Defining preinvasive remodelling of CD4 T cell differentiation to intercept lung squamous carcinogenesis

Teerapon Sahwangarrom

University College London

PhD Supervisor: Dr James L Reading

A thesis submitted for the degree of

Doctor of Philosophy

University College London

September 2025

Declaration

I Teerapon Sahwangarrom confirm that the work presented in this thesis is my own, and Grammarly was used solely to assist with grammar, spelling, and language clarity. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Lung cancer is the leading cause of cancer death worldwide, with late-stage disease accounting for the majority of diagnoses, underscoring the urgent need for improved early detection and prevention technologies. T cell responses are initiated and dysregulated during early carcinogenesis, providing a basis for immune-focused early cancer detection and interception strategies. Here, we profiled pre-invasive lesions of human and murine airways to examine whether early T cell reprogramming during lung squamous cell carcinoma (LUSC) progression could be exploited for the early detection and precision immunoprevention of lung cancer.

Here, I performed scRNAseq analysis of human pulmonary carcinoma-in-situ (CIS) lesions and adjacent normal lung tissue. In our primary UCLH bronchoscopy cohort, we discovered that CIS lesions are characterised by an early expansion of highly suppressive *BATF*+*Helios*+*OX40*+*GITR*+*ICOS*+CD177+ Treg cells ('BATF Tregs', n=16 samples, 7 patients). We validated this result in a scRNAseq meta-analysis of 457 samples from 251 patients, and bulk gene expression data of 122 bronchial biopsies from 77 patients. BATF+Tregs were also associated with resistance to anti-PD-1 in the advanced setting using publicly available data, independent of mutational burden and overall infiltration (n=195). Within CIS lesions, clonal tracking identified that BATF+Tregs likely emerged from local expansion, *in situ* differentiation from stem-like progenitor pools and infiltration of effector Tregs from the circulation. This result led to my team finding that circulating effector Tregs could be used to track and predict preinvasive disease progression in the blood of individuals undergoing bronchoscopy and CT screening, providing a platform for early detection under patent.

Using scRNAseq to analyse lung, blood and draining lymph nodes from a carcinogen-driven mouse model of LUSC carcinogenesis, I found that Batf+Tregs and systemic effector Tregs were significantly increased during preinvasive progression *in vivo*, recapitulating our results from human LUSC development. Finally, I discovered pathways to support immune interception of Batf+Treg differentiation, one of which has been functionally validated by my team. These data suggest that preinvasive BATF+Tregs could be tracked and targeted to provide a novel platform for interception of LUSC in those at risk.

Impact Statement

Lung cancer remains the foremost cause of cancer-related mortality globally, with most cases detected at advanced stages, highlighting the pressing demand for more effective approaches to early detection and prevention. Importantly, T cell activity is both triggered and disrupted during the initial phases of carcinogenesis, forming a critical foundation for developing immune-based strategies for early cancer detection and interception. An understanding of immune cell interaction with other cell types and T cell dynamics in the pre-invasive microenvironment is elusive, hindering the development of cancer interception and early detection strategies.

This thesis overcomes that limitation by performing scRNAseq with TCRseq data of human CIS and paired lung tissues. This allows me to identify BATF+Tregs that are significantly enriched in human CIS, as well as CD4 subsets that can differentiate into BATF+Tregs within the CIS lesion based on the number of shared TCRs. Moreover, we also performed scRNAseq with TCRseq data of human peripheral blood mononuclear cells (PBMCs). This allows me to identify PBMC CD4 subsets that shared TCRs with CIS BATF+Tregs, as well as markers for early detection. Moreover, this thesis demonstrates crosstalk axes that could support BATF+Treg differentiation in human CIS and a validation cohort to verify that BATF+Tregs were increased during LUSC carcinogenesis by integrating the lung cancer atlas with scRNAseq of our UCLH cohort.

In addition, our lab, in collaboration with Professor Sam Janes' group, took advantage of a murine model of LUSC driven by chemical carcinogen exposure, mirroring the multistage process of human LUSC. I then performed analysis of scRNAseq with TCRseq data from this mouse model, allowing me to verify results from our human analysis in a relevant experimental model. Herein, I compared the transcriptional profile of Batf+Tregs between human and mouse preinvasive LUSC, refining clinically and experimentally actionable immune interception targets with direct translational relevance. This work highlights the remodelling of T cells during LUSC

development and potential targets for the interception of LUSC, which could impact how we understand and intervene in lung cancer.

Acknowledgements

I would like to thank my principal supervisor, Dr James L Reading, for his continued guidance and support and for giving me the opportunity to complete my PhD at UCL within the Pre-Cancer Immunology laboratory. During my first year as a PhD student, we met very regularly, until I developed my own pipelines for scRNAseq. Moreover, I also gained immunology knowledge whenever we had a 1-1 meeting, even if it was little by little.

Moreover, I would like to thank my peers and co-authors of our manuscript in revision: Samuel Gamble, Claudia Peinador-Marin, and Dr Zoe Whiteman. We have been working hard together with James for years to gather the story for the paper. Hopefully, we will get this work accepted and celebrate together soon.

In addition, I would like to thank Dr Sandra Gomez-Lopez, Dr Zoe Whiteman, Dr Marta Lebrusant-Fernandez, and Abigail Shurr for performing the NTCU mouse experiments at the 18-week and the 24-week timepoints. These required enormous energy and incredibly hard work. However, you guys have done very well, and we got new data to support our findings for the CD4 paper and my thesis.

I would like to thank everyone else in the Reading lab, Andrei Enica, Amber Rogers, Anakin Seng Kuong Ung, Dr Hongui Cha, Dr Constantin Ahlmann-Eltze, Daniel Osei-Bordom, Dr Marco De Donatis, and Dr Petros Fessas, for all the feedback during my PhD.

Finally, I would like to thank my family members (my mom and dad) for all their support during my PhD. Also, I would like to thank Chulabhorn Royal Academy and the Royal Thai government, who gave me a PhD studentship to support my PhD study.

Table of Contents

Table of figures

List of tables

Abbreviations

A2A	Adenosine 2A receptor
AAH	Atypical Adenomatous Hyperplasia
ADCC	Antibody-Dependent Cellular Cytotoxicity
AFB	Autofluorescence bronchoscopy
APCs	Antigen-presenting cells
ATP	Adenosine Triphosphate
BATF	Basic Leucine Zipper ATF-Like Transcription Factor
BCR	B cell receptor
cDC1	conventional dendritic cell type 1
cDC3	conventional dendritic cell type 3
CIS	Carcinoma In Situ
CP	Cardiac puncture
CTLs	Cytotoxic T lymphocytes
DCs	Dendritic cells
DEG	Differentially expressed gene
dLN	draining lymph node
eTregs	effector Tregs
FACS	Fluorescence-activated cell sorting
FDR	False Discovery Rate
FOXP3	Forkhead box protein P3
FVB/N	Friend Virus B strain, originally derived from NIH Swiss mice
GLM	Generalized Linear Model
GSEA	Gene set enrichment analysis
HLA	Human Leukocyte Antigen
HNSCC	Head and Neck Squamous Cell Carcinoma
HSP	Heat Shock Protein
ICB	Immune checkpoint blockade
IHC	immunohistochemistry

IPEX	Immune dysregulation, Polyendocrinopathy, Enteropathy, X-linked syndrome.
iTregs	induced Tregs
KNN	k-nearest neighbour
LDCT	Low-dose computed tomography
LLL	Left Lower Lobe
LMB	Left Mainstem Bronchus
LOH	loss of heterozygosity
LT	Left side of the thorax
LUAD	Lung Adenocarcinoma
LUSC	Lung Squamous Cell Carcinoma
MAIT	Mucosal-Associated Invariant T cell
MDSCs	Myeloid-Derived Suppressor Cells
MHC	Major Histocompatibility Complex
mIF	multiplex immunofluorescence
MNU	Methylnitrosourea
NES	Normalised enrichment score
NINJA	iNducible iNtron-encoded Junctional Antigen
NK	Natural Killer
NSCLC	non-small cell lung cancer
NTCU	N-Nitroso-Tris-Chloroethyurea
OR	Odds ratio
PCA	Principal Component Analysis
pDC	plasmacytoid dendritic cell
PID	Pre-invasive disease
PMLs	pre-malignant lesions
QC	Quality control
RIB_POST	Posterior Rib
RLL	Right Lower Lobe
RNA	Ribonucleic Acid
RUL_CAR	Right Upper Lobe Carina
SCC	Squamous cell carcinoma
TAAs	Tumour-associated antigens
TAMs	Tumour-associated macrophages

Tconv	conventional T cells
TCR	T cell receptor
Tcm	**central memory T cells**
Tem	effector memory T cells
Tex	exhausted T cells
Tfh	follicular helper T cells
Th1	helper 1 T cells
Th17	helper 17 T cells
Th2	helper 2 T cells
Th9	helper 9 T cells
TIL	Tumour-Infiltrating Lymphocyte
TLSs	Tertiary Lymphoid Structures.
Tm	memory T cells
TMB	Tumour Mutational Burden
TMM	trimmed mean of M-values
TNF	Tumour necrosis factor
Tregs	regulatory T cells
Trm	resident memory T cells
TSDR	Treg-specific demethylated region
tTregs	thymus-derived Tregs
UCL	University College London
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique Molecular Identifier
WES	Whole-Exome Sequencing
WLB	White light bronchoscopy

Chapter 1. Introduction

1.1 Lung Cancer

Lung cancer is the leading cause of cancer death worldwide, due mostly to late-stage diagnosis. In general, lung cancer is classified into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), with the latter comprising approximately 85% of cases (1). Within NSCLC, adenocarcinoma has overtaken lung squamous cell carcinoma (LUSC) as the most common histological subtype worldwide (2). Despite therapeutic advances, approximately three-quarters of patients are diagnosed at stage III or IV disease, at which point metastatic dissemination to lymph nodes or distant organs renders curative treatment unfeasible. By contrast, patients diagnosed earlier show markedly superior outcomes. For example, patients with stage IA1 disease (tumours <1 cm in diameter, without bronchial invasion or nodal involvement) achieve five-year survival rates exceeding 90%, whereas only 10% of patients with stage IV disease survive more than five years (3).

Due to these survival disparities, the UK Lung Cancer Coalition set up the 25 by 25 initiative (4), and links this aim to the NHS's long-term plan that aims to diagnose 75% of cancers at an early stage. Its goal is to increase five-year lung cancer survival to 25 per cent by 2025. Though little progress has been made, five-year survival rates increased from 9% in 2005 to 16% (5), yet these five-year survival outcomes are worse in the UK compared to other developed countries (5). Related to this, 15% of patients with NSCLC in the UK are diagnosed at stage I, as opposed to 24% in the United States; such differences have resulted in substantial numbers of avoidable deaths (6). Causes of this difference are considered to be a combination of geographic variation in access to curative resection or radical radiotherapy, as well as diagnostic delays at the levels of patient presentation, primary care, and secondary care.

Multiple strategies have been made to overcome these deficiencies. Public health campaigns such as Got a cough, get a check (7) and targeted educational programmes have been proposed to promote symptom identification, and general practice-based intervention for promoting adherence to chest X-ray referral guidelines

(7). In addition, the National Lung Cancer Audit helped to support systematic assessment of nationwide results and improved care by systematically appraising national results and making reforms. The recent data shows that there has been a 5.4% increase in the number of lung cancer operations and importantly less variation between UK trusts in provision of care. The improvements have also been supported by therapeutic innovation. For example, stereotactic ablative radiotherapy (SABR) now provides a potentially curative option for patients who are unfit for surgery (8), and the development of molecularly targeted therapies and immune checkpoint inhibitors has expanded treatment options. However, such systemic agents remain restricted to advanced disease and confer only modest extensions in survival, typically measured in months not years (9).

Low-dose computed tomography (LDCT) screening for high-risk individuals represents an important advancement in shifting the detection of lung cancer towards earlier stages, particularly for peripheral adenocarcinomas. However, this approach has limited applicability to central airway tumours (10), including early-stage LUSC, which frequently arise from preinvasive squamous lesions. These lesions constitute the histopathological precursors of LUSC and therefore represent an important window for early detection and therapeutic interception.

Improving outcomes for patients with LUSC requires the development of reliable biomarkers that can identify preinvasive lesions and stratify their malignant potential. Understanding the molecular and cellular processes that control lesion progression will be crucial for the discovery of biomarkers and chemopreventive methods aimed at halting malignant transformation. The current fundamental clinical dilemma lies with whether preinvasive lesions justify active therapeutic intervention. Bronchoscopic, tissue-preserving therapies offer an attractive alternative to radical surgery, but their effectiveness requires a thorough prospective evaluation before they can be implemented into everyday practice.

1.2 Preinvasive Squamous Airway Disease

The human airway epithelium forms a pseudostratified barrier that protects the lungs against pathogens and inhaled particulate matter through the mucociliary clearance system. It consists of three main cell types: ciliated cells, secretory cells, and basal cells. In the proximal airway, goblet cells are the predominant secretory cells, while club cells and neuroendocrine cells are more abundant in the distal airways (11). Basal cells are a stem population that maintains the differentiated luminal cells when cells are lost during both normal homeostasis and states of injury (12). Importantly, basal cells are thought to represent the cell of origin for LUSC, owing to their high proliferative capacity and increased vulnerability to mutation during cell division (13). Supporting this, basal cell markers such as keratin 5 (KRT5) are found in approximately 87% of LUSC tumours (14), and mouse models have demonstrated that basal cells can give rise to LUSC-like tumours (15). For example, in mouse models, biallelic inactivation of *Lkb1* and *PTEN* in basal cells gives rise to LUSC that closely mirror human disease (15).

LUSC is a subtype of NSCLC, accounting for 20-30% of cases globally (16–18). Despite its prevalence as the second most common lung cancer, LUSC is characterised by a dismal prognosis and a lack of effective targeted therapies, unlike LUAD, which is better understood genetically and catered for therapeutically (19,20). LUSC progresses through a series of pre-invasive lesions (PIDs) that progressively transform normal bronchial epithelium into invasive carcinoma (21–23). This stepwise process is characterised by histological, genetic, and transcriptomic changes, many of which are already detectable in early lesions and may serve as biomarkers of malignant potential (Fig. 1.1).

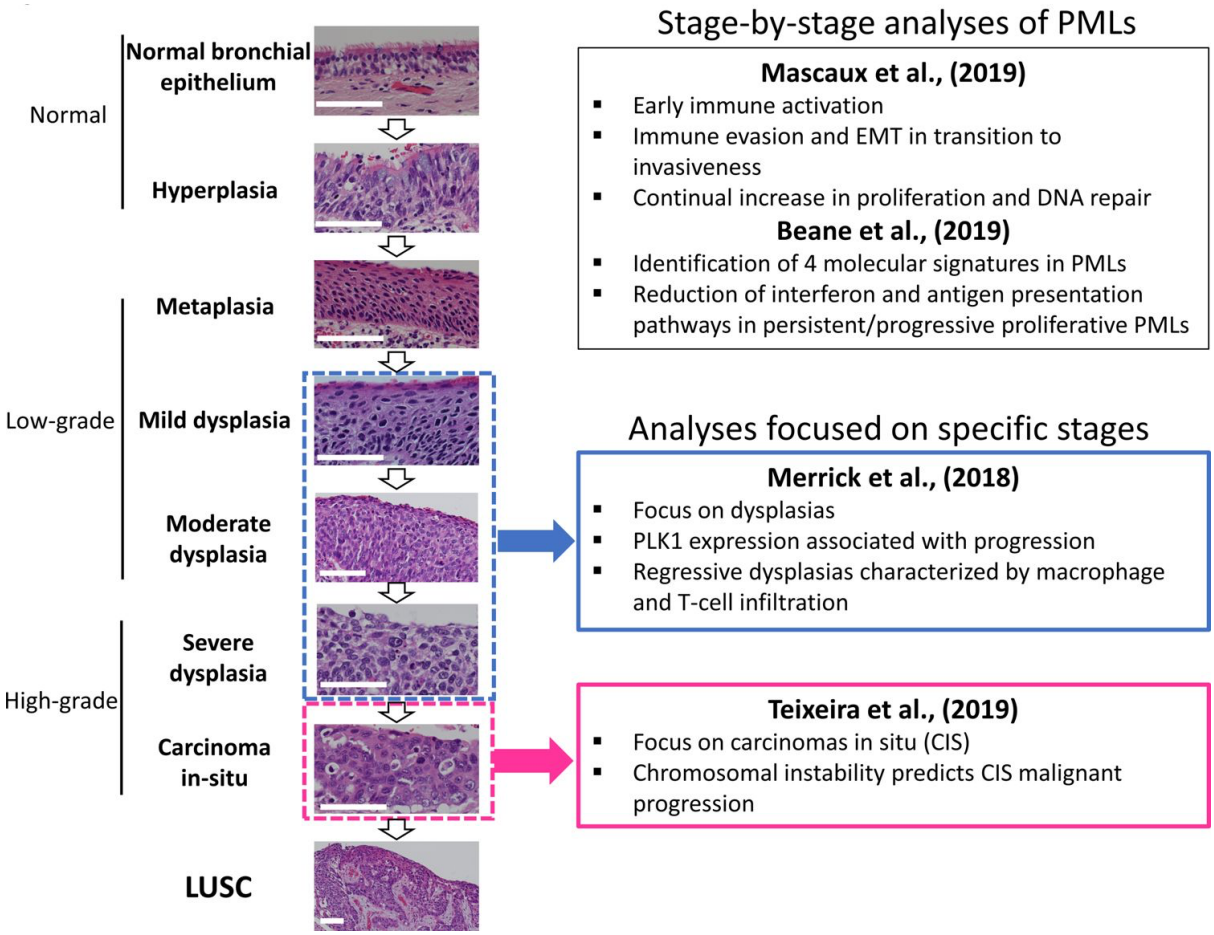

Figure 1.1: The developmental trajectory of pre-invasive lesions in lung squamous cell carcinoma (LUSC), accompanied by representative haematoxylin–eosin stained histological sections and a summary of findings from the four studies integrated within the XTABLE (Exploring Transcriptomes of Bronchial Lesions) resource. PIDs are generally categorised as normal epithelium (including hyperplastic changes), low-grade, or high-grade lesions. Of the included studies, Mascaux et al. (2019) and Beane et al. (2019) profiled gene expression across multiple developmental stages, whereas Merrick et al. (2018) and Teixeira et al. (2019) concentrated on dysplastic lesions (indicated in blue) and carcinoma in situ (CIS; indicated in pink), respectively. The figure also highlights the principal findings of each study. Figure adapted from Roberts et al. 2023.

1.2.1 Normal Epithelium and Hyperplasia

The bronchial epithelium in healthy individuals is a pseudostratified columnar epithelium comprising basal cells, ciliated cells, and goblet cells. It is the usual lining of the airway that serves as a protective barrier. This normal epithelium maintains tissue homeostasis and functions properly in protecting the lungs from pathogens and particulates such as bacteria, viruses, fungi, allergens, and pollutants.

The earliest morphological changes in LUSC development are seen as epithelial hyperplasia, where the number of basal cells expands abnormally, leading to a thickening of this lining (24). The earliest molecular changes in hyperplasia include an increase in proliferation and DNA repair that reflected in the ascending gene module associated with proliferation, which evolves from normal tissue to cancer (24). Also, a transitory increase in metabolism and early immune sensing through activation of resident immune cells is observed in low-grade preinvasive lesions, which come after hyperplasia (24). This indicates that hyperplasia molecularly resembles normal tissue more closely and shows less immune activation compared to the subsequent low-grade lesions.

1.2.2 Squamous Metaplasia

Squamous metaplasia corresponds to stage 3 in the developmental stages of LUSC, and it is classified under the low-grade lesions along with mild and moderate dysplasia. Squamous metaplasia is part of the low-grade lesions that show an increase in metabolism and early immune sensing (24). There is a biphasic gene-expression trajectory with certain gene modules peaking in low-grade lesions, including metabolism-related genes involved in fatty acid metabolism, oxidative phosphorylation, and the citric acid cycle, showing an increase in expression (24). Moreover, early immune sensing occurs at this stage with activation of resident immune cells. Immune functions are modulated modestly compared to high-grade lesions; specifically, some immune functions are downregulated, such as those

involved in the negative regulation of the immune system, including TNFRSF14, CD200, CD59, TGFB3, and HLA-G, which are downregulated in low-grade lesions (24). There is an increase in proliferation and DNA repair pathways from normal tissue through to cancer, but the more pronounced activation of proliferation pathways is observed progressively through the stages. The upregulation of immune escape mechanisms, such as immune checkpoints and suppressive interleukins, is not prominent yet in squamous metaplasia but becomes significant in high-grade lesions (24).

1.2.3 Dysplasia (Mild, Moderate, Severe)

Dysplasia refers to an abnormal development or growth of cells within tissues, often considered a precancerous condition (24). Dysplasia represents a stage in the progression of pre-invasive lesions in the bronchial epithelium. Mascaux et al. classified dysplasia into mild, moderate, and severe grades, with severe dysplasia being part of the high-grade lesions that precede invasive LUSC. Moreover, dysplasia in the lung bronchial epithelium indicates atypical cellular changes that reflect increasing abnormality and risk for progression toward invasive cancer (24). Cellular abnormalities and polymorphisms increase with severity, with mild dysplasia involving changes in the lower third of the epithelium and severe dysplasia involving alterations throughout the entire epithelial layer. There is a proliferation of basal and basal-like cell populations with increasing atypia (24). Dysplastic lesions show loss of differentiation markers, such as a loss of ciliated cells (loss of FOXJ1+ cells) and secretory cell fates (loss of SCGB1A1+ club cells and MUC5AC+ goblet cells) (25). There is an increase in proliferating squamous cells (KRT13+), proliferating basal cells (TP63+, KRT+, MKI67+), and peri-goblet cell intermediates (CEACAM5+) (25). Epithelial to mesenchymal transition (EMT) increases as lesions progress.

The molecular changes in dysplasia during LUSC development can be divided into low-grade (mild and moderate dysplasia) and high-grade dysplasia (severe dysplasia and carcinoma in situ). Genomic alterations include loss of heterozygosity (LOH) on chromosome arms 3p, 9p, 8p, 13q, and 17p (25). Also, mutations commonly found include TP53, CDKN2A, and SOX2, with TP53 mutations present in nearly all

invasive LUSC tumours (26). Loss-of-function mutations in NOTCH1 and FAT1 are more prominent in dysplasia before advancing to carcinoma in situ (CIS). Moreover, epigenetic dysregulation, such as promoter hypermethylation, occurs; for example, NKX2-1 (TTF1) is hypermethylated and downregulated in progressive CIS lesions compared to regressive ones. Transcriptional alterations include upregulation of cell cycle and DNA replication genes (e.g., MKI67, PLK1) (25), increased oxidative phosphorylation metabolism, suppressed DNA damage response, altered cell adhesion, and increased proliferation. Progressive lesions show reduced expression of genes regulating antitumour immunity, including downregulation of interferon-gamma signalling and antigen presentation (suppression of MHC class I and II) (25). In addition, immune microenvironment changes include decreased infiltration of pro-inflammatory macrophages and CD8+ T lymphocytes, with increased regulatory T cells (e.g., regulatory T cells), tumour-associated macrophages (TAMs), and neutrophils, contributing to immune suppression and evasion (25).

1.2.4 Carcinoma In Situ (CIS)

CIS represents the final pre-invasive stage of LUSC. CIS is a premalignant lesion characterised by extensive cytological changes that resemble carcinomas throughout the epithelial layer, but with an intact basement membrane and no stromal invasion (25). It represents a stage where abnormal cells are present but have not yet invaded surrounding tissues. There is a proliferation of basal and basal-like cells with increasing atypia, loss of ciliated cell fate and secretory cell fates, increased proliferation of squamous cells, proliferating basal cells, and peri-goblet cell intermediates, increasing EMT in precancerous lesions as they progress to CIS and invasive LUSC (25).

In terms of molecular changes in CIS, genetic alterations were found, including mutations in TP53, CDKN2, and SOX2, which are commonly found in CIS lesions, paralleling those mutations seen in invasive LUSC tumours (25). Individuals who smoke often have loss of function alterations in NOTCH1 and FAT1, found both in normal epithelia and dysplastic lesions prior to CIS (25). Epigenetic dysregulation is

evident with progression; for example, NKX2-1 is hypermethylated with reduced expression in progressive CIS lesions compared to regressive ones (25). Transcriptional signatures in CIS demonstrate increased expression of genes involved in cell cycle and proliferation (e.g., MKI67 and PLK1), altered DNA damage response, metabolism, cell adhesion pathways, and chromosomal instability (25). The immune microenvironment in CIS shows increasing immunosuppression with fewer CD8+ T cells, suppressed IFNγ signalling and antigen presentation (via MHC-I and MHC-II), and increased regulatory T cells, tumour-associated macrophages (CD163+ TAMs), and neutrophils (25).

1.2.5 Invasive Carcinoma

Invasive carcinoma is a stage of cancer where malignant cells have breached the basement membrane of the epithelium and invaded the surrounding stromal tissue, distinguishing it from pre-invasive or in situ lesions where the basement membrane remains intact. In invasive carcinoma, the malignant cells show extensive cytological abnormalities throughout the epithelial layer (25). Unlike CIS, invasive carcinoma exhibits stromal invasion beyond the epithelial basement membrane. There is a loss of normal cellular differentiation and organisation. Increased proliferation of malignant cells is evident, often accompanied by morphological changes such as loss of ciliated and secretory cell fates and an increase in squamous and proliferating basal cells in squamous cell carcinoma (25). The tumour microenvironment changes are found, including immune cell infiltration alterations, with increased tumour-promoting inflammation and immune evasion mechanisms (25).

In terms of molecular changes in invasive carcinoma, there is an accumulation of genetic alterations, including mutations, copy number variations, and loss of heterozygosity in key oncogenes and tumour suppressor genes. For LUSC, common genetic alterations include mutations in TP53, CDKN2A, SOX2 amplification (3q gain), and loss-of-function mutations in NOTCH1 and FAT1 (25). For LUAD, mutations in EGFR, KRAS, BRAF, TP53, MAP2K1, and alterations in HLA leading to immune evasion (25). Moreover, epigenetic dysregulation such as promoter hypermethylation

of tumour suppressor genes (e.g., CDKN2A) is frequent (25). Transcriptional signatures show increased cell cycle activity, altered epithelial differentiation, increased metabolism, and suppression of IFNγ signalling and antigen-presentation pathways in progressing lesions (25). The immune microenvironment shifts toward immunosuppression with decreased CD8+ T cells, increased regulatory T cells, and tumour-associated macrophages (25).

While low-dose CT screening shows promise for detecting small peripheral adenocarcinomas at an earlier stage, it is limited in its ability to detect central airway tumours or preinvasive disease (27). Conventional white light bronchoscopy (WLB) also has poor sensitivity for identifying preinvasive lesions (28). Autofluorescence bronchoscopy (AFB), which illuminates the epithelium with blue light, can improve the detection of preinvasive disease because the epithelium shows slightly weaker red fluorescence but proportionally much weaker green fluorescence than normal tissues (29).

AFB-based studies have enabled estimates of preinvasive lesion prevalence. For example, in a cohort of 511 smokers aged 40–74 years with a smoking history of at least 20 pack-years, the prevalence of preinvasive lesions was: mild dysplasia (44%), moderate dysplasia (13%), severe dysplasia (6%), and CIS (1.6%) (30,31). A later study by Ishizumi et al. reported lower rates—moderate dysplasia (9%), severe dysplasia (1.9%), and CIS (0.8%)—in a larger cohort of 1,581 smokers (21). This decline may reflect the introduction of cigarette filters, changes in smoking behaviours, and is consistent with the overall reduction in LUSC incidence worldwide.

Longitudinal surveillance studies of preinvasive lesions have provided insights into how preinvasive disease behaves over time. With rare exceptions, most studies have small numbers of subjects and/or a short follow-up duration (32–34). Definitions of high-grade lesions have varied, and progression rates to LUSC are often difficult to assess, since many high-grade lesions are treated at diagnosis (33,34) or at three months later if CIS persists (32). In a unique cohort of 53 untreated preinvasive lesions (including 36 high-grade disease) followed for 12–85 months, 17% of high-grade

lesions progressed to invasive cancer, 19% regressed spontaneously, and 64% remained unchanged microscopically; notably, none of the low-grade lesions progressed to carcinoma (35).

The largest longitudinal study to date, involving 164 patients, demonstrated that approximately one-third developed lung cancer over ten years (61 cancers detected in 55 patients) with the median time-to-event of 16.5 months (36). Of these, 35 were endobronchial (25 occurring at the previously known index lesion and 10 cancers developed at new endobronchial lesions). Index lesion refers to the disease being present when the patient's airway was first imaged using bronchoscopy. Twenty of the cancers developed in parenchymal region, and three patients developed multiple cancers. Importantly, lesion-specific progression was significantly more likely in patients with high-grade lesions (hazard ratio 1.84, 95% CI 1.05–3.22). This finding has been shown by several studies (32,35). Collectively, longitudinal analyses demonstrate that high-grade lesions are at significant risk of progression to invasive cancer and that they also serve as markers of lung cancer risk elsewhere, or within the airway and lung parenchyma.

1.3 CD4 T cells in Cancer

The field of cancer immunology has emphasised the role of CD8+ cytotoxic T lymphocytes (CTLs) in tumour surveillance and immunotherapy. However, increasing evidence underscores that CD4+ T cells are equally indispensable in orchestrating both innate and adaptive immune responses against tumours (37–39). Apart from their helper functions, CD4+ T cells exhibit direct effector functions, regulate the tumour microenvironment (TME), and can either promote tumour eradication or facilitate immune evasion (40). This dualistic capacity reflects their intrinsic heterogeneity, driven by lineage polarisation, functional plasticity, and dynamic adaptation to the TME (41).

1.3.1 Differentiation and Functional Polarisation

Naive CD4+ T cells recognise antigens presented by APCs via peptide-major histocompatibility complex (MHC) class II molecules (39). Upon recognition, these naïve CD4+ T cells become primed and undergo clonal expansion and differentiation. This differentiation is accompanied by epigenetic changes that lead to the formation of distinct CD4+ helper T cell subsets. Each subset of CD4+ helper T cells is characterised by specific transcription factors and effector functions, and is mainly defined by the production of distinct cytokines (39,41).

The differentiation pathway of CD4 T cells depends on several key factors. Cytokine millieu, the local cytokine environment during priming, is important. Different cytokines promote differentiation into specific CD4+ helper T cell subsets. The strength of the T cell receptor (TCR) stimulation, the intensity of TCR signalling, influences the fate of differentiation (41). The type and functional state of the APC, as well as the different APCs and their activation status, can direct T cell polarisation.

There are six major CD4+ T cell subsets in cancer: Th1, Th2, Th9, Th17, Tregs, and follicular helper T (Tfh) cells (41). Th1 subset is characterised by its production of the cytokine interferon-gamma (IFNγ), along with tumour necrosis factor (TNF) and interleukin-2 (IL-2). Their differentiation is driven by polarising cytokines such as interleukin-12 (IL-12) and IFNγ (37). They express the transcription factor T-bet, which controls their development and function. In cancer, Th1 cells are considered to be beneficial because they activate and maintain antitumour immunity. They support the activation and cytotoxic function of CD8+ T cells, natural killer (NK) cells, and myeloid cells through the secretion of IFNγ and TNF (39). These cytokines not only enhance immune cell functions, but can also have direct effects on tumour and stromal cells. For example, IFNγ and TNF can induce tumour cell death mechanisms such as apoptosis, necroptosis, and ferroptosis, and can modulate the tumour microenvironment by inhibiting angiogenesis and normalising tumour vasculature, which restricts tumour growth (41). Moreover, Th1 cells also provide important help in the priming phase of anti-tumour immunity by licensing APCs, especially classical type

1 dendritic cells (cDC1) via CD40-CD40L interactions to efficiently activate cytotoxic CD8+ T cells (42). Their effector cytokines can act both locally at the immunological synapse and at a distance within the TME, broadly reprogramming immune cells and enhancing tumour control. However, Th1 cells can also have some unfavourable effects. Th1-derived cytokines, such as IFNγ and TNF, can induce the upregulation of immunosuppressive molecules on tumour cells and other cells within the TME (41). Among these molecules is programmed cell death ligand 1 (PD-L1), which interacts with PD-1 on T cells to inhibit their activity, thereby dampening the immune response (43). Moreover, IFNγ and TNF can upregulate the expression of MHC class I and II molecules and chemokines (e.g., CXCL9), which have immunostimulatory effects but also contribute to modulating the immune landscape in complex ways (44).

The Th2 subset of CD4+ T cells is characterised by differentiation under the influence of the polarising cytokine IL-4 and expression of the transcription factor GATA-3 (45). These cells produce effector cytokines IL-4, IL-5, and IL-13 (45). In cancer, Th2 cells have a complex role with both beneficial and unfavourable effects. In terms of beneficial effects, they can induce in situ inflammatory immune responses (45). However, they also support pro-tumour properties of macrophages, which can help tumour progression (45).

The Th9 subset of CD4+ T cells is characterised by its differentiation under the influence of polarising cytokines IL-4 and TGFβ and expression of the transcription factor Pu.1 (46). These cells produce the effector cytokine IL-9 (46). In cancer, Th9 cells have a complex role with both beneficial and unfavourable effects. They can support cytotoxic CD8 T cells and NK cells and enhance anti-tumour immunity by recruitment of DCs (47). However, they promote survival and proliferation of tumour cells and support immunosuppression by Tregs (41).

Moreover, the Th17 subset is characterised by its differentiation under the influence of polarising cytokines such as IL-6, IL-1β, and TGFβ, and they express the transcriptional factor ROR-γt (48). Their main effector cytokines include IL-17 and IL-22 (48). In cancer, Th17 cells promote DC recruitment and support CD8+ T cell-mediated anti-tumour immunity. Both Th1 and Th17 cells have been shown to promote tumour regression (41). However, Th17 cells can also support tumour progression and

metastasis, as well as promote tumour angiogenesis, which facilitates tumour growth (48).

The Tfh subset of CD4+ T cells is specialised in providing help to B cells (49). They are characterised by expression of the transcriptional factor BCL-6 and produce the effector cytokine IL-21 (49). In cancer, Tfh cells support the anti-tumour B cell response and promote the formation of ectopic lymphoid structures, which are associated with positive clinical outcomes in several cancers. However, Tfh cells can also have unfavourable functions, such as contributing to immunosuppression and supporting B cell lymphomas. In Hodgkin lymphoma, for example, a type of CD4+ T cells with Tfh-like features, producing CXCL13, form rosettes around malignant Hodgkin-Reed-Sternberg cells, potentially supporting tumour growth (50). In addition, Tfh cells can proliferate upon immune checkpoint blockade therapy and may suppress anti-tumour T cell immunity, limiting the efficacy of such treatments (51).

The Treg subset is characterised by the expression of the transcription factor FOXP3 (52). They play a critical role in maintaining immune tolerance and preventing autoimmunity by exerting anti-inflammatory functions (52). However, in the context of cancer, Tregs often have an immunosuppressive effect that can promote tumour progression (53). Tregs differentiate under the influence of cytokines such as TGFβ and IL-2, with FOXP3 as their transcription factor (52). They produce TGFβ, which has anti-inflammatory and immunosuppressive effects. In tumours, Tregs suppress anti-tumour immune responses, which can allow tumours to grow and metastasise (54). They contribute to creating a tolerogenic environment that inhibits effective immune surveillance against cancer cells.

1.3.2 CD4+ T Cells in Tumour-Draining Lymphoid Tissues

The immune response begins in the tumour-draining lymph nodes, specialised secondary lymphoid organs where naive CD4+ T cells continuously scan for their cognate antigens presented by APCs (42). Tumour-associated antigens (TAAs) reach these lymphoid tissues either through the uptake and migration of DCs that have captured tumour antigens and the Drainage of soluble tumour antigens into the lymph nodes (42). Within the lymph nodes, lymph node-resident DCs present these TAAs on MHC class II molecules to naive CD4+ T cells, leading to their priming, clonal expansion, and differentiation (42). One of the critical roles of CD4+ T cells in the lymphoid tissues is to license APCs, particularly classical type 1 dendritic cells (cDC1), to effectively prime CD8+ cytotoxic T lymphocytes (CTLs) (42). This licensing occurs via CD40 ligand (CD40L) expressed on activated CD4+ T cells, engaging CD40 on DCs. Licensed DCs upregulate adhesion molecules, costimulatory molecules (CD80, CD86, CD70), and produce key cytokines such as IL-12 and IL-15, which are essential for robust activation and differentiation of CD8+ T cells (42). IL-12, for example, is important for equipping CD8+ T cells with the ability to produce IFNγ, a key anti-tumour cytokine. The CD40–CD40L interaction is also crucial for B cell activation by Tfh cells, supporting humoral responses that can contribute to anti-tumour immunity (55).

CD4+ T cells also help support B cell activation, which can present tumour antigens on MHC class II molecules, especially after B cell receptor (BCR)-mediated uptake of tumour antigens (55). Activated B cells can stimulate cytotoxic lymphocyte responses, as seen in the Epstein–Barr virus (EBV)-associated tumours (55). Tertiary lymphoid structures (TLSs), which are lymph node-like structures formed within tumours, often contain activated B cells and CD4+ T cells. The presence of TLSs correlates with better clinical outcomes in several cancers, including colorectal carcinoma, lung squamous cell carcinoma, and breast cancer. CD4+ T cells within TLSs contribute to anti-tumour immunity by helping B cells and supporting the local immune response (56).

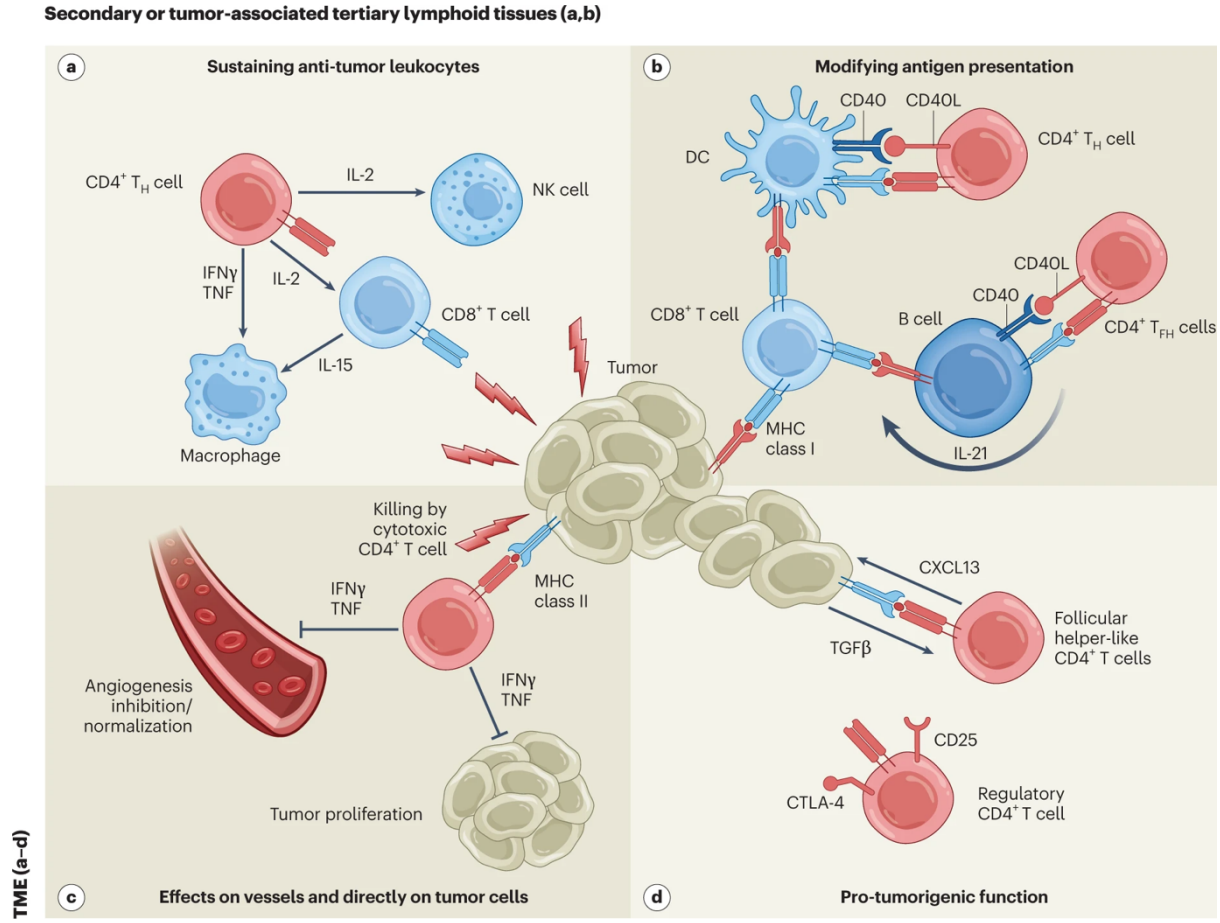

Figure 1.2 - The functional flexibility of CD4+ T cells shapes both innate and adaptive immunity within the tumour microenvironment (TME), tumour-draining lymph nodes, and tertiary lymphoid structures (TLSs). a) Through secretion of cytokines such as IL-2, IFN-γ, and TNF, CD4+ T cells provide essential support to other anti-tumour immune cells, including NK cells, CD8+ T cells, and myeloid populations. b) In surrounding lymphoid tissues, they also regulate antigen presentation by dendritic cells and B cells, performing canonical T helper (TH) and Tfh roles through molecules such as CD40L and IL-21. c) Within the TME itself, CD4+ T cells can restrain tumour growth by producing effector cytokines and exerting direct cytotoxic activity, impacting not only tumour cells but also local vasculature. d) Conversely, CD4+ T cells can facilitate tumour progression, particularly through Treg and Tfh functions, by depleting IL-2, limiting antigen presentation via CTLA-4, and providing aberrant help to B cell lymphomas. Figure adapted from Speiser et al. 2023

1.3.3 Pro-Tumorigenic CD4+ T Cell Subsets

Pro-tumorigenic CD4+ T cell subsets primarily include Tregs and Tfh cells that can support tumour growth and persistence (41). Treg cells are characterised by expression of *FOXP3, CD25, CTLA-4*, and other markers such as *ICOS, OX40, 4-1BB, LAG3, TIGIT, GITR*, and *PD-1* (57). They have immunosuppressive and anti-inflammatory functions that can inhibit effective anti-tumour immunity. Treg cells often correlate with poor prognosis in many cancers (e.g., gastrointestinal tumours, ovarian carcinoma, breast carcinoma) when they outnumber CD8+ T cells (52,57,58). They may be induced either in the thymus against self-antigens present in the tumour tissue or in tumour-draining lymph nodes against tumour antigens. Treg cells can promote tumour growth and metastasis and may directly increase the survival of metastatic tumour cells. Their suppressive functions include depletion of IL-2 and interference with CD80–CD86-dependent T cell priming, which compromises effective anti-tumour immune responses (57). Tumour antigen-specific CD4+ T cell priming can also lead to anergy, a state that can predispose cells to Treg development. High expression of chemokine receptors CCR4 and CCR8 on Treg cells is reported in various human tumours, such as breast and colorectal cancers (58). In some studies, Tfh-like CD4+ T cells can have pro-tumorigenic functions, including support for B cell lymphomas. For example, in classic Hodgkin lymphoma, the TME is dominated by TGFβ, which induces CXCL13-producing CD4+ T cells with Tfh features that rosette around malignant Hodgkin–Reed–Sternberg cells, thereby nurturing the lymphoma (50). Moreover, Tfh cells may also contribute to immunosuppression in the TME. Therefore, while CD4+ T cells have many anti-tumour roles, certain subsets like Treg cells and some Tfh-related populations can promote tumour development by suppressing effective immune responses and supporting tumour survival and growth (41,50).

1.4 Regulatory T Cells in Cancer

Tregs are an immunosuppressive subset of CD4+ T cells that are indispensable for immune tolerance and homeostasis. They are characterised by expression of the transcription factor forkhead box protein P3 (FOXP3) (59), Tregs suppress

autoreactive and excessive immune responses that would otherwise culminate in autoimmunity, allergy, or transplant rejection. Their importance is underscored by the observation that loss-of-function mutations in FOXP3 in humans cause the severe autoimmune condition immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome, while Foxp3-deficient scurfy mice develop fatal systemic autoimmunity (59,60). Moreover, Tregs suppress various immune effector cells, including CD4+ T helper cells and CD8+ cytotoxic T cells.

In cancer, tumours exploit the suppressive functions of Tregs to evade immune surveillance, restrain effector T cell cytotoxicity, and create a tolerogenic TME (61). High infiltration of Tregs has been consistently associated with poor prognosis in most solid tumours, including melanoma, NSCLC, ovarian, and gastric carcinomas (62). By contrast, their prognostic impact in colorectal cancer (CRC) remains more nuanced due to the presence of FOXP3+ non-Treg subsets with divergent functional programs (63).

The pivotal role of Tregs in cancer immunology is further highlighted by their intimate interaction with other immunosuppressive networks, including myeloid-derived suppressor cells (MDSCs), tumour-associated macrophages (TAMs), and inhibitory immune checkpoints such as CTLA-4 and PD-1 (64,65). Together, these suppressive elements form the cornerstone of tumour immune escape. As a result, Tregs have become attractive targets for therapeutic intervention, particularly in the context of immune checkpoint blockade (ICB) and combination immunotherapies (66).

1.4.1 Development and Subsets of Tregs

1.4.1.1 Origin of Tregs: Thymic vs Peripheral Lineages

Tregs arise from two distinct developmental routes based on their developmental origin (Fig. 1.3): thymus-derived Tregs (tTregs, also termed natural Tregs) and peripherally induced Tregs (iTregs) (67). tTregs are generated in the

thymus from CD4+CD8- thymocytes that engage self-peptide–MHC class II molecules on thymic APCs with relatively high-affinity TCR binding to self-peptide-MHC complexes (68,69). Unlike thymocytes that undergo deletion via negative selection due to excessive self-reactivity, a fraction is instead diverted into the Treg lineage (67,69). This process depends on IL-2 receptor (IL-2R) signalling, STAT5 activation, and stabilisation of FOXP3 transcription through epigenetic modifications, particularly demethylation of conserved non-coding sequence 2 (CNS2), also known as the Treg-specific demethylated region (TSDR) (70). Co-stimulatory signals delivered via CD28 also contribute critically to lineage commitment by reinforcing FOXP3 expression and establishing stable suppressive function (71).

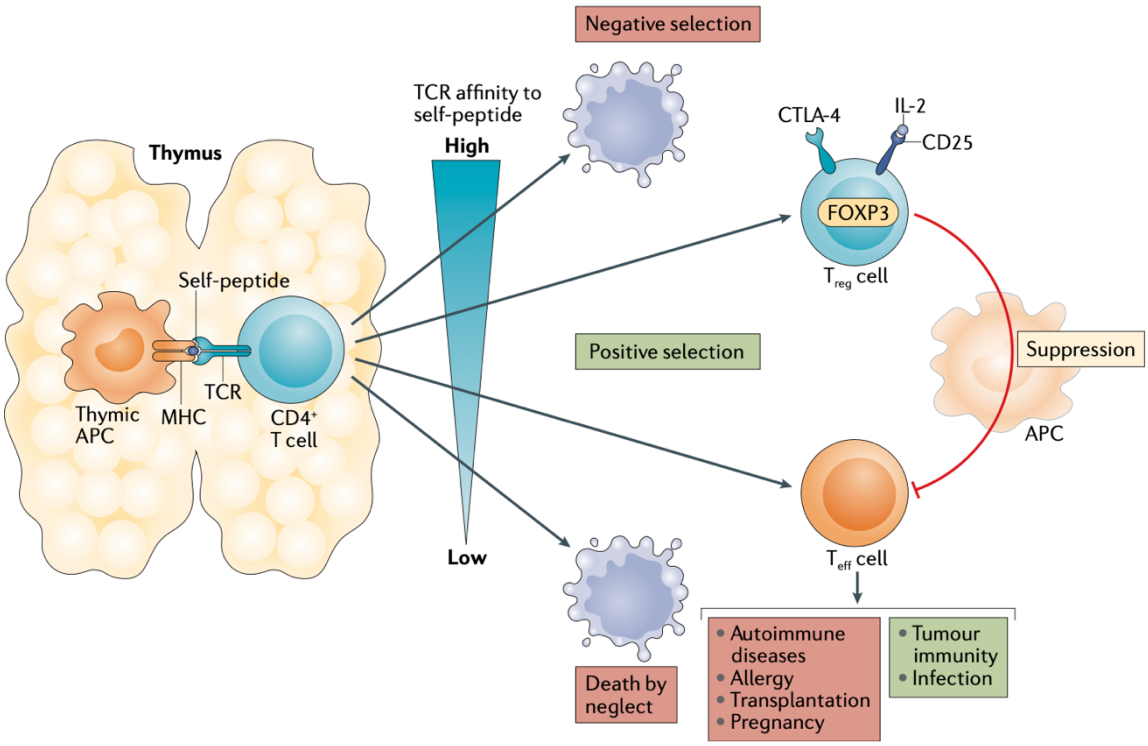

Figure 1.3 - The development of regulatory T (Treg) cells begins in the thymus, where the classical, natural Treg subset arises. Their differentiation is initiated by T cell receptor (TCR) signals generated when CD4+ thymocytes recognise self-peptides presented on MHC class II molecules of thymic antigen-presenting cells. Among CD4+CD8- thymocytes, those with intermediate-to-high affinity binding to self-peptide–MHC complexes are positively selected. When TCR stimulation reaches a sufficiently strong threshold, these precursors upregulate the IL-2 receptor α-chain (CD25), engage in IL-2 signalling, and induce expression of the transcription factor FOXP3, which programs their commitment to the Treg lineage. By

contrast, thymocytes with low-affinity interactions with self-peptide-MHC survive but develop into conventional effector T (Teff) cells, whereas those with excessively high-affinity recognition are deleted through negative selection to prevent severe autoimmunity. Cells with too little affinity fail to receive TCR signals and are eliminated by death by neglect. Although Teff cells generally maintain tolerance, some retain self-reactivity or alloreactivity, which may underlie autoimmune disease, pregnancy complications, or graft rejection. Treg cells play a critical role in suppressing these potentially harmful immune responses. However, this same regulatory capacity can limit beneficial immunity: Treg cells may dampen Teff responses against cancers that express self-antigens. Molecules such as CTLA-4 further contribute to this suppressive activity. Figure adapted from Togashi, Shitara, and Nishikawa (2019).

However, iTregs are generated in the periphery under conditions of chronic antigen exposure and tolerogenic cues. Their induction often requires TCR stimulation in the presence of TGFβ or retinoic acid, with microbial metabolites and specific cytokine environments serving as additional modulators (72). FOXP3 induction in human conventional CD4+ T cells (Tconv) by TGFβ in vitro often yields unstable populations that secrete inflammatory cytokines rather than acquire durable suppressive functions (73). Thus, although iTregs may contribute to tolerance at mucosal barriers or in chronic infection, the majority of tumour-infiltrating Tregs in humans are thought to be of thymic origin (74).

1.4.1.2 Epigenetic and Transcriptional Basis of Treg Lineage Stability

The stability of the Treg lineage is indeed tightly linked to the epigenetic landscape of the FOXP3 locus. Specifically, the conserved non-coding sequence 2 (CNS2) enhancer, also known as the Treg cell-specific demethylation region, is essential for maintaining stable FOXP3 expression through cell divisions, thereby ensuring lineage stability (70). FOXP3 functionally enforces immune suppression only when stably and epigenetically fixed; transient FOXP3 expression reflects activation-associated regulation and lacks durable suppressive activity. Genome-wide analyses have identified other loci, such as Ctla4 exon 2 and Ikzf4 (EOS) intron 1 in mice, that are

specifically demethylated in thymus-derived Tregs (tTregs) but not in conventional T cells, providing markers that distinguish bona fide Tregs from activated conventional T cells (75). These epigenetic modifications act as locks that maintain the suppressive phenotype of true Tregs even in inflammatory environments. However, iTregs often exhibit incomplete demethylation at these loci, particularly CNS2, which may lead to instability of FOXP3 expression (76). This instability can cause iTregs to lose FOXP3 expression and adopt effector-like functions, rather than maintaining their suppressive activity. Therefore, the epigenetic status at these key regulatory regions underpins the functional stability and identity of Treg cells.

1.4.1.3 Phenotypic Classification of Human FOXP3^{+}CD4^{+} T Cells

In humans, FOXP3 expression alone does not identify suppressive Tregs because activated conventional (non-Treg) CD4+ T cells can transiently upregulate FOXP3 without acquiring immunosuppressive function. A functional classification based on the expression levels of CD45RA (a marker of naïve T cells) and FOXP3 is used to better distinguish Treg subsets. This classification divides FOXP3+ CD4+ T cells into three fractions: naïve Tregs, effector Tregs, and non-Treg FOXP3+ cells (77).

Naïve Tregs are defined by the expression of CD45RA, with low expression of FOXP3 and low expression of CD25. Naive Tregs are cells that have recently egressed from the thymus and have not yet been activated in the periphery (67). They are considered to be tTregs that have undergone positive selection in the thymus based on high-affinity interactions between their TCRs and self-peptide-MHC complexes presented by thymic APCs (60). Naive Tregs possess weak immunosuppressive activity compared with their effector counterparts. Upon TCR stimulation in the periphery, naive Tregs proliferate rapidly and differentiate into effector Tregs, which have high FOXP3 expression and strong immunosuppressive functions (77). These cells play a crucial role in maintaining self-tolerance and preventing autoimmune disease by suppressing autoreactive T cells (67). They serve as a precursor pool for the generation of the more functionally active effector Treg

cells. The naive Treg population is distinct from FOXP3+ non-Treg cells, which have low FOXP3 and CD25 expression but lack suppressive activity and instead produce inflammatory cytokines like IFNγ and IL-17 (77).

Effector Tregs (eTregs) are a subpopulation of Tregs characterised by a highly immunosuppressive phenotype and function. They are different from naïve Tregs and non-Tregs based on specific markers and functional properties. eTregs are characterised by the expression profile CD45RA- FOXP3hiCD25hi CD4+ (67). Moreover, they express high levels of immunosuppressive molecules such as *CTLA-4, PD-1, LAG3, TIM3, ICOS, TNFRSF4, TNFRSF18*, and are positive for *MKI67* (indicating active proliferation) (78). They also express the apoptosis marker FAS, suggesting they are prone to apoptosis (67). eTregs produce immunosuppressive cytokines (e.g., IL-10 and TGFβ) (79). eTregs differentiate from naïve Tregs (CD45RA+ FOXP3lo CD25lo CD4+) upon TCR stimulation (77). Naïve Tregs have weak immunosuppressive activity but upon activation, they rapidly proliferate and differentiate into eTregs with strong suppressive capacity. eTregs have potent immunosuppressive effects on effector T cells, including CD8+ cytotoxic T cells and CD4+ helper T cells (67). They suppress immune responses via multiple mechanisms, such as consumption of IL-2, CTLA-4-mediated suppression of APCs, secretion of immunosuppressive cytokines (IL-10, IL-35, TGFβ), conversion of ATP into adenosine via CD39 and CD73 (80), and direct killing of effector cells through granzyme and perforin (81). eTregs are enriched in the TME, often comprising 10-50% of CD4+ T cells in tumours, compared with 2-5% in peripheral blood (67). Moreover, their accumulation in tumours is associated with poor prognosis in various cancers such as melanoma, NSCLC, gastric, and ovarian cancers (67). They exhibit an activated phenotype distinct from peripheral tissue Tregs, supporting their strong immunosuppressive capacity within tumours. Apoptotic eTregs in the TME can have even stronger immunosuppressive effects due to their conversion of ATP to adenosine, further inhibiting local immune responses.

FOXP3+ non-Tregs are a distinct subpopulation within the broader group of FOXP3-expressing CD4+ T cells in humans (77). Unlike classical Tregs, which are characterised by immunosuppressive functions, FOXP3+ non-Tregs do not exhibit

immunosuppressive capacity (67). FOXP3+ non-Tregs are characterised as CD45RA-FOXP3lo CD25lo CD4+ T cells. They do not suppress immune responses. Instead, they produce inflammatory cytokines such as IL-2, IFNγ, and Il-17, which are typically associated with immunostimulatory rather than immunosuppressive functions (67). This population is heterogeneous, meaning it contains various subsets with potentially different roles and characteristics. FOXP3 expression in non-Tregs can be induced transiently in conventional CD4+ T cells upon TCR stimulation, especially in the presence of TGFβ (73). However, this FOXP3 expression is unstable and does not confer regulatory function. Therefore, FOXP3+ non-Tregs are not considered true Tregs because they lack the stable immunosuppressive phenotype characteristic of tTregs or iTregs with immunosuppressive activity. FOXP3+ non-Tregs predominate in autoimmune diseases, where their lack of suppressive function and production of inflammatory cytokines may contribute to pathology (61). In addition, FOXP3+ non-Tregs have been detected in the TME, for example, in CRC, where their presence correlates with a favourable prognosis, possibly due to their immunostimulatory nature (82). Distinguishing FOXP3+ non-Tregs from effector Tregs in tumour tissues is challenging, but important for understanding their respective roles in cancer immunity.

1.4.2 Mechanisms of Treg-Mediated Immunosuppression

Tregs mediate immunosuppression through multiple intricate cellular and molecular mechanisms that collectively dampen immune responses, particularly those of effector T (Teff) cells. (i) IL-2 consumption and cytokine deprivation, Treg cells constitutively express high levels of the interleukin-2 receptor subunit alpha (CD25), which forms part of the high-affinity IL-2 receptor (83). Although Treg cells produce little IL-2 themselves, they are highly dependent on it for survival. By binding and consuming IL-2 from the microenvironment, Treg cells effectively reduce the availability of IL-2 to Teff cells, limiting their activation and proliferation (83). (ii) CTLA-4 Mediated Suppression of APCs,
CTLA-4 is constitutively expressed on Treg cells and plays a pivotal role in immunosuppression. CTLA-4 binds to the co-stimulatory molecules CD80 and CD86

(also known as B7-1 and B7-2) on APCs with higher affinity than the activating receptor CD28 on Teff cells (78). This binding inhibits the maturation and antigen-presenting capacity of APCs by transmitting inhibitory signals, thereby reducing their ability to prime and activate Teff cells. Moreover, CTLA-4 can physically remove CD80/CD86 molecules from APCs through a process called trogocytosis, further diminishing co-stimulatory signals needed for effective Teff cell activation (78). (iii) production of immunosuppressive cytokines

Tregs secrete various inhibitory cytokines that suppress immune responses, including IL-10 (79), which limits APC activation and inflammatory responses, TGF-β, which can suppress Teff cell function and promote immune tolerance, and IL-35 (84), which contributes to Treg-mediated suppression and inhibits Teff cells. (iv) Metabolic disruption via adenosine production, Treg cells express ectonucleotidases CD39 and CD73 (80), which convert extracellular ATP, often released during tissue damage or inflammation, into adenosine. Adenosine engages the adenosine A2A receptor on Teff cells and APCs, resulting in immunosuppressive signalling that inhibits Teff cell activation and cytokine production (80). (v) cytotoxic through granzyme and perforin secretion, Tregs can directly kill target immune cells by releasing cytotoxic molecules such as granzyme and perforin (81). This mechanism leads to the apoptosis of Teff cells or other immune cells involved in the anti-tumour response. These mechanisms together enable Treg cells to maintain immune homeostasis by preventing excessive or autoreactive immune responses, but can also hinder effective anti-tumour immunity by suppressing the activity of tumour-specific effector T cells.

1.4.3 Accumulation and Characteristics of Tregs in the TME

1.4.3.1 Characteristics of tumour-infiltrating Tregs (TIL-Tregs)

TIL-Tregs typically display an activated phenotype distinct from peripheral Tregs, characterised by high expression of *CTLA-4, TIGIT, LAG3, TIM3, ICOS, TNFRSF4*, and *TNFRSF18* (53). These markers indicate strong immunosuppressive capacity. TIL-Tregs produce cytokines such as IL-10, TGF-β, and IL-35, which contribute to the suppression of effector T cell and APC functions (58). Moreover, TIL-

Tregs are proliferative (MKI67+) but also prone to apoptosis due to expression of death receptors like FAS (67). Apoptotic Tregs retain or even enhance their immunosuppressive functions. They have a distinct and stable epigenetic profile characterised by DNA hypomethylation of genes like *FOXP3, CTLA4*, and *IKZF4 (EOS)*, which maintain lineage stability and suppressive function (85). However, the detailed epigenomic landscape of TIL-Tregs remains largely unexplored and requires further investigation. In addition, the TCR repertoire of TIL- Tregs is skewed toward self-antigens and shows clonal expansion, which differs from conventional T cells that are more diverse (67). This suggests that Tregs in tumours recognise specific self-antigens and contribute to immune tolerance within the TME. They also suppress antitumour immunity by inhibiting effector T cells and APCs, limiting immune surveillance and enabling tumour progression. A high ratio of Tregs to conventional T cells in the TME generally correlates with poor prognosis in many cancers, although exceptions exist, such as certain colorectal cancers where non-suppressive FOXP3+ non-Tregs might predominate (63).

Dykema et al. explored heterogeneity with Treg populations using scRNAseq coupled with TCRseq on Tregs isolated from resected human lung cancers and adjacent normal lung tissue (86). Samples were obtained from both treatment-naïve patients and those treated with anti-PD1 therapy. They also studied a murine tumour model, which allowed for the tracking of tumour-associated antigen (TAA)- specific Tregs. Integration of these datasets revealed multiple distinct Treg subsets with divergent functions and transcriptional profiles, including an activated, highly suppressive subset marked by high expression of *TNFRSF4 (OX40)* and *TNFRSF18 (GITR)*, along with genes implicated in Treg suppressive function such as *LAG3, ENTPD1*, and *EBI3* (86). This OX40hi GITRhi Treg subset represented about 20% of total TIL-Tregs and was significantly enriched in tumours compared to adjacent normal lung tissue (86). Functional assays demonstrated that the OX40hi GITRhi Tregs are markedly more suppressive ex vivo than other Treg subsets. Furthermore, this activated subset was enriched in tumours from patients who did not respond to PD-1 blockade, suggesting a role in mediating resistance (86). RNA velocity and TCR clonotype sharing analyses indicated that in nonresponding tumours, the Tregs differentiate into and maintain this activated, suppressive state, whereas in responding tumours, there is a differentiation flux away from this subset (86).

Moreover, Shan et al. investigated TIL-Tregs within the TME, focusing on their gene expression programs, functional states, and regulatory networks, particularly in head and neck squamous cell carcinoma (HNSCC) but also extending to other solid tumours, including lung cancer (87). They performed analysis of scRNAseq of CD4+ T cells from tumours and peripheral blood of patients with HNSCC, as well as from tonsil tissues and healthy donor blood. They identified a subpopulation of highly activated intratumoural Tregs that express multiple tumour necrosis factor receptor (TNFR) family genes such as *TNFRSF4, TNFRSF9*, and *TNFRSF18* (87). They called these cells TNFR+ Tregs. These TNFR+ Tregs were highly enriched in the TME compared to non-tumour tissues and peripheral blood, indicating a tumour-restrict phenotype (87). Moreover, the presence and enrichment of TNFR+ Tregs in tumours correlated with worse progression-free survival in patients with HNSCC (87). This association was also confirmed in other solid tumour types, including NSCLC and melanoma, indicating that these highly suppressive Tregs are clinically relevant across cancers. In addition, through gene regulatory network (GRN) analysis using SCENIC and causal modelling (mixed graphical models and FCI-MAX), the transcription factor BATF emerged as a central regulator of the TNFR+ Treg transcriptional program (87). BATF was connected to multiple key genes involved in Treg activation, survival, migration, and suppressive function, including costimulatory/inhibitory receptors (*TNFRSF4, TNFRSF9, ICOS, CTLA4*), chemokine receptor *CXCR6*, and apoptosis regulators (*CFLAR, PIM3*) (87).

In addition, Itahashi et al. investigated TIL-Tregs in the TME of human lung cancers, particularly NSCLC, focusing on their epigenetic profiles, differentiation mechanisms and functional roles (88). Using ATAC-seq, they showed that TIL-Tregs in the TME have a unique and distinct open chromatin profile compared with other T cell subsets, including peripheral Tregs, conventional CD4+ T cells, and CD8+ T cells. This distinct epigenetic landscape indicates that TIL-Tregs undergo dynamic chromatin remodelling during differentiation and activation in the tumour setting. Motif enrichment and footprint analyses revealed several transcriptional factors important for the differentiation and activation of Tregs in the TME, including BATF, IRF4, NF-kB (REL), NR4A family members, NFAT, MEF, SOX4, and TBX21 (88). Among these, BATF was highlighted as a key epigenetic regulator that preferentially binds to

chromatin regions specific to TIL-Tregs and upregulates their accessibility, promoting the activated phenotype of Tregs in the TME. Moreover, their scRNA-seq with ATAC-seq analyses showed that BATF expression is upregulated early during Treg differentiation in tumours and tissues, preceding the increase of other transcription factors involved in later activation stages (88). They also studied BATF deficiency in Tregs in mouse tumour models. They found a significant reduction in Treg numbers within tumours, decreased expression of activation and suppressive markers (e.g., *CTLA4, ICOS, GITR, OX40, PD-1, TIGIT, CCR8*), and resulted in suppressed tumour growth (88). This suggests that BATF is a crucial transcription factor that epigenetically and transcriptionally programs TIL-Tregs, enabling their accumulation, activation, and suppressive function (88). Moreover, their findings also highlight BATF as a promising therapeutic target to modulate Treg-mediated immunosuppression in cancer.

All of these studies indicate that TIL-Tregs or BATF+Tregs are established in invasive cancers (e.g., NSCLC, HNSCC). Moreover, this raises the question of whether BATF+Tregs are also present in pre-invasive cancers, particularly in lung cancer.

1.4.3.2 Accumulation of TIL-Tregs in the TME

Tumours and associated stromal cells produce various chemokines that mediate Treg chemotaxis via their respective receptors, and different chemokine-receptor axes are involved depending on tumour type: (i) CCL22–CCR4: a dominant pathway in many solid tumours, with CCL22 produced by tumour-associated macrophages (TAMs) and dendritic cells (89). (ii) CCL17–CCR4: enriched in certain tumours, also produced by myeloid cells (89). (iii) CCL28–CCR10: implicated in hypoxic tumours such as ovarian carcinoma (90). (iv) CCL5–CCR5: widely expressed across inflamed tumours (91). (v) CCL1–CCR8: an emerging axis particularly relevant to TIL-Tregs. Moreover, dysfunctional or exhausted CD8+ T cells within tumours can produce chemokines like CCL1 and CCL22, thereby contributing Treg recruitment and reinforcing immune suppression (92). This overlap in chemotactic signals presents a

challenge in selectively targeting Treg infiltration without affecting the recruitment of conventional effector T cells.

Moreover, tumour-intrinsic factors, including driver gene alterations such as EGFR mutations, MYC activation, and aberrations in the WNT–β-catenin pathway, influence the recruitment and accumulation of regulatory T (Treg) cells in the tumour microenvironment (TME) (93). Activating alterations in EGFR drive MAPK/PI3K-STAT3 signalling that induces CCR-binding chemokines (e.g., CCL22/CCL17) and TGF-β**,** promoting preferential Treg recruitment and stabilisation of FOXP3 expression while limiting effective CD8+ T-cell infiltration. MYC activation amplifies an immunosuppressive transcriptional and metabolic state characterised by increased IL-10, TGF-β, PDL1, reduced antigen presentation, and nutrient competition, conditions under which metabolically flexible, FOXP3hi Tregs preferentially survive and acquire heightened suppressive capacity. Aberrant WNT–β-catenin signalling, in contrast, establishes an immune-excluded phenotype by suppressing dendritic-cell-recruiting chemokines and impairing T-cell priming, thereby skewing the effector-to-Treg balance toward functional dominance of Tregs even without large numerical expansion. Together, these tumour-intrinsic drivers converge to create a TGF-β-rich, low-inflammation environment that not only attracts FOXP3+ cells but reinforces their epigenetic stability and suppressive identity, enabling durable immune tolerance within the tumour. Moreover, loss of tumour suppressors like PTEN or LKB1 also promotes an immunosuppressive environment, favouring Treg dominance (94). Furthermore, hyperactivation of focal adhesion kinase (FAK) in tumour cells drives the production of chemokines such as CCL5, which recruits Treg cells and is associated with exclusion or exhaustion of CD8+ T cells, leading to increased tumour growth and metastasis (95). These oncogenic pathways thus remodel the immune contexture to support immune tolerance and tumour progression.

1.4.4 Antigen Specificity and TCR Repertoire of TIL-Tregs

TIL-Tregs exhibit a TCR repertoire that is clonally expanded and predominantly skewed towards recognition of self-antigens, reflecting their thymic origin. This repertoire is largely distinct from that of conventional effector T cells (Tconv cells), which display a higher diversity and are more focused on tumour-specific neoantigens (96). The preferential recognition of self-antigens by Tregs confers a competitive advantage in the tumour microenvironment, enabling them to be preferentially activated and expanded upon encountering abundant self-peptides presented by dying or proliferating tumour cells (97). Consequently, tumours expressing shared self-antigens can more effectively recruit and activate Tregs, facilitating immune evasion by suppressing antitumour immune responses (97). These findings suggest that the immunosuppressive activity of Tregs is more pronounced against T cells targeting shared self-antigens than those targeting neoantigens

Thymus-derived regulatory T (tTreg) cells are intrinsically skewed toward recognising self-antigens due to their development being driven by high-affinity interactions between their T cell receptors (TCRs) and self-peptide–MHC class II complexes presented by thymic antigen-presenting cells (69). This self-reactivity bias is maintained in the periphery and is particularly pertinent within the tumour microenvironment, where tumours frequently overexpress or aberrantly express self-antigens (98). As a result, tumour-infiltrating Treg cells are thought to preferentially recognise shared self-antigens, including differentiation antigens, cancer-testis antigens, and overexpressed tissue-restricted proteins (67).

This predominance of self-antigen recognition by Treg cells has important clinical implications. Effector T cells (Teffs) specific for shared self-antigens tend to be more susceptible to suppression by Treg cells. Treg cells induce anergy in these Teffs by modulating antigen-presenting cells (APCs) to provide limited co-stimulatory signals, thereby dampening their activation (99). In contrast, Teffs targeting neoantigens—non-self antigens arising from oncogenic mutations or viral proteins—are relatively resistant to Treg-mediated immunosuppression in both animal models and human studies (67). This differential susceptibility helps explain why immune

checkpoint inhibitors (ICIs) demonstrate greater efficacy in tumours with a high mutational burden, where neoantigens are abundant and drive more robust antitumour immune responses (67). Overall, the preferential recognition of self-antigens by tumour-infiltrating Treg cells contributes to the selective suppression of shared antigen-specific Teffs, thereby influencing the effectiveness of immunotherapies targeting neoantigen-specific responses.

Tregs play a crucial role in maintaining immune homeostasis by suppressing autoreactive immune responses against self-antigens (61). Tumours exploit this physiological function of Tregs to establish an immunosuppressive microenvironment that facilitates immune evasion (53). Specifically, TIL-Tregs preferentially recognise and expand in response to shared self-antigens, which are endogenous proteins abnormally or highly expressed by tumour cells (67). This selective expansion enables Tregs to suppress effector T cells that target these shared antigens, effectively silencing or anergising tumour-reactive T cells that would otherwise mediate antitumour immunity (67). This mechanism of immune suppression is particularly problematic in tumours with a low mutational burden, where neoantigens—non-self antigens derived from somatic mutations or oncogenic viral proteins—are scarce. In such tumours, the antigenic landscape is dominated by shared self-antigens, making Treg-mediated suppression more effective and comprehensive. Animal model studies suggest that CD8+ T cells recognising self-antigens are rendered anergic due to limited co-stimulatory signalling from antigen-presenting cells (APCs), a process controlled by Tregs. Conversely, CD8+ T cells specific for neoantigens appear more resistant to Treg-mediated suppression, highlighting the differential impact of Tregs on antitumour responses depending on antigen specificity (100). Consequently, Treg-mediated immunosuppression poses a significant barrier to effective antitumour immunity, particularly in cancers with low neoantigen loads. This contributes to resistance against immune-checkpoint inhibitors (ICIs) such as anti-PD-1 and anti-CTLA-4 therapies (101), which otherwise aim to reinvigorate exhausted effector T cells. The predominance of Treg suppression in tumours may underlie the poor clinical outcomes and limited therapeutic efficacy observed with ICIs. Therefore, understanding the preferential expansion of Tregs against shared self-antigens underscores the importance of developing therapeutic strategies that selectively target Tregs within the tumour microenvironment. Such approaches could relieve the

suppression of tumour-specific effector T cells, especially in tumours with low mutational burdens, thereby overcoming a critical mechanism of tumour immune evasion and enhancing responses to immunotherapy.

1.4.5 Tregs and Immune Checkpoint Therapy

Because of their central role in enforcing immunosuppression within the TME, Tregs are an attractive therapeutic target. However, systemic depletion of Tregs carries a substantial risk of autoimmunity, reflecting their essential role in maintaining peripheral tolerance (67).Therefore, successful strategies must selectively target tumour-infiltrating eTregs while sparing peripheral or naïve Tregs. Multiple approaches are under investigation, ranging from direct depletion to functional modulation and indirect interference with Treg-supportive pathways.

1.4.5.1 Direct Targeting of Tregs

CD25 (IL-2Rα) is constitutively expressed at high levels on Tregs, making it a logical therapeutic target (67). Monoclonal antibodies such as daclizumab and basiliximab were initially developed for autoimmune diseases and transplant rejection (67). In cancer, CD25-directed antibodies have been evaluated for the depletion of Treg cells. However, these antibodies may also impair activated effector T cells, which transiently upregulate CD25 during proliferation (102). To overcome this limitation, non–IL-2–blocking anti-CD25 antibodies (anti-CD25NIBs) are in development, designed to selectively deplete Tregs without interfering with IL-2 signalling in Teffs (103).

CCR4: A subset of eTreg cells with high immunosuppressive activity expresses CCR4. The anti-CCR4 monoclonal antibody mogamulizumab has been approved in Japan for CCR4-expressing T cell leukaemia and lymphoma and has been shown to selectively deplete CCR4+ eTreg cells in cancer patients, augmenting tumour antigen-specific T cell responses (104). Clinical trials combining mogamulizumab with immune

checkpoint inhibitors (ICIs) like nivolumab are ongoing, showing promising early results with manageable toxicity.

CTLA-4: Anti-CTLA-4 antibodies (e.g., ipilimumab) not only block inhibitory signals on effector T cells but also deplete CTLA-4-expressing Treg cells within tumours via antibody-dependent cellular cytotoxicity (ADCC) (105). This depletion increases the CD8+ T cell to Treg cell ratio and is important for the antitumor efficacy of these agents. However, some clinical data have reported a lack of Treg cell depletion post-treatment, indicating variable effects depending on the cancer types and treatments.

Other Immune Checkpoint Molecules: Molecules such as OX40, ICOS, and GITR, members of the TNF receptor superfamily, are constitutively expressed on Treg cells (67). Agonistic antibodies targeting these receptors are under clinical investigation. Activation of these receptors reportedly reduces Treg immunosuppressive activity while stimulating conventional T cells, potentially shifting the immune balance toward tumour rejection (67). Clinical trials have shown some disease stabilisation and partial responses, but it remains to be determined whether these agents effectively deplete or inhibit Treg cells within the TME.

1.4.5.2 Targeting Tregs via Intracellular Signalling Pathways

Phosphoinositide 3-kinase delta (PI3Kδ) is critical for Treg survival and function but less essential for effector T cells. Selective PI3Kδ inhibitors (e.g., idelalisib, experimental agents such as AMG319) selectively impair Tregs, leading to enhanced anti-tumour immunity in murine models (106). Clinical trials in head and neck squamous cell carcinoma (HNSCC) demonstrated transient tumour regression following PI3Kδ inhibition, though autoimmune adverse events (colitis, rash) were observed, highlighting the challenge of balancing efficacy with safety (106).

TCR Signalling: Treg cells require continuous T cell receptor (TCR) signalling for survival. Tyrosine kinase inhibitors (TKIs) that inhibit TCR-associated kinases (e.g.,

dasatinib) have shown off-target effects that reduce Treg cell numbers and may improve clinical outcomes in some cancers (107).

1.4.5.3 Indirect Targeting Tregs

IDO1 (Indoleamine 2,3-dioxygenase 1) Inhibition: IDO1 expression in tumours depletes tryptophan, leading to effector T cell anergy and promoting Treg cell differentiation and activation. IDO1 inhibitors (e.g., epacadostat) have shown some promising activity in early trials but failed to demonstrate progression-free survival benefits in a large phase III trial when combined with pembrolizumab (108). Detailed analyses of Treg cell modulation in these trials are pending.

VEGF/VEGFR Signalling Blockade: VEGF promotes immunosuppression and Treg cell accumulation in tumours. VEGF/VEGFR2 inhibitors (e.g., bevacizumab, ramucirumab) reduce Treg cells in the TME and improve antitumor immunity (67). Combinations of VEGF pathway inhibitors with ICIs have shown promising clinical efficacy (67).

TGF-β Pathway Inhibition: TGF-β suppresses effector T cell functions and promotes tumour progression. TGF-β receptor kinase inhibitors (e.g., galunisertib) and bifunctional fusion proteins targeting PD-L1 and TGF-β (e.g., M7824) have demonstrated enhanced antitumor immunity in preclinical models and early clinical trials (67). These agents might reduce Treg-mediated immunosuppression.

Heat Shock Protein (HSP) Inhibition: HSP70 and HSP90 modulate Treg cell functions. HSP90 inhibitors deplete Treg cells and myeloid-derived suppressor cells (MDSCs) in the TME and enhance effector T cell responses (67). Clinical trials of HSP90 inhibitors combined with ICIs are ongoing.

1.5 Tregs in Early Cancer and Disease Interception

While most research has focused on advanced cancers, the role of Tregs in early and pre-invasive disease is gaining attention. In carcinoma-in-situ (CIS) lesions of the lung and other tissues, Treg accumulation may represent one of the earliest immunosuppressive adaptations, facilitating progression to invasive cancer. Understanding Treg plasticity and function in pre-invasive contexts could open a window of interception.

1.5.1 Immunology in Pre-Invasive Disease

Mascaux et al. (2019) conducted one of the most comprehensive investigations to date, combining bulk transcriptomic microarray data with multispectral imaging across 122 biopsies from 77 patients, spanning nine morphological stages (normal epithelium to invasive LUSC). In Normal Bronchial Tissue, which includes stages 0 to 2 (normal histology with normal fluorescence (stage 0), hypo-fluorescence (stage 1), and hyperplasia (stage 2) (109). At these earliest stages of bronchial tissue development, there is evidence of immune sensing through the activation of resident immune cells. This is suggested by the presence of immune-related gene expression and early immune signalling pathways being engaged. Genes involved in the negative regulation of the immune system (e.g., *TNFRSF14 (HVEM), CD200, CD59, TGFB3,* and *HLA-G*) are significantly downregulated in low-grade lesions, which include these early normal and hyperplasia stages (109). This downregulation suggests that the immune system is unleashing or experiencing reduced suppression, allowing it to be more responsive at these early stages of transformation. Resting mast cells are more abundant in these early stages compared to later stages, indicating a resting or less activated immune microenvironment. There is also a higher abundance of naive B cells and naive CD4 T cells, with a transient influx of naive CD4 T cells occurring around mild dysplasia (which is just beyond this normal tissue category) (24). This implies that the immune system is in a more naive/resting state in early stages, but poised for activation. There is a transitory increase in metabolic pathways in the tumour cells such as fatty acid metabolism, oxidative phosphorylation, and the citric

acid cycle in low-grade lesions, which may reflect shifts in cellular function and potentially influence the immune microenvironment. Gene modules with biphasic expression patterns peak in low-grade lesions (which include hyperplasia), associated with metabolism and early immune sensing. Unlike later stages, few immunosuppressive genes and immune checkpoint molecules are upregulated at this point, consistent with an immune microenvironment that has not yet shifted toward immune escape.

In low-grade pre-invasive lesions (stages 3–5: metaplasia, mild and moderate dysplasia), the immunological changes show a distinct pattern characterised by early immune sensing and activation, but relatively limited immune suppression compared to more advanced stages (109). There is a transitory increase in metabolism and early immune sensing at the low-grade stage. This is reflected by the activation of resident immune cells and a biphasic gene-expression pattern, where certain immune-related genes peak in expression in these lesions. Gene-expression modules indicative of immune response show increased expression starting at low-grade lesions, suggesting that the immune system is recognising early abnormal changes and responding to them. This includes a transient influx of naive CD4 T cells at the mild dysplasia stage (stage 4), which is followed by a shift towards activated memory CD4 T cells in subsequent stages. There is an increase in the abundance of certain immune cells, including activated T cells (CD4 memory), macrophages (M0), memory B cells, follicular T-helper cells, and dendritic cells, although these increases become more pronounced in high-grade lesions. Resting mast cells are more abundant in early developmental stages (including low-grade lesions), whereas activated mast cells increase in higher-grade lesions. Naive B cells decrease while memory B cells increase, indicating the beginning of an adaptive immune memory response. Immune-related genes involved in negative regulation of the immune system (such as *TNFRSF14, CD200, CD59, TGFB3*, and *HLA-G*) are downregulated in low-grade lesions, indicating a relative unleashing of the immune response at this early stage. This means the immune suppressive pathways are less active here, facilitating immune activation. Conversely, co-inhibitory molecules and suppressive interleukins (such as *IDO1, PD-L1, CTLA4, TIGIT, IL10*, and *IL6*) show no significant upregulation in low-grade lesions, unlike in high-grade lesions, where these immunosuppressive pathways become more prominent. Only a few immune functions are modulated

specifically in low-grade lesions, including upregulation of a small number of immune processes and downregulation of others, such as the response to TGFβ. Immune-cell densities such as CD4+ and CD8+ T cells show a transitory increase in low-grade lesions, but the more robust immune infiltration and spatial reconfiguration of the tumour microenvironment (such as segregation of epithelial cells from T cells) occur later in high-grade lesions.

In high-grade pre-invasive lesions of lung squamous cell carcinoma (stages 6 and 7, severe dysplasia and carcinoma in situ), significant immunological changes occur that reflect both immune activation and immune escape, prior to tumour invasion (109). Gene expression profiling shows a marked increase in genes associated with immune response activation in high-grade lesions. There is a significant enrichment of immune functions among upregulated genes, including antigen processing and peptide antigen presentation, signalling pathways such as TNFα via NF-κB, IL6 JAK STAT3, and interferon-gamma response. Moreover, Activated T cells, total neutrophils, M1 macrophages, and myeloid signatures are significantly increased in abundance, indicating an active immune sensing and response in these lesions. Quantitative deconvolution of gene expression data reveals an increase of myeloid-derived cells, neutrophils, and macrophage subtypes (especially M1 macrophages) in high-grade lesions. There is co-regulation and correlation in abundance of immune cells from both innate and adaptive immunity, such as activated CD4 memory T cells, macrophages (M0), memory B cells, follicular T-helper cells, and dendritic cells. A shift from resting to activated immune statuses is observed; for example, resting mast cells decrease while activated mast cells increase. In addition, they also found that naive B cells decrease while memory B cells increase. Naive CD4 T cells influx at mild dysplasia (stage 4) is followed by a sharp decline in their abundance, with a concurrent increase in activated CD4 memory T cells in the high-grade stages. Genes involved in the negative regulation of the immune system are significantly upregulated in high-grade lesions, indicating the onset of immune escape mechanisms. Immunomodulatory genes, including immune checkpoint molecules (co-inhibitory molecules) such as *IDO1, PD-L1 (CD274), TIGIT, CTLA4, TIM3*, and suppressive interleukins like IL-10 and IL-6, show significantly increased expression from severe dysplasia onwards. Stimulatory molecules such as *TNFRSF9 (CD137), TNFRSF18 (GITR), ICOS, CD80, CD86*, and others also increase, suggesting a complex immune

regulatory environment. Moreover, protein-level confirmation via immunohistochemistry shows significant increases in *CTLA4, IDO1*, and *PD-L1* expression in high-grade lesions and SCC compared to normal tissue. Furthermore, Multispectral imaging reveals increased densities of immune cells, including CD4 and CD8 T cells, myeloid cells, neutrophils, and macrophages in the stroma and epithelium of high-grade lesions. PD-L1+ cell densities increase significantly in high-grade lesions and invasive cancer. Spatial analysis shows segregation between epithelial cells and CD3+ T cells in high-grade lesions, indicating a physical separation that may contribute to immune evasion by limiting immune cell access to transformed epithelial cells.

In the invasive carcinoma stage (Stage 8: invasive lung squamous cell carcinoma, LUSC), several detailed immunological changes occur as part of the tumour microenvironment evolution (109). Immune activation and immune escape mechanisms are both present at this stage. There is a significant upregulation of immunomodulatory molecules, including co-inhibitory immune checkpoints and suppressive interleukins, which indicate active immune suppression, allowing tumour evasion from immune destruction. The expression of several immune checkpoint molecules, such as *PD-L1 (CD274), IDO1, CTLA4, TIGIT*, and *TIM3*, is significantly increased in invasive carcinoma compared to earlier stages and normal tissue. These molecules contribute to immune evasion by inhibiting effective anti-tumour immune responses. Immunosuppressive interleukins like *IL6, IL10*, and *TGFB1* are also elevated, further contributing to an immunosuppressive microenvironment. There is an increased density of immune cells in the tumour microenvironment, especially of myeloid-derived cells, including macrophages and neutrophils, as well as activated T cells (both CD4+ and CD8+). However, despite increased immune infiltration, the tumour has evolved mechanisms to evade immune attack. The immune status shifts towards activation; for example, activated mast cells and memory B cells become more abundant, while naive immune cells decrease. Spatial analysis reveals that in high-grade lesions and invasive carcinoma, epithelial tumour cells (CK+) become spatially segregated from CD3+ T cells. This spatial separation suggests a physical exclusion of T cells from the tumour cell clusters, which is a potential immune evasion mechanism. PD-L1+ cells, mainly immune cells in the stroma, increase significantly, consistent with the immune checkpoint upregulation. Besides immune changes,

invasive carcinoma shows activation of epithelial–mesenchymal transition (EMT) pathways, which are associated with tumour invasion and metastasis, and may be linked to changes in immune interactions. The presence of immune checkpoint molecules and an immunosuppressive environment at this stage supports the use of checkpoint inhibitor immunotherapy as a treatment strategy for advanced lung SCC.

1.5.2 Tregs in Pre-Invasive Disease: Drivers of Tumour Progression and Opportunities for Interception in LUSC

According to Mascaux et al. (2019), those cells with the CD3+ FOXP3+ phenotype (which correspond to regulatory T cells, Tregs) were rarely found in the epithelium at early stages of development (stages 0–5, which include low-grade lesions). This suggests that Tregs are not abundant or are infrequent in low-grade pre-invasive lesions, and their infiltration or functional activity likely increases in later high-grade or invasive stages (109).

However, in high-grade pre-invasive lesions, there is both activation of immune responses and upregulation of immune escape mechanisms, including immune checkpoints and suppressive interleukins (109). This suggests an immunosuppressive microenvironment begins to form at these stages. High-grade lesions show significant upregulation of co-inhibitory molecules such as CTLA4 and PD-L1, as well as suppressive interleukins like IL10 and TGFB1, which are commonly associated with Treg function and immune suppression. These molecules are known to be expressed by or target Tregs and contribute to their suppressive activity. The functional and phenotype panels used for multispectral imaging included markers such as FOXP3 (a canonical Treg marker), and immunohistochemistry confirmed increased densities of suppressive checkpoint molecules in high-grade lesions and invasive cancer. Although the document does not explicitly describe Treg densities or spatial distribution, the presence of FOXP3+ cells was assessed. There is a notable spatial reconfiguration of the tumour microenvironment in high-grade lesions, including segregation of epithelial cells from CD3+ T cells, which likely includes Tregs. This

spatial arrangement may reflect mechanisms of immune evasion that involve Tregs. The study observed increases in activated T cells in high-grade lesions before invasion. Although not detailed for Tregs specifically, this activation shift likely includes the expansion or increased activity of regulatory T cells to modulate immune responses. The study shows that immune escape mechanisms are activated before tumour invasion, with a significant increase in the expression of co-inhibitory molecules and suppressive interleukins from high-grade lesions onwards, continuing into invasive carcinoma (Stage 8). Many immunomodulatory molecules, including immune checkpoints such as *CTLA4, PD-L1, IDO1, TIGIT*, and *TIM3,* are upregulated at the invasive stage. *CTLA4* is a well-known marker expressed by Tregs and plays a role in immune suppression.

Moreover, according to Pennycuick et al., CIS represents a critical sweet spot for studying LUSC interception because it is a high-grade pre-invasive lesion stage where the disease is at a pivotal point: around half of CIS lesions progress to invasive LUSC, whereas about a third spontaneously regress (110). This intermediate stage could allow us to observe and compare the biological and immunological differences between lesions that will progress and those that will regress, providing valuable insights into mechanisms of immune surveillance and immune escape that occur early in carcinogenesis. Studying CIS offers the opportunity to understand immune surveillance mechanisms that lead to regression, as Pennycuick et al. found that regressive CIS lesions show higher immune cell infiltration, particularly CD8+ cytotoxic T cells, compared to progressive lesions (110). Also, CIS offers the chance to identify immune escape mechanisms present before invasive LUSC develops, such as impaired antigen presentation through genomic and epigenetic changes, upregulation of immune modulatory pathways (e.g., CCL27/CCR10 signalling), and downregulation of immunostimulatory molecules (e.g., TNFSF9) (110). It could help us investigate potential therapeutic targets and interventions at a stage where they might effectively prevent progression to invasive LUSC, thus enabling early interception strategies.

1.6 Research Gap

Previous studies have established that immune activation and immune suppression emerge before tumour invasion during LUSC progression, highlighting the importance of the immune microenvironment in pre-invasive disease. Mascaux *et al.* demonstrated that high-grade pre-invasive lesions show increased immune checkpoint expression, suppressive cytokines, and spatial reorganisation of immune-epithelial interactions, suggesting that immune escape precedes malignant invasion.

However, these observations were largely descriptive and based on bulk transcriptomic analyses and multiplex imaging, leaving unresolved the identity, transcriptional state, and mechanistic role of the immune populations driving immune suppression at this critical stage. It remains unclear which regulatory T-cell programmes operate in pre-invasive lesions, how they are shaped by local antigen-presenting cells, and whether specific tumour-adapted Treg subsets actively orchestrate immune escape. Addressing this gap requires single-cell and functional approaches capable of resolving immune cell states, lineage relationships, and cellular crosstalk with pre-invasive lesions. This thesis directly addresses these unresolved questions by studying T cell dynamics in human CIS lesions to identify which T cell subsets are significantly higher in human CIS lesions compared to adjacent normal tissues using scRNAseq data. Moreover, this allows us to study T cell differentiation in pre-invasive disease, as well as identify potential markers for early detection via human PBMC using paired scRNAseq-TCRseq data to investigate T cell subsets that are significantly higher in human CIS lesions compared to adjacent normal tissues have shared TCRs with T cell subsets in human PBMC.

1.7 Hypotheses and Aims

I hypothesise that Tregs might be associated with an immune escape mechanism at the stage of CIS or high-grade dysplasia in pre-invasive LUSC. This may provide a critical window to intercept LUSC development before invasive transformation occurs.

My aims are

- To study T cell dynamics at the single-cell level during LUSC development with a focus on high-grade pre-invasive lesions from both human and murine models. Also, to examine whether BATF+Tregs accumulate in high-grade pre-invasive lesions compared to adjacent normal airways
- To identify CD4 subsets in PBMC that have shared TCRs with BATF+Tregs in human PID lesions, and possible markers for early detection of LUSC
- To identify biological molecular pathways in BATF+Tregs and crosstalk axes between APCs as senders and BATF+Tregs as receivers from both human and murine models, which may help to find targets to disrupt the BATF+Treg pool.

Chapter 2. Methodology

2.1 single-cell RNA-sequencing (scRNAseq) analysis

2.1.1 Single-cell data pre-processing and quality control (scRNA-TCR-seq)

Cell Ranger (7.1.0) , is the official software pipeline developed by 10x Genomics for processing and analysing scRNAseq data generated on the Chromium platform. The Cell Ranger was used to demultiplex the FASTQ reads, align them to the GRCh38 human transcriptome for human UCLH Bronch. Fellow PhD student Amber Rogers from my lab sorted live single cells from single-cell digests of lung lesions via Fluorescence-activated cell sorting (FACS) from normal samples (n = 6), low-grade samples (n = 1), and high-grade samples (n = 10). Samples are shown with clinical information in Table 3.1 below.

Table 2.1 Clinical information of patients' bronchial biopsies used in scRNAseq analysis

Patient ID	Site	Histology	Sex	Age	Pack years
P149	RLL	CIS	M	71	50
P149	LLL	Normal			
P152	LT	CIS	M	81	120
P152	LLL	Normal			
P153	LMB	CIS	F	77	48
P153	LLL	sqM			
P155	RUL	Normal	F	74	45
P155	MC	Normal			
P157	LT	CIS	M	70	40
P157	LT	SD			
P157	MT	Low-Grade			
P158	RIB-POST	CIS			

P158	RUL-CAR	CIS	M	70	40
P158	RUL-ORI	CIS			
P704	LLL	Normal	M	74	18

Surveillance data and GRCm38 for NTCU mouse data, and extract their cells and unique molecular identifier (UMI) barcodes. The output of this pipeline is a digital gene expression matrix for each sample, which records the number of UMIs for each gene that are associated with each cell barcode. Cell quality was evaluated using two main criteria: the number of genes detected per cell and the proportion of reads mapping to mitochondrial genes. The genes-per-UMI ratio served as an indicator of dataset complexity, with higher values reflecting greater transcript diversity. The mitochondrial read fraction represented the percentage of transcripts derived from mitochondrial genes, a marker often associated with stressed or dying cells. These metrics were used to determine which cells were of insufficient quality and, therefore, excluded from downstream analysis.

- UMI counts per cell: Cells should ideally have more than 500 detected transcripts. Counts between 500 and 1000 are considered acceptable, though they may indicate that the sequencing depth was suboptimal.

- Genes detected per cell: This metric follows the same principle as UMI counts, though the values are typically slightly lower. In high-quality datasets, the distribution should display a single dominant peak, corresponding to successfully captured cells.

- UMIs vs. genes detected: The number of UMIs and genes per cell is typically assessed together. To visualise this, I plotted genes against UMIs, with mitochondrial read fraction shown by colour. Cells with low counts and few detected genes often display disproportionately high mitochondrial fractions, suggesting damage or cell death where

cytoplasmic mRNA has leaked but mitochondrial transcripts remain. Such cells are excluded based on gene and UMI thresholds. This joint visualisation highlights how low-quality cells cluster in the bottom-left quadrant, whereas high-quality cells exhibit both higher gene detection and higher UMI counts.

- Mitochondrial count ratio: This metric helps detect excessive mitochondrial content, which is often associated with dead or dying cells. Samples are considered low quality if the mitochondrial fraction exceeds 20%, unless elevated mitochondrial reads are expected for the specific cell type under study.

Low-quality cells were filtered if the number of detected genes was below 250 or above 3 times the median absolute deviation away from the median gene number of all cells. Cells were filtered out if the proportion of mitochondrial gene counts was higher than 10%. Low-quality genes were filtered out if genes had zero expression in all cells.

2.1.2 Single-cell data integration and clustering

Seurat (v4.4.1) (111) was used to normalise the raw count data, identify highly variable features, scale features, and integrate samples. PCA was performed based on the 3,000 most variable features identified using the vst method implemented in the Seurat package (111). Gene features associated with immunoglobulin genes, specific mitochondrial-related genes, and type I Interferon response were excluded from clustering to avoid cell subsets driven by those genes (86). Highly variable genes were selected for each cell cluster by fitting a locally weighted scatterplot smoothing regression of standard deviation against the mean for each gene and identifying genes with positive residuals. For each sample, all cell clusters were then concatenated by retaining each cluster's HVGs to produce a concatenated gene expression vector consisting of all highly variable features identified from different cell clusters. Dimension reduction was done using the RunUMAP function. Cell markers were identified by using a two-sided Wilcoxon rank sum test. Genes with adjusted $P<0.05$

were retained. Canonical correlation between the first two PCs and the covariate of interest (phenotype status, including CIS and normal) was calculated. Clusters were labelled based on normalised enrichment scores (NES) using immune and non-immune gene signatures and T cell gene signatures. Moreover, differential expression gene analysis was performed using the *FindMarkers* function.

2.1.3 Cell type annotation for scRNAseq data

To perform cell type annotation on scRNAseq from human PID and NTCU mouse data, I developed a pipeline for a robust method of cell type annotation. First, I performed differential expressed gene analysis (DEG) of a cell cluster vs other cell clusters in the pre-invasive TME to get DEGs that were highly expressed in each cell cluster. Then, I performed gene set enrichment analysis (GSEA) (112) to calculate the normalised enrichment score (NES) on each cell cluster and guide roughly which cell types they could be on each cell cluster. The Normalised Enrichment Score (NES) represents the enrichment score of a gene set, scaled by the expected score for gene sets of the same size under random permutations. By adjusting for gene set size and dataset-specific biases, the NES provides a robust and comparable measure of pathway enrichment across multiple gene sets, forming the core statistic used in GSEA (112). Finally, I confirmed the temporary cell type annotation based on GSEA using curated gene markers of T cells and whole cell types from Zheng et al. (113), CellTypist (114), and Azimuth (115), which were plotted on dotplot using ggplot.

2.1.4 Integration of scRNAseq of human PID and lung cancer atlas.

To integrate scRNAseq of T cells between human PID and lung cancer atlas from Salcher et al., I combined raw count matrices of scRNAseq of human PID and lung cancer from Salcher et al. (116) and then created a Seurat object. Next, I performed scRNAseq data integration and clustering in a similar way as I did for scRNAseq analysis of human PID.

To calculate the proportion of BATF+Tregs as T cells, I first subset all T cells from scRNAseq's metadata. Then, I created a dataframe showing the number of cells on each T cell subset (for each patient and for each pathology) using a for-loop. Next, the proportion of BATF+Tregs as T cells was calculated by the number of BATF+Tregs on each patient with pathology, divided by the total number of T cells on each patient with pathology. An unpaired t-test was used to compare the proportion of BATF+Tregs between normal and pre-LUSC, and between normal and LUSC.

2.2 MHC II signalling pathway analysis

MHC class II genes, including *HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DMA, HLA-DMB, HLA-DQA2, HLA-DOA, HLA-DOB, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, and HLA-DRB5*, were used to calculate MHC II score using the UCell package (v2.6.2) (117). UMAPs showing UCell score of MHC II were constructed to visualise MHC II expression between normal and CIS using scRNA-seq of human UCLH Bronch. Surveillance data. UCell computes single-cell signature scores using the Mann–Whitney U statistic applied to within-cell gene ranks, with tail-rank truncation to mitigate sparsity. This yields composition-invariant, robust scores that are fast and memory-efficient to compute, compare favourably with existing approaches (e.g., *AddModuleScore, AUCell*), and integrate directly with Seurat, making UCell well-suited for large, heterogeneous scRNA-seq datasets and for reproducible cell-state annotation.

CellChat (v2.1.1) (118) provides a powerful framework for inferring and analysing intercellular communication in single-cell transcriptomic data. By combining a rigorously curated ligand–receptor database with probabilistic modelling and systems-level network analysis, CellChat surpasses earlier methods that relied on simplified ligand–receptor definitions. Its ability to handle both discrete cell types and continuous trajectories, alongside robust visualisation and comparative tools, makes it particularly suited for investigating the complex cellular ecosystems of cancer and pre-

invasive disease. It was used to analyse MHC II signalling in CIS and normal tumour microenvironment. First, ligand and receptor interaction strengths were calculated for all pairs of cell types in CIS and normal TME separately. Next, ligand and receptor interaction in the similar signalling pathways were merged and computed cell-cell signalling scores. A dotplot was constructed to visualise crosstalk axes of MHC II between basal cells, suprabasal cells, B cells, cDC2, fibroblasts, and cDC1 as senders and BATF+ Tregs as receivers.

2.3 Cell-cell communication analysis

Since I found that the expression of MHC-II genes is highly enriched on suprabasal and basal cells under normal conditions, but on immune cells such as fibroblasts and B cells in human CIS. This suggested that I need to study MHC-II signalling and crosstalk axes on these cells (as senders) and BATF+Tregs (as receivers) to see which crosstalk axes would facilitate BATF+Treg proliferation.

CellPhoneDB (v4.0.1) (119) provides a robust, statistically principled framework for inferring cell–cell communication from scRNA-seq data. By uniquely modelling the multimeric architecture of receptor–ligand complexes, it surpasses earlier tools that relied on binary one-to-one interactions. However, CellPhoneDB is primarily focused on the detection of enriched ligand–receptor pairs in discrete cell clusters, whereas CellChat offers a more comprehensive systems-level analysis that integrates cofactors, pathway classification, and network-level properties. In practice, CellPhoneDB remains an invaluable tool for defining high-confidence ligand–receptor interactions, but CellChat may offer broader biological interpretability for complex tissue ecosystems

Currently, CellChat (118) and CellPhoneDB (119) still do not have a function that compares gene expression amongst two or multiple conditions to find DEG genes that are highly expressed in CIS compared to normal tissue. Therefore, I need to perform CellChat and CellPhoneDB separately between CIS and normal conditions.

First, I subset the interesting cell types on the T cell Seurat object (including basal cells, suprabasal, cDC1, cDC2, B cells, fibroblasts, pDC, and BATF+Tregs) in CIS and normal tissues. The subset object was used to create a CellChat object using the *createCellChat* function. Next, the ligand and receptor interaction database was set using CellChatDB.human. Next, differential expressed gene analysis was performed using the *identifyOverExpressedGenes* function. Then, the communication probability score on the signalling pathway level is calculated by summarising the communication probabilities of all ligand-receptor interactions associated with each signalling pathway using the *computeCommunProb* function. Finally, CellChat.norm and CellChat.cis objects were the output from CellChat analysis.

In order to compare the strength of interaction of MHC-II signalling between CIS and normal, I subset the MHC-II signalling pathway from the CellChat.norma and CellChat.cis. Then, I rescaled those probability scores of MHC-II signalling from CellChat.cis and CellChat.norm. Next, dotplots showing MHC-II signalling between CIS and normal were created using ggplot.

CellPhoneDB (v4.0.1) (119) was used to analyse crosstalk interaction between BATF+ Treg and all cell types in the preinvasive tumour microenvironment. The crosstalk analysis between CIS and normal was performed separately. The Seurat object containing BATF+ Treg and all cell types in both CIS and normal was used in the analysis. BATF+ Treg cluster was set as the receiver, whereas all remaining cell types were set as the sender populations. For the receiver cell population, the differential expression gene analysis was performed to find genes significantly enriched in each cell type, with the key criteria being set as follows: genes expressed in at least 10% of the cells of the respective whole cell type clusters and filtered after the DEG test for a P-value of less than 0.05 and average log fold change more than 0.25. For sending the cell population, all ligands expressed in at least 5% of the cells in all remaining cell types were considered. CellPhoneDB analysis was performed.

To identify specific crosstalk axes that are highly enriched in CIS compared to normal, crosstalk axes were selected with the key consideration being set as follows: crosstalk axes need to have higher interaction strength between ligands and receptors in CIS compared to normal. A chord diagram was constructed to visualise the

interaction between ligands (all cell types) and receptors BATF^{+} Tregs using the circlize package (v0.4.15). Crosstalk axes were validated using bulk transcriptome data from human biopsies by Mascaux *et al*. (109)

2.4 Multivariate analysis of paired WES and bulk RNAseq of metastatic NSCLC treated with anti-PD1 treatment

In order to expand my question, whether expression of the BATF+Treg gene signature was associated with anti-PD1 resistance in metastatic NSCLC independently from CD4 and CD8 infiltration, TMB. Multivariate analysis was performed. First, I calculated expression of BATF+Treg gene signature, CD4 (*CD4, CD3D, CD3E, CD3G*), CD8 (*CD8A, CD8B, CD3D, CD3E, CD3G*) on each patient using geometric mean (exp(mean(log(data)))). Next, high- and low-expressions of the BATF+Treg gene signature, CD4, CD8, and TMB were classified using the median. Next, I performed multivariate analysis (discrete model) using the Logistic regression model to examine whether expression of the BATF+Treg gene signature was associated with anti-PD1 resistance independently of CD4, CD8 infiltration, and TMB. Finally, a forest plot was created using ggplot

2.5 Gene set enrichment analysis (GSEA)

To perform GSEA, I first identified gene markers that were highly enriched in each cell type using the *FindAllMarkers* function provided by the Seurat package (120). Also, I used gene markers as a reference for each cell type from Dykema et al. (86) and Guo et al (121). Then, we created a for-loop to perform GSEA analysis on each cell type. First, gene markers were selected for each cell type, and then average log2FC values were arranged from highest to lowest. Next, GSEA was performed using the fgsea function with the default parameters from the fgsea package (112). Next, we created a dataframe that includes pathway (referenced cell types), normalised enrichment scores (NES), size (gene markers that overlapped between our

cell types and referenced cell types), and our cell types. Next, we subset our BATF+Treg cluster to create a dotplot showing NES scores with referenced cell types.

2.6 Normalisation (Galon's microarray data)

I used TMM (trimmed mean of M-values) normalisation from the edgeR package to compute normalisation factors that represent sample-specific biases. These factors are multiplied by the library size to yield the effective library size. TMM assumes that most genes are not differentially expressed between samples. If many genes are uniquely or highly expressed in one experimental condition, it will affect the accurate quantification of the remaining genes. To adjust this, one sample is chosen as a reference sample. The fold changes and absolute expression levels of other samples within the dataset are then calculated relative to the reference sample. Next, the genes in the dataset are trimmed to remove differentially expressed genes using these two values. The trimmed mean of the fold changes is then found for each sample. Finally, read counts are scaled by this trimmed mean and the total count of their sample.

2.7 Bulk transcriptome analysis

Bulk transcriptome data of human biopsies were downloaded from the NCBI repository (GSE33479) (109). The dataset includes 122 biopsies: 13 biopsies with normal histology and normofluorescent (normal), 14 with normal histology and hypofluorescent (normal), 15 hyperplasia (low grade), 15 metaplasia (low grade), 13 mild dysplasia (low grade), 13 moderate dysplasia (low grade), 12 severe dysplasia (high grade), 13 CIS (high grade), and 14 squamous cell carcinoma (SCC) from 77 patients. The DEGs of $BATF^{+}$ Tregs, were used to calculate the representative expression of the cell types using geometric mean. The Wilcoxon rank sum test was used to compare the difference in the expression of those cell types between normal and high-grade, high-grade and low-grade, and high-grade and squamous cell carcinoma.

2.8 Pseudotime trajectory analysis (monocle2)

Monocle2 (1.3.4) (122) was used for pseudotime analysis. CCA batch corrected PC values for scRNA-seq of human UCLH Bronch. Surveillance and mouse NTCU were used as input for the UMAP dimension reduction-based construction of pseudotime trajectories for CD4 subsets. Following the construction of the pseudotime trajectory, based on the function cluster_cells, CD4 subset clusters were divided into large and well-separated groups called partitions, within each of which a principal graph was fitted using the function *learn_graph*. The principal graph was shown on the UMAP as skeleton lines, indicating differentiation trajectories. Assigning CD4 subset cluster to the nearest principal graph nodes, the principal graph node containing the highest fraction of Naïve-like CD4 T cells was specified as the root, then pseudotime was calculated using the *order_cells* function.

Since I found shared TCRs between BATF+Tregs and progenitor CD4 cells in CIS, this suggests that BATF+Tregs could be differentiated from progenitor CD4 cells. In order to examine and validate whether BATF+Tregs were differentiated from naïve-like CD4, monocle3 was performed. First, we subset all CD4 subsets in CIS from the T cell Seurat object. Then, the subset object of CD4 T cells was converted to a monocle3 object using the as.cell_data_set function. Next, the relationships between CD4 subsets, as a trajectory of gene expression changes, were analysed using the learn_graph function. Next, I ordered cells using order_cells (reduction_method = “UMAP”, root_cells = “Naive-like CD4”).

2.9 Differential Abundance of T cell Types and Whole Cell Types using miloR

I subset the original single cell object to include P149_CIS, P149_Normal, P152_CIS, P152_Normal, P155_Normal, P157_CIS, P704_Normal, and P704_CIS for miloR analysis (123). Next, the Seurat object was converted to a sce object using the *as function.SingleCellExperiment* function before converting to a Milo object using the

Milo function. Then, miloR started to build a k-nearest neighbour graph first using the buil*dGraph* function (k = 30 nearest neighbours to consider for the graph building; d = 25 PC dimensions to use). Next, miloR used makeNhoods function to define neighbourhoods on a graph with parameters (prop = 0.3, k = 30, d = 25, refined = TRUE, reduced_dims = “PCA”). This function randomly samples vertices on a graph to define neighbourhoods. These are then refined by computing the median profile for the neighbourhood in reduced-dimensional space and selecting the nearest vertex to this position. Next, I used the countCells function to quantify the number of cells in each neighbourhood according to an input experimental design. Our experimental design was based on patients with pathology (e.g. P149_CIS, P149_Normal). Then, the calcNhoodDistance function was used to calculate Euclidean distances between single cells in a neighbourhood using the same dimensionality as was used to construct the graph. Next, the testNhoods function was used to perform differential neighbourhood abundance testing. I then matched cell type annotation using the annotateNhoods function. Finally, a Beeswarm plot was created using ggplot

2.10 Bootstrap Version of Differential Abundance of T cells using miloR

In order to address an imbalanced dataset, I developed a bootstrap version of the differential abundance of cell types using miloR (Fig. 2.1) (123). As there was a difference in the number of cells between NTCU and control mice dLN at week 18 and also between human PID and normal, I downsampled by sampling based on Cell ID to have a similar number of cells between NTCU and control mice dLN. Then, miloR was performed using similar parameters to the normal version of miloR. Next, all significant neighbourhoods with the cutoff of logFC <= -0 for control / normal and logFC >= 0 for NTCU / human CIS were used to calculate the average logFC on each iteration. Finally, the beeswarm plot was plotted to visualise the results of bootstrap miloR.

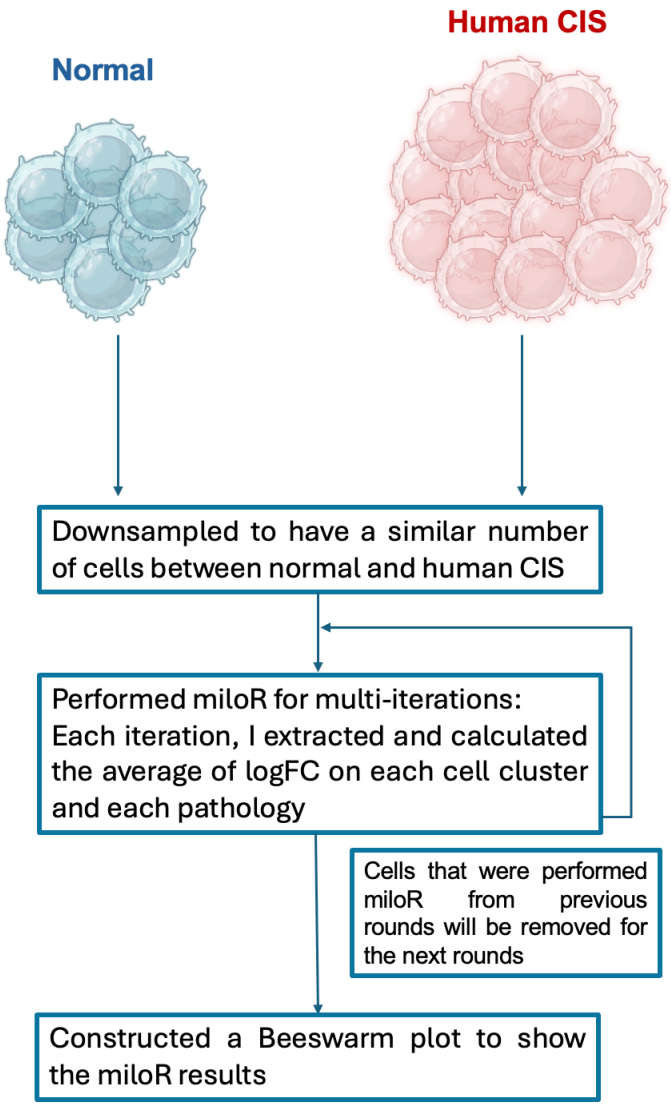

Figure 2.1 – The overview of bootstrap miloR as I developed. It begins by downsampling to achieve a similar number of cells between normal and human CIS/controls, as well as NTCU mice. Then miloR was performed for multi-iterations: each iteration, the average of logFoldChange was calculated on each cell cluster and each pathology. Next, cells that were performed miloR from previous rounds were removed for the next rounds. Finally, a Beeswarm plot was constructed to show the miloR results.

2.11 STARTRAC analysis (Expansion, Transition)

STARTRAC (124) integrates scRNA-seq with paired TCR sequencing to derive quantitative indices of clonal expansion, migration, and developmental transitions. By

leveraging clonotype sharing as a lineage marker, STARTRAC goes beyond static enumeration of expanded clones and instead reconstructs the dynamic fates of T cells across tissues and phenotypic states. This makes STARTRAC a powerful tool for dissecting TCR-driven trajectories in tumour immunology, including expansion of exhausted CD8+cells, transition of effector to memory states, and conversion of conventional CD4+T cells into regulatory T cells.

The transition index (STARTRAC-tran) quantifies how strongly two transcriptionally defined T cell clusters are linked by clonotype sharing. The rationale is that if two subsets contain a higher-than-expected number of cells with identical TCR clonotypes, it is likely that they represent different phenotypic states of the same clonal lineage (i.e., evidence of developmental transition).

The STARTRAC-expa index quantifies the extent of clonal proliferation within a transcriptionally defined T cell cluster. The rationale is that, if a cluster contains many cells belonging to the same TCR clonotypes, then those cells likely underwent antigen-driven expansion.

Integrating the TCR data and cluster assignment of the T cells. The STARTRAC indexes quantify the magnitude of clonal expansion. The clonal expansion index uses Shannon entropy to quantify the distribution evenness of the TCR repertoire (all clonotypes) of the T cell cluster. Then, the expansion index is calculated as 1-evenness, similar to standard TCR clonality measurement, and then normalised to the range of 0 to 1, with a high value indicating high clonality of the cluster. A heatmap was used to visualise the results of START

2.12 NTCU Mouse Model of Lung Squamous Cell Carcinoma

Dr Zoe Whiteman and Dr Sandra Gomez-Lopez used female FVB/n mice, aged 6 weeks at the start of treatment. Prior to chemical application, the dorsal fur of

each mouse was shaved one day in advance. Subsequently, 75 µL of 13 mM N-nitroso-tris-chloroethylurea (NTCU), dissolved in acetone, was topically applied to the shaved dorsal skin twice weekly for a duration of 12 weeks. Following the treatment phase, mice were maintained for an additional observational period of up to 12 weeks to allow for the progression of pre-invasive lesions to invasive carcinoma. Throughout the study, mice were housed in individually ventilated cages under controlled environmental conditions (12-hour light/dark cycle, 25 °C temperature, 45-65% relative humidity) with ad libitum access to food and water. Body weights were recorded weekly to monitor health status, ensuring weight loss did not exceed 10% of initial body weight or fall below 15-20% relative to age-matched controls. Any adverse reactions or side effects were closely monitored in accordance with ethical guidelines. Tissue collection was performed at study termination via terminal anaesthesia and transcardial perfusion with PBS.

Chapter 3. BATF+ Tregs Expand During Human LUSC Carcinogenesis

3.1 Background

In the UK, recent data show that five-year survival for lung cancer has seen modest but encouraging progress. For patients diagnosed between 2016 and 2020, approximately 65% of stage I cases and 40% of stage II cases survive for at least five years, though survival drops sharply to around 15% for stage 3 and just 5% for stage 4 (Cancer Research UK, 2022). Overall, only 45% of lung cancer patients live at least one year post-diagnosis (Cancer Research UK, 2022), while just roughly 20% reach the five-year mark, and around 10% make it to ten years. Despite these seemingly low figures, mortality rates from lung cancer in the UK have improved significantly over the past few decades: between 1993 and 2018, male death rates fell by over 53%, and female rates by around 21% (125). These gains reflect a combination of factors: reduced smoking, earlier detection, and more effective treatments, including immunotherapy and targeted therapy. Interestingly, most gains have been seen in earlier-stage disease. Though survival remains lower for many other cancers, this highlights the urgent need for continued investment in early diagnosis and innovative therapies to capitalise and further develop on these advances.

Early detection of lung cancer offers a critical window of opportunity for disease interception by identifying and halting cancer at its earliest and most treatable stages. Unlike advanced-stage disease, early lesions such as CIS (in LUSC) or atypical adenomatous hyperplasia (AAH in LUAD) are often confined to the airway epithelium or small nodules without invasive spread. With the advent of low-dose CT (LDCT) screening and the rollout of the targeted NHS lung health check pioneered in the UK by the SUMMIT study led by Prof Sam Janes, more patients are likely to be diagnosed at stage I, when curative surgery or localised treatment is feasible (126). This ‘stage-shift’ in detection opens the door to early treatment, but finding precancerous lesions could enable interception strategies by intervening before malignant progression through targeted therapies, immunoprevention, or surveillance. Ongoing efforts, in addition to LDCT, including bronchoscopic sampling, the discovery of circulating

biomarkers, and molecular imaging, may help enable real-time assessment of lesion biology. Recent work in the Janes lab has pioneered clinical and basic research that hopes to foster new approaches to detect pre-invasive central (pre-LUSC) or peripheral (pre-LUAD) pulmonary lesions and monitor them longitudinally. Clinically, these studies are highlighting how early detection may improve survival and can transform lung cancer into a disease that can be intercepted, rather than merely treated, but also highlight molecular cancer cell intrinsic and immunological discoveries that frame my project. Our collective efforts are aimed at achieving a paradigm shift that mirrors interception strategies in other cancers, like removing precancerous polyps by colonoscopy and polypectomy in the colon (127) and giving HPV vaccination in cervical cancer (128), where detecting precancerous changes leads to significant mortality reduction.

CIS represents a critical inflexion point in the progression of lung squamous cancer and serves as one of the most promising target lesions for LUSC interception (129). As a late pre-malignant lesion that has not yet invaded the basement membrane, CIS offers a narrow but actionable window for therapeutic intervention before full-blown malignancy and metastatic potential arise. At this stage, preventive strategies ranging from immunoprevention and chemoprevention to surgical excision can in theory be precisely deployed to halt disease progression. It is our hope that molecular and microenvironmental changes within CIS can be exploited as biomarkers for early detection and as actionable targets for risk-adapted treatment.

It is important to consider that failing to detect or act on CIS early enough risks transitioning to invasive disease, where treatment becomes significantly less effective and more morbid. But, conversely, without accurate risk stratification, early interventions may lead to overtreatment of indolent lesions that might otherwise regress, exposing patients to unnecessary physical, psychological, and financial burdens. Therefore, understanding CIS as a biological and clinical fulcrum may allow us to balance timely intervention against the harms of late treatment and overtreatment, shifting lung cancer toward a precision prevention paradigm.

My central hypothesis is that the immune microenvironment is a decisive factor in determining whether CIS lesions in the lung regress or progress to invasive disease.

According to Pennycuick et al., transcriptomic and immune profiling of endobronchial CIS lesions show regressive lesions harbour a significantly more active immune compartment (110). These lesions exhibited higher infiltration of CD8+ cytotoxic T cells. Also, the expression of the HLA-A gene was significantly reduced in progressive lesions compared to regressive lesions, indicating that regressive lesions maintain higher levels of the molecule, which is necessary for presenting antigens to T cells. However, progressive lesions displayed signs of immune evasion: (i) diminished T cell infiltration, their quantifications of infiltrating immune cells by immunohistochemistry (IHC) showed that regressive lesions had higher frequencies of intra-lesional cytotoxic CD8 T cells compared to progressive lesions and analysis of lymphocytes from H&E stained slides showed more infiltrating lymphocytes in regressive lesions compared to progressive lesions, (ii) reduced expression of HLA class I genes, expression of HLA-A gene was significantly reduced in progressive lesions compared to regressive lesions and differential methylation analysis showed hypermethylation in chromosome 6, which covers the major HLA genes, leading to epigenetic silencing of multiple antigen presentation genes (such as HLA-A, HLA-B, HLA-C, TAP1), this hypermethylation was inversely correlated with gene expression, supporting silencing of these antigen presentation genes in progressive lesions, and (iii) upregulation of immunosuppressive pathways such as PD-L1, expression of PD-L1 was not significantly upregulated at the gene expression level; however, IHC identified three progressive lesions with more than 25% of epithelial cells expressing PD-L1, suggesting some progressive lesions may use this pathway for immune escape, and CCL27/CCR10, the ligand and receptor expression ratio analysis of cytokine pairs identified significant upregulation of CCL27:CCR10 axis in progressive lesions, CCL27:CCR10 signalling has been previously associated with immune escape via PI3K/Akt activation in melanoma and showed correlation with PIK3CA and AKT1 expression in CIS lesions (130). Importantly, multiplex immunofluorescence confirmed that signs of immune exclusion and immune escape were already present at the CIS stage. They found that regressive CIS lesions harboured significantly more infiltrating lymphocytes, particularly CD8+ cytotoxic T cells, compared to progressive lesions, indicating immune infiltration differences at this early stage (130). Moreover, some progressive lesions showed an immune cold phenotype with low immune cell infiltration, whereas others were immune hot but still progressed, suggesting the presence of immune escape mechanisms beyond immune cell exclusion (130). These

data suggest that immune surveillance can mediate the spontaneous regression of pre-invasive lesions, but failure of this surveillance allows progression to malignancy (110). Studying these early immune changes at the cell type-specific level is critical not only for identifying biomarkers predictive of lesion outcome, but also for designing immunopreventive interventions (e.g., precision, low-toxicity checkpoint blockade or immunomodulators) that could restore anti-tumour immunity and intercept lung cancer before it becomes invasive.

At around the same time, Mascaux *et al*. highlighted a dynamic remodelling of the immune microenvironment during the progression from normal epithelium to invasive LUSC, suggesting that T cells may play a central role. They observed a gradual increase in T cell infiltration as lesions progressed from normal to high-grade pre-invasive stages (including severe dysplasia and CIS), followed by a reduction in T cell-associated gene expression in invasive LUSC using bulk gene expression microarray data. Notably, there was a distinct shift in T cell functional states, marked by an early enrichment of immune activation genes in low-grade lesions and a transition to immune suppressive and exhausted phenotypes in high-grade stages. Among CD4+ T cells, the study used bulk RNA deconvolution to reveal a skewing in differentiation at high-grade lesions, with significant upregulation of genes associated with Tconv memory cells and Tregs. Specifically, high-grade lesions showed enrichments of *PD-L1 (CD274), TIGIT, CTLA4*, and *IL10*, which are immune suppressive markers (109). These findings support the rationale for studying and possibly targeting CD4+ T cell subsets during the pre-invasive window to prevent immune escape and tumour evolution.

Mascaux et al., found naive CD4 T cells also increased transiently during mild dysplasia and then declined in later stages, suggesting initial sensing and priming of naive CD4+ T cells in early lesions. Moreover, they found co-inhibitory molecules (e.g., *CTLA4, PD-L1, TIGIT*, and *IDO1*), commonly associated with Tregs, are upregulated in high-grade lesions and LUSC. Stimulatory markers like *TNFRSF9, ICOS*, and *GITR*, which are expressed on activated CD4+ subsets (Tregs and effector T cells), were also increased (109). These suggest that Tregs may be involved in immune evasion in high-grade premalignant lesions during LUSC carcinogenesis. These insights were made using bulk transcriptomics of 122 human bronchial biopsies from 77 smoker

patients. However, if we can study immunodynamics at the single-cell level, particularly early T cell dynamics and T cell differentiation in pre-invasive disease, we could better define the landscape for tissue biomarkers of progression, premalignancy and targets for interception. My study, therefore, starts by asking which immune cell types are significantly enriched in high-grade lesions compared to normal lung tissue. I first performed scRNAseq analysis of human bronchial lesions from 7 patients to address this.

3.2 Aims

- To identify T cell clusters that are significantly enriched in high-grade lesions compared to normal airways
- To compare the phenotype and transcriptional profile of Tregs which I found to those Treg subsets that were found in established NSCLC
- To investigate transition states of CD4 subsets in human lesions and PBMC that have shared TCRs with BATF+Tregs in human lesions, so I could explore the potential for these cells in early detection via the blood

3.3 Results

3.3.1 BATF+Tregs are significantly increased in human CIS

To deeply study the T cell compartment during LUSC carcinogenesis in human pre-invasive lesions, fellow PhD student Amber Rogers from my lab sorted live single cells from single-cell digests of lung lesions via FACS from normal samples (n = 6), low-grade samples (n = 1), and high-grade samples (n = 10). Samples collected during autofluorescence bronchoscopies were cut using sterile scalpel blades and sequenced by the Novogene company. I processed raw files for human scRNAseq analysis using CellRanger v.7.0.1 (see Materials and Methods section scRNAseq analysis section). After QC, I obtained 33,556 cells in total. Then I performed clustering via Seurat v.4.1.0 (see methods, scRNAseq analysis section). I next identified T cells from whole cell types using the Gene Set Enrichment Analysis (GSEA) method based on the expression of T cell markers (CD3D, CD3E, CD3G), which yielded 6,285 T cells (Fig. 3.1A-B).

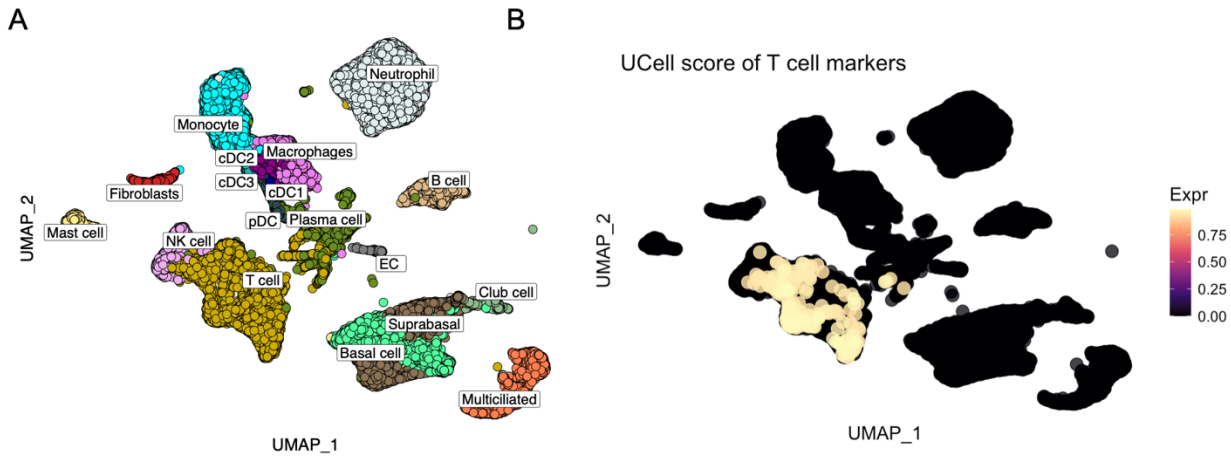

Figure 3.1 Selection of T cell cluster from scRNAseq of whole cell types in human pre-invasive pulmonary disease. A) UMAP showing whole cell types in pre-invasive disease (n = 17 samples, 7 patients, 33556 cells in total). B) UMAP presenting UCell score of T cell gene markers (*CD3D*, *CD3E*, *CD3G*) (n = 17 samples, 7 patients, 33556 cells in total). A T cell cluster was used to perform re-clustering again by increasing the resolution of clustering to study T cell immunology in pre-invasive disease.

Next, I performed re-clustering of T cells and used GSEA to identify five CD8 T cell subsets defined by key marker genes: Effector memory (Tem) CD8 (*EOMES, DUSP2, GZMK*), memory (Tm) CD8 (*CD55, ANXA1)*, mucosal-associated invariant T cells (MAIT) (*SLC4A10, ZBTB16*), tissue-resident memory (Trm) CD8 (*HOPX, XCL1, ZNF683)*, and exhausted (Tex) CD8 (*CXCL13, LAG3, PDCD1*) (Fig. 3.2C). I also resolved four FOXP3 negativeCD4 Tconv cell subsets: Progenitor CD4 (*CCR7, TCF7, SELL*), T helper 17 (Th17) CD4 (*RORA, IL17A, IL17F*), T follicular helper (Tfh) CD4 (*TSHZ2, CXCR5, CXCL13*), and proliferating (*MKI67, MCM5*). Most relevantly, I found two distinct *FOXP3*+ Treg subsets: BATF+ Tregs and Tcm-like Tregs (Fig. 3.2A). Tcm-like Tregs were labelled as such based on high expression of *S1PR1, AES*, and *KLF2*, whereas BATF+Tregs were identified by high expression of *FOXP3, BATF, CCR8, IL2RA, TNFRSF18, TNFRSF4, IRF4, CTLA4, ENTPD1, CD177, EBI3*, and *LAYN* (Fig. 3.2B).

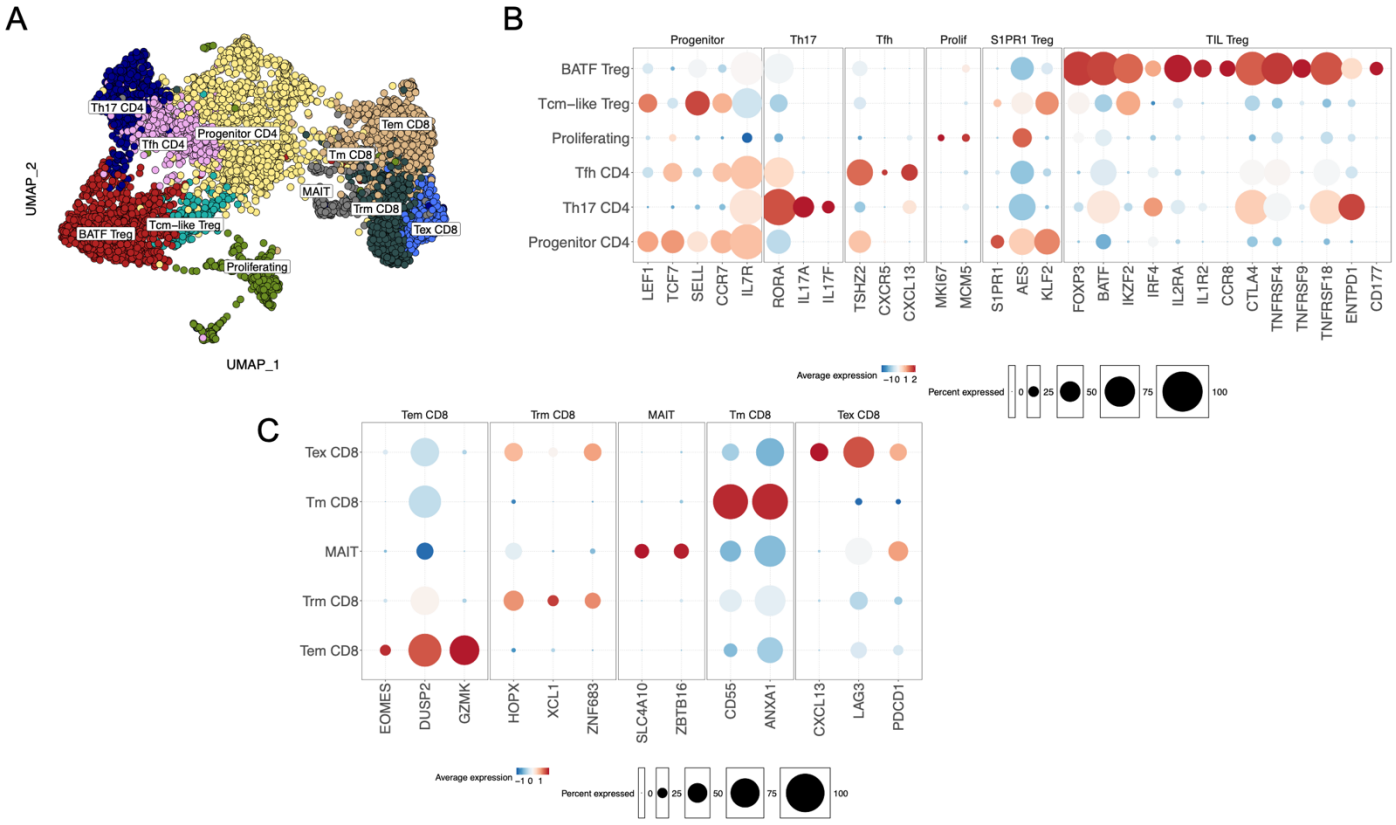

Figure 3.2 – Study of T cells in pre-invasive pulmonary disease using scRNAseq of human bronchial lesions. A) UMAP showing T cells in pre-invasive disease (n = 17 samples, 7 patients, 6,285 cells in total). B) Dotplot portraying expression of gene modules amongst all CD4 T cell subsets in pre-invasive disease. The gene modules includes progenitor module (*LEF1, TCF7, SELL, CCR7, IL7R*), Th17 module (*RORA, IL17A, IL17F*), Tfh module (*TSHZ2, CXCR5, CXCL13*), Proliferation module (*MKI67, MCM5*), S1PR1 Treg module (*S1PR1, AES, KLF2*), and TIL Treg module (*FOXP3, BATF, IKZF2, IRF4, IL2RA, IL1R2, CCR8, CTLA4, TNFRSF4, TNFRSF9, TNFRSF18, ENTPD1, CD177*). C) Dotplot presenting expression of gene modules amongst all CD8 T cell subsets in pre-invasive disease. The gene modules include Tem CD8 (*EOMES, DUSP2, GZMK*), Trm CD8 (*HOPX, XCL1, ZNF683*), MAIT (*SLC4A10, ZBTB16*), Tm CD8 (*CD55, ANXA1*), Tex CD8 (*CXCL13, LAG3, PDCD1*).

Next, to examine which T cell clusters were enriched in the CIS vs normal tissue I calculated the proportion of each subset amongst T cells. BATF+Tregs were significantly increased in high-grade pre-malignant lesions compared to normal (p = 3.61e-04, q= 3.97e-03, where q is the FDR corrected value, Fig 3.3A). Interestingly, no other subset reached significance with or without FDR correction in this data set. This result suggests that BATF+Tregs were increased in human central airway CIS and thus during LUSC carcinogenesis. To confirm whether BATF+Tregs were enriched in high-grade lesions compared to normal, amongst other T cell populations, I used an alternative statistical method. I performed differential abundance analysis of T cells between high-grade lesions and normal tissues using miloR, a method that utilises neighbourhoods of cells identified independently and then maps these back onto predefined clusters. It firstly constructs a k-nearest neighbour (KNN) graph to represent high-dimensional cell relationships and defines overlapping neighbourhoods around selected index cells using a refined sampling strategy to ensure coverage and reduce multiple testing burden. Cell counts within each neighbourhood, stratified by experimental samples, are modelled using a negative binomial generalised linear model (GLM) that accommodates complex experimental designs and adjusts for technical covariates (123) (see methods). BATF+Tregs were significantly enriched in human CIS compared to normal via MiloR (p= 2.65e-05, avglogFC= 5.13, Fig. 3.3B). This was visually confirmed when mapping MiloR neighbourhoods back onto the T cell UMAP (Fig. 3.3C).

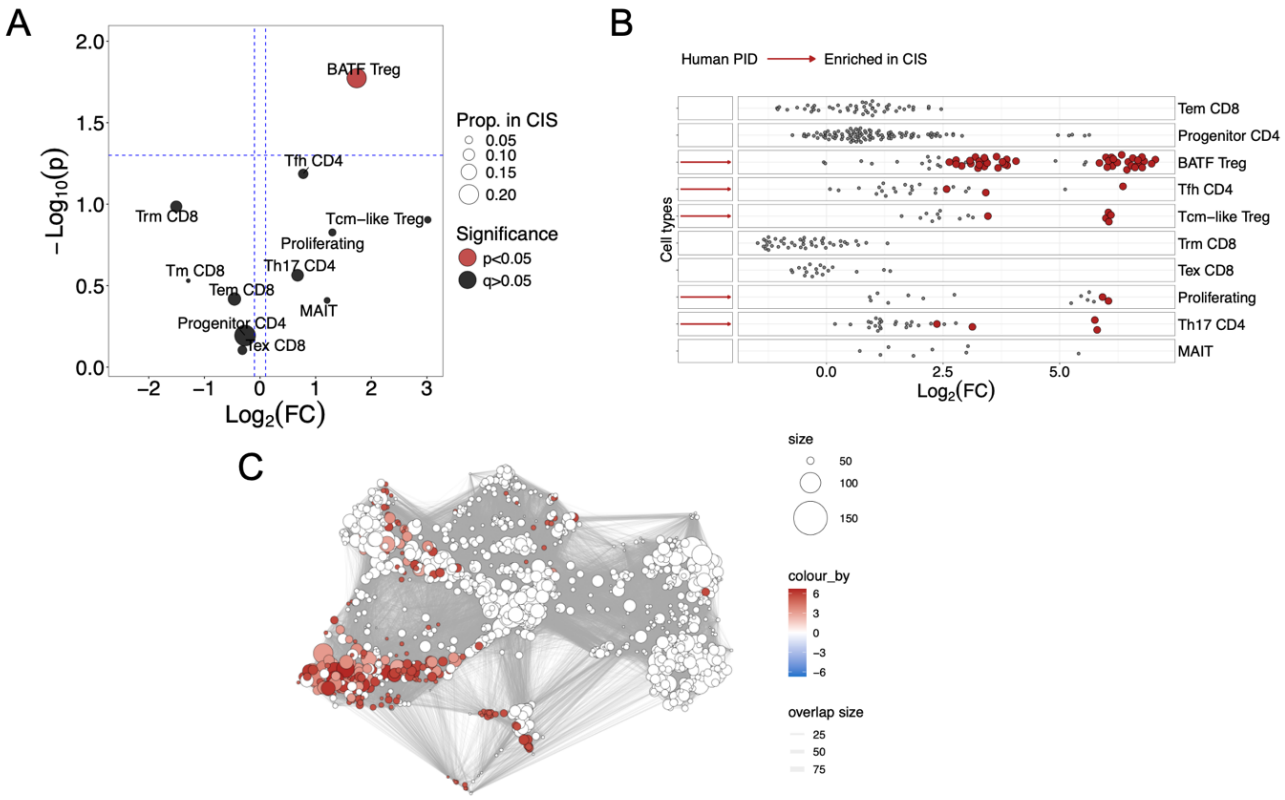

Figure 3.3 – Early accrual of BATF+Tregs during human pulmonary carcinogenesis. A) Volcano plot showing the proportion of T cell subsets between human CIS and normal lung. An unpaired t-test was used. B) Beeswarm plot displaying differential abundance of T cells between normal and human CIS samples using miloR: a significant neighbourhood that was enriched in CIS shown in red colour, and those enriched in normal are shown in blue. C) MiloR projection onto T cell UMAP embedding and a graph representation of neighbourhoods. Each node represents a neighbourhood, coloured by the log fold change (logFC) in abundance between CIS (red) and normal (blue). The colour gradient indicates differential abundance, with neighbourhoods enriched in CIS shown in red and those enriched in normal shown in blue. Basically, miloR builds many small, overlapping local neighbourhoods of cells in a low-dimensional space (for example, a k-nearest neighbour graph), which allows it to capture subtle and continuous changes in cell states. For each neighbourhood, miloR counts how many cells come from each condition and uses a generalised linear model to test whether these counts differ while accounting for sample-level variation and covariates. Neighbourhoods coloured in red indicate a significantly higher abundance in the CIS condition, whereas neighbourhoods coloured in blue indicate a significantly higher abundance in the normal condition

Next, to confirm whether the gene signature of these BATF+Tregs was enriched considerably in human CIS at the gene level, I performed pseudobulk analysis of CD4 T cells between human CIS and control using edgeR (see methods). This was an important third confirmation as it meant that cell subset number differences did not affect the result, since each sample is treated as a single bulk RNAseq replicate. *BATF, FOXP3, IKZF2, IL2RA, CD177, TNFRSF4, TNFRSF18, CCR8, CTLA4, and IL10* were significantly enriched in human CIS compared to normal (Fig. 3.4). This indicates that BATF+Tregs and their signature were significantly increased in T cells of human CIS compared to normal at both the cell proportion and bulk gene expression levels. These data represent the first identification of a highly suppressive Effector Treg subset as a key feature of human T cell remodelling in preinvasive lesions during LUSC development.

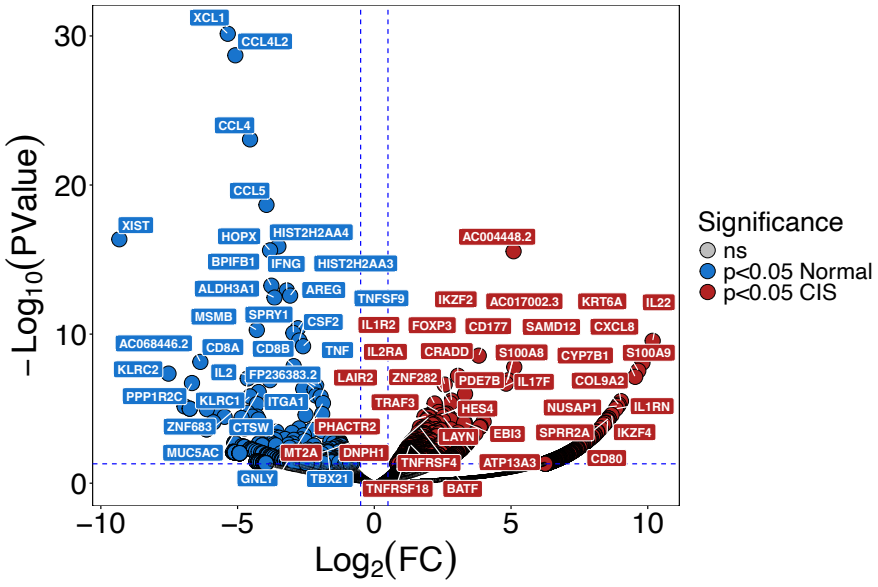

Figure 3.4 – Differentially expressed gene analysis of CD4 T cells between human CIS and normal lung tissue using scRNAseq of human pre-invasive disease. Volcano plot showing differentially expressed genes in CD4 T cells between normal (blue) and CIS (red) samples using edgeR. Significant genes in both human CIS and normal lung tissues based on logFoldChange more than 0.5 (enriched in human CIS) and lower than 0.5 (enriched in normal lung tissues) with p-value lower than 0.05.

3.3.2 BATF+Treg signatures derived from human CIS associate with anti-PD1 resistance in external NSCLC data sets.

BATF+Tregs are found in late stages of cancers (e.g., NSCLC), where they suppress effector T cells *in vitro* (67), enhance tumour growth *in vivo* (131), and were associated with poor survival and anti-PD1 resistance in NSCLC (86). To investigate phenotypes of BATF+Tregs and compare transcriptional profiles of BATF+Tregs to other Treg subsets, I performed pseudobulk analysis of BATF+Tregs vs other CD4 subsets in scRNAseq data to derive a BATF+Treg CIS signature. I found 415 significant genes based on the cutoff (logFC = 0.5, p-value = 0.05). Next, I rearranged those significant genes based on their logFC values, and I selected the Top 100 genes to represent BATF+Tregs in human PID. To compare the transcriptional profile of BATF+Tregs in human PID to other Treg subsets that were published, I performed GSEA analysis on BATF+Tregs using gene signatures of Treg subsets from Dykema et al. (86) and Guo et al. (132) , two landmark scRNAseq studies of T cells in NSCLC. I compared the transcriptome of BATF+Tregs to over 44 human CD4 T cell signatures derived from scRNAseq studies of human NSCLC and corrected for multiple tests with FDR correction (see methods GSEA section). BATF+Tregs showed significant enrichment ($q<0.05$) of six highly suppressive and / or effector CD4 Treg signatures, but in contrast showed no enrichment or negative enrichment for 14 transitional or early progenitor Treg signatures derived from cells with lower suppressive potential. Notably, BATF+Tregs showed the most substantial enrichment of C2-Activated (3) Tregs (NES = 2.57, q = 1.63e-03), C0-Activated (1) OX40hi/GITRhi Tregs (NES = 2.51, q= 2.61e-02), Suppressive TIL Tregs (NES = 2.1, q = 3.42e-02) (Fig 3.5A), which are overlapping, similar clusters from these studies associated with the most potently suppressive effector Treg cell types and which defined adverse clinical outcomes (86). CIS, in particular, the OX40hi GITRhi Tregs in the Dykema study were associated with anti-PD1 resistance in NSCLC (86). Next, I wanted to verify whether BATF+Tregs were associated with anti-PD1 resistance in NSCLC in an independent cohort where I could also test effects of potential confounding variables like total infiltration and tumour mutational burden.

To investigate the association between BATF+Tregs and anti-PD1 resistance in NSCLC, a large-scale study of anti-PD1 in metastatic NSCLC with paired WES and RNAseq data (N = 195 patients) (133) was utilised. I calculated the gene signature of BATF+Tregs found in human CIS using geometric mean, and then univariate analysis was performed. I found that the gene signature of BATF+Tregs defined non-responders to anti-PD1 treatment (OR= 0.45, p = 0.032). To examine whether this association was independent of TMB, CD8, and CD4 infiltration, multivariate analysis was performed (see methods). The gene signature of BATF+Tregs was associated with anti-PD1 resistance in metastatic NSCLC independently of TMB, CD8, and CD4 infiltration (OR = 0.87, p = 0.026, Fig 3.5B). This data indicates that BATF+Tregs infiltrating CIS lesions were associated with anti-PD1 resistance in metastatic NSCLC, independent of infiltration or TMB, implying that the presence of these cells, may plausibly impede use of PD-1 targeting agents in the interception setting. Although I note that functional data would be required to confirm this, BATF+Treg signatures could also be of use to predict who might benefit from PD-1 mediated interception, should further clinical and experimental data support these observations. Such biomarkers will be key because of the mentioned potential in interception to overtreat, the emphasis in this new setting is on first doing no harm.

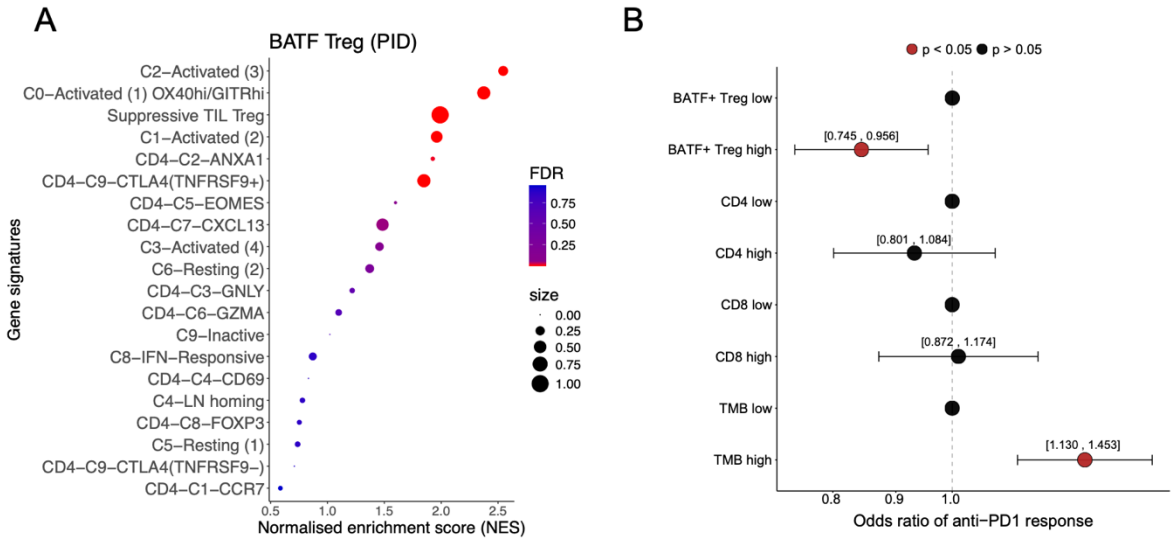

Figure 3.5 – An association of BATF+Tregs for anti-PD1 resistance in metastatic NSCLC. A) Dotplot showing normalised enrichment score of BATF+Tregs using gene signatures from Guo et al., and Dykema et al. The *FindAllMarkers* function was used to extract differentially expressed genes in BATF+Tregs, and then GSEA was used to compare the

transcriptional profiles of Treg subsets found in NSCLC at the cellular level. FDR was calculated by Benjamini-Hochberg, whereas size was the scale of overlapped genes between differentially expressed genes from BATF+Treg and signature genes of Treg subsets from Guo et al. and Dykema et al.. B) Forest plot presenting a discrete model of multivariate analysis of BATF+Tregs adjusted by TMB, CD4, and CD8 using paired WES and bulk RNAseq of metastatic lung cancer treated with anti-PD1 treatment. The expression of the BATF+ Treg gene signature, CD4, and CD8 infiltration was calculated using the geometric mean. Then, the expression of BATF+Treg gene signature, CD4, CD8 infiltration and TMB was classified into low and high using the median function.

3.3.3 BATF+Tregs in human CIS are clonally expanded and share TCRs with local progenitor CD4 T cells

To investigate which CD4 subsets in human CIS could differentiate into CIS BATF+Tregs, I calculated shared TCRs between BATF+Tregs and other CD4 subsets in human CIS lesions using scRNAseq+TCRseq, in 7 samples from 3 patients where we had the 10x VDJ kit output. I found that most BATF+Treg clones were private (unshared), and that 7.72 % of BATF+Tregs shared TCRs with any other cells. 92.28% of BATF+Tregs that shared clones with another cell did so within the BATF+Treg cluster, highlighting that these cells are i) oligoclonally expanded within human high-grade premalignant lesions and ii) likely not fully dependent on differentiation from another cell type and therefore are iii) probably undergoing in situ antigen or cytokine-driven expansion/maintenance, perhaps via contact with MHCII APCs. 7.72% of BATF+Tregs did share TCRs with other non-BATF+Treg cells in the TME.

I broke this down to understand possible differentiation within the lesions. 3.86% shared with progenitor CD4, 2.98% with Th17 CD4, 0.7% with Tcm-like Treg, and 0.18% with Tfh CD4 (Fig. 3.6A). This suggests that, based on within CIS tissue overlap, BATF+Tregs are mostly maintained via in situ turnover, but that a notable fraction could be differentiated from progenitor CD4, Th17 CD4, and Tfh CD4 in the human CIS environment. This could mean that a subset of this cluster forms via iTreg

conversion from Tconv CD4 under skewing conditions (tolerogenic APC, TGF-β, for example (41)) or that both Tconv and BATF+Treg lineages are established from the same naïve precursor clone. In either case, this supports that these cells are partly induced from the naïve or non-Treg pool (i.e. not entirely from pre-committed thymic Tregs). To assess whether in situ conversion may be predicted via an unsupervised model, I performed differentiation prediction using TCR overlapping T cell populations with trajectory and pseudotime analysis (monocle2) (see methods). Progenitor CD4 T cells in the human CIS lesion were predicted to be an early stage of BATF+Treg differentiation in this model (Fig. 3.6B-C). Interestingly, the data also show heterogeneity in the BATF+Treg pool which suggests that minor subclusters may also be at earlier stages of differentiation, possibly explained by recent migrants into the TME which have yet to fully differentiate, which I will explore downstream.

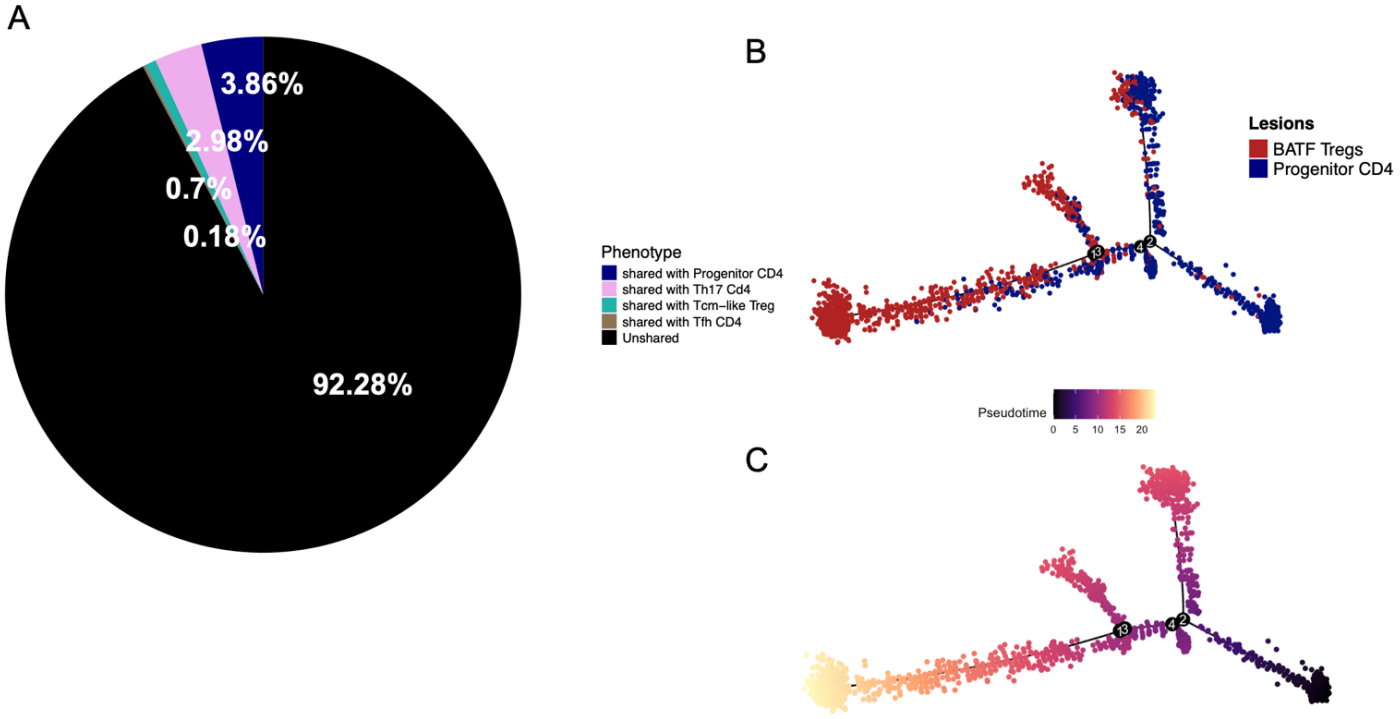

Figure 3.6 – TCR overlap between BATF+Tregs and all CD4 subsets in human CIS lesions. A) pie chart showing the percentage of CD4 T cells that shared TCRs with BATF+Tregs in human CIS (7 samples from 3 patients). B) trajectory analysis between BATF+Tregs and progenitor CD4 using monocle2. C) Pseudotime analysis between BATF+Tregs and progenitor CD4, which had the highest shared TCRs with BATF+Tregs in human CIS using monocle2 (7 samples from 3 patients). Monocle2 works by ordering individual cells along a continuous trajectory called pseudotime, which represents the

progression of cells through a biological process rather than actual clock time. Monocle2 reduces the complexity of gene expression data and reconstructs branching trajectories, allowing it to identify points where cells diverge into different fates.

Next, I performed differentially expressed gene analysis of progenitor CD4 between CIS and normal using the *FindMarkers* function in Seurat, because I sought to examine whether progenitor CD4 could differentiate into BATF+Tregs in human CIS, in which case these cells may show a Treg-biased programme even in their progenitor state. The strengths of *FindMarkers* in Seurat are that it is very sensitive to detecting cell state differences, and good for identifying transcriptional programs that define cell states, and does so even if some patients contribute very few cells. I found that *BATF, FOXP3, TNFRSF4*, and *TNFRSF18* were significantly increased in CIS vs normal progenitor CD4 T cells (Fig. 3.7). *FOXP3* is the transcriptional factor that programs Tregs. Induction of *FOXP3* in progenitor CD4 T cells could drive their differentiation into iTregs (under TGF-β and IL-2 conditions). Moreover, *BATF* is the transcriptional factor that plays a role in the differentiation and activation of Tregs within the TME. *BATF* acts as a key epigenetic regulator that controls the chromatin remodelling of TIL-Tregs and promotes their activation-associated gene expression and suppressive function. Together with the clonal overlap analysis, this suggests that CIS progenitor CD4 could differentiate into CIS BATF+Tregs.

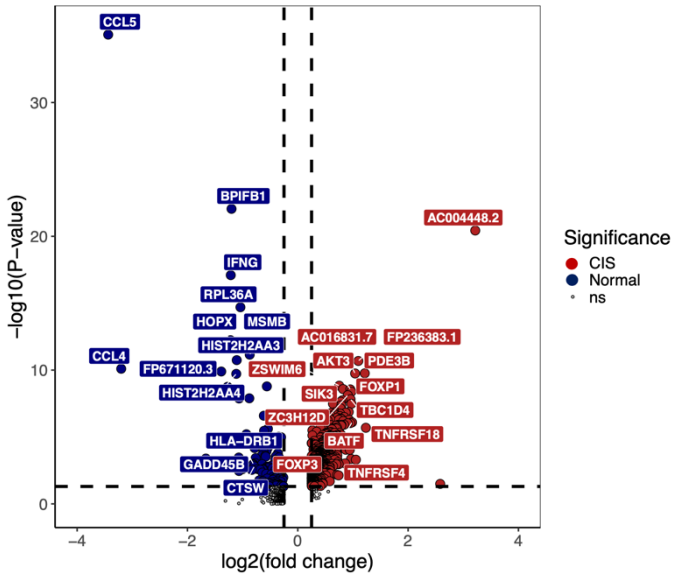

Figure 3.7 Differentially expressed gene analysis of progenitor CD4 between CIS and normal. Volcano plot showing differentially expressed gene analysis of the progenitor

CD4 cluster between human CIS and normal lung tissues (7 samples from 3 patients that have TCR data) using the *FindMarkers* function in the Seurat package. *FindMarkers* function in Seurat is used to identify genes that are differentially expressed between groups of cells, such as distinct cell states or stages of differentiation, in scRNAseq data. By statistically comparing gene expression levels between predefined cell populations, *FindMarkers* highlights genes that are upregulated or downregulated as cells transition from one state to another. These marker genes help define the molecular identity of each cell state and provide insight into the biological processes underlying cell differentiation and also functional specialisation. Significant genes in both human CIS and normal lung tissues based on logFoldChange more than 0.5 (enriched in human CIS) and lower than 0.5 (enriched in normal lung tissues) with p-value lower than 0.05.

3.3.4 Remodelling of MHC class II signalling networks in human CIS may explain BATF+Treg expansion

Since most BATF+Tregs seemed to be maintained via in situ expansion, I reasoned that I could find an APC that might drive BATF+Treg stimulation via MHCII. To identify crosstalk axes that could support BATF+Treg formation, I analysed all cell types of human lesions (n= 17 samples). After QC, I obtained a total of 33,556 cells. More cells were derived from CIS (total= 26,703 cells) than normal (total = 2226 cells) and low-grade lesions (total = 4,627) with an average of 4,793 cells per patient. After clustering, 18 distinct cell types were identified by using GSEA with 59 gene signatures from scRNAseq of human lung cancer atlas (116) (Fig. 3.8A). This empirical method was used alongside manually identified differences in marker genes inspected via gene expression dot plot (Fig. 3.8B). Next, to identify which cell types are significantly enriched in human CIS compared to normal airways, differential abundance analysis of whole cell types was performed using miloR (see methods). B cell, T cell, suprabasal, neutrophil, monocyte, basal cell, cDC2, plasma cell, macrophage, mast cell, NK cell, fibroblast, pDC, cDC3 were significantly increased in human CIS (Fig 3.8C). This suggests that an increased B cell and cDC2 pool might better support BATF+Treg formation in human CIS.

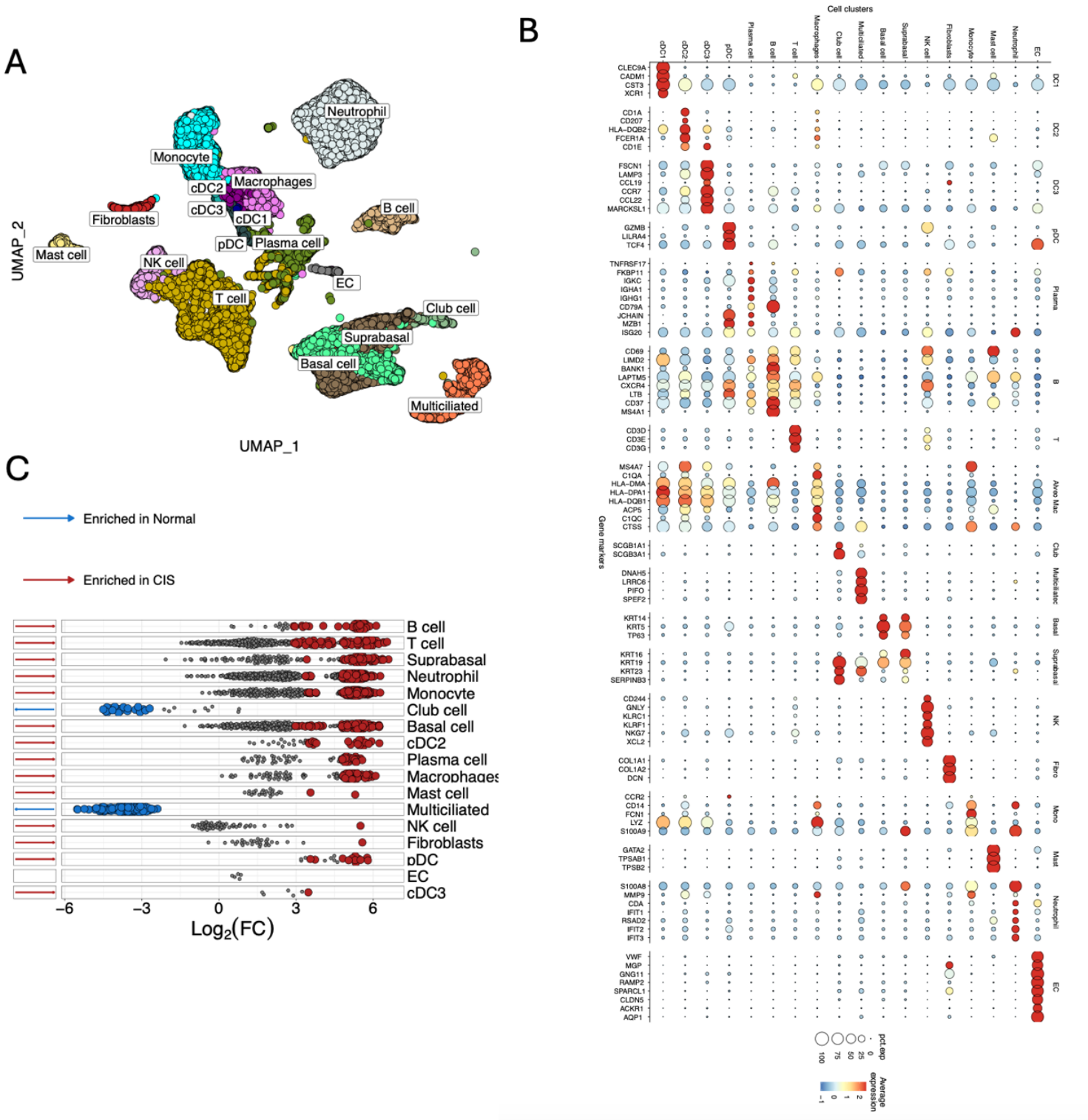

Figure 3.8 – Study of whole cell types in pre-invasive tumour microenvironment using scRNAseq data. A) UMAP showing whole cell types in human PID (N=17 samples, 7 patients, 33,556 cells in total). B) Dotplot presenting expression of gene modules amongst whole cell types in human PID (N = 17 samples, 7 patients, 33,556 cells in total). Each dot

size shows the percentage of expression of genes on each cell type and the colour gradient presents the average of expression of genes on each cell type: high expression in red, whereas low expression in blue. C) Beeswarm plot showing the differential abundance of whole cell types between human CIS and normal using miloR. Each node represents a neighbourhood, coloured by the log fold change (logFC) in abundance between CIS (red) and normal (blue). The colour gradient indicates differential abundance, with neighbourhoods enriched in CIS shown in red and those enriched in normal shown in blue.

There are four major signals that could support BATF+Treg differentiation based on what we know of Treg maintenance within lung tissue and NSCLC: Signal 1 (MHC class II (MHC-II) signalling), Signal 2 (co-stimulatory and co-inhibitory signalling), Signal 3 (cytokine stimulation), and Signal 4 (chemotaxis). To investigate which cell types were enriched in the expression of the MHC-II gene signature in human CIS compared to normal at the cellular level, I calculated the UCell score of the MHC-II gene signature (*HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DMA, HLA-DMB, HLA-DQA2, HLA-DOA, HLA-DOB, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5*) on whole cell types between human CIS and normal in each patient. Then, I calculated logFC on each whole cell type (the expression of MHC-II in human CIS divided by the expression of MHC-II in normal lung tissue). Next, I calculated the average logFC of the differential abundance of whole cell types using miloR on each whole cell type. Then, a two-way volcano plot was constructed. I found that fibroblasts, plasma cells, monocytes, B cells, and cDC2 were enriched in expression of the MHC-II gene signature and in differential abundance in human CIS, whereas basal cells were enriched in expression of the MHC-II gene signature and in differential abundance in normal lung tissue (Fig. 3.9). This suggests that an enriched B-cell, fibroblast, plasma, monocytes, and cDC2 pool might support BATF+Treg formation in the human CIS lesions.

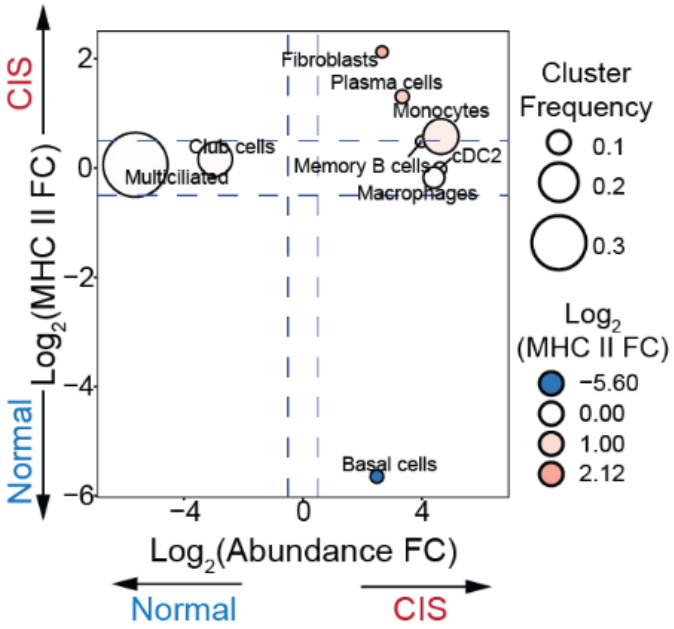

Figure 3.9 – Enrichment of MHC-II signalling of whole cell types between CIS and normal lung tissue from the UCLH Bronchoscopy Surveillance Cohort (N = 17 samples, 7 patients). The two-way volcano plot displays the log fold change in MHCII between cell types in normal and CIS on the y-axis, with the log fold change in cell abundance using miloR on the x-axis at the cellular level. UCell was used to calculate MHC-II expression. MHC-II gene signature includes *HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DMA, HLA-DMB, HLA-DQA2, HLA-DOA, HLA-DOB, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5.*

Next, I analysed MHC-II signalling between antigen-presenting cells, basal, suprabasal cells, and fibroblasts as senders and BATF+Tregs as receivers using CellChat (see methods of CellChat section) (118). CellChat is an open-source computational tool designed to infer, analyse, and visualise intercellular communication networks from scRNAseq data. The purpose of CellChat is to identify how different cell types communicate via ligand-receptor interactions, which is essential for understanding processes such as development, immune responses, and disease progression (118). I subset basal cells, fibroblasts, antigen-presenting cells, and BATF+Tregs, and then CellChat analysis was performed on human CIS and normal separately. I found that there were 12 MHC class II signalling axes from basal and suprabasal cells that were significantly enriched in normal compared to human CIS, whereas 12 MHC class II signalling axes from cDC1 and 11 MHC class II

signalling from pDC were significantly enriched in human CIS compared to normal (Fig. 3.10). This suggests that antigen-presenting cells (e.g., pDC, cDC1) would support BATF+Treg formation in human pre-malignant lesions during LUSC carcinogenesis.

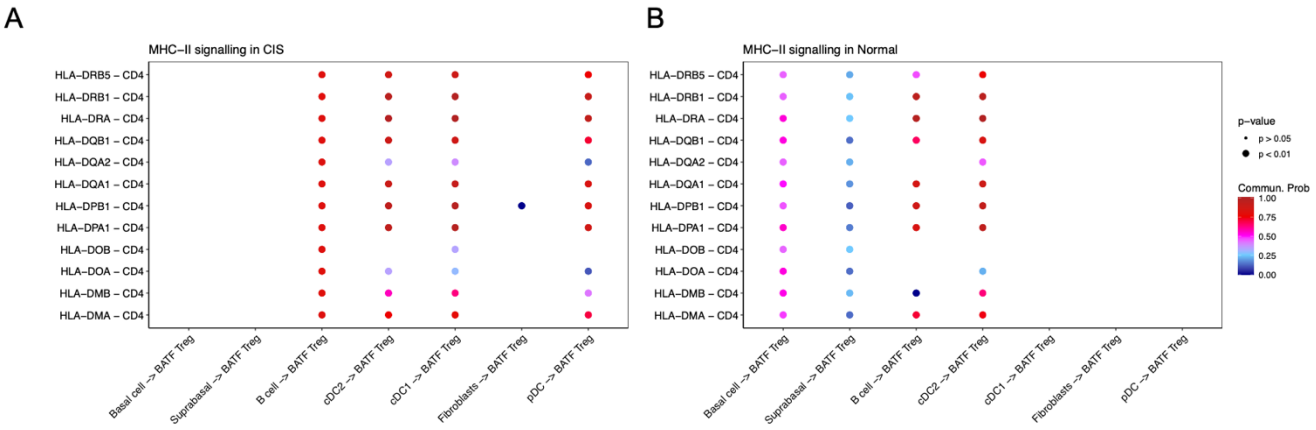

Figure 3.10 – Crosstalk axes of MHC class II signalling of antigen-presenting cells, basal, suprabasal, fibroblasts as senders, and BATF+Tregs as receivers between human CIS and normal using CellChat (N = 17 samples, 7 patients). A) Dotplot showing crosstalk axes of MHC class II signalling of antigen-presenting cells, basal, suprabasal, fibroblasts as senders and BATF+Tregs as receivers in human CIS. B) Dotplot presenting crosstalk axes of MHC class II signalling of antigen-presenting cells, basal, suprabasal, fibroblasts as senders and BATF+Tregs as receivers in normal. The colour gradient presents the communication probability that shows the interaction strength of ligands from senders (antigen-presenting cells, basal, suprabasal, fibroblasts) and receptors from receivers (BATF+Tregs) that was calculated by CellChat: high communication of probability shown in red, whereas low communication of probability shown in blue. CellChat estimates the probability and strength of signalling interactions between cell populations and summarises these interactions at both individual ligand and receptor and pathway levels.

While scRNA-seq robustly identifies an enrichment of BATF+Tregs in high-grade lesions (including CIS), this approach inherently lacks spatial resolution. Because tissue dissociation disrupts native architecture, these data cannot determine whether BATF+Tregs are located within the dysplastic epithelial compartment (intralesional) or predominantly in the adjacent stromal that surrounds CIS. Consequently, the observed increase in BATF+Tregs should be interpreted primarily

as a lesion-associated expansion rather than definitive evidence of intraepithelial immune suppression.

This spatial uncertainty is particularly relevant when interpreting early immune escape mechanisms. For example, BATF+Treg enrichment could reflect (i) direct suppression of effector cells inside CIS epithelium, (ii) establishment of a suppressive perilesional barrier that limits immune entry or function at the epithelial-stromal interface, or (iii) broader field-level immunoregulation in chronically inflamed airway mucosa. These scenarios have different implications for how and when immune escape is initiated, and they cannot be distinguished using scRNAseq alone.

Next, I performed crosstalk axes analysis between antigen-presenting cells, suprabasal, basal and fibroblast cells as senders and BATF+Tregs as receivers using CellPhoneDB (119). This type of cell-cell interaction enables an estimate of specific receptor-ligand pairs based on human databases. Firstly, I performed differential expressed gene analysis of whole cell types to identify differentially expressed genes of those cell types on each pathology separately and then CellPhoneDB was performed (see methods). I observed 53 crosstalk axes that were enriched in human CIS compared to normal, based on crosstalk interaction strength between human CIS and normal (Fig. 3.11A). Next, I also validated 53 crosstalk axes using bulk transcriptomics of 122 human bronchial biopsies from 77 smoker patients. I found potential 13 crosstalk axes that were enriched in human CIS compared to normal (Fig. 3.11B-C) (Table 3.1). These results underpinned a CRUK biology to prevention grant awarded to my team in 2023 to functionally explore additional immune interception targets for LUSC.

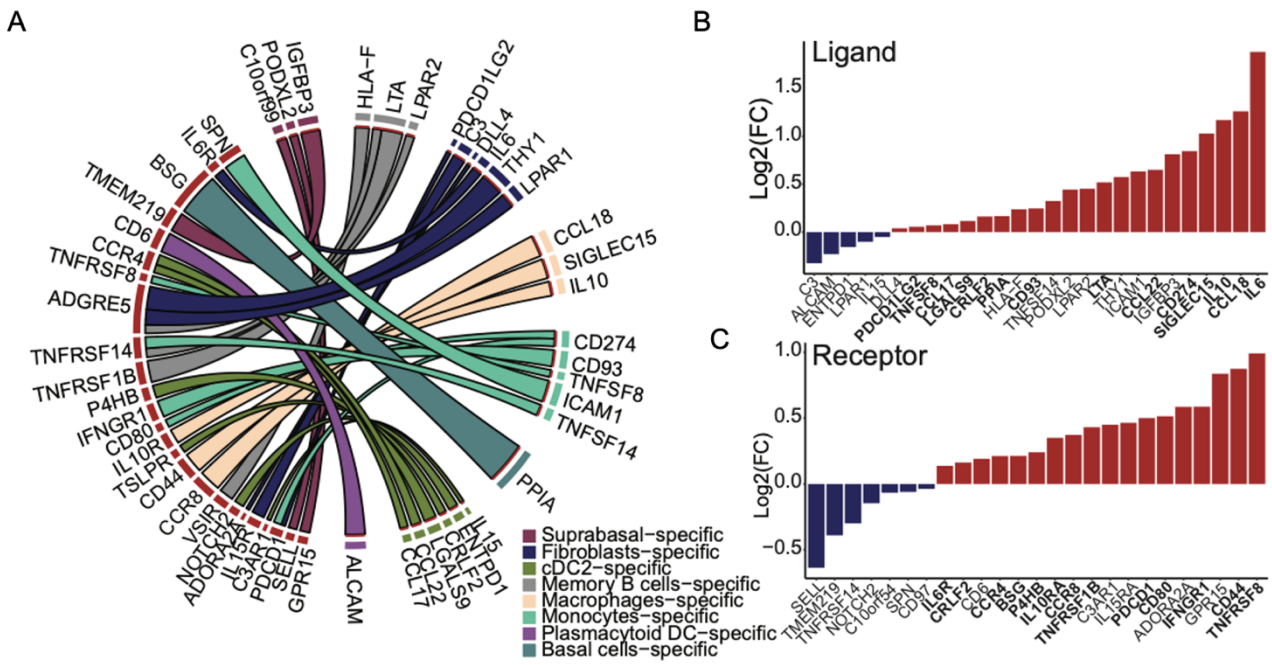

Figure 3.11 – Enriched crosstalk axes in human CIS between antigen-presenting cells, basal, suprabasal, fibroblasts as senders and BATF+Tregs as receivers using CellPhoneDB. A) Chord diagram showing cell-cell communication between antigen-presenting cells, basal, suprabasal, fibroblasts as senders and BATF+Tregs as receivers using CellPhoneDB. Each colour on the chord diagram represents ligand-specific senders (suprabasal, fibroblasts, cDC2, B cells, macrophages, monocytes, pDC, basal cells) and receptors on BATF+Tregs, shown in red. B) Barplot presenting validation of ligands by logFC using bulk transcriptomics of 122 human biopsies from 77 smoker patients: enriched in high-grade lesions in red; enriched in normal in blue.C) Barplot showing validation of receptors by logFC using the bulk transcriptomics data: enriched in high-grade lesions in red; enriched in normal in blue. The logFC was calculated by log2 of the average expression of the ligands/receptors in human high-grade lesions, divided by the average expression of ligands/receptors in human normal lung tissues.

Table 3.1 – Potential crosstalk axes that are enriched in high-grade lesions using the bulk transcriptomics data from Mascaux et al. for validation

Ligand	Receptor	Ligand-specific
PDCD1LG2	PDCD1	Fibroblasts-specific
CCL18	CCR8	Macrophages-specific
SIGLEC15	CD44	Macrophages-specific

CD274	CD80	Monocytes-specific
CD274	PDCD1	Monocytes-specific
CD93	IFNGR1	Monocytes-specific
LGALS9	P4HB	cDC2-specific
LTA	TNFRSF1B	B cells-specific
TNFSF8	TNFRSF8	Monocytes-specific
CCL22	CCR4	cDC2-specific
CCL17	CCR4	cDC2-specific
PPIA	BSG	Basal cells-specific
IL6	IL6R	Fibroblasts-specific

3.3.5. MHCII loss on basal cells and gain on APCs as an immune escape mechanism in LUSC carcinogenesis

To investigate which cell types were significantly enriched in the expression of the MHC-II gene signature in human CIS compared to normal at the patient level, I calculated the UCell score of the MHC-II gene signature (*HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DMA, HLA-DMB, HLA-DQA2, HLA-DOA, HLA-DOB, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5*) on whole cell types between human CIS and normal in each patient. I found that fibroblasts and B cells were significantly higher in expression of the MHC-II gene signature in CIS, whereas suprabasal and basal cells were significantly increased in expression of the MHC-II gene signature in normal (Fig. 3.12A). When UMAPs of whole cell types were facetted, it was visibly obvious that MHC-II expression in basal cells was decreased in human CIS lesions, whereas MHC-II expression in APCs was increased in human CIS lesions (3.12A-B). The MHC-II expression reduced in basal cells in human CIS indicates that effector CD4 T cells could not optimally detect tumour-associated antigens on CIS basal cells, whereas there is still a high chance for CD4 priming, expansion and differentiation (especially Tregs) in tolerogenic TME. Together, this may support the expansion of the Treg pool and unchecked tumour growth. Also, this suggests that an increased B-cell and fibroblast pool might support BATF+Treg formation in the human CIS lesions.

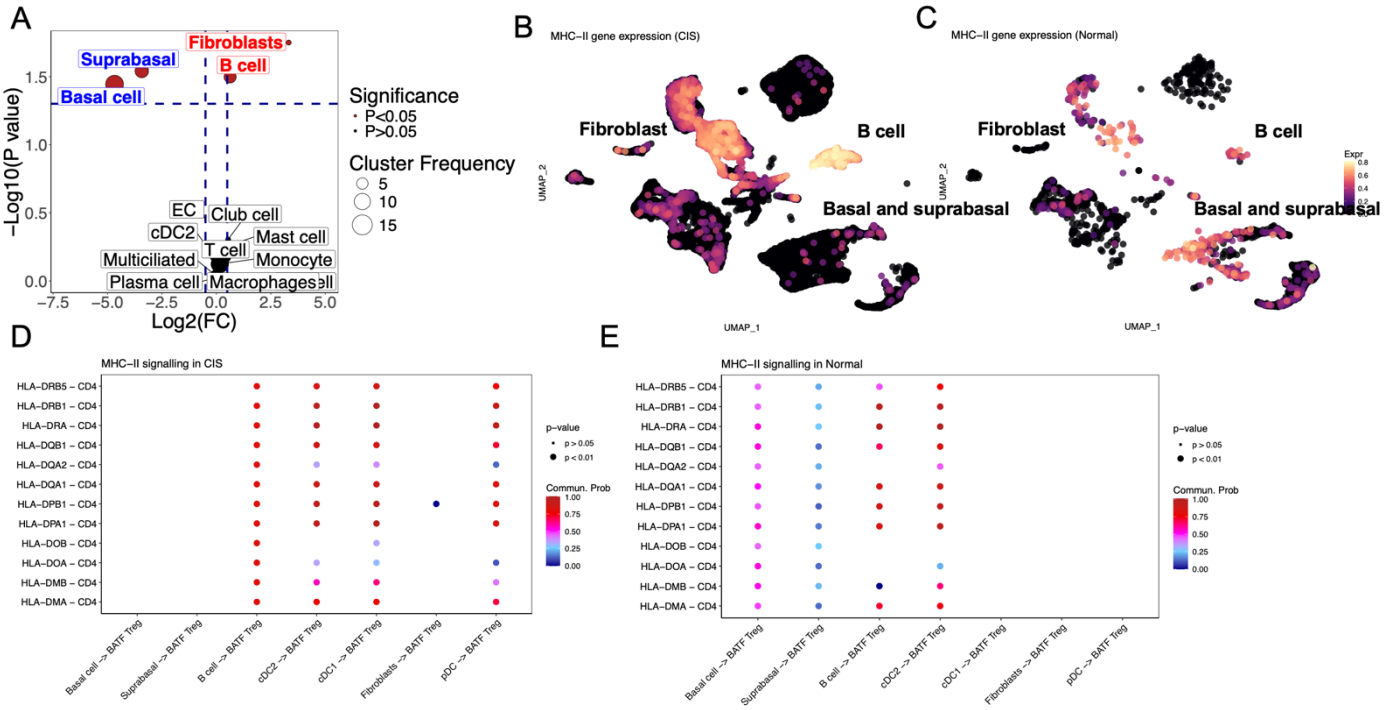

Figure 3.12 MHC-II expression loss on basal cells and gain on APCs as an immune escape mechanism in pre-invasive LUSC. A) The volcano plot shows the average MHC-II expression on each whole cell type at the patient level, which was calculated using the UCell package and an unpaired t-test was used for the statistical test. B,C) Faceted UMAPs of whole cell types showing UCell score of MHC-II expression between CIS and normal (N=17 samples, 7 patients). D,E) Crosstalk axes of MHC class II signalling of antigen-presenting cells, basal, suprabasal, fibroblasts as senders, and BATF+Tregs as receivers between human CIS and normal using CellChat. D) Dotplot showing crosstalk axes of MHC class II signalling of antigen-presenting cells, basal, suprabasal, fibroblasts as senders and BATF+Tregs as receivers in human CIS. E) Dotplot presenting crosstalk axes of MHC class II signalling of antigen-presenting cells, basal, suprabasal, fibroblasts as senders and BATF+Tregs as receivers in normal lung tissues. The colour gradient presents the communication probability that shows the interaction strength of ligands from senders (antigen-presenting cells, basal, suprabasal, fibroblasts) and receptors from receivers (BATF+Tregs) that was calculated by CellChat: high communication of probability shown in red, whereas low communication of probability shown in blue.

3.3.6 BATF+Tregs recirculate between the CIS lesions and blood, providing a basis for early detection

So far we have followed my analysis of BATF+Tregs within the lesion, where I have found evidence supporting possible differentiation and *in situ* maintenance via APCs. However, effector Tregs in human tissues and tumours are increasingly recognised as coming from or migrating to the blood (134,135). I therefore wondered if peripheral T cells may overlap with BATF+Tregs in the lesion. To study the T cell compartment during LUSC development in human PBMC, Amber Rogers sorted total non-naïve live T cells from human PBMC (n=5), and scRNAseq was performed. I processed raw files as above for human scRNAseq analysis using CellRanger v.7.0.1 (see methods of scRNAseq analysis section). After QC, I obtained 80,396 cells in total. Then I performed clustering via Seurat v.4.1.0 (see methods). Using the GSEA method, I identified three CD8 T cell clusters: Tem CD8 cells expressing *EOMES, SUB1,* and *GZMK*; GZMB+ late Tem/Temra CD8 cells characterised by *GZMB, KLRG1,* and *FGFBP2*; and a cycling population marked by *MKI67* and *MCM5*. In addition, five CD4 T cell clusters were identified, including naïve CD4 (*CCR7, SELL, LEF1, TCF7*), Tcm CD4 (*TOB1, FOXP1*), Tm CD4 (*S100A10, LGALS1, ANXA1*), Tem CD4 (*GZMK, GZMA, CCL5*), and Th17 CD4 T cells expressing *RORC, CCR6, CXCR6* with two distinct Treg subsets: effector Tregs and Tcm-like Tregs (Fig. 3.13A). Tcm-like Tregs were identified as they have high expression of *CD27* and *CARD16*, whereas Effector Tregs were identified as they have high expression of *FOXP3, BATF, IL2RA, CTLA4, TIGIT, IKZF2, TNFRSF9, and ENTPD1* (Fig. 3.13B).

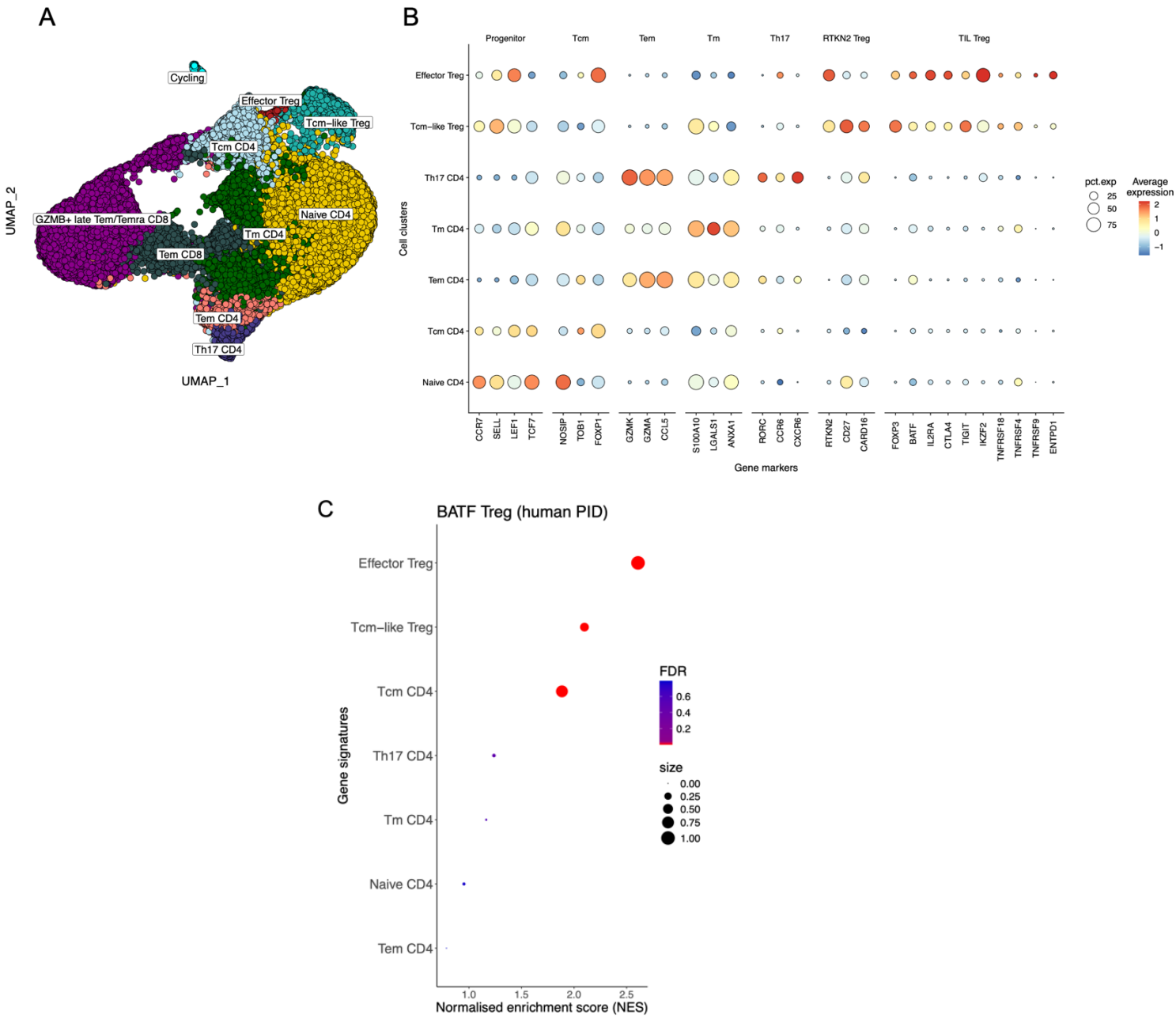

Figure 3.13 – Study of T cells in human PBMC using scRNAseq (N = 5 samples, 4 patients). A) UMAP showing T cell clusters in human PBMC (N = 5 samples, 4 patients, 80,396 cells in total). B) Dotplot presenting expression of gene modules amongst all CD4 subsets in human PBMC. Each dot size shows the percentage of expression of genes on each cell type and the colour gradient presents the average of expression of genes on each cell type: high expression in red, whereas low expression in blue. C) Dotplot showing normalised enrichment score amongst all CD4 subsets in human PBMC using differentially expressed genes from BATF Tregs in human PID. The fgsea package was used to calculate the normalised enrichment score. FDR was calculated by the Benjamini-Hochberg, whereas size

was the scale of overlapped genes between differentially expressed genes from BATF+Treg and signature genes of CD4 subsets in human PBMC.

Next, I compared the transcriptional profile of CD4 subsets in human PBMC and BATF+Tregs in human CIS lesions, and I found that effector Tregs had the strongest normalised enrichment score with BATF+Treg in human CIS (NES = 2.56, q = 1.43e0-4) (Fig. 3.13C). This suggests that gene signature of effector Tregs could be used for early detection, as they had a similar transcriptional profile to BATF+Tregs in human CIS

To investigate which CD4 subsets in human PBMC could differentiate into BATF+Tregs, I calculated shared TCRs between BATF+Tregs and other CD4 subsets in human PBMC. I found that 55.56% of Tcm-like Tregs and 34.92% of Effector Tregs shared TCRs with BATF+Tregs in human CIS (Fig. 3.14A). This suggests that BATF+Tregs could be differentiated from Tcm-like Tregs and Effector Tregs in the human PBMC. To assess whether this may be observed via an unsupervised model, I performed TCR overlapping T cell populations using trajectory and pseudotime analyses (monocle2) (see methods, pseudotime section). Tcm-like Tregs in human PBMC were predicted to be an early stage of BATF+Treg differentiation in human CIS lesions (Fig. 3.14B-C). Taken together, these data suggest BATF+Tregs and Eff.Tregs in PBMCs are end points of local and peripheral differentiation, and that progenitor CD4 T cells and Tcm-like Tregs are upstream of these. Having the TCRseq data also shows that this is a clonal differentiation program, and highlights that Tcm-like Tregs and effector Tregs are both possible contributors to the BATF+Tregs pool, together with the progenitor CD4 T cells in CIS lesions.

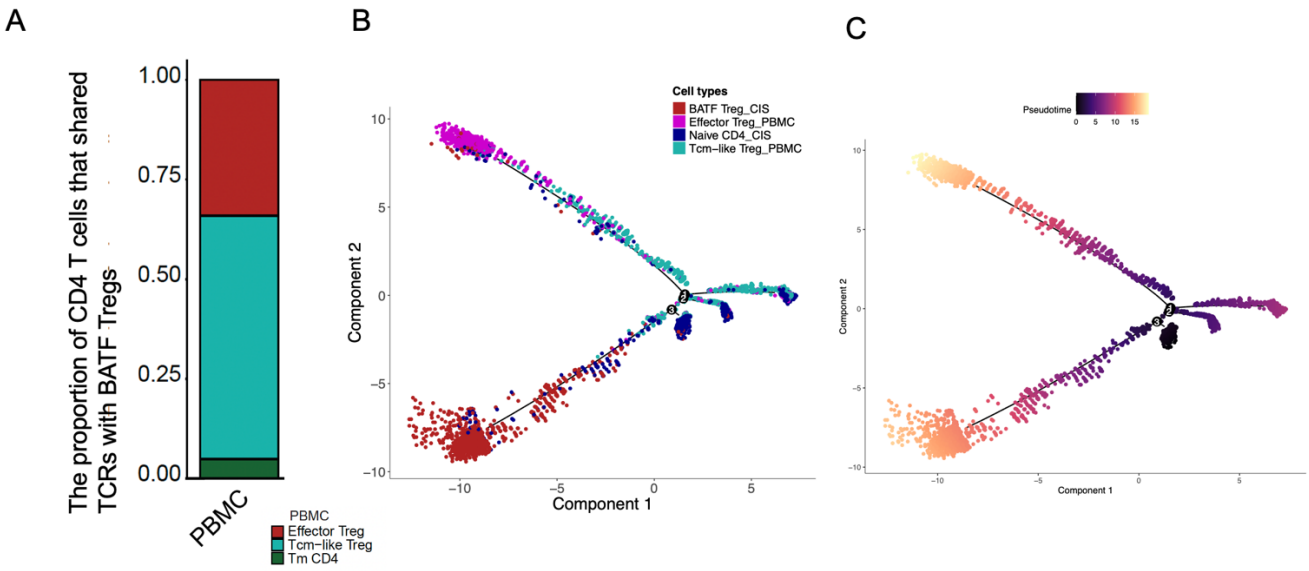

Figure 3.14 – The differentiation of CD4 subsets in human PBMC into BATF+Tregs in human CIS. A) Stacked barplot showing the proportion of CD4 subsets in human PBMC that shared TCR clones with BATF+Treg in human CIS lesions. B) Trajectory analysis between effector Tregs and Tcm-like Tregs in human PBMC, and Naïve-like CD4 and BATF+Tregs in human CIS using monocle2. C) Pseudotime analysis between effector Tregs and Tcm-like Tregs in human PBMC, and Naïve-like CD4 (only CD4 subsets that shared TCRs with BATF+Tregs in human CIS) and BATF+Tregs in human CIS using monocle2.

3.3.7 Validation of proportion of BATF+Tregs during LUSC carcinogenesis

The limitation of T cells from our scRNAseq data of pre-invasive disease is that we have a few patients, and the number of cells is limited. To overcome this obstacle and increase statistical power, I integrated our scRNAseq of T cells with scRNAseq of T cells from a publicly available lung cancer atlas. This atlas contains 250,214 cells from 271 donors across 24 studies of 123 diseases (116). In addition, I integrated data generated from an overlapping set of 11 samples (similar histology, but from different sites) from 3 patients in the UCLH bronchoscopy surveillance study (our UCLH cohort and PCGA cohort), which were generated via Dr Lukas Kalinke as part of the pre-invasive atlas initiative.

After QC, I obtained 250,214 T cells in total from 458 samples and 271 patients. Then I performed clustering via Seurat v.4.1.0 (see methods). Using the GSEA method, I identified eight CD8 T cell clusters: Tem CD8 (*EOMES, GZMK, SUB1*), MAIT (*SLC4A10, ZBTB16*), Tk CD8 (*FOSL2, REL, FAM3C*), Temra CD8 (*GZMB, NKG7, XCL1*), Tm CD8 (*CD69, CD103*), Tex CD8 (*LAG3, PDCD1*), Naive CD8 (*CCR7, TCF7, LEF1*), and Prolif Tex CD8 (*MKI67, MCM5, CXCL13*), five CD4 T cells: naïve-like CD4 (*LEF1, TCF7, SELL, CCR7*), Tm CD4 (*S100A10, LGALS1, ANXA1*), Tcm CD4 (*NOSIP, TOB1*), Tfh CD4 (*TSHZ2, CCR5, CXCL13*), and Th17 CD4 (*RORA, IKZF3, RUNX1*), with two distinct Treg subsets: BATF+Tregs and Tcm-like Tregs (Fig. 3.15A). Tcm-like Tregs were identified as they have high expression of *RTKN2*, whereas BATF+Tregs were identified as they have high expression of *FOXP3, BATF, IKZF2, IRF4, IL2RA, CTLA4, TNFRSF4, TNFRSF18, ENTPD1, CD177* (Fig. 3.15B).

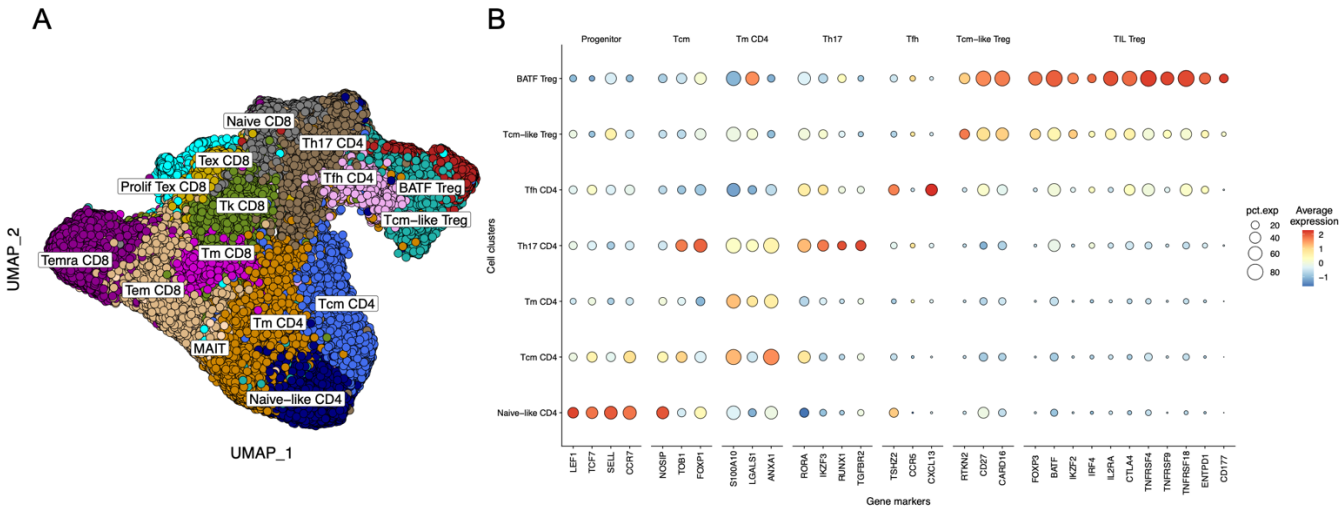

Figure 3.15 – Study of T cells during lung cancer development using integrated scRNAseq of human PID and scRNAseq of lung cancer atlas (N = 458 samples, 271 patients, 250,214 cells in total). A) UMAP showing T cell clusters in meta-analysis object of lung cancer (N = 458 samples, 271 patients, 250,214 cells in total). B) Dotplot presenting the expression of gene modules amongst all CD4 subsets in the meta-analysis of lung cancer. Each dot size shows the percentage of expression of genes on each cell type and the colour gradient presents the average of expression of genes on each cell type: high expression in red, whereas low expression in blue.

Then, I calculated proportion of T cells between normal, pre-LUSC, and LUSC and I found that BATF+Tregs were significantly increased in pre-LUSC compared to normal (p = 0.0067, Fig. 3.12A). Also, BATF+Tregs were significantly increased in LUSC compared to normal (p = 0.0086, Fig. 3.16A).

In addition, the difference lies in the number of cells between human CIS (5,172 cells in total) and normal (327cells in total). I developed Bootstrap miloR custom script to address an imbalanced dataset (see methods). In this method, I sampled T cells from human CIS for 327 cells, which is equal to the number of cells from normal on each of five independent iterations, and miloR was performed recursively. I found that BATF+Tregs were still significantly increased in human CIS (p= 3.24e-05, logFC =2.32, Fig. 3.16C). This suggests that BATF+Tregs were significantly increased in human CIS compared to normal.

Finally, to confirm that BATF+Tregs were increased dramatically in human CIS during LUSC carcinogenesis. I performed a pseudobulk of BATF+Tregs vs other CD4 subsets to create a gene signature of BATF+Tregs. I calculated the expression of the BATF+Treg gene signature using bulk transcriptomics of 122 human bronchial biopsies from 77 smoker patients (see methods). I found that BATF+Tregs were significantly increased in high-grade stage compared to normal (p = 1.56e-05, Fig. 3.16B).

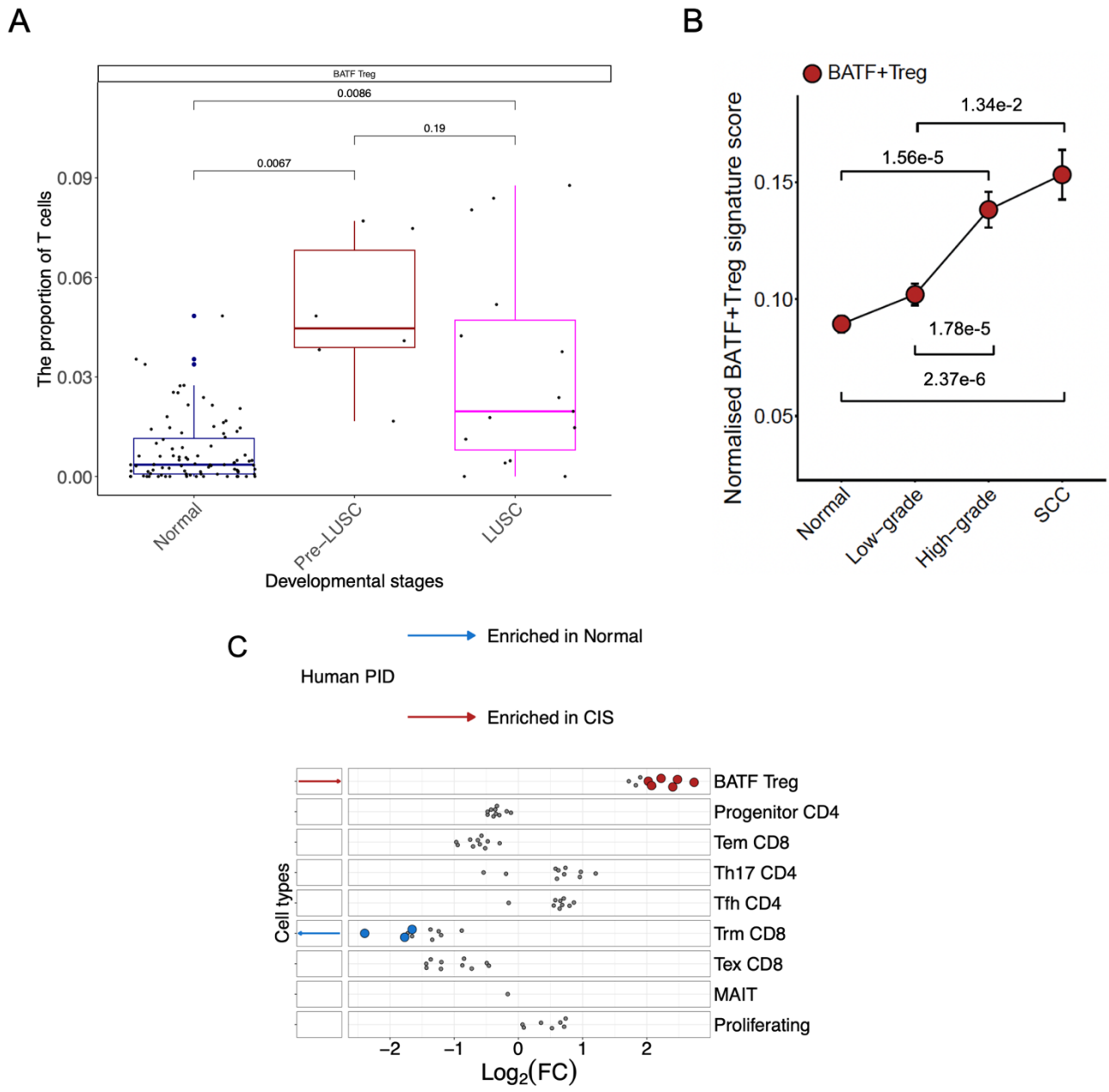

Figure 3.16 – The validation of BATF+Tregs increases during LUSC carcinogenesis (N = 458 samples, 271 patients). A) Boxplot showing the proportion of BATF+Tregs between normal, pre-LUSC, and LUSC using the integrated scRNAseq of human PID and lung cancer atlas. An unpaired t-test was used for the statistical test. B) Line plot showing the normalised BATF+Treg gene signature score using bulk transcriptomics of 122 human biopsies from Mascaux et al. An unpaired t-test was used for the statistical test. C) Beeswarm plot presenting the differential abundance of T cells using scRNAseq of human PID using bootstrap miloR for 10 iterations. Each node represents a neighbourhood, coloured by the log fold change (logFC)

in abundance between CIS (red) and normal (blue). The colour gradient indicates differential abundance, with neighbourhoods enriched in CIS shown in red and those enriched in normal shown in blue

3.4 Discussion

In this study, I wanted to explore the T cell landscape during squamous carcinogenesis in the human lung. Lung cancer is the leading cause of cancer death worldwide because the disease is usually undetected until an advanced stage. Earlier detection and interception of lung cancer could therefore have a significant effect on survival and disease-related treatment and morbidity. My study used rare scRNAseq data from pre-invasive bronchoscopy between normal tissues and high-grade lesions (CIS). I aimed to study T cell dynamics at the single-cell level during LUSC development with a focus on high-grade premalignant lesions. Then, transcriptional profiles of BATF+Tregs – the only subset enriched in CIS lesions - were compared to Treg subsets that were found in established disease, where we saw that these cells were similar to subsets which define anti-PD-1 failure and reproduced these results independently. Next, I expanded opportunities for LUSC interception by identifying crosstalk axes between antigen-presenting cells as senders and BATF+Tregs as receivers, found evidence of within and across tissue seeding of the BATF+Treg pool, and finally I validated these findings in external data sets with alternative models. During this analysis I also found a major remodelling of the MHCII landscape with implications for early immune escape.

I found that BATF+Tregs were significantly enriched in human CIS compared to normal tissues using scRNAseq data. In terms of transcriptional profile, BATF+Tregs that were found in human PID had a similar transcriptional profile to OX40hi/GITRhi Tregs that were found in Dykema et al. (86), and suppressive TIL Tregs that were found in Guo et al. (132). OX40hi/GITRhi Tregs that were found in Dykema et al., were highly suppressive Tregs and associated with anti-PD1 resistance in NSCLC. Next, I validated this result on paired WES and RNAseq data of anti-PD1 in metastatic NSCLC (N = 195 patients) and found that BATF+Tregs were associated with anti-PD1 resistance in NSCLC independently from TMB, CD8, and CD4

infiltration. Moreover, Van Gulijk et al. also found activation and enhanced suppressive capacity of Tregs in anti-PD-L1 therapy-resistant solid tumour-bearing mice. Tumour-infiltrating Tregs in human patients with skin cancer, and in patients with NSCLC, upregulated a suppressive transcriptional gene program after ICB treatment, which correlated with lack of treatment response. This suggests treatment with anti-PD1 and anti-PD-L1 unleashes the immunosuppressive role of Tregs, resulting in therapy resistance and that ongoing trials targeting PD-1 for prevention could benefit by blocking BATF+Tregs, or that BATF+Treg interception alone might be potent.

In addition, I observed that MHC class II expression in suprabasal and basal cells was significantly decreased in human CIS, whereas the proportion of suprabasal cells significantly increased in human CIS. This suggests that there might be some transcriptional and signalling suppression of the MHC class II transactivator in human bronchial epithelial cells in human CIS. Ning et al. identified gene regulatory functions for YAP/TAZ-TEAD-TP63 in early stages of lung cancer development, repressing the MHC class II transactivator CIITA, which is repressed in progressive bronchial premalignant lesions *in vitro* (136). CIITA repression is associated with decreased CD8 T cell recruitment and decreased Th1 T cell responses that accompany bronchial premalignant progression. It may be that this switch in MHC class II expression is also enhancing Tregs, being a double hit to lose Th1 and enhance Tregs (136).

Within human CIS lesions, BATF+Tregs showed the highest sharing of TCR clones with themselves and progenitor CD4 cells. Moreover, progenitor CD4 in human CIS was predicted to be at an early stage of BATF+Treg differentiation during LUSC carcinogenesis. This suggests that progenitor CD4+ cells in human CIS can differentiate into BATF+ Tregs in human CIS. Similar to Wang et al., who analysed scRNA-TCRseq of early-stage LUAD (GGO nodules), there was 21.5% of shared TCR clones between naïve CD4 T cells and Tregs (137). This suggests a potential model of progenitor CD4 cells in human CIS becoming BATF+Tregs in human CIS.

Within human PBMC, BATF+Tregs in human CIS had the highest shared TCR clones with Tcm-like Tregs and also effector Tregs in human PBMC. Also, Tcm-like Tregs were predicted to be an early stage of BATF+Tregs differentiation during LUSC carcinogenesis via pseudotime. This indicates that Tcm-like Tregs and effector Tregs

in human CIS could differentiate into BATF+Tregs in human CIS. According to Ahmadzadeh et al., they found that there were significantly more shared TCR clonotypes between intratumoral and circulating Tregs, implying the migration and expansion of tumour-specific Tregs from the periphery (135). This suggests the potential model of effector Tregs in human CIS becomes BATF+Tregs in human CIS.

In this study, I also explore crosstalk axes between antigen-presenting cells, basal, suprabasal, and fibroblasts as senders and BATF+Tregs as receivers that could support BATF+Treg formation using CellPhoneDB. Then, I validated those crosstalk axes using bulk transcriptomics of 122 human bronchial biopsies from 77 smoker patients. I found potential 13 crosstalk axes that were enriched in human CIS compared to normal. For example, CCL17- CXCR4 interaction could be used to recruit Tregs in NSCLC. High expression of CCL17 in the tumour microenvironment of NSCLC has been shown to attract Tregs, thereby dampening local antitumour immunity (67). CD274 – PDCD1 interaction could promote Treg differentiation or expansion. The PD-1/PD-L1 axis has been shown to promote induced Treg development by inhibiting PI3K/Akt/mTOR pathways and enhancing PTEN signalling that supports Treg immunosuppressive function and tolerance (138). Blocking PD-1/PD-L1 (e.g., nivolumab) can suppress Treg expansion as shown in glioma models (139). These results underpinned a CRUK biology to prevention grant awarded to my team in 2023 to functionally explore additional immune interception targets for LUSC. My hope is that my data will guide validation of a key target with clinical relevance.

A key limitation of dissociation-based scRNAseq is the inability to map immune subsets to their native microanatomical niches. In CIS, this is not a minor technicality: whether BATF+Tregs reside intralesionally (intercalated among dysplastic basal cells) vs stromally or perilesionally (e.g., concentrated beneath the basement membrane or at the lesion margin) changes the biological interpretation of early immune escape. If BATF+Tregs are truly intralesional, their presence would support a model in which CIS establishes local, direct immunosuppression early, potentially blunting cytotoxic activity, antigen-presentation, or effector priming within the lesion itself. In contrast, if BATF+Tregs are primarily stromal, the data may instead indicate formation of a suppressive perimeter that restricts immune infiltration, traps effector cells outside the epithelium, or enforces tolerogenic signalling at the epithelial-stromal interface. Both

models can yield the same bulk outcome in scRNA-seq (increased BATF+Tregs in human CIS lesions) while implying different timing and mechanisms: direct intralesional suppression vs impaired immune access/positioning. This uncertainty also affects how epithelial changes are interpreted (e.g., reduced antigen presentation programs or altered immune-epithelial crosstalk). Without spatial mapping, it cannot be determined whether such epithelial states co-occur with BATF+Tregs in the same microregions (supporting a tightly coupled escape niche) or whether they are uncoupled across the lesion and surrounding tissue (supporting parallel but spatially separated processes). Therefore, conclusions about early immune escape should be framed as: BATF+Tregs are strongly associated with high-grade pre-invasive disease, but the data do not yet resolve whether they mediate escape through intralesional suppression or perilesional exclusion/barrier formation.

Future work should explicitly resolve BATF+Treg localisation relative to (i) dysplastic epithelium, (ii) basement membrane integrity, and (iii) spatial distributions of effector T cells, antigen-presenting cells, and key inhibitory ligands. Spatial resolved methods (e.g., multiplex immunofluorescence/immunohistochemistry, spatial transcriptomics, or targeted in situ hybridisation for BATF/FOXP3 and associated markers (e.g., CD177)) would distinguish intralesional vs stromal niches and test whether BATF+Tregs co-localise with immune exclusion patterns or epithelial immune-evasion states. Establishing this microanatomical context is essential for translating a BATF+Treg-associated CIS signal into a mechanistic model of early immune escape and for prioritising interception strategies aimed at either breaking perilesional barriers or neutralising intralesional suppression

The other limitation of this study is that the number of whole cell types and T cells is limited. However, I developed bootstrap miloR, integrated our scRNAseq of T cells with scRNAseq of T cells from lung cancer atlas (116), and also calculated expression of BATF+Treg gene signature using bulk transcriptomics of 122 human bronchial biopsies from 77 smoker patients (109) to validate the accumulation of BATF+Tregs during LUSC carcinogenesis. As the number of T cells is limited, this could affect the number of TCRs that matched to BATF+Tregs and other CD4 subsets. Therefore, it would be better to have more normal samples to balance the number of

normal cells with the number of CIS cells, but this is a limitation of research in a cutting-edge translational cohort, where such samples are not always abundant. Also, Mascaux et al faced a similar issue, reporting that in normal tissue of central airways Tregs are rare (109).

3.5 Summary

- BATF+Tregs were significantly increased in human CIS lesions compared to normal

- BATF+Tregs had a similar transcriptional profile to highly suppressive Tregs that were found in established disease

- BATF+Tregs were associated with anti-PD1 resistance in NSCLC

- BATF+Tregs could be differentiated from Tcm-like Tregs and effector Tregs from human PBMC and progenitor CD4 from human CIS lesions and MHCII+ APCs *in situ*

Chapter 4. Batf+Tregs as interception targets in murine LUSC carcinogenesis

4.1 Background

To study the complex immune response to NSCLC *in vivo* mouse models must recapitulate histological and molecular features of the disease, whilst showing the inflamed immunotype of the NSCLC TME present in most patients (140). Crucially for T cell-focused analysis and immunotherapy studies the mouse model must be immunogenic and drive T cell responses we can measure and manipulate.

Mouse models of NSCLC have historically been created by introducing one or more tumour-specific driver mutations into the appropriate target cell to reproduce the signalling pathways and dependencies inferred from human disease (141).These include among others *Kras, Braf, Egfr, Lkb1, Rac1, NfkappaB*, and *p53* (142). The use of multiple mutations in combination can be particularly instructive as such models mimic common patterns of mutations and have been used to reveal synergistic effects, or dependencies, that lead to shorter or longer latency periods or an altered disease phenotype (142).

Although these are not models which I have studied in this thesis its important to bring context for our model of choice and highlight their strengths and weaknesses. Some of the key models to study NSCLC include:

i) The Kras-Lox-STOP-Lox-G12D p53 flox/flox (KP) model: The majority of studies to date have been performed using the Lox-Stop-Lox conditional KrasG12D mutation engineered in the endogenous Kras locus (143). In this model introduction of Cre recombinase in cells of LSL-KrasG12D mice results in expression of the mutant allele at endogenous levels. The Kras-Lox-STOP-Lox-G12D cassette included a Lox-STOP-Lox (LSL) conditional allele of oncogenic Kras G12D, which remains inactive until Cre recombinase removes the stop codon, allowing expression of mutant Kras, a driver of oncogenesis; Trp53 flox/flox means conditional knockout alleles of Trp53 or p53 tumour suppressor gene (144). These exons are flanked by loxP sites and are

deleted upon Cre expression, resulting in p53 loss. The KP mouse model has been widely used as an inducible tumour model that recapitulates the cardinal features of the adenocarcinoma subtype in humans. This model has enabled the observation of disease progression from the event of tumour transformation and has provided valuable insight into questions of cancer development. KP mice are excellent for studying cancer cell intrinsic changes, but present difficulties for studying immunology as they contain few neoantigens and elicit poor T cell responses (145).

KP iNducible iNtron-encoded Junctional Antigen (NINJA): Since there is the limitation of studying immunology in the KP mouse model as it contains few neoantigens and elicits a weak T cell response Professor Nikhil S. Joshi's group developed the KP NINJA mouse model to allow controlled neoantigen expression in autochthonous lung tumours, which enables tracking of antigen-specific T cell responses using tetramers or TCR transgenics (e.g., H-2D^{b}-GP33-41 tetramers to detect endogenous GP33-specific CD8+ T cells), mimics tumour immunoediting and immune escape and testing checkpoint blockade (PD-1, CTLA-4), TCR therapy, and immune memory (146). The NINJA system is a genetic cassette that enables tight control of neoantigen expression within tumours (145). Neoantigens are encoded in an intron and are not expressed until induced (146). This is important because the NINJA system was explicitly engineered to avoid attacking the body's own cells or self-tolerance. In the NINJA system, neoantigen is encoded in a silent intron, and there is no expression in the thymus or peripheral tissue. Neoantigen is only expressed after tumour induction to ensure the immune system has not seen the antigen before. Expression restricted to EPCAM+ lung epithelial tumour cells to prevent off-target expression that could cause tolerance. Also, neoantigen-specific T cells are naïve and functional in the periphery to be ready to mount a strong response when tumours express the antigen (146). The KP NINJA model is the latest iteration of several modifications to enhance immunogenicity in the KP model, others included (briefly mention here the KP-LUC and KP-OVA), but these lacked many of the key advantages of the KP NINJA mentioned above (avoiding tolerance, temporal uncoupling of neoantigen expression and disease initiation and IO response).

KM is a genetically engineered mouse model designed to express oncogenic Kras under Cre control and overexpress the Myc oncogene either via a doxycycline-

inducible system (e.g., Rosa26-rtTA; TetO-Myc), or inserted directly at the Rosa26 locus (e.g., Rosa26-LSL- Myc) (147). The KM mouse model is useful for studying oncogenic cooperation (Kras + Myc), immunosuppressive tumours, fast-growing lung cancer, and resistance to immunotherapy (147). As Myc suppresses MHC-I and antigen presentation, it is not suitable for studying about immunogenicity.

KMA (Kras + Myc + APOBEC3B), generated by co-expression of oncogenic KrasG12D, MYC, and conditional expression of human APOBEC3B from the Rosa26 locus, serve as a novel immune-visible model of KRAS-driven autochthonous lung adenocarcinoma. This model exhibits a transient but significant CD8+ T cell-mediated immune response that extends survival compared to KRAS and MYC-driven mice lacking APOBEC3B. Despite early immune infiltration, KMA tumours eventually evade immune surveillance, partly through upregulation of ERBB signalling pathways, including increased expression of EGFR/ERBB ligands such as AREG. This immune evasion mechanism leads to reduced T cell infiltration and tumour progression. Importantly, transient blockade of ERBB receptors with Afatinib restores CD8+ T cell infiltration and, when combined with PD1 immune checkpoint blockade, significantly improves survival. The KMA model thus provides a valuable platform to investigate the interplay between oncogene-driven tumour evolution, APOBEC3B-induced immune visibility, ERBB-mediated immune evasion, and therapeutic strategies to overcome resistance in an immunocompetent setting (Laing et al., bioRxiv, 2023).

Chemically-induced cancer has proven a robust approach to study carcinogenesis across a range of studies, including in NSCLC. Carcinogen-driven models closely mimic chronic environmental carcinogen exposure and drive diverse diseases in terms of penetrance and molecular heterogeneity, which may better reflect the variability of human disease. This comes as the cost of longer experimental time, intra- and inter-assay variation, which necessitates larger numbers and investment in mice and resources (149). Another limitation is that the C57Bl/6 mouse is relatively resistant to carcinogen-driven cancer (150), meaning that alternative strains such as the FVB mouse must be used, which lacks the range of established knock-out models to study immunological mechanisms through reporters and cell or gene-specific knockouts (e.g. the FOXP3-GFP, FOXP3-DTR and Batf flox / flox) (151).

However, a key advantage of carcinogen-driven mouse models is that they allow the thorough study of pre-invasive progression and therefore permit clinically relevant interception experiments along a more physiological timeline. One example is the Methylnitrosourea (MNU) model, which causes predominantly LUAD but also LUSC. The study of pre-invasive LUSC progression requires a different model.

The N-Nitroso-Tris-Chloroethyurea (NTCU) model has been shown to induce LUSC in mice with comparable histology to human LUSC (151). According to Wang et al., NTCU can introduce the same sequence of LUSC lesions seen in humans, namely the full continuum of normal-hyperplasia-metaplasia-dysplasia-LUSC (152). Moreover, RNA sequencing of LUSC induced in the animal model using NTCU showed a high percentage of mutation similarity to human LUSC (153). This suggests that this model closely resembles human LUSC and has been proposed as a key resource to understand LUSC biology with respect to tumour evolution and potentially the TME. In this regard, NTCU mice were found to have a high tumour mutational burden (154), suggesting that this mouse model could generate neoantigens and elicit a strong neoantigen-specific T cell response, suitable to study immune stimulation and regulation during LUSC carcinogenesis.

NTCU is a nicotine-derived nitrosamine, chemically similar to 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), which is found in tobacco smoke (155). It is chemically related to other DNA-damaging agents such as NNK, Urethane (156). The mechanism by which NTCU causes mutations is that it alkylates DNA, causing somatic mutations that lead to cellular dysplasia and neoplastic transformation, particularly in airway epithelial cells (152). The NTCU model was first described in Wang et al., 2004 as a method to replicate human central airway LUSC development in mice, which had previously lacked a robust model (152). Moreover, according to Gomez-Lopez et al., during and after the NTCU treatment, there is progressive accumulation of mutations in airway basal cells, as shown by whole-genome sequencing of microdissected epithelial samples. Mutations include a distinct signature attributable to the alkylating agent NTCU, along with other mutational signatures such as clock-like aging signatures (e.g., SBS5) and chemotherapy-related

signatures (e.g., SBS11, SBS32, SBS36). Clonal expansions arise from highly mutated basal cell clones that spread along the airways (154).

Dr Zoe Whiteman and Dr Sandra Gomez-Lopez from Professor Sam Janes's group established the NTCU progressive airway neoplasia mouse model at UCL in 2019. In terms of mouse selection and preparation, they tested Bagg Albino Laboratory-Bred, substrain "c" (BALB/c) and Friend Virus B strain, originally derived from NIH Swiss mice (FVB/N female) and C57BL/6 female (B6 or Black 6 is the most widely used inbred mouse strain as it serves as the genetic backbone for transgenic, knockout, and cancer models) that were 6-8 weeks old at the start of experiment. Based on these observations, they decided to use FVB/N females because this strain is more sensitive to disease than C57BL/6 and generated preinvasive and invasive lesions. Female mice were more susceptible than males, for reasons that are incompletely understood (157).

Having selected the FVB/N female mouse, Drs. Whiteman and Gomez-Lopez applied NTCU agent (0.04-0.1 mg NTCU in 25- 100 µL solvent) topically to the shaved dorsal skin of mice twice weekly. They found that mice developed squamous metaplasia and low-grade dysplasia at 10-12 weeks, high-grade dysplasia and carcinoma in situ (CIS) at 16-20 weeks, and 80% developed invasive LUSC from 24 weeks in the presence of low and high-grade lesions (154).

The NTCU model set up at UCL by the Janes team recapitulated stepwise histological progression from normal epithelium to invasive carcinoma. Similar to human pre-invasive disease, NTCU-induced pre-invasive lesions can be separated histologically into three grades: flat atypia, low-grade, and high-grade. Flat atypia is defined as a single cell layer with flattened, enlarged nuclei and an increased nuclear-cytoplasmic ratio (Fig 4.1). Low-grade lesions are comprised of a well-ordered multi-layered epithelium. High-grade lesions are defined as a multi-layered disorganised epithelium with enlarged nuclei. Invasive disease develops within the alveolar space, having broken through the basement membrane (158). Both pre-invasive and invasive NTCU-induced disease express keratin 5 (KRT5), a hallmark of LUSC. This is in contrast to healthy murine lungs, where expression of KRT5 is limited to the trachea and main-stem bronchi and does not extend to intrapulmonary regions (159).

However, in NTCU-treated mice, KRT5 is found in intrapulmonary lesions, matching the pathological spread of human LUSC (159).

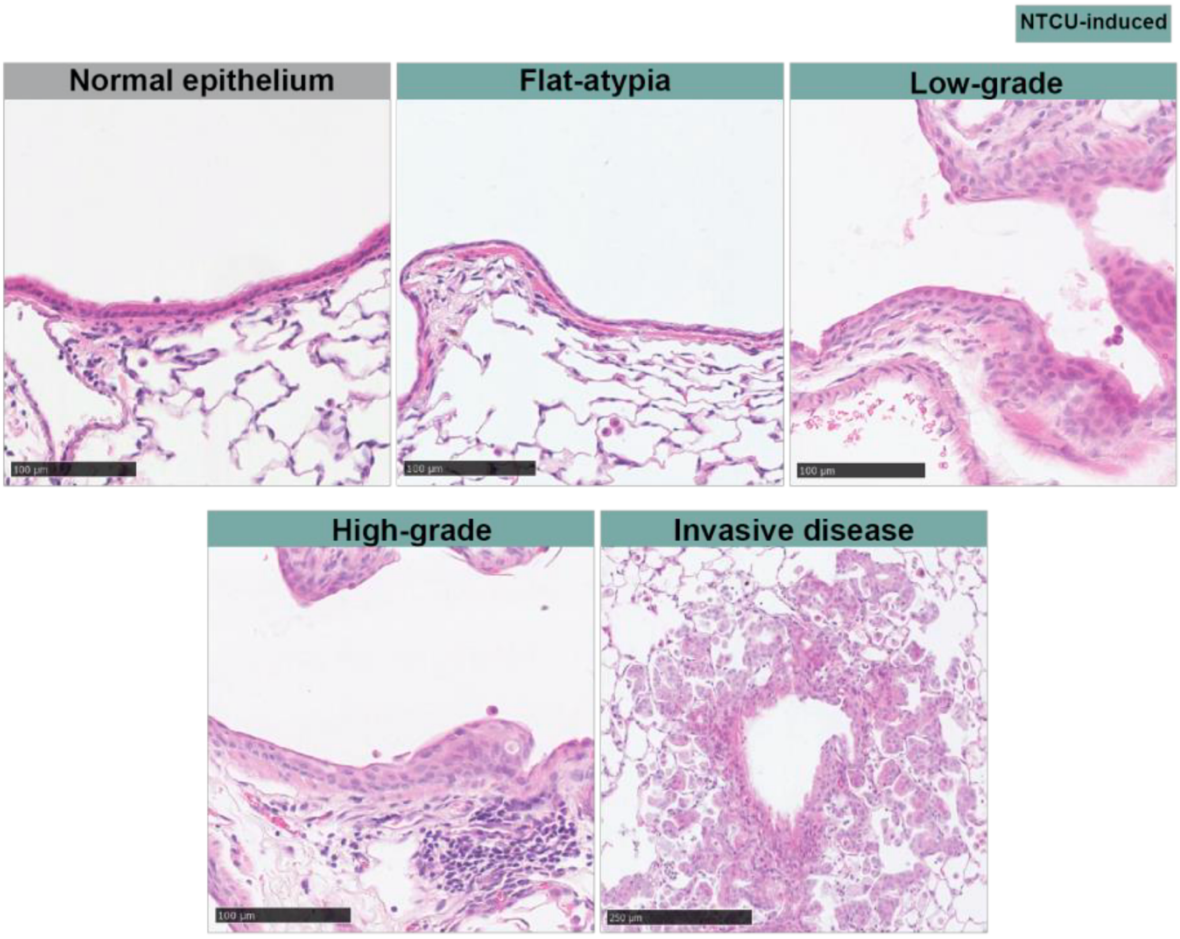

Figure 4.1 – The development of pre-invasive and invasive disease in the NTCU mouse experiment developed by Dr Zoe Whiteman and Dr Sandra Gomez-Lopez using H&E staining. Normal epithelium refers to histologically normal airway epithelium from an age-matched mouse that has not been exposed to the NTCU treatment. This normal epithelium is characterised by the absence of keratin 5 (KRT5) expression in the intrapulmonary regions of the lung. In healthy murine lungs, KRT5 expression is limited to the trachea and main-stem bronchi but does not extend into the intrapulmonary airways. Flat-atypia is defined as a single cell layer of epithelium characterised by flattened, enlarged nuclei and an increased nuclear-cytoplasmic ratio. Low-grade lesions are comprised of a well-ordered multi-layered epithelium. This contrasts with flat atypia, which is defined as a single cell layer with flattened, enlarged nuclei and an increased nuclear-cytoplasmic ratio, and high-grade lesions, which are defined as a multi-layered, disorganised epithelium with enlarged nuclei. The invasive disease is

distinguished histologically from pre-invasive lesions by the loss of the basement membrane barrier and invasion of cancerous cells into the lung parenchyma.

Moreover, Dr Zoe Whiteman and Dr Sandra Gomez-Lopez performed a cross-sectional time series analysis of the immune microenvironment of NTCU-induced disease using T-cell immunofluorescence staining in both parenchyma and in proximity to the bronchial tree of the lung between NTCU-treated mice and age-matched controls. They quantified CD4+FOXP3- (Tconv), CD4+FOXP3+ (Treg), and CD8+ subsets. They found that Tregs were significantly increased in NTCU-treated mice compared to age-matched controls at 11- and 24-week timepoints in the lung parenchyma. In addition, they also found an increase in the association of Tregs to the bronchial tree in NTCU-treated mice compared to age-matched controls at 15, 18, and 24-week timepoints. This suggests that Tregs accumulate in proximity to the bronchial tree during LUSC progression in NTCU-treated mice (Nature manuscript in revision: 2024-04-07961, Whiteman, Z. Thesis (2022)). As the NTCU mouse model could recapitulate the stepwise histological progression from normal epithelium to invasive carcinoma, similar to human pre-invasive disease (159), with a concurrent increase in Tregs in the bronchial tree, I hypothesised that BATF+Tregs would also be found as a major Treg subset in the NTCU model.

In this chapter, I wanted to understand the phenotype and transcriptional profile of CD4+FOXP3+ T cells (Fig 4.2) that the Janes team found in NTCU-treated mice and compare transcriptional profiles with BATF+Tregs I discovered in human pre-malignant lesions. Moreover, I proposed that using scRNAseq and TCRseq across tissues, and studying clonal trajectories and cell-cell interactions could help us understand the ontogeny of these cells. I believed that this data could provide us with potential molecular pathways, interactions and trafficking dynamics that could be used to harness Tregs as the basis to track and target LUSC development for lung cancer interception.

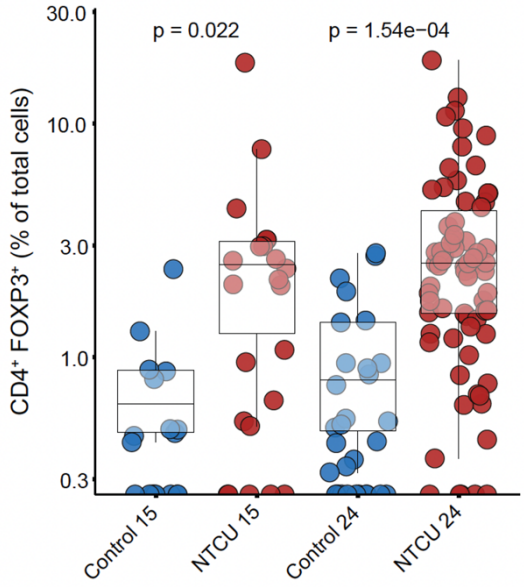

Figure 4.2 – The proportion of CD4+ FOXP3+ (Tregs) in the NTCU mouse (total lungs) experiment. Dr. Zoe Whiteman and Dr Sandra Gomez-Lopez quantified the frequency of Tregs using IF staining between control and NTCU-treated mice at the 15-week and 24-week timepoints.

4.2 Aims

- To identify the phenotype of Tregs in NTCU-treated mice at the 24-week timepoint
- To compare the transcriptional profile of Tregs between NTCU-treated mice at the 24-week timepoint and BATF+Tregs in human pre-malignant lesions
- To identify biological molecular pathways (Treg intrinsic signalling) in BATF+Tregs, both in mouse and human models
- To identify crosstalk axes (juxtacrine and paracrine signalling) between antigen-presenting cells as senders and Batf+Tregs as receivers, which may help to find targets to disrupt the Batf+Treg pool.

4.3 Results

4.3.1 Batf+Tregs are significantly enriched in the airways of the NTCU mouse model according to scRNAseq analysis.

To explore Treg expansion and differentiation in pre-clinical pulmonary pre-invasive squamous carcinogenesis, Dr Zoe Whiteman and Dr Sandra Gomez-Lopez from Professor Sam Janes's group conducted an experiment in the NTCU progressive airway neoplasia mouse model (152). Three mice were treated with NTCU twice weekly for 12 weeks, followed by lesion progression from week 12 to week 24. Previous work from Dr Zoe Whiteman and Dr Sandra Gomez-Lopez has shown that premalignant lesions started developing at week 15 (low-grade) and week 18 (high-grade) (Fig. 4.3). At week 24, 80% of mice had invasive LUSC, and all mice had low grade and high-grade lesions (Fig. 4.3) (154) (Whiteman, Z. (2022)). Although the team also discovered that Foxp3+CD4+Tregs significantly increased in the bronchial tree of NTCU-treated mice, compared to controls (Nature manuscript: 2024-04-07961, Whiteman, Z. (2022)) the full phenotypes and transcriptional identities of Tregs, as well as CD4 and CD8 T cells, were not clear from this multiplex immunofluorescence experiment (mIF).

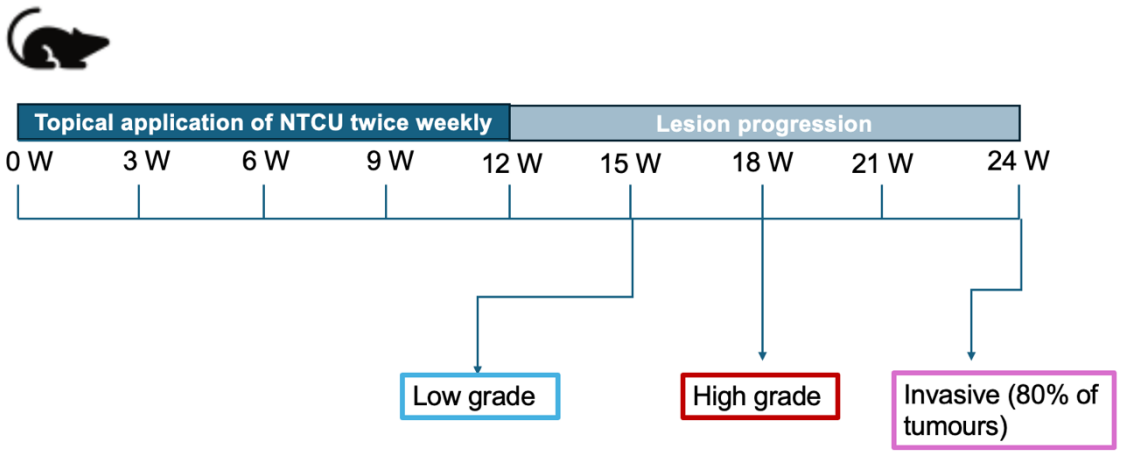

Figure 4.3 – Timeline of NTCU mouse experiment. This timeline was created by Dr Zoe Whiteman and Dr Sandra Gomez-Lopez from Professor Sam Janes' team.

To deeply study the T cell compartment via scRNAseq live, single CD45+ cells were isolated from single cell digests of lungs via Fluorescence-activated cell sorting (FACS) sorting from 24-week control (n=3) or NTCU-treated mice (n=3) and scRNAseq was performed. I processed raw files as above for murine scRNAseq analysis using CellRanger v.7.0.1 (see methods scRNAseq analysis section). After QC, I obtained 64,783 cells in total. I next identified T cells from whole mouse cell types using Gene Set Enrichment Analysis (GSEA) method based on the expression of T cell markers (*Cd3d*, *Cd3e*, *Cd3g*), which yielded 19,659 number of T cells. I performed clustering via Seurat v.4.1.0 (see methods scRNAseq analysis section). Using GSEA (see methods), I identified five CD8 T cell subsets defined by key marker genes: Naïve (*Ccr7*, *Lef1*, *Tcf7*), Mait (*Klrb1a*, *Slc4a10*, *Zbtb16*), Tk CD8, Temra Cd8 (*Ascl2*, *Klf2*, *Tbx21*), and Cycling (*Mki67*, *Mcm2*, *Mcm4*)). I also resolved five Foxp3- CD4 T cell subsets: Naïve-like (*Ccr7, Lef1*, *Tcf7*), Tcm (*Tob1*, *Foxp1*, *Nosip*), Tfh (*Satb1*, *Tox2*, *Tshz2*), Th17 (*Il23r*, *Rora*, *Rorc*), Tm (*Anxa2*, *Itgb1*, *Lgals1*). Most relevantly, I identified two distinct Treg subsets: Batf+ Tregs and Tcm-like Tregs, which resemble populations found in human CIS/LUSC (see chapter 3) (Fig. 4.4A). Tcm-like Tregs were labelled as such based on high expression of *Cd74*, *Cpne2*, and *Klf3*, whereas Batf+Tregs were identified by high expression of *Foxp3*, *Batf*, *Ccr8*, *Il2ra*, *Tnfrsf18*, and *Tnfrsf4* (Fig. 4.4B-C).

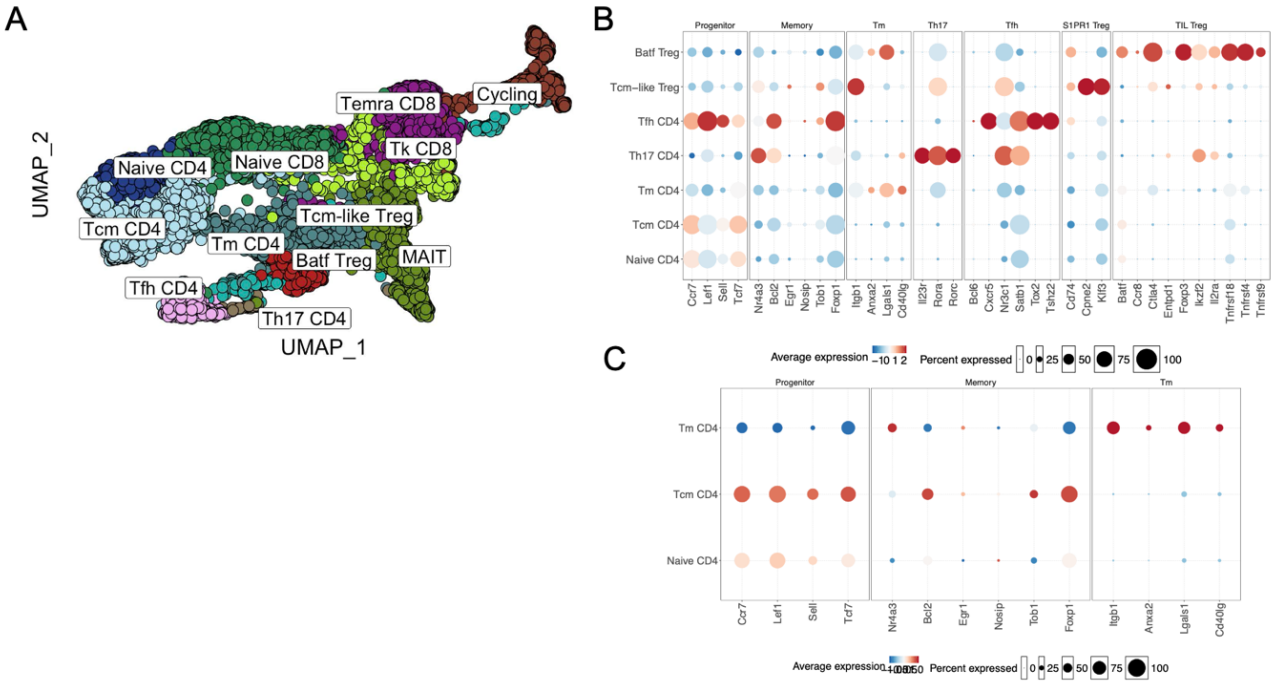

Figure 4.4 – Study of T cells in NTCU mouse model at the 24-week timepoint. A) UMAP showing T cells in the NTCU mouse model at the 24-week timepoint (n = 6, 3 NTCU-treated mice with 3 age-matched controls, 15,901 cells in total). B) Dotplot presenting expression of gene modules amongst all CD4 subsets in the NTCU mouse model at the 24-week timepoint. C) Dotplot portraying expression of gene modules amongst memory CD4 T cell clusters in the NTCU mouse model at the 24-week timepoint. Each dot size shows the percentage of expression of genes on each cell type and the colour gradient presents the average of expression of genes on each cell type: high expression in red, whereas low expression in blue.

Next, to examine whether *Batf*+Tregs were enriched in the NTCU vs control Treg pool. I calculated the proportion of each subset amongst total Tregs. Batf+Tregs were significantly increased in NTCU-treated mice compared to control (p = 0.041, Fig 4.5). In contrast, Tcm-like Tregs were decreased (although this is expected given proportional analysis). This result suggests that Batf+Treg cells were enriched in NTCU during LUSC carcinogenesis, mirroring the significant increase in BATF+Tregs seen in human CIS compared to normal lesions (see Chapter 3).

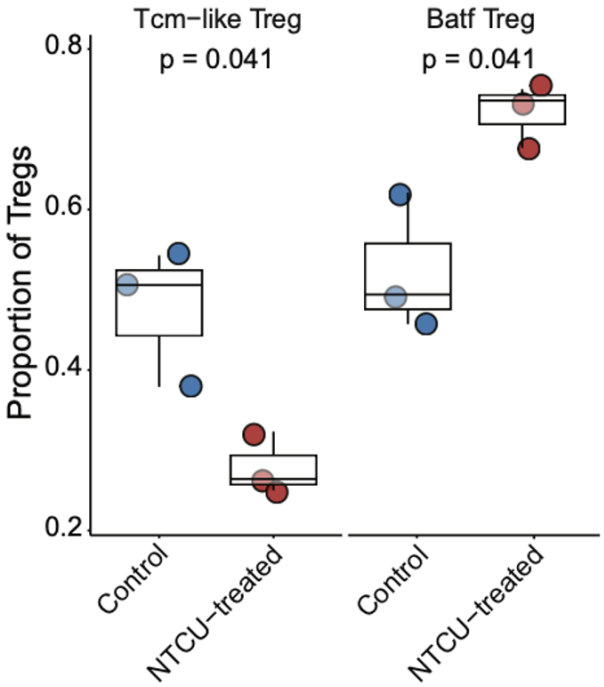

Figure 4.5 – The proportion of Treg subsets in the NTCU experiment between NTCU-treated mice and age-matched controls at the 24-week timepoint. Boxplot showing proportion of Treg subsets: Batf+Tregs and Tcm-like Tregs between NTCU-treated mice and age-matched controls at the 24-week timepoint. An **u**npaired t-test was used to perform a statistical test.

4.3.2 Batf+Tregs found in the NTCU mouse model show strong transcriptional similarity to those in human CIS

According to the dotplot of gene expression amongst all CD4 subsets, Batf+Tregs from NTCU mice showed high expression of *Batf, Ccr8, Ctla4, Entpd1, Foxp3, Ikzf2, Il2ra, Tnfrsf18, Tnfrsf4, and Tnfrsf9*, which appeared similar to BATF+ Tregs found in human CIS (Fig. 3.2B). To formally examine whether Batf+Tregs were similar to BATF+Tregs in human CIS, I performed GSEA analysis using gene signatures of BATF+Tregs from human CIS and other Tregs from a study of human Tregs in NSCLC by Dykema et al. (86). I compared the transcriptome of Batf+Tregs to 17 different human CD4 T cell signatures derived from scRNAseq studies of human NSCLC. Batf+Tregs showed significant enrichment ($q<0.05$) for 7 CD4 Treg signatures, but in contrast showed no enrichment or negative enrichment for 10 effector CD4 T cell signatures. Remarkably, of all human CD4 T cell signatures tested murine Batf+Tregs showed the strongest enrichment for BATF+Tregs derived from our human CIS samples (NES = 2.1, q =3.27e-12, Fig 4.6). This suggests that Batf+Tregs found in the NTCU mouse model had a markedly similar transcriptional profile to BATF+Tregs found in human CIS and they therefore could be used to identify and target pathways that may be clinically relevant for human LUSC immune interception.

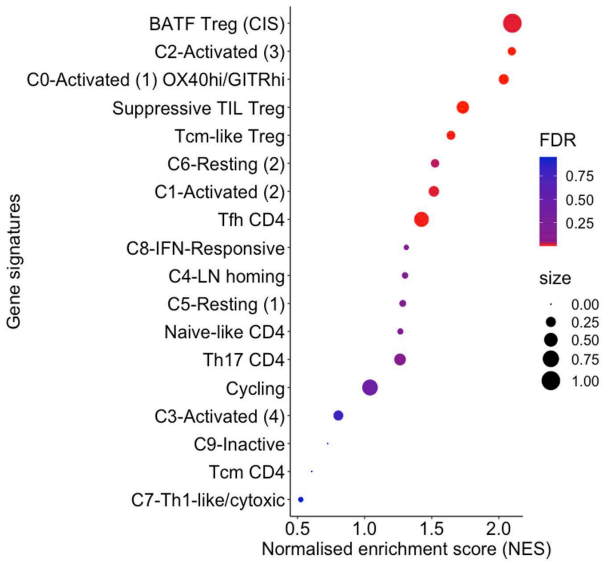

Figure 4.6 – Comparison of transcriptional profiles of Batf+Tregs in NTCU mouse model at the 24-week timepoint. Normalised enrichment score (NES) was calculated using the fgsea package. FDR was calculated by the Benjamini-Hochberg, whereas size was the scale of overlapped genes between differentially expressed genes from Batf+Treg and signature genes of CD4 subsets in human PID and Treg subsets from Dykema et al.

4.3.3 Targeting the PI3Kδ pathway intercepts LUSC *in vivo*

I next explored targetable pathways that were conserved in human and mouse Batf+Tregs to find a possible interception agent which could be tested in the NTCU mouse model. Targeting BATF+Tregs in patients with CIS may be valuable since I found that their gene signature was associated with anti-PD-1 response failure and with poor early stage LUSC outcome (see chapter 3). In order to identify potential interception targets, GSEA analysis was performed using gene signatures of all 1,309 biological pathways from the Reactome database (160); an open-access, manually curated and peer-reviewed database of signalling and metabolic molecules and their relations organised into biological pathways. I specifically screened for pathways that were enriched in mouse and human BATF Tregs but were not enriched in any other CD4 T cell subsets. I found that Batf+Treg (from mouse n = 6 pathways FDR<0.05) (Fig. 4.7A) and BATF+Treg (from human n = 10 pathways FDR<0.05) (Fig. 4.7B) have significantly strong enrichment of multiple pathways after FDR correction (FDR<0.05).

This included 12 pathways enriched for BATF+Tregs, but not other CD4 T cell subsets, in both mouse and human. One of the four pathways was the PI3K-Akt-MTOR pathway (Fig. 4.7C), suggesting that inhibiting the PI3K-Akt pathway could be used to target Batf+Tregs.

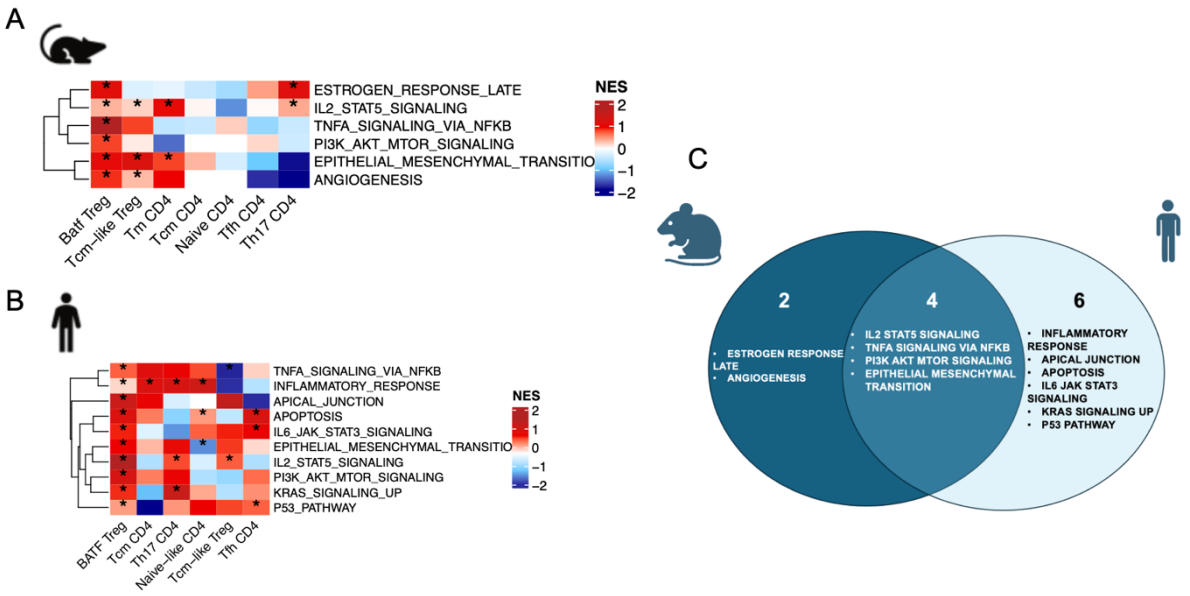

Figure 4.7 – The biological pathways of CD4 subsets between human and mouse models. A) The biological pathways of CD4 subsets in the NTCU mouse model. The heatmap shows biological pathways from the Reactome database, which were significantly enriched in Batf+Tregs and other CD4 subsets in the NTCU mouse model. B) The biological pathways of CD4 subsets in human PID. The heatmap shows biological pathways from the Reactome database, which were significantly enriched in BATF+Tregs and other CD4 subsets in human PID cohort. C) Venn diagram showing the biological pathways of Batf+Tregs that overlapped between human and mouse. Asterisks denote pathways that are significantly enriched (GSEA, adjusted p-value lower than 0,05) in the indicated cell type.

This led us to explore what was known about the role of the PI3K pathway in Tregs, via a discussion with Prof Bart Vanhaesebroeck (FRS) at UCL, who is a global expert in this pathway. Interestingly, it has previously been shown by Prof Vanhaesebroeck's group that targeting an isoform of PI3Kδ breaks immune tolerance by disabling Tregs, which allows effector T cells to destroy tumours in d910A mice, which were tested by inoculating them with different syngeneic cancer cell lines,

including B16 melanoma, Lewis lung carcinoma, EL4 thymoma, and 4T1 breast cancer cells (161). The D910A mice refer to a genetically engineered knock-in mouse strain where the catalytic activity of the P110δ isoform of PI3K has been rendered inactive by a single point mutation: aspartic acid (D) at position 910 is replaced by alanine (A) Moreover, AMG319, which is a highly selective PI3Kδ inhibitor, significantly reduced tumour-infiltrating (TIL) Tregs and bolstered cytotoxic CD8+ T cell activity in the clinic in when used to treat head and neck squamous cell carcinoma (HNSCC) in the neoadjuvant setting (106). Therefore, we decided to test a PI3Kδ inhibitor to target Batf+Tregs. The resulting interception experiment was conducted by Dr Zoe Whiteman and Dr Sandra Gomez-Lopez and is featured in Dr Zoe Whiteman's thesis and our joint manuscript in revision (Nature manuscript in revision: 2024-04-07961). The results showed Batf+Tregs could be targeted by PI3Kδ inhibitor (see Fig. 4.10D in discussion). Thus, my analysis helped to identify an interception opportunity and a possible drug to target Batf+Tregs that was functionally validated.

4.3.4 Potential targets for LUSC interception beyond the PI3Kδ inhibitor

To identify crosstalk axes that could support Batf+Treg formation, I analysed all cell types of 24-week mouse lungs from NTCU-treated (n=3) and control mice (n=3). After QC I obtained a total of 64,783 cells. After clustering, 13 distinct cell types were identified by using GSEA with 24 gene signatures from scRNAseq of the lung cancer atlas (Fig. 4.8A). This empirical method was used alongside manually identified differences in marker genes inspected via gene expression dot plot (Fig. 4.8B). I resolved T cells, monocytes, NK cells, Neutrophils, cDC1, cDC2, Endothelial cells, Fibroblasts, B cells, Macrophages, ILC, TRM-macrophages, and Glial cells (Fig. 4.8A-B). Next, I calculated the proportion of whole cell types between NTCU-treated and control conditions and found that cDC2 cells were significantly increased in NTCU-treated mice (Fig. 4.8C). This suggests that an increased cDC2 pool might better support Batf+Treg formation in the lungs of NTCU-treated mice.

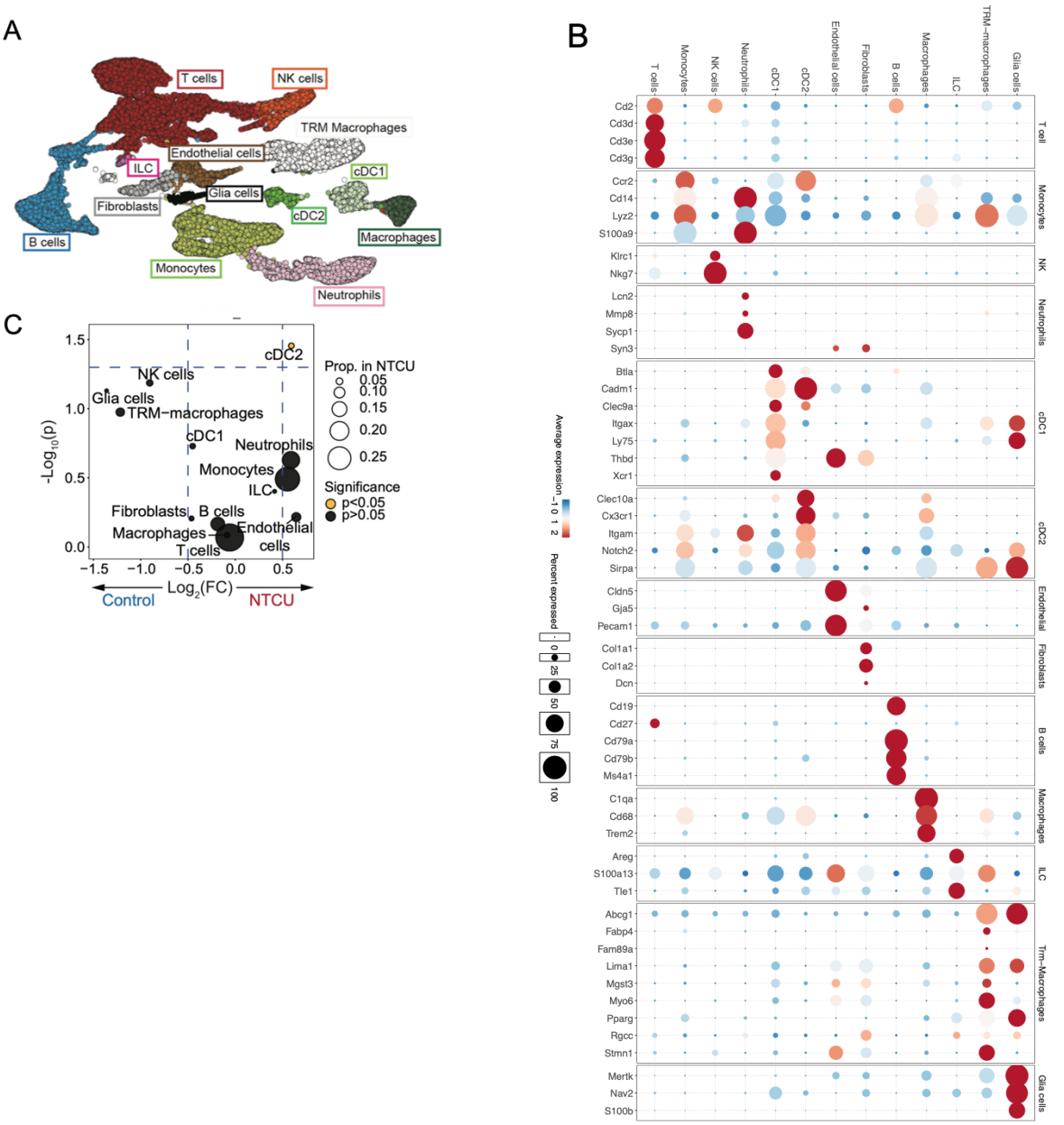

Figure 4.8 – Study of whole cell types in the NTCU mouse experiment at the 24-week timepoint. A) UMAP showing whole cell types in the NTCU mouse experiment (N = 3 NTCU, N = 3 age-matched, 46,069 cells). B) Dotplot presenting expression of gene modules amongst whole cell types in the NTCU mouse model. Each dot size shows the percentage of expression of genes on each cell type and the colour gradient presents the average of expression of genes on each cell type: high expression in red, whereas low expression in blue. C) Volcano plot showing the proportion of whole cell types between NTCU-treated and control mice at the 24-week timepoint. An unpaired t-test was used for the statistical test.

Next, I performed crosstalk axes analysis between immune cells and fibroblast cells as senders and Batf+Tregs as receivers using CellPhoneDB (see methods CellPhoneDB section). This type of cell-cell interaction enables an estimate of specific receptor-ligand pairs based on human databases, which then require transcription into mouse homologues. I found 4 potential crosstalk axes that were also found in human PID: Icam1 – Spn, Cd274 – Pdcd1, Icam1 – Itgal, and Entpd1 – Adora2a (Fig. 4.9A-B).

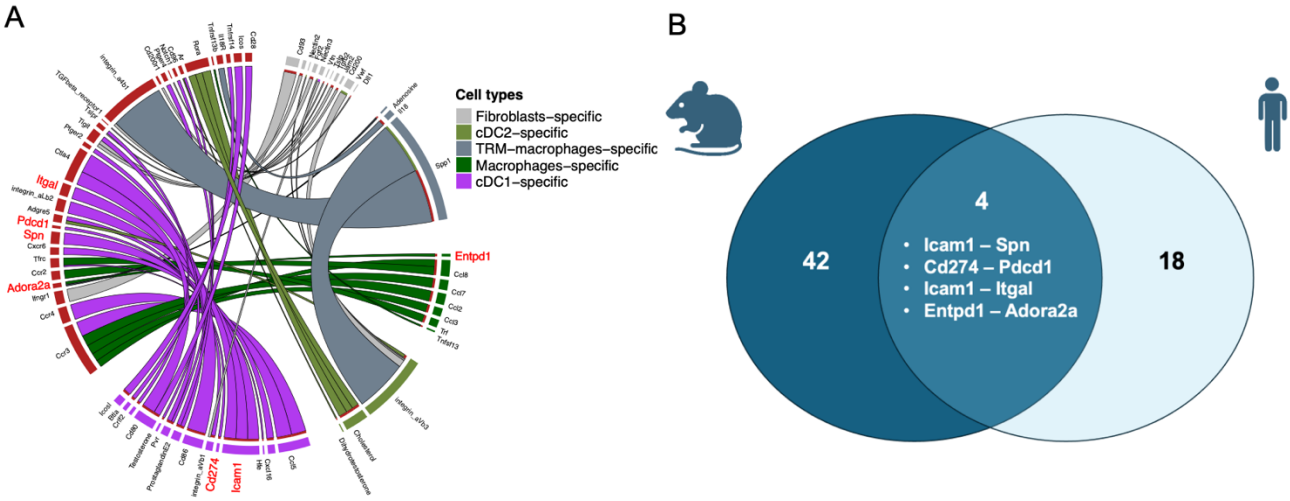

Figure 4.9 – The crosstalk axes between other cell types and Batf+Tregs in the NTCU mouse model at the 24-week timepoint (N =3 NTCU, N= 3 age-matched controls). A) Chord diagram showing crosstalk axes that were enriched in NTCU-treated mice between fibroblasts, cDC2, TRM-macrophages, macrophages, and cDC1 as senders and Batf+Tregs as receivers using CellPhoneDB. B) Venn diagram showing the crosstalk axes that were enriched in the NTCU-treated mice that overlapped with crosstalk axes that were enriched in human CIS.

4.4 Discussion

In this study, I concentrated on studying LUSC progression using the NTCU mouse model because the NTCU model could recapitulate stepwise histological progression from normal epithelium to invasive carcinoma, which is similar to human pre-invasive disease (152). I aimed to identify the phenotype of regulatory T cells in NTCU-treated mice at the 24-week timepoint. Then, transcriptional profiles of regulatory T cells were compared between NTCU-treated mice at the 24-week timepoint and BATF+Tregs in human premalignant lesions. To target Batf+Tregs, biological molecular pathways were identified both in mouse and human PID disease (paracrine signalling). Next, I expanded further opportunities for LUSC interception by identifying crosstalk axes (juxtacrine signalling) between antigen-presenting cells as senders and Batf+Tregs as receivers. scRNAseq was an appropriate method in this study as it could reveal heterogeneity within cell populations and understand biological processes that occur at the single-cell level (162). Therefore, I can gain insights into cell differentiation, disease mechanisms (162).

I observed that Batf+Tregs were significantly enriched in the airways of the NTCU mouse model using scRNAseq. In terms of transcriptional profile, Batf+Tregs found in the NTCU mouse model had a similar transcriptional profile to BATF+Tregs in human CIS. This indicates that targeting Batf+Tregs would likely yield similar immunological consequences in the NTCU mouse model and human CIS. Moreover, Batf+Tregs could be targeted via the PI3Kδ pathway, as Batf+Tregs had significantly strong enrichment of the PI3K-AKT-MTOR pathway compared to other CD4 subsets.

Batf+Tregs were significantly enriched in NTCU-treated mice compared to age-matched controls. This suggests that Batf+Tregs were accumulated during LUSC carcinogenesis. Similar to human pre-malignant lesions, BATF+Tregs were significantly enriched in CIS compared to normal tissues. Also, Batf+Tregs had significantly strong enrichment of gene signature of BATF+Tregs in human CIS and OX40hi/GITRhi Tregs in NSCLC from Dykema et al. (86) ,which were identified as highly suppressive regulatory T cells that were associated with antiPD1 resistance (86). This suggests that Batf+Treg in the NTCU mouse model had a similar

transcriptional profile to BATF+Tregs in human CIS, and they were conserved across species. To target Batf+Tregs, Batf+Tregs had significantly strong enrichment of PI3K-AKT-MTOR pathway both in human lesions and the NTCU mouse model. This suggests that we can target Batf+Tregs via the PI3K-AKT-MTOR pathway in the NTCU mouse model for LUSC interception.

The details of this experiment have previously been featured in Dr Zoe Whiteman's thesis. They are also presented in Fig.5 of our manuscript in revision (Nature manuscript: 2024-04-07961). Dr Sandra Gomez-Lopez performed the NTCU mouse experiment, treated with PI3Kδ inhibitor PI-3065 that preferentially inhibits trafficking and maintenance of Tregs compared to other populations of T cells and also has demonstrated in an established tumour model for anti-cancer activity, alongside the PI3Kα inhibitor BYL719 as a control (Fig. 4.10A). They administered PI-3065 and BYL719 via the diet at the 15-week timepoint and maintained this treatment until the period of harvest at the 24-week timepoint in order to model interception at the progressive preinvasive stage of pulmonary carcinogenesis (Fig. 4.10A). PI3065 treatment had an effect on tumour incidence, but BYL719 had no effect on tumour incidence ($p = 0.022$, Fig. 4.10B-C). They found that tumours developing in the presence of PI-3065 but not BYL719 were smaller, with the complete loss of the largest 30% of tumours ($p = 0.022$, Fig. 4.10C). The percentage of CD4+ FOXP3+ cells was significantly decreased in the NTCU condition treated with PI-3065 compared to vehicle ($p = 0.007$, Fig. 4.10D), whereas CD8 T cells were significantly increased in proximity to PI-3065 and BYL719-treated tumours ($p = 0.003$, $p = 0.00011$, Fig. 4.10E). This suggests that Batf+ Tregs could be targeted by PI-3065.

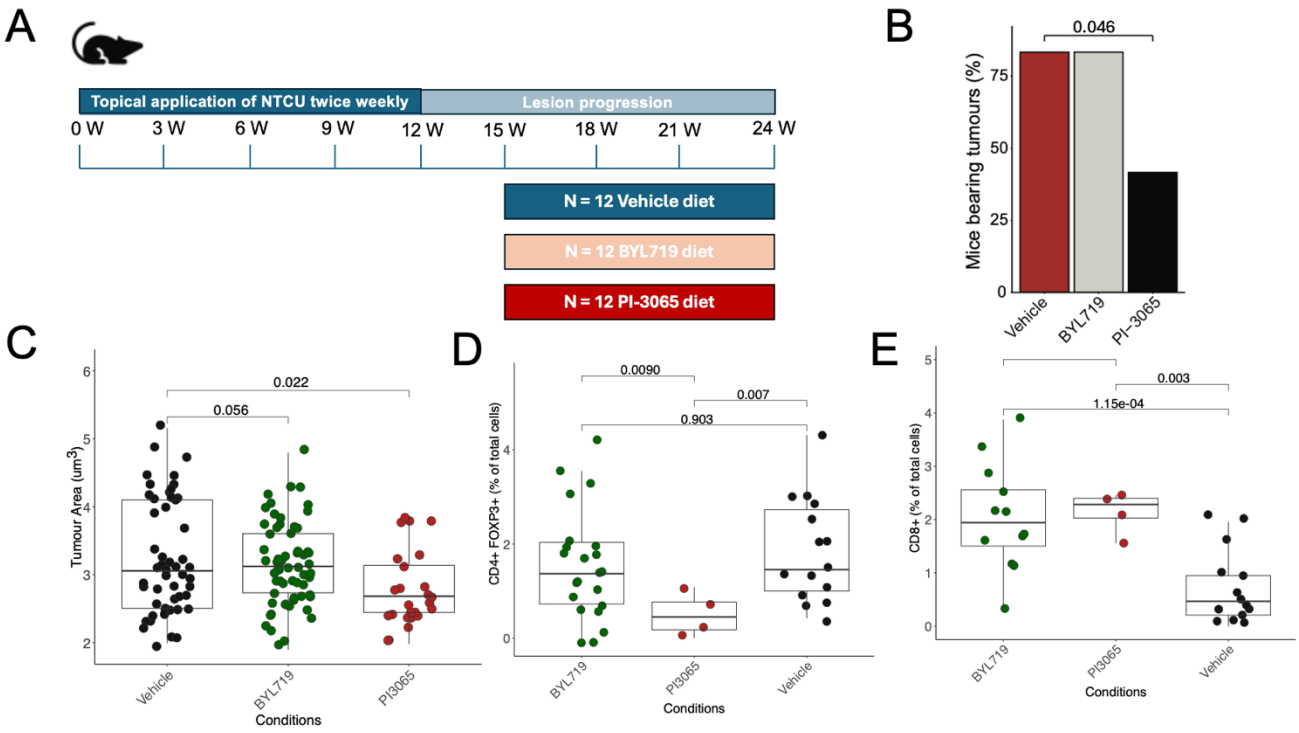

Figure 4.10 – Targeting the Batf+Tregs via the PI3K pathway. A) Timeline of the NTCU mouse mice treated by BYL719 and PI-3065, which was performed by Dr Zoe Whiteman and Dr Sandra Gomez-Lopez. B) Barplot showing the percentage of tumours between vehicle (N = 12), NTCU mice treated with BYL719 (N = 12), and NTCU mice treated with PI-3065 (N = 12). C) Boxplot showing tumour area (um^3) tumours between vehicle (N = 12), NTCU mice treated with BYL719 (N = 12), and NTCU mice treated with PI-3065 (N = 12). D) The frequency of Tregs tumours between vehicle (N= 12), NTCU mice treated with BYL719 (N = 12), and NTCU mice treated with PI-3065 (N = 12). E) The frequency of CD8+ T cells tumours between vehicle (N= 12), NTCU mice treated with BYL719 (N = 12), and NTCU mice treated with PI-3065 (N = 12).

In my thesis I wanted to explore further opportunities for LUSC interception beyond the PI3Kδ inhibitor. To consider alternative methods to disrupt BATF Tregs, I performed receptor-ligand ('crosstalk') analysis at the 24-week timepoint of the mouse scRNAseq data. I found 4 potential crosstalk axes that were also found in human PID: Icam1 – Spn, Cd274 – Pdcd1, Icam1 – Itgal, and Entpd1 – Adora2a. Cd274 – Pdcd1 interaction could indirectly promote Treg differentiation, particularly peripheral Tregs, by creating a tolerogenic environment and modulating TCR and cytokine signalling (163). According to Francisco et al., they found that Cd274 – Pdcd1 axis promoted

conversion of naive CD4+ T cells to Foxp3+ Tregs in the periphery as blocking PD-L1 reduced peripheral Tregs induction *in vitro* and *in vivo* (163). Icam1 – Itgal (LFA-1) interaction plays an indirect role in promoting Treg differentiation, particularly peripheral Tregs by supporting stable cell-cell contacts between T cells and antigen-presenting cells and enhancing the context for tolerogenic signalling. This interaction is essential for naive T cell and antigen-presenting cell conjugation as it could promote Foxp3 induction in the presence of TGFβ (164). Moreover, blockade of Icam1–Itgal could disrupt Treg induction in vitro and reduce oral tolerance in vivo (165). According to Salomon et al., they found that LFA-1-deficient mice showed defects in Treg development and peripheral tolerance (166). Entpd1 – Adora2a axis plays an important immunoregulatory role and could promote Treg differentiation, particularly in the peripheral Tregs (167). This axis could generate extracellular adenosine that acts on A2A receptors to enhance Foxp3 expression, suppress competing effector lineages, and stabilise the suppressive Treg phenotype (168). In addition, Adenosine signalling via A2A could promote Treg expansion and suppressive function, especially in tumours (80). In mucosal tissues and cancer, the CD39-CD73-adenosine-A2A axis was important for peripheral tolerance and Treg induction (168). This suggests that the Entpd1-Adoraa2a axis could be a potential axis that could be targeted to disrupt Treg differentiation.

The limitation of this study is that the number of Batf+Tregs is limited. This could affect the number of TCRs that matched to Batf+Tregs. To investigate the clonal origin and differentiation of Batf+Tregs using scRNA-TCRseq, we need a greater number of Batf+Tregs. To overcome this limitation, Dr Marta Lebrusant-Fernandez and Abigail Shurr sorted total CD3+ live cells enriched for CD4+CD25+Treg populations in the NTCU mouse model at the 18-week timepoint, which Dr Zoe Whiteman found that NTCU mice developed high-grade lesions at the 18-week timepoint (Whiteman, Z. (2022)).

4.5 Summary

- Batf+Tregs were significantly enriched in NTCU-treated mice at the 24-week timepoint

- Batf+Tregs in the NTCU model had a similar transcriptional profile to BATF+Tregs found in human CIS

- Batf+Tregs could be targeted via the PI3Kδ pathway

Chapter 5. Local and peripheral differentiation of Tregs during murine LUSC carcinogenesis.

5.1 Background

In invasive tumours, Tregs can be maintained by several factors in the TME, such as chemokine recruitment from tumours. In terms of antigen, they are stimulated MHCII APCs or tumour cell MHCII interactions locally, primed in local dLN and recruited from the blood, converted from Tconv, including naïve or Th subsets or a mixture of these processes. Increasing data also suggests tissue and tumour Tregs migrate backwards into the dLN and blood. In humans, I found CIS Tregs may derive from effector Tregs in human PBMC and progenitor CD4 T cells in human CIS lesions (see Chapter3). However, the origin of Tregs across tissues in preinvasive pulmonary disease in vivo remains unclear, knowing this may help direct Treg-based LUSC interception.

There are several strategies to deplete or functionally suppress regulatory T cells, including monoclonal antibodies (e.g. CD25, CTLA-4, CCR4, FR4, and GITR), Toxin-based therapies (e.g. Denileukin diftitox) and drugs with Treg-selective effects (e.g. Cyclophosphamide, Lenalidomide, and Imatinib).

Each of these strategies works via different mechanisms, yet there are no approved clinical oncology agents that effectively treat solid tumours by exclusively depleting Tregs, suggesting that we need a better understanding of their origins and differentiation.

In pre-clinical models, some anti-CTLA4 antibodies have been shown to deplete regulatory T cells via Fc-mediated mechanisms (169). This mechanism may also be relevant for clinical activity in humans (105). However, Tregs repopulate rapidly after CTLA4 blockade (170), and CTLA4 blockade can also promote regulatory T cell expansion as it activates CD4+ and CD8+ effector T cells, leading to increased

IL-2 production (171). Regulatory T cells constitutively express the high-affinity IL-2 receptor, so they outcompete effector T cells for IL-2 (67). Similarly, anti-CCR8 antibodies, targeting the CCR selectively expressed on tissue-Tregs, work through Fc-mediated depletion of CCR8+ tumour Tregs via ADCC/ADCP and have shown promise *in vivo*, leading to several ongoing trials, but this depletion can be compensated for by iTreg induction (172) and no clinical activity has been reported as yet (67).

Targeting a broader network of Tregs, including those in the dLN, may be more effective and lead to longer-lasting depletion. For example, IL-2 non-blocking anti-CD25 antibodies generated by Prof Quezada's team show effective tumour control in vivo with a single dose in CT26 colorectal carcinoma model (in Balb/C mice), MC38 tumour model (in C57BL6 mice), and MCA205 tumour model (in C57BL6 mice) (173) and have been shown to be effective in GBM models, in work which I have recently co-authored not featured in this thesis (103) . However, anti-CD25 antibodies rarely deplete all regulatory T cells, especially those with low CD25 expression. These residual regulatory T cells undergo homeostatic proliferation in response to elevated IL-2 levels, which rise following partial regulatory depletion due to reduced IL-2 consumption (174). This expansion can restore the regulatory T cell compartment within days to weeks (103), suggesting that this may be a method to deplete Tregs in multiple tissue compartments, but could required repeat dosing. Treg depletion also needs the correct Fc-receptors to be present on innate effector cells such as macrophages and NK cells, because antibody-bound Tregs are eliminated through FcγR-mediated mechanisms. Without sufficient FcγR-bearing effector cells in the tumour microenvironment, antibodies may block receptor signalling but fail to actually deplete Tregs (102), meaning different TMEs may or may not support this mechanism of depletion.

If we are to develop mAbs agents to target Tregs for LUSC interception, it is therefore important that we understand Treg origins and differentiation during pre-invasive progression, across different tissues so specific strategies to block Tregs most effectively can be pursued.

Blocking the supply of Tregs migrating into the TME from the dLN or the blood rather than just in the tumour may offer a more effective, long-lasting interception. Seeing as human BATF+Tregs overlapped with blood effector Tregs, I wanted to understand the clonal origins and differentiation of Batf+Tregs in high-grade preinvasive lesions from the NTCU model to explore if we could inhibit Tregs both in the tumour and periphery to block the ongoing infiltration of Tregs into high-grade lesions.

As stated, at week 24 80% of NTCU mice developed invasive and pre-invasive disease, but we aim to give interception earlier (in high or low grade preinvasive disease). To better understand the landscape for interception (rather than treatment) I wanted to focus on mice with only high-grade premalignant lesions and extensively characterise the Treg pool, both in terms of gene expression and their clonal origin. Dr Zoe Whiteman had previously found that NTCU-treated mice developed high-grade pre-malignant lesions at week 18 (Fig 5.1), in the absence of invasive disease, suggesting that harvesting Tregs at this time would enable the analysis of Tregs in the lungs of mice with (NTCU) or without high-grade disease (age-matched control). To ensure we had sufficient cells for this analysis, Dr Marta Lebrusant-Fernandez and Abigail Shurr sorted total CD3+ live cells enriched for CD4+CD25+Treg populations from the lungs of NTCU-treated mice at 18 weeks (n=3) alongside age-matched controls (n=3), but preserving the original ratios in each condition based on population frequencies from FACS during cell sorting. The team also sorted live T cells from dLN (n=2 per group) and cardiac puncture (CP -as source of peripheral T cells taken peri-mortem, n=3 per group), and we performed scRNAseq with TCR-seq.

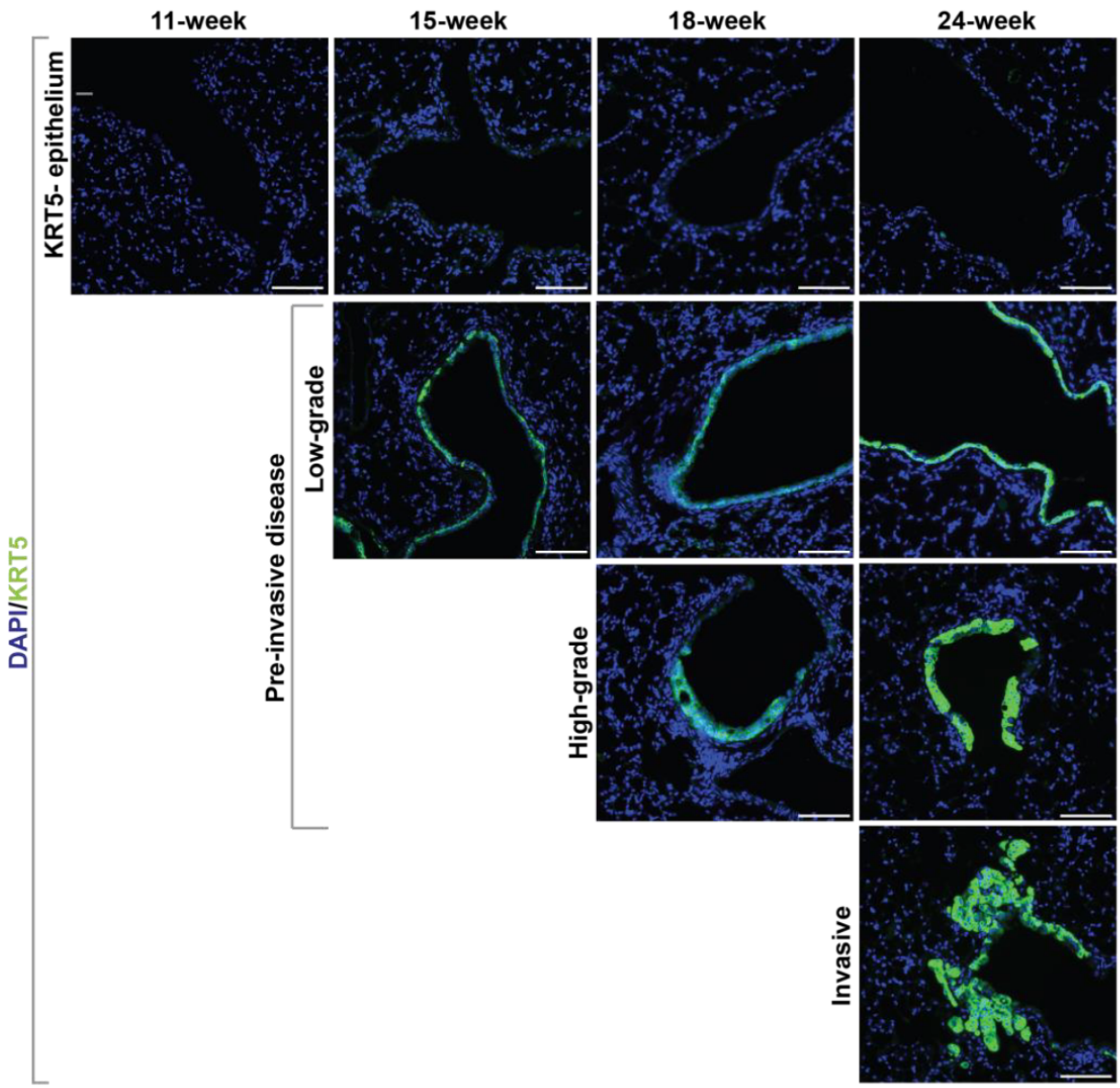

Figure 5.1 – The histological disease extents at different timepoints after the NTCU treatment in mice lungs, focusing on the development of pre-invasive and invasive LUSC lesions, which was performed by Dr Zoe Whiteman, using immunofluorescence staining. At the 18-week timepoint, the pre-invasive lesions cover a vast extent of the bronchial airway surface. These lesions range from low-grade to high-grade pre-invasive disease. This means that by 18 weeks post NTCU initiation, there is widespread presence of abnormal epithelium along the bronchial tree, with areas showing more severe dysplasia (high-grade lesions) as well as less severe abnormalities (low-grade lesions). This shows a significant progression compared to earlier timepoints like 11 weeks (very little to no intrapulmonary KRT5+ lesions) and 15 weeks (development of low-grade lesions mainly in proximal lung regions. At the 24-week timepoint, the bronchial airway surface is extensively

covered with pre-invasive lesions, indicating a range of grades from low- to high-grade lesions. Importantly, approximately 80% of the mice at this timepoint have developed invasive disease. This invasive disease is characterised by cells breaking through the basement membrane into the alveolar space, which signifies progression beyond pre-invasive stages.

In this chapter, I explore the clonal origins and differentiation of Batf+Tregs in high-grade premalignant lesions in the NTCU model to explore how the ongoing infiltration and differentiation of Tregs into high-grade lesions could be most effectively targeted for interception.

5.2 Aims

- To characterise the CD4 T cell compartment in the lung, dLN, and cardiac puncture of the NTCU model at the 18-week timepoint.
- To investigate CD4 subsets in the lung, dLN, and cardiac puncture that shared TCRs with Batf+Tregs in the lung
- To understand transition states of CD4 subsets in the lung, dLN, and cardiac puncture of the NTCU mouse at 18 weeks that have shared TCRs with Batf+Tregs in the lung using monocle2

5.3 Results

5.3.1 Batf+Tregs were significantly increased in the NTCU lung at the 18-week timepoint

To deeply study the T cell compartment during LUSC development in the NTCU mouse experiment at week 18, Dr Marta Lebrusant-Fernandez and Abigail Shurr sorted total CD3+ live cells enriched for CD4+CD25+Treg populations from the lungs of NTCU-treated mice at 18 weeks (n=3) alongside age-matched controls (n=3), and scRNAseq was performed. I processed raw files as above for murine scRNAseq analysis using CellRanger v.7.0.1 (see methods scRNAseq analysis section). After QC, I obtained 40,193 cells in total. I performed clustering via Seurat v.4.1.0 (methods). Using GSEA method (see methods GSEA section). I identified four CD8 T cell clusters: Tem CD8 (*Klrg1, Cd27, Cd28*), Tcm CD8 (*Foxp1, Bcl2, Nosip*), cytotoxic Temra CD8 (*Cx3cr1, Nkg7, Fgfbp2*), and Cycling (*Mki67, Mcm2, Mcm5*)), four CD4 T cell clusters: Naïve-like CD4 (*Ccr7*, *Lef1*, *Tcf7*), Tm CD4 (*Bcl2, Il7r*, *Foxp1*), Tem CD4 (*Runx3*, *Tbx21*, *Prf1*), and Th17 CD4 (*Il23r*, *Rora*, *Rorc*)) with two distinct Treg subsets: Batf+ Tregs and Tcm-like Tregs (Fig. 5.2A). Tcm-like Tregs were labelled based on high expression of *Ccr7*, *Sell*, whereas Batf+Tregs were identified as they have high expression of *Foxp3*, *Batf*, *Ccr8*, *Il2ra*, *Tnfrsf18*, *Tnfrsf4* (Fig. 5.2B).

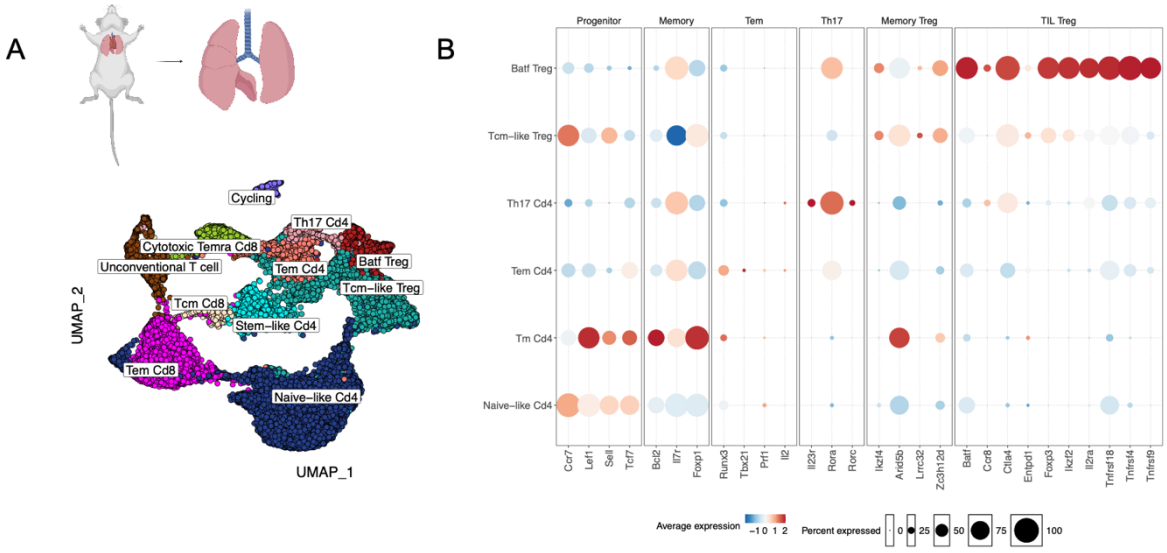

Figure 5.2 – Study of T cells in NTCU mouse at the 18-week timepoint using scRNAseq of the NTCU mouse lung. A) UMAP showing T cells in NTCU lung at the 18-week timepoint (N = 3 NTCU, N = 3 age-match controls, 40,193 cells). B) Dotplot presenting expression of gene modules amongst all CD4 T cell subsets in the NTCU lung at the 18-week timepoint. Each dot size shows the percentage of expression of genes in each cell type, and the colour gradient presents the average expression of genes in each cell type: high expression in red, whereas low expression in blue.

Next, to examine whether Batf+Tregs were increased in the NTCU lung vs control Treg pool, I calculated the proportion of each subset amongst total Tregs and found that Batf+Tregs were significantly enriched in NTCU-treated mice compared to control at the 18-week timepoint (p = 0.005, Fig. 5.3A). In contrast, Tcm-like Tregs were decreased, as expected given the proportional analysis. This result suggests that Batf+Treg were increased in NTCU during LUSC carcinogenesis, mirroring the significant increase in BATF+Tregs seen in human CIS vs normal lesions (see chapter 3) and my results from week 24 (see Chapter 4). Because flow cytometry sorting enriched for Tregs (CD4+CD25+) I also confirmed that Batf+Tregs were increased within the Treg pool in NTCU vs control as seen at 24 weeks (Fig. 4.5). Moreover, to confirm whether Batf+Tregs were enriched in NTCU-treated mice compared to other CD4 T cell populations, I used an alternative method and performed differential abundance analysis of CD4 T cells between NTCU-treated and control mice using miloR (see methods). Batf+Tregs were significantly increased in the NTCU-treated condition compared to control mice (p = 3.42e-05, logFC = 2.57, Fig. 5.3B). As there

is a different number of cells between NTCU-treated and control mice, I used the bootstrap miloR method to address the imbalanced dataset (see methods). I found that Batf+Tregs were still significantly enriched in NTCU-treated mice (Fig. 5.3C). This recapitulated what I saw in human CIS lesions (Fig. 3.3A) and NTCU mice at 24 weeks (Fig. 4.5).

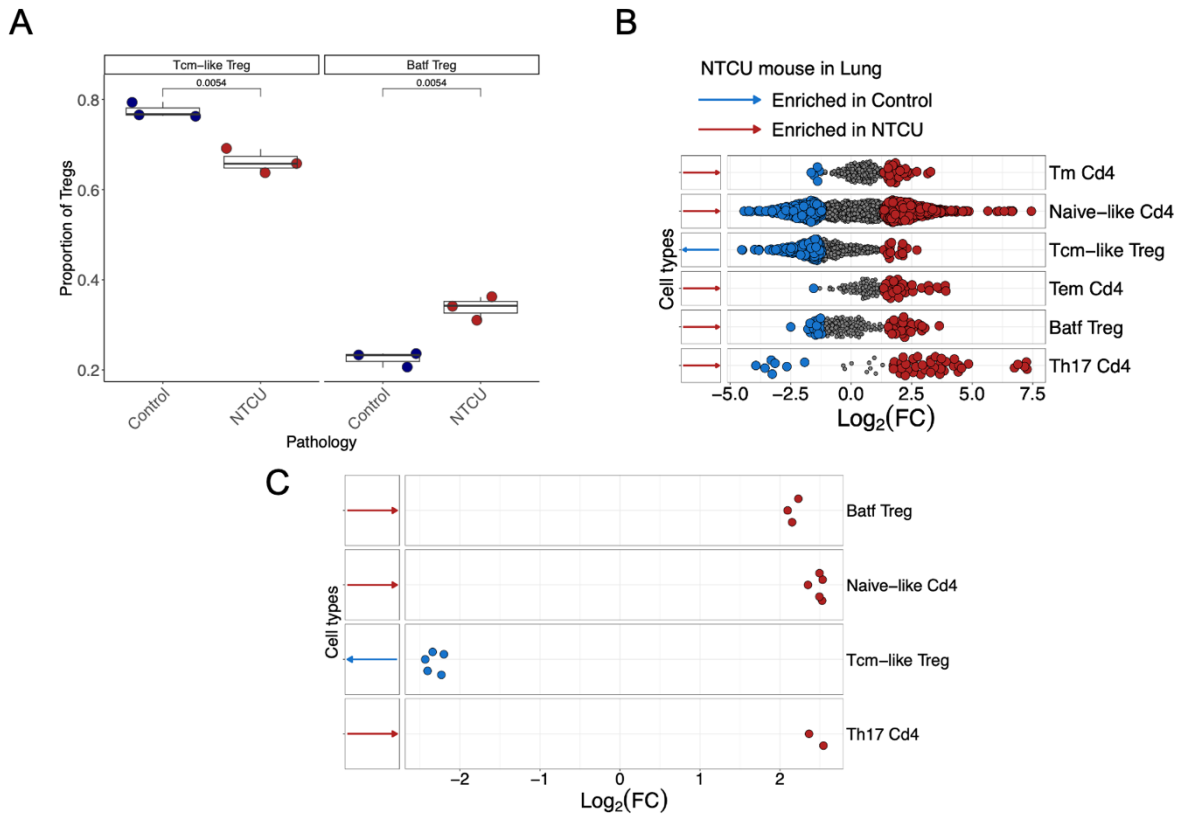

Figure 5.3 – The abundance of T cells in the NTCU lung at the 18-week timepoint (N = 3 NTCU, N = 3 age-matched controls). A) Boxplot showing the proportion of Treg subsets (Tcm-like Tregs and Batf+Tregs) between NTCU-treated (N = 3) and control mice (N = 3). An unpaired t-test was used for the statistical test. B) Beeswarm plot presenting the abundance of CD4 T cells between NTCU-treated and control mice at the 18-week timepoint using miloR. Each dot represents a neighbourhood, coloured by the log fold change (logFC) in abundance between CIS (red) and normal (blue). The colour gradient indicates differential abundance, with neighbourhoods enriched in CIS shown in red and those enriched in normal shown in blue. C) Beeswarm plot showing the abundance of CD4 T cells between NTCU-treated and control mice at the 18-week timepoint using bootstrap miloR. Each dot shows a neighbourhood, coloured by the log fold change (logFC) in abundance between CIS (red) and normal (blue).

To investigate which CD4 subsets in the NTCU lung could differentiate into Batf+Tregs, I calculated shared TCRs between Batf+Tregs and other CD4 subsets in the NTCU-treated condition. I observed that most (84.6%) Batf+Treg clones were private (unshared), suggesting that Batf+Tregs were mostly maintained in situ, likely by clonal expansion within high-grade premalignant tissue in the NTCU. Of the 15.4% which were shared, 7.4 % were shared with Naïve-like CD4, 4.5% with Tcm-like Tregs, and 3.5% of Th17 CD4 shared TCRs with Batf+Tregs in NTCU lungs (Fig. 5.4A). Moreover, to investigate the different transcriptional profiles between those Batf+ Tregs which shared TCRs with other CD4 subsets, I ran DEG analysis between Batf+Tregs that shared TCRs with different subsets. I found that Batf+ Tregs sharing TCRs with Th17 CD4 and Naïve-like CD4 had high expression of TIL Treg gene signature (Fig. 5.4B). This suggests that a minor population of Batf+Tregs could be differentiated from local conversion from Naïve-like CD4 and Th17 CD4 T cells in the NTCU preinvasive lung environment. To assess whether this trajectory was predicted via an alternative model, I analysed TCR overlapping T cell populations using trajectory and pseudotime analyses (monocle2) (see methods pseudotime section). Naïve-like CD4 in NTCU lungs were predicted to be a precursor of Batf+Treg differentiation in NTCU lungs (Fig. 5.4C-D).

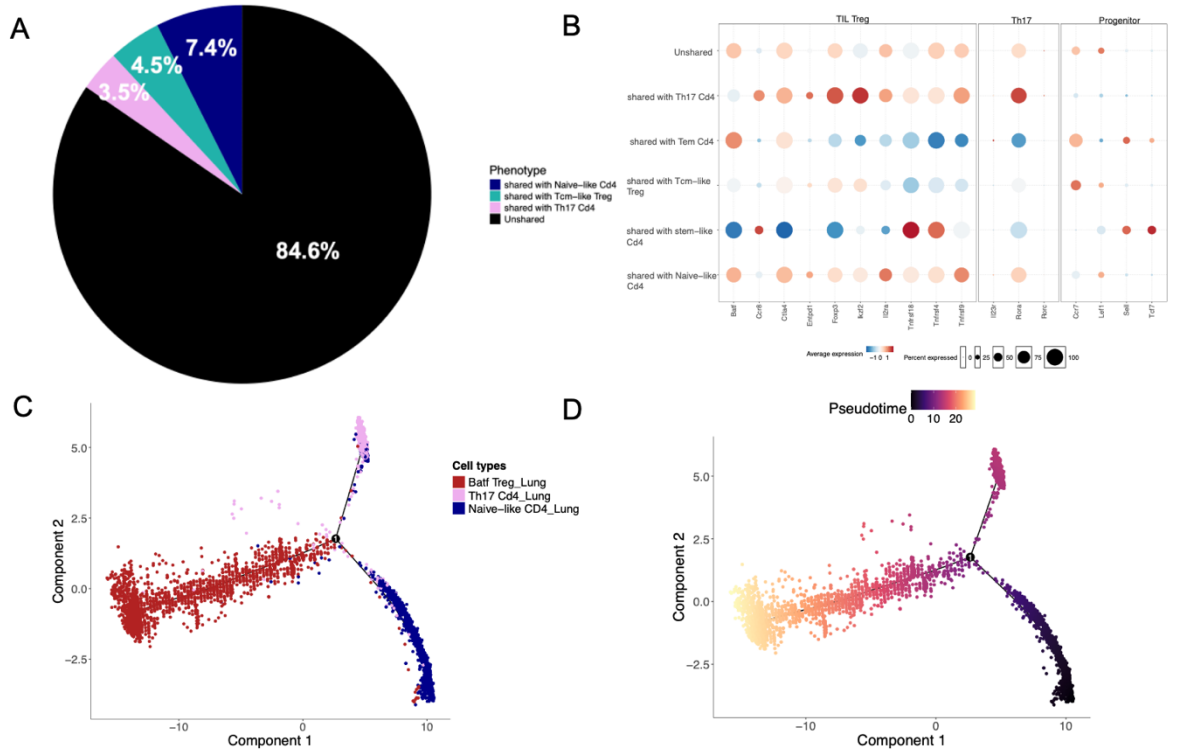

Figure 5.4 – The differentiation of CD4 subsets into Batf+Tregs in the NTCU lung at the 18-week timepoint. A) The percentage of shared TCR clones of CD4 subsets with Batf+Tregs in the NTCU lung at the 18-week timepoint. B) Dotplot showing the expression of gene modules amongst those CD4 subsets that shared TCRs with Batf+Tregs in the NTCU lung at the 18-week timepoint. Each dot size shows the percentage of expression of genes in each cell type, and the colour gradient presents the average expression of genes in each cell type: high expression in red, whereas low expression in blue. C) Trajectory analysis of Batf+Tregs differentiation using monocle2 only those shared TCRs with Batf+Tregs in NTCU lung. Th17 CD4 and naïve-like CD4 were chosen because they have high expression of TIL-Treg gene signature on Fig. 5.4B. D) Pseudotime analysis of Batf+Tregs differentiation using monocle2.

5.3.2 Effector Tregs were significantly increased in NTCU CP and could differentiate into Batf+Tregs in NTCU lung at the 18-week timepoint

To study the blood T cell compartment during LUSC development in the NTCU mouse experiment (CP) at week 18, Dr Marta Lebrusant-Fernandez and Abigail Shurr sorted total CD3+ live cells Treg populations from the cardiac puncture of NTCU-treated mice at 18 weeks (n=3) alongside age-matched controls (n=3), and scRNAseq was performed. I processed raw files as above for murine scRNAseq analysis using CellRanger v.7.0.1 (see methods, scRNAseq analysis section). After QC, I obtained 31,021 cells in total. I performed clustering via Seurat v.4.1.0 (methods). Using GSEA (see methods, GSEA section) I identified three CD8 T cell clusters; Tem CD8 (*Klrg1, Cd27, Cd28*), Tcm CD8 (*Foxp1, Bcl2, Nosip*), cytotoxic Temra CD8 (*Cx3cr1, Nkg7, Fgfbp2*). Three CD4 T cell clusters: Naïve CD4 (*Ccr7, Lef1, Tcf7*), Tcm CD4 (*Bcl2, Il7r, Foxp1*), and Th17 CD4 (*Il23r, Rora, Rorc*)) with two distinct Treg subsets: Effector Tregs and Tcm-like Tregs (Fig. 5.5A-B). Tcm-like Tregs were labelled based on high expression of *Ccr7*, *Sell*, whereas Effector Tregs were identified as they have high expression of *Foxp3*, *Batf*, *Ikzf2*, *Tnfrsf18*, *Tnfrsf4* (Fig. 5.5A-B).

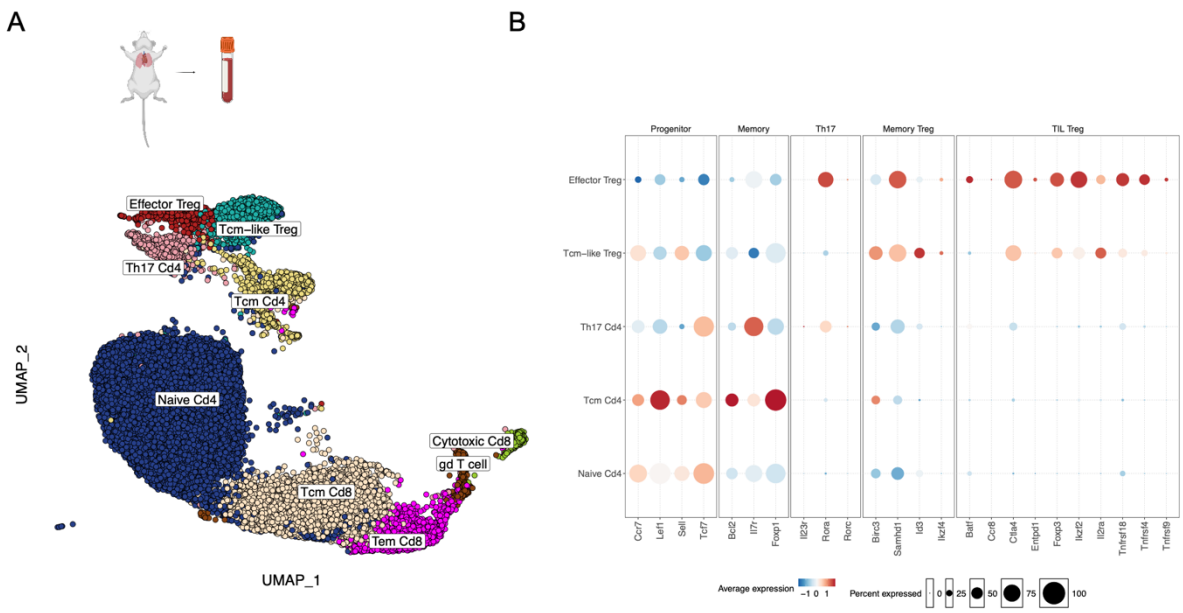

Figure 5.5 – Study of T cells in the NTCU mouse at the 18-week timepoint using scRNAseq of the NTCU mouse CP (N = 3 NTCU, N=3 age-matched controls). A) UMAP showing all T cells in the NTCU CP at the 18-week timepoint (N = 3 NTCU, N = 3 age-matched controls, 31,021 cells in total). B) Dotplot showing expression of gene modules amongst CD4 subsets in the NTCU CP. Each dot size shows the percentage of expression of genes in each cell type, and the colour gradient presents the average expression of genes in each cell type: high expression in red, whereas low expression in blue.

Next, to examine whether Effector Tregs were enriched in NTCU-treated CP mice compared to other CD4 populations, I performed differential abundance of CD4 T cells between NTCU-treated and control mice using miloR (see methods, miloR section). I found that Effector Tregs were significantly increased in NTCU-treated condition compared to control mice (Fig. 5.6A). Moreover, I performed my bootstrapped down-sampled miloR analysis for 5 iterations, confirming that Effector Tregs were still significantly enriched in NTCU-treated mice in CP when adjusting for cell number (Fig. 5.6B).

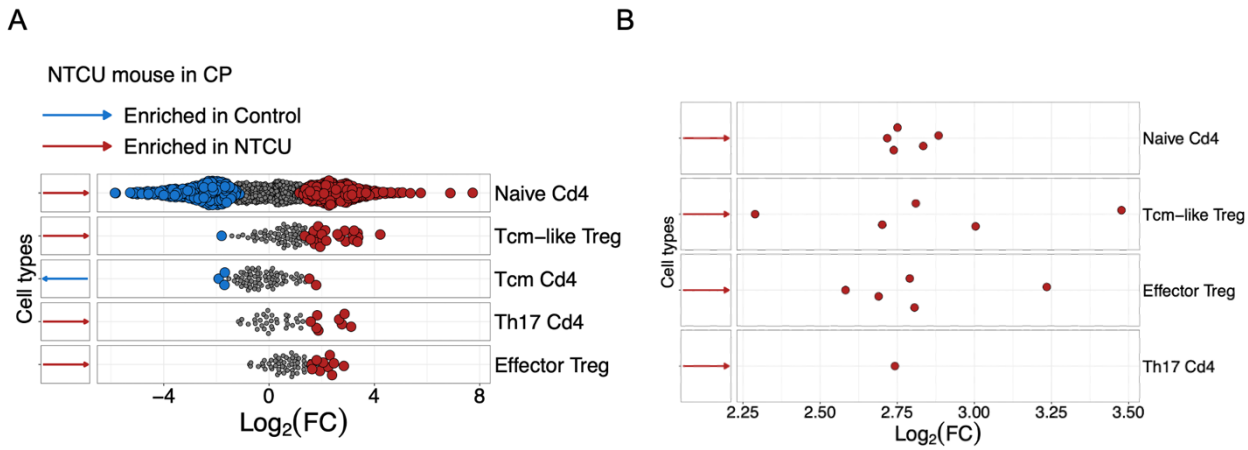

Figure 5.6 – The abundance of CD4 T cell subsets in the NTCU CP at the 18-week timepoint (N = 3 NTCU, N = 3 age-matched controls). A) Beeswarm plot showing the differential abundance of CD4 subsets between NTCU-treated and age-matched controls in CP using miloR. B) Beeswarm plot presenting the abundance of CD4 subsets between NTCU-treated and age-matched controls in CP using bootstrap miloR for five iterations. Each dot shows a neighbourhood, coloured by the log fold change (logFC) in abundance between CIS (red) and normal (blue).

To investigate which CD4 subsets in NTCU CP could differentiate into Batf+Tregs, I calculated shared TCRs between Batf+Tregs and other CD4 subsets in NTCU CP. 12.9% Batf+Tregs in NTCU with TCR data shared any TCRs with the blood CD4 pool (Fig. 5.7A). Of the Batf+TCRs which were shared with the blood 7.2% were shared with Effector Tregs, 4.8% with Naive CD4, and 0.9% with Th17 CD4 (Fig. 5.7A). Moreover, when investigating the difference in transcriptional profiles of those Batf+ Tregs which shared TCRs with other CD4 subsets, I found that Batf+Tregs which shared with Effector Tregs had high expression of the TIL Treg gene signature (Fig. 5.7B). This suggests that Batf+Tregs could be differentiated from Effector Tregs in NTCU CP. Additionally, Effector Tregs in NTCU CP were predicted to represent an early stage of Batf+Treg differentiation in NTCU lungs (Fig. 5.7C-D). Taken with the data above, this suggests that both local (Naïve-like CD4) and blood (Eff.Tregs) CD4 T cells contribute to the Batf+Treg pool (Fig. 5.7E).

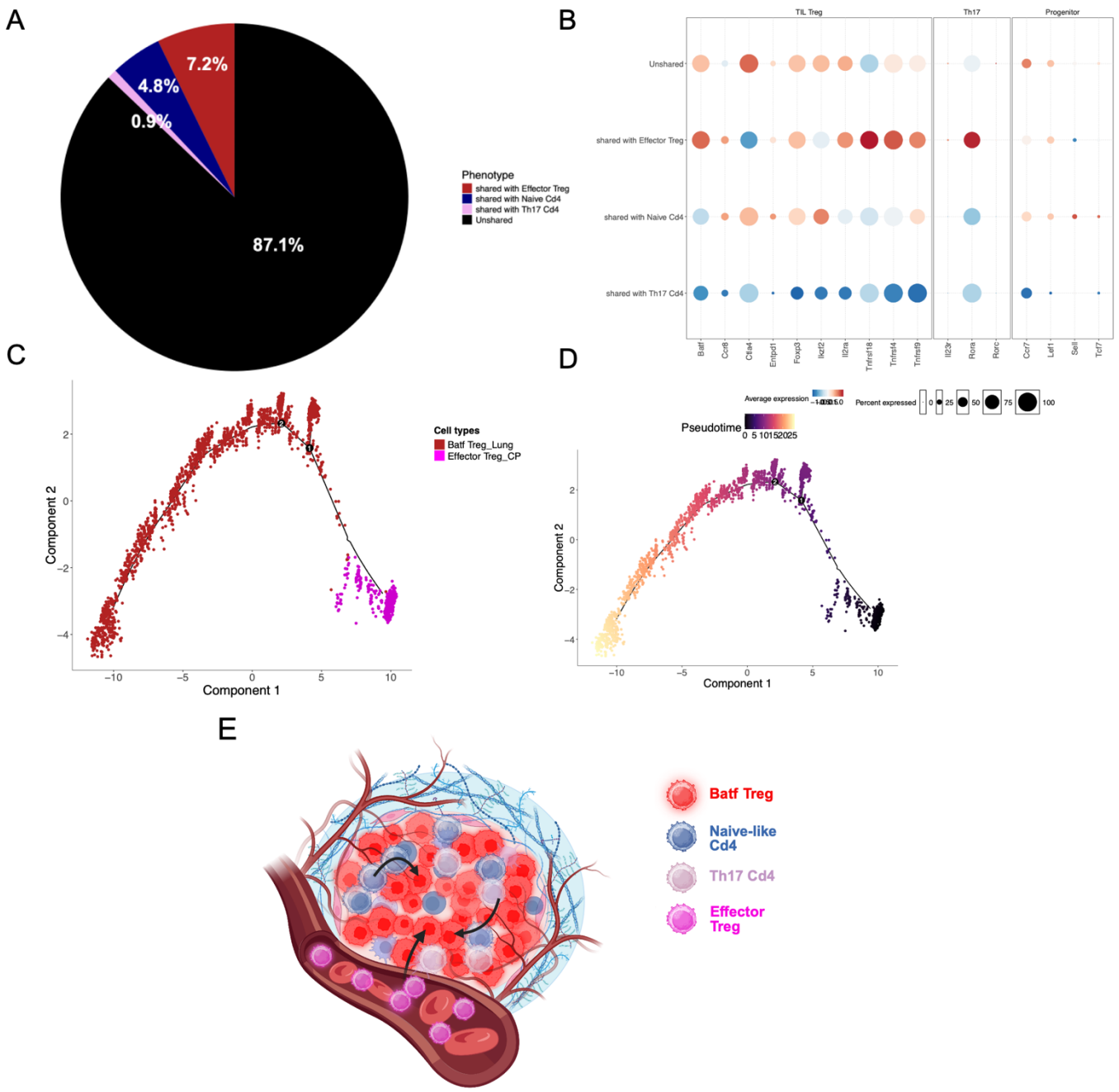

Figure 5.7 – The differentiation of CD4 subsets in the NTCU CP into Batf+Tregs in the NTCU lung at the 18-week timepoint. A) Pie chart showing the percentage of CD4 subsets in the NTCU CP shared TCR clones with Batf+Tregs in the NTCU lung at the 18-week timepoint. B) Dotplot presenting expression of gene modules of those CD4 subsets in the NTCU CP that shared TCRs with Batf+Tregs in the NTCU lung. Each dot size shows the percentage of expression of genes in each cell type, and the colour gradient presents the average expression of genes in each cell type: high expression in red, whereas low expression

in blue. C) Trajectory analysis between effector Tregs in the NTCU CP, which shared TCRs with Batf+Tregs in the NTCU lung, and Batf+Tregs in the NTCU lung using monocle2. D) Pseudotime analysis between effector Tregs in the NTCU CP and Batf+Tregs in the NTCU lung using monocle2. E) Diagram showing the potential model of Batf+Tregs differentiation in the NTCU lung and the NTCU CP in the NTCU microenvironment.

5.3.3 Batf+ Tregs were the most expanded compared to other CD4 subsets in NTCU dLN at week 18

To study the T cell compartment during LUSC development in the NTCU mouse experiment (dLN) at week 18, Dr Marta Lebrusant-Fernandez and Abigail Shurr sorted total CD3+ live cells enriched for CD4+CD25+Treg populations from the dLN of NTCU-treated mice at 18 weeks (n=2) alongside age-matched controls (n=2), and scRNAseq was performed. After processing and QC, I obtained 18,578 cells in total and performed clustering and GSEA to annotate clusters as above. I identified Two CD8 T cell clusters: Tcm-like CD8 (*Foxp1, Bcl2, Nosip*) and gd T cells (*Trdc, Trdv1, Trgc1*)), five CD4 T cell clusters: Naïve CD4 (*Ccr7, Lef1, Tcf7*), Tm CD4 (1) (*Tpt1, Eef1b2*), Tm CD4 (2) (*Nr4a3, Zfp36, Junb*), Tm CD4 (3) (*Cd40lg, S100a10*), and Early activated CD4 (*Bcl2, Il7r, Foxp1*)) with two distinct Treg subsets: Batf+ Tregs and Tcm-like Tregs (Fig. 5.8A). Tcm-like Tregs were labelled based on high expression of *Ccr7*, *Sell*, whereas Batf+Tregs were identified as they have high expression of *Foxp3*, *Batf*, *Ccr8*, *Il2ra*, *Tnfrsf18*, *Tnfrsf4* (Fig. 5.8B).

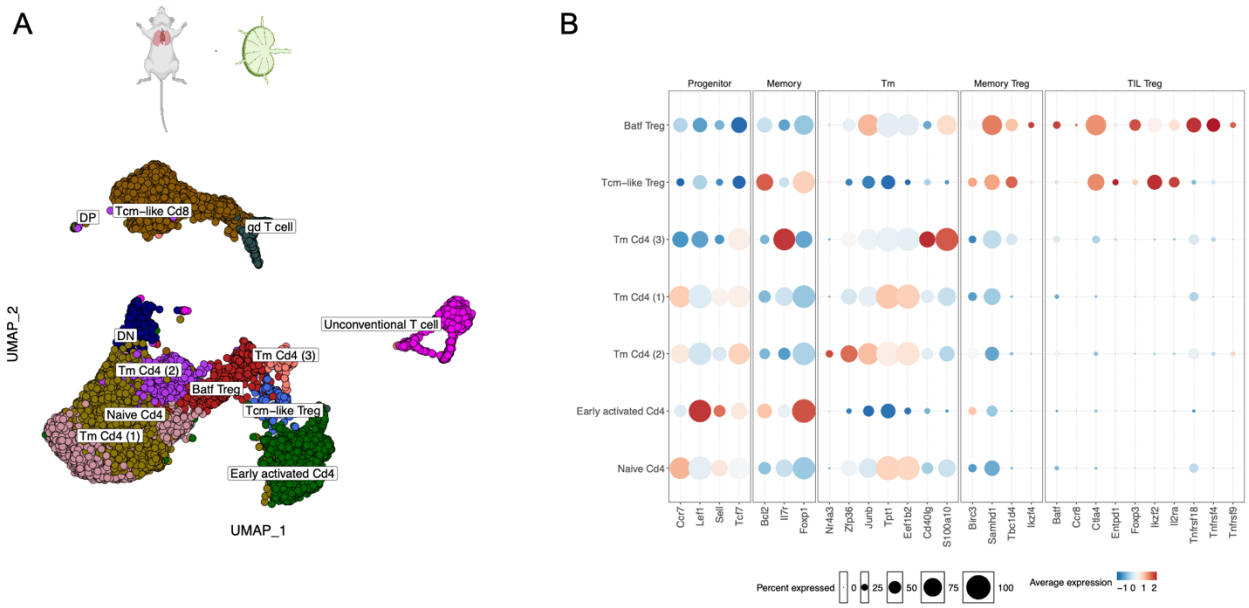

Figure 5.8 – Study of T cells in the NTCU mouse at the 18-week timepoint using scRNAseq of the NTCU dLN (N=2 NTCU, N=2 age-matched controls). A) UMAP showing T cell clusters in the NTCU dLN at the 18-week timepoint. B) Dotplot presenting expression of gene modules amongst all CD4 subsets in the NTCU dLN scRNAseq data. Each dot size represents the percentage of gene expression in each cell type, and the colour gradient indicates the average expression of genes in each cell type: high expression is represented by red, while low expression is represented by blue.

Next, to examine whether Batf+Tregs were enriched in NTCU-treated dLN mice compared to other CD4 populations, I performed differential abundance of CD4 T cells between NTCU-treated and control mice using miloR (see methods, miloR section). I found that Batf+ Tregs were not significantly increased in NTCU-treated condition compared to control mice (Fig. 5.9A), suggesting that the increase in Batf+Tregs in the lung and Effector Tregs in the blood of NTCU mice was not simply due to more Batf+Tregs being present in the dLN. I reasoned that there is more Batf+Treg clonal expansion in the NTCU dLN to balance the increased migration of these cells to the lung lesions, which would explain the lack of increase in number. When testing clonal expansion, I found that although the number of these cells is not higher in NTCU dLN,

they are significantly more clonally expanded (Fig. 5.9B). This may potentially balance the egress of these cells from the dLN to supply the Batf+Treg pool.

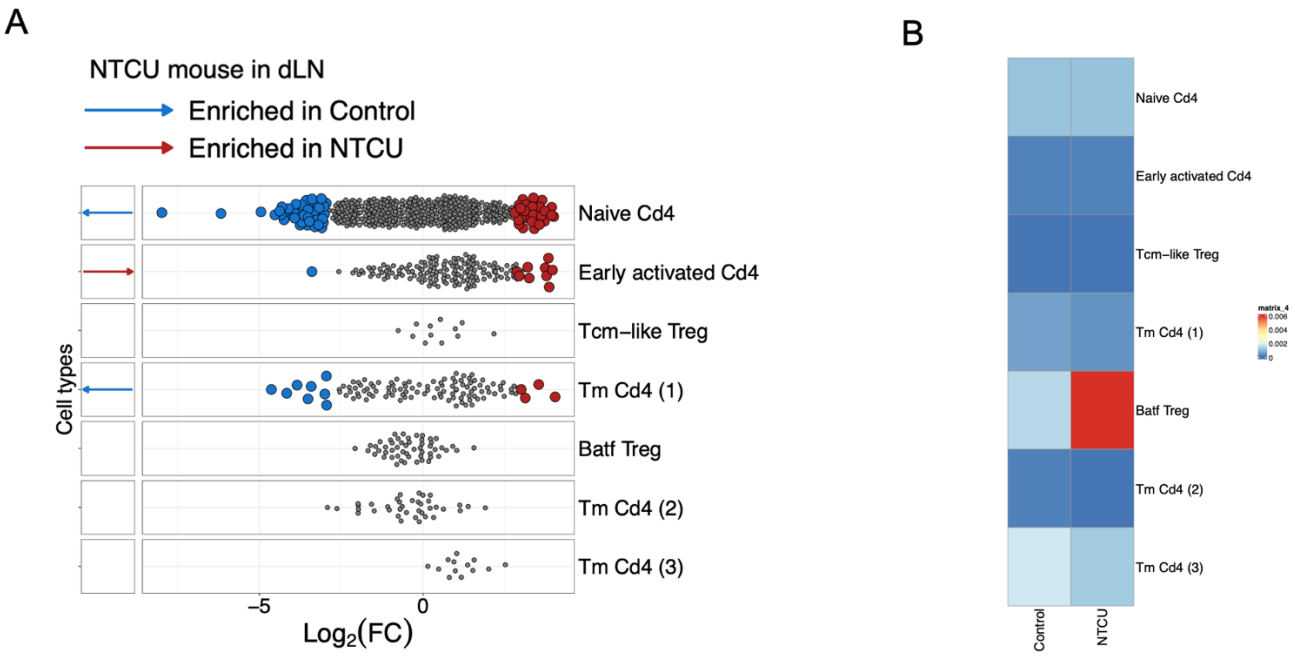

Figure 5.9 – The abundance of CD4 subsets in the NTCU dLN at the 18-week timepoint. A) The differential abundance of CD4 subsets in the NTCU dLN at the 18-week timepoint using miloR. Each dot shows a neighbourhood, coloured by the log fold change (logFC) in abundance between CIS (red) and normal (blue). B) Heatmap presenting STARTRAC expansion score of CD4 subsets in dLN between the NTCU-treated and control mice using STARTRAC package. STARTRAC-expansion is used to quantify clonal expansion of T cells using scRNAseq that include paired TCR information. It assesses how often identical TCR clonotypes are found within a given cell population, based on the principle that cells sharing the same TCR originate from a common ancestor. A high STARTRAC-expansion score indicates strong antigen-driven proliferation, whereas a low score reflects a more diverse, unexpanded T cell repertoire. By comparing expansion scores across cell states, STARTRAC-expansion provides a way to assess how immune activation and differentiation shape T cell populations in control and NTCU conditions.

To investigate which other CD4 subsets in NTCU dLN could differentiate into Batf+Tregs, I calculated shared TCRs between Batf+Tregs and other CD4 subsets in

NTCU dLN. Of those which were shared, 1.5% were Naive CD4, 1% were Tm CD4 (1), and 0.6% were Tm CD4 (2) (Fig. 5.10A). Moreover, to investigate the difference in transcriptional profiles between those Batf+ Tregs that shared TCRs with other CD4 subsets, I found that Naïve CD4, Tm CD4 (1), and Tm CD4 (2) had low expression of TIL Treg gene signature (Fig. 5.10B). This suggests that T cells in NTCU dLN seemed to have low activation and Batf+Tregs in dLN clonally expanded within NTCU dLN because they had the highest expansion score from STARTRAC (see methods) (Fig. 5.9B). Tm CD4 in NTCU dLN were predicted to be an early stage of Batf+Treg differentiation in NTCU lungs (Fig. 5.10C-D).

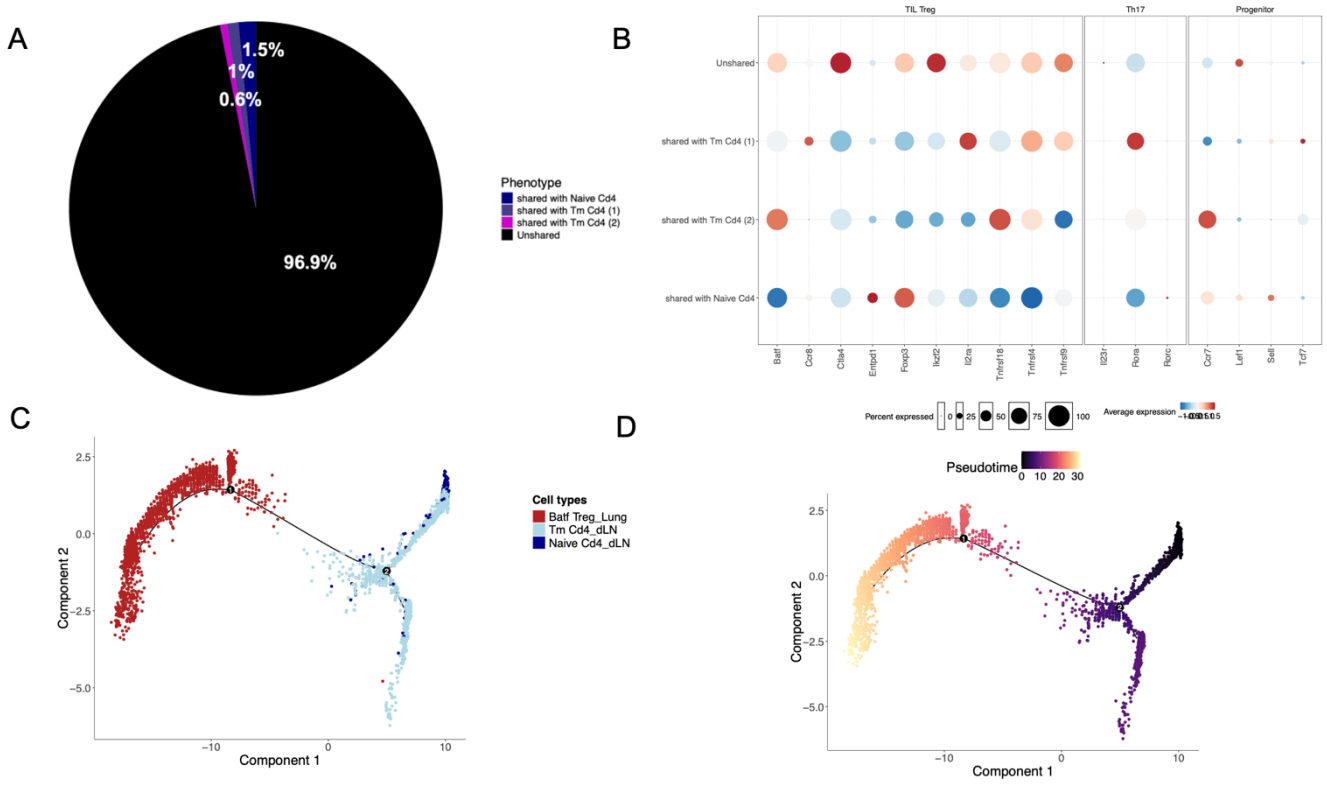

Figure 5.10 – The differentiation of CD4 subsets in dLN into Batf+Tregs in the NTCU dLN at the 18-week timepoint (N=2 NTCU, N = 2 age-matched controls). A) Pie chart showing the percentage of CD4 subsets in the NTCU dLN that shared TCR clones with Batf+Tregs in the NTCU lung. B) Dotplot presenting expression of gene modules amongst those CD4 subsets in the NTCU dLN that shared TCRs with Batf+Tregs in the NTCU lung. Each dot size shows the percentage of expression of genes in each cell type, and the colour gradient presents the average expression of genes in each cell type: high expression in red, whereas low expression in blue. C) Trajectory analysis between Tm CD4 and Naïve CD4 in the NTCU dLN and Batf+Tregs in the NTCU lung using monocle2. D) Pseudotime analysis

between Tm CD4 and Naïve CD4 in the NTCU dLN and Batf+Tregs in the NTCU lung using monocle2.

5.3.4 Batf+Tregs in NTCU lung primarily originate from effector Tregs in NTCU CP

I previously showed scRNAseq-TCR analysis of Batf+Tregs and other CD4 subsets amongst three different sites separately. I observed Batf+Tregs in NTCU lung had shared TCR clones with naïve-like CD4 in NTCU lung, effector Tregs in NTCU CP, and Naïve CD4 in NTCU dLN. To explore how those CD4 subsets in three different locations could contribute to Batf+Treg differentiation, I performed an analysis of STARTRAC-transition amongst all CD4 subsets in NTCU dLN, NTCU CP, and NTCU lung (see methods). I found that effector Tregs had the highest pairwise transition score with Batf+Tregs in NTCU lung (Fig. 5.11A). This suggests that Batf+Tregs in NTCU lung primarily originate from effector Tregs in NTCU CP. To assess whether this may be observed via an unsupervised model, I performed TCR overlapping T cell populations using trajectory and pseudotime analyses from monocle2 (see methods). Effector Tregs in NTCU CP were predicted to be an early stage of Batf+Treg differentiation in NTCU lungs (Fig. 5.11B-C).

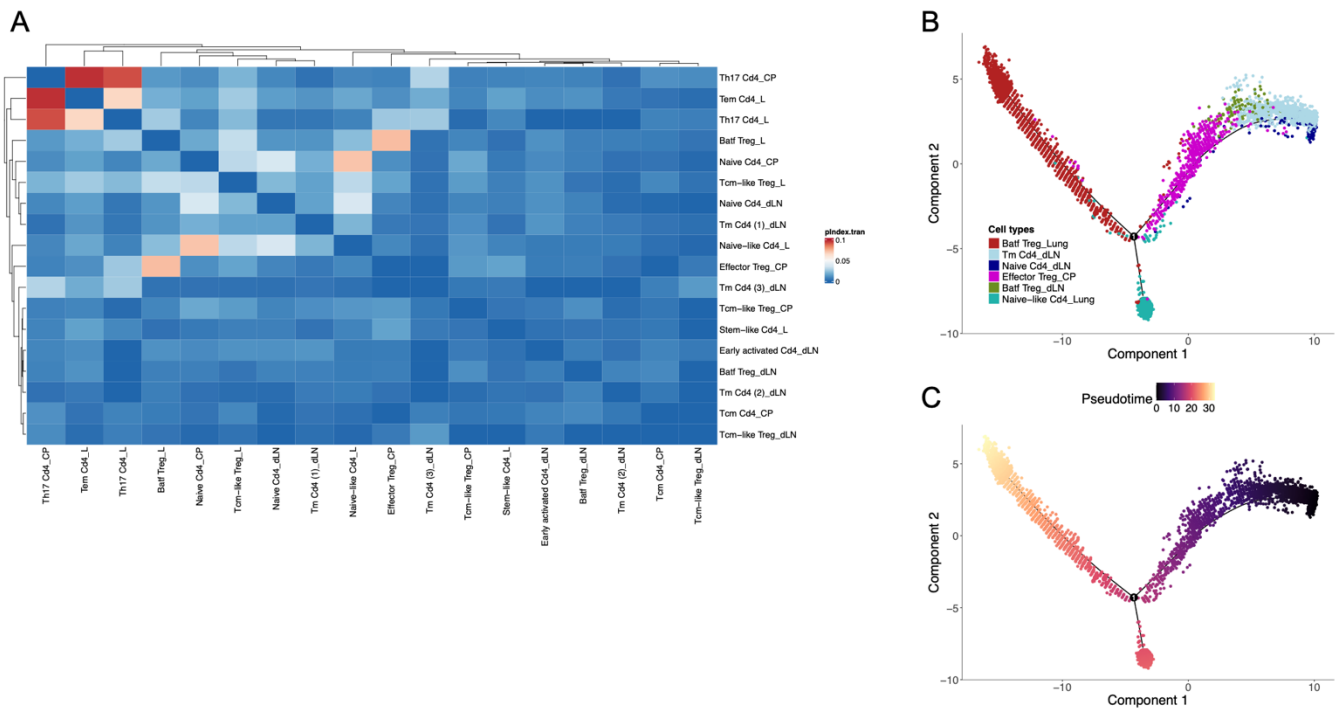

Figure 5.11 – The differentiation of CD4 subsets amongst all tissues in the NTCU-treated mice at the 18-week timepoint. A) Heatmap showing pairwise STARTRAC-transition between CD4 subsets across all tissues in the NTCU-treated mice at the 18-week timepoint. Pairwise STARTRAC-transition is used to quantify how frequently T cell clones are shared between two different cell states or populations using scRNAseq with paired TCR information. It measures the extent to which cells with identical TCRs are found in both states, which suggests that individual clones have transitioned from one state to another during differentiation or activation. A high STARTRAC-transition score indicates strong clonal connectivity and likely developmental relatedness between the two populations, whereas a low score suggests limited state transition. B) Trajectory analysis between those CD4 subsets that shared TCRs with Batf+Tregs in the NTCU lung and Batf+Tregs in the NTCU lung using monocle2. C) Pseudotime analysis between those CD4 subsets that shared TCRs with Batf+Tregs in the NTCU lung and Batf+Tregs in the NTCU lung using monocle2.

5.4 Discussion

In this study, I aimed to understand the clonal origins and differentiation of Batf+Tregs in high-grade preinvasive lesions from the NTCU mouse model. Dr Marta Lebrusant-Fernandez and Abigail Shurr sorted total CD3+ live cells enriched for CD4+CD25+Treg populations from the lungs of NTCU-treated mice at 18 weeks (n=3) alongside age-matched controls (n=3). The team also sorted live T cells from dLN (n=2 per group) and cardiac puncture (n=3 per group), and then we performed scRNAseq with TCR-seq. This is the current best method to study the clonal origins and T cell differentiation, as it allows me to be able to track TCR clones of Batf+Tregs and examine whether there are CD4 subsets from dLN, CP , and lung that have shared TCR clones with Batf+Tregs in lung by coupling together single-cell gene expression with TCR data. This study is the first ever analysis of T cell clonal dynamics during high-grade preinvasive lung disease *in vivo* and provides a translational framework for understanding the immune response occurring just prior to invasion. These data can inform long-term research into immune-driven early detection and prevention or interception of lung cancer as a valuable resource. Here, these data showed overall that there is a complex integration of blood, dLN and lung Treg pools, suggesting that maximal Treg targeting may require a systemic approach rather than local. This fits with the data of my colleagues who showed that using a systemic acting mAb (anti-CD25) but not a locally depleting mAb (anti-CTLA-4) can reduce Treg numbers, proposing that we might need to target the broader Treg network to block accumulation in preinvasive lesions.

I observed that Batf+Tregs were significantly increased in high-grade lesions in NTCU-treated mice at the 18-week timepoint compared to age-matched controls, which is similar to what I found in NTCU-treated mice at the 24-week timepoint. This suggests that the abundance of Batf+tregs increased over time during LUSC development, consistent with total Foxp3+CD4 numbers at the bronchial tree when stained by mIF by my colleague Dr Zoe Whiteman in her PhD thesis and our joint manuscript in revision. This is also similar to human pre-malignant lesions, where BATF+Tregs were significantly increased in human CIS vs normal tissue (Fig. 3.3A). Also, Batf+Tregs had significantly strong enrichment of gene signature of BATF+Tregs

in human CIS, Batf+Tregs in NTCU model at 24-week timepoint, and OX40hi/GITRhi Tregs in NSCLC from Dykema et al. OX40hi/GITRhi Tregs were identified as highly suppressive Tregs that were associated with antiPD1 resistance (175).This suggests that Batf+Tregs in the NTCU mouse model at the 18-week timepoint had a similar transcriptional profile to BATF+Tregs in human CIS and Batf+Tregs in the NTCU mouse model at the 24-week timepoint. This indicates that Batf+Treg identity is established early, maintained into invasion and is conserved between mouse and human models.

Within just the lung, Batf+Tregs in the NTCU model had the highest sharing of TCR clones with naive-like CD4 (excluding sharing amongst themselves, which occurs in NTCU because Batf+Tregs are clonally expanded). Moreover, naïve-like CD4 in NTCU lung was predicted to be at an early stage of Batf+Treg differentiation during LUSC carcinogenesis (Fig. 5.3A). This suggests that naïve-like CD4 in the NTCU lung could differentiate into Batf+Tregs in the NTCU lung, an event called iTreg conversion or differentiation, which is known to occur in response to (i) TGFβ signalling, which is the canonical pathway driving Foxp3 induction in peripheral naïve CD4+ T cells, (ii) IL-2 availability, which is essential survival and expansion signal for Tregs, in lesions, IL2 often comes from activated effector CD4/CD8 T cells, but Tregs consume it preferentially, reinforcing their own expansion, (iii) co-inhibitory interactions (e.g., PD-1/PD-L1 and CTLA-4/CD80/CD86 ligation both favour conversion and stabilisation of iTregs, (iv) chronic antigen presentation and MHCII exposure in invasive tumours. This also reflects what I observed in human CIS lesions (Chapter 3), and together this is the first known example of iTreg differentiation from naïve precursors in squamous pulmonary pre-malignancy ever shown. This is important because blocking this conversion event could reduce Tregs accumulating in high-grade preinvasive disease and possibly therefore, intercept LUSC. According to Wang et al., who performed scRNA-TCRseq of early-stage LUAD (GGO nodules) there were 21.5% of shared TCR clonotypes between naïve CD4+ T cells and Tregs also in this context (137). They proposed that naïve CD4 T cells preferentially differentiated into Tregs during pre-invasive stages (176).This indicates the potential possibility model of naïve-like CD4 in NTCU lung becomes Batf+Tregs in the NTCU lung and that this may be common in LUAD, representing a major axis for us to target in the future.

In terms of overlap with the blood, Batf+Tregs in the NTCU lung had the highest shared TCR clones with effector Tregs in NTCU CP (Fig. 5.6A). Also, effector Tregs in NTCU CP were predicted to be an early stage of Batf+Tregs differentiation during LUSC progression via pseudotime. This suggests that effector Tregs in NTCU CP (or generally the circulation) could differentiate into Batf+Tregs in NTCU lung, supporting that both local and circulating cells could converge into the Batf+Treg pool. Ahmadzadeh et al. recently aimed to determine whether tumour-infiltrating Tregs are antigen-driven, clonally expanded, and neoantigen reactive in invasive NSCLC. They found that there were significantly more shared TCR clones between intratumoral and circulating Tregs, that with Tregs and Tconv; implying the migration and expansion of tumour-specific Tregs from the periphery (135). This suggests the potential possibility model of effector Tregs in NTCU cardiac puncture become Batf+Tregs in NTCU lung. However, more recently, several papers have suggested that Eff.Tregs in the blood could arise from (‘retrograde’) migration from the tumour and that Tregs are highly mobile across tissues (unlike Tex CD8 T cells) (134). Its therefore possible that eff.Tregs come from Batf+Tregs rather than vice versa.

Batf+Treg in the NTCU lung had the highest shared TCR clones with naïve CD4 in the NTCU dLN. Naïve CD4 in NTCU dLN was also predicted to be an early stage of Batf+Tregs differentiation in the lung during LUSC carcinogenesis. This suggests that Naïve CD4 in NTCU dLN could differentiate into Batf+treg in NTCU lung. There are no direct studies showing naïve CD4 T cells in dLN transitioning into suppressive Tregs in pre-invasive disease. This could be the possibility model of Naïve CD4 in the NTCU dLN become Batf+Tregs in the NTCU lung. However, a small number of Batf+Tregs in the lung also shared TCRs with Batf+Tregs in the dLN and this only occurred in the NTCU but not control animals. Also, Batf+tregs in NTCU dLN but not control were clonally expanded. So, in a complex model, it seems possible that during preinvasive disease i) local naïve CD4, ii) blood eff.Tregs iii) dLN naïve CD4 and iv) Batf+Tregs in the dLN can all contribute to Batf+Tregs in the lung of NTCU animals in preinsaisve disease. Yet, its important to remember that still the vast majority (84.6%) of these clones in the lung do not share TCRs with any other subsets and are likely locally expanded and maintained in situ (possibly via antigens within the lesions but not being used in the dLN for priming). Interestingly, it’s been shown that under homeostatic conditions, Eff.tregs and naïve CD4 may convert to tissue-Tregs

(essentially Batf+Tregs) (88), Therefore the increase in Effector Tregs and Batf+Tregs may mostly be an increase in an existing differentiation axis, due to inflammation or increased local APCs with MHCII.

I combined TCR clones of CD4 subsets from dLN, CP, and lung and then performed pairwise transition analysis using STARTRAC (124) to identify which CD4 subsets had the highest sharing of TCR clones with Batf+Tregs in the NTCU lung. Effector Tregs in NTCU CP had the highest score of pairwise transition with Batf+ Tregs in the NTCU lung compared to other CD4 subsets from other tissues. This suggests that the non-private Batf+Tregs in NTCU lung primarily originate from effector Tregs in NTCU CP more than from the other sources. Similarly, Ahmadzadeh et al (135), compared TCRβ clones across FOXP3+Tregs in tumour and peripheral blood lymphocytes and FOXP3- conventional T cells in tumour and blood. They found that there was significant overlap of TCRβ clones with FOXP3+Tregs in blood, whereas there was minimal overlap of TCRβ clones with FOXP3- conventional T cells either in tumour or blood. This indicates that the non-private Batf+Tregs primarily originate from effector Tregs rather than being locally induced from conventional CD4+ T cells, and that this resembles data from invasive NSCLC.

The limitation of this study is that the number of mice, especially dLN (n = 2 per group). Moreover, as we only focus on the NTCU mouse model, the TPA skin model could also be an additional model to study squamous cancer progression in another tissue (177). TPA (12-O-tetradecanoylphorbol-13-acetate) is a tumour promoter used in classic two-stage skin carcinogenesis models. This model mimics stepwise progression from hyperplasia, papilloma, to invasive cancer and also shows strong infiltration of neutrophils, macrophages, and Tregs (177). This could also help us study immunosuppressive mechanisms and CD4+ T cell differentiation (e.g., Treg) and help to see if my findings are generalisable for multi-cancer interception (178).

In this chapter, I highlighted the potential possibility for models of CD4+ conventional (e.g., naïve-like CD4) and effector Tregs differentiation pathways to become Batf+Tregs in NTCU lung. My work suggests that to optimise Treg-depletion therapy, it could be beneficial to target both Batf+Tregs and other CD4 subsets (e.g.,

effector Tregs), which might prevent repopulation of the tumour by Batf+Tregs after depletion and therefore delay or block re-establishment of immunosuppression in the preinvasive TME. However, this approach should be taken with caution to avoid systemic autoimmunity and loss of peripheral immune tolerance. This could be done by finding unique targets on these populations in NTCU-treated mice vs aged-matched controls or by intermittent dosing of mAbs, such as has been shown to minimise toxicity and preserve function of the PI3Kδ (179). Future work is also needed to better define possible Treg-mediated suppression of B cells, which could be a new axis of preinvasive adaptive immune suppression to target for LUSC interception.

5.5 Summary

- Batf+Tregs were significantly increased in NTCU lung at the 18-week timepoint
- Batf+Tregs clonally expanded in dLN and lung during LUSC carcinogenesis
- Naïve-like CD4 in NTCU lung, effector Tregs in NTCU CP, and naïve CD4 in NTCU dLN could differentiate into Batf+Tregs in NTCU lung
- The minor population of non-private Batf+Tregs in NTCU lung primarily originate from effector Tregs in NTCU CP

Chapter 6. Discussion and Future Work

Lung cancer is the leading cause of cancer-related death worldwide (1). It is broadly categorised into small-cell lung cancer, which is characterised by small, round cells, and NSCLC. NSCLC accounts for approximately 85% of all lung cancer cases, with LUSC being the common subtype amongst smokers. The high mortality rate of lung cancer is primarily due to the lack of symptoms, diagnosis, and treatment until later stages of progression. Approximately three-quarters of patients with lung cancer are diagnosed with incurable disease at stage III or IV (2), at which point the disease has spread to lymph nodes or other organs. Earlier detection at stages I and II is associated with significantly higher rates of survival, as a wider variety of treatments can be used, such as surgical resection, which is standard of care for early disease.

CIS represents a critical stage in the progression from premalignant lesions to invasive LUSC. It is considered to be the sweet spot to study for LUSC interception because CIS is a late premalignant lesion stage that directly precedes invasive carcinoma in the stepwise progression from normal epithelium through hyperplasia, metaplasia, dysplasia, and CIS to invasive cancer. Also, CIS is the high-grade pre-invasive lesion stage where the disease is at a pivotal point: roughly half of CIS lesions progress to invasive LUSC, whereas about a third spontaneously regress (180). This intermediate stage allows me to observe and compare the biological and immunological differences between CIS lesions and normal lung tissue that could provide valuable insights into mechanisms of immune surveillance and immune escape that occur early in LUSC carcinogenesis.

In addition, Mascaux et al performed analyses on bulk transcriptomics of human pre-malignant lesions; they found that CIS lesions show a significant upregulation of immune-related genes, particularly those involved in immune response pathways, compared to lower-grade lesions. There is an increased expression of co-inhibitory immune checkpoint molecules and suppressive interleukins such as *IDO1, PD-L1 (CD274), TIGIT, CTLA4, ICOS, IL10*, and *IL6,* which are indicative of immune suppression mechanisms being activated at the CIS stage (24). Simultaneously, stimulatory molecules (e.g., *TNFRSF9, TNFRSF18, ICOS, CD80, CD86, CD70,*

TNFSF9, TNFRSF25 also exhibit increased expression, reflecting a complex immune microenvironment with both activation and suppression signals (24). Moreover, gene ontology analysis revealed that CIS lesions have a large number of upregulated immune functions, including negative regulation of the immune system, antigen processing, and presentation of peptide antigens, which contribute to the immune escape mechanism. In addition, genes involved in the negative regulation of immune responses, such as *TNFRSF14, CD200, CD59, TGFB3*, and *HLA-G*, are upregulated in the high-grade lesions, including CIS, contrary to their downregulation in the low-grade lesions, indicating a shift toward immune evasion (180). Since Mascaux et al. used bulk transcriptomics of human biopsies before the tumour evasion and used CIBERSORTX to estimate the abundance of immune cells during LUSC development, they were unable to conclude which T cell subsets exactly increased during LUSC carcinogenesis at the single-cell level. This analysis of early immunodynamics at the single-cell level could identify targetable dysfunctional (e.g., exhausted CD8 or Tregs) and specific checkpoints or pathways that are druggable for cancer prevention, especially in patients with high-grade lesions that have a major risk of developing LUSC. This work could provide a precise method to intercept cancer progression and also identify markers for early detection.

6.1 Thesis Conclusion

Using scRNAseq data, I analysed human CIS lesions and adjacent lung tissue to define early immune escape mechanisms during LUSC progression. In a primary UCLH bronchoscopy cohort, CIS lesions were characterised by an early and consistent expansion of a highly suppressive BATF+Helios+OX40+GITR+ICOS+CD177+ Treg population (BATF+Tregs; n = 16 samples from 7 patients). This finding was independently validated in a large-scale scRNAseq meta-analysis comprising 457 samples from 251 patients, as well as in bulk gene expression profiling of 122 bronchial biopsies from 77 patients. BATF+Tregs were additionally associated with resistance to anti-PD1 therapy in metastatic lung cancer using publicly available datasets (n = 195), independent of tumour mutational burden and overall immune infiltration. Clonal tracking within CIS lesions indicated that BATF+Tregs arise through a combination of local clonal expansion, in situ

differentiation from stem-like progenitor pools, and recruitment of effector Tregs from the circulation. Building on this observation, my team demonstrated that circulating effector Tregs can be leveraged as a minimally invasive biomarker to track and predict preinvasive disease progression in individuals undergoing bronchoscopy and CT screening, forming the basis of a patented early-detection platform. To establish evolutionary conservation, I performed scRNAseq of lung, blood, and dLNs in a carcinogen-driven mouse model of LUSC, revealing a parallel expansion of Batf+Tregs and systemic effector Tregs during pre-invasive progression in vivo. Finally, I identified candidate pathways that regulate Batf+Treg differentiation, one of which has been functionally validated by my team, collectively supporting a model in which BATF+Tregs represent an early, trackable, and therapeutically targetable axis of immune escape that can be exploited for immune interception of LUSC in high-risk individuals.

6.2 Divergent Treg States in Pre-invasive LUSC and Checkpoint Blockade Response

I found an early accumulation of BATF+Tregs in human CIS. These cells show a transcriptional program characterised by *FOXP3, BATF, IL2RA (CD25), CTLA4, OX40 (TNFRSF4), and GITR (TNFRSF18)*, as well as clonal expansion of BATF+Tregs within the CIS lesions. Their accumulation at the pre-invasive stage points to an immunosuppressive event established before tumour invasion, which could dictate whether lesions regress or progress, and shape subsequent responsiveness to immune checkpoint blockade.

These findings align closely with the recent study by Dykema et al. (2023), who performed integrated scRNA-seq and TCR-seq profiling of TIL-Tregs from treatment-naïve and anti–PD-1–treated NSCLCs (86). They identified a transcriptionally distinct OX40hi/GITRhi Treg subset, representing ~20% of TIL-Tregs, that had the high expression levels of suppressive genes (*IL2RA, ENTPD1/CD39, EBI3, LAYN*)

and was functionally the most immunosuppressive *in vitro* (86). Critically, this subset was enriched in tumours resistant to PD-1 blockade, indicating that it is a key mediator of ICB failure (86). The gene signature of BATF+Treg observed in human CIS had a strong enrichment score with the gene signature of OX40hi/GITRhi Treg subset using GSEA, suggesting that the OX40hi/GITRhi suppressive program emerges early in pre-invasive lesions and persists into invasive disease.

Importantly, Dykema et al. also uncovered a contrasting Treg trajectory. Using a murine tumour model, they tracked tumour antigen–specific Tregs and showed that these cells can lose *FOXP3*, downregulate *CD25*, and acquire Th1-like properties, including T-bet (*TBX21*), IFN-γ, *CCL4*, and *CCL5* expression. This Th1-like Treg was found to be a subcluster (SC0) enriched in PD-1–responsive NSCLC (86). Thus, intratumoral Tregs are not monolithic suppressors but exist along divergent paths: one toward an activated OX40hi GITRhi suppressive fate associated with resistance, and another toward a Th1-like state associated with response (86). Since our scRNAseq of human pre-malignant lesions does not contain many cells, it would be beneficial to identify Th1-like Tregs in our study once we obtain more cells and patient samples.

Moreover, TCR overlap analysis revealed the progenitor CD4, Th17 CD4, and Tfh CD4 subsets as potential precursors to BATF+ Tregs in human CIS. In PBMC, BATF+Tregs overlapped clonally with Tcm-like and effector Tregs, suggesting peripheral contributions to BATF+Treg differentiation. This dual origin supports a model where BATF+Tregs arise both locally and peripherally, creating a dynamic and replenishable suppressive pool. The presence of circulating BATF+Tregs also raises the possibility of using them as a liquid biomarker for early detection.

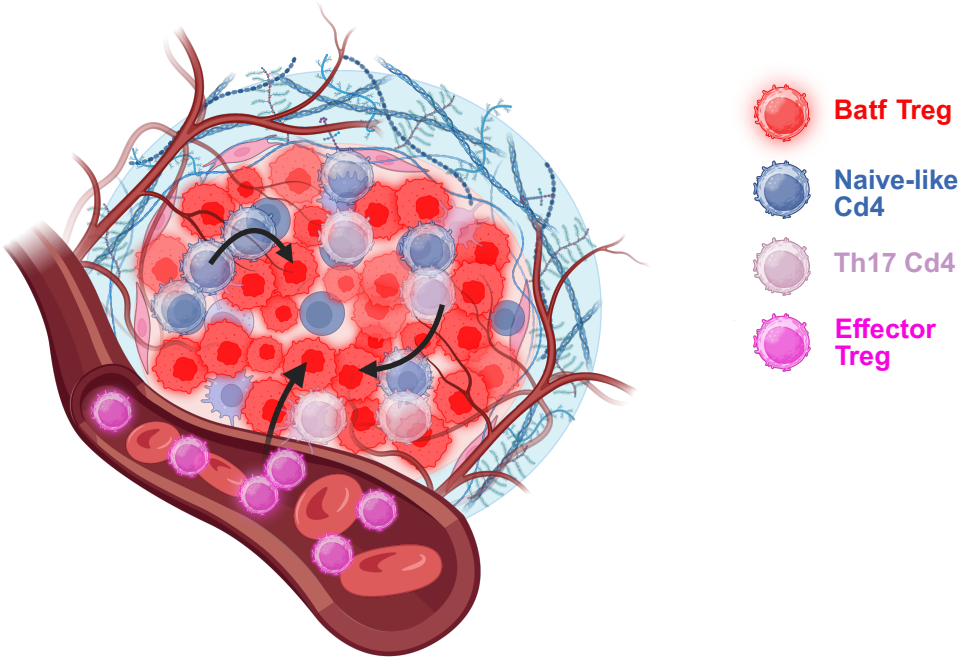

Figure 6.1 – The potential models of BATF+Treg differentiation during LUSC carcinogenesis. Within human CIS lesions, BATF+Tregs were clonally expanded, and Naïve-like CD4 and Th17 CD4 cells could differentiate into BATF+Tregs. Moreover, effector Tregs, which had similar transcriptional profiles to BATF+Tregs, had highly shared TCR clones with BATF+Tregs.

6.3 MHC Class II-Mediated Immune Crosstalk in Human CIS

A key theme emerging from my thesis is that immune escape mechanisms were observed in human CIS using our scRNAseq of human pre-invasive lesions. My results showed that MHC-II expression was decreased in basal cells in human CIS compared to basal cells in normal lung tissue. This suggests that antigen processing and presentation declined in human CIS. However, the MHC-II expression was increased in APCs (e.g., B cells) in human CIS compared to normal lung tissue. This indicates that APCs in human CIS lesions may support BATF+Treg differentiation in the tolerogenic TME, and this could prevent effective antitumour immune responses.

In addition, Ning et al. (2023) have provided important mechanistic insight into epithelial contributions to immune evasion in premalignant lesions (136). Their study demonstrated that the YAP/TAZ–TEAD–TP63 transcriptional complex in basal bronchial epithelial cells directly represses CIITA, the master transactivator of MHC class II genes (136). They demonstrated that this repression results in reduced MHC-II gene expression in progressive bronchial premalignant lesions, thereby contributing to immune evasion by diminishing epithelial-immune communication, particularly with Th1 CD4+ T cells (136). The study found that low CIITA and MHC-II expression levels are associated with lesion progression and decreased Th1 cell infiltration across multiple datasets, whereas regressive lesions tend to maintain higher epithelial MHC-II expression (136).

Bringing these findings together, a striking picture emerges in which MHC-II operates in two opposing but complementary ways in human CIS:

1. MHC-II expression increased in APCs in human CIS (B cells, cDC2, pDC): In my CIS dataset, MHC-II+ dendritic subsets were associated with BATF+ Treg expansion. Rather than priming effective effector CD4+ responses, these interactions appear to reinforce suppressive immunity, favouring BATF+ Treg proliferation and stability.
2. MHC-II expression decreased in epithelial cells in human CIS (basal cells): Ning et al. show that expression of MHC-II in epithelial cells is repressed in progressive premalignant lesions via YAP/TAZ–TEAD–TP63–mediated silencing of CIITA (136). The consequence is reduced presentation of epithelial-derived antigens to CD4+ effector T cells and a loss of Th1 infiltration, which is linked to lesion progression.

This combined mechanism leads to diminished adaptive immune responses and could help to explain why CIS lesions persist or progress despite the presence of immune surveillance, favouring tolerance over immune-mediated clearance.

In addition, my cell-cell interaction analyses showed that antigen-presenting cell populations, such as cDC1, cDC2, and pDCs in human CIS, provided enriched

MHC-II signals to BATF+Tregs, whereas epithelial-derived MHC-II signals were decreased compared to normal. This shift from epithelial to immune APC dominance likely reflects a rewiring of antigen presentation capacity to favour tolerance.

Ligand–receptor analyses identified multiple axes enriched in human CIS, including: CCL17–CCR4 and CCL22 – CCR4: the CCL17-CCR4 axis has been quite well described in the context of lung cancer immunology. CCR4 is one of the hallmark chemokine receptors for Tregs, especially activated/effector Tregs. Its ligands are CCL17 and CCL22, both produced by dendritic cells, macrophages, and tumour cells (67). Engagement of CCR4 by CCL17/CCL22 promotes Treg chemotaxis into tissues and enhances their survival and suppressive activity (67). In cancer, this pathway is a central mechanism for recruiting immunosuppressive Treg cells to the tumour microenvironment. In human NSCLC, several studies show tumour-associated macrophages and DCs secrete CCL17/CCL22, creating a gradient that attracts CCR4+FOXP3+ Tregs into tumours (67). Higher CCL17/CCR4 expression correlates with increased Treg infiltration, poor effector T-cell activity, and worse prognosis (67). Blockade of CCR4 (e.g., with mogamulizumab, a CCR antibody used in T-cell lymphoma) reduces TIL-Tregs and restores CD8+ T-cell activity in preclinical lung cancer models (67).

CD274 – PDCD1 (PD-L1 – PD-1): the CD274-PDCD1 axis is not just about CD8 exhaustion; it also has profound effects on Tregs. Tregs constitutively express PD-1, and many TIL-Tregs are PD-1 high. PD-L1 (CD274) is expressed by tumour cells, dendritic cells, macrophages, and even epithelial cells in dysplastic lesions (181). PD-1/PD-L1 signalling has two key effects on Tregs: (i) supports Treg stability and expansion: PD-1 signalling prevents Treg apoptosis and promotes FOXP3 expression (181). (ii) enhances suppressive function: PD-1 engagement increases IL-10, TGFβ, and CTLA-4 activity in Tregs. Some studies also show that blocking PD-1 can destabilise Tregs and reduce their suppressive capacity, even if effector T cells are simultaneously reinvigorated (182). NSCLC tumours are rich in PD-1hi effector Tregs, and high PD-L1 expression by tumour cells or myeloid cells correlates with greater Treg density and immunosuppressive activity in the TME. In murine models of lung cancer, PD-L1 deficiency reduces Treg accumulation, while PD-1 agonism can expand TIL Tregs (183). Moreover, in clinical settings, anti-PD-1/PD-L1 therapies not

only reinvigorate CD8+ T cells but also indirectly affect Treg homeostasis – sometimes reducing their suppressive capacity (184).

ICAM1 – ITGAL (LFA-1): The ICAM1-ITGAL axis is less commonly discussed than CCR4 or PD-L1 when it comes to Tregs, but it plays an important adhesion and co-stimulatory role that can strongly influence Tregs in both lung cancer and pre-invasive lesions (185). ITGAL encodes CD11a, part of the LFA-1 integrin, expressed on T cells, including Tregs. ICAM1 is broadly expressed on endothelium, epithelial cells, APCs, and tumour cells (185). In terms of functional consequences for Tregs, ICAM1-LFA-1 interactions allow Tregs to adhere to tumour vasculature and migrate into tissues. ICAM1 on DCs provides stabilisation for TCR-MHC interactions during antigen recognition, promoting Treg priming and expansion (185). Tregs use LFA-1/ICAM1 to form stable immunological synapses with DCs or effector T cells, strengthening suppression (185). Also, LFA-1 signalling provides co-stimulation that can help maintain FOXP3 expression.

ENTPD1-ADORA2A: the ENTPD1-ADORA2A axis represents a metabolic immune checkpoint that is particularly important for Treg function and stability in both invasive lung cancer and pre-invasive settings (186). ENTPD1 encodes CD39, an ectonucleotidase highly expressed on Tregs (especially TIL-Tregs, fragile Tregs, and sometimes exhausted CD8s). CD39 converts extracellular ATP/ADP into AMP. NT5E, often co-expressed on Tregs, endothelial cells, tumour cells) then converts AMP into adenosine (187). ADORA2A encodes the adenosine A2A receptor, expressed on T cells (including Tregs), NK cells, and other lymphocytes (187). In terms of functional effects, Adenosine signalling through ADORA2A increases intracellular cAMP, which (i) enhances FOXP3 stability and Treg lineage commitment. (ii) Boosts Treg suppressive activity (IL-10, TGFβ production) (67). (iii) Inhibits effector T-cell proliferation and cytokine production. Moreover, it creates a strongly immunosuppressive microenvironment, particularly in hypoxic or inflamed tissues like tumours and dysplastic epithelium (187).

LGALS9-CD44: the LGALS9-CD44 axis contributes to Treg stability and suppressive activity and may already be active in pre-invasive lesions. LGALS9 encodes Galectin-9, a beta-galactoside-binding lectin secreted by tumour cells, epithelial cells, and APCs (188). CD44 is a glycoprotein expressed on activated T cells, memory T cells, and particularly TIL-Tregs (188). In terms of functional consequences for Tregs, Galectin-9 binds to glycosylated CD44, which stabilises FOXP3 expression and promotes Treg lineage maintenance, enhances Treg survival in hypoxic and stressed environments, boosts Treg migratory capacity and tissue retention (188). Moreover, Galectin-9 also interacts with TIM-3 (HAVCR2) on effector T cells, inducing exhaustion/apoptosis, indirectly favouring Treg dominance (188). In NSCLC tumours, tumour cells and tumour-associated macrophages often overexpress Galectin-9 (LGALS9). CD44hiFOXP3+ Tregs are enriched in NSCLC TME and show stronger suppressive activity than CD44lo counterparts. In terms of functional studies, blocking Galectin-9 reduces Treg stability and suppressive function while enhancing CD8 responses. Moreover, High Galectin-9 correlates with poor prognosis and resistance to immune checkpoint blockade (188).

In human CIS, both B cells and fibroblasts show increased expression of MHC-II expression. However, fibroblasts are usually not considered to be key players in immune regulation; they may function as atypical APCs that could promote BATF+Treg differentiation and formation. Those interesting crosstalk axes results underpinned a CRUK biology to prevention grant awarded to my team in 2023 to functionally explore additional immune interception targets for LUSC.

6.4 Targeting BATF+Tregs for Immune Interception in Pre-Invasive LUSC

Targeting BATF+Tregs could be a promising target for immune interception in pre-invasive LUSC because BATF+Tregs were found across species (both in human and mouse models). In our NTCU mouse model, single-cell analysis revealed a significant increase of Batf+Tregs in the bronchial tree at week 24 compared with age-matched controls, paralleling their rise in human CIS. Transcriptional profile

comparison showed that murine Batf+Tregs closely resemble BATF+Tregs derived from human CIS biopsies. These findings suggest that the Batf+Tregs were also found before tumour invasion and that the murine pre-invasive model can be used to identify clinically relevant targets for interception in humans.

GSEA analysis identified a key signalling pathway: Batf+Tregs in both the mouse model and human CIS showed selective enrichment for PI3K–AKT–mTOR signalling compared to other CD4+ T cell subsets. This finding suggested that pharmacologically targeting PI3K signalling might specifically impair Batf+ Tregs within pre-invasive lesions without affecting other CD4+ lineages. Based on this insight, Dr. Zoe Whiteman and Dr. Sandra Gomez-Lopez conducted an experiment in NTCU-treated mice. Treatment with the PI3Kδ-selective inhibitor PI-3065 led to reduced tumour incidence and elimination of the largest tumour fraction, decreased FOXP3+CD4+ Tregs, and enhanced proximity of CD8+ T cells within lesions. In contrast, the PI3Kα inhibitor BYL719 did not produce these effects. These results provide functional evidence that the Batf+ Treg in the pre-invasive LUSC could be targeted by the PI3Kδ inhibitor. However, this effect was demonstrated at the single timepoint, and there was no evidence of a survival benefit or durable tumour suppression.

The single-cell interactome analysis reveals that the pre-invasive niche is already structured by cell–cell interactions that support regulatory T cells (Tregs), including the PD-L1–PD-1 checkpoint, ICAM1–ITGAL (LFA-1) adhesion, and the ENTPD1–ADORA2A (CD39–A2A) metabolic pathway, along with an increased population of cDC2 cells. These pathways likely work together with PI3K signalling to stabilise Batf+ Tregs by enhancing their recruitment and retention, promoting tolerogenic priming, and supporting adenosine-driven lineage maintenance. The presence of these circuits in carcinoma *in situ* (CIS) suggests that PI3Kδ inhibition operates within an already immunosuppressive environment. This insight supports the rationale for combination therapies—such as pairing PI3Kδ inhibitors with A2A antagonists to disrupt adenosine-mediated Treg stability or temporarily targeting LFA-

1–ICAM1 interactions to decrease Treg retention in epithelial tissue—while carefully considering potential risks related to impaired cell trafficking or activation.

Tracking of T cell clones across blood, draining lymph nodes, and lung reveals that the Batf+ Treg population is primarily expanded within the CIS lesions, with a smaller contribution from local conversion of naïve-like CD4 cells within the lesion microenvironment. Rather than fully explaining differential therapeutic efficacy, this organisational structure constrains possible mechanisms of Treg maintenance, indicating that Batf+Treg accumulation is largely sustained by in situ proliferation of established clones. This provides a plausible rationale for why systemic treatments, such as anti-CD25 antibodies or small-molecule PI3Kδ inhibitors, which impair Treg survival or function, are more effective than local Treg depletion approaches in reducing Treg accumulation in pre-invasive lesions. Furthermore, it implies that combining functional inhibition of Tregs via PI3Kδ blockade with strategies that limit their replenishment—such as suppressing cDC2-driven conditioning or adenosine signalling, which could more effectively and durably reduce the Batf+ Treg differentiation.

A key limitation of this study is the relatively small number of whole-cell populations and T cells available for analysis in human PID. To overcome this, I employed a bootstrap version of miloR, integrated our single-cell RNA-seq dataset of T cells with reference scRNA-seq data from the lung cancer atlas, and further validated the results by examining the expression of a BATF+ Treg gene signature in bulk transcriptomic profiles from 122 bronchial biopsies obtained from 77 smokers. These complementary methods confirmed the accumulation of BATF+ Tregs during LUSC carcinogenesis. However, the limited number of T cells restricted the detection of TCRs associated with BATF+ Tregs and other CD4+ subsets. To enhance future analyses, increasing the number of normal samples would be beneficial to achieve a more balanced comparison between normal and CIS-associated T cell populations.

6.5 Future Work

Future work will be required to define the precise spatial localisation of BATF+Tregs within CIS lesions, particularly in relation to the dysplastic epithelium, basement membrane integrity, and the distribution of effector T cells, antigen-presenting cells, and key inhibitory ligands. Because scRNAseq lacks spatial resolution, it remains unclear whether BATF+Tregs reside within the epithelial compartment or are primarily confined to the surrounding stroma. Spatially resolved approaches, including multiplex immunofluorescence or immunohistochemistry, spatial transcriptomics, and targeted *in situ* hybridisation for BATF, FOXP3, and related markers (e.g., OX40, GITR, and CD177), would allow intralesional and stromal niches to be distinguished and determine whether BATF+Tregs co-localise with immune exclusion or epithelial immune-evasion programs. Establishing this microanatomical context is essential for translating the observed association between BATF+Tregs and CIS into a mechanistic understanding of early immune escape, and for guiding immune-interception strategies aimed either at disrupting suppressive perilesional barriers or counteracting intralesional T-cell suppression.

Moreover, my work underscores practical challenges that will influence future research directions: Since I found BATF+Tregs, cDC2, B cells were significantly enriched in human CIS lesions, it would be great to examine whether accumulation of BATF+Tregs, cDC2, and B cell infiltration, MHC-II processing and presentation mutation, loss of heterozygosity at human leukocyte antigens, MHC-II germline homozygosity, and high TCR clonality in blood or lesions associate with poor progression-free survival and faster tumour growth rate in early stage of lung cancer in the future. Moreover, it would be interesting to identify a key epigenetic regulator of BATF+Tregs, the DNA regulatory elements, and transcription factor networks that control the differentiation and activation of BATF+Tregs in human PID. Furthermore, because Tregs function systemically, designing treatment regimens requires careful balancing to achieve therapeutic efficacy without disrupting immune homeostasis. Nevertheless, by grounding target selection in transcriptional patterns conserved across species and confirming findings through targeted pharmacologic interventions,

my story demonstrates an approach for cancer interception in LUSC via BATF+Treg targeting. PI3Kδ inhibition emerges as a promising primary strategy to target BATF+Tregs.

Chapter 7. Appendix

7.1 Peer-reviewed publications authored during this PhD

Regulatory T cell depletion promotes myeloid cell activation and glioblastoma response to anti-PD1 and tumour-targeting antibodies

Galvez-Cancino F, Navarrete M, Beattie G, Puccio S, Conde-Gallastegi E, Foster K, Morris Y, Sahwangarrom T, et al.
Immunity 58, 1236-1253 (2025).
https://www.cell.com/immunity/fulltext/S1074-7613(25)00162-1

CD4 T cell-directed interception of lung squamous carcinogenesis

Sahwangarrom T*, Gamble S*, Whiteman Z*, Peinador-Marin C*, et al. Under revision at Nature

A Pan-Cancer Dendritic Cell Atlas Reveals Routes of cDC2 Maturation as a Key Factor for Response to Immune Checkpoint Blockade

Qian D, Padmasri S, Murai J, Sivakumar M, **Sahwangarrom T**, et al. Under review in Immunity

Bibliography

1. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. Transl Lung Cancer Res. 2016 June;5(3):288–300.

2. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. Lancet Lond Engl. 2021 Aug 7;398(10299):535–54.

3. 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 Off Publ Int Assoc Study Lung Cancer. 2016 Jan;11(1):39–51.

4. Wise J. UK lung cancer care must improve drastically, says report. BMJ. 2016 Oct 18;355:i5633.

5. Arnold M, Rutherford MJ, Bardot A, Ferlay J, Andersson TML, Myklebust TÅ, et al. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995-2014 (ICBP SURVMARK-2): a population-based study. Lancet Oncol. 2019 Nov;20(11):1493–505.

6. Andreano A, Peake MD, Janes SM, Valsecchi MG, Pritchard-Jones K, Hoag JR, et al. The Care and Outcomes of Older Persons with Lung Cancer in England and the United States, 2008-2012. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 2018 July;13(7):904–14.

7. Kennedy MPT, Cheyne L, Darby M, Plant P, Milton R, Robson JM, et al. Lung cancer stage-shift following a symptom awareness campaign. Thorax. 2018 Dec;73(12):1128–36.

8. Hamaji M, Matsuo Y, Chen-Yoshikawa TF, Mizowaki T, Date H. Surgery and stereotactic body radiotherapy for early stage non-small cell lung cancer: review of meta-analyses. J Thorac Dis. 2019 Aug;11(Suppl 13):S1646–52.

9. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med. 2016 Nov 10;375(19):1823–33.

10. Ishizumi T, McWilliams A, MacAulay C, Gazdar A, Lam S. Natural history of bronchial preinvasive lesions. Cancer Metastasis Rev. 2010 Mar;29(1):5–14.

11. Rock JR, Hogan BLM. Epithelial progenitor cells in lung development, maintenance, repair, and disease. Annu Rev Cell Dev Biol. 2011;27:493–512.

12. Rock JR, Randell SH, Hogan BLM. Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. Dis Model Mech. 2010;3(9–10):545–56.

13. Voynow JA, Fischer BM, Roberts BC, Proia AD. Basal-like cells constitute the proliferating cell population in cystic fibrosis airways. Am J Respir Crit Care Med. 2005 Oct 15;172(8):1013–8.

14. Miettinen M, Sarlomo-Rikala M. Expression of calretinin, thrombomodulin, keratin 5, and mesothelin in lung carcinomas of different types: an immunohistochemical analysis of 596 tumors in comparison with epithelioid mesotheliomas of the pleura. Am J Surg Pathol. 2003 Feb;27(2):150–8.

15. Xu C, Fillmore CM, Koyama S, Wu H, Zhao Y, Chen Z, et al. Loss of Lkb1 and Pten leads to lung squamous cell carcinoma with elevated PD-L1 expression. Cancer Cell. 2014 May 12;25(5):590–604.

16. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer. 2014 Aug;14(8):535–46.

17. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209–49.

18. Bray F, Ferlay J, Laversanne M, Brewster DH, Gombe Mbalawa C, Kohler B, et al. Cancer Incidence in Five Continents: Inclusion criteria, highlights from Volume X and the global status of cancer registration. Int J Cancer. 2015 Nov 1;137(9):2060–71.

19. Hirsch FR, Suda K, Wiens J, Bunn PA. New and emerging targeted treatments in advanced non-small-cell lung cancer. Lancet Lond Engl. 2016 Sept 3;388(10048):1012–24.

20. Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. Lung Cancer Amst Neth. 2001;31(2–3):139–48.

21. Ishizumi T, McWilliams A, MacAulay C, Gazdar A, Lam S. Natural history of bronchial preinvasive lesions. Cancer Metastasis Rev. 2010 Mar;29(1):5–14.

22. Kadara H, Scheet P, Wistuba II, Spira AE. Early Events in the Molecular Pathogenesis of Lung Cancer. Cancer Prev Res Phila Pa. 2016 July;9(7):518–27.

23. Thakrar RM, Pennycuick A, Borg E, Janes SM. Preinvasive disease of the airway. Cancer Treat Rev. 2017 July;58:77–90.

24. Mascaux C, Angelova M, Vasaturo A, Beane J, Hijazi K, Anthoine G, et al. Immune evasion before tumour invasion in early lung squamous carcinogenesis. Nature. 2019 July;571(7766):570–5.

25. Mazzilli SA, Rahal Z, Rouhani MJ, Janes SM, Kadara H, Dubinett SM, et al. Translating premalignant biology to accelerate non-small-cell lung cancer interception. Nat Rev Cancer. 2025 May;25(5):379–92.

26. Teixeira VH, Pipinikas CP, Pennycuick A, Lee-Six H, Chandrasekharan D, Beane J, et al. Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. Nat Med. 2019 Mar;25(3):517–25.

27. McWilliams A, Mayo J, MacDonald S, leRiche JC, Palcic B, Szabo E, et al. Lung cancer screening: a different paradigm. Am J Respir Crit Care Med. 2003 Nov 15;168(10):1167–73.

28. Chhajed PN, Shibuya K, Hoshino H, Chiyo M, Yasufuku K, Hiroshima K, et al. A comparison of video and autofluorescence bronchoscopy in patients at high risk of lung cancer. Eur Respir J. 2005 June;25(6):951–5.

29. Hung J, Lam S, LeRiche JC, Palcic B. Autofluorescence of normal and malignant bronchial tissue. Lasers Surg Med. 1991;11(2):99–105.

30. Lam S, leRiche JC, Zheng Y, Coldman A, MacAulay C, Hawk E, et al. Sex-related differences in bronchial epithelial changes associated with tobacco smoking. J Natl Cancer Inst. 1999 Apr 21;91(8):691–6.

31. Lam S, MacAulay C, leRiche JC, Palcic B. Detection and localization of early lung cancer by fluorescence bronchoscopy. Cancer. 2000 Dec 1;89(11 Suppl):2468–73.

32. Bota S, Auliac JB, Paris C, Métayer J, Sesboüé R, Nouvet G, et al. Follow-up of bronchial precancerous lesions and carcinoma in situ using fluorescence endoscopy. Am J Respir Crit Care Med. 2001 Nov 1;164(9):1688–93.

33. Breuer RH, Pasic A, Smit EF, van Vliet E, Vonk Noordegraaf A, Risse EJ, et al. The natural course of preneoplastic lesions in bronchial epithelium. Clin Cancer Res Off J Am Assoc Cancer Res. 2005 Jan 15;11(2 Pt 1):537–43.

34. Hoshino H, Shibuya K, Chiyo M, Iyoda A, Yoshida S, Sekine Y, et al. Biological features of bronchial squamous dysplasia followed up by autofluorescence bronchoscopy. Lung Cancer Amst Neth. 2004 Nov;46(2):187–96.

35. Jeremy George P, Banerjee AK, Read CA, O'Sullivan C, Falzon M, Pezzella F, et al. Surveillance for the detection of early lung cancer in patients with bronchial dysplasia. Thorax. 2007 Jan;62(1):43–50.

36. van Boerdonk RAA, Smesseim I, Heideman DAM, Coupé VMH, Tio D, Grünberg K, et al. Close Surveillance with Long-Term Follow-up of Subjects with Preinvasive Endobronchial Lesions. Am J Respir Crit Care Med. 2015 Dec 15;192(12):1483–9.

37. Ruterbusch M, Pruner KB, Shehata L, Pepper M. In Vivo CD4+ T Cell Differentiation and Function: Revisiting the Th1/Th2 Paradigm. Annu Rev Immunol. 2020 Apr 26;38:705–25.

38. Kiner E, Willie E, Vijaykumar B, Chowdhary K, Schmutz H, Chandler J, et al. Gut CD4+ T cell phenotypes are a continuum molded by microbes, not by TH archetypes. Nat Immunol. 2021 Feb;22(2):216–28.

39. Zhou L, Chong MMW, Littman DR. Plasticity of CD4+ T cell lineage differentiation. Immunity. 2009 May;30(5):646–55.

40. Borst J, Ahrends T, Bąbała N, Melief CJM, Kastenmüller W. CD4+ T cell help in cancer immunology and immunotherapy. Nat Rev Immunol. 2018 Oct;18(10):635–47.

41. Speiser DE, Chijioke O, Schaeuble K, Münz C. CD4+ T cells in cancer. Nat Cancer. 2023 Mar;4(3):317–29.

42. Ferris ST, Durai V, Wu R, Theisen DJ, Ward JP, Bern MD, et al. cDC1 prime and are licensed by CD4+ T cells to induce anti-tumour immunity. Nature. 2020 Aug;584(7822):624–9.

43. Mumberg D, Monach PA, Wanderling S, Philip M, Toledano AY, Schreiber RD, et al. CD4(+) T cells eliminate MHC class II-negative cancer cells in vivo by indirect effects of IFN-gamma. Proc Natl Acad Sci U S A. 1999 July 20;96(15):8633–8.

44. Tate DJ, Patterson JR, Velasco-Gonzalez C, Carroll EN, Trinh J, Edwards D, et al. Interferon-gamma-induced nitric oxide inhibits the proliferation of murine renal cell carcinoma cells. Int J Biol Sci. 2012;8(8):1109–20.

45. Ellyard JI, Simson L, Parish CR. Th2-mediated anti-tumour immunity: friend or foe? Tissue Antigens. 2007 July;70(1):1–11.

46. Purwar R, Schlapbach C, Xiao S, Kang HS, Jiang X, Jetten AM, et al. Robust tumor immunity to melanoma mediated by interleukin-9-producing T cells. Nat Med. 2012 Aug;18(8):1248–53.

47. Végran F, Berger H, Boidot R, Mignot G, Bruchard M, Dosset M, et al. The transcription factor IRF1 dictates the IL-21-dependent anticancer functions of TH9 cells. Nat Immunol. 2014 Aug;15(8):758–66.

48. Numasaki M, Watanabe M, Suzuki T, Takahashi H, Nakamura A, McAllister F, et al. IL-17 enhances the net angiogenic activity and in vivo growth of human non-small cell lung cancer in SCID mice through promoting CXCR-2-dependent angiogenesis. J Immunol Baltim Md 1950. 2005 Nov 1;175(9):6177–89.

49. Crotty S. T Follicular Helper Cell Biology: A Decade of Discovery and Diseases. Immunity. 2019 May 21;50(5):1132–48.

50. Aoki T, Chong LC, Takata K, Milne K, Marshall A, Chavez EA, et al. Single-cell profiling reveals the importance of CXCL13/CXCR5 axis biology in lymphocyte-rich classic Hodgkin lymphoma. Proc Natl Acad Sci U S A. 2021 Oct 12;118(41):e2105822118.

51. Zappasodi R, Merghoub T, Wolchok JD. Emerging Concepts for Immune Checkpoint Blockade-Based Combination Therapies. Cancer Cell. 2018 Apr 9;33(4):581–98.

52. Sakaguchi S, Mikami N, Wing JB, Tanaka A, Ichiyama K, Ohkura N. Regulatory T Cells and Human Disease. Annu Rev Immunol. 2020 Apr 26;38:541–66.

53. De Simone M, Arrigoni A, Rossetti G, Gruarin P, Ranzani V, Politano C, et al. Transcriptional Landscape of Human Tissue Lymphocytes Unveils Uniqueness of Tumor-Infiltrating T Regulatory Cells. Immunity. 2016 Nov 15;45(5):1135–47.

54. Halvorsen EC, Franks SE, Wadsworth BJ, Harbourne BT, Cederberg RA, Steer CA, et al. IL-33 increases ST2+ Tregs and promotes metastatic tumour growth in the lungs in an amphiregulin-dependent manner. Oncoimmunology. 2019;8(2):e1527497.

55. Choi IK, Wang Z, Ke Q, Hong M, Paul DW, Fernandes SM, et al. Mechanism of EBV inducing anti-tumour immunity and its therapeutic use. Nature. 2021 Feb;590(7844):157–62.

56. Zhang B, Kracker S, Yasuda T, Casola S, Vanneman M, Hömig-Hölzel C, et al. Immune surveillance and therapy of lymphomas driven by Epstein-Barr virus protein LMP1 in a mouse model. Cell. 2012 Feb 17;148(4):739–51.

57. Oliveira G, Stromhaug K, Cieri N, Iorgulescu JB, Klaeger S, Wolff JO, et al. Landscape of helper and regulatory antitumour CD4+ T cells in melanoma. Nature. 2022 May;605(7910):532–8.

58. Kos K, de Visser KE. The Multifaceted Role of Regulatory T Cells in Breast Cancer. Annu Rev Cancer Biol. 2021 Mar;5:291–310.

59. Sakaguchi S, Sakaguchi N, Asano M, Itoh M, Toda M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. J Immunol Baltim Md 1950. 1995 Aug 1;155(3):1151–64.

60. Sakaguchi S, Miyara M, Costantino CM, Hafler DA. FOXP3+ regulatory T cells in the human immune system. Nat Rev Immunol. 2010 July;10(7):490–500.

61. Togashi Y, Nishikawa H. Regulatory T Cells: Molecular and Cellular Basis for Immunoregulation. Curr Top Microbiol Immunol. 2017;410:3–27.

62. Tada Y, Togashi Y, Kotani D, Kuwata T, Sato E, Kawazoe A, et al. Targeting VEGFR2 with Ramucirumab strongly impacts effector/ activated regulatory T cells and CD8+ T cells in the tumor microenvironment. J Immunother Cancer. 2018 Oct 11;6(1):106.

63. Saito T, Nishikawa H, Wada H, Nagano Y, Sugiyama D, Atarashi K, et al. Two FOXP3(+)CD4(+) T cell subpopulations distinctly control the prognosis of colorectal cancers. Nat Med. 2016 June;22(6):679–84.

64. Williams JB, Horton BL, Zheng Y, Duan Y, Powell JD, Gajewski TF. The EGR2 targets LAG-3 and 4-1BB describe and regulate dysfunctional antigen-specific CD8+ T cells in the tumor microenvironment. J Exp Med. 2017 Feb;214(2):381–400.

65. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science. 2011 Mar 25;331(6024):1565–70.

66. Romano E, Kusio-Kobialka M, Foukas PG, Baumgaertner P, Meyer C, Ballabeni P, et al. Ipilimumab-dependent cell-mediated cytotoxicity of regulatory T cells ex vivo by nonclassical monocytes in melanoma patients. Proc Natl Acad Sci U S A. 2015 May 12;112(19):6140–5.

67. Togashi Y, Shitara K, Nishikawa H. Regulatory T cells in cancer immunosuppression - implications for anticancer therapy. Nat Rev Clin Oncol. 2019 June;16(6):356–71.

68. Weissler KA, Caton AJ. The role of T-cell receptor recognition of peptide:MHC complexes in the formation and activity of Foxp3^{+} regulatory T cells. Immunol Rev. 2014 May;259(1):11–22.

69. Jordan MS, Boesteanu A, Reed AJ, Petrone AL, Holenbeck AE, Lerman MA, et al. Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. Nat Immunol. 2001 Apr;2(4):301–6.

70. Feng Y, Arvey A, Chinen T, van der Veeken J, Gasteiger G, Rudensky AY. Control of the inheritance of regulatory T cell identity by a cis element in the Foxp3 locus. Cell. 2014 Aug 14;158(4):749–63.

71. Zhang R, Huynh A, Whitcher G, Chang J, Maltzman JS, Turka LA. An obligate cell-intrinsic function for CD28 in Tregs. J Clin Invest. 2013 Feb;123(2):580–93.

72. Coombes JL, Siddiqui KRR, Arancibia-Cárcamo CV, Hall J, Sun CM, Belkaid Y, et al. A functionally specialized population of mucosal CD103+ DCs induces Foxp3+ regulatory T cells via a TGF-beta and retinoic acid-dependent mechanism. J Exp Med. 2007 Aug 6;204(8):1757–64.

73. Tran DQ, Ramsey H, Shevach EM. Induction of FOXP3 expression in naive human CD4+FOXP3 T cells by T-cell receptor stimulation is transforming growth factor-beta dependent but does not confer a regulatory phenotype. Blood. 2007 Oct 15;110(8):2983–90.

74. Walker MR, Carson BD, Nepom GT, Ziegler SF, Buckner JH. De novo generation of antigen-specific CD4+CD25+ regulatory T cells from human CD4+CD25- cells. Proc Natl Acad Sci U S A. 2005 Mar 15;102(11):4103–8.

75. Floess S, Freyer J, Siewert C, Baron U, Olek S, Polansky J, et al. Epigenetic control of the foxp3 locus in regulatory T cells. PLoS Biol. 2007 Feb;5(2):e38.

76. Li X, Liang Y, LeBlanc M, Benner C, Zheng Y. Function of a Foxp3 cis-element in protecting regulatory T cell identity. Cell. 2014 Aug 14;158(4):734–48.

77. Miyara M, Yoshioka Y, Kitoh A, Shima T, Wing K, Niwa A, et al. Functional delineation and differentiation dynamics of human CD4+ T cells expressing the FoxP3 transcription factor. Immunity. 2009 June 19;30(6):899–911.

78. Wing K, Onishi Y, Prieto-Martin P, Yamaguchi T, Miyara M, Fehervari Z, et al. CTLA-4 control over Foxp3+ regulatory T cell function. Science. 2008 Oct 10;322(5899):271–5.

79. Jarnicki AG, Lysaght J, Todryk S, Mills KHG. Suppression of antitumor immunity by IL-10 and TGF-beta-producing T cells infiltrating the growing tumor: influence of tumor environment on the induction of CD4+ and CD8+ regulatory T cells. J Immunol Baltim Md 1950. 2006 July 15;177(2):896–904.

80. Deaglio S, Dwyer KM, Gao W, Friedman D, Usheva A, Erat A, et al. Adenosine generation catalyzed by CD39 and CD73 expressed on regulatory T cells mediates immune suppression. J Exp Med. 2007 June 11;204(6):1257–65.

81. Grossman WJ, Verbsky JW, Barchet W, Colonna M, Atkinson JP, Ley TJ. Human T regulatory cells can use the perforin pathway to cause autologous target cell death. Immunity. 2004 Oct;21(4):589–601.

82. Saito T, Nishikawa H, Wada H, Nagano Y, Sugiyama D, Atarashi K, et al. Two FOXP3(+)CD4(+) T cell subpopulations distinctly control the prognosis of colorectal cancers. Nat Med. 2016 June;22(6):679–84.

83. Takahashi T, Kuniyasu Y, Toda M, Sakaguchi N, Itoh M, Iwata M, et al. Immunologic self-tolerance maintained by CD25+CD4+ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state. Int Immunol. 1998 Dec;10(12):1969–80.

84. Collison LW, Workman CJ, Kuo TT, Boyd K, Wang Y, Vignali KM, et al. The inhibitory cytokine IL-35 contributes to regulatory T-cell function. Nature. 2007 Nov 22;450(7169):566–9.

85. Ohkura N, Hamaguchi M, Morikawa H, Sugimura K, Tanaka A, Ito Y, et al. T cell receptor stimulation-induced epigenetic changes and Foxp3 expression are independent and complementary events required for Treg cell development. Immunity. 2012 Nov 16;37(5):785–99.

86. Dykema AG, Zhang J, Cheung LS, Connor S, Zhang B, Zeng Z, et al. Lung tumor-infiltrating Treg have divergent transcriptional profiles and function linked to checkpoint blockade response. Sci Immunol. 2023 Sept 15;8(87):eadg1487.

87. Shan F, Cillo AR, Cardello C, Yuan DY, Kunning SR, Cui J, et al. Integrated BATF transcriptional network regulates suppressive intratumoral regulatory T cells. Sci Immunol. 2023 Sept 15;8(87):eadf6717.

88. Itahashi K, Irie T, Yuda J, Kumagai S, Tanegashima T, Lin YT, et al. BATF epigenetically and transcriptionally controls the activation program of regulatory T cells in human tumors. Sci Immunol. 2022 Oct 14;7(76):eabk0957.

89. Curiel TJ, Coukos G, Zou L, Alvarez X, Cheng P, Mottram P, et al. Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. Nat Med. 2004 Sept;10(9):942–9.

90. Facciabene A, Peng X, Hagemann IS, Balint K, Barchetti A, Wang LP, et al. Tumour hypoxia promotes tolerance and angiogenesis via CCL28 and T(reg) cells. Nature. 2011 July 13;475(7355):226–30.

91. Tan MCB, Goedegebuure PS, Belt BA, Flaherty B, Sankpal N, Gillanders WE, et al. Disruption of CCR5-dependent homing of regulatory T cells inhibits tumor growth in a murine model of pancreatic cancer. J Immunol Baltim Md 1950. 2009 Feb 1;182(3):1746–55.

92. Hoelzinger DB, Smith SE, Mirza N, Dominguez AL, Manrique SZ, Lustgarten J. Blockade of CCL1 inhibits T regulatory cell suppressive function enhancing tumor immunity without affecting T effector responses. J Immunol Baltim Md 1950. 2010 June 15;184(12):6833–42.

93. Spranger S, Bao R, Gajewski TF. Melanoma-intrinsic β-catenin signalling prevents anti-tumour immunity. Nature. 2015 July 9;523(7559):231–5.

94. Peng W, Chen JQ, Liu C, Malu S, Creasy C, Tetzlaff MT, et al. Loss of PTEN Promotes Resistance to T Cell-Mediated Immunotherapy. Cancer Discov. 2016 Feb;6(2):202–16.

95. Serrels A, Lund T, Serrels B, Byron A, McPherson RC, von Kriegsheim A, et al. Nuclear FAK controls chemokine transcription, Tregs, and evasion of anti-tumor immunity. Cell. 2015 Sept 24;163(1):160–73.

96. Hindley JP, Ferreira C, Jones E, Lauder SN, Ladell K, Wynn KK, et al. Analysis of the T-cell receptor repertoires of tumor-infiltrating conventional and regulatory T cells reveals no evidence for conversion in carcinogen-induced tumors. Cancer Res. 2011 Feb 1;71(3):736–46.

97. Nishikawa H, Kato T, Tawara I, Saito K, Ikeda H, Kuribayashi K, et al. Definition of target antigens for naturally occurring CD4(+) CD25(+) regulatory T cells. J Exp Med. 2005 Mar 7;201(5):681–6.

98. Hsieh CS, Liang Y, Tyznik AJ, Self SG, Liggitt D, Rudensky AY. Recognition of the peripheral self by naturally arising CD25+ CD4+ T cell receptors. Immunity. 2004 Aug;21(2):267–77.

99. Pace L, Tempez A, Arnold-Schrauf C, Lemaitre F, Bousso P, Fetler L, et al. Regulatory T cells increase the avidity of primary CD8+ T cell responses and promote memory. Science. 2012 Oct 26;338(6106):532–6.

100. Maeda Y, Nishikawa H, Sugiyama D, Ha D, Hamaguchi M, Saito T, et al. Detection of self-reactive $CD8^+$ T cells with an anergic phenotype in healthy individuals. Science. 2014 Dec 19;346(6216):1536–40.

101. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med. 2014 Dec 4;371(23):2189–99.

102. Arce Vargas F, Furness AJS, Solomon I, Joshi K, Mekkaoui L, Lesko MH, et al. Fc-Optimized Anti-CD25 Depletes Tumor-Infiltrating Regulatory T Cells and Synergizes with PD-1 Blockade to Eradicate Established Tumors. Immunity. 2017 Apr 18;46(4):577–86.

103. Galvez-Cancino F, Navarrete M, Beattie G, Puccio S, Conde-Gallastegi E, Foster K, et al. Regulatory T cell depletion promotes myeloid cell activation and glioblastoma response to anti-PD1 and tumor-targeting antibodies. Immunity. 2025 May 13;58(5):1236-1253.e8.

104. Sugiyama D, Nishikawa H, Maeda Y, Nishioka M, Tanemura A, Katayama I, et al. Anti-CCR4 mAb selectively depletes effector-type FoxP3+CD4+ regulatory T cells, evoking antitumor immune responses in humans. Proc Natl Acad Sci U S A. 2013 Oct 29;110(44):17945–50.

105. Arce Vargas F, Furness AJS, Litchfield K, Joshi K, Rosenthal R, Ghorani E, et al. Fc Effector Function Contributes to the Activity of Human Anti-CTLA-4 Antibodies. Cancer Cell. 2018 Apr 9;33(4):649-663.e4.

106. Eschweiler S, Ramírez-Suástegui C, Li Y, King E, Chudley L, Thomas J, et al. Intermittent PI3Kδ inhibition sustains anti-tumour immunity and curbs irAEs. Nature. 2022 May;605(7911):741–6.

107. Imagawa J, Tanaka H, Okada M, Nakamae H, Hino M, Murai K, et al. Discontinuation of dasatinib in patients with chronic myeloid leukaemia who have maintained deep molecular response for longer than 1 year (DADI trial): a multicentre phase 2 trial. Lancet Haematol. 2015 Dec;2(12):e528-535.

108. Beatty GL, O'Dwyer PJ, Clark J, Shi JG, Bowman KJ, Scherle PA, et al. First-in-Human Phase I Study of the Oral Inhibitor of Indoleamine 2,3-Dioxygenase-1 Epacadostat (INCB024360) in Patients with Advanced Solid Malignancies. Clin Cancer Res Off J Am Assoc Cancer Res. 2017 July 1;23(13):3269–76.

109. Mascaux C, Angelova M, Vasaturo A, Beane J, Hijazi K, Anthoine G, et al. Immune evasion before tumour invasion in early lung squamous carcinogenesis. Nature. 2019 July;571(7766):570–5.

110. Pennycuick A, Teixeira VH, AbdulJabbar K, Raza SEA, Lund T, Akarca AU, et al. Immune Surveillance in Clinical Regression of Preinvasive Squamous Cell Lung Cancer. Cancer Discov. 2020 Oct 1;10(10):1489–99.

111. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018 June;36(5):411–20.

112. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics J Integr Biol. 2012 May;16(5):284–7.

113. Zheng L, Qin S, Si W, Wang A, Xing B, Gao R, et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. Science. 2021 Dec 17;374(6574):abe6474.

114. Domínguez Conde C, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science. 2022 May 13;376(6594):eabl5197.

115. Zhang Y, Aevermann B, Gala R, Scheuermann RH. Cell type matching in single-cell RNA-sequencing data using FR-Match. Sci Rep. 2022 June 15;12(1):9996.

116. Salcher S, Sturm G, Horvath L, Untergasser G, Kuempers C, Fotakis G, et al. High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer. Cancer Cell. 2022 Dec 12;40(12):1503-1520.e8.

117. Andreatta M, Carmona SJ. UCell: Robust and scalable single-cell gene signature scoring. Comput Struct Biotechnol J. 2021;19:3796–8.

118. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021 Feb 17;12(1):1088.

119. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc. 2020 Apr;15(4):1484–506.

120. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018 June;36(5):411–20.

121. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med. 2018 July;24(7):978–85.

122. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014 Apr;32(4):381–6.

123. Dann E, Henderson NC, Teichmann SA, Morgan MD, Marioni JC. Differential abundance testing on single-cell data using k-nearest neighbor graphs. Nat Biotechnol. 2022 Feb;40(2):245–53.

124. Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. Nature. 2018 Dec;564(7735):268–72.

125. Shelton J, Zotow E, Smith L, Johnson SA, Thomson CS, Ahmad A, et al. 25 year trends in cancer incidence and mortality among adults aged 35-69 years in the UK, 1993-2018: retrospective secondary analysis. BMJ. 2024 Mar 13;384:e076962.

126. Bhamani A, Horst C, Bojang F, Quaife SL, Dickson JL, Tisi S, et al. The SUMMIT Study: Utilising a written 'Next Steps' information booklet to prepare participants for potential lung cancer screening results and follow-up. Lung Cancer Amst Neth. 2023 Feb;176:75–81.

127. Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med. 2012 Feb 23;366(8):687–96.

128. Lei J, Ploner A, Elfström KM, Wang J, Roth A, Fang F, et al. HPV Vaccination and the Risk of Invasive Cervical Cancer. N Engl J Med. 2020 Oct 1;383(14):1340–8.

129. Umar A, Dunn BK, Greenwald P. Future directions in cancer prevention. Nat Rev Cancer. 2012 Dec;12(12):835–48.

130. Pennycuick A, Teixeira VH, AbdulJabbar K, Raza SEA, Lund T, Akarca AU, et al. Immune Surveillance in Clinical Regression of Preinvasive Squamous Cell Lung Cancer. Cancer Discov. 2020 Oct 1;10(10):1489–99.

131. Alvisi G, Brummelman J, Puccio S, Mazza EM, Tomada EP, Losurdo A, et al. IRF4 instructs effector Treg differentiation and immune suppression in human cancer. J Clin Invest. 2020 June 1;130(6):3137–50.

132. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med. 2018 July;24(7):978–85.

133. Shim JH, Kim HS, Cha H, Kim S, Kim TM, Anagnostou V, et al. HLA-corrected tumor mutation burden and homologous recombination deficiency for the prediction of response to PD-(L)1 blockade in advanced non-small-cell lung cancer patients. Ann Oncol Off J Eur Soc Med Oncol. 2020 July;31(7):902–11.

134. Beumer N, Imbusch CD, Kaufmann T, Schmidleithner L, Gütter K, Stüve P, et al. DNA hypomethylation traits define human regulatory T cells in cutaneous tissue and identify their blood recirculating counterparts. Nat Immunol. 2025 Aug;26(8):1315–28.

135. Ahmadzadeh M, Pasetto A, Jia L, Deniger DC, Stevanović S, Robbins PF, et al. Tumor-infiltrating human CD4+ regulatory T cells display a distinct TCR repertoire and exhibit tumor and neoantigen reactivity. Sci Immunol. 2019 Jan 11;4(31):eaao4310.

136. Ning B, Tilston-Lunel AM, Simonetti J, Hicks-Berthet J, Matschulat A, Pfefferkorn R, et al. Convergence of YAP/TAZ, TEAD and TP63 activity is associated with bronchial premalignant severity and progression. J Exp Clin Cancer Res CR. 2023 May 8;42(1):116.

137. Wang Z, Yang L, Wang W, Zhou H, Chen J, Ma Z, et al. Comparative immunological landscape between pre- and early-stage LUAD manifested as ground-glass nodules revealed by scRNA and scTCR integrated analysis. Cell Commun Signal CCS. 2023 Nov 13;21(1):325.

138. Cai J, Wang D, Zhang G, Guo X. The Role Of PD-1/PD-L1 Axis In Treg Development And Function: Implications For Cancer Immunotherapy. OncoTargets Ther. 2019;12:8437–45.

139. DiDomenico J, Lamano JB, Oyon D, Li Y, Veliceasa D, Kaur G, et al. The immune checkpoint protein PD-L1 induces and maintains regulatory T cells in glioblastoma. Oncoimmunology. 2018;7(7):e1448329.

140. Hynds RE, Frese KK, Pearce DR, Grönroos E, Dive C, Swanton C. Progress towards non-small-cell lung cancer models that represent clinical evolutionary trajectories. Open Biol. 2021 Jan;11(1):200247.

141. Kellar A, Egan C, Morris D. Preclinical Murine Models for Lung Cancer: Clinical Trial Applications. BioMed Res Int. 2015;2015:621324.

142. Meuwissen R, Berns A. Mouse models for human lung cancer. Genes Dev. 2005 Mar 15;19(6):643–64.

143. Deng H, Ge H, Dubey C, Losmanova T, Medová M, Konstantinidou G, et al. An optimized protocol for the generation and monitoring of conditional orthotopic lung cancer in the KP mouse model using an adeno-associated virus vector compatible with biosafety level 1. Cancer Immunol Immunother CII. 2023 Dec;72(12):4457–70.

144. Drosten M, Guerra C, Barbacid M. Genetically Engineered Mouse Models of K-Ras-Driven Lung and Pancreatic Tumors: Validation of Therapeutic Targets. Cold Spring Harb Perspect Med. 2018 May 1;8(5):a031542.

145. Fitzgerald B, Connolly KA, Cui C, Fagerberg E, Mariuzza DL, Hornick NI, et al. A mouse model for the study of anti-tumor T cell responses in Kras-driven lung adenocarcinoma. Cell Rep Methods. 2021 Sept 27;1(5):100080.

146. Cui C, Wang J, Fagerberg E, Chen PM, Connolly KA, Damo M, et al. Neoantigen-driven B cell and CD4 T follicular helper cell collaboration promotes anti-tumor CD8 T cell responses. Cell. 2021 Dec 9;184(25):6101-6118.e13.

147. Neidler S, Kruspig B, Hewit K, Monteverde T, Gyuraszova K, Braun A, et al. Identification of a Clinically Relevant Signature for Early Progression in KRAS-Driven Lung Adenocarcinoma. Cancers. 2019 Apr 29;11(5):600.

148. Young NP, Jacks T. Tissue-specific p19Arf regulation dictates the response to oncogenic K-ras. Proc Natl Acad Sci U S A. 2010 June 1;107(22):10184–9.

149. Witschi H, Espiritu I, Dance ST, Miller MS. A mouse lung tumor model of tobacco smoke carcinogenesis. Toxicol Sci Off J Soc Toxicol. 2002 Aug;68(2):322–30.

150. Sahu P, Donovan C, Paudel KR, Pickles S, Chimankar V, Kim RY, et al. Pre-clinical lung squamous cell carcinoma mouse models to identify novel biomarkers and therapeutic interventions. Front Oncol. 2023;13:1260411.

151. Zakaria MA, Rajab NF, Chua EW, Selvarajah GT, Masre SF. NTCU induced pre-malignant and malignant stages of lung squamous cell carcinoma in mice model. Sci Rep. 2021 Nov 18;11(1):22500.

152. Wang Y, Zhang Z, Yan Y, Lemon WJ, LaRegina M, Morrison C, et al. A chemically induced model for squamous cell carcinoma of the lung in mice: histopathology and strain susceptibility. Cancer Res. 2004 Mar 1;64(5):1647–54.

153. Zakaria MA, Rajab NF, Chua EW, Selvarajah GT, Masre SF. NTCU induced pre-malignant and malignant stages of lung squamous cell carcinoma in mice model. Sci Rep. 2021 Nov 18;11(1):22500.

154. Gómez-López S, Alhendi ASN, Przybilla MJ, Bordeu I, Whiteman ZE, Butler T, et al. Aberrant basal cell clonal dynamics shape early lung carcinogenesis. Science. 2025 June 12;388(6752):eads9145.

155. Costantino JP, Gail MH, Pee D, Anderson S, Redmond CK, Benichou J, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst. 1999 Sept 15;91(18):1541–8.

156. Sotomayor RE, Collins TF. Mutagenicity, metabolism, and DNA interactions of urethane. Toxicol Ind Health. 1990 Jan;6(1):71–108.

157. Riolobos L, Gad EA, Treuting PM, Timms AE, Hershberg EA, Corulli LR, et al. The Effect of Mouse Strain, Sex, and Carcinogen Dose on Toxicity and the Development of

Lung Dysplasia and Squamous Cell Carcinomas in Mice. Cancer Prev Res Phila Pa. 2019 Aug;12(8):507–16.

158. Succony L, Gómez-López S, Pennycuick A, Alhendi ASN, Davies D, Clarke SE, et al. Lrig1 expression identifies airway basal cells with high proliferative capacity and restricts lung squamous cell carcinoma growth. Eur Respir J. 2022 Mar;59(3):2000816.

159. Rock JR, Onaitis MW, Rawlins EL, Lu Y, Clark CP, Xue Y, et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. Proc Natl Acad Sci U S A. 2009 Aug 4;106(31):12771–5.

160. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011 Jan;39(Database issue):D691-697.

161. Ali K, Soond DR, Pineiro R, Hagemann T, Pearce W, Lim EL, et al. Inactivation of PI(3)K p110δ breaks regulatory T-cell-mediated immune tolerance to cancer. Nature. 2014 June 19;510(7505):407–11.

162. Zhang Y, Wang D, Peng M, Tang L, Ouyang J, Xiong F, et al. Single-cell RNA sequencing in cancer research. J Exp Clin Cancer Res CR. 2021 Mar 1;40(1):81.

163. Francisco LM, Salinas VH, Brown KE, Vanguri VK, Freeman GJ, Kuchroo VK, et al. PD-L1 regulates the development, maintenance, and function of induced regulatory T cells. J Exp Med. 2009 Dec 21;206(13):3015–29.

164. Josefowicz SZ, Lu LF, Rudensky AY. Regulatory T cells: mechanisms of differentiation and function. Annu Rev Immunol. 2012;30:531–64.

165. Chen W, Jin W, Hardegen N, Lei KJ, Li L, Marinos N, et al. Conversion of peripheral CD4+CD25- naive T cells to CD4+CD25+ regulatory T cells by TGF-beta induction of transcription factor Foxp3. J Exp Med. 2003 Dec 15;198(12):1875–86.

166. Salomon B, Rhee L, Bour-Jordan H, Hsin H, Montag A, Soliven B, et al. Development of spontaneous autoimmune peripheral polyneuropathy in B7-2-deficient NOD mice. J Exp Med. 2001 Sept 3;194(5):677–84.

167. Ohta A, Sitkovsky M. The adenosinergic immunomodulatory drugs. Curr Opin Pharmacol. 2009 Aug;9(4):501–6.

168. Deaglio S, Dwyer KM, Gao W, Friedman D, Usheva A, Erat A, et al. Adenosine generation catalyzed by CD39 and CD73 expressed on regulatory T cells mediates immune suppression. J Exp Med. 2007 June 11;204(6):1257–65.

169. Selby MJ, Engelhardt JJ, Quigley M, Henning KA, Chen T, Srinivasan M, et al. Anti-CTLA-4 antibodies of IgG2a isotype enhance antitumor activity through reduction of intratumoral regulatory T cells. Cancer Immunol Res. 2013 July;1(1):32–42.

170. Sharma A, Subudhi SK, Blando J, Scutti J, Vence L, Wargo J, et al. Anti-CTLA-4 Immunotherapy Does Not Deplete FOXP3+ Regulatory T Cells (Tregs) in Human Cancers. Clin Cancer Res Off J Am Assoc Cancer Res. 2019 Feb 15;25(4):1233–8.

171. Hannani D, Vétizou M, Enot D, Rusakiewicz S, Chaput N, Klatzmann D, et al. Anticancer immunotherapy by CTLA-4 blockade: obligatory contribution of IL-2 receptors and negative prognostic impact of soluble CD25. Cell Res. 2015 Feb;25(2):208–24.

172. Whiteside SK, Grant FM, Gyori DS, Conti AG, Imianowski CJ, Kuo P, et al. CCR8 marks highly suppressive Treg cells within tumours but is dispensable for their accumulation and suppressive function. Immunology. 2021 Aug;163(4):512–20.

173. Solomon I, Amann M, Goubier A, Arce Vargas F, Zervas D, Qing C, et al. CD25-Treg-depleting antibodies preserving IL-2 signaling on effector T cells enhance effector activation and antitumor immunity. Nat Cancer. 2020 Dec;1(12):1153–66.

174. Hayes ET, Hagan CE, Khoryati L, Gavin MA, Campbell DJ. Regulatory T Cells Maintain Selective Access to IL-2 and Immune Homeostasis despite Substantially Reduced CD25 Function. J Immunol Baltim Md 1950. 2020 Nov 15;205(10):2667–78.

175. Dykema AG, Zhang J, Zhang B, Cheung LS, Zeng Z, Cherry CM, et al. Lung tumor-infiltrating T $_{reg}$ have divergent transcriptional profiles and function linked to checkpoint blockade response [Internet]. Immunology; 2022 Dec [cited 2023 Jan 9]. Available from: http://biorxiv.org/lookup/doi/10.1101/2022.12.13.520329

176. Wang Z, Yang L, Wang W, Zhou H, Chen J, Ma Z, et al. Comparative immunological landscape between pre- and early-stage LUAD manifested as ground-glass nodules revealed by scRNA and scTCR integrated analysis. Cell Commun Signal CCS. 2023 Nov 13;21(1):325.

177. Demehri S, Turkoz A, Kopan R. Epidermal Notch1 loss promotes skin tumorigenesis by impacting the stromal microenvironment. Cancer Cell. 2009 July 7;16(1):55–66.

178. Chaudhry A, Rudra D, Treuting P, Samstein RM, Liang Y, Kas A, et al. CD4+ regulatory T cells control TH17 responses in a Stat3-dependent manner. Science. 2009 Nov 13;326(5955):986–91.

179. Eschweiler S, Ramírez-Suástegui C, Li Y, King E, Chudley L, Thomas J, et al. Intermittent PI3Kδ inhibition sustains anti-tumour immunity and curbs irAEs. Nature. 2022 May;605(7911):741–6.

180. Pennycuick A, Teixeira VH, AbdulJabbar K, Raza SEA, Lund T, Akarca AU, et al. Immune Surveillance in Clinical Regression of Preinvasive Squamous Cell Lung Cancer. Cancer Discov. 2020 Oct;10(10):1489–99.

181. Francisco LM, Salinas VH, Brown KE, Vanguri VK, Freeman GJ, Kuchroo VK, et al. PD-L1 regulates the development, maintenance, and function of induced regulatory T cells. J Exp Med. 2009 Dec 21;206(13):3015–29.

182. Woods DM, Ramakrishnan R, Laino AS, Berglund A, Walton K, Betts BC, et al. Decreased Suppression and Increased Phosphorylated STAT3 in Regulatory T Cells are Associated with Benefit from Adjuvant PD-1 Blockade in Resected Metastatic Melanoma. Clin Cancer Res Off J Am Assoc Cancer Res. 2018 Dec 15;24(24):6236–47.

183. Wei T, Zhang J, Qin Y, Wu Y, Zhu L, Lu L, et al. Increased expression of immunosuppressive molecules on intratumoral and circulating regulatory T cells in non-small-cell lung cancer patients. Am J Cancer Res. 2015;5(7):2190–201.

184. Forde PM, Chaft JE, Smith KN, Anagnostou V, Cottrell TR, Hellmann MD, et al. Neoadjuvant PD-1 Blockade in Resectable Lung Cancer. N Engl J Med. 2018 May 24;378(21):1976–86.

185. Reina M, Espel E. Role of LFA-1 and ICAM-1 in Cancer. Cancers. 2017 Nov 3;9(11):153.

186. Van Kerkhove O, Verfaillie S, Maes B, Cuppens K. The Adenosinergic Pathway in Non-Small Cell Lung Cancer. Cancers. 2024 Sept 13;16(18):3142.

187. Ohta A, Sitkovsky M. Extracellular adenosine-mediated modulation of regulatory T cells. Front Immunol. 2014;5:304.

188. Wu C, Thalhamer T, Franca RF, Xiao S, Wang C, Hotta C, et al. Galectin-9-CD44 interaction enhances stability and function of adaptive regulatory T cells. Immunity. 2014 Aug 21;41(2):270–82.