgD^{hi} $CD24^{hi}$ $CD38^{hi}$ $CD10^{hi}$ CD5+) clusters (cluster 20) comprised 7.8% and 0.99% of the total B cell population respectively. Ig negative Plasma cells characterised by low/dim CD19 expression and CD138 positivity (clusters 3, 7 and 15) comprised 1.98%, 1.38% and 0.39% of the total B cell population respectively. Memory B cells were observed at various stages of maturation: early memory B cells (cluster 18, 0.32%), fully affinity matured, class-switched B cells ($CD27^{hi}$ IgD- IgM-) (cluster 14, 12.13%) and double negative memory B cells ($CD27^{lo}$ IgD- IgM-) (cluster 13, 2.53%). Several B regulatory cell clusters were also identified, characterised using surface markers such as CD5, CD24, CD25, CD27, CD38, CD1d, TIM-1, PD1, PDL-1, TGF-beta and intracellular cytokine expression of IL-10, were observed to varying frequencies (0.21%-2%). Clusters 1 and 9 are both likely to represent $ki67^{hi}$ IL10+ CD27+ CD38+ plasmablasts. Clusters 2 and 8 represent immature PDL1+ IL10+ Bregs with cluster 8 also being $CD5^{hi}$. Cluster 19 is a PD1+ $CD5^{hi}$ $CD25^{hi}$ IL10+ Breg.

All B cells were unstimulated and thus IL-10 expression was representative of the *in vivo* immune milieu of NSCLC patients and healthy donors.

We performed comparative analyses between the paired blood and tumour samples from each patient (n=48) (Figure 3.5). On visual inspection, there are clear differences in cluster expression between the environments.

In blood, there is higher expression of the following.

Cluster 12 – Follicular (light green)

Cluster 13 –Double negative Memory (Teal)

Cluster 14 – Class-switched Memory (Aquamarine)

Cluster 16 – Activated (Mustard)

Cluster 17 – Germinal Centre (Lilac)

Cluster 19 – PD1+ Breg (Dark Grey)

Cluster 20 – Transitional (Light Grey)

In tumour, there is higher expression of the following.

Cluster 3 – CD138+ Plasma (Royal Blue)

Cluster 7 – Ig- $CD38^{hi}$ Plasma (Orange)

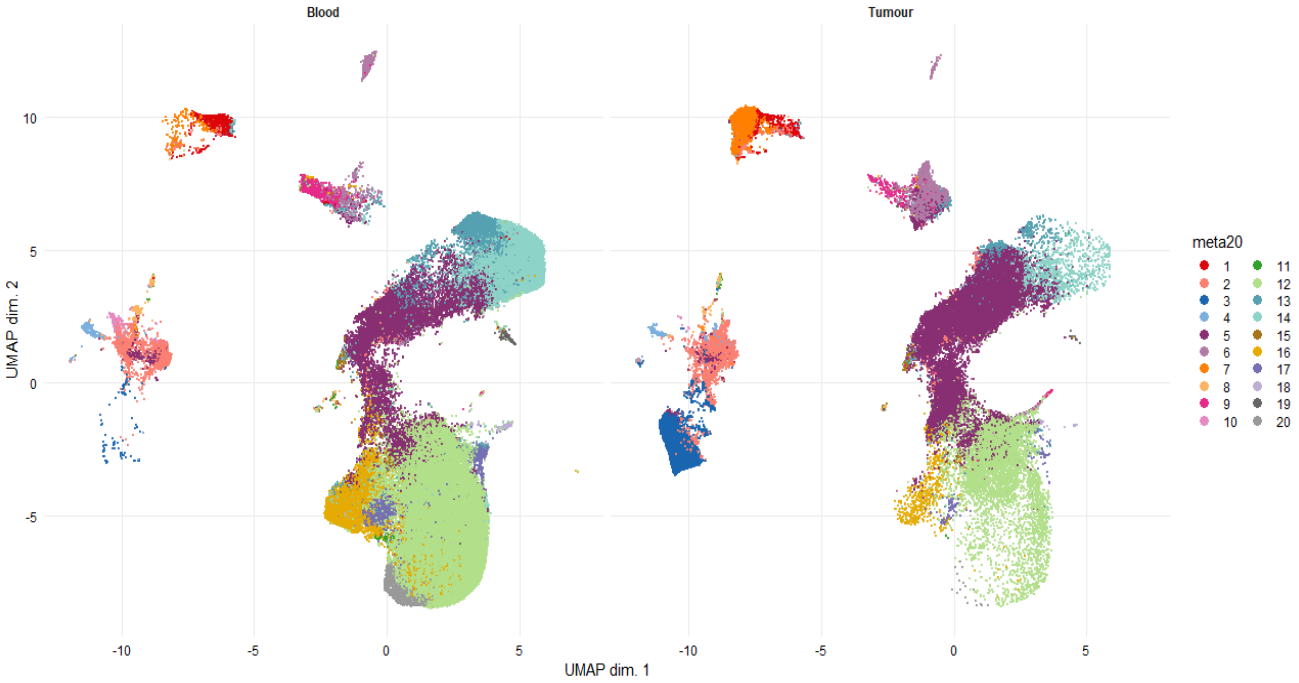


Figure 3.5. A dimensionality reduction method known as UMAP, employed to display high-dimensional clusters from Figure 3.3 in two dimensions. This map segregates populations in accordance with phenotypic similarities and abundance. Comparative cluster expression between blood and tumour is shown.

3.3.2.1.1 Differential Abundance Analysis

We performed a differential abundance (DA) analysis of the defined cell populations, reporting on all B cell clusters in the population. This method compares the proportions of cell types between the two clinical conditions and aims to highlight the populations that are present at significantly different ratios. In order to gain power to detect differences between conditions, we utilised a mixed model to model the response and patients were treated as a random effect thus formally accounting for patient to patient variability as described by Nowicka et al (237,276). DA analysis of the overall cell population identified eight clusters as significantly differentially abundant between the two environments (Figure 3.6). These are as follows:

Cluster	Predominant Abundance	P value
3 - CD138+ Plasma	Tumour	$7.7*10^{-6}$
5 - Ig- CD38^{lo} Plasma	Tumour	$1.4*10^{-9}$
7 - Ig- CD38^{hi} Plasma	Tumour	$9.2*10^{-21}$
12 - Follicular	Blood	$5.7*10^{-11}$
13 - Double negative Memory	Blood	$1.2*10^{-3}$
14 - Class-switched Memory	Blood	$2.0*10^{-12}$
16 - Activated	Blood	$3.3*10^{-14}$
17 - Germinal Centre	Blood	$9.7*10^{-17}$

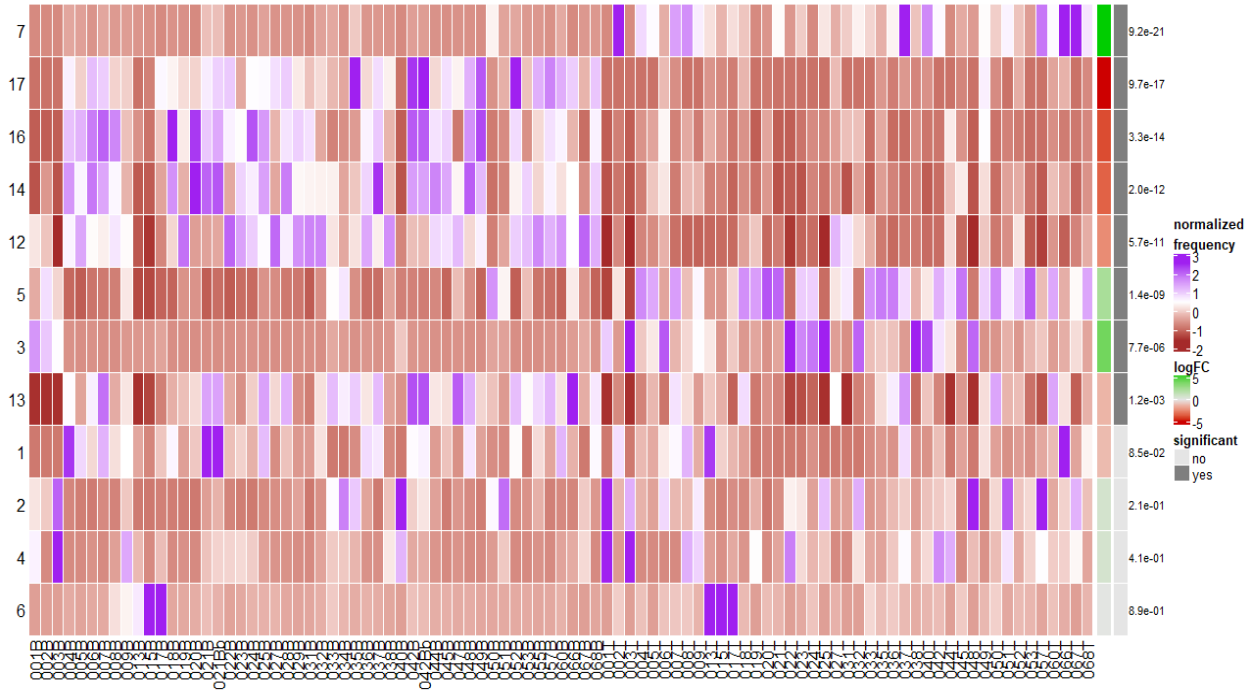


Figure 3.6. A differential abundance heatmap demonstrating the significantly abundant clusters in accordance with tissue compartment. Tissue type is shown along the bottom x axis (B – blood, T – tumour). The grey bars on the right-hand side indicate a $p<0.05$ accounting for multiple correction testing with Benjamini Hochberg. The log fold change is with respect to tumour. The table above summarises the differentially abundant populations according to tissue compartment along with corresponding p-value.

Figure 3.7 below is a graphical representation of the differing abundances of each cluster described in table 3.2 across the two tissue compartments (blood and tumour).

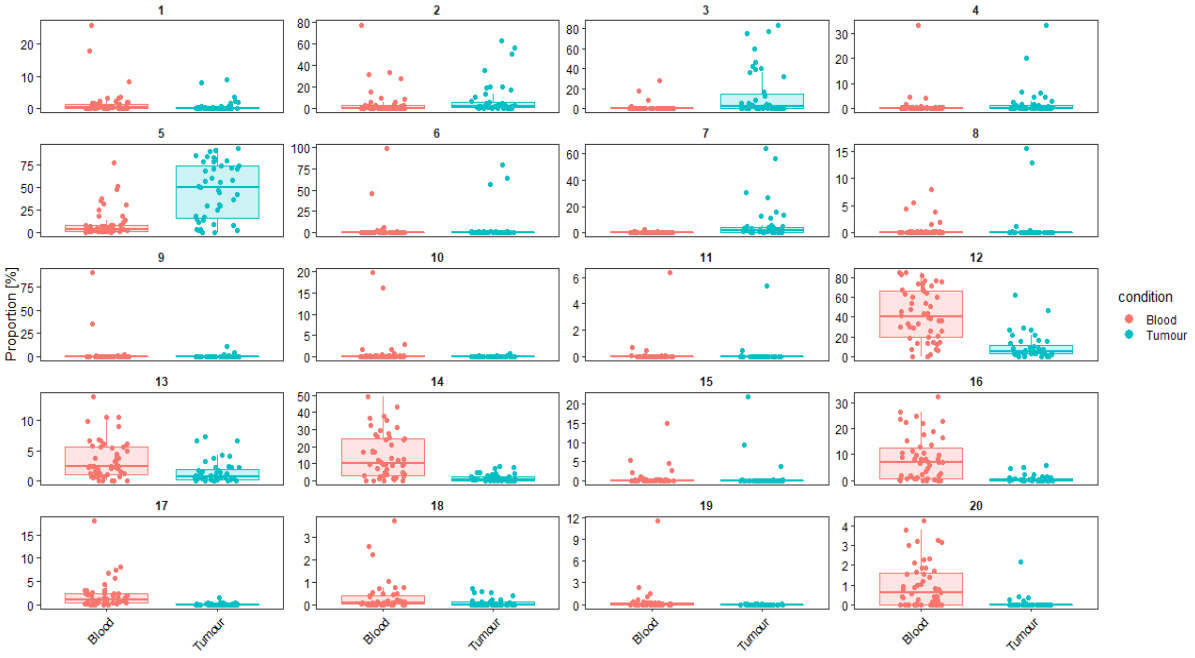


Figure 3.7. Boxplots to illustrate the compartmental difference in cluster abundance. Those that are significantly differentially abundant on the differential abundance heatmap (Figure 3.6) are shown above – clusters 3, 5, 7, 12, 13, 14, 16 and 17. Clusters 18 and 20 although are visually more abundant in the circulation, these did not fall out as significant on differential abundance testing.

3.3.2.1.2 Interrogation of a Public Mass Cytometry Dataset

Lavin et al [PMID 28475900] demonstrated significantly different innate cell compartments in lung adenocarcinoma between healthy tissue and cancerous tissues (277). We used the data from this group and gated for CD19+ B cells and performed an unsupervised comparative principal component analysis between healthy lung tissue, blood, and tumour samples from stage I lung adenocarcinoma patients. The MDS plot (Figure 3.7) illustrates clear separation between the three tissue types suggesting differences in the CD19+ B cell compartment between a) healthy and cancer tissue and b) blood and tumour of NSCLC patients. This supports the gross differences we observed in our dataset. Owing to the significant

heterogeneity between the antibody panel used by this group and ours, we were unable to drill deeper into the observed phenotypes. However, differential abundance analysis revealed several clusters which were significantly higher in the blood than in the tumour microenvironment.

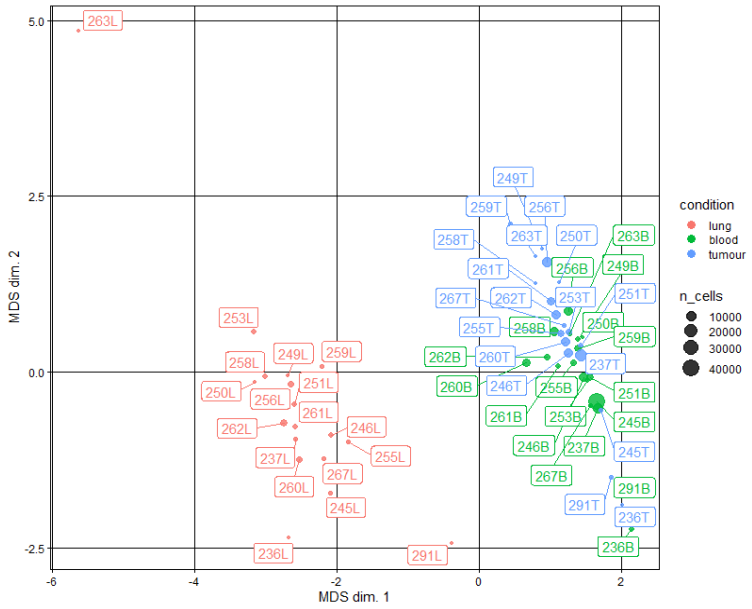


Figure 3.8. A Multi-dimensional scaling plot demonstrating compartmental differences in a public dataset based on principal component analysis.

3.3.2.2 Histology

We examined the differences in the B cell profile between different histological subtypes (adenocarcinoma and squamous cell carcinoma) in the blood and in the tumour. On assessment of the baseline characteristics of these patients, we noted no significant differences in demographics, onco-pathological factors, and outcome between these two histological subtypes. Although there was a preponderance of males with squamous cell carcinoma (82%), this difference was not significant after correction for multiple testing (p=0.362).

3.3.2.2.1 Blood

Comparative expression of B cell populations between the histological subtypes revealed visual differences albeit subtle. Cluster 3, 11 and 18 were deemed the only clusters significant on differential abundance analysis however this was without considering patient to patient variability indicating that a small number of patients were driving these differences. All three clusters; transitional cell population (3), $CD138^{hi}$ plasma cell population (11) and $CD5^{hi}$ $CD25^{hi}$ $PD1^{int}$ Breg (18) were more abundant in adenocarcinoma patients.

3.3.2.2.2 Tumour

Despite visually striking differences in cluster 9, CD138+ plasma cell population; higher in adenocarcinoma, this was not significant on differential abundance testing. Of note, cluster 11, CD5+ CD25+ PD1+ Breg was higher intra-tumourally in squamous cell carcinoma, the inverse relationship to what was seen in the blood (p=NS).

3.3.2.3 Stage

We examined the differences in the B cell profile between the different stages of cancer in the blood and in the tumour. On assessment of the baseline characteristics of these patients, we noted a significantly higher mortality rate (p=0.045) as well as a lower median disease-free survival (671 days) in stage III cancers (p=0.041) although these differences were not significant after correction for multiple testing.

3.3.2.3.1 Blood

No B cell populations were deemed to be significantly differentially abundant between cancer stages in the circulation.

3.3.2.3.2 Tumour

Significantly higher levels of intratumoural IgG+ CD138+ terminally differentiated plasma cells (cluster 15) were found in higher stage tumours (2 and 3) with respect to stage 1 tumours (p=0.016) (Figure 3.8).

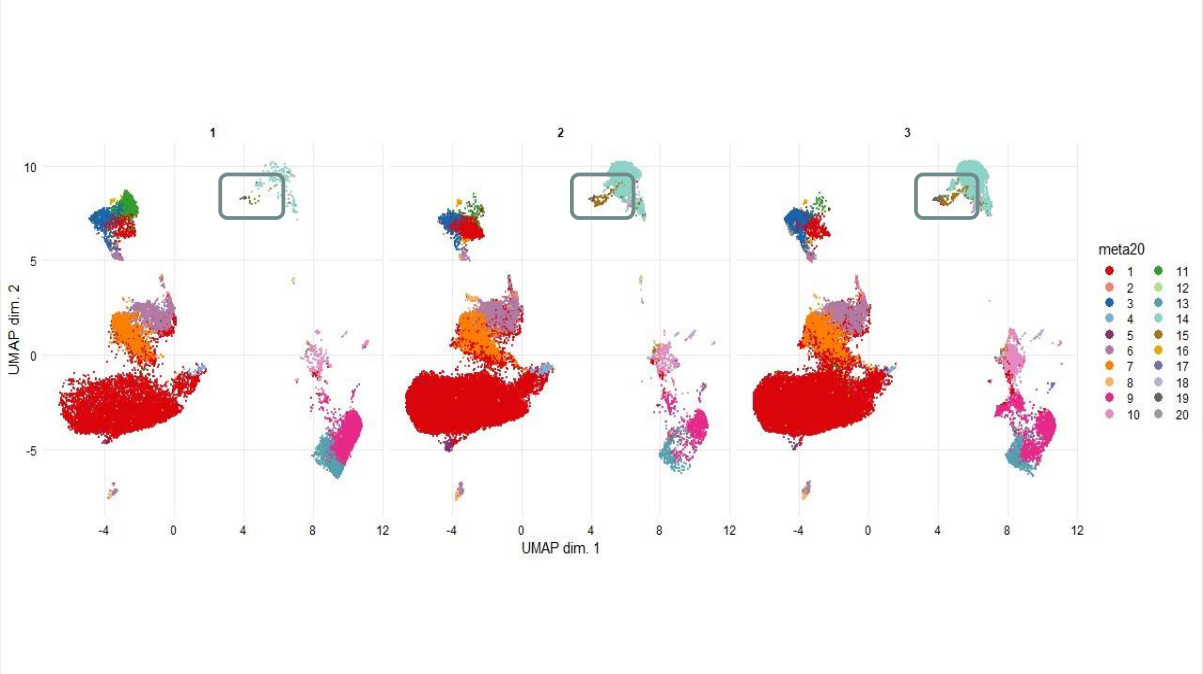


Figure 3.9. Comparative UMAP demonstrating higher visual intra-tumoural expression of cluster 15 (CD138+, brown cluster in the grey box) in higher stage tumours. Each number 1, 2, 3 along the top row corresponds to stage of cancer. In early-stage tumours (1), there is low visual presence of the brown cluster 15, as demonstrated by the grey box. However, as stage increases, the presence of this brown cluster increases. These cells represent a terminally differentiated plasma cell population that is Ig producing, i.e., an effector population.

3.3.2.4 Lymphovascular Invasion

We examined the differences in the B cell profile between the tumours that displayed pathological evidence of lymphovascular invasion post-resection. On assessment of the baseline characteristics of these patients, we noted a significantly higher rate of visceral pleural invasion (p=0.016) in cancers with lymphovascular invasion, although these differences were not significant after correction for multiple testing.

3.3.2.4.1 Blood

No B cell populations were deemed to be significantly differentially abundant between the two pathological groups in the circulation.

3.3.2.4.2 Tumour

No B cell populations were deemed to be significantly differentially abundant between the two pathological groups in the tumour.

3.3.2.5 Recurrence

We examined the differences in the B cell profile between those patients that recurred and those that did not. On assessment of the baseline characteristics of these patients, we noted a significantly higher mortality rate (58% versus 16%) in those patients who developed post-operative recurrence (p=0.006).

3.3.2.5.1 Blood

Significantly higher levels of circulating ki67+ CD27^{hi} CD38^{hi} CD95^{hi} IL10^{int} plasmablasts were noted in recurrence patients (p=0.02) (Figure 3.9).

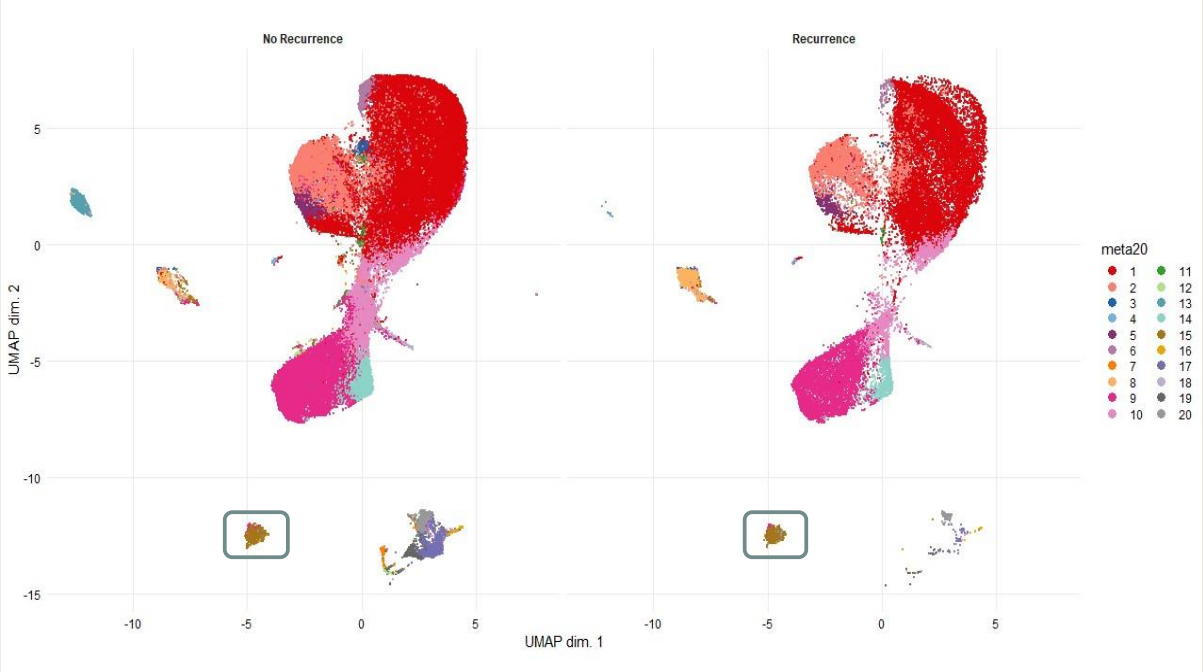


Figure 3.10. Comparative UMAP demonstrating higher visual circulating expression of cluster 15 (IL10+ plasmablasts, brown cluster in the grey box) in recurrence patients. Although the visual difference is not particularly striking, the differential abundance based on overall abundance was significant.

3.3.2.5.2 Tumour

Significantly higher levels of intratumoural CD138+ IgG+ CD25^{hi} early plasma cell/plasmablasts were noted in non-recurrence patients (p=0.039) (Figure 3.11). This population was weakly PDL1+. Overall expression of CD138 and IgG (plasma cell markers) within the tumour were also noted to be higher in non-recurrence patients. Of note, the following populations were visually more differentially abundant in non-recurrence patients, albeit not significant (Figure 3.11).

Cluster 9 – CD138+ CD25^{hi} IgG+ PDL1- IL10- Plasma cell (pink cluster)

Cluster 10 – CD138int IgG^{int} PDL1^{int} IL10^{int} Plasma cell/regulatory suppressive (light pink cluster)

Cluster 11 – ki67^{hi} CD5+ CD10+ CD24^{hi} CD25+ CD27+ PD1+ Transitional (green cluster)

Cluster 13 – CD138^{hi} IgG^{hi} PDL1l^{o} IL10- Plasma cell (aquamarine cluster)

Cluster 14 – CD19- CD38^{hi} IgG+ Early Plasma cell (teal cluster)

Of the above populations, only cluster 13 was deemed to be significantly more abundant in non-recurrence patients (Figure 3.11, grey box) (p=0.03) on differential abundance analysis. However, there is visual attenuation of all clusters above in the recurrence cohort. Cluster 14 shows early features of plasma cell differentiation with CD19, high CD38 expression and early IgG expression. Cluster 9 then shows CD138 expression with higher levels being expressed by cluster 13 which is a true effector terminally differentiated plasma cell population. Cluster 13

also shows early PDL1 expression which represents likely early transitioning to a regulatory plasma cell phenotype; cluster 10 displays higher levels of PDL1 as well as producing IL10 becoming a suppressive plasma cell population. Cluster 11 represents a very early transitional cell population which has not yet become suppressive (IL10-, CD10+, PD1+). This focused assessment of the TME reinforces the phenotypic spectrum of infiltrating early/late plasma cells, with effector populations conferring less disease aggression and better outcomes and suppressive populations behaving like true B regulatory cells (Bregs) as in the circulation and conferring poorer outcomes.

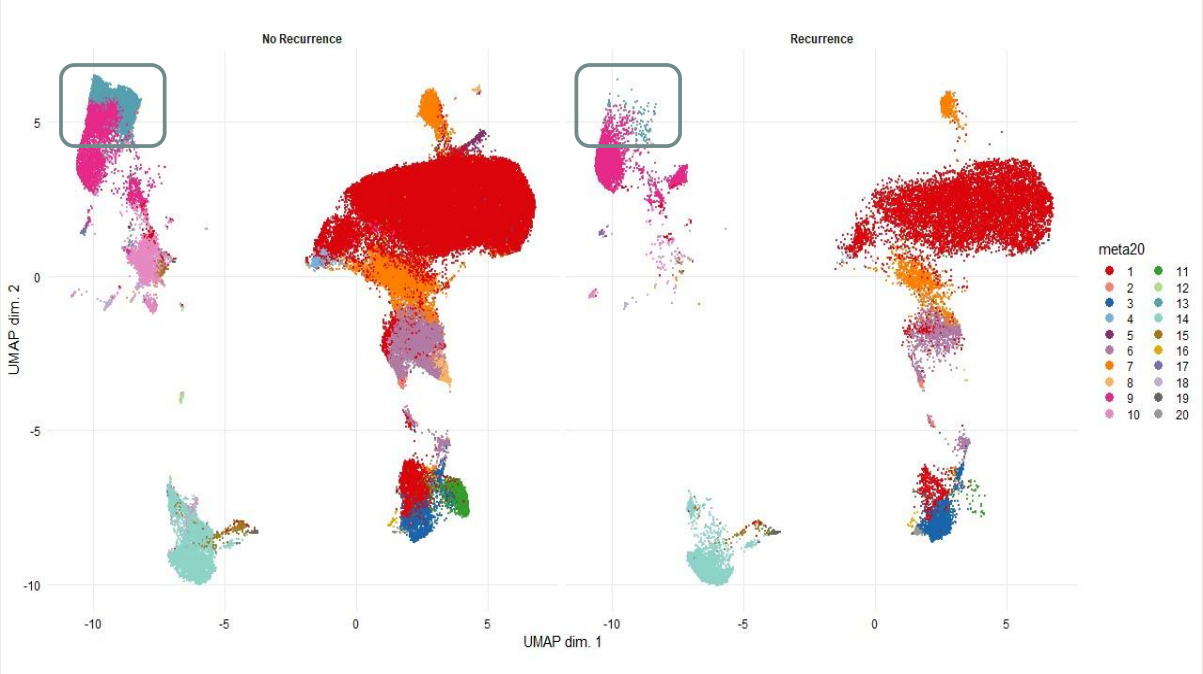


Figure 3.11. Comparative UMAP demonstrating higher visual intra-tumoural expression of cluster 13 (CD138+, aquamarine cluster in the grey box) in non-recurrence patients. There is visually higher expression of other clusters which were not deemed significant in non-recurrence patients: Cluster 14 shows early features of plasma cell differentiation with CD19, high CD38 expression and early IgG expression. Cluster 9 then shows CD138 expression with higher levels being expressed by cluster 13 which is a true effector terminally differentiated plasma cell population. Cluster 13 also shows early PDL1 expression which represents likely early transitioning to a regulatory plasma cell phenotype; cluster 10 displays higher levels of PDL1 as well as producing IL10 becoming a suppressive plasma cell population. Cluster 11 represents a very early transitional cell population which has not yet become suppressive (IL10-, CD10+, PD1+).

3.4 Quantitative Multispectral Tissue Imaging (VECTRA Polaris)

Multiplexed immunofluorescent assays for CD4, CD8, CD20, CD138, IL-10 and BCL-6 were employed to visualise changes in immune infiltrate composition across NSCLC and with reference to those patients who developed post-operative recurrence. The proportion of various cell phenotypes (CD4+ T cells, CD8+ T cells, CD20+ B cells, CD20+ BCL6+ B cells, CD138+ Plasma cells, CD138+ IL10+ suppressive Plasma cells and CD20+ IL10+ B regulatory cells) in each slice was calculated (% = cell phenotype count/total number of cells in each slice x 100%). A significantly higher proportion of suppressive B cells (CD138+ IL10+ and CD20+ IL10+) infiltrate the tumour stroma as opposed to the tumour nest (p<0.0001) (Figure 3.12).

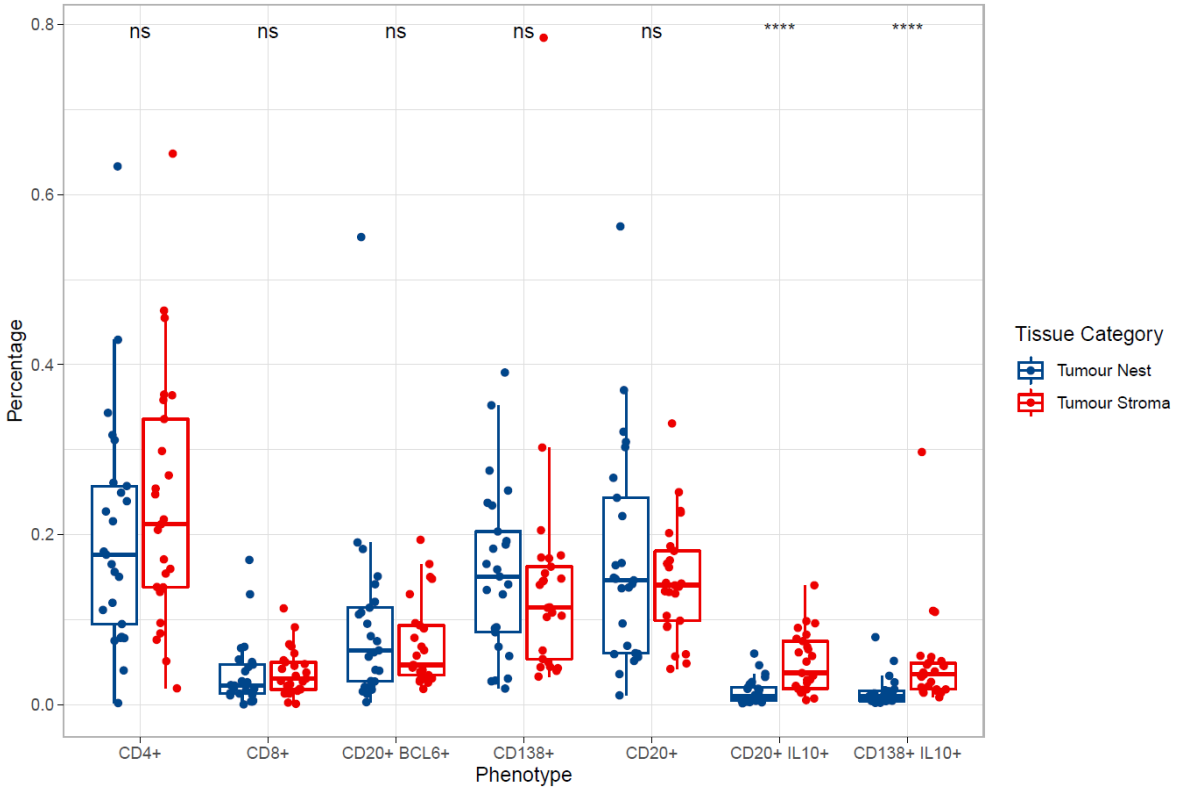


Figure 3.12. Boxplot demonstrating the differences in phenotype infiltration across tumour compartments. Pairwise Wilcoxon Rank Sums Test was used to determine statistical significance. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

Visual representation of these differences is shown below in Figure 3.13. Multispectral Immunofluorescence tissue slices illustrate the preponderance of suppressive CD138+ and CD20+ populations at the tumour stroma surrounding the nest.

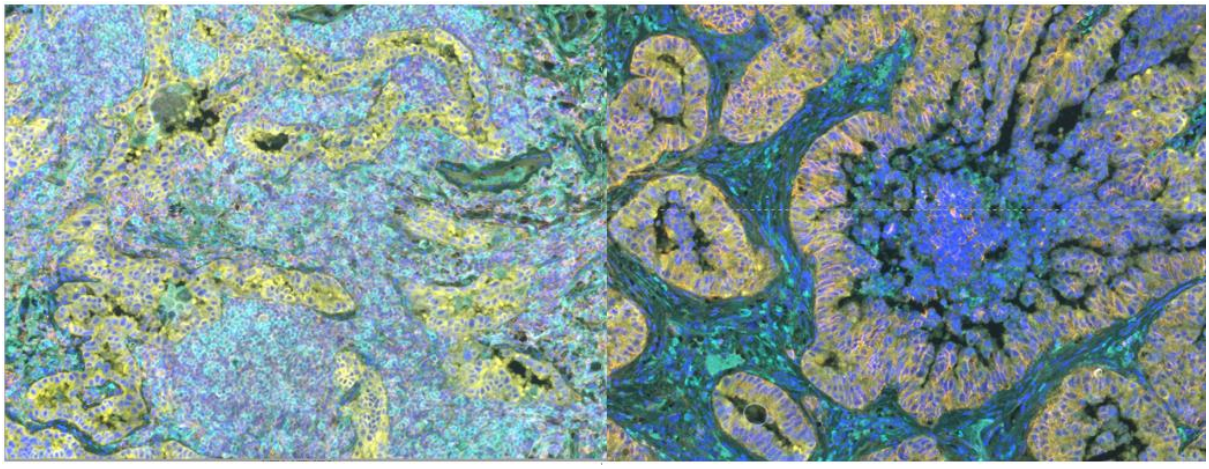


Figure 3.13. Non-small cell lung cancer tissue slices illustrating B cell geospatial distribution. The left-hand panel shows the populations lighting up as yellow, are IL-10 producing cells which are CD20+. They aggregate to the peripheral stroma of the tumour. The right-hand panel similarly shows yellow populations which represent CD138+ IL10+ cells collecting around the tumour nest in the stroma.

There were not any overall or compartmental differences in phenotype when stratified according to histology, stage, presence of lymphovascular invasion, presence of visceral pleural invasion or mortality. Similarly, despite a slight increase in the presence of suppressive B cells, both CD138+ and CD20+, in those patients who recurred, this was not deemed significant. Univariate and Multivariate testing by cox proportional hazards did not identify any significant independent predictors of recurrence or death. Correlation analyses did identify a strong positive correlation between CD20+ and CD138+ cells in both the tumour nest and stroma (Figure 3.12). A weaker positive correlation in the tumour stroma was seen between CD4+ and CD8+ T cells ($r=0.6$, $p<0.05$).

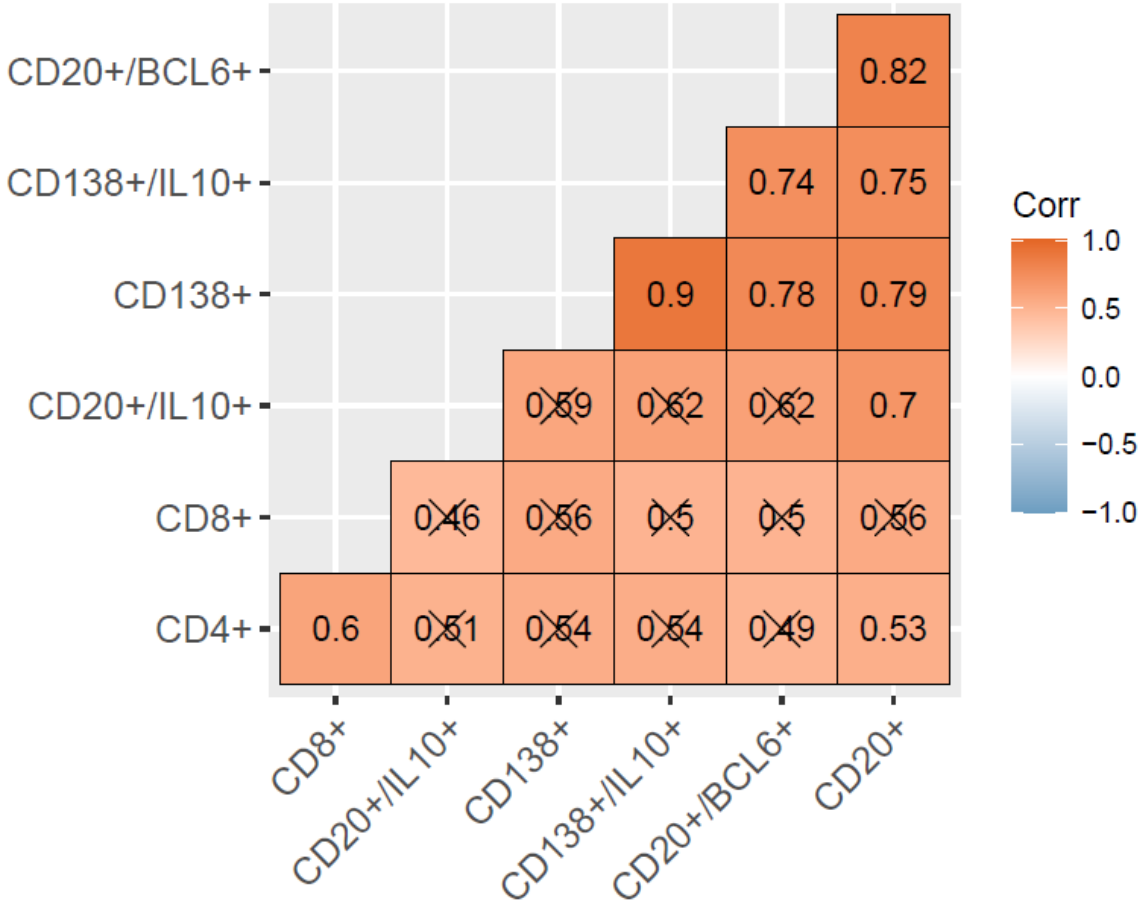


Figure 3.14. Correlation analysis for cellular phenotypes detected in the tumour stroma. This matrix indicates correlation between the cellular phenotypes studied along the x and y axes. The numbers in the boxes indicate correlation strength. They range from -1 to +1. A positive number indicates a positive correlation whereas a negative number, a negative correlation. Those numbers that are not crossed out indicate a significant correlation i.e. a p value of <0.05. The correlation between CD138+ and CD138+ IL10+ cells in the tumour stroma is very strong at 0.9. Similarly CD138+ cells associate strongly with CD2+ BCL6+ cells albeit less strongly with an R value of 0.78.

3.5 Functional Genomics Analysis

Numerous exploratory strategies have been taken to investigate the immune milieu in cancer, as such as we performed bulk RNA sequencing on a subset of our tumour samples (n=27) as well as interrogating public RNAseq transcriptomic datasets in NSCLC as a way of exploring any predictive biomarkers for outcome and validating the single cell findings from our high-dimensional data (section 3.3).

Owing to the delicate nature of the NSCLC tumour tissue and PBMCs, RNA extraction was sensitive, and yield was low despite good RNA Integrity Number (RIN) values for most samples. Thus, we proceeded with a stranded analysis on the 3' QuantSeq mRNA kit from Lexogen on Illumina NextSeq platform [FWD kit] (section 2.4.2). This allowed us to perform a genome wide analysis of gene expression, which proved to be cost efficient alternative to microarray and standard RNA-sequencing. This method uses the first 50 base pairs only from the polyA tail so cannot use information for mapping or mutations only gene expression.

3.5.1 Transcriptomics Cohort

The cohort consisted of 27 tumour samples taken from the initial NSCLC cohort. Of these, 14 were adenocarcinoma, 12 squamous cell carcinomas and 1 was a large cell carcinoma. Eight patients recurred. Eighteen patients displayed lymphovascular invasion on the post-resection tumour samples. The tumour stage (T stage) was as follows: T1 (n=2), T2 (n=10), T3 (n=6) and T4 (n=9).

3.5.2 Data Integrity

Overall, the data integrity was below average owing to numerous factors including a lower-than-expected yield of uniquely mapped reads (60-70%). There were a few samples which had higher than expected proportions of unmapped reads (up to 50% in some cases) and in others the proportion of fragments mapping to multiple loci exceeded 10%. This is likely explained by DNA and rRNA contamination. Owing to the problems on different levels, affecting the samples in a non-uniform way, most probably introducing a bias to the gene counts, we elected to optimise the subsequent analysis by excluding some samples from the analysis.

3.5.3 Recurrence

We analysed the data by specifically looking at the genomic signatures between those tumours that recurred and those that did not. We performed Differential Gene Expression (DGE) analysis on the bulk raw data and identified 293 differentially expressed genes ($p<0.05$) between recurrence and non-recurrence patients. Of these, 88 were positively upregulated in recurrence patients and 205 were down regulated with respect to recurrence patients (Figure 3.13). However, there was no unification of said genes into any specific ontological pathway. Gene Ontology analysis did not demonstrate any significant ontological correlates at $p<0.05$.

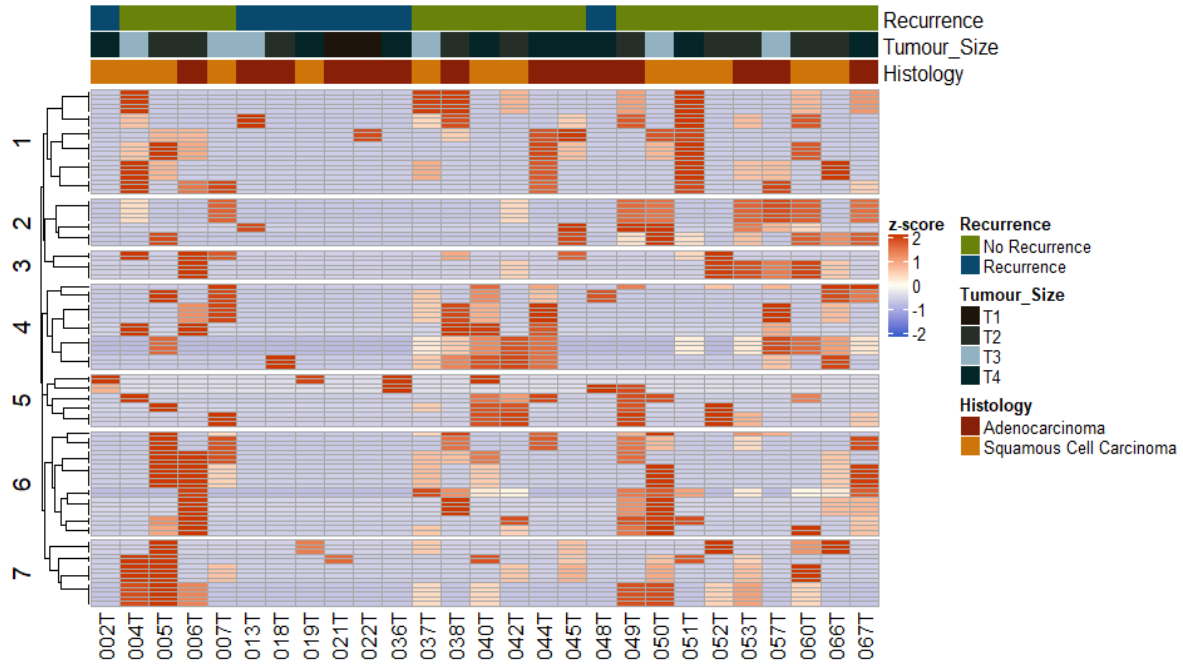


Figure 3.15. Bulk RNA sequencing heatmap stratified between various parameters. Top 100 differentially expressed genes displayed.

Insight into specific gene pathways with KEGG analysis showed in non-recurrence patients, a significantly higher expression of genes associated with Lysine degradation. Groups have demonstrated links between lysine demethylation and B cell proliferation and humoural

responses (278), however no tangible links have yet been demonstrated to unify said pathways, B cells and outcomes in cancer.

3.5.4 NSCLC [Public RNAseq Dataset]

Rousseaux S et al (GSE30219) analysed the transcriptomic profile in 174 NSCLC patients as a means of identifying differentially expressed genes associated with early recurrence and aggressive lung cancers following surgery (279). We performed Differential Gene Expression (DGE) analysis on the bulk raw data and identified 298 differentially expressed genes ($p<0.05$) between recurrence and non-recurrence patients. Of these, 252 were positively upregulated in recurrence patients and 46 were down regulated with respect to recurrence patients. Insight into specific gene pathways with KEGG analysis showed in recurrence patients, a significantly higher expression of genes associated with cell cycle and oocyte meiosis. Gene Ontology analysis corroborated these findings, highlighting genes associated with the nuclear and cell division as the most significantly enriched in recurrence patients (Figure 3.14). No specific genomic correlation between recurrence and B cell function was found on interrogation of this dataset.

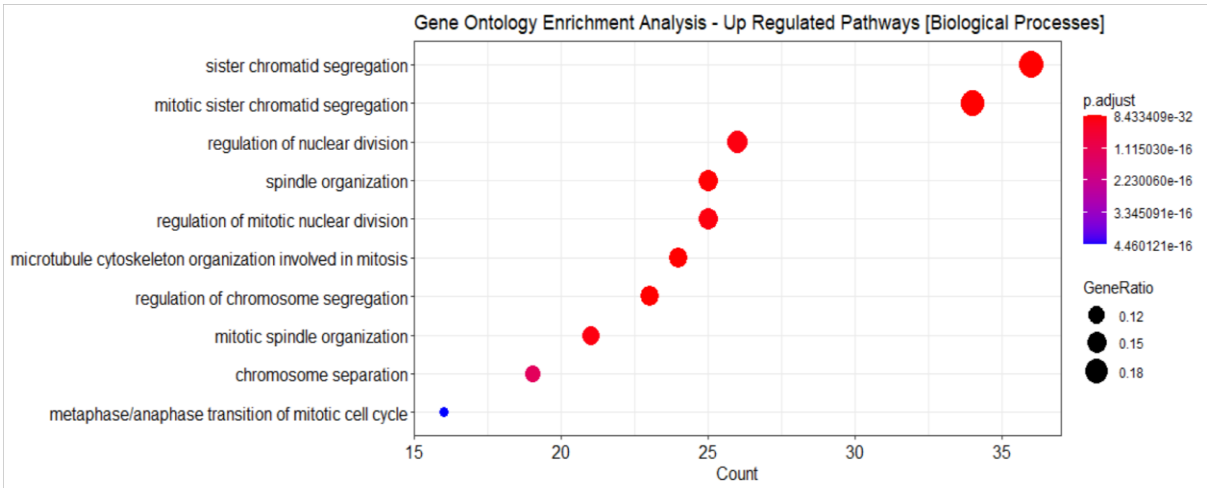


Figure 3.16. Gene ontological assessment of public RNAseq database (GSE30219), demonstrating significantly upregulated biological pathways in NSCLC patients who suffer from early recurrence.

3.6 Discussion

B cells and plasma cells located in the tumour microenvironment or in the circulation play important roles in shaping the anti-tumour effector immune response. Tumour-infiltrating B cells can exert both pro- or anti-tumour responses depending on the composition of the tumour microenvironment, the B cell phenotypes (suppressive or effector) present and any associated antibody production.

Our interrogation into the B cell repertoire of early-stage NSCLC patients, utilised deep phenotyping using mass cytometry, bulk RNA sequencing and multispectral tissue imaging using immunofluorescence. A clear difference in B cell surrogacy between tumour and blood was demonstrated using mass cytometry where a higher CD138+ plasma cell infiltrate was noted in tumour tissue. This tumour-infiltrating, antibody-producing cells associated with a lower degree of recurrence ($p<0.05$). In the circulation, a higher suppressive B cell presence was linked to recurrence ($p<0.05$) and at the tissue level, these suppressive B cells (CD138+ or CD20+) were noted to reside predominantly in the tumour stroma as opposed to infiltrating the nest.

Analysis of RNA sequencing data from The Cancer Genome Atlas (TCGA) database revealed that high levels of expression of B cell and plasma cell signature genes correlated with increased overall survival in patients with melanoma, lung adenocarcinoma, pancreatic adenocarcinoma, and head and neck squamous cell carcinoma but contrastingly had the inverse correlation in glioblastoma and clear cell renal cell carcinoma (156). IgG+ plasma cells with focused specificity and the ability to expand into large hypermutating clones facilitate tumour cell killing via complement fixation, promotion of antigen presentation via dendritic cell and drive cytotoxic T cell responses (131,155,157,189,192). Effector B cells which produce

inflammatory cytokines and upregulate MHC class I and II, also exert direct tumour cell killing, form tertiary lymphoid structures and shift the balance toward tumouricidal M1 macrophage phenotypes (131,157). Conversely, immunosuppressive B cells residing in the circulation and IgA+ plasma cells act in the opposite way, shifting the balance towards suppressive T regulatory cells, M2 macrophages and dampen Th1 responses thus exerting an overall pro-tumour effect (120,160).

Tumour-Infiltrating B cells (TILBs) and plasma cells have become recently identified as important components of TME and are linked to outcomes in lung cancer and responses to checkpoint blockade in advanced disease (131,157,192). Recent spatial mapping of lung adenocarcinoma has shown highly enriched populations of plasma and memory B cells in tumour tissues with high levels of differentiation and somatic hypermutation (280). High levels of CXCL13 production were noted in the tumour tissue and this evolved with cancer progression suggesting increased trafficking of these cells into the TME from tumour derived signals (280). Plasma cell preponderance was noted in the TME in our study with follicular and memory residing in the circulation with trafficking into the TME dependent on the appropriate chemotactic/antigen-specific signal.

CD138+ plasma cell infiltration has correlated with both good and poor outcomes in NSCLC depending on the study (144–146), with a predilection for a favourable prognosis in adenocarcinoma subtypes. CD138 expression has been associated with increased survival across all NSCLC stages (145,146). Immunohistochemistry data from a large clinical cohort (281) has demonstrated a significant link between CD138+ plasma cell infiltration and better survival. CD138+ cells co-localised with p46+ NK cells; NK:B cell interactions have been shown

in vivo whereby NK cells can induce the plasma cell effector phenotype with subsequent Ig production (282).

Increased intratumoural plasma cell infiltration has been reported to be associated with extended overall survival in NSCLC patients receiving anti-PDL1 treatment (283). Single cell RNA sequencing data from the POPLAR trial (284) showed that the status of an immune module, determined by the high correlation found among activated T cells, IgG+ Plasma cells and macrophages, termed lung cancer activated molecule (LCAM1) is associated with better progression free survival in patients treated with anti-PDL1. The $LCAM1^{hi}$ status showed a trend towards better overall survival in similarly treated patients (285). Murine models of castrate resistant prostate cancer refractory to Oxaliplatin treatment have demonstrated an increased presence of IgA+ Plasma cells that induce exhaustion of CD8+ T cells through PDL1 expression as well as TGF-β and IL-10 production, a true regulatory suppressive plasma cell phenotype (286). Removal of this population re-enables the control of large tumours by Oxaliplatin. This same suppressive population has been shown to accumulate in animal and human cases of inflammatory liver diseases impeding anti-cancer effector T cell responses (287). Blockade of PDL1 in advanced stage lung cancer could be dampening down the suppressive plasma cell phenotypes at play in the TME and thus shifting the balance towards the effector IgG producing phenotypes. Patil et al intimate that the benefit of PDL1 blockade is a result of direct effector plasma cell behaviour that is T cell independent (283). In serous ovarian cancer, Kroeger et al (288) demonstrated that plasma cells are required to induce CD8+ effector T cell responses, with adaptive immune response coordination occurring in Tertiary Lymphoid Structures (TLS). In reality, there is likely to be multi-level plasma cell:T cell cross-talk facilitating effector tumour responses.

We have demonstrated an infiltration of plasma cells into the TME that exist across a phenotypic spectrum. Suppressive regulatory B cells can develop at any stage of B cell maturation; the finding of CD138+ IL-10 producing plasmablasts provided evidence that this could occur even at the terminally differentiated end of the spectrum, with these cells demonstrating a BLIMP-1^{LO} phenotype and expressing switched IgG isotypes (180,289). Suppressive Bregs can suppress through cytokine signalling (IL-10, IL-35, TGF-β) or independently through PD-1/PDL-1 interactions. A recent description of neurotransmitter, GABA release by activated B and plasma cells promotes induction of IL-10+ suppressive macrophage populations which subsequently dampens anti-tumour CD8+ T cell responses (290).

Further questions remain such as what type or stage of B cell lymphocyte gives rise to a particular plasma cell subset? What are the extracellular or tumour derived signals inducing the regulatory phenotype? What are the core transcriptional elements that drive plasma cell phenotype?

There is a clear heterogeneity in the literature which describes the role of CD138+ plasma cells and CD20+ B cells in solid tumours. This is likely due to the study of limited markers, additional deeper study of plasma cells and matched antibody isotypes, account for concurrent T cell subset and address the impact of unique histological and molecular subtypes as well the impact of prior treatments (125). Our data has demonstrated the unique roles of effector and suppressive B cell subsets in NSCLC and the clear difference in surrogacy between tissue compartments. The lesser infiltration of suppressive phenotypes into the tumour nest may be to prevent the subsequent infiltration of effector plasma and effector B cells into the tumour nest. This tumour line of defence would be in keeping with the observation that

higher levels of suppressive B cell subsets in the circulation links to higher rates of recurrence and a higher tumour plasma cell infiltrate correlates with protection against recurrent disease.

The ambiguous role of B cells in the tumour environment predetermines the multidirectional development of immunotherapeutic approaches, either supporting positive B cell types or inhibiting the suppressive B cell phenotypes (156).

Chapter 4. Objective 2

B cell changes are predictive of outcome and immune related adverse events (irAE) in NSCLC patients receiving checkpoint blockade therapy

4.1 Immunotherapy in NSCLC

Checkpoint blockade immunotherapy has transformed the management of many common cancers. Five year overall survival (OS) for combined nivolumab/ipilimumab in advanced melanoma was 52% and for nivolumab alone 44% (60). The complete response rate with combination immunotherapy was 22%. The median OS for advanced NSCLC treated with best supportive care is less than 5 months (291); the median OS in patients with tumour proportion score (TPS) of >50% treated with pembrolizumab was 30 months (66). Five year OS for treatment naïve patients treated with pembrolizumab was 23%, and for those with TPS >50%, OS was 29.6% (65). In patients with PD-L1 negative non-squamous lung cancer the addition of pembrolizumab to chemotherapy improved the objective response rate (ORR) from 14.3% for chemotherapy alone to 32.3%, and more than doubled the 24 month survival rate from 15.5% to 38.5% (68).

Whilst checkpoint blockade primes naïve T cells and abrogates T cell exhaustion in the tumour microenvironment to mediate the immunological regression of cancer it is increasingly clear that B cells are also crucial cellular components contributing to the efficacy of checkpoint blockade. Higher expression of B cell related genes were seen in responders compared to non-responders, as was the density of micro-environmental B cells and tertiary lymphoid structures (TLSs) in the tumour tissue of melanoma patients (157). Enrichment of a TLS signature (TLSs were B cell rich) predicted improved survival with immunotherapy in melanoma (131). Sarcoma immune class (SIC) E is characterised by high expression of a B cell

lineage signature and these patients had the highest response rate (RR) and progression free survival (PFS) when treated with pembrolizumab (192).

However, the utility of these agents is marred by the development of immune related adverse events (irAE) which are common toxicities with checkpoint blockade. In a meta-analysis of 7936 patients in 48 clinical trials the most common all grade irAEs with nivolumab were skin (24.28%), gastro-intestinal (10.73%) and endocrine (10.09%) and the most common grade 3 or higher irAEs were hepatitis (1.26%) gastrointestinal (1.2%) and skin (0.99%) (292). Pneumonitis was reported in 3% of patients. For the nivolumab/ipilimumab combination, the most common all grade irAEs were skin (50.56%), gastro-intestinal (33.55%) and endocrine (27.55%) and grade 3 or greater toxicity, hepatic (10.06%), gastrointestinal (9.93%) and endocrine (4.07%). Pneumonitis was reported in 7.23% and grade 3 or greater colitis in 5.21%. The most common cause of treatment-related death with anti-PD-1/PD-L1 is pneumonitis (35%) whereas with CTLA4 blockade, 70% of all fatal irAEs are caused by colitis (293). The rate of pneumonitis with combined pemetrexed or platinum or pembrolizumab was 4.9% (68) and pulmonary irAEs negatively correlate with objective response in patients treated with nivolumab (292). Given the very widespread use of these agents, the management of severe irAEs represents a significant health care burden. Thus, predictors of those at high risk of high-grade toxicity would be of huge value to adequately assess the risk:benefit ratio for immunotherapy, to develop vigilant surveillance policies in high-risk patients and possibly develop preventive strategies.

We are aware of only one study that analyses the role of B cells in the development of irAEs secondary to checkpoint blockade (2). On-treatment changes in peripheral B cells were analysed in 39 melanoma patients, 23 of whom were treated with combination

immunotherapy. This group utilised a B cell parameter comprising a decline in B cell number >70% and a 2-fold or greater increase in either $CD21^{lo}$ B cells or plasmablasts. The $CD21^{lo}$ B cells were associated with greater clonality, activation and PD1 expression. None of the patients with B cell changes treated with combination therapy were free of grade 3 or greater irAEs at 6 months compared with 87% of those without such B cell changes, and such patients were more likely to develop multi-organ autoimmune toxicities (2).

In order to explore the relationship between B cells and outcome on checkpoint blockade therapy, we utilised deep phenotyping techniques with two B cell specific CyTOF panels, functional B cell assays, robust high dimensional display techniques, and regression models to analyse the importance of pre-treatment specific B cell immunophenotypes on response and the development of severe irAEs in lung cancer patients treated with anti-PD-1/PD-L1 or combination chemotherapy/PD-1 blockade. We explored the changes between pre- and post-treatment B cells in circulation as well as publicly available genomics datasets in both melanoma and NSCLC to corroborate our findings.

4.2 Advanced Disease NSCLC Cohort

Our cohort consisted of 46 patients; full demographic data is shown in Table 4.1. Median follow-up was 399 days with each patient having completed at least twelve cycles of treatment by the time of analysis with time on treatment across the entire cohort ranging from 331 days to 590 days.

Eight patients (17.4%) experienced severe grade 3 or greater irAE as defined in the ESMO Clinical Practice Guidelines (62) post checkpoint blockade treatment. Of these, four patients received pembrolizumab monotherapy and four were on combination treatment with chemotherapy. All irAE were diagnosed using clinico-radiological correlates and required

biopsy verification in the case of colitis. These patents constitute the toxicity cohort. The median time to severe toxicity was 42.5 days, range 31 days (Polymyalgia Rheumatica) – 115 days (colitis). All patients in the toxicity group discontinued immunotherapy. The remaining no toxicity patients did not experience grade 3/4 toxicity, however 4 patients of the non-toxicity group developed a self-limiting grade 1 skin rash after the first cycle of therapy which resolved with emollients and topical steroids and did not require stopping the immunotherapy. At the time of peripheral blood sampling, all patients were treatment naïve (baseline samples, day 0). Post-treatment samples were taken after the first cycle of checkpoint blockade on day 21. There was no significant history of previous cancers, autoimmune phenomena, allergic conditions, current steroid use or recent (within previous 6 months) vaccination history in our cohort. There was no significant difference in age or gender between the toxicity cohort (n=8) and the non-toxicity cohort (n=38). Clinical responses such as complete response (CR) (n=2), partial response (PR) (n=16), stable disease (SD) (n=8), and progressive disease (PD) (n=20) (disease progression within 180 days) were determined based on Response Evaluation Criteria in Solid Tumours (RECIST) 1.1 (294). For the purposes of data analysis, patients were further categorized as responders (n=26) based on progression-free survival for at least 180 days, while non-responders were patients who had disease progression within 180 days (n=20).

Table 4.1. Cohort Characteristics		
	Advanced NSCLC Cohort (n=46)	**Control (n=5)**
Age (years)	72.3 +/- 1.8	70.7 +/- 0.8
Gender (% male)	73%	60%
Advanced NSCLC Cohort (n=46)		
Mutation Status*		
PD-L1 < 50%	56.5% (n=26)	
PD-L1 > 50%	43.5% (n=20)	
Treatment Arm		
Anti-PD1 (Pembrolizumab or Nivolumab)	32.6% (n=15) [12 Pembrolizumab, 3 Nivolumab]	
Anti-PD1 + Chemotherapy	52.2% (n=24)	
Anti-PDL-1 (Atezolizumab)	6.5% (n=3)	
Anti-PDL-1 + Avastin + Chemotherapy	8.7% (n=4)	
Tumour Stage		
IASLC Stage IIIb	8.7% (n=4)	
IASLC Stage IIIc	26.1% (n=12)	
IASLC Stage IVa	41.3% (n=19)	
IASLC Stage IVb	23.9% (n=11)	
Histology		
Adenocarcinoma	67.4% (n=31)	
Squamous Cell Carcinoma	32.6% (n=15)	
Immune-Related Adverse Events**		
Grade 3-4	17.4% (n=8)	

Target Organ	
Colon	6.5% (n=3)
Lungs	2.2% (n=1)
Kidneys	2.2% (n=1)
Other***	6.5% (n=3)
Number of Organs affected	
0	82.6% (n=38)
1	15.2% (n=7)
2	2.2% (n=1)

*All patients were eGFR, ALK, ROS, BRAF wildtype except those who underwent QUAD therapy treatment with Atezolizumab, Avastin and Chemotherapy following TKI failure (eGFR+ve)

**No significant baseline differences between those who developed toxicity and those who did not

***Other includes arthritis, arthralgias, polymyalgia rheumatica and dry eyes

Table 4.1. Baseline characteristics of the advanced disease NSCLC patients who received immunotherapy.

4.3 Deep B cell phenotype in NSCLC

We performed large-scale mass cytometry analysis of 46 peripheral blood samples from patients with advanced stage NSCLC (IASLC stage IIIb or higher) and five healthy age-matched donor samples (Table 4.1). This dataset was analysed using an internal test/validation method, the first 20 prospectively samples were analysed, and our findings were corroborated and replicated using the second set (n=26) of prospectively collected samples. We present the analysis of all samples here. All patient samples were taken before commencing treatment checkpoint blockade as either anti-PD-1/PD-L1 monotherapy or pembrolizumab in combination with chemotherapy. Post-treatment samples were taken after the first cycle of immunotherapy, on day 21 post baseline draw (day 0).

4.3.1 Antibody Panels

Cells were stained with two antibody panels created for this study (Table 2.1). There is a paucity of data in the literature demonstrating deep B cell phenotyping. The first panel was designed to detect the expression of B cells at various stages of maturation (activated, transitional, marginal zone, follicular, germinal centre, class-switched and plasma) as well as rarer B cell populations such B regulatory (Breg) cells. All patients were run on this first "base" B cell panel. The second panel was designed to interrogate Breg phenotypes further and run on a subset of patients (n=8) in order to validate and further characterise the phenotypes identified by the first panel. Both panels included markers for natural killer cells, T cells and granulocytes.

4.3.2 B cell Milieu in Pre-Treatment Advanced Disease NSCLC

In order to generate a comprehensive picture of the B cell landscape we created two-dimensional maps of the data, using dimensionality reduction algorithms (UMAP and Diffusion Maps). UMAP was used in preference to t-SNE owing to faster run times, especially with large perplexity hyper-parameters and better preservation of the global cellular architecture with competitive visual quality (240). The arrangement of cells along their development path, with bifurcations along points of differentiation was achieved using diffusion maps. Distinct cellular phenotypes were determined following application of the FlowSOM clustering algorithm which employs a Self-Organising Map to order determine how markers are behaving on cells in high dimensional space thus creating well partitioned cellular subsets (236).

CD3+ T cells accounted for 54.4% (mean) of the total live cell population across all samples. The total CD19+ B cell population accounted for 5.3% of the total live cell population on

average across all samples. CD19+ cells were taken forward into all further downstream analyses.

In order to map cell phenotypes, FlowSOM analyses were repeatedly performed for reliability and expression of B cell clusters were visualised as a heatmap (Figure 4.1) with heterogeneity in marker level displayed at single cell level using UMAP (Figure 4.2).

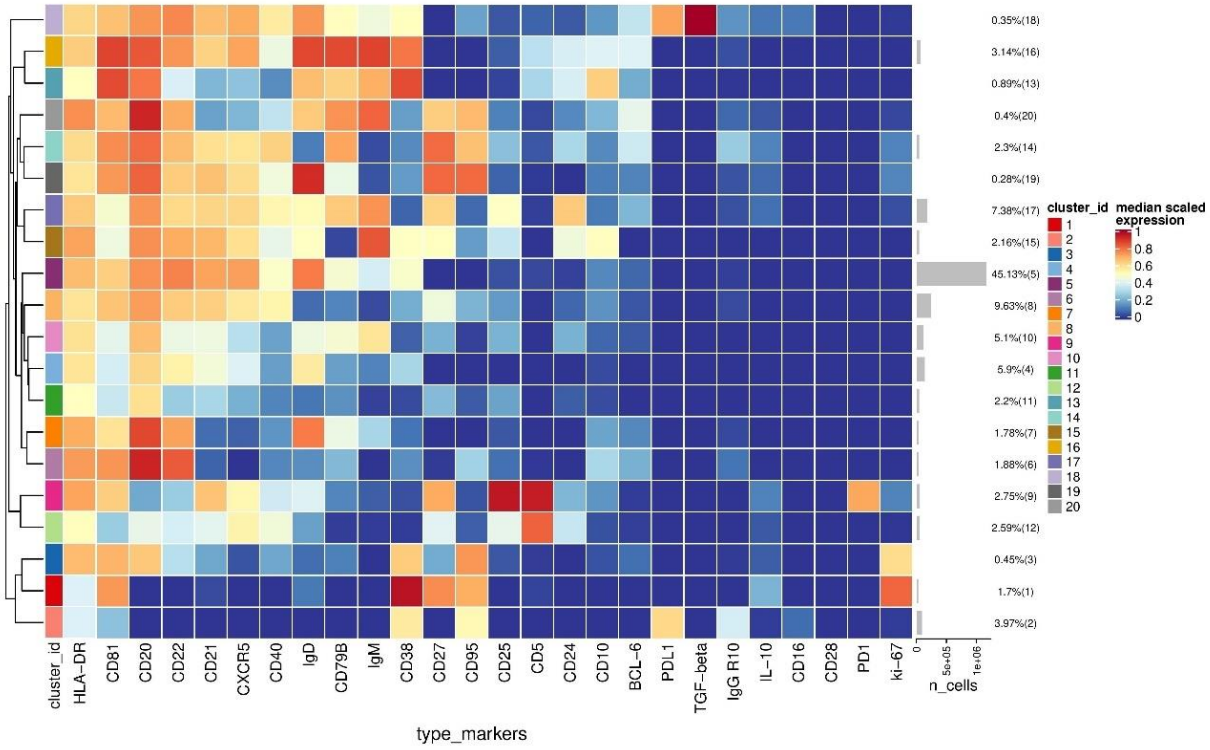


Figure 4.1. Base B cell phenotyping heatmap illustrating 20 B cell phenotypes as clusters (1-20).

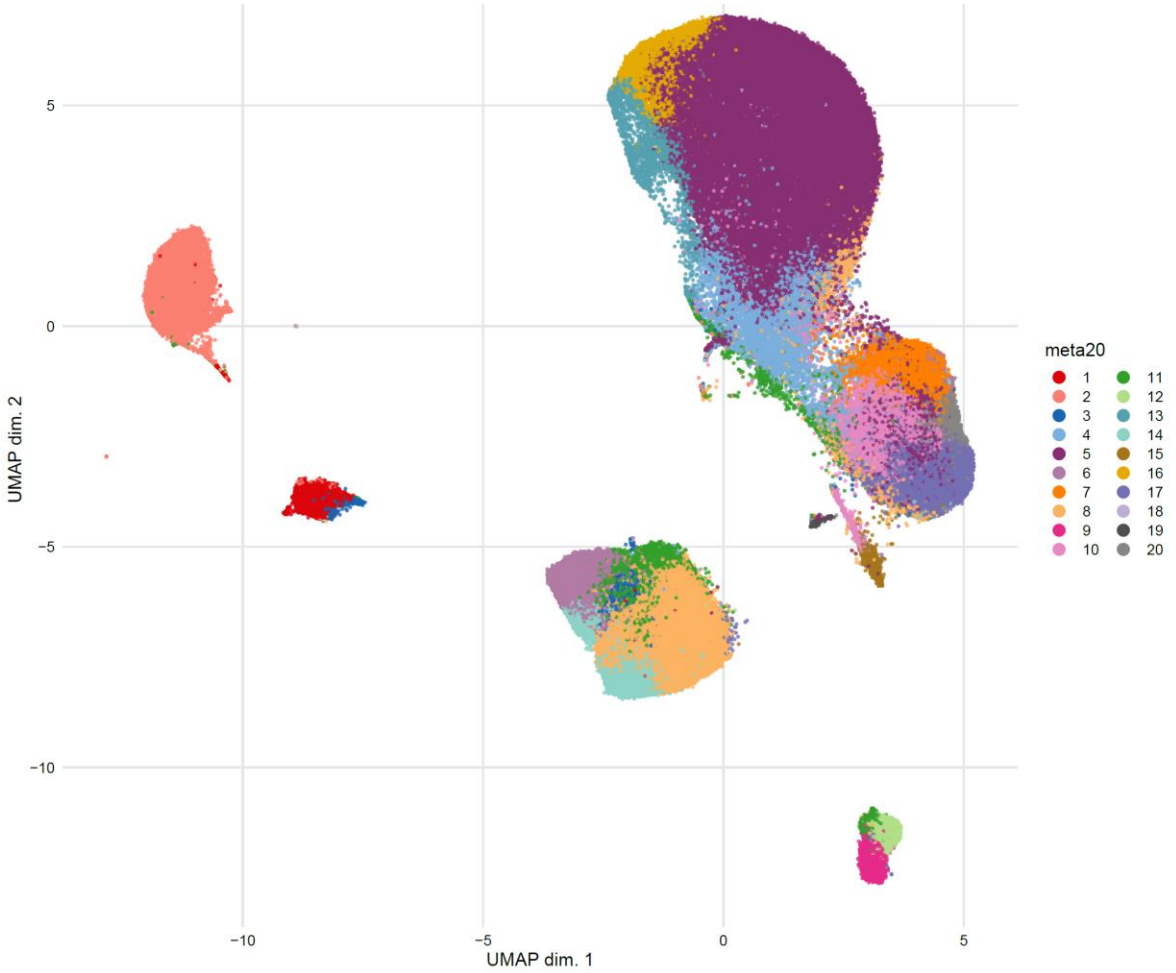


Figure 4.2. UMAP plot displaying all cells from each FlowSOM-derived cluster identified in (A) coloured by cluster.

We defined 20 distinct B cell clusters at various stages of maturation. The mostly frequently observed cluster was of the follicular B cell lineage (clusters 5 (Figure 4.1)) characterised by high expression of CD20, CD22 and IgD and with lower levels of IgM. Marginal zone (IgM+ IgD^{lo} $CD21^{int}$ $CD27^{lo}$) (cluster 10) and Transitional B cell (IgM^{hi} $CD24^{hi}$ $CD38^{hi}$ CD5+) clusters (cluster 16) comprised 5.1% and 3.14% of the total B cell population respectively. Clusters 13 (0.89%) and 15 (2.16%) are also likely to represent transitional B cell populations with high levels of CD10, CD24 and CD38. Both clusters are also high expressers of IgM and IgD, with cluster 13 being $CD21^{lo}$ and cluster 15 being $CD21^{hi}$. Memory B cells were observed at various

stages of maturation: early memory B cells (cluster 19, 0.28%), fully affinity matured, class-switched B cells (CD27hi IgD- IgM-) (cluster 14, 2.3%) and double negative memory B cells (CD27lo IgD- IgM-) (cluster 11, 2.22%).

Seven B regulatory cell clusters, characterised using surface markers such as CD5, CD24, CD25, CD27, CD38, CD1d, TIM-1, PD1, PDL-1, TGF-β and intracellular cytokine expression of IL-10, were observed to varying frequencies (0.35% to 7.38%) (Clusters 1, 2, 3, 9, 12, 17 and 18). All B cells were unstimulated and thus IL-10 expression was representative of the *in vivo* immune milieu of NSCLC patients and healthy donors. These same populations were also seen in our Objective 1 cohort with the PDL1+ populations driving suppression and tumour progression in the early-stage setting.

4.3.3 Deep B cell Phenotype associates with high-grade irAEs secondary to Checkpoint Blockade

We performed comparative analyses between toxicity patients (n=8), no toxicity (NT) patients (n=38) and healthy controls (n=5) (Figure 4.3 and 4.4).

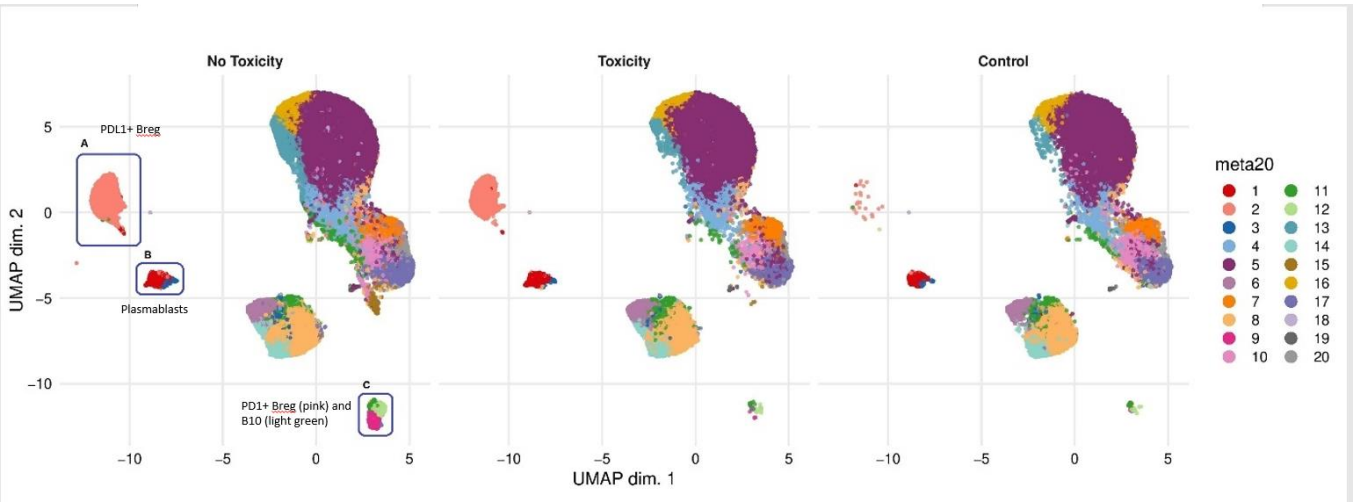


Figure 4.3. UMAP plots stratified according to condition, "no toxicity", "toxicity" and "healthy controls". Clusters highlighted by blue boxes correspond to clusters 1, 2, 3, 9, and 12 identified from Figure 4.1. These clusters are attenuated in the "toxicity" cohort in this unsupervised analysis. The box labelled "A" is encircling cluster 2 (peach). The box labelled "B" is encircling clusters 1 (red) and 3 (blue). The box labelled "C" is encircling clusters 9 (pink) and 12 (light green). All samples are randomly downsampled to account for equally representative populations across samples.

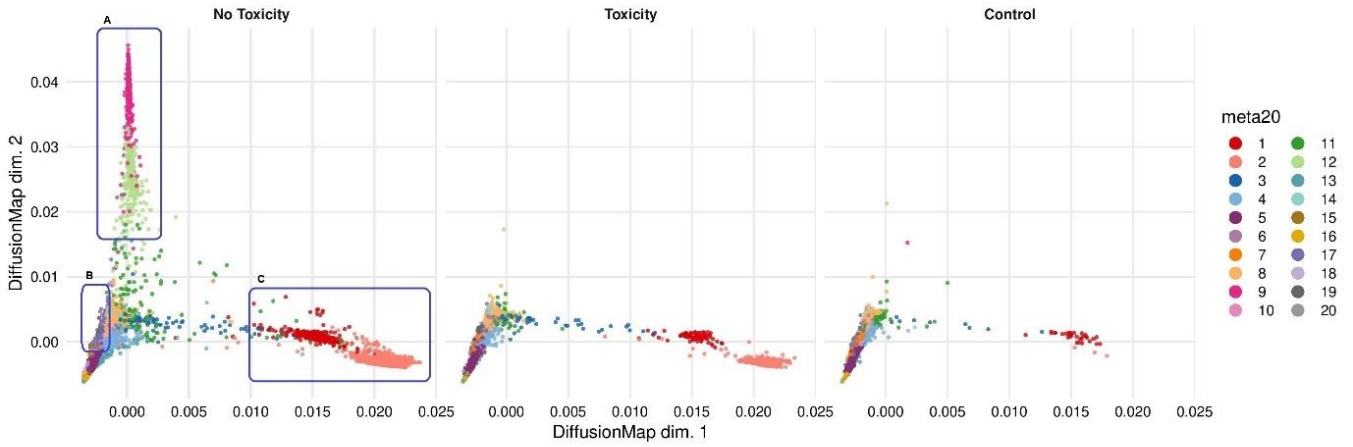


Figure 4.4. Diffusion map stratified according to condition, "no toxicity", "toxicity" and "healthy controls". Clusters highlighted by blue boxes correspond to clusters 1, 2, 9, 12, 17 and 18 identified from Figure 4.1. These clusters are attenuated in the "toxicity" cohort in this unsupervised analysis. The box labelled "A" is encircling clusters 9 (pink) and 12 (light green). The box labelled "B" is encircling clusters 17 (deep lilac) and 18 (light lilac). The box labelled "C" is encircling clusters 1 (red)

and 2 (peach). All samples are randomly downsampled to account for equally representative populations across samples.

The UMAP data (Figure 4.3) clearly shows loss of clusters 9 and 12 together with loss of cluster 15 and significant attenuation of clusters 1,2 and 3 in the toxicity patients. The diffusion map (Figure 4.4) also very clearly shows loss of clusters 2 and 9, significant attenuation of clusters 1, 2 and 3 and loss of clusters 17 and 18. The loss of these latter two clusters is not immediately obvious on the UMAP as they are small subsets and lost in the group of similarly differentiated subsets. Thus, the proportions of the seven suppressive B cell clusters (Clusters 1, 2, 3, 9, 12, 17 and 18) were severely attenuated in the toxicity cohort relative to the NT cohort. On deeper interrogation of the clusters attenuated in the toxicity cohort (Toxicity Linked Clusters; TLCs), the most notable feature was IL-10 expression seen in clusters 1, 3, 9, 12, 17 and 18. In total, there are ten IL-10+ clusters in the entire B cell population (with all the other 10 B cell clusters being completely negative for IL-10 expression), with the expression of IL-10 residing within six out of the seven of the suppressive clusters of interest (Clusters 1, 3, 9, 12, 17 and 18). Cluster 2 was clearly attenuated. Even though this is not an IL-10+ population, this is a $PDL1^{hi}$ B cell subset which has been described as suppressing via PDL1 independently of IL-10 (197). Whilst there is presence of IL-10 in the TLCs in the NT patient group (Figure 4.5, TLCs demarcated by red arrows with IL-10 positive regions shown as yellow) there is clearly less IL-10 expression in the corresponding TLCs in the toxicity patients.

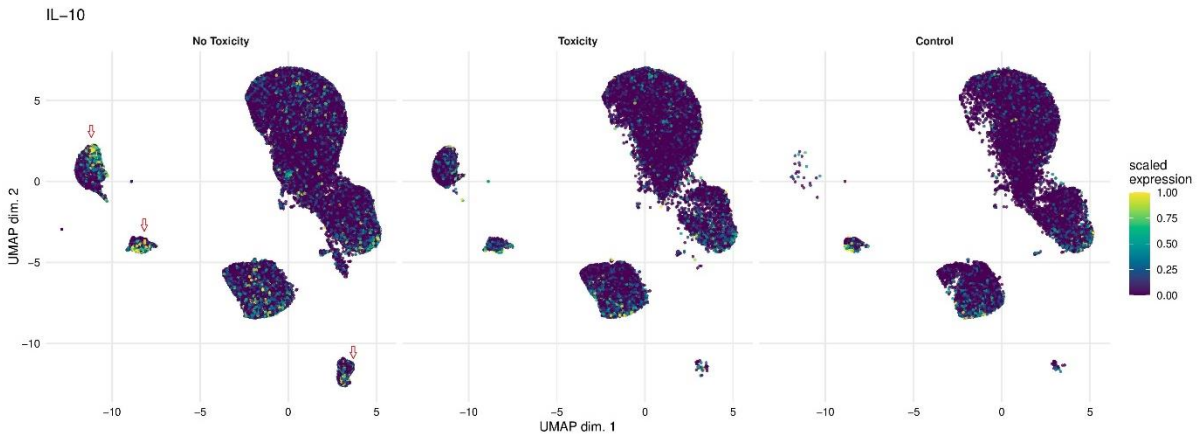


Figure 4.5. UMAP plots stratified according to overall population IL-10 expression and by condition. Expression is largely confined to the Breg clusters demarcated by the red arrows. Matched IL10 expression in the corresponding clusters in the toxicity patients is reduced. The left hand most red arrow depicts cluster 2 with subpopulations of IL10 expressing cells (areas of light green and yellow), said cells are not seen in the corresponding cells in the toxicity patients.

4.3.3.1 Breg phenotypes associated with autoimmune toxicity in NSCLC patients on CCB

IL-10 is an immunosuppressive cytokine and a canonical marker for the B regulatory cell (Breg) population. This diverse array of cells work to limit excessive inflammation and autoimmunity through IL-10 production (104). Analysis of individual marker expression patterns within the most strongly IL-10+ clusters that were significantly reduced in the toxicity cohort allowed for partitioning into four major phenotypes (Figure 4.1) which are based on the unsupervised partition analysis (Figure 4.2).

One of these clusters was PD1+ (cluster 9) and two were PDL1+ (clusters 2 and 18 (Figure 4.1): these were the only PD-1 positive or PD-L1 positive clusters of the entire 20 B cell subsets. Cluster 18 was the only population to express high levels of TGF-β, a known immunosuppressive cytokine. Of the remaining four clusters, we observed high CD5

expression as well as expression of CD24, CD25 and CD27 in clusters 12 and 17, and high CD38 and CD27 expression in clusters 1 and 3. Thus, phenotypically these seven attenuated clusters were characterised as (Figure 4.1):

A. $Ki67^{hi}$ $CD38^{hi}$ $CD27^{hi}$ $CD95^{hi}$ $IL\text{-}10^{int}$ Red Cluster (Cluster 1)
B. $PDL1^{hi}$ $CD38^{int}$ $CD95^{int}$ $TGF\text{-}\beta^{-ve}$ $IL\text{-}10^{lo}$ Peach Cluster (Cluster 2)
C. $Ki67^{int}$ $CD38^{int}$ $CD27^{int}$ $CD95^{hi}$ $IL\text{-}10^{lo}$ Royal Blue Cluster (3)
D. $PD1^{hi}$ $CD5^{hi}$ $CD25^{hi}$ $CD27^{hi}$ $CD24^{lo}$ $CD38^{lo}$ $IL\text{-}10^{int}$ Pink Cluster (9)
E. $CD5^{hi}$ $CD24^{lo}$ $CD25^{lo}$ $CD27^{lo}$ $IL\text{-}10^{lo}$ Light Green Cluster (12)
F. $CD24^{hi}$ $CD25^{hi}$ $CD27^{hi}$ $IL\text{-}10^{lo}$ Deep Lilac Cluster (17)
G. $PDL1^{hi}$ $CD38^{int}$ $CD95^{lo}$ $TGF\beta^{hi}$ $IL\text{-}10^{lo}$ Light Lilac Cluster (18)

These clusters all map to the following Breg populations as described in the literature:

1. $PD\text{-}1^{hi}$ Breg which as here (cluster 9) was $CD5^{hi}$ $CD24^{lo}$ and $CD38^{lo}$ and which produced significant IL-10 on PD-1 triggering (295).
2. $PD\text{-}L1^{hi}$ Breg which like our clusters 2 and 18 subset were CD10 and CD27 negative (197).
3. The other high IL-10 expressing clusters were Clusters 1 and 3 which were highly proliferative and had phenotypic similarity to the previously described plasmablast phenotype (CD27+ CD38+ IL-10+) (180).
4. Finally, clusters 12 and 17 expressed CD24, CD25 and CD27 and cluster 12 was one of only 2 clusters besides the PD-1+ cluster expressing high levels of CD5. Furthermore, clusters 12 and 17 also expressed int/hi levels of CD1d. B10 cells are characterised as $CD1d^{hi}$ CD5+ (173). Whilst IL-10 expression is the best marker of human B10 plus B10 progenitor cells (296), these cells form a subset of $CD24^{hi}$ CD27+ B cells (296) and both

of these markers were present on cluster 17, with the expression of CD24 being the highest in all of the strong IL-10 expressing cells in this cluster.

4.3.3.2 Differential Abundance Analysis

We performed a differential abundance (DA) analysis of the defined cell populations, reporting on all B cell clusters in the population. This method compares the proportions of cell types between the two clinical conditions and aims to highlight the populations that are present at significantly different ratios. In order to gain power to detect differences between conditions, we utilised a mixed model to model the response and patients were treated as a random effect thus formally accounting for patient to patient variability as described by Nowicka et al (237,276). DA analysis of the overall cell population identified five clusters as significantly more abundant in the non-toxicity cohort as shown by the grey bars in Figure 4.6. Four of these are the described suppressive Breg populations (clusters 2, 9, 12 and 18), which are as follows:

- $PDL1^{hi}$ $CD38^{int}$ $CD95^{int}$ $TGF\text{-}\beta^{-ve}$ $IL\text{-}10^{lo}$ Breg (p=0.011) (Cluster 2 from initial unsupervised analysis)
- $PD1^{hi}$ $CD5^{hi}$ $CD25^{hi}$ $CD27^{hi}$ $CD24^{lo}$ $CD38^{lo}$ $IL\text{-}10^{int}$ Breg (p=0.011) (Cluster 9 from initial unsupervised analysis)
- $CD5^{hi}$ $CD24^{lo}$ $CD25^{lo}$ $CD27^{lo}$ $IL\text{-}10^{lo}$ Breg (p=0.011) (Cluster 12 from initial unsupervised analysis)
- $PDL1^{hi}$ $CD38^{int}$ $CD95^{lo}$ $TGF\beta^{hi}$ $IL\text{-}10^{lo}$ Breg (p=0.042) (Cluster 18 from initial unsupervised analysis)

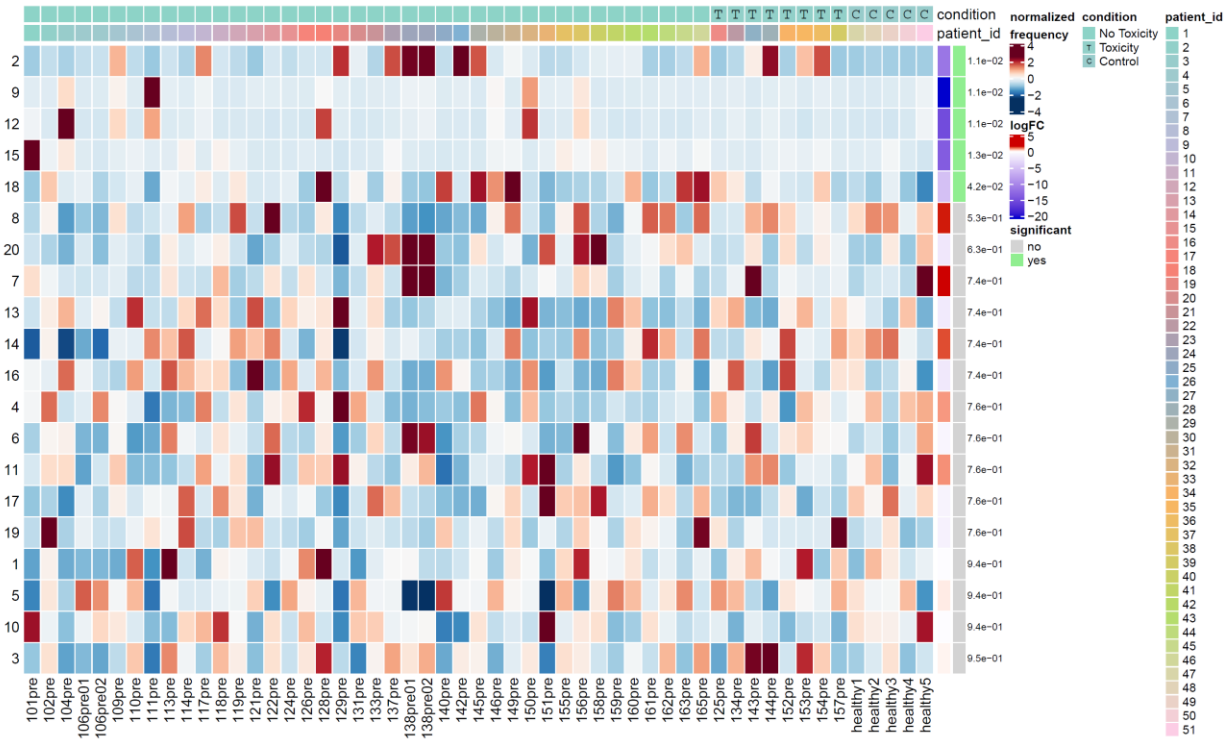


Figure 4.6. Differential Abundance Heatmap illustrating 20 previously identified clusters (2A) (left hand column) with relative normalised abundance of each cluster by individual patient and healthy control sample (main panel). Patient to patient variability was treated as a random effect to improve the robustness of the model. A generalised linear mixed regression model was applied to determine significance of differential abundance between conditions (toxicity and no-toxicity); the top five clusters were of statistical significance as shown by the green bars. Of these, four are the identified Breg populations. PDL1hi CD38int CD95int TGFβ-ve IL-10-ve Breg (p=0.011) (Cluster 2 from initial unsupervised analysis), PD1hi CD5hi CD25hi CD27hi CD24lo CD38lo IL-10int Breg (p=0.011) (Cluster 9 from initial unsupervised analysis), CD5hi CD24lo CD25lo CD27lo IL-10lo Breg (p=0.011) (Cluster 12 from initial unsupervised analysis) and PDL1hi CD38int CD95lo TGFβhi IL-10lo Breg (p=0.042) (Cluster 18 from initial unsupervised analysis). Cluster 15, an identified Transitional B cell population (IgMhi CD5- CD21hi CD10hi CD24hi CD38hi) was also significantly more abundant in non-toxicity patients (p=0.013).

4.3.3.3 Breg/Tfh Panel Sub-Group Analysis

We analysed a subgroup of eight patients (5 no-toxicity, 3 toxicity) on a more targeted Breg-specific panel. The data were analysed as above using dimensionality reduction plots to

display clusters by condition (Figure 4.7). Clusters 13, 14 and 16 are all significantly deficient in toxicity patients based on DA testing ($p<0.01$). Clusters 13 and 16 are both PD1+, with cluster 13 being $PD1^{int}$ and cluster 16 being $PD1^{lo}$. Both clusters also express TIGIT and CD1d, with higher levels of both in cluster 13. Cluster 14 is a B10 type cluster. Importantly, it is a significant lack of all three types of clusters, which seems to be the consistent feature amongst the toxicity patients identified.

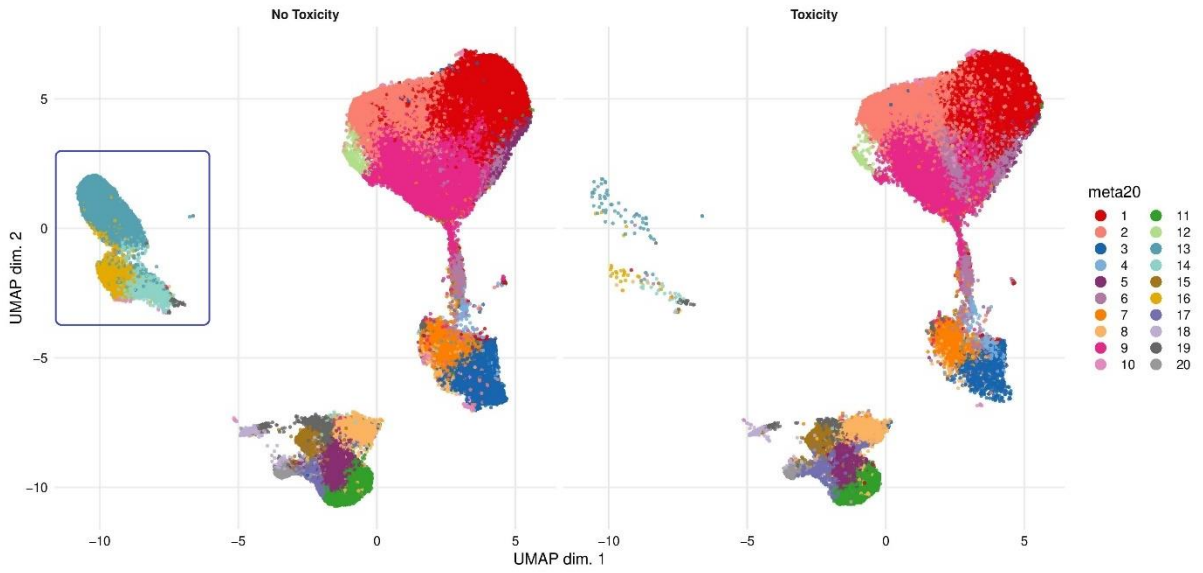


Figure 4.7. Comparative UMAP stratified according to toxicity in a subcohort of 8 patients using a refined Breg/Tfh panel

4.3.3.4 Time to Toxicity Analysis

We assessed the association between cluster abundance and time to toxicity in the form of Kaplan Meier analyses focusing on the seven identified main Breg subsets which were identified as differentially abundant either visually and/or through generalised mixed modelling based on median cluster abundance and expression. Clusters were delineated as high or low abundance according to measures and cut-off points taken from individual

bootstrapped ROC curve analyses. Of these, only clusters 9 and 17 showed significantly reduced time to toxicity with low Breg abundance (Figure 4.8). The cohort of patients with low abundance of specific Breg clusters had a greater risk of high-grade irAEs.

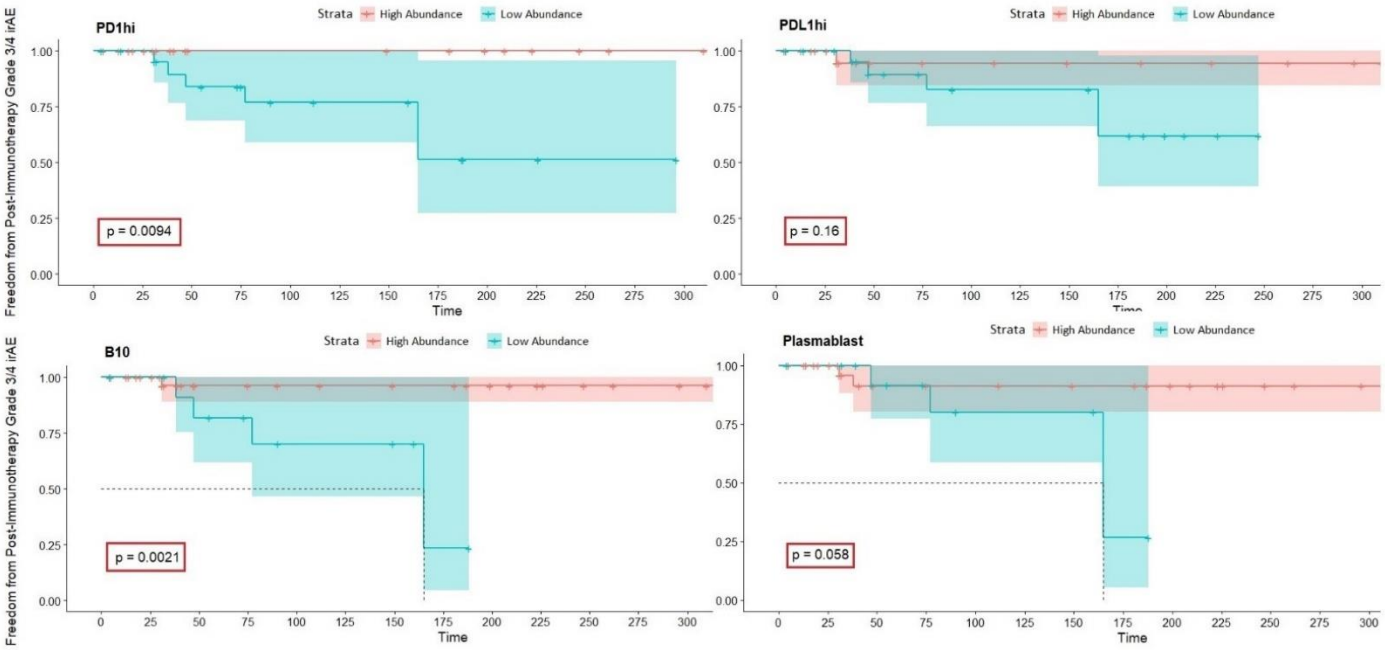


Figure 4.8. Kaplan-Meier Time to Toxicity analyses stratified to four pertinent suppressive Breg clusters identified in phenotyping analyses.

The rates of freedom from high-grade irAEs at 6 months were 100% (high) versus 58% (low) for the PD1^{hi} subset (cluster 9) (p=0.0061). For the PDL1^{hi} TGF-β^{-ve} subset (cluster 2), 96% (high) versus 61% (low) (p=0.0912). For the B10-like subset (cluster 17), 98% (high) versus 73% (low) (p=0.0038) and 89% (high) versus 55% (low) for the PDL1^{hi} TGF-β^{hi} subset (cluster 18) (p=0.0387 Fisher's exact test).

4.3.3.5 Deep B cell phenotyping in an independent cohort of NSCLC patients using Mass Cytometry

4.3.3.5.1 NSCLC

We validated our findings using an entirely separate cohort of patients from a different institution (PAIR Study, REC reference 17/LO/1950 Guy's and St Thomas's NHS Foundation Trust and King's College London. Credit to Dr. Zena Willsmore who facilitated sample handling and CyTOF data acquisition in NSCLC and Melanoma cohorts). This was a small but well dichotomised group of patients (n=8) with two of the patients experiencing high-grade toxicity on checkpoint blockade therapy. Using the same base B cell mass cytometry panel, we created a comprehensive overview of the B cell landscape. In this independent cohort, nine truly IL10+ Breg phenotypes were identified (clusters 4, 9, 12, 14, 16, 17, 18, 19 and 20), in the same proportions and expression of canonical Breg markers as previously described (Figure 4.9). Of these, the six most strongly IL10+ clusters are summarised below:

$PDL1^{hi}$ $CD38^{int}$ $CD95^{lo}$ $TGF\beta^{hi}$ $IL\text{-}10^{lo}$ Light Blue Cluster (4)

$PD1^{hi}$ $CD5^{lo}$ $CD25^{hi}$ $CD27^{hi}$ $CD24^{lo}$ $CD38^{lo}$ $IL\text{-}10^{lo}$ Light Green Cluster (12)

$Ki67^{hi}$ $CD38^{hi}$ $CD27^{hi}$ $CD95^{hi}$ $IL\text{-}10^{hi}$ Mustard Cluster (16)

$Ki67^{hi}$ $CD38^{hi}$ $CD27^{int}$ $CD95^{hi}$ $IL\text{-}10^{int}$ Deep Lilac Cluster (17)

$Ki67^{lo}$ $CD38^{hi}$ $CD27^{hi}$ $CD95^{hi}$ $IL\text{-}10^{lo}$ Light Lilac Cluster (18)

$PD1^{hi}$ $CD5^{hi}$ $CD25^{hi}$ $CD27^{hi}$ $CD24^{lo}$ $CD38^{lo}$ $IL\text{-}10^{lo}$ Light Grey Cluster (20)

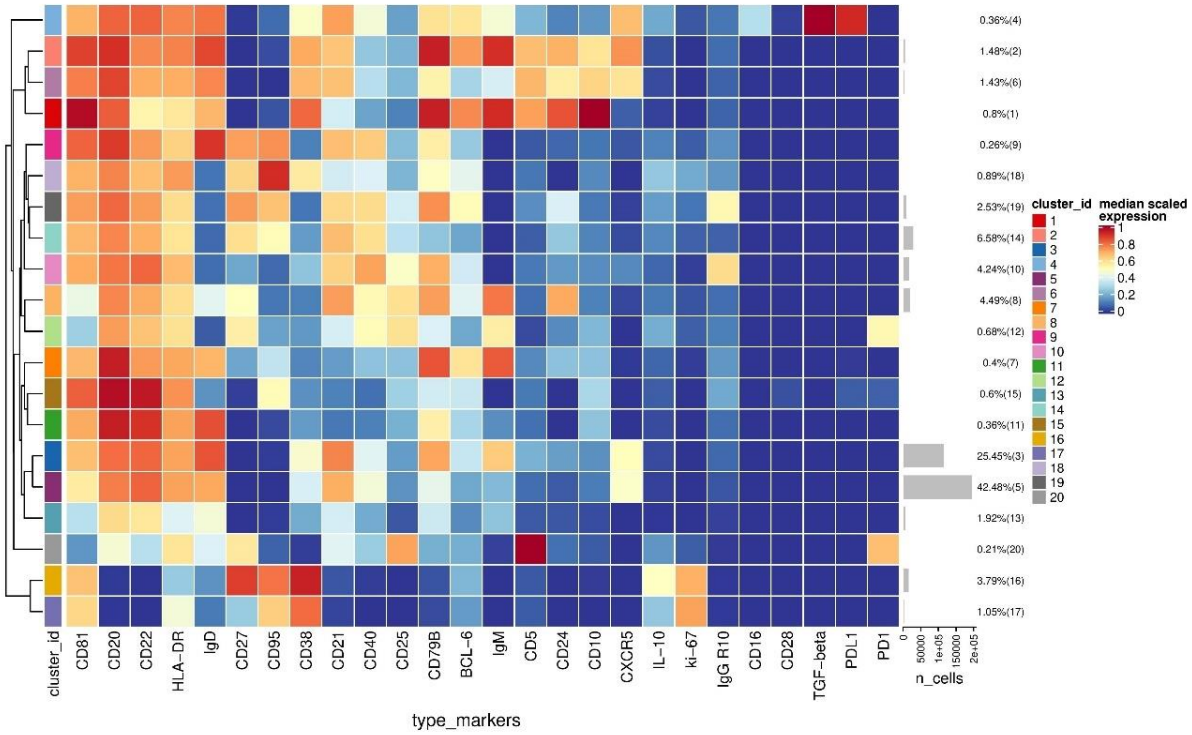


Figure 4.9. Heatmap showing normalised expression of the base B cell panel markers for the 20 B cell clusters identified with FlowSOM. The cluster IDs and relative frequencies are displayed as a bar graph on the right-hand side, along with the median scaled expression.

These nine clusters fit into the already described populations, namely plasmablasts (clusters 16, 17 and 18), B10-type cells (clusters 9, 14 and 19), PD1^{hi} Bregs (clusters 12 and 20) and the PDL1+ Immature Bregs (cluster 4). In this cohort, we noted two PD1+ populations largely delineated from one another by levels of CD5 expression (295) and three plasmablast populations with varying degrees of proliferation (ki67 positivity) and IL-10 expression. Cluster 16 was the most strongly IL-10 positive population. The PDL1+ population (cluster 7) displayed high levels of PDL1 expression, which has been observed in the original description of this heterogeneous population (197).

Comparative expression of the B cell populations between toxicity and non-toxicity patients was displayed using dimensionality reduction methods (Figures 4.10 and 4.11).

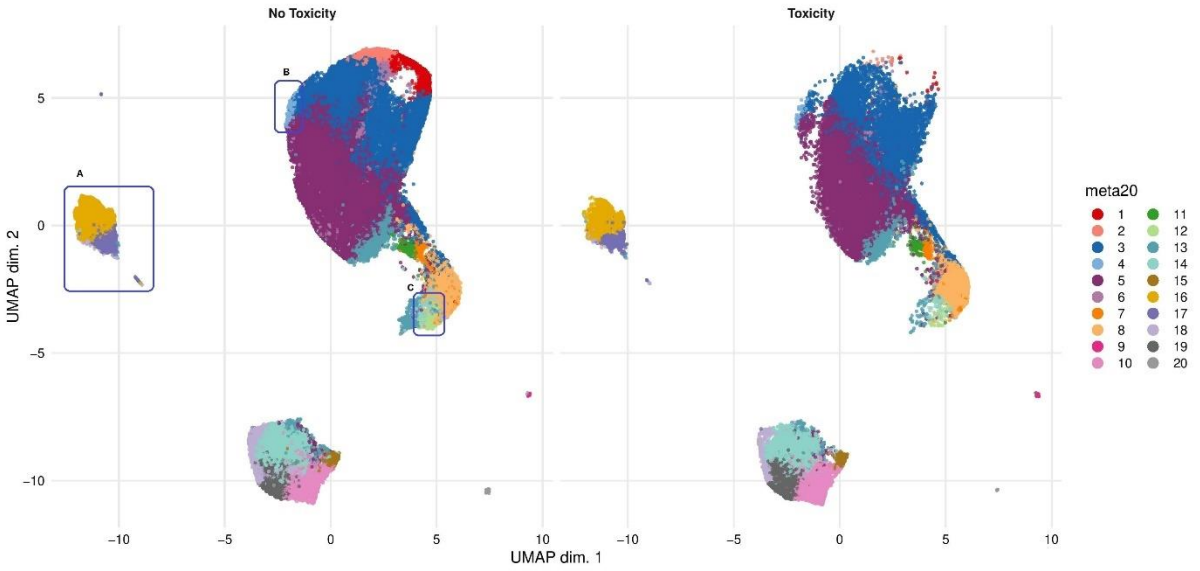


Figure 4.10. UMAP plots stratified according to condition, "no toxicity", "toxicity" and "healthy controls". Clusters highlighted by blue boxes correspond to clusters identified from (4.9). These clusters are attenuated in the "toxicity" cohort in this unsupervised analysis. The box labelled "A" is encircling clusters 16 (mustard) and 17 (lilac). The box labelled "B" is encircling cluster 4 (light blue). The box labelled "C" is encircling cluster 12 (light green). All samples are randomly downsampled to account for equally representative populations across samples.

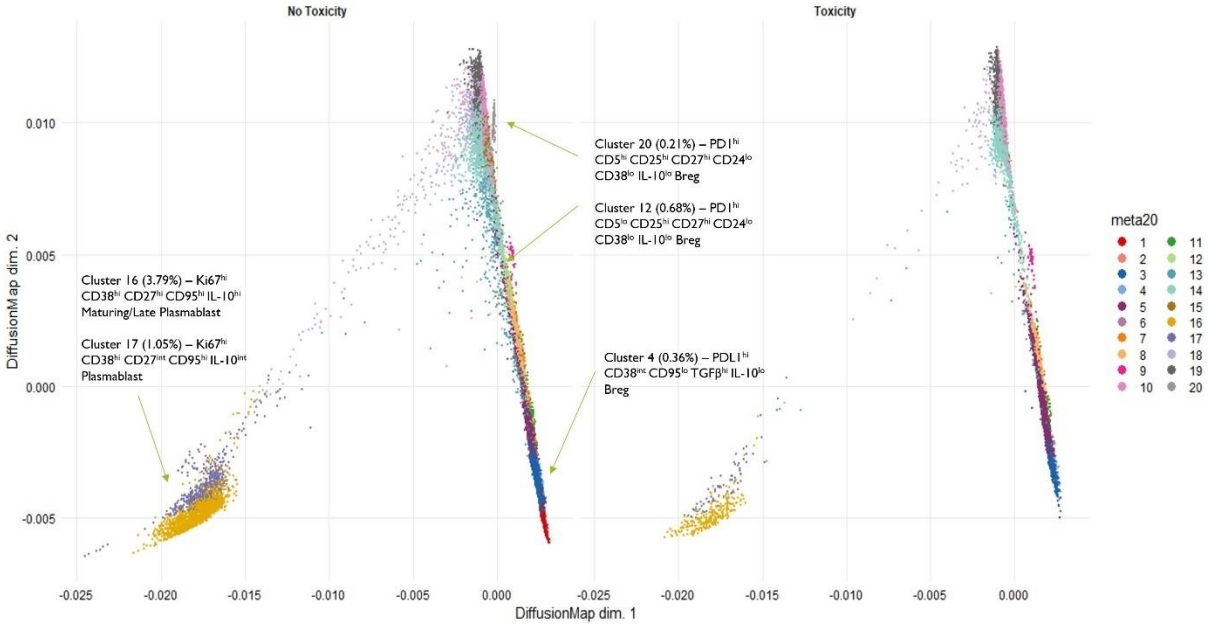


Figure 4.11. Diffusion map stratified according to condition, "no toxicity", "toxicity" and "healthy controls". Clusters labelled are 4, 12, 16, 17, and 20 are attenuated in the "toxicity" cohort in this unsupervised analysis. Cluster 4 is shown by the thin "light blue" line of cells to the right of the darker blue cells in cluster 3. All samples are randomly downsampled to account for equally representative populations across samples.

Comparative expression with UMAP (Figure 4.10) shows a striking loss of clusters 1 and 2 in the toxicity cohort. These map to transitional cell clusters 13 ($CD21^{lo}$) and 15 ($CD21^{hi}$ CD27+) respectively from our initial cohort, which were also deficient in toxicity patients; cluster 15 was significantly attenuated in toxicity patients on differential abundance testing (p=0.013). Cluster 4 was significantly attenuated in toxicity patients (Figure 4.10) and this maps to the similarly deficient $PDL1^{hi}$ $CD38^{int}$ $CD95^{lo}$ $TGF\beta^{hi}$ $IL\text{-}10^{lo}$ population, cluster 18 from our initial cohort. The diffusion map (Figure 4.11) shows a clear loss of clusters 12 and 20 ($PD1^{hi}$) and a significant attenuation of clusters 16 and 17 (plasmablast) in the toxicity cohort. Clusters 12 and 20 map to cluster 9 ($PD1^{hi}$) from our initial cohort. Finally, clusters 16 and 17 map to

clusters 1 and 3 respectively in our initial cohort, thus providing validation of our initial findings.

Bulk expression of IL-10 (Figure 4.12) shows predominant expression focused to the Breg clusters of interest (notably clusters 12, 16, 17 and 18) and despite the presence of these Breg phenotypes in toxicity patients (albeit lower compared to non-toxicity patients), there is diminished scale expression of IL-10 in these populations in toxicity patients. Figure 4.10 also shows a striking loss of clusters 1, 2 and 6 in toxicity patients: phenotypically these are likely to represent IgM+ IgD+ CD5+ CD21+ CD10+ CD24+ CD38+ transitional B cells which represent B cells at a key stage in their developmental pathway and encompass a range of transitional B cell subsets with unique regulatory functional profiles including those which suppress via IL-10 (297). These three clusters (1, 2, and 6) mirror the transitional cell populations, clusters 13, 15 and 16 defined in our initial NSCLC cohort respectively. In the initial cohort cluster 15 was one of the 5 significantly reduced B cell clusters in toxicity patients. Cluster 1 in this validation cohort and cluster 13 from the original are both $CD21^{lo}$ which is known to be associated with high PD-1 expression.

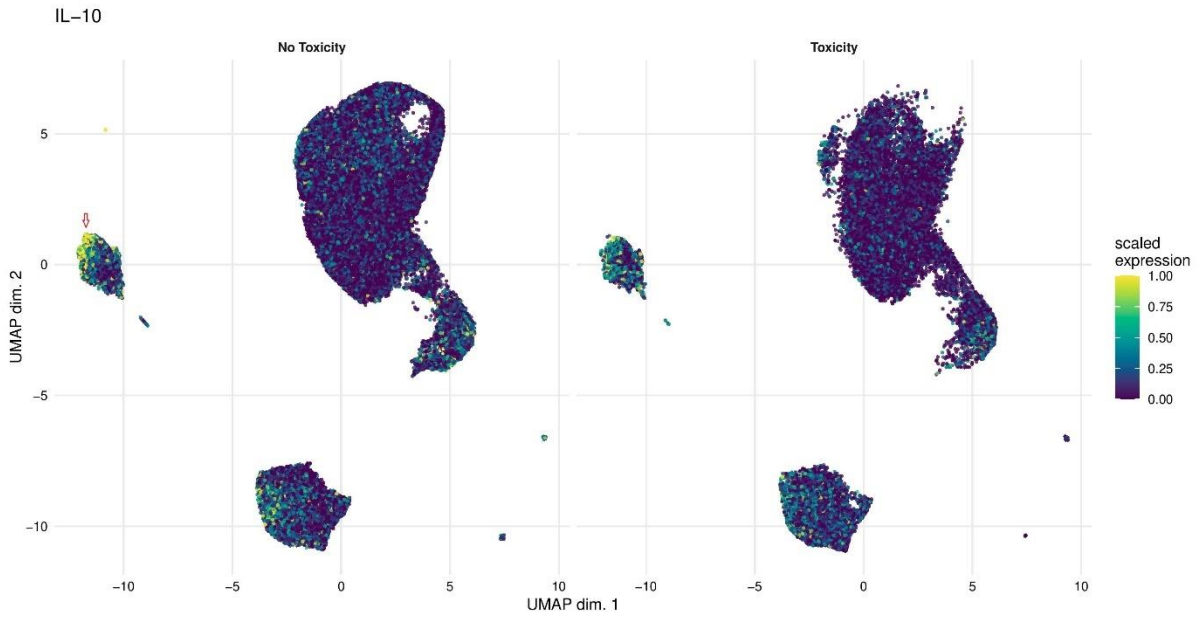


Figure 4.12. Bulk expression UMAP stratified according to IL-10 expression, which demonstrates less intense IL-10 expression in the toxicity cohort compared to the non-toxicity patients. Expression is largely confined to the Breg clusters demarcated by the red arrows. Matched IL10 expression in the corresponding clusters in the toxicity patients is reduced.

4.3.3.5.2 Melanoma

We performed a similar validation in an independent cohort of melanoma patients from a different institution (King's College London) and where the CyTOF analysis was performed. This was to assess whether the noted B cell changes in NSCLC were disease specific. This cohort contained 15 treatment naïve patients, all of whom were treated with anti-PD1 checkpoint blockade therapy. Six patients developed high-grade irAE. FlowSOM clustering in the same manner revealed IL10 positive populations with significant similarity to the populations seen in our NSCLC cohorts, both in terms of phenotype and proportion (Figure

4.13). Eight total IL10 positive clusters were detected with six matching the phenotypes already described.

A. Cluster 2 (2.16%) – ki67^{int} IL10^{int} IgG^{hi} CD24+ CD25^{hi} CD27^{hi} Br1 Breg

B. Cluster 3 (2.19%) – ki67^{int} IL10^{int} CD24+ CD27^{hi} CD38^{hi} CD95^{hi} Plasmablast

C. Cluster 4 (1.61%) – ki67^{hi} IL10^{int} CD24^{lo} CD27^{hi} CD38^{hi} CD95^{hi} Plasmablast

D. Cluster 9 (0.49%) – IL10^{lo}, CD25^{int} CD38^{hi} CD95^{hi} CD24- CD27^{lo} Br1 subtype

E. Cluster 16 (2.17%) – PDL1^{hi} TGF-β^{hi} CD24+ CD38+ Immature Breg

F. Cluster 20 (0.55%) – CD5+ CD25+ PD1+ IL10$^{lo/-}$ Breg

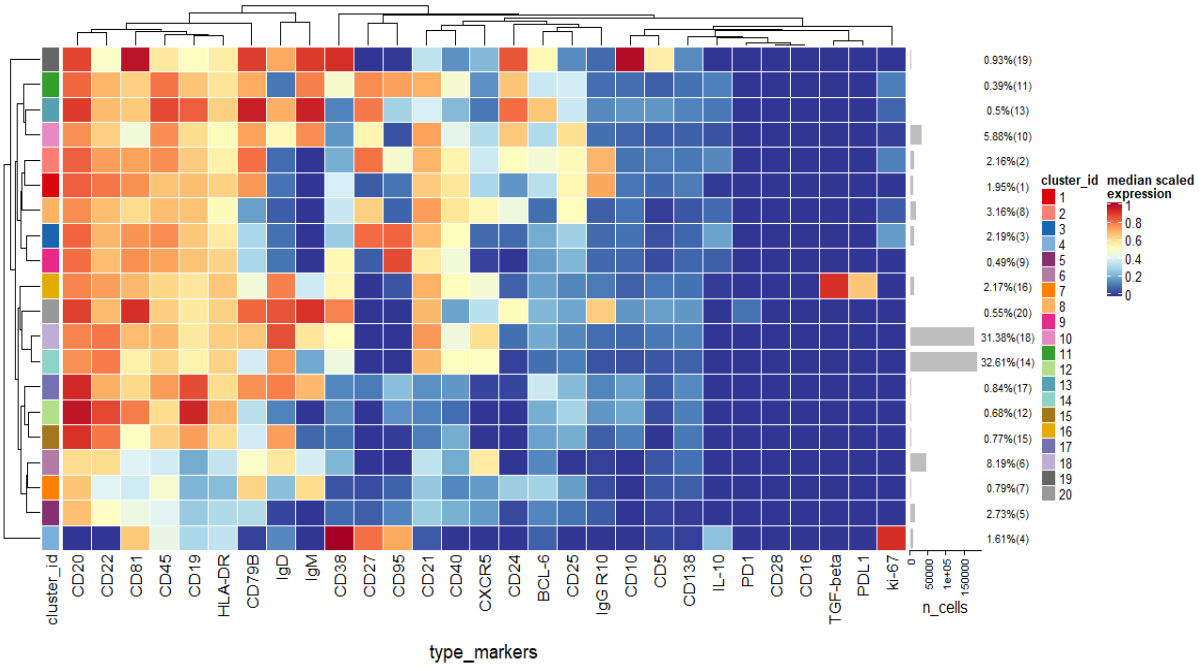


Figure 4.13. Heatmap showing normalised expression of the base B cell panel markers for the 20 B cell clusters identified with FlowSOM in melanoma patients. The cluster IDs and relative frequencies are displayed as a bar graph on the right-hand side, along with the median scaled expression.

Again, using comparative expression UMAP, diffusion map and boxplots between those who did and did not experience high-grade toxicity on checkpoint blockade, we demonstrated increased expression of these six pertinent IL10 clusters in non-toxicity patients (Figure 4.14).

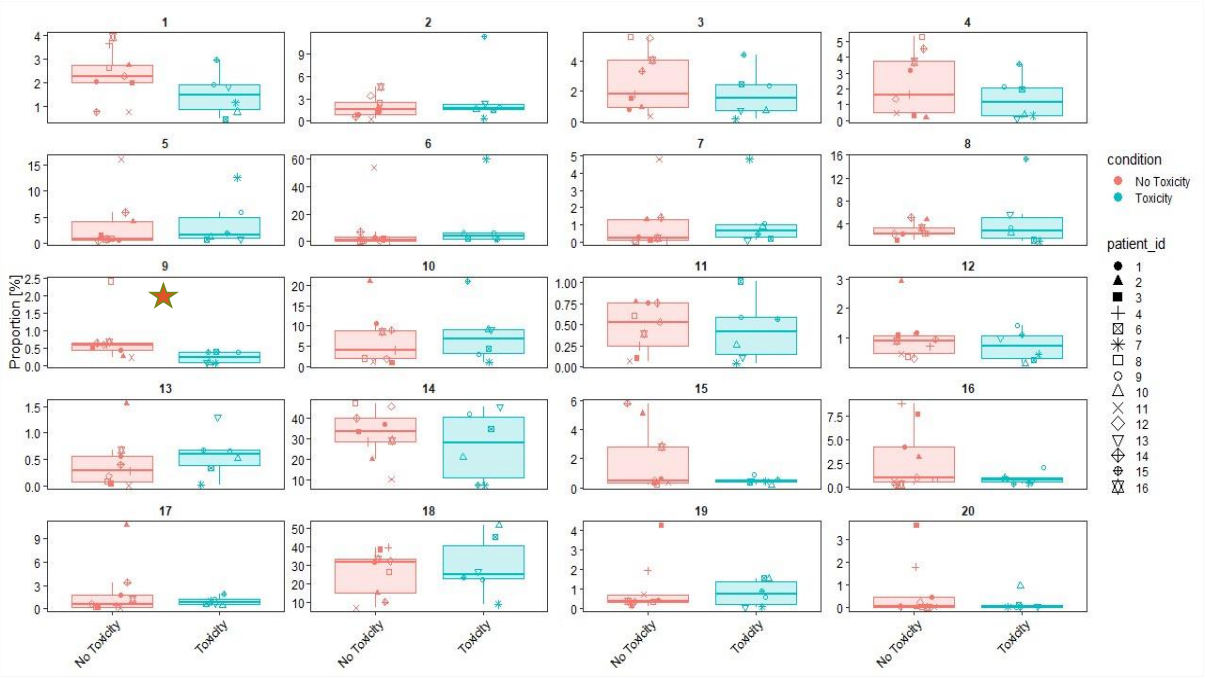


Figure 4.14. Boxplot representation highlighting differential cluster abundance between toxicity and non-toxicity melanoma patients. Suppressive clusters include 2, 3, 4, 9, 16 and 20. Cluster 9 which is demarcated by the red star, is the only statistically significant cluster that was differential abundant between the two conditions. This represented a suppressive B10 population which was more abundant in non-toxicity patients (p=0.043).

Loss of the PD1+ cluster is most striking on the comparative UMAP (Figure 4.15) and the added information from the diffusion map shows lower abundance of the Plasmablast population (cluster 4 which is most IL10 positive population), as well as cluster 20 (light grey, PD1+), cluster 9 (Br1 – pink cluster) and cluster 3 (dark blue, also Plasmablast).

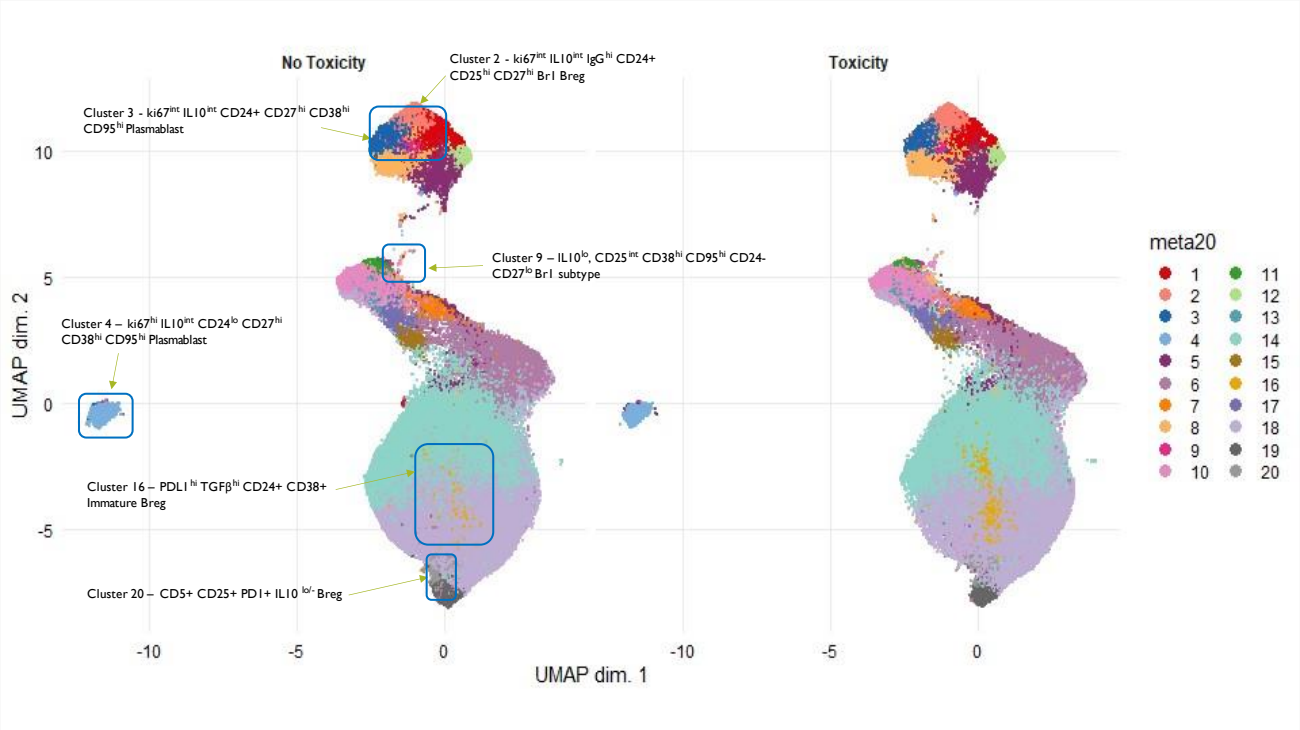


Figure 4.15. UMAP plots stratified according to condition, "no toxicity" and "toxicity". Clusters highlighted by blue boxes correspond to clusters identified from (4.13). These clusters are attenuated in the "toxicity" cohort in this unsupervised analysis. All samples are randomly downsampled to account for equally representative populations across samples.

Differential abundance testing revealed significantly higher expression of cluster 9 (Br1 subtype) in non-toxicity patients (p=0.043). Although visually the dimensionality reduction maps are not as striking as the comparative maps in NSCLC, despite well dichotomised cohorts, one must consider the biologically very different disease processes. The NSCLC cohorts had a median age of 65-70 compared to a much younger melanoma cohort of 50 years. Furthermore, all patients in the NSCLC cohorts were either ex- or current smokers which was not the case for the melanoma patients. These factors will likely influence the degree of inflammation in the tumour microenvironment and the subsequent *in vivo*

induction of these suppressive B cell phenotypes. The differences in Breg biology between different cancers will certainly require further exploration.

To the best of our knowledge, these data provide the first detailed analysis of the association of circulating IL10 positive B regulatory cells with the development of severe autoimmunity following immunotherapy in NSCLC patients. Patients developing grade 3/4 irAEs have a significant phenotypic defect in their regulatory B cell repertoire prior to commencing immunotherapy. Given the variability in these differences on an individual patient level, we sought to explore the functional B cell biology in these patients with respect to suppressive function. This is detailed in section 4.4.

4.3.4 Deep B cell Phenotype associates with Response to Checkpoint Blockade Treatment [Pre-Treatment]

We performed comparative analyses between responders (n=26) and non-responders (n=20) using dimensionality reduction methods. The most noticeable differences related to two clusters both of which were PDL1 positive, and likely suppressive phenotypes as shown by IL10 and TGF-β expression.

A. $PDL1^{int/hi}$ $CD38^{hi}$ $CD95^{hi}$ $TGF\beta^{-ve}$ $IL\text{-}10^{lo}$ - Mustard Cluster (16 – 3.86%)

B. $PDL1^{hi}$ $CD38^{int}$ $CD95^{int}$ $TGF\beta^{int}$ $IL\text{-}10^{lo}$ - Deep Lilac Cluster (17 – 0.35%)

Subtle blunting of these two clusters was demonstrated on UMAP and diffusion map in responders. Stratification according to IL-10 and PDL1 expression (Figure 4.16) showed noticeably higher expression of PDL1 in cluster 16 in non-responders and slightly higher IL10 in clusters 19 and 20 (grey island on UMAP) also in non-responders, although not as striking.

Clusters 19 and 20 are both ki67^{hi} Plasmablast populations as previously described and expressed the highest amount of IL10 in the B cell repertoire. Differential abundance analysis indicated significantly higher expression of the likely immature PDL1 Breg phenotypes; 16 and 17 in non-responders (p=0.002 and p=0.00022 respectively).

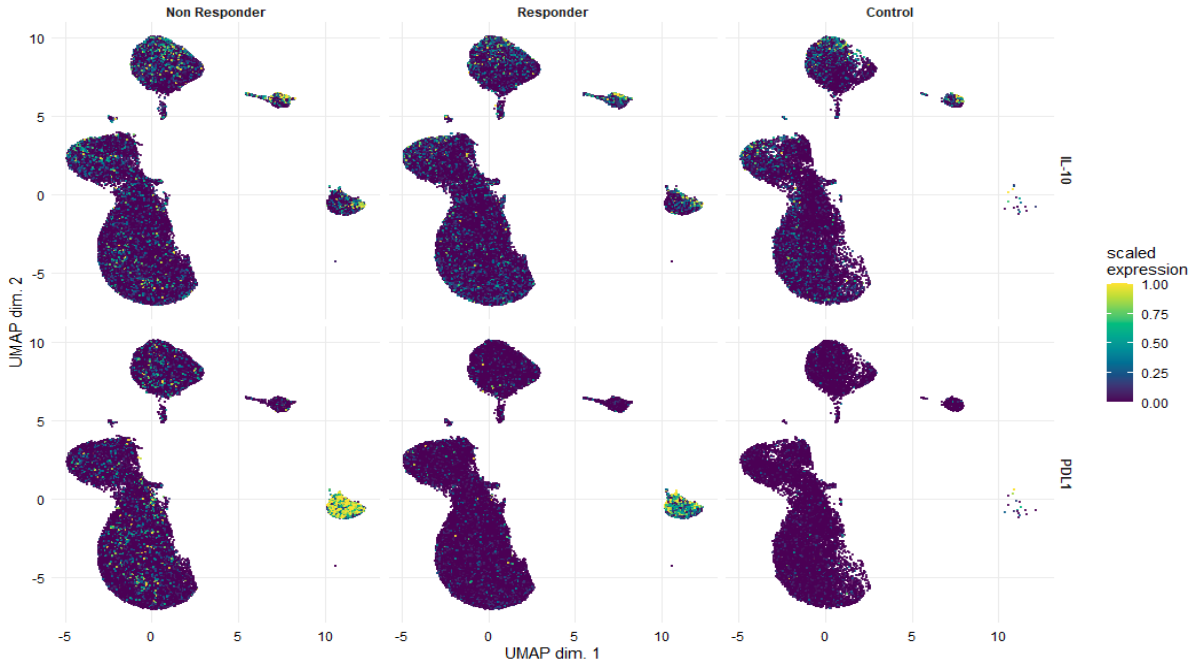


Figure 4.16. UMAP plots stratified according to overall population IL-10 and PD-L1 expression and by condition. Expression is largely confined to the suppressive B cell clusters demarcated by the red arrows, however there is no major difference in IL-10 expression between conditions, but there is higher expression of PD-L1 in non-responders. Both PDL1+ clusters (green box) were significantly more abundant in non-responders when analysed with differential abundance analysis (p<0.05, Benjamini-Hochberg correction).

4.3.4.1 Validation of B cell changes associated with Response to CCB in Melanoma [Analysis of a public dataset]

We sought to seek corroboration of our findings from 4.3.4 in a publicly available mass cytometry dataset in melanoma (298). We gated the files for CD19+ positive cells and utilised

the panel's B cell markers to cluster the cells. The cohort was split into those who received Pembrolizumab (anti-PD1, n=40) or Ipilimumab (anti-CTLA4, n=24) as first-line therapy.

In the Pembrolizumab cohort, the pre-treatment circulating B cell profile demonstrated presence of PDL1+ cells and PD1+ cells. Differential abundance testing revealed significantly higher expression of PD1+ CD25+ CD38+ IL10+ Bregs in non-responders, which phenotypically mapped to cluster 9 from our toxicity analysis (p=0.00069) but no other clusters. This cluster existed in those patients who went on to receive Ipilimumab therapy and was indeed higher in non-responders albeit not significant on statistical modelling. Although the exact clusters were not shown to be significant in this separate dataset, the presence of suppressive IL10 positive B cell populations in those who patients who fail to respond to treatment provides validation that suppression of key anti-tumour responses (whether they are T or B cell driven) is driving poor outcomes in these patients.

From our dataset, aside from the PDL1+ cells, no further link with TLCs and response to treatment was observed. The relationship between toxicity and response to treatment was not significant (p=0.121 by Chi Squared with Yates' correction).

4.3.5 Longitudinal Changes on Treatment with Checkpoint Blockade

We performed longitudinal analysis comparing the circulating B cell profiles between paired pre- and post-treatment peripheral blood samples (n=19) from our initial NSCLC cohort. All patients received anti-PD1 therapy. Comparative diffusion maps did not display any grossly obvious differences between pre- and post-treatment samples however, we did note significant on treatment changes in circulating Breg profiles. We interrogated the Breg phenotypes identified.

4.3.5.1 Longitudinal changes on immunotherapy in Breg subsets that distinguish between toxicity and non-toxicity patients

We assessed the effect of immunotherapy on the identified Breg populations (Figure 4.17). In those patients who did not experience irAE, treatment induced a significant increase in circulating PD1^{hi} and PDL1^{hi} TGF-β^{hi} Breg cells with no change in the circulating PDL1^{hi} TGF-β^{-ve} or B10 type populations. However, in those who did experience severe toxicity, there was no significant increase in any of the Breg subsets and the level of the PDL1^{hi} TGF-β^{hi} and the B10-type populations decreased.

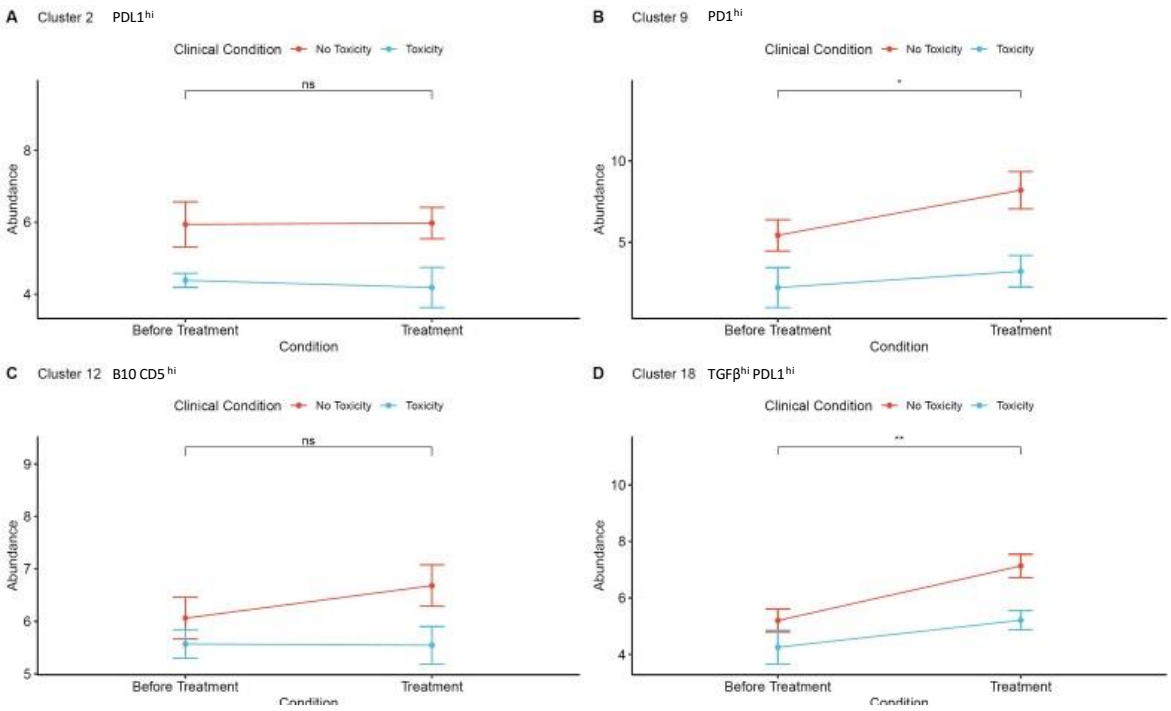


Figure 4.17. Longitudinal analyses demonstrate significant increases in circulating Breg number post-treatment with checkpoint blockade therapy. Line plots indicate the fold change in circulating Breg cluster abundance before and after treatment with sub-stratification according to those who developed high-grade irAE after treatment. Changes are shown for four major Breg phenotypes identified in the initial unsupervised analysis. None of the patients who developed toxicity post-treatment patients experienced significant increases in circulating population number. Statistical

significance was determined by pairwise Wilcoxon signed rank test (significance level p<0.05). *P<0.05, **P<0.01, ***P<0.001.

4.4 Functional B cell Assays

We sought to further explore the biology of Breg deficiency in toxicity patients by carrying out functional ex vivo assays to determine the nature of the cytokine response in these advanced disease patients. We induced Bregs ex vivo and cultured/stimulated them alone and in co-culture with autologous CD4+ T cells from toxicity and non-toxicity patients as defined in the key methods section 2.3.3.

4.4.1 Ex vivo induction of primary B cell derived cytokines is reduced in patients who develop high-grade irAE on combination checkpoint blockade

4.4.1.1 Anti-inflammatory IL-10 deficit

We analysed all 8 toxicity patients, 29 non-toxicity patients and 8 healthy controls from the initial cohort (investigators were blinded to the outcome of these patients). TLR9 ligation has been shown to induce a regulatory phenotype in B cells with subsequent IL-10 production (299). Subsequent stimulation of B cells ex vivo with Phorbol 12-myristate 13-acetate (PMA) and Ionomycin resulted in a potent induction at 40 hours post initial stimulation, measured through intracellular cytokine staining. There was a considerable decrease in B cell derived IL-10 production in those patients developing high-grade post-treatment irAE (Figure 4.18). We detected significantly lower levels of free IL10 in the cell culture supernatants from toxicity patients as measured through IL-10 Quantikine ELISA (Figure 4.19). This was the case both in terms of IL-10 absorbance and concentration (pg/ml).

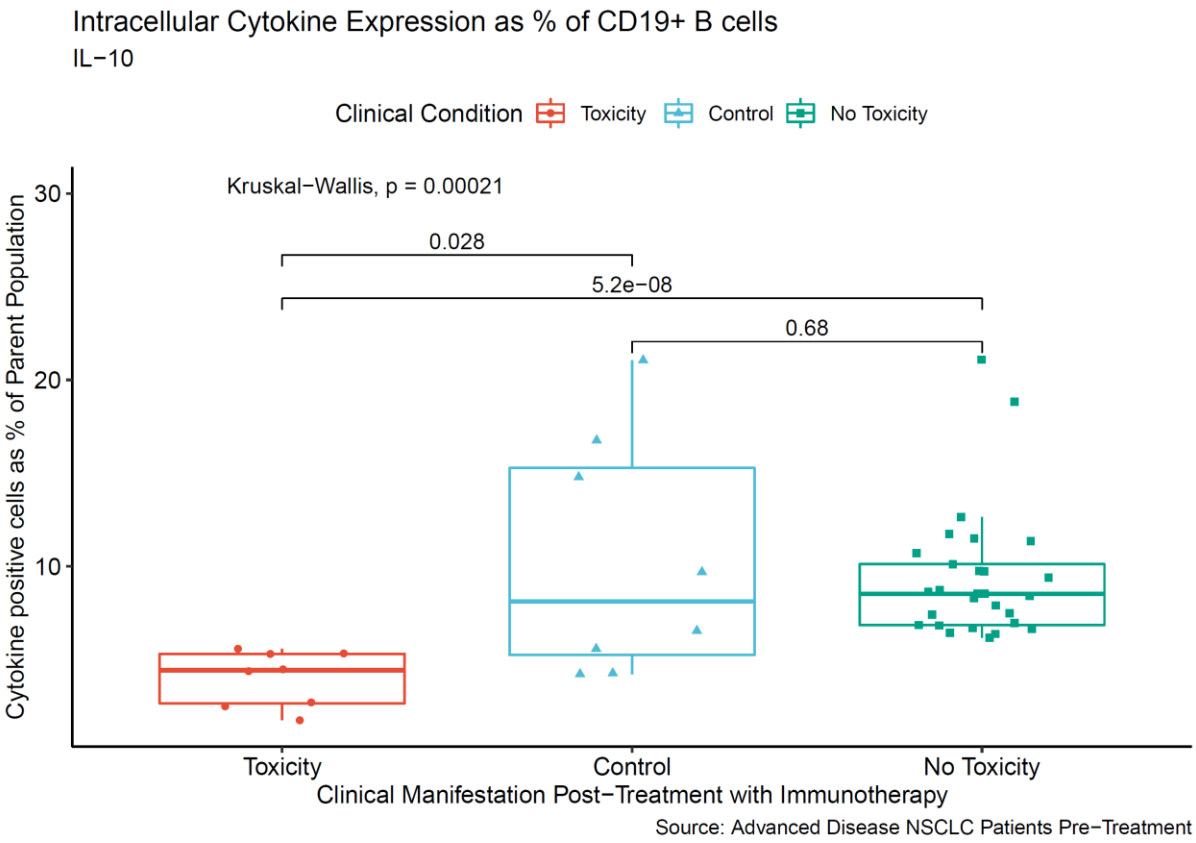


Figure 4.18. CD19+ B cell IL-10 expression in NSCLC patients stratified by the development of high-grade post-treatment irAE. Data are presented as % of total CD19+ population, with points indicating individual patients. Statistical analysis was conducted using Wilcoxon Rank Sums test (to test for specific inter-group differences) and Kruskal Wallis (to generally test for overall differences between all three groups). All analyses were conducted using a Benjamini-Hochberg multiple comparisons correction, comparing all conditions as indicated. A significance level of <0.05 was adopted.

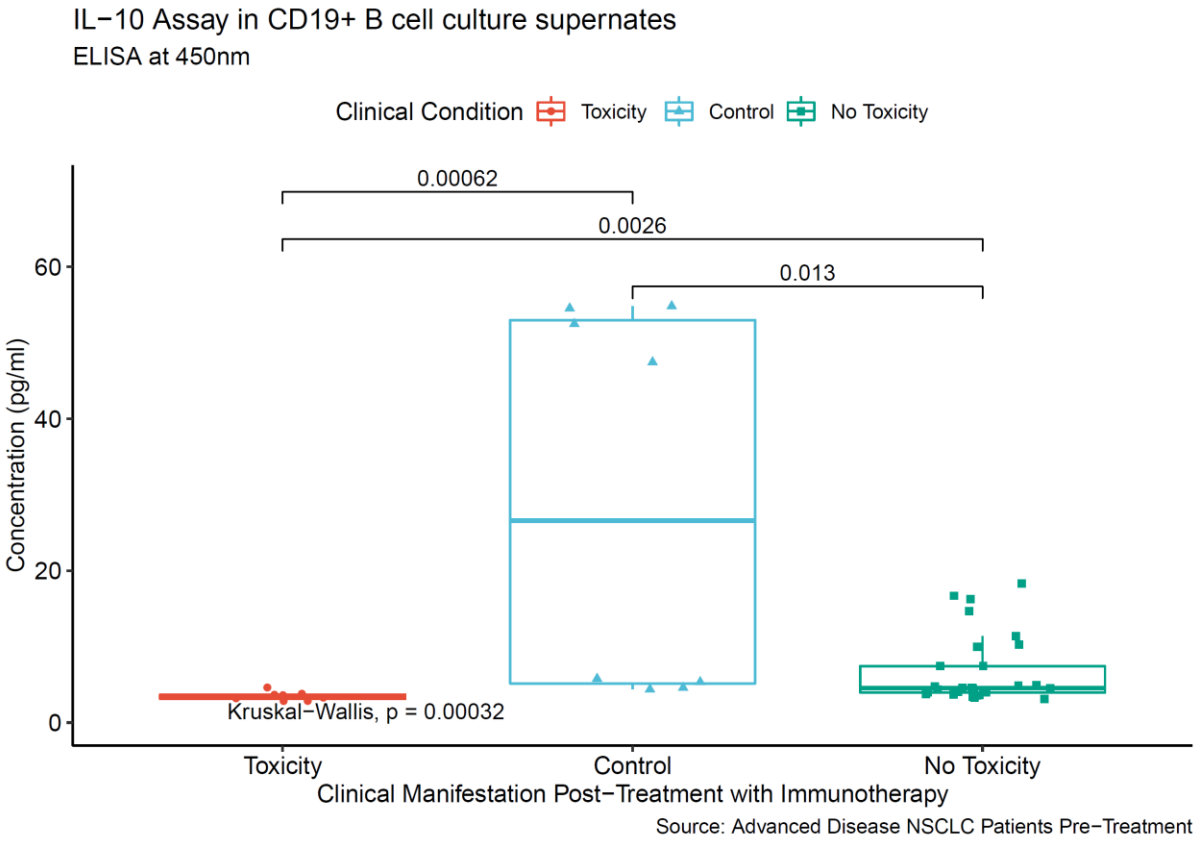


Figure 4.19. ELISA of CD19+ B cell culture supernatants in NSCLC patients stratified by the development of high-grade post-treatment irAE. Data are presented as absorbance (nm) and concentration (pg/ml), with mean values portrayed. Individual points indicating individual patients. Statistical analysis was conducted using Wilcoxon Rank Sums test and Kruskal Wallis followed by a Benjamini-Hochberg multiple comparisons correction, comparing all conditions as indicated; **P<0.05, **P<0.01, ***P<0.001.*

4.4.1.2 Pro-inflammatory cytokine failure in B cells

On further investigation, we detected a significantly lower expression of a range of pro-inflammatory cytokines from the B cells of patients who experienced high grade toxicity. We examined IFN-γ, TNF-α, IL-6, IL-17 and IL-2 (Figure 4.20). The pattern for IFN-γ expression was similar to that of IL-10 - significantly higher expression in non-toxicity patients when compared with toxicity patients and no difference between healthy donors and no toxicity patients (Figure 4.21). There was also statistically significantly increased expression of IL-6

and IL-17 in the non-toxicity cohort compared with the toxicity cohort albeit the differences were less marked when compared to IL-10. Healthy controls produced the most TNF-α with no significant difference between toxicity and non-toxicity patients. As shown in Figure 4.20, there is quite some variability in the different cytokine production across the different conditions. This is not only true for the toxicity and non-toxicity conditions (cancer patients) but also in healthy controls. It has been demonstrated that healthy control PBMCs when stimulated ex vivo produce cytokines across a spectrum of concentrations, from low natural producers to very high natural producers (111). This natural variation in cytokine production could well be what we are seeing in our healthy controls but also in the cancer patients, the same variation is likely to be exaggerated if the tumours are inducing or indeed inhibiting cytokine responses.

Thus, there appears to be a poly-functional failure in B cells derived from toxicity patient that affects both pro- and anti-inflammatory cytokines. Although this seems to be most marked for IL-10, the global cytokine deficit suggests a form B cell exhaustion in these patients.

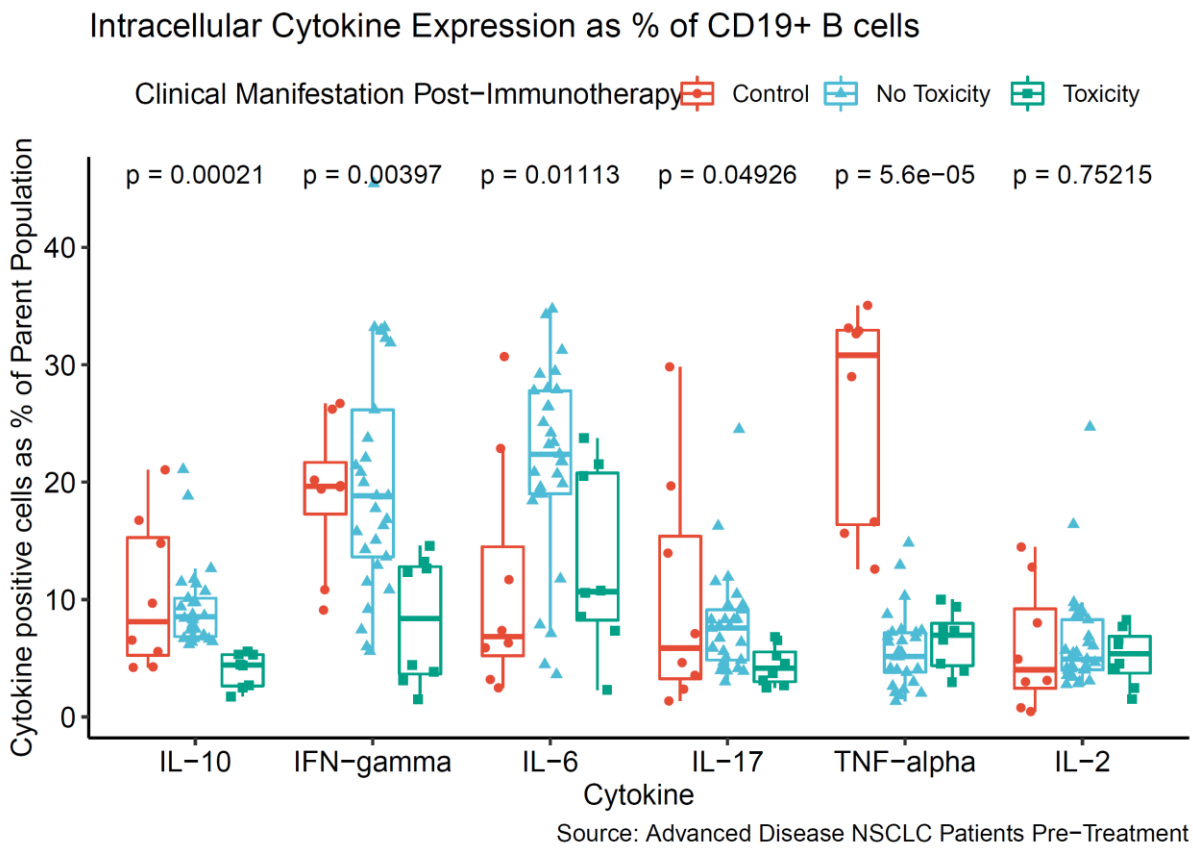


Figure 4.20. CD19+ B cell pan-cytokine expression in NSCLC patients stratified by the development of high-grade post-treatment irAE. Data are presented as % of total CD19+ population, (mean), with points indicating individual patients. Statistical analysis was conducted using Wilcoxon Rank Sums test and Kruskal Wallis followed by a Benjamini-Hochberg multiple comparisons correction, comparing all conditions as indicated; **P<0.05, **P<0.01, ***P<0.001.*

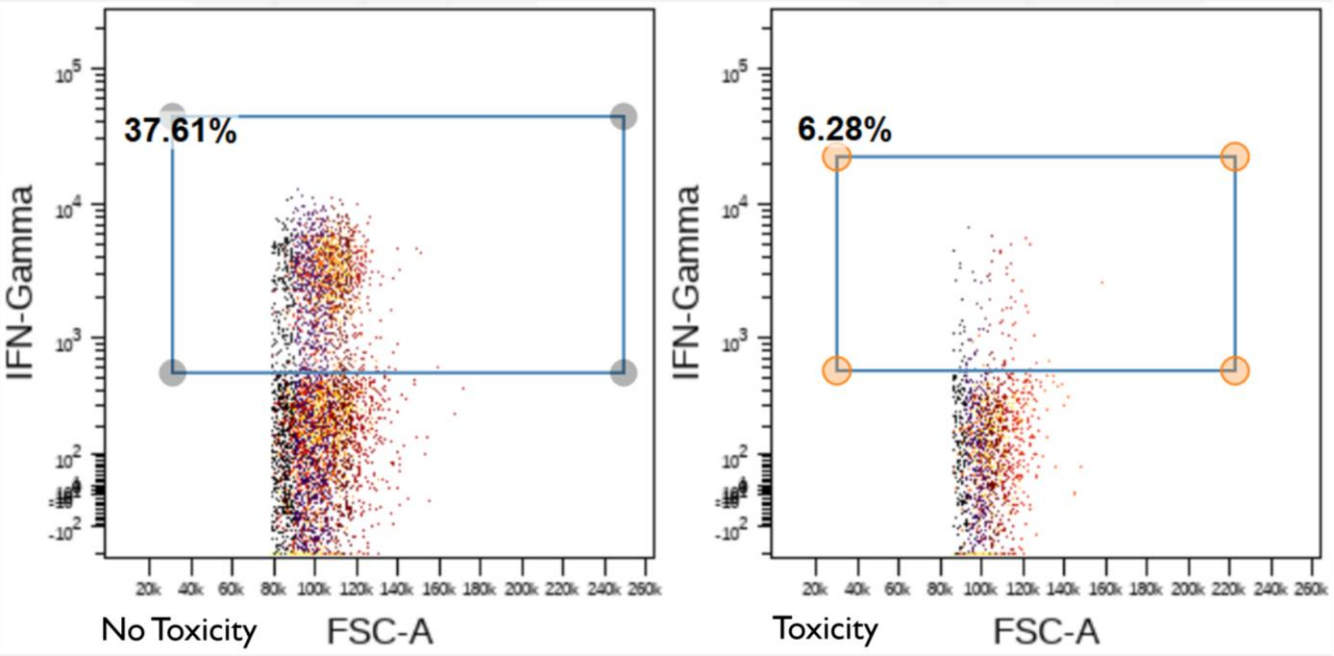


Figure 4.21. Representative and quantitative plots of CD19+ B cell IFN-γ production after culture alone. Total $CD19^{+}$ B cells were stimulated with CpG (TLR9), and IL-2. Cells were incubated for 40 hours, including re-stimulation with PMA and Ionomycin in the final 3 hours before harvesting and analysis by flow cytometry.

4.4.2 B cell suppression of CD4+ Th1 responses is diminished in patients who develop high-grade irAE on CCB

Human B cells are known to exhibit suppressive capacity *in vitro*. The can suppress both innate and adaptive cells, including monocyte pro-inflammatory cytokine release, and a number of Th cytokines, mediated through the release of IL-10 (104). TLR9-activated B cells were able to suppress IFNγ production by T cells activated with αCD3/28 in healthy cohorts and in patients who did not develop high-grade irAE on treatment with checkpoint blockade therapy in a dose-dependent manner (Figures 4.22).

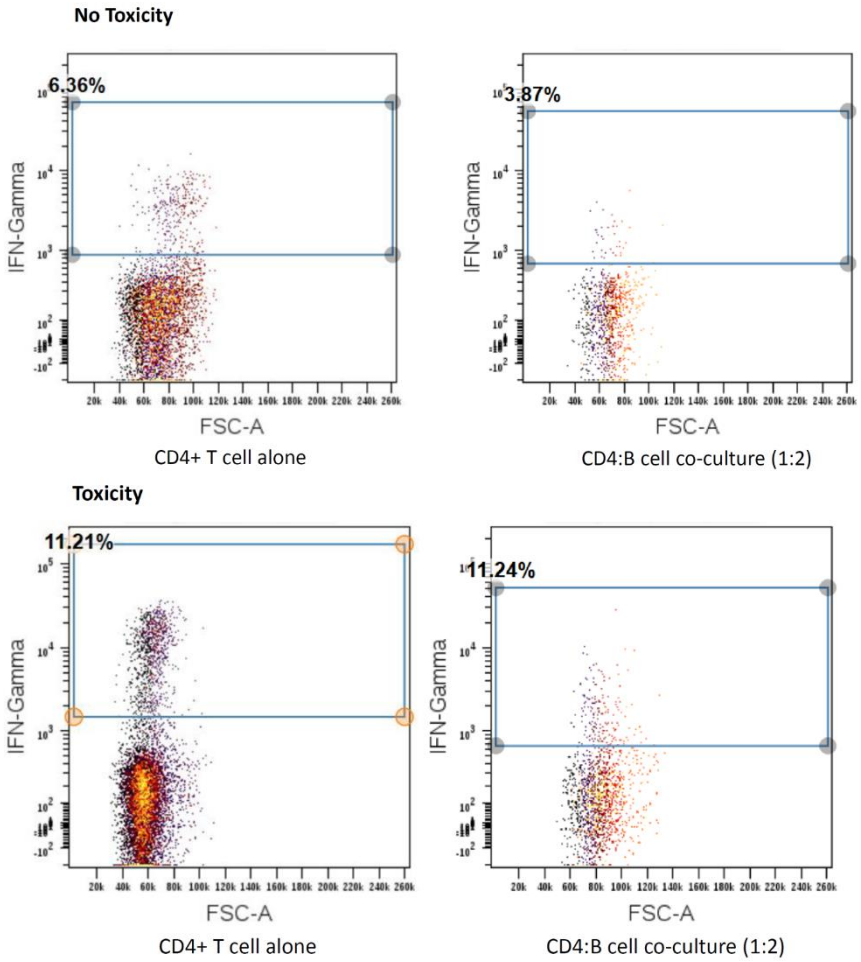


Figure 4.22. Representative and quantitative plots of CD4+ T cell IFNγ production after culture alone (left hand panels), co-culture with CD19+ B cells (right hand panels) for both non-toxicity (top row) and toxicity (bottom row) patients. Cell trace violet (CTV)-labelled CD4+ T cells stimulated with αCD3/28 for 4 days, including re-stimulation with PMA+IO in the final 5 hours before harvesting and analysis of IFNγ production, or CTV dilution by flow cytometry (n=16, total). IFNγ production assessed after co-culture with autologous CD19+ B cells in all patients.

However, this suppression was not seen in patients who developed high-grade irAEs at low (1:1) and high (1:4) B cell concentrations. Figure 4.23 shows a stepwise decrease in IFNγ production with increasing concentration of CD19+ B cells in co-culture, whilst this trend is also relatively preserved in toxicity patients, there is no inter-group or overall statistical significance. Whereas there is statistical significance in the non-toxicity and control cohorts.

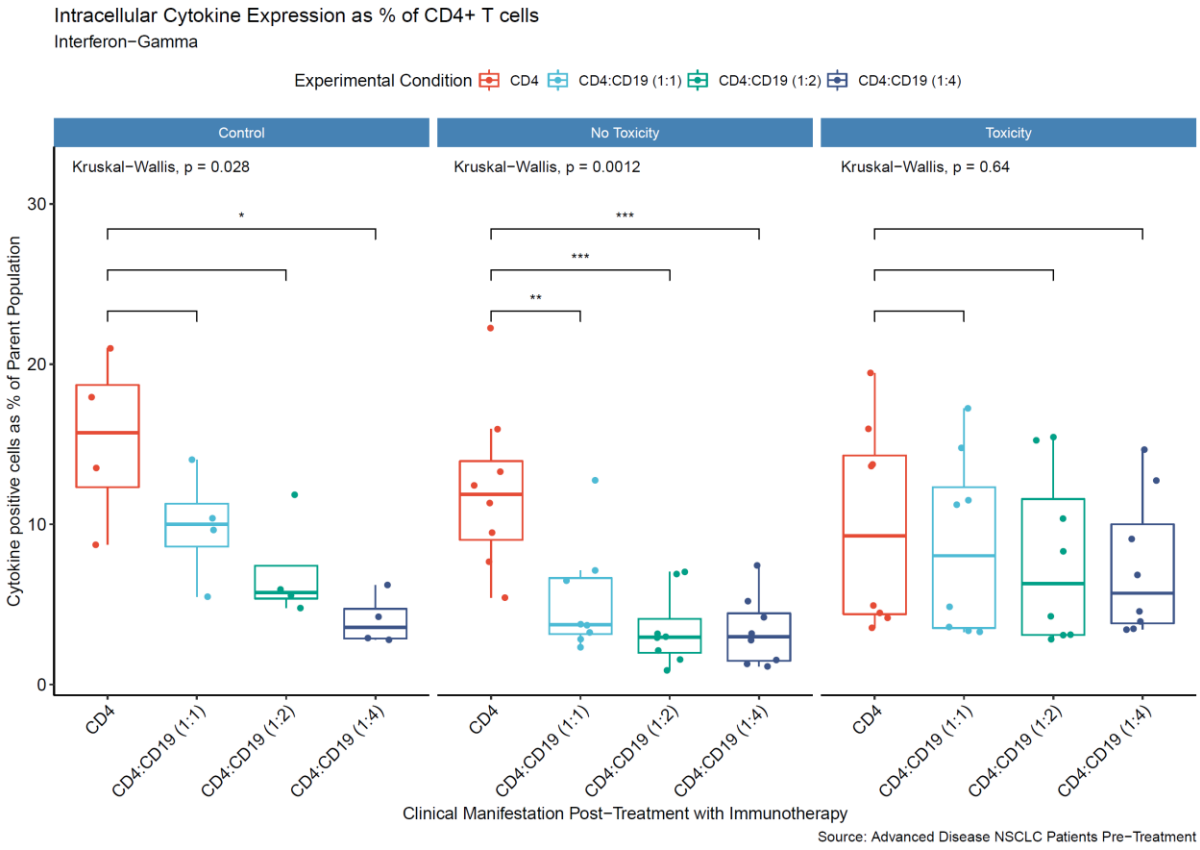


Figure 4.23. IFNγ production by maximally stimulated CD4+ cells alone and in co-culture with increasing doses of paired autologous primary CD19+ B cells. Data points indicate individual patients. Statistical analysis was conducted using Wilcoxon Rank Sums test and Kruskal Wallis followed by a Benjamini-Hochberg multiple comparisons correction, comparing all conditions as indicated; *P<0.05, **P<0.01, ***P<0.001.

Similarly, significantly higher IFNγ suppression indices were seen in non-toxicity and control patients with respect to toxicity patients at all B cell concentrations (Figure 4.24). Neutralisation of IL-10 within similar co-cultures has resulted in a loss of suppression (300,301) confirming the dependency on IL-10 to suppress T cell responses, and it was also demonstrated that concentrations of IL-10 upon TLR9 activation (>1ng/ml) were sufficient to directly suppress IFNγ production in CD4+ T cells. These data demonstrate an inherent functional defect in Breg IL10 effector response in lung cancer patients who experience high-

grade autoimmune toxicity on checkpoint blockade therapy, which results in an inability to suppress autologous CD4+ Th1 responses.

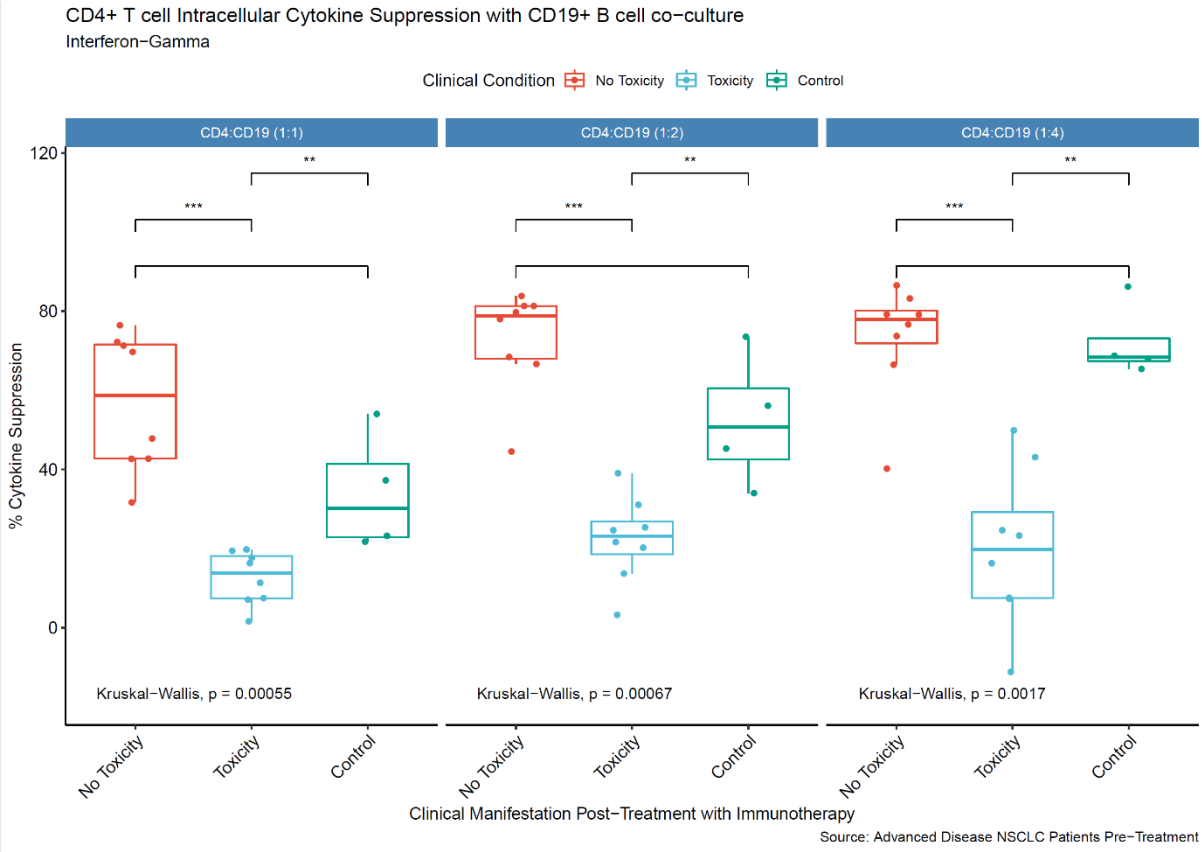


Figure 4.24. IFNγ suppression calculated by ((x-y/x) *100, where x = IFNγ production by maximally stimulated and activated CD4+ T cells, y = IFNγ production in co-cultured T cells at varying CD4:CD19+ B cell ratios. Data points indicate individual patients. Statistical analysis was conducted using Wilcoxon Rank Sums test and Kruskal Wallis followed by a Benjamini-Hochberg multiple comparisons correction, comparing all conditions as indicated; **P<0.05, **P<0.01, ***P<0.001.*

4.5 Functional Genomics Analysis [Public RNAseq Datasets]

Numerous exploratory strategies have been taken to improve the response to checkpoint blockade in cancer. We sought to interrogate public bulk RNA sequencing transcriptomics datasets in NSCLC and melanoma as a means of exploring any predictive biomarkers for

response to checkpoint blockade and validate the single cell findings from our high-dimensional data in NSCLC (section 4.3.4).

4.5.1 NSCLC

Cho J-W et al (GSE126044) analysed the tumour DNA methylome in NSCLC as a means of identifying differentially methylated promotors or enhancers associated with response to anti-PD1 therapy (302). We performed Differential Gene Expression (DGE) analysis on the bulk raw data and identified 458 differentially expressed genes ($p<0.05$) between responders and non-responders. Of these, 287 were positively upregulated in responders and 171 were down regulated with respect to responders. Insight into specific gene pathways with KEGG analysis showed in responders, a significantly higher expression of genes associated with antigen presentation and co-stimulation of lymphoid cells; MHC class II, CD80/86 and indeed genes associated with cytotoxic T lymphocytes such as Granzyme B and Perforin. Gene Ontology analysis corroborated these findings, highlighting genes associated with the adaptive immune response, neutrophil and granulocyte activation and regulation of the inflammatory response as the most significantly enriched in responders. No specific genomic correlation between response and suppressive B cell function was found on interrogation of this dataset.

4.5.2 Melanoma

Auslander et al (GSE115821) analysed similar “omics” and response data in melanoma patients (303). In a similar manner to the above, we performed DGE analysis on the bulk raw data and identified data and identified 668 differentially expressed genes ($p<0.05$) between responders and non-responders. Of these, 446 were positively upregulated in responders and 222 were down regulated with respect to responders. KEGG analysis identified B cell specific

lineage factors as upregulated in responders, such as CD19, CD20, CD21 and B cell survival factor, BAFF. As with NSCLC, the gene ontology analysis found genes associated with the adaptive immune response, B cell activation and B cell/Ig-mediated immunity to be significantly enriched in responders (gene ontology figures). The emphasis from the genomic data revealed biomarkers strongly associated with response as opposed to those associated with poor response. This was the case for both disease types.

4.6 Discussion

As the use of checkpoint inhibitors in cancer has increased, the incidence of irAEs has become a significant health care issue and predictive biomarkers of those at risk of developing irAEs is of crucial importance in limiting excessive toxicity and maximising the risk:benefit equation for the use of these agents. Numerous studies have demonstrated the role of B regulatory cells (Bregs) in limiting immune reactivity. Bregs can suppress immune reactivity against exogenous and endogenous antigen and in particular act as a cellular brake on the development and severity of autoimmunity often via IL-10 production. Through testing in independent cohorts, we show here that lung cancer patients developing severe toxicity from anti PD-1 +/- chemotherapy had significant gaps in their B cell repertoire specifically affecting IL-10 expressing B cell subsets. IL-10 production is a cardinal feature of Bregs, and all subsets described have strong phenotypic parallels with previously described Breg subsets. Patients' not developing severe toxicity had preservation of these subsets which were functionally competent, and, in these patients, levels of these cells increased on checkpoint blockade whereas in toxicity patients they either remained static or fell.

Of these Bregs subsets, we noted in our NSCLC cohort a separate validation cohort in melanoma, preponderance of PD1+ and PDL1+ Bregs in non-responders. The pro-

tumourigenic effects of Bregs on the anti-tumour microenvironment has been well described; via the production of IL10, IL35 and TGF-β, these suppressive cells enhance the tolerability of the TME to the host immune response (304–306)

Aberrancy in Breg number and function has been identified in numerous immune-related pathologies such as autoimmune disease, chronic infection, cancer and in transplant rejection (104,307). The hallmark of Breg identification and enumeration is considered to be IL-10 production which is the major effector mechanism of suppression. Unlike Tregs there is no unequivocal lineage marker by which to capture this heterogeneous subset (104). The original description of these suppressive cells was noted by Janeway et al who noted that B-cell deficient mice developed an exacerbated form of experimental autoimmune encephalomyelitis (EAE) (308). The protective role of IL10 producing B cells was then confirmed in an EAE murine model by Fillatreu and colleagues (105), after which various other regulatory B cell subsets were described (110,180). Unique Breg subsets have been implicated in maintaining immune tolerance in allergen-induced hypersensitivities (179) as well as protecting against allograft rejection in murine models of transplantation and using IL-10 as a biomarker for predicting clinical outcomes in patients undergoing transplantation (309). More recently, reduced numbers of IL10 producing B cells have been suggested as contributory to severity of disease in SARS-CoV-2 infection (310,311).

Through testing in independent cohorts, we show here that lung cancer patients developing severe toxicity from anti PD-1/PD-L1 blockade had significant gaps in their B cell profile at diagnosis, specifically affecting suppressive B cell subsets (IL-10+, TGF-β+ or PDL-1+). All subsets detected in this context display strong phenotypic parallels with previously described Breg subsets. Patients who did not subsequently develop severe toxicity had preservation of

these subsets at diagnosis and in these patients, the levels of these cells increased on checkpoint blockade whereas in toxicity patients Breg populations they either remained static or fell. Our phenotyping data suggests there is an attenuated presence of circulating Breg populations in toxicity patients and our functional data suggests that the B cells in toxicity patients not only fail to mount an adequate anti-inflammatory cytokine (IL-10) response but also an adequate pro-inflammatory cytokine (IFN-γ, IL-6, IL-17) response. Pro-inflammatory cytokine production has been demonstrated by B cells ex vivo and in the setting of checkpoint blockade in melanoma (312,313). TLR9 ligation upregulates B cell IL-6 production and PMA/ionomycin drives IFNγ production (314). Both of these components are part of the Breg IL-10 stimulation cocktail and neither IL-10, IL-6 or IFNγ production was seen in B cells from patients developing toxicity. This pan-cytokine failure in B cells from toxicity patients may explain the lack of a reciprocal relationship between toxicity and response to treatment. The nature and mechanistic underpinning of this polyfunctional defect and which implies B cell exhaustion in toxicity patients is unknown and is a key part of our ongoing studies.

B cell depletion significantly enhanced the local reaction to applied antigen in a contact hypersensitivity model where inflammation is T cell dependent (173). This was due to the significant reduction of a splenic IL-10 producing $CD1d^{hi}$ CD5+ B cell subset - B10 cells. Suppression of T cell driven inflammation was IL-10 dependent and in this model, IL-10 production was restricted to this particular B cell subset. Adoptive transfer of $CD1d^{hi}$ CD5+ B cells into $CD19^{-/-}$ mice inhibited hypersensitivity responses but transfer of $CD1d^{hi}$ CD5+ B cells from unsensitised mice or from mice sensitised with an unrelated antigen had no effect suggesting that these B10 cells were likely to be antigen specific. B cell depletion prior to EAE induction with myelin oligodendrocyte glycoprotein (MOG) significantly exacerbated the resulting disease (315). B cell depletion prior to immunisation led to significant increases in

MOG-specific T effector cells and IFN-γ and IL-17 producing CD4+ cells. Adoptive transfer of modest numbers of B10 cells completely reversed the accelerated disease of B cell depleted mice prior to MOG immunisation. B10 function required IL-21R signalling, CD40 interaction and class II expression (316). A model was proposed whereby B cells capture autoantigen thus triggering BCR signalling and driving IL-10 producing capacity. These B cells present antigen via class II in a cognate interaction to CD4+ cells, which produce local IL-21 inducing B cell IL-10 production and subsequent T cell suppression. In humans, B10+ and B10pro cells are a subset of $CD24^{hi}$ and CD27+ B cells. Clusters 12 and 17 (from the initial NSCLC dataset, n=46) described herein were $CD1d^{int/hi}$ $CD5^{hi}$ $CD24^{lo}$ $CD25^{lo}$ $CD27^{lo}$ IL-10^{lo} and $CD24^{hi}$ $CD25^{hi}$ $CD27^{hi}$ IL-10^{lo} respectively, the latter of which have the highest CD24 expression of our identified Bregs. These B cell subsets were significantly reduced in toxicity patients.

Three of our patients developed severe colitis, one of the most serious autoimmune complications of checkpoint blockade therapy. In the DSS model of ulcerative colitis (UC), $CD19^{-/-}$ mice were more susceptible to DSS-induced intestinal damage and there was greater intestinal infiltration of neutrophils and T cells (317). The proportion of splenic B cells was inversely proportional to the degree of intestinal damage and the proportion of circulating B10 cells increased during DSS-induced damage. The administration of IL-10 to $CD19^{-/-}$ mice significantly decreased intestinal damage and adoptive transfer of B10 cells to $CD19^{-/-}$ mice significantly reduced colitis. Co-administration of peritoneal IL-10 competent B cells also reduced the severity of colitis induced by the intraperitoneal administration of $CD25^{-}CD45RB^{hi}$ CD4+ T cells(318). CpG-induced B cell production of IL-10 and the frequency of IL-10 producing B cells ($CD19^{hi}$ $CD1d^{hi}$) was significantly reduced in patients with Crohn's disease (CD) or UC compared with B cells from healthy donors (319).

Other groups have shown that the main Breg subset supressing MOG-induced EAE are the CD138+ plasmablasts which regulate the generation of MOG-specific Th1 and Th17 cells by suppressing DCs through high IL-10 production (180). It was suggested that the protective effect of splenic B cells was due to Plasmablast differentiation in DLNs: CD1d^{hi} CD5+ B10 cells differentiate into plasmablasts in culture (289). The phenotype of human IL-10 producing plasmablasts was CD27^{int} CD38+cells. Clusters 1 and 3, depleted in toxicity patients, are highly proliferative B cell subsets, which have a strong phenotypic resemblance to this subset with a Ki67+ CD38+ CD27+ CD95^{hi} IL-10+ phenotype. This mirrors clusters 5, 15 and 20 from the independent NSCLC validation cohort (section 4.3.3.5.1).

Hepatocellular carcinoma (HCC) tissues contain a significant population of PD-1^{hi} B cells whose presence correlated with disease progression and with levels of plasma IL-10 (295). These exhibited a CD5^{hi} CD24$^{-/+}$ CD27$^{hi/+}$ CD38^{dim} phenotype. Our PD-1+ cluster, significantly reduced in toxicity patients, has strong phenotypic resemblance to this Breg subset, being PD1^{hi} CD5^{hi} CD25^{hi} CD27^{hi} CD24^{lo} CD38^{lo} IL-10^{int}. Incubation with HCC cell supernatants induced B cell PD-1 expression via TLR4 ligation. PD-1 triggering and PD-L1/PD-1 interaction induced marked IL-10 production and CD5^{hi} PD1^{hi} B cells were major IL-10 producers. This PD-1 driven IL-10 production suppressed T cell immunity thus accelerating tumour progression (295). The increased expression of this subset in our non-responder cohort of NSCLC patients fits with the findings from this group where high infiltration of these PD-1^{hi} B cells in HCC correlated with disease stage and early recurrence. PD-1+ B cells from thyroid cancer patients suppressed T cell proliferation and decreased T cell viability, effects which could be reversed with PD-L1 blockade (320). Patients with oesophageal squamous cell carcinoma had significant elevation of peripheral CD19+CD24^{hi}CD27+ B10 cells with elevated peripheral B cell IL-10 production (321). B cells could internalise exosomes and treatment with exosomes

from these patients could significantly increase the B10 population and stimulate both IL-10 and PD-1 expression, an effect apparently mediated via TLR4 and MAPK pathway activation.

Finally, we also found deficiency in toxicity patients and in treatment responders of a PDL1^{hi} CD38^{int} CD95^{int} TGF-beta^{lo} IL-10^{int} subset. PDL1^{hi} Bregs limit the expansion of the humoural immune response in part by blocking the differentiation of T_{FH} cells (197). T_{FH} cells typically express PD-1 and interact with B cells in the B cell follicle to initiate the GC reaction, whereby B cells undergo expansion and affinity maturation. Elevated levels of T_{FH} cells are well described in autoimmune diseases characterised by autoantibody formation particularly the rheumatic diseases (322). Autoimmune thyroiditis is a common irAE. The percentage of CXCR5+ circulating T_{FH} cells was elevated in patients with Hashimoto's thyroiditis compared with controls as were the percentages of PD-1+ T_{FH} cells (323). The percentage of PD-1+ T_{FH} cells positively correlated with anti-thyroglobulin antibody levels and negatively correlated with FT3 levels. PD-L1 blockade of PD-L1+ Bregs from untreated Rheumatoid Arthritis (RA) patients increased the proliferation and cytokine production of CD8+ T cells (324). These PD-L1+ Bregs were significantly decreased in untreated patients compared with controls and were increased upon successful treatment of RA. In the 4T1 breast cancer model PD-L1 + B cells suppressive of T cell proliferation were increased in bone marrow, peripheral blood and spleen (325). These PD-L1+ CD19+ cells were also up regulated in the peripheral blood of breast cancer patients. Myeloid derived suppressor cells (MDSCs) appeared to co-localise with marginal zone B cells suggesting the potential for crosstalk between these two cell types within the tumour microenvironment. Indeed, co-culture of murine MDSCs with B cells up-regulated PD-L1 expression and drove the B cells to become T cell suppressive (325).

Compelling evidence shows that Bregs directly suppress auto-reactive T cells, suppress TFH cell generation thereby limiting autoreactive B cells, and thus serve as a brake on autoimmunity. Thus, it is entirely plausible that a defective Breg brake could predispose to the development of autoimmunity upon the use of checkpoint blockade. A deficient Breg checkpoint will fail to limit the enhanced activity of self-reactive T cells and auto-antibody formation enabled by PD-1/PD-L1 blockade, a situation in which anti-PD-1 will also directly reduce the suppressive functions of PD-1+ and PD-L1+ Bregs. The paradox is that by negatively regulating T cell function, Bregs are likely to blunt the anti-tumour response and hence result in disease progression.

We show here that toxicity patients have a clear defect in their functional Breg repertoire, with significant reduction of separate IL-10 expressing B cell subsets, which closely map to Breg subsets described in the literature. A limitation of this study is its size and so these results, while striking and biologically plausible, will need to be validated in more patients possibly using a narrower spectrum flow cytometry panel. The high-dimensional nature of the data allied with the analysis enabled for robust exploration, biomarker discovery, annotation and differential abundance analysis. However, the technique and analysis will be limited by manual gating methods, which are open to subjectivity, bias towards well-known subtypes and inefficiency in larger datasets. The development of integrated machine learning methods will help to bridge the gap with other OMICS data analyses and help to infer developmental trajectories directly from cytometry data (326).

Although, many irAEs occur early, longer-term follow-up will be required to ensure that the no toxicity patients remain free of severe irAEs. Furthermore, although we performed high dimensional analyses in lung cancer and melanoma patients treated with PD-1/PD-L1

blockade, larger scale validation in other cancers including melanoma will need to be performed and in those treated with anti-CTLA4 blockade. If validated, pre-therapy functional Breg profiling could become an important tool to help identify patients who may be at high risk of developing severe irAEs on checkpoint blockade.

Chapter 5. Objective 3

The circulating tumour-associated auto-antibody signature in NSCLC is predictive of outcome post-surgical resection

5.1 Characterising the Proteome in Non-Small Cell Lung Cancer

Worldwide, lung cancer is the leading cause of malignancy-related death in men and the second in women. Only 18% of patients at initial presentation are suitable for curative treatment, mainly surgical resection. The overall survival in treated patients is 20-30%, which can improve to 60% in early stage disease (327,328). However, there are still a considerable proportion of patients with early-stage lung cancers who relapse very quickly post resection. The behaviour of these cancers does not obey the expected outcomes based on disease patterns set out by prognostic scores such as the TNM staging system. Biologically, these cancers are very different and this warrants development of scoring systems that delve deeper into the biological nature of these cancers and are therefore better able to prognosticate surgically resected disease.

Lung cancers are clinically detected late in their natural history. Detection and screening is at present based on clinico-radiological assessment. Modalities such as low-dose CT scanning are albeit highly sensitive at detecting small lung nodules but at the cost of a high false positive rate which will affect their utility as routine screening tools (329). There have been attempts to utilise common tumour markers, such as carcino-embryonic antigen (CEA), cytokeratin-19 fragments and neuron-specific enolase (NSE), but these methods lack adequate sensitivity and/or specificity (330). The discovery of novel biomarkers might aid early diagnosis, screening and importantly provide predictive insight into the response to treatment and natural progression of the disease post-resection (330,331).

In an era of patient-tailored immuno-therapeutics and molecular profiling of cancers, biomarker-based stratification of post-operative outcome is vital to tackling malignancy. This is especially true for lung cancer, where disease recurrence confers little chance for meaningful therapeutic intervention.

5.1.1 Biomarker Research in NSCLC

The current clinical TNM staging system (13) (IASLC 8th TNM edition) allows us to group patients in categories and allocate them to surgery & non-surgical treatment. But we recognize the imperfections of this crude system and the need to revise the stage of patients up or down following intervention and after obtaining full operative and pathological data. This has marked effect on operative mortality and survival results.

It is important to make the distinction between diagnostic biomarker signatures in NSCLC, of which there are many and prognostic biomarker signatures, of which there are fewer. Whilst the aim of our study was to discover biomarkers for prognostic purposes, we discuss current diagnostic panels for early lung cancer detection further below.

Current prognostic biomarkers for early stage lung cancers have been described but are limited in their utility owing to the lack of proper validation and lack of adequate sensitivity and/or specificity (330). In the past few decades, numerous prognostic biomarkers have been proposed but few have been validated. Genomic biomarker investigation has advanced our understanding of lung cancer and indeed mutations in p53, KRAS, eGFR and BRCA are well established aberrancies seen in early stage cancer and confer significantly reduced overall 5-year survival (331). However assessing tissue level DNA requires access to tissue and biopsies are not always truly representative of the tumour landscape owing to intra-tumoural heterogeneity (34,332). Multiplex gene panels in breast cancer such as MammaPrint and

OncotypeDX similarly require fresh frozen tissue collected in RNA preservation solution which adds to the complexity (333). Disparities also exist among different racial groups in accessing these platforms which again limits clinical utility (334). Transcriptomic and epigenetic biomarker research is an evolving field; JAK-STAT pathway mRNA is being explored as an NSCLC biomarker and associations are being made between global CpG methylation patterns and outcome in adenocarcinoma (335,336). However, these fields are still within the preliminary phases, and translation into the patient clinic is not without its challenges. Immuno-histological markers such as tumour-infiltrating lymphocytes (TILs) (CD8+, T regulatory cells) and the presence of lymphovascular invasion have also been explored (34,333,334). More recently, the paradigm of checkpoint blockade has resulted in tumour stratification according to CTLA-4 and PD-L1 status, with discordant results (332,335). Proteomic based research exploring autoantibodies in the serum is an attractive option, collecting serum in the pre-treatment phase is an easily implementable intervention and can be carried out in the clinic or bedside. Few proteomics signatures stratifying outcome in NSCLC have been validated. Mass spectrometry defined signatures has demonstrated poor survival in early stage lung cancer histological specimens, but despite this the mechanistic relationship between proteomic signatures and disease outcome is poorly understood (331,337).

5.1.2 Proteomic Signatures in NSCLC

Autoantibody (AAb) profiling is a promising approach that incorporates the immune recognition of a myriad of aberrant cancer proteins into a single diagnostic test. In an era of precision medicine and increasing health care-associated costs, the potential predictive value of peripheral blood AAb for mapping the clinical course of disease pre- and post-treatment is increasingly important and warrants investigation (329). Autoantibodies (AAbs) reflect the

initial humoral immune response against a tumour and their increased levels can be detectable months to years prior to clinical evidence of a primary tumour (338) or indeed recurrence post-resection of a primary tumour. While the mechanisms involved in the production of AAbs in cancer patients (338), remain speculative, AAbs are well known to be sensitive biomarkers in the detection and surveillance of many types of tumours (338,339).

In depth meta-analytical work identified 52 studies which investigated 64 single tumour AAbs and 20 tumour AAb panels in the early detection of lung cancer (diagnostic) (329). Autoantibody panels outperformed the predictive capabilities of single AAbs, and although there was a wide range in specificity and sensitivity for both types of panels, this difference was narrower in the multi-AAb panels. This relationship holds true for prognostic AAb panels in early stage NSCLC, where panels of AAbs are more reliable in predicting long term survival in radically treated patients (340). Combined CYFRA 21-1 and CEA have a stronger association with prognosis than the individual measures taken alone (341,342).

High-density protein microarrays allow the functional testing of thousands of proteins simultaneously, increasing the chance of discovery of new autoantibody signatures (343). These microarrays, in which proteins are immobilised in their natural conformations, enable the identification of AAb profiles within patient sera (343,344). Here we utilised the Sengenics Immunome™ Protein Array [Sengenics, Singapore] containing 1627 proteins, to screen sera from a total of 157 NSCLC patients. After which, we adopted a machine learning approach in order to investigate the utility of using the pre-resection samples in the context of malignancy to identify sera-based proteomic changes specifically associated with outcome in non-small cell lung cancer (NSCLC) following surgery, with a view to using them as predictive surveillance markers. This yielded predictive AAb panels which were able to determine outcome in

resected NSCLC patients with considerable accuracy in the test and validation cohorts. It is important to recognise that the biomarker panels we identified are based on the circulating autoantibodies in pre-operative patient serum so whilst they may be a reflection of the *in vivo* B cell response to tumour associated antigen, there may also be long-lived memory B cell populations which are generating said antibodies and therefore we cannot assume a direct 1:1 correlation between autoantibody level and complementary protein antigen expression. Furthermore, determining whether these antibodies are functional i.e., inhibitory, or stimulatory would help to shed more light on the overall biological interactions between host and tumour. Is there a protective humoural arm or a suppressive arm being induced by the tumour itself? These questions are still to answered and were not addressed by this array as it was not designed for that purpose however it will form part of the further external validation work we are hoping to undertake in the near future.

5.2 Carcinoma of Lung Biomarkers (CLUB) Study Cohort

The CLUB study was a historical study that aimed to assess the sera proteomics of participants with lung cancer. The working hypothesis from this trial was that “sera from patients with lung cancer have proteomic profile(s) different to those from individuals without lung cancer”. The primary objective of this study was to detect biological marker(s) in the sera of patients with lung cancer and correlate them with clinical outcomes (detection of primary tumours, recurrence, survival, and prognosis). Information collected included demographic data such as age, gender, race, occupation, smoking, alcohol, any co-morbidity, nutritional status (weight, BMI), hormonal influences such as time into menstrual cycle in addition to haematological and biochemical investigations. Histo-pathological diagnosis was also collected as an essential component of the patient clinical data to be collected.

The patient selection criteria are summarised below

Inclusion criteria

i. Radiologically confirmed lung mass/nodule/disease referred for surgical exploration, sampling or resection
ii. No previous malignancy (other than adequately treated BCC of the skin).
iii. Age greater than 18 years
iv. Life expectancy ≥12 weeks
v. Fit for surgery
vi. Written informed consent

This historical trial generated a biorepository of pre- and post-operative serum samples taken from NSCLC patients who underwent surgical resection for stage I-III disease. We utilised these samples for further proteomics analysis in order to stratify between survivorship in this clinically robust cohort. Patient samples in the initial stage were selected based on robust clinical annotation and sample availability.

5.2.1 Study Design

A total of 157 study participants' (NSCLC Stage I-III) pre-operative serum samples were utilised in the proteomics analysis, taken from the repository of trial patients. This was determined using a power calculation based on the standard deviations of each protein in the immunome array and sample size estimates were calculated across a range of power values (90-99%), finally settling on a power value of 95%. Once this overall cohort size was determined, a random set of patients (investigators were blinded to clinical metadata) was selected in order to train the machine learning model and subsequently tune the model hyperparameters using k-fold cross validation (345). This training cohort is known as cohort 1. A smaller independent,

separate cohort of patients was randomly selected to provide an unbiased evaluation of the final model. This entirely independent cohort used to validate the model is known as cohort 2. Cohort sizes were determined using a stratified random sample-based approach to split the overall dataset. For reasonable sized datasets (n > 100), this commonly used approach in machine learning settings has been shown to be close to optimal when allocating 66-70% of the samples to the training set (cohort 1) (346).

5.3 Sengenics Platform

The Sengenics Immunome protein array (Sengenics Inc, Singapore) contains over 1600 correctly folded and functional human proteins spotted in quadruplicate. This employs patented KREX functional proteomics technology which utilises the BCCP folding marker for the production of correct folded and functional proteins. For further details on the platform, sample preparation, experimental methods, data acquisition and analysis, see sections 2.4.6 and 2.5.6.

5.4 Results

5.4.1 Pilot Set

In order to determine the utility of the platform in biomarker discovery in NSCLC, we first selected a well-defined cohort of early-stage NSCLC patients (n=42). These patients were all stage I patients with 50:50 split between survivors and non-survivors. A simple ROC analysis was performed on this curated early-stage NSCLC dataset, the top performing individual AAbs are shown below in Figure 5.1. The top performing individual AAb was EAPP with an AUC 0.641.

Name	Cutoff	Sensitivity	Specificity	PPV	NPV	Youden	AUC
VPS4A	115.627	0.824	0.579	0.636	0.786	0.402	0.604
EAPP	131.532	0.647	0.737	0.688	0.700	0.384	0.641
FIP1L1	108.733	0.647	0.737	0.688	0.700	0.384	0.625
PTK7_int	117.301	0.588	0.789	0.714	0.682	0.378	0.610
NFIB	46.732	1.000	0.368	0.586	1.000	0.368	0.616
TACC1	435.235	0.471	0.895	0.800	0.654	0.365	0.616
CT83	57.663	0.941	0.421	0.593	0.889	0.362	0.616
DBNL	69.014	0.941	0.421	0.593	0.889	0.362	0.613
DDX43	80.461	0.941	0.421	0.593	0.889	0.362	0.632
KATNB1	49.944	0.941	0.421	0.593	0.889	0.362	0.601
PTK2_iso2	47.803	0.941	0.421	0.593	0.889	0.362	0.560
PTK6	69.241	0.941	0.421	0.593	0.889	0.362	0.632
SOX15	80.188	0.941	0.421	0.593	0.889	0.362	0.656
ZNF136	52.821	0.941	0.421	0.593	0.889	0.362	0.601
AK3	37.503	0.882	0.474	0.600	0.818	0.356	0.598
CTCFL	63.913	0.882	0.474	0.600	0.818	0.356	0.588
EPAS1	43.841	0.882	0.474	0.600	0.818	0.356	0.579
EPHA2	33.277	0.882	0.474	0.600	0.818	0.356	0.567
FLT1	44.105	0.882	0.474	0.600	0.818	0.356	0.582
HDAC4	96.926	0.882	0.474	0.600	0.818	0.356	0.588
HEYL	54.141	0.882	0.474	0.600	0.818	0.356	0.545
IP6K1	50.357	0.882	0.474	0.600	0.818	0.356	0.579
KRT4	51.500	0.882	0.474	0.600	0.818	0.356	0.582
OSBPL9	41.200	0.882	0.474	0.600	0.818	0.356	0.594
PHF1	60.971	0.882	0.474	0.600	0.818	0.356	0.545
RPS6KA3	60.483	0.882	0.474	0.600	0.818	0.356	0.594
SSX5	58.990	0.882	0.474	0.600	0.818	0.356	0.573
STK4	52.820	0.882	0.474	0.600	0.818	0.356	0.598
TAF6L	60.483	0.882	0.474	0.600	0.818	0.356	0.570
TRAIP	67.610	0.882	0.474	0.600	0.818	0.356	0.619
AIM2	44.610	0.824	0.526	0.609	0.769	0.350	0.604

Figure 5.1. Top performing AAbs in the pilot dataset of stage I NSCLC

Survival analysis was performed on the AAbs with only 16 AAbs showing significantly different survival pattern between high and low expressers. However, owing to the small sample size in this dataset, the analyses were biased towards the number of observations in each cohort. The top 15 AAbs from the ROC analysis were thence taken forward into a ROC-RFE analysis to identify the best performing panel combination (Figure 5.2).

Final panel number	Accuracy-Testing Set	Sens - Testing Set	Spec – Testing Set	ROC – Training Set	No Biomarkers	Sens – Training Set	Spec – Training Set	biomarker.panel
80	0.583	0.714	0.400	0.921	2	0.933	0.800	PTK7_int,NFIB
189	0.417	0.571	0.200	0.850	3	0.933	0.800	PTK7_int,AK3,VPS4A
451	0.333	0.429	0.200	0.950	3	0.933	0.700	PTK7_int,NFIB,FIP1L1
518	0.500	0.571	0.400	0.932	3	0.933	0.700	PTK7_int,NFIB,VPS4A
171	0.500	0.714	0.200	0.907	3	0.933	0.700	PTK7_int,NFIB,AK3
1757	0.417	0.571	0.200	0.907	4	0.933	0.700	PTK7_int,NFIB,FIP1L1,VPS4A
6691	0.417	0.571	0.200	0.871	6	0.933	0.700	PTK7_int,NFIB,FIP1L1,AK3,TACC1,KATNB1
86	0.500	0.571	0.400	0.850	2	0.933	0.700	PTK7_int,PTK2_iso2
1812	0.500	0.571	0.400	0.857	4	0.933	0.700	PTK7_int,NFIB,KATNB1,TACC1
2759	0.500	0.714	0.200	0.843	5	0.933	0.700	PTK7_int,NFIB,FIP1L1,AK3,ZNF136
4508	0.417	0.571	0.200	0.821	5	0.933	0.700	PTK7_int,NFIB,VPS4A,FIP1L1,EAPP
4425	0.333	0.429	0.200	0.857	5	0.933	0.700	PTK7_int,NFIB,VPS4A,TACC1,DDX43
6085	0.500	0.571	0.400	0.850	6	0.933	0.700	PTK7_int,NFIB,PTK6,AK3,TACC1,DBNL
1388	0.417	0.429	0.400	0.843	4	0.933	0.700	PTK7_int,NFIB,VPS4A,DBNL
2864	0.417	0.571	0.200	0.850	5	0.933	0.700	PTK7_int,NFIB,AK3,VPS4A,PTK2_iso2
8890	0.583	0.714	0.400	0.843	6	0.933	0.700	PTK7_int,NFIB,KATNB1,PTK2_iso2,DBNL,TACC1
2863	0.500	0.571	0.400	0.814	5	0.933	0.700	PTK7_int,NFIB,AK3,TACC1,PTK2_iso2
7758	0.417	0.429	0.400	0.779	6	0.933	0.700	PTK7_int,NFIB,FIP1L1,CT83,EAPP,KATNB1
6510	0.417	0.571	0.200	0.786	6	0.933	0.700	PTK7_int,NFIB,FIP1L1,VPS4A,AK3,EAPP
12867	0.500	0.571	0.400	0.829	7	0.933	0.700	PTK7_int,NFIB,AK3,PTK2_iso2,TACC1,ZNF136,KATNB1
8108	0.500	0.571	0.400	0.779	6	0.933	0.700	PTK7_int,NFIB,PTK6,CT83,KATNB1,TACC1
1240	0.417	0.429	0.400	0.814	4	0.933	0.700	PTK7_int,NFIB,DDX43,DBNL
5819	0.417	0.571	0.200	0.850	6	0.933	0.700	PTK7_int,NFIB,FIP1L1,AK3,EAPP,DBNL
4690	0.333	0.429	0.200	0.807	5	0.933	0.700	PTK7_int,NFIB,VPS4A,FIP1L1,KATNB1
15507	0.583	0.714	0.400	0.779	7	0.933	0.700	PTK7_int,NFIB,TACC1,ZNF136,PTK2_iso2,KATNB1,DBNL
2662	0.500	0.571	0.400	0.786	5	0.933	0.700	PTK7_int,NFIB,AK3,EAPP,PTK2_iso2
24162	0.417	0.429	0.400	0.771	9	0.933	0.700	PTK7_int,NFIB,PTK6,TACC1,AK3,CT83,DDX43,PTK2_iso2,SOX15
5474	0.417	0.571	0.200	0.843	6	0.933	0.700	PTK7_int,NFIB,FIP1L1,CT83,AK3,TACC1
20642	0.583	0.714	0.400	0.789	8	0.933	0.700	PTK7_int,NFIB,PTK6,CT83,PTK2_iso2,KATNB1,TACC1,DBNL
12394	0.500	0.571	0.400	0.729	7	0.933	0.700	PTK7_int,AK3,DDX43,PTK2_iso2,TACC1,SOX15,KATNB1

Figure 5.2. ROC-RFE analysis of the top 15 AAbs identified 30 AAb panels with high sensitivity and specificity in the training set. The figures show the AUC, sensitivity, and specificity values for both the training (2/3rd of pilot set) and testing cohorts (1/3rd of pilot set).

Panels containing PTK7_int and NFIB as the core component were able to stratify between survivors and non-survivors with high sensitivity (0.933) in the training set (blue) but this decreased in the testing sets (orange). This is likely due to the small numbers in the latter. Figure 5.3 below shows the ROC performance of top performing panels in terms of AUC (0.85-0.95) in the training set. Biologically, PTK7_int and NFIB have demonstrated strong associations with outcomes in NSCLC (347–350).

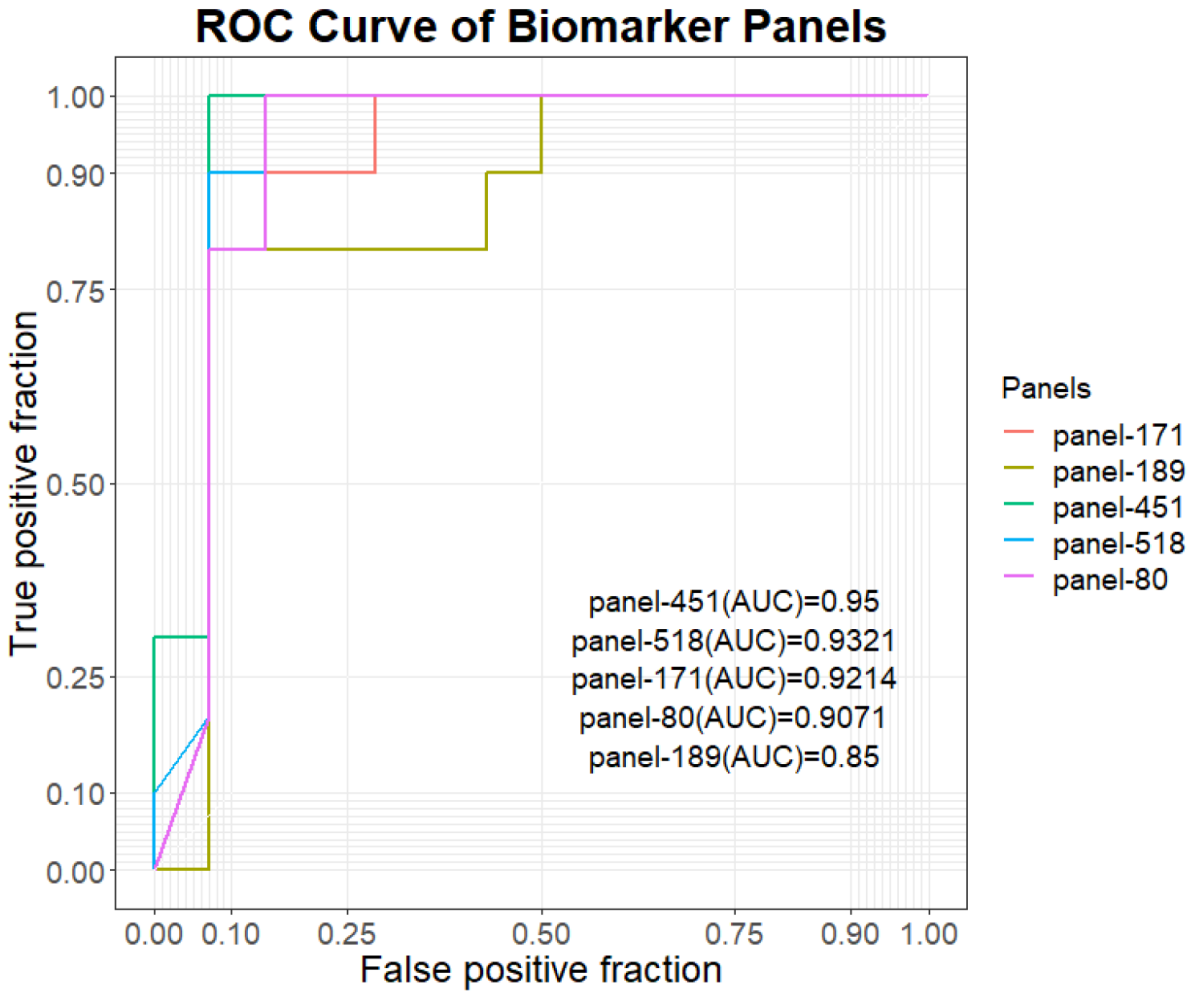


Figure 5.3. ROC curve of the top performing AAb panels (5 panels) in the pilot dataset of stage I NSCLC

This preliminary dataset demonstrated the robustness of the platform, and the data analysis processes. As such we proceeded with a larger dataset and split the cohorts as described above (section 5.2.1).

5.4.2 Study Cohorts

Cohort 1 consisted of 111 NSCLC patients (65 survivors, 46 non-survivors). Cohort 2 consisted of 46 NSCLC patients (27 survivors, 19 non-survivors). Survivors were defined as patients who were alive and recurrence-free at follow-up. Median follow-up of the entire recurrence free population was 1825 days (range of 1195-2555 days). Non-survivors were defined as patients who died from post-operative recurrence within 12 months. The participant characteristics are summarised in Table 5.1. There was no significant difference between cohorts 1 and 2 in terms of age, gender and stage. There was a higher preponderance of adenocarcinomas in cohort 2 (60.9% versus 49.5%), and a higher preponderance of squamous cell carcinomas in cohort 1 (50.5 versus 39.1%).

Group	Cohort 1	Cohort 2
Total Cohort Number	111	46
Male, n (%)	63 (56.8%)	28 (60.9%)
Female, n (%)	48 (43.2%)	18 (39.1%)
Mean Age +/- SD (years)	72.4 +/- 10.4	73.5 +/- 7.87
Adenocarcinoma, n (%)	55 (49.5%)	28 (60.9%)
Squamous Cell Carcinoma, n (%)	56 (50.5%)	18 (39.1%)
Tumour Size, n (%)		
T1 (0-3cm)	56 (50.5%)	16 (34.8%)
T2 (3-5cm)	34 (30.6%)	21 (45.7%)
T3 (5-7cm)	15 (13.5%)	4 (8.7%)
T4 (>7cm)	6 (5.4%)	5 (10.8%)
Nodal Status, n (%)		
N0	74 (66.7%)	34 73.9%)
N1	21 (18.9%)	7 (15.2%)
N2	16 (14.4%)	5 (10.9%)
Presence of Lymphovascular Invasion, n (%)	52 (46.8%)	20 (43.5%)
IASLC Stage, n (%)		
I	71 (64.0%)	30 (65.2%)
II	17 (15.3%)	10 (21.7%)
III	22 (19.8%)	5 (10.9%)
IV	1 (0.9%)	1 (2.2%)
Adjuvant Therapy, n (%)	40 (36.0%)	18 (39.1%)
Mortality, n (%)	46 (41.4%)	19 (41.3%)
Recurrence, n (%)	46 (41.4%)	19 (41.3%)
Median Follow-up (days)	1825	1805

Table 5.1 Clinico-pathological characteristics of the study cohorts

Survival distribution of the total study population is displayed in Figure 5.4. Cox Proportional Multivariate Hazards Analysis identified IASLC stage and the presence of lymphovascular invasion as significant independent negative prognostic risk factors (HR 1.72, p<0.001 and HR 2.03, p=0.006 respectively); histology was not deemed significant.

Characteristic	Post-operative Mortality			Time to Death		
	OR[1]	95% CI[1]	p-value	HR[1]	95% CI[1]	p-value
Age	0.97	0.94, 1.01	0.15	0.98	0.95, 1.00	0.073
IASLC_Stage	2.19	1.43, 3.44	<0.001	1.72	1.31, 2.25	<0.001
Vascular_Invasion						
No	—	—		—	—	
Yes	2.28	1.15, 4.60	0.019	2.03	1.22, 3.37	0.006

[1] OR = Odds Ratio, CI = Confidence Interval, HR = Hazard Ratio

Figure 5.4. Multivariate Analysis using logistic and cox proportional hazards regression to assess outcome in terms of post-operative mortality and time to death respectively. Stepwise backward elimination was employed to remove the least significant independent predictors. All variables (age, gender, histology, IASLC stage, nodal status, lymphovascular invasion and adjuvant chemotherapy use) were entered into the model and successively removed depending on significance in the model.

All samples were assayed using the Sengenics Immunome Protein Array containing 1600+ proteins spotted in quadruplicates (Sengenics Inc, Singapore).

5.4.3 Identification of Predictive AAbs

The final AAb panel was selected based on an iterative applied machine learning pipeline as specified in the methodology, the algorithm for which is shown in Figure 2.6. This bespoke algorithm comprised machine learning elements that have been previously validated in other cancer classification datasets (267,351)

Initial data processing involved filtering according to the penetrance fold change analysis in order to avoid biasing subsequent model generation. 1355 AAbs remained which were taken forward into the deeper analysis. Within the remaining AAb data, >93% displayed collinearity of r>0.75 on Spearman Rank Correlation analysis, hence the reason for proceeding with recursive feature elimination by random forest modelling.

The AAbs, which appeared most frequently with the highest importance values across 100 randomly seeded iterations, are listed in Table 5.2. Corroborative regression and genomics analysis methods were performed and indicate the AAbs, which were common to all analytical techniques. Overall, 60 AAbs (RFE set) were identified as the most stable with no improvement in predictive performance beyond this number.

AAb	**Repeated k-fold cross validation**	**LOOCV**	**Lasso Regression**	**DESeq 2**
SPATA19	✓	✓		
CASP7	✓	✓		
TSPY3	✓	✓		
GLS2	✓	✓	✓	✓
TCEA2	✓	✓		
CTNNA2	✓	✓		✓
ITPKB	✓	✓		
AFF4	✓	✓		
MAGEB2	✓	✓	✓	
C1orf174	✓	✓	✓	
TSGA10	✓	✓		
TYRO3_int	✓		✓	
DCBLD2	✓		✓	✓
PCLAF	✓			
SPO11	✓	✓		
BPIFA1	✓	✓		
MAGEB4	✓	✓	✓	
HMGN5	✓		✓	
MAEL	✓			
LUZP4	✓		✓	
HDAC4	✓			
SOX15	✓			
HOOK1	✓			

CDK16	✓		✓	
CSAG1	✓			
SPACA3	✓			
IMPDH1	✓			
MAGEB5	✓			
TXN2	✓		✓	✓
NFYA	✓			
PHF7	✓		✓	✓
HIST1H1C	✓			
IP6K1	✓			
TFG	✓			
AIM2	✓			
SGO1	✓			
PYCR1	✓			
FAM50B	✓			
HK2	✓			
ERBB3_int	✓			
TBL1X	✓		✓	
ZNF207	✓		✓	
EEF1D	✓			
PPP2R1A	✓			
MAP2K7	✓			
RPL7A	✓			?
CBLC	✓			
COX6B2	✓			
ACTB	✓			
CA9	✓			
FLCN	✓			
GAGE2	✓			
ARAF	✓			
AK3	✓			
HMG20B	✓			
CNN1	✓			
EPAS1	✓			
EAPP	✓		?	?
TSSK6	✓			
GRK6	✓			

Table 5.2. Topmost stable AAbs as determined by recursive feature elimination

5.4.4 Additive Predictive Modelling

The RFE set of AAbs was used to generate AAb panels by additively selecting the top-ranking AAbs in a cumulative fashion. These inputs were used to determine the ROC metrics at each

additive iteration for cohort 1 (Figure 5.5). An upward linear trend in all three parameters (AUC< sensitivity, specificity) was noted as more AAbs were added. This progressive increase peaked at 44 cumulative AAbs (AUC 0.44-0.975; Sensitivity 67.4%-87%; Specificity 46.2%-98.5%). Beyond this, the predictive metrics become rather unstable and less uniform hence the decision to proceed with the top 44 AAbs for deeper analysis.

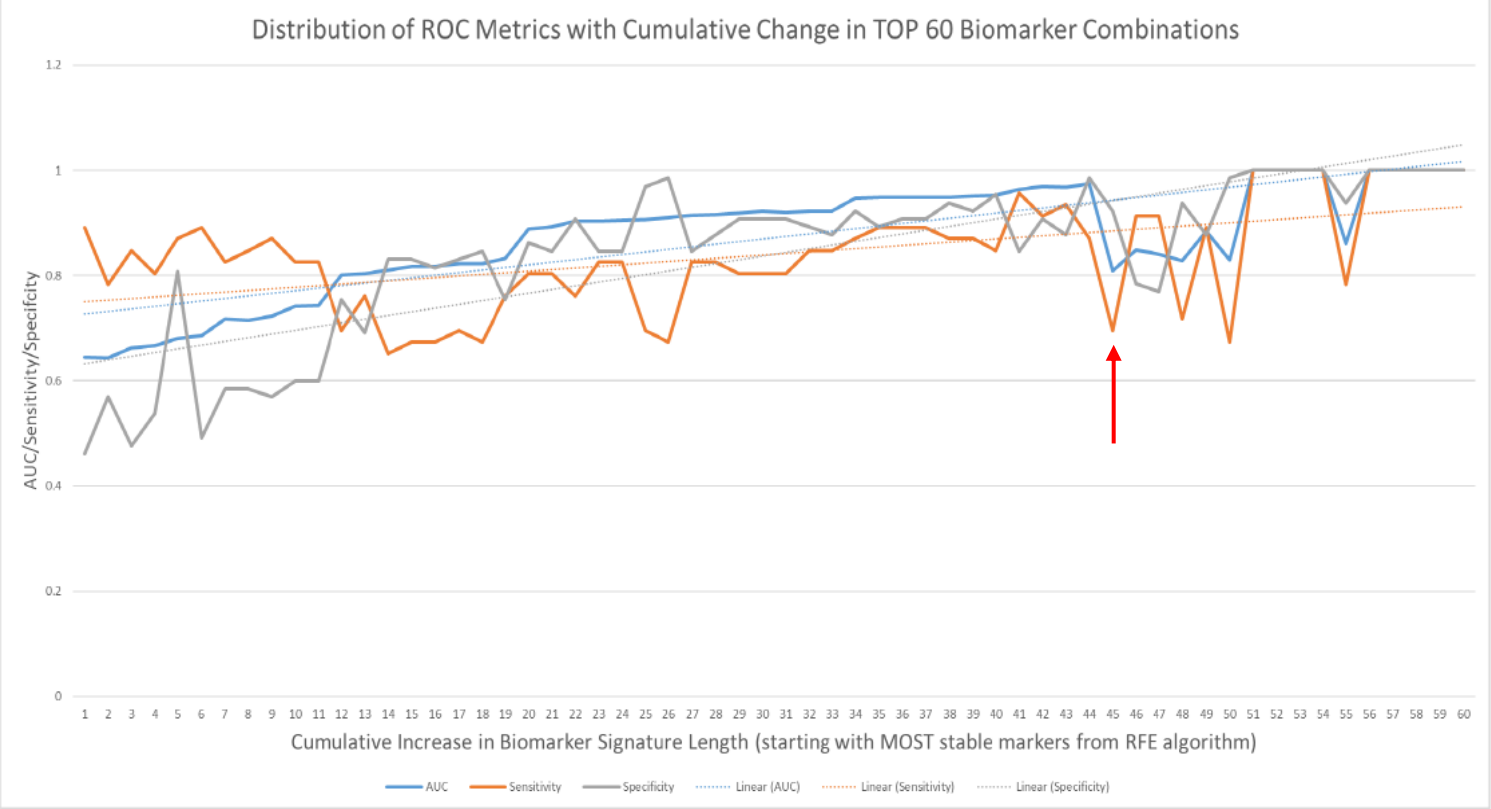


Figure 5.5. ROC metrics displayed for additive modelling of AAbs in the RFE set (n=60). There is a progressive linear increase in AUC (blue), sensitivity (orange) and specificity (grey) with each cumulative addition of AAbs. The solid lines show the interval change in each metric with each AAb addition and the dashed lines represent the overall trend. There is a progressive improvement in all parameters up to 44 AAbs, beyond which, there is a decay in performance for each metric as shown by the sharp negative deflection (red arrow). The performance of the model deteriorated beyond 44 cumulative AAbs indicating no added value.

5.4.5 Multi-model Inference Approach

Given that a 60-AAb diagnostic scoring system would be cumbersome and impractical, we utilised an information-theoretic approach to determine the AAb combination with the highest diagnostic potential in the most parsimonious model. We employed the Akiake Information Criterion method (AICc) in order to estimate the "goodness of fit" of statistical models and thereby compare multiple models with one another. The AICc avoids overfitting the model in smaller sample sizes. Based on the cumulative ROC analysis, we proceeded with the top 44 AAbs in this downstream analysis. Following stepwise backward elimination of these markers in a multivariate logistic regression model, with survivorship as the dependent variable, 18 AAbs were determined to be the most significant and were therefore used in the multi-model inference analysis. Any further addition of more AAbs did not lead to significant improvements of model performance but did contribute to significant increases in computational time.

5.4.6 Assessing Model Performance

Panel A, the most parsimonious and best performing model, comprised 13 AAbs - SPATA19, TSPY3, GLS2, TCEA2, TSGA10, HMGN5, LUZP4, HDAC4, SPACA3, IMPDH1, TXN2, TFG and PPP2R1A (Table 5.2). This refined model was assessed in cohort 1 (AUC 0.918, Sensitivity 89.1%, Specificity 80.1%) and validated in the independent cohort 2 (AUC 0.842, Sensitivity 84.2%, Specificity 74.1%) (Figure 5.6). There was no significant difference in the ROC metrics between the two cohorts, indicating good performance in the validation cohort.

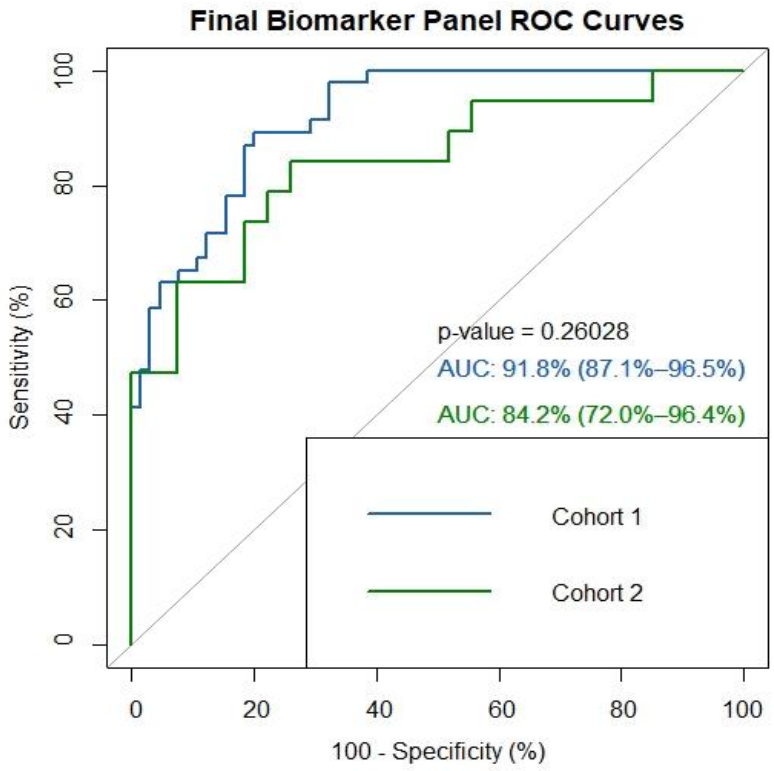


Figure 5.6. ROC curves demonstrating performance of Panel A (13 AAbs) in both cohort 1 (test) and cohort 2 (validation). P value indicates no significant difference in performance of this model between cohorts. AUC 95% confidence intervals displayed in brackets.

We noted a preponderance of bona-fide Cancer Testis Antigens (CTAGs) in the RFE AAb set (16/60 (26.7%)). We thus elected to explore two further CTAG specific panels in order to determine the prognostic relevance of these highly conserved AAbs in NSCLC. We refer to the final AAb panel we defined as Panel A (13 AAbs). Panel B refers to the CTAGs extracted from the RFE set (16 AAbs) and Panel C refers to the CTAGs extracted from Panel A (6 AAbs).

The strong CTAG presence in Panel A was comprised of six AAbs, SPATA19, SPACA3, TSPY3, TCEA2, TSGA10 and LUZP4, all with established pro-tumourigenic roles in cancer (Table 5.2). CTAGs trigger unprompted humoral immunity and immune responses in malignancies, altering tumour cell physiology and neoplastic behaviours. Their limited expression in normal

somatic tissues coupled with recurrent up-regulation in epithelial carcinomas makes them highly attractive AAb and vaccine targets. We explored performance of all three panels in cohort 1 and 2 (Figure 5.7 and 5.8). Panel A performed significantly better in cohort 1 (test) than both panels B and C (CTAG panels). However, in cohort 2 (validation), the differences between panel A and the CTAG panels (B and C) was not significant. Panel B (16 CTAG panel) outperformed panel A in cohort 2 (AUC 0.875 versus 0.842, p=NS) but panel C underperformed compared to panel A in cohort 2 (AUC 0.69 versus 0.842, p=NS). The increased predictive performance of panel B (16 CTAG panel) reaffirms the importance of CTAGs in discriminating between survivorship in lung cancer. In spite of CTAG preponderance, this data shows that the non CTAG antigens in panel A, which are critical mediators of Wnt signalling and phosphatase activity are clearly biologically important in their ability to prognosticate in lung cancer. We therefore developed global score indexes for each panel and explored the survival implications of each panel in multivariate models.

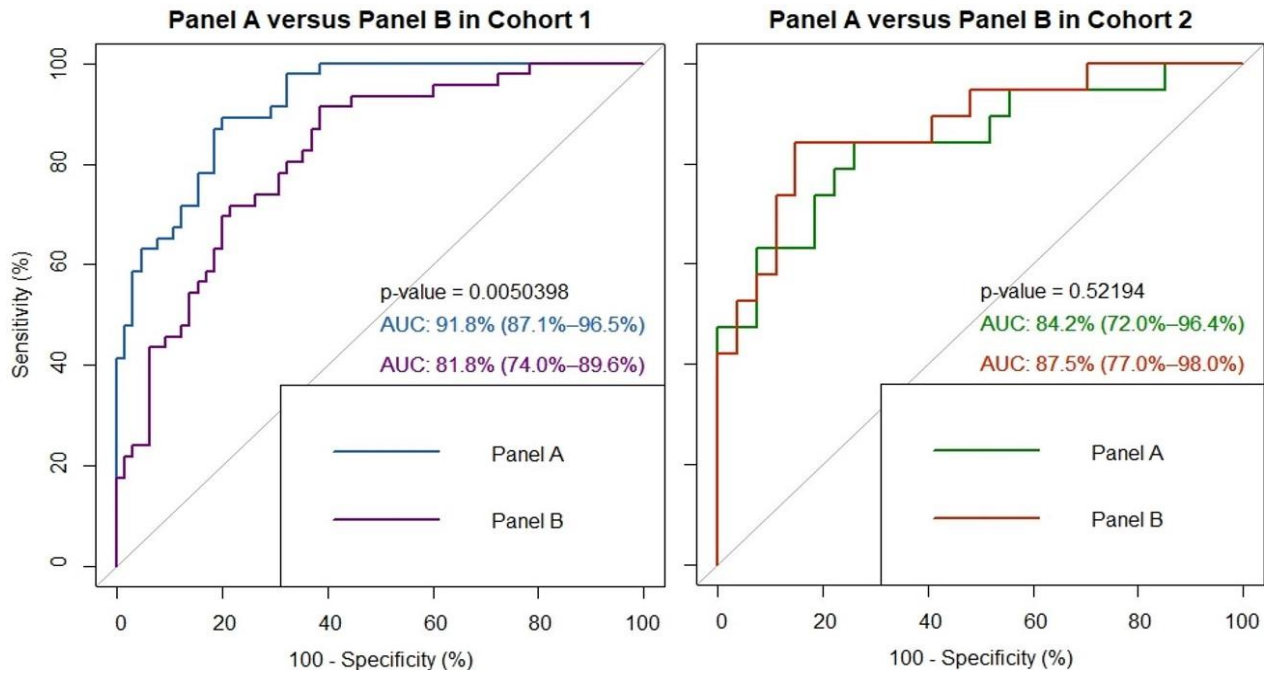


Figure 5.7. ROC curves demonstrating performance of Panel A (13 AAbs) and Panel B (16 CTAG AAbs from RFE set) in cohort 1 (test) (left panel) and cohort 2 (validation) (right panel). AUC confidence intervals displayed in brackets. P values indicate significant differences in performance in cohort 1 but not cohort 2.

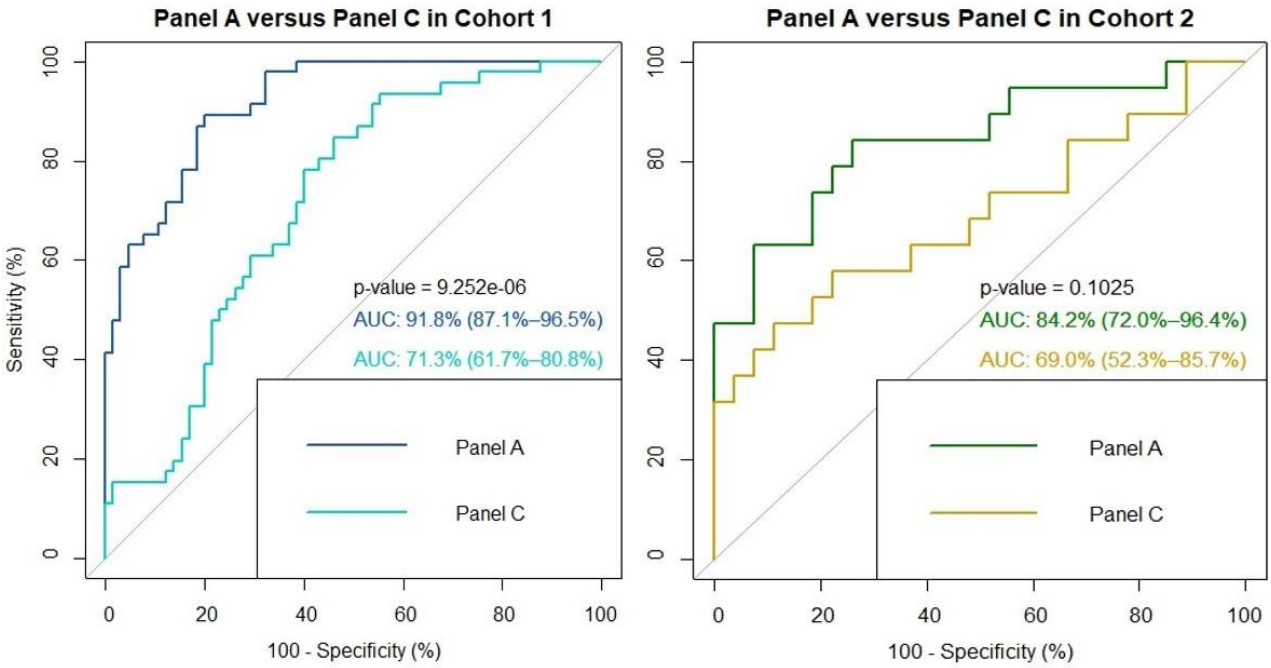
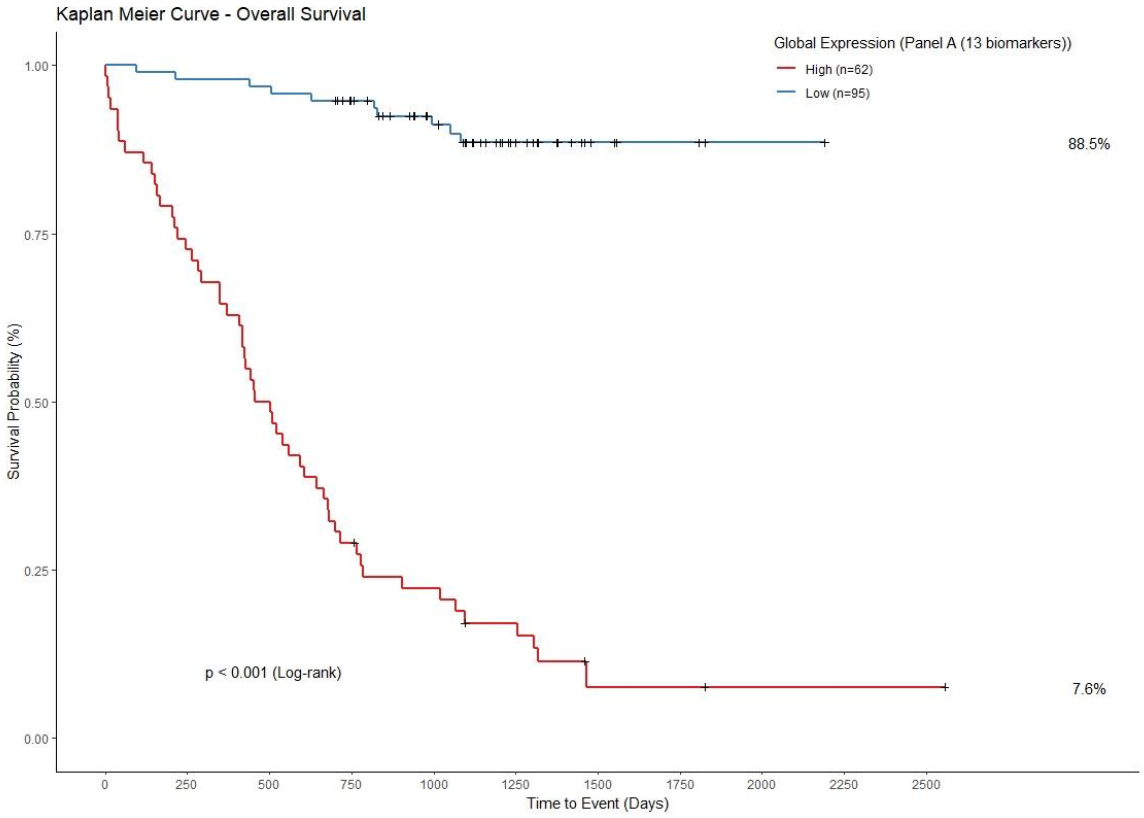


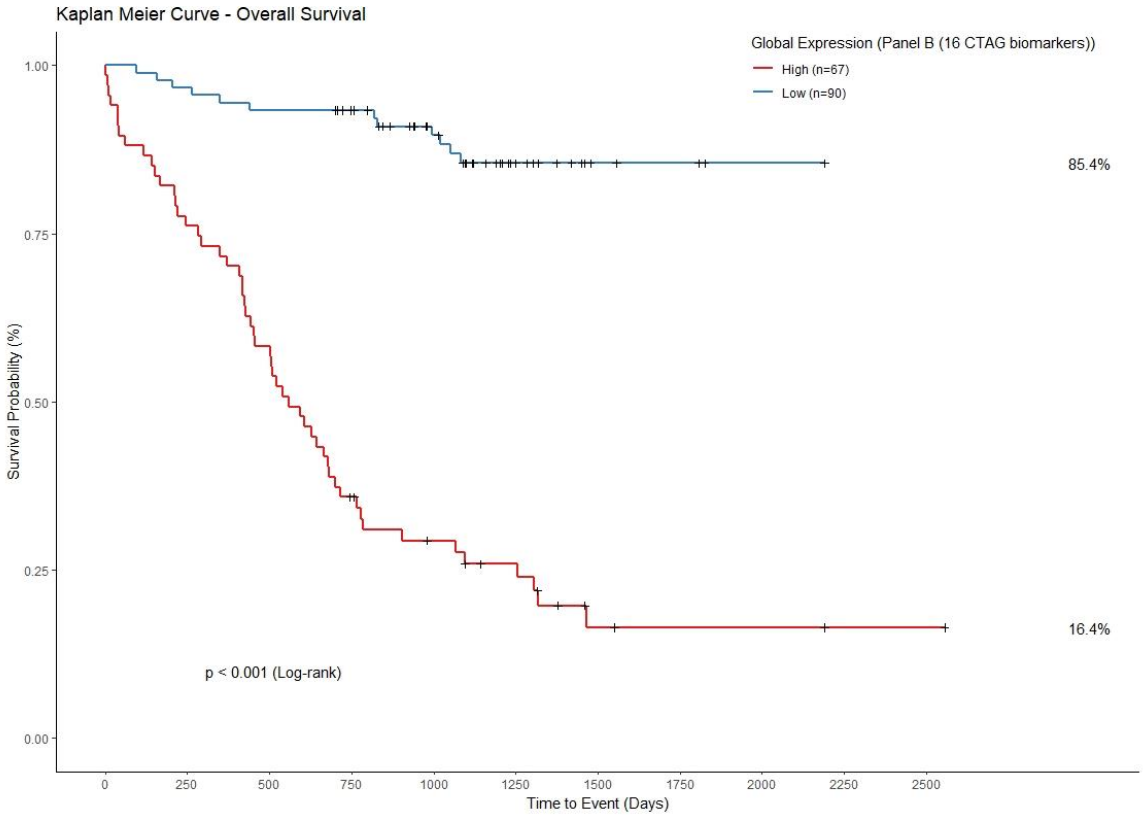
Figure 5.8. ROC curves demonstrating performance of Panel A (13 AAbs) and Panel C (6 CTAG AAbs from Panel A) in cohort 1 (test) (left panel) and cohort 2 (validation) (right panel). AUC confidence intervals displayed in brackets. P values indicate significant differences in performance in cohort 1 but not cohort 2.

5.4.7 Survival Analysis

Further interrogation of these signatures was carried out by generating a continuous risk score for every individual on the basis of model coefficients. The resultant predicted risk scores from cohort 1 (training) were divided using optimal cut-off points determined through ROC analysis in order to further dichotomise the patient cohorts as “high expressers” and “low expressers”. We performed this for all three panels (A, B and C), the scores were inferred directly from the AAb signal intensities. Using these individual risk scores, we carried out survival analyses (Figure 5.9) and multivariate cox proportional hazards modelling (Figure 5.10) in the entire NSCLC cohort. Patient age, gender, histology, nodal status, IASLC stage, lymphovascular invasion and whether patients underwent adjuvant chemotherapy were all predictors that were entered into the model alongside all the panel scores. All panels were

able to effectively dichotomise between survivor statuses in our cohort, with high expression conferring a significantly worse outcome ($p<0.001$), reaffirming findings from the ROC analysis. Five-year survival in high expressers of panels A, B and C was 7.6%, 16.4% and 19.9% respectively and high expressers of panel A had a median survival of just under 16 months which for early stage resected lung cancer is very low. On multivariate testing, only panels A and B were deemed significant independent predictors of survival, HR 19.6 and 7.22 respectively ($p<0.05$). IASLC stage was still deemed an independent predictor of outcome albeit not significant (HR 1.24, $p=0.11$). Panel C was deemed a significant independent predictor of outcome only when entered into a multivariate model without panels A and B. This reaffirmed the findings that the CTAGs alone from panel A were not sufficiently predictive enough when compared with panels A and B but are still significant predictors independent of age, gender, IASLC stage, lymphovascular invasion, histology, nodal status and whether patients underwent adjuvant chemotherapy.


Kaplan Meier Curve - Overall Survival
Global Expression (Panel A (13 biomarkers))
High (n=62)
Low (n=95)
88.5%
7.6%
p < 0.001 (Log-rank)
Survival Probability (%)
1.00
0.75
0.50
0.25
0.00
0
250
500
750
1000
1250
1500
1750
2000
2250
2500
Time to Event (Days)


Kaplan Meier Curve - Overall Survival
Global Expression (Panel B (16 CTAG biomarkers))
High (n=67)
Low (n=90)
85.4%
16.4%
p < 0.001 (Log-rank)
Survival Probability (%)
1.00
0.75
0.50
0.25
0.00
0
250
500
750
1000
1250
1500
1750
2000
2250
2500
Time to Event (Days)

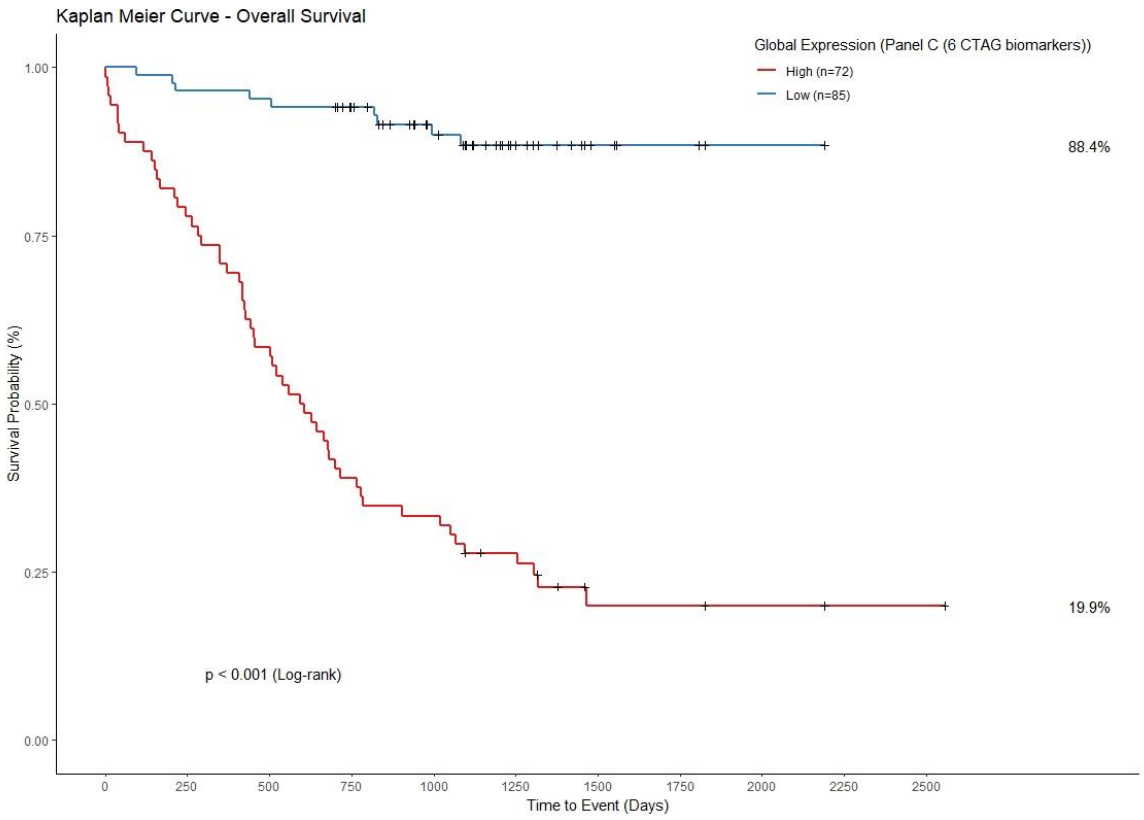


Figure 5.9. Kaplan Meier Survival Analysis for Panel A (13 AAbs), Panel B (16 CTAG AAbs from RFE set) and Panel C (6 CTAG AAbs from panel A). All curves demonstrate significantly worse outcome in patients with high global expression of each AAb signature. The number of patients in each expression group is shown in brackets in the legend for each Kaplan-Meier plot. Percentage survival out to 5 years is displayed to the right-hand side for each expression group in each plot. The median survival for high expressers of Panel A, B and C is 479 days, 558 days and 600.5 days respectively.

Characteristic	HR[1]	95% CI[1]	p-value
`Panel A`	19.6	4.73 to 81.6	**<0.001**
`Panel B`	7.22	1.23 to 42.4	**0.025**
IASLC_Stage	1.24	0.95 to 1.62	0.11

[1] HR = Hazard Ratio, CI = Confidence Interval

Figure 5.10. Multivariate Cox Proportional Hazards modelling in entire NSCLC cohort (n=157). Stepwise backward elimination was employed to remove the least significant independent predictors.

5.4.8 Gene Ontology Analysis

Of the identified AAbs in the RFE set (n=60), all are known for their role in biological processes heavily interlinked with neoplasia and malignant transformation. Processes such as chromosomal organisation, cellular component homeostasis, ribosome function, transcription regulation, DNA repair and regulation of AAb phosphotransferase activity, namely MAPK activation and the MAP/RAF kinase cascade (352) thus reaffirming their biological relevance. This is consistent with gene ontological analysis where the most significant pathways related to the RFE set, altered chromosome organisation [gene ontology (GO) term GO:0051276, false discovery rate (FDR) = $5.24*10^{-3}$] and phosphotransferase activity (GO:0016776, FDR = $3.89*10^{-2}$) in NSCLC.

The generated STRING AAb association network of the top 60 AAbs (section 2.5.6.3) demonstrated that the overall interaction enrichment for this selected number of AAbs was significantly higher than would be expected for a random set of AAbs of similar size, drawn from the genome (p=0.0022) suggesting biological interaction as a group and consistent with the preponderance of CTAGs. Further underscoring this; most of the sero-reactive AAbs are intracellular antigens (52/60) interacting with membrane and non-membrane bound organelles such as ribosomes (4/60) with the majority residing within the nucleus (37), a

usually immuno-privileged site. This pattern has been observed in autoantibody studies in melanoma (339). Despite this, autoantibodies generated against autologous nuclear antigens are frequently found in cancer patient sera (338). Nuclear antigens however do not undergo antigen presentation during the negative selection of self-reactive lymphocytes largely because of their intrinsic proteolytic instability, which affects the binding kinetics with MHC class II receptors (338). Exposure of the nuclear antigens to one's immune system and the resultant generation of autoantibodies is therefore thought to occur following tumour cell death and release of the intracellular contents into the circulation (353).

5.5 Biomarker Discovery

Historically, the majority of autoantibody-based biomarker research has concentrated on diagnosis of disease states or early detection of cancers as opposed to trying to map the course of disease post-treatment. This is true for NSCLC and melanoma (354). Overall, this balance is likely to shift in favour of the latter. Results from the NLST and European NELSON trials (355,356), strongly supported lung cancer screening. Widespread use of CT coronary angiography to assess inpatient chest pain as well as use of whole-body CT use in the assessment of Trans-catheter valve intervention results in a high detection rate of incidental findings, a large proportion of which are lung malignancies. The combination of screening strategies and increased use of CT scanning for non-cancer related conditions will result in a surge in the detection rate of early-stage lung cancers and therefore an increased surgical resection rate. Early-stage lung cancers confer a multitude of outcomes ranging from indolent disease with high post-operative disease-free survival rates at 5 years to highly aggressive disease with relapse in the first 12 months post-resection. Identifying patients at the greatest

risk of recurrence and aspects of the biology of poor outcome disease through biomarker-based stratification is becoming an ever-more pressing need.

Current prognostic biomarkers for early-stage lung cancers have been described but are limited in their utility owing to the lack of proper validation and lack of adequate sensitivity and/or specificity. The key points in evaluating biomarker studies in early-stage lung cancer include well-defined objectives and study populations, robust specimen storage and use, and use of a clinically applicable assay that is validated in an independent cohort of patients. Critical appraisal of published prognostic signatures in early-stage lung cancers found that adherence to these criteria was poor with overt flaws in study design. Subramanian et al published a set of guidelines to inform prognostic biomarker studies in lung cancer, and although all the studies pertain to gene expression microarray data, we have adhered to these criteria as closely as possible (357). We validated our signature in one completely independent dataset which is a feature that is lacking in many prognostic signatures in early stage lung cancer (357). A 14 gene, qt-PCR derived expression signature was previously validated in two independent large stage I, non-squamous NSCLC datasets which demonstrated robustness of the statistical design. This signature showed poor survival in high-risk patients based on gene expression in both validation cohorts and although the AUC values were significantly higher for this signature than standard NCCN risk criteria in both validation sets, the absolute values were still relatively low (0.60 and 0.61), compared with the AUC values from our panels A and B (AUC 0.842 and 0.875) (358). Furthermore, this study did not assess the therapeutic relevance of the genes identified in the final signature.

Most studies in the last decade that have explored serum or blood based antibodies targeting tumour-associated antigens have been for early lung cancer detection and have employed

ELISA as the primary detection method as well as Western blotting, AAb chip and SDS-PAGE (329). Studies investigating single AAbs have included Cyclin B1, p53, NY-ESO-1, MUC1, MDM2, p16, APE1, CD25, Cathepsin D, ABCC3, IGFBP-2, BARD1, BRAF, Dickkopf-1, c-Myc and a range of heat shock AAbs (329). Sensitivities and specificities for lung cancer detection have ranged from 0-90.3% and 0 to 100% respectively (reviewed in 23). Studies investigating panels of AAbs have commonly utilised p53, cyclin B1, MDM2, IMPDH, NY-ESO-1, CAGE, GAGE and MAGE family AAbs, SOX2 and c-Myc. Sensitivities and specificities for lung cancer detection have ranged from 0-92.2% and 79.5 to 92.2% respectively. Commonly used AAbs in clinical practice include total PSA for prostate cancer (AUC 0.71) (359), pre-operative serum CEA for colorectal cancer (AUC 0.543) (360), NY-ESO-1 and Neuron-specific enolase (NSE) in small cell lung cancer (AUC 0.619 and 0.773 respectively) (361), and a panel of serum autoantibodies including NY-ESO-1, p53, MMP-7 and HSP70 in oesophageal adenocarcinoma (AUC 0.815) (362). Much like the markers delineated in our dataset, these tumour-associated antigens are a combination of tumour suppressor genes and oncogenes, with roles in cell cycle regulation, DNA replication and apoptosis. These are processes that are commonly deregulated in various solid tumours such as breast, bladder, colon, oesophageal and prostate (329,363–365).

Circulating AAbs have also been investigated as prognostic AAbs in early-stage lung cancer, the most common of which are CEA and CYFRA 21-1. The largest study exploring the role of CEA found that elevated pre-operative levels conferred poor 5-year survival (366). The body of data exploring this AAb in prognosticating NSCLC is conflicted, with some studies reporting no association with post-operative survival (340), this is also true for NSE. Meta-analytical data demonstrated that elevated pre-treatment CYFRA 21-1 is an independent negative prognostic risk factor at 12 months (HR 1.88) in 2063 patients (367). Combined AAb panels of CEA and CYFRA 21-1 have demonstrated prognostic value in some studies (342) but not others

(341). Combined pre-operative elevation of NSE and CA125 conferred reduced overall and disease free survival in NSCLC (368).

Rather than focus on a single uniquely predictive marker, antibody profiling offers higher predictive power that is predicated on combining numerous tumour-associated antibodies. Given the complexity and multi-factorial nature of the anti-tumour immune response and tumour immune evasion mechanisms in cancers that are not solely reliant on single oncogenic drivers, combination AAb signatures would prove more valuable (329). None of the prognostic studies mentioned offered any predictive assessment of their panels but instead used hazard ratios (measures of association, not predictive power) with no separate test/validation. Our panels are significant independent predictors of outcome with robust ROC metrics and outperform the commonly used diagnostic panels used in clinical practice today (AUC 0.845-0.875). Rousseaux et al (279) described a 26-gene signature which was an independent predictor of poor prognosis in early stage lung cancer. Much like our cohorts, they described a 152-patient population with early-stage lung cancer that they followed up for a minimum of 5 years. The cohort was not exclusively NSCLC however, but comprised of small cell, carcinoid, and large cell cancers. Gene expression was assessed using Affymetrix microarray technology on resected tumour specimens and analysis identified 26 germline restricted genes whose aberrant expression was individually associated with lower survival probability. The expression pattern of the genes was used to dichotomise the cohort into 3 groups ranging from absolutely no expression of any of the 26 genes to expression of 3 or more of the 26 genes. Our approach assessed the proteomic profile by utilising the raw data in predictive models and by incorporating overall signature scores which preserves the overall biological expression of each marker. Their dataset lacked internal testing and validation on a separate cohort but utilised orthogonal techniques to explore the expression of the 26

genes (i.e qt-PCR). Our blood-based AAb panel demonstrates robust prognostic capabilities in separate test and validation cohorts, which in themselves are well dichotomised.

Using the approach adopted by Gnjatic and colleagues (344); identifying AAbs based on RFU levels and positive seroreactivity in survivors versus non-survivors, we successfully identified 60 prognostic AAbs with individual ROC and survival data metrics. This collection of AAbs demonstrates biological interaction as a group, partaking in key cellular processes that are often unregulated in tumours and are the inciting insult in tumorigenesis.

Three broad categories of genes comprised our final panel (panel A), namely CTAG expression, Wnt signalling AAb aberrancy and Serine/Threonine AAb phosphatase deregulation. CTAGs are united by their role in embryonic development and restriction of expression to male germ cells. Ectopic re-expression of these antigens has been seen in a variety of somatic solid tumours and in triple negative breast cancers, high expression correlated with worse survival in multivariate analysis (HR 2.02, 95% CI 1.27-3.20; p=0.003) (369). Ectopic gene signatures of normally silenced CTAG genes that are expressed in cancer associated with a highly aggressive lung cancer phenotype and independently predicted poor outcome (279). We identified 16 CTAGs (27%) in our RFE set (S2) as being highly discriminatory for survivorship in this distinct cohort of NSCLC patients (SPATA19, SPACA3, TSGA10, TSPY3, LUZP4, TCEA2, CTNNA2, MAGEB2, SPO11, MAGEB4, MAEL, CSAG1, MAGEB5, COX6B2, GAGE2, TSSK6). This CTAG only model displayed high predictive power in the validation cohort (AUC 0.875, sensitivity 84.2%) and was a significant independent predictor of poor outcome. Clonal and subclonal CTAG expansion is generally uniform in tumour cells with variations in behaviour tightly regulated by epigenetic alterations (370). CTNNA2, a known regulator of cell-cell adhesion and tumour suppressor gene, is frequently mutated in laryngeal carcinoma (371) and genomics analyses

have identified CTNNA2 mutation as a driver event in gastric carcinogenesis as well as being a late expanded subclonal population in various other cancer types (372). DNA methylation and post-translational histone modification are the most common epigenetic driver events (373). Co-expression of HDAC4 in the RFE set provides a mechanistic relationship between these two types of tumour-associated antigens possibly working in synergy with one another.

This has prompted their investigation as therapeutic targets and AAbs of disease. Owing to their highly restricted expression patterns in normal tissues and ectopic expression in tumour types, their utility as individual diagnostic markers is limited but makes them highly sought after as targets for cancer vaccines (374). CTAGs such as MAGE, GAGE, NY-ESO-1 are well described in historical AAb signatures in lung cancer as well as other solid tumours and MAGE, PRAME and NY-ESO-1 CTAGs are being investigated as targets for CAR-T cell based therapies in solid tumours (375). Numerous trials are underway exploring NY-ESO-1 specific TCR gene transduced lymphocytes, humanised antibodies, peptide conjugation into oncolytic viral vectors and antibody drug conjugates targeting this CTAG (376). Toxic payload delivery via TCR conjugates has successfully targeted intracellular NY-ESO-1 and LAGE1 in vitro and in vivo abrogating tumour growth (377).

A key element to the success of CTAG-dependent vaccine therapy is in appropriately identifying CTAG-expressing cancer cells that are abundant in tumours, rarely expressed in normal tissue, and have defined functional characteristics such that targeting results in the absolute attenuation of tumourigenic potential. Whilst peptide-based vaccine therapies alone have been met with challenges; MAGEA3 targeting although elicited CD8+ T cell clones, showed no measurable clinical benefit (206,370). Combining said therapies with immunogenic adjuvants, adoptive T cell transfer and even polyepitopic RNA based vaccines

hold a lot of promise. The Lipo-MERIT trial demonstrated strong CD4+ and CD8+ T cell induction along with durable objective clinical benefit in unresectable melanoma patients treated with a poly-antigenic liposomal RNA vaccine with or without combination with anti-PD1 checkpoint blockade therapy (378). The RNA vaccine targeted four main CTAGs; NY-ESO-1, MAGEA3, TPTE and Tyrosinase (378).

Aberrant activation of the Wnt/B-catenin signalling pathway is causally linked to many cancers, in terms of recurrence, immune evasion and metastasis. A number of the identified tumour-associated antigens are known to signal via this cascade, HMGN5, TFG, MAEL, SOX15 and Dicckopf-1, the latter of which has been investigated in numerous other AAb signatures (379,380) and like TFG interacts via the Wnt co-receptor LRP6 (381). AAbs such as MAEL and PTK7 both signal via this cascade and expression was significantly associated with poor outcome in our cohort. MAEL, also a CTAG, has been shown to be critical for cancer cell survival and is over-expressed in bladder and gastric cancers (382). Functional experiments have determined that MAEL AAb exerts its oncogenic dominance through degradation of the AAb phosphatase ILKAP (382). Mutating or silencing of phosphatase activity is a well-known tumour escape mechanism and forms one of the core themes of our identified AAb panel, along with aberrant CTAG expression and dysregulated Wnt Signalling. MAEL provides a unique link between all three biological pathways. These molecules provide unique therapeutic targets as demonstrated by an antibody drug conjugate against the Wnt-signalling PTK7 tyrosine kinase molecule which elicited potent anti-tumour activity in low-passage patient derived solid tumour xenograft models (347). In solid tumours like NSCLC, high PTK7 expression confers significantly reduced overall survival. Targeting this molecule in early phase I human clinical trials have just completed recruiting in December 2020 [NCT02222922]. Other agents targeting parts of the Wnt signalling axis that have entered or

completed phase I clinical trials include Vantictumab, a monoclonal antibody against the Fzd receptor (NCT01345201, NCT02005315, NCT01957007, and NCT01973309). OMP-54F28 acts as a decoy receptor binding all Wnt ligands to suppress downstream signalling has entered phase II trials (NCT02069145, NCT02092363, NCT02050178, and NCT01608867). Porcupine enzyme inhibitors which suppress the production of bioactive Wnt ligands have also entered phase I trials (LGK974,NCT01351103; and ETC-1922159, NCT02521844) (379,383).

Cellular responses to DNA damage are integral to maintaining the genome and preventing cancer progression; Serine-Threonine phosphatases like AAb Phosphatase 2 play a key role in the DNA damage response through regulation of important cell cycle AAbs and tumour suppressor genes such as ATM, Chk1, Chk2, p53 and BRCA1 (384). Cancer cells tend to evade the activation of DNA repair pathways through copy number alterations of Ser/Thr phosphatases, missense mutations and increased mutant gene expression. Identifying aberrancy of these important AAbs and utilising early antigen expression is key to disease surveillance and therapeutics. Following the exploitation of BCR/ABL kinase inhibition in chronic myeloid leukaemia, efforts have been made to explore PP2A phosphatase reactivation/inhibition in anti-tumour therapy. PPP2R1A dysregulation was individually a significant independent predictor of poor survival in both cohorts, belonging to the PP2A enzyme family, these complexes exert control over oncogenic signalling pathways (MEK/ERK, Srk-Jnk) and over collateral resistance phosphorylation pathways. Their inhibition in a KRAS-mutant human lung cancer cell line resulted in improved responses with MEK inhibitors (385). Mutations of PPP2R1A significantly enhance cancer cell migration in endometrial and ovarian carcinomas (386), whereas allosteric activation of this wild-type complex induces cell cycle arrest with broad anti-tumour activity (387). Current phase 2 trials in recurrent glioblastoma (NCT03027388) are investigating the role of PP2A inhibitor, LB100. TPTE, a CTAG also exerts

PTEN-related tyrosine phosphatase activity and was one of the targets of the liposomal RNA vaccine used in the lipo-MERIT study (378) thus demonstrating the key role of AAb phosphatases in tumorigenesis and how tumours exert oncogenic control through dysregulation of this proliferative brake.

The current era in lung cancer research utilises promising molecular AAbs including auto-antibodies in the blood, complement fragments, circulating microRNAs, circulating tumour DNA, DNA methylation status of tumour tissue, direct profiling of tumour-associated antigens in serum and RNA airway and nasal signatures (330). With the development of artificial intelligence technologies and biological integration, future AAbs prospects will encompass areas such as metabolomics, exhaled breath AAbs, integrating low-dose CT scanning with sputum cell image analysis as well as the use of next generation RNA sequencing for circulating tumour DNA in early lung cancer detection (330). Due to the sheer mass of AAbs in need of clinical validation, standardised metrics of clinical utility are required as well as the use of newer AI-based and machine learning technologies to help select the most robust combinations.

Whilst the aim of this study was to identify a highly prognostic panel for early-stage lung cancer, the biology of the final markers suggests a line of sight to the clinic in terms of adjuvant therapies, in which high expression of our prognostic AAb might indicate a very poor outcome group of patients whose survival; might be improved by targeting some of the AAbs that the autoantibodies have formed against particularly the CTAGs. In our dataset, low CTAG expressers (Figure 5.9) had good outcomes, with 85.4% 5-year overall survival ($p<0.001$), targeting this group is unlikely therefore to be of benefit however high expressers who suffer poor outcomes post-resection, may well be suitable for a CTAG-based polyepitopic RNA

vaccine as an adjunct to standard adjuvant chemotherapy to further eliminate micro-metastatic deposits and cells with high biological propensity for aggressive disease. One must consider the limitations with polyepitopic vaccines in that a dominant epitope may drive immune responses and according to vaccine timing and location, different Th1/Th2 responses will be elicited.

This unique dataset with its robust clinical stratification (death from recurrent disease in 12 months versus long-term disease-free survivorship at greater than 5-year follow-up) and long-term follow-up is well placed to explore proteomic-based differences. This analysis utilised novel machine learning approaches to derive unique AAb signatures with the highest predictive capability in our patient cohorts. The AAbs individually are highly relevant to cancer biology with important roles in key mechanisms that underlie tumorigenesis and are fuelling current clinical trials in cancer medicine. Elucidating the role of CTAGs in cancer medicine and exploring links with Wnt Signalling and phosphatase cascades is of vital importance. These tightly regulated antigens with their almost uniform expression in epithelial carcinomas provide an excellent target for not only prognosticating disease but also as a therapeutic vaccine target, clearly exemplified by the data from the LipoMERIT study (378). Our broad unsupervised interrogation of >1600 AAbs in a robust clinical dataset only serves to reinforce this.

5.5.1 Limitations and Further Work

Despite good performance, the panel identified should also be employed to determine clinical utility in larger external cohorts in a pragmatic setting involving NSCLC patients as well as other cancers to ascertain if the signature is disease specific.

Chapter 6. Discussion

The B cell repertoire in patients with Non-Small Cell Lung cancer plays a clear role at the level of the circulation and in the tumour microenvironment. This is clear in treatment naïve patients and in predicting response to therapy with checkpoint blockade. The dual pro-tumour and anti-tumour B cell responses in NSCLC, underpin the need for proper systematic characterization of antibody expression, spatial localisation, and functional phenotypes of B cells in order to ultimately guide the multi-faceted development of B cell derived immunotherapies.

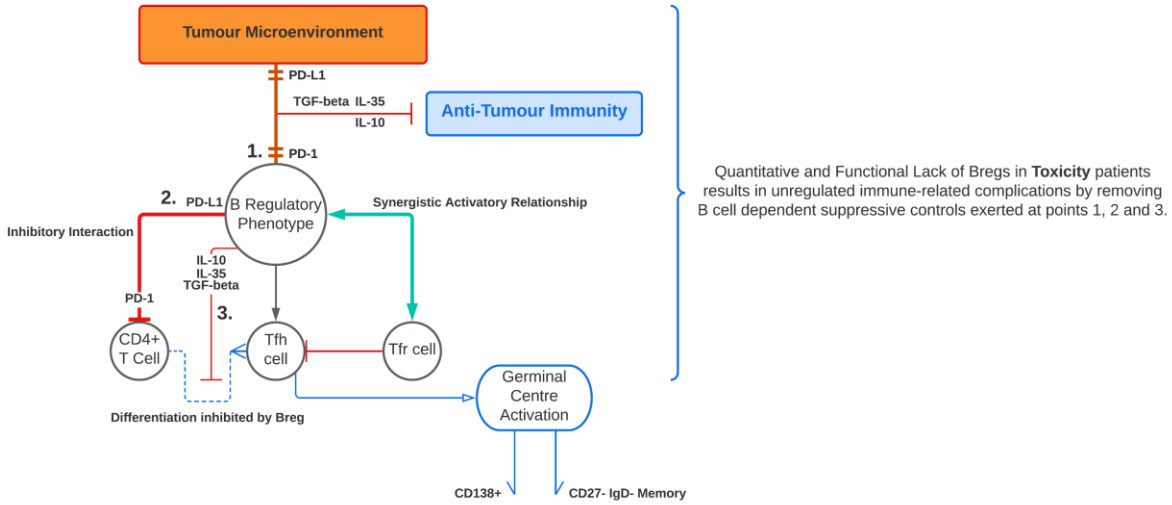


Figure 6.1. This schematic illustrates the interplay between Bregs, CD4+ T cells, Tfh, and Tfr cells and the role of suppressive cytokines and the PD1-PD-L1 axis in mediating downstream responses. Bregs can be PD1 or PDL1 positive. Interactions with tumour cell PD-L1 (1.) or upon encountering PD-L1+ cells, this PD-1 triggering results in acquisition of suppressive functions with suppresses tumour-specific T-cell immunity and promotes cancer growth via IL-10 production. Interaction with CD4+ specific PD-1 via B-cell specific PD-L1 (2.) inhibits Tfh maturation and differentiation from CD4+ T cells and subsequent entry into the germinal centre where Tfh help induces B cell maturation into memory and terminally differentiated plasma cells. These direct PD1-PD-L1 interactions between CD4+ and Breg cells respectively results in Stat5 upregulation and subsequent Tfh suppression. CXCR5+ CD25+ Foxp3+ T follicular regulatory cells work synergistically with Breg cells to skew the balance away from

Tfh maturation and expansion. Tfh suppression also occurs via suppressive cytokine signalling (3.). These three mechanisms are thought to be the B cell exerted brakes on autoimmunity which prevents unregulated immune related adverse events following checkpoint blockade.

6.1 Differential intra-tumoural and circulating B cell phenotype expression exists in the setting of early-stage NSCLC and is predictive of post-operative outcome.

Our interrogation into the B cell repertoire of early-stage NSCLC patients, utilised deep phenotyping using mass cytometry, bulk RNA sequencing and multispectral tissue imaging using immunofluorescence. A clear difference in B cell surrogacy between tumour and blood was demonstrated using mass cytometry where a higher CD138+ plasma cell infiltrate was noted in tumour tissue. These tumour-infiltrating, antibody-producing cells associated with a lower degree of recurrence ($p<0.05$). In the circulation, a higher suppressive B cell presence was linked to recurrence ($p<0.05$) and at the tissue level, these suppressive B cells (CD138+ or CD20+) were noted to reside predominantly in the tumour stroma as opposed to infiltrating the nest.

The ambiguous role of B cells in the tumour environment predetermines the multidirectional development of immunotherapeutic approaches, either supporting positive B cell types or inhibiting the suppressive B cell phenotypes (156).

The lack of a universally accepted method for evaluating antitumor versus pro-tumour B cells makes it difficult to evaluate the role of B cells cancer. In general, the presence of CD20+ B cells or IgG+ plasma cells confers prognostic benefit, however there are discrepancies with this rule (94). This has been observed the with CD138 expression also (154). Suppressive B cells (Bregs) are deleterious in lung cancer and given the common surface markers which are universally shared with effector type B cells, the expansion of robust multi-OMICS techniques will help to serve to delineate these niche populations effectively.

6.1.1 Theoretical Implications

Mimicking the impact of effector B cells in NSCLC is an attractive immunotherapeutic strategy. Whether this is through adoptive transfer of Ig producing B cells or enhancing B cell activation with exogenous ligands, remains to be seen. The latter has been shown to be effective in restricting tumour growth and metastasis in murine models (113,119). Similarly CD138+ induction with Reishi polysaccharide fraction has been shown to suppress tumour in an antibody mediated cytotoxic fashion in the Lewis lung tumour mouse model (388). Our current understanding indicates that tumour-infiltrating B cells are important regulators of lung cancer progression. B cells are multi-faceted; either pro- or anti-tumour as determined by the signals delivered by the TME. Further exploration into this area will serve in developing B cell based immunotherapeutic strategies.

6.2 Circulating B cell changes are predictive of outcome and immune related adverse events (irAE) in NSCLC patients receiving checkpoint blockade therapy

Through testing in independent cohorts, we showed that lung cancer patients developing severe toxicity from anti PD-1/PD-L1 blockade had significant gaps in their B cell profile at diagnosis, specifically affecting suppressive B cell subsets (IL-10+, TGF-β+ or PDL-1+). All subsets detected in this context display strong phenotypic parallels with previously described Breg subsets. Patients who did not subsequently develop severe toxicity had preservation of these subsets at diagnosis and in these patients, the levels of these cells increased on checkpoint blockade whereas in toxicity patients Breg populations either remained static or fell. Our phenotyping data suggests there is circulating presence of Breg populations in toxicity patients however our functional data suggests that the B cells in toxicity patients not only fail to mount an adequate anti-inflammatory cytokine (IL-10) response but also an

adequate pro-inflammatory cytokine (IFN-γ, IL-6, IL-17) response. This polyfunctional defect implies B cell exhaustion in toxicity patients likely driven by the tumour microenvironment, the signal for which remains to be elucidated. This pan-cytokine failure in B cells from toxicity patients explains the lack of a reciprocal relationship between toxicity and response to treatment.

6.2.1 Theoretical Implications

In summary, we show that patients who have inherent functional defects in their Breg repertoire as well as deficits in specific peripheral Breg phenotypes are predisposed to developing severe autoimmunity. Our findings, despite the small size of the external validation cohort, significantly enhance our understanding of the aetiopathogenesis of severe autoimmunity under checkpoint blockade. They have significant implications for our ability both to predict the likelihood of severe immune toxicity with checkpoint blockade, and ultimately also to develop tolerable checkpoint blockade regimens in such high-risk patients. Future translational applications are both potentially diagnostic and therapeutic. Narrower spectrum flow cytometry assays at diagnosis would help formulate an idea of the circulating Breg milieu in patients and when combined with functional IL-10 assays, would stratify risk of toxicity in patients undergoing checkpoint blockade therapy. By stratifying susceptible patients based on Breg profile and comparing to those patients included for treatment according to standard criteria as parallel trial arms, we could gain useful insight into the utility of this type of cellular risk stratification. Similarly, developing a large prospective cohort of patients would inform us as to whether specific Breg signatures or deficiencies correlate with toxicity at anatomical sites. As was shown in murine models of intestinal damage, administering an IL-10 adjuvant either as a stable mRNA vaccine or via CD19+ adoptive B cell transfer, to patients experiencing toxicity, tissue damage could be considerably limited. This

could be a significant therapeutic application if not prophylactic in those who are deemed highly susceptible to toxicity.

6.3 Circulating tumour-associated auto-antibody signatures in NSCLC are predictive of post-operative outcome

We utilised the Sengenics Immunome™ AAb Array [Sengenics, Singapore] containing 1627 AAbs, to screen sera and profile the circulating auto-antibody signatures from a total of 157 non-small cell lung cancer (NSCLC) patients across two independent cohorts. We set out to identify a high-risk sub-group of surgically resectable NSCLC patients who may benefit from adjuvant therapy and explore the biological significance of the identified AAbs along with information relevant to therapeutic application. We implemented a bespoke machine learning approach to investigate the utility of using the pre-resection samples in the context of malignancy to identify sera-based proteomic changes specifically associated with outcome in NSCLC following surgery. This yielded predictive AAb panels which were able to predict outcome in resected NSCLC patients with considerable accuracy, provide improved prognostic accuracy beyond the IASLC staging system and differentiate high and low-risk patients based on panel expression within all disease stages.

6.3.1 Theoretical Implications

Three broad categories of genes comprised our final panel, namely CTAG expression, Wnt signalling AAb aberrancy and serine/threonine AAb phosphatase deregulation. Biologically these pathways are all heavily implicated in tumour progression and potentially the markers identified could bear significant therapeutic implications. Adjuvant therapies could be delivered to those patients determined as high risk by this AAb panel for post-operative recurrence. For example, the highly restricted expression patterns of CTAGs in normal tissues

and ectopic expression in tumour types, makes them highly sought after as targets for cancer vaccines (374). The Lipo-MERIT trial demonstrated strong CD4+ and CD8+ T cell induction along with durable objective clinical benefit in unresectable melanoma patients treated with a poly-antigenic liposomal RNA vaccine with or without combination with anti-PD1 checkpoint blockade therapy (389). High CTAG expressers who suffer poor outcomes post-resection, may well be suitable for a CTAG-based polyepitopic RNA vaccine as an adjunct to standard adjuvant chemotherapy to further eliminate micro-metastatic deposits and cells with high biological propensity for aggressive disease.

6.4 Concluding Remarks

Throughout the course of this project, the generated data has shed light into the role of B cells in the immunobiology of NSCLC, both in early and advanced disease. We have demonstrated clear differences in the surrogate nature of B cells between the circulation and the tumour microenvironment. We have determined a highly predictive prognostic auto-antibody signature which we will be seeking to validate in an independent external cohort. Lastly, we have demonstrated insight into the role of the suppressive B cell and the aetiopathogenesis of immune related adverse events following checkpoint blockade.

B cells exist on a phenotypic spectrum but also a functional spectrum. At any stage of maturation, B cells given the right environmental stimulus can either be effector (supporting anti-cancer CT8+ T cells and produce Ig) or suppressive and produce IL-10, IL-35 or TGF-β. The suppressive B cells can suppress independently of cytokine production and instead dampen immune responses through direct PD-1/PDL-1 interactions. In early-stage disease, we have demonstrated that plasma cells in the TME exist on this functional spectrum with effector phenotypes associating with positive outcomes and vice-versa with suppressive phenotypes.

The natural regulatory plasma cells gather around the tumour nest preventing effector T cell infiltration presumably and inducing the suppressive phenotype in trafficking B cells from the circulation. Suppressive cells however are not for all intents and purposes entirely deleterious in cancer, as we have shown in the context of checkpoint blockade these can be protective against autoimmune complications. In patients who lack these suppressive responses, we postulate that B cells exhibit an exhausted phenotype which pre-dominantly affects the suppressive cytokines but also pro-inflammatory IL-6, IFN-γ and TNF-α, which would explain the lack of reciprocity with response to treatment. The humoural response to tumour associated antigen is still poorly understood, we have attempted to study the autoantibodies in the circulation of patients with early-stage lung cancer and derived a predictive biomarker for relapse. These anti-antibodies are either produced by the effector terminally differentiated plasma cells which we have seen are protective (Objective 1) in response to tumour-associated antigen. It is unclear if these are inhibitory or activatory and whether they are produced solely in response to the tumour or by circulating memory B cells constitutively.

Understanding the environmental signals that induce either phenotype, whether it be locally produced TME stimuli or host stimuli, remains the key to understanding how to support anti-cancer effector responses and in the right situation either support or inhibit the suppressive B cell responses.

There is much to still investigate, however. In early-stage disease, determining clonal differences in B cell populations between recurrence and non-recurrence patients is of key importance, and in particular determining if there are key tumour associated antigens which are driving any clonal changes. We hope to investigate this through single cell RNA sequencing of the B cell receptor in a wider cohort. In the advanced disease setting, we aim to

prospectively validate a bespoke Breg flow cytometry panel on a larger scale to further investigate the role of Bregs in determining outcome post-checkpoint blockade. Moreover, investigating this in other disease settings such as melanoma will be important. Some irAEs may develop on the basis of pre-existing autoimmunity to autoantigens expressed in the target organ (and tumour) and baseline autoantibody may be of value in predicting those at risk of these irAEs. A lack of adequate Breg induction may be permissive of checkpoint-related autoimmunity mediated by activated CD4 cells and exacerbating auto-antibody formation. Much of the recent literature has demonstrated a role for the microbiome in Breg induction and whether there are such differences in NSCLC patients receiving checkpoint blockade who develop toxicity and those who do not, remains to be seen. The microbiome appears crucial not only in driving colitis secondary to checkpoint blockade but other severe irAEs and may be important in determining Breg induction. Exploring the gut and lung microbiome in this setting through murine models will form another facet of our ongoing work. Finally, the role of the functional in vivo B cell response in determining outcome is yet to be properly investigated, we cover this final phase in Chapter 7.

Chapter 7. Further Work

Determining the Functional in vivo B cell response to vaccination with Pneumovax (PPSV23) in NSCLC patients

7.1 Background and Rationale

Our retrospective analysis [Patel AJ et al, *J Thorac Dis* 2020 – Appendix 1] at QEHB to ascertain the proportion of NSCLC patients who are hospitalised due to respiratory infection/bacterial LRTI showed that within the NSCLC cohort at the QEHB, there were 480 admissions over 2 years of which 165 were for respiratory infection (34%). There were 110 in-hospital deaths (22.9% mortality rate during admission) of which 65% of all mortality were due to respiratory infection. Whilst infection appears an enormous unmet need, prophylactic antibiotics may well be detrimental as this inhibits the activity of PD1 checkpoint blockade agents which are rapidly becoming the mainstay of treatment in this patient group (390). Thus, infection prevention is key, and vaccination is a useful tool especially as 18% of all audited respiratory admissions had a positive pneumococcal culture (the most prevalent micro-organism out of all those who were culture positive – Figure 7.1 below) and this is likely to be an under representation as pneumococcus is notoriously difficult to culture from sick patients. Given that these patients are older adults with COPD and are current smokers, it is likely that they will benefit from vaccination from a clinical perspective whilst allowing the investigation of the utility of a functional in vivo B cell model to predict outcome from disease. Antibody responses to pneumonia vaccination have never been undertaken in NSCLC despite multiple pragmatic guidelines recommending this action.

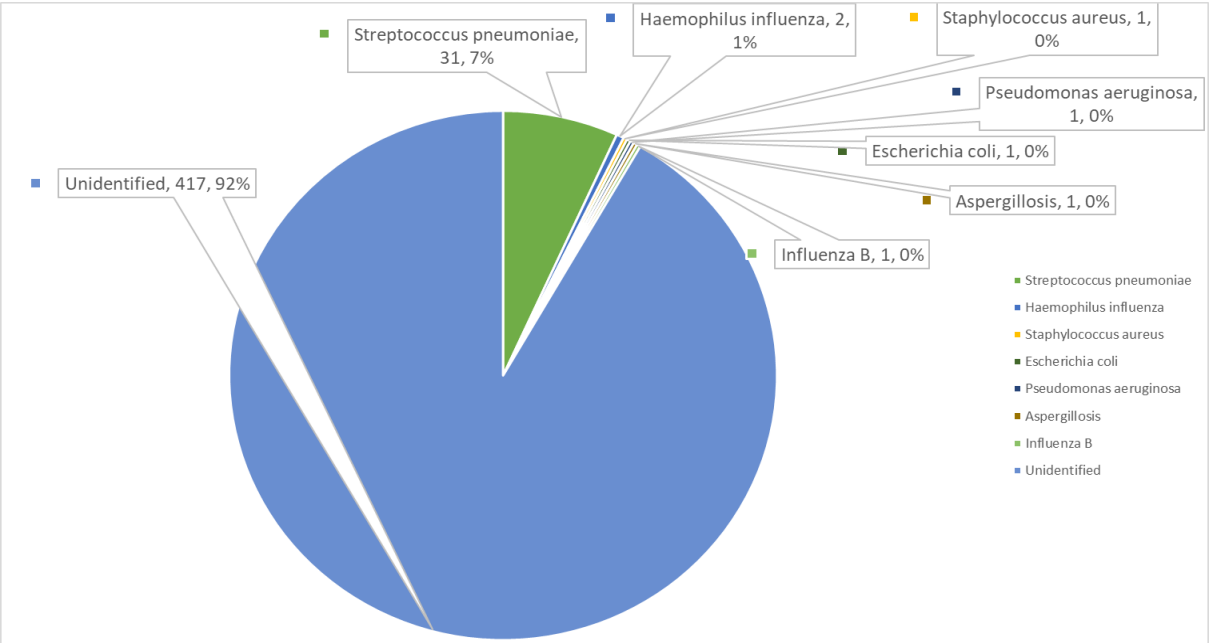


Figure 7.1. Microbiological Repertoire of sputum growth from NSCLC cohort

7.2 Study Design

This study is a basic science investigation involving human participants within University Hospitals Birmingham NHS Trust, with patient recruitment from two sites within the trust; Birmingham Heartland's hospital (surgical NSCLC cohort) and from Queen Elizabeth Hospital Birmingham (QEHB) (advanced disease cohort).

The study has been discussed and presented to a Patient and Public Involvement (PPI) group (below). The results from this cohort will be compared during data analysis with healthy volunteers who are being recruited for vaccination as part of the Time of Day (ToD) study (REC Reference: 17/WM/0426, IRAS ID: 235222).

7.2.1 Study Participants

The study participants will be patients with NSCLC from two cohorts.

Cohort 1: acute surgical NSCLC patients, informed of study at visit 1 (first surgical assessment), recruited and vaccinated at visit 2 (pre-operative consultation).

Cohort 2: advanced disease NSCLC patients, informed of study at visit 1 (first oncology assessment), recruited and vaccinated at visit 2 (pre-treatment consultation).

Participants will receive PPSV23 vaccine in their non-dominant arm by intra-muscular injection in the deltoid muscle bulk. This will take place in the pre-operative setting or pre-initiation of chemo/immunotherapy.

7.2.2 Planned Size of recruitment target

The aim is to recruit 40 patients across the cohort; this is in keeping with one of our primary endpoints of feasibility. The recruitment period is 12 months from the start of the study.

7.2.3 Follow-up duration

12 months

7.2.4 Planned Study Period*

April 2019-October 2021

*Owing to the development and resurgence of the COVID-19 pandemic, this study has been paused by the R&D department at the trust until COVID-19 pressures ease and we are able to resume recruitment.

7.2.5 Research Question/Aim(s)

The over-arching aim of the study is to determine the functional in vivo B cell response in NSCLC patients through vaccination. We plan to achieve this by performing functional antibody testing on NSCLC patients by assessing their response to vaccination in the pre-treatment phase (pre-surgery or pre-treatment with immunotherapy +/- chemotherapy).

7.2.6 Patient & Public Involvement Group

The study has been discussed and presented to the RESOLVE group, which is the national thoracic surgical patient and public involvement group. This took place at the national society of cardiothoracic surgery conference in London (March 2019). Their input has been most favourable and helped to design this study. A member of the RESOLVE group will provide input either in person or via telephone conference at all future study management meetings.

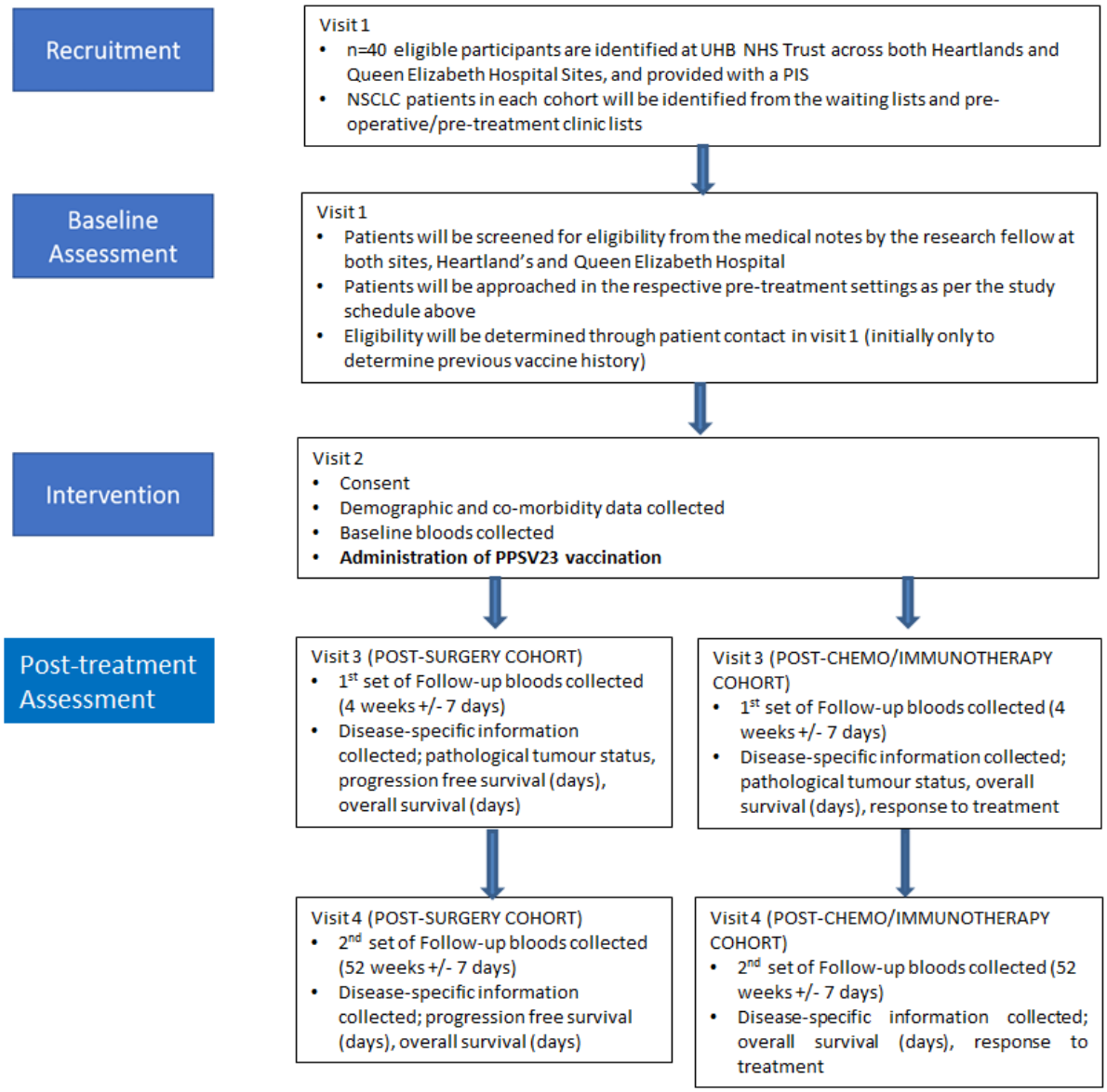
7.2.7 Design and Timing Rationale

This is a high risk cohort of patients particularly for post-operative infection; patients who undergo thoracic surgical resection are much more likely to develop a post-operative pulmonary complication such as pneumonia (391). This provides good rationale to vaccinate;

we are electing not to vaccinate in the immediate pre-operative period as some patients may develop a reaction to the vaccine and become unwell which would preclude them from undergoing their lung cancer resection. Furthermore, through work examining the innate immune responses to pulmonary resection; it was found that greater pro- as well as anti-inflammatory responses were detected via thoracotomy as compared with minimally invasive surgery (392). When the body is stressed by surgery, there is a propensity for a more exaggerated immune response which correlates with the invasiveness of the surgery. Therefore, vaccinating in the immediate post-operative period will make functional B cell responses largely uninterpretable owing to the confounding from the post-surgical stress response.

Thus, vaccinating all patients pre-operatively, allowing a minimum of 2-4 weeks from the time of vaccination to surgery will confer added protection against the development of post-operative pulmonary complications whilst still allowing for any vaccine induced reactions to occur and settle prior to surgery. Similarly in the advanced disease cohort, vaccinating patient prior to commencing any therapy within a similar time frame will offer maximal protection whilst still allowing the standard of care to take place. This will maintain the ability to examine the functional B cell response in the context of NSCLC in this unique cohort of previously unstudied patients. The results from this cohort will be compared during data analysis with healthy volunteers who are being recruited for vaccination as part of the Time of Day (ToD) study (REC Reference: 17/WM/0426, IRAS ID: 235222).

7.2.8 Study Flow Chart



7.2.9 Pneumococcus

Streptococcus pneumoniae (S. pneumoniae), a bacterium also known as a pneumococcus, is a major cause of pneumonia, sepsis and meningitis. Mortality rates remain high despite modern antibiotics, so successful vaccination is vital in order to prevent infection. Research suggests that there is a correlation between invasive pneumococcal disease (IPD) and dysfunctional B-cell responses (393,394). In 2003, a single dose of Pneumovax-23 (PPSV23) was recommended for all individuals aged 65 and over, and individuals below the age of 65 with

a long-term health condition such as chronic obstructive pulmonary disease (COPD), diabetes, chronic kidney or liver failure or immunosuppression (The Green Book, 2017). This is the vaccine which participants in this study will be receiving. PPSV23 contains purified capsular polysaccharide to 23 different pneumococcal capsular serotypes (1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19F, 19A, 20, 22F, 23F, 33F).

7.2.10 PPSV23 Treatment

The PPSV23 vaccination is an NHS standard of care treatment available to all adults aged 65 and over, as well as anyone from the ages of 2 to 65 years with a long-term health condition such as COPD, diabetes, chronic kidney or liver failure or immunosuppression.

People over 65 only need a single vaccination at any time which will protect for life; those with a long-term health condition may just need a single one-off vaccination or five-yearly vaccination depending on their underlying health problem (https://www.nhs.uk/conditions/vaccinations/pneumococcal-vaccination/). As the vaccination can be offered at any time on the NHS in accordance with the above criteria; participants will receive PPSV23 vaccine in their non-dominant arm by intra-muscular injection in the deltoid muscle bulk. This will take place during visit 2.

7.3 Study Objectives

This is an exploratory physiological investigation which seeks to determine the functional *in vivo* B cell response in NSCLC.

7.3.1 Primary Objectives

- Explore the functional *in vivo* B cell response at 4 weeks (+/- 7 days) and 52 weeks (+/- 7 days) post vaccination (nationally recommended, standard of care vaccination, PPSV23) in NSCLC patients who are due to undergo treatment by surgery and those who are due to start first-line immunotherapy +/- chemotherapy.
- Assess the feasibility of carrying out this study using the proposed research methods in the NSCLC cohorts described. Feasibility is defined in terms of study uptake rates (% of those approached who are recruited), recruitment rates (n/month) and retention rates (% of those who provide samples and data at follow-up appointments).

7.3.2 Secondary Objectives

- Determine if the functional *in vivo* B cell response correlates with outcome or prognosis in this cohort of patients.
- Compare the functional *in vivo* B cell response in NSCLC patients with trends seen in the healthy volunteer population from the Time-of-Day study, as an exploratory endpoint (REC reference: 17/WM/0426, IRAS ID: 235222).
- Report on trends seen in morbidity and mortality secondary to respiratory infection in this cohort of patients, as an exploratory endpoint, if time permits.
- Through serial blood samples and laboratory testing of AAbs of specific adaptive immunity, investigate the mechanism by which NSCLC determines a functional immune response.

7.3.3 Laboratory Outcomes

- Vaccine specific serum antibody concentration (IgG, IgA and IgM) following a single dose of PPSV23 vaccination at 4 weeks (+/- 7 days) and 52 weeks (+/- 7 days) post

vaccination to determine the functional B cell response (quality, quantity and longevity of the antibody response) in the context of NSCLC.

- Ig levels by class and by antibodies specific for bacterial target antigens using standard laboratory ELISA and multiplex immunoassays including functional antibody tests (Luminex platform, Opsonophagocytic killing assays (OPKA) and total IgG, A and M).
- Antigen specific B cell levels, to be measured by ELISpot and Cytometry by Time of Flight (CyTOF) method.
- Investigation of AAbs of inflammation, innate and specific immunity will be made to determine how the NSCLC microenvironment influences immune response. This will include numbers and activation status of neutrophils, monocytes, dendritic cells, natural killer cells, T and B lymphocyte subsets measured in serial blood samples at 4 weeks (+/- 7 days) and 52 weeks (+/- 7 days) post vaccination.

7.3.4 Disease Specific Outcomes

- Overall and Progression-free survival in days for both treatment cohorts.
- Pathological staging of disease (i.e. Tumour status).

7.3.5 Proposed Analysis Methods

1. Descriptive stats for patients (split by cohort of NSCLC patients and healthy controls), use plots and tables where appropriate.
2. Investigate results using Chi-square (or Fisher's Exact for low numbers).
3. Compare pre- and post-vaccination (4 and 52 week) blood results of NSCLC and control patients using mixed effects models with adjustment for patient characteristics, baseline blood measurements and prognostic factors.

4. Kaplan-Meier survival plots (for PFS and OS) produced as appropriate for data (e.g., curves comparing B cell signatures of the NSCLC and outcomes).
5. Cox proportional hazard regression models constructed for each outcome: PFS and OS, using NSCLC and control groups, and B cell signatures as categorical explanatory variables, adjusting for patient characteristics.

7.4 Eligibility Criteria

7.4.1 Inclusion criteria

- Patient is able to give informed consent and undertake the patient assessment as detailed in the protocol.
- Patient has resectable NSCLC and be scheduled for surgical resection of the NSCLC from the time of baseline blood withdrawal OR has advanced disease NSCLC that is eligible for chemo/immunotherapy.
- Patient is aged 65 or over OR patient is aged between 18 to 65 years AND have a long-term health condition such as COPD, diabetes, chronic kidney disease or liver disease, and therefore meeting standard of care requirements for PPSV23 vaccination.

7.4.2 Exclusion criteria

- Patient has been previously vaccinated with the PPSV23 vaccine in the last 5 years.
- Patient takes long-term oral steroids or other form of long-term immunosuppression that will affect vaccine response (a 14-day washout period is required for a short course of steroids or immunosuppression).
- Patient has a suppressed immune system caused by a health condition such as HIV.
- Patient is known to be allergic to components of the study vaccines.

- Patient is suffering from an acute infective illness (as determined by the clinician at visit 2).

7.5 Recruitment Target

7.5.1 Size of recruitment target

The analyses of this study are exploratory, and as one of the objectives is to assess feasibility, a partially qualitative outcome, it is difficult to effectively power the study. However, the following considerations have been taken into account. In 1995 Browne recommended a sample of 30 subjects or more as a general rule for pilot studies when the aim is to obtain variance estimates for a sample size calculation sample size for a bigger Randomised Controlled Trial. Kieser and Wassmer (395) investigated the theoretical basis for Browne's recommendation and concluded that a sample size of 40 will be adequate for applying Browne's method in the subsequent sample size calculation. We have therefore chosen an intended sample size of 40 patients for a given disease area. This sample size will allow the mean value for a standardised, normally distributed primary outcome measure to be estimated to within ±4% with 90% confidence, i.e., 90% confidence interval (CI) width of 8%, assuming a relative standard deviation of 10%. Recruitment of at least 44 patients per group will allow for the sample size of 40 to be met whilst allowing for 10% dropout. A control group of health elderly patients (no NSCLC) who will receive vaccination as part of the Time-of-Day study (REC reference: 17/WM/0426, IRAS ID: 235222) will be used to compare the trends in the incidence of respiratory infection as well as their functional B cell response to vaccination.

7.6 Trial Delivery

Once the COVID-19 pressures have eased, we will carry out this study in line with the remits set out in the MHRA ethics approval (REC 19/WM/0182, IRAS 262536, ISRCTN 49173282).

References

1. Chevrier S, Levine JH, Zanotelli VRT, Silina K, Schulz D, Bacac M, et al. An Immune Atlas of Clear Cell Renal Cell Carcinoma. Cell. 2017 May 4;169(4):736-749.e18.

2. Das R, Bar N, Ferreira M, Newman AM, Zhang L, Bailur JK, et al. Early B cell changes predict autoimmunity following combination immune checkpoint blockade. J Clin Invest. 2018 Jan 8;128(2):715–20.

3. Osorio JC, Ni A, Chaft JE, Pollina R, Kasler MK, Stephens D, et al. Antibody-mediated thyroid dysfunction during T-cell checkpoint blockade in patients with non-small-cell lung cancer. Ann Oncol Off J Eur Soc Med Oncol. 2017 01;28(3):583–9.

4. Kuehn HS, Ouyang W, Lo B, Deenick EK, Niemela JE, Avery DT, et al. Immune dysregulation in human subjects with heterozygous germline mutations in CTLA4. Science. 2014 Sep 26;345(6204):1623–7.

5. Bade BC, Dela Cruz CS. Lung Cancer 2020: Epidemiology, Etiology, and Prevention. Clin Chest Med. 2020 Mar;41(1):1–24.

6. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries - Sung - - CA: A Cancer Journal for Clinicians - Wiley Online Library [Internet]. [cited 2021 Apr 12]. Available from: https://acsjournals.onlinelibrary.wiley.com/doi/full/10.3322/caac.21660

7. Lung cancer statistics [Internet]. Cancer Research UK. 2015 [cited 2021 Jun 16]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer

8. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet Lond Engl. 2018 Mar 17;391(10125):1023–75.

9. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The Eighth Edition Lung Cancer Stage Classification. Chest. 2017 Jan;151(1):193–203.

10. Detterbeck FC, Gibson CJ. Turning Gray: The Natural History of Lung Cancer Over Time. J Thorac Oncol. 2008 Jul 1;3(7):781–92.

11. Janssen-Heijnen MLG, van Erning FN, De Ruysscher DK, Coebergh JWW, Groen HJM. Variation in causes of death in patients with non-small cell lung cancer according to stage and time since diagnosis. Ann Oncol. 2015 May;26(5):902–7.

12. Lung cancer [Internet]. nhs.uk. 2017 [cited 2021 Apr 12]. Available from: https://www.nhs.uk/conditions/lung-cancer/

13. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WEE, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. J Thorac Oncol. 2016 Jan;11(1):39–51.

14. Types of lung cancer | Cancer Research UK [Internet]. [cited 2021 Apr 12]. Available from: https://www.cancerresearchuk.org/about-cancer/lung-cancer/stages-types-grades/types

15. The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. N Engl J Med. 2011 Aug 4;365(5):395–409.

16. Pastorino U, Silva M, Sestini S, Sabia F, Boeri M, Cantarutti A, et al. Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy. Ann Oncol Off J Eur Soc Med Oncol. 2019 Jul 1;30(7):1162–9.

17. Oudkerk M, Devaraj A, Vliegenthart R, Henzler T, Prosch H, Heussel CP, et al. European position statement on lung cancer screening. Lancet Oncol. 2017 Dec;18(12):e754–66.

18. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. N Engl J Med. 2020 Feb 6;382(6):503–13.

19. Field JK, Duffy SW, Baldwin DR, Whynes DK, Devaraj A, Brain KE, et al. UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. Thorax. 2016 Feb;71(2):161–70.

20. Lung cancer: MedlinePlus Genetics [Internet]. [cited 2021 Apr 12]. Available from: https://medlineplus.gov/genetics/condition/lung-cancer/

21. El-Telbany A, Ma PC. Cancer Genes in Lung Cancer: Racial Disparities: Are There Any? Genes Cancer. 2012 Jul;3(7–8):467.

22. Modrek B, Ge L, Pandita A, Lin E, Mohan S, Yue P, et al. Oncogenic activating mutations are associated with local copy gain. Mol Cancer Res MCR. 2009 Aug;7(8):1244–52.

23. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004 May 20;350(21):2129–39.

24. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004 Jun 4;304(5676):1497–500.

25. Aviel-Ronen S, Blackhall FH, Shepherd FA, Tsao MS. K-ras mutations in non-small-cell lung carcinoma: a review. Clin Lung Cancer. 2006 Jul;8(1):30–8.

26. Marchetti A, Martella C, Felicioni L, Barassi F, Salvatore S, Chella A, et al. EGFR mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. J Clin Oncol Off J Am Soc Clin Oncol. 2005 Feb 1;23(4):857–65.

27. Ma PC, Jagadeeswaran R, Jagadeesh S, Tretiakova MS, Nallasura V, Fox EA, et al. Functional expression and mutations of c-Met and its therapeutic inhibition with SU11274 and small interfering RNA in non-small cell lung cancer. Cancer Res. 2005 Feb 15;65(4):1479–88.

28. Wadowska K, Bil-Lula I, Trembecki Ł, Śliwińska-Mossoń M. Title: Genetic Markers in Lung Cancer Diagnosis: A Review. Int J Mol Sci [Internet]. 2020 Jun 27 [cited 2021 Apr 14];21(13). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7369725/

29. Olivier M, Hollstein M, Hainaut P. TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. Cold Spring Harb Perspect Biol [Internet]. 2010 Jan [cited 2021 Apr 14];2(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2827900/

30. Rurańska B, Stawicka M, Godlewski D. The role of p53 gene in lung cancer with special emphasis on hereditary types. Rep Pract Oncol. 1997 Jan 1;2(2):56.

31. Sanchez-Cespedes M, Parrella P, Esteller M, Nomoto S, Trink B, Engles JM, et al. Inactivation of LKB1/STK11 is a common event in adenocarcinomas of the lung. Cancer Res. 2002 Jul 1;62(13):3659–62.

32. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007 Aug 2;448(7153):561–6.

33. Shinmura K, Kageyama S, Tao H, Bunai T, Suzuki M, Kamo T, et al. EML4-ALK fusion transcripts, but no NPM-, TPM3-, CLTC-, ATIC-, or TFG-ALK fusion transcripts, in non-small cell lung carcinomas. Lung Cancer Amst Neth. 2008 Aug;61(2):163–9.

34. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. N Engl J Med. 2017 Jun 1;376(22):2109–21.

35. Bailey C, Black JRM, Reading JL, Litchfield K, Turajlic S, McGranahan N, et al. Tracking Cancer Evolution through the Disease Course. Cancer Discov. 2021 Apr;11(4):916–32.

36. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature. 2017 Apr 26;545(7655):446–51.

37. Biswas D, Birkbak NJ, Rosenthal R, Hiley CT, Lim EL, Papp K, et al. A clonal expression AAb associates with lung cancer mortality. Nat Med. 2019 Oct;25(10):1540–8.

38. The TRACERx consortium, Rosenthal R, Cadieux EL, Salgado R, Bakir MA, Moore DA, et al. Neoantigen-directed immune escape in lung cancer evolution. Nature. 2019 Mar;567(7749):479–85.

39. AbdulJabbar K, Raza SEA, Rosenthal R, Jamal-Hanjani M, Veeriah S, Akarca A, et al. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. Nat Med. 2020 Jul;26(7):1054–62.

40. Kelley MJ, McCrory DC. Prevention of lung cancer: summary of published evidence. Chest. 2003 Jan;123(1 Suppl):50S-59S.

41. Dela Cruz CS, Tanoue LT, Matthay RA. Lung Cancer: Epidemiology, Etiology, and Prevention. Clin Chest Med. 2011 Dec;32(4):605–44.

42. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, et al. The LLP risk model: an individual risk prediction model for lung cancer. Br J Cancer. 2008 Jan 29;98(2):270–6.

43. Strand TE, Rostad H, Møller B, Norstein J. Survival after resection for primary lung cancer: a population based study of 3211 resected patients. Thorax. 2006 Aug 1;61(8):710–5.

44. Sihoe ADL, Van Schil P. Non-small cell lung cancer: When to offer sublobar resection. Lung Cancer. 2014 Nov 1;86(2):115–20.

45. Ginsberg RJ, Rubinstein LV. Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. Lung Cancer Study Group. Ann Thorac Surg. 1995 Sep;60(3):615–22; discussion 622-623.

46. Duma N, Santana-Davila R, Molina JR. Non–Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. Mayo Clin Proc. 2019 Aug 1;94(8):1623–40.

47. Song WA, Zhou NK, Wang W, Chu XY, Liang CY, Tian XD, et al. Survival benefit of neoadjuvant chemotherapy in non-small cell lung cancer: an updated meta-analysis of 13 randomized control trials. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 2010 Apr;5(4):510–6.

48. Felip E, Rosell R, Maestre JA, Rodríguez-Paniagua JM, Morán T, Astudillo J, et al. Preoperative chemotherapy plus surgery versus surgery plus adjuvant chemotherapy versus surgery alone in early-stage non-small-cell lung cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2010 Jul 1;28(19):3138–45.

49. Pignon JP, Tribodet H, Scagliotti GV, Douillard JY, Shepherd FA, Stephens RJ, et al. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. J Clin Oncol Off J Am Soc Clin Oncol. 2008 Jul 20;26(21):3552–9.

50. Antonia SJ, Villegas A, Daniel D, Vicente D, Murakami S, Hui R, et al. Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. N Engl J Med. 2018 Dec 13;379(24):2342–50.

51. Faivre-Finn C, Vicente D, Kurata T, Planchard D, Paz-Ares L, Vansteenkiste JF, et al. Four-Year Survival With Durvalumab After Chemoradiotherapy in Stage III NSCLC-an Update From the PACIFIC Trial. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 2021 Jan 19;

52. Hegde PS, Chen DS. Top 10 Challenges in Cancer Immunotherapy. Immunity. 2020 Jan 14;52(1):17–35.

53. Parrott DV, De Sousa MA, East J. Thymus-dependent areas in the lymphoid organs of neonatally thymectomized mice. J Exp Med. 1966 Jan 1;123(1):191–204.

54. van den Broek T, Borghans JAM, van Wijk F. The full spectrum of human naive T cells. Nat Rev Immunol. 2018 Jun;18(6):363–73.

55. Kumar BV, Connors TJ, Farber DL. Human T Cell Development, Localization, and Function throughout Life. Immunity. 2018 Feb 20;48(2):202–13.

56. Vignali DAA, Collison LW, Workman CJ. How regulatory T cells work. Nat Rev Immunol. 2008 Jul;8(7):523–32.

57. Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. Nat Rev Immunol. 2020 Nov;20(11):651–68.

58. Fife BT, Bluestone JA. Control of peripheral T-cell tolerance and autoimmunity via the CTLA-4 and PD-1 pathways. Immunol Rev. 2008 Aug;224:166–82.

59. Fritz JM, Lenardo MJ. Development of immune checkpoint therapy for cancer. J Exp Med. 2019 Jun 3;216(6):1244–54.

60. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Rutkowski P, Lao CD, et al. Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. N Engl J Med. 2019 17;381(16):1535–46.

61. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med. 2010 Aug 19;363(8):711–23.

62. Haanen JB a. G, Carbonnel F, Robert C, Kerr KM, Peters S, Larkin J, et al. Management of toxicities from immunotherapy: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol Off J Eur Soc Med Oncol. 2017 Jul 1;28(suppl_4):iv119–42.

63. Kumar V, Chaudhary N, Garg M, Floudas CS, Soni P, Chandra AB. Current Diagnosis and Management of Immune Related Adverse Events (irAEs) Induced by Immune Checkpoint Inhibitor Therapy. Front Pharmacol. 2017;8:49.

64. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ, Wu YL, et al. Lung cancer: current therapies and new targeted treatments. The Lancet. 2017 Jan 21;389(10066):299–311.

65. Garon EB, Hellmann MD, Rizvi NA, Carcereny E, Leighl NB, Ahn MJ, et al. Five-Year Overall Survival for Patients With Advanced Non–Small-Cell Lung Cancer Treated With Pembrolizumab: Results From the Phase I KEYNOTE-001 Study. J Clin Oncol Off J Am Soc Clin Oncol. 2019 Oct 1;37(28):2518–27.

66. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Updated Analysis of KEYNOTE-024: Pembrolizumab Versus Platinum-Based Chemotherapy for Advanced Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score of 50% or Greater. J Clin Oncol Off J Am Soc Clin Oncol. 2019 01;37(7):537–46.

67. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer. N Engl J Med. 2015 Oct 22;373(17):1627–39.

68. Gadgeel S, Rodríguez-Abreu D, Speranza G, Esteban E, Felip E, Dómine M, et al. Updated Analysis From KEYNOTE-189: Pembrolizumab or Placebo Plus Pemetrexed and Platinum for Previously Untreated Metastatic Nonsquamous Non-Small-Cell Lung Cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2020 Mar 9;JCO1903136.

69. Paz-Ares L, Vicente D, Tafreshi A, Robinson A, Soto Parra H, Mazières J, et al. A Randomized, Placebo-Controlled Trial of Pembrolizumab Plus Chemotherapy in Patients With Metastatic Squamous NSCLC: Protocol-Specified Final Analysis of KEYNOTE-407. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 2020 Oct;15(10):1657–69.

70. Paz-Ares L, Luft A, Vicente D, Tafreshi A, Gümüş M, Mazières J, et al. Pembrolizumab plus Chemotherapy for Squamous Non-Small-Cell Lung Cancer. N Engl J Med. 2018 Nov 22;379(21):2040–51.

71. Borghaei H, Gettinger S, Vokes EE, Chow LQM, Burgio MA, de Castro Carpeno J, et al. Five-Year Outcomes From the Randomized, Phase III Trials CheckMate 017 and 057: Nivolumab

Versus Docetaxel in Previously Treated Non-Small-Cell Lung Cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2021 Mar 1;39(7):723–33.

72. Jänne PA, Yang JCH, Kim DW, Planchard D, Ohe Y, Ramalingam SS, et al. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. N Engl J Med. 2015 Apr 30;372(18):1689–99.

73. Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, et al. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. Science. 2007 May 18;316(5827):1039–43.

74. Piotrowska Z, Thress KS, Mooradian M, Heist RS, Azzoli CG, Temel JS, et al. MET amplification (amp) as a resistance mechanism to osimertinib. J Clin Oncol. 2017 May 20;35(15_suppl):9020–9020.

75. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med. 2010 Oct 28;363(18):1693–703.

76. Shaw AT, Kim DW, Nakagawa K, Seto T, Crinó L, Ahn MJ, et al. Crizotinib versus Chemotherapy in Advanced ALK-Positive Lung Cancer. N Engl J Med. 2013 Jun 20;368(25):2385–94.

77. Lin JJ, Riely GJ, Shaw AT. Targeting ALK: Precision Medicine Takes on Drug Resistance. Cancer Discov. 2017 Feb;7(2):137–55.

78. Shaw AT, Ou SHI, Bang YJ, Camidge DR, Solomon BJ, Salgia R, et al. Crizotinib in ROS1-rearranged non-small-cell lung cancer. N Engl J Med. 2014 Nov 20;371(21):1963–71.

79. Katayama R, Kobayashi Y, Friboulet L, Lockerman EL, Koike S, Shaw AT, et al. Cabozantinib overcomes crizotinib resistance in ROS1 fusion-positive cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 2015 Jan 1;21(1):166–74.

80. Kinno T, Tsuta K, Shiraishi K, Mizukami T, Suzuki M, Yoshida A, et al. Clinicopathological features of nonsmall cell lung carcinomas with BRAF mutations. Ann Oncol Off J Eur Soc Med Oncol. 2014 Jan;25(1):138–42.

81. Drilon A, Laetsch TW, Kummar S, DuBois SG, Lassen UN, Demetri GD, et al. Efficacy of Larotrectinib in TRK Fusion-Positive Cancers in Adults and Children. N Engl J Med. 2018 Feb 22;378(8):731–9.

82. Suda K, Tomizawa K, Mitsudomi T. Biological and clinical significance of KRAS mutations in lung cancer: an oncogenic driver that contrasts with EGFR mutation. Cancer Metastasis Rev. 2010 Mar;29(1):49–60.

83. Suzuki K, Kachala SS, Kadota K, Shen R, Mo Q, Beer DG, et al. Prognostic immune markers in non-small cell lung cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 2011 Aug 15;17(16):5247–56.

84. Stankovic B, Bjørhovde HAK, Skarshaug R, Aamodt H, Frafjord A, Müller E, et al. Immune Cell Composition in Human Non-small Cell Lung Cancer. Front Immunol. 2019 Feb 1;9:3101.

85. Galon J, Pagès F, Marincola FM, Thurin M, Trinchieri G, Fox BA, et al. The immune score as a new possible approach for the classification of cancer. J Transl Med. 2012;10(1):1.

86. Ruffini E, Asioli S, Filosso PL, Lyberis P, Bruna MC, Macrì L, et al. Clinical significance of tumor-infiltrating lymphocytes in lung neoplasms. Ann Thorac Surg. 2009 Feb;87(2):365–71; discussion 371-372.

87. Dieu-Nosjean MC, Antoine M, Danel C, Heudes D, Wislez M, Poulot V, et al. Long-term survival for patients with non-small-cell lung cancer with intratumoral lymphoid structures. J Clin Oncol Off J Am Soc Clin Oncol. 2008 Sep 20;26(27):4410–7.

88. Pollard JW. Tumour-educated macrophages promote tumour progression and metastasis. Nat Rev Cancer. 2004 Jan;4(1):71–8.

89. Ohri CM, Shikotra A, Green RH, Waller DA, Bradding P. Macrophages within NSCLC tumour islets are predominantly of a cytotoxic M1 phenotype associated with extended survival. Eur Respir J. 2009 Jan;33(1):118–26.

90. Ohtaki Y, Ishii G, Nagai K, Ashimine S, Kuwata T, Hishida T, et al. Stromal macrophage expressing CD204 is associated with tumor aggressiveness in lung adenocarcinoma. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 2010 Oct;5(10):1507–15.

91. Shimizu K, Nakata M, Hirami Y, Yukawa T, Maeda A, Tanemoto K. Tumor-infiltrating Foxp3+ regulatory T cells are correlated with cyclooxygenase-2 expression and are associated with recurrence in resected non-small cell lung cancer. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 2010 May;5(5):585–90.

92. Sharma S, Yang SC, Zhu L, Reckamp K, Gardner B, Baratelli F, et al. Tumor cyclooxygenase-2/prostaglandin E2-dependent promotion of FOXP3 expression and CD4+ CD25+ T regulatory cell activities in lung cancer. Cancer Res. 2005 Jun 15;65(12):5211–20.

93. Palucka AK, Coussens LM. The Basis of OncoImmunology. Cell. 2016 Mar 10;164(6):1233–47.

94. Wang S si, Liu W, Ly D, Xu H, Qu L, Zhang L. Tumor-infiltrating B cells: their role and application in anti-tumor immunity in lung cancer. Cell Mol Immunol. 2018 Apr 8;1.

95. Burger JA, Wiestner A. Targeting B cell receptor signalling in cancer: preclinical and clinical advances. Nat Rev Cancer. 2018 Jan 19;18(3):148–67.

96. Pagès F, Sanchez-Cabo F, Mlecnik B, Kirilovsky A, Nilsson M, Damotte D, et al. Effector Memory T Cells, Early Metastasis, and Survival in Colorectal Cancer. N Engl J Med. 2005;13.

97. Gottlin EB, Bentley RC, Campa MJ, Pisetsky DS, Herndon JE, Patz EF. The Association of Intratumoral Germinal Centers with early-stage non-small cell lung cancer. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 2011 Oct;6(10):1687–90.

98. Drayton DL, Liao S, Mounzer RH, Ruddle NH. Lymphoid organ development: from ontogeny to neogenesis. Nat Immunol. 2006 Apr;7(4):344–53.

99. Janeway CA, Ron J, Katz ME. The B cell is the initiating antigen-presenting cell in peripheral lymph nodes. J Immunol Baltim Md 1950. 1987 Feb 15;138(4):1051–5.

100. Bouaziz JD, Yanaba K, Venturi GM, Wang Y, Tisch RM, Poe JC, et al. Therapeutic B cell depletion impairs adaptive and autoreactive CD4+ T cell activation in mice. Proc Natl Acad Sci. 2007 Dec 26;104(52):20878–83.

101. Rodríguez-Pinto D. B cells as antigen presenting cells. Cell Immunol. 2005 Dec;238(2):67–75.

102. Shaffer AL, Lin KI, Kuo TC, Yu X, Hurt EM, Rosenwald A, et al. Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. Immunity. 2002 Jul;17(1):51–62.

103. LeBien TW, Tedder TF. B lymphocytes: how they develop and function. Blood. 2008 Sep 1;112(5):1570–80.

104. Rosser EC, Mauri C. Regulatory B cells: origin, phenotype, and function. Immunity. 2015 Apr 21;42(4):607–12.

105. Fillatreau S, Sweenie CH, McGeachy MJ, Gray D, Anderton SM. B cells regulate autoimmunity by provision of IL-10. Nat Immunol. 2002 Oct;3(10):944–50.

106. Carter NA, Vasconcellos R, Rosser EC, Tulone C, Munoz-Suano A, Kamanaka M, et al. Mice Lacking Endogenous IL-10-Producing Regulatory B Cells Develop Exacerbated Disease and Present with an Increased Frequency of Th1/Th17 but a Decrease in Regulatory T Cells. J Immunol. 2011 May 15;186(10):5569–79.

107. Flores-Borja F, Bosma A, Ng D, Reddy V, Ehrenstein MR, Isenberg DA, et al. CD19+CD24hiCD38hi B Cells Maintain Regulatory T Cells While Limiting TH1 and TH17 Differentiation. Sci Transl Med. 2013 Feb 20;5(173):173ra23-173ra23.

108. Tian J, Zekzer D, Hanssen L, Lu Y, Olcott A, Kaufman DL. Lipopolysaccharide-Activated B Cells Down-Regulate Th1 Immunity and Prevent Autoimmune Diabetes in Nonobese Diabetic Mice. J Immunol. 2001 Jul 15;167(2):1081–9.

109. Parekh VV, Prasad DVR, Banerjee PP, Joshi BN, Kumar A, Mishra GC. B Cells Activated by Lipopolysaccharide, But Not By Anti-Ig and Anti-CD40 Antibody, Induce Anergy in CD8+ T Cells: Role of TGF- 1. J Immunol. 2003 Jun 15;170(12):5897–911.

110. Shen P, Roch T, Lampropoulou V, O'Connor RA, Stervbo U, Hilgenberg E, et al. IL-35-producing B cells are critical regulators of immunity during autoimmune and infectious diseases. Nature. 2014 Feb 23;507(7492):366–70.

111. Lighaam LC, Unger PPA, Vredevoogd DW, Verhoeven D, Vermeulen E, Turksma AW, et al. In vitro-Induced Human IL-10+ B Cells Do Not Show a Subset-Defining Marker Signature and Plastically Co-express IL-10 With Pro-Inflammatory Cytokines. Front Immunol. 2018 Sep 5;9:1913.

112. Olkhanud PB, Damdinsuren B, Bodogai M, Gress RE, Sen R, Wejksza K, et al. Tumor-Evoked Regulatory B Cells Promote Breast Cancer Metastasis by Converting Resting CD4+ T Cells to T-Regulatory Cells. Cancer Res. 2011 May 15;71(10):3505–15.

113. Bodogai M, Lee Chang C, Wejksza K, Lai J, Merino M, Wersto RP, et al. Anti-CD20 antibody promotes cancer escape via enrichment of tumor-evoked regulatory B cells expressing low levels of CD20 and CD137L. Cancer Res. 2013 Apr 1;73(7):2127–38.

114. Visser KE de, Korets LV, Coussens LM. De novo carcinogenesis promoted by chronic inflammation is B lymphocyte dependent. Cancer Cell. 2005 May 1;7(5):411–23.

115. Andreu P, Johansson M, Affara NI, Pucci F, Tan T, Junankar S, et al. FcRgamma activation regulates inflammation-associated squamous carcinogenesis. Cancer Cell. 2010 Feb 17;17(2):121–34.

116. Affara NI, Ruffell B, Medler TR, Gunderson AJ, Johansson M, Bornstein S, et al. B Cells Regulate Macrophage Phenotype and Response to Chemotherapy in Squamous Carcinomas. Cancer Cell. 2014 Jun;25(6):809–21.

117. Schioppa T, Moore R, Thompson RG, Rosser EC, Kulbe H, Nedospasov S, et al. B regulatory cells and the tumor-promoting actions of TNF-α during squamous carcinogenesis. :6.

118. DiLillo DJ, Yanaba K, Tedder TF. B Cells Are Required for Optimal CD4+ and CD8+ T Cell Tumor Immunity: Therapeutic B Cell Depletion Enhances B16 Melanoma Growth in Mice. J Immunol. 2010 Apr 1;184(7):4006–16.

119. Sorrentino R, Morello S, Forte G, Montinaro A, De Vita G, Luciano A, et al. B cells contribute to the antitumor activity of CpG-oligodeoxynucleotide in a mouse model of metastatic lung carcinoma. Am J Respir Crit Care Med. 2011 May 15;183(10):1369–79.

120. Bodogai M, Moritoh K, Lee-Chang C, Hollander CM, Sherman-Baust CA, Wersto RP, et al. Immunosuppressive and Prometastatic Functions of Myeloid-Derived Suppressive Cells Rely upon Education from Tumor-Associated B Cells. Cancer Res. 2015 Sep 1;75(17):3456–65.

121. Tao H, Lu L, Xia Y, Dai F, Wang Y, Bao Y, et al. Antitumor effector B cells directly kill tumor cells via the Fas/FasL pathway and are regulated by IL-10. Eur J Immunol. 2015 Apr;45(4):999–1009.

122. Li Q, Lao X, Pan Q, Ning N, Yet J, Xu Y, et al. Adoptive transfer of tumor reactive B cells confers host T cell immunity and tumor regression. Clin Cancer Res Off J Am Assoc Cancer Res. 2011 Aug 1;17(15):4987–95.

123. Ribas A, Medina T, Kummar S, Amin A, Kalbasi A, Drabick JJ, et al. SD-101 in Combination with Pembrolizumab in Advanced Melanoma: Results of a Phase Ib, Multicenter Study. Cancer Discov. 2018 Oct;8(10):1250–7.

124. Weihrauch MR, Richly H, von Bergwelt-Baildon MS, Becker HJ, Schmidt M, Hacker UT, et al. Phase I clinical study of the toll-like receptor 9 agonist MGN1703 in patients with metastatic solid tumours. Eur J Cancer. 2015 Jan 1;51(2):146–56.

125. Wouters MCA, Nelson BH. Prognostic Significance of Tumor-Infiltrating B Cells and Plasma Cells in Human Cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 2018 Dec 15;24(24):6125–35.

126. Nelson BH. CD20+ B Cells: The Other Tumor-Infiltrating Lymphocytes. J Immunol. 2010 Nov;185(9):4977–82.

127. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity. 2013 Oct 17;39(4):782–95.

128. Hennequin A, Derangère V, Boidot R, Apetoh L, Vincent J, Orry D, et al. Tumor infiltration by Tbet+ effector T cells and CD20+ B cells is associated with survival in gastric cancer patients. Oncoimmunology. 2016 Feb;5(2):e1054598.

129. Nielsen JS, Sahota RA, Milne K, Kost SE, Nesslinger NJ, Watson PH, et al. CD20+ Tumor-Infiltrating Lymphocytes Have an Atypical CD27- Memory Phenotype and Together with CD8+ T Cells Promote Favorable Prognosis in Ovarian Cancer. Clin Cancer Res. 2012 Jun 15;18(12):3281–92.

130. Brunner SM, Itzel T, Rubner C, Kesselring R, Griesshammer E, Evert M, et al. Tumor-infiltrating B cells producing antitumor active immunoglobulins in resected HCC prolong patient survival. Oncotarget. 2017 Sep 19;8(41):71002–11.

131. Cabrita R, Lauss M, Sanna A, Donia M, Larsen MS, Mitra S, et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature. 2020 Jan 15;1–5.

132. Da Gama Duarte J, Peyper JM, Blackburn JM. B cells and antibody production in melanoma. Mamm Genome Off J Int Mamm Genome Soc. 2018 Dec;29(11–12):790–805.

133. Eerola A, Soini Y, Paakko P. Tumour infiltrating lymphocytes in relation to tumour angiogenesis, apoptosis and prognosis in patients with large cell lung carcinoma. Lung Cancer. 1999 Nov;26(2):73–83.

134. Eerola AK, Soini Y, Pääkkö P. A high number of tumor-infiltrating lymphocytes are associated with a small tumor size, low tumor stage, and a favorable prognosis in operated small cell lung carcinoma. Clin Cancer Res Off J Am Assoc Cancer Res. 2000 May;6(5):1875–81.

135. Hernández-Prieto S, Romera A, Ferrer M, Subiza JL, López-Asenjo JA, Jarabo JR, et al. A 50-gene signature is a novel scoring system for tumor-infiltrating immune cells with strong correlation with clinical outcome of stage I/II non-small cell lung cancer. Clin Transl Oncol. 2015 Apr;17(4):330–8.

136. Schmidt M, Hellwig B, Hammad S, Othman A, Lohr M, Chen Z, et al. A comprehensive analysis of human gene expression profiles identifies stromal immunoglobulin κ C as a compatible prognostic marker in human solid tumors. Clin Cancer Res Off J Am Assoc Cancer Res. 2012 May 1;18(9):2695–703.

137. Germain C, Gnjatic S, Tamzalit F, Knockaert S, Remark R, Goc J, et al. Presence of B cells in tertiary lymphoid structures is associated with a protective immunity in patients with lung cancer. Am J Respir Crit Care Med. 2014 Apr 1;189(7):832–44.

138. Kinoshita T, Muramatsu R, Fujita T, Nagumo H, Sakurai T, Noji S, et al. Prognostic value of tumor-infiltrating lymphocytes differs depending on histological type and smoking habit in completely resected non-small-cell lung cancer. Ann Oncol. 2016 Nov;27(11):2117–23.

139. Al-Shibli KI, Donnem T, Al-Saad S, Persson M, Bremnes RM, Busund LT. Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 2008 Aug 15;14(16):5220–7.

140. Pelletier MP, Edwardes MD, Michel RP, Halwani F, Morin JE. Prognostic markers in resectable non-small cell lung cancer: a multivariate analysis. Can J Surg J Can Chir. 2001 Jun;44(3):180–8.

141. Schalper KA, Brown J, Carvajal-Hausdorf D, McLaughlin J, Velcheti V, Syrigos KN, et al. Objective Measurement and Clinical Significance of TILs in Non–Small Cell Lung Cancer. JNCI J Natl Cancer Inst [Internet]. 2015 Mar [cited 2018 Aug 21];107(3). Available from: https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/dju435

142. Griss J, Bauer W, Wagner C, Simon M, Chen M, Grabmeier-Pfistershammer K, et al. B cells sustain inflammation and predict response to immune checkpoint blockade in human melanoma. Nat Commun. 2019 Dec;10(1):4186.

143. Tarella C, Passera R, Magni M, Benedetti F, Rossi A, Gueli A, et al. Risk Factors for the Development of Secondary Malignancy After High-Dose Chemotherapy and Autograft, With or Without Rituximab: A 20-Year Retrospective Follow-Up Study in Patients With Lymphoma. J Clin Oncol [Internet]. 2010 Dec 28 [cited 2019 Oct 9]; Available from: https://ascopubs.org/doi/pdf/10.1200/JCO.2010.28.9777

144. Kurebayashi Y, Emoto K, Hayashi Y, Kamiyama I, Ohtsuka T, Asamura H, et al. Comprehensive Immune Profiling of Lung Adenocarcinomas Reveals Four Immunosubtypes with Plasma Cell Subtype a Negative Indicator. Cancer Immunol Res. 2016 Mar;4(3):234–47.

145. Fujimoto M, Yoshizawa A, Sumiyoshi S, Sonobe M, Kobayashi M, Koyanagi I, et al. Stromal plasma cells expressing immunoglobulin G4 subclass in non-small cell lung cancer. Hum Pathol. 2013 Aug;44(8):1569–76.

146. Lohr M, Edlund K, Botling J, Hammad S, Hellwig B, Othman A, et al. The prognostic relevance of tumour-infiltrating plasma cells and immunoglobulin kappa C indicates an important role of the humoral immune response in non-small cell lung cancer. Cancer Lett. 2013 Jun 10;333(2):222–8.

147. Crescioli S, Correa I, Karagiannis P, Davies AM, Sutton BJ, Nestle FO, et al. IgG4 Characteristics and Functions in Cancer Immunity. Curr Allergy Asthma Rep. 2016;16:7.

148. Karagiannis P, Gilbert AE, Josephs DH, Ali N, Dodev T, Saul L, et al. IgG4 subclass antibodies impair antitumor immunity in melanoma. J Clin Invest. 2013 Apr;123(4):1457–74.

149. Wang H, Xu Q, Zhao C, Zhu Z, Zhu X, Zhou J, et al. An immune evasion mechanism with IgG4 playing an essential role in cancer and implication for immunotherapy. J Immunother Cancer. 2020 Aug 1;8(2):e000661.

150. Hald SM, Bremnes RM, Al-Shibli K, Al-Saad S, Andersen S, Stenvold H, et al. CD4/CD8 co-expression shows independent prognostic impact in resected non-small cell lung cancer patients treated with adjuvant radiotherapy. Lung Cancer. 2013 May;80(2):209–15.

151. Suzuki K, Kadota K, Sima CS, Nitadori J ichi, Rusch VW, Travis WD, et al. Clinical Impact of Immune Microenvironment in Stage I Lung Adenocarcinoma: Tumor Interleukin-12 Receptor β2 (IL-12Rβ2), IL-7R, and Stromal FoxP3/CD3 Ratio Are Independent Predictors of Recurrence. J Clin Oncol. 2013 Feb;31(4):490–8.

152. Banat GA, Tretyn A, Pullamsetti SS, Wilhelm J, Weigert A, Olesch C, et al. Immune and Inflammatory Cell Composition of Human Lung Cancer Stroma. PloS One. 2015;10(9):e0139073.

153. Bruno TC, Ebner PJ, Moore BL, Squalls OG, Waugh KA, Eruslanov EB, et al. Antigen-Presenting Intratumoral B Cells Affect CD4 $^{+}$ TIL Phenotypes in Non–Small Cell Lung Cancer Patients. Cancer Immunol Res. 2017 Oct;5(10):898–907.

154. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med. 2015 Aug;21(8):938–45.

155. Sautès-Fridman C, Petitprez F, Calderaro J, Fridman WH. Tertiary lymphoid structures in the era of cancer immunotherapy. Nat Rev Cancer. 2019 Jun;19(6):307–25.

156. Sharonov GV, Serebrovskaya EO, Yuzhakova DV, Britanova OV, Chudakov DM. B cells, plasma cells and antibody repertoires in the tumour microenvironment. Nat Rev Immunol. 2020 May;20(5):294–307.

157. Helmink BA, Reddy SM, Gao J, Zhang S, Basar R, Thakur R, et al. B cells and tertiary lymphoid structures promote immunotherapy response. Nature. 2020 Jan 15;1–7.

158. Mohr E, Serre K, Manz RA, Cunningham AF, Khan M, Hardie DL, et al. Dendritic cells and monocyte/macrophages that create the IL-6/APRIL-rich lymph node microenvironments where plasmablasts mature. J Immunol Baltim Md 1950. 2009 Feb 15;182(4):2113–23.

159. MacLennan ICM, Toellner KM, Cunningham AF, Serre K, Sze DMY, Zúñiga E, et al. Extrafollicular antibody responses. Immunol Rev. 2003 Aug;194:8–18.

160. Pinto D, Montani E, Bolli M, Garavaglia G, Sallusto F, Lanzavecchia A, et al. A functional BCR in human IgA and IgM plasma cells. Blood. 2013 May 16;121(20):4110–4.

161. Zhou J, Min Z, Zhang D, Wang W, Marincola F, Wang X. Enhanced frequency and potential mechanism of B regulatory cells in patients with lung cancer. J Transl Med. 2014 Nov 11;12:304.

162. Liu J, Wang H, Yu Q, Zheng S, Jiang Y, Liu Y, et al. Aberrant frequency of IL-10-producing B cells and its association with Treg and MDSC cells in Non Small Cell Lung Carcinoma patients. Hum Immunol. 2016 Jan;77(1):84–9.

163. Fillatreau S. Regulatory plasma cells. Curr Opin Pharmacol. 2015 Aug;23:1–5.

164. Blair PA, Noreña LY, Flores-Borja F, Rawlings DJ, Isenberg DA, Ehrenstein MR, et al. CD19+CD24hiCD38hi B Cells Exhibit Regulatory Capacity in Healthy Individuals but Are Functionally Impaired in Systemic Lupus Erythematosus Patients. Immunity. 2010 Jan;32(1):129–40.

165. Bosma A, Abdel-Gadir A, Isenberg DA, Jury EC, Mauri C. Lipid-antigen presentation by CD1d(+) B cells is essential for the maintenance of invariant natural killer T cells. Immunity. 2012 Mar 23;36(3):477–90.

166. Shao Y, Lo CM, Ling CC, Liu XB, Ng KTP, Chu ACY, et al. Regulatory B cells accelerate hepatocellular carcinoma progression via CD40/CD154 signaling pathway. Cancer Lett. 2014 Dec 28;355(2):264–72.

167. Olkhanud PB, Baatar D, Bodogai M, Hakim F, Gress R, Anderson RL, et al. Breast cancer lung metastasis requires expression of chemokine receptor CCR4 and regulatory T cells. Cancer Res. 2009 Jul 15;69(14):5996–6004.

168. Wang WW, Yuan XL, Chen H, Xie GH, Ma YH, Zheng YX, et al. CD19+CD24hiCD38hiBregs involved in downregulate helper T cells and upregulate regulatory T cells in gastric cancer. Oncotarget. 2015 Oct 20;6(32):33486–99.

169. Qian L, Bian GR, Zhou Y, Wang Y, Hu J, Liu X, et al. Clinical significance of regulatory B cells in the peripheral blood of patients with oesophageal cancer. Cent-Eur J Immunol. 2015;40(2):263–5.

170. Wei X, Jin Y, Tian Y, Zhang H, Wu J, Lu W, et al. Regulatory B cells contribute to the impaired antitumor immunity in ovarian cancer patients. Tumour Biol J Int Soc Oncodevelopmental Biol Med. 2016 May;37(5):6581–8.

171. Zhang C, Xin H, Zhang W, Yazaki PJ, Zhang Z, Le K, et al. CD5 Binds to Interleukin-6 and Induces a Feed-Forward Loop with the Transcription Factor STAT3 in B Cells to Promote Cancer. Immunity. 2016 Apr;44(4):913–23.

172. Lee-Chang C, Bodogai M, Martin-Montalvo A, Wejksza K, Sanghvi M, Moaddel R, et al. Inhibition of breast cancer metastasis by resveratrol-mediated inactivation of tumor-evoked regulatory B cells. J Immunol Baltim Md 1950. 2013 Oct 15;191(8):4141–51.

173. Yanaba K, Bouaziz JD, Haas KM, Poe JC, Fujimoto M, Tedder TF. A regulatory B cell subset with a unique CD1dhiCD5+ phenotype controls T cell-dependent inflammatory responses. Immunity. 2008 May;28(5):639–50.

174. Mizoguchi A, Mizoguchi E, Takedatsu H, Blumberg RS, Bhan AK. Chronic intestinal inflammatory condition generates IL-10-producing regulatory B cell subset characterized by CD1d upregulation. Immunity. 2002 Feb;16(2):219–30.

175. Wang HM, Zhang XH, Feng MM, Qiao YJ, Ye LQ, Chen J, et al. Interleukin-35 Suppresses the Antitumor Activity of T Cells in Patients with Non-Small Cell Lung Cancer. Cell Physiol Biochem. 2018;47(6):2407–19.

176. Pylayeva-Gupta Y, Das S, Handler JS, Hajdu CH, Coffre M, Koralov SB, et al. IL35-Producing B Cells Promote the Development of Pancreatic Neoplasia. Cancer Discov. 2016 Mar 1;6(3):247–55.

177. Iwata Y, Matsushita T, Horikawa M, Dilillo DJ, Yanaba K, Venturi GM, et al. Characterization of a rare IL-10-competent B-cell subset in humans that parallels mouse regulatory B10 cells. Blood. 2011 Jan 13;117(2):530–41.

178. Lindner S, Dahlke K, Sontheimer K, Hagn M, Kaltenmeier C, Barth TFE, et al. Interleukin 21-induced granzyme B-expressing B cells infiltrate tumors and regulate T cells. Cancer Res. 2013 Apr 15;73(8):2468–79.

179. van de Veen W, Stanic B, Yaman G, Wawrzyniak M, Söllner S, Akdis DG, et al. IgG4 production is confined to human IL-10–producing regulatory B cells that suppress antigen-specific immune responses. J Allergy Clin Immunol. 2013 Apr;131(4):1204–12.

180. Matsumoto M, Baba A, Yokota T, Nishikawa H, Ohkawa Y, Kayama H, et al. Interleukin-10-Producing Plasmablasts Exert Regulatory Function in Autoimmune Inflammation. Immunity. 2014 Dec;41(6):1040–51.

181. Saze Z, Schuler PJ, Hong CS, Cheng D, Jackson EK, Whiteside TL. Adenosine production by human B cells and B cell-mediated suppression of activated T cells. Blood. 2013 Jul 4;122(1):9–18.

182. Liu J, Zhan W, Kim CJ, Clayton K, Zhao H, Lee E, et al. IL-10-producing B cells are induced early in HIV-1 infection and suppress HIV-1-specific T cell responses. PloS One. 2014;9(2):e89236.

183. Nouël A, Pochard P, Simon Q, Ségalen I, Le Meur Y, Pers JO, et al. B-Cells induce regulatory T cells through TGF-β/IDO production in A CTLA-4 dependent manner. J Autoimmun. 2015 May;59:53–60.

184. Zhang Y, Morgan R, Chen C, Cai Y, Clark E, Khan WN, et al. Mammary-tumor-educated B cells acquire LAP/TGF-β and PD-L1 expression and suppress anti-tumor immune responses. Int Immunol. 2016 Sep;28(9):423–33.

185. Yang C, Lee H, Pal S, Jove V, Deng J, Zhang W, et al. B Cells Promote Tumor Progression via STAT3 Regulated-Angiogenesis. Viola JPB, editor. PLoS ONE. 2013 May 29;8(5):e64159.

186. Sarvaria A, Madrigal JA, Saudemont A. B cell regulation in cancer and anti-tumor immunity. Cell Mol Immunol. 2017 Aug;14(8):662–74.

187. Neyt K, Perros F, GeurtsvanKessel CH, Hammad H, Lambrecht BN. Tertiary lymphoid organs in infection and autoimmunity. Trends Immunol. 2012 Jun;33(6):297–305.

188. Litsiou E, Semitekolou M, Galani IE, Morianos I, Tsoutsa A, Kara P, et al. CXCL13 Production in B Cells via Toll-like Receptor/Lymphotoxin Receptor Signaling Is Involved in Lymphoid Neogenesis in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2013 Mar 22;187(11):1194–202.

189. Sautès-Fridman C, Cherfils-Vicini J, Damotte D, Fisson S, Fridman WH, Cremer I, et al. Tumor microenvironment is multifaceted. Cancer Metastasis Rev. 2011 Mar;30(1):13–25.

190. Wang Z, Cheng Q, Tang K, Sun Y, Zhang K, Zhang Y, et al. Lipid mediator lipoxin A4 inhibits tumor growth by targeting IL-10-producing regulatory B (Breg) cells. Cancer Lett. 2015 Aug 10;364(2):118–24.

191. Damsky W, Jilaveanu L, Turner N, Perry C, Zito C, Tomayko M, et al. B cell depletion or absence does not impede anti-tumor activity of PD-1 inhibitors. J Immunother Cancer. 2019 Dec;7(1):153.

192. Petitprez F, Reyniès A de, Keung EZ, Chen TWW, Sun CM, Calderaro J, et al. B cells are associated with survival and immunotherapy response in sarcoma. Nature. 2020 Jan 15;1–5.

193. Yuan J, Adamow M, Ginsberg BA, Rasalan TS, Ritter E, Gallardo HF, et al. Integrated NY-ESO-1 antibody and CD8+ T-cell responses correlate with clinical benefit in advanced melanoma patients treated with ipilimumab. Proc Natl Acad Sci U S A. 2011 Oct 4;108(40):16723–8.

194. Varn FS, Wang Y, Cheng C. A B cell-derived gene expression signature associates with an immunologically active tumor microenvironment and response to immune checkpoint blockade therapy. OncoImmunology. 2019 Jan 2;8(1):e1513440.

195. Zhang M, Xia L, Yang Y, Liu S, Ji P, Wang S, et al. PD-1 blockade augments humoral immunity through ICOS-mediated CD4+ T cell instruction. Int Immunopharmacol. 2019 Jan 1;66:127–38.

196. Achour A, Simon Q, Mohr A, Séité JF, Youinou P, Bendaoud B, et al. Human regulatory B cells control the TFH cell response. J Allergy Clin Immunol. 2017 Jul 1;140(1):215–22.

197. Khan AR, Hams E, Floudas A, Sparwasser T, Weaver CT, Fallon PG. PD-L1hi B cells are critical regulators of humoral immunity. Nat Commun. 2015 May;6(1):5997.

198. Schildberg FA, Klein SR, Freeman GJ, Sharpe AH. Coinhibitory pathways in the B7-CD28 ligand-receptor family. Immunity. 2016 May 17;44(5):955–72.

199. Thibult ML, Mamessier E, Gertner-Dardenne J, Pastor S, Just-Landi S, Xerri L, et al. PD-1 is a novel regulator of human B-cell activation. Int Immunol. 2013 Feb;25(2):129–37.

200. Sowerby L, Dewan AK, Granter S, Gandhi L, LeBoeuf NR. Rituximab Treatment of Nivolumab-Induced Bullous Pemphigoid. JAMA Dermatol. 2017 Jun 1;153(6):603–5.

201. Shaikh H, Daboul N, Albrethsen M, Fazal S. A case of autoimmune haemolytic anaemia after 39 cycles of nivolumab. BMJ Case Rep. 2018 Apr 18;2018.

202. Ito M, Fujiwara S, Fujimoto D, Mori R, Yoshimura H, Hata A, et al. Rituximab for nivolumab plus ipilimumab-induced encephalitis in a small-cell lung cancer patient. Ann Oncol. 2017 Sep 1;28(9):2318–9.

203. Shiuan E, Beckermann KE, Ozgun A, Kelly C, McKean M, McQuade J, et al. Thrombocytopenia in patients with melanoma receiving immune checkpoint inhibitor therapy. J Immunother Cancer. 2017;5:8.

204. Hasanov M, Konoplev SN, Hernandez CMR. Nivolumab-induced cold agglutinin syndrome successfully treated with rituximab. Blood Adv. 2018 Aug 2;2(15):1865–8.

205. Mauri C, Menon M. Human regulatory B cells in health and disease: therapeutic potential [Internet]. 2017 [cited 2019 Nov 4]. Available from: https://www.jci.org/articles/view/85113/pdf

206. Vansteenkiste JF, Cho BC, Vanakesa T, De Pas T, Zielinski M, Kim MS, et al. Efficacy of the MAGE-A3 cancer immunotherapeutic as adjuvant therapy in patients with resected MAGE-A3-positive non-small-cell lung cancer (MAGRIT): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol. 2016 Jun;17(6):822–35.

207. Foster AD, Sivarapatna A, Gress RE. The aging immune system and its relationship with cancer. Aging Health. 2011 Oct 1;7(5):707–18.

208. Ferrara R, Mezquita L, Auclin E, Chaput N, Besse B. Immunosenescence and immunecheckpoint inhibitors in non-small cell lung cancer patients: Does age really matter? Cancer Treat Rev. 2017 Nov 1;60:60–8.

209. Cuppens K, Oyen C, Derweduwen A, Ottevaere A, Sermeus W, Vansteenkiste J. Characteristics and outcome of unplanned hospital admissions in patients with lung cancer: a longitudinal tertiary center study. Towards a strategy to reduce the burden. Support Care Cancer [Internet]. 2016 Jan 27 [cited 2018 May 14]; Available from: http://link.springer.com/10.1007/s00520-016-3087-4

210. Chiou WY, Hung SK, Lai CL, Lin HY, Su YC, Chen YC, et al. Effect of 23-Valent Pneumococcal Polysaccharide Vaccine Inoculated During Anti-Cancer Treatment Period in Elderly Lung Cancer Patients on Community-Acquired Pneumonia Hospitalization: A Nationwide Population-Based Cohort Study. Medicine (Baltimore). 2015 Jul;94(26):e1022.

211. Bandura DR, Baranov VI, Ornatsky OI, Antonov A, Kinach R, Lou X, et al. Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. Anal Chem. 2009 Aug 15;81(16):6813–22.

212. Olsen LR, Leipold MD, Pedersen CB, Maecker HT. The anatomy of single cell mass cytometry data. Cytom Part J Int Soc Anal Cytol. 2019 Feb;95(2):156–72.

213. Stern AD, Rahman AH, Birtwistle MR. Cell size assays for mass cytometry. Cytom Part J Int Soc Anal Cytol. 2017 Jan;91(1):14–24.

214. Quatromoni JG, Eruslanov E. Tumor-associated macrophages: function, phenotype, and link to prognosis in human lung cancer. Am J Transl Res. 2012;4(4):376–89.

215. Leipold MD, Ornatsky O, Baranov V, Whitfield C, Nitz M. Development of mass cytometry methods for bacterial discrimination. Anal Biochem. 2011 Dec 1;419(1):1–8.

216. Edgar LJ, Vellanki RN, Halupa A, Hedley D, Wouters BG, Nitz M. Identification of hypoxic cells using an organotellurium tag compatible with mass cytometry. Angew Chem Int Ed Engl. 2014 Oct 20;53(43):11473–7.

217. Behbehani GK, Bendall SC, Clutter MR, Fantl WJ, Nolan GP. Single-cell mass cytometry adapted to measurements of the cell cycle. Cytom Part J Int Soc Anal Cytol. 2012 Jul;81(7):552–66.

218. Ornatsky OI, Kinach R, Bandura DR, Lou X, Tanner SD, Baranov VI, et al. Development of analytical methods for multiplex bio-assay with inductively coupled plasma mass spectrometry. J Anal At Spectrom. 2008;23(4):463–9.

219. Finck R, Simonds EF, Jager A, Krishnaswamy S, Sachs K, Fantl W, et al. Normalization of mass cytometry data with bead standards. Cytom Part J Int Soc Anal Cytol. 2013 May;83(5):483–94.

220. Kleinsteuber K, Corleis B, Rashidi N, Nchinda N, Lisanti A, Cho JL, et al. Standardization and quality control for high-dimensional mass cytometry studies of human samples. Cytom Part J Int Soc Anal Cytol. 2016 Oct;89(10):903–13.

221. Mihaylova M. QuantSeq FWD for Illumina Publications [Internet]. Lexogen. [cited 2020 May 29]. Available from: https://www.lexogen.com/publications/quantseq-publications/

222. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

223. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011 May 2;17(1):10–2.

224. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015 Apr;12(4):357–60.

225. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012 Aug 15;28(16):2184–5.

226. Dobin A. alexdobin/STAR [Internet]. 2021 [cited 2021 Mar 23]. Available from: https://github.com/alexdobin/STAR

227. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013 May;41(10):e108.

228. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015 Jan 15;31(2):166–9.

229. Rathinam S, Ward DG, James ND, Rajesh PB. Proteomic analysis of resectable non-small cell lung cancer: post-resection serum samples may be useful in identifying potential markers. Interact Cardiovasc Thorac Surg. 2011 Jul;13(1):3–6.

230. Sumera A, Anuar ND, Radhakrishnan AK, Ibrahim H, Rutt NH, Ismail NH, et al. A Novel Method to Identify Autoantibodies against Putative Target AAbs in Serum from beta-Thalassemia Major: A Pilot Study. Biomedicines. 2020 Apr 26;8(5).

231. Duarte J, Serufuri JM, Mulder N, Blackburn J. AAb Function Microarrays: Design, Use and Bioinformatic Analysis in Cancer AAb Discovery and Quantitation. In: Wang X, editor. Bioinformatics of Human Proteomics [Internet]. Dordrecht: Springer Netherlands; 2013 [cited 2021 Mar 6]. p. 39–74. (Translational Bioinformatics). Available from: https://doi.org/10.1007/978-94-007-5811-7_3

232. Mak A, Kow NY, Ismail NH, Anuar ND, Rutt NH, Cho J, et al. Detection of putative autoantibodies in systemic lupus erythematous using a novel native-conformation AAb microarray platform. Lupus. 2020 Dec;29(14):1948–54.

233. ParkerICI/premessa [Internet]. Parker Institute for Cancer Immunotherapy; 2019 [cited 2019 Nov 19]. Available from: https://github.com/ParkerICI/premessa

234. ParkerICI/premessa [Internet]. Parker Institute for Cancer Immunotherapy; 2020 [cited 2021 Mar 17]. Available from: https://github.com/ParkerICI/premessa

235. Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, Sarkar D, et al. flowCore: a Bioconductor package for high throughput flow cytometry. BMC Bioinformatics. 2009 Apr 9;10:106.

236. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytom Part J Int Soc Anal Cytol. 2015 Jul;87(7):636–45.

237. Nowicka M, Krieg C, Crowell HL, Weber LM, Hartmann FJ, Guglietta S, et al. CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets. F1000Research. 2017;6:748.

238. Chevrier S, Crowell HL, Zanotelli VRT, Engler S, Robinson MD, Bodenmiller B. Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. Cell Syst [Internet]. 2018 Mar [cited 2018 Apr 26]; Available from: http://linkinghub.elsevier.com/retrieve/pii/S2405471218300632

239. Weber LM, Nowicka M, Soneson C, Robinson MD. diffcyt: Differential discovery in high-dimensional cytometry via high-resolution clustering [Internet]. Bioinformatics; 2018 Jun [cited 2019 Oct 22]. Available from: http://biorxiv.org/lookup/doi/10.1101/349738

240. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018 Dec 3;

241. Zhao K, Lu Z xiang, Park JW, Zhou Q, Xing Y. GLiMMPS: robust statistical model for regulatory variation of alternative splicing using RNA-seq data. Genome Biol. 2013 Jul 22;14(7):R74.

242. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec;15(12):550.

243. Love MI, Soneson C, Patro R. Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification. F1000Research. 2018 Oct 1;7:952.

244. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research. 2015 Dec 30;4:1521.

245. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013 Jan;31(1):46–53.

246. Pagès H, Carlson M, Falcon S, Li N. AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor [Internet]. Bioconductor version: Release (3.12); 2021 [cited 2021 Mar 17]. Available from: https://bioconductor.org/packages/AnnotationDbi/

247. Durinck S, Huber W, Davis S, Pepin F, Buffalo VS, Smith M. biomaRt: Interface to BioMart databases (i.e. Ensembl) [Internet]. Bioconductor version: Release (3.12); 2021 [cited 2021 Mar 17]. Available from: https://bioconductor.org/packages/biomaRt/

248. ComplexHeatmap Complete Reference [Internet]. [cited 2021 Mar 17]. Available from: https://jokergoo.github.io/ComplexHeatmap-reference/book/

249. EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling [Internet]. [cited 2021 Mar 17]. Available from: https://bioconductor.org/packages/release/bioc/vignettes/EnhancedVolcano/inst/doc/EnhancedVolcano.html

250. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics. 2009 May 27;10(1):161.

251. Luo W. pathview: a tool set for pathway based data integration and visualization [Internet]. Bioconductor version: Release (3.12); 2021 [cited 2021 Mar 17]. Available from: https://bioconductor.org/packages/pathview/

252. Yu G, Wang LG, Dall'Olio G. clusterProfiler: statistical analysis and visualization of functional profiles for genes and gene clusters [Internet]. Bioconductor version: Release (3.12); 2021 [cited 2021 Mar 18]. Available from: https://bioconductor.org/packages/clusterProfiler/

253. ggplot2 Based Publication Ready Plots • ggpubr [Internet]. [cited 2021 Mar 29]. Available from: https://rpkgs.datanovia.com/ggpubr/

254. Bommert A, Sun X, Bischl B, Rahnenführer J, Lang M. Benchmark for filter methods for feature selection in high-dimensional classification data. Comput Stat Data Anal. 2020 Mar;143.

255. Wright MN, Ziegler A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. J Stat Softw. 2017;77(1).

256. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Ser B Methodol. 1996;58(1):267–88.

257. Multimodel Inference: Understanding AIC and BIC in Model Selection - Kenneth P. Burnham, David R. Anderson, 2004 [Internet]. [cited 2021 Mar 6]. Available from: https://journals.sagepub.com/doi/10.1177/0049124104268644

258. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In: Parzen E, Tanabe K, Kitagawa G, editors. Selected Papers of Hirotugu Akaike [Internet]. New York, NY: Springer; 1998 [cited 2021 Mar 6]. p. 199–213. (Springer Series in Statistics). Available from: https://doi.org/10.1007/978-1-4612-1694-0_15

259. Cooper JD, Han SYS, Tomasik J, Ozcan S, Rustogi N, van Beveren NJM, et al. Multimodel inference for AAb development: an application to schizophrenia. Transl Psychiatry. 2019 Feb 11;9(1):1–10.

260. López-Ratón M, Rodríguez-Álvarez MX, Suárez CC, Sampedro FG. **OptimalCutpoints** : An *R* Package for Selecting Optimal Cutpoints in Diagnostic Tests. J Stat Softw. 2014;61(8).

261. Wickham H. ggplot2. Cham: Springer International Publishing; 2016.

262. Kassambara A, Kosinski M, Biecek P, others. survminer: Drawing Survival Curves using'ggplot2'. R Package Version 03. 2017;1.

263. Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, et al. Mlr: Machine Learning in R. J Mach Learn Res. 2016 Jan;17(1):5938–5942.

264. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformatics. 2008 Dec 11;9(1).

265. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.

266. Kursa MB. praznik: Tools for Information-Based Feature Selection. 2020.

267. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28(5):1–26.

268. Qiu W. Sample Size and Power Calculation in Microarray Studies Using the sizepower package. :8.

269. Scherer R. shearer/samplesize [Internet]. 2019 [cited 2021 Mar 6]. Available from: https://github.com/shearer/samplesize

270. Presentation-Ready Data Summary and Analytic Result Tables • gtsummary [Internet]. [cited 2021 Mar 23]. Available from: http://www.danieldsjoberg.com/gtsummary/

271. Lackey A, Donington JS. Surgical Management of Lung Cancer. Semin Interv Radiol. 2013 Jun;30(2):133–40.

272. Lang-Lazdunski L. Surgery for nonsmall cell lung cancer. Eur Respir Rev. 2013 Sep 1;22(129):382–404.

273. Uramoto H, Tanaka F. Recurrence after surgery in patients with NSCLC. Transl Lung Cancer Res. 2014 Aug;3(4):242–9.

274. al-Kattan K, Sepsas E, Fountain SW, Townsend ER. Disease recurrence after resection for stage I lung cancer. Eur J Cardio-Thorac Surg Off J Eur Assoc Cardio-Thorac Surg. 1997 Sep;12(3):380–4.

275. Nedergaard BS, Ladekarl M, Nyengaard JR, Nielsen K. A comparative study of the cellular immune response in patients with stage IB cervical squamous cell carcinoma. Low numbers of several immune cell subtypes are strongly associated with relapse of disease within 5 years. Gynecol Oncol. 2008 Jan;108(1):106–11.

276. Nowicka M, Krieg C, Crowell HL, Weber LM, Hartmann FJ, Guglietta S, et al. CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets. F1000Research [Internet]. 2019 May 24 [cited 2019 Oct 18];6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5473464/

277. Lavin Y, Kobayashi S, Leader A, Amir EAD, Elefant N, Bigenwald C, et al. Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. Cell. 2017 May 4;169(4):750-765.e17.

278. Jiang Y, Li C, Wu Q, An P, Huang L, Wang J, et al. Iron-dependent histone 3 lysine 9 demethylation controls B cell proliferation and humoral immune responses. Nat Commun. 2019 Jul 3;10(1):2935.

279. Rousseaux S, Debernardi A, Jacquiau B, Vitte AL, Vesin A, Nagy-Mignotte H, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. Sci Transl Med. 2013 May 22;5(186):186ra66.

280. Hao D, Han G, Sinjab A, Gomez-Bolanos LI, Lazcano R, Serrano A, et al. The Single-Cell Immunogenomic Landscape of B and Plasma Cells in Early-Stage Lung Adenocarcinoma. Cancer Discov. 2022 Sep 13;CD-21-1658.

281. Backman M, La Fleur L, Kurppa P, Djureinovic D, Elfving H, Brunnström H, et al. Infiltration of NK and plasma cells is associated with a distinct immune subset in non-small cell lung cancer. J Pathol. 2021;255(3):243–56.

282. Blanca IR, Bere EW, Young HA, Ortaldo JR. Human B cell activation by autologous NK cells is regulated by CD40-CD40 ligand interaction: role of memory B cells and CD5+ B cells. J Immunol Baltim Md 1950. 2001 Dec 1;167(11):6132–9.

283. Patil NS, Nabet BY, Müller S, Koeppen H, Zou W, Giltnane J, et al. Intratumoral plasma cells predict outcomes to PD-L1 blockade in non-small cell lung cancer. Cancer Cell. 2022 Mar 14;40(3):289-300.e4.

284. Leader AM, Grout JA, Maier BB, Nabet BY, Park MD, Tabachnikova A, et al. Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification. Cancer Cell. 2021 Dec 13;39(12):1594-1609.e12.

285. Teillaud JL, Dieu-Nosjean MC. Intratumoral plasma cells: More than a predictive marker of response to anti-PD-L1 treatment in lung cancer? Cancer Cell. 2022 Mar 14;40(3):240–3.

286. Shalapour S, Font-Burgada J, Di Caro G, Zhong Z, Sanchez-Lopez E, Dhar D, et al. Immunosuppressive plasma cells impede T-cell-dependent immunogenic chemotherapy. Nature. 2015 May 7;521(7550):94–8.

287. Shalapour S, Lin XJ, Bastian IN, Brain J, Burt AD, Aksenov AA, et al. Inflammation-induced IgA+ cells dismantle anti-liver cancer immunity. Nature. 2017 16;551(7680):340–5.

288. Kroeger DR, Milne K, Nelson BH. Tumor-Infiltrating Plasma Cells Are Associated with Tertiary Lymphoid Structures, Cytolytic T-Cell Responses, and Superior Prognosis in Ovarian Cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 2016 Jun 15;22(12):3005–15.

289. Maseda D, Smith SH, DiLillo DJ, Bryant JM, Candando KM, Weaver CT, et al. Regulatory B10 cells differentiate into antibody-secreting cells after transient IL-10 production in vivo. J Immunol Baltim Md 1950. 2012 Feb 1;188(3):1036–48.

290. Zhang B, Vogelzang A, Miyajima M, Sugiura Y, Wu Y, Chamoto K, et al. B cell-derived GABA elicits IL-10+ macrophages to limit anti-tumour immunity. Nature. 2021 Nov;599(7885):471–6.

291. Cullen MH, Billingham LJ, Woodroffe CM, Chetiyawardana AD, Gower NH, Joshi R, et al. Mitomycin, ifosfamide, and cisplatin in unresectable non-small-cell lung cancer: effects on survival and quality of life. J Clin Oncol Off J Am Soc Clin Oncol. 1999 Oct;17(10):3188–94.

292. Xing P, Zhang F, Wang G, Xu Y, Li C, Wang S, et al. Incidence rates of immune-related adverse events and their correlation with response in advanced solid tumours treated with NIVO or NIVO+IPI: a systematic review and meta-analysis. J Immunother Cancer. 2019 Dec 4;7(1):341.

293. Wang DY, Salem JE, Cohen JV, Chandra S, Menzer C, Ye F, et al. Fatal Toxic Effects Associated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis. JAMA Oncol. 2018 01;4(12):1721–8.

294. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009 Jan 1;45(2):228–47.

295. Xiao X, Lao XM, Chen MM, Liu RX, Wei Y, Ouyang FZ, et al. PD-1hi Identifies a Novel Regulatory B-cell Population in Human Hepatoma That Promotes Disease Progression. Cancer Discov. 2016;6(5):546–59.

296. Iwata Y, Matsushita T, Horikawa M, DiLillo DJ, Yanaba K, Venturi GM, et al. Characterization of a rare IL-10-competent B-cell subset in humans that parallels mouse regulatory B10 cells. Blood. 2011 Jan 13;117(2):530–41.

297. Simon Q, Pers JO, Cornec D, Le Pottier L, Mageed RA, Hillion S. In-depth characterization of CD24(high)CD38(high) transitional human B cells reveals different regulatory profiles. J Allergy Clin Immunol. 2016 May;137(5):1577-1584.e10.

298. Subrahmanyam PB, Dong Z, Gusenleitner D, Giobbie-Hurder A, Severgnini M, Zhou J, et al. Distinct predictive AAb candidates for response to anti-CTLA-4 and anti-PD-1 immunotherapy in melanoma patients. J Immunother Cancer. 2018 06;6(1):18.

299. van der Vlugt LEPM, Haeberlein S, de Graaf W, Martha TED, Smits HH. Toll-like receptor ligation for the induction of regulatory B cells. Methods Mol Biol Clifton NJ. 2014;1190:127–41.

300. Bibby JA, Purvis HA, Hayday T, Chandra A, Okkenhaug K, Rosenzweig S, et al. Cholesterol metabolism drives regulatory B cell IL-10 through provision of geranylgeranyl pyrophosphate. Nat Commun. 2020 Jul 8;11(1):3412.

301. Perucha E, Melchiotti R, Bibby JA, Wu W, Frederiksen KS, Roberts CA, et al. The cholesterol biosynthesis pathway regulates IL-10 expression in human Th1 cells. Nat Commun. 2019 Jan 30;10(1):498.

302. Cho JW, Hong MH, Ha SJ, Kim YJ, Cho BC, Lee I, et al. Genome-wide identification of differentially methylated promoters and enhancers associated with response to anti-PD-1 therapy in non-small cell lung cancer. Exp Mol Med. 2020 Sep;52(9):1550–63.

303. Auslander N, Zhang G, Lee JS, Frederick DT, Miao B, Moll T, et al. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. Nat Med. 2018 Oct;24(10):1545–9.

304. Patel AJ, Richter A, Drayson MT, Middleton GW. The role of B lymphocytes in the immuno-biology of non-small-cell lung cancer. Cancer Immunol Immunother CII. 2020 Mar;69(3):325–42.

305. Michaud D, Steward CR, Mirlekar B, Pylayeva-Gupta Y. Regulatory B cells in cancer. Immunol Rev. 2021 Jan;299(1):74–92.

306. Mauri C. Novel Frontiers in Regulatory B cells. Immunol Rev. 2021 Jan;299(1):5–9.

307. Mauri C, Blair PA. The incognito journey of a regulatory B cell. Immunity. 2014 Dec 18;41(6):878–80.

308. Wolf SD, Dittel BN, Hardardottir F, Janeway CA. Experimental autoimmune encephalomyelitis induction in genetically B cell-deficient mice. J Exp Med. 1996 Dec 1;184(6):2271–8.

309. Cherukuri A, Mohib K, Rothstein DM. Regulatory B cells: TIM-1, transplant tolerance, and rejection. Immunol Rev. 2021 Jan;299(1):31–44.

310. Woodruff MC, Ramonell RP, Nguyen DC, Cashman KS, Saini AS, Haddad NS, et al. Extrafollicular B cell responses correlate with neutralizing antibodies and morbidity in COVID-19. Nat Immunol. 2020 Dec;21(12):1506–16.

311. Menon M, Hussell T, Ali Shuwa H. Regulatory B cells in respiratory health and diseases. Immunol Rev. 2021 Jan;299(1):61–73.

312. Matsushita T. Regulatory and effector B cells: Friends or foes? J Dermatol Sci. 2019 Jan 1;93(1):2–7.

313. de Jonge K, Tillé L, Lourenco J, Maby-El Hajjami H, Nassiri S, Racle J, et al. Inflammatory B cells correlate with failure to checkpoint blockade in melanoma patients. Oncoimmunology. 2021 Feb 2;10(1):1873585.

314. Barr TA, Brown S, Ryan G, Zhao J, Gray D. TLR-mediated stimulation of APC: Distinct cytokine responses of B cells and dendritic cells. Eur J Immunol. 2007 Nov;37(11):3040–53.

315. Matsushita T, Yanaba K, Bouaziz JD, Fujimoto M, Tedder TF. Regulatory B cells inhibit EAE initiation in mice while other B cells promote disease progression. J Clin Invest. 2008 Oct;118(10):3420–30.

316. Yoshizaki A, Miyagaki T, DiLillo DJ, Matsushita T, Horikawa M, Kountikov EI, et al. Regulatory B cells control T-cell autoimmunity through IL-21-dependent cognate interactions. Nature. 2012 Nov 8;491(7423):264–8.

317. Yanaba K, Yoshizaki A, Asano Y, Kadono T, Tedder TF, Sato S. IL-10-producing regulatory B10 cells inhibit intestinal injury in a mouse model. Am J Pathol. 2011 Feb;178(2):735–43.

318. Maseda D, Candando KM, Smith SH, Kalampokis I, Weaver CT, Plevy SE, et al. Peritoneal cavity regulatory B cells (B10 cells) modulate IFN-γ+CD4+ T cell numbers during colitis development in mice. J Immunol Baltim Md 1950. 2013 Sep 1;191(5):2780–95.

319. Oka A, Ishihara S, Mishima Y, Tada Y, Kusunoki R, Fukuba N, et al. Role of regulatory B cells in chronic intestinal inflammation: association with pathogenesis of Crohn's disease. Inflamm Bowel Dis. 2014 Feb;20(2):315–28.

320. Wang X, Wang G, Wang Z, Liu B, Han N, Li J, et al. PD-1-expressing B cells suppress CD4+ and CD8+ T cells via PD-1/PD-L1-dependent pathway. Mol Immunol. 2019 May 1;109:20–6.

321. Mao Y, Wang Y, Dong L, Zhang Q, Wang C, Zhang Y, et al. Circulating exosomes from esophageal squamous cell carcinoma mediate the generation of B10 and PD-1high Breg cells. Cancer Sci. 2019 Sep;110(9):2700–10.

322. Deng J, Wei Y, Fonseca VR, Graca L, Yu D. T follicular helper cells and T follicular regulatory cells in rheumatic diseases. Nat Rev Rheumatol. 2019 Aug;15(8):475–90.

323. Zhao J, Chen Y, Zhao Q, Shi J, Yang W, Zhu Z, et al. Increased circulating Tfh17 and PD-1+Tfh cells are associated with autoantibodies in Hashimoto's thyroiditis. Autoimmunity. 2018;51(7):352–9.

324. Zacca ER, Onofrio LI, Acosta CDV, Ferrero PV, Alonso SM, Ramello MC, et al. PD-L1+ Regulatory B Cells Are Significantly Decreased in Rheumatoid Arthritis Patients and Increase After Successful Treatment. Front Immunol. 2018;9:2241.

325. Shen M, Wang J, Yu W, Zhang C, Liu M, Wang K, et al. A novel MDSC-induced PD-1-PD-L1+ B-cell subset in breast tumor microenvironment possesses immuno-suppressive properties. Oncoimmunology. 2018;7(4):e1413520.

326. Palit S, Heuser C, de Almeida GP, Theis FJ, Zielinski CE. Meeting the Challenges of High-Dimensional Single-Cell Data Analysis in Immunology. Front Immunol [Internet]. 2019 [cited 2020 May 16];10. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2019.01515/full

327. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010 Dec 15;127(12):2893–917.

328. International Early Lung Cancer Action Program Investigators, Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, et al. Survival of patients with stage I lung cancer detected on CT screening. N Engl J Med. 2006 Oct 26;355(17):1763–71.

329. Yang B, Li X, Ren T, Yin Y. Autoantibodies as diagnostic AAbs for lung cancer: A systematic review. Cell Death Discov. 2019 Dec;5(1):126.

330. Seijo LM, Peled N, Ajona D, Boeri M, Field JK, Sozzi G, et al. AAbs in Lung Cancer Screening: Achievements, Promises, and Challenges. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 2019 Mar;14(3):343–57.

331. Burotto M, Thomas A, Subramaniam D, Giaccone G, Rajan A. AAbs in Early-Stage Non–Small-Cell Lung Cancer: Current Concepts and Future Directions. J Thorac Oncol. 2014 Nov;9(11):1609–17.

332. Vargas AJ, Harris CC. AAb development in the precision medicine era: lung cancer as a case study. Nat Rev Cancer. 2016 Aug;16(8):525–37.

333. Xin L, Liu YH, Martin TA, Jiang WG. The Era of Multigene Panels Comes? The Clinical Utility of Oncotype DX and MammaPrint. World J Oncol. 2017 Apr;8(2):34–40.

334. Roberts MC, Weinberger M, Dusetzina SB, Dinan MA, Reeder-Hayes KE, Carey LA, et al. Racial Variation in the Uptake of Oncotype DX Testing for Early-Stage Breast Cancer. J Clin Oncol. 2016 Jan 10;34(2):130–8.

335. Shinjo K, Okamoto Y, An B, Yokoyama T, Takeuchi I, Fujii M, et al. Integrated analysis of genetic and epigenetic alterations reveals CpG island methylator phenotype associated with distinct clinical characters of lung adenocarcinoma. Carcinogenesis. 2012 Jul;33(7):1277–85.

336. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012 Sep 14;150(6):1121–34.

337. Yanagisawa K, Tomida S, Shimada Y, Yatabe Y, Mitsudomi T, Takahashi T. A 25-signal proteomic signature and outcome for patients with resected non-small-cell lung cancer. J Natl Cancer Inst. 2007 Jun 6;99(11):858–67.

338. Zaenker P, Gray ES, Ziman MR. Autoantibody Production in Cancer--The Humoral Immune Response toward Autologous Antigens in Cancer Patients. Autoimmun Rev. 2016 May;15(5):477–83.

339. Zaenker P, Lo J, Pearce R, Cantwell P, Cowell L, Lee M, et al. A diagnostic autoantibody signature for primary cutaneous melanoma. Oncotarget. 2018 Jul 17;9(55):30539–51.

340. Crosbie PAJ, Shah R, Summers Y, Dive C, Blackhall F. Prognostic and predictive AAbs in early stage NSCLC: CTCs and serum/plasma markers. Transl Lung Cancer Res [Internet]. 2013 Oct [cited 2021 Mar 22];2(5). Available from: https://tlcr.amegroups.com/article/view/1611

341. Blankenburg F, Hatz R, Nagel D, Ankerst D, Reinmiedl J, Gruber C, et al. Preoperative CYFRA 21-1 and CEA as prognostic factors in patients with stage I non-small cell lung cancer: external validation of a prognostic score. Tumour Biol J Int Soc Oncodevelopmental Biol Med. 2008;29(4):272–7.

342. Tomita M, Shimizu T, Ayabe T, Yonei A, Onitsuka T. Prognostic significance of tumour marker index based on preoperative CEA and CYFRA 21-1 in non-small cell lung cancer. Anticancer Res. 2010 Jul;30(7):3099–102.

343. Ramachandran N, Raphael JV, Hainsworth E, Demirkan G, Fuentes MG, Rolfs A, et al. Next-generation high-density self-assembling functional AAb arrays. Nat Methods. 2008 Jun;5(6):535–8.

344. Gnjatic S, Wheeler C, Ebner M, Ritter E, Murray A, Altorki NK, et al. Seromic analysis of antibody responses in non-small cell lung cancer patients and healthy donors using conformational AAb arrays. J Immunol Methods. 2009 Feb 28;341(1–2):50–8.

345. Kuhn M, Johnson K. Over-Fitting and Model Tuning. In: Kuhn M, Johnson K, editors. Applied Predictive Modeling [Internet]. New York, NY: Springer; 2013 [cited 2021 Mar 6]. p. 61–92. Available from: https://doi.org/10.1007/978-1-4614-6849-3_4

346. Dobbin KK, Simon RM. Optimally splitting cases for training and testing high dimensional classifiers. BMC Med Genomics. 2011 Apr 8;4(1):31.

347. Damelin M, Bankovich A, Bernstein J, Lucas J, Chen L, Williams S, et al. A PTK7-targeted antibody-drug conjugate reduces tumor-initiating cells and induces sustained tumor regressions. Sci Transl Med. 2017 Jan 11;9(372):eaag2611.

348. Chen R, Khatri P, Mazur PK, Polin M, Zheng Y, Vaka D, et al. A Meta-analysis of Lung Cancer Gene Expression Identifies PTK7 as a Survival Gene in Lung Adenocarcinoma. Cancer Res. 2014 May 15;74(10):2892–902.

349. Sachdev JC, Maitland M, Sharma M, Moreno V, Boni V, Kummar S, et al. A phase 1 study of PF-06647020, an antibody-drug conjugate (ADC) targeting AAb tyrosine kinase 7 (PTK7), in patients with advanced solid tumors including platinum resistant ovarian cancer (OVCA). Ann Oncol [Internet]. 2016 Oct [cited 2019 Jul 30];27(suppl_6). Available from: http://academic.oup.com/annonc/article/doi/10.1093/annonc/mdw435.29/2800538/A-phase-1-study-of-PF06647020-an-antibodydrug

350. Park M, Yoon H joon, Kang MC, Kwon J, Lee HW. PTK7 regulates radioresistance through nuclear factor-kappa B in esophageal squamous cell carcinoma. Tumor Biol. 2016 Oct;37(10):14217–24.

351. Kai-Bo Duan, Rajapakse JC, Haiying Wang, Azuaje F. Multiple SVM-RFE for gene selection in cancer classification with expression data. IEEE Trans NanoBioscience. 2005 Sep;4(3):228–34.

352. Jeggo PA, Pearl LH, Carr AM. DNA repair, genome stability and cancer: a historical perspective. Nat Rev Cancer. 2016 Jan;16(1):35–42.

353. Carl PL, Temple BRS, Cohen PL. Most nuclear systemic autoantigens are extremely disordered AAbs: implications for the etiology of systemic autoimmunity. Arthritis Res Ther. 2005;7(6):R1360-1374.

354. Gray ES, Rizos H, Reid AL, Boyd SC, Pereira MR, Lo J, et al. Circulating tumor DNA to monitor treatment response and detect acquired resistance in patients with metastatic melanoma. Oncotarget. 2015 Dec 8;6(39):42008–18.

355. Yousaf-Khan U, van der Aalst C, de Jong PA, Heuvelmans M, Scholten E, Lammers JW, et al. Final screening round of the NELSON lung cancer screening trial: the effect of a 2.5-year screening interval. Thorax. 2017 Jan;72(1):48–56.

356. Horeweg N, Scholten ET, de Jong PA, van der Aalst CM, Weenink C, Lammers JWJ, et al. Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers. Lancet Oncol. 2014 Nov;15(12):1342–50.

357. Subramanian J, Simon R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? J Natl Cancer Inst. 2010 Apr 7;102(7):464–74.

358. Kratz JR, He J, Van Den Eeden SK, Zhu ZH, Gao W, Pham PT, et al. A practical molecular assay to predict survival in resected non-squamous, non-small-cell lung cancer: development and international validation studies. Lancet Lond Engl. 2012 Mar 3;379(9818):823–32.

359. Oto J, Fernández-Pardo Á, Royo M, Hervás D, Martos L, Vera-Donoso CD, et al. A predictive model for prostate cancer incorporating PSA molecular forms and age. Sci Rep. 2020 Feb 12;10(1):2463.

360. Huang CS, Chen CY, Huang LK, Wang WS, Yang SH. Prognostic value of postoperative serum carcinoembryonic antigen levels in colorectal cancer patients who smoke. PLOS ONE. 2020 Jun 5;15(6):e0233687.

361. Yang J, Jiao S, Kang J, Li R, Zhang G. Application of serum NY-ESO-1 antibody assay for early SCLC diagnosis. Int J Clin Exp Pathol. 2015;8(11):14959–64.

362. Xu YW, Chen H, Guo HP, Yang SH, Luo YH, Liu CT, et al. Combined detection of serum autoantibodies as diagnostic AAbs in esophagogastric junction adenocarcinoma. Gastric Cancer. 2019 May 1;22(3):546–57.

363. Zhao J, Wang Y, Wu X. HMGN5 promotes proliferation and invasion via the activation of Wnt/β-catenin signaling pathway in pancreatic ductal adenocarcinoma. Oncol Lett. 2018 Sep;16(3):4013–9.

364. Wu J, Wang J. HMGN5 expression in bladder cancer tissue and its role on prognosis. Eur Rev Med Pharmacol Sci. 2018 Feb;22(4):970–5.

365. Li Q, Wei P, Huang B, Xu Y, Li X, Li Y, et al. MAEL expression links epithelial-mesenchymal transition and stem cell properties in colorectal cancer. Int J Cancer. 2016 Dec 1;139(11):2502–11.

366. Okada M, Nishio W, Sakamoto T, Uchino K, Yuki T, Nakagawa A, et al. Prognostic significance of perioperative serum carcinoembryonic antigen in non-small cell lung cancer: analysis of 1,000 consecutive resections for clinical stage I disease. Ann Thorac Surg. 2004 Jul;78(1):216–21.

367. Pujol JL, Molinier O, Ebert W, Daurès JP, Barlesi F, Buccheri G, et al. CYFRA 21-1 is a prognostic determinant in non-small-cell lung cancer: results of a meta-analysis in 2063 patients. Br J Cancer. 2004 Jun 1;90(11):2097–105.

368. Yu D, Du K, Liu T, Chen G. Prognostic value of tumor markers, NSE, CA125 and SCC, in operable NSCLC Patients. Int J Mol Sci. 2013 May 27;14(6):11145–56.

369. Karn T, Pusztai L, Ruckhäberle E, Liedtke C, Müller V, Schmidt M, et al. Melanoma antigen family A identified by the bimodality index defines a subset of triple negative breast cancers as candidates for immune response augmentation. Eur J Cancer Oxf Engl 1990. 2012 Jan;48(1):12–23.

370. Li XF, Ren P, Shen WZ, Jin X, Zhang J. The expression, modulation and use of cancer-testis antigens as potential AAbs for cancer immunotherapy. Am J Transl Res. 2020 Nov 15;12(11):7002–19.

371. Fanjul-Fernández M, Quesada V, Cabanillas R, Cadiñanos J, Fontanil T, Obaya A, et al. Cell-cell adhesion genes CTNNA2 and CTNNA3 are tumour suppressors frequently mutated in laryngeal carcinomas. Nat Commun. 2013;4:2531.

372. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci Transl Med. 2015 Apr 15;7(283):283ra54.

373. Fratta E, Coral S, Covre A, Parisi G, Colizzi F, Danielli R, et al. The biology of cancer testis antigens: putative function, regulation and therapeutic potential. Mol Oncol. 2011 Apr;5(2):164–82.

374. Simpson AJG, Caballero OL, Jungbluth A, Chen YT, Old LJ. Cancer/testis antigens, gametogenesis and cancer. Nat Rev Cancer. 2005 Aug;5(8):615–25.

375. Jakobsen MK, Gjerstorff MF. CAR T-Cell Cancer Therapy Targeting Surface Cancer/Testis Antigens. Front Immunol [Internet]. 2020 [cited 2021 Feb 17];11. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2020.01568/full

376. Raza A, Merhi M, Inchakalody VP, Krishnankutty R, Relecom A, Uddin S, et al. Unleashing the immune response to NY-ESO-1 cancer testis antigen as a potential target for cancer immunotherapy. J Transl Med. 2020 Mar 27;18(1):140.

377. Qiu CX, Bai XF, Shen Y, Zhou Z, Pan LQ, Xu YC, et al. Specific Inhibition of Tumor Growth by T Cell Receptor-Drug Conjugates Targeting Intracellular Cancer-Testis Antigen NY-ESO-1/LAGE-1. Bioconjug Chem. 2020 Dec 16;31(12):2767–78.

378. Sahin U, Oehm P, Derhovanessian E, Jabulowsky RA, Vormehr M, Gold M, et al. An RNA vaccine drives immunity in checkpoint-inhibitor-treated melanoma. Nature. 2020 Sep;585(7823):107–12.

379. Wang Z, Li Z, Ji H. Direct targeting of β-catenin in the Wnt signaling pathway: Current progress and perspectives. Med Res Rev [Internet]. [cited 2021 Feb 1];n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/med.21787

380. Kim JH, Kwon J, Lee HW, Kang MC, Yoon HJ, Lee ST, et al. AAb tyrosine kinase 7 plays a tumor suppressor role by inhibiting ERK and AKT phosphorylation in lung cancer. Oncol Rep. 2014 Jun 1;31(6):2708–12.

381. Semënov MV, Tamai K, Brott BK, Kühl M, Sokol S, He X. Head inducer Dickkopf-1 is a ligand for Wnt coreceptor LRP6. Curr Biol CB. 2001 Jun 26;11(12):951–61.

382. Zhang X, Ning Y, Xiao Y, Duan H, Qu G, Liu X, et al. MAEL contributes to gastric cancer progression by promoting ILKAP degradation. Oncotarget. 2017 Dec 26;8(69):113331–44.

383. Gurney A, Axelrod F, Bond CJ, Cain J, Chartier C, Donigan L, et al. Wnt pathway inhibition via the targeting of Frizzled receptors results in decreased growth and tumorigenicity of human tumors. Proc Natl Acad Sci U S A. 2012 Jul 17;109(29):11717–22.

384. Peng A, Maller JL. Serine/threonine phosphatases in the DNA damage response and cancer. Oncogene. 2010 Nov;29(45):5977–88.

385. Kauko O, O'Connor CM, Kulesskiy E, Sangodkar J, Aakula A, Izadmehr S, et al. PP2A inhibition is a druggable MEK inhibitor resistance mechanism in KRAS-mutant lung cancer cells. Sci Transl Med. 2018 Jul 18;10(450).

386. Jeong AL, Han S, Lee S, Su Park J, Lu Y, Yu S, et al. Patient derived mutation W257G of PPP2R1A enhances cancer cell migration through SRC-JNK-c-Jun pathway. Sci Rep. 2016 Jun 7;6(1):27391.

387. Morita K, He S, Nowak RP, Wang J, Zimmerman MW, Fu C, et al. Allosteric Activators of AAb Phosphatase 2A Display Broad Antitumor Activity Mediated by Dephosphorylation of MYBL2. Cell. 2020 Apr 30;181(3):702-715.e20.

388. Liao SF, Liang CH, Ho MY, Hsu TL, Tsai TI, Hsieh YSY, et al. Immunization of fucose-containing polysaccharides from Reishi mushroom induces antibodies to tumor-associated Globo H-series epitopes. Proc Natl Acad Sci U S A. 2013 Aug 20;110(34):13809–14.

389. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature. 2017 Jul;547(7662):222–6.

390. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, et al. Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors. Science. 2018 Jan 5;359(6371):91–7.

391. Lugg ST, Agostini PJ, Tikka T, Kerr A, Adams K, Bishay E, et al. Long-term impact of developing a postoperative pulmonary complication after lung surgery. Thorax. 2016 Feb;71(2):171–6.

392. Jones RO, Anderson NH, Murchison JT, Brittan M, Simon EJ, Casali G, et al. Innate immune responses after resection for lung cancer via video-assisted thoracoscopic surgery and thoracotomy. Innov Phila Pa. 2014 Apr;9(2):93–103; discussion 103.

393. Leggat DJ, Khaskhely NM, Iyer AS, Mosakowski J, Thompson RS, Weinandy JD, et al. Pneumococcal polysaccharide vaccination induces polysaccharide-specific B cells in adult peripheral blood expressing CD19+CD20+CD3−CD70−CD27+IgM+CD43+CD5+/−. Vaccine. 2013 Sep;31(41):4632–40.

394. Leggat DJ, Thompson RS, Khaskhely NM, Iyer AS, Westerink MAJ. The Immune Response to Pneumococcal Polysaccharides 14 and 23F Among Elderly Individuals Consists Predominantly of Switched Memory B Cells. J Infect Dis. 2013 Jul 1;208(1):101–8.

395. Kieser M, Wassmer G. On the Use of the Upper Confidence Limit for the Variance from a Pilot Sample for Sample Size Determination. Biom J. 1996;38(8):941–9.

396. Akinosoglou KS, Karkoulias K, Marangos M. Infectious complications in patients with lung cancer. Eur Rev Med Pharmacol Sci. 2013 Jan;17(1):8–18.

397. Putinati S, Trevisani L, Gualandi M, Guerra G, Rossi MR, Sartori S, et al. Pulmonary infections in lung cancer patients at diagnosis. Lung Cancer. 1994 Sep;11(3–4):243–9.

398. Kohno S, Koga H, Oka M, Kadota J, Kaku M, Soda H, et al. The pattern of respiratory infection in patients with lung cancer. Tohoku J Exp Med. 1994 Aug;173(4):405–11.

399. Berghmans T, Sculier JP, Klastersky J. A Prospective Study of Infections in Lung Cancer Patients Admitted to the Hospital. Chest. 2003 Jul;124(1):114–20.

400. Schmedt N, Heuer OD, Häckl D, Sato R, Theilacker C. Burden of community-acquired pneumonia, predisposing factors and health-care related costs in patients with cancer. BMC Health Serv Res. 2019 Jan 14;19(1):30.

401. Nwulu U, Brooks H, Richardson S, McFarland L, Coleman JJ. Electronic risk assessment for venous thromboembolism: investigating physicians' rationale for bypassing clinical decision support recommendations. BMJ Open. 2014 Sep 26;4(9):e005647–e005647.

402. Leffondré K, Abrahamowicz M, Siemiatycki J, Rachet B. Modeling Smoking History: A Comparison of Different Approaches. Am J Epidemiol. 2002 Nov 1;156(9):813–23.

403. Perlin E, Bang KM, Shah A, Hursey PD, Whittingham WL, Hashmi K, et al. The impact of pulmonary infections on the survival of lung cancer patients. Cancer. 1990 Aug 1;66(3):593–6.

404. Read WL, Tierney RM, Page NC, Costas I, Govindan R, Spitznagel ELJ, et al. Differential Prognostic Impact of Comorbidity. J Clin Oncol. 2004 Aug;22(15):3099–103.

405. Piccirillo JF, Tierney RM, Costas I, Grove L, Spitznagel EL. Prognostic importance of comorbidity in a hospital-based cancer registry. JAMA. 2004 May 26;291(20):2441–7.

406. Vincent JL, Dubois MJ, Navickis RJ, Wilkes MM. Hypoalbuminemia in Acute Illness: Is There a Rationale for Intervention?: A Meta-Analysis of Cohort Studies and Controlled Trials. Ann Surg. 2003 Mar;237(3):319–34.

407. Attiê R, Chinen LTD, Yoshioka EM, Silva MCF, de Lima VCC. Acute bacterial infection negatively impacts cancer specific survival of colorectal cancer patients. World J Gastroenterol WJG. 2014 Oct 14;20(38):13930–5.

408. Schmid M, Hanske J, Ravi P, Krishna N, Reznor G, Meyer CP, et al. Relationship between androgen deprivation therapy and community-acquired respiratory infections in patients with prostate cancer. Int J Urol. 2016;23(4):305–11.

409. Chang Y, Huh JW, Hong SB, Lee DH, Suh C, Kim SW, et al. Outcomes and prognostic factors of patients with lung cancer and pneumonia-induced respiratory failure in a medical intensive care unit: A single-center study. J Crit Care. 2014 Jun;29(3):414–9.

410. Hampson P, Dinsdale RJ, Wearn CM, Bamford AL, Bishop JRB, Hazeldine J, et al. Neutrophil Dysfunction, Immature Granulocytes, and Cell-free DNA are Early AAbs of Sepsis in Burn-injured Patients: A Prospective Observational Cohort Study. Ann Surg. 2017;265(6):1241–9.

411. Andruska A, Micek ST, Shindo Y, Hampton N, Colona B, McCormick S, et al. Pneumonia Pathogen Characterization Is an Independent Determinant of Hospital Readmission. Chest. 2015 Jul;148(1):103–11.

412. Lim WS, Baudouin SV, George RC, Hill AT, Jamieson C, Le Jeune I, et al. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. Thorax. 2009 Oct;64 Suppl 3:iii1-55.

413. Bartlett JG, Mundy LM. Community-Acquired Pneumonia. N Engl J Med. 1995;333(24):7.

414. Wong A, Marrie TJ, Garg S, Kellner JD, Tyrrell GJ, the SPAT Group. Increased risk of invasive pneumococcal disease in haematological and solid-organ malignancies. Epidemiol Infect. 2010 Dec;138(12):1804–10.

415. Garcia-Vidal C, Ardanuy C, Gudiol C, Cuervo G, Calatayud L, Bodro M, et al. Clinical and microbiological epidemiology of Streptococcus pneumoniae bacteremia in cancer patients. J Infect. 2012 Dec;65(6):521–7.

416. Kouranos V, Dimopoulos G, Vassias A, Syrigos KN. Chemotherapy-induced neutropenia in lung cancer patients: The role of antibiotic prophylaxis. Cancer Lett. 2011 Dec;313(1):9–14.

417. Cullen M, Gaunt C, Hastings M, Stuart N, Fernando I, Stanley A. Antibacterial Prophylaxis after Chemotherapy for Solid Tumors and Lymphomas. N Engl J Med. 2005;11.

418. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, et al. Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors. Science. 2018 Jan 5;359(6371):91–7.

Appendices

Appendix 1. Characterising the Impact of Pneumonia on outcome in NSCLC [Patel AJ et al, *J Thorac Dis* 2020 PMID 32642129]

Background

Lung cancer patients often present with pneumonia which remains an important cause of morbidity and mortality during the disease course (396). The pathogenesis of infection is multi-factorial and is influenced by the underlying disease, co-morbidities, and iatrogenic immunosuppression. In lung cancer this can range from infections resulting from tumour obstruction of the airways to infective exacerbations of chronic obstructive pulmonary disease (COPD), to opportunistic infections due to immunosuppression (396). The frequency of respiratory infection in Unplanned Hospital Admissions (UHAs) in the existing literature, is between 21.9% and 49% (209,397–399). Studies examining this are largely historical, variable in sample size and may be difficult to extrapolate from given the changes in the epidemiology and treatment of lung cancer. A retrospective study characterising lung cancer patients in the Netherlands highlighted the potential impact of respiratory infection on a nationwide level (11). The burden of community acquired pneumonia is high in cancer in general, with variations by cancer subtype and particularly high rates in lung cancer (400).

Retrospective Cohort Analysis

[Published analysis; Patel AJ et al, *J Thorac Dis* 2020 PMID 32642129]

Knowing that Lung cancer is a significant global burden, there is strong potential for improving survival if we can identify and treat modifiable risk factors for infection. Specifically, with reference to the uptake of Pneumovax (PPSV23) vaccination which is an NHS standard of care vaccine designed to protect against pneumonia and therefore an easily implementable strategy. We hypothesised that pneumonia presents a significant burden and is strongly linked to outcome in lung cancer patients who present to hospital and significantly more so

when compared to other cancer subtypes. We carried out an observational study, characterising the impact of pneumonia in three cohorts of patients (see below) and specifically assessing the importance of PPSV23 vaccination in NSCLC patients.

1. Non-Small Cell Lung Cancer (NSCLC) patients

2. Breast, Colorectal and Prostate cancer patients

3. All other patients without any form of malignancy

We analysed the proportion of UHAs from a single UK centre within these patient cohorts. NSCLC represents 85% of lung cancer and is pathologically and clinically very different to small cell lung cancer as a result this study is limited to the larger NSCLC cohort.

Data Synthesis Methods

This retrospective observational analysis was conducted at a single centre. The University Hospitals Birmingham is one of the largest teaching hospital trusts in England, serving a regional, national and international population and treating almost 2.2 million patients per year. Between April 2016 and June 2018, data was gathered concerning all UHAs to our centre from the entire cohort of NSCLC patients referred to our centre for diagnosis and treatment (n=605); this figure was calculated at the end of our study period and therefore accounts for all new cases of NSCLC found on admission. The time period was dictated by the availability and completeness of data from the electronic patients' records (EPR). The study was conducted with the approval of the UHB Trust and registered with the Queen Elizabeth Hospital audit department (audit code 14333).

The total number of NSCLC patients being treated at QEH within our time frame, both in the inpatient and outpatient setting, were identified through an informatics search for the ICD-10 code C349 and cross referenced for accuracy with the EPR 'prescribing information and communications system' (PICS), a technique which has shown validity in other studies (401). Our study cohort included all patients with known NSCLC who had a UHA to our centre; we retrospectively ascertained clinical and demographic data at admission and during their inpatient stay in this time period. The data included; demographics, past medication history, 5-year vaccine status (Pneumovax - PPSV23), lung cancer staging (TNM 8th Edition (9)), smoking status and co-morbidities. Smoking status was categorised into current, never smoked and ex-smokers who were further sub-stratified according to those who quit within the last 2 years, and those who quit greater than 2 years ago. This method was employed in an epidemiology study which described how to model smoking history (402). Admission data included haematological, biochemical and microbiological parameters, radiography, treatment initiated upon admission, in-hospital mortality and length of stay as well as 30-day readmission to hospital.

The type of admission was categorised according to those who were admitted with a pneumonia (Pn) and all other admissions (AOA). This was ascertained by scrutinising the presenting complaints from the patient (i.e. "difficulty in breathing", "shortness of breath", fever, purulent/productive cough, wheeze, limited exercise capacity) and the working diagnosis as documented in the post-take medical discharge letter. In addition, correlation was made with infection AAbs (white cell count, neutrophil count, C-reactive AAb (CRP), and albumin levels), positive microbiological growth from sputum and blood cultures. All chest

radiographs on admission were individually reviewed for evidence of consolidation or pneumonic changes to enrich the clinical data. With cases where there was a degree of ambiguity as to the aetiology of the changes on the "admission chest radiograph", relevant cases were discussed with radiology colleagues with further review of cross-sectional imaging (baseline or most recent CT thorax where available).

Where appropriate, the cause of death was obtained from the medical death certificate and hospital records. Deaths were classified as complications from respiratory infection (e.g., pneumonia, chest sepsis and respiratory or multi-system organ failure secondary to chest sepsis) whether from their initial admission or 30-day readmission, cancer related for those patients who died of disease progression, non-cancer related and if the cause of death was uncertain when records were not available classified as unclear.

A more limited data set were collected, over the same time period as cohort 1, for cohort 2 (Breast, Colorectal or Prostate cancer patients) and cohort 3 (all other patients without any form of malignancy), focusing on the type of UHA (Pn or AOA) and outcome data (length of in-hospital stay and mortality) for inter-cohort comparison.

Statistical Analysis

For continuous variables, results are expressed as means and standard deviations and for categorical variables, as counts and percentages. For continuous data, group comparison was carried out using a t-test or Mann-Whitney test depending on the distribution of data. Group differences for categorical data were assessed using the chi squared test of independence. Ordinal data were further assessed using Kendall's tau-b statistic. Inter-cohort comparison of continuous data was assessed using Kruskal-Wallis analysis. Univariate analyses of risk factors associated with respiratory related UHA were assessed by performing individual unadjusted

logistic regression analysis with inclusion of one covariate per model. A backward elimination, stepwise multivariate logistic regression analysis was performed to identify the independent predictors of in-hospital mortality within this dataset as a whole and sub-stratified according to those who presented with a UHA secondary to pneumonia. The tests were considered significant at $p<0.05$. Missing data were excluded from analyses. All analyses were performed using the IBM SPSS Statistics for Windows, Version 22.0 (IBM Corp. Released 2013. Armonk, NY: IBM Corp. and SAS 9.3 statistical package version (SAS Institute, Inc, Carry, North Carolina, USA).

Results

The informatics search identified 455 NSCLC, 1190 other cancer and 54,158 non-cancer patient UHAs to be included in analysis. Over the 26-month study period, the UHA rate for patients with NSCLC was 75.2%; there were 455 separate patient UHAs from a total cohort of 605 patients.

The NSCLC cohort are described in more detail showing that these patients are elderly (mean age 71.9 years) with a small male preponderance (56.7%) and a significant number of co-morbidities including 17.6% of patients with a coded diagnosis of COPD. The NSCLC patient UHAs were further stratified into those that presented with a pneumonia (Pn) and all other admission (AOA) cause (Table 6.1).

Variables	Value	Entire UHA Cohort (n=455)	Pneumonia (Pn) Cohort (n=164)	All Other Causes (AOC) Cohort (n=291)	p Value
Age	Mean (+/- SD) [years]	71.9 (+/- 11.3)	72.4 (+/- 10.5)	71.7 (+/- 11.8)	0.534
Gender	Male	258 (56.7%)	99 (60.4%)	159 (54.6%)	0.278
	Female	197 (43.3%)	65 (39.6%)	132 (45.4%)	
Chronic Obstructive Pulmonary Disease	Yes	80 (17.6%)	36 (22.0%)	44 (15.1%)	0.073
Smoking Status	Current	66 (14.5%)	26 (15.8%)	40 (13.7%)	0.003
	Ex-smoker (1 day – 2 years)	176 (38.7%)	65 (39.6%)	111 (38.2%)	
	Ex-smoker (>2 years)	70 (15.4%)	36 (22.0%)	34 (11.7%)	
	Never	143 (31.4%)	37 (22.6%)	106 (36.4%)	
Pneumovax (PPSV23) status in the last 5 years	Yes	147 (32.3%)	66 (40.2%)	81 (27.8%)	0.009
	No	308 (67.7%)	98 (59.8%)	210 (72.2%)	
Type 2 Diabetes Mellitus	Non-Insulin Dependent	44 (9.7%)	16 (9.8%)	28 (9.6%)	0.815
	Insulin Dependent	7 (1.5%)	2 (1.2%)	5 (1.7%)	
	No	404 (88.8%)	146 (89%)	258 (88.7%)	
Ischaemic Heart Disease	Acute MI within 30 days pre-admission	14 (3.0%)	3 (1.8%)	11 (3.8%)	0.32
	Stable angina	28 (6.2%)	13 (8.0%)	15 (5.2%)	

	No	413 (90.8%)	148 (90.2%)	265 (91%)	
Gastro-oesophageal reflux disease	Yes	4 (0.9%)	0	4 (1.4%)	0.302
Atrial Fibrillation	Yes	32 (7.0%)	13 (8.0%)	19 (6.5%)	0.572
Hypertension	Yes	90 (19.8%)	33 (20.1%)	57 (19.6%)	0.903
Chronic Renal Failure (stage III or greater)	Yes	6 (1.3%)	1 (0.6%)	5 (1.7%)	0.426
Abdominal Aortic Aneurysm	Yes	9 (1.9%)	3 (1.8%)	6 (2.1%)	0.583
Cerebrovascular Disease	Transient Ischaemic attack	8 (1.8%)	4 (2.4%)	4 (1.4%)	0.53
	Stroke	9 (2.0%)	2 (1.2%)	7 (2.4%)	
	No	438 (96.2%)	158 (96.4%)	280 (96.2%)	
Hypercholesterolaemia	Yes	12 (2.6%)	3 (1.8%)	9 (3.1%)	0.55
Peripheral Vascular Disease	Yes	9 (1.9%)	4 (2.4%)	5 (1.7%)	0.728
NSCLC Staging	Ia	67 (14.7%)	18 (11.0%)	49 (16.8%)	0.855
	Ib	32 (7.0%)	13 (7.9%)	19 (6.5%)	
	IIa	17 (3.7%)	5 (3.0%)	12 (4.1%)	
	IIb	33 (7.3%)	11 (6.7%)	21 (7.2%)	
	IIIa	62 (13.6%)	31 (18.9%)	33 (11.3%)	
	IIIb	39 (8.6%)	14 (8.5%)	25 (8.5%)	
	IIIc	13 (2.9%)	8 (4.9%)	5 (1.7%)	
	IVa	175 (38.5%)	58 35.4%)	117 (40.2%)	
	IVb	17 (3.7%)	6 (3.7%)	11 (3.7%)	
Histology*	Adenocarcinoma	198 (43.5%)	67 (40.9%)	131 (45%)	0.767
	Adenosquamous	7 (1.5%)	4 (2.4%)	3 (1.1%)	
	Squamous Cell Carcinoma	111 (24.4%)	42 (25.6%)	69 (23.7%)	

	Carcinoid	3 (0.7%)	1 (0.6%)	2 (0.7%)	
	Large Cell Neuroendocrine	10 (2.2%)	3 (1.8%)	7 (2.4%)	
	Synovial Sarcoma	2 (0.4%)	1 (0.6%)	1 (0.3%)	
	Unknown	124 (27.3%)	46 (28.1%)	78 (26.8%)	N/A
Previous Lung Resection*	Yes	42 (9.2%)	14 (8.5%)	28 (9.6%)	0.759
	No	348 (76.5%)	124 (75.6%)	224 (77%)	
	Unknown	65 (14.3%)	26 (15.9%)	39 (13.4%)	N/A
Recurrence of Disease following Lung Resection	Yes	21 (4.6%)	6 (3.7%)	15 (5.2%)	0.513
	No	21 (4.6%)	8 (4.9%)	13 (4.5%)	

*Missing/'Not available' data accounted for

Table 1: Baseline characteristics for UHA in NSCLC patients including univariate analyses of risk factors associated with Admission for Respiratory Infection

Within the NSCLC cohort, 164 UHAs were as a direct result of pneumonia (36.0%). The proportion of these UHAs which were associated with a new lung cancer diagnoses following admission was 3.1% (n=14) so 96.9% of UHAs were in known NSCLC patients (64% of new lung cancer diagnoses were stage III disease or higher (n=9)). Differences in demographic and clinical data between Pn and AOA sub-groups were explored (Tables 6.1 and 6.2); the respiratory cohort were more likely to have had the pure polysaccharide vaccine, Pneumovax (PPSV23) vaccine (p=0.009) and a positive smoking history (p=0.003). Co-morbidities, age; cancer staging, and histological diagnosis were not significantly different between the cohorts. Comparison of blood parameters and treatment profiles (Table 6.2) found CRP to be higher (p=0.001) and albumin to be significantly lower in the Pn sub-group (p<0.0001).

Variables	**Value**	**Pneumonia (Pn) Cohort (n=164)**	**All Other Causes (AOC) Cohort (n=291)**	**p Value**
White Cell Count*	Median [*10^6 cells/ml]	10.4	9.7	0.432
Neutrophil Count*	Median [*10^6 cells/ml]	8.3	7.5	0.243
Lymphocyte Count*	Median [*10^6 cells/ml]	1.1	1.3	0.172
CRP*	Median [mg/L]	60	17	0.001
Albumin	Mean (+/- SD) [grams/L]	36.4 (+/- 5.95)	39.03 (+/- 5.06)	<0.0001

Total AAb	Mean (+/- SD) [grams/L]	66.1 (+/- 8.53)	67.95 (+/- 7.84)	0.066
Clarithromycin	Yes	68 (41.5%)	33 (11.3%)	<0.0001
Co-amoxiclav	Yes	90 (54.9%)	85(29.2%)	<0.0001
Tazocin	Yes	68 (41.5%)	67 (23%)	<0.0001
Doxycycline	Yes	32 (19.5%)	14 (4.8%)	<0.0001
Ciprofloxacin	Yes	12 (7.3%)	7 (2.4%)	0.015

*log values analysed to portray normality of data.

Table 2. UHA Clinical data for NSCLC patients

Streptococcus pneumoniae was cultured from the secretions of 31 patients (18.9%) and in 2 cases (1.2%); there was co-growth of Haemophilus influenza. In-hospital mortality (p<0.0001) and length of hospital stay (p=0.031) were significantly higher in the Pn sub-group (Table 6.3).

Variables	**Value**	**Pneumonia (Pn) Cohort (n=164)**	**All Other Causes (AOC) Cohort (n=291)**	**p Value**
In-hospital Mortality	Yes	72 (43.9%)	38 (13.1%)	<0.0001
	No	92 (56.1%)	253 (86.9%)	
Total Length of Hospital Stay	Mean (+/- SD) [days]	11.76 +/- 13.4	9.35 +/- 10.1	0.031
30-day readmission	Yes	4 (2.4%)	19 (6.5%)	0.056
	No	160 (97.6%)	272 (93.5%)	

Table 3. Outcome data following Hospital Admission in NSCLC cohort

The overall 30-day readmission rate was 5.1% (n=23), 4 from the Pn and 19 from the non-respiratory sub-group (p=0.056). Eight readmissions were due to new pneumonia cases (34.8%). In-hospital mortality following readmission was 65.2% (n=15). Factors related to in-

hospital mortality within the NSCLC patient cohort were further explored by Multivariate analysis (Tables 6.4 and 6.5); patient age, PPSV23 status, Pn admission, Tazocin administration on admission, tumour stages IIb, IIb, IVa and IVb and smoking status were all significant independent predictors of in-hospital mortality. NSCLC staging was also significantly associated with in-hospital mortality on univariate analysis (p=0.017).

Variables	Estimate	Standard Error	p Value	Odds Ratio	95% Confidence Interval
CONSTANT	-9	1.366	<0.0001	0.0001	-
Age	0.065	0.015	<0.0001	1.067	1.035 – 1.099
Positive 5-year Pneumovax (PPSV23) status	-1.83	0.377	<0.0001	0.16	0.077 – 0.366
Pneumonia (Pn) admission	2.254	0.324	<0.0001	9.522	5.051 – 17.954
Tazocin administered on Admission	1.882	0.338	<0.0001	6.57	3.385 – 12.751
Current Smoker (versus never)	2.016	0.451	<0.0001	7.508	3.104 – 18.161
Ex-smoker [1 day – 2 years] (versus never)	1.1	0.371	0.003	3.004	1.451 – 6.217
Tumour Staging IIb (versus stage Ia disease)	2.75	0.724	<0.0001	15.646	3.785 – 64.670
Tumour Staging IIIb (versus stage Ia disease)	1.529	0.696	0.028	4.614	1.180 – 18.037
Tumour Staging IVa (versus stage Ia disease)	1.311	0.557	0.019	3.709	1.245 – 11.046
Tumour Staging IVb (versus stage Ia disease)	2.26	0.809	0.005	9.584	1.961 – 46.804

Table 4. Significant Independent Risk factors associated with In-hospital Mortality in the NSCLC Cohort on Multivariate Logistic Regression Analysis

Variables	Estimate	Standard Error	p Value	Odds Ratio	95% Confidence Interval
CONSTANT	-6.004	1.881	0.001	0.002	-
Age	0.069	0.024	0.003	1.072	1.023 – 1.122
Positive 5-year Pneumovax (PPSV23) status	-1.312	0.497	0.008	0.269	0.102 – 0.713
Raised Neutrophil Count on Admission	0.09	0.043	0.038	1.094	1.005 – 1.191
Current Smoker (versus never)	1.899	0.724	0.009	6.678	1.615 – 27.614

Table 5. Significant Independent Risk factors associated with In-hospital Mortality in the Pneumonia (Pn) sub-group within the NSCLC Cohort on Multivariate Logistic Regression Analysis

Within the NSCLC cohort, odds ratios of 0.160 (95% CI 0.077 – 0.366; p<0.0001) for positive PPSV23 status and 9.522 (95% CI 5.051 – 17.954; p<0.0001) for Pn status indicate that for patients admitted to hospital with pneumonia without previous PPSV23 in the last 5 years, the odds of death were almost 60-fold higher. There were no significant associations between other blood parameters, tumour histology or comorbidity and in-hospital mortality.

Inter-cohort Comparison

There was a significantly higher incidence of pneumonia in the NSCLC cohort (cohort 1) compared with cohorts 2 and 3, p<0.0001 (36.0% versus 1.3% versus 2.2% respectively). Clinical and demographic data for the NSCLC cohort show there was a significantly higher incidence of COPD in the NSCLC cohort compared with cohorts 2 and 3 (p<0.0001). When compared with the outcome data in cohorts 2 and 3, NSCLC patients with pneumonia had a

significantly higher rate of in-hospital mortality and length of in-hospital stay compared with pneumonia in cohorts 2 and 3 ($p<0.0001$ and $p=0.011$ respectively) (Table 6.6 and Figure 6.1).

Variables	**Value**	**Pneumonia (Pn) NSCLC Cohort only (n=164)**	**Pneumonia (Pn) in Other Malignancy Cohort (breast, colon, prostate) (n=15)**	**Pneumonia (Pn) in Non-Malignancy Cohort (n=1185)**	**p Value**
In-hospital mortality	Yes	72 (43.9%)	6 (40.0%)	215 (18.1%)	<0.0001
Length of In-hospital stay	Median number of days	8	7	5	0.011
Readmission within 30 days	Yes	4 (2.4%)	0 (0.0%)	9 (0.8%)	0.141

Table 6. Outcome data following Hospital Admission in three UHA cohorts

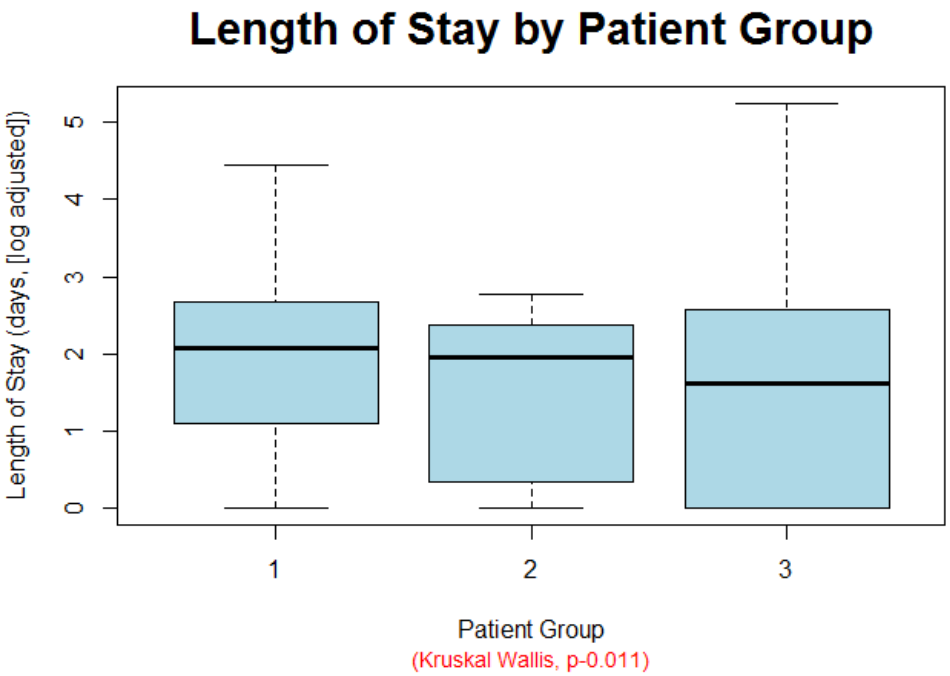


Figure 1. Length of in-hospital stay by patient group

Discussion

Our study has shown a high incidence of UHA in patients with NSCLC and admission for pneumonia affects over one third of all NSCLC patients admitted to hospital which is higher than for other cancer cohorts (breast, colorectal, prostate) and non-cancer patients. In our NSCLC cohort, we found that 36% of UHAs were due to pneumonia which is similar to 34.3%, a rate from a previously reported study examining broncho-alveolar lavage (BAL) growth in all lung cancer admissions over a retrospective period (397). Similarly, the incidence of community-acquired pneumonia from a population of 89,007 cancer patients revealed 21-fold higher rate in lung cancer patients compared with other non-malignancy, propensity-scored, age-matched cohorts (400). Advanced stage lung cancer in particular is difficult to treat and given the high prevalence of Stage III disease or higher in the respiratory cohort (71.3%, n=117), the control of infection may be an important clinical step in improving the prognosis of lung cancer.

Survival outcome from pneumonia in NSCLC is significantly worse than for other cancer and non-cancer patients. For NSCLC patients, pneumonia is associated with significantly longer lengths of in-hospital stay and is an independent risk factor for in-hospital mortality as is the need for readmission and not having had the PPSV23 pneumonia vaccination. Historical, retrospective data have shown median survival of all lung cancer (small and non-small cell) patients following hospital admission with a respiratory infection was significantly lower than those who did not contract respiratory infection during their inpatient stay (4.2 months versus 12.9 months; $p<0.05$) (403). In the context of a relatively lethal disease such as lung cancer, intervening to reduce the mortality from the tumour itself or from comorbidities is challenging. This study shows that inter-current events such as infection have a detrimental

effect on mortality and infection prevention and early management could contribute large gains to overall survival (11,404,405).

The physiological status of patients presenting with pneumonia in the NSCLC cohort was worse than in those who did not. Hypoalbuminaemia is known to be an independent predictor of poor outcome in hospitalised patients, and this association appears to be independent of both inflammation and nutritional status (406). Our study demonstrated that albumin levels on admission to hospital were significantly lower in patients with pneumonia, and CRP levels higher. Neutrophil count emerged as an independent predictor of mortality in the respiratory admission sub-group only; higher counts on admission conferred a poorer in-hospital outcome. This relationship has also been demonstrated in colorectal cancer patients suffering from bacterial infections with higher neutrophil counts conferring poorer cancer-specific survival ($p=0.02$) (407). In prostate cancer, exposure to medical androgen receptor therapy is associated with a significantly higher risk of community-acquired pneumonia ($p< 0.001$) presumably owing to poorer physiological reserve (408). Chang et al postulated that prognosis of lung cancer patients with severe pneumonia is mainly dependent on the underlying lung injury, which can be surmised through admission criteria, i.e. higher inflammatory markers connote a more intense systemic inflammatory response indicative of greater on-going parenchymal injury and factors such as previous radiotherapy or chemotherapy also contribute to lung injury (409). Higher neutrophil counts on admission would indicate a robust and active innate immune response with greater degrees of on-going inflammation and injury (410).

Streptococcus pneumoniae was the most frequently cultured organism from sputum in our NSCLC pneumonia sub-group, at a rate of 18.9%. Few studies have managed to address the microbiological repertoire of culprit organisms in pulmonary infection in the setting of NSCLC and culture rates vary considerably. Diagnosing pneumonia in this cohort of patients can prove to be difficult given that clinical signs such as pyrexia and productive cough, allied with raised inflammatory markers such as CRP and white cell count, may be a manifestation of the cancer itself as opposed to pulmonary sepsis. Combining clinical criteria with radiological findings and microbiology is essential. Determining the exact culprit organism in pneumonia is often difficult and unreliable; sputum culture examination has a low specificity and BAL fluid is frequently contaminated with upper airway bacterial commensals (397,398). A 6-year retrospective analysis evaluated 9624 patients with a discharge diagnosis of pneumonia, 46.1% of these patients were culture-negative, of the ones that were culture-positive (51.2%), pathogens that were commonly grown include S. pneumoniae (11.4%), H. influenza (16.7%) and Methicillin Sensitive Staphylococcus Aureus (MSSA) (26.0%) (411). The S. pneumoniae positive cases are potentially vaccine preventable infections. This large study demonstrated that culture-negative pneumonia was independently associated with lower risk for 90-day readmission (411); our study did not demonstrate a significant link between pneumonia and readmission, the numbers were small therefore perhaps not adequately powered to look at this specifically.

This analysis has highlighted the significant difference in mortality between those who have received PPSV23 (3.96%) and those who have not (20.2%). Five-year positive PPSV23 status was significantly higher in the NSCLC Pn sub-group ($p=0.009$), likely owing to the fact this was a higher risk group and therefore warranted vaccination. Multivariate analysis identified 5-year PPSV23 status as an independent predictor of mortality ($p<0.0001$); in the entire NSCLC

cohort, not having had the vaccine conferred 6.25-fold higher odds of in-hospital death. *S. pneumoniae* is the most prevalent organism (60-75%) responsible for cases of community acquired pneumonia (CAP) in the UK (412).

In the United Kingdom and in the USA, the administration of PPSV23 is a national health directive led recommendation for all adults above the age of 65 and those below the age of 65 with a long-term health condition. However, for patients with cancer, there has been little investigation into the efficacy of PPSV23. Given that lung cancer is responsible for the highest burden of cancer-related deaths globally, identifying preventable risk factors for early morbidity and mortality in this cohort of patients should be a clear public health aim (1). Preventing pneumonia and reinforcing the need for smoking cessation are crucial. Owing to intensive anti-cancer therapies, lung cancer patients have a degree of immune compromise making them significantly more susceptible to invasive pneumococcal disease (IPD) than the overall adult population (413,414). Studies have shown that between 69% and 72.4% of cancer patient IPD cases were caused by serotypes covered in the 23-valent PPSV23 (414,415). The level of information in the literature regarding the effectiveness of PPSV23 in cancer patients is limited however a nationwide population based cohort study from Taiwan examined the effects of PPSV23 in 157 newly diagnosed elderly lung cancer patients (all comers) (210). Two-year cumulative CAP hospitalisation rates were 37.1% versus 55.4% for lung cancer patients with and without PPSV23 ($p<0.001$) and overall survival rates were 46.6% versus 26.2% for lung cancer patients with and without PPSV23 ($p<0.001$) (210).

Our data demonstrated significantly higher antibiotic administration on admission in the NSCLC Pn sub-group; in lung cancer patients in whom infection is suspected, empiric antibiotic

therapy is often administered, and more so in the setting of chemotherapy induced neutropenia (396,416). Prophylactic administration of fluoroquinolones post-chemotherapy in lung cancer patients has been shown to reduce the incidence of fever, probable infection and hospitalisation (417). However, advocating increased antibiotic usage in an era of antimicrobial resistance needs challenging. Moreover, specific to NSCLC, in the era of immunotherapy, Microbiome data has demonstrated that the routine use of antibiotics in NSCLC reduces the efficacy of checkpoint blockade agents conferring poorer median overall survival (8.3 months versus 15.3 months, p=0.001) (418).

Conclusions

Pneumonia contributes a large burden to the comorbidity of NSCLC. Given that a number of new lung cancer diagnoses were made on admission with pneumonia, being more vigilant and screening for infection in this “at risk” cohort may help to identify disease at an earlier stage and thus bode better patient outcomes. There is a higher burden of pneumonia in NSCLC patients compared with other cohorts, and these patients have a poorer physiological status as demonstrated by our biochemical admission data. This, coupled with poor vaccine uptake in the cohort in general and high antibiotic use in pneumonia patients are highly conducive to poor outcome in this group .Therefore, vigilance for infection, optimising vaccination, judicious use of antibiotics particularly in the era of immunotherapy, patient education, early diagnosis with adequate assessment and efforts to identify a culprit organism should be a priority in both the inpatient and outpatient setting to improve outcome in NSCLC and reduce the incidence of hospital admission. These are all low cost, easy to implement strategies that could be rapidly adopted to make significant advances in reducing mortality from infection in NSCLC.

Study Strengths and Limitations

This study has a number of limitations; firstly, it was undertaken in a large tertiary centre in the UK and so it is possible that the UHA rate has been underestimated as patients have attended elsewhere and this study doesn't include primary care contacts. The study is retrospective making it potentially prone to interpretation or selection bias and the data was not recorded in a systematic, planned manner. A prospective study could have gathered more complete data on performance status and lung function which could contribute to infection risk. More comprehensive data could also then be gathered for NSCLC patients who did not have a UHA to identify risk factors for admission. Similarly, owing to lack of data contemporaneity, it was difficult to perform further sub-group analyses for the "other cancer" and "non-cancer" cohorts as was done for the NSCLC cohort. Despite these limitations this large study does demonstrate the frequency of UHAs for NSCLC patients in a single centre and highlights the importance of pneumonia in the morbidity and mortality of this disease, particularly when compared with other cancer types and indeed non-malignancy patients.

Study Rationale

The literature suggests that functional antibody assessment to vaccination in NSCLC patients has not yet been done. The paradigm of immune incompetence as a function of the lung cancer suggests that response to vaccine may also be impaired and assessing immune competence in these patients could be used to stratify or indicate prognosis/outcome. We intended to study this through design and conduct of an observational study in NSCLC patients being vaccinated against pneumonia with PPSV23. The outlines and details of this study are discussed in the "Further Work" section. Owing to the outbreak of the global COVID-19

pandemic, recruitment was halted just as the study had opened at both sites. We intend to continue with this work once the COVID-19 pressures ease and the study is once again given the “green light”.