UNIVERSITY OF LONDON

Spatial Interrogation of Tumour Microenvironment using Artificial Intelligence

Yeman Brhane Hagos

Supervisor: Prof. Yinyin Yuan

Department of Molecular Pathology
The Institute of Cancer Research
University of London

A thesis submitted for the degree of
Doctor of Philosophy in Computational Pathology

February 2023

ICR The Institute of Cancer Research

I would like to dedicate this thesis to my loving parents ...

Declaration

I hereby declare that this submission is my own work and it does not contain another person's work except where specific reference is made. This thesis contain materials taken only from my first author publications listed bellow:

1. Hagos, Yeman Brhane, Catherine SY Lecat, Dominic Patel, Anna Mikolajczak, Simon P Castillo, Kane Foster, Thien-An Tran et al. "The mosaic microenvironment of myeloma bone marrow trephine biopsies mapped by deep learning." Cancer Research. In revision. February 2023.

2. Hagos, Yeman Brhane, Faranak Sobhani, Simon P. Castillo, Allison H. Hall, Khalid AbdulJabbar, Roberto Salgado, Bryan Harmon et al. "DCIS AI-TIL: Ductal Carcinoma In Situ Tumour Infiltrating Lymphocyte Scoring Using Artificial Intelligence." In Artificial Intelligence over Infrared Images for Medical Applications and Medical Image Assisted Biomarker Discovery: First MICCAI Workshop, AIIIMA 2022, and First MICCAI Workshop, MIABID 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18 and 22, 2022, Proceedings, pp. 164-175. Cham: Springer Nature Switzerland, 2022.

3. Hagos, Yeman Brhane, Ayse U. Akarca, Alan Ramsay, Riccardo L. Rossi, Sabine Pomplun, Alessia Moioli, Andrea Gianatti et al. "DeepMIF: Deep Learning Based Cell Profiling for Multispectral Immunofluorescence Images with Graphical User Interface." In Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV, pp. 140-149. Cham: Springer Nature Switzerland, 2022.

4. Hagos, Yeman Brhane, Ayse U. Akarca, Alan Ramsay, Riccardo L. Rossi, Sabine Pomplun, Victoria Ngai, Alessia Moioli et al. "High inter-follicular spatial co-localization of CD8+ FOXP3+ with CD4+ CD8+ cells predicts favorable outcome in follicular lymphoma." Hematological Oncology 40, no. 4 (2022): 541-553.

5. Hagos, Yeman Brhane, Catherine SY Lecat, Dominic Patel, Lydia Lee, Thien-An Tran, Manuel Rodriguez-Justo, Kwee Yong, and Yinyin Yuan. "Cell abundance aware deep learning for cell detection on highly imbalanced pathological data." In 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 1438-1442. IEEE, 2021.

6. Hagos, Yeman Brhane, Priya Lakshmi Narayanan, Ayse U. Akarca, Teresa Marafioti, and Yinyin Yuan. "ConCORDe-Net: cell count regularized convolutional neural network for cell detection in multiplex immunohistochemistry images." In Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22, pp. 667-675. Springer International Publishing, 2019.

Yeman Brhane Hagos
February 2023

Acknowledgements

Firstly, I would like to express my utmost gratitude to my supervisor, Prof Yinyin Yuan, for the opportunity to pursue my PhD study. Thank you for your trust, patience, and guidance throughout my journey. I had the opportunity to learn so much from your kind, sincere, and generous advice and support, which helped me to accomplish this milestone. Thank you very much!

I am deeply grateful to my thesis examiners Dr Constantino Carlos Reyes Aldasoro and Prof. Fayyaz Minhas for accepting to examine my thesis and for providing me with constructive feedback on my thesis.

I am very thankful to all my colleagues: Khalid Abdul Jabbar, Simon P. Castillo, Faranak Sobhani, Hanyun Zhang, Xiaoxi Pan, and Anca Grapa. I feel fortunate to have had the opportunity to work with such unique and talented individuals who will continue to inspire me.

I am also incredibly thankful to my collaborators: Prof. Teresa Marafioti, Prof. Kwee Yong, Ayse U Akarca, Catherine SY Lecat, Prof. E. Shelley Hwang, Dr Roberto Salgado, Dr Hugo M. Horlings and others. You have been incredibly supportive, and I am grateful to have had the opportunity to collaborate with you.

I would like to thank also the European Union and the Institute of Cancer Research for funding my PhD. For the first three years of my PhD, I received a prestigious and highly competitive scholarship from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 766030. I was part of the Computational ONcology TRaining Alliance Innovative Training Network (CONTRA), an alliance of multiple universities across Europe. This program gave me a great opportunity to meet talented like-minded professionals, share research ideas and present my work in workshops organised twice a year. I would like to thank also the ESRs and PIs in the CONTRA consortium for making the program diverse and enjoyable, it has helped me to grasp the different horizons of oncology. The ICR funded my fourth year of PhD. I would like to say thank you for this opportunity.

I also give special thanks to my Tigrayan community in London: Miliete Brhanu, Samsom Kahsay, Aklil Zenebe, Fisseha Zeray, Abraham Mebrahtu, Solomon Gebrehiwet and the

Tigrayan community in London. Thank you for your support and for being like a family to me when I was not able to talk with my family in Tigray for about two years due to the Tigray war.

Finally, and most importantly, I would like to thank my parents and siblings for always believing in me. Dad and mum, you have taught me love, hard work and most importantly to have a purpose in life. You have always been motivational and caring for my future more than you care for yourself. I always wanted to make you proud and I hope I did it.

Abstract

The tumour microenvironment can provide crucial information for disease diagnosis, treatment planning, and prognosis. However, the complexity of its morphological, cellular, and spatial architecture hinders accurate evaluation and quantification. Deep mining of its content using knowledge-driven artificial intelligence methods can significantly benefit clinicians and patients by uncovering new disease biology and generating objective assessments in the decision-making process.

In this PhD thesis, we developed deep learning based image analysis pipelines to spatially interrogate the role of the tumour microenvironment in various cancer types, including follicular lymphoma, multiple myeloma, and ductal carcinoma *in situ*, using multispectral immunofluorescence (MIF), multiplex immunohistochemistry (MIHC), and hematoxylin and eosin (H&E) tissue staining technologies.

Firstly, we developed new deep learning-based pipelines to detect and classify single cells and segment different tissue compartments on MIF and MIHC images. The deep learning models were trained and validated using expert pathologists' annotations. Secondly, we developed tissue morphology and single-cell spatial analysis methods tailored to the tissue structures' complexity to identify spatially resolved phenotypes and spatial topography of cells to predict disease prognosis. We showed the significance of the architectural distribution of tumour infiltrating lymphocytes (TILs) on the prediction of disease outcome. Finally, we implemented an automated TILs scoring pipeline from H&E images that account for ductal carcinoma *in situ* spatial infiltration pattern and mimic pathologists' TILs scoring procedure. The spatial scores were associated with patient response to treatment and risk of recurrence.

In conclusion, we built new deep learning based image analysis pipelines that dissect tissue structures and spatially map cell phenotypes in histopathology images and identified novel spatial prognostic features in multiple cancer types. Once validated, these methods could be utilised in clinics as decision support for the diagnosis and prognosis of cancer patients for precision medicine.

Table of contents

List of figures

List of tables

List of abbreviations

AI	Artificial Intelligence.
ASCT	Autologous Stem Cell Transplant.
AUC	Area Under The Curve.
BM	Bone Marrow.
CI	Confidence Interval.
CNN	Convolutional Neural Networks.
ConCORDe-Net	Cell COunt RegularizeD Convolutional Neural Network.
DCIS	Ductal Carcinoma *In situ*.
DCISr	Ductal Carcinoma *In situ* Recurrence.
DL	Deep Learning.
ER	Estrogen Receptor.
FL	Follicular Lymphoma.
FLIPI	Follicular Lymphoma Interntional Prognostic Index.
GAN	Generative Adversarial Network.
GUI	Graphical User Interface.
H&E	Hematoxylin And Eosin.
HER2	Human Epidermal Growth Factor Receptor 2.
HR	Hazard Ratio.
IBCr	Invasive Breast Carcinoma Recurrence.
MGUS	Monoclonal Gammopathy Of Unknown Significance.
MIF	Multispectral Immunofluorescence.
MIHC	Multiplex Immunohistochemistry.
MM	Multiple Myeloma.
NDMM	Newly Diagnosed Multiple Myeloma.
no-IBE	No Ipsilateral Breast Event.
PD1	Programmed Cell Death Receptor 1.

PDL1	Programmed Cell Death Receptor Ligand 1.
PR	Progesterone Receptor.
ReLU	Rectified Linear Unit.
RFS	Relapse Free Survival.
ROC	Receiver Operating Characteristic Curve.
SMM	Smouldering Multiple Myeloma.
TIL-WG	International Immuno-Oncology Biomarker Working Group.
TILs	Tumour Infiltrating Lymphocytes.
TME	Tumour Microenvironment.
TNBC	Triple-negative Invasive Breast Cancer.
Treg	T Regulatory.
Umap	Manifold Approximation And Projection.
WSI	Whole Slide Image.

Chapter 1

Introduction

1.1 Tumour and tumour microenvironment

Cancer is a disease characterised by unregulated cell growth with the capability of spreading throughout the body [1, 2]. Under normal conditions, cells divide to create new cells, and aged cells die in a programmed way [1]. When this control is lost, aged cells continue to live and cell divide without control and this forms a tumour (Figure 1.1A) [1, 2].

Although the tumour is one of the earliest documented diseases in history, we still wonder about its initiation, development, and invasion. The name "cancer" was coined by Hippocrates, a Greek physician, around 400 BC [3, 4]. However, the earliest description of human cancer could be traced back to 3000 BC, as found in the Edwin Smith Papyrus documenting the first human breast cancer case [3]. These ancient Egyptian reports described cancer as a deadly, incurable sickness, and they believed it to be "the curse of the gods" [3, 4].

Cancer was long thought to be a disease composed solely of abnormal cells with autonomous proliferative, and survival abilities [4, 2]. Thus, cancer treatment has been limited to targeting cancer cells. For example, according to DiLonardo et al. [4], ancient Egyptians used heated substances to burn cancerous tissue. However, tumours develop within a body surrounded by a complex and heterogeneous multicellular environment influencing their survival. For example, abnormal cells could be created due to uncontrolled genetic changes. However, can these cells multiply, grow, and become a tumour without interaction with the surrounding environment? The surrounding environment is highly likely to influence the tumours, either supporting or opposing their growth and survival. This surrounding environment is called the "tumour microenvironment". Advances in cancer biology have revealed that the tumour microenvironment (TME) plays an equal, if not greater, role in cancer cell initiation, progression, and survival [5]. According to DiLonardo et al. [4], the

importance of TME has triggered a shift in how cancer biology and treatment are perceived; instead of considering cancer cells fully autonomous cells and a cancer-centred treatment approach, cancer has begun to be viewed as a disease evolving in complex multicellular tissue, determining its fate.

The TME is a rich resource, everything except the cancer cells. As shown in Figure 1.1B, it includes fibroblasts, blood vessels, stromal cells, immune cells, and antigen-presenting cells such as macrophages and dendritic cells, among others [9, 8, 10]. While some of these cell types promote the growth and survival of cancer cells, others suppress them. Moreover, cancer cells could hijack the control system of the host immune system, and even cells thought to help the body will be recruited in favour of tumour invasion [5, 8, 10]. This helps the tumour to alter the microenvironment to promote its growth, for instance, by forming blood vessels towards the tumour mass [11]. In contrast, immune cells could suppress tumour proliferation or be recruited by tumour cells to facilitate immune evasion [5, 10, 12].

The interaction between tumour and TME was first coined by Stephen Paget in his "seed and soil theory", seeds as tumour cells and soil as the microenvironment [13]. However, Paget's work did not get much attention until mid-20^{th} century. In the 1960s, cancer researchers started to study immunology in relation to tumour development [13]. This paved the way for the development of innovative cancer treatment strategies such as immunotherapy [8, 5].

It also led to the creation of advanced technologies to decipher the molecular, cellular, and tissue architectural features of tumours and the microenvironment. For instance, molecular profiling technologies like whole genome sequencing and ribonucleic acid sequencing have helped us to learn more about cancer biology. These technologies provide a high-dimensional profile of cancer and normal cells, revealing the underlying genetic modifications and functional changes during cancer progression. While these technologies initially did not preserve the tissue context, recent molecular profiling technologies such spatial transcriptomics preserve the spatial context [14].

Moreover, spatial histology tissue staining technologies such as multiplex immunohistochemistry (MIHC) and multispectral immunofluorescence (MIF) allow the examination of cancer cells and TME at a single cell level while preserving spatial tissue context. The MIHC and MIF technologies use antibodies and proteins to stain millions of cells within a tissue section while preserving the tissue's spatial context [15]. These technologies allow investigation of cellular composition, cells spatial organisation, and tissue morphology in normal vs tumour tissue, patients with the same cancer type but different clinical outcomes, and patients with different cancer types. However, these technologies generate vast amounts of data, making analysing them a bottleneck. Nevertheless, these limitations could be ad-

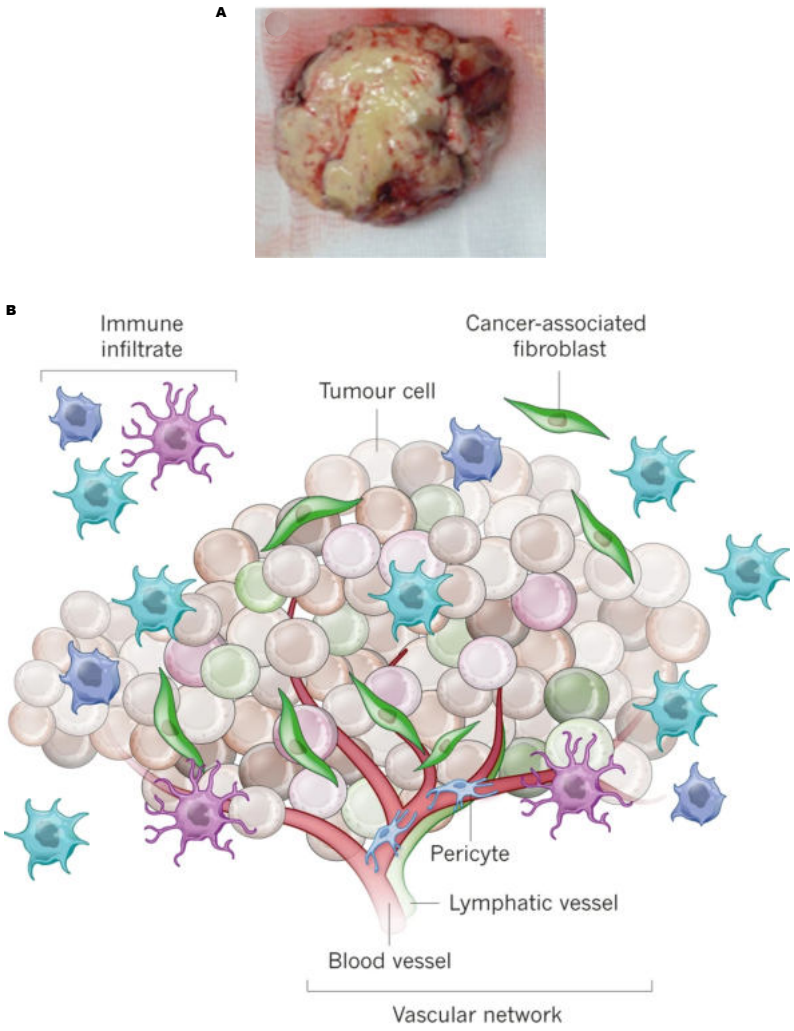


Figure 1.1: **Tumour microenvironment. A)** Tumour block removed from a patient with glioblastoma. Used with permission of Royal Society of Chemistry, from Ciasca et al. [6]; permission conveyed through Copyright Clearance Center, Inc. The tumour block contains cancer cells and non-cancerous cells such immune cells. **B)** A cartoon showing cancer cells initiate, grow and survive surrounded by various cell types and tissues in the TME such as blood vessels, immune cells and cancer-associated fibroblast. These cells could promote or impede the growth of the tumour and they possess an elastic behaviour. Pericytes are cell types important for blood vessel formation and control of blood flow [7]. The main function of blood vessels is to transport nutrients to tissues and waste materials outside the tissue or organ. The lymphatic system transports white blood cells and fluid molecules to maintain the cells. Tumour cells promote the growth of lymphatic and blood vessels in their area to maintain their survival and fast growth. The image is taken from Junttila et al. [8] with permission from the Springer Nature.

dressed using advanced computing resources and deep learning-powered image analysis tools [16–18].

The focus of this thesis is to study the tissue microenvironment and spatial immune landscape of multiple cancer types including follicular lymphoma (FL), multiple myeloma (MM) and ductal carcinoma *in situ* (DCIS) using spatial histopathology staining technologies and deep learning based computational methods. In the following sections, we will have a brief look at the TME and clinical management of FL, MM and DCIS, and applications of deep learning in digital pathology highlighting the gaps in the literature which will be addressed in this thesis.

1.2 Follicular lymphoma

1.2.1 What is follicular lymphoma?

Lymphoma is a type of blood cancer that affects white blood called lymphocytes [19]. Lymphomas are categorised into Hodgkin lymphomas or non-Hodgkin lymphomas based on cell content. Hodgkin lymphomas are characterised by the presence of large, multinucleated lymphocytes, also known as Reed-Sternberg cells [19, 20]. Lymphocytes are grouped into T and B cells. Follicular lymphoma is a slow-growing (indolent) non-Hodgkin lymphoma that affects B cells [19, 21].

The B cells are created in the bone marrow (BM) from lymphoid progenitors. Towards their maturity, these B cells start to develop B cell receptors and they leave the BM and travel through the bloodstream to other parts of the body such as lymph nodes and liver [21, 23] (Figure 1.2A). In the lymph node, B cells reside and grow in a specific region called the germinal centre, and their primary function is to generate antibodies that help our body fight infections [23]. When these cells accumulate abnormal genetic alterations such as t(14;18) translocation [21], they become neoplastic and start to multiply out of control and invade the germinal centre. During the initial stage of FL, neoplastic cells penetrate existing lymphoid follicles, interact with the germinal centre microenvironment, and emit signals that promote survival and proliferation of B cells [21]. In the case of FL, the cancer cells develop in a clustered or non-diffused way and form follicular or nodular structure (Figure 1.2B), and hence the disease is named follicular lymphoma [21]. During the neoplastic progression, neoplastic follicles destroy the natural lymph node tissue architecture and invade nearby adipose tissues [21].

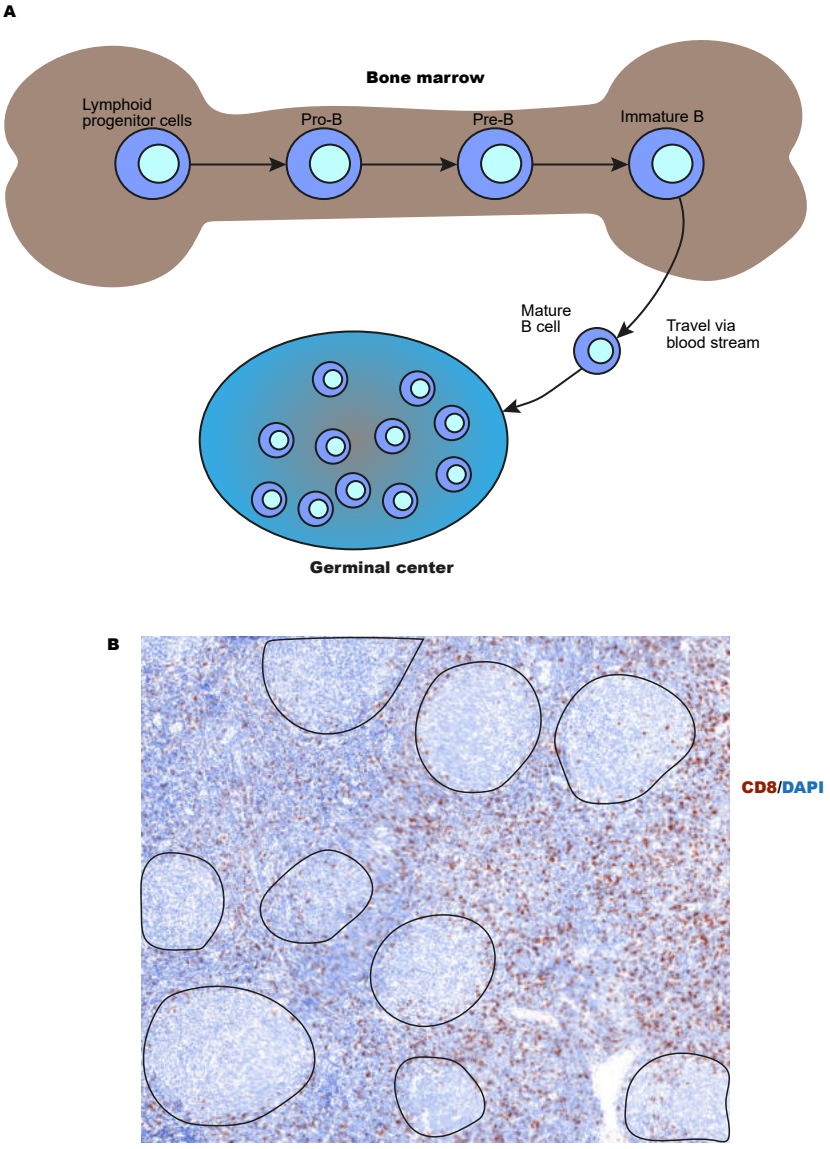


Figure 1.2: **B cells maturation and follicle histology image. A)** A cartoon showing B cell development from bone marrow to germinal centre. B cells develop in the bone marrow from lymphoid progenitor cells. The stages of development include Pro-B, Pre-B, immature B and mature B cells [22]. Mature B cells leave the bone marrow and travel to other organs such as lymph nodes and reside in the germinal centre. **B)** Histology image showing annotation of neoplastic follicles on FL tissue sample of lymph node. The black colour annotations show the neoplastic follicles. DAPI (4', 6-diamidino-2-phenylindole) is a nuclear stain.

1.2.2 Diagnosis and treatment follicular lymphoma

The diagnosis rate of FL is higher in developed counties than in developing countries, in older than young people, and in men than women [21]. Follicular lymphoma accounts for 5% of all haematological neoplasms and 20%–25% of non-Hodgkin lymphomas in western countries [21].

The diagnosis of FL involves a combination of lymph node biopsy for morphological analysis, blood test, and imaging test [21, 24]. Expert pathologists or haematologists analyse these samples for prognostic features and genetic changes that help the disease staging, grading and deciding treatment [25]. A blood test is used to count the proportion of clonal B cells in the blood. In contrast, non-invasive imaging technologies such as computed tomography are used for staging (I-IV) to identify areas affected by lymphoma and to guide biopsy sampling [19, 21, 25]. Stage III and IV are considered advanced stages [19]. Moreover, FL grading involves examining the presence of large lymphocytes using tissue biopsy of the lymph node. The FL grading system includes 1, 2, 3A or 3B, in increasing order of the number of large lymphocytes. Grade 1, 2, and 3A are slow-growing FL, while grade 3B is classified as fast-growing and treated in the same way as high-grade non-Hodgkin lymphoma [19].

After the diagnosis, pathologists generate a prognostic score to estimate the best-suited treatment. In the UK, Follicular Lymphoma Interntional Prognostic Index (FLIPI) is one of the most commonly used prognostic scores. This score is computed based on clinical information including age, haemoglobin levels, number of involved nodal areas, stage and lactate dehydrogenase levels [19].

Treatment of FL depends on the stage and grade of the disease. For grade 1-3A (slow growing FL), treatment is not always suggested [19]. For early stage FL, radiation therapy of the affected area is recommended [19, 26]. For advanced stage FL, chemo-immunotherapy treatment is applied [19]. The most commonly used chemo-immunotherapy treatment regimes include bendamustine, CVP (cyclophosphamide, vincristine and prednisolone), CHOP (cyclophosphamide, hydroxydaunorubicin, Oncovin, prednisolone) and rituximab [19, 27]. The advent of rituximab treatment, which targets the B cell marker CD20, has significantly improved patients' overall survival [28, 21]. Thanks to the development of this drug about 50% of FL patients survive $>$ 10 years [21, 29].

Montoto et al. [30] evaluated the prognostic value of FLIPI in a cohort of study containing 103 patients with FL. The patients received different types of treatments including monotherapy with alkylating agents, CHOP chemotherapy, and CVP chemotherapy. They found that the FLIPI score of the patients is associated with the survival of patients after the progression of the disease. Another study by Nooka et al. [31] examined the prognostic values

of FLIPI in patients who received rituximab and non-rituximab treatment regimens. The FLIPI risk groups showed an association with the patient's overall survival and recurrence free survival for both treatment regimens. However, this index fails to predict the risk of relapse in some patients [32].

However, FL remains incurable cancer. Particularly, about 20% of patients progress or relapse in the first two years of treatment [21]. Thus, identifying these groups of patients at diagnosis is crucial so that alternatives to the current treatment standard can be administered [21]. It is also important to note that many individuals die from therapy-related toxicity or secondary cancers [33]. Therefore, our priority should be on managing high-risk patients, employing existing and novel medicines properly, and reducing therapy for low-risk patients [21].

1.2.3 Follicular lymphoma microenvironment

Follicular lymphoma differs from other types of lymphoma by forming a nodular structure, perhaps due to unique neoplastic B cells and TME interaction [21]. The FL microenvironment comprises lymphoid cells, stromal cells, and extracellular matrix components in addition to malignant cells [21]. The TME and neoplastic B cells engage in reciprocal signalling crosstalk using cytokines and chemokines, neoplastic B cells thriving for their survival and proliferation by recruiting cells that are supposed to fight cancer [34].

Despite the slow progression of FL, a significant proportion of patients experience relapse or transform into high-grade lymphoma, which is associated with a poor prognosis [35, 36]. The disease has an indolent remitting and relapsing course, but there is a lot of individual variation [37, 38]. While most patients respond to various chemotherapy regimens, some develop de novo resistance. Some patients achieve remission, but relapse early and have a poor prognosis. On the other hand, some patients experience remission that lasts for many years and can be life-long [21, 39]. Over the years, several attempts have been made to decipher biologic and genetic alterations [40–42], and morphological [43–45] characteristics of FL that predict disease prognosis.

Dave et al. [40] from the Staudt laboratory analysed the gene expression profile of samples from patients with FL to identify immune response (IR) signatures associated with patient survival. They identified two immune signatures, IR1 and IR2, which predicted long and short survival, respectively, in follicular lymphoma. The IR1 gene expression profiling signature contained genes that encoded both T-cell and macrophage markers, whereas the IR2 signature contained genes that were preferentially expressed in macrophages, dendritic cells, or both. Cell sorting experiments revealed that these signatures were produced by infiltrating immune cells rather than malignant lymphoma cells. Furthermore, subsequent

gene expression profiling-based studies suggested the potential importance of immune surveillance in this disease, raising the prospect of novel immune approaches [40, 46].

Moreover, to define which immune cells influence FL prognosis, either by supporting tumour cell growth or by causing tumour cell death, several qualitative and quantitative immunohistochemistry or immunofluorescence studies have been performed. Initial studies assessed single markers such as CD3 or CD68 found that high levels of T cells were generally associated with a good prognosis [47–49], whereas macrophage infiltrates were usually associated with a poor prognosis [48, 49]. However, Taskinen et al. [50] found that macrophage infiltrates were associated with favourable outcomes and no association with outcome was observed in another study by Canioni et al. [51].

Recently, multiplex staining has enabled the identification of multiple immune cells with high accuracy [45, 52, 53]. The T cells associated with a good prognosis in FL predominantly have a cytotoxic CD8+ phenotype [47], and CD4+ cells with a follicular regulatory phenotype also appear to be associated with a favourable prognosis [45, 54, 55]. On the other hand, the extent of infiltration by CD4+ cells with a follicular helper phenotype does not appear to be prognostic of patient outcome [48]. Multiple studies have suggested that the CD68+ M1 macrophages are associated with a poor prognosis [48], but this has not been found in all situations [50]. Similarly, the presence of large numbers of CD163+, CD68- M2 macrophages has been reported to be associated with opposing prognostic impact in different patient cohorts [56].

There are multiple reasons for the frequent discrepancies between studies. Firstly the patient characteristics in the different cohorts may vary, and the different treatments administered may impact prognostic factors for patient outcomes. In one study, for instance, the density of CD68+ cells in the inter-follicular areas was associated with a poor prognosis in the cohort of patients treated with fludarabine but a good prognosis in those treated with cyclophosphamide vincristine and prednisone [27]. The study reported by Kridel et al. [56] found that the presence of CD68- M2 macrophages was associated with a poor prognosis in patients from the British Columbia Cancer Agency treated with rituximab and a non-anthracycline regimen. However, the presence of CD68- M2 macrophages was associated with a favourable prognosis in those patients treated with rituximab and an anthracycline-containing regimen such as R-CHOP (R stands for rituximab) in the PRIMA trial [56].

Technical issues may also account for the different observations concerning prognostic biomarkers. There may be a lack of rigorous intra- and inter-laboratory standardisation for both staining specific antigens and expert histopathologists' interpretation of staining patterns. The Lunenberg Lymphoma Biomarker Consortium accessed the reproducibility of manual scoring of immune cell markers from histology images [57]. They unexpectedly found that

there is high variability among pathologists in the scoring of nearly all cell markers in a study of diffuse large B-cell lymphoma [57]. The same group also evaluated the reliability of the immunohistochemical analysis of the TME in FL and again found significant discordance among expert laboratories [57], when small sub-populations in an immune infiltrate are assessed. These problems are further compounded by the fact that the number of cells considered can be statistically insufficient unless large areas of a biopsy are evaluated, which can be very time-consuming.

Another potential difficulty in comparing studies is that some examine the cellular composition of entire slides, whereas others concentrate on a specific region such as a neoplastic region only. In FL, the intra-follicular areas containing neoplastic cells are morphologically distinct from the inter-follicular areas [44]. It has been shown that the non-neoplastic immune infiltrates are quantitatively and qualitatively different between these areas [27, 44, 58, 59] and that the nature of the infiltrate in both sites may be predictive of outcome.

1.2.4 The spatial ecosystem of follicular lymphoma

High-throughput multiplex imaging technologies such as MIF and MIHC that provide unprecedented spatial resolution are revolutionising histopathology and spatial biology. These technologies enable capturing multiple proteins that show the function of cells within a tissue section while preserving the tissue's spatial context. In addition to the spatial cell distribution, tissue structures such as neoplastic follicles could be identified either manually by expert pathologists or automated image analysis methods as shown in Figure 1.2B. In FL, once the neoplastic follicles are identified and cells are spatially mapped within the tissue section, the spatial organisation cells within the neoplastic follicles, outside the neoplastic follicles or across the entire tissue section could be analysed (Figure 1.3). As the number of markers increase, morphological pattern quantification by human become more challenging [45]. Thus, hypothesised or exploratory automated image analysis methods are crucial to identifying spatial patterns or biomarkers of tumour-infiltrating cells. Previous works on the spatially resolved analysis of FL TME will be discussed below.

A cell could have a different function depending on the microenvironment [62]. In colorectal cancer, Schurch et al. [62] showed that a patient with a separated tumour and immune compartments showed higher survival compared with patients who have an overlapping tumour and immune compartments, which could be due to the regulation of the anti-tumour function of immune cells by the tumour. Thus, in the case of FL, analysing the immune infiltration pattern of the intra-follicular (a region which mainly contains neoplastic cells) and inter-follicular regions separately could provide more precise insight into the prognostic value

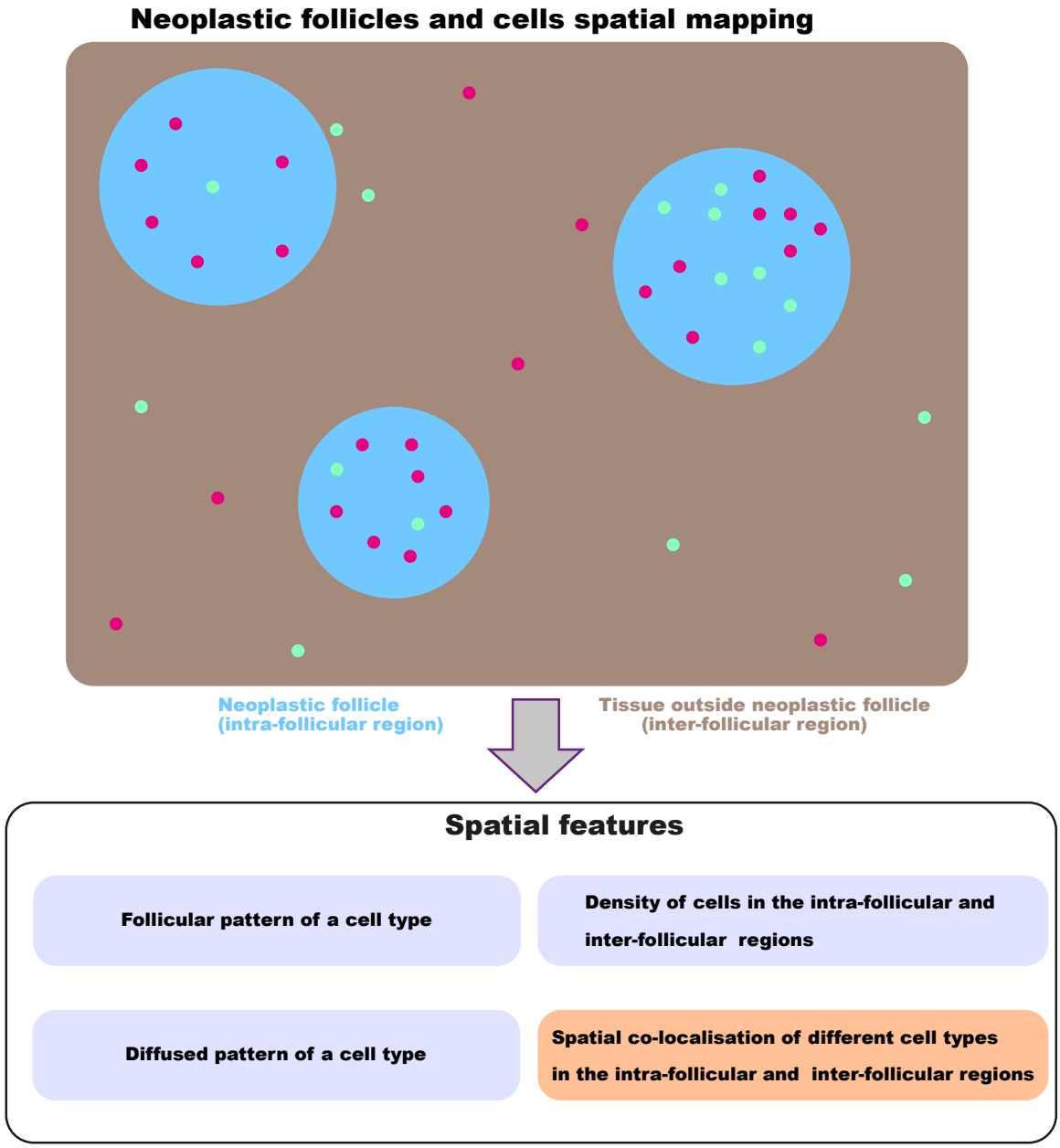


Figure 1.3: **A cartoon showing the analysis of the spatial immune landscape of follicular lymphoma:** A cartoon showing identification of neoplastic follicles and cell types, and spatial features that could be measured in a FL tissue section. The coloured dots represent cells and one colour represents one cell type. Segmentation of the follicles enables analysis of the spatial organisation of cells in the intra-follicular and inter-follicular regions. The follicular pattern of cells indicates a distribution of cells mainly in the intra- and per-follicular region of follicles, while diffused pattern indicates a distribution of cells in the intra-follicular and inter-follicular regions [44]. While the prognostic value of follicular and diffused pattern of cells and density of cells have been studied [27, 43, 44, 47], the prognostic value of spatial co-localisation of different cell types was not studied in the context of FL tissue compartments.

Table 1.1: **Prognostic value of infiltration patterns of immune cells in follicular lymphoma.** Good and poor represent association with good and poor prognosis, respectively. CHOP: doxorubicin hydrochloride (hydroxydaunorubicin), vincristine sulfate (oncovin), and prednisone; CVP: cyclophosphamide, vincristine sulfate, and prednisone; IAF: intrafollicular; IEF: inter-follicular; na:no association; PFS: progression free survival; OS: overall survival; Treg: T regulatory cells; *: prognosis depends on the treatment administered; –: analysis was not done. The R in R-CHOP and R-CVP stands for Rituximab. The endpoint is the clinical variable used to evaluate the prognostic value.

Pattern	[27]	[35]	[43]	[44]	[45]	[47]	[54]	[60]	[61]
CD4+ dense	na	–	na	na	–	–	Good	na	Good
CD4+ sparse	na	–	na	na	–	–	Poor	na	Poor
CD4+ IEF	na	na	na	–	–	–	na	Good	Good
CD4+ IAF	–	Good	–	–	–	–	–	–	–
CD8+ dense	na	–	–	na	–	Good	na	na	Good
CD8+ sparse	na	–	–	na	–	Poor	na	na	Poor
CD8+ IEF	na	na	–	–	–	na	na	na	–
CD8+ IAF	–	na	–	–	–	–	–	–	–
Treg dense	*	–	Good	Poor	Poor	–	Good	na	Good
Treg sparse	na	–	Poor	Good	Good	–	Poor	na	Poor
Treg IEF	Good	na	Good	–	–	–	Good	Good	Good
Treg IAF	na	na	Good	–	–	–	–	na	–
CD68+ dense	*	–	–	Poor	–	–	na	na	Good
CD68+ sparse	na	–	–	Good	–	–	na	na	Poor
CD68+ IEF	*	na	–	–	–	–	na	–	–

Information about the data used by the studies

Paper	**Endpoint**	**Treatment type**
[27]	PFS	CVP or fludarabine
[35]	PFS	Various treatment regimens
[43]	OS	Various treatment regimens, mostly CHOP
[44]	PFS and OS	Multiagent chemotherapy and radiation
[45]	OS	Rituximab plus other regimens
[47]	OS	Various treatment regimens
[54]	OS	Various treatment regimens
[60]	Transformation	Various treatment regimens, mostly CVP
[61]	OS and PFS	CVP, CHOP, R-CVP or R-CHOP

of immune cells in FL. Table 1.1 summarises the prognostic value of the spatial distribution of immune cells from previous studies.

To dissect the inter- and intra-follicular region of FL, manual annotation or automated methods have been adopted. Samsi et al. [63] developed a watershed based image analysis algorithm to segment follicles on CD10 antibody stained MIHC images. Colour and texture features were employed to locate possible follicular locations achieving around 87% overlap with expert annotation [63]. Senaras et al. [64] proposed U-Net [65] based deep learning algorithm to delineate follicles on CD8 antibody stained immunohistochemical whole slide images, achieving dice similarity coefficient around 86%. However, since no marker stains the follicles accurately, most previous studies relied on the manual evaluation of infiltration patterns in the inter- and intra-follicular regions of FL [27, 43, 44, 47, 54, 60, 61].

Daphne de Jong et al. [27] using immunohistochemical staining of immune T cells markers (CD4, CD8, CD69 and FOXP3), and macrophages marker (CD68) and evaluated the prognostic value of infiltration pattern of different cell types. The infiltration patterns were scored manually by experts. They showed that high inter-follicular infiltrate of FOXP3+ T regulatory (Treg) cells was associated with improved clinical outcomes. However, the prognostic value of dense Treg cells and CD68+ cells and the inter-follicular pattern of CD68+ cells was dependent on the type of treatment administered. As shown in Table 1.1, other studies have also investigated the prognostic impact of the architectural pattern of immune T cells and macrophages.

Some studies showed that spare Treg cells are associated with favourable prognosis [44, 45], while others show poor prognosis [54, 43]. Farinha et al. [44] utilised immunohistochemical staining of immune cells and investigated the association between follicular patterns (follicular and peri-follicular infiltration) or diffused pattern immune cells with patients' overall survival. The binary classification of the follicular or diffused pattern was done manually, and diffused pattern of FOXP3+ Treg cells was associated with more prolonged overall survival [44]. Nelson et al. performed a triplex MIHC staining of CD3, FOXP3 and CD69 to the association between infiltration pattern of immune cells and overall survival [45]. Similar to [44], FOXP3+ Treg cells diffused infiltration pattern measured by hypothesised interaction distribution [66] was associated with favourable overall survival [45]. However, other multiple studies found diffused infiltration was associated with poor prognosis [43, 54, 61]. Moreover, Mondello et al. employed [35] MIF panel containing ten immune-related markers and studied the association between the abundance of immune cells in the intra- and inter-follicular regions and early relapse. They found that reduced intra-follicular CD4+ T cell infiltration was associated with early relapse in FL patients [35].

The discordance in the prognostic value of the cell types could be due to the varying precision of T-cell subset identification, treatment administered and endpoint clinical variable. Moreover, these studies were conducted on a different cohort of patients with different characteristics [57].

1.2.5 Limitations of previous studies

As discussed above, while some studies investigated the prognostic value of spatial infiltration patterns of various immune cells in the inter- and intra-follicular regions of FL, these studies have some limitations:

- Some of the previous studies employed a tissue microarray with a small number of cores. For example, the studies by Farinha et al. [44] and Mondello et al. [35] used tissue microarrays with only two cores and five cores per patient, respectively. The use of a small number of cores could introduce bias and shows a small spectrum of the spatial heterogeneity of FL microenvironment.

- The previous studies contained a limited number of immune cell markers in their assay, which shows again a limited spectrum of the cell content of FL, although these studies could have been interested in specific cell phenotypes only.

- With exception of a recent study by Mondello et al. [35], quantification of the spatial distribution patterns was performed manually by expert pathologist(s) from MIHC stained images [27, 44, 54]. While this could be achievable in studies employing tissue microarray and MIHC, it is challenging on the whole slide tissue section and highly multiplex images such MIF containing a large number of markers. It has been also shown that manual immune scoring lacks consistency even among experienced pathologists [57] and automated image analysis could generate objective quantification.

- The studies presented in Table 1.1 analysed the spatial infiltration pattern only at the single cell type level. In solid tumours such as oestrogen receptor-positive breast cancer, it has been demonstrated that the spatial relationship of immune cells and tumour cells was prognostic rather than the density of immune cells [67]. Thus, analysing the relative spatial co-localisation of multiple cells using multiplex staining technologies in the intra-follicular and inter-follicular regions of FL might provide new biological insight into the interaction between the neoplastic and immune cells in the TME.

In this study, to investigate the spatial immune landscape of FL microenvironment, we employed multiple MIF staining panels to capture a wide spectrum of immune cells on tissue

sections including immune T cells, myeloid cells, natural killer T cells and macrophages. We developed deep learning based automated image analysis method to spatially localise cell types on multiplex images and a spatial analysis method to investigate the association between spatial co-localisation of different cell types in the intra-follicular and inter-follicular regions with the patient's clinical outcome.

1.3 Multiple myeloma

1.3.1 What is multiple myeloma?

Multiple myeloma (MM) is a blood cancer of plasma cells that develops in the BM [68–70]. The BM is a spongy material in the middle of a bone. It is the factory where body cells are created and differentiated into different types. Plasma cells are white blood cells that develop from B lymphocytes [71]. Under normal conditions, plasma cells produce antibodies to fight against diseases and infections. When the normal plasma cells undergo unwanted genetic change, the plasma cells start to multiply faster, outnumbering the normal cells in the marrow [68]. Then, they start producing harmful monoclonal proteins, also known as M-Protein, and abnormal antibodies that damage the bones and other organs, such as the kidney [69]. This leads to the development of MM.

1.3.2 The development of multiple myeloma

Cancer is generally known as a multi-stage disease caused by the heterogeneous accumulation of genetic changes in cells [72, 73, 69]. While many human cancers do not have clinically known phases, MM have clinically recognised stages with less adverse characteristics. It is preceded by asymptomatic stages of monoclonal gammopathy of unknown significance (MGUS) and/or smouldering multiple myeloma (SMM) also known as pre-cursor conditions [69] as shown in Figure 1.4. Although these stages lack the clinical features of MM such as organ damage, they share some genetic alteration with MM, making diagnosis and treatment of MM challenging [74]. In precursor diseases, abnormal plasma cells reside within the BM. During the disease progression, cells start to proliferate to tissue outside the BM through the bloodstream [68, 75]. These cells mark the end stage of the transformation of normal plasma cells to myeloma cells [68, 75].

The fundamental idea behind the onset and progression on MM is that distinct pathways deregulate the plasma cell's intrinsic biology, causing myeloma-like characteristics [69]. As shown in Figure 1.4 many genetic and microenvironmental mechanisms accompanying transformation have been identified [68, 74]. Myeloma is caused by the non-linear and

brunching combination of epigenetic and inherited genetic events that, when combined with normal physiological processes needed to generate antibody diversity, lead to genetic modifications that immortalise a myeloma-propagating cell [76]. Hyperdiploidy [69, 74], MYC mutation [69, 76], copy number changes [69], and chromosomal translocation [74, 77, 69] are frequently observed in MM patients.

Myeloma cell proliferation requires a close connection with the bone marrow microenvironment [78, 74]. The myeloma plasma cells need a supportive microenvironment, such as an increase in myeloid-derived suppressor cells and Treg cells [69]. Moreover, the myeloma cells disrupt the normal cellular composition of the microenvironment to allow the myeloma cells to immortalise [74]. These immortalised cells gain genetic alterations over time, leading to clinically recognisable myeloma clinical features and treatment-resistant clonal expansion into peripheral blood, which could lead to the leukaemic phase [74, 68].

1.3.3 Diagnosis of multiple myeloma

Multiple myeloma is difficult to diagnose because some people show little to no symptoms, especially during the early phase of the disease [71]. It is mainly found in the ageing population [79]. When suspected, usually urine and blood test are conducted to examine the presence of MM associated M-proteins and antibodies [71]. During MM development, the abnormal plasma cells outnumber other cell types. Thus, quantifying the proportion of abnormal plasma cells in the BM or blood is one of the diagnostic methods. To assess the proportion of abnormal plasma cells in the marrow, bone marrow aspirate and trephine biopsy are taken [80, 81]. Figure 1.5 shows an illustrative image of BM trephine tissue sampling from the pelvis. An aspirate sample is a liquid blood sample, while trephine is a tissue sample taken from bone marrow.

From the aspirate biopsy, the abundance of cells and the presence of myeloma-associated proteins are assessed [81]. The BM trephine sample is used to assess the morphology of the cells and the bone microenvironment. Magnetic resonance and computed tomography imaging are also used to assess any bone damage in arms, legs and pelvis [82].

The MGUS and SMM patients have 1% and 1-10% annual risk of progression to MM, respectively [82, 84]. According to the International Myeloma Working Group guidelines revised in 2014 [84], MGUS is characterised by < 3 g/dL monoclonal protein and $< 10\%$ clonal plasma cells on BM biopsy. The MM is characterised by ≥ 3 g/dL monoclonal protein, $\geq 60\%$ clonal plasma cells on BM biopsy and damage of organs such as kidney and liver [84]. SMM shows level of monoclonal protein and clonal plasma cells between MGUS and MM [84].

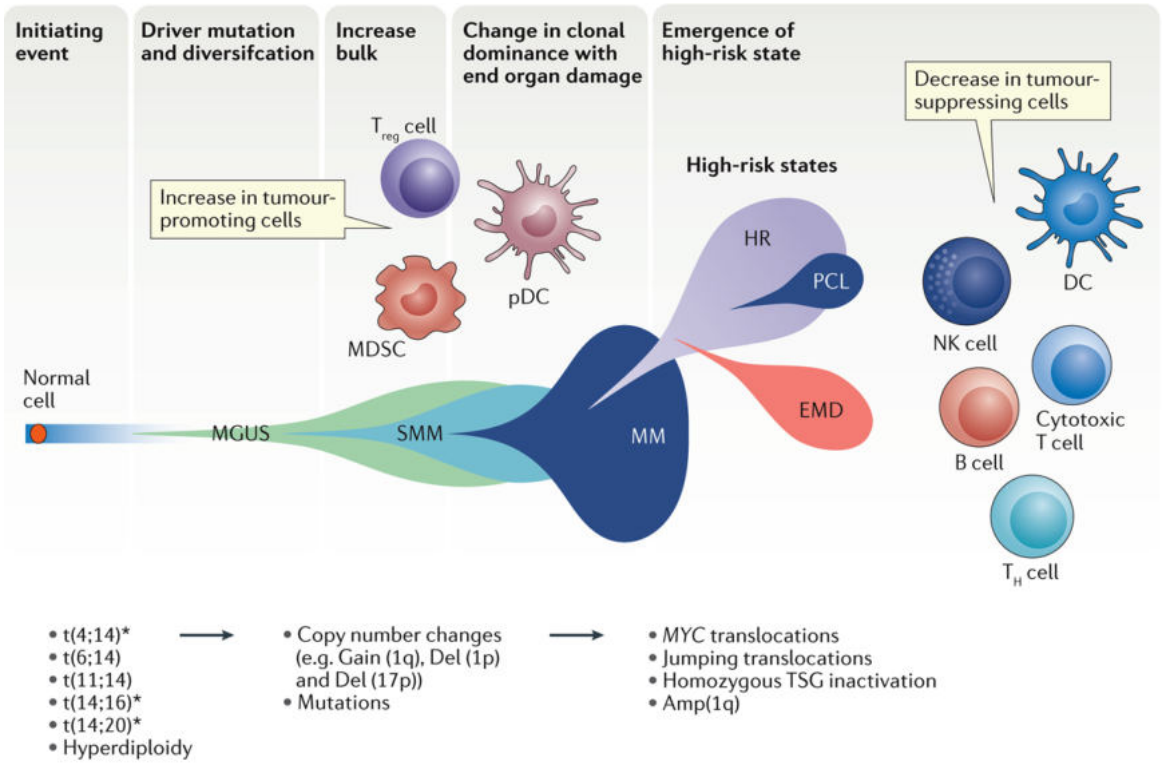


Figure 1.4: **The genetic and cellular composition changes accompanying multiple myeloma progression.** The image is taken from Pawlyn et al. [69] with permission from the Springer Nature. MM develops from normal plasma cells through precursor stages known as MGUS and SMM [69]. The onset conditions include chromosomal translocation and hyperdiploidy. The t(A:B) represents translocation between chromosomes A and B. Translocations with an asterisk (*) are those with high risk. The progression of MM is accompanied by co-evolution of genetic alteration in plasma cells such as genes translocation, hyperdiploidy, copy number alteration, mutations, and a shift in the microenvironmental cellular composition of BM [69]. The TME of an advanced stage of MM is characterised by an increase in tumour promoting and a decrease in tumour suppressing cells [69]. Amp: amplification; DC: dendritic cell; Del: deletion; EMD: extramedullary disease; HR: high-risk; MDSC: myeloid-derived suppressor cell; MGUS: Monoclonal gammopathy of undetermined significance; MM: multiple myeloma; NK cell: natural killer cell; PCL: plasma cell leukaemia; pDC: plasmacytoid dendritic cell; SMM: smouldering multiple myeloma; T_{reg} cell: T regulatory cell; T_H cell: T helper cell; TSG: tumour suppressor gene.

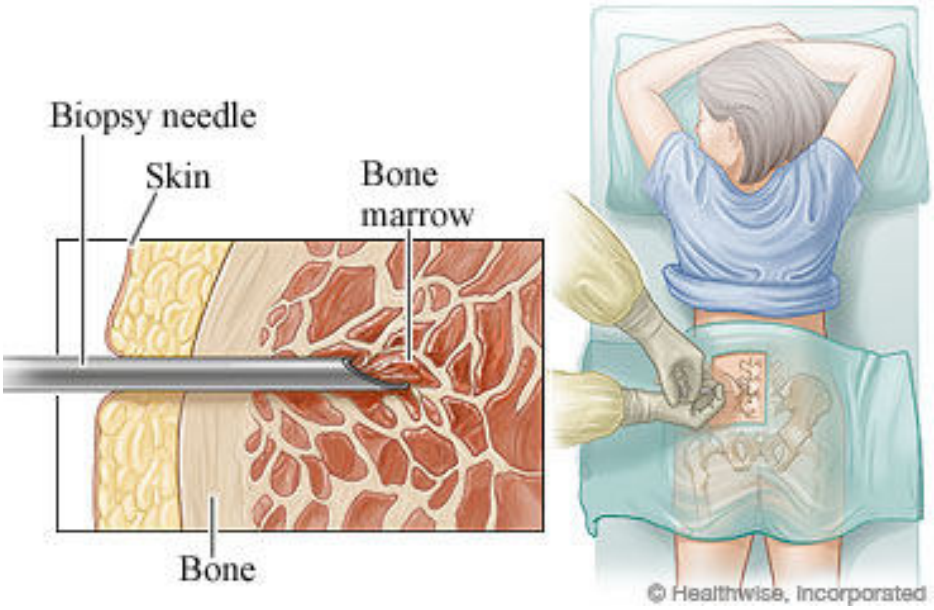


Figure 1.5: **Bone marrow trephine biopsy sampling.** A biopsy needle is inserted through the skin and the bone into the bone marrow to remove a small amount of tissue from the bone marrow. Then, the tissue will be dissected into sections to be examined under a microscope or converted into digital images to be analysed by a computer algorithm. The image was taken from Alberta Health Services without any modification following the terms and conditions mentioned in the website [83].

1.3.4 Multiple myeloma treatment

Currently, there is no cure for MM. However, the advent of innovative chemoimmunotherapy and autologous stem cell transplant (ASCT) has improved the survival of MM patients to a median of 5-to-7-year depending on factors such as tumour stage [82].

A collection of genetic events determine the clinical behaviour of myeloma, and a batch of features needs to be reported to perform effective risk stratification and identify high-risk behaviour [74, 85]. Immunomodulatory drugs like thalidomide and lenalidomide are effective in low-risk groups, allowing for long-term survival and cure [74]. However, high-risk groups do not benefit as much from these drugs [74]. The recent development of ASCT treatment has shown increased patient survival. However, there remain patients who have not benefited [69, 86]. Current risk stratification and treatment strategies are based solely on molecular profiling, which limits its efficacy as it shows only a limited spectrum of the malignancy and its TME [87, 88]. In the study by Chiecchio et al. [89], MM and its asymptomatic predecessors, MGUS and SMM, were also found to have similar genetic aberration patterns.

1.3.5 Multiple myeloma permissive microenvironment

A tumour grown to a clinically relevant size has already learned how to evade the host immune system by modulating the cellular composition of the microenvironment and seizing immune suppressive cells to modulate immune infiltration in favour of their growth [70, 73, 69, 86, 90]. The BM microenvironment is a rich ecosystem of diverse cells such as immune cells, dendritic cells, macrophages, osteoclast, osteoblast and others, which either promote or inhibit MM cells' survival and progression either individually or through a bi-directional signalling network [69, 70, 91].

Previous studies have shown that MM cells modulate the BM microenvironment by upregulating osteoclast cells and deregulating osteoblast and immune-related cells [92–94]. Osteoblasts are bone-forming cells, while osteoclasts are bone-resorbing cells [94]. Like any tissue, bone is a dynamic tissue that undergoes continuous changes throughout human life [94]. Thus, damaged or old bones are removed, and new ones are created in a controlled manner. An increase in the number and activity of osteoclast and early recruitment of osteoblast by MM cells contributing to bone resorption have been observed in MM patients [95, 96]. From histopathology images, features such as bone trephine thickness could be measured using image processing algorithms to understand how bone physiology changes during myeloma progression from precursors.

In the study by Darena et al. [97], they investigated the role of Treg cells in MGUS and MM patients using flow cytometry of peripheral blood. The Treg cells showed immune sup-

pressive phenotype in both patient groups and they did not observe a statistically significant difference in the abundance of Treg cells between the patient groups. Similarly, another study by Prabhala et al. [98] that was conducted using peripheral blood showed a similar abundance of Treg cells in MGUS and MM patients. However, in some studies, a reduced number of Treg have been observed in MM patients [98, 99].

Some studies compared the cellular composition of paired diagnostic and post-treatment samples. Lee et al. [100] studied the immune microenvironment of BM of diagnostic MM samples and post-treatment using flow cytometry technique. The patients received induction chemotherapy and ASCT, and the post-treatment samples were taken 100 days after ASCT. They reported that the percentage of CD8+ cells significantly increased while the percentage of CD4+ cells significantly decreased post-treatment compared with at diagnosis. However, there was no significant difference in the percentage of Treg cells between the two groups. Another study by Lucas et al. [101] investigated the cellular composition of diagnostic MM and post-treatment samples using flow cytometry technique on peripheral blood mononuclear cells. Similar to [100], the patient received induction chemotherapy and ASCT, but post-treatment samples were taken after 90 days from ASCT. Compared with diagnostic samples, post-treatment samples showed a significant increase in CD8+ cells and a significant decrease in CD4+ cells at post-treatment. Moreover, they reported that the percentage of Treg cells significantly increased after treatment compared with diagnosis.

The discrepancy in the abundance of Treg cells could be partly due to the variation in patient characteristics and treatment administered in the different studies. Moreover, variation in sampling sites, which included BM, peripheral blood, and whole blood could be one factor [102]. For example, the frequency of Treg cells has been found higher in BM compared to peripheral blood [103]. Furthermore, these studies have been conducted using BM aspirate liquid samples and looking at the Treg cells in the spatial tissue context of the microenvironment could give more insight.

1.3.6 Spatial context in multiple myeloma

The BM is the primary site for blood cell formation, and it contains a unique milieu that allows the continuous production and differentiation of cells [104]. Understanding the mechanisms behind the spatial organisation and geographical closeness of immune cells with tumour cells through visualisation and quantification could offer essential insights on MM progression and treatment response [105, 104].

Although the majority MGUS and SMM patients do not progress to MM in their lifetime, some patients do progress to MM with a higher chance for SMM [68, 74, 82, 84, 106]. To date, no biomarker determines which MGUS or SMM individuals may eventually develop

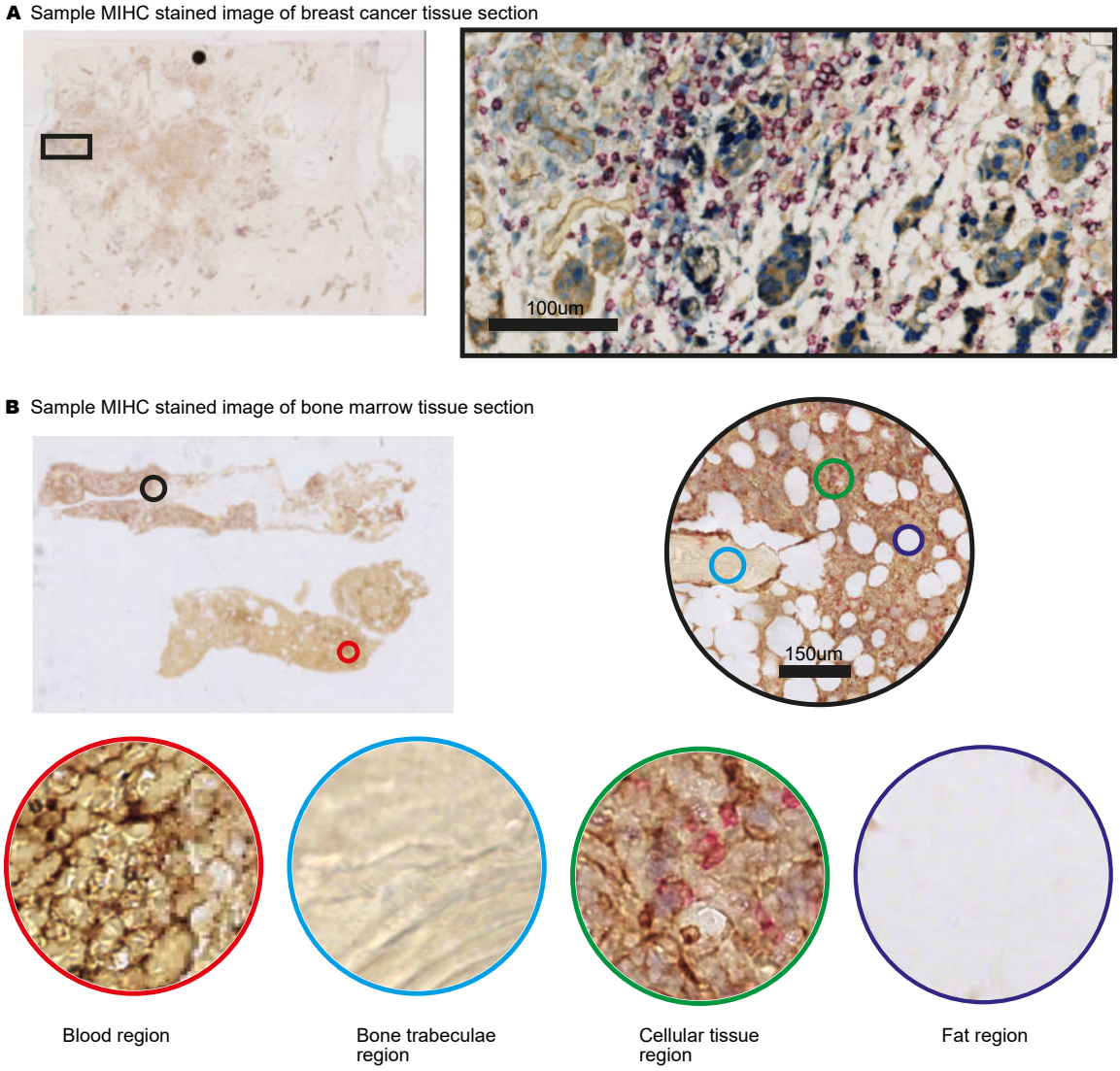


Figure 1.6: **Illustrative image showing morphologic diversity of bone marrow trephine tissue section compared to breast cancer tissue sample: A)** Multiplex immunohistochemistry taken from breast cancer tissue samples. The different colours indicate the marker or protein expression status of cells. **B)** Multiplex immunohistochemistry image of bone marrow trephine sample taken from a patient with multiple myeloma. The trephine sample contains a more mosaic microenvironment of diverse tissue compartments or types compared to the breast cancer sample in (A). The digital image of the trephine sample is composed of blood, bone trephine, cellular tissue and fat regions.

MM. In clinical practice, it is recommended to have a regular follow-up to check for MM-induced end-organ damages before treatment [107].

In previous studies, understanding of MM disease stages and transformation of precursor diseases to malignancy was mainly focused on genetic alteration in malignant cells and cellular composition of the TME using bone marrow or peripheral blood aspirate [68, 74]. Genetic aberrations associated with progression have been found, but their clinical adoption is limited due to patient heterogeneity [69, 105]. In some studies, genetic aberrations observed in MM were also found in the precursors [69, 105]. This suggests that in addition to genetic alteration in MM cells, the TME could significantly influence the disease progression and treatment search. In various human cancer types, in addition to cellular composition, the spatial organisation and interplay between tumour cells and immune cells, or among different immune cell types was found prognostic [108, 109, 67, 110]. Despite the genetic similarity between MM and precursor conditions, the spatial organisation of different cell types and the tissue morphology of the microenvironment could be different. Salama et al. [111] suggested that in haematopoietic malignancies such as MM, analysing malignant plasma cells in the context of their hematopoietic niche could generate novel therapeutic targets. For multiple myeloma, this remains under-explored, except for some recent works exploring the cellularity of bone marrow using histology images [112–114].

A pilot study by Walters et al. [105] showed that multiplex staining technologies could be applied to BM trephine samples following specialised decalcification and sample processing, and these technologies could be used to understand the spatial interplay of malignant myeloma cells and immune cells. Moreover, Walters et al. [105] shared a preliminary result showing more CD8+ T cells in proximity with malignant plasma cells in MM sample compared with MGUS.

1.3.7 Limitations of previous studies

In previous studies, exploration of TME of MM patients mainly involved liquid biopsy called aspirate. In addition to the aspirate samples of bone marrow, whole blood and peripheral blood were also employed to understand the cellular composition of TME [81]. However, this approach loses the spatial tissue context, which could generate new biological insight about the disease [111].

The BM morphological and immunological architecture can only be investigated using BM trephine tissue biopsy. Trephine biopsy contains cells intact with the tissue structural context, enabling spatial interrogation of the microenvironment. To capture the different cell types and tissue structures intact, high throughput imaging technologies such MIF and MIHC

could be used [15]. The investigation of the spatial phenotyping of samples from patients with MM is lagging behind solid tumours for several reasons:

- Obtaining BM trephine tissue sample is difficult compared to aspirate liquid sample;
- BM trephine samples need specialised tissue preprocessing due to the spongy nature of the bone marrow tissue;
- The mosaic architecture of the tissue and low tissue integrity of the BM trephine tissue hinder the adoption of ecological spatial methods developed on solid tumours. Figure 1.6 shows sample MIHC stained images of tissue sections from breast tumour resection, and BM trephine tissue sample. The BM tissue sample is a mosaic tissue that contains blood, bone trabeculae, cellular tissue and fat areas. Solid tumour (for example, breast cancer) tissue sections often have continuous tissue integrity. Thus, the spatial methods developed on such tissue might not be robust when adapted to a fragmented microenvironment like in BM trephine samples.

In this study, using carefully obtained BM trephine tissue samples, MIHC staining of immune cells and myeloma plasma cells and deep learning based computational methods, we explored the spatial interaction between myeloma cells and Treg cells with immune T cells in MGUS, MM and post-treatment samples to understand disease biology.

1.4 Ductal carcinoma *in situ*

1.4.1 What is ductal carcinoma *in situ*?

Ductal carcinoma *in situ* (DCIS) is a pre-invasive breast lesion, which is characterised by the presence of abnormal cells within the breast duct. Figure 1.7 presents cartoons showing the architecture of normal breast ducts and breast ducts with DCIS and invasive breast cancer. In DCIS, the abnormal cells are isolated from the stromal region by a nearly continuous basement membrane of the ducts [115, 116]. In the case of invasive breast cancer, the natural basement membrane of ducts collapse and tumour cells migrate into the stromal tissue surface [115, 116].

The introduction of organised breast cancer screening in the 1980s using mammography increased the detection of DCIS [117]. Though this has contributed to the early detection of breast cancer, it has lead also over-diagnosis and over-treatment in breast cancer [118]. In the United States, DCIS accounts for about 20% of all breast cancer [118]. According to Cancer Research UK, about 6,900 cases are diagnosed with DCIS in the UK each year

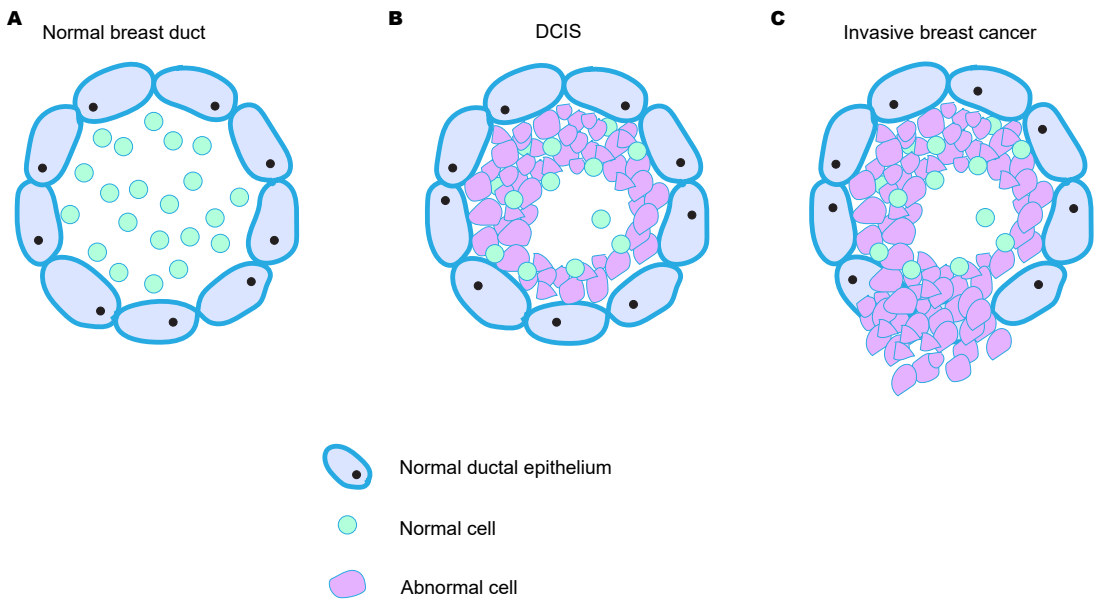


Figure 1.7: **Cartoons showing the difference between normal duct, a duct with carcinoma *in situ* and invasive breast cancer: A)** A cartoon of a normal breast duct. The duct contains only normal cells. **B)** A cartoon of a duct with ductal carcinoma *in situ* (DCIS). The abnormal cells are contained within the duct. The abnormal cells are separated from the stromal region (a region outside the duct) by a continuous myoepithelium layer, and basement membrane [115, 116]. **C)** A cartoon showing a duct with invasive breast cancer. The basement membrane is broken, and the tumour cells spread into the surrounding stromal region [115, 116].

[119]. Similar to invasive breast cancer, the chance of developing DCIS increases with age [118, 120]. With frequent detection of DCIS, this has led to more discussions on the initiation, development and heterogeneity of the disease [119].

Though DCIS is considered a non-invasive lesion, if left untreated, it is likely to develop into aggressive or invasive breast cancer [118, 120]. In the UK, about 55,920 new cases of invasive breast cancer were diagnosed each year from 2016-2018 with 11,499 breast cancer related death [119]. It is predicted that the incidence of breast cancer will increase by about 2% from 2014 to 2035 [119]. This indicates that the mortality rate of breast cancer is high, and early detection and treatment could save many lives.

1.4.2 Pathologic and clinical characteristics of ductal carcinoma *in situ*

Pathologists sometimes misinterpret DCIS as atypical ductal hyperplasia or invasive malignancy [121]. This is due to histopathologic heterogeneity of DCIS and undersampling during core-needle biopsies can potentially add to diagnostic ambiguity if the pathologist

has insufficient representative material [121]. An insufficient amount of tissue biopsy could lead to uncertainty during diagnosis and thus inter-observer variability [121]. The histologic characteristics of DCIS involve DCIS tissue architectural, and nuclear features. DCIS is characterised by heterogeneous histopathologic phenotypes. The fundamental classification method includes mitotic figures, nuclear proliferation, and architecture [117]. Architectural grouping involves assessing the malignant regions' tissue structure; the commonly known architectural subtypes are micropapillary, comedo, solid, cribriform, and mixed [122]. Nuclear grading involves identifying mitotic cells and nuclear deformation of cells [122, 118]. In terms of nuclear grade, the prevalence of high grade (42%) and intermediate grade (43%) is similar, with less prevalent low grade DCIS, which accounts for about 14% [117]. The nuclear grade is one of the most important clinical parameters for the diagnosis of DCIS because it predicts disease prognosis and the risk of progression to aggressive cancer [117, 123, 124].

Another clinical parameter is hormone receptors status [125]. Miligy et al. [125] investigated the impact of estrogen receptor (ER) expression in predicting recurrence in a cohort of 643 DCIS patients. About 74% of the DCIS cases were ER positive, and ER positivity was strongly associated with favourable outcome [125]. Moreover, another recent study by Thorat et al. [126] assessed the prognostic value of expression of ER and the impact of multi-clonal expression for recurrence in DCIS. Formalin-fixed paraffin-embedded were collected for 755 cases from the UK/ANZ DCIS trial, and ER expression was assessed using immunohistochemistry staining [126]. The ER positive patients accounted for about 70% of the cases, and ER positivity was a predictor of low risk of recurrence [126].

1.4.3 Ductal carcinoma *in situ* treatment

Currently, breast-conserving treatment is commonly advised [127]. However, if the DCIS is too widespread to permit breast conservation, then a mastectomy is the recommended treatment [127]. Furthermore, employing mastectomy significantly reduces the likelihood of recurrence [128]. After five years of follow-up, for both invasive and DCIS, mastectomy was found to reduce the rate of recurrence by about ten times and five times, compared with breast-conserving surgery alone and breast-conserving surgery and radiotherapy combined, respectively [128, 129].

Similarly, Elshof et al. [130] showed that after a median follow-up of ten years, the rates of invasive recurrence after mastectomy, breast-conserving surgery and radiotherapy combined, and breast-conserving surgery alone were 1.9%, 8.8%, and 15.4%, respectively. Another study by Darby et al. [131] investigated the magnitude of reduction in the rate of recurrence for breast-conserving surgery followed by radiotherapy on 10,801 women in 17 randomised trials. Overall, They found that radiation decreased the chance of any initial

recurrence, which included distant or local recurrence, within ten years from 35.0% to 19.3%, and it decreased the risk of dying from breast cancer within 15 years from 25.2% to 21.4% [131].

Another treatment regimen administered for DCIS is endocrine therapy. Compared to either a mastectomy or breast-conserving surgery alone, endocrine therapy was most commonly used among women with DCIS who underwent breast-conserving surgery plus radiation therapy [132]. However, in many countries, postmenopausal women with DCIS are rarely treated with endocrine therapy due to side effects and uncertain clinical trial results [128]. Nevertheless, despite a lack of consensus on its effectiveness, in the United States, endocrine therapy is more common than in other countries, and about 50% of ER+ patients receive adjuvant tamoxifen (a drug that inhibits the oestrogen receptor) [133].

The goal of DCIS treatment is to stop DCIS progression into aggressive cancer and, thus, to reduce breast cancer related deaths [118]. However, how do we know which DCIS will progress into invasive breast cancer in future? Numerous research and clinical trials have assessed DCIS prognostic variables and invasive propensity using clinical data, histopathologic characteristics, and molecular features [118].

1.4.4 Progression from ductal carcinoma to invasive cancer

Currently, the biology of DCIS is poorly understood. Despite DCIS being a pre-invasive lesion, untreated DCIS cases possess a high chance of progressing into aggressive breast cancer [118, 120]. Collins et al. [134] reported that about 46% of $1,877$ cases initially diagnosed with DCIS had developed into aggressive breast cancer upon follow-up (without any treatment). Moreover, patients with DCIS treated with breast-conserving surgery followed by radiation therapy showed about 26-36% risk of developing local recurrence after 13-20 years of follow-up [135] and 6% of developing invasive recurrence [136]. The progression of DCIS into aggressive breast cancer with and without treatment makes it challenging to understand what is driving the process.

From the molecular and histopathologic point of view, DCIS and invasive breast cancer within the same histopathologic grade have shown molecular and histochemical similarities [128]. This supports the theory that the progression of DCIS to invasive cancer might involve parallel genetic pathways [137, 138, 128]. In addition, this highlights that DCIS is not an obligate precursor of invasive breast cancer [139]. Both DCIS and aggressive breast cancer could evolve separately in the same section, probably due to the microenvironmental effect that favours carcinogenesis [117]. In this sense, DCIS and invasive ductal carcinoma should have limited genetic overlap because they evolve separately, and neoplastic cells develop within ducts until a mutation or other epigenomic event gives rise to a new population of

cells that could break the natural basement membrane of ducts and invade the stroma area [140]. The study by Kim et al. [141] showed that copy number changes and well-known mutations associated with carcinogeneses such as TP53, PIK3CA, and AKT1 were present in pure DCIS. However, driver genes and the co-occurrence of mutations were substantially less common compared with invasive breast cancer [141].

In recent years, biomarkers that could stratify newly diagnosed DCIS lesions according to the risk of recurrence have developed. Commercially available tools like the Oncotype DX Breast DCIS Score [142] have shown promising potential for predicting the risk of recurrence of DCIS after treatment [143–145]. The Oncotype DX Breast DCIS score is a continuous score ranging between 0 and 100, and the score is computed based on the expression of 21 specific genes found in the surgically excised breast tumour tissue [142]. Rakovitch et al. [145] conducted a population-based validation study of the Oncotype DX DCIS Score to evaluate the prognostic significance of this score in predicting the likelihood of recurrence in DCIS patients who only received breast-conserving surgery. Their study consisted of 828 patients with a median follow-up of 9.6 years. They found that in a population of individuals with pure DCIS treated by breast-conserving alone, the score independently predicts and quantifies recurrence risk at an individual level [145].

Moreover, recent research has indicated that understanding the interactions between abnormal cells and the TME such as stromal tumour infiltrating lymphocytes (TILs) is important to predict the likelihood of recurrence and to unveil the mechanism behind the process [120, 146–150]. Stromal TILs are lymphocytes found outside the DCIS ducts.

1.4.5 Tissue staining modality for tumour infiltrating lymphocyte assessment

Discussion on the different histopathology staining modalities can be found in Section 1.5.2 and this section discusses the histopathology staining modality suggested by the International Immuno-Oncology Biomarker working Group (TIL-WG) for stromal TILs scoring in DCIS.

The TIL-WG [151] emphasise the use of hematoxylin and eosin (H&E) stained formalin-fixed paraffin-embedded sections for the evaluation of stromal TILs since this modality is cheap, widely available, and it clearly shows various tissue architecture [152–154]. Some studies have used multiplexed whole tissue section staining techniques such as MIHC [155, 156]. However, this multiplex staining is not routinely used in diagnostic applications since it is expensive. Immunohistochemistry staining uses a specific marker to identify specific cell type(s). Thus, no single immunohistochemistry marker identifies all mononuclear lymphocytes and allows limited visualisation of tissue structures [152]. On the other hand,

H&E staining is routinely used in hospitals all over the world. Hence, currently, the gold standard for TILs assessment is H&E staining [152, 157]. However, multiplex staining is important for sub-typing the cells, providing information on the activation and functional status of cells [154].

1.4.6 Prognostic value of stromal tumour infiltrating lymphocytes in ductal carcinoma *in situ*

Invasive breast cancer has been the primary focus of research on the immune system's role in the disease's progression [158]. High numbers of immunological infiltrate, in particular effector immune cells like CD8+ T cells, are detected more commonly in human epidermal growth factor receptor 2 (HER2) positive and triple-negative invasive breast cancer (TNBC) breast cancers than in other subtypes [159, 160, 147]. Recently, the existence and potential clinical importance of the immune infiltrate in individuals with DCIS have garnered more attention in recent years [158]. Similar to breast cancer, extensive immune infiltrates are mainly found in DCIS that is HER2 positive and TNBC [158]. A detailed analysis of the immune cell subsets linked to DCIS revealed that increased CD8+ T cell subsets [161] and increased B lymphocytes [162] were linked to local recurrence, supporting the idea that TILs play a crucial role in the development of DCIS.

Preventing invasive cancer is the primary objective of treatment for DCIS [120]. Because low-grade DCIS is less likely to progress into high-grade aggressive breast cancer [120] and therefore, biomarkers that identify a group of patients that are likely to progress into invasive cancer are needed. There is growing evidence that indicates the presence of a strong immune system in the host is necessary for improved outcomes, particularly in HER2 positive and TNBC patients [163]. TILs are becoming more recognised as a promising biomarker in breast cancer with the potential to contribute to the clinical decision-making process regarding treatment [163, 148]. TILs are also increasingly being used as an important biomarker in immunotherapy clinical trials, since the prognostic role of TILs is becoming obvious [164].

However, in order to assume that high DCIS TILs will prevent DCIS from progressing into invasive breast cancer or recurrence as invasive breast cancer after treatment, this might only make sense if there is a linear evolution between DCIS and invasive cancer [163]. However, DCIS is a precursor that may not necessarily lead to invasive cancer [118, 117] because the presence of parallel evolution of both DCIS and invasive cancer within the same tissue section [128, 137, 138]. Moreover, previous studies have shown about 46% of DCIS cases do not develop into invasive breast cancer [134]. These findings might suggest that

the host's immune microenvironment could have a decisive impact on the progression of the disease [163].

Some recent works interrogating the prognostic role of stromal TILs in DCIS have shown promising results [120, 146–150]. The TIL-WG on breast cancer, which is a group of expert pathologists, clinicians, and researchers in the field of immuno-oncology biomarkers of breast cancer, has developed a set of guidelines to manually score stromal TILs on H&E digital images of DCIS resections [151]. This guideline was developed to ensure standardised reporting and reproducibility in stromal TILs assessment in DCIS [151]. The detail of the guidelines for stromal TILs scoring in DCIS can be found in Dieci et al. [151].

Numerous studies have found a link between TILs and local recurrence. However, the precise function of the immune system in the development of ductal carcinoma *in situ* remains to be fully understood [158]. The predictive effect of TILs in DCIS patients was investigated in the randomised SweDCIS radiotherapy study by Schiza et al. [165]. TILs were evaluated using H&E stained tissue sections for 711 cases following the guidelines provided by the TIL-WG and dichotomised into high ($> 5\%$) and low ($\leq 5\%$) score [165]. Majority of women (61.9%) had low stromal TILs [165]. After five years follow-up from surgery, DCIS cases with a high TILs prevalence had a significantly elevated cumulative ipsilateral breast events incidence [165]. For HER2 negative patients, high TILs was correlated with developing ipsilateral breast events over five years follow-up [165]. This might suggest that TILs are a response to extremely aggressive cancer cells that are present in the DCIS and the cancer cells in the DCIS might have taken control of the immune system [163].

Agahozo et al. [166], using H&E stained tissue sections from 473 individuals, examined the association between stromal TILs abundance and biomarker groups. The patients were grouped based on their biomarker status: ER, progesterone receptor (PR), and HER2. About 28% of the patients had high TILs (>30%) [166]. Though the overall TILs composition was not different between the different subtypes, the proportion of high TILs cases was significantly higher in HER2 positive and TNBC DCIS with the majority expressing CD4+ antibody [166].

Though more research works are using the guideline by TIL-WG [151], some works employed an extended version of this guideline. The prognostic impact of TILs touching TILs [146, 147] and circumferential TILs [167] were also investigated. Touching TILs are lymphocytes that are found within the one lymphocyte thickness distance from the basement membrane of the breast duct. In contrast, circumferential TILs are TILs in about three layers of TILs from the DCIS boundary.

Toss et al. [146] investigated different stromal TILs scoring using manual pathologists scoring, including stromal TILs touching the basement membrane of DCIS ducts and distant

TILs and evaluated their prognostic values. For the distant TILs, they investigated TILs within {0.2, 0.5, and 1}mm distance from the DCIS border. This study employed H&E images of 150 patients for training and H&E images of 666 cases for validation. They found that only touching TILs was associated with the patient's outcome and showed the highest concordance rate among expert observers [146]. Dense touching TILs was associated with shorter recurrence-free survival independent of clinical parameters [146]. They reported a lower concordance rate among observers for {0.2, 0.5, and 1}mm. This could be due to the approximate nature of the scoring by the observers, and it is not easy to accurately define larger distances such as 0.5mm compared to touching TILs. This suggests computerised algorithms could be suited to objectively quantify stromal TILs for larger boundaries since an exact distance could be set, which the algorithm will consistently use.

Another study by Xu et al. [147] also evaluated the prognostic value of touching TILs in 129 patients with DCIS, with 98 of those patients receiving whole breast irradiation treatment. A high number of touching TILs was associated with ipsilateral breast tumour recurrence [147]. Farolfi et al. studied the association of abundance of TILs level and second breast event in 496 DCIS patients after a median follow-up of 8.5 years [150]. However, when the whole cohort was considered, the 5-year cumulative incidence of second breast event did not differ between patients with TILs $\leq$5% and >5%. However, in the cases which did not receive radiotherapy treatment, high TILs was correlated lower risk of the second event [150].

Another study by Badve et al. [167] explored the prognostic relevance of touching TILs and circumferential TILs in predicting the risk of developing a second breast event. They utilised a multi-national cohort of 266 patients, of which 70 patients had an observed event and H&E staining. They found that higher circumferential TILs was associated with a low risk of developing a second breast event; however, there was no association between touching TILs and risk of developing a second breast event [167].

1.4.7 Limitations of previous studies

As discussed in the previous section, to facilitate the standardisation of TILs assessment on H&E stained tissue sections, an international group of pathologists, physicians, and researchers also known as TIL-WG developed a set of guidelines for breast cancer [151]. All the studies discussed in section 1.4.6, which assessed the prognostic value of stromal TILs in DCIS, were based on manual scoring from H&E whole slide image by expert pathologists.

However, the visual evaluation of gigapixel whole tissue section digital images by humans has innate limitations that hinder the clinical adoption of this approach. Firstly, inter-observer variability and bias due to perceptual limitation of human visual capability [168–172]. Toss

et al. [146] evaluated the concordance between multiple expert pathologists on stromal TILs scoring for different stromal boundary widths. They found that a high concordance rate among expert observers was obtained for touching TILs, but not for TILs scores obtained using larger stromal boundaries. Secondly, there is a shortage of experienced pathologists, especially in developing countries and remote areas [173, 174]. Moreover, there is a time limitation to generate a comprehensive assessment of numerous gigapixels of whole slide images to meet research and clinical needs.

The TIL-WG in their recent work [152] suggested that developing an automated image analysis method for stromal TILs assessment could address these limitations. For ease of interpretation, they suggested the development of the algorithm should follow the guidelines set for manual scoring when possible to facilitate ease of interpretation [152]. However, to the best of my knowledge, there is no automated machine learning algorithm that follows the guidelines for stromal TILs scoring in DCIS, despite the immense application of machine learning in digital pathology.

Recent developments in powerful computing devices such as graphical processing unit hardware and highly accurate methodologies have boosted the application of machine learning in digital pathology [152, 175]. Deep neural networks, especially convolutional neural networks (CNN) have revolutionised the field of computer vision, giving a promising future for digital pathology. Compared to traditional computational histopathology analysis, remarkable performance boosts have been observed in digital pathology tasks such as cell detection and classification [108, 176, 177], tissue segmentation [178–180], and predicting patient outcome [18, 181–183]. Classical computational histology image analysis involves hard-coded computer algorithms which require immense guidance from an expert and often require features extracted by an expert to make predictions. Developing such algorithms requires careful implementation, and it often suffers from generalizability to unseen data due to some fixed parameters optimised using a limited amount of data. Moreover, this approach is particularly time-consuming to apply on giga-pixel whole slide images in histology. On the contrary, deep learning based approaches show outstanding diagnostic performance and higher generalisation with the requirement of expert annotation only during model development or training.

In this study, we implemented a fully automated deep learning based image analysis pipeline for stromal TILs scoring in DCIS using single cell and tissue level algorithms to follow the essential guidelines set by the TIL-WG [151].

1.5 Histology staining technologies

1.5.1 Brief introduction on microscopy and digital pathology

Histopathology is the study of the symptoms and causes of diseases such as cancer using tissue samples [184, 187]. In histopathology assessment, samples are first extracted from the suspected area [186]. The sample could be a biopsy or surgical sample. During tissue section preparation, thin slices are extracted and put on a glass.

In traditional histopathology, the sample is put in a glass slide and visually examined by an expert histopathologist(s) using a microscope (Figure 1.8). Pathologists look for morphologic abnormalities such as the abundance of specific cell types, cells morphology (for example, nuclear shape and size), tissue abnormalities (for example, the architecture of ducts in the breast for breast cancer and structure of lymph nodes for lymphoma cancers), or cell infiltration patterns. However, the advancement of technology has brought the realisation of digital pathology.

The introduction of digital images to pathology has transformed the historical microscopic tissue assessment into what is now known as digital pathology [187]. In digital pathology, a glass slide which contains a thin tissue slice is converted into a digital image which is processed using computer algorithms [184, 188] (Figure 1.8). This has a lot more benefits than traditional tissue assessment. Digital pathology allows the remote sharing of images, enabling expert pathologists to collaborate all over the world to diagnose disease, facilitate research on disease biology, and improve the precision of diagnostic procedures [188]. Moreover, the technology allowed the assessment of morphologic heterogeneities among different tissue slides beyond a single glass slide [187]. Nowadays, digital pathology has empowered computational pathology. Computational pathology utilises computerised image analysis algorithms to harness diagnostic and prognostic features from digitised tissue specimens [184]. Moreover, these digital images could be used to develop fully automated machine learning methods to extract features of interest and evaluate the association between these features with patients' clinical data such as survival, treatment response, and molecular data [184], which will be discussed in the coming sections.

Due to the advancements in hardware and software in the past few decades, digital microscopy has emerged as an effective diagnostic mechanism in pathology [189]. The digital version of a glass slide is called whole slide image (WSI) [189]. Compared to natural images, the size of WSI is enormous and could reach up to 10GB. These images are visualised using specialised image viewers, which allow zooming in and out and sliding to explore the tissue at different scales. WSIs have gained a surge of interest in research and diagnostics. However, still, it is not fully incorporated in pathology workflow due to costs associated

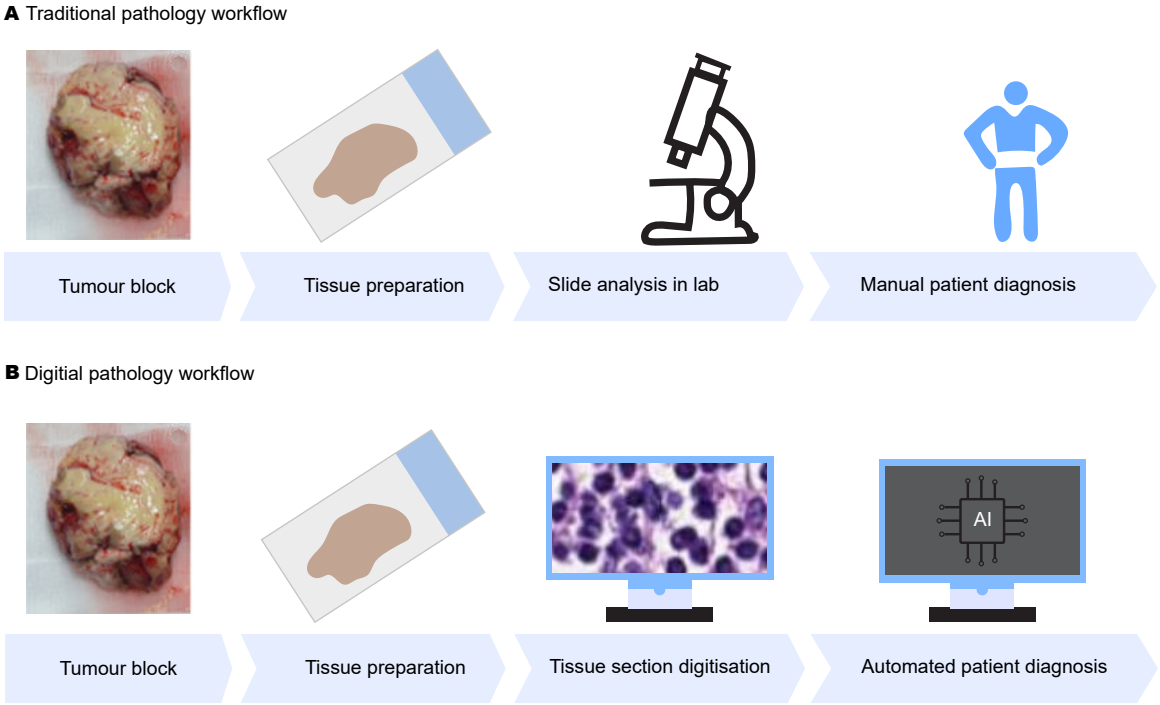


Figure 1.8: **A schematic diagram comparing traditional histology and digital pathology workflows.** In both workflows, tissue sections or thin slices are extracted from a tumour block. In traditional pathology workflow, the tissue sections are analysed by an expert pathologist(s) under a microscope to diagnose a patient (A) [187]. In digital pathology, the tissue sections are converted into digital images and analysed by algorithms to automatically generate diagnostic features from the images (B) [187].

with scanning all materials, limitation of storage, compromised image quality, and regulatory hurdles [189, 190].

The adoption of standardised diagnostic language and standards, as well as the development of digital tools, including the problems of massive data management and image processing, are the two main aspects that will increase accurate diagnosis [191]. With the aid of these tools, all accessible clinical history data may be gathered and combined to help make a diagnosis and connect pathologists for a second opinion [189]. Numerous possible benefits of automatic image analysis include less inter-observer difference, increased uniformity, and increased productivity [189]. Furthermore, the wealth of knowledge included in WSI offers enormous prospects for the development and evaluation of novel, more efficient therapies that may completely transform the treatment of patients with cancer and other disorders [189]. However, the development of image processing and machine learning techniques for decision-support diagnostic methods and validation for diagnostic applications is an ongoing research topic in digital pathology.

One of the main challenges in replacing conventional microscopy with WSI image in the diagnostic environment include the lack of experience of pathologists in using WSI, compromised view of the tissue as some tissue structures are missed in digital images, which could lead to the difference in diagnostic results among pathologist [189, 192]. However, the digitisation of tissue slides has facilitated research in histopathology, and this was engined by the development of advanced microscopic imaging technologies and image analysis algorithms [188]. The following sections will briefly describe commonly used whole tissue section staining modalities.

1.5.2 Histopathology spatial staining technologies

Hematoxylin and eosin staining

The standard histopathology staining in medical diagnosis is H&E [193]. H&E is composed of hematoxylin and eosin stains. Cell nuclei are stained with haematoxylin in dark blue to purple, while cytoplasmic regions and extracellular matrix are stained in pink to dark red [194]. In H&E, different tissue structures are represented in different shades of these colours [195].

One of the main advantages of H&E is that it is fast [196], cheap, and shows a significant amount of microscopic tissue [197]. Moreover, the output images are not greatly influenced by tissue processing or modest variations in laboratory procedure and staining device [198]. It allows the identification of histopathologic information that could help in disease prognosis, diagnosis and treatment planning and follow-up [199]. In addition to the visualisation of cells

and tissue structures, H&E allows the identification of cell types such as cancer, lymphocytes and fibroblasts [108] (Figure 1.9A, B). However, from H&E the molecular sub-types of cells could not be identified even by expert pathologists. Thus, it is frequently necessary to seek special staining such as immunohistochemical staining to support the original diagnosis on H&E stained tissue sections [193].

Immunohistochemistry staining

Immunohistochemistry is a bright field staining technique used in the diagnostic process that allows staining for a specific protein using antibody reagents at single cell level on a whole tissue section while preserving the tissue structure [199]. In immunohistochemical staining, these terminologies - markers, proteins and antibodies are often used interchangeably. As shown in Figure 1.9C, a conventional immunohistochemistry stains one marker per tissue section. This is subject to limitations, including high inter-observer variability, and it requires multiple tissue sections to assess multiple proteins [15]. The number of tissue sections that can be extracted from a biopsy or tumour surgical block is limited. To assess multiple antibodies, consecutive sections are used. This makes it challenging to evaluate the co-localisation of cells expressing these proteins due to the heterogeneity of tumours and computational complexity of aligning the digital images of tissue sections [15]. Therefore, using conventional immunohistochemistry, some crucial diagnostic and prognostic information could be missed [15]. This has led to the development of multiplexing technologies.

Multiplex staining technologies

Multiplexing technologies enable the simultaneous detection of multiple proteins on the same tissue section at a resolution similar to H&E and conventional immunohistochemistry (Figure 1.9D, E). The localisation of multiple proteins and their contemporaneous interaction with particular tissue compartments or cell types are made more accessible by multiplexing [189, 15]. Furthermore, these technologies enable the identification of protein co-expression, showing cells' functional/molecular identity, and quantifying cellular composition and cell-cell spatial interactions. Thus, multispectral assays pave an excellent potential for understanding cancer biology and for building effective therapeutic methods [15, 189, 200].

Several highly multiplexed tissue imaging techniques have emerged, enabling thorough analyses of cell makeup, function, and cell-cell interactions that point to better diagnostic benefit [15]. The MIF and MIHC are among the most commonly used multiplex staining technologies. They have been also shown to be cost-effective, and reproducible with standardised analysis pipeline [15].

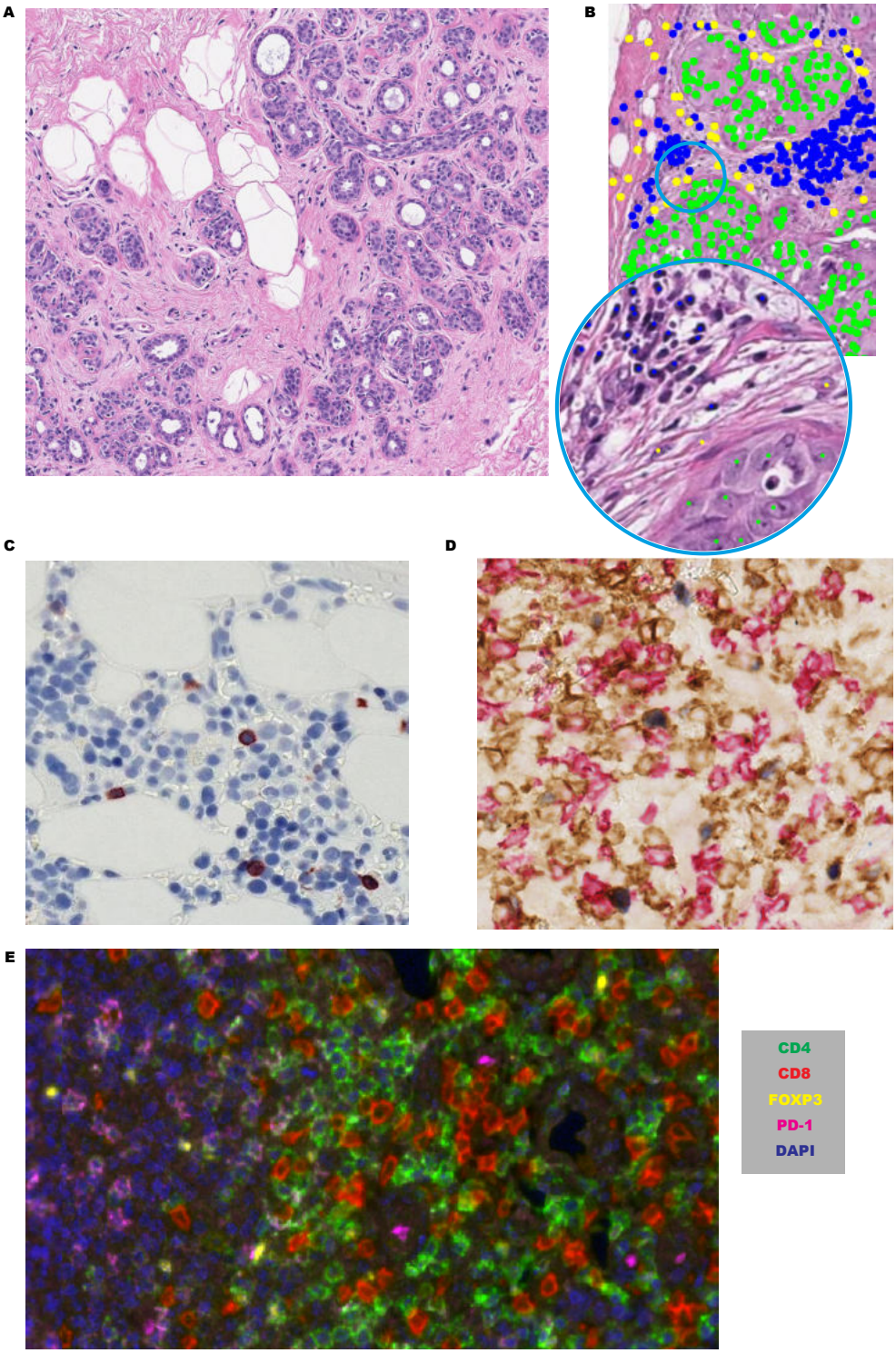


Figure 1.9: **Example of images from commonly used histopathology staining technologies:**

Figure 1.9: **A)** A sample H&E image of ductal carcinoma *in situ* tissue section. **B)** Pathologist annotation of different cell types from H&E images. The labels are, blue: lymphocyte; green: cancer cell and yellow: fibroblast. **C)** A sample image of conventional immunohistochemical staining containing only one marker, CD20 (brown). **D)** A sample image showing MIHC staining of multiple markers. The labels are, blue: FOXP3; brown: CD4; and red: CD8. **E)** A sample MIF images captured using the Vectra 3 multiplex imaging platform. DAPI stands for 4',6-diamidino-2-phenylindole.

In clinical and research areas, these high-plex and high-resolution imaging technologies could play a crucial role in the development of cancer immunotherapy [201]. Lu et al. [201] conducted a systematic review and meta-analysis on 8,135 patients samples belonging to various solid tumours from different studies. The patients received programmed cell death receptor 1 (PD1) or programmed cell death receptor ligand 1 (PDL1) immune checkpoint inhibitor treatment. Lu et al. [201] showed that compared to single staining immunohistochemistry and molecular profile-driven phenotypes, including gene expression profile and tumour mutation burden, multiplex staining was found to be highly accurate in predicting response to treatment [15].

As mentioned above, one of the main disadvantages of immunohistochemistry is that one marker is used per tissue section. However, a given cell can express multiple markers showing the multi-functionality of the cell. For instance, CD8+ tumour infiltrating lymphocytes could be identified using CD8, CD20 or CD3 antibody staining [202, 15]. Moreover, expression of markers such as PD1 or PDL1 on the surface of these CD8+ T cells of cancer patients could be a good indication of effective PD1 or PDL1 inhibitor immunotherapy treatment [203, 204]. In addition, a better knowledge of cancer initiation and progression could be attained from research on the pattern of markers expression and the relative spatial organisation of immune cells, cancer cells, and stromal cells [2, 15, 205]. Thus, multiplex spatial staining technologies provide a broader spectrum regarding the cellular microenvironment of a tissue section compared to standard H&E and conventional immunohistochemistry.

In the market, there are different multiplex spatial staining technologies. Tan et al. [15] presented an overview of different multiplex imaging modalities and vendors. These technologies differ in their tissue processing, the number of markers that could be stained, speed, cost and resolution of the output data [15].

In comparison to MIHC, MIF offers more optimal and increased multiplexing and contrast capabilities [15]. Most studies using MIHC could go up to 3-plex, while MIF could provide up to 50-plex markers [15].

Imaging mass cytometry is another high throughput spatial imaging technology [206, 207]. However, compared with MIF and MIHC, it is slower and has lower sensitivity and

specificity [15]. Moreover, its resolution is limited to a maximum of about $1\mu m$ per pixel, while MIF and MIHC staining technologies could achieve up to $0.25\mu m$ resolution per pixel [15].

In general, these technologies capture multiple markers on the same tissue section with a subcellular resolution while preserving spatial tissue architecture. However, the aggregation of multiple colours/features in one image makes it challenging to analyse the images manually for pathologists or analysis using traditional image processing algorithms used for other imaging modalities. Moreover, the output data is even bigger than H&E in dimension since multiple channels are generated for each marker, for example in MIF staining. When 10s of markers are aggregated in one high-resolution image, it is beyond human perception to estimate the spatial pattern of multiple expression profiles. Thus, the different multiplex staining technologies come up with installed image processing algorithms. For example, the Vectra platform has a built-in image analysis pipeline called inForm cell profiling. Moreover, there are commercial and publicly available platforms to analyse multiple images.

The most commonly used commercial platforms to analyse MIF and MIHC images are HALO and Oncotopix [15]. HALO and Oncotopix are developed by Indica Labs and Visionpharm, respectively. Both platforms provide cell profiling, tissue segmentation and cells neighbourhood analysis packages, and the platform can read different MIF and MIHC image formats.

QuPath is a freely available open-source software for the visualisation and analysis of biomedical imaging data including H&E and multiplex images [208]. Qupath provides comprehensive image analysis tools based on traditional pattern recognition algorithms to detect and classify single cells and tissue segmentation algorithms. Moreover, it allows adding user-developed scripts to the existing functionalities and batch processing. Moreover, spatial data and morphologic features could be exported in a tabular form for downstream analysis [208]. Recent versions of QuPath also allow using deep learning based image analysis algorithms such as StarDist [209] as plug-ins. However, the development of QuPath is still in progress, and the traditional machine learning algorithms installed in the software might contain user pre-defined parameters such as intensity threshold, which might not work well on new images obtained using different devices and settings. In addition, this makes the algorithm susceptible to the batch effect. Thus, users, most of the time, use separate scripts with user-defined parameters which inherit QuPath functionality. While this allows flexibility, but limits scalability to a large number of images.

In recent years, the development of highly accurate deep learning based computer vision algorithms has impacted digital pathology image analysis. Thus, digital pathology has witnessed a surge of interest in the application of deep learning by customising the algorithms

that were developed on natural images to digital pathology problems. The next section gives a brief overview of deep learning and its application in histopathology image analysis.

1.6 Deep learning for histology image analysis

Recently, artificial intelligence (AI) has gained a surge of interest in almost every field, including in science [210], economics [211], and security [212]. AI is the engineering and science of constructing intelligent machines with computer algorithms that demonstrate human-like learning from data, decision-making, and problem-solving [213]. AI is not only about replicating what we human beings can do; AI have been used in situations that are risky for humans and computational-intensive environments. In fact, AI has become an integral part of large search engines like Google. AI is a vast field that could range from a simple room temperature controller using a user-defined algorithm to a self-driving car using deep-learned data features to make decisions.

Deep learning is a subfield of machine learning, and machine learning is a subfield of AI. Machine learning enables computing devices to decide without strict instructions from humans. On the other hand, deep learning introduces more freedom to computer algorithms and uses multi-layered architecture to learn more abstract features at different scales. Deep learning is inspired by biological neuron function, which is computationally realised by artificial neurons.

1.6.1 Artificial neural networks

An artificial neuron is the building block of deep learning models. The computations in artificial neurons are inspired by the function of a biological neuron [214]. Figure 1.10A shows a schematic diagram of an artificial neuron [215]. Fully connected artificial neural networks consist of interconnected artificial neurons that aim to solve a particular set of problem(s). The network accepts a vector of input data and predicts an output (Figure 1.10B).

The algebraic representation of the artificial neuron displayed in Figure 1.10A can be written as shown in Equation (1.1).

$$\hat{y} = f(z) = f\left(w_0 + \sum_{i=1}^{i=n} w_i\, x_i\right) \tag{1.1}$$

where x_i and w_i represent the i^{th} input feature and it corresponding weight, respectively. The w_0 represents a bias parameter. The function f is a non-linear activation function, which has z as an input. The $\hat{y}$ is the output of the neuron. The bias parameter is used as an

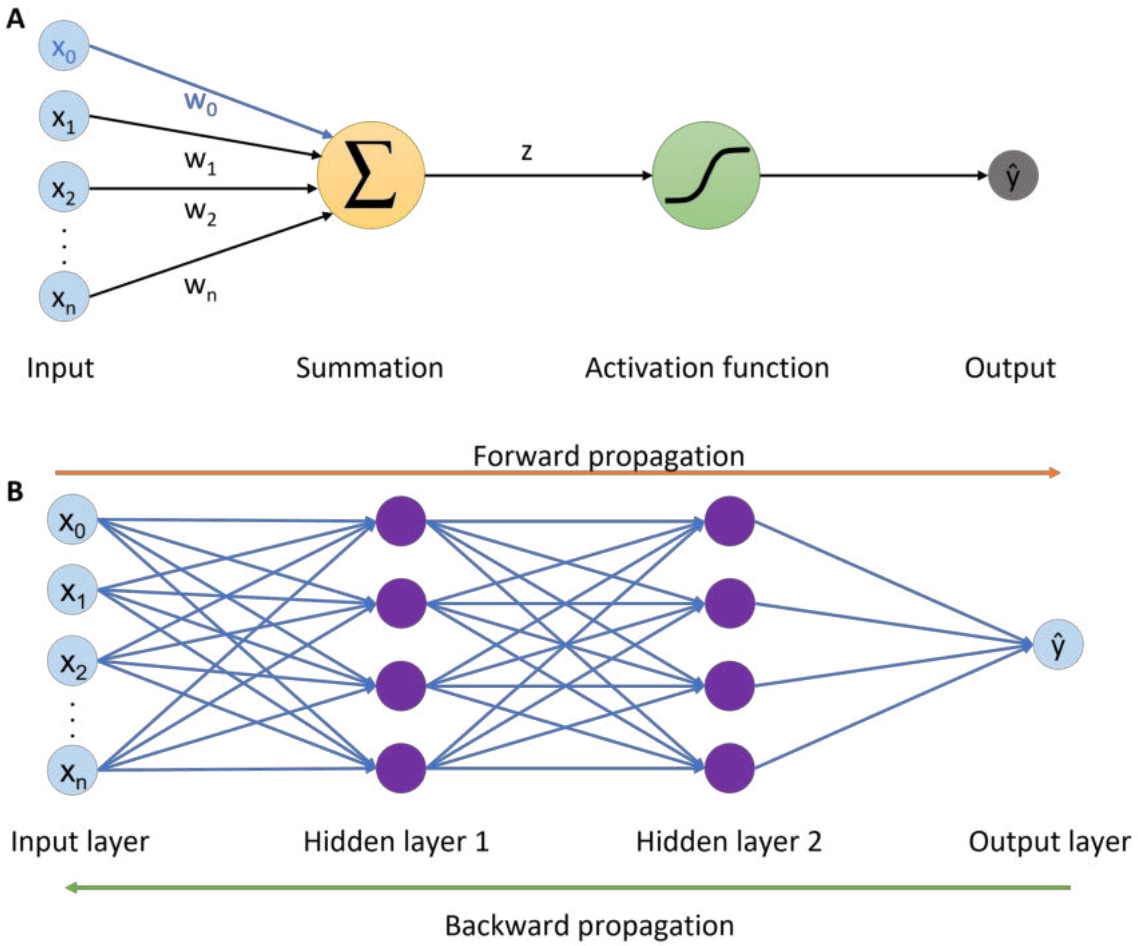


Figure 1.10: **Structure of artificial neural networks: A)** A schematic diagram of an artificial neuron. The $x_0, x_1, x_2, ..., x_n$ represents a set of input features. $x_0 = 1$ and w_0 is called a bias. The parameters $w_1, w_2, ..., w_n$ are called weights. The circle with the summation sign represents the weighted summation of the input features, which outputs z. The weighted sum, z, is fed to a non-linear activation function to generate the neuron output, $\hat{y}$. **B)** Schematic diagram of a fully connected artificial neural network which contains an input layer, two hidden layers and an output layer. The purple-coloured circles represent neurons of the hidden layers. Every neuron in the hidden layers has input features, summation block and activation functions.

offset value which sets the threshold for the non-linear activation function f. In a real-world application, the relation between an input variable and the output variable could be non-linear. The activation functions are included to model the non-linear association between the input and output variables. The most commonly used non-linear activation functions include rectified linear unit, sigmoid, and hyperbolic tangent. The advantages and disadvantages of different activation functions could be found here [216].

In a multi-layer artificial neural network as shown in Figure 1.10B, the output of a given neuron at any layer could be computed similarly. For example, the output of the second neuron of the first hidden layer in Figure 1.10B, $h_2^{(1)}$, can be computed using Equation (1.2).

$$h_2^{(1)} = f\left(w_{0,2}^{(1)} + \sum_{i=1}^{i=n} w_{i,2}^{(1)}\, x_i\right) \tag{1.2}$$

where the superscripts in weight parameters and h indicate the hidden layer number. The $w_{i,2}^{(1)}$ represents the multiplicative weight of the i^{th} feature from the input layer (layer 1), x_i to the second neuron in the first hidden layer. Similarly, the output of the remaining neurons in the hidden layers and output layer could be computed.

The weight and bias parameters are learned or optimised from data. The process of optimising the parameters from the data is called model training. In order to train the artificial neural network, an objective function, also known as a loss function, should be defined. The objective function computes the difference between the actual value and the value predicted by the network. The choice of loss function depends on the task. For example, for the classification problem, categorical cross entropy and its variants are commonly used, while mean square error is commonly used for regression-related tasks. Model training aims to find model parameters that minimise the error and it is an iterative process. The input data is iteratively fed to the network to compute the predicted value. This is called forward propagation. Subsequently, model parameters are updated to minimise the error in the model prediction using a process called backward propagation. A detailed explanation of forward propagation and backward propagation can be found in these books [217, 218].

Fully connected multi-layer artificial neural networks are suited for one-dimensional data. However, they are not efficient for multidimensional data such as images. If a pixel is considered a feature in an image, a small-sized image, for example, a 100×100 pixels RGB image (coloured image), will result in a $30,000$ dimensional feature after flattening. Moreover, in an image, the two-dimensional or 3-dimensional spatial organisation of pixels is important. Convolutional neural networks are specifically designed to address these issues.

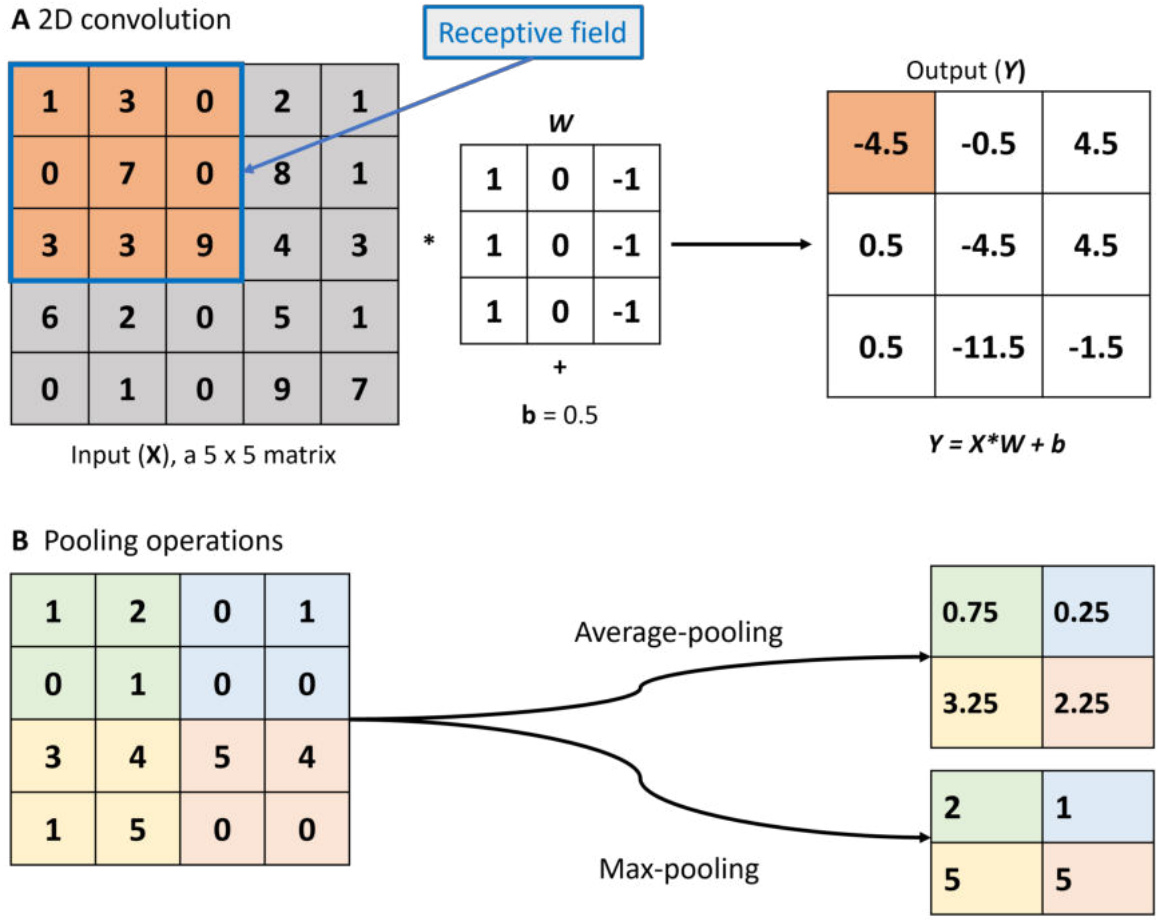


Figure 1.11: **Working principle of convolutional neural networks: A)** Schematic diagram showing convolution operation on a 2-dimensional array, X. The weight (w) and bias (b) are the parameters of the convolution operation. The receptive field is a region where convolution is applied. The asterisk (*) represents the convolution operation. The shaded area in the output (Y) represents convolution operation output over the shaded receptive field in X after adding the bias. The convolution was applied using a sliding window stride of 1. **B)** A schematic diagram showing max- and average-pooling operations. The receptive field of pooling operation is a non-overlapping 2×2 window.

1.6.2 Convolutional neural networks

Similar to fully connected artificial neural networks, CNN consists of an input layer, multiple hidden layers, and an output layer. The most commonly used hidden layers include the convolution layer and pooling layers. Convolution layers contain learnable convolution filters and biases. The convolution is computed only from neighbouring neurons. Figure 1.11A shows a diagram of the convolution operation. The neighbourhood region is known as the receptive field [217]. In the case of fully connected layers, as shown in Figure 1.10B, the receptive field is the entire previous layer's output. On the other hand, in CNN, the receptive field is much smaller than the entire image region.

Moreover, the filter is shared across the entire dimension of the image. Thus, CNN create sparse connections and significantly reduces the number of parameters compared to fully connected artificial neural networks. For example, as shown in Figure 1.11A, for a 5×5 input matrix (which can be considered as an image), a 3×3 filter is used. To find the output, a sliding window of different receptive fields was considered. This results in only ten trainable parameters: nine weights and one bias. If a fully connected artificial neural network is applied, 26 parameters will be needed: 25 weights and one bias. This makes CNN memory efficient for image data. Moreover, such implementation of convolution layers enables to detection of objects within an image irrespective of their location.

Another essential layer in CNN is the pooling or subsampling layer. It is usually placed after the convolution layer. The most commonly used pooling operations are max- and average-pooling, shown in Figure 1.11B. Pooling helps reduce the feature map's size and learn multi-scale features. A feature map is the output of a neuron in the hidden layers. Pooling also serves as a regularisation method to avoid over-fitting. Similar to artificial neural networks, activation functions are used in CNN as well.

A CNN consists of cascaded convolution and pooling layers to automatically extract multi-scale features from images [219]. A CNN could be applied for an image-to-image translation, to predict a single value, such as the class of the image, or to regress a specific value, for example, the number of objects in an image. Thus, the last layers of CNNs are task-specific. For the classification task, the cascaded convolution and polling layers are used for feature extraction. The extracted features are flattened and fed to fully connected layers to predict class probability [220]. A detailed explanation of the working principle of CNN can be found in the book by Goodfellow et al. [217] and a detailed review of various CNN architectures can be found in [221].

In the last decade, deep learning techniques have been enormously utilised in histopathology image analysis. The following section gives an overview of the application of CNN in histopathology image analysis.

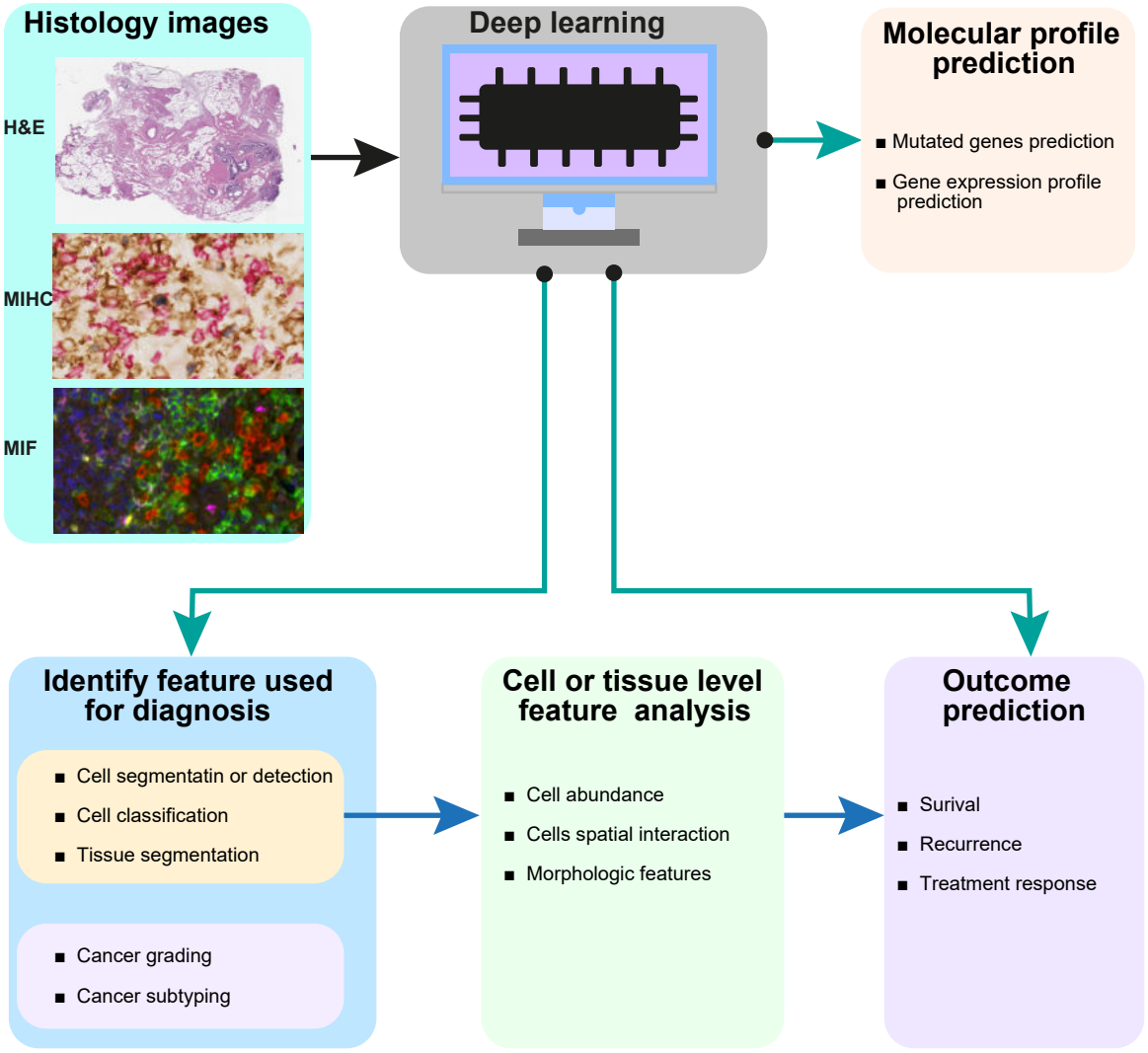


Figure 1.12: **Overview on the application of deep learning in digital pathology**. MIF = multiplex immunofluorescence; MIHC = multiplex immunohistochemistry and WSI = whole slide image.

1.6.3 Application of deep learning in digital pathology

Pathology is the backbone of cancer diagnosis, pharmaceutical research, clinical research, and treatment selection and follow-up [186]. The current gold standard for histopathology evaluation is based on a manual assessment of samples by expert pathologists. For standardised reporting, a lot of robust guidelines have been developed by a group of international expert pathologists [151]. However, there is a shortage of expert pathologists all over the world [222]. Moreover, manual assessment suffers from inter-observer and intra-observer variability due to a subjective estimation of parameters and visual perception limitations of humans, especially for gigapixel histopathology images [223]. These hamper the fulfilment of current research and clinical needs, clinical adoption, and innovation of new prognostic biomarkers [186, 152]. Furthermore, as we are moving towards individualised cancer therapy, the need for an accurate histopathologic biomarker assessment is growing [186].

The advent of devices to convert glass slides into high-resolution digital images has sped up research and training in pathology [186]. This has also enabled the development of computer algorithms to analyse these digital images, which could be used as decision support for pathologists. Due to the availability of increased histopathology images and computing power, deep learning-powered digital image analysis offers hope for increasing the scope of digital pathology and accuracy of cellular and morphologic biomarkers evaluation [152, 186]. Moreover, digital pathology is gaining popularity due to the development of high-plex and high-resolution spatial molecular profiling technologies such as MIF that give quantitative results with minimal inter-observer variability [186]. While the adoption of deep learning models in hospitals remains a milestone to be addressed in the future due to a lack of interpretability and trust, it has been extensively used in research to address diverse methodological and clinical problems in cancer research as shown in Figure 1.12.

Recent AI breakthroughs could change how pathologists identify and stratify cancer, and deep learning techniques constitute a milestone in this change, as they are behind major advancements in histology [186]. While various types of deep neural networks exist, CNNs are the most commonly used architecture in digital pathology image analysis [224]. A detailed review of the application of CNN in digital pathology could be found in [220, 224]. Recently, in digital pathology-focused research, a wide range of deep learning architectures have been explored [225] aimed at various clinical and methodological goals (Figure 1.12). Among them are below:

- Developing cell and tissue level deep learning models to identify diagnostic features;
- Predicting patient outcome;
- Predicting molecular features from digital pathology images.

Developing cell and tissue level deep learning models to identify diagnostic features

The field of digital pathology has recently witnessed a surge of interest in the application of deep learning for cell classification [176], cell detection [226], cell counting [227–229], cell segmentation [178, 230, 231], tissue segmentation [178, 232]. However, automated cell detection and classification remain challenging due to variations in slide preparation and cell morphological diversity in shape and size. For example, closely located cells with weak boundaries are often difficult to discern [226, 227, 233].

Mercan et al. [234] developed and evaluated a deep learning based WSI analysis decision support pipeline for the diagnosis of breast cancer. The study consisted of 240 breast biopsies categorised into benign, atypia, DCIS, and invasive cancer. Three expert pathologists performed the diagnosis. First, to segment the tissue area into necrosis, stroma, background, normal epithelium, malignant epithelium and blood, a ResNet based CNN was employed [232]. Then, morphologic tissue texture and structure features were extracted from these different compartments and used to predict a diagnosis of patients. The authors showed that the deep learning-based approach is promising in assisting pathologists in differentiating atypical samples from DCIS samples [234].

Campanella et al. [235] developed a multi-instance learning based weakly supervised deep learning on H&E WSI to predict cancer versus normal on a large scale dataset consisting of 44,732 whole slide images from 15,187 patients. The dataset belonged to breast metastasis to lymph nodes, skin cancer, and prostate core biopsies. The authors evaluated the performance of different CNN models, including ResNets [236], VGGs [237] and DenseNet201 [238] for the classification task. Their CNN based classifier achieved above 0.98 area under the curve (AUC) for all cancer types on a test set. They claimed that the deep learning model could be used as a decision support system and could enable pathologists to exclude about 60-70% of patients with 100% sensitivity [235]. Adopting a weakly supervised method enabled them to avoid the need for time-consuming pixel-level annotation.

Another study developed a CNN based pipeline to identify and classify tumour-associated stroma in diagnostic breast biopsies [239]. The algorithm was evaluated on 2,387 H&E stained tissue sections from 882 women aged 40–65 years. The deep learning method was trained to differentiate stroma regions in proximity to invasive regions and stroma from benign samples. Using a stromal features-based binary classifier, they achieved an AUC of 0.96 for benign vs invasive classification. They showed that the morphology of the stromal region could help classify breast biopsies and understand TME of breast cancer.

In the study by Nagpal et al. [240], they developed a deep learning algorithm to predict the Gleason score of prostate cancer. The deep learning model was trained and validated on 12 million manually annotated patches from 1226 and 331 slides, respectively. Compared to a

standard reference score, the automated deep learning method has significantly outperformed 29 pathologists (accuracy of 70% vs 60%). The authors suggested that the deep learning method could be used in a situation where there is a smaller number of pathologists. Abdul-Jabbar et al. [108] developed a deep learning methodology to interrogate the spatial immune cell distribution in lung cancer. They developed CNN based deep learning pipelines to differentiate the tissue regions from the background, detect cells, and classify cell phenotypes on H&E and MIHC images. The H&E and MIHC images were from the same tumour block, and AbdulJabbar et al. showed the density of tumour cells from H&E images and density of TTF1+ tumour cells from MIHC images was highly correlated, showing the reproducibility of their automated computational pipeline. Moreover, tumours with more than one immune cold region had a high risk of relapse independent of standard lung cancer clinical variables. The immune cold regions were identified using immune hotspot analysis [241]. They showed that immune geographic variability reveals tumour ecological restrictions that may promote immune evasion and aggressive phenotypes.

Predicting patient outcome

In recent years, deep learning algorithms have been employed to perform complex tasks from high-resolution digital pathology images, such as predicting patient survival and treatment outcome. Deep learning models could be trained using expert annotations for tasks such as single cell detection, cell classification, and tissue segmentation. However, the pathologist could not tell at least with high certainty the survival of patients solely from H&E images. However, end-to-end deep learning algorithms have been employed to regress patients' survival from only H&E WSI without any prior input from pathologists.

One of the main challenges in this approach is that WSI are too big to load into memory and train deep learning models. Some images could reach up to 10GB in size, making it difficult to fit into GPU for model training. Thus, patch-based approach is mainly adopted. Moreover, it is not known prior which part of the gigapixel image is associated with patients' survival. The algorithm should learn this automatically.

Yamashita et al. [242] proposed a deep learning method to score the risk of relapse of hepatocellular carcinoma after surgery. About 70% of hepatocellular carcinoma patients relapse within 5 years from surgery [242]. Yamashita et al. [242] first divided WSI into patches and trained CNN based classifier to identify patches highly likely to be from the tumour region. These patches were used to train another deep learning model to predict the risk of relapse. They achieved a concordance score of 0.724 and 0.683 on internal and external validation data, which was higher than the Tumor-Node-Metastasis standard classification system. They suggested deep learning could be used to refine the clinical management of

hepatocellular carcinoma. The drawback of this approach is that the relevance of the stromal region is excluded during model training.

Courtiol et al. [243] developed a deep learning pipeline to predict the overall survival of mesothelioma patients from H&E WSI that utilises patches in both tumour and stromal regions. The deep learning model achieved a concordance index of 0.654 on separately held test data. Another study by Wang et al. [244] demonstrated that recurrence in early-stage non-small cell lung cancer could be predicted using automatically extracted nuclear features from H&E tissue micro-array images. CNN was used to segment cell nuclei and nuclear shape. Subsequently, texture features were extracted. To predict recurrence based on these features, a support vector machine was used [244]. The model was developed using a cohort of 70 (cohort 1) patients and then validated on two other cohorts, cohort 2 (119) and cohort 3 (116) patients. Their model achieved an accuracy of recurrence prediction of 81% on the training set, and 82% and 75% on cohort 2 and cohort 3, respectively. They found that nuclear histomorphologic features could robustly predict recurrence status in non-small cell lung cancer.

Predicting patient outcomes from histopathology images is very complex. As presented above, some studies trained a deep learning model to predict the patient outcome from the whole region of the tissue section. In contrast, others trained deep learning models only on the tumour regions of the tissue section. While patch prediction ranking was used to identify regions or patches predictive of the outcome [242, 243], this approach is less interpretable since the patches contain a composite of tissue and cellular features. Others used deep learning as an intermediate step to extract human interpretable features such as nuclear morphologic features [244], and these features were interrogated with patient outcomes. This approach provides better interpretability than the earlier approaches. In this study, we adopted this approach.

Predicting molecular profile from digital pathology images

Tumour molecular profiling is the characterisation of the genomic make-up of tumour [245]. In the era of personalised medicine, molecular profiling has become crucial for informing therapeutic decisions, tumour diagnosis and prognosis in daily practice [246]. The number of druggable tumour-related molecular abnormalities has increased significantly over the past ten years, and biomarker-matched medicines have significantly improved survival in many cancer types [246]. For example, HER2 targeting drugs have significantly improved the prognosis of HER2 positive breast cancer patients [247]. However, tumour molecular profile is expensive and time-consuming. Due to these factors, molecular testing is challenging to scale up to clinical needs. H&E histopathologic images are routinely available ubiquitously.

Though pathologists are not able to estimate molecular profiles by simply looking at histology images, some recent works based on deep learning have shown promising results.

Coudray et al. [248] trained CNN to predict the most commonly mutated genes in lung adenocarcinoma from H&E images. They showed that mutation of six genes, including STK11, EGFR, FAT1, SETBP1, KRAS and TP53, could be predicted from H&E images. They trained a binary classifier, mutation exists or absent and obtained area under the curve ranging from 0.733 to 0.85 on these genes. Using a similar approach, Anand et al. [249] predicted BRAF mutation in thyroid cancer. In another study by Barbera et al. [250], it was demonstrated that deep learning could predict HER2 status from H&E histopathology images. Their pipeline mimics the pathologist strategy by first segmenting the tumour region, and the status of HER2 was predicted from the tumour region. Similarly, Shamai et al. [251] developed a ResNet-based deep learning classifier to detect ER status from H&E images. They trained and validated their algorithm on $20,600$ digitised images of $5,356$ patients with breast cancer from two cohorts. Their model achieved about 91% ER detection, and they suggested that histopathology and deep learning could be adopted for mass molecular profiling to meet the current clinical needs.

Another study by Kather et al. [18] developed a deep residual network to predict microsatellite instability in gastrointestinal cancer directly from histology images with an AUC of 0.84. They applied a multi-stage approach, where tumour vs normal tissue patches was separated, and another classifier was applied to predict the microsatellite instability status. Microsatellite instability determines gastrointestinal cancer patients' response to immunotherapy, and Kather et al. claimed that their approach could be used as a support system to guide whether immunotherapy could be a better treatment for a given patient solely from histology images. Xu et al. [252] postulated that the underlying genetic drivers are responsible for the aberrant alterations in tumour cell nuclei morphologies and the TME and that these features in histological images might therefore be used to predict genetic status such as tumour mutation burden. Transfer learning based CNN model was applied to predict mutation burden on 253 patients with bladder cancer from The Cancer Genome Atlas (TCGA). Their model could differentiate low vs high tumour mutation burden with an accuracy of 73%. Moreover, the survival of a patient with a predicted low mutation burden and high mutation burden was significantly different.

All the above studies rely on CNN to directly predict molecular features. As mentioned in the previous section on patient outcomes prediction, such approaches lack interpretability. Dia et al. [253] proposed a novel approach to predict molecular phenotypes from WSI by extracting human interpretable image features. They used more than 1.6 million single cell annotations from board-certified pathologists from about $5,700$ images to train deep learning

models for high-resolution cell classification and cell detection. After segmenting cells and tissue types, they extracted 607 histomorphologic features. These features were found to be associated with molecular features such as checkpoint proteins. Another study by Lu et al. [254] proposed a cell graph based graph deep learning approach that captures the whole image instead of a small region to predict HER2 and PR status in breast cancer patients. WSI could have hundreds of thousands of cells, and the cell graph was built by connecting neighbouring cells with an edge. The authors showed that a graph based method outperforms other deep learning approaches in predicting HER2 and PR status.

Gamble et al. [255] developed three separate deep learning models to predict ER/PR/HER2 biomarkers directly from H&E images. Their model was trained and validated on $3,274$ slides from $1,249$ patients collected from 37 sites. AUC values of 0.86, 0.75, and 0.60 were obtained for ER, PR and HER2, respectively. They used saliency analysis and pathologist reviews of clustered patches to interpret model learnt predictive features. Interpretability analyses reveal established relationships between biomarkers and histomorphology, such as those between low-grade and lobular histology, and ER/PR positive and higher inflammatory infiltrates [255].

1.7 Thesis objective

The focus of this thesis is the study of the complex tissue microenvironment in multiple cancer types that have unique tissue architectural features. In particular, we illustrate how tailored deep learning and spatial statistical algorithms can be used to understand disease biology and identify prognostic features in FL, MM and DCIS (Figure 1.13). Following the overview of these diseases, their biology and clinical management in **Chapter 1**, here I outline the specific objectives of this thesis along with the methodological and clinical questions addressed.

1.7.1 Deep learning methods development for cell phenotyping on multiplex images

In oncology, spatial context is crucial to understand cells' spatial organisation and interaction within a tissue. As discussed in Section 1.5.2, MIF is a spatial imaging technology that enables spatial single cell phenotyping of whole tissue sections at high resolution. This technology is rapidly growing, and it needs advanced image analysis algorithms to fully harness the available high-dimensional spatial data. Currently, there is no generally accepted single pipeline to analyse MIF images due to the images' complexity associated with the

high dimensionality of the data, intermixing of multiple markers and variability in colour and number of markers used. Some studies used commercially available software such as HALO and Visiopharm. However, these tools are costly and lack flexibility, especially for exploratory research. To resolve this situation, we aim to develop a new, fully automated, deep learning based computational pipeline. While developing the pipeline, we aim to answer the following crucial questions from domain knowledge of the data and its application:

- What is the efficient way of collecting single cell annotations to train and validate deep learning models for MIF images?
- How to ensure the model generalises to multiple panels which use different proteins, numbers of proteins and staining colours?
- How to include domain knowledge into deep learning model development to improve performance?

Chapter 2 attempts to address these central questions to develop a scalable deep learning based fully automated workflow called **DeepMIF** for spatial single cell phenotyping from MIF image.

1.7.2 Spatially resolved analysis of follicular lymphoma microenvironment using multiplex staining and deep learning

As discussed in Section 1.2.1, one of the distinguishing characteristics of FL from other types of lymphoma is that cancer cells in FL grow in clumps and form follicle-like tissue structures due to its non-diffused nature. The intra- and inter-follicular regions of FL have distinct cellular composition and morphological appearance. However, most previous studies analysed FL as one homogeneous tissue ecosystem. Moreover, there is a lack of literature on the topological organisation of immune cells in these tissue compartments and its association with patients' prognosis. These could be due to a lack of computational pipelines and spatial statistical methods tailored to the tissue architecture of FL. To address these gaps, we have been collaborating with Prof. Teresa Marafioti's lab from University College London and Dr. Giuseppe Gritti from the Ospedale Papa Giovanni, Italy to select immune cell markers and MIF staining of tissue samples. Using MIF spatial imaging, the **DeepMIF** pipeline developed in **Chapter 2** and FL tissue compartments tailored spatial ecological analysis, we aim to answer the following clinical questions.

- Are neoplastic follicle morphological features associated with patients' prognosis?

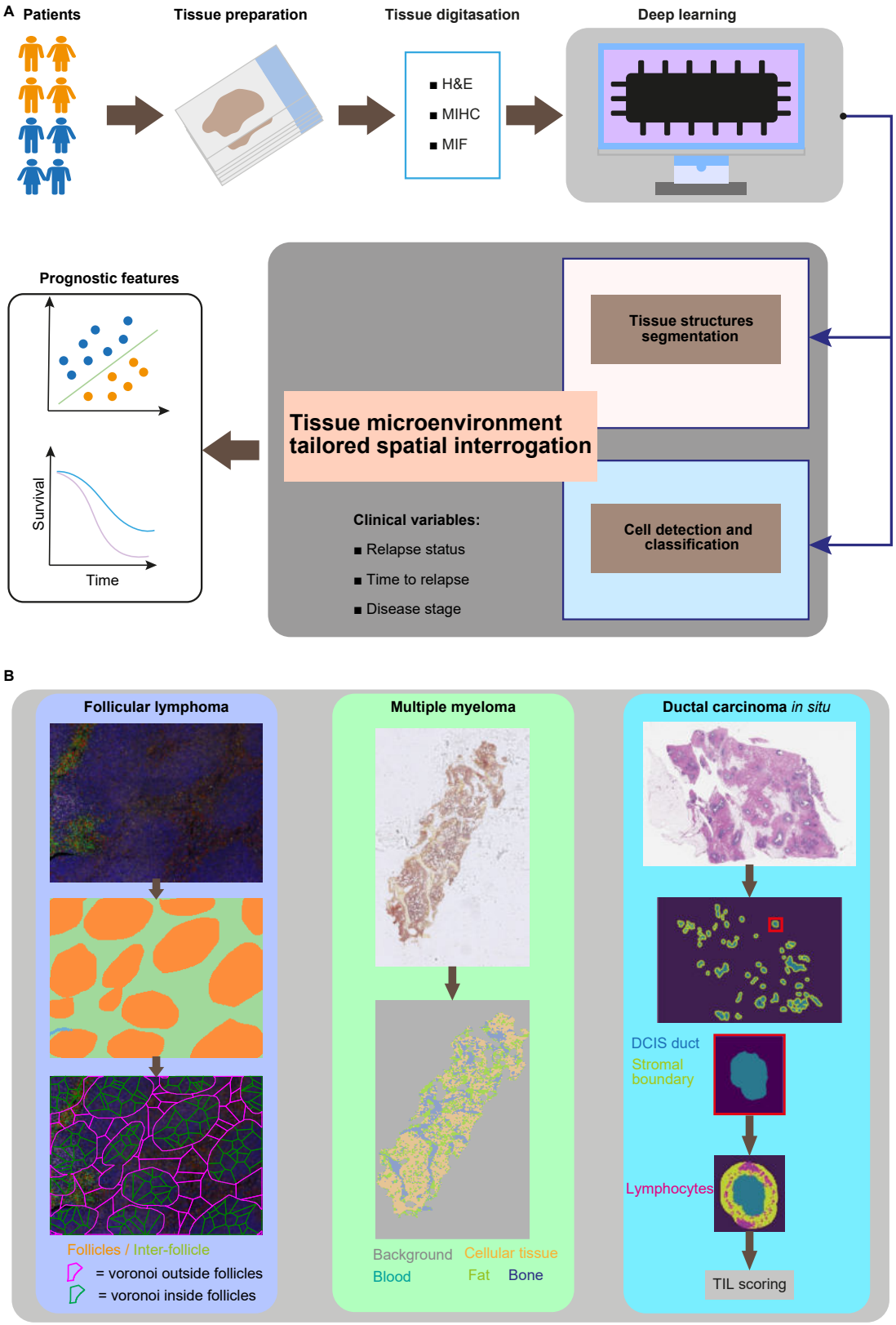


Figure 1.13: **Thesis graphical abstract:**

Figure 1.13: **A)** A diagram showing the process of spatial interrogation of the TME using spatial histology staining technologies and automated computational methods. Under the patients panel, different colours represent patients with different clinical outcomes. First, tissue samples were taken from the suspected organ or tissue of the patients. Then, the tissue section is divided into thin slices and put on glasses for digitisation using different histology staining technologies. The digitised images were processed using deep learning based computational methods developed in this thesis to identify spatial prognostic features. **B)** Tumour types studied in this thesis, along with their distinct tissue microenvironment. Voronois are polygons. DCIS: ductal carcinoma *in situ*; H&E: Hematoxylin and eosin; MIF: multiplex immunofluorescence; MIHC: multiplex immunohistochemistry; and TIL: tumour infiltrating lymphocyte.

- Is there specific spatial immune cell phenotype in the intra- or inter-follicular region of FL which is associated with disease prognosis?
- Is the spatial organisation of cell phenotypes more prognostic than the abundance of cells in FL?

With these questions in mind, in **Chapter 3**, spatial analysis methods tailored to the tissue structure of FL microenvironment were proposed to identify prognostic morphological and spatial features in the intra- and inter-follicular region of FL.

1.7.3 Spatial mapping of bone marrow trephine using deep learning

MM is a multi-stage disease that develops in the BM as described in Section 1.3. The trephine biopsy of BM allows the evaluation of cellular, morphological, and spatial architecture of the BM microenvironment of MM patients. However, as explained in Section 1.3.6, the morphological and spatial microenvironment of BM tissue in MM has been under-explored. This is due to its specialised tissue sampling and processing requirement, the mosaic nature of the tissue landscape and the co-existence of rare and abundant cells in the marrow. These tissue and cell level challenges hinder the development of unbiased automated computational methods to interrogate the dynamic changes in the spatial microenvironment during MM progression and post-treatment. Thus, in this thesis, we aim to address the methodological and translational questions below:

- How to develop deep learning methods that accurately identify rare and abundant cells and dissect the mosaic tissue compartments from MIHC stained WSI?
- Is there a change in the bone trabeculae physiology, and spatial organisation of myeloma plasma cells and immune T cells during the progression of MM?

- Is there a change in the bone trabeculae physiology and spatial organisation of myeloma plasma cells and immune T cells after treatment?

We have been collaborating with Prof. Kwee Yong's lab at University College London to address these central questions. In **Chapter 4**, we developed automated machine learning based image analysis workflows to dissect the mosaic tissue habitats (**MoSaicNet**), a cell imbalance aware deep learning pipeline (**AwareNet**) to enable accurate detection and classification of rare cell types using weighting mechanism. Moreover, we developed methods to interrogate the topological organisation of cells in the BM tissue sections and bone morphology to answer the above questions.

1.7.4 Evaluation of morphological features and spatial immune infiltration patterns as a biomarker for recurrence in ductal carcinoma in situ

In **Chapters 2**, **3** and **4**, we focused on developing machine learning tools to analyse high throughput multiplex histology images and identify prognostic spatial and morphological features using algorithms tailored to the tissue architecture of the tumours. While high throughput multiplex technologies are great tools for exploratory study, their clinical translation is currently limited since the technologies are expensive and not ubiquitous. The H&E staining is cheap, ubiquitous and routinely used for cancer diagnosis. Thus, we extended the idea of spatial interrogation of the TME in relation to tissue structures to DCIS using H&E stained WSIs. As discussed in Section 1.4, stromal TILs are gaining interest in DCIS study and the TIL-WG on breast cancer developed a set of guidelines for manual stromal TILs scoring on H&E images of DCIS [151]. However, manual assessment has inherent limitations such as intra- and inter-observer variability [172] and shortage of experienced pathologists [174]. As suggested by the TIL-WG [152], an automated image analysis pipeline could alleviate these limitations. Thus, in this thesis, we aim to address the methodological and translational questions below:

- How to develop an image analysis pipeline that follows the essential guidelines outlined by the TIL-WG?
- Are stromal TILs in proximity to the DCIS duct more prognostic than distant TILs in predicting recurrence?
- Is there association between DCIS duct morphology, stromal TILs score, and DCIS mutation burden?

To address these challenges, we have been collaborating with Prof. Shelley Hwang's lab from the Duke University School of Medicine, Prof. Hugo M. Horlings from The Netherlands Cancer Institute, Prof. Robert West from Stanford University Medical Center, Prof. Carlo Maley's lab from Arizona State University and Dr. Roberto Salgado from GZA-ZNA Hospitals (Belgium). In **Chapter 5**, we developed an automated image analysis workflow that captures the main concepts laid by the TIL-WG and interrogates the DCIS TME addressing the above questions.

Chapter 6 covers general discussion and conclusions. We discuss the main contributions of this thesis, the limitations and future directions.

Chapter 2

Deep learning based cell profiling for multiplex images

2.1 Overview

Recently, spatial biology has gained a surge of interest in oncology to understand disease biology. This is driven by the advent of high throughput spatial staining technologies such as MIF that allow the examination of multiple proteins on a single tissue section without losing spatial context. The MIF analysis is used to investigate the cellular landscape of tissue sections in terms of abundance and spatial organization of cells. In this process, cell detection and cell phenotype classification are often prerequisites to quantifying cell abundance and exploring spatial heterogeneity of the cellular landscape. However, these tasks are particularly challenging for MIF images due to high levels of variability in staining, expression intensity, intermixing of colours and inherent noise as a result of preprocessing artefacts. The complex makeup of markers in the multispectral images hinders the accurate quantification of cell phenotypes.

Spatial staining technologies such as the Vectra 3 from Akoya Biosciences generates a multiplex image and de-convoluted image (one image per marker). We developed a new deep learning based image analysis pipeline with a graphical user interface (GUI) to detect and classify cell phenotypes on MIF images (DeepMIF) and visualize WSI and cell phenotypes. The cell detection algorithm was first developed on MIHC images and adopted to the de-convoluted images of MIF data. This idea was inspired by the "divide and conquer" algorithm design paradigm. Once the cells are detected on the de-convoluted images, an algorithm that evaluates the co-expression of multiple markers was developed to identify cells expressing single or multiple markers. We trained, tested and validated our model

on $> 50k$ expert single-cell annotations from multiple immune cells panels on 15 samples of follicular lymphoma patients. Our algorithm obtained a cell classification accuracy and AUC ≥ 0.98 on an independent validation panel. The cell phenotype identification took on average 27.5 minutes per WSI, and reconstruction of the WSI from tiles took on average 0.07 minutes. DeepMIF is optimized to run on local computers or high-performance clusters independent of the host platform using docker packaging. These suggest that the DeepMIF is an accurate and efficient tool for the analysis and visualization of MIF images, leading to the identification of novel prognostic cell phenotypes in tumours.

2.2 Introduction

Recent advances in multiplex staining technologies enabled us to study the spatial interaction of cell types in the tumour microenvironment [15, 256, 257]. These technologies allow the detection of multiple proteins at a single cell level on a tissue section at high resolution while preserving their spatial position [15, 258, 259].

For single cell spatial analysis, cell detection and classification are the first key steps [260]. Combined with accurate cell detection and classification techniques, MIF has the potential to allow detailed investigation of cell spatial organization to study tumour spatial heterogeneity [260]. The field of digital pathology has recently witnessed a surge of interest in the application of deep learning in cell detection and classification [175]. However, automated cell detection and classification remain challenging due to variations in slide preparation and cell morphological diversity in shape and size. For example, closely located cells with weak boundaries are often difficult to discern [261, 227, 233, 226]. Moreover, often a parameter such as a kernel size needed to be fixed [261], which cannot cater for cells with a range of sizes and shapes. Furthermore, the need to differentiate cells with a subtle difference in marker expression intensity, adds another layer of complexity to multiplex image analysis.

Moreover, multiplex images come with complexity for analysis due to the intermixing of the markers as a result of co-expression of multiple markers by single cell and weak signals [262–264]. It has been shown that automated machine learning methods such as deep learning excel at objectively identifying cell phenotypes and generating quantitative features from images [259]. These methods can be trained end-to-end to extract features which are robust to the heterogeneity of signal in MIF images [265, 266]. Previously, commercial softwares have been used to analyse MIF images, such as InForm image analysis software [267, 268] and Visiopharm [259]. Other studies have used deep learning to analyse MIF images [259, 269, 270]. In [259, 269], to detect cell nuclei, 4'-6-diamidino-2-phenylindole

(DAPI) DNA staining images were used. However, DAPI staining is not adequate to capture all cell nuclei. Maric et al. [269] proposed to use a combination of DNA staining markers, which is costly.

Here, to address the above stated challenges, we developed a multi-stage workflow that accurately identifies cell phenotypes on MIHC and MIF. This work has the following main contributions:

- We developed Cell COunt RegularizeD Convolutional neural Network (ConCORDe-Net) inspired by Inception-V3 [271] and U-Net [65] architecture which incorporates cell counter regularisation and designed for cell detection in MIHC without the need of pre-specifying parameters such as cell size.
- The model parameters of ConCORDe-Net were optimized using a new objective function that combines conventional Dice overlap and a new cell count loss function which regularizes the network parameters to detect closely located cells and/or weakly stained cells.
- The quantitative experiments support that ConCORDe-Net outperformed the state-of-the-art methods at detecting closely located as well as weakly stained cells.
- We developed an accurate deep learning based computational pipeline to identify cell phenotypes on MIF (DeepMIF) from its de-convoluted images instead of using DAPI that generalizes across multiple panels.
- We developed a whole slide MIF viewer and DeepMIF could be used from the graphical user interface. Thus the algorithm will be widely accessible to researchers with less/no programming skills.
- DeepMIF is easily customizable to allow users to specify cell phenotype of interest in a configuration file or from the graphical user interface.
- DeepMIF could easily run on local computers or high-performance clusters independent of the platform and allows parallelization of tasks to speed up execution.

2.3 Materials

2.3.1 Gal8 dataset

The Gal8 dataset was a pilot study consisting of six ER negative breast cancer samples aimed at assessing the association between galectin-8 (Gal8) positive tumour cells and immune

infiltration in ER negative breast cancer patients. The samples were MIHC stained for CD8, Gal8 and phosphorylated signal transducer and activator of transcription 1 (pSTAT1) (a surrogate marker of immune activation and production). A sample MIHC image from this dataset is shown in Figure 2.1A.

The MIHC were scanned at Prof. Marafioti's lab at University College London. The images were scanned at 40× resolution. The pSTAT1 protein presents a range of expression profiles, and its expression was classified into strong, medium, and weak.

2.3.2 Follicular lymphoma dataset

The follicular lymphoma dataset contains MIF images of 39 patients diagnosed with follicular lymphoma. The patients were diagnosed at Papa Giovanni XXIII Hospital (Bergamo, Italy) with grade I-IIIa FL between 01-Jan-2006 and 31-Dec-2015 and treated with standard R-CHOP or R-CVP.

Multispectral immunofluorescence and staining assessment

Two to four micron thick formalin-fixed paraffin-embedded tissue sections were extracted from patients with FL and normal tonsils and subjected to MIF staining. The MIF staining was carried out by using the Vectra 3 platform (AKOYA Biosciences).

The tissue fixation, staining, and MIF panel optimisation were performed by Dr. Ayse U Akarca from Prof. Marafioti's lab at University College London.

The details of the antibodies used can be found in Table S2.1. Four MIF panels were developed to study specific immune-cell populations. The panels, along with the markers/proteins in each panel, are listed below:

1. **Immune T cells markers:** CD4/CD8/FOXP3/PD1;
2. **Tumour-associated macrophages markers:** CD68/CD163/CD206/PDL1;
3. **Myeloid cells markers:** CD8/CD11b/CD14/CD15;
4. **Natural killer T cells markers:** CD8/Granzyme B (Granz B)/Granulysin/CD16/CD56.

To each panel, 4',6-diamidino-2-phenylindole (DAPI) was added for nuclear counter staining.

Tissue sections were dewaxed, rehydrated and then submitted to the heat-induced antigen retrieval solution either Tris EDTA (tris(hydroxymethyl) aminomethane–ethylenediaminetetraacetic acid) (pH 9.0) or sodium citrate buffer (pH 6.0) on the Leica BOND RXm automated immunostainer (Leica Microsystems, Milton Keynes, UK). Following the pre-treatment, the tyramide signal amplification (TSA)-based Opal staining method was used according to the

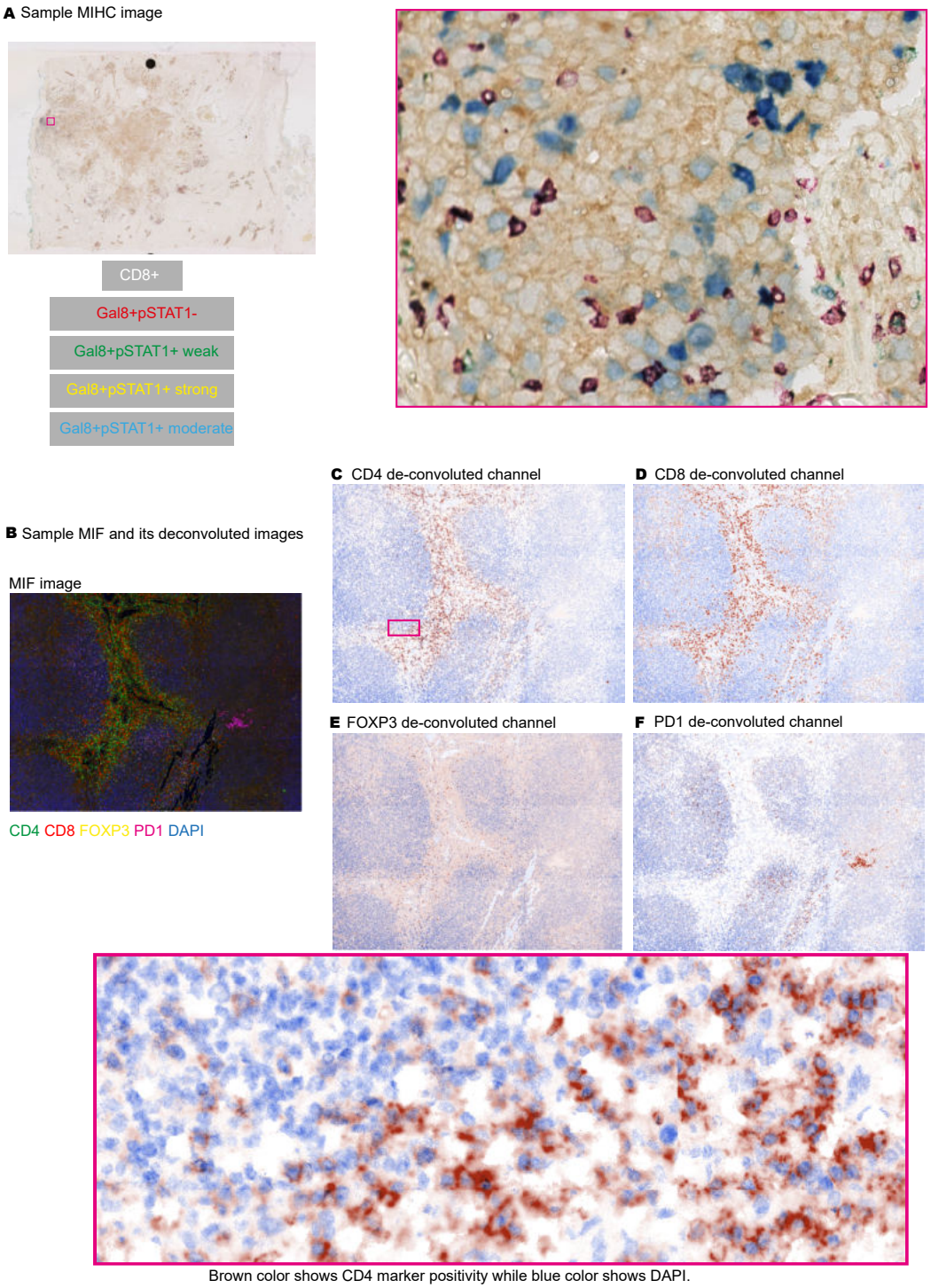


Figure 2.1: **Sample multiplex immunohistochemistry and multiplex immunofluorescence images:**

Figure 2.1: **A)** Sample multiplex immunohistochemistry (MIHC) image taken from our Gal8 dataset. The colours in the image show the positivity of the cells for a specific protein or marker. **B)** Sample multiplex immunofluorescence (MIF) image taken from our follicular lymphoma dataset. The MIF image contains CD4, CD8, FOXP3 and PD1 markers, represented by different colours. **C-D)** De-convoluted images of the multiplex image in (B). The MIF image in (B) shows all markers in one image, while the de-convoluted images show a specific marker positivity status for each cell in the image. The MIF images were scanned using Vectra 3 platform from Akoya Biosciences and the de-convoluted images were generated by the platform. One de-convoluted image was generated for each marker used in the MIF staining. In the de-convoluted images, brown colour represents marker positivity while blue represents 4',6-diamidino-2-phenylindole (DAPI), a nuclear marker.

manufacturer's recommendations (Opal 7-Colour Automation IHC Kit, Akoya Biosciences, Marlborough, MA, USA; Catalogue No. NEL811001KT).

The MIF protocol was optimised for each panel and validated against conventional single immunohistochemistry on sequential tissue sections of human reactive tonsils and two other FL samples (retrieved from the files of the Department of Histopathology, UCLH, London, UK) which were used as lymphoma controls to determine optimal antibody conditions and to establish the appropriate antibody staining order. For detection, each antibody was paired with an individual Opal fluorophore that allowed its visualisation. For molecules expected to co-localise in the same cellular compartment (e.g. cytoplasm, nuclei), each marker was paired with spectrally separated Opals. Those antibodies showing weak protein expression were paired with Opals that would emit a stronger spectral signal. The Opal fluorophores were used at 1 in 50 to 1/200 dilutions. The optimal conditions were set by firstly performing single multispectral fluorescence to QC the consistency/reproducibility of the staining pattern, to check the specificity of the protein expression and to monitor any non-specific signal and/or background. The staining pattern observed for each single MIF was compared and accepted when the results were similar to those obtained with single chromogenic immunostaining.

Image acquisition

Slides were scanned on the Vectra 3 platform at $20\times$ magnification using appropriate exposure times. A sample MIF image from the follicular lymphoma dataset is shown in Figure 2.1B. The Vectra 3 platform provides a protocol to scan either the whole slide or multispectral regions of interest. For whole slide scanning, the platform allows scanning images at $4\times$ or $10\times$ magnification with $2.5\mu m$ or $1\mu m$ pixel resolution, respectively. Under the multispectral regions-based procedure, the platform can acquire images with magnification up to $40\times$ and pixel resolution of $0.25\mu m$. The whole slide MIF scanning is also slower than multispectral region-based approaches. Thus, in this study, we used the latter approach. The regions of

interest, or multispectral regions, were selected by expert pathologists, and the regions were exported in TIF image format. In addition to the MIF images, the Vectra 3 platform generates de-convoluted images. A de-convoluted image contains the expression status of all cells within the region of interest for a given marker or protein used in the MIF panel. Figure 2.1C-F show de-convoluted images for the CD4, CD8, FOXP3, and PD1 markers used in the MIF image depicted in Figure 2.1B.

2.4 Methods

For single cell spatial analysis, cell detection and classification are the first key steps [260]. Multiplex images that allow capturing multiple proteins' expression at a single cell level enable spatial single cell phenotyping. It has been shown that automated machine learning methods such as deep learning excel at objectively identifying cell phenotypes and generating quantitative features from images [259]. Here, we developed ConCORDe-Net, a cell count regularised convolutional neural network to detect single cells from WSI MIHC images, which was specifically designed to address the challenges of detecting weakly stained and touching cells. The ConCORDe-Net architecture was trained and validated using the Gal8 dataset described in Section 2.3.1.

As shown in Figure 2.1, the intensity profile of the de-convoluted images resembles that of the MIHC images. This makes easy to adopt ConCORDe-Net developed for MIHC to the de-convoluted images. Thus, ConCORDe-Net that was initially developed for MIHC was trained on the de-convoluted images from MIF and developed DeepMIf, which allows single cell spatial mapping on MIF images. The DeepMIf pipeline was trained and validated using the Follicular lymphoma dataset, which is described in Section 2.3.2.

The next sections explain the details of ConCORDe-Net and DeepMIF deep learning models.

2.4.1 ConCORDe-Net pipeline

ConCORDe-Net is a deep learning model that detects the position of single cells on MIHC stained WSI. To improve cell detection accuracy, the model uses cell count as a regularisation mechanism. The model was initially developed using the Gal8 dataset (Section 2.3.1). To train and validate the model, single cell annotations were collected as described in the following sections.

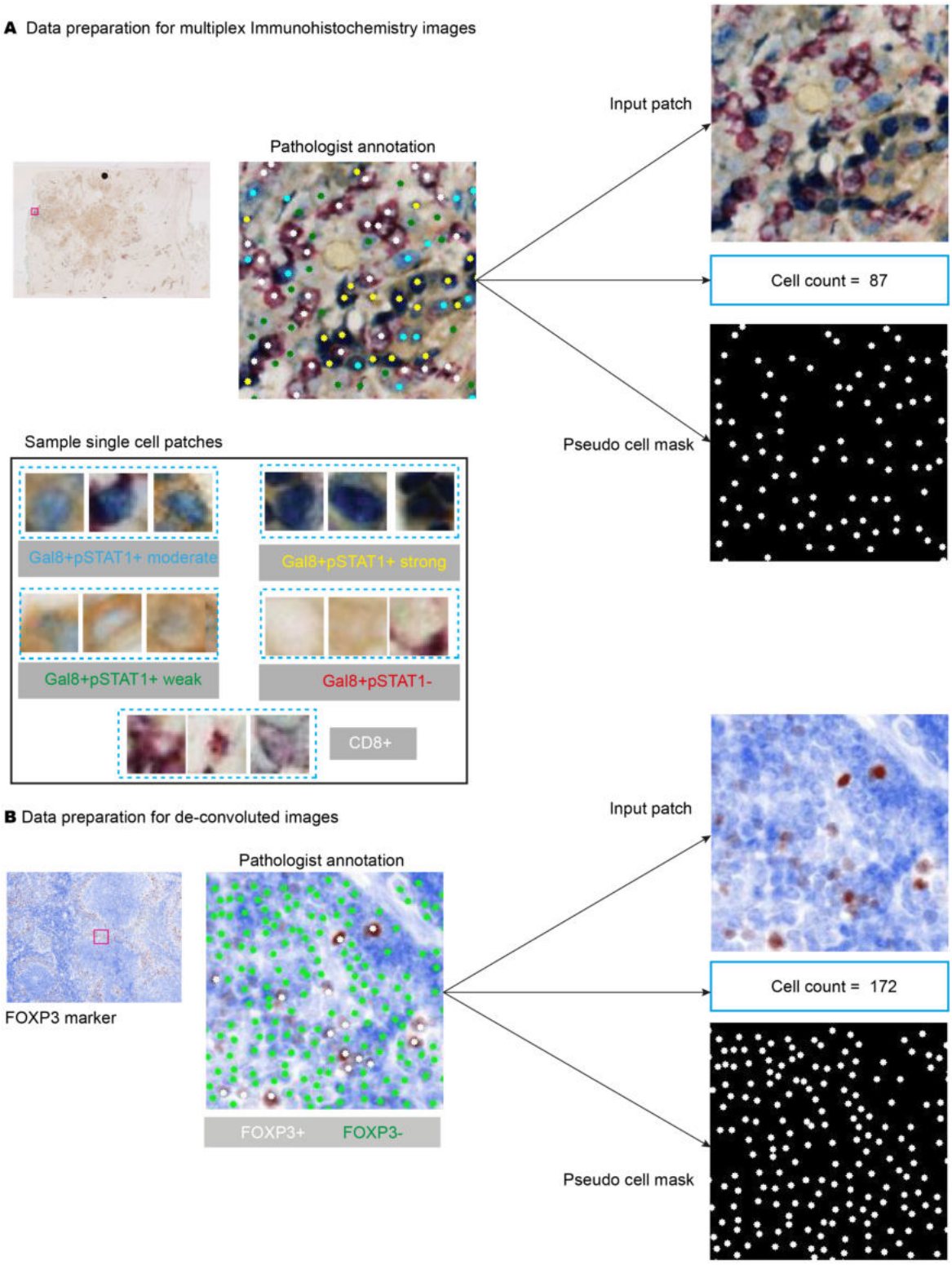


Figure 2.2: **Training data preparation from MIHC and de-convoluted images of MIF images: A)** Sample image showing the preparation of training data for cell detection and classification models from MIHC image.

Figure 2.2: Single cell dot annotation at the centre of the cell nucleus was obtained from a pathologist using a unique colour for each cell type. For cell detection model development, the annotated regions were divided into $224 \times 224 \times 3$ patches. From the dot annotation, a pseudo cell mask is generated with a radius of 4 pixels around the pathologist annotation and the number of cells with the region was also counted. For cell classification, a $28 \times 28 \times 3$ dimensional image centred on the pathologist dot annotation was extracted. **B)** Sample image showing the preparation of training data for cell detection and classification models from de-convoluted images of MIF data. A dot annotation at the centre of the nucleus of cells was collected on the de-convoluted images, using white colour for cells positive for a marker and green for cells negative for a marker. The input patch, pseudo cell mask images and single cell patches were generated as explained in (A).

Gal8 single cell annotation

The Gal8 dataset as described in Section 2.3.1 contains MIHC whole-tumour slide images from patients with breast cancer, and the images were scanned at $40\times$ magnification with $0.227\mu m$ per pixel resolution. To train and validate ConCORDe-Net, a total of 175 regions were annotated from different regions of six whole tumour images by an expert pathologist using a web-based software called Polyscope developed in our lab. To incorporate the variation present in the data, the annotations were collected from different regions of the slides. These annotated regions were extracted from the software, using a separate script. The annotated regions were then randomly split into three categories: training (120), validation (28), and testing (27). Within these regions, a total of $20,477$ cells were annotated, and these belonged to five different types of cells: CD8+, GAL8+pSTAT1-, GAL8+pSTAT1+ strong, GAL8+pSTAT1+ moderate, and GAL8+pSTAT1+ weak. An illustrative example of patches is shown in Figure 2.2A. Moreover, the distribution of the data for each cell is presented in Table 2.1.

Some previous works on cell detection relied on single cell segmentation annotation to detect cells [272]. However, collecting single cell segmentation annotation is laborious. Thus, here, we proposed to use dot annotation at the centre of the nucleus of the cells, which is faster and easier than careful segmentation of cell extent. Then, cell nucleus pseudo segmentation was generated as described in the next Section to train ConCORDe-Net.

Dot annotation to cell pseudo segmentation transformation

The reference ground truth annotation obtained was a dot at the centre of a cell. However, to train the proposed cell detection pipeline, a reference cell mask image (R) and scalar cell count (C_t) were needed as targets. C_t is simply the number of annotated cells in the input patch. The R was generated from dot annotation using Equation (2.1) as cell pseudo

Table 2.1: **Distribution of training, validation and testing Gal8 dataset single cell annotations.**

Cell type	Training	Validation	Test
CD8	2,971	653	624
GAL8+pSTAT1-	4,118	881	903
GAL8+pSTAT1+ strong	919	183	200
GAL8+pSTAT1+ moderate	1,558	295	279
GAL8+pSTAT1+ weak	4,770	1,038	1,102

segmentation by dilating the single cell annotation.

$$R(i,j) = \begin{cases} 1 & \text{if } d < r \\ 0 & otherwise \end{cases} \tag{2.1}$$

where $R(i,j)$ is pixel intensity value at (i,j) of pseudo-segmentation image (R), and d is Euclidean distance between pixel location (i,j) and any of cell dot annotations location. r is distance threshold which was empirically set to 4 pixels. While choosing the value of r, we made sure, the pseudo mask of neighbouring cells does not touch. Figure 2.2A shows a sample annotation image and its corresponding cell pseudo mask image.

Cell Counter from digital images

Cell counting is an essential part of the histopathologic assessment for diagnosis and treatment planning [273, 274]. There have been previously developed deep learning based approaches to counting cells from histopathology images [227, 233, 226, 228, 229] and these could be categorized into two based on their approach. In [227, 233, 226, 228], cell density prediction followed by post-processing image analysis was employed to count cells, while others directly trained CNN to regress cell count from an input patch [229]. Here, similar to Xue et al. [229], we implemented a CNN based network to regress the number of cells and incorporated the cell counter model to regularise cell detection model. The cell counting model was integrated into the cell detection model to improve its performance. The idea of incorporating the cell counter model in the cell detection network was inspired by the idea of providing "hint" while solving a complex computational problem or puzzle. The hint here will be the number of cells in the image.

Figure 2.3A shows an image containing an array of handwritten digits. To find the position of a given number (for example, the number 2) in the image, a question could be asked in these ways:

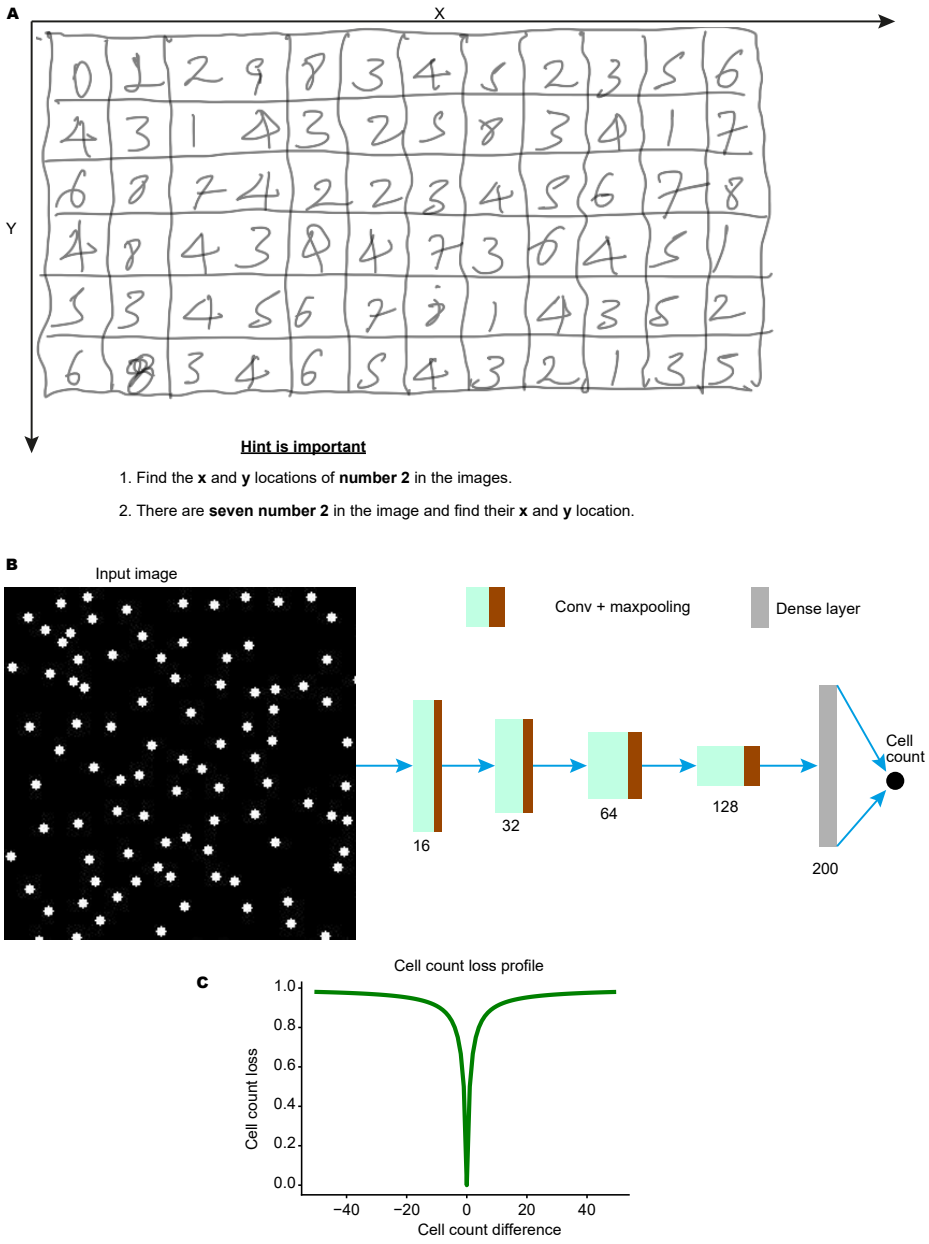


Figure 2.3: **Deep convolutional neural network based cell counter: A)** An image containing handwritten digits showing how providing **hint** improves human performance during problem-solving that inspired incorporating cell count hypothesis in ConCORDe-Net model. **B)** Model architecture. The numbers below the boxes indicate the number of neurons in the layer. **C)** Cell counter model loss function profile.

1. Find the x and y locations of number 2 in the image.

2. There are **seven** number 2 in the image and find their x and y location.

For a human, asking using the second way could improve the accuracy of the answer since a hint is given, there are seven numbers. We hypothesized that incorporating such hint or domain knowledge into the cell count model development could improve the performance and ConCORDe-Net incorporates this idea in the model architecture. To incorporate this concept, a model which can count cells is needed. Thus, a cell counter model was developed. Cell count from binary images such as the cell pseudo mask image in Figure 2.2A could be performed using image processing algorithms such as counting connected component analysis. However, this is not possible especially from the output of model CNN models during training since the prediction is noisy during the initial epochs. Thus, we developed a CNN based cell counter model.

Our proposed cell counter network is shown in Figure 2.3B. It is a CNN architecture that predicts the number of cells in an input image. Mathematically, it can be considered as a mapping function, $f : \mathbb{R}^{k\times k} \rightarrow \mathbb{R}^1$, where k is the size of the input patch, which is 224 in our case.

The cell counter CNN consists of feature extraction and regression parts. The feature extraction section is composed of four consecutive convolutional layers of 3×3 filter size, and *same* padding. The number of neurons in these layers are $\{16, 32, 64, 128\}$ respectively. Every convolutional layer was followed by a max-pooling layer of size (2×2) with stride 2 to reduce the dimensionality of features in the previous layer by half. The regressor part has a series of two dense layers of $\{200, 1\}$ neurons. The output dense layer has one neuron which is the estimated number of cells in the input image. The activation of all convolutional and dense layers was set to rectified linear unit (ReLU).

Parameters of all layers were randomly initialized using uniform Glorot initialization [275]. The parameters were optimised using Adam optimisation [276] using a learning rate of 10^{-4}. Initially, we experimented with mean square error as an objective function. However, as shown in Figure 2.4A, this results in an error that could reach up to 100 during the initial epochs. As we will describe in section 2.4.1, to integrate a cell counter in our cell detection model, which employs Dice overlap loss (bounded between 0 and 1), we intend to develop a new loss function which has loss profile bounded between 0 and 1. Thus, we came up with a new cell count loss (C_l) function in Equation (2.2).

$$C_l = 1 - \frac{1}{1 + \frac{1}{n}\sum_{j=1}^{n} |C_{pj} - C_{tj}|} \tag{2.2}$$

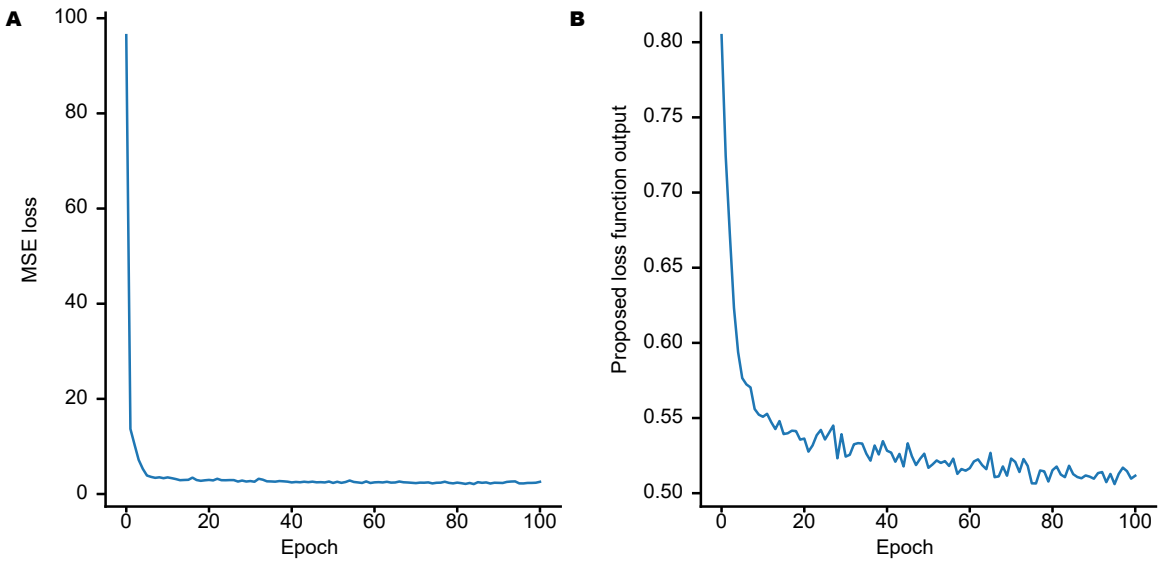


Figure 2.4: **Comparison between training loss history of mean square loss and our proposed loss function:** Line plots of training loss histories of cell count prediction from an input image obtained using mean square error (MSE) loss function (**A**) and using our proposed loss function (**B**). The values of our proposed loss function are bounded between 0 and 1, while MSE loss values could reach up to 100 during the initial epochs. To record both loss histories, the same data was used.

where n is batch size, C_{pj} and C_{tj} are predicted and true number of cells in the j^{th} image, respectively. Figure 2.3C shows the profile of C_l as a function of cell count difference $(C_p - C_t)$. The loss is bounded between 0 and 1 (Figure 2.4B).

Before integrating the cell counter model into the cell detection pipeline, it was trained and evaluated using pseudo-segmentation and the number of cells (Figure 2.2A) as an input and output, respectively. To increase the amount of data, horizontal and vertical flipping was applied to all input training patches. The pseudo-segmentation is a binary image, however, when the cell counter model is integrated with the cell detection model, during model training noisy (especially during a few initial epochs) tensors of floating value will be fed. To realize such an environment, the following morphological and intensity deformation was applied to the cell pseudo mask during cell counter model training;

- Morphological erosion using rectangular structuring element of width $w = 2$ was performed to every patch with a probability $p = 0.4$, where p and w were empirically chosen.

- The images were multiplied by a random matrix of the same size as the image with an empirically chosen probability $p = 0.4$. All elements in the random matrix are in the range $[0.7, 1]$ to set pixel values between 0.7 and 1.

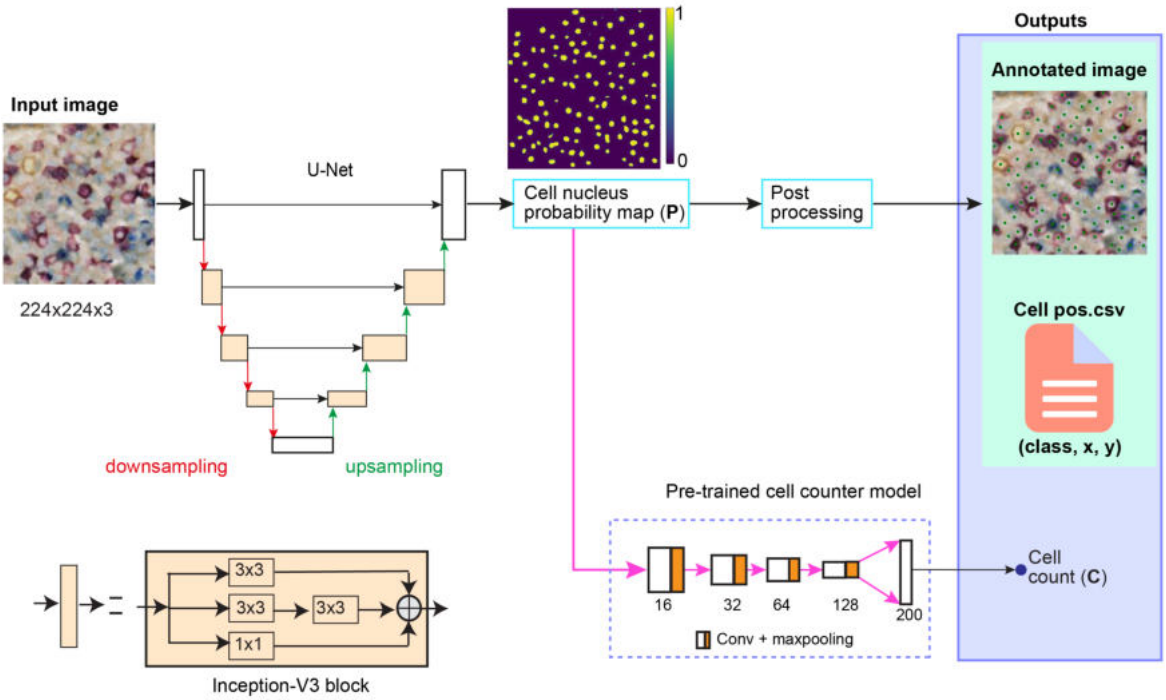


Figure 2.5: **Schematic diagram of ConCORDe-Net architecture**. The network has two outputs, a cell nucleus probability map and a predicted number of cells (C_p). The likelihood map was thresholded using an empirically optimized threshold $T = 0.85$ to convert to a binary mask. The centre of every binary object represents the centre of a cell. In the Inception-V3 module, 3 x 3 and 1x1 indicate filter size of convolutional layers. The numbers in the pre-trained cell counter model indicate the number of neurons.

Cell detection on multiplex immunohistochemistry

Figure 2.5 shows the proposed ConCORDe-Net cell detection convolutional neural network. The input is $224 \times 224 \times 3$ size patch. The network has three parts; an encoder, a decoder and a pre-trained cell counter. The encoder-decoder section is an extended version U-Net [65]. The standard U-Net architecture [65] uses VGG-style in its encoder and decoder section. We have proposed to use the Inception-V3 module instead of the VGG block. The inception module contains multiple parallel kernels of different sizes allowing the model to learn multi-scale features at a given layer. The encoder contains three inception modules and the first two modules were followed by 2D max-pooling layers. The decoder is composed of transposed convolution, concatenation, and inception modules. The 1×1 filter size convolutional layer at the end of the decoder is used to reduce the dimension of the tensor from $224 \times 224 \times 32$ to $224 \times 224 \times 1$. The decoder generates a cell nucleus prediction map (P). The cell nucleus prediction map is connected to the pre-trained cell counter model, which generates the predicted number of cells (C_p). Therefore, the cell detection architecture has two outputs, a

cell location prediction map and a predicted number of cells. Activation of all layers was set to ReLU, but sigmoid for the last layer in the decoder section.

The parameters of the cell counter model were transfer learned from cell pseudo-segmentation as explained in Section 2.4.1. Parameters of the other layers were randomly initialized using uniform Glorot initialization [275], and optimized using Adam [276], a learning rate of 10^{-4} and an objective function shown in Equation 2.3. The cell detection loss (D_l) in Equation 2.3 has two parts. The first part is Dice overlap loss, and the second part is cell count loss.

$$D_l = \left(1 - \frac{2\sum(R \odot P) + \varepsilon}{\sum R + \sum P + \varepsilon}\right) + \lambda \left(1 - \frac{1}{1 + \frac{1}{n}\sum_{j=1}^{n} |C_{pj} - C_{tj}|}\right) \tag{2.3}$$

where R and P in the first part of the equation are the reference image and predicted cell centre probability map image, respectively. Both R and P contain the batch size of images. The $\odot$ operator represents element-wise matrix multiplication. The summation operator adds all the elements of the matrix and generates a single value. The value $\varepsilon = 10^{-5}$ was added to ensure computational stability when the denominator is zero. The second part is the same as Equation (2.2). The value of Dice overlap loss ranges between 0 and 1. Using mean square loss functions for cell count loss could result in a loss value that could reach up to 100 as shown in Figure 2.4A. Thus, this will make the model to focus on cell count but not on the spatial map of the cell detection map. To ensure fair learning from the cell detection overlap and cell count losses, a cell count loss was designed to have a loss profile which ranges between 0 and 1, similar to Dice overlap loss (Figure 2.3C and Figure 2.4B). For cell detection, the spatial location matters more than cell count, and $\lambda = 0.3$ was used to weight the cell count loss. The value of λ was optimized on the validation set with λ value (0, 1] as a search space.

Cell probability map post processing

The model generates a predicted cell nucleus centre probability map image. To convert the probability map to a binary image, we applied a threshold of 0.85. To fill holes in the binary image, we applied morphological closing as follows:

$$I_{out} = (I_{in} \oplus s) \ominus s \tag{2.4}$$

where I_{in}, I_{out}, s, $\oplus$ and $\ominus$ denote the input image, output image, disk structuring element, dilation operator and erosion operator, respectively. The size of the structuring element was set to 5 pixels. We excluded objects of area ≤ 10 $pixel^2$. The centre of every object in the binary image is a centre of a cell. All the threshold parameters above were optimized on the validation dataset maximizing cell detection F1-score.

Cell Classification

The Gal8 dataset contains five types of cells: CD8, GAL8+pSTAT1-, GAL8+pSTAT1+ strong, GAL8+pSTAT1+ moderate, and GAL8+pSTAT1+ weak. GAL8+pSTAT1+ cells were divided based on the expression level of pSTAT1 into strong, moderate, and weak. Discriminating among GAL8+pSTAT1+ cells is challenging, even for experts. Inspired by the principle of the divide and conquer algorithm, we convert the problem into multi-stage classification. The first classifier (**Classifier 1**) differentiates between CD8, GAL8+pSTAT1-, and all GAL8+pSTAT1+ cells. Then, a second classifier (**Classifier 2**) was trained to further divide GAL8+pSTAT1+ cells into GAL8+pSTAT1+ strong, GAL8+pSTAT1+ moderate, and GAL8+pSTAT1+ weak.

Both classifiers were trained using $28 \times 28 \times 3$ patches which can cover the whole cell area for the majority of the cells. A similar network architecture was used for both classifiers.

The classifier has feature extraction and classification sections. The feature extraction part is a modified version of VGG architecture [237] consisting of four convolutional layers of $\{32,\ 64\ 128\ 128\}$ neurons with filters size 3 x 3, stride 1 and *same* padding. Each convolutional layers were followed by 2 x 2 max-pooling layer. The classification layer consisted of two dense layers of $\{200,\ 3\}$ neurons with a dropout layer, $rate = 0.3$ in between. Softmax activation was applied to the last dense layer and ReLU for other layers. Uniform Glorot [276] was applied to initialize model parameters. We used Adam [276] optimizer with a learning rate of 10^{-4}. A categorical cross-entropy objective function was applied. To handle class imbalance, in each mini-batch, an equal number of patches from all cell types were fed to the network, and the number of iterations was determined by the number of patches in the most underestimated class. Moreover, runtime augmentation of flipping, and zooming with scale $s = [0.85\ 1.15]$ was applied with a probability of $p = 0.4$, where s and p were empirically optimized.

2.4.2 DeepMIF pipeline

For cell phenotype identification on MIF images, DeepMIF was proposed. The DeepMIF pipeline has four main sections: cell detection, cell classification, co-expression analysis, and WSI viewer (Figure 2.6). Our MIF images were generated using Vectra 3 platform. The Vectra 3 platform generates MIF and de-convoluted images corresponding to every protein/marker used in the MIF (Figure 2.1). While the MIF image contains complex aggregate features of all markers used, the de-convoluted images have simple features (negative or positive for a marker) Figure 2.1B. Thus, we applied deep learning to the de-convoluted images followed by co-expression analysis to identify cells expressing either single or multiple markers on MIF images (Figure 2.6). Moreover, the de convoluted images look like a single marker immunohistochemistry image. Thus, we adopted ConCORDe-Net for cell detection on the de-convoluted images. In multiplex staining images, colour is one of the main discriminating features to identify cell phenotypes. Since the colours in the MIHC image and de-convoluted images were different (Figure 2.1), new annotations were collected for training and evaluation.

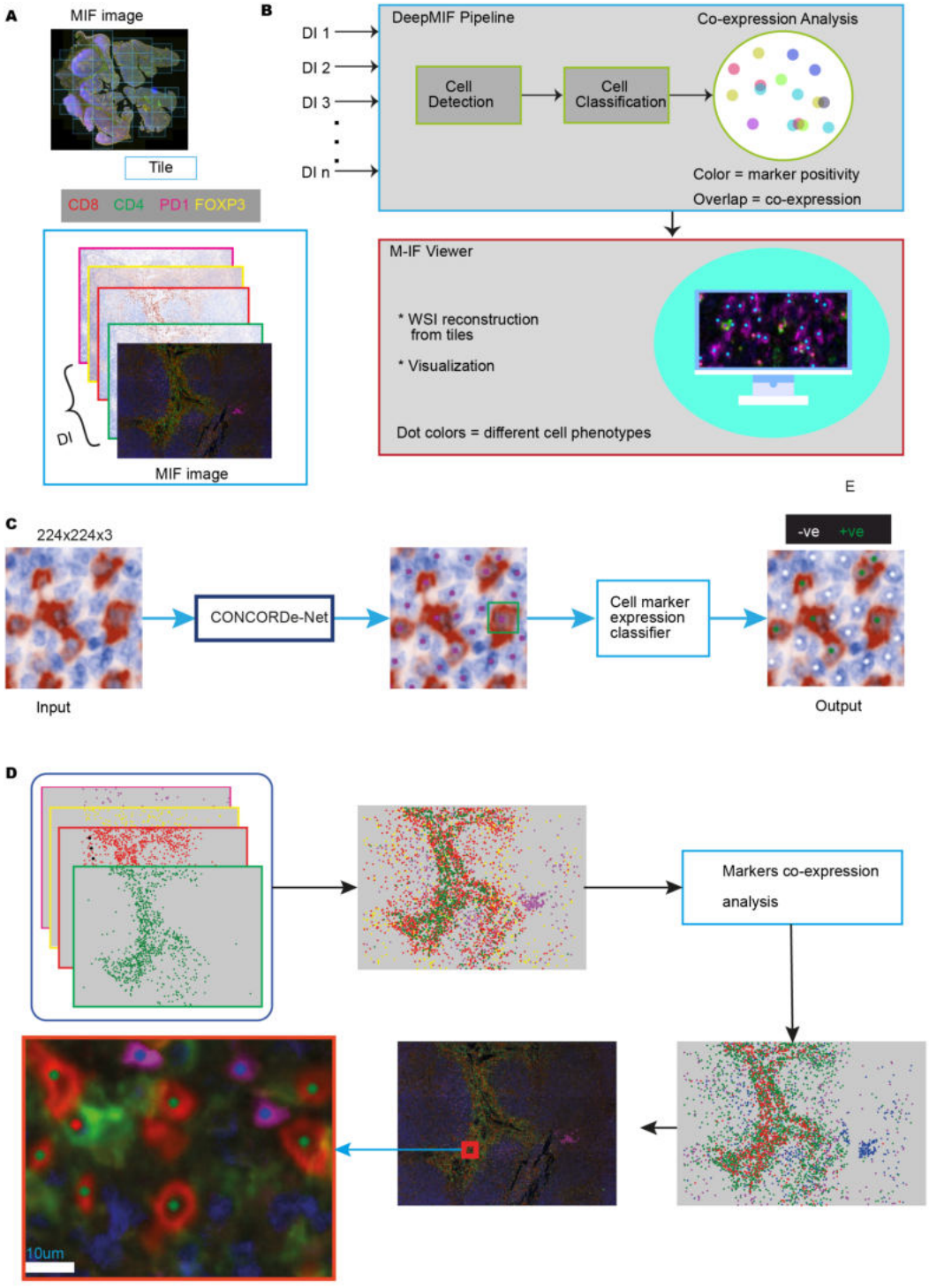


Figure 2.6: **The overview of DeepMIF pipeline: A**) MIF image and its corresponding de-convoluted images. **B)** Overview of DeepMIF pipeline. **C)** Single cell detection and classification on the de-convoluted images.

Figure 2.6: ConCORDe-Net generates the x and y locations of the centre of the nucleus of the cells. A patch centred on these cells' location was extracted and fed to a VGG style CNN based cell classifier (conv = convolution, maxp = maxpooling layer). The classifier generates the probability of the input patch being positive (p_{+ve}) and negative (p_{-ve}) for a marker. The class of the cell was assigned to positive ($p_{+ve} \geq p_{-ve}$) or negative otherwise. This repeats for the *n* de-convoluted images. **D)** Marker co-expression analysis. The positive cells from the *n* de-convoluted images were mapped onto one plane for co-expression analysis. Then, cells were spatially mapped on MIF images. The bar indicates a 10μm resolution. DI = de-convoluted image; WSI = whole slide image

Single cell annotation from de-convoluted images

To train and validate ConCORDe-Net on the de-convoluted images of MIF images, we used diagnostic MIF images of follicular lymphoma patients, as explained in Section 2.3.2. The MIF images were obtained using the Vectra 3 platform at $20\times$ magnification and $0.5\mu m$ per pixel resolution. The output of the Vectra 3 platform contains the MIF image and de-convoluted images (Figure 2.1B). The de-convolution images were obtained from Vectra 3 platform. To optimize model parameters and test the model, $40,327$ single cells were annotated by experts from 10 samples (Table 2.2). The single cell annotations were collected from the de-convoluted images (Figure 2.2B). To capture the tissue heterogeneity, the annotations were collected from different regions of the slides. To make sure data from the same patient is not used for training and testing, the training, validation and testing split was done at patient level 60%, 20%, and 20%, respectively. These $40,327$ single cell annotations were collected from the immune T cells panel. To evaluate the generalizability of the model to other panels, we validated the model on $10,038$ cells collected from two independent MIF panels from 5 samples (Table 2.2).

Table 2.2: **Distribution of training, validation and testing single cell annotation dataset from de-convoluted images of MIF images.** The model optimization data was extracted from the immune T cells panel. Annotations were collected from de-convoluted images of non-nuclear (CD8, CD4, PD1, CD16 and CD206) and nuclear (FOXP3 and Granulysin) markers. natural killer T = Natural killer T cells panel

	Model optimizing data			**Model validation data**	
	Training	**Validation**	**Testing**	**natural killer T**	**Macrophages**
Positive	5,088	2,147	1,287	4,021	239
Negative	16,000	6,958	8,847	5,188	590
De-convoluted images	CD8, CD4, FOXP3 and PD1			CD16 and Granulysin	CD206

Cell detection on de-convoluted images

To detect cells on the de-convoluted images, we used ConCORDe-Net explained in Section 2.4.1 since the model was designed to give attention to weakly signal markers and touching cells, one of the challenges in MIF due to co-expression of multiple markers. The model input is a $224 \times 224 \times 3$ pixels image. To discern weakly stained cells and touching cells, the model uses cell count in the training data. Here, the model was trained from scratch using human annotations (Table 2.2).

Cell classification on de-convoluted images

To predict cell marker positivity on de-convoluted images, separate models were trained for nuclear and non-nuclear markers. The number of single-cell annotations collected from non-nuclear markers was much higher than that of nuclear markers. To minimize the effect of the imbalance, we developed separate CNN of the same architecture for nuclear markers and non-nuclear markers.

Patches of size $20 \times 20 \times 3$ and $28 \times 28 \times 3$ pixels were extracted for nuclear (e.g FOXP3) and non-nuclear (e.g CD4) markers, respectively. We used smaller patch sizes for nuclear markers to minimize the effect of background noise.

InceptionV3 [277] and VGG [278] are among the most commonly used classification CNNs. These architectures are deep and have a large number of parameters, which could result in overfitting for a small dataset. Thus, we custom designed shallower versions of these CNN with a depth of 5 layers. The first model (**Our model 1**) uses the Inception module for the feature learning section, while the second model (**Our model 2**) uses the VGG module. The Inception module uses multiple kernel sizes at a given layer to extract multi-scale features, followed by a max-pooling layer, while the VGG module uses a series of two convolutions of the same kernel size, followed by a max-pooling layer.

The feature learning section starts with a convolution layer of 16 neurons, and the number of neurons increases by 16 for every layer added. The classification section consists of two dense layers of {200, 2$ neurons, with a dropout $(rate = 0.3)$ layer in between. The ReLU activation was used in all layers, but softmax in the last layer to generate a probability. Model parameters were randomly initialized using uniform Glorot [275] and optimized using Adam [276], learning rate of 10^{-4} and binary cross-entropy loss function. The model was trained for 500 epochs with $patience = 50$ epochs.

The deep learning-based cell detection and classification enabled us to spatially map cells positive for all markers on the deconvoluted images. We then applied co-expression analysis to identify cells expressing a single marker or co-expressing multiple markers in the MIF images (Figure 2.6).

Markers co-expression analysis

To identify cells co-expressing multiple markers, we first spatially mapped the location of cells positive for the markers onto a single plane (Figure 2.6D). A cell is said to be co-expressing multiple markers if detections on de-convoluted images overlap after mapping them onto a single plane. The

cell detection algorithm is optimized to find the centre of the nucleus of cells. However, due to the variation in the nature of markers (nuclear or cytoplasmic), the predicted location might slightly vary. We empirically set threshold distance $r = 1.5\mu m$ (about a quarter of lymphocyte diameter). In the rest of the paper, overlapping markers mean within a distance of r. Co-expression analysis was performed as follows.

Suppose we have MIF panel with n markers $\{m_i : \ i \ \in \{1,2,...n\}\}$, which will generate n de-convoluted images. Suppose, we want to identify cells $(C_{phenotype})$ co-expressing k markers, $M_p = \{m_i : \ i \in \{1,2,...,k\}, \ k \leq n\}$ and negative for l markers $M_n = \{m_j : \ j \in \{1,2,...,l\}, \ l < n\}$. Distance can only be computed between two points. So, k points are said to be overlapping with each other if every pair of points is overlapping. The number of combinations, (N), is computed as follows:

$$\mathrm{N} = {}^{k}C_2 = \frac{k!}{2!(k-2)!} \quad = \frac{k(k-1)}{2} \tag{2.5}$$

The combinations are $\{(m_1,m_2), \ (m_1,m_3), \ (m_1,m_4), \ ..., (m_{k-1},m_k)\}$. The complexity of iterations is $\mathcal{O}(k^2)$. To speed up the computation, we used vectorized forms instead of single cell level looping. The co-expression analysis have N iterations, each with 4 main steps.

Let $CoExp$ is a dictionary of length k with marker name and the location of cells co-expressing the k markers as a key and value, respectively. Initially, the values are set to empty and updated as follows:

Iteration 1:$(m_1, \ m_2)$ combination.

Step 1: Get m_1 and m_2 positive location. Suppose the de-convoluted images for marker m_1 has n cells positive for m_1, $U = \{u_i \ \in \mathbb{R}^2, i \ \varepsilon \ \{1,2...n\}\}$ and the de-convoluted images for m_2 has a m cells positive for m_2, $V = \{v_i \in \mathbb{R}^2, i \ \varepsilon \ \{1,2...m\}\}$. The u_i and v_i have its own (x,y) location in the image space.

Step 2: Compute distance matrix, $D \ \in \ \mathbb{R}^{n \times m}$. The distance is defined as follows:

$$D_{ij} = \|u_i - v_j\|_2 = \sqrt{(x_{u_i} - x_{v_j})^2 + (y_{u_i} - y_{v_j})^2} \tag{2.6}$$

Step 3: Identify cells $Q \subseteq U$ and $P \subseteq V$ which co-express m_1 and m_2. The Q and P are the subset of the original cell collection in which the items in i^{th} location of Q and P are overlapping. Mathematically,

$$\underset{Q \in U, P \in V}{\arg\min} \, D(U, \ V) \ : \ \|q_i - p_i\|_2 \ \leq \ r, \ for \ q_i \in Q \ and \ p_i \in P \tag{2.7}$$

Step 4: Update $CoExp$ for m_1 and m_2, i.e, $CoExp \ [m_1] = Q$ and $CoExp \ [m_2] = P$. The Q and P contain a list of the (x,y) location of cells positive for marker m_1 and m_2, respectively.

Steps 1 to 4 will be subsequently applied for the other combinations. If a marker in the current combination was considered in the previous combination, cells positive for that marker which co-express all previous markers will be only considered. For example, for the second combination,

(m_1, m_3), Q will be used instead of the original collection U for m_1. Again from (m_1, m_3) combination, a subset of Q will co-express m_3. Then, $CoExp$ will be updated with values for m_1 and m_3. Remove cells form $CoExp$ for m_1, that did not express m_3 and their corresponding m_2 expressing cells. If any of the combinations do not have overlapping markers, the iteration stops and there are no cells that co-express the k markers.

After identifying the cells co-expressing the k markers based on the marker status on the deconvoluted images, we computed the centre position of the cells co-expressing the markers. The centre location of the j^{th} cells in $CoExp$ was computed using Equation (2.8).

$$(x_j, y_j) = \left(\frac{\sum_{i=1}^{i=k} x_{j,i}}{k}, \frac{\sum_{i=1}^{i=k} y_{j,i}}{k} \right) \tag{2.8}$$

where the $(x_{j,i}, y_{ji})$ is the predicted (x,y) location of the j^{th} cell in $CoExp$ for the i^{th} marker. Finally, $C_{phenotype}$ is a subset of C that doesn't express any of the l markers in M_n. Mathematically,

$$C_{phenotype} = \{c_i \in C,\ i \in \{1,2,...,n\}\} : \|c_i - z_j\|_2 > r | \forall z_j \in Z \tag{2.9}$$

where Z is a set of locations positive for markers in M_n.

DeepMIF graphical user interface

To make DeepMIF easily usable and interactive for pathologists and the wider research community, we developed a GUI. The GUI has two main components: the whole slide MIF image viewer and the deep learning pipeline (Figure 2.7 and Figure S1.1). The viewer reconstructs the WSI from its tiles and displays cell phenotypes using the location of tiles in their filename and position of the cells generated using DeepMIF (Figure 2.6A, C, and Figure S1.1). The image viewer was developed using an OS-independent PyQT Python package. It is interactive and allows batch processing of files. To allow multi-tasking of rendering, visualization, and running the deep learning pipeline in parallel, threading and multiprocessing are employed.

2.4.3 Model performance evaluation metrics

In machine learning, model selection is based on the performance evaluation on validation data. For cell detection and classification, the most commonly used metrics include accuracy, precision, recall, F1-score, and AUC (for classification). For a binary classifier, give a randomly selected negative (0) and a positive (1), AUC shows the probability that the model prediction for the positive sample will be higher than the negative sample. Considering a binary classifier (positive vs negative), here are terms used in the computation of these metrics:

- True positive (TP): actual value (positive) and model prediction (positive)

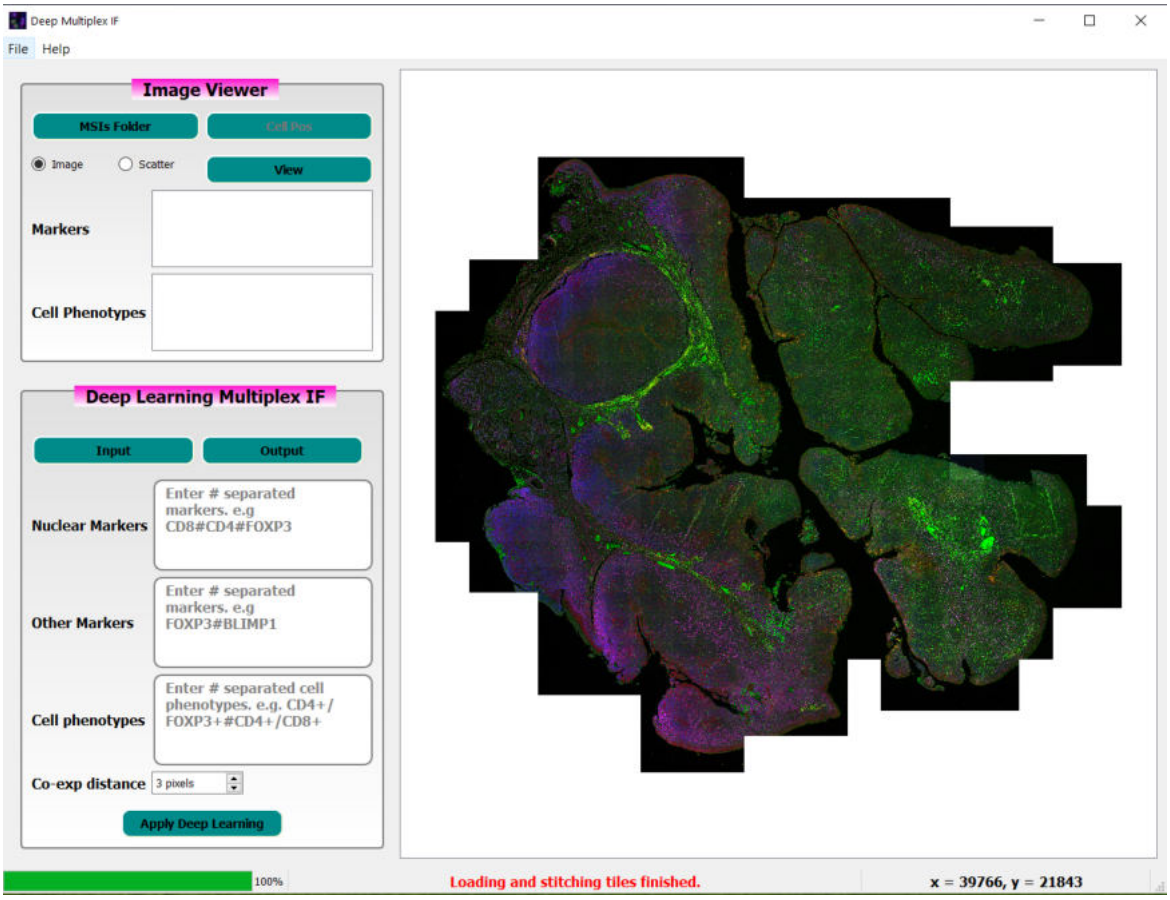


Figure 2.7: **The DeepMIF pipeline graphical user interface.** The graphical user interface has two main components: image viewer and deep learning based spatial cell phenotyping component. The image viewer can be used to visualise MIF images or cell phenotypes as scatter plots. The viewer also displays the markers and cell phenotypes. To use the DeepMIF pipeline from the graphical user interface, a user needs to input the input folder which contains the images, and the output directory using their corresponding buttons. The user should input also the nuclear markers and non-nuclear markers in the MIF panel, cell phenotypes of interest. The co-expression distance is a distance in pixels in order to consider to markers are overlapping.

- False negative (FN): actual value (positive) and model prediction (negative)
- False positive (FP): actual value (negative) and model prediction (positive)
- True negative (TN): actual value (negative) and model prediction (negative)

The formula to compute accuracy, precision, recall and F1-score is shown below.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{2.10}$$

$$Precision = \frac{TP}{TP+FP} \tag{2.11}$$

$$Recall = \frac{TP}{TP+FN} \tag{2.12}$$

$$F1-score = 2\,\frac{Precision*Recall}{Precision+Recall} \tag{2.13}$$

2.4.4 Implementation and code availability

Both ConCORDe-Net and DeepMIF were implemented in Python. Tensorflow 2.0 [279] was used for the development of the deep learning pipeline. The graphical user interface was developed using the PyQT5 Python package. The implementation of CONCORDe-Net and DeepMIF are available in GitHub.

2.5 Results and discussion

2.5.1 ConCORDe-Net performance evaluation on Gal8 dataset

Deep convolutional network accurately predicts number cells in an image

To investigate if CNN can regress the number of cells from an input image, the proposed cell counter model was trained and evaluated on the test cell pseudo-segmentation images before integrating to ConCORDe-Net. Pearson correlation $r = 0.999$ was obtained between the true and predicted number of cells (Figure 2.8). The high correlation supports that the proposed cell counter network could be reliably used as a cell count approximation function.

ConCORDe-Net outperforms state-of-the-art cell detection methods on MIHC images

Quantitatively, I evaluated ConCORDe-Net using standard metrics: precision, recall and F1-score. A detection was considered true positive if it lies within a Euclidean distance of 8 pixels to a ground truth annotation. Moreover, I compared ConCORDe-Net with state-of-the-art methods, MapDe [261]

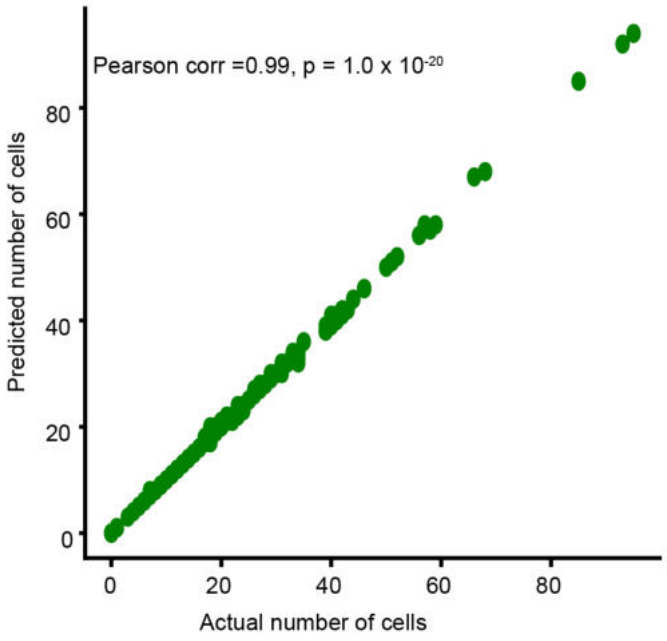


Figure 2.8: **Cell counter performance evaluation.**

and U-Net [65] as shown in Table 2.3. These models were trained from scratch using the data as ConCORDe-Net.

ConCORDe-Net achieved the highest recall and F1-score compared to state of the art methods, MapDe [261] and U-Net [65] Table 2.3. MapDe [261] was proposed for H&E stained images of patients with lung cancer, and I used parameters that were specified in the paper and re-trained from scratch. The precision of ConCORDe-Net was lower than the three other methods due to the following reasons since ConCORDe-Net identifies weakly stained cells that were missed by other methods, which could be missed by an expert too.

Figure 2.9 shows a visual output of ConCORDe-Net followed by cell classification and comparison with MapDe [261] and U-Net [65] which uses Dice overlap loss as an objective function. ConCORDe-Net is better in discerning touching cells with weak boundary gradient and weakly stained GAL8+ pSTAT1- cells compared to MapDe [261] and U-Net [65].

Cell classifier performance evaluation

To visualize the separation of the different classes, 128 dimensional deep learnt features were extracted from the classifier and visualized in 2 dimensions after t-SNE dimensionality reduction. Both features learnt by both **Classifier 1** and **Classifier 2** have shown the separation of classes Figure 2.10 A, B.

Then, the performance of both classifier models was qualitatively evaluated using receiver operating characteristic curve (ROC), the AUC, accuracy, precision, recall, and F1-score on a separately

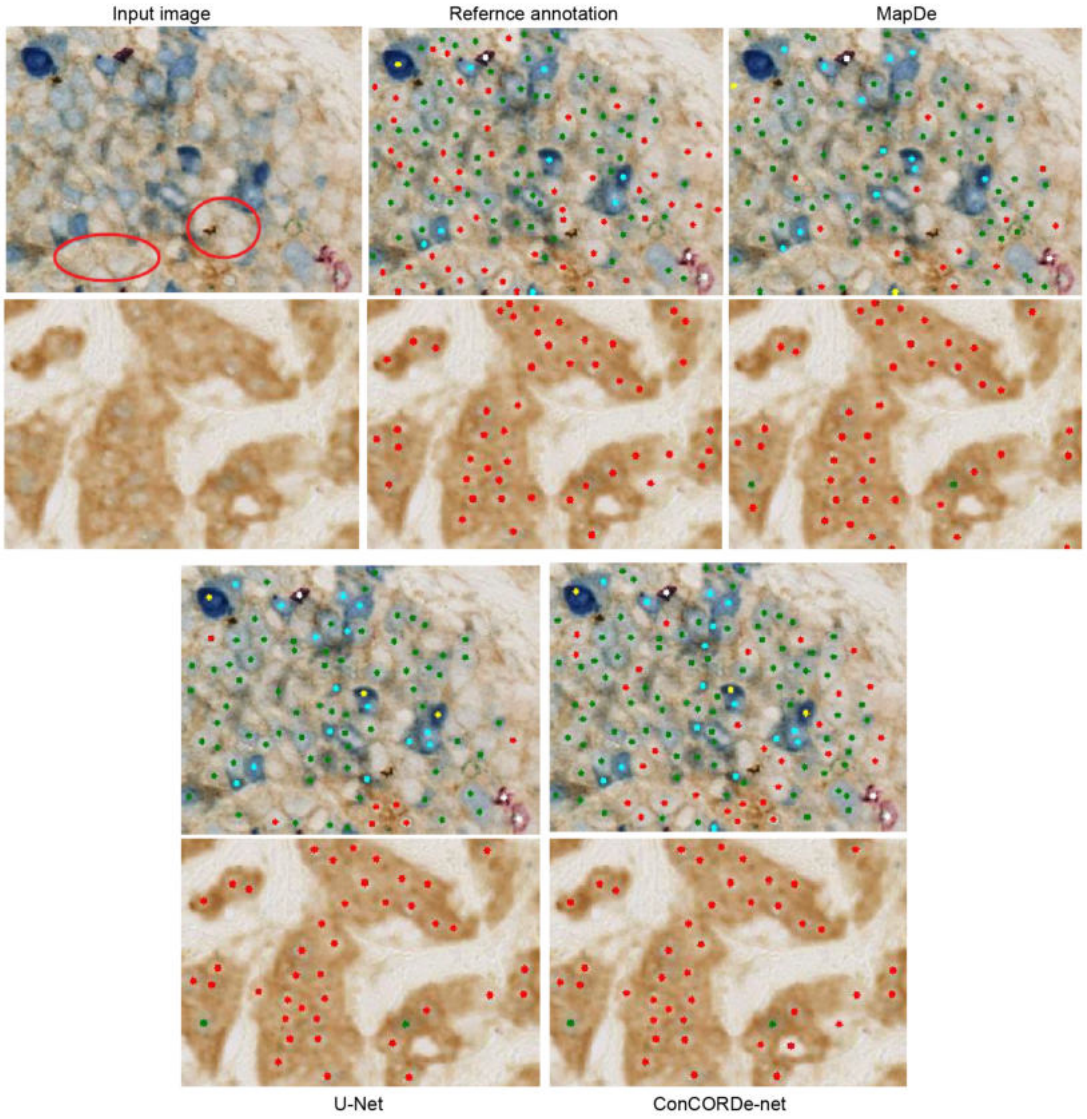


Figure 2.9: **Visual images comparing the performance of ConCORDe-Net and state-of-the-art methods.** Illustrative examples of the proposed unified cell detection and classification on test data. and comparison with state-of-the-art method, MapDe [261] and U-Net [65]. White, red, yellow, cyan and dark green coloured points represent CD8, GAL8+ pSTAT1-, GAL8+ pSTAT1+ strong, GAL8+ pSTAT1+ moderate, and GAL8+ pSTAT1+ weak cells, respectively. The red circles on the top left input images highlight cells that were missed by MapDe [261] and U-Net [65], but detected using ConCORDe-Net.

Table 2.3: **Cell detection performance comparison of different models on the Gal8 dataset.** Model_1 stands for ConCORDe-Net after cell counter removed.

Method	Precision	Recall	F1-score
ConCORDe-Net	0.854	**0.892**	**0.873**
U-Net [65] + Cell Counter	0.872	0.837	0.854
Model_1	**0.908**	0.80	0.845
U-Net [65]	**0.908**	0.785	0.841
MapDe [261]	0.804	0.876	0.838

held test data shown in Table 2.1. The ROC and AUC of **Classifier 1** are presented in Figure 2.10C and AUC value of greater than 0.99 was achieved for all cell types. For Classifier 1, the overall accuracy computed on the original distribution of data was 98.1%. Figure 2.10D shows ROC and AUC of this **Classifier 2**. For all cell types, AUC value was higher than 0.97 and overall accuracy of 93% was obtained. The accuracy of Classifier 2 is lower than Classifier 1 (Figure 2.10C, D), due to the intrinsic challenge in the data. All cells input to Classifier 2 expresses pSTAT1 at different levels (weak, moderate and strong). There is no clear threshold for separating these classes even for pathologists. Thus, there could be bias in the annotation data which could confuse the model to decide the class of the cell. After cascading the two classifiers, overall accuracy of 96.5%, and precision, recall and F1-score of 0.98 was achieved. These show the proposed cell classification models were able to identify the marker status of the cells with high accuracy.

The limitation of the Gal8 dataset is that it contains a small number of patients/samples. The dataset contains a pilot study of 6 patients and a limited number of human annotations were used to train and validate ConCORDe-Net and classifier models. ConCORDe-Net was trained on a large amount of human annotation data and was extended beyond MIHC images to detect cells on MIF images.

2.5.2 DeepMIF performance evaluation: cell detection and classification

To enable the automated detection and classification of diverse cell types in MIF images, we first applied ConCORDe-Net on the de-convoluted images. The number of cells detected by the proposed deep learning method significantly correlated with the cells annotated by the expert pathologists (Spearman $r = 0.94, p = 1.82x10^{-12}$ (Figure 2.11A). Moreover, to evaluate the performance of ConCORDe-Net single-cell detection on the de-convoluted images, we used recall, precision and F1-score. On separately held test data, we obtained recall, precision, and F1-score of 0.85, 0.86, and 0.86, respectively.

To evaluate the performance of cell classifier models, we used precision, recall, accuracy and AUC. The performance of our proposed models (Our model 1 and Our model2) was compared with ImageNet pre-trained models including VGG16, InceptionV3, ResNet50 and Xception models. These

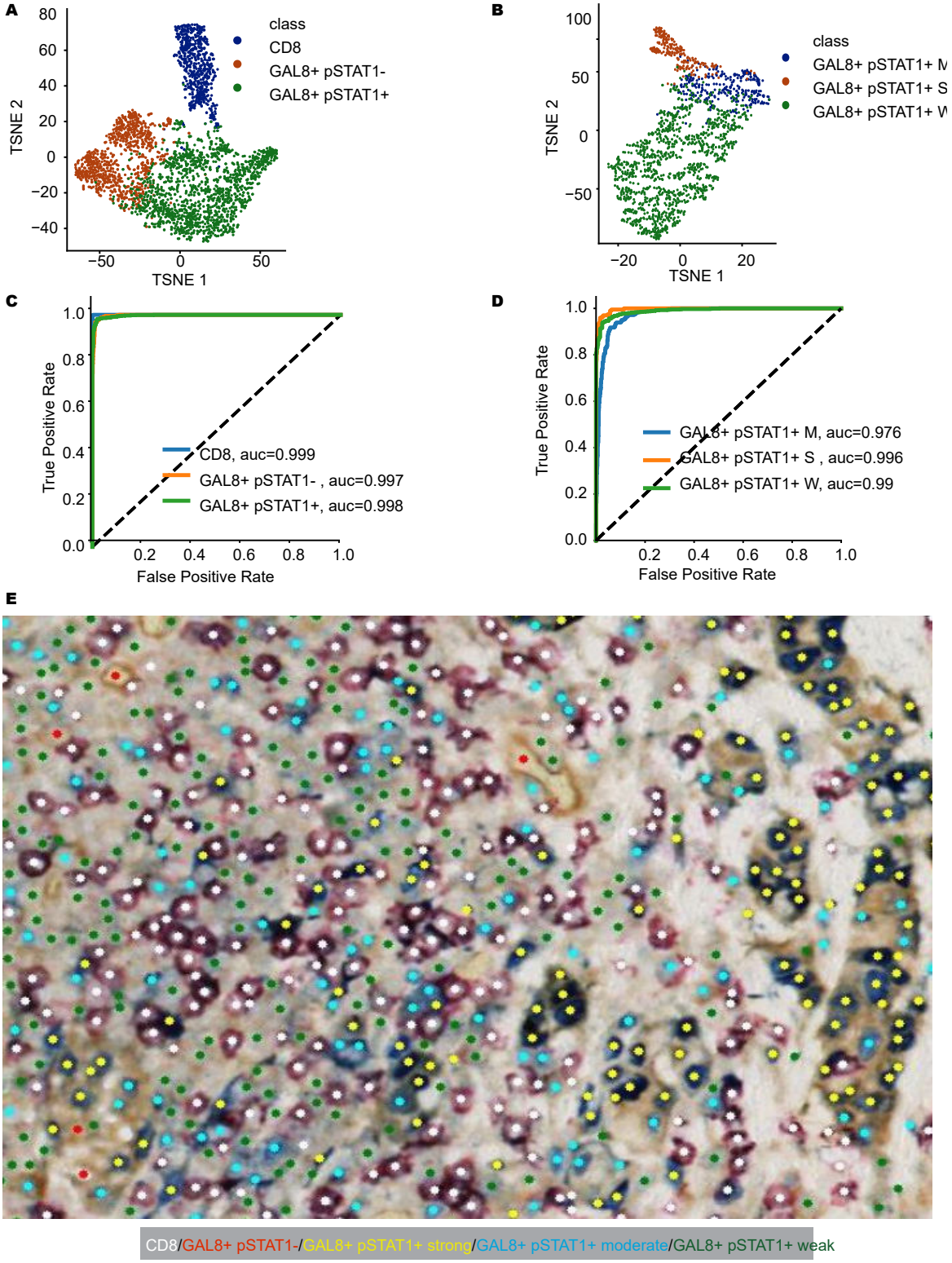


Figure 2.10: **Cell classification models performance evaluation:** t-SNE feature visualization of Classifier1 (**A**) and Classifier2 (**B**) learned features. A 128 dimensional features vectors were extracted from the output of the first dense layer of the classifiers and reduced to a two-dimensional vector using t-SNE. ROC and AUC of classifier1 (**C**) and classifier2 (**D**). **E**. Sample patch showing DL spatial mapping of single cells. S=Strong, W=Weak, M=Moderate.

models were fine-tuned using our single cell annotation. All models were trained on the same data from the immune T cells panel (Table 2.2).

On a separately held $> 10k$ test data (Table 2.2), **Our model 2** achieved recall (0.96), the highest among all models (Figure 2.11B) and precision of 0.96, same as **Our model 1**, and VGG16. Using 1000 bootstraps taking 60% of the test data in each bootstrap, the recall estimate of VGG16 and **Our model 2** was 0.96, 95% CI ($0.95 - 0.96$), and 0.965, 95% CI ($0.96 - 0.97$), respectively. These models achieved the same value of AUC (0.96), compared to Xception (0.91), InceptionV3 (0.89) and ResNet50 (0.94) (Figure S1.2A). **Our model 2** and **Our model 1** achieved 0.98 accuracy (Figure S1.2B), higher than ImageNet pre-trained models. Overall, **Our model 2** outperformed all the other models. This could be due to the less number of parameters (Figure 2.11B) compared to the other models and thus, less chance of over-fitting. Moreover, due to the reduced number of parameters in **Our model 2**, it takes less time and memory during training and inference.

Moreover, visualization of the features learned by the CNN using uniform manifold approximation and projection (UMAP) dimensionality reduction demonstrated that cells of different classes are separated (Figure 2.11C).

Subsequently, we validated **Our model 2** on single cell annotation collected from panels and markers not seen during model optimisation. On 10038 single cell annotations collected from natural killer T cells and macrophages panels data, **Our model 2** achieved AUC, precision, recall and accuracy values ≥ 0.98 (Figure 2.11B, D and Figure S1.2C-D). This shows **Our model 2** is generalizable to other panels and it could be reliably used to classify cells on de-convoluted images.

2.5.3 MIF markers co-expression analysis

The co-expression algorithm developed in Section 2.4.2 was tested on simulated data before applying it to real MIF images (Figure 2.12A). A set of points in a 2D plane representing marker positivity with semi-random (x,y) location were generated for $k = 4$ markers. During the simulation, some points were set to overlap. On this simulated data, our co-expression algorithm was able to locate overlapping markers (open circles).

Then, the algorithm was applied on real MIF data (Figure S1.3A, B). Although the model was trained on images from immune T cell panel only, DeepMIF was able to accurately identify cell phenotype on natural killer T cell panel images which have completely different markers and cell phenotypes (Figure 2.12B, C). This shows DeepMIF pipeline is generalizable to multiple panels.

Though, MIF is a high-throughput approach to characterise immune phenotypes landscape in tissue sections, the intermixing of colours deters accurate identification of cells and discernment of touching cells. In a computerized analysis of MIF, colour (intensity) is the main discriminating feature between different cell phenotypes. Multiple panels could have different colours, and a supervised model trained on MIF data from one panel might not be generalized to another. However, irrespective of the number of markers/colours used in the MIF panels, the de-convoluted images in all panels have only brown (positive) and blue (DAPI, negative) colours. This suggests that a model trained on

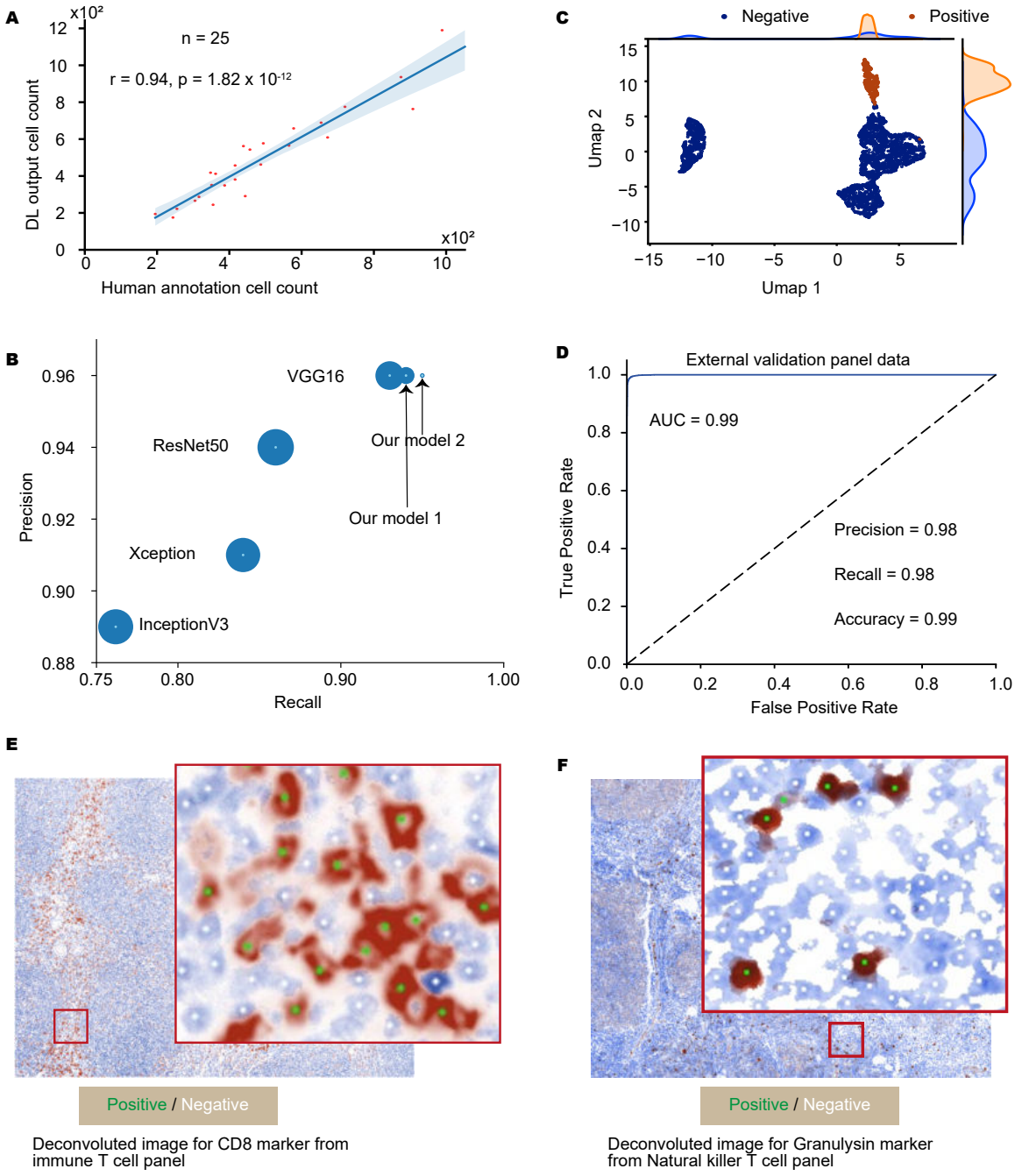


Figure 2.11: **DeepMIF cell detection and classification models performance evaluation: A)** Correlation between the number of cells annotated by an expert and the number of cells detected by deep learning. A dot represents a human-annotated rectangular region, which contains around 450 cells. **B)** Scatter plot showing the comparison of precision and recall values for the ImageNet pre-trained model and our models. size=number of parameters. **C)** Two-dimensional representation of deep learned features after Uniform manifold approximation and projection (UMAP) dimensionality reduction along with their marginal distributions. Negative and positive classes represent cells negative and positive for a marker, respectively. **D)** Performance evaluation of **Our model 2** on external validation panel. **E-F)** Illustrative images showing cell detection and classification results from the immune T cell panel (the panel used for model development) (E) and from the natural killer T cells panel (used for as external validation) (F).

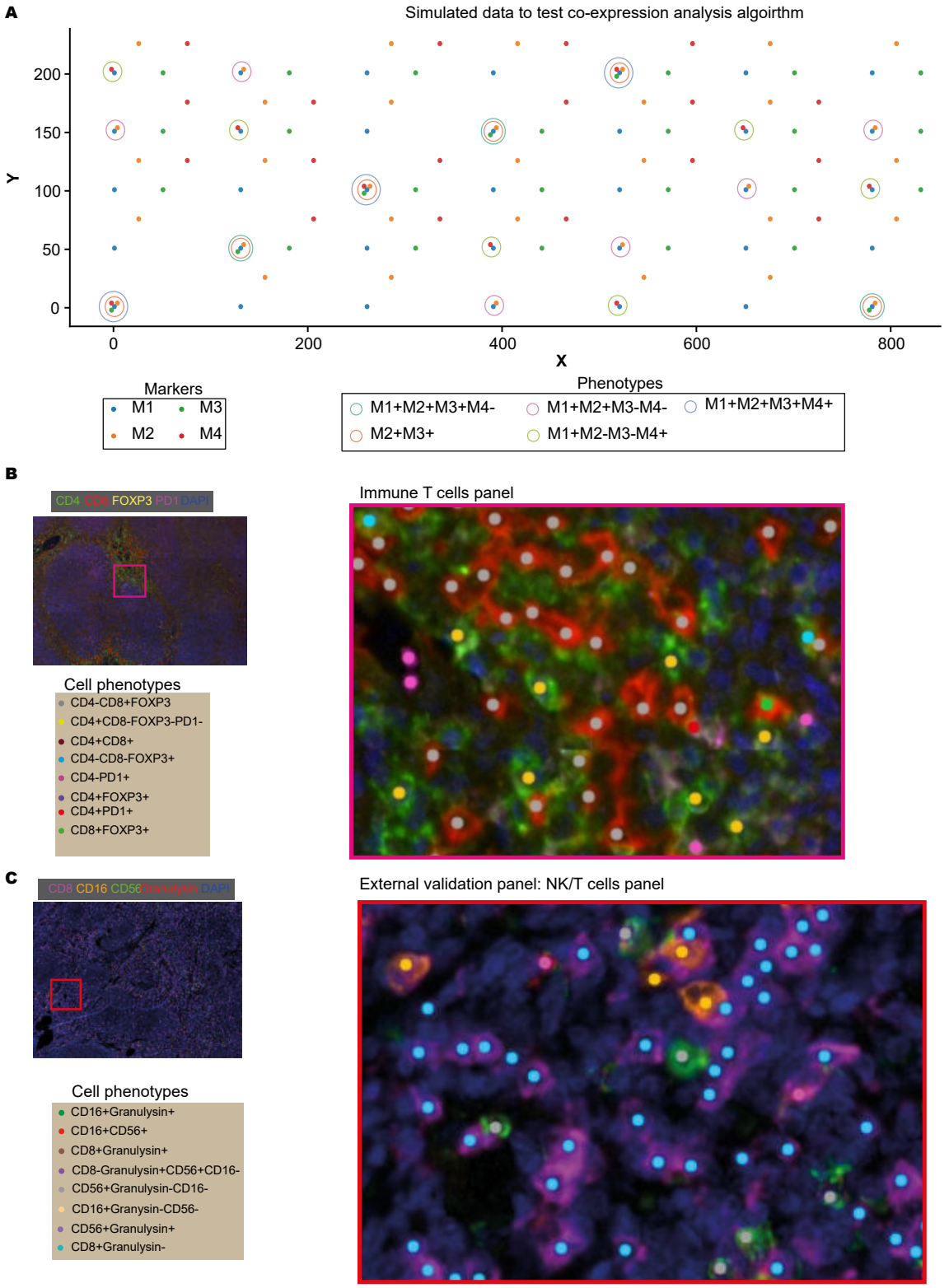


Figure 2.12: **Visual illustrative images of markers co-expression identification: A)** Simulated data to test co-expression analysis. **C-D)** Illustrative images for cell phenotype identification on immune T cells panel **(C)** and validation panel image **(D)**.

de-convoluted images from one-panel data could be generalized to the other panels. Thus, our newly developed deep learning image analysis pipeline systematically detects and classifies cells in MIF images from the de-convoluted images (Figure 2.6A) and this makes DeepMIF generalizable across panels.

DeepMIF could be effectively used for exploratory analysis or hypothesis-driven research using MIF data. In our DeepMIF pipeline, once cells positive for each marker are localised on the de-convoluted images, the user can request to identify cell phenotypes expressing any combination of the markers in the panel. Moreover, our approach enables exploring even rare cell types. In our immune T cell pipeline, we were able to identify clinically relevant rare cell types such as CD8+FOXP3+ cells which will be discussed in the next chapter.

Once cell phenotypes are identified for ease of downstream analysis, DeepMIF generates an excel file with the cell phenotype name, x position and y position of the cell. This allows us to perform spatial interrogation on cell-cell interaction and associate it with clinical data.

2.5.4 DeepMIF cell identification runtime evaluation

To statistically evaluate the speed of the DeepMIF pipeline, the execution time for 24 slides was recorded. The algorithm was executed on a desktop using 12 cores. The speed evaluation was conducted on the immune T cells panel which contains four markers and eight cell phenotypes of interest (Figure 2.12B). On average the process from cell detection on the de-convoluted images to cells co-expression analysis took 27.5 (range: 9.4 to 57.7) minutes per WSI (Figure 2.13A). The variation in run time per slide is due to the variation in image size and the number of cells (Figure 2.13B). In one of the images, about 90 million cells were detected. Since the samples are from lymph nodes, the tissue is densely populated with cells.

We then evaluated the whole slide MIF image reconstruction time from tiles. Upon evaluation on 24 slides, the average reconstruction was 0.07(range: 0.02 to 0.14) minutes for images size ranging from 48.5 million and 388.2 million pixels, respectively (Figure 2.13C, D). This shows the image viewer is fast.

In addition to the GUI, DeepMIF is available in Docker and it can run on high-performance clusters using Docker or Singularity. Moreover, DeepMIF can be applied to any multichannel spatial transcriptomics data, such as image mass cytometry (IMC) and CODEX data. These technologies could have up to 40 markers/channels. After detecting cells positive for the markers on their respective channel, our algorithm could be used to identify cells expressing single or multiple markers.

2.5.5 Limitation of DeepMIF and future directions

The DeepMIF pipeline has some limitations. Data generated from a small number of patients were used to train and validate the algorithm. The MIF images were also from one cancer type generated from one lab. Multi-centre MIF data incorporating inherent histopathology data variation in tissue

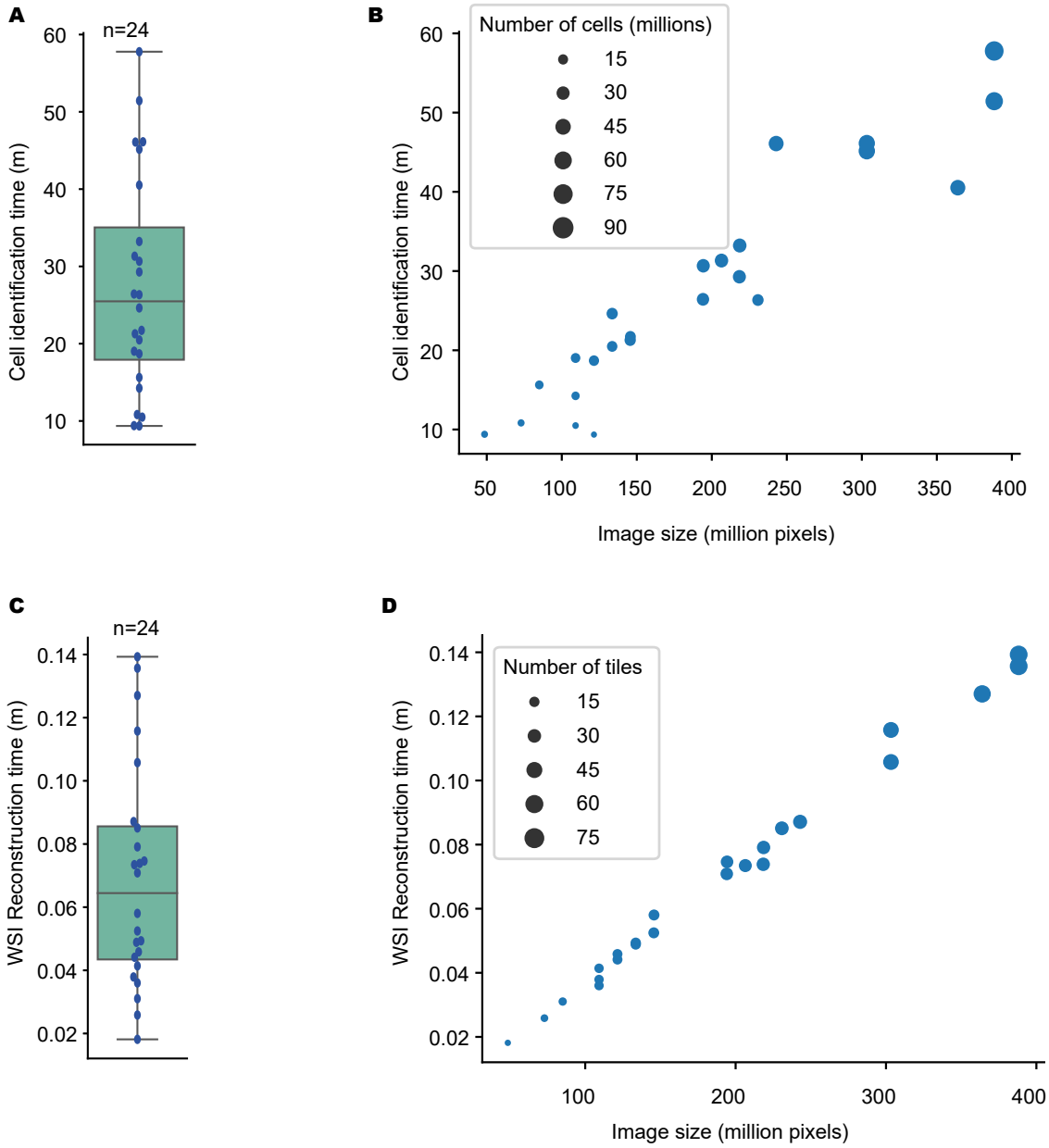


Figure 2.13: **DeepMIF pipeline speed evaluation: A)** Distribution of time taken for cell phenotype identification for n=24 slides. **B)** Scatter plot showing the time taken for cell identification as a function of the number of cells and image size. **C)** Distribution of whole slide image (WSI) reconstruction time from tiles. **D)** Scatter plot showing the distribution of WSI reconstruction time as a function of image size and the number of tiles. A point represents a WSI. Time is measured in minutes. The *n* indicates the number of slides.

processing, scanning device, scanning setting and data from different cancer types could improve the robustness of the DeepMIF pipeline. Having said that the cell detection performance of ConCORDe-Net is evaluated on MIHC and de-convoluted images from the MIF data. Moreover, cell segmentation instead of cell centre detection could allow more flexibility on downstream analysis [280]. But, this needs a laborious single cell manual segmentation annotation to train and validate the algorithm.

2.6 Conclusion

We proposed a deep learning based unified cell detection and classification method for whole tissue section multiplex images. Cell count regularised CNN was employed for cell detection, followed by CNN based single cell classification. The parameters in the cell detection architecture were learnt using a new objective function which optimizes Dice overlap and cell count. Our experiment shows that incorporating problem-specific knowledge such as cell count improves the performance of the cell detection algorithm. Moreover, we developed a deep learning based cell phenotype spatial mapping method, DeepMIF with a GUI for MIF images analysis. On separately held test data, the DeepMIF pipeline achieved a cell detection F1-score of 0.86 and cell classification accuracy and AUC of ≥ 0.98.

This suggests DeepMIF could be reliably used to analyse MIF images to identify novel prognostic cell phenotypes in the tumour microenvironment. Thus, in **Chapter 3**, DeepMIF was applied to multiple immune cell panels to spatially map single cell phenotypes and understand the immune landscape of follicular lymphoma.

Chapter 3

Spatial interrogation of follicular lymphoma tumour microenvironment

3.1 Overview

As described in Section 1.2.1, in follicular lymphoma, cancer cells develop in clumps and form follicle-like tissue structures. The heterogeneity of the inter- and intra-follicular regions of the lymphoid tissue in follicular lymphoma presents challenges to studying its immune microenvironment. We investigated the spatial interplay of T cells, macrophages, myeloid cells, and natural killer T cells using multispectral immunofluorescence images of diagnostic biopsies from 32 patients.

A deep learning-based image analysis pipeline was tailored to the needs of follicular lymphoma spatial histology research, enabling the identification of different immune cells within and outside neoplastic follicles. Neoplastic follicles were manually segmented by accredited hematopathologists. We analysed neoplastic follicles' morphologic features and the density and spatial co-localisation of immune cells in the inter-follicular and intra-follicular regions of follicular lymphoma.

The low inter-follicular density of CD8+FOXP3+ cells and co-localisation of CD8+FOXP3+ with CD4+CD8+ cells were significantly associated with relapse (BH corrected $p = 0.0057$ and BH corrected $p = 0.0019$, respectively) and shorter time to progression after first-line treatment (Logrank $p = 0.0097$ and log-rank $p = 0.0093$, respectively). A low inter-follicular density of CD8+FOXP3+ cells is associated with increased risk of relapse independent of FLIPI ($p = 0.038$, hazard ratio (HR) $= 0.42$, 95% $[0.19, 0.95]$, but not independent of the co-localisation of CD8+FOXP3+ with CD4+CD8+ cells ($p = 0.43$). Co-localisation of CD8+FOXP3+ with CD4+CD8+ cells is predictors of time to relapse independent of the FLIPI score and density of CD8+FOXP3+ cell density ($p = 0.027$, HR $= 0.0019$, 95% $[7.19x10^{-6}, 0.49]$. This suggests a potential role of inter-follicular CD8+FOXP3+ and CD4+CD8+ cells in the disease progression of FL, warranting further validation on larger patient cohorts.

3.2 Introduction

In the western world, FL is the second most common subtype of non-Hodgkin lymphoma, accounting for between 20% and 25% of cases [281, 21]. The disease tends to follow an indolent, remitting, and relapsing course, with great individual variability. While patients achieving a sustained response to first-line treatment show prolonged survival, those who fail to achieve a response or relapse early after the end of the therapy have an adverse outcome [282, 283, 39, 284]. Early identification of refractory/early relapsing cases and investigation of the biological basis is currently a major challenge [285].

The TME plays a key role in the clinical course of FL. Two immune response (IR) gene expression signatures, IR1 and IR2, were identified to be predictive of long and short survival, respectively, in FL [40]. The IR1 signature included genes encoding both T-cells and macrophage molecules, whereas the IR2 signature comprised genes expressed in macrophages, dendritic cells, or both. This and subsequent molecular studies [40, 60, 46], suggested the potential importance of immune surveillance in FL raising the possibility of novel immune approaches. The role of immune T cells [47–49, 27, 286, 59, 35, 287], macrophages [50, 51, 288, 27, 286], natural killer T (NKT) cells [289, 290] and myeloid cells [291, 292] were investigated in FL generating inconsistent results. These studies were conducted on a different cohort of patients who might have different characteristics and were analysed using different computational pipelines that could potentially hamper consistency and comparison of the prognostic power of the different immune cell groups. Moreover, the composition of the intra-follicular areas, which contain neoplastic cells, is distinct from the inter-follicular areas. However, most of the previous studies considered the FL TME as one homogenous ecosystem. The pattern of immune infiltration in these two sites is predictive of outcome [286, 59, 44]. Thus, investigating the spatial interaction of immune cells in the two regions could provide new insight into the biology of FL. However, no computational image analysis software tailored to these cell compartments is available.

Recently, deep learning has gained a surge of interest in digital pathology [293] demonstrating its relevance to predict the diagnosis of several malignant diseases, including Lymphoma [294–296]. It has been shown that this technology also serves as a discovery tool to identify novel cell populations associated with tumour progression. Automated microscopy analysis is a more reliable approach to enumerate infiltrating cell populations, but there has been limited use of deep-learning analysis to study the microenvironment in FL [64, 57].

Thus, we decided to use MIF tissue spatial staining technology to study the morphological features of neoplastic follicles and the spatial immune landscape of FL. We aim to:

- Develop a multi-panel MIF tissue spatial staining to capture a multitude of cell phenotypes and a deep learning-based method to identify cell phenotypes;
- Study the morphologic features of neoplastic follicles;

- Develop cell distribution and spatial analysis pipelines tailored to the tissue compartments of FL;
- Identify novel immune phenotypes associated with risk of relapse in FL.

3.3 Materials

3.3.1 Patients studied

Patients diagnosed at Papa Giovanni XXIII Hospital (Bergamo, Italy) with grade I-IIIa FL between 01-Jan-2006 and 31-Dec-2015, treated with standard R-CHOP or R-CVP and with the availability of the diagnostic surgical biopsy were eligible for this study. Clinical information of 39 patients was gathered from the electronic charts. The diagnosis of FL was confirmed by three haematopathologists (Prof. Teresa Marafioti, Prof. Alan Ramsay, and Prof. Sabine Pomplun from the Department of Histopathology, University College Hospitals London, London, UK). They independently reviewed the morphology using H&E staining. The relevant immunostaining evaluated included CD20, CD3, BCL-2, BCL-6, CD10, CD21, MIB-1. All cases expressed BCL-2, CD10 and BCL-6, and no areas of diffuse growth pattern were present. The diagnosis of follicular lymphoma followed the criteria of the revised 4^{th} edition of the World Health Organisation classification of tumours of haematopoietic and lymphoid tissues [297].

The exclusion criteria included stage I disease, bendamustine therapy, and rituximab maintenance. Seven cases were excluded, of which six showed suboptimal tissue sections affecting staining, and the additional case had received bendamustine treatment. The final number of analysed cases was 32 (Figure 3.1).

The study was approved by the Ethics Committee (approval number REG. 197/17) and performed following the ethical standards of the 1964 Helsinki declaration and its later amendments. All patients provided written informed consent. The time to relapse, measured from diagnosis to relapse/progression time, was used as a clinical endpoint.

3.3.2 Clinical characteristics

The clinical characteristics of the 32 patients included in this study are summarised in Table 3.1. The histological grade for all patients who relapsed varied between grades 1 and 2, but only one patient showed a focal grade 3A pattern. After a median follow-up of 10.4 years (range 0.25–15.2 years), 23 patients remained alive. A total of nine deaths occurred. The causes were related to the progression of FL (three cases); transformation to diffuse large B-cell lymphoma (one case); secondary cancer (one case); unknown (occurring >10 years post-treatment, two cases); complication of allogeneic stem-cell transplantation (one case); and acute hepatitis (one case). Fifteen patients relapsed after a

Table 3.1: **Characteristics of follicular lymphoma patients**.

Clinical Characteristics	All patients(N, %)
All patients	32 (100)
Age	
Median (range)	50.9 (30.5 – 77.9)
>60	10 (31.3)
Gender	
Male	16 (50)
Ann Arbor stage	
III – IV	29 (90.6)
Bone marrow involvement	
Yes	21 (65.6)
FLIPI	
Low	3(9.4)
Intermediate	16 (50)
High	13 (40.6)
Treatment	
R-CHOP	26 (81.3)
R-CVP	6 (18.7)

median of 2.83 years (range 0.6-14.8 years). The remaining 17 patients did not relapse after a median observation of 11.5 years (range 0.25-14.8 years), with four deaths unrelated to disease progression.

3.4 Methods

3.4.1 Multiplex immunofluorescence images and cell phenotypes

To study the immune landscape of FL, we used four MIF panels staining different immune cell markers from formalin-fixed paraffin-embedded diagnostic tissue sections of 32 FL (Figure 3.1A, B). These MIF images were generated for this study and were not previously published. The immune spatial phenotypes were investigated using four MIF panels for T cells, macrophages, myeloid cells, and NKT cells. The markers used in these panels include:

1. **Immune T cells markers:** CD4/CD8/FOXP3/PD1;
2. **Tumour-associated macrophages markers:** CD68/CD163/CD206/PDL1;
3. **Myeloid cells markers:** CD8/CD11b/CD14/CD15;
4. **Natural killer T cells markers:** CD8/Granzyme B (Granz B)/Granulysin/CD16/CD56.

The panels were optimised by Dr. Ayse U Akarca from Prof. Teresa Marafioti's lab in University College London. The details on the panels optimization could be found in Section 2.3.2.

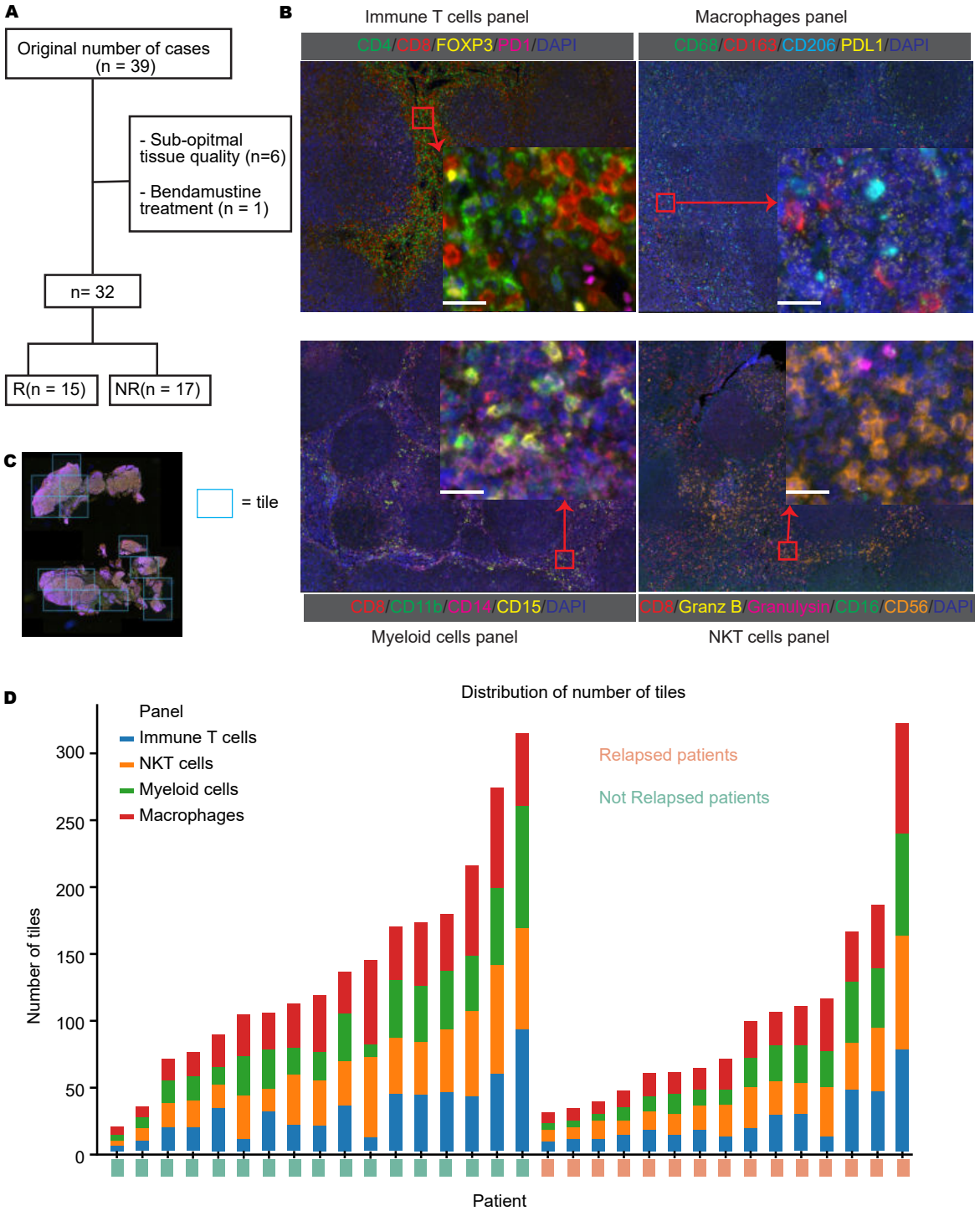


Figure 3.1: **Details of the study cohort and MIF images: A)** Consolidated Standards of Reporting Trials (CONSORT) Diagram. R = Relapsed; NR = Not relapsed. **B)** MIF images were acquired using Vectra 3 platform with multiple markers in each panel. The scale bar indicates $10\mu m$. **C)** Example image showing the multispectral regions of interest (tile) selection from a whole tissue section. The tiles were selected by a pathologist from different parts of the image to include heterogeneity of tissue. Each tile was exported as a TIF file. **D)** Stacked bar plots showing the variation of the number of tiles obtained from the Vectra 3 platform for the 32 cases involved in this study for the four panels.

The MIF images of diagnostic biopsies were acquired using the Vectra 3 platform as explained in Section 2.3.2. The multispectral regions of interest or "tiles" (Figure 3.1C) were selected from different region of the whole slide tissue section by expert pathologist and exported as TIF files. The Vectra 3 platform's Phenochart viewer allows pathologists to view a whole slide contextual scan as H&E image and select different regions of interest for acquisition. The tiles were selected from different areas to capture the heterogeneity in the tissue section. Depending on the size of the available tissue section and region of interest decided by the pathologist, the number of tiles selected on each sample could vary as shown in Figure 3.1D. A typical tile was $4032 \times 3012 \times 3$ pixel size at $20\times$ magnification with $0.5\mu m$ per pixel resolution. Different tissue sections were used for each panel.

I have been collaborating with Prof. Teresa Marafioti (histopathologist) from the University College London, United Kingdom and Dr. Giuseppe Gritti (haematologists) from Ospedale Papa Giovanni XXIII, Bergamo, Italy. Based on literature and my collaborators expertise in FL, we were interested in a set of cell phenotypes from the four immune cell panels. The cell phenotypes of interest are shown in Table 3.2.

3.4.2 Cell phenotyping on multiplex immunofluorescence images

The DeepMIF pipeline developed in Chapter 2 was used to spatially map the cell phenotypes on these four panels (Figure 3.2A). First, cell positive for each markers were identified on the de-convoluted images. Then, cells expressing single or multiple markers were identified as follows: Firstly, for a given tile, we mapped the location of the positive markers from its de-convoluted images onto a single plane. Then, to identify overlapping and non-overlapping markers, we computed a Euclidean distance between the markers in the image space. If the distance between detected markers on the de-convoluted images is less than 1.5µm, the markers are co-expressed on a cell. The distance value was empirically set. The detail of DeepMIF algorithm could be found in Section 2.4.2.

3.4.3 Deep learning model validation

The deep learning models of DeepMIF pipeline were trained on data from the immune T cells panel and applied to all four panels. Validation ensures the cell detection and classification models training on immune T cell data generalises to the other panels. Validation was performed using two types of data. As explained in in Section 2.5.2, the DeepMIF cell classifier was validated on macrophages and natural killer T cells panels single cell annotation data. Moreover, CD8 marker is included in immune T cell, myeloid cells and NKT panel. This enabled us to verify the density of the CD8+ cells in these panels. The correlation of density of CD8+ cells between the panels was used to measure the generalizability of the deep learning models to the other panels. We expect a strong correlation of CD8+ cell density between the panels if the deep learning method is generalizable to all panels.

Table 3.2: **Cell phenotypes of interest in each panel.** The cell phenotypes were selected based on literature and recommendations from haematologists with expertise in FL.

Immune T cell	**Macrophages**
CD4-CD8+FOXP3-PD1-	CD68+CD206-
CD4+CD8-FOXP3-PD1-	CD163+PDL1-
CD4-CD8-FOXP3+	CD68-CD206+
CD4-CD8-PD1+	CD163-PDL1+
CD4+CD8+	CD68+CD206+
CD4+FOXP3+	CD163+PDL1+
CD8+PD1+	
CD4+PD1+	
CD8+FOXP3+	
Myeloid cells panel	**Natural killer T cells panel**
CD8+CD11B-	CD8+Granulysin-Granz B-
CD8-CD11B+CD14-CD15-	CD16+Granulysin-CD56-
CD11B-CD14+CD15-	CD56+Granulysin-CD16-
CD11B-CD14-CD15+	CD8-Granz B+
CD11B+CD15+	CD8-Granulysin+CD56-CD16-
CD11B+CD14+	CD8+Granulysin+
CD8+CD11B+	CD8+Granz B+
CD15+CD14+	CD56+Granulysin+
	CD16+CD56+
	CD16+Granulysin+

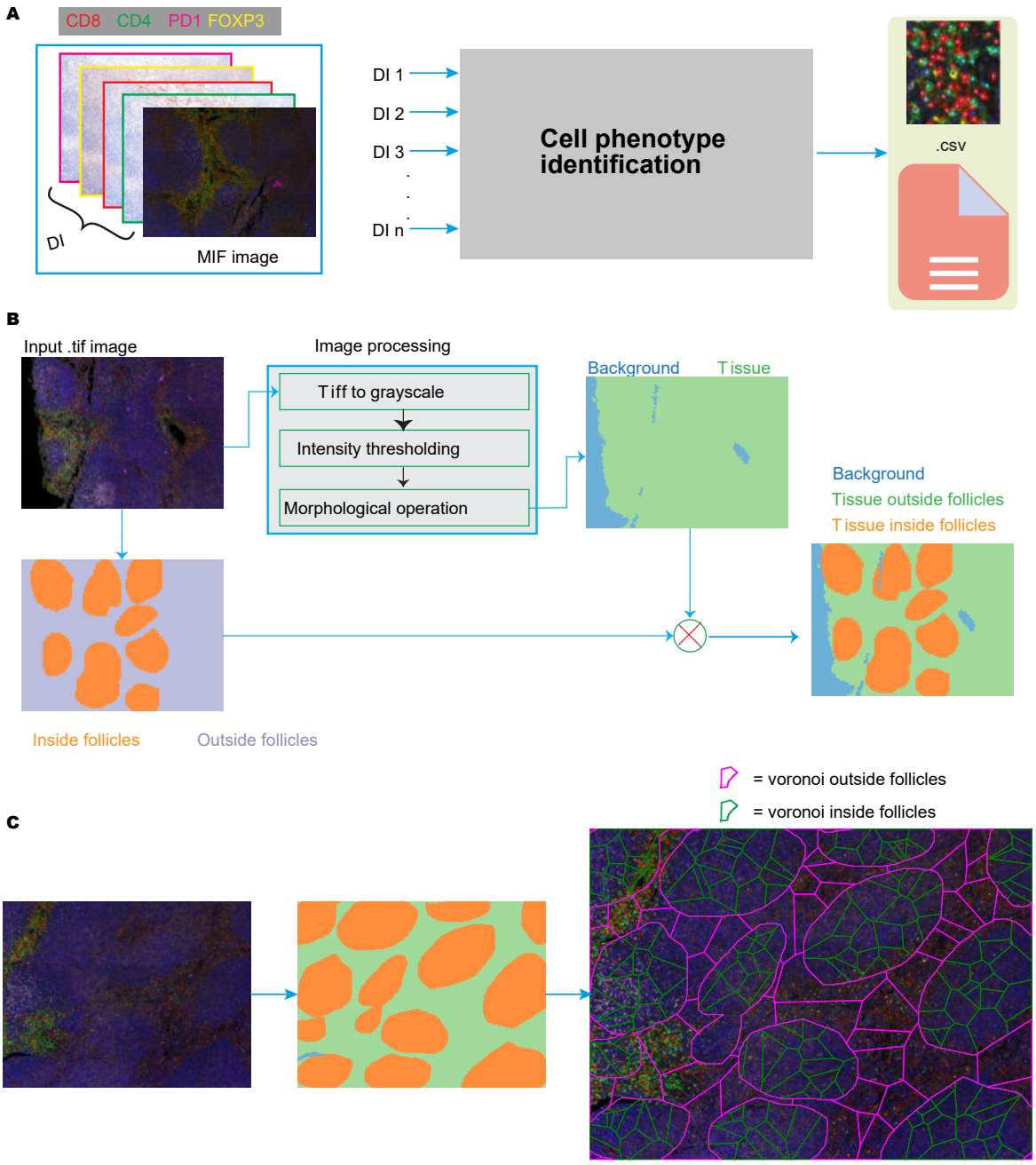


Figure 3.2: **Computational deep learning and image processing pipelines: A)** DeepMIF pipeline for cell identification on MIF image. Details on DeepMIF could be found in Section 2.4.2. **B)** Schematic diagram showing tissue and follicles segmentation. Follicles are manually segmented by an expert pathologist. **C)** Spatial Voronoi tessellation of within and outside follicles tissue compartments for Morisita-horn index spatial analysis. Following the follicle segmentation in (**B**), the area within the follicle and outside follicle were divided into smaller polygons called "Voronoi". Using cells location data obtained in (**A**), cells were mapped onto these cells to apply spatial clustering analysis.

3.4.4 Tissue and follicles segmentation

To segment the tissue from the background, we first converted the MIF image into a greyscale image with an intensity range from 0 to 1. A threshold, T = 0.03 was applied to convert the greyscale image into a binary image. The value of T was optimised from the intensity profile of the greyscale images, and by visual inspection of the segmentation results. To smooth and fill holes in the binary image, we applied morphological closing (dilation followed by erosion) operations using disk structuring-element of radius 10 pixels ($5\mu m$). Mathematically, let I_{in} be the input image, and S be the structuring element. The output image, I_{out} is computed as,

$$I_{out} = (I_{in} \oplus) \ominus S \tag{3.1}$$

, where $\oplus$ and $\ominus$ are the dilation and erosion operations, respectively.

Follicles were manually delineated by three accredited hematopathologists (Prof. Alan Ramsay, Prof. Sabine Pomplun, and Teresa Marafioti, from Department of Histopathology, University College Hospitals London, London, UK). Two regions of interest were annotated for each FL tissue section: the region representing the neoplastic follicle (called within follicle) and the areas between neoplastic follicles (called outside follicles). To annotate the follicles, we used PD1, CD68, CD14, and Granulysin de-convoluted images for immune T cells panel, macrophage panel, myeloid panel and natural killer T cells panel, respectively. The selection of these de-convoluted images was made because follicles were more visible on these images than other de-convoluted images.

Finally, the output segmentation (S_{out}) was obtained by combining tissue segmentation (S_T) and follicle segmentation (S_F) as shown in Figure 3.2B and Equation (3.2).

$$S_{out} = S_L \odot S_T \tag{3.2}$$

where $\odot$ represents element-wise matrix multiplication.

3.4.5 Image and spatial analysis tailored to follicular lymphoma tissue compartments

The composition and spatial organization of immune cells in follicular lymphoma was analysed in the intra-follicular (within follicles) and inter-follicular (outside follicles) regions, by developing a tissue and follicles segmentation pipeline (Figure 3.2B). This approach was designed to investigate whether distinct patterns of immune cell infiltrates in the two micro-ecosystems represent a robust tool to predict clinical outcome. To quantify cells spatial co-localization and immune cell composition, we applied a Morisita-Horn index [298, 299, 260] to the regions within and outside the follicles separately (Figure 3.2C) and demonstrated differences between the two cellular compartments.

3.4.6 Neoplastic follicles morphologic features

Following the neoplastic follicles segmentation, morphological features were measured. Here is the definition of morphological features of neoplastic follicles used:

- **Total follicle tissue area:** it is computed as sum of neoplastic follicles tissue area measured in μm^2 in all tiles of a given case. It is normalised by the total tissue area of the slide.
- **Number of follicles.** It is count of neoplastic follicles normalised by total tissue area.
- **Mean follicle area.** It is the average area of each neoplastic follicle measured in μm^2.
- **Mean follicle solidity.** Solidity of a neoplastic follicle measures its density or the extent to which the neoplastic follicle covers its convex hull [300, 301] (Figure 3.3A). A given slide has many neoplastic follicles. A slide level score was generated as a mean of solidity of all neoplastic follicles in the slide. Its value is ranges between 0 and 1. Solidity is a measure of irregularity of a shape [301]. A value of 1 indicates a solid follicle, while a value less than 1 indicates follicle with irregularities.
- **Mean follicle eccentricity.** It measures the elongation of the neoplastic follicles. An eccentricity of fully circle is 0 and eccentricity of elliptical object is greater than 0 but less than 1 [302] (Figure 3.3B). Similar to solidity, mean value was taken to get score at case level.

3.4.7 Cell density in the different tissue compartments of follicular lymphoma

Figure 3.1D shows the variation of number of tiles between different cases. The difference is due to variation in the amount tissue in the sample or variation in the number of region of interest during MIF image scanning in the Vectra 3 platform. In such scenario, cell density is robust to variation in the amount of tissue compared to the abundance of cells since the earlier is the normalised by tissue area. Thus, we statistically compared cell densities instead of cell abundance between patient groups. To identify prognostic cell types within and outside the neoplastic follicles, we mapped cells to their respective region (within or outside follicle) and we computed cell density (cells per mm^2) following the tissue and follicle segmentation results. For a given cell type, cell density within follicles is computed by dividing the number of cells within follicles by follicles tissue area (mm^2). A similar approach was applied for the outside follicles areas.

3.4.8 Tessellation of FL cellular compartments and Morisita-Horn index

Morisita-Horn index is a measure of co-localization of two spatial variables used in ecological and immunological studies [298, 299, 260]. In FL, the inter- and intra-follicular regions have distinct

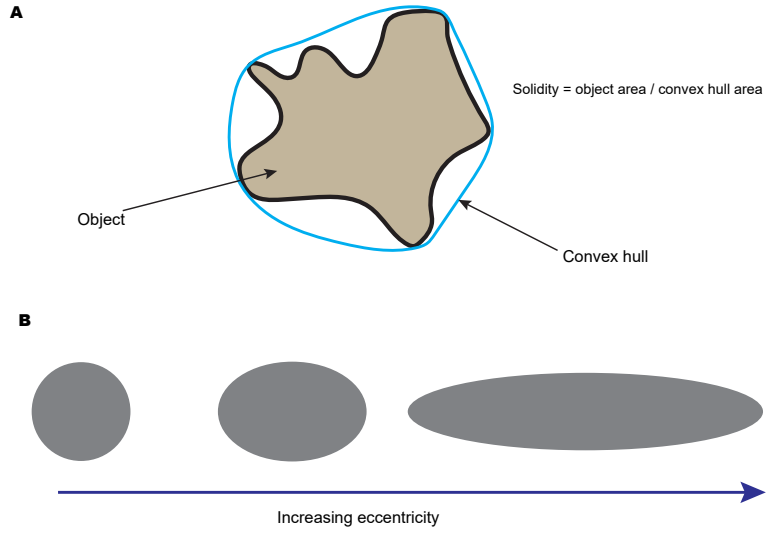


Figure 3.3: **Morphological shape descriptors: A)** A cartoon showing computation of solidity shape feature of an object (for example, a neoplastic follicle). Small value of solidity indicates shape irregularity. The value of solidity is bounded between 0 and 1. **B)** A set of cartoon objects illustrating eccentricity shape descriptor measure. Eccentricity measures the elongation of the object. The value of eccentricity is bounded between 0 and 1.

morphological and immune infiltration patterns [27, 286, 59, 44]. Here, we hypothesised, the two tissue compartments differ in their cellular structure and thus we analysed the co-localization of cells in these regions separately.

To compute the Morisita-Horn index, we first tessellated the tissue area into smaller regions as shown in Figure 3.2C. The most commonly used tessellation strategies are square and Voronoi tessellation [260]. We chose Voronoi tessellation because it mimics the natural distribution of spatial point patterns [260]. In Voronoi tessellation, an image is divided into a set of polygons using randomly selected seed points as a centre. The number of polygons (n) is determined by the tissue area as shown in Equation (3.3) for an image at 20$\times$ resolution [303]. To generate and visualise tessellation, we used scipy.spatial [304] Python packages.

$$n = \frac{\sqrt{Tissue\ area}}{48} \tag{3.3}$$

We computed the Morisita-Horn index for the tissue within follicles as follows. Let **Z** be the number of tiles (**T**), collected from a given patient tissue section and let Y be the number of follicles (**F**) in i^{th} tile, T_i. Let L be the number of polygons in the j^{th} follicle, F_j obtained using Equation (3.3), which depends on the area of follicle F_j. Since the tiles are non-overlapping, the polygons within the follicles from the Z tiles could be combined into a set P to apply the spatial analysis.

$$P = \{P_{111}, P_{112}, ..., P_{2jk}, ..., P_{3jk}, P_{ijk}, ..., P_{ZYL}\} \tag{3.4}$$

where $P_i jk$ is the k_{th} polygon in follicle F_j of Tile T_i.

Then, we computed the number and proportion of each cell types in each polygon using the location and class labels of cells obtained from the output of DeepMIF pipeline.

We then computed the spatial co-localization measure of Morisita-Horn index (M) for a pair of cell types, C and C' as follows,

$$M = 2\frac{\sum_k X_k X_k^{'}}{\sum_k (X_k)^2 + \sum_k (X_k^{'})^2} \tag{3.5}$$

where X_k and X_k' are the proportion of C and C' in the k^{th} polygon, and $1 \leq \mathrm{k} \leq |P|$, where $|P|$ is the number of polygons within follicles in Equation (3.4). A similar procedure was followed for the region outside follicles. The value of m ranges from 0 (spatial segregation) to 1 (high spatial co-localization).

3.4.9 Statistical analysis

All image and statistical analyses were carried out using the Python programming language. All correlation values were measured using the non-parametric Spearman test. The p-values were computed using the two-sided unpaired, non-parametric Wilcoxon method, considering $p < 0.05$ as significant. To correct for multiple testing, we applied the Benjamini-Hochberg (BH) method.

As the main measure for prognostic analysis, we used relapse status and relapse free survival (RFS), which is defined as the period of time from diagnosis to relapse or progression, with censoring at death or the last follow-up date. Kaplan–Meier method was applied to estimate RFS. To quantify the hazard ratio for the effect of biomarker groups, the Cox proportional hazard model was used. The multivariate analysis included biomarkers identified from our analysis and FLIPI, a standard clinical variable in the diagnosis of FL. The Kaplan–Meier analysis and multivariate Cox regression analysis were performed using the Lifelines (v0.25.4) Python package [305]. Moreover, we measured the concordance index (c-index) which is a measure of concordance between the observed and predicted survival times. It is the fraction of individuals whose expected survival times are correctly arranged out of all individuals that can actually be ordered [306] and it was computed using the Lifeline Python package [305].

3.5 Results

3.5.1 Deep learning models accurately map single cells in multiplex immunofluorescence images

To spatially map single cells on MIF images, we applied DeepMIF developed in Chapter 2. DeepMIF's single cell detection and classification were trained on the immune T cells panel data (Table 2.2).

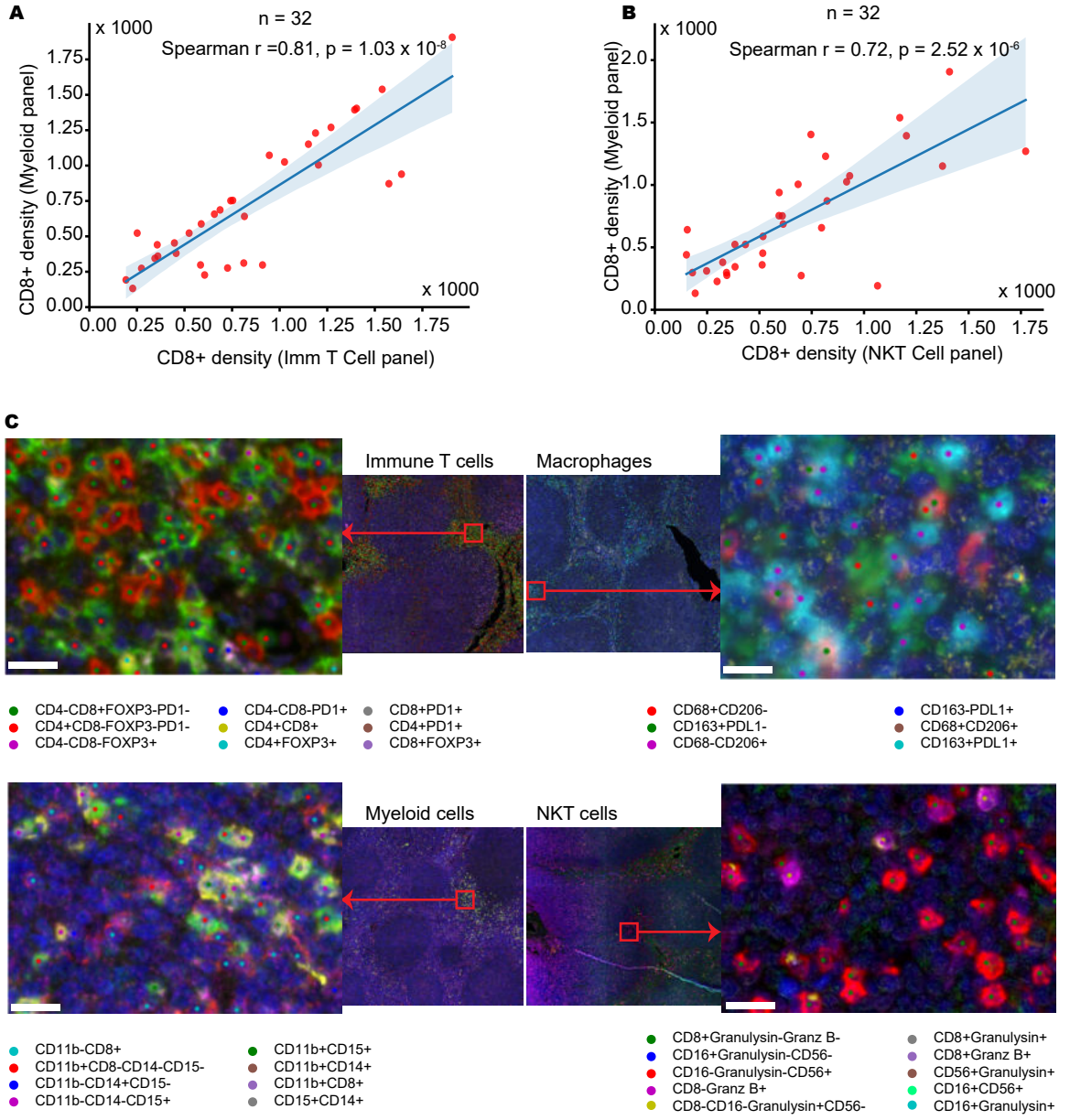


Figure 3.4: **Evaluation and validation of deep learning models: A,B)** Deep learning model validation. The deep learning models were trained on immune T cell panel data. The trained model was then applied to all panels. The density of CD8+ cells (cells per $1000\mu m^2$) across different panels was significantly correlated. A dot represents a sample or patient. All correlation values were computed using a non-parametric Spearman correlation. **C)** After detecting cells on the de-convoluted images, using the proposed co-expression analysis (see Method), we were able to spatially map cells expressing single or multiple markers in all panels, which allows us to visually validate the deep learning models and co-expression analysis on MIF images. The scale bar is 10μm.

In Section 2.5.2, we showed that the models generalise to other panels using manual annotation of single cells from images in the macrophages and natural killer T cells panel. Moreover, the immune T cells panel, myeloid cells panel and natural killer T cells panel contain CD8 marker allowing us the abundance of CD8+ cells in these panels. As shown in Figure 3.1D, the number of "Tiles" are different between the panels. To overcome the amount of tissue bias, we correlated the density of CD8+ cells instead of cell count. The density of CD8+ cells in the myeloid cells panel was significantly correlated with the density of CD8+ cells in immune T cells and natural killer T cells (spearman $r = 0.81,\ p = 1.03 \times 10^{-8}$; spearman $r = 0.72,\ p = 2.54 \times 10^{-4}$, Figure 3.4A, B and Table S2.3). Moreover, the density of CD8+ cells in the immune T cells and natural killer T cells panel was also significantly correlated (spearman $r = 0.55$ $p = 8.42 \times 10^{-4}$, Figure S2.1D and Table S2.3). Figure 3.4C shows sample DeepMIF pipeline output images for the four panels. These indicate DeepMIF could be reliably used to identify single cell phenotypes on MIF images.

3.5.2 Neoplastic follicles morphometric features and relapse status

We hypothesised that specific neoplastic follicle morphological features could be associated with relapse after R-CHOP or R-CVP chemoimmunotherapy. Thus, after delineating the neoplastic follicles, we measured solidity, area, eccentricity, total neoplastic area, and the number of follicles (Section 3.4.6). We observe that neoplastic follicles in the diagnostic samples of relapsed patients show a borderline decrease in solidity score (increase in shape irregularity) compared with neoplastic follicles from the diagnostic samples of patients who did not experience relapse ($p = 0.02$ (no correction applied), Figure S2.11A), though this was not significant after the multiple testing correction (BH corrected $p = 0.1$, Figure 3.5A). Illustrative images showing tiles which have neoplastic follicles with low and high solidity scores are displayed in Figure S2.10A, B. The tile with low solidity contains merged neoplastic follicles, creating shape irregularity (Figure S2.10A). The neoplastic follicles area, eccentricity, total neoplastic area and the number of neoplastic follicles were not different between the relapsed and not relapsed patients (Figure 3.5B-E).

3.5.3 Decreased inter-follicular CD8+FOXP3+ cells is associated with relapse

To identify prognostic cell types, we first computed the density of cell phenotypes listed in Table 3.2 in the inter- and intra-follicular regions of the neoplastic follicles. Relapse status (relapsed or not relapsed), and RFS were used for prognosis analysis. The prognostic significance of the cell phenotypes of interest with respect to relapse status and RFS are summarised in Table S2.4 and 2.4) and (Figures S2.3 and S2.5. After applying multiple test corrections, only the density of CD8+FOXP3+ cells outside the neoplastic follicles was prognostic.

A significantly lower density of CD8+FOXP3+ cells outside the neoplastic follicles was found in diagnostic samples of patients who later relapsed, compared to those patients who did not relapse (BH

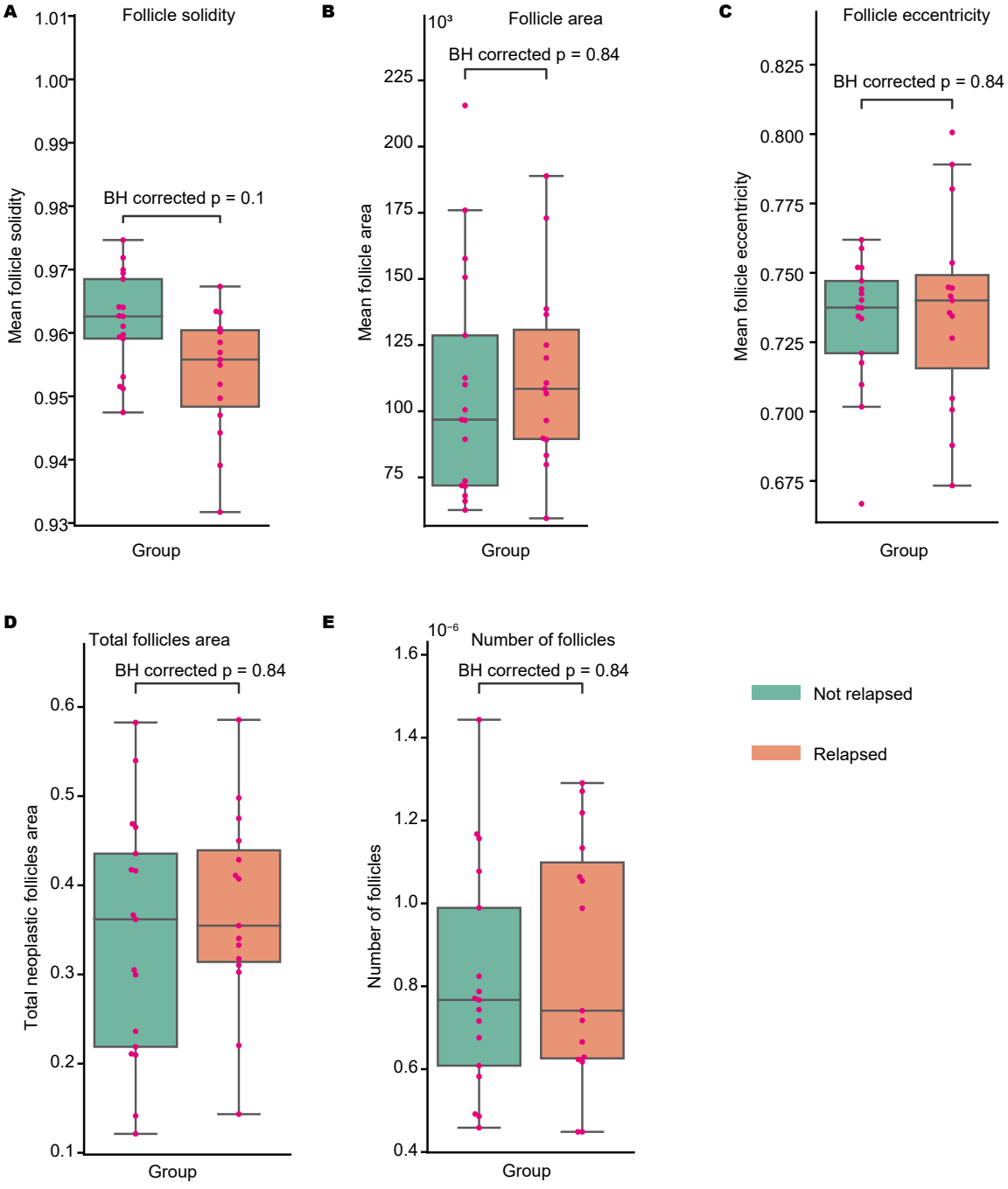


Figure 3.5: **Neoplastic follicles morphological features and follicular lymphoma prognosis. A-C)** Boxplots showing the difference in neoplastic follicles solidity (A), average area (B), and eccentricity (C) between relapsed (n = 15) and not relapsed (n = 17). **D-E)** Boxplots showing the difference in total follicle area (D), and the number of follicles (E) between relapsed (n = 15) and not relapsed (n = 17). These features were normalised by the total amount of tissue in the slide. The area was measured in μm^2. For statistical comparisons among groups, a two-sided, nonparametric, unpaired, Wilcoxon signed-rank test was used.

corrected p = 0.0057, Figure 3.6A). However, the density of CD8+FOXP3+ cells in the intra-follicular region was not different between the relapsed and not relapsed cases (BH corrected $p = 0.142$, Figure 2.2). Using Kaplan-Meier estimates, increased CD8+FOXP3+ cells outside the neoplastic follicles was significantly associated with improved RFS using a median split (high 50% vs low 50%: Logrank p = 0.0097 and c-index = 0.68, Figure 3.6B).

The CD8+FOXP3+ cells accounted for 1.6% and 3.4% of CD8 marker and FOXP3 marker expressing immune cells, respectively (Figure 3.6C, D). Moreover, in about 88% of the patients, the density of CD8+FOXP3+ cells in the inter-follicular area of the tissue was higher than the density in the intra-follicular region (Figure S2.7). This shows CD8+FOXP3+ cells are predominantly found in the inter-follicular microenvironment.

We then asked if the CD8+FOXP3+ cells infiltration is associated with morphological irregularity of neoplastic follicles measured by solidity. We found that there was no association between CD8+FOXP3+ cells infiltration and neoplastic follicles morphological irregularity (spearman $r = 0.145, p = 0.428$, Figure S2.8A, B).

3.5.4 Clinical relevance of immune cells co-localisation in follicular lymphoma

To understand the spatial interaction of the inter-follicular CD8+FOXP3+ cells with the other T cell subsets shown in Table 3.2 in the TME, we first explored their spatial neighbourhood using nearest neighbour (NN) analysis (Figure 3.7A-B). For each CD8+FOXP3+ cell, we identified the NN cell phenotype and computed the distance in the tissue space (Figure 3.7A). In the inter-follicular region, CD4+CD8+ and CD4+FOXP3+ NN cells tend to localise closer to CD8+FOXP3+ cells than other T cell subsets including CD4-CD8+FOXP3-, CD4+CD8-FOXP3-, and CD4-CD8-FOXP3+ cells.

We then asked if the co-localisation of these T cell subsets with CD8+FOXP3+ cells in the inter-follicular region is associated with relapse and RFS. To quantify spatial co-localisation, we computed the Morisita-Horn index, which increases in value if there is a high degree of spatial colocalization between two variables **(Methods)**. The inter-follicular co-localisation of CD8+FOXP3+ with CD4+FOXP3+ cells was not associated with relapse status (BH corrected p = 0.142, Figure S2.5A) and patient RFS (Logrank $p = 0.06$, Figure S2.5B) using Kaplan-Meier estimates. However, a lower degree of inter-follicular co-localisation of CD8+FOXP3+ with CD4+CD8+ cells was associated with relapse (BH corrected p = 0.0019, Figure 3.7C). Using Kaplan-Meier estimates, a higher degree of co-localisation of CD8+FOXP3+ with CD4+CD8+ cells was associated with longer RFS (Logrank $p = 0.0093$ and c-index = 0.67, Figure 3.7D).

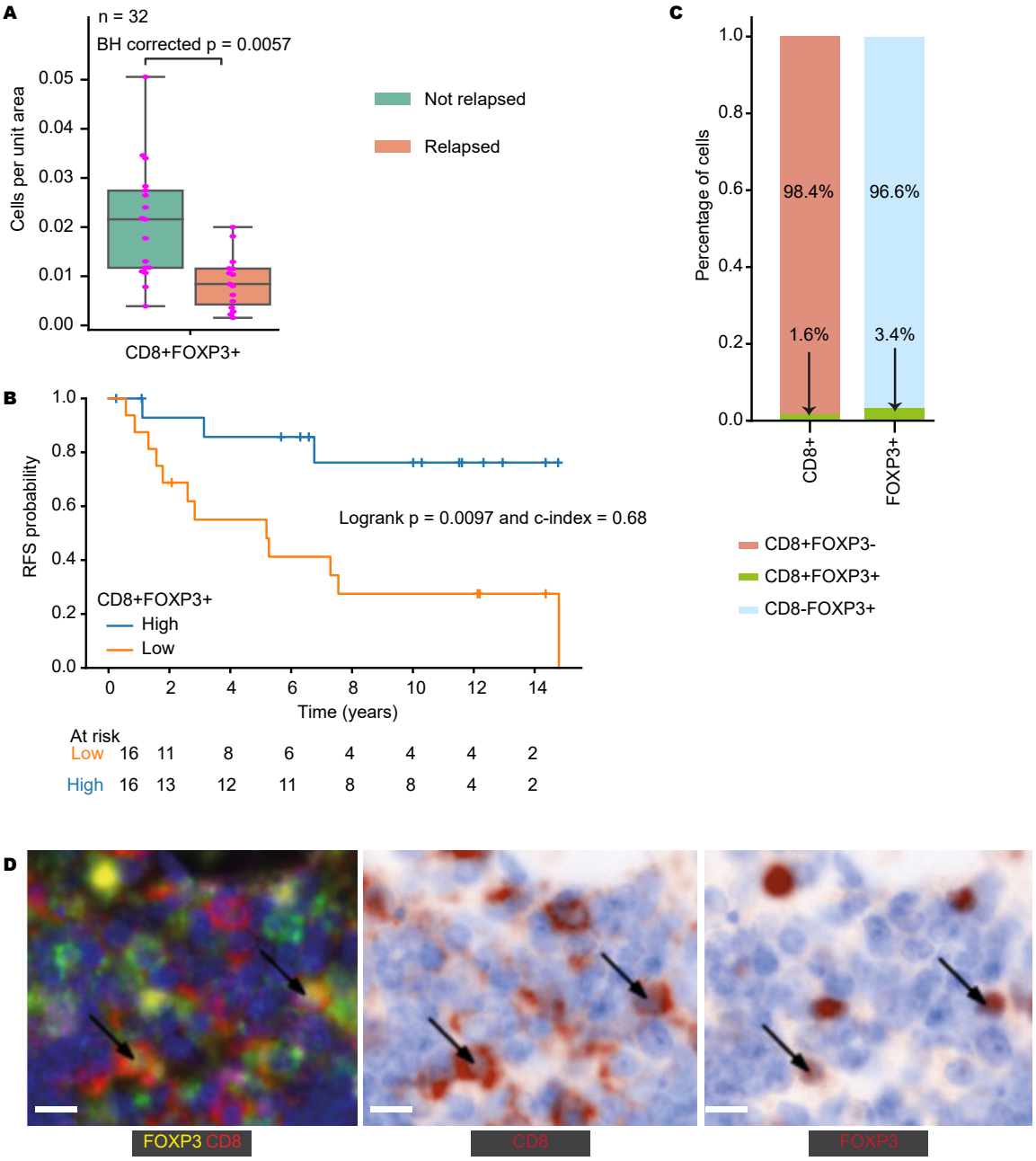


Figure 3.6: **Prognostic cell subsets outside the neoplastic follicles. A)** Boxplot showing difference in density of CD8+FOXP3+ cells (cells/1000μm^2) outside follicles between relapsed (n = 15) and not relapsed (n = 17). **B)** Kaplan-Meier curves illustrating recurrence free survival (RFS) of patients dichotomised using median CD8+FOXP3+ cells density outside follicles. The c-index indicates the concordance index between the observed survival times and predicted survival times. **C)** The percentage of CD8+ and FOXP3+ T cells expressing both CD8 and FOXP3 markers. **D)** Sample illustrative image containing CD8+FOXP3+ cells. The arrows point to the centre position of CD8+FOXP3+ cells detected by our deep learning method on MIF and de-convoluted images. The scale bar represents 10μm. For statistical comparisons among groups, a two-sided, nonparametric, unpaired, Wilcoxon signed-rank test was used, unless stated otherwise. To correct for multiple testing, we applied Benjamini-Hochberg (BH).

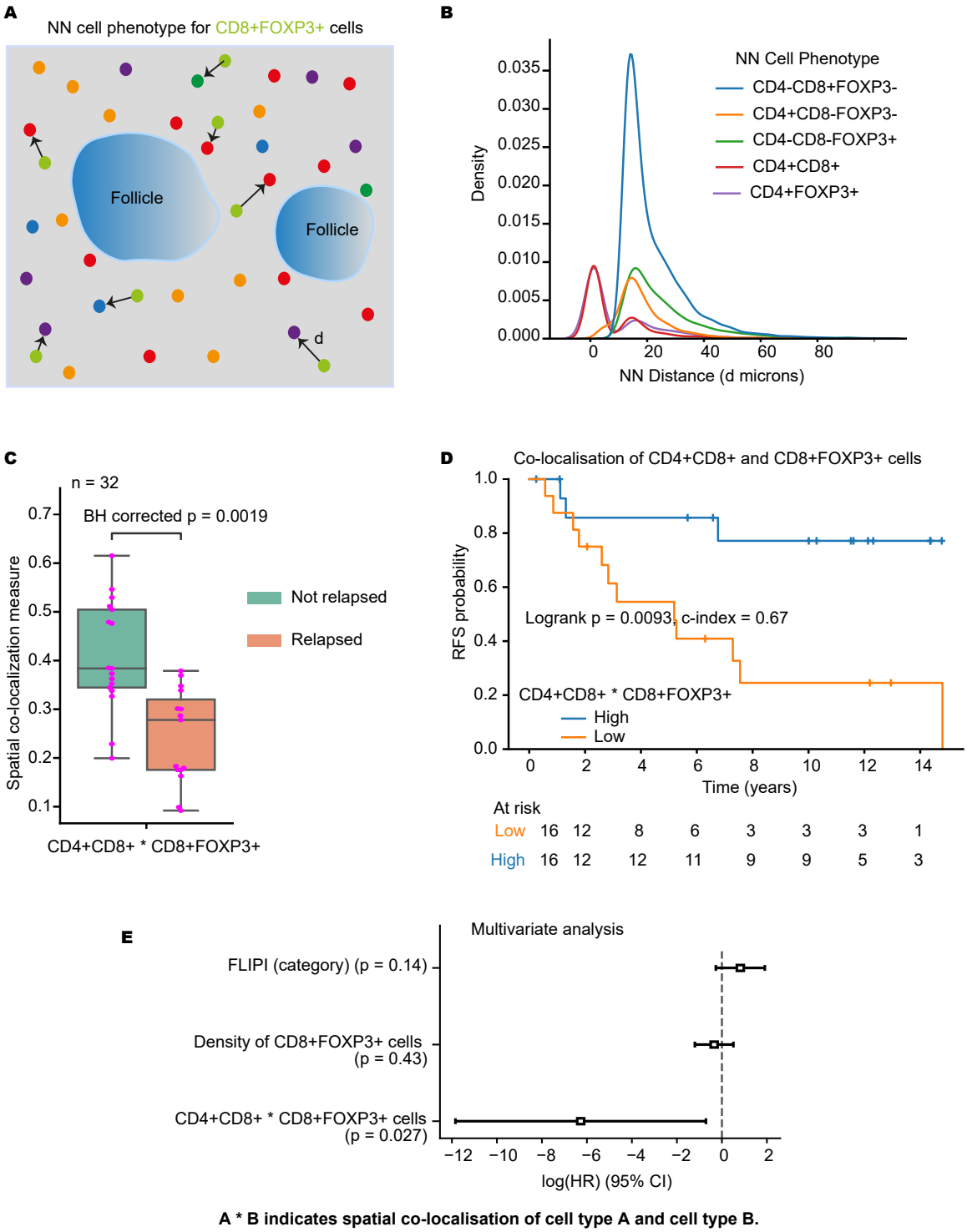


Figure 3.7: **Prognostic spatial co-localization of cell populations outside the neoplastic follicles. A)** Graphical representation of CD8+FOXP3+ cells nearest neighbour (NN) cells outside the neoplastic follicles. A dot represents a cell and the different colours indicate different cell phenotypes as shown in the legend in (B).

Figure 3.7: **B)** The distribution of the distance of NN cells of different phenotypes. **C)** Boxplot showing the difference in co-localisation of CD8+FOXP3+ with CD4+CD8+ cells outside follicles between relapsed (n=15) and not relapsed (n=17). **D)** Kaplan-Meier curves illustrating recurrence free survival (RFS) of patients dichotomised using median co-localisation of CD8+FOXP3+ with CD4+CD8+ cells outside follicles. The c-index indicates the concordance index between the observed survival times and predicted survival times. **E)** Forest plots showing multivariate Cox regression analyses. Continuous values were used for the density and spatial localization parameters. Follicular lymphoma international prognostic index (FLIPI). For statistical comparisons among groups, a two-sided, nonparametric, unpaired, Wilcoxon signed-rank test was used, unless stated otherwise. To correct for multiple testing, we applied Benjamini-Hochberg (BH)

3.5.5 Multivariate analysis to predict the risk of relapse in follicular lymphoma

To investigate whether the inter-follicular density of CD8+FOXP3+ and co-localisation of CD8+FOXP3+ with CD4+CD8+ are predictors of RFS independent of FLIPI, we applied multivariate Cox regression analysis. For the regression analysis, continuous values of the density and spatial co-localisation scores were used. Tumours with low inter-follicular co-localisation of CD8+FOXP3+ cells with CD4+CD8+ were at a significantly increased risk of relapse compared with tumours with a higher inter-follicular co-localisation of these cell types ($p = 0.027$, $HR = 0.0019$, 95% confidence interval (CI) [7.19×10^{-6}, 0.49], Figure 3.7E) that was independent of FLIPI and density of CD8+FOXP3+ cells. Moreover, both inter-follicular CD8+FOXP3+ cells density and co-localisation of CD8+FOXP3+ cells with CD4+CD8+ were not associated with FLIPI scores (Figure S2.6A, B). However, there is a positive correlation between CD8+FOXP3+ density and co-localisation of CD8+FOXP3+ with CD4+CD8+ (Figure S2.6C). Similarly, a low inter-follicular density of CD8+FOXP3+ was associated with an increased risk of relapse independent of FLIPI ($p = 0.038$, $HR = 0.42$, 95% CI [0.19, 0.95], Figure S2.6D), but not independent of co-localisation of CD8+FOXP3+ with CD4+FOXP3+ cells ($p = 0.42$, Figure 3.7E).

3.6 Discussion

In this chapter, we employed a deep learning based image processing pipeline and spatial statistical analysis for MIF images to decipher the immune microenvironment in FL. To the best of our knowledge, this is the first study to analyse the distribution and spatial interaction of immune cells in the inter-and intra-follicular compartments of FL using high throughput MIF images and deep learning image analysis. In FL, the abundance and distribution of immune cells within and outside neoplastic follicles are distinct and heterogeneous [44], and thus the spatial interaction of the cells. The combination of a high-plex MIF technique, deep learning-based image analysis, and spatial ecological measures focused on the intra- and inter-follicular tissue compartments enabled us to

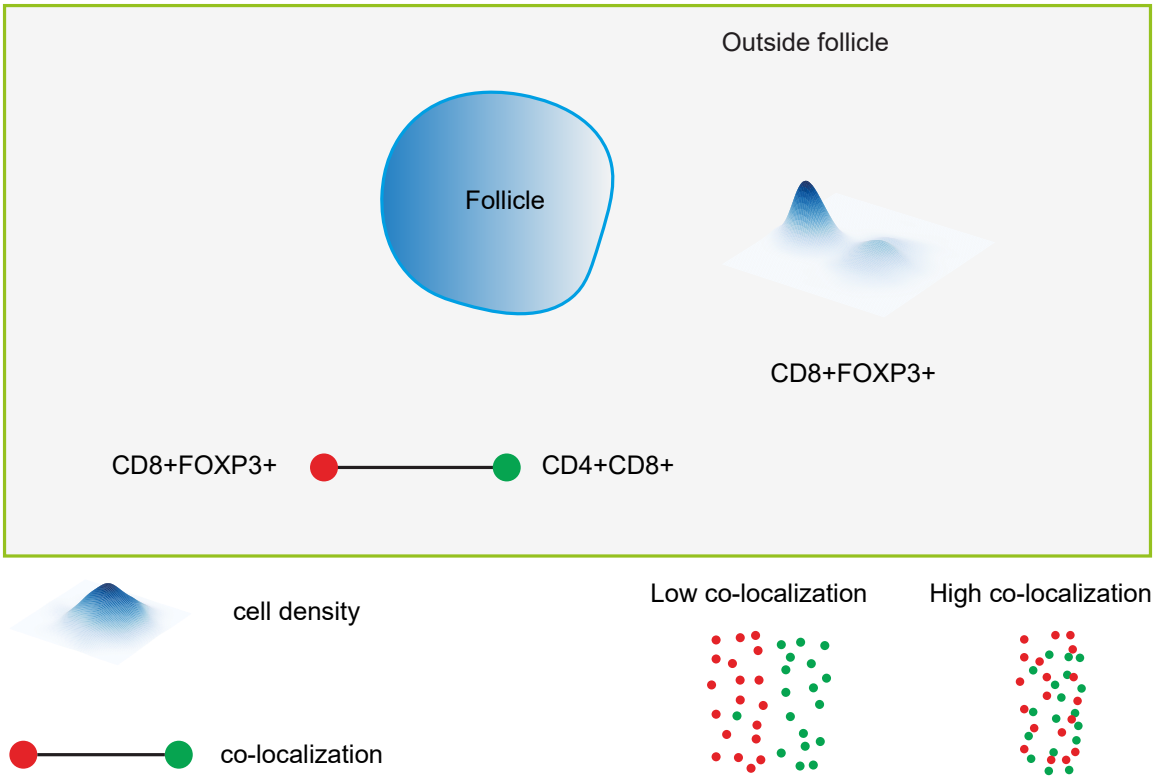


Figure 3.8: **The cartoon is a map of the two FL immune ecosystems:** within and outside the neoplastic follicles indicating the clinically relevant cell co-localization and the density of immune cell subsets.

identify prognostic neoplastic morphometric features, cell populations, and cells spatial patterns in FL.

During the development of FL, neoplastic follicles invade the lymph node capsule, extend into the perinodal adipose tissue, and replace the lymph node structure [21]. In contrast to the typical reactive follicles engaged in immunological reactions, neoplastic follicles are rounded and uniform in appearance [21]. However, neighbouring neoplastic follicles sometimes combine to create a diffused pattern of disease [21], which could be a sign of aggressive lymphoma [307]. Our data shows that there is a borderline increase in the neoplastic follicles' shape irregularity in patients with adverse clinical outcomes and these shape irregularities were due to the merging of expanding neoplastic follicles (Figure S2.10A, B). Though the current FL diagnosis is mainly based on the cell morphology [21, 307], the neoplastic follicles morphologic assessment could give additional insight into the diagnosis of FL.

In terms of immune cell infiltration, our study shows that the inter-follicular CD8+FOXP3+ T cells are prognostic and positively correlate with patients RFS. Even though these cells account for a small fraction of CD8+ immune T cells in the TME of FL, it has been shown that rare cell types such as antigen-specific T cells can play a crucial role in the development of cancer [308, 309]. In 1970, Gershon and Kondo described a pool of CD8+ regulatory T cells that support tumorigenesis [310]. This type of cells were subsequently described in prostate [311, 312], colon [313] and non-small cell lung cancer [312]. In a mouse model, Mayer et al. also showed that CD8+FOXP3+ cells have a light suppressive function [314]. However, other studies supported our results and showed that CD8+FOXP3+ cells have anti-tumour cytotoxic activity. Using flow cytometry on mice treated with Granulocyte-macrophage colony-stimulating factor secreting HER-2/neu vaccine, CD8+FOXP3+ T cells were abundantly found in regressing and immunogenic tumours [315]. Mayer et al. [314] and Le et al. [315] showed that CD8+FOXP3+ is a phenotype for anti-tumour T cells, and such cells have a similar expression profile to activated T cells.

Triggering an effective immune response promotes the expansion of CD8+FOXP3+ lymphocytes [315]. In a mouse model, Le et al. demonstrated that CD4+ T cells promote the expansion of tumour-specific T cells such as CD8+FOXP3+ cells by secreting stimulatory cytokines like interleukin-2 and transforming growth factor- β [315]. Moreover, K. Y et al. [311] showed that CD8+FOXP3+ T cells are immunosuppressive, but their inhibitory function could be altered using toll-like receptors-8 signalling [311, 316] suggesting this could be utilised by immunotherapeutic strategies in cancer [311, 316]. Furthermore, it is reported that toll-like receptor signalling pathways interact with R-CHOP immunochemotherapy that is used in FL [317]. Further functional studies are needed to understand whether the CD8+FOXP3+ T cells in the TME of FL have an "innate" anti-tumour function or if this is modulated by exposure to the immunochemotherapy treatment.

To investigate the spatial interaction of inter-follicular CD8+FOXP3+ cells with other cell types identified by our approach, we applied spatial co-localization analysis. We found that higher co-localization of CD8+FOXP3+ cells with CD4+CD8+ in the inter-follicular regions is associated with favourable outcome in FL. Previous studies described CD4+CD8+ cells as effector anti-tumour T cells

in a series of tumour studies including cutaneous T-cell lymphoma [318, 319], nodular lymphocyte-predominant Hodgkin lymphoma [320], and melanoma [319]. Nana et al. [318] showed that the CD4+CD8+ T cells have a high interleukin-2 cytokine secretion profile [318] and interestingly, a high level of interleukin-2 is reported to enhance the cytotoxic activity of CD8+ Tregs cells [321]. These results might suggest that CD4+CD8+ cells might have boosted the anti-tumoral activity of CD8+FOXP3+ cells through an interleukin-2 dependent pathway and thus resulting in a favourable patient outcome. Interleukin-2 treatment was approved by the U.S. food and drug administration in 1998 to treat advanced stage myeloma [322]. In a study by Smith et al [323], high-dose bolus intravenous interleukin-2 was administered to 684 individuals with metastatic melanoma either alone or in combination with other melanoma vaccinations. The response rate was about 13% in the patients who received interleukin-2 alone, and they observed a 3% increase in the response rate for the patients who received additional vaccines. In another study by Davar et al. [324], for patients with advanced myeloma treated by interleukin-2, the overall response rate was 18%. Though it is beyond the scope of this PhD study, studying the efficacy of interleukin-2 treatment for FL patient and how this is associated with the spatial organization of CD8+FOXP3+ cells and CD4+CD8+ could be interesting future directions.

This study has some limitations. The sample size is small. FL is characterised by indolent nature with highly variable clinical spectrum [325] and our results should be interpreted taking into consideration the sample size. Secondly, the manual annotation of the neoplastic follicles is laborious which could limit scalability to a larger cohort of studies. Thus an automated deep learning methodology to segment neoplastic follicles is a valuable development.

3.7 Conclusion

We showed that the combination of MIF, deep learning and regional spatial analysis is a promising strategy to identify novel immune cell phenotypes in FL that could stratify relapsed versus not relapsed FL patients, and predict the risk of relapse. We showed that low inter-follicular density of CD8+FOXP3+ cells or low inter-follicular co-localisation of CD8+FOXP3+ cells with CD4+CD8+ cells (Figure 3.8) is associated with relapse and shorter RFS in FL patients treated with R-CHOP or R-CVP. The inter-follicular density of CD8+FOXP3+ and co-localisation of CD8+FOXP3+ with CD4+CD8+ appear to be predictive of RFS independent of FLIPI score, and combining these features with FLIPI scores could improve FL prognostication. These findings require validation on a large cohort of FL patients treated with the same or different treatment regimens.

In this chapter, we showed that the spatial organization of immune cells is more prognostic than cell abundance using multiplex staining technology and spatial analysis algorithms tailored to the complexity of FL tissue compartments. In the next chapter, we will look at spatial mapping of tissue architectures and immune landscape of multiple myeloma across disease stages and post-treatment.

Chapter 4

The mosaic microenvironment of myeloma bone marrow trephine biopsies mapped by deep learning

4.1 Overview

As discussed in Section 1.3, BM trephine biopsy is crucial for diagnosing and studying multiple myeloma. However, the complexity of tissue preparation for BM trephine biopsy and the heterogeneity of cellular, morphological, and spatial architecture preserved in trephine samples hinders accurate evaluation.

We used multi-panel multiplex immunohistochemistry staining of CD4/CD8/FOXP3/BLIMP1 markers to identify regulatory T cells, effector T cells and myeloma tumour cells. To dissect the diverse cellular communities and mosaic tissue habitats, we developed a superpixel-inspired deep learning method (MoSaicNet) that adapts to the complex tissue architectures, and a cell imbalance aware deep learning pipeline (AwareNet) to enable accurate detection and classification of rare cell types using a cell weighting mechanism. MoSaicNet and AwareNet achieved an area under the curve of >0.98 for classification on separately held test data. We demonstrated how the application of MoSaicNet and AwareNet enabled novel investigation of bone heterogeneity and thickness using an autoencoder-based method, as well as spatial histology analysis of BM trephine samples from Monoclonal Gammopathies of Undetermined Significance (MGUS), paired Newly Diagnosed Multiple Myeloma (NDMM) and post-treatment patients.

The most significant difference between MGUS and NDMM samples was not related to cell density, but spatial heterogeneity. We also observed significantly fewer BLIMP1+ cells in spatial proximity to CD8+ cells in MGUS compared with NDMM samples ($p = 0.036$). Compared to NDMM samples, bone heterogeneity was decreased in post-treatment samples ($p = 0.01$), with a concomitant reduction in FOXP3+CD4+ regulatory T ($p = 0.004$) and BLIMP1+ plasma ($p = 0.013$) cells.

Spatial analyses of BM trephine using histology, deep learning, and tailored spatial statistics algorithms allow us to address questions about tumour topography and explore how the spatial distribution of immune cells may relate to disease progression and treatment response. In summary, deep-learning-based spatial mapping of BM complements can provide new insights into the myeloma marrow microenvironment.

4.2 Introduction

As discussed in Section 1.3, MM is an incurable haematological malignancy characterised by the uncontrolled proliferation of abnormal plasma cells in BM [71, 326, 84]. According to the International Myeloma Working Group, the current diagnosis of MM is based on the presence of clonal neoplastic plasma cells and organ dysfunction, of which the most common is bone destruction [84]. This is mostly investigated by BM aspirate, trephine biopsy samples, and whole body non-invasive imaging [84].

Increasingly, there is growing appreciation that myeloma is not driven by malignant plasma cells in isolation, but tumour growth is accompanied by global immune dysregulation in MM [327, 328]. These include impaired T cell effector function [329] and antigen presentation [330] and an increase in suppressor cells such as Treg cells [331–333]. These studies were based on MM blood/BM aspirates or MM cell lines employing flow cytometry and gene expression analysis, and not using biopsies that preserve the architecture of the BM. Visualisation and quantification of the spatial arrangement of immune cells and malignant plasma cells in the BM microenvironment and understanding the complicated mechanisms behind these realities could give new insight MM [105].

BM trephine biopsy is a tissue sample that allows examination of BM cellular and morphological environment. Multiplex staining of such tissue samples could allow spatial mapping of the BM of MM patients. Moreover, deep learning methods, specifically CNNs, have been shown to accurately identify complex visual patterns and single cell phenotypes in histopathological images without handcrafted features [114, 334, 335]. This offers a unique opportunity to harness the cellular and non-cellular mosaic spatial ecology of BM. Despite many studies in solid tumours employing multiplex spatial staining techniques, this has not yet been realised in MM [105]. Thus, the spatial relationship between BM cell types in MM has not yet been studied. This is due to the unique tissue integrity and morphology of BM trephine samples. The BM trephine samples are intrinsically different from those of solid tumours and require a specialised sampling process and decalcification (Figure 4.1A, B) [105].

The BM also has a highly organised structure, being a specialised haemopoietic and immunological organ [336]. Thus, the spatial context of cell-to-cell interactions is likely to be crucially important in the development of immunity. The BM is one of the priming sites of T cells and contains both rare and abundant cell types (Figure 4.1C) [337]. Deep learning methods are sensitive to the biases in the data unless carefully designed. Thus, there are new challenges in developing reliable automated

analysis for BM trephine samples due to possible biases in cell abundance and tissue architecture complexity.

Here, we proposed to use MIHC staining of immune T cells and MM tumour cells in situ on BM trephine biopsy samples. Subsequently, we developed new deep learning-based image analysis and spatial statistical analysis pipelines addressing the above challenges to explore the spatial heterogeneity of MM. Thus, in this Chapter we aimed:

- To implement cell detection and classification deep learning framework that uses cell weighting mechanism to accurately identify rare and abundant cell types on MIHC images;
- To develop a machine learning workflow to dissect the mosaic tissue microenvironment of BM trephine samples;
- To develop an automated computational image analysis pipeline to analyse bone heterogeneity and bone thickness from digitised trephine biopsy and understand the effect of our therapies on bone density and heterogeneity;
- To develop single cell spatial statistical analysis methods tailored to the tissue complexity of BM trephine biopsies to generate new insights into biology and function of tumour and non-tumour cells in-situ, unveiling their dynamic changes across disease stages and post-treatment.

4.3 Materials

4.3.1 Patients studied

All patients were managed at University College London Hospital (UCLH). BM trephine biopsies from two cohorts of patients were studied: nine patients with MGUS and ten with MM. Patients with suboptimal tissue samples (less amount of tissue) were excluded. For the second group, we studied MM sections at diagnosis prior to treatment initiation and also at repeat BM biopsies taken at 100 days (D100) following ASCT. All patients provided written informed consent for this project. Ethical approval was granted by the Health Research Authority, U.K. (Research ethics committee reference: 07/Q0502/17).

Patient characteristics for the MGUS group are shown in Table 4.1. The median age was 61 years, and 56% were male. The majority had IgG MGUS (56%), three had IgA MGUS (33%), and one had kappa light chain MGUS (11%). Five patients were deemed to have a low risk of MM progression (56%), two (22%) had intermediate risk, and two (22%) had a high risk [338].

The characteristics of the ten patients in the MM group are described in Table 4.2. The median age at MM diagnosis was 56 years, consistent with an age group that would usually proceed with treatment following induction therapy. Six (60%) patients were male, five had IgG disease (50%), and half had standard cytogenetic risk by IMWG criteria. Four patients (40%) had ISS stage I disease,

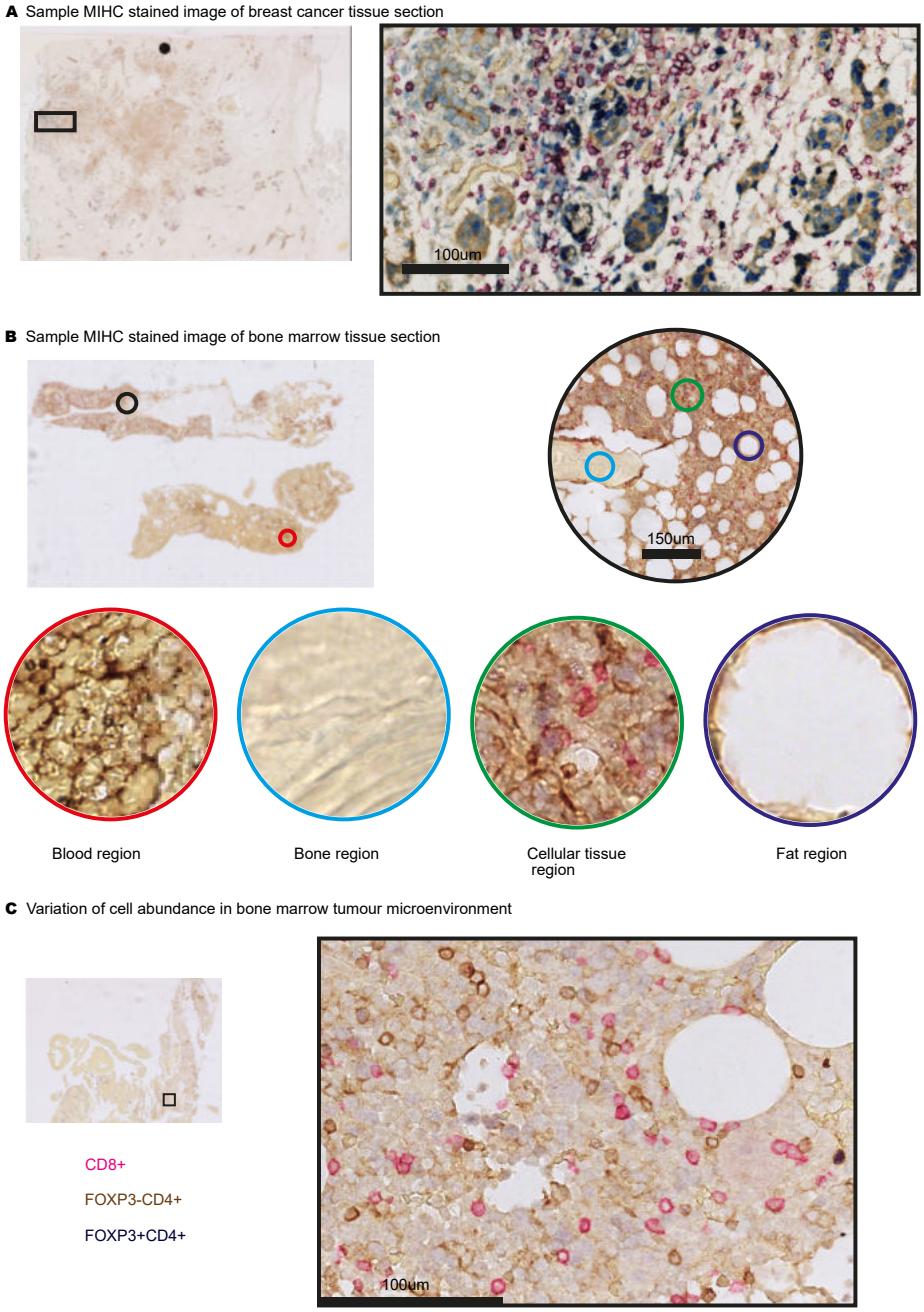


Figure 4.1: **Challenges in analysing tissue sections of BM images: A,B)** Sample MIHC images comparing the tissue composition of tissue sections from breast cancer sample (A) and BM trephine sample taken from a patient with multiple myeloma (B). The breast cancer tissue section was taken from our Gal8 dataset described in Section 2.3.1. These sample images show the complexity of BM trephine tissue architecture compared to tissue samples from solid tumours (e.g., breast). The glsbm trephine tissue image is a mosaic of blood, bone, cellular tissue and fat areas. **C)** The BM is a habitat for rare and abundant cell types. For example, FOXP3+CD4+ cells are rare compared with CD8+ and FOXP3-CD4+ cells.

Table 4.1: **Patient characteristics: MGUS**

Patient characteristics (n=9)	**Patient no. (%)**
Age at diagnosis	
Median (range)	61 (54, 89)
Gender	
Male	5 (56)
Immunoglobulin (Ig) isotype	
IgG	5 (56)
IgA	3 (33)
Light chains only	1 (11)
Light chain isotype	
Kappa	5 (56)
Lambda	3 (33)
Polytypic	1 (11)
IMWG Cytogenetics risk	
Standard risk	5 (56)
High risk	1 (11)
Unknown	3 (33)
Risk categories for progression to MM	
Low	5 (56)
Intermediate	2 (22)
High	2 (22)

five (50%) had stage II, and one (10%) had stage III [339]. All patients received first-line carfilzomib, cyclophosphamide and dexamethasone (KCD) induction therapy, followed by Melphalan 200mg/m2 as a conditioning regimen prior to ASCT.

4.3.2 Tissue processing

BM samples were collected and processed as per ICSH guidelines [340]. They were first fixed in neutral buffered formalin and then decalcified with formic acid. After decalcification, biopsy specimens were embedded in paraffin wax and cut on a microtome at 2–3μm. Serial sections were cut and mounted on glass slides.

4.3.3 Immunohistochemistry staining

The MIHC staining was performed using the fully automated Leica Bond slide stainer. Each slide was serially stained to identify three different antigens using different membranous or nuclear stains. The details of antibodies used are in Table S3.1. Two MIHC multiplex panels were used in this study. Panel 1 included T cell markers CD4 and CD8, as well as FOXP3, a transcription factor specifically expressed by CD4+ Treg cells [341]. A sample image of panel 1 is shown in Figure 4.2A. Panel

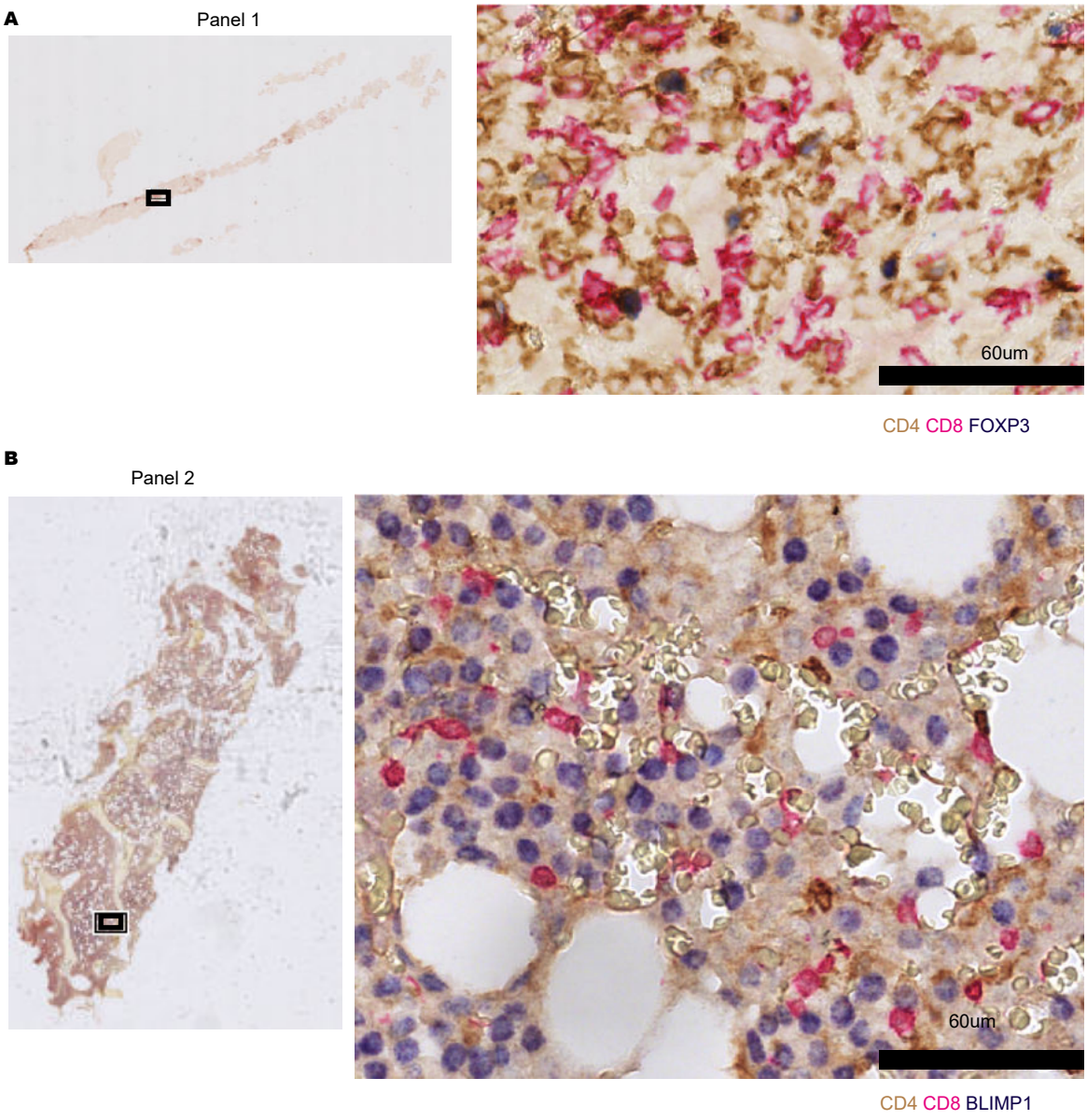


Figure 4.2: **Sample multiplex immunohistochemistry stained bone marrow trephine samples: A)** Panel 1: CD4/CD8/FOXP3 markers. **B)** Panel 2: CD4/CD8/BLIMP1 markers.

Table 4.2: **Patient characteristics: Paired diagnostic and post-treatment samples.** Carfilzomib, cyclophosphamide, dexamethasone (KCD); plasma cells (PC)

Patient characteristics (n=10)	**Patient no.** (%)
Age at diagnosis	
Median (range)	56 (53, 63)
Gender	
Male	6 (60)
Immunoglobulin (Ig) isotype	
IgG	5 (50)
IgA	2 (20)
Light chains only	3 (30)
Light chain isotype	
Kappa	7 (70)
Lambda	3 (30)
IMWG Cytogenetics risk	
Standard risk	5 (50)
High risk	5 (50)
IMWG ISS staging	
I	4 (40)
II	5 (50)
III	1 (10)
PC % in diagnostic bone marrow biopsy	
Median (range)	70%(13%, 80%)
Line of therapy at treatment	
1	10 (100)
Induction therapy	
KCD	10 (100)
PC % at D100 bone marrow biopsy post-treatment	
Median (range)	0.5% (0%, 10%)

2 comprised CD4, CD8 and B lymphocyte-induced maturation protein-1 (BLIMP1), an important transcriptional repressor required for plasma cell formation that is also expressed on MM cells [342] and their survival [343, 344]. We have used BLIMP1 to identify MM tumour cells as the more commonly used antigen, CD138, is cytoplasmic and so unsuitable for combining with CD4 and CD8 stains for T cells. A sample image of panel 2 is shown in Figure 4.2B. Staining protocols can be found in Table S3.2 and S3.3. Stained slides were then scanned using the Hamamatsu Nanozoomer s360 scanner and analysed by the deep learning models. The optimisation of the panels and slides scanning were done by my collaborators from Prof. Kwee Yong's lab at University College London Cancer Institute: Catherine SY Lecat, University College London Cancer Institute, Research Department of Haematology and Dominic Patel, University College London Cancer Institute, Research Department of Pathology.

In MIHC, colour is the main discriminant feature used by machine learning algorithms. To avoid collecting new single-cell annotations and training separate models for each panel, we used the same colours for protein expression on the MIHC panels. The markers and their corresponding col can be found in Table S3.2 and S3.3.

4.4 Methods

Unlike solid tumours, BM trephine sections consist of isolating structural elements over different spatial scales, reflecting a mix of cellular communities and mosaic habitats. To dissect this complex tissue landscape and detect rare cells in MIHC (Figure 4.1), we specifically designed two deep learning methods, MoSaicNet (**Mo**rphological analysis with **S**uperpixel-based h**abi**tat detection **Net**work) to dissect the mosaic landscape of BM tissue (Figure 4.3A) and AwareNet to detect and classify cells (Figure 4.3B). First, to dissect the MM tissue into blood, bone, fat, and cellular tissue patches/habitats, a superpixel-based deep learning method was designed to capture the complex landscape (Figure 4.3A). To train and validate MoSaicNet, we collected expert segmentation annotations for 260 regions, which resulted in 69884 superpixels (Table 4.3). Subsequently, we were able to quantify the amount of cellular tissue, which served as an important quality control parameter, to determine whether a slide would be considered for further analysis.

To optimally detect and classify cells within BM trephine samples, that contain both rare (e.g FOXP+CD4+) and abundant cells (Figure 4.1C). Thus, to optimally detect and classify these cell types, we developed AwareNet [345]. Subsequently, we analysed the BM spatial microenvironment in terms of cell density, cell ratio, cell spatial proximity and clustering, and bone physiology in terms of bone density heterogeneity, and bone thickness (Figure 4.3C). The next sections provide the details of methods developed in this Chapter.

4.4.1 Preprocessing of whole slide images

The MIHC were scanned at 40× magnification with a pixel resolution of $0.23\mu m$ per pixel. A representative image has a $40,000 \times 40,000$ pixel size at 40× magnification. For efficient image processing, the images were downscaled to 20× magnification. The images were further divided into "tiles" of size $2,000 \times 2,000$ pixels, which could be loaded into memory.

4.4.2 MoSaicNet: Segmenting BM trephine components using deep learning and superpixel

The digital image of the BM trephine is a mosaic landscape of background, blood, bone, cellular tissue, and fat region (Figure 4.1B). To automatically segment these regions, we developed MoSaicNet (Figure 4.3A). MoSaicNet contains superpixel extraction and a CNN superpixel classifier. The next sections describe the development of the MoSaicNet pipeline.

MoSaicNet training and validation data preparation

To train, validate, and test MoSaicNet, we collected 260 regions of interest from 19 samples (Table 4.3) annotated by expert pathologists (Figure S3.1A) from the different regions whole tissue section MIHC images. These annotated regions were split into training, validation and testing sets. The training (47%), validation (31%), and testing (22%) split were randomly done at the patient level. These annotated regions were extracted from the WSIs and divided into superpixels using the simple linear iterative clustering (SLIC) superpixels algorithm [346] (Figure 4.3A). The SLIC algorithm groups neighbouring pixels with similar pixel intensity into one superpixel. The shape of the superpixels is controlled by the compactness (C) parameter of the SLIC algorithm. The number of superpixels depends on the size of the image and the parameter k as shown in Equation 4.1. The values C and k are optimised by a user to ensure superpixels are capturing homogeneous pixels and bounding to region boundaries in the image under consideration depending on the scenario [346, 347]. The number of superpixels (n) was computed using Equation 4.1.

$$n = \left\lceil \frac{Image\ area}{k} \right\rceil \tag{4.1}$$

where the symbol $\lceil \rceil$ represents ceil operator. Upon visual assessment, superpixels with $k = 2,000$ and $C = 30$ best adhere to the boundaries of tissue and fat regions. This resulted in about 40×40 pixel ($18.4\mu \times 18.4\mu$) sized superpixel regions (Figure 4.3A). After applying SLIC, we generated $69,884$ superpixels from the 260 regions (Table 4.4). These superpixels belonged to blood, bone, fat and tissue classes. The class label of the superpixels was the class of the region it belongs to. Then, we implemented and trained a custom-designed CNN to automatically classify these superpixel regions.

Table 4.3: **Number of manual region segmentation annotation data.**

	No. patients	**No. regions**
Training	9	126
Validation	6	83
Test	4	51

Table 4.4: **Number of superpixels extracted from the human annotation of blood, bone, fat, and tissue regions.**

Category	**No. Patients**	**No. regions**	**Number of superpixels**			
			Blood	**Bone**	**Fat**	**Tissue**
Training	9	126	4,560	12,991	10,523	14,338
Validation	6	83	1,913	5,642	4,484	6,103
Test	4	51	1,091	2,275	1,626	4,338

Deep superpixel classifier

For a superpixel-based classifier, Konstantinos et al. [347] showed a shallow model performs comparably with VGG [237], InceptionV3 [271], and ResNet50 [348] models while reducing the number of parameters by about 10 times. Thus, we implemented and trained a custom-designed network to automatically predict the class of the superpixel regions. The proposed CNN based superpixel classifier consists of convolution layers with $\{16, 32, 64\}$ neurons followed by two dense layers with $\{200, 4\}$ neurons. All convolutions were performed using a 3×3 filter and followed by a max-pooling layer with a receptive field of 3×3 pixels. To minimise the chance of overfitting, a dropout layer ($rate = 0.3$) was used between the dense layers. ReLU activation was applied to all layers, but Softmax was for the last layer to transform the tensors to probabilities. Parameters were initialised using uniform Glorot [275] and optimised using Adam [276] using a learning rate of 10^{-4}. We used categorical cross-entropy loss with class weighting. The model was trained for 500 epochs with the patience of 50 epochs.

MoSaicNet pipeline post-processing method

The output of the classifier is a 4-dimensional class probability vector. The output class/label will be the class with the maximum probability. To convert class labels into segmentation images, all the pixels within the superpixels were assigned the same label/colour. The pixels within a superpixel have similar intensity values, and the superpixels are irregularly shaped polygons. To smooth the prediction, we applied a morphological closing operation (Equation (4.2)) with a structuring element (s) of a disk with a radius of 20 pixels.

$$I_{out} = (I_{in} \oplus) \ominus s \tag{4.2}$$

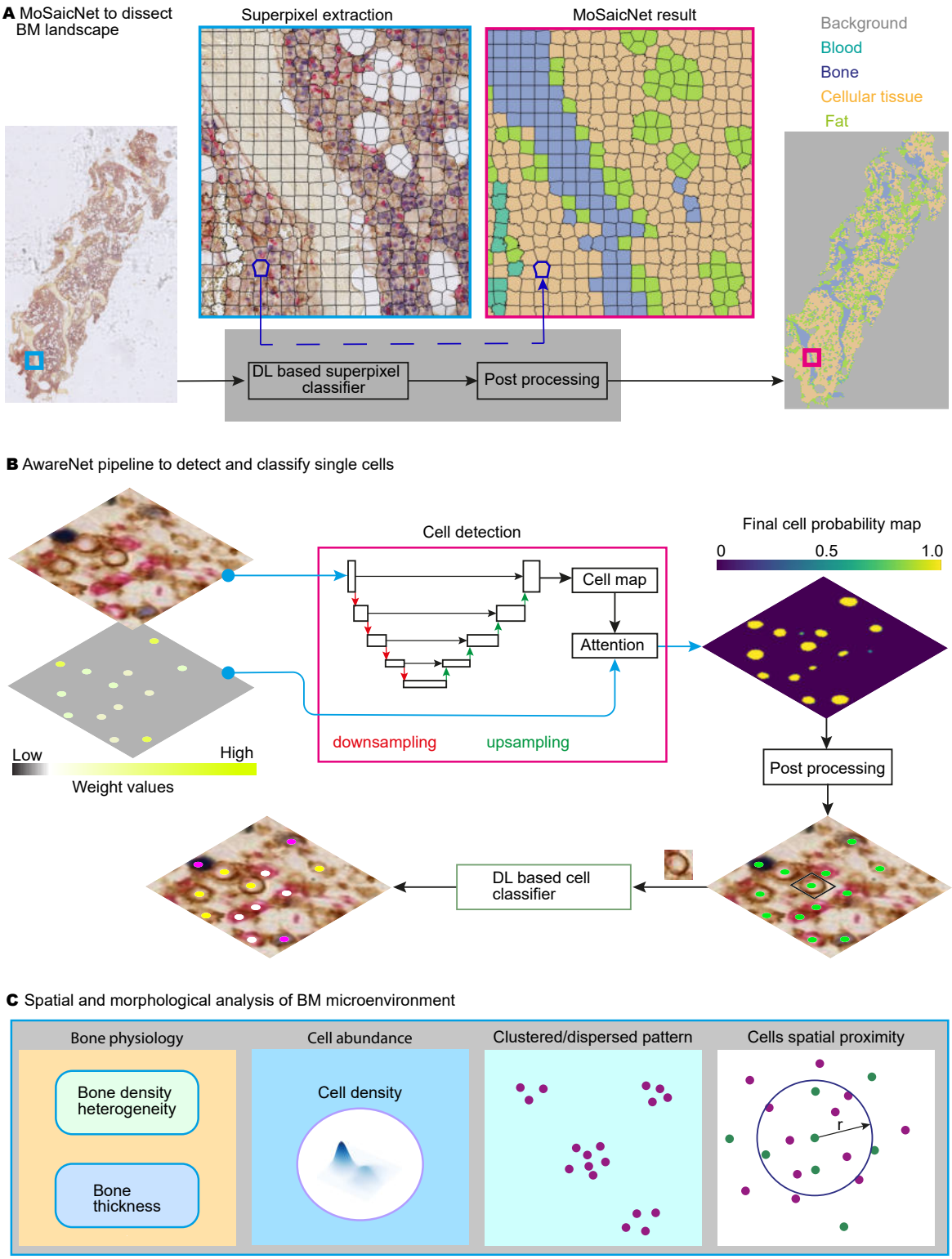


Figure 4.3: **Overview of computational deep learning and image processing pipelines for BM MIHC images: A)** MoSaicNet pipeline. The polygons (black) indicate superpixels. MoSaicNet dissects a tissue section into bone, blood, fat, and cellular tissue regions.

Figure 4.3: **B)** AwareNet method which utilises cell weights for cell detection and classification. The weight image pixel values were generated from the abundance and spatial location of the cell types using the pathologist's manual annotation. A weight image was applied to the objective function during model parameter optimisation to regularise the algorithm by assigning high weight to rare cell types. The cell detection algorithm generates a cell probability map. A post-processing algorithm was developed to find the cell nucleus centre, (x, y) location, from the probability map. A patch centred on each cell was extracted and fed to deep learning (DL) based classifier to infer its class. **C)** Spatial and morphological analysis of BM trephine samples. Bone density heterogeneity was investigated using an auto-encoder-based machine learning method. We used spatial proximity analysis to study the spatial relations of cells. r = radius. Cell density refers to the number of cells per unit of tissue area.

where I_{in}, I_{out}, $\oplus$ and $\ominus$ are input image, output image, dilation operation, and erosion operation, respectively.

The trained model was applied to WSIs to quantify the amount of cellular tissue in the image. The amount of tissue present in the image was used as quality control for further analysis. Moreover, to speed up the processing time, cell detection and classification models were applied only to the cellular tissue region of the image.

Bone density representation learning using a convolutional auto-encoder

To understand bone density heterogeneity, the bone region of BM WSI was divided into superpixels and transformed into feature vectors that represent the semantic information. Auto-encoder and patch-based approaches were used to learn the representation of WSIs. However, pixels within a patch might have non-homogenous pixel values. Here, we divided the WSI into superpixels which contain homogenous pixel intensity instead of patches and applied a convolutional auto-encoder to learn the lower-dimensional representation of the superpixels.

The convolutional auto-encoder learns a low dimensional representation of superpixels such that it can recover the input from the representation (Figure S3.1B). The convolutional autoencoder consists of encoder and decoder parts. The encoder transforms the superpixels into a low dimensional latent variable (learned representation) and the decoder reconstructs the input superpixels from the latent variable. The encoder consists of 4 convolutional layers with $\{8, 16, 32, 64\}$ neurons. Each layer is composed of a 2D convolution layer ($filter = 3 \times 3$, and $stride = 2$), LeakyReLU activation, and batch normalization. The decoder section consists of four layers with a reversed order of the number of neurons with a transposed 2D convolution layer instead of a 2D convolution layer. We experimented with $\{2, 8, 32, 64\}$ latent variable dimensions. To optimise model parameters, we applied the mean squared error loss function.

Table 4.5: **Distribution of training, validation and testing single cell annotation data.**

	No. slides	CD8+	FOXP3-CD4+	FOXP3+CD4+
Training	5	2,244	1,000	243
Validation	3	1,555	689	140
Test	3	1,304	692	135

We used a learning rate of 10^{-4} and a batch size of 64. Parameters were initialised using uniform Glorot [275] and optimised using Adam [276]. The model was trained for 500 epochs with a patience of 50 epochs.

4.4.3 AwareNet: weight-based deep convolutional network for cell detection and classification

The BM microenvironment is a home for both rare and abundant cell types [337]. To accurately detect and classify single cells on MIHC images, we developed a cell abundance aware weight-based deep learning based single cell identification model. AwareNet is published in IEEE International Symposium on Biomedical Imaging (ISBI) 2021 [349]. A schematic overview of AwareNet is displayed in Figure 4.3B. The models were trained and validated using a manual pathologist's annotation.

Single cell annotation

To train, validate and test our proposed deep learning-based single cell detection and classification models, we first collected 8,004 single cell annotations from 11 samples by expert pathologists (Figure S3.1A) using web-based WSI viewer and annotation tool developed in our lab (not published). The single cells annotations belonged to CD8+, FOXP3-CD4+ and FOXP3+CD4+ cells. Figure 4.4A shows sample cells belonging to these classes. The training (46%), validation (27%) and test (27%) split was done randomly at the patient level to ensure that cells from the same patients are not included in different categories (Table 4.5).

Data preparation for cell detection

For training, the annotated regions were divided into $256 \times 256 \times 3$ patches. Let n be the number of training patches, the training data, T_d was represented by a set

$$T_d = \{\mathbf{I},\ \mathbf{R},\ \mathbf{W})\} = \{(I^1, R^1, W^1), (I^2,\ R^2,\ W^2), (I^3,\ R^3,\ W^3), ..., (I^n, R^n, W^n)\}$$

, where $I^i \in R^{256\times256\times3}$, $R^i \in R^{256\times256\times1}$, and $W^i \in R^{256\times256\times1}$ are the i^{th} input, reference and weight images, respectively. Sample I, R and W images are shown in Figure 4.4B. The weight and reference images were generated from single cell nucleus dot annotation.

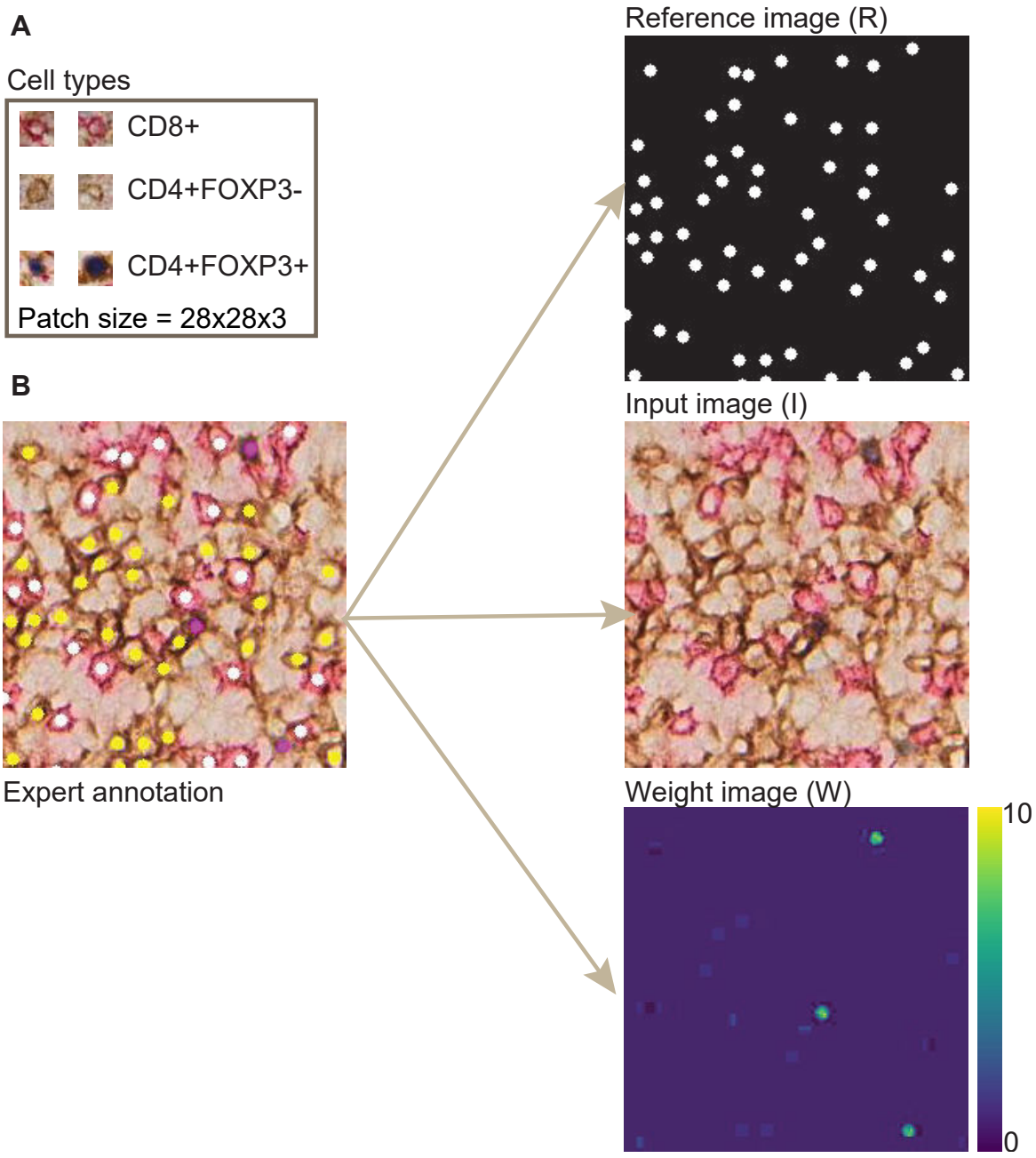


Figure 4.4: **AwareNet training data preparation: A)** Sample patches for all cell types. **B)** Sample annotated, reference (R) and weight (W) images for an input image (I). In W, a less abundant cell type is assigned a larger weight. FOXP3+CD4+ cells have a larger weight than FOXP3-CD4+ and CD8 cells.

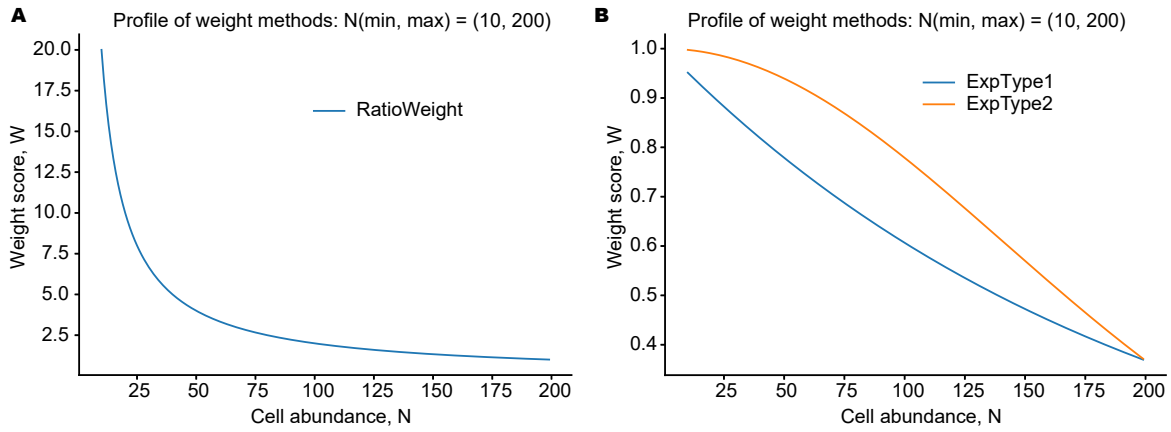


Figure 4.5: **Profile of cell weight image generating methods: A)** Simulated profile of ratio weight image generating method. The x-axis represents the abundance of cells ranging from 10 up to 200. This is for simulation purposes only, in reality, the number of cells in each class is more than the maximum value in this range. **B)** Simulated profile of ExpType1 and ExpType2 weight image generating method.

Reference image

The reference image is the ground truth binary image with a pixel value of 1 at the nucleus centre and zeros elsewhere. It is an artificial image generated from the expert single cell dot annotation using Equation (4.3).

$$R(i,j) = \begin{cases} 1 & \text{if } d < r \\ 0 & otherwise \end{cases} \tag{4.3}$$

where $R(i,j)$ is pixel value at (i,j) and d is an Euclidean distance from (i,j) to the closest cell centre. The value of r was set to 4 pixels ($1.768\mu m$). The value of r was chosen empirically, making sure blobs in R do not touch each other (Figure 4.4B).

Weight image

The idea of weight image was inspired by the class weighting method in classification tasks. In the class-weighted classification approach, class weights are generated based on the abundance of data in each class [350]. Higher weight is assigned to the under-represented class. Then, while optimising the classifier, the error in each sample is scaled by the weight assigned to its class.

Here, we extended this idea to cell detection on MIHC images. Similar to the classification task, the weights were inferred from the relative abundance of each cell type in the training data of the pathologist's annotation. Rare cells are given larger weight. Let n be the number of cell types in the dataset and suppose $N = \{N^1, N^2, N^3, ..., N^n\}$ represents a set of the abundance of the n cell types in

the training set. Then, using the spatial location of the cells, and the weight scores, we generated a weight image. We implemented three different cell abundance weighting strategies:

- **RatioWeight**: This approach assigns weight 1 for the most abundant cell class and a weight great than 1 to other classes (Figure 4.5A). Once these scalar weights are computed for each class, to create a weight image (W), which has the same size as the input image I, we applied Equation 4.4.

$$W(i,j) = \begin{cases} \frac{max(N)}{N^k} & \text{if } d^k < r \\ 1 & otherwise \end{cases} \tag{4.4}$$

  where, the list (i,j) represents pixel location and d^k is a Euclidean distance from k^{th} cell type centre to pixel location (i,j). This is iteratively applied to all annotated cells in the image. The value of r was set to 4 pixels (1.768 μm). Pixels outside the cell nucleus centre are assigned a value of 1, the same weight as the cells of the most abundant class.

- **ExpType1**: This is a negative exponential weighting that squashes the weights to a range between 0.37 and 1 as shown in Equation 4.5 and Figure 4.5B. Pixels outside the cell nucleus centre are assigned a value of $exp^{-1} = 0.37$, the same weight as the cells of the most abundant class.

$$A(i,j) = \begin{cases} \exp -\frac{N^k}{max(N)} & \text{if } d^k < r \\ \exp -1 & otherwise \end{cases} \tag{4.5}$$

- **ExpType2**: This is a negative square exponential weight method which as lower decay rate as shown in Equation 4.6 and Figure 4.5B. Pixels outside the cell nucleus centre are assigned a value of $exp^{-1} = 0.37$, the same weight as the cells of the most abundant class.

$$W(i,j) = \begin{cases} \exp -(\frac{N^k}{max(N)})^2 & \text{if } d^k < r \\ \exp -1 & otherwise \end{cases} \tag{4.6}$$

For our dataset, the cell classes are CD8+, FOXP3-CD4+, and FOXP3+CD4+, and the abundance of these classes was $2,244$, $1,000$ and 243, respectively. These cell abundance values were used to generate weight image (W) using Equations (4.4 - 4.6). The weight image shown in Figures 4.4B was generated using the RatioWeight method in Equation 4.4.

4.4.4 AwareNet cell detection model architecture

The schematics diagram of the AwareNet cell detection pipeline is shown in Figure 4.6. It is a U-Net [65] based CNN inspired by InceptionV3 [271]. We applied inceptionV3 blocks which extract multi-scale features at a given layer. The model has an encoder and decoder part. The encoder learns

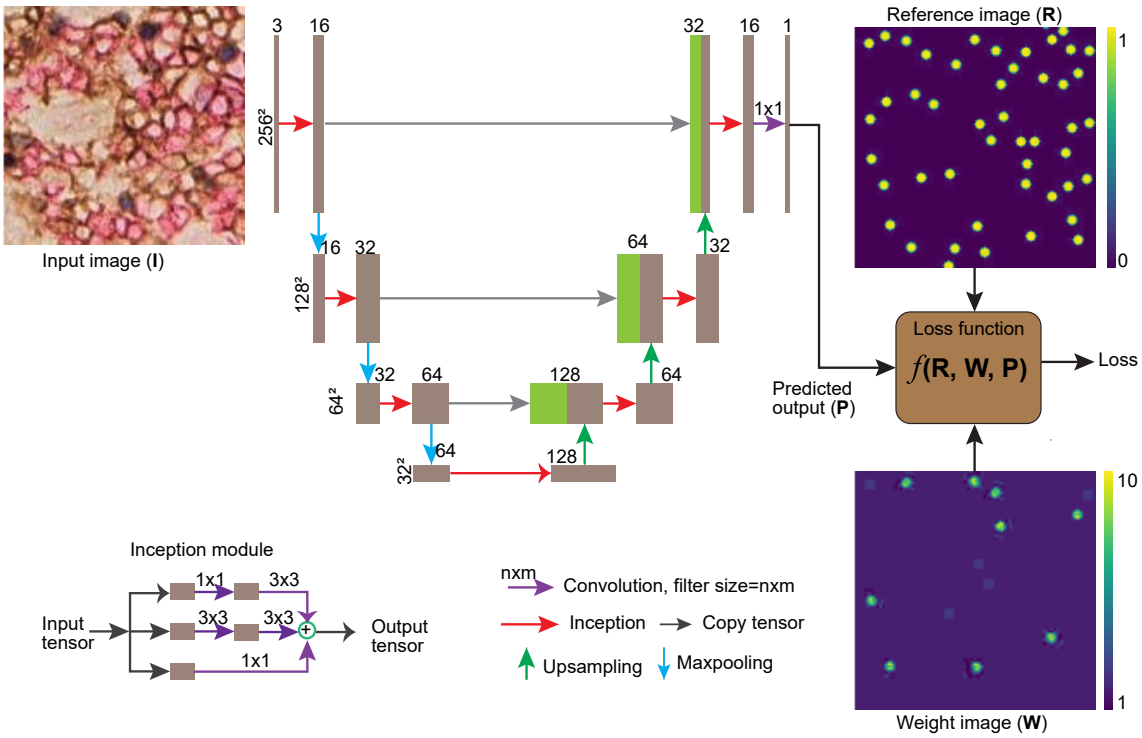


Figure 4.6: **Schematic of AwareNet cell detection model:** The number on the top and side of the blocks indicate the size and spatial dimension of the features, respectively. The reference image (R), which is a binary image, is a ground truth cell nucleus map. The weight image (W) is a cell weight image to penalise model prediction error during training. The output of U-Net (P) architecture is a cell nucleus centre probability map. During model training, the prediction error (loss) was computed as a function of W, P and R.

a low-dimensional representation of the input image, and the decoder reconstructs a target image. The 1×1 convolutional layer at the end of the architecture transforms $256 \times 256 \times 16$ dimensional features to $256 \times 256 \times 1$, size of the reference image (R). Parameters were initialised using uniform Glorot [275]. We used Adam optimizer [276] with a learning rate of 10^{-4} and a Dice overlap loss weighted by the weight image. The output (Y) in Figure 4.6 is a cell centre probability map.

4.4.5 AwareNet training objective function

The AwareNet cell detection model was trained using a reference image (R) as a ground truth cell nucleus map. The output of the AwareNet model (P) is the cells nucleus centre probability map which has the same dimension as the input image. During model training, the prediction error (loss) was computed using P, W and R as shown in Equation 4.7. The error is weighted Dice overlap loss. The weight image (W) penalises model prediction error during training. It applies a high penalty to errors in less abundant cell types, minimising the effect of cell class imbalance during cell detection.

$$D_l = 1 - \frac{2\sum(W \odot R \odot P) + \varepsilon}{\sum(W \odot R) + \sum(W \odot P) + \varepsilon} \tag{4.7}$$

where W, R and P are the weight, reference and predicted output images, respectively. The $\odot$ operator represents element-wise matrix multiplication. The summation operator adds all the elements of the matrix and generates a single value. The value $\varepsilon = 10^{-5}$ was added to ensure computational stability when the denominator is zero.

AwareNet post-processing: From cell probability map image to cell location

As shown in Figure 4.3B, AwareNet generates cells nucleus centre probability map image. Then, using the image post-processing algorithm, we extracted the spatial coordinates of the centre of the cell's nucleus.

To convert the probability map image to a binary image, we applied a threshold of 0.8. To fill holes in the binary image, we applied morphological closing as shown in Equation 4.8 using disk structuring element (S) of radius 5 pixels.

$$I_{out} = (I_{in} \oplus) \ominus S \tag{4.8}$$

, where I_{out} and I_{in} are are input image and output images, respectively. The $\oplus$ and $\ominus$ are dilation and erosion operations, respectively. After the threshold, sometimes some artefacts are detected as the cell nucleus. To remove these false positive predictions, we excluded objects with an area smaller than 10 $pixel^2$. In the cleaned image, each object represents a cell. An object is a set of connected pixels with a pixel value of 1. Then, the centre x and y location of each cell were extracted using *region_props* function of Scikit-image Python library [351].

All the threshold hyper-parameters were optimised on validation data by maximizing the cell detection F1-score. The AwareNet cell detection pipeline generates the (x,y) positions of the cells in the image space and saves it to an excel file.

To identify the type of the detected cell, we extracted a $28 \times 28 \times 3$ patch centred on the cell nucleus (Figure 4.3B) and applied a custom-designed CNN classifier which will be explained in the next section.

4.4.6 Cell classification model

To train a cell classification model, we extracted $28 \times 28 \times 3$ patches as shown in Figure 4.4A from pathologist annotation. For cell classification, we applied a custom-designed shallow VGG [237] style architecture, which contains three convolutional layers with $\{16, 32, 64\}$ neurons followed by two dense layers with $\{200, 3\}$ neurons. Each convolution layer was configured as follows: filter size (3×3), stride (1) and ReLU activation function. To reduce the dimensionality of the feature map, each convolution layer was followed by a max-pooling layer with 2×2 receptive field and stride of 2. ReLU activation was used for the first dense layer and Softmax for the last layer to transform the tensors into probabilities. Parameters were initialised using uniform Glorot [275] and optimised using Adam [276] with a learning rate of 10^{-4}. We applied categorical cross-entropy loss with class weighting explained in Equation (4.4 - 4.6).

Visualisation of features learned by classifier

A CNN classifier model has feature extraction and classification parts. To visualise the features learned by a classifier model, we took the features at the output of the first dense layer of the classifier section, which is a 200-dimensional vector. To reduce the dimensions of features into 2D without losing information, we applied manifold approximation and projection (Umap).

4.4.7 Cell density

We used cell density to statistically compare the abundance of cells between patients in the different clinical groups. Cell density is measured as the number of cells per unit of tissue area. Cell abundance could be confounded by the amount of cellular tissue present in a given tissue slide, which could bias the analysis. Cell density normalises the raw cells count by the amount of available tissue (Equation 4.9), removing the bias.

$$Cell\ density = \frac{Cell\ count}{Tissue\ area} \tag{4.9}$$

4.4.8 Cells proximity analysis

We investigated the spatial proximity of a pair of cell types (e.g., BLIMP1+ MM plasma cells and CD8+ T cells) within the BM microenvironment as follows (Figure 4.3C): Consider a tissue section that contains k number of type A cells located at $\{a_i, i \in \{1,2,3,\ldots,k\}\}$ and m number of type B cells located at $\{b_j, j \in \{1,2,3,\ldots,m\}\}$. Each cell has an (x,y) position. The number of type B cells within a distance r from type A cell was computed using Equation (4.10 and 4.11).

$$N_{prox(b\to a)} = \frac{1}{k}\sum_{i=1}^{k}\frac{\sum_{j=1}^{m}\Omega}{\Phi_i} \tag{4.10}$$

$$\Omega = \begin{cases} 1, & \text{if } d(a_i,b_j) \le r \\ 0, & \text{otherwise} \end{cases} \tag{4.11}$$

where d is the Euclidean distance function between the two cells, a_i and b_j. The Φ_i is a normalizing factor, which is the total number of cells (all types) within r distance from a_i. In BM trephine samples, there is a huge variation in the tissue architecture caused by the prevalence of non-cellular regions such as bone and fat regions (Figure 4.1B). Moreover, in single cell-based spatial analysis, the density of cells could be a confounding factor. Incorporating Φ_i is aimed at correcting for these factors while doing spatial analysis.

4.4.9 Bone density heterogeneity

To learn the low dimensional representation of bone superpixels, we custom-designed a convolutional auto-encoder (Figure S3.1B). For ease of visualisation and applying unsupervised clustering algorithms on the representation of bone superpixels, we applied Umap dimensionality reduction.

Then, we applied a clustering algorithm to divide the latent representation space into smaller regions. Kmeans and Gaussian Mixture Models (GMM) are the most commonly used clustering algorithms. We applied GMM to detect bone superpixel clusters in the embedding space due to its flexibility to cluster shapes [352]. To determine the number of clusters, we used the Akaike information criterion and the Bayesian information criterion from the Scikit-Learn python package [353] (Figure S3.3E-F). A cluster contains superpixels with similar bone density. The clustering enabled us to identify artefact bone superpixels with input from an expert histopathologist with whom I have been collaborating on this project (Dr. Manuel Rodriguez- Justo, Manuel Rodriguez- Justo; Research Department of Pathology, University College London Cancer Institute). These clusters were excluded from further analysis.

To quantify the heterogeneity (H) of bone density within a slide, we computed the maximum variance (Var) of the latent representations of all superpixels within the slide using Equation (4.12).

$$H = max(Var(Umap1), Var(Umap2)) \tag{4.12}$$

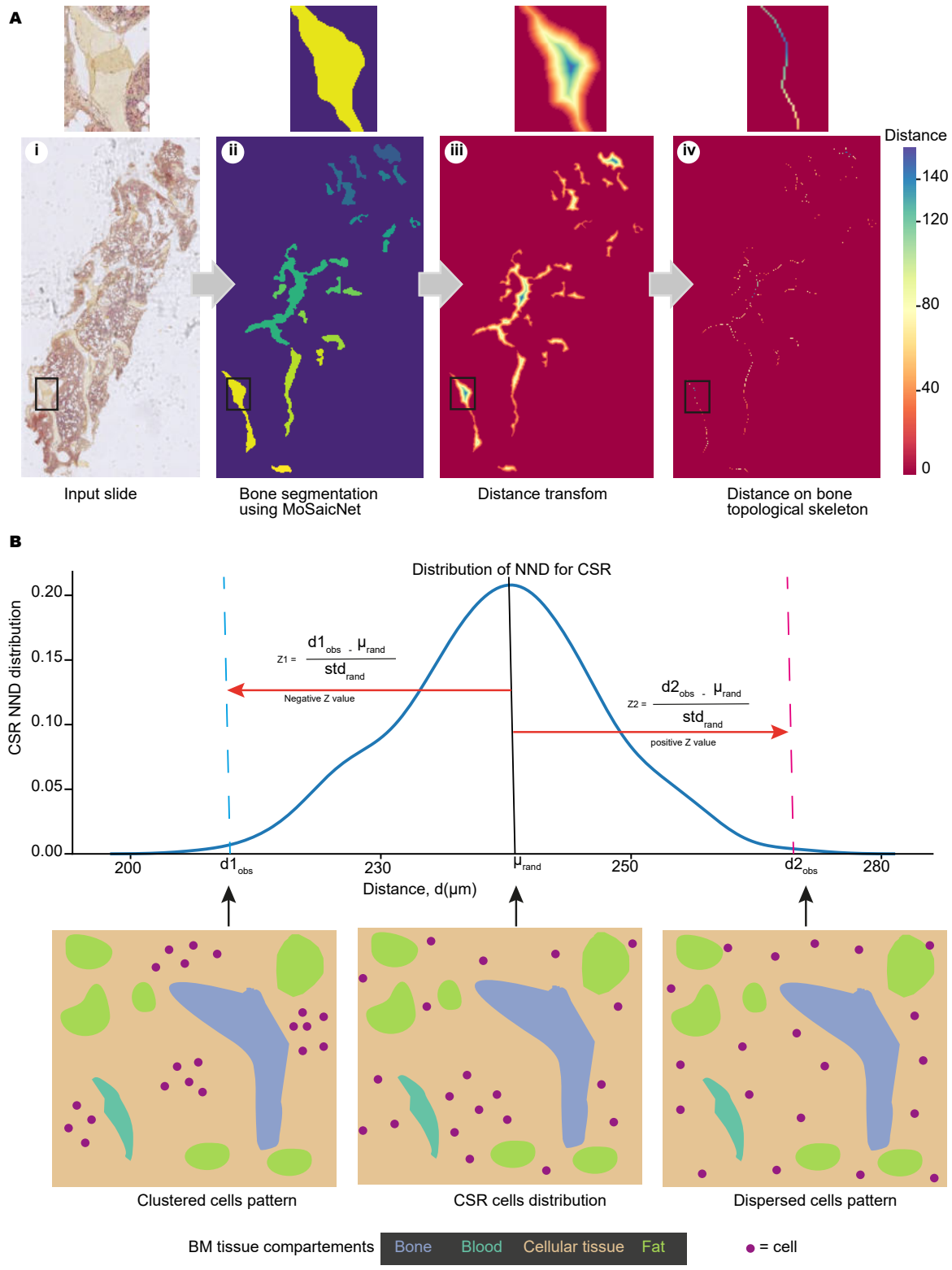


Figure 4.7: **Computational methods for bone thickness and cells clustering analysis: A)** Image analysis to estimate bone thickness. The bone segmentation (ii) is an output of MoSaicNet, and each bone is displayed in a different colour. The colour bar shows the pixel intensity of the image in (iii and iv).

Figure 4.7: The pixel intensity on the skeleton indicates half of the bone thickness. **B)** Cells infiltration pattern analysis using nearest neighbour distance (NND) and the null hypothesis of complete spatial randomness (CSR). The blue curve shows the distribution of NND for randomly distributed cells. It is a normal distribution with mean NND of μ_{rand}. When there is clustered pattern of cells compared to CSR, the mean NND is less than μ_{rand} (for example, $d1_{obs}$, obs: observed). This results in a negative z-score. On the other hand, when there is a scattered pattern of cells compared to CSR, the mean NND is higher than μ_{rand} (for example, $d2_{obs}$). This results in a positive z-score. The $d1_{obs}$ and $d2_{obs}$ represent the mean NND of the cell distribution in their respective panel. The values $z1$ and $z2$ are the z-scores for $d1_{obs}$ and $d2_{obs}$, respectively. $Z < -1.96$, $Z > 1.96$, and $-1.96 \leq Z \leq 1.96$ indicates a clustered, dispersed, and random distribution of observed cells, respectively. std: standard deviation; μ: mean NND of CSR; d_{obs}: observed distance.

4.4.10 Automated machine learning algorithm to quantify bone thickness

The proposed method to quantify bone thickness is shown in Figure 4.7A. We extracted the bone regions from the output of MoSaicNet. To compute bone thickness, first, we applied distance transform [354], and medial axis transform [355] (Figure 4.7A). The Distance transform (**DT**) computes the minimum distance from bone pixels to non-bone pixels. The medial axis transform (**MAT**) generates the topological skeleton of the bone. The topological skeleton of a bone is a series of bone pixels which have more than one closest equ-distant non-bone pixel. The distance values on the topological skeleton show half the thickness (width) of the bone across its length. Within a given WSI, there could be multiple bones. For instance, the WSI in Figure 4.7A has 20 bone regions. The thickness of the bone could also vary along its length. The thickness of a bone is estimated as the mean thickness along its length (skeleton). Then, the bone thickness (B_T) for a given tissue sample is computed as the mean of the mean thicknesses of all bones within the sample. Mathematically, suppose a given tissue section has n bones. Each bone will have its corresponding topological skeleton after applying **MAT**. The thickness (T_j) of the j^{th} bone (B_j) could be computed using Equation 4.13.

$$T_j = \frac{2}{|L|} \sum DT(B_j) \odot MAT(B_j) \tag{4.13}$$

where $\odot$ represents element-wise matrix multiplication. $MAT(B_j)$ generates a binary image with pixel values of 1 on the skeleton of B_j and 0 elsewhere. $|L|$ is skeletal length of B_j and computed using Equation 4.14.

$$|L| = \sum MAT(B_j) \tag{4.14}$$

Thus, multiplying the result of DT and MAT yields an image with distance values in the topological skeleton of the bones as shown in Figure 4.7A(iv). These distance values represent half of the bone thickness along its axis and this was multiplied by 2 to get the full thickness. Finally, B_T is the average thickness of all bones present in the sample, which is computed using Equation 4.15.

$$B_T = \frac{1}{n}\sum_{j=1}^{n} T_j \tag{4.15}$$

4.4.11 Cell infiltration patterns: spatial clustering, dispersion, or random

Quantifying the degree of clustering or dispersion of cells in BM trephine samples is challenging as it can be confounded by the mosaic tissue architecture of the BM trephine (presence of non-cellular tissues, Figure 4.1B), cell abundance, and the amount of cellular tissue area.

In ecology, it was shown that the mean nearest neighbour distance (NND) of all pairs of variables shows the spatial organization of the variables [356]. The NND is the distance from a spatial point to its closest neighbour. Under the null hypothesis or complete spatial randomness (CSR), the distribution of NND is normal [356] (Figure 4.7B). Here, we used the concept of NND and the null hypothesis to identify the infiltration pattern of cells.

Let a given tissue section has k cells of type A (e.g BLIMP1+ cells), $C = \{c_i : i \in \{1,\ 2,\ 3,\ k\}\}$. Each cell has an (x, y) position attribute. The NND for cell c_i, the i^{th} element of C, is computed using Equation (4.16).

$$NND_i = d_{c_i,N(c_i)} \ \leq \ d_{c_i,c_j}\ \forall c_j \in C - c_i \tag{4.16}$$

where d_{c_i,c_j} is a Euclidean distance between the i^{th} and j^{th} cells. The $N(c_i)$ is the nearest cell to c_i. Then, the slide level observed NND was computed as mean NND over the k cells as shown in Equation 4.17.

$$NND_{obs} = \frac{1}{k}\sum_{i=1}^{k} NND_i \tag{4.17}$$

Under a null hypothesis, the k cells could be at any location in the cellular tissue space (T) (Figure 4.7B). Thus, the k cells, $C = \{c_i : i \in \{1,\ 2,\ 3,\ k\}\}$, where the position of the i^{th} cell $(x_i, y_i) \in$ T, which is a set of k randomly distributed cells across the cellular tissue region. The cellular tissue region was segmented using MoSaicNet. To get the distribution of NND for CSR of cells over the tissue region, we computed NND for 300 CSR iterations. Once the CSR cells are generated, NND computation in Equation 4.16 and 4.17 could be applied. We computed Z-score, to measure the difference between the NND for CSR cells and the NND of observed cells pattern (Equation (4.18)) (Figure 4.7B).

$$Z = \frac{NND_{obs} - \mu_{CSR}}{std_{CSR}} \tag{4.18}$$

where μ_{CSR} and std_{CSR} are the mean and standard deviation of NND for the CSR of cells. The value of Z shows the cell infiltration pattern relative to the random distribution of cells. $Z < -1.96$, $Z > 1.96$, and $-1.96 \leq Z \leq 1.96$ indicate clustered, dispersed and random distribution of observed

cells, respectively. P values computed from spatial proximity data using different distance values were corrected using the Benjamini-Hochberg correction method.

4.4.12 Statistical analysis

All statistical analyses were carried out using the Python programming language. All correlation values were measured using the non-parametric Spearman test. The p-values were computed using a two-sided unpaired (for MGUS vs NDMM) or paired (for newly diagnosed multiple myeloma (NDMM) vs post-treatment), non-parametric Wilcoxon method, considering $p < 0.05$ as significant.

4.4.13 Implementation and code availability

All methods and analyses were implemented in Python. For reproducibility and ease of sharing, the code and its dependencies are packed into a Docker container. The code runs on both local and high-performance clusters. Python implementation of AwareNet, MoSaicNet, and spatial statistical analysis could be found in this GitHub repository (BM-Spatial-Analysis).

4.5 Results

4.5.1 High accuracy of MoSaicNet classification model

To evaluate the performance of the MoSaicNet classification model, we used $9,330$ superpixels extracted from separately held manually annotated samples. The class labels include blood, bone, fat and cellular tissue. To measure the classifier's performance, we used accuracy, AUC, precision, recall and F1-score. The performance evaluation of the classification model is shown in Table 4.6. To estimate the 95% CI, we applied $1,000$ bootstraps, each bootstrap taking 80% of the $9,330$ superpixels using random sampling with replacement. A confusion matrix was used to visualise the proportion of correct and incorrect predictions by the model.

Taking all classes together, the superpixel classifier model achieved an AUC value of 0.99, 95% CI [0.989, 0.991] as shown in Table 4.6. Moreover, at the individual class level, the mean bootstrap AUC was > 0.984 for all the classes with a minimum 95% CI lower bound of AUC score of 0.983 for the bone and cellular tissue classes (Table 4.6, Figure 4.8A). The overall accuracy (unweighted) was 0.937, 95% CI [0.935, 0.94].

Out of the $9,330$ superpixels, 585 superpixels were misclassified. Out of the 585 misclassified superpixels, 208 tissue superpixels were misclassified as bone, and 122 bone superpixel patches were misclassified as tissue (Figure S3.2A). This is also evident reflected in the low precision value on bone class 0.88, 95% CI [0.87, 0.89], low recall value in bone class (0.933, 95% CI [0.93, 0.94]) and low recall value in cellular tissue class (0.932, 95% CI [0.93, 0.94]) (Table 4.6). Moreover, 88 tissue superpixels and 29 bone superpixels were misclassified as a fat class, and the precision score for the

Table 4.6: **MoSaicNet superpixel classifier performance evaluation.** The 95% CI was computed using 1000 bootstraps. Each bootstrap contained a randomly sampled 80% of the instances with replacement. The metrics are reported with the mean value from the 1000 bootstraps and 95% CI. All classes indicate blood, bone, fat, and tissue classes combined.

Metric	Mean, 95% CI	Class name
AUC	0.984, [0.983, 0.985]	Bone
Precision	0.88, [0.87, 0.89]	Bone
Recall	0.933, [0.93, 0.94]	Bone
F1-score	0.906, [0.9, 0.91]	Bone
AUC	0.999, [0.999, 0.999]	Blood
Precision	1.0, [1.0, 1.0]	Blood
Recall	0.933, [0.93, 0.94]	Blood
F1-score	0.966, [0.96, 0.97]	Blood
AUC	0.984, [0.983, 0.985]	Tissue
Precision	0.958, [0.95, 0.96]	Tissue
Recall	0.932, [0.93, 0.94]	Tissue
F1-score	0.944, [0.94, 0.95]	Tissue
AUC	0.993, [0.992, 0.994]	Fat
Precision	0.933, [0.93, 0.94]	Fat
Recall	0.954, [0.95, 0.96]	Fat
F1-score	0.943, [0.94, 0.95]	Fat
AUC	0.99, [0.989, 0.991]	All classes
Precision	0.943, [0.94, 0.945]	All classes
Recall	0.938, [0.935, 0.942]	All classes
F1-score	0.94, [0.935, 0.945]	All classes

fat class was 0.933, 95% CI [0.93, 0.94] (Table 4.6). The mean value of the F1-score of 0.91 was obtained for the bone class, and for the other classes, the mean value F1-score was 0.943. Taking all classes together, F1-score of 0.94, 95% CI [0.935, 0.945] was obtained (Table 4.6).

To visualise the separation of the classes using the features learnt by the classifier, we extracted the deep learnt features and applied Umap dimensionality reduction. This enables the inspection of misclassified superpixels in the embedding space. Most of the tissue superpixels misclassified as bone were superpixels with poor tissue quality, non-cellular regions, and bone bordering areas (Figure 4.8B). Most of the 122 bone superpixels that were misclassified as tissue were a result of background staining of the bordering area (Figure 4.8B).

4.5.2 AwareNet accurately detects and classifies rare and abundant cells

To accurately detect rare and abundant cells within the tissue section of MIHC stained BM trephine sample, we proposed AwareNet. AwareNet uses an weight mechanism to regularise cell detection model training. We evaluated the cell detection performance of our proposed weight generation

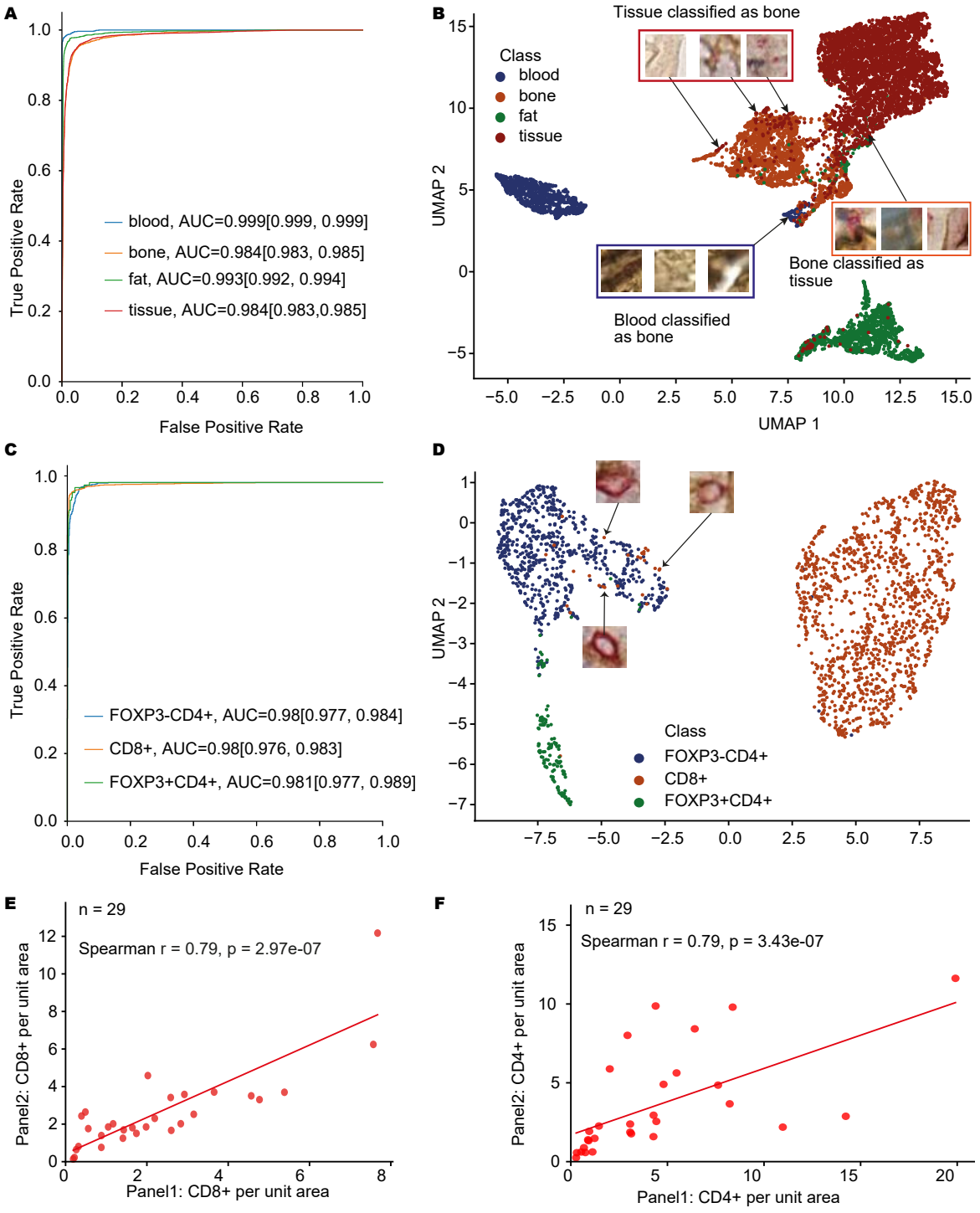


Figure 4.8: **Performance evaluation of MoSaicNet and AwarNet deep learning models:**

Figure 4.8: **A)** ROC curves and AUC scores of MoSaicNet superpixel classifier. AUC scores are displayed with mean and 95% CI. **B)** 2-dimensional mapping of superpixels using MoSaicNet learned 200-dimensional features after dimensionality reduction by Umap. **C)** ROC curves and AUC scores of classifier model on separately held test data. glsauc scores are displayed with mean and 95% CI. **D)** Umap features visualisation of deep learned features by AwareNet single cell classifier CNN. **E-F)** validation of AwareNet model using correlation of density of CD8+ **(E)** and CD4+ cells **(F)** in panel 1 and panel 2.

Table 4.7: **Cell detection performance of different models**. U-Net [65] model is a model in Figure 4.6 trained without applying weights. CONCORDe-Net [334] is our model developed in Chapter 2. Both CONCORDe-Net and U-Net models were trained using the data as the AwareNet models.

Method	**Precision**	**Recall**	**F1-score**
ExpType1 based AwareNet	**0.82**	0.75	**0.78**
RatioWeight based AwareNet	0.80	0.75	0.78
ExpType2 based AwareNet	0.78	**0.77**	0.77
CONCORDe-Net [334]	0.81	0.72	0.76
U-Net [65]	0.80	0.70	0.75

strategies and compared their performance with state-of-the-art methods such as U-Net [65] and CONCORDe-Net [334]. The comparison was done using precision, recall, and F1-score on separately held test images. CONCORDe-Net (described in Section 2.4.1) is cell count regularised CNN designed for cell detection for MIHC images[334].

Overall, the ExpType1 weight based AwarNet model that employs a negative exponential weighting strategy outperformed the other models. An F1-score of 0.78 was obtained using ExpType1 and RatioWeight based AwareNet models, a 3% increase compared to U-Net [65] and a 2% increase compared to CONCORDe-Net [334] as shown in Table 4.7. Moreover, the recall of the ExpType2 based AwareNet model was higher than U-Net [65] and CONCORDe-Net [334] baseline models by at least 5%. For the ExpType1 based AwareNet model, a detection was considered true positive if it is within 10 pixels ($4.42\mu m$) Euclidean distance to a ground truth annotation. For all models, the distance was optimised independently, maximising the F1-score. We used a distance range of $[3, 12]$ pixels as a search space. This suggests that class weighting improves cell detection performance.

AwareNet excels in detecting CD4+ FOXP3+ cells, which are rare in BM trephines (representing 7% of the training data) (Figure 4.9, Table S3.4). In our dataset, compared to CD8+ cells, there is less number of FOXP3+CD4+ cells. The visualisation in Figure 4.9 indicates a model with ExpType1 detected FOXP3+CD4+ cells, which were under-represented in the training dataset, while the model trained without any weight (U-Net) missed some of these cells. The detection results for CD8+ and FOXP3-CD4+ cells remain similar with and without weight.

To measure classification performance, we estimated accuracy, AUC, precision, recall and F1-score on separately held $2,131$ test images of cells. These cells belonged to CD8+, FOXP3-CD4+ and

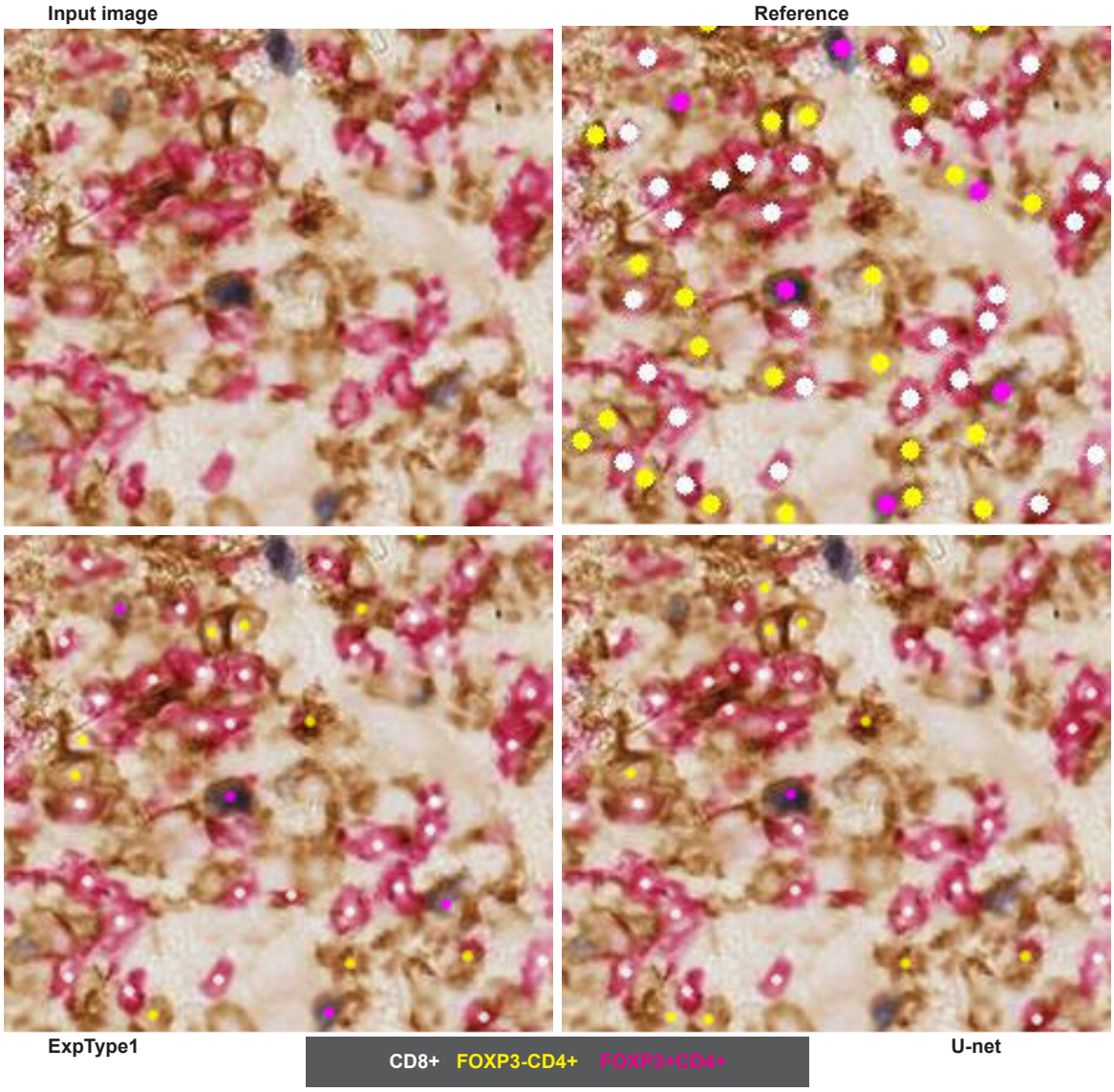


Figure 4.9: **Samples results from different cell detection methods.**

Table 4.8: **AwareNet single cell classification evaluation metrics.** The 95% CI was computed using 1,000 bootstraps. Each bootstrap contained randomly sampled 80% of the instances. The metrics are reported with the mean value from the 1,000 bootstraps and 95% CI. All classes indicate CD8+, FOXP3-CD4+ and FOXP3+CD4+ classes combined.

Metric	**Mean, 95% CI**	**Class name**
AUC	0.981, [0.977, 0.989]	FOXP3+CD4+
Precision	0.857, [0.83, 0.89]	FOXP3+CD4+
Recall	0.92, [0.9, 0.94]	FOXP3+CD4+
F1-score	0.887, [0.87, 0.91]	FOXP3+CD4+
AUC	0.98, [0.976, 0.983]	CD8+
Precision	0.98, [0.98, 0.98]	CD8+
Recall	0.98, [0.98, 0.98]	CD8+
F1-score	0.98, [0.98, 0.98]	CD8+
AUC	0.98, [0.977, 0.984]	FOXP3-CD4+
Precision	0.964, [0.96, 0.97]	FOXP3-CD4+
Recall	0.949, [0.94, 0.96]	FOXP3-CD4+
F1-score	0.956, [0.95, 0.96]	FOXP3-CD4+
AUC	0.98, [0.977, 0.984]	All classes
Precision	0.933, [0.923, 0.942]	All classes
Recall	0.949, [0.94, 0.96]	All classes
F1-score	0.941, [0.93, 0.95]	All classes

FOXP3+CD4+ classes (Table 4.5). For cell classification, there was no significant difference on AUC for the different weighting strategies. Figure 4.8C shows the performance of the cell classifier with ExpType1 weighting. Similar to the MoSaicNet classification performance evaluation, to estimate the 95% CI, 1,000 bootstraps were applied, with each bootstrap taking 80% of the data with replacement.

Taking all the three classes together, the single classifier model of AwareNet achieved an AUC value of 0.98, 95% CI [0.977, 0.984] as shown in Table 4.8. Moreover, at the individual cell class level, the mean bootstrap AUC value was > 0.98 for all the classes with a minimum 95% CI lower bound of AUC score of 0.976 for the CD8+ class (Table 4.8, Figure 4.8C). The overall accuracy was 0.965, 95% CI [0.962, 0.969].

Only 74 cells were misclassified out of 2131 cells (Figure S3.2B), resulting in 0.965, 95% CI [0.962, 0.969] accuracy (unweighted). 11 cells out of 135 FOXP3+CD4+ cells were misclassified as FOXP3-CD4+, and 12 FOXP3-CD4+ cells were misclassified as FOXP3+CD4+ cells (Figure S3.2B). This resulted in Precision (0.857, 95% CI [0.83, 0.89]), Recall (0.92, 95% CI [0.9, 0.94]), and F1-score (0.887, 95% CI [0.87, 0.91]) score for the FOXP3+CD4+ class (Table 4.8). For the FOXP3-CD4+ and CD8+ class, the F1-score was 0.956, 95% CI [0.95, 0.96], and 0.98, 95% CI [0.98, 0.98], respectively (Table 4.8). Moreover, when all classes combined, the classifier obtained an F1-score of 0.941, 95% CI [0.93, 0.95].

A Umap-based inspection of the misclassified FOXP3-CD4+ and CD8+ cells revealed that these cells were mainly cells co-expressing both CD8 and CD4 proteins (Figure 4.8D). We also found these rare cell types in follicular lymphoma [357], which is described in the previous Chapter but not studied in myeloma. Overall, the proposed cell detection model achieved high cell classification accuracy for both abundant and rare cell types.

AwareNet was trained on single cell data from CD4/CD8/FOXP3 panel data and directly applied to both panels, CD4/CD8/FOXP3 and CD4/CD8/BLIMP1. After applying the model to both panels, the numbers of CD8+ cells and CD4+ cells in both panels were significantly correlated ($r = 0.79$, $p = 2.97 \times 10^{-7}$ and $r = 0.79$, $p = 3.43 \times 10^{-7}$, Figure 4.8E-F, respectively), validating the reliability of AwareNet.

4.5.3 MoSaicNet reveals changes in bone physiology post-treatment

Using MoSaicNet, we quantified the proportion (%) of blood, bone, fat, and cellular regions in all sections (Figure 4.10A). In the NDMM group, trephine samples taken post-treatment contained a greater proportion of bone (%bone) when compared with diagnostic samples ($p = 0.037$, Figure 4.10B). There was also a borderline decrease in the %bone with age (Figure 4.10D). There was, however, no difference in the %bone between MGUS and NDMM (Figure 4.10C) or between male and female patients (Figure 4.10E). The proportion of fat (%fat) showed a borderline increase post-treatment compared with %fat at diagnosis (p = 0.05, Figure S3.3A) but was not different between MGUS patients and NDMM patients without effect of age or gender (Figure S3.3B-D).

To investigate the heterogeneity of bone structure in BM samples, we used a convolutional auto-encoder to learn the embedding of 177.6 thousand bone superpixels extracted from nine MGUS (27.8%), ten NDMM (34.4%) and ten post-treatment (37.8%) WSIs (Figure S3.3G). Bone superpixels were mapped into 32 feature vectors and clustered into 17 groups (Figure 4.10F, Figure S3.3E-F). Based on this grouping, there was a positive trend on the similarity of bone superpixels from MGUS to bone superpixels from post-treatment samples, however this was not significant (r = 0.4, p = 0.12, Figure 4.10G).

We then asked if the bone density differed between the patient groups. The intra-and inter-sample bone density heterogeneity in NDMM was significantly higher at diagnosis compared to post-treatment (p = 0.0098, Figure 4.10H-I). We observe a borderline difference in the heterogeneity of bone between NDMM and MGUS samples (p = 0.086, Figure 4.10H,J), but no difference between MGUS and post-treatment samples (Figure 4.10H and p = 0.87, Figure S3.3H).

Furthermore, to analyse bone thickness, we developed an automated image analysis algorithm (Figure 4.7A). The bone thickness of NDMM samples was similar to post-treatment samples (p = 0.23, Figure 4.10K) and MGUS (p = 0.37, Figure 4.10L) patients. The bone thickness in patients aged $\leq$ 58 years (median) was significantly higher compared with that in patients aged >58 years (p = 0.018, Figure 4.10M), without variation between gender (p = 1.0, Figure 4.10).

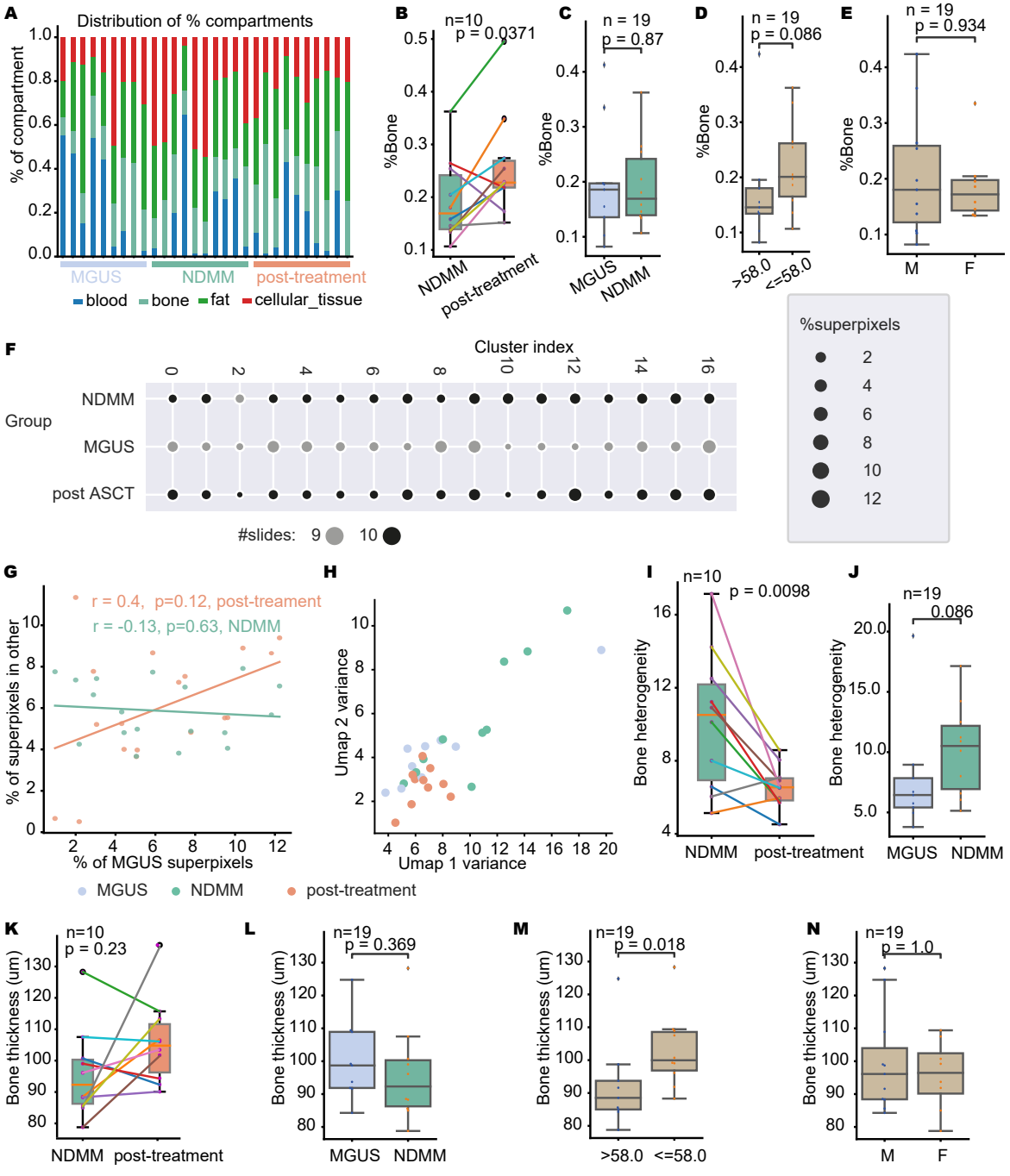


Figure 4.10: **Studying bone physiology using MoSaicNet: A)** Proportion of different compartments of BM trephine digital images. One stacked bar represents a sample.

Figure 4.10: **B-E)** Boxplots showing the difference in %bone between samples from NDMM and post-treatment **(B)**, MGUS and NDMM **(C)**, different age groups **(D)** (median age=58.0 years), and gender groups **(E)**. **F)** Scatter plot showing the distribution of the number of bone superpixels in 17 clusters from MGUS, NDMM and post-treatment samples. The size of the dot represent the percentage of bone superpixels in each cluster. To consider the abundance of bone superpixels in each group, the percentage was computed per group. For example, the size of the dot in cluster 0 for NDMM group indicates the percentage of bone superpixels from NDMM samples belonging to cluster 0. The clusters contain bone superpixels from different slides and the colour represents the number of slides in each cluster. **G)** correlation of percentage of superpixels in each cluster between different patient groups. A point represents a cluster. **H)** Scatter plot of slide-level heterogeneity of bone features measured by features variance. A point represents a patient/slide. **I,J)** box plots showing differences in bone density heterogeneity between NDMM and post-treatment **(I)**, and between MGUS and NDMM **(J)**. **K-L)** Boxplots showing the difference in bone thickness between samples from NDMM and post-treatment **(K)**, MGUS and NDMM **(L)**, and different age groups (median age=58.0 years) **(M)** and gender **(N)**.

4.5.4 Decreased FOXP3+CD4+ and BLIMP1+ cell density post-treatment

When comparing cells density on the NDMM and post-treatment samples, we observed a decrease in Treg cells (FOXP3+CD4+ cells), CD8+ T cells and BLIMP1+ myeloma cells following treatment ($p = 0.0039$, $p = 0.0039$ and $p = 0.013$, respectively, Figure 4.11A-C, K,J). However, the density of FOXP3-CD4+ T cell did not change post-treatment compared with diagnostic samples ($p = 0.77$, Figure 4.11D). We then compared the ratio of different cell types in the NDMM and post-treatment samples. The FOXP3+CD4+:FOXP3-CD4+ ratio is significantly reduced post-treatment compared with the ratio at diagnosis ($p = 0.0137$, Figure S3.5A), largely due to the reduction of the density of FOXP3+CD4+ cells post-treatment. However, the FOXP3-CD4+:CD8+ ratio, the FOXP3+CD4+:CD8+ ratio, CD8+:BLIMP1+, and CD4+:BLIMP1+ were not different between the two-time points (Figure S3.5B-E, respectively). The density of FOXP3+CD4+ cells was significantly correlated with the density of BLIMP1+ cells in the post-treatment (Spearman $r = 0.79$ and $p = 0.006$, Figure S3.4C) samples but not in NDMM samples (Spearman $r = 0.20$, $p = 0.58$, Figure S3.4C).

4.5.5 Increased spatial proximity between BLIMP1+ cells and CD8+ cells in NDMM compared to MGUS

The density of FOXP3+CD4+, FOXP3-CD4+ and CD8+ cells were not significantly different between MGUS and NDMM (Figure 4.11E, Figure S3.4A, B, respectively). Moreover, the FOXP3+CD4+:FOXP3-CD4+ ratio, FOXP3-CD4+:CD8+ ratio, and FOXP3+CD4+:CD8+ ratio were not significantly different between MGUS and NDMM (Figure S3.5F-H, respectively). However, there was a borderline increase in tumour burden measured by density of BLIMP1+ cells and BLIMP1+:CD4+ ratio in the NDMM sample compared to MGUS samples ($p = 0.08$, Figure 4.11F, and $p = 0.08$, Figure S3.5I,

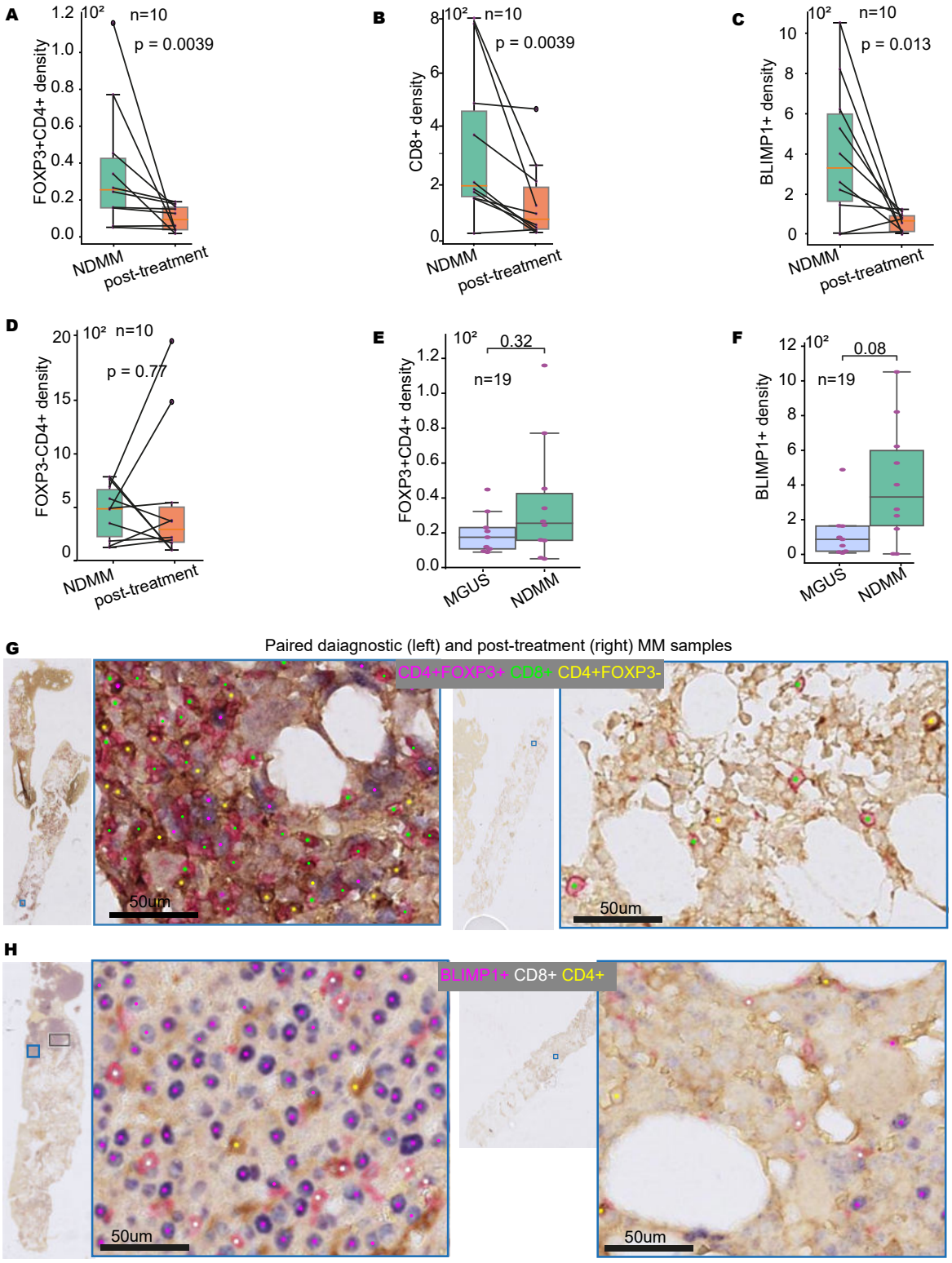


Figure 4.11: **Density of immune T cells and plasma cells in MGUS, NDMM and post-treatment samples:**

Figure 4.11: **(A-D)** Boxplots showing the difference in density of FOXP3+CD4+ cells **(A)**, CD8+ cells **(B)**, BLIMP1+ cells **(C)**, and FOXP3-CD4+ cells **(D)** between paired NDMM samples and post-treatment samples (n=10 pairs). **E-F)** Boxplots showing the difference in the density of FOXP3+CD4+ (**E**) and BLIMP1+ (**F**) cells between MGUS and NDMM samples (n=19). **G-H)** Sample images showing the reduction of the density of FOXP3+CD4+ (**G**) and BLIMP1+ cells **(H)** at post-treatment compared to paired NDMM samples. Cell density is presented per 1 mm^2 tissue area.

respectively), largely due to the increase in tumour burden. However, Furthermore, the ratio of the number of BLIMP1+ cells to CD8+ cells did not differ between MGUS and NDMM ($p = 0.165$, Figure S3.5J). The density of FOXP3+CD4+ cells was not correlated with the density of BLIMP1+ cells in MGUS (Spearman $r = 0.47$, $p = 0.205$, Figure S3.4C)

Next we asked if the spatial proximity between immune cells and BLIMP1+ plasma cells differed according to disease state and treatment. To demonstrate that the spatial analysis result is not dependent on the distance threshold chosen, cell proximity was calculated for a range of distances with the maximum distance set at the cell-cell communication distance of $250\mu m$ [358, 108] (30, 50, 100, 150, 200, 250)μm. Cell proximity data was corrected for cell abundance (Figure S3.7A-D). The number of FOXP3+CD4+ cells in proximity to FOXP3-CD4+ cells decreased at post-treatment compared with the paired diagnostic samples (BH corrected $p = 0.023$ for $r = 30 - 250\mu m$ Figure S3.8A). However, the number of FOXP3+CD4+ cells in proximity to CD8+ cells was not different between NDMM samples and paired post-treatment samples (BH corrected $p > 0.05$ for $r = 30 - 250\mu m$ Figure S3.8B). The number of BLIMP1+ cells in proximity to CD8+ and CD4+ cells significantly reduced after treatment (BH corrected $p = 0.02$ and $p = 0.027$ for $r = 30 - 250\mu m$, Figure 4.12A and Figure S3.8C, respectively), indicating a significant change in the immune microenvironment post-treatment. However, the number of FOXP3+CD4+ cells in proximity to FOXP3-CD4+ and CD8+ cells and the number of BLIMP1+ cells in proximity to CD4+ cells was not different between NDMM and MGUS samples (Figure S3.8D-F). Interestingly, despite similar cell density, the number of BLIMP1+ cells in proximity to CD8+ cells in MGUS samples was significantly lower than in NDMM samples (BH corrected $p = 0.036$ for $r = 30 - 250\mu m$ Figure 4.12B, C), which may indicate variability in anti-tumour immune activity in the precursor stage compared with the malignant stage.

4.5.6 Significant spatial clustering of CD8+ cells in NDMM samples compared with post-treatment

We next asked how cells distribute within the BM tissues; do they display a dispersion or spatial clustering pattern? To identify the spatial pattern of a specific cell type, we compared the observed nearest neighbour distance with spatial randomness of the cell type within the tissue section. In most MGUS, NDMM, and post-treatment samples, we observed clustered patterns (Z-score < -1.96) of CD8+, BLIMP1+ and FOXP3-CD4+ cells compared to spatial randomness (Figure 4.12D-H), but

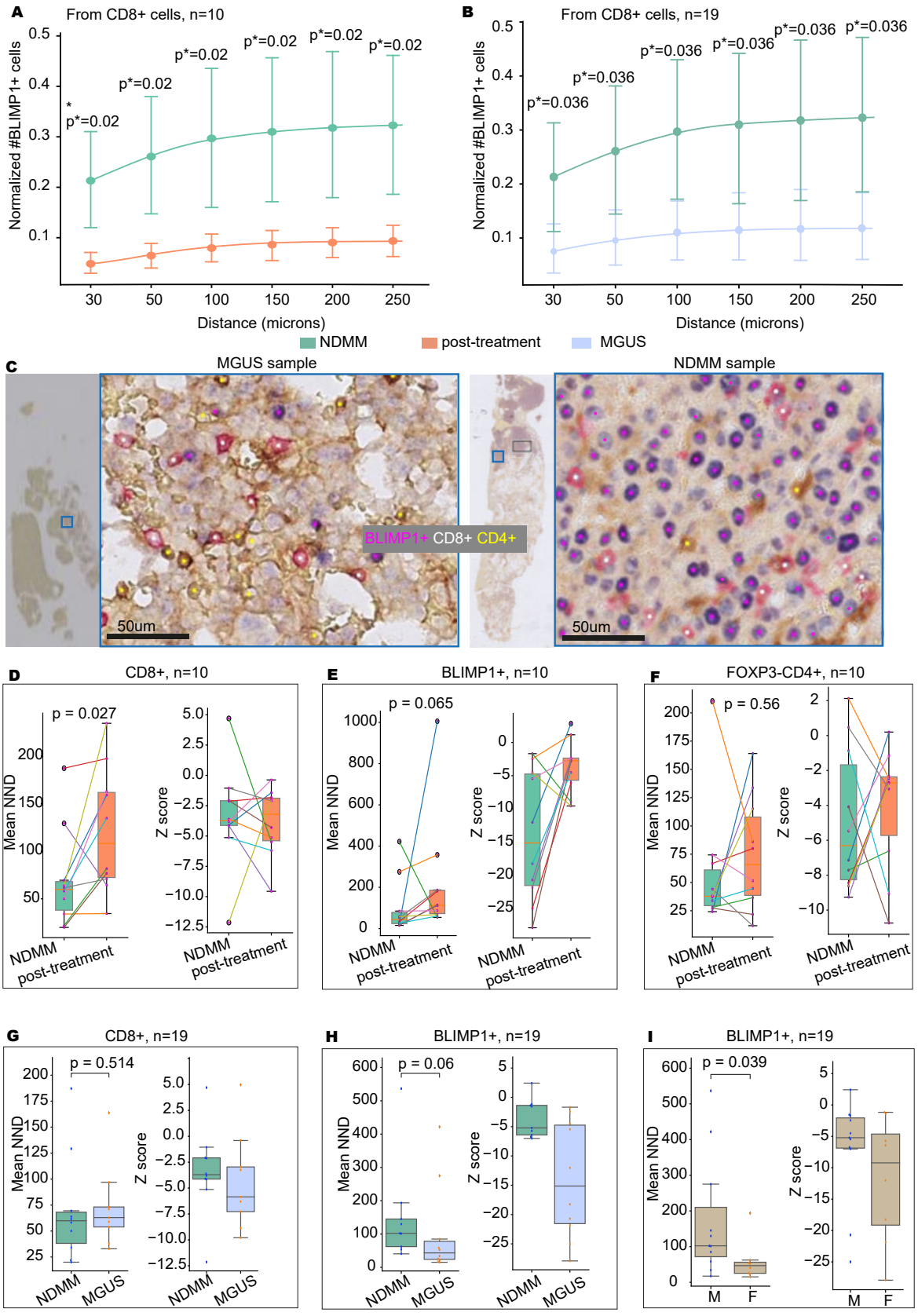


Figure 4.12: **Spatial neighbourhood of immune and tumour cells:**

Figure 4.12: **Spatial neighbourhood of immune and tumour cells: A-B)** Point plots showing the difference in the number of BLIMP1+ cells within a range of distance ($r = 30 - 250\mu m$) from CD8+ cells between NDMM and post-treatment **(A)** and between MGUS and NDMM **(B)**. The p* indicate p values after multiple testing correction using the Benjamini-Hochberg method. The points represent the mean and the bars are 95% confidence intervals indicating uncertainty. **C)** Sample images showing an increased number of BLIMP1+ cells in the neighbourhood with CD8+ on NDMM samples compared with MGUS samples. **D-I)** Clustered or dispersed pattern of immune and tumour cells in BM trephine sample. Boxplots showing the difference in NND and Z score between NDMM and post-treatment for CD8+ cells **(D)**, BLIMP1+ cells **(E)**, FOXP3-CD4+ cells **(F)**. Boxplots showing the difference in NND and Z score between NDMM and MGUS for CD8+ cells **(G)** and BLIMP1+ cells **(H)**, and between male and female for BLIMP1+ cells **(I)**. The Z score shows the significance of the difference between the NND distribution for a given cell type from a complete spatial random distribution and the observed NND.

not for FOXP3+CD4+ cells (Figure S3.9A,B). The degree of clustering of CD8+ cells in the NDMM was significantly higher at diagnosis than in post-treatment samples (p = 0.027, Figure 4.12D) but not compared to MGUS samples (p = 0.514, Figure 4.12G). There was a borderline increase in the clustering of BLIMP1+ cells in the NDMM samples compared with their paired post-treatment, and with MGUS samples (p = 0.065 and p = 0.06, Figure 4.12B,H, respectively). The degree of clustering of BLIMP1+ cells in female samples was significantly higher than in male patients (p = 0.039, Figure 4.12I) but not different between age groups (Figure S3.9D).

4.6 Discussion

Myeloma, like many other blood cancers, initiates and evolves largely in the BM. The BM ecological niche is highly organised, where haemopoietic, including immune cells, osteoblasts, osteoclasts, adipocytes, and other cells interact and co-evolve with neoplastic cells [359, 360]. The BM milieu and its architectural pattern are therefore crucial to the decoding of neoplasm evolution for many blood cancers. Analysis of the intact BM niche has been limited in the past, both due to the difficulty in preserving epitopes and nucleic acid during the processing of BM trephines, and the lack of specialised computational methods that are capable of removing sample artefacts and dissecting BM components.

Here, we demonstrate that through the generation of carefully preserved BM trephine tissue sections and the development of spatial histology methods based on deep learning and spatial statistics, new biological insights on MM neoplastic progression and treatment response can be derived. The spatial architecture of MM BM was interrogated by establishing fully automated computational pipelines to analyse immune cell topography and spatial aggregation, bone density heterogeneity and thickness, in addition to the changes in tumour load and BM components during neoplastic progression and treatment. Previously, spatial topography of stromal components in BM using 3D microscopy in a mouse model [361] and spatial topography of BM adipose tissue and hematopoietic

stem cells in rhesus macaques were studied [362]. To the best of our knowledge, this is the first study to use spatial histology based on deep learning to explore spatial topological patterns in human BM trephine samples that inform changes in disease status in MM. This is in contrast to the many machine learning methods available for BM aspirate-derived cell suspensions for cell counts and bone marrow evaluation [340, 363]. Methods developed in our study may impact the study of many other diseases by unlocking the potential of deep learning and spatial tissue architecture, thus generating new insights from routine BM trephine samples.

BM trephine tissue is a mosaic landscape of blood, bone, cellular tissue, and fat. To dissect the complex mosaic tissue microenvironment into individual components in MIHC images, MoSaicNet was developed. Instead of a standard application of CNNs to generate patch-level [364] or pixel-level classification [365, 272], MoSaicNet can efficiently define the highly irregular tissue component boundary without requiring large amounts of expert annotation training, thus combining the best of two approaches. Patch-based approaches use rigid image patches as units for classification tasks, requiring fewer annotations but cannot generate a detailed mapping of the tissue. In comparison, pixel-based algorithms such as U-Net [365] or Micro-Net [272] generate detailed contour, but such algorithms often require large amounts of training data. MoSaicNet combines a machine learning-based approach, superpixel segmentation, and deep learning classification to efficiently map out the MM BM tissue landscape using superpixels as spatial units, classifying them into cellular components, blood, bone, fat, and background.

Building on MoSaicNet, a new autoencoder-based approach was developed to study bone physiology. This was inspired by the potential role of bone and related cells, such as osteoblasts and osteoclasts, in regulating BM remodelling [335, 91] and MM dormancy and proliferation [366]. Autoencoder is an effective method for dimension reduction and denoising. Here we demonstrated its value in bone density heterogeneity analysis, using feature extraction based on autoencoder and unsupervised clustering of the bone superpixels. We observed that the amount of bone in the biopsies taken post-treatment was greater than those taken at diagnosis, reflecting the destructive effect of MM tumour cells on bone. The bone density of NDMM samples was also more heterogeneous when compared to matched post-treatment samples, and also with samples from MGUS patients, again reflecting an effect of the disease process on bone physiology that occurs in a spatial heterogeneous manner [367]. Moreover, a novel method was developed to study bone thickness using distance transform and topological analysis. In agreement with the bone trabecular surface analysis on lymphoid cancer samples [114], bone% and bone thickness decreased with ageing but were not different between male and female samples. Taken together, our data indicate that bone analytical methods may be useful for the study of bone degeneration during MM progression and treatment, and bone heterogeneity may be a useful marker for disease activity.

Subsequently, AwareNet, developed specifically to identify rare immune cell types, enabled us to dissect the hematopoietic ecosystem of BM, in the context of MM. Deep learning models are often sensitive to class imbalance, resulting in lower accuracy in detecting rare cell types such as FOXP3+CD4+ Treg cells in our samples. To resolve this, cell segmentation-based spatial cell

weighting was proposed [365, 368]. AwareNet extends cell segmentation-based spatial cell weighting [365, 368] by using cell identification instead of segmentation, which is less costly. Furthermore, giving a higher weight score to rare cell types improved the detection of rare cell types compared to U-Net [365] and CONCORDe-Net [334].

Using AwareNet, we observed a reduction in the density of BLIMP1+ tumour cells, and of the immune cell subsets, CD8 and Treg cells in post-treatment BM, compared with diagnostic samples from paired NDMM. While the reduction in tumour cell density is expected, the decrease in immune cell subsets may suggest an alteration in immune function such as anti-tumour responses. Several studies have reported on the changes in frequency or proportion of T cell subsets in post-treatment BM or blood. However, all these studies have hitherto studied BM aspirate samples and assessed immune cell subsets as a percentage of the CD138-negative fraction of mononuclear cells, while our study quantified cell density as a function of tissue surface area. Thus, although we have reported an increase in CD8+ T cells as a fraction of CD3+ cells in post-treatment BM aspirates compared to pre-treatment samples [100], it is not possible to directly compare these data. T-regulatory cells have attracted a great deal of attention in MM, and most studies, including our previous work in BM aspirates, concur in reporting an increased abundance of these cells in MM patients compared with healthy controls [333, 369, 370]. Hence, our observation in this study of a greater density of Treg cells in NDMM samples compared with post-treatment samples is consistent with previous studies [371]. On the other hand, our observation that the density of these cells falls following treatment may be at odds with studies using aspirate samples, for the reasons described above, as well as variation in sampling time and site, but the actual treatments received and type of transplant are also likely to influence the results [331, 327, 328].

Importantly, new insights were derived from the topological analysis of MM plasma cells and immune T cells. In solid tumours such as oestrogen receptor-positive breast [372] and lung tumours [108], spatial scores were found to be more prognostic than cell counts. In MM, however, the spatial relationship of cells and their prognostic value have remained unexplored. Our approaches control for cell abundance and take into account the local tissue architecture and cell distribution. Interestingly, the number of BLIMP1+ cells in spatial proximity with CD8+ cells was significantly greater in diagnostic MM samples compared with MGUS and post-treatment samples. Given reports of tumour-reactive CD8+ T cell populations in MM patients [373], the proximity of CD8+ T cells to tumour cells may represent increased immune activity in MM, and the "homing" of CD8+ T cells to tumour sites. This is consistent with the clustered pattern of CD8+, CD4+ and BLIMP1+ cells in most cases. We observed a dispersed pattern of FOXP3+CD4+ Treg cells. The expansion of Treg cells has been found to contribute to the growth, proliferation, and survival of myeloma plasma cells [331]. Thus, the dispersed pattern of Treg cells may be a phenotype of expansion, which may promote the invasion and differentiation of MM plasma cells.

Limitations of this study include the limited number of samples. More samples are needed to capture the full cellular and non-cellular region heterogeneity, and the results should be interpreted with this consideration. Secondly, the samples used in this study were collected from one site, which

could introduce tissue processing and staining bias. Thirdly, inter-and intra-observer variation was not analysed for cell detection and classification tasks.

4.7 Conclusion

We demonstrated how spatial and machine learning methods can be used to dissect the mosaic tissue microenvironment of BM trephine samples (MoSaicNet) and accurately identify immune T and MM plasma cells (AwareNet). Despite the limited sample size, bone trabeculae morphologic and cell topological spatial analyses enabled the deep mine of both cellular and non-cellular parts of the BM niche. Possible future works include:

- Adopting MoSaicNet and AwareNet to routinely available H&E stain of BM trephine samples to further explore bone remodelling;
- Exploring the association of bone morphologic features and cells spatial organisation features with patients' clinical outcomes such as treatment response, and survival.
- Integrating morphologic and spatial features with molecular features to identify genetic aberrations associated with morphologic/spatial phenotypes in the BM niche.
- Identifying morphologic and spatial features of progressor and non-progression MM precursors [374] to help refine risk models. Insights generated from this study warrant further validation and investigation in larger cohorts, which is in progress.

Chapter 5

Evaluation of morphological features and spatial immune infiltration patterns as a biomarker for recurrence in ductal carcinoma *in situ*

5.1 Overview

As we discussed in Section 1.4, DCIS is a pre-invasive lesion of tumour cells within the duct [116]. One of the distinguishing histologic features of DCIS from invasive breast cancer is that the tumour cells are isolated from the breast stromal region by basement membrane and myoepithelium layer, whereas in invasive breast cancer, these separating membranes are broken, and the tumour cells have direct contact with stomal cells [116]. The prognostic evaluation of morphological features of the DCIS duct, which contains the tumour cells, and the spatial organisation of stromal TILs in the surrounding regions of DCIS are gaining interest [151, 116]. Currently, manual assessment of TILs by expert pathologists is considered a gold standard and the TIL-WG on breast cancer has developed a set of guidelines for manual stromal TILs scoring for DCIS on H&E images [151]. However, the manual assessment has inherent limitations associated with the shortage of expert pathologists, inter-observer variability, and a comprehensive assessment of WSI H&E images is also time-consuming. Thus, the TIL-WG has envisioned that machine learning methods could alleviate the limitation of visual assessment, and when possible, the methods should follow the guidelines [152]. However, a reliable automated scoring method following the guidelines is yet to be developed due to the inherent complexity of DCIS duct morphology and the assessment strategy.

Here, we aimed to simultaneously interrogate the association of DCIS duct morphological features and the spatial organisation of stromal TILs (AI-TIL) in the surrounding region of DCIS duct with recurrence. First, we developed an automated image analysis pipeline comprising of single cell level

and tissue level algorithms to automate the stromal TILs scoring guideline by the TIL-WG [151]. The pipeline consists of tissue segmentation, cell detection, cell classification, DCIS duct segmentation, and scoring the spatial TILs pattern at a different distance around the DCIS duct. Moreover, we analysed the association between DCIS morphological features, spatial stromal TILs and mutation burden. We implemented a generative adversarial network (GAN) based segmentation method to identify and segment DCIS duct automatically. To identify lymphocytes, we used a pre-trained deep learning model. We collected single cell and tissue level annotations and slide level stromal TILs scores from expert breast pathologists to train and validate our algorithms.

Our DCIS segmentation model achieved a dice overlap of 0.94 (± 0.01), and the cell classifier model achieved 92% accuracy compared to pathologists' annotations. We observe a higher correlation between pathologists' scores and AI-TIL computed using broad stomal boundaries (Spearman $r = 0.67$, $p = 3.83 \times 10^{-7}$, $W = 0.7mm$) compared with circumferential AI-TIL (cAI-TIL) computed using $W = 0.03mm$ stromal boundary (Spearman $r = 0.28$, $p = 0.12$). Using multivariate COX proportional hazard regression analysis, a low cAI-TIL was associated with an increased risk of recurrence independent of standard biomarkers and clinical variables (HR $= 0.56$, 95% CI: $0.39 - 0.80$, $p = 0.002$). cAI-TIL was not correlated with DCIS duct morphologic features or mutation burden. In Invasive breast carcinoma recurrence (IBCr) patients, DCIS duct area was positively correlated with mutation burden (Spearman $r = 0.78$, $p = 0.0021$), but not in the No ipsilateral breast event (no-IBE) and ductal carcinoma *in situ* recurrence (DCISr) patients.

Taken all together, the spatial organisation of stromal TILs computed using an automated method has prognostic relevance in DCIS, and this should be validated on larger cohorts of patients for use in a clinical setting.

5.2 Introduction

As we discussed in Section 1.4, DCIS is a pre-invasive lesion of abnormal cells within the breast duct that are separated from the stromal region by myoepithelium and basement membrane [375, 376, 116]. Before the development of breast cancer screening, DCIS was rarely recognised; however, it now accounts for approximately 20% of all breast cancer [128, 120]. Usually, treatment is recommended since there is about a 50% chance of transforming the DCIS into invasive cancer [120].

Over the past decade, sequencing-based methods have been employed to find molecular processes leading to the transformation of DCIS into invasive breast cancer [141, 377, 378]. Molecular alterations such as mutation of well-known genes including TP53, PIK3CA and AKT1, and copy number variation were found to correlate with the progression of DCIS to invasive cancer [141, 378]. However, these studies lack consistency. In the current clinical management of DCIS, all patients are recommended to undergo treatment similar to invasive breast cancer, which could lead to overtreatment [116]. Thus, the unmet clinical question is which DCIS patients are likely to progress into aggressive breast cancer over time. Despite all the knowledge about DCIS tumour cells' molecular

features, histopathology remains the only reliable approach to detect DCIS since the tumour cells are restricted within the duct [116]. Thus, morphological analysis of the DCIS duct and stromal TILs spatial organisation relative to DCIS duct are gaining interest.

A recent work by Risom et al. [116] showed that invasive recurrence in DCIS is accompanied by changes in DCIS morphology and the structure and composition of the TME. Furthermore, the spatial distribution of stromal TILs in the vicinity of DCIS duct has been found prognostic in DCIS [120, 146–150].

Thus, to improve the reproducibility and standardise reporting of stromal TILs, the TIL-WG on breast cancer has developed a guideline for manual stromal TILs scoring for DCIS on H&E images [151]. However, even studies following the guidelines have shown inconsistency. The inconsistency could be partly attributed to the inherent limitations of manual assessment such as intra- and inter-observer variability [172, 170, 171]. Moreover, the shortage of experienced pathologists compared to the demand limits the scalability of the manual scoring in research and clinical environment [173, 174]. Thus, the TIL-WG suggested that machine learning based assessment of stromal TILs might overcome these limitations and advised the algorithm should comply with the guideline where possible for ease of interpretation [152, 174]. To the best of our knowledge, there is no published computational method that adheres to the essential concepts in the DCIS stromal TILs scoring guidelines set by the TIL-WG.

Thus, here, we proposed a fully automated multi-stage image analysis workflow consisting of single-cell level and tissue level algorithms to quantify spatial patterns of stromal TILs following the essential steps set by the TIL-WG [151], and morphological features of DCIS duct from diagnostic H&E WSI. Using our workflow and statistical analysis method, we evaluated the following:

- The prognostic relevance of DCIS duct morphological features and spatial stromal TILs patterns as biomarkers of recurrence using diagnostic H&E images of DCIS patient;
- The concordance between the automated and pathologists' stromal TILs score, estimated following the TIL-WG guidelines [151];
- The association between DCIS duct morphology, automated spatial stromal TILs score, andDCIS mutation burden.

5.3 Materials

5.3.1 Patients studied

The samples were collected from a retrospective multi-centre study activated at 12 participating TBCRC (Translational Breast Cancer Consortium) sites, which identified women treated for DCIS at one of the enrolling institutions between 01/01/1998 and 02/29/2016. The TBCRC and the Department of Defense (DOD) approved this study for the collection of archival tissues. Duke University School

of Medicine, Durham, USA, served as the initiating and central site for all data, samples, assays, and analysis. The study was approved by the Duke Health Institutional Review Board (Protocol ID: Pro00068646) as well as the institutional review boards of each participating institution. In addition, individual sites reviewed medical records to identify patients eligible for the study.

Study inclusion eligibility criteria included: Women aged 38-76 years at diagnosis of DCIS without invasion; no prior treatment for breast cancer; and definitive surgical excision with no ink on tumour margins and treated with mastectomy, lumpectomy with radiation, or lumpectomy. All cases consisted of an initial diagnosis of pure DCIS, with ipsilateral recurrence occurring no less than 12 months from the date of the primary diagnosis. Clinical data, including treatment data, were collected at each site, and standardised data points were entered into a web-based portal. Tumour tissue was collected from formalin-fixed paraffin-embedded blocks and cut into $5\mu m$ sections. All slides were scanned and reviewed centrally by a breast pathologist (Allison H. Hall, from the Department of Pathology, Duke University School of Medicine, Durham, North Carolina, USA) to confirm the diagnosis.

In total, a subset of 165 patients from the TBCRC study with primary DCIS was included in this study. The clinical characteristics of the patients are summarised in Table 5.1. The cohort consisted of three groups of patients:

1. no-IBE: patients with primary DCIS with no recurrence after ≥ 5 years follow-up;
2. DCISr: patients primary DCIS with DCIS recurrence, recurrence diagnosed ≥ 12 months after initial diagnosis;
3. IBCr: patients with primary DCIS cases with invasive and or metastatic recurrence, ≥ 12 months after initial diagnosis.

After a median follow-up of 69 months (range: 12-228 months) after the initial diagnosis, $115(70\%)$ patients were diagnosed with a recurrence. These consisted of $64(39\%)$ patients with DCISr, and $51(31\%)$ patients with IBCr. The remaining $50(30\%)$ patients had no recurrence after a median observation of 105 months (range: 60-228 months). The ER status, PR status, DCIS grade, presence of necrosis, treatment administered and type of surgical procedure is shown in Table 5.1.

5.3.2 Sectioning and data acquisition

When a suspected DCIS case was detected, diagnostic H&E slides from the lumpectomy/mastectomy resection containing the DCIS component were examined by the pathologists. Additionally, H&E slides for recurrent surgery were sent to the pathologists for assessment and slides from a node as a control sample. The healthy/control (normal tissue) was taken from a normal node or breast tissue with no cancer. The control block was confirmed to be devoid of a tumour by pathologists. Two blocks of tumours (Block A and Block B) were extracted from the tumour resection, each containing $\geq 2mm$

Table 5.1: **The characteristics of the TBCRC cohort data included in this study.** no-IBE = no ipsilateral breast event; DCISr = DCIS recurrence; IBCr = invasive breast carcinoma recurrence; ER = estrogen receptors, PR = progesterone receptors.

Parameter	**No. patients (%)**
Age at diagnosis	
Median (range)	52 (38, 76)
H&E image available	
Yes	127 (77)
Mutation data available	
Yes	81 (49)
Relapse group	
no-IBE	50 (30)
DCISr	64 (39)
IBCr	51 (31)
ER status	
Positive	98 (59)
Negative	35 (21)
Unknown	32 (20)
PR status	
Positive	83 (50)
Negative	46 (28)
Unknown	36 (22)
Grade	
I	12 (7)
II	62 (38)
III	91 (55)
Radiation therapy	
Yes	90 (54)
No	69 (42)
Unknown	6 (4)
Hormonal therapy	
Yes	54 (33)
No	92 (56)
Unknown	9 (11)
Necrosis	
Yes	127 (77)
No	35 (21)
Unknown	3 (2)
Surgical procedure	
Mastectomy	46 (28)
Lumpectomy	119 (72)

DCIS area. The aim of having the two tumour blocks was to analyse the intra-tumour heterogeneity of the DCIS. Then, these tumour blocks were sectioned into 25-30 slides of five microns thick tissue.

The first and last section was used for H&E staining. The slides were scanned at $20\times$ magnification with $0.5\mu m$ resolution using Aperio AT2 brightfield imaging. Our dataset contains 243 diagnostic H&E WSI from 127 DCIS patients, with some patients having only H&E images from one block. Some of the middle sections were used for DNA whole exome sequencing of the DCIS. Moreover, the normal tissue was sectioned into 10 slides, each five microns thick. Again the first section was used for H&E staining and the rest for control DNA sequencing. The DNA extraction and sequencing were conducted at Prof. Maley's lab (Arizona Cancer Evolution Center, Arizona State University) using formalin-fixed paraffin-embedded GeneRead DNA Kit as described in their recent work on a similar study [379]. As shown in Table 5.1, 81 cases(49%) had DNA sequencing data.

5.4 Methods

We developed an automated image analysis pipeline to quantify the spatial distribution of stromal TILs in the vicinity of DCIS duct following the guidelines by TIL-WG and quantify DCIS morphological features (Figure 5.1). The pipeline combines machine learning and classical image processing/analysis algorithms that operate at the single cell and tissue levels. As shown in Figure 5.1, the automated image analysis pipeline consists of the following algorithms:

1. Image preprocessing to divide the WSI into smaller images or patches, also known as "tiling";
2. Tissue segmentation to delineate the area of WSI which contains tissue sample;
3. Cell detection and classification to identify the position of lymphocytes in the WSI;
4. Immune hotspot removal (as suggested by the working group and expert pathologists);
5. DCIS duct segmentation to extract DCIS duct boundary;
6. Finally, image analysis method to generate slide level of the spatial stromal TILs score and quantify DCIS duct morphological features.

We used the TBCRC dataset described in Section 5.3. We collected single cell and tissue level annotations and slide level stromal TILs scores from expert breast pathologists to train and validate the algorithms.

5.4.1 A brief overview of the TIL-WG guideline for stromal TILs scoring in DCIS

According to the TIL-WG guidelines [151], pathologists look at each DCIS duct within a WSI and estimate stromal TILs score at duct level. Then, the slide level score is generated by averaging the duct

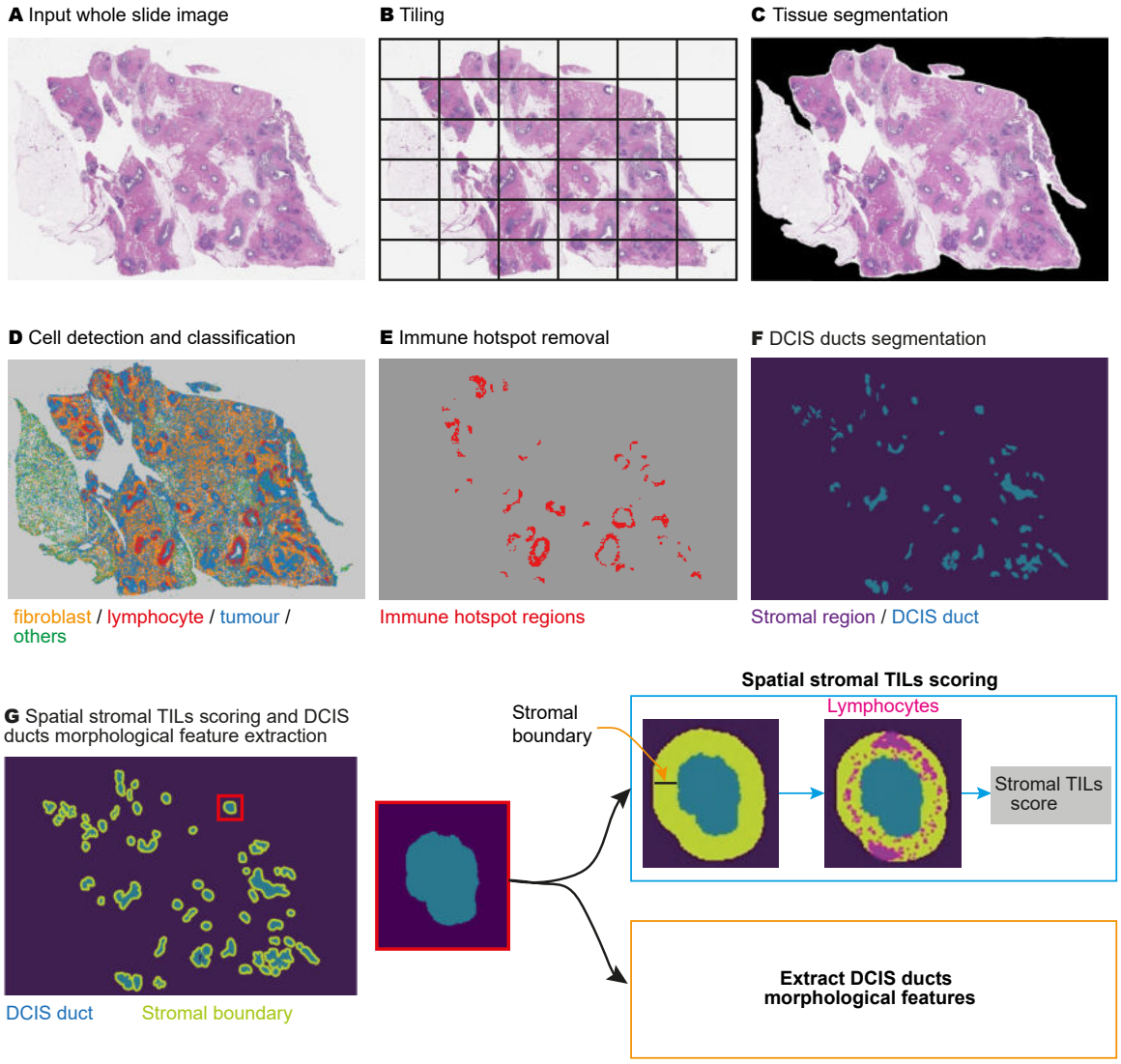


Figure 5.1: **Overview of the fully automated image analysis pipeline for spatial stromal TILs scoring and DCIS morphological feature extraction:** **A)** The input image is a H&E stained WSI of diagnostic slide of DCIS patient. **B)** Dividing a gigapixel WSI into smaller images or patches, which can be loaded into memory easily. **C)** Tissue segmentation to remove slide area without tissue. **D)** Cell detection and classification to spatially map lymphocytes and other cell types within the WSI. **E)** Immune hotspot region identification. **F)** DCIS duct segmentation to localise and segment individual DCIS duct. **G)** Image analysis to compute spatial stromal TILs for a given stromal boundary width and estimate DCIS duct morphological features. This step utilises the outputs of cell detection and classification (D) and DCIS duct segmentation (F) algorithms. The displayed boundary is $0.2mm$. The features were computed at the duct level, and mean aggregation was applied to generate a score for WSI.

level scores. In DCIS, stromal TILs are TILs situated in a "specialised breast stromal region" around each DCIS duct [151]. Though there is no currently agreed size for this region, the TIL-WG suggested using a boundary that could extend up to $0.7mm$ until more data is found that supports a specific value. The duct level TILs score is estimated as a percentage of the surrounding stromal area of the duct covered by lymphocytes, excluding immune hotspots. Immune hotspots are specific regions within a WSI with spatially clustered lymphocytes compared to the rest of tissue [241]. Moreover, TILs within the DCIS duct and necrosis region should be excluded. A detailed explanation of the DCIS stromal TILs scoring guideline could be found in Dieci et al. [151]. Sample images showing stromal TILs scores by pathologists are presented in Figure S4.1.

5.4.2 Pathologists' input for models training and validation

DCIS duct segmentation annotation

To train and validate our DCIS segmentation algorithm, we collected DCIS duct segmentation annotation for 43 H&E WSI. The images were annotated by two expert pathologists in breast cancer: Hugo M. Horlings from the Netherlands Cancer Institute and Robert West from Stanford University Medical Center. The annotations were collected using a web-based WSI viewer and annotation tool developed in our lab (unpublished). From these annotations, we extracted $2,550$ "tiles" of $1,024 \times 1,024 \times 3$ pixels size. Sample H&E images with DCIS duct segmentation are shown in Figure 5.2.

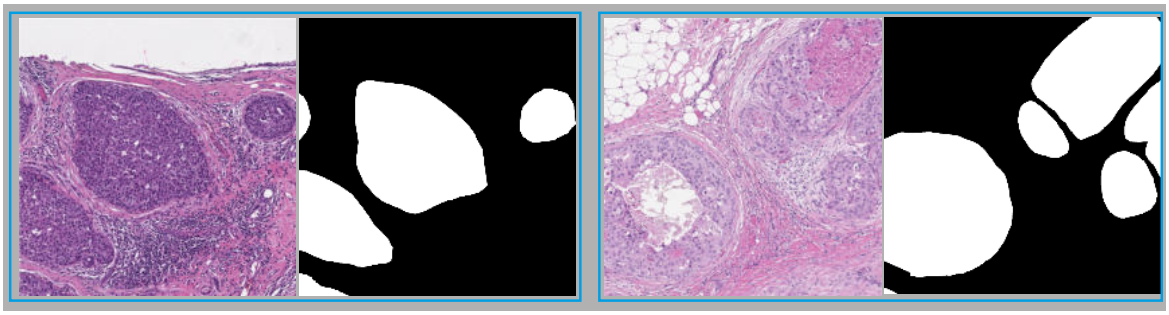


Figure 5.2: **Sample pair of images showing DCIS H&E images and DCIS duct segmentation mask generated from pathologist annotation.** The DCIS mask image shows the DCIS duct (white) and stromal regions (black).

Single cell annotations

For cell identification on H&E, we employed an existing pipeline developed in our lab [108]. To validate the performance of the pipeline in our DCIS dataset, we collected $10,486$ single cell dot annotations by an expert pathologist, Allison Hall, from Duke University School of Medicine. These annotations fall into three classes: epithelial cells $(n = 6,821)$, fibroblasts $(n = 1,075)$, and lymphocytes $(n = 2,590)$.

Pathologist DCIS stromal TILs scores

To validate the automated stromal TILs score with pathologists' scores, we asked two pathologists (Hugo M. Horlings from The Netherlands Cancer Institute and Roberto Salgado from GZA-ZNA Hospitals, Antwerp, Belgium) to independently estimated DCIS stromal TILs for 46 samples. The scores were estimated following the guideline by the TIL-WG [151]. The scores are continuous values ranging between 0% and 100%.

5.4.3 Tiling

Tiling is the process of dividing the high-resolution WSI into smaller images or patches (Figure 5.1A, B). The H&E WSI were scanned at $20\times$ magnification with a pixel resolution of $0.5\mu m$ per pixel. A representative image has a $53,784 \times 37,284$ pixel size at $20\times$ magnification. Since the WSI are too big to load into memory, they were divided into $2,000 \times 2,000$ pixel size images, which could be loaded into memory for further processing.

5.4.4 Tissue segmentation

The H&E images contain both tissue or biopsy area and glass area (background pixels). Thus, accurately removing the background region has advantages such as improving computational time and avoiding false positive detections due to noisy pixels. For a multi-stage pipeline, removing the background region at an early stage improves the speed of the workflow since the subsequent stages will analyse only tissue pixels. Thus, after tiling, tissue segmentation was applied.

Tissue segmentation is a binary pixel classification, tissue and non-tissue (Figure 5.1C). This could be achieved using traditional image processing algorithms or deep learning based approaches. Some studies employed traditional image processing algorithms such as Otsu's adaptive thresholding [380] of the grayscale version of H&E image [179] and watershed segmentation [381] to segment tissue from WSI to segment tissue region from H&E images. However, these methods might not robustly learn weak features and often need preprocessing steps that induce bias in the model. This process reduces the generalisability of the model to new data [272].

Here, we applied a pre-trained Micro-Net [272] to segment tissue on the H&E images. Micro-Net uses multi-resolution images to capture multi-scale features, which enables the network to extract varying scale features to differentiate tissue regions from artefact and background regions [108, 272]. This enables the model to capture essential features to identify tissue regions. Since an approximate tissue boundary is enough, the segmentation was applied at $1.25\times$ magnification. After tissue segmentation, subsequent analyses were applied to tissue regions only.

5.4.5 Cell detection and classification on whole slide H&E images

One of the essential steps in our automated spatial stromal TILs scoring is identifying lymphocytes' location within the WSI. This was achieved using cell detection followed by cell classification. Cell detection aims to find the spatial position of cells within the segmented tissue region, while cell classification aims to identify the type of cells.

To detect and classify single cells, we used a well-established pipeline which was trained and validated on 26,960 single cell annotations from 53 WSI [108] (Figure 5.1D). To handle stain variability, the pipeline uses Reinhard colour normalisation [382]. The Reinhard colour normalisation transforms the source image colour space into the target image colour space [383]. The target image was selected from the images used during model development. The pipeline uses spatially constrained CNN for cell detection and classification [108, 176]. For every pixel in the tissue region, the algorithm computes the probability of the pixel belonging to the cell nucleus. Then, post-processing methods such as thresholding, local peak intensities and pixel grouping were applied to identify the centre of the nucleus [176]. The classifier network uses neighbouring ensemble prediction combined with the standard softmax function for the classification [176]. The classifier categorises the detected cells into four classes: epithelial cells, fibroblasts, lymphocytes, and others (Figure 5.1D). The details of training parameters and hyper-parameters of cell detection and classification pipeline can be found in [176, 108].

5.4.6 Immune hotspot identification

According to the stromal TILs scoring guidelines [151], immune hotspots should be excluded during the DCIS stromal TILs reporting. Immune hotspots are specific regions within a tissue section with spatially clustered lymphocytes compared to the rest of tissue [241]. To identify immune hotspots, we applied Getis–Ord spatial statistics proposed by Nawaz et al. [241] (Figure 5.1E). Once the immune hotspot regions were identified, lymphocytes residing in these regions were excluded from the automated stromal TILs computation.

5.4.7 Automated DCIS duct segmentation

The DCIS duct segmentation is an integral part of the spatial stromal TILs scoring, and DCIS duct morphological phenotype quantification (Figure 5.1F). For H&E images scanned at 20× magnification, the size of DCIS duct could range from thousands to millions of pixels. Thus, a larger receptive field (patch) is required to fully capture the histological feature of a DCIS duct and to train a deep learning algorithm to segment DCIS. Image segmentation is an image-to-image translation task. Recently, adversarial learning has gained a surge of interest in image-to-image translation involving semantic labelling [384, 385]. However, training GAN on high-resolution images could be unstable [384, 386, 387]. Wang et al. [384] address these issues using a robust adversarial learning objective together with a new multi-scale generator and discriminator architectures. For high-resolution images,

conditional GAN proposed in [384] was found to outperform U-Net for segmentation task [65]. A recent work from our lab by Sobhani et al. [388] showed that the conditional GAN proposed by Wang et al. [384] performs well for DCIS duct segmentation on MIHC images. For DCIS duct segmentation on H&E, we adopted the DCIS segmentation model on MIHC images [388] to H&E images. For DCIS duct segmentation on H&E, I collaborated with Dr. Sobhani (first author of [388]). The model was trained from scratch using pathologist DCIS segmentation annotation on H&E images described in Section 5.4.2.

Image segmentation aims to transform an image from H&E image domain to a binary mask image. The model comprises a generator and a discriminator. The generator's objective is to generate a DCIS binary mask from H&E stained histology image. At the same time, the discriminator learns to distinguish the ground truth (human annotation) image from the generated mask. It is trained in a supervised way, and the training dataset is prepared as sets of pairs $(x,y) = \{(x_i,y_i) : i \in \{1,2, ...,n\}\}$, where x_i and y_i are $1,024 \times 1,024 \times 3$ histology image and $1,024 \times 1,024 \times 1$ DCIS ground truth mask, respectively (Figure 5.2). The objective of the network is to model the distribution of the DCIS mask using Equation (5.1), given the input H&E image.

$$\min_{G}\Big(\max_{D1,D2,D3}\sum_{k=1}^{3}\big\{E_{(x,y)}\left[log\ D_k(x,y)\right]+E_x\left[log\left(1\ -\ D_k\ (x,G(x))\right)\right]\big\}\Big) \tag{5.1}$$

where G denotes the generator, which contains a global generator to generate an initial DCIS map and a local generator to enhance the quality of the output of the first generator [384]. The D denotes the discriminator part that employs multiple discriminators with identical network structures that operate at different scales of an image pyramid [384]. A three-level ($k = 3$) image pyramid was used as proposed in the original paper by Wang et al. [384]. These multi-scare networks regularise the generator to learn the coarse and fine features of the image [384].

Model parameters were randomly initialized using uniform Glorot [275] and optimized using Adam [276] with a learning rate of (0.0002) and beta1 (0.5). The model was trained for 100 epochs.

The segmentation model was trained and validated using pathologist annotations described in Section 5.4.2. The annotation consists of $2,550$ tiles from 43 WSI/patients. To investigate the generalizability of our segmentation model, we performed a 5-fold cross-validation. Around $1,900$ tiles from 36 slides were used for training and the remaining for testing in each round. To remove false-positive prediction from artefact regions and normal duct, we excluded predicted DCIS duct with an area smaller than $1,2000\mu m^2$.

5.4.8 Estimating pathologist's stromal boundary width

As we stated above, the TIL-WG suggested using a stromal boundary width that could extend up to $0.7mm$ from the DCIS border [151]. We asked a pathologist to annotate DCIS duct and stromal boundary for some slides that he will use for stromal TILs scoring as shown in Figure 5.3. We extracted the annotation and quantified the width as described below. The pathologist manually annotated

individual DCIS duct border and stromal boundary. Multiple DCIS ducts were annotated within a slide. The annotations are saved as polygons with x and y. Let $A_b \in \mathbb{R}^{n\times 2}$ and $A_s \in \mathbb{R}^{m\times 2}$ represent a matrix notation of coordinates of the DCIS duct border annotation and the stromal boundary annotation of the i^{th} annotated DCIS duct. n and m represent the length of the annotations in pixels. Every element of A_b and A_s have (x,y) attributes, which allows distance computation.

Suppose $D \in \mathbb{R}^{n\times m}$ denotes distance matrix and $d_{i,j}$ denotes the value of the element $D[i,j]$. The $d_{i,j}$ is a Euclidean distance between the i^{th} element of A_b and the j^{th} element of A_s. Then, the average thickness or width of DCIS duct level boundary (W) for one DCIS duct was computed using Equation 5.2.

$$W = \frac{1}{n}\sum_{i=1}^{n}\left(\min_i d_{ij}\right),\ j = 1,2,3,...,m \tag{5.2}$$

The *min* operation is applied to every row of the matrix, D, and it finds the minimum distance for every point in A_b to A_s, generating a $n \times 1$ distance array. As can be seen in Figure 5.3, the boundary width is not constant across the DCIS border. Thus, the boundary width from the pathologist annotation (W) for i^{th} DCIS duct was computed as the mean of the width across its border. This enabled us to evaluate the variability of stromal boundaries utilised by pathologists within a slide and across slides.

5.4.9 Image analysis algorithm to estimate spatial stromal TILs score

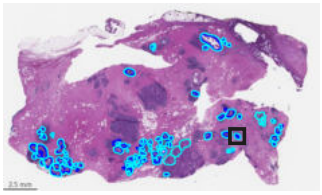
Following the identification of lymphocytes and DCIS duct segmentation described above, we developed an image analysis algorithm (Figure 5.1G) to generate a stromal TILs score (AI-TIL) considering only TILs in the vicinity of DCIS duct as proposed by the TIL-WG [151].

To compute the AI-TIL score for a given H&E WSI, we first computed stromal TILs for every DCIS duct. Suppose a given WSI has n DCIS ducts, $\{D_i : i \in \{1,2,\ ...,n\}\}$.

We first defined the width of the stroma boundary in millimetres (W). Then, the stromal TILs score for the i^{th} DCIS duct was computed as a ratio of the number of lymphocytes in the stromal boundary and the defined stromal area. While lymphocytes in the neighbourhood of DCIS duct (within a distance (W)) could be identified by simply calculating a Euclidean distance between the duct boundary and the lymphocytes, computing the stromal boundary area needs demarcating the stromal tissue area. To delineate the stromal boundary, we applied a morphological dilation operation over the DCIS duct binary mask image. The morphological operation is in image space using pixel units. For ease of interpretation, the boundary width is in standard units in the tissue space, such as *mm* [146]. For the desired boundary width W in millimetres, the width in pixels was computed using Equation (5.3).

$$W_{pixels} = \frac{W}{mmp} \times 10^3 \tag{5.3}$$

, where mpp stands for microns per pixel image resolution from the metadata of WSI images (e.g. 0.5 in our images). To demarcate the stromal boundary in Figure 5.1C, we applied morphological



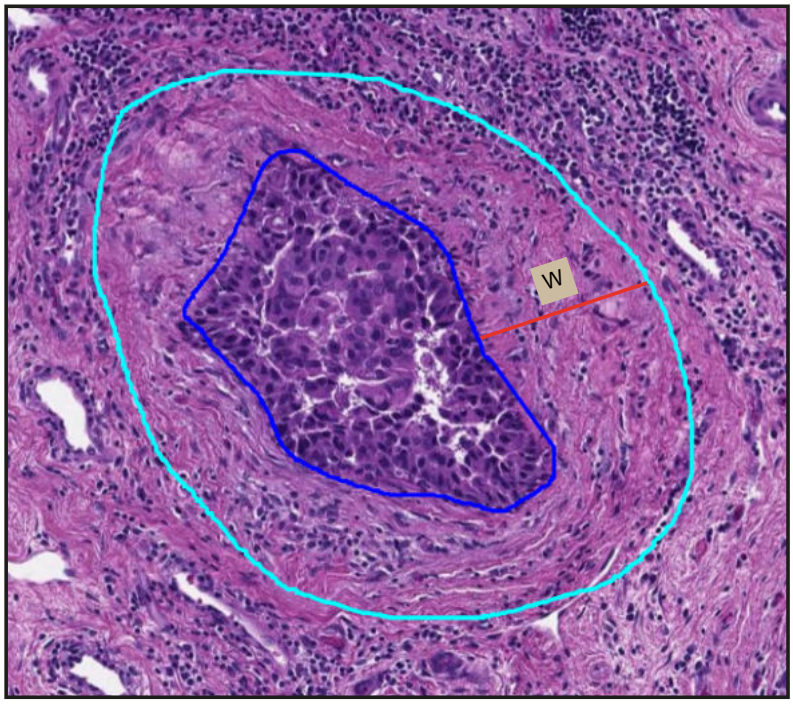


Figure 5.3: **Sample image showing DCIS expert segmentation annotation and stromal TILs scoring boundary.**

dilation operation as shown in Equation (5.4).

$$I_{SB} = \left(I_{D_i} \oplus W_{pixels}\right) - I_{D_i} \tag{5.4}$$

, where I_{D_i}, $\oplus$, and I_SB are cropped binary images of duct D_i (Figure 5.1C), dilation operation, and stromal boundary image, respectively. While cropping, a margin of $W_{pixels} + 10pixels$ was included beyond the bounding box of D_i for the dilation operation. Cropping the duct image before the dilation operation increases the algorithm's speed. A dilated region that overlaps with a neighbouring DCIS duct tissue was excluded while computing the stomal area. Then, stromal TILs score for duct D_i, $score_i$, was computed using Equation (5.5).

$$score_i = \frac{\#L}{stromal\ boundary\ area} \tag{5.5}$$

, where $\#L$ is the number of lymphocytes in the stromal boundary and the stomal boundary area is measured in μm^2. This is applied to all n ducts in each slide, resulting in a vector of $(1 \times n)$ duct level TILs scores. TILs between adjacent ducts were counted in each duct. Finally, the slide level stromal TILs score was computed as the mean of individual DCIS duct AI-TIL score using Equation (5.6). We also analysed the heterogeneity of duct level AI-TIL measured as variance of AI-TIL within the slide.

$$AI\text{-}TIL = \frac{\sum_{i=1}^{n} score_i}{n} \tag{5.6}$$

The AI-TIL indicates the density of TILs in the stromal boundary and its unit is cells per unit μm^2 tissue area. The statistical analyses were performed at the patient level. An average value of AI-TIL was computed when there are multiple H&E images from one patient. Moreover, we analysed the distance between the AI-TIL score of the slides from one patient.

While computing AI-TIL, the user defines the value of W. This allows assessing the prognostic value of TILs depending on their proximity to the DCIS duct border. Here, we computed AI-TIL for a set of stromal boundary widths, W = $\{0.03, 0.05, 0.067, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}mm$. Previous studies utilising the TIL-WG guidelines for stromal TILs manual scoring in DCIS analysed the prognostic relevance of circumferential TILs [167, 389]. Circumferential TILs were defined as dense infiltrate of about three layers of lymphocytes around DCIS [167, 389]. The diameter of lymphocytes ranges from $7 - 15\mu m$ depending on their type [390]. Therefore, a $0.03mm$ $(30\mu m)$ stromal boundary approximates the circumferential region of DCIS duct. Thus, in our automated stromal TILs scoring, AI-TIL computed using $W = 0.03mm$ boundary was called circumferential AI-TIL (**cAI-TIL**).

5.4.10 DCIS duct morphologic features

DCIS has diverse nuclear atypia and architectural patterns [139, 391]. These morphological features have implications for recurrence risk, post-surgical recurrence time, and progression into invasive cancer [391]. Here, after segmenting the DCIS duct, we analysed the following histomorphologic features:

- **Total DCIS tissue area.** It is the total tissue area of DCIS normalised by the total amount of tissue within in the WSI.
- **Number of DCIS ducts.** It is count of DCIS ducts within in a WSI normalized by the amount tissue within in the WSI.
- **Mean DCIS duct area.** It is the average DCIS duct area measured in μm^2.
- **Mean DCIS solidity.** The solidity of a DCIS duct measures the extent to which the DCIS duct covers its convex hull [300, 301]. A graphical illustration of the solidity measure can be found in Figure 3.3A. Its value ranges between 0 and 1. It is a measure of shape irregularity [301]. A value of 1 indicates a solid DCIS duct, while a value less than 1 indicates a DCIS duct with irregularities. An average value was reported for a slide level DCIS duct solidity score.
- **Mean DCIS duct eccentricity.** It is a measure of elongation of DCIS duct. Its value ranges between 0 and 1 as shown in Figure 3.3B. A value close to 0 indicates circular DCIS duct, and for elongated DCIS duct, their eccentricity value increases. An average value was reported for a slide level DCIS duct eccentricity score.

5.4.11 Mutation data preprocessing

To investigate the intra-tumour heterogeneity of mutation within a tumour, two blocks (block A and block B) were created from the tumour resection. Tissue sections were extracted from these blocks, and DNA whole exome was extracted and sequenced [379]. Single nucleotide variants were called, and intra-tumour heterogeneity was estimated using the intra-tumour heterogeneity estimation pipeline under the set of parameter values optimised using 28 technical replicates on similar DCIS samples [379]. In short, this method uses Platypus to generate candidate single nucleotide variants, asymmetrically filters them to determine if they are shared or private, and then refines them considering several quality metrics, including their coverage in normal and tumour samples, patient-specific germline variants (using normal tissue DNA sequencing) and known germline variants in human populations. The asymmetric filter applies the notion that the presence of a high-confidence variant in a sample increases the confidence of that variant in the other sample [379]. Thus, the mutation data contains only somatic mutations.

Then, mutation burden, common mutations and mutation divergence % were recorded. Here, a mutation denotes a single nucleotide variant. These features were computed as follows. Let A_m and

B_m represent a set of mutations in blocks A and B, respectively. The computation of patient-level mutation burden, common mutation, and mutation divergence % are shown in Equations (5.7-5.9), respectively.

$$common\ mutations = A_m \cap B_m \tag{5.7}$$

$$Mutation\ burden = A_m \cup B_m \tag{5.8}$$

$$Divergence\% = \frac{A_m \setminus B_m \cup B_m \setminus A_m}{Mutation\ burden} \tag{5.9}$$

Where $\cap$, $\cup$ and $\setminus$ are the intersection, union and relative difference set operations, respectively. The number of common mutations indicates the number of mutations shared by the two tumour blocks, while mutation divergence % is the proportion of mutations that are not shared between the two tumour blocks. The whole exome sequencing, single nucleotide variants calling and computation of mutation burden, common mutations, and divergence %, were performed by Dr . Diego Mallo from Prof. Carlo Maley's lab at Arizona State University (unpublished).

Using these mutation features, we investigated the following key aspects; (1) the prognostic impact of these mutation features and (2) the association of these features with DCIS histomorphologic features and AI-TIL. The mean histomorphologic feature of the WSI from block A and B was compared with mutation burden. The histomorphologic feature distance was measured as the absolute value of the difference between the histomorphologic feature from the two blocks. The same applies to AI-TIL. The histomorphologic feature distance was compared with divergence %.

5.4.12 Statistical analysis

All correlations were computed using the Spearman method. The p-values were computed using the two-sided unpaired non-parametric Wilcoxon method (if not stated), considering $p < 0.05$ as significant.

As a primary measure for prognostic analysis, we used recurrence status (no-IBE, DCISr and IBCr) and RFS, which is defined as the period of time from initial surgery to recurrence with censoring at death or last follow-up date. The time-to-event curves were calculated using the Kaplan–Meier methods and compared with the Log-rank test. Moreover, we measured the concordance index (C-index) which is a measure of concordance between the observed and predicted survival times. It is the fraction of individuals whose expected survival times are correctly arranged out of all individuals that can actually be ordered [306].

The Cox proportional hazard method was used to quantify the hazard ratio for the effect of biomarker groups and clinical parameters. The biomarker groups and clinical variables include DCIS size, age (≤ 50 or > 50 [167]) years, grade, PR status, ER status, necrosis, surgical procedure,

hormonal treatment, and radiation treatment. Only variables with a significant p-value in the univariate Cox proportional hazards ($p < 0.05$) were included for multivariate analysis [392]. The multivariate analysis only included cases with complete data.

5.5 Results

5.5.1 Deep learning models performance evaluation

The cell classifier model achieved an overall accuracy of 0.92 (Table 5.2) and a sample image showing single cell detection and classification is presented in Figure 5.4A. In our AI-TIL scoring pipeline, lymphocytes are of particular interest. For the lymphocyte class, the model achieved sensitivity (0.92), specificity (0.98) and accuracy (0.95) (Table 5.2). This shows that the model could be reliably used.

The performance of our proposed DCIS segmentation algorithm was evaluated using 5-fold cross-validation. The segmentation model achieved a dice overlap score of 0.94($\pm$0.01), precision of 0.95($\pm$0.012), recall of 0.94($\pm$0.012), and specificity of 0.98($\pm$0.004) (Tables 5.3). The values are represented as mean $\pm$ standard deviation. Illustrative images of pathologist DCIS annotation and predicted DCIS masks are shown in Figure 5.4B. The small deviation in metrics between folds shows that the model is less sensitive to variation in the training and testing samples.

Table 5.2: **Cell classification model performance evaluation.**

	Fibroblast	**Lymphocyte**	**Tumour**
Sensitivity	0.77	0.92	0.98
Specificity	0.98	0.98	0.93
Accuracy	0.88	0.95	0.95

Table 5.3: **DCIS duct segmentation model performance evaluation.**

Fold	**Dice overlap**	**Precision**	**Recall**	**Sensitivity**	**Specificity**
Fold 1	0.93	0.93	0.93	0.93	0.98
Fold 2	0.95	0.96	0.95	0.95	0.98
Fold 3	0.94	0.94	0.93	0.93	0.98
Fold 4	0.94	0.94	0.94	0.94	0.97
Fold 5	0.96	0.96	0.96	0.96	0.98

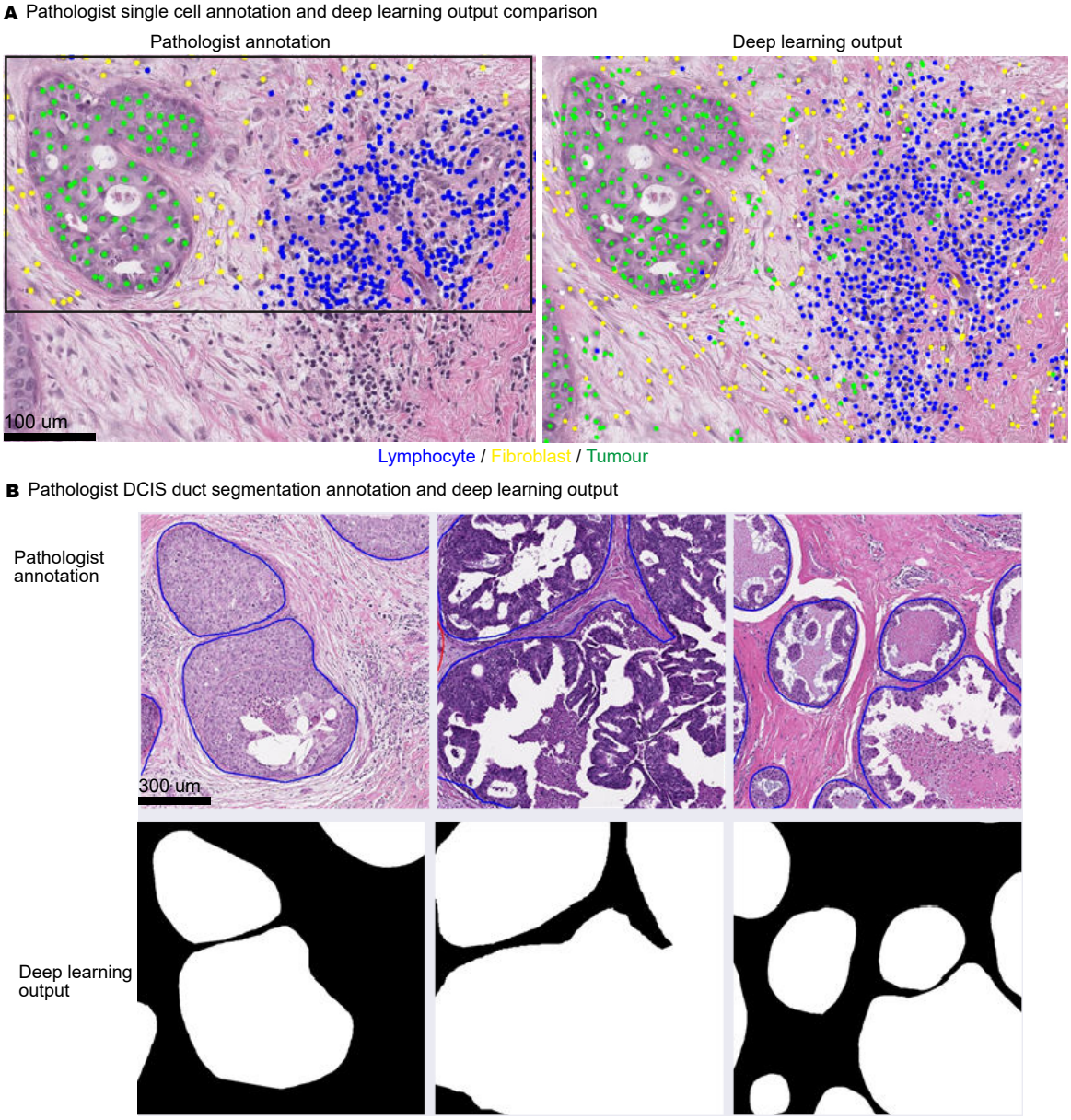


Figure 5.4: **Illustrative images comparing pathologist annotation and automated methods for single cell identification and DCIS duct segmentation: A)** Sample pair of images showing pathologist single cell annotation and deep learning single cell detection and classification output. A pathologist annotated cells only within the black rectangular box. **B)** Pairs of images illustrating ground truth (blue colour) and predicted DCIS mask by the proposed method. In the deep learning output, the white and black pixels represent, DCIS ducts and stromal regions, respectively

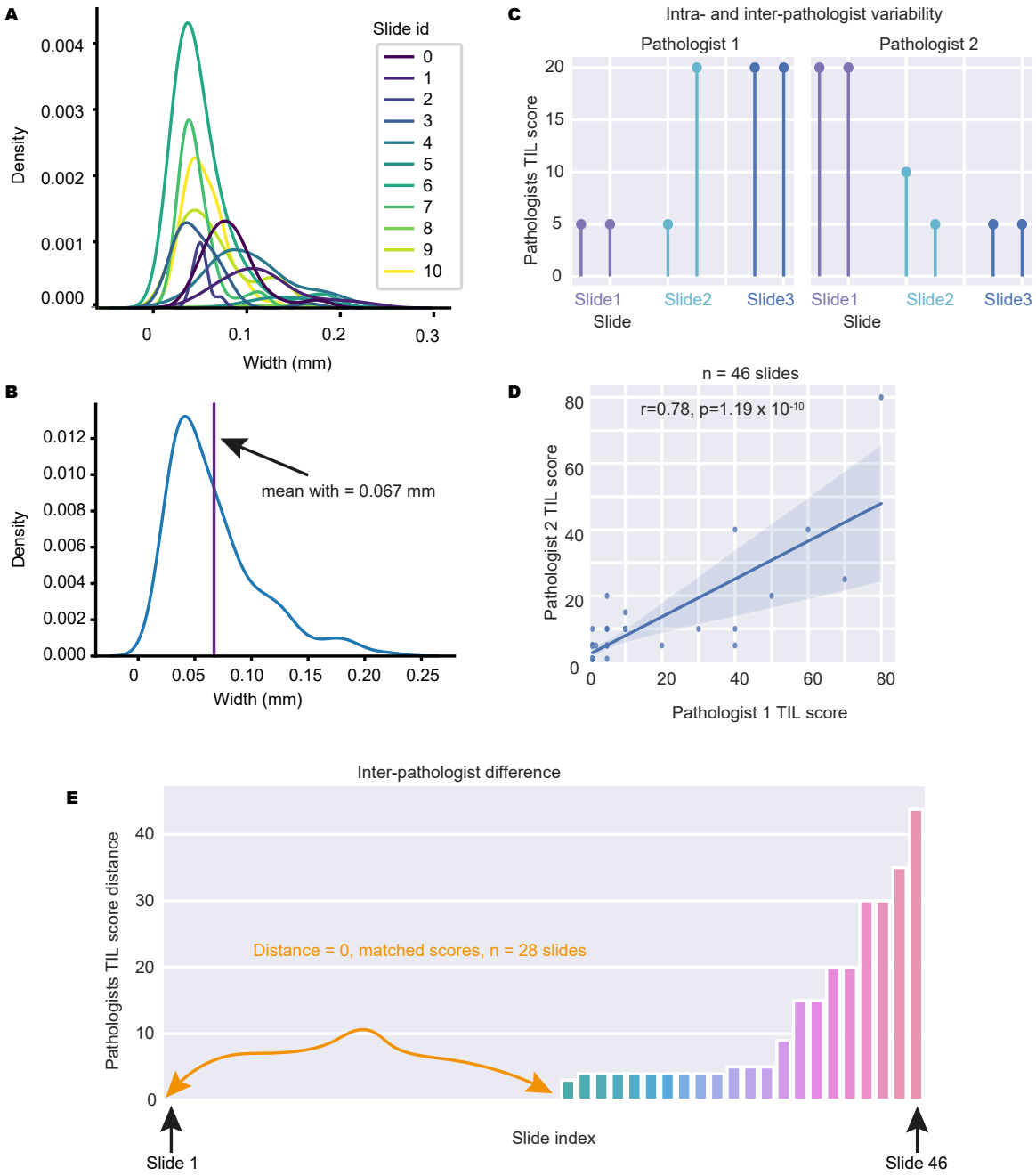


Figure 5.5: **Inter-pathologists stromal TILs scoring association: A)** Density plot showing the inter-and intra-slide variation of TILs scoring boundary width (W) drawn by a pathologist for 11 slides. **B)** Distribution of W after combining annotation from the 11 slides in (A). Each DCIS duct was considered independently. **C)** Line plots showing intra-and inter-observer variability of stromal TILs scoring for three slides or cases. **D)** Regression plot showing inter-pathologists stromal TILs score correlation. The shaded area indicates the 95% confidence interval of the regression plot. A point represents a slide. **E)** Distance between pathologists TILs scores for 46 slides. The maximum distance was 45%. The distance was computed as the absolute value of the difference between the scores by the two pathologists.

5.5.2 Pathologists stromal TILs scoring analysis

In DCIS, there is no single agreed stromal boundary for stromal TILs estimation [151]. To get an idea of the boundary width that expert pathologists use for TILs, we collected manual DCIS duct and stromal boundary annotations for 11 WSI as shown in Figure 5.3. On average, 39 DCIS ducts were annotated per WSI. For every duct, the average width of the stromal boundary, W, was computed (Methods). There was high intra-and inter-slide variation in the value of W (Figure 5.5A). At slide level, the average standard deviation of W was $0.032mm$. Taking each DCIS duct individually, the average W was $0.067mm$ with a standard deviation of $0.041mm$. The values of W ranged from $0.014mm$ to $0.228mm$ (Figure 5.5B). This shows high variability in the stromal boundary width annotation by a pathologist.

Then, we investigated intra- and inter-pathologist variation of stromal TILs scoring on 46 slides. Pathologists' scores are continuous values between 0 and 100% (Figure S4.2A). The pathologists were given three slides with duplicates among the 46 slides without letting them know the existence of duplicates. The scores by both pathologists for the three slides are shown in Figure 5.5C. For two slides, there was an intra-pathologist agreement but an inter-pathologist variation. In one slide, both pathologists showed intra-pathologists variation. Overall, the scores of the two pathologists were significantly correlated (Spearman $r = 0.78$, $p = 1.19 \times 10^{-10}$, Figure 5.5D). The two pathologists' scores showed agreement in 52.2% of the 46 slides (Figure 5.5E and Figure S4.2B). In the remaining 47.8% slides, the difference in the pathologists stromal TILs score ranged between 3% and 45%. The pathologists show high concordance for slides with low stromal TILs. 16 slides out of the 46 slides scored 1% by both pathologist (Figure S4.2B). High inter-pathologists variation was observed mostly in WSI with high stromal TILs. For example, for the sample image shown in Figure S4.3, the stromal TILs score by the two pathologists was 25% and 70%. These results suggest there is intra- and inter-pathologist variation for DCIS stromal TILs scoring.

5.5.3 The association between manual and automated stomal TILs scoring methods

We analysed the correlation between pathologists' scores and AI-TIL generated using our automated pipeline. Our pipeline allows estimation of stromal TILs infiltration at different boundary scales (Figure 5.6A). We computed AI-TIL for W = {0.03, 0.05, 0.067, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8}mm stomal boundaries. The AI-TIL values computed using {0.03, 0.05, 0.067, 0.08, 0.1}mm boundaries were correlated (Figure S4.4). Moreover, the AI-TIL values computed using {0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 }mm boundary widths were also correlated (Figure S4.4). The correlation coefficient between AI-TIL score and pathologists' average score asymptotically increased until $W = 0.2mm$ and converges to about Spearman$r = 0.65$, $p = 1.12 \times 10^{-6}$ (Figure 5.6B, C, and Table S4.1). The pathologists' score was not correlated with cAI-TIL, computed using $W = 0.03mm$, (Spearman $r = 0.23$, $p = 0.12$, Figure 5.6B, and Table S4.1). The higher correlation at higher W compared to

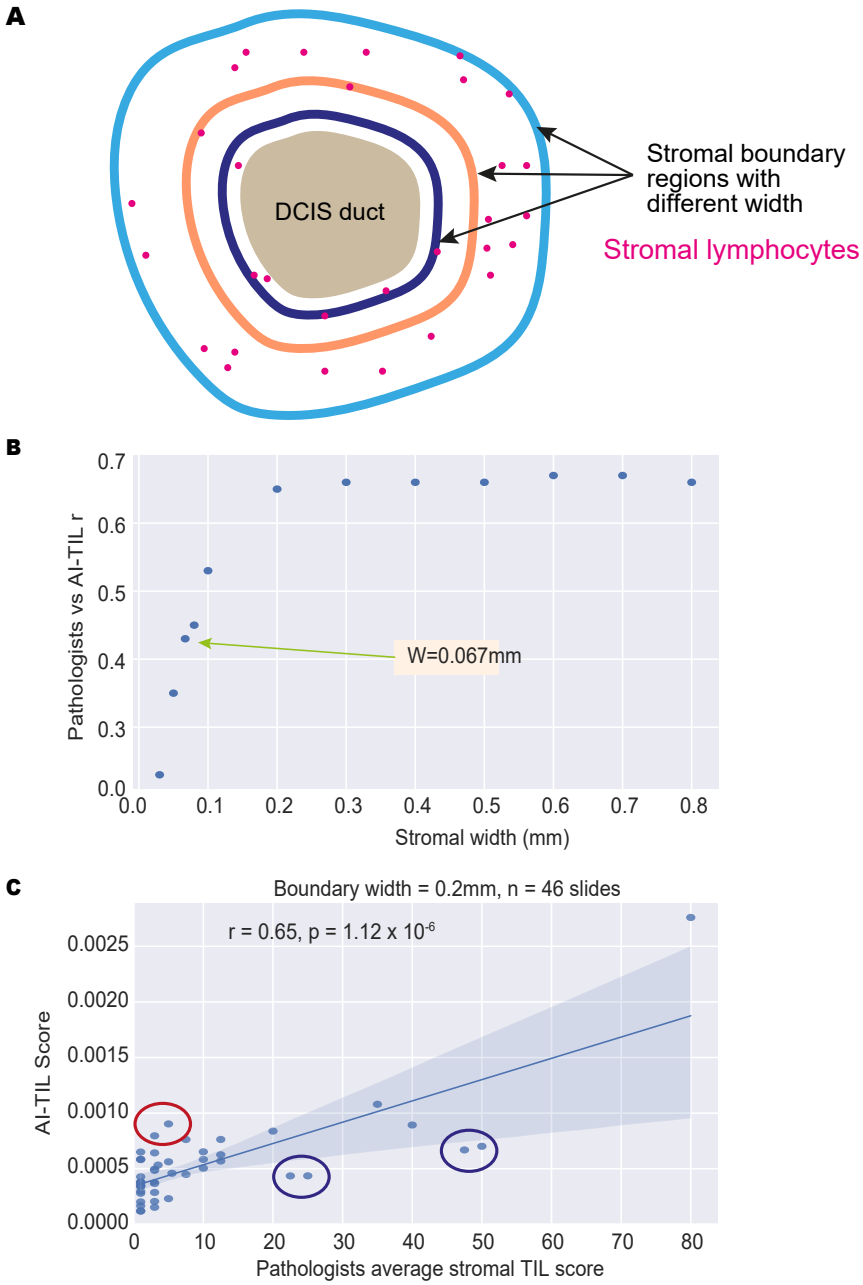


Figure 5.6: **Comparison between manual stromal TILs score and AI-TIL: A)** A cartoon showing DCIS duct and stromal boundary annotation with different widths. **B)** Correlation (r) between pathologists' score and AI-TIL for $W = \{0.03, 0.05, 0.067, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}mm$. The $W = 0.067mm$ represents the width obtained from a pathologist annotation presented in Figure 5.5B. **C)** Regression plot showing correlation between pathologists (average) TILs scoring and AI-TIL for $W = 0.2mm$. The shaded area indicates the 95% confidence interval of the regression plot. A point represents a slide. Outlier slides are annotated by the dark red circle (slides with matched pathologists' stromal TILs scores) and dark blue (slides with high inter-pathologist variability).

lower W value suggests that pathologists visually capture a large stromal TILs boundary, even though exact annotation shows an average of $W = 0.067mm$ (Figure 5.5B).

A perfect correlation between the automated and manual scoring was not expected since the automated method quantifies the number of stromal TILs per unit tissue area, whereas the pathologists report the percentage of the stromal area covered by TILs. Moreover, the manual TILs score is compared against the mean score of the two pathologists' scores, and cases where the two pathologists disagree will impact the degree of correlation. Similar to the inter-pathologists agreement, we observed disagreement between the manual and automated method for WSI with high stromal TILs (Figure S4.3). A previous study by Thagaard et al. [174] comparing automated stromal TILs, and manual scoring also observed discrepancies on slides with high TILs. In Figure 5.5C, for the four outliers slide (outside the 95% CI of the regression plot) annotated by dark blue circles, the difference between the two pathologists' manual scores ranged from 20% to 45%. We also observed disagreement between pathologists and the automated method on slides with inter-pathologist matched low stromal TILs (dark red annotation in Figure 5.5C). These slides were characterised by the presence of scarce stomal area or immune hotspots in the vicinity of DCIS duct, which could have been classified as non-immune hotspots by the automated method (Figure S4.5).

5.5.4 The association between spatial stromal TILs score, DCIS duct morphological features and clinical variables

The percentage of lymphocyte infiltration was not different between no-IBE, DCISr and IBCr patients (Figure S4.6A). We computed AI-TIL for $W = \{0.03, 0.05, 0.067, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}mm$ stomal boundaries. Using our approach, which assesses TILs in the neighbourhood of DCIS duct, the stromal TILs score computed using a smaller boundary was more prognostic than stromal TILs score computed using a larger stromal boundary (Figure 5.7A and Figure S4.6B, C). Diagnostic samples of no-IBE patients showed higher cAI-TIL (computed using $W = 0.03mm$) compared with DCISr patients and IBCr patients (p=0.0083 and p=0.041, Figure 5.7A, respectively). There was no difference in cAI-TIL between the diagnostic samples of DCISr and IBCr patients. Moreover, after combining invasive and DCIS recurrence groups, no-IBE patients showed significantly higher cAI-TIL compared with recurrence samples but not for larger boundaries ($p = 0.011$, Figure 5.7C and Figure S4.6C). For $W = 0.05mm$, diagnostic samples of no-IBE patients showed higher AI-TIL compared with IBCr patients (p=0.04, Figure 5.7A). However, for $W = 0.05mm$, there was no difference in AI-TIL when comparing DCISr vs IBCr patient and patients with recurrence vs no-IBE patient (Figure 5.7A and Figure S4.6C), respectively). For stromal boundaries $> 0.05mm$, AI-TIL was not different among the recurrence groups (Figure 5.7A and Figure S4.6B, C). Despite a higher correlation of AI-TIL with pathologists' scores for higher W (Figure 5.6B), AI-TIL was better associated with recurrence status for smaller W values. Thus, for the subsequent analyses, we only considered cAI-TIL (computed using $W = 0.03mm$). Using 5% threshold on pathologist score (the average of the two pathologists' stromal TILs score) [146], the cAI-TIL score for pathologists' high

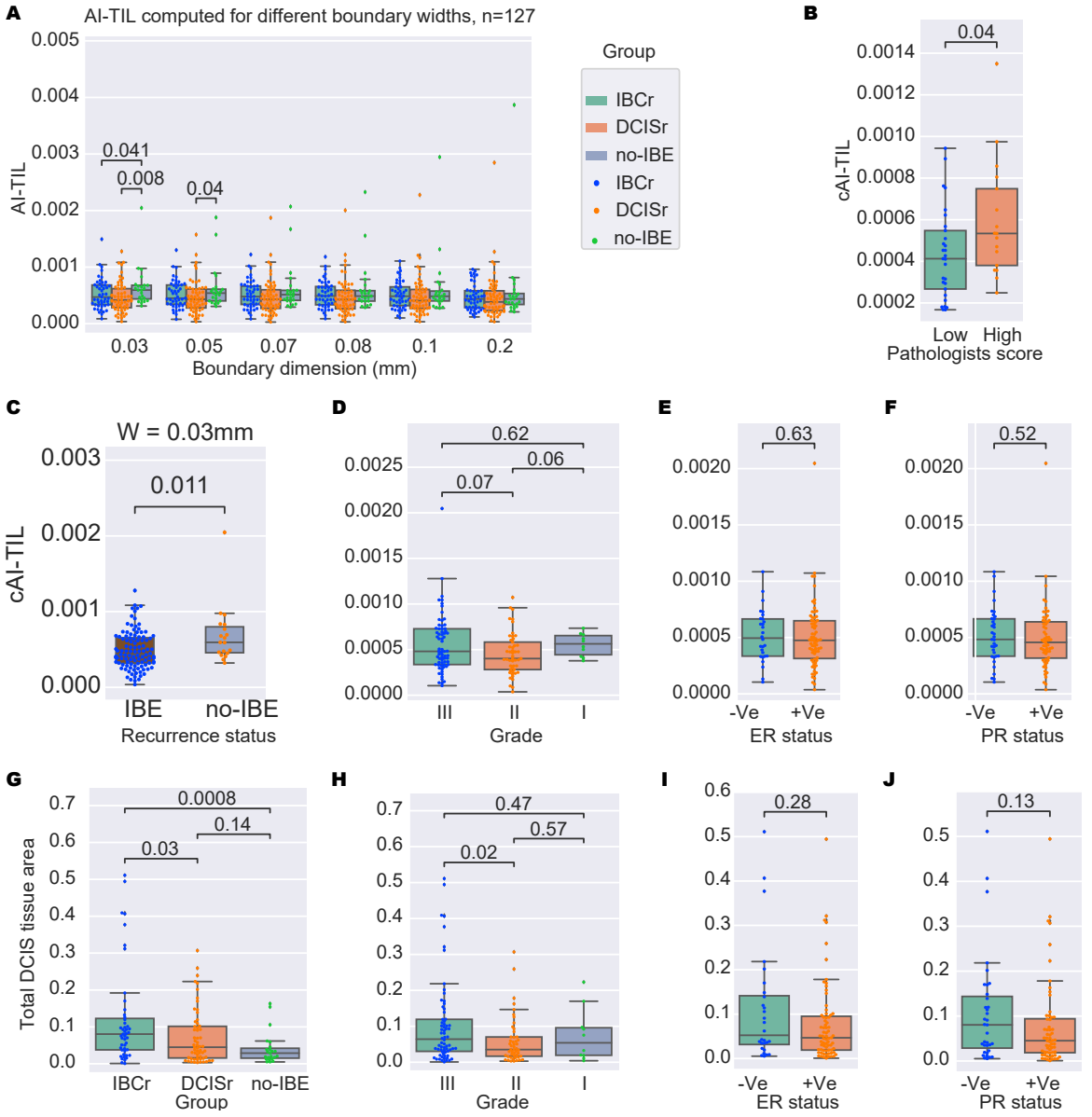


Figure 5.7: **Association of morphologic features and automated stromal TILs with recurrence status, biomarkers and clinical variables: A** Boxplots showing the difference in AI-TIL between recurrence groups for various TILs scoring boundary size. Only P values < 0.05 are displayed. **B)** Boxplots showing the difference in cAI-TIL (stromal boundary width = $30\mu m$) between Low and High stromal TILs groups from pathologists' scores; Low (stromal TILs $\leq 5\%$) and High (stromal TILs $> 5\%$). **C-J)** Boxplots showing difference in AI-TIL (**C-F**) and total DCIS tissue area (**G-J**) between different patient groups. The label of the y-axis is shared among all plots in a row. The number of patients: recurrence groups (no-IBE (n=19), DCISr (n=61) and IBCr (n=47)); Grade (III (n=68), II (n=49), and I (n=10)); ER status (+ve (n=71) and -Ve (n=28), and PR status (+Ve (n=60) and -Ve (n=35)). ns = not significant; PR = progesterone receptor; ER = estrogen receptor; +Ve = Positive; -Ve = Negative. The p-values were computed using a two-sided unpaired non-parametric Wilcoxon method considering p<0.05 significant.

score group was significantly higher than the cAI-TIL score for the pathologists' low group ($p = 0.04$, Figure 5.7B).

Then, we analysed the association of cAI-TIL, DCIS duct solidity and DCIS area with biomarkers including DCIS grade, ER and PR. cAI-TIL was not associated with DCIS grade, ER, and PR status (Figure 5.7C-F). Moreover, DCIS duct solidity was not associated with DCIS recurrence, grade, ER, and PR status (Figure S4.7E-H). Patients with DCIS and invasive recurrence showed higher total DCIS tissue area compared to patients with no-IBE ($p = 8.0 \times 10^{-4}$ and $p = 0.031$, Figure 5.7G, respectively). However, the average area of DCIS duct was not associated with recurrence (Figure S4.7A). Grade III patients also showed higher total DCIS tissue area and average DCIS duct area compared to grade II ($p = 0.022$, Figure 5.7H and $p = 0.003$, Figure S4.7B). However, total tissue area and average DCIS duct area were not associated with ER and PR status (Figure 5.7I, J and Figure S4.7C, D).

Our automated pipeline allows evaluation of slide level heterogeneity of cAI-TIL and DCIS duct morphologic features. Figure S4.8 shows the density plots of DCIS duct level cAI-TIL distribution for multiple slides. As can be seen in the density plot, there is intra- and inter-slide cAI-TIL variability or heterogeneity among DCIS duct. We then asked whether this heterogeneity is associated with the morphological phenotype of the DCIS duct. To test this hypothesis, we recorded cAI-TIL and morphological features, including area, solidity, and eccentricity at DCIS duct level. Then, we analysed the correlation between cAI-TIL and morphologic features, treating every DCIS duct independently. However, we did not observe a correlation between cAI-TIL and DCIS duct morphological features such as area, solidity and eccentricity (Figure S4.9). This suggests the circumferential TILs is not associated with DCIS duct morphology.

We then explored if the heterogeneity of DCIS duct level cAI-TIL and DCIS duct morphology are associated with recurrence, grade, ER status, and PR status. The heterogeneity was measured by the variance of DCIS duct level features within a WSI. When there are multiple WSI per patient, an average value of the variance was computed. Grade III DCIS showed significantly higher heterogeneity of cAI-TIL and DCIS duct area compared to grade II DCIS ($p = 0.012$ and $p = 0.019$ Figure S4.10B, F, respectively). However, heterogeneity of cAI-TIL, DCIS duct area and DCIS solidity were not associated with recurrence, ER status or PR status (Figure S4.10). Furthermore, 103 out of 127 patients had two H&E slides from two tumour blocks. The cAI-TIL and morphometric distance measured by the absolute value of the difference in cAI-TIL, DCIS duct area and DCIS solidity were not associated with recurrence, grade, ER status or PR status (Figure S4.11).

5.5.5 Correlation between histologic features

Subsequently, the correlation of DCIS morphologic features and immune infiltration was analysed. Expectedly, the number of DCIS ducts was significantly correlated with the total DCIS area, both normalised by the amount tissue (Spearman $r = 0.85,\ p < 0.001$, Figure 5.8A). DCIS solidity was negatively correlated with eccentricity (Spearman $r = -0.62,\ p < 0.001$ Figure 5.8A). The cAI-TIL

was not correlated with DCIS morphology features such as tissue area, the number of DCIS ducts, duct area, solidity and eccentricity (Figure 5.8A). However, it was significantly correlated with lymphocyte percentage in the WSI, irrespective of the recurrence status (Spearman $r = 0.6,\ p < 0.001$, Figure 5.8A, B).

5.5.6 Association between AI-TIL and recurrence-free survival

To investigate the association between AI-TIL and RFS, Kaplan–Meier estimation statistics was applied using different commonly used stratification quantiles (25%, 33%, 50% 67%, and 75%)(Figure 5.9A). First, using our automated image analysis pipeline, we computed AI-TIL for different stromal boundary values, $W = \{0.03, 0.05, 0.067, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}mm$. Then, to investigate if AI-TIL computed using a specific stromal boundary width is associated with RFS, we stratified the patients into groups with high and low AI-TIL scores. We evaluated different stratification quantiles as shown in Figure 5.9A. The RFS curves of these groups were estimated using the Kaplan–Meier method and compared using the Log-rank test. As shown in Figure 5.9A, cAI-TIL (AI-TIL computed using stromal boundary width of 0.03mm) consistently resulted in improved RFS prediction for 25%, 33%, 50%, and 67% stratification quantiles. Using median cAI-TIL to stratify patients into high and low cAI-TIL values, a higher cAI-TIL value was associated with longer RFS (Log-rank $p = 0.036$, Figure S4.13). The cAI-TIL achieved concordance index of 0.58 (Table S4.2) in predicting RFS. Even though the concordance index of cAI-TIL is low, it achieved the highest concordance index compared with the AI-TIL computed using wider stromal boundaries (Table S4.2).

5.5.7 Univariate and multivariate analyses

To evaluate and compare the prognostic relevance of cAI-TIL as a continuous variable, DCIS morphological features, clinical variables and biomarkers, we conducted univariate Cox proportional hazards regression analysis. The results for the univariate analyses are shown in Table 5.4. Using continuous values of cAI-TIL, higher cAI-TIL was associated with significantly prolonged RFS ($HR = 0.62$, 95% CI: 0.44-0.88 $p = 0.008$, Table 5.4). Moreover, ER+ status, PR+ status, and hormonal therapy were associated with prolonged RFS. However, grade III and the presence of necrosis were associated with shorter RFS (Table 5.4). The size of DCIS, number of DCIS ducts, total DCIS duct area, average DCIS duct area, lymphocyte % and type of surgical procedure (lumpectomy vs mastectomy) were not associated with RFS (Table 5.4).

Then, to explore if cAI-TIL predicts the risk of recurrence independent of the standard biomarkers and clinical variables, we conducted a multivariate COX proportional hazard regression analysis. Only variables with significant p-values from the univariate analysis were included in the multivariate analysis [392]. After the multivariate analysis, cAI-TIL was the only parameter prognostic for RFS ($HR = 0.56$, 95% CI: 0.39-0.80, $p = 0.002$, Figure 5.9B). Thus, cAI-TIL is a predictor of an improved RFS in DCIS independent of grade, ER status, PR status, necrosis and hormonal therapy.

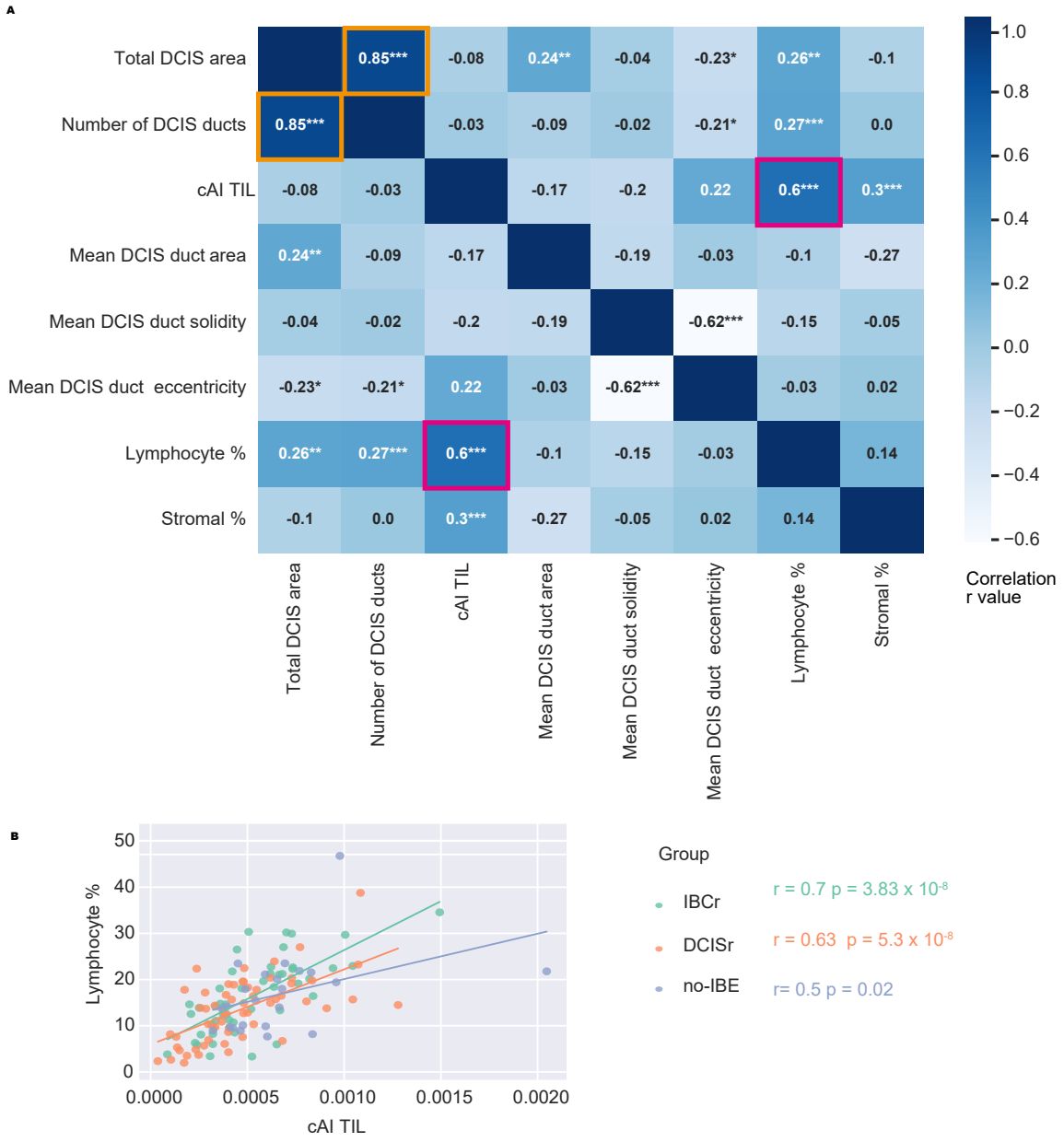


Figure 5.8: **Correlation between histologic features: A)** Spearman correlation matrix of histomorphologic features. The correlation was computed at the patient level (n = 127). *(p<0.01), **(p<0.001), and ***(p<0.0001). To correct for multiple testing, we applied the Benjamini-Hochberg method. **B)** Scatter plot showing correlation between cAI-TIL (boundary = 0.03*mm*) and lymphocyte percentage.

Table 5.4: **Univariate analysis of cAI-TIL, lymphocyte percentage, DCIS duct morphological features, clinical parameters and biomarkers for RFS.** The number of DCIS ducts represents an increment of 1 DCIS duct per mm^2 area. Continuous values of cAI-TIL were used but normalised to increment of 300 stromal TILs per mm^2 stromal area for ease of interpretation [174].

Feature	**HR**	**95% CI**	**P value**
cAI-TIL	0.62	0.44-0.88	**0.008****
Lymphocyte %	1.00	0.97-1.04	0.93
Size of DCIS (cm)	1	1.000-1.001	0.6
Number of DCIS ducts	1.4	0.77-2.4	0.29
Total DCIS area	2.26	0.27-19.13	0.45
Mean DCIS duct area	1	0.999-1.00	0.32
Mean DCIS duct eccentricity	0.0339	0-5.36	0.19
Age			
>50	1.07	0.66-1.74	0.78
Grade			
Grade II	1.73	0.66-4.56	0.27
Grade III	2.6	1.02-6.64	**0.046***
ER status			
Positive	0.46	0.27-0.78	**0.0043****
PR status			
Positive	0.54	0.33-0.88	**0.013***
Necrosis			
Yes	2.15	1.12-4.15	**0.022***
Hormonal therapy			
Yes	0.59	0.36-0.97	**0.037***
Radiation therapy			
Yes	0.74	0.46-1.21	0.24
Surgical procedure			
Mastectomy	0.85	0.47-1.56	0.61

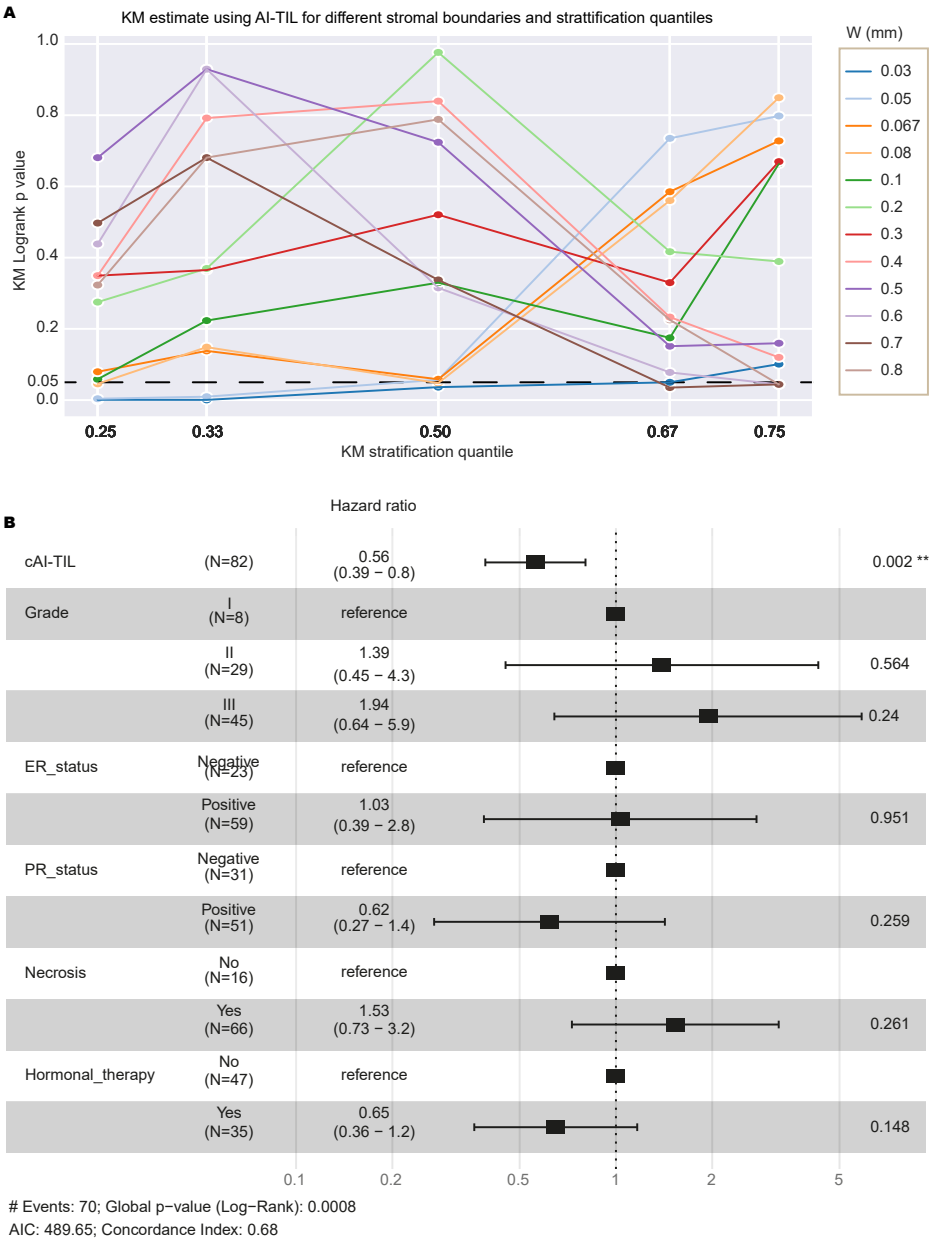


Figure 5.9: **Low cAI-TIL is associated with risk of recurrence: A)** Line plots showing the variation of RFS Kaplan–Meier (KM) estimator Log-rank p values as a function of AI-TIL computed using different stromal boundary widths (W, in millimetre) and stratification quantiles. The stratification quantiles were used to group the patients into high and low scores. The RFS curves of these groups were estimated using the KM estimator and compared using the Log-rank test. The black horizontal line shows Log-rank p $= 0.05$, a threshold for a significant difference. **B)** Forest plot showing multivariate COX proportional hazard regression analysis. Continuous values of cAI-TIL were used. cAI-TIL is stromal TILs score computed using stromal boundary width of 0.03mm. PR = progesterone receptor; ER = estrogen receptor

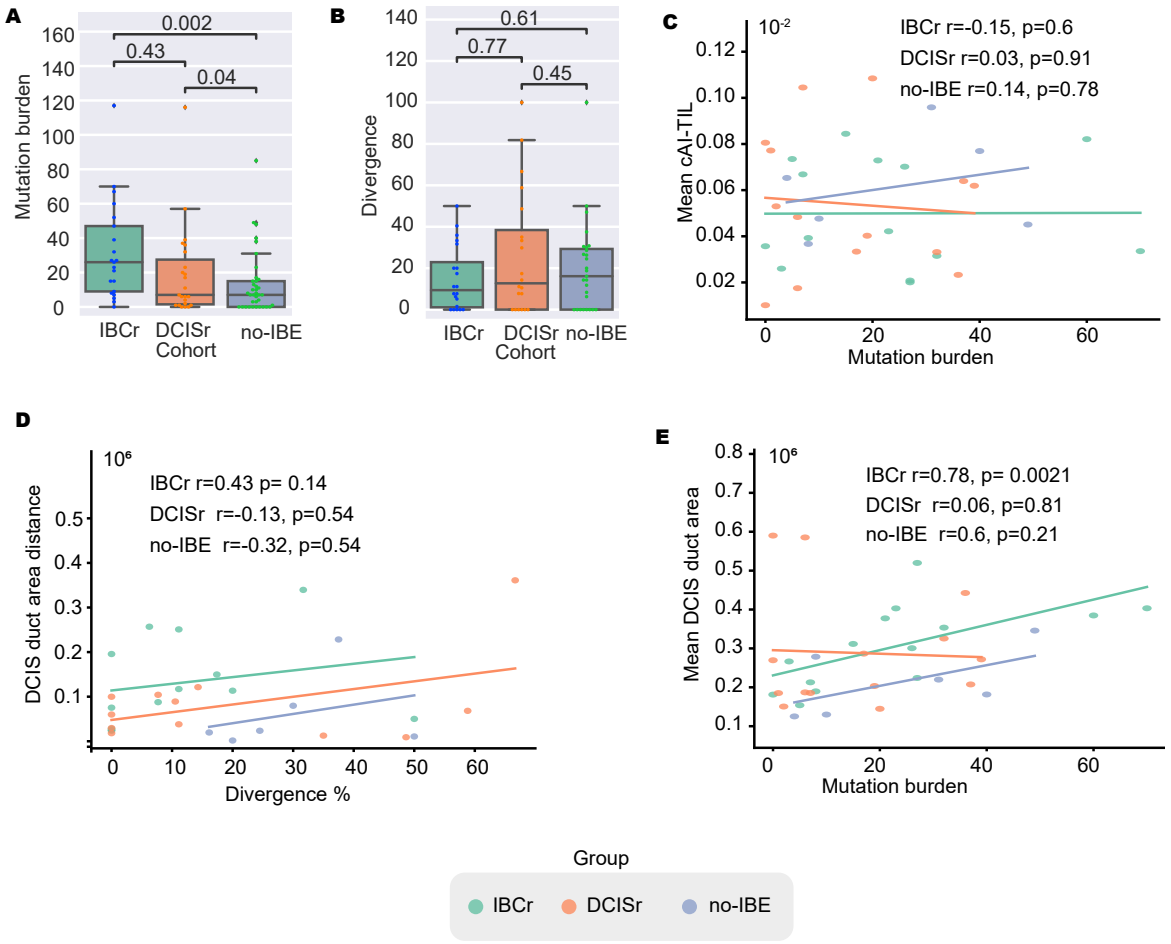


Figure 5.10: **Association between mutation data, DCIS duct morphology, cAI-TIL and recurrence: A)** Boxplots showing the difference in mutation burden between no-IBE (n=37), DCISr (n=23) and IBCr (n=21) patients. **B)** Boxplots showing difference in mutation divergence % between no-IBE (n=27), DCISr (n=20) and IBCr (n=20) patients. For mutation divergence, patients with zero mutation were excluded. **C)** Scatter plots showing the correlation between mean cAI-TIL and mutation burden. no-IBE (n=6), DCISr (n=14) and IBCr (n=14). **D)** Scatter plots showing the correlation between DCIS duct area distance and mutation divergence %. no-IBE (n=6), DCISr (n=12) and IBCr (n=13). **E)**Scatter plots showing the correlation between DCIS duct area and mutation burden. no-IBE (n=6), DCISr (n=14) and IBCr (n=14).

5.5.8 High mutation burden is associated with recurrence but not mutation divergence

Next, we analysed the association of mutation burden and mutation divergence with recurrence, clinical variables and biomarkers. The IBCr and DCISr patients showed higher mutation burden compared with no-IBE patients ($p = 0.002$ and $p = 0.04$, Figure 5.10A, respectively). However, mutation divergence % was not different between these groups (Figure 5.10B). Similarly, after combining invasive and DCIS recurrence groups, a higher mutation burden was associated with recurrence but not mutation divergence (Figure S4.16A, B). The number of shared mutations between the pair tumour block was significantly correlated with the mutation burden of the patients (Spearman$r = 0.98,\ p = 3.7 \times 10^{-57}$, Figure S4.15A). We observe a similar pattern while comparing the recurrence groups based on shared mutations (Figure S4.15B). Both mutation burden and mutation divergence % were not associated with DCIS grade, ER status, and PR status (Figure S4.16D-J).

Using Kaplan–Meier estimate, both mutation burden and divergence were not associated with RFS using median stratification (Log-rank $p = 0.09$ and Log-rank $p = 0.07$, Sup. Figure 5.10A, B, respectively). Moreover, we conducted univariate analysis on both mutation burden and divergence % as continuous variables. Similarly, both mutation burden and divergence were not associated with RFS (Table S4.5).

5.5.9 Association between cAI-TIL, DCIS duct morphological features, and mutation burden

Then, we asked if DCIS morphologic features and cAI-TIL are associated with mutation burden and mutation divergence %. The cAI-TIL and DCIS duct solidity were not associated with mutation burden or divergence % in IBCr, DCISr and no-IBE patients (Figure 5.10C, Figure S4.17D-F and Table S4.3 and S4.4). Moreover, there was no association between DCIS duct area distance and mutation divergence % in the IBCr, DCISr and no-IBE patients (Figure 5.10D). The number of DCIS ducts was not also associated with mutation burden or divergence % (Table S4.3 and S4.4). However, in IBCr patients, DCIS duct area was positively correlated with mutation burden (Spearman $r = 0.78,\ p = 0.0021$, Figure 5.10E), but not in the no-IBE and DCISr patients (Spearman $r = 0.6,\ p = 0.21$; Spearman $r = 0.06,\ p = 0.81$, Figure 5.10E, respectively). This might suggest that morphologic phenotypes associated with genomic changes might be visible in patients with adverse clinical outcomes.

5.6 Discussion

Morphological assessment of DCIS and spatial organisation of stromal TILs are gaining interest in DCIS progression study [116, 151]. Previous research on the progression of DCIS to invasive

breast cancer revealed that the process is accompanied by changes in the morphological phenotypes of membranes holding the tumour cells within the duct [116] and spatial organization of TILs in surrounding stromal regions of DCIS duct [392, 151, 167, 389]. Here, we aim to 1) develop a computational pipeline to simultaneously quantify these features from H&E image, 2) evaluate their prognostic relevance in DCIS progression, and 3) assess the association of these features and mutation burden.

The TIL-WG on breast cancer has developed a guideline for manual quantification of stromal TILs on H&E images of DCIS to ensure reproducibility and standardized reporting [151]. According to the guideline, TILs should be reported only in "specialised regions" around the DCIS duct instead of the whole stromal tissue area [151]. However, due to human visual limitations, intra- and inter-observer variability have been reported, even when following the guideline [152, 172]. Here, we compared stromal TILs scoring of 46 slides by two pathologists following the TIL-WG guidelines and the disagreement ranged between 3% and 45% (Figure 5.5E). The inter-pathologist variability was high on slides that have high TILs, which is consistent with previous studies [172, 174]. Locy et al. [172] reported that using MIHC staining reduces inter- and intra-observer variability on TILs scoring. However, MIHC staining is expensive, and there is no single marker capable of capturing all single nucleated immune cells [152]. As suggested by the TIL-WG, machine learning methods could help alleviate the perceptual and practical limits of manual scoring [152, 174]. They also suggested that these machine learning methods should follow the TIL-WG to facilitate interpretation and clinical adoption [152, 174].

Here, we present a fully automated image analysis workflow that incorporates cell and tissue level computational algorithms to mimic the TIL-WG essential guidelines to score stromal TILs on DCIS H&E images. Our single cell level and tissue level algorithms and stromal TILs scoring achieved high concordance with expert pathologists' inputs. Machine learning methods allow exact quantification of the spatial distribution of TILs in the stromal regions at different distances, which is difficult for pathologists. We observed high inter- and intra-slide variability on the stromal boundary width annotation that will be used for TILs scoring by a pathologist (range:0.014 - 0.228mm, Figure 5.5A, B). Toss et al. [146] also noted that manual scoring of larger stromal boundaries is challenging and lacks consistency compared to touching TILs [146]. We found that the correlation between pathologists and our automated pipeline scores was high for stromal boundaries $\geq 0.2mm$, though the pathologist's exact annotation was $0.067mm$. This could be explained by the variability of stromal boundary width annotation explained above. Pathologists might also look at a wider stromal boundary while scoring the slides due to the short time they spend on a WSI.

Using our automated image analysis pipeline, we demonstrated that the continuous value of cAI-TIL, which is computed using stromal boundary width of 0.03mm, is an independent predictor of the risk of recurrence in DCIS ($HR = 0.56$, 95% CI: 0.39-0.80, $p = 0.002$, Figure 5.9B), after correcting for clinical variables and biomarkers. Prior to our work, the prognostic relevance of TILs distribution in DCIS was assessed using stromal TILs%, immune hotspot, touching TILs, and circumferential TILs [167, 389, 392]. Immune hotspot and stromal TILs% were not associated with risk of relapse in DCIS

[146, 167]. However, some studies have shown that touching and circumferential TILs are associated with a risk of recurrence in DCIS following standard treatment. Xu et al. [392] and Toss et al. [146] independently reported that a high touching TILs ($> 5\%$) is associated with shorter recurrence-free survival in cohorts of 129 patients and 816 patients, respectively. In Xu et al. [392], 16 out of 129 patients showed ipsilateral breast tumour recurrence after a median follow-up of 53 months. However, Badve et al. [167] showed that circumferential TILs are associated with a favourable outcome but not touching TILs. In a study by Farolfi et al. [150], a high stromal TILs level estimated using TIL-WG guideline (boundary width not mentioned), was associated with a reduced risk of recurrence only in the group of patients who did not receive the radiation therapy. In others, no association between TILs level and the outcome was reported [148, 149].

These previous studies were based on manual TILs scoring following the guidelines by TIL-WG [151]. The discrepancies in these studies could be partly due to inter-observer variability, event rates, and endpoint variability, including ipsilateral recurrence and invasive carcinoma, and low event rates generating weak statistical evaluation [167]. Moreover, the variation could be due to the difference in the spatial configuration of stromal TILs scoring, such as the duct boundary width or the number of ducts considered. For example, in [146], a maximum of 20 ducts were considered, where small or big ducts were excluded. With automated image analysis methods like ours, the biases that come from computation and boundary configuration could be avoided. Moreover, the computational methods allow quantification of the morphology of DCIS duct.

Given that in invasive breast cancer, the basement membrane of the duct is broken and DCIS tumour cells spread into the stroma, we hypothesised that the morphological disruption in DCIS duct structure could be associated with recurrence and molecular characteristics of DCIS. Our findings show no link between these features and recurrence. However, DCIS duct area was positively associated with mutation burden in patients who developed invasive recurrence, but no association in patients with DCIS recurrence and patients with no recurrence event. This could suggest that morphologic phenotypes linked to changes in the genome are only seen in people with worse clinical outcomes.

Our work has some limitations associated with the computational pipeline development and study population, which should be considered while interpreting the results. Firstly, the sample size used in this study is small, which might introduce bias during our statistical analysis. Thus, a follow-up study with large cohorts of patients is needed to validate our results. This is currently in progress, and the image analysis pipelines developed here will be used. Secondly, regarding DCIS segmentation, limited pathologist annotation was used to train and validate our deep learning model. Since accurate DCIS duct segmentation is the heart of the automated spatial stromal TILs scoring and DCIS duct morphology estimation, more annotations from multi-centre studies capturing the inherent variation in slide digitisation might be needed to ensure the adoption of our model to other studies and in a clinical setting. Thirdly, according to TIL-WG [151], stromal TILs in necrosis and regressive hyalinisation regions should be excluded. This was not considered in our automated model development and future work will focus on automating the segmentation of these regions. Fourthly, similar to Badve et al. [167], due to longer follow-up time, the event rate in our dataset is higher compared to some previous

studies such as the study by Xu et al. [392], which consisted of only 16 recurrence events from 129 patients studied. Thus, this could introduce bias when comparing our findings with previous studies. Additionally, we had pathologist stromal TILs scoring for a limited number of patients, making it difficult to assess the additional prognostic value of the automated TILs score to pathologists' assessment. Finally, while we find that high circumferential TILs are associated with a low risk of recurrence, the type and functional status of the cells cannot be determined from this study using H&E images. Knowledge of the cells' types and functional status is essential for developing therapeutic mechanisms for DCIS. High throughput spatial staining techniques such as MIF or MIHC image could be employed to unveil the phenotypes and functional status of these cells *in situ* [116, 151]. In Chapter 2 and 4, we developed deep learning based single cell and tissue analysis methods to process MIF and MIHC images, and these methods could be employed here.

5.7 Conclusion

In this chapter, we developed an automated image analysis pipeline consisting of cell and tissue level algorithms to simultaneously interrogate the association of morphological phenotype of DCIS and spatial organisation of TILs in the surrounding stromal region of DCIS duct. Our stromal TILs scoring pipeline follows most of the essential steps set by the TIL-WG guideline for stromal TILs scoring in DCIS [151]. We found that high circumferential TILs is correlated with a low risk of recurrence independent of clinical variables and biomarkers. However, there was no association between DCIS morphological phenotype and recurrence. Taken all together, the spatial organisation of stromal TILs computed using an automated method have prognostic relevance in DCIS, and this should be validated on a larger cohort of patients for use in a clinical setting.

Chapter 6

General discussion and conclusions

In this thesis, we studied the complex spatial microenvironment of FL, MM and DCIS using tailored algorithms. These cancer types develop in different organs with unique tissue architectures; as a result, the tissue architecture of these tumours is distinct. We developed deep learning and spatial statistical methods tailored to the complexity of the tissue of these tumours to look at their spatial microenvironment and find phenotypes that were linked to the patient's prognosis. We developed deep learning based computational methods to delineate tissue compartments and spatially map cell types. Subsequently, spatial statistical methods were used to identify spatial phenotypes associated with patient treatment outcomes and recurrence free survival. By studying these multiple cancer types, we found that to evaluate the prognostication of cell phenotypes, algorithms considering the complexity of the tissue architecture of the microenvironment are essential.

The contribution of this thesis could be viewed in two ways, methodological developments, and biological insights and biomarkers development. In terms of methodological contributions, we developed novel deep learning based image analysis workflows to identify single cells and dissect tissue comportments on H&E, MIHC and MIF whole tissue section histology images. We also implemented spatial statistical analysis methods tailored to the nature of the tissue and cellular architectures to quantify cells' spatial organisation in different tissue compartments available in the tumour resection. As biological insights and biomarkers development, we identified spatial immune phenotypes and tissue morphological features that correlate with disease prognosis and provide new insights into disease biology. In the following sections, we will discuss the main findings and contributions, the limitations of this thesis, and future directions.

6.1 Methodological contributions

In **Chapter 2** and **Chapter 3**, we developed DeepMIF, a fully automated, deep learning based MIF images analysis workflow to spatially map single cell phenotypes on high-resolution image and to

interrogate the spatial immune landscape of the intra- and inter-follicular region of FL. The main methodological contribution in these chapters could be summarized as follows:

1. We developed ConCORDe-Net, a cell count regularised CNN to detect cells on MIHC images and demonstrated how to incorporate problem domain knowledge into deep learning model development to improve performance;

2. We developed the DeepMIF pipeline, a highly accurate, fast and cost-effective deep learning method to identify cell phenotypes on MIF images from its de-convoluted images, which generalises across multiple panels. To make the pipeline easily usable by the wider research community, a GUI was developed which allows batch MIF images analysis, whole slide MIF image reconstruction from tiles and visualisation. The pipeline's code and dependency libraries were packaged using Docker and it runs in both local and high-performance clusters;

3. We implemented an automated spatial analysis pipeline to quantify spatial co-localisation of cells in the intra- and inter-follicular regions of FL whole slide tissue section.

The development of ConCORDe-Net and DeepMIF took into consideration several methodological challenges in annotation data collection and model development. Training and validating deep learning models need huge amounts of data, which is laborious. For cell detection, several previous works relied on single cell segmentation annotation. ConCORDe-Net uses single cell dot annotation at the centre of the cell's nucleus, which is less time costly than single cell segmentation. Moreover, inspired by the significance of providing hints while solving a challenging problem, ConCORDe-Net incorporates cell count as a regulariser during cell detection that improved cell detection performance compared to state-of-the-art methods. Moreover, DeepMIF, which was extended from ConCORDe-Net, utilises de-convoluted images of MIF data to ensure the model's generalisability and flexibility on cell phenotype exploration. Once cells positive for each marker are detected on their corresponding de-convoluted images, a user could request to quantify and visualise cell phenotype expressing any combination of markers used in the MIF panel. Furthermore, in **Chapter 3**, a pipeline that enables spatial interrogation of cells' spatial interaction within and outside the neoplastic follicles of FL was developed, which enabled the identification of clinically relevant spatial phenotypes of FL.

Though ConCORDe-Net developed in **Chapter 2** showed significant improvement in cell detection accuracy on MIHC images, it was not robust for the detection of rare cell types which is present in the BM trephine samples of MM patients. From a tissue perspective, while lymph node biopsies of FL cases contain only the intra-follicle (neoplastic) and the inter-follicle regions, the BM trephine tissue sample is a mosaic habitat of cellular tissue (neoplastic and normal tissue), fat, bone, and blood. Considering this limitation in ConCORDe-Net and tissue architecture complexity, in **Chapter 4**, AwareNet and MoSaicNet were proposed. Moving forward, image analysis algorithms were developed to interrogate the spatial cell-cell interaction and bone morphology of MM patients. The methodological significance of **Chapter 4** could be summarized as follows:

1. We developed AwareNet, which employed a cell abundance weighting mechanism that demonstrated better cell detection of rare and abundant cells compared to other state-of-the-art methods;

2. AwareNet and MoSaicNet addresses the computational challenges of analysing BM trephine images that could have broad implications on studies of other haematologic cancers beyond MM;

3. We implemented an automated computational image processing pipeline to analyse bone heterogeneity and bone thickness from digitised trephine biopsy, providing insight into the effect of therapies on bone density and heterogeneity;

4. We implemented spatial statistical analyses of bone marrow trephine biopsies considering the tissue complexity to generate new insights into the biology and function of tumour and non-tumour cells *in situ*, unveiling their dynamic changes across disease stages and after treatment.

Using multiplex imaging technologies and machine learning algorithms (ConCORDe-Net, DeepMIF, AwareNet, MosaicNet) developed in **Chapter 3** and **Chapter 4**, we found that spatial organisation of immune cells and/or tumour cells in distinct tissue structures is prognostic in MM and FL. While multiplex technologies are great tools for exploratory study and research, their clinical translation is limited (at least up to now) because these technologies are expensive and thus not available in many hospitals. The H&E staining is cheap, ubiquitous and routinely used for cancer diagnosis. In **Chapter 5**, we extended the idea of spatial interrogation of the TME in relation to tissue structures to H&E WSI in DCIS. In DCIS, tumour cells are contained within the DCIS ducts. Recent studies have shown that stromal TILs (TILs outside the DCIS ducts) correlates with patients' prognosis of recurrence. Thus, the TIL-WG on breast cancer developed a set of guidelines for stromal TILs manual scoring from H&E images in DCIS [151]. Briefly, the stromal TILs score is the percentage of the stromal area around the DCIS ducts covered by stromal TILs where areas of normal ducts, necrosis, immune hotspots, and intra-ductal TILs are excluded [151]. Pathologists ideally generate a score per DCIS duct and report the average score over all DCIS ducts in the H&E image. The guideline is developed to ensure standardised reporting and reproducibility and to facilitate clinical adoption of stromal TILs [151, 174]. However, manual assessment has inherent limitations such as intra- and inter-observer variability [172] and shortage of experienced pathologists [393, 174]. Whole tissue section H&E are giga-pixel images, and the intra- and inter-observer variability could be due to the perceptual limitation of human observers [152] and pathologists put an estimated score, not an exact value. Thus, to alleviate these limitations, the TIL-WG suggested the development of an automated image analysis pipeline which adheres to the guidelines [152].

To automate the stromal TILs scoring and follow the essential steps in TIL-WG guideline, in **Chapter 5**, we implemented an automated image analysis pipeline which contains cell level and tissue level algorithms. The methodological significance of **Chapter 5** could be summarised as follows:

1. We implemented a GAN based DCIS segmentation workflow to localise and segment individual DCIS ducts. This workflow enables quantifying DCIS duct morphological features.

2. We optimised and validated TILs identification pipeline on H&E WSI, which contains methods for tiling, tissue segmentation, cell detection and cell classification to spatially localise the TILs within the WSI, and exclude immune hotspot. We used pre-trained deep learning models, and for scalability to a large dataset and ease of sharing, all the codes and dependencies of the pipeline were packaged using a Docker container.

3. We implemented an image analysis algorithm which delineates a stromal area of specific width around each segmented DCIS duct considering neighbouring DCIS ducts and computes duct level stromal TILs score, the density of stromal TILs in this area. Slide level score was computed as the average of scores over all DCIS ducts within the image. This algorithm mimics the pathologists' strategy as suggested by TIL-WG [151]. However, beyond the pathologists' capability, it accurately counts TILs and generates stromal TILs scores for different boundary scales.

The methodological contributions discussed above enabled the identification of features that provide new biological insight and correlate with treatment response and recurrence-free survival in FL, MM and DCIS as discussed below.

6.2 Biological insights and biomarker development

Following the development of the deep learning based single cell phenotyping method in **Chapter 2**, the model was applied to diagnostic samples of FL patients. Lymph node biopsies of FL contain morphologically distinct regions: intra-follicular (neoplastic region) and inter-follicular region. A disruption in the topographic distribution of immune cells in the intra- or inter-follicular region of the lymph node reflects a dysregulated activity of the cells and microenvironmental deviation from the normal condition [394]. Thus, we investigated if there are specific cells spatial co-localisation biomarker in the intra- and inter-follicular regions of FL that predicts the risk of recurrence after standard treatment using R-CHOP or R-CVP.

DeepMIF was employed to spatially localise different immune cell phenotypes from multiple MIF panels including immune T cells, macrophages, natural killer T cells and myeloid cells as presented in **Chapter 3**. Follicles regions were annotated by expert haematologists. Spatially resolved cells spatial co-localisation analysis revealed that co-localisation of CD8+FOXP3+ cells with CD4+CD8+ cells outside the neoplastic follicles is a predictor of a favourable outcome in FL patients. FLIPI is a clinically used prognostic score commonly used to decide the better treatment strategy and predict the likely outcome of a treatment [395]. This score is computed based on clinical information including age, haemoglobin levels, number of involved nodal areas, stage and lactate dehydrogenase levels [19, 396].

Interestingly, the spatial co-localisation of CD8+FOXP3+ with CD4+CD8+ outside the neoplastic follicles was a predictor of the risk of recurrence in FL independent of FLIPI and the abundance of CD8+FOXP3+ and CD4+CD8+ cells. High spatial co-localisation of CD8+FOXP3+ with CD4+CD8+ outside the neoplastic follicles was associated with a lower risk of recurrence in FL patients who were treated with R-CHOP or R-CVP chemoimmunotherapy drugs. Though this prognostic cell phenotype was explored using MIF staining (which is mainly employed in the research area), after validating the efficacy of this biomarker on large cohorts of patients, a triplex MIHC (which is less expensive and available in more places) could be used to stain for CD4, CD8, and FOXP3 markers. AwareNet (described in **Chapter 4**) which was optimized to detect cells on MIHC images which contain CD4, CD8, and FOXP3 markers could be employed here.

Next, we explored the association between the spatial interaction of immune cells and myeloma cells and bone physiology with MM stages and post-treatment. The cells' spatial proximity and image processing algorithms to quantify bone physiology are described in **Chapter 4**. Previous research showed that the cellular composition of the BM is different across MM disease stages and MM treatment drugs influence the cellular composition of BM microenvironment. In line with the previous works, there were a reduction of Treg cells and myeloma (cancer) cells after treatment compared with their paired diagnostic samples. Of particular interest was that the difference between MGUS and MM was in the spatial proximity of immune T cells and myeloma cells, but not the abundance of either immune T cells or myeloma cells. Moreover, our newly developed image analysis pipeline to explore bone density heterogeneity and bone thickness convincingly showed that bone heterogeneity was significantly reduced after treatment revealing the effect of therapies on bone density and bone reconstitution. Overall, the proposed methods to spatially interrogate the spatial and morphological features of BM of MM patients generated insightful views on MM disease stages and post-treatment, adding to the current literature, which is mainly focused on aspirate blood samples (without spatial context).

Similarly, we explored the prognostic relevance of the spatial pattern of stromal TILs in DCIS using a fully automated image analysis approach. There is growing evidence supporting that the host immune system is crucial in fighting cancer, and stromal TILs are increasingly being used as an important biomarker in immunotherapy clinical trials [164]. Thus, to ensure standardized reporting and reproducibility, The TIL-WG has developed a guideline to score stromal TILs in DCIS from routinely available H&E images [151]. In **Chapter 5**, we demonstrated that high stromal TILs in the neighbourhood of DCIS ducts (within $0.03mm$ distance from DCIS duct boundary) is associated with a low risk of recurrence after correcting the effect of clinical variables and biomarkers used in DCIS diagnosis. It is also worth mentioning that the total percentage of lymphocytes was not associated with recurrence status. Previous studies based on manual scoring showed conflicting results [136, 397–400]. This could be in part due to the subjective nature of manual scoring, but our automated stromal TILs scoring method generates unbiased assessment of stromal TILs.

On the debate about the role of cells in the TME on patient prognosis, this study suggests that the cells should be assessed in the context of tissue architectures they reside in and their neighbourhood

cells. I believe that the advancement of high-plex and high-resolution imaging technologies coupled with deep learning based image analysis could significantly improve our understanding of the TME and guide the identification of innovative therapeutic mechanisms.

6.3 Limitations

The limitations of this study have been discussed throughout the thesis. These could be summarised into three categories. Firstly, the datasets used contain a relatively small number of patients. This could limit the statistical power of the clinical analysis, and the effect of sampling bias could not be completely ignored, even though we used stringent statistical measures. Secondly, a limited number of human annotations and data from one centre were used to train and validate the proposed deep learning models. Having said that, our deep learning models were validated on separately held data and compared against the existing state-of-the-art methods. However, to adopt our deep learning methods and clinical findings in a clinical environment, multi-centre cohorts of data capturing the heterogeneous patient characteristics, intrinsic data variation associated with staining devices, staining settings, and tissue preprocessing are crucial. Lastly, from a clinical perspective, the scope of this study was to find spatial cell phenotypes and morphological features of TME that are associated or correlated with patients' outcomes. Causality studies and biological mechanisms were beyond the scope of this study.

6.4 Future directions

This section describes suggested research directions that could be pursued in continuation of this study's methodological and biomarker advancement.

Spatial biology technologies have brought together years of advancements in tissue microscopy, genomics, and transcriptomics, enabling simultaneous spatial interrogation of TME to assess tissue structure and cellular heterogeneity in tissue context [401–404]. These technologies powered by AI hold massive potential in understanding healthy tissue, tumour initiation and progression, drug development and prognostic feature identification. This thesis provides AI based computational tools to identify cell phenotypes, dissect tissue compartments and quantify morphological and spatial context heterogeneity from such high-plex and high-resolution MIF and MIHC imaging data. As described in **Chapter 2**, DeepMIF detects the location of cells and identifies the markers' expression status (positive or negative). The algorithm is fast and generalises to multiple panels. However, cell segmentation could provide more flexibility for downstream analysis instead of cell detection. Another interesting continuation of DeepMIF is developing a machine learning model to identify cell types and functional status on H&E stained image using co-registered MIF image. The MIF staining is expensive, but H&E is routinely available. Thus, predicting cell types and functions from H&E images will be economical.

In our FL study, using high throughput MIF spatial staining and DeepMIF, we demonstrated that co-localisation of CD8+FOXP3+ cells with CD4+CD8+ cells in the inter-follicular region is a predictor of the risk of recurrence in FL as described in **Chapter 3**. For the first time, this study reported the prognostic association of these cell phenotypes with FL recurrence. Future studies could be pursued to understand these cell types' functions and how they interact with each other and with tumour cells in the context of FL.

In our work on MM in **Chapter 4**, we developed computational methods tailored to the tissue integrity and cellular ecosystem of BM to interrogate the cells' spatial organisation and bone physiology with MM disease stage and post-treatment. The proposed methods showed new insights, unveiling the cellular organisation and bone morphological dynamic changes across disease stages and after treatment. In the current management of MM, asymptomatic MGUS and SMM precursors of MM, are not treated until end-organ damage is observed. Identifying high-risk patients and treating these patients before the event will be of great value [69]. The current risk model focuses mainly on clinical metrics, and molecular profiling BM [87, 88]. Moreover, sequencing based studies have identified some genomic alteration associated with the progression of MM from precursors but lacked consistency [405]. The proposed methods in **Chapter 4** could be used on a cohort(s) containing progressor and regressor patients to explore bone morphological features and cellular spatial phenotypes associated with the risk of progression and to evaluate if incorporating these features could improve the risk prediction accuracy. Moreover, another fascinating research direction is exploring the association of bone morphologic features and cells spatial organisation metrics with patients' clinical outcomes such as treatment response, and survival.

Finally, in our work on DCIS, we developed a computational image analysis pipeline to evaluate the prognostic value of the spatial pattern of stromal TILs and DCIS morphological features in predicting recurrence. Some recent studies [136, 397–400], which are based on manual stromal TILs scoring and according to the TIL-WG [151], TILs within a "specialised stromal region" around the DCIS ducts have been found to be associated with prognosis of DCIS. However, the precise size of the region has yet to be determined [151]. Manual assessments are prone to intra- and inter-observer variability [172]. Moreover, with visual assessment, it is difficult to accurately define the size of stromal boundary and count TILs, especially for large stromal boundaries [397]. Our proposed computational methods provide accurate stromal TILs quantification at various scales, as presented in **Chapter 5**, enabling assessment of spatial differential immune infiltration. The proposed method could be employed for a large multi-centre cohort(s) of patients to objectively compare the prognostic value of stromal TILs at different scales and thereby define a specific distance for standardisation of the assessment. Our data shows that circumferential TILs are prognostic but not stromal TILs far away from DCIS ducts. This raises an interesting research question: are the type and functional status of circumferential TILs different from distant TILs? This could also provide new insight into the biology of DCIS tumour cells and their interaction with immune cells, which could pave the way to developing novel immunotherapeutic techniques. MIF staining technologies could provide

information on the functional status and phenotypes of the cells, and the DeepMIF method presented in **Chapter 2** could be employed here to map the cell phenotypes spatially.

References

[1] Sibaji Sarkar, Garrick Horn, Kimberly Moulton, et al. Cancer Development, Progression, and Therapy: An Epigenetic Overview. *International Journal of Molecular Sciences*, 14(10):21087, oct 2013.

[2] Douglas Hanahan and Robert A. Weinberg. Hallmarks of Cancer: The Next Generation. *Cell*, 144(5):646–674, mar 2011.

[3] Steven I. Hajdu. A note from history: Landmarks in history of cancer, part 1. *Cancer*, 117(5):1097–1102, mar 2011.

[4] Anna di Lonardo, Sergio Nasi, and Simonetta Pulciani. Cancer: We should not forget the past. *Journal of Cancer*, 6(1):29–39, 2015.

[5] Roghayyeh Baghban, Leila Roshangar, Rana Jahanban-Esfahlan, et al. Tumor microenvironment complexity and therapeutic implications at a glance. *Cell Communication and Signaling 2020 18:1*, 18(1):1–19, apr 2020.

[6] Gabriele Ciasca, Tanya Enny Sassun, Eleonora Minelli, et al. Nano-mechanical signature of brain tumours. *Nanoscale*, 8(47):19629–19643, dec 2016.

[7] David Attwell, Anusha Mishra, Catherine N. Hall, et al. What is a pericyte? *Journal of Cerebral Blood Flow & Metabolism*, 36(2):451, feb 2016.

[8] Melissa R. Junttila and Frederic J. De Sauvage. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature 2013 501:7467*, 501(7467):346–354, sep 2013.

[9] Girieca Lorusso and Curzio Rüegg. The tumor microenvironment and its contribution to tumor evolution toward metastasis. *Histochemistry and cell biology*, 130(6):1091–1103, dec 2008.

[10] Alireza Labani-Motlagh, Mehrnoush Ashja-Mahdavi, and Angelica Loskog. The Tumor Microenvironment: A Milieu Hindering and Obstructing Antitumor Immune Responses. *Frontiers in Immunology*, 11:940, may 2020.

[11] Pia Nyberg, Tuula Salo, and Raghu Kalluri. Tumor microenvironment and angiogenesis. *Frontiers in bioscience : a journal and virtual library*, 13(17):6537–6553, may 2008.

[12] T. L. Whiteside. The tumor microenvironment and its role in promoting tumor growth. *Oncogene*, 27(45):5904, oct 2008.

[13] Isaac P. Witz. The Tumor Microenvironment: The Making of a Paradigm. *Cancer Microenvironment*, 2(Suppl 1):9, 2009.

[14] Vivien Marx. Method of the Year: spatially resolved transcriptomics. *Nature Methods 2021 18:1*, 18(1):9–14, jan 2021.

[15] Wei Chang Colin Tan, Sanjna Nilesh Nerurkar, Hai Yun Cai, et al. Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. *Cancer Communications*, 40(4):135–153, apr 2020.

[16] Amelie Echle, Niklas Timon Rindtorff, Titus Josef Brinker, et al. Deep learning in cancer pathology: a new generation of clinical biomarkers. *British Journal of Cancer 2020 124:4*, 124(4):686–696, nov 2020.

[17] Miriam Hägele, Philipp Seegerer, Sebastian Lapuschkin, et al. Resolving challenges in deep learning-based analyses of histopathological images using explanation methods. *Scientific Reports 2020 10:1*, 10(1):1–12, apr 2020.

[18] Jakob Nikolas Kather, Jefree Schulte, Heike I. Grabsch, et al. Deep learning detects virus presence in cancer histology. *bioRxiv*, page 690206, 2019.

[19] Lymphoma Action, Follicular lymphoma, https://lymphoma-action.org.uk/types-lymphoma-non-hodgkin-lymphoma/follicular-lymphoma, Accessed on October 2022, 2022.

[20] Hatem Kaseb and Hani M. Babiker. Hodgkin Lymphoma. *StatPearls*, jul 2022.

[21] Antonino Carbone, Sandrine Roulland, Annunziata Gloghini, et al. Follicular lymphoma. *Nature Reviews Disease Primers*, 5(1), dec 2019.

[22] Tucker W. Lebien and Thomas F. Tedder. B lymphocytes: how they develop and function. *Blood*, 112(5):1570–1580, sep 2008.

[23] Dimitar G. Efremov, Sven Turkalj, and Luca Laurenti. Mechanisms of B Cell Receptor Activation and Responses to B Cell Receptor Inhibitors in B Cell Malignancies. *Cancers 2020, Vol. 12, Page 1396*, 12(6):1396, may 2020.

[24] Sarah M. Choi, Bryan L. Betz, and Anamarija M. Perry. Follicular Lymphoma Diagnostic Caveats and Updates. *Archives of pathology & laboratory medicine*, 142(11):1330–1340, nov 2018.

[25] M. Dreyling, M. Ghielmini, S. Rule, et al. Newly diagnosed and relapsed follicular lymphoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 32(3):298–308, mar 2021.

[26] Connie L. Batlevi, Fushen Sha, Anna Alperovich, et al. Follicular lymphoma in the modern era: survival, treatment outcomes, and identification of high-risk subgroups. *Blood Cancer Journal 2020 10:7*, 10(7):1–12, jul 2020.

[27] Daphne De Jong, Ad Koster, Anton Hagenbeek, et al. Impact of the tumor microenvironment on prognosis in follicular lymphoma is dependent on specific treatment protocols. *Haematologica*, 94(1):70–77, jan 2009.

[28] Camille Golfier and Gilles Salles. Antibody Therapy Maintenance in Follicular Lymphoma. *Hematology/Oncology Clinics of North America*, 34(4):689–699, aug 2020.

[29] Massimo Federico, Monica Bellei, Luigi Marcheselli, et al. Follicular lymphoma international prognostic index 2: a new prognostic index for follicular lymphoma developed by the international follicular lymphoma prognostic factor project. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27(27):4555–4562, sep 2009.

[30] S. Montoto, A. López-Guillermo, A. Altés, et al. Predictive value of Follicular Lymphoma International Prognostic Index (FLIPI) in patients with follicular lymphoma at first progression. *Annals of Oncology*, 15(10):1484–1489, oct 2004.

[31] A. K. Nooka, C. Nabhan, X. Zhou, et al. Examination of the follicular lymphoma international prognostic index (FLIPI) in the National LymphoCare study (NLCS): a prospective US patient cohort treated predominantly in community practices. *Annals of Oncology*, 24(2):441–448, feb 2013.

[32] Matthew J. Maurer, Emmanuel Bachy, Hervé Ghesquières, et al. Early event status informs subsequent outcome in newly diagnosed follicular lymphoma. *American Journal of Hematology*, 91(11):1096–1101, nov 2016.

[33] Clémentine Sarkozy and Matthew J. Maurer. Cause of Death in Follicular Lymphoma in the First Decade of the Rituximab Era: A Pooled Analysis of French and US Cohorts. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 37(2):144–152, jan 2019.

[34] A. Carbone, A. Gloghini, H. J. Gruss, and A. Pinto. CD40 ligand is constitutively expressed in a subset of T cell lymphomas and on the microenvironmental reactive T cells of follicular lymphomas and Hodgkin's disease. *The American Journal of Pathology*, 147(4):912, 1995.

[35] Patrizia Mondello, Angelo Fama, Melissa C. Larson, et al. Lack of intrafollicular memory CD4 + T cells is predictive of early clinical failure in newly diagnosed follicular lymphoma. *Blood cancer journal*, 11(7), jul 2021.

[36] Sandra J. Horning and Saul A. Rosenberg. The Natural History of Initially Untreated Low-Grade Non-Hodgkin's Lymphomas. *http://dx.doi.org/10.1056/NEJM198412063112303*, 311(23):1471–1475, jan 2010.

[37] Carla Casulo and Paul M. Barr. How I treat early-relapsing follicular lymphoma. *Blood*, 133(14):1540–1547, apr 2019.

[38] Thomas D. Rodgers and Paul M. Barr. Management of relapsed follicular lymphoma. *Annals of Lymphoma*, 5(0):14–14, jun 2021.

[39] Carla Casulo, Michelle Byrtek, Keith L. Dawson, et al. Early relapse of follicular lymphoma after rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone defines patients at high risk for death: An analysis from the National LymphoCare Study. *Journal of Clinical Oncology*, 33(23):2516–2522, aug 2015.

[40] Sandeep S. Dave, George Wright, Bruce Tan, et al. Prediction of Survival in Follicular Lymphoma Based on Molecular Features of Tumor-Infiltrating Immune Cells. *New England Journal of Medicine*, 351(21):2159–2169, nov 2004.

[41] Alessandro Pastore, Vindi Jurinovic, Robert Kridel, et al. Integration of gene mutations in risk prognostication for patients receiving first-line immunochemotherapy for follicular lymphoma: A retrospective analysis of a prospective clinical trial and validation in a population-based registry. *The Lancet Oncology*, 16(9):1111–1122, sep 2015.

[42] Sarah Huet, Bruno Tesson, Jean Philippe Jais, et al. A gene-expression profiling score for prediction of outcome in patients with follicular lymphoma: a retrospective training and validation analysis in three international cohorts. *The Lancet Oncology*, 19(4):549–561, apr 2018.

[43] Joaquim Carreras, Armando Lopez-Guillermo, Bridget C. Fox, et al. High numbers of tumor-infiltrating FOXP3-positive regulatory T cells are associated with improved overall survival in follicular lymphoma. *Blood*, 108(9):2957–2964, nov 2006.

[44] Pedro Farinha, Abdulwahab Al-Tourah, Karamjit Gill, et al. The architectural pattern of FOXP3-positive T cells in follicular lymphoma is an independent predictor of survival and histologic transformation. *Blood*, 115(2):289–295, jan 2010.

[45] Lilli S. Nelson, James R. Mansfield, Roslyn Lloyd, et al. Automated prognostic pattern detection shows favourable diffuse pattern of FOXP3+ Tregs in follicular lymphoma. *British Journal of Cancer*, 113(8):1197–1205, oct 2015.

[46] Richard J. Byers, Ebrahim Sakhinia, Preethi Joseph, et al. Clinical quantitation of immune signature in follicular lymphoma by RT-PCR-based gene expression profiling. *Blood*, 111(9):4764–4770, may 2008.

[47] Björn Engelbrekt Wahlin, Birgitta Sander, Birger Christensson, and Eva Kimby. CD8+ T-cell content in diagnostic lymph nodes measured by flow cytometry is a predictor of survival in follicular lymphoma. *Clinical Cancer Research*, 13(2 I):388–397, jan 2007.

[48] Takumi Sugimoto and Takashi Watanabe. Follicular Lymphoma: The Role of the Tumor Microenvironment in Prognosis. *Journal of clinical and experimental hematopathology : JCEH*, 56(1):1–19, 2016.

[49] Jordi Ochando and Mounia S. Braza. T follicular helper cells: A potential therapeutic target in follicular lymphoma. *Oncotarget*, 8(67):112116–112131, 2017.

[50] Minna Taskinen, Marja Liisa Karjalainen-Lindsberg, Heidi Nyman, et al. A high tumor-associated macrophage content predicts favorable outcome in follicular lymphoma patients treated with rituximab and cyclophosphamide- doxorubicin-vincristine-prednisone. *Clinical Cancer Research*, 13(19):5784–5789, oct 2007.

[51] Danielle Canioni, Gilles Salles, Nicolas Mounier, et al. High numbers of tumor-associated macrophages have an adverse prognostic value that can be circumvented by rituximab in patients with follicular lymphoma enrolled onto the GELA-GOELAMS FL-2000 trial. *Journal of Clinical Oncology*, 26(3):440–446, jan 2008.

[52] Angela R. Dixon, Cédric Bathany, Michael Tsuei, et al. Recent developments in multiplexing techniques for immunohistochemistry. *Expert Review of Molecular Diagnostics*, 15(9):1171–1186, sep 2015.

[53] Takahiro Tsujikawa, Sushil Kumar, Rohan N. Borkar, et al. Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. *Cell Reports*, 19(1):203–217, apr 2017.

[54] Abigail M. Lee, Andrew J. Clear, Maria Calaminici, et al. Number of CD4+ cells and location of forkhead box protein P3-positive cells in diagnostic follicular lymphoma tissue microarrays correlates with outcome. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 24(31):5052–5059, nov 2006.

[55] Armando Lopez-Guillermo, Joaquim Carreras, Giovanna Roncador, et al. High numbers of tumor-infiltrating programmed cell death 1-positive regulatory lymphocytes are associated with improved overall survival in follicular lymphoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27(9):1470–1476, mar 2009.

[56] Robert Kridel, Luc Xerri, Bénédicte Gelas-Dore, et al. The prognostic impact of CD163-positive macrophages in follicular Lymphoma: A study from the BC cancer agency and the lymphoma study association. *Clinical Cancer Research*, 21(15):3428–3435, aug 2015.

[57] Birgitta Sander, Daphne de Jong, Andreas Rosenwald, et al. The reliability of immunohistochemical analysis of the tumor microenvironment in follicular lymphoma: A validation study from the Lunenburg Lymphoma Biomarker Consortium. *Haematologica*, 99(4):715–725, 2014.

[58] Björn Engelbrekt Wahlin, Mohit Aggarwal, Santiago Montes-Moreno, et al. A unifying microenvironment model in follicular lymphoma: Outcome is predicted by programmed death-1-positive, regulatory, cytotoxic, and helper T cells and macrophages. *Clinical Cancer Research*, 16(2):637–650, jan 2010.

[59] Majda Saifi, Aurélie Maran, Pierre Raynaud, et al. High ratio of interfollicular CD8/FOXP3-positive regulatory T cells is associated with a high FLIPI index and poor overall survival in follicular lymphoma. *Experimental and Therapeutic Medicine*, 1(6):933–938, nov 2010.

[60] Annuska M. Glas, Laurent Knoops, Leonie Delahaye, et al. Gene-expression and immunohistochemical study of specific T-cell subsets and accessory cell types in the transformation and prognosis of follicular lymphoma. *Journal of Clinical Oncology*, 25(4):390–398, feb 2007.

[61] Tomás Álvaro-Naranjo, Marylène Lejeune, Maria Teresa Salvadó, et al. Immunohistochemical patterns of reactive microenvironment are associated with clinicobiologic behavior in follicular lymphoma patients. *Journal of Clinical Oncology*, 24(34):5350–5357, dec 2006.

[62] Christian M. Schürch, Salil S. Bhate, Graham L. Barlow, et al. Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell*, 182(5):1341–1359.e19, sep 2020.

[63] Siddharth Samsi, Gerard Lozanski, Arwa Shana'ah, et al. Detection of Follicles from IHC Stained Slides of Follicular Lymphoma Using Iterative Watershed. *IEEE transactions on bio-medical engineering*, 57(10):2609, oct 2010.

[64] Caglar Senaras, Muhammad Khalid Khan Niazi, Vidya Arole, et al. Segmentation of follicles from CD8-stained slides of follicular lymphoma using deep learning. In John E. Tomaszewski and Aaron D. Ward, editors, *Medical Imaging 2019: Digital Pathology*, volume 10956, page 25. SPIE, mar 2019.

[65] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9351, pages 234–241. Springer, may 2015.

[66] Chris J. Rose, Khimara Naidoo, Vanessa Clay, et al. A statistical framework for analyzing hypothesized interactions between cells imaged using multispectral microscopy and multiple immunohistochemical markers. *Journal of Pathology Informatics*, 4(Suppl):4, mar 2013.

[67] Andreas Heindl, Ivana Sestak, Kalnisha Naidoo, et al. Relevance of spatial heterogeneity of immune infiltration for predicting risk of recurrence after endocrine therapy of ER+ breast cancer. *Journal of the National Cancer Institute*, 110(2):166–175, feb 2018.

[68] Shaji K. Kumar, Vincent Rajkumar, Robert A. Kyle, et al. Multiple myeloma. *Nature Reviews Disease Primers 2017 3:1*, 3(1):1–20, jul 2017.

[69] Charlotte Pawlyn and Gareth J. Morgan. Evolutionary biology of high-risk multiple myeloma. *Nature Reviews Cancer*, 17(9):543–556, 2017.

[70] Kimberly Noonan and Ivan Borrello. The immune microenvironment of myeloma. *Cancer Microenvironment*, 4(3):313–323, 2011.

[71] Niels W.C.J. van de Donk, Charlotte Pawlyn, and Kwee L. Yong. Multiple myeloma. *The Lancet*, 397(10272):410–427, jan 2021.

[72] Jun Yokota. Tumor progression and metastasis. *Carcinogenesis*, 21(3):497–503, mar 2000.

[73] Vitor H. Teixeira, Christodoulos P. Pipinikas, Adam Pennycuick, et al. Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. *Nature Medicine*, 25(3):517–525, 2019.

[74] Gareth J. Morgan, Brian A. Walker, and Faith E. Davies. The genetic architecture of multiple myeloma. *Nature Reviews Cancer 2012 12:5*, 12(5):335–348, apr 2012.

[75] Yamni Nigam. The lymphatic system 2: structure and function of the lymphoid organs, 2020.

[76] K. K. Jovanović, C. Roche-Lestienne, I. M. Ghobrial, et al. Targeting MYC in multiple myeloma. *Leukemia*, 32(6):1295–1306, jun 2018.

[77] Phuc H. Hoang, Alex J. Cornish, Sara E. Dobbins, et al. Mutational processes contributing to the development of multiple myeloma. *Blood Cancer Journal*, 9(8):60, 2019.

[78] Giada Bianchi, Shaji Kumar, Irene M. Ghobrial, and Aldo M. Roccaro. Cell Trafficking in Multiple Myeloma. *Open journal of hematology*, 3(Suppl 1):1, feb 2012.

[79] Catherine Thieblemont and Bertrand Coiffier. Lymphoma in older patients. *Journal of Clinical Oncology*, 25(14):1916–1923, may 2007.

[80] Bridget S. Wilkins. Pitfalls in bone marrow pathology: avoiding errors in bone marrow trephine biopsy diagnosis. *Journal of clinical pathology*, 64(5):380–386, may 2011.

[81] Kosei Matsue, Yuya Matsue, Kaoru Kumata, et al. Quantification of bone marrow plasma cell infiltration in multiple myeloma: usefulness of bone marrow aspirate clot with CD138 immunohistochemistry. *Hematological oncology*, 35(3):323–328, sep 2017.

[82] Charlotte Cosemans, Bénedith Oben, Ingrid Arijs, et al. Prognostic Biomarkers in the Progression From MGUS to Multiple Myeloma: A Systematic Review. *Clinical Lymphoma, Myeloma and Leukemia*, 18(4):235–248, apr 2018.

[83] Healthwise Staff. Bone Marrow Aspiration and Biopsy: Before Your Procedure, MyHealth Alberta, https://myhealth.alberta.ca/health/AfterCareInformation/pages/conditions.aspx?hwid=ud1728&, Accessed on January 20, 2023, 2021.

[84] S. Vincent Rajkumar, Meletios A. Dimopoulos, Antonio Palumbo, et al. IMWG criteria for the diagnosis of multiple myeloma, 2014.

[85] Fenghuang Zhan, Erming Tian, Klaus Bumm, et al. Gene expression profiling of human plasma cell differentiation and classification of multiple myeloma based on similarities to distinct stages of late-stage B-cell development. *Blood*, 101(3):1128–1140, feb 2003.

[86] Marina Bolzoni, Denise Toscani, Federica Costa, et al. The link between bone microenvironment and immune cells in multiple myeloma: Emerging role of CD38. *Immunology Letters*, 205(April 2018):65–70, 2019.

[87] Alessandro Laganà, Itai Beno, David Melnekoff, Violetta Leshchenko, Deepu Madduri, Dennis Ramdas, Larysa Sanchez, Scot Niglio, Deepak Perumal, Brian A. Kidd, Riccardo Miotto, Jane Houldsworth, Rita Shaknovich, Ajai Chari, Hearn Jay Cho, Bart Barlogie, Sun and Samir Parekh. Precision Medicine for Relapsed Multiple Myeloma on the Basis of an Integrative Multiomics Approach. *JCO Precision Oncology*, 2(I), 2018.

[88] Rachel Rosenthal, Elizabeth Larose Cadieux, Roberto Salgado, et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature*, 567(7749):479–485, mar 2019.

[89] Laura Chiecchio, Gian Paolo Dagrada, Ashraf H. Ibrahim, et al. Timing of acquisition of deletion 13 in plasma cell dyscrasias is dependent on genetic context. *Haematologica*, 94(12):1708–1713, dec 2009.

[90] Jakob Nikolas Kather, Meggy Suarez-Carmona, Pornpimol Charoentong, et al. Topography of cancer-associated immune cells in human solid tumors. *eLife*, 7, 2018.

[91] S. Manier, A. Sacco, X. Leleu, et al. Bone marrow microenvironment in multiple myeloma progression. *Journal of Biomedicine and Biotechnology*, 2012, 2012.

[92] Shmuel Yaccoby. Advances in the understanding of myeloma bone disease and tumour growth. *British Journal of Haematology*, 149(3):311–321, 2010.

[93] G D Roodman. SPOTLIGHT REVIEW Pathogenesis of myeloma bone disease. *Leukemia*, 23:435–441, 2009.

[94] Jung Min Kim, Chujiao Lin, Zheni Stavre, et al. Osteoblast-Osteoclast Communication and Bone Homeostasis. *Cells 2020, Vol. 9, Page 2073*, 9(9):2073, sep 2020.

[95] R. Bataille, D. Chappard, C. Marcelli, et al. Recruitment of new osteoblasts and osteoclasts is the earliest critical event in the pathogenesis of human multiple myeloma. *The Journal of clinical investigation*, 88(1):62–66, 1991.

[96] Claire M. Edwards, Junling Zhuang, and Gregory R. Mundy. The Pathogenesis of the Bone Disease of Multiple Myeloma. *Bone*, 42(6):1007, jun 2008.

[97] Giovanni D'arena, Giovanni Rossi, Luca Laurenti, et al. Circulating Regulatory T-Cells in Monoclonal Gammopathies of Uncertain Significance and Multiple Myeloma: In Search of a Role. *Journal of immunology research*, 2016, 2016.

[98] Rao H. Prabhala, Paola Neri, Jooeun E. Bae, et al. Dysfunctional T regulatory cells in multiple myeloma. *Blood*, 107(1):301–304, jan 2006.

[99] Ritu Gupta, Prasanth Ganeshan, Mrinali Hakim, et al. Significantly reduced regulatory T cell population in patients with untreated multiple myeloma. *Leukemia Research*, 35(7):874–878, jul 2011.

[100] Lydia Lee, Nouf Alrasheed, Garima Khandelwal, et al. Increased Immune-Regulatory Receptor Expression on Effector T Cells as Early Indicators of Relapse Following Autologous Stem Cell Transplantation for Multiple Myeloma. *Frontiers in immunology*, 12, feb 2021.

[101] Fabienne Lucas, Michael Pennell, Ying Huang, et al. T Cell Transcriptional Profiling and Immunophenotyping Uncover LAG3 as a Potential Significant Target of Immune Modulation in Multiple Myeloma. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation*, 26(1):7–15, jan 2020.

[102] Sylvia Feyler, Marie Von Lilienfeld-Toal, Sarah Jarmin, et al. CD4(+)CD25(+)FoxP3(+) regulatory T cells are increased whilst CD3(+)CD4(-)CD8(-)alphabetaTCR(+) Double Negative T cells are decreased in the peripheral blood of patients with multiple myeloma which correlates with disease burden. *British journal of haematology*, 144(5):686–695, mar 2009.

[103] James Favaloro, Ross Brown, Esther Aklilu, et al. Myeloma skews regulatory T and pro-inflammatory T helper 17 cell balance in favor of a suppressive state. *Leukemia & lymphoma*, 55(5):1090–1098, 2014.

[104] Daniel Lucas. Structural organization of the bone marrow and its role in hematopoiesis. *Current opinion in hematology*, 28(1):36–42, jan 2021.

[105] Denise K. Walters and Diane F. Jelinek. Multiplex Immunofluorescence of Bone Marrow Core Biopsies: Visualizing the Bone Marrow Immune Contexture. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 68(2):99–112, feb 2020.

[106] Robert A. Kyle and S. Vincent Rajkumar. Monoclonal gammopathies of undetermined significance. *Reviews in clinical and experimental hematology*, 6(3):225–252, sep 2002.

[107] ROBERT A. KYLE. "Benign" monoclonal gammopathy–after 20 to 35 years of follow-up. *Mayo Clinic proceedings*, 68(1):26–36, 1993.

[108] Khalid AbdulJabbar, Shan E.Ahmed Raza, Rachel Rosenthal, et al. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nature Medicine 2020 26:7*, 26(7):1054–1062, may 2020.

[109] Qingzhu Jia, Aoyun Wang, Yixiao Yuan, et al. Heterogeneity of the tumor immune microenvironment and its clinical relevance. *Experimental Hematology & Oncology 2022 11:1*, 11(1):1–14, apr 2022.

[110] Chiara Molinari, Giorgia Marisi, Alessandro Passardi, et al. Heterogeneity in Colorectal Cancer: A Challenge for Personalized Medicine? *International Journal of Molecular Sciences*, 19(12), dec 2018.

[111] Mohamed E Salama, Nicholas Landis, Holger Lange, and Steven J Potts. Characterization of Spatial Distribution of Myeloma Cells and Their Microenvironment in Bone Marrow; A Computer Assisted Study Using Novel Whole Slide Image Analysis Technology. *Blood*, 118(21):2889–2889, nov 2011.

[112] A. S. Hagiya, A. Etman, I. N. Siddiqi, et al. Digital image analysis agrees with visual estimates of adult bone marrow trephine biopsy cellularity. *International Journal of Laboratory Hematology*, 40(2):209–214, apr 2018.

[113] Frederik Skou Nielsen, Mads Jozwiak Pedersen, Mathias Vassard Olsen, et al. Automatic Bone Marrow Cellularity Estimation in H&E Stained Whole Slide Images. *Cytometry Part A*, 95(10):1066–1074, oct 2019.

[114] Leander van Eekelen, Hans Pinckaers, Michiel van den Brand, et al. Using deep learning for quantification of cellularity and cell lineages in bone marrow biopsies and comparison to normal age-related variation. *Pathology*, 54(3):318–327, apr 2022.

[115] Anita Bane. Ductal Carcinoma In Situ: What the Pathologist Needs to Know and Why. *International Journal of Breast Cancer*, 2013:1–7, 2013.

[116] Tyler Risom, David R. Glass, Inna Averbukh, et al. Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. *Cell*, 185(2):299–310.e18, jan 2022.

[117] Lars J. Grimm, Habib Rahbar, Monica Abdelmalak, et al. Ductal Carcinoma in Situ: State-of-the-Art Review. *Radiology*, 302(2):246–255, feb 2022.

[118] Ujas Parikh, Chloe M. Chhor, and Cecilia L. Mercado. Ductal Carcinoma In Situ: The Whole Truth. *American Journal of Roentgenology*, 210(2):246–255, oct 2017.

[119] Cancer Research UK. In situ breast carcinoma incidence statistics, https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-in-situ, Accessed on October 2022, 2021.

[120] Rafael Caparica, Marco Bruzzone, Elisa Agostinetto, et al. Tumour-infiltrating lymphocytes in non-invasive breast cancer: A systematic review and meta-analysis. *The Breast*, 59:183–192, oct 2021.

[121] Joann G. Elmore, Gary M. Longton, Patricia A. Carney, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA*, 313(11):1122–1132, mar 2015.

[122] Jaafar Makki. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. *Clinical Medicine Insights. Pathology*, 8(1):23, 2015.

[123] Meagan E. Brennan, Robin M. Turner, Stefano Ciatto, et al. Ductal carcinoma in situ at core-needle biopsy: meta-analysis of underestimation and predictors of invasive breast cancer. *Radiology*, 260(1):119–128, jul 2011.

[124] Lotte E. Elshof, Konstantinos Tryfonidis, Leen Slaets, and A. Elise Van Leeuwen-Stok. Feasibility of a prospective, randomised, open-label, international multicentre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ - The LORD study. *European Journal of Cancer*, 51(12):1497–1510, jul 2015.

[125] Islam M. Miligy, Michael S. Toss, Sho Shiino, et al. The clinical significance of oestrogen receptor expression in breast ductal carcinoma in situ. *British Journal of Cancer 2020 123:10*, 123(10):1513–1520, aug 2020.

[126] Mangesh A. Thorat, Pauline M. Levey, J. Louise Jones, et al. Prognostic value of ER and PgR expression and the impact of multi-clonal expression for recurrence in ductal carcinoma in situ: Results from the UK/ANZ DCIS trial. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 27(10):2861, may 2021.

[127] N. Bijker, M. Donker, J. Wesseling, et al. Is DCIS breast cancer, and how do I treat it? *Current treatment options in oncology*, 14(1):75–87, mar 2013.

[128] Maartje van Seijen, Esther H. Lips, Alastair M. Thompson, et al. Ductal carcinoma in situ: to treat or not to treat, that is the question. *British Journal of Cancer 2019 121:4*, 121(4):285–292, jul 2019.

[129] Alastair M. Thompson, Karen Clements, Shan Cheung, et al. Management and 5-year outcomes in 9938 women with screen-detected ductal carcinoma in situ: the UK Sloane Project. *European Journal of Cancer*, 101:210–219, sep 2018.

[130] Lotte E. Elshof, Michael Schaapveld, Marjanka K. Schmidt, et al. Subsequent risk of ipsilateral and contralateral invasive breast cancer after treatment for ductal carcinoma in situ: incidence and the effect of radiotherapy in a population-based cohort of 10,090 women. *Breast cancer research and treatment*, 159(3):553–563, oct 2016.

[131] S. Darby, P. McGale, C. Correa, et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10801 women in 17 randomised trials. *Lancet*, 378(9804):1707, nov 2011.

[132] Beth A Virnig, Michael T Torchia, Stephanie L Jarosek, et al. Use of endocrine therapy following diagnosis of ductal carcinoma in situ or early invasive breast cancer. *Data Points Publication Series*, 2012.

[133] Elizabeth M. Ward, Carol E. DeSantis, Chun Chieh Lin, et al. Cancer statistics: Breast cancer in situ. *CA: a cancer journal for clinicians*, 65(6):481–495, nov 2015.

[134] Laura C. Collins, Rulla M. Tamimi, Heather J. Baer, et al. Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. *Cancer*, 103(9):1778–1784, may 2005.

[135] Andrea V Barrio and Kimberly J Van Zee. Controversies in the treatment of ductal carcinoma in situ. *Annual review of medicine*, 68:197–211, 2017.

[136] Rafael Caparica, Marco Bruzzone, Elisa Agostinetto, et al. Tumour-infiltrating lymphocytes in non-invasive breast cancer: A systematic review and meta-analysis. *The Breast*, 59:183–192, 2021.

[137] Maria A. Lopez-Garcia, Felipe C. Geyer, Magali Lacroix-Triki, et al. Breast cancer precursors revisited: molecular features and progression pathways. *Histopathology*, 57(2):171–192, aug 2010.

[138] Jia Min B. Pang, Kylie L. Gorringe, and Stephen B. Fox. Ductal carcinoma in situ - update on risk assessment and management. *Histopathology*, 68(1):96–109, jan 2016.

[139] Lucia Salvatorelli, Lidia Puzzo, Giada Maria Vecchio, et al. Ductal Carcinoma In Situ of the Breast: An Update with Emphasis on Radiological and Morphological Features as Predictive Prognostic Factors. *Cancers*, 12(3), mar 2020.

[140] Anna K. Casasent, Mary Edgerton, and Nicholas E. Navin. Genome evolution in ductal carcinoma in situ: invasion of the clones. *The Journal of pathology*, 241(2):208–218, jan 2017.

[141] Shinn Young Kim, Seung Hyun Jung, Min Sung Kim, et al. Genomic differences between pure ductal carcinoma in situ and synchronous ductal carcinoma in situ with invasive breast cancer. *Oncotarget*, 6(10):7597–7607, 2015.

[142] Sharon Nofech-Mozes, Wedad Hanna, and Eileen Rakovitch. Molecular Evaluation of Breast Ductal Carcinoma in Situ with Oncotype DX DCIS. *The American Journal of Pathology*, 189(5):975–980, may 2019.

[143] Miglena Komforti, Sunil S. Badve, Bryan Harmon, et al. Tumour-infiltrating lymphocytes in ductal carcinoma in situ (DCIS)—assessment with three different methodologies and correlation with Oncotype DX DCIS Score. *Histopathology*, 77(5):749–759, nov 2020.

[144] Lutske Lodewijk, Jakob W Kist, Gerlof D Valk, et al. Risk of recurrence of ductal carcinoma in situ by oncotype Dx technology: Some concerns. *Cancer*, 120(7):1085–1085, apr 2014.

[145] Eileen Rakovitch, Sharon Nofech-Mozes, Wedad Hanna, et al. A population-based validation study of the DCIS Score predicting recurrence risk in individuals treated by breast-conserving surgery alone. *Breast Cancer Research and Treatment*, 152(2):389, jul 2015.

[146] Michael S. Toss, I. Miligy, Abdubaqi Al-Kawaz, et al. Prognostic significance of tumor-infiltrating lymphocytes in ductal carcinoma in situ of the breast. *Modern Pathology 2018 31:8*, 31(8):1226–1236, mar 2018.

[147] Fei Fei Xu, Sai Fang Zheng, Cheng Xu, et al. Prognostic and predictive significance of tumor infiltrating lymphocytes for ductal carcinoma in situ. *Oncoimmunology*, 10(1), 2021.

[148] G. Pruneri, Matteo Lazzeroni, V. Bagnardi, et al. The prevalence and clinical relevance of tumor-infiltrating lymphocytes (TILs) in ductal carcinoma in situ of the breast. *Annals of Oncology*, 28(2):321–328, feb 2017.

[149] Shona Hendry, Jia Min B. Pang, David J. Byrne, et al. Relationship of the Breast Ductal Carcinoma In Situ Immune Microenvironment with Clinicopathological and Genetic Features. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 23(17):5210–5217, sep 2017.

[150] Alberto Farolfi, Elisabetta Petracci, Luigi Serra, et al. Tumor-Infiltrating Lymphocytes (TILs) and Risk of a Second Breast Event After a Ductal Carcinoma in situ. *Frontiers in Oncology*, 10:1486, aug 2020.

[151] Maria Vittoria Dieci, Nina Radosevic-Robin, Susan Fineberg, et al. Update on tumor-infiltrating lymphocytes (tils) in breast cancer, including recommendations to assess tils in residual disease after neoadjuvant therapy and in carcinoma in situ: a report of the international immuno-oncology biomarker working group on breast cancer. In *Seminars in cancer biology*, volume 52, pages 16–25. Elsevier, 2018.

[152] Mohamed Amgad, Elisabeth Specht Stovgaard, Eva Balslev, et al. Report on computational assessment of Tumor Infiltrating Lymphocytes from the International Immuno-Oncology Biomarker Working Group. *npj Breast Cancer*, 6(1):1–13, dec 2020.

[153] Lironne Wein, Peter Savas, Stephen J. Luen, et al. Clinical Validity and Utility of Tumor-Infiltrating Lymphocytes in Routine Clinical Practice for Breast Cancer Patients: Current and Future Directions. *Frontiers in oncology*, 7(AUG):156, aug 2017.

[154] Shona Hendry, Roberto Salgado, Thomas Gevaert, et al. Assessing Tumor-infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method From the International Immunooncology Biomarkers Working Group: Part 1: Assessing the Host Immune Response, TILs in Invasi. *Advances in anatomic pathology*, 24(5):235–251, 2017.

[155] Anna Niwińska and Wojciech P. Olszewski. The role of stromal immune microenvironment in the progression of ductal carcinoma in situ (DCIS) to invasive breast cancer. *Breast Cancer Research*, 23(1):1–12, dec 2021.

[156] Milim Kim, Yul Ri Chung, Hyun Jeong Kim, et al. Immune microenvironment in ductal carcinoma in situ: A comparison with invasive carcinoma of the breast. *Breast Cancer Research*, 22(1):1–12, mar 2020.

[157] C. Göranzon, A. K. Kumawat, E. Hultgren-Hörnqvist, et al. Immunohistochemical characterization of lymphocytes in microscopic colitis. *Journal of Crohn's and colitis*, 7(10), nov 2013.

[158] Marie Colombe Agahozo, Dora Hammerl, Reno Debets, et al. Tumor-infiltrating lymphocytes and ductal carcinoma in situ of the breast: friends or foes? *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 31(7):1012–1025, jul 2018.

[159] Rico D. Bense, Christos Sotiriou, Martine J. Piccart-Gebhart, et al. Relevance of Tumor-Infiltrating Immune Cell Composition and Functionality for Disease Outcome in Breast Cancer. *Journal of the National Cancer Institute*, 109(1), jan 2016.

[160] Yan Mao, Qing Qu, Xiaosong Chen, et al. The Prognostic Value of Tumor-Infiltrating Lymphocytes in Breast Cancer: A Systematic Review and Meta-Analysis. *PloS one*, 11(4), apr 2016.

[161] Michael J. Campbell, Frederick Baehner, Tess O'Meara, et al. Characterizing the immune microenvironment in high-risk ductal carcinoma in situ of the breast. *Breast cancer research and treatment*, 161(1):17–28, jan 2017.

[162] Islam Miligy, Priya Mohan, Ahmed Gaber, et al. Prognostic significance of tumour infiltrating B lymphocytes in breast ductal carcinoma in situ. *Histopathology*, 71(2):258–268, aug 2017.

[163] Roberto Salgado, Susan Fineberg, Alex De Caluwe, and Evandro de Azambuja. Tumour infiltrating lymphocytes and ductal carcinoma in situ: The art of thinking counterintuitively. *European Journal of Cancer*, 168:138–140, jun 2022.

[164] S. Loi, S. Michiels, S. Adams, et al. The journey of tumor-infiltrating lymphocytes as a biomarker in breast cancer: clinical utility in an era of checkpoint inhibition. *Annals of Oncology*, 32(10):1236–1244, oct 2021.

[165] Aglaia Schiza, Viktoria Thurfjell, Axel Stenmark Tullberg, et al. Tumour-infiltrating lymphocytes add prognostic information for patients with low-risk DCIS: findings from the SweDCIS randomised radiotherapy trial. *European Journal of Cancer*, 168:128–137, jun 2022.

[166] Marie Colombe Agahozo, Mieke R. van Bockstal, Floris H. Groenendijk, et al. Ductal carcinoma in situ of the breast: immune cell composition according to subtype. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 33(2):196–205, feb 2020.

[167] Sunil S. Badve, Sanghee Cho, Xiaoyu Lu, et al. Tumor Infiltrating Lymphocytes in Multi-National Cohorts of Ductal Carcinoma In Situ (DCIS) of Breast. *Cancers 2022, Vol. 14, Page 3916*, 14(16):3916, aug 2022.

[168] Tad T. Brunyé, Ezgi Mercan, Donald L. Weaver, and Joann G. Elmore. Accuracy is in the eyes of the pathologist: The visual interpretive process and diagnostic accuracy with digital whole slide images. *Journal of biomedical informatics*, 66:171–179, feb 2017.

[169] Carsten Denkert, Stephan Wienert, Audrey Poterie, et al. Standardized evaluation of tumor-infiltrating lymphocytes in breast cancer: results of the ring studies of the international immuno-oncology biomarker working group. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 29(10):1155–1164, oct 2016.

[170] Shannon K Swisher, Yun Wu, Carlos A Castaneda, et al. Interobserver agreement between pathologists assessing tumor-infiltrating lymphocytes (tils) in breast cancer using methodology proposed by the international tils working group. *Annals of surgical oncology*, 23(7):2242–2248, 2016.

[171] Thaer Khoury, Xuan Peng, Li Yan, et al. Tumor-infiltrating lymphocytes in breast cancer: Evaluating interobserver variability, heterogeneity, and fidelity of scoring core biopsies. *American Journal of Clinical Pathology*, 150(5):441–450, 2018.

[172] Hanne Locy, Stefaan Verhulst, Wilfried Cools, et al. Assessing Tumor-Infiltrating Lymphocytes in Breast Cancer: A Proposal for Combining Immunohistochemistry and Gene Expression Analysis to Refine Scoring. *Frontiers in Immunology*, 13:82, feb 2022.

[173] Charles L. Hitchcock. The future of telepathology for the developing world. *Archives of pathology & laboratory medicine*, 135(2):211–214, feb 2011.

[174] Jeppe Thagaard, Elisabeth Specht Stovgaard, Line Grove Vognsen, et al. Automated Quantification of sTIL Density with HE-Based Digital Image Analysis Has Prognostic Potential in Triple-Negative Breast Cancers. *Cancers*, 13(12), jun 2021.

[175] Andrew Janowczyk and Anant Madabhushi. J Pathol Inform Deep learning for digital pathology image analysis : A comprehensive tutorial with selected use cases. *Journal of pathology informatics*, 2016.

[176] Korsuk Sirinukunwattana, Shan E Ahmed Raza, Yee-Wah Tsang, et al. Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. *IEEE Transactions on Medical Imaging*, 35(5):1196–1206, may 2016.

[177] Ching Wei Wang, Sheng Chuan Huang, Yu Ching Lee, et al. Deep learning for bone marrow cell detection and classification on whole-slide images. *Medical image analysis*, 75, jan 2022.

[178] Shan E Ahmed Raza, Linda Cheung, Muhammad Shaban, et al. Micro-net: A unified model for segmentation of various objects in microscopy images. *Medical image analysis*, 52:160–173, 2019.

[179] Mahendra Khened, Avinash Kori, Haran Rajkumar, et al. A generalized deep learning framework for whole-slide image segmentation and analysis. *Scientific Reports 2021 11:1*, 11(1):1–14, jun 2021.

[180] Hyeongsub Kim, Hongjoon Yoon, Nishant Thakur, et al. Deep learning-based histopathological segmentation for whole slide images of colorectal cancer in a compressed domain. *Scientific Reports 2021 11:1*, 11(1):1–14, nov 2021.

[181] Cheng Lu, Kaustav Bera, Xiangxue Wang, et al. A prognostic model for overall survival of patients with early-stage non-small cell lung cancer: a multicentre, retrospective study. *The Lancet Digital Health*, 2(11):e594–e606, nov 2020.

[182] Sergey Klimov, Yue Xue, Arkadiusz Gertych, et al. Predicting Metastasis Risk in Pancreatic Neuroendocrine Tumors Using Deep Learning Image Analysis. *Frontiers in Oncology*, 10:3336, feb 2021.

[183] Jie Hao, Sai Chandra Kosaraju, Nelson Zange Tsaku, et al. PAGE-Net: Interpretable and integrative deep learning for survival analysis using histopathological images and genomic data. *Pacific Symposium on Biocomputing*, 25(2020):355–366, 2020.

[184] Stephan W. Jahn, Markus Plass, and Farid Moinfar. Digital Pathology: Advantages, Limitations and Emerging Perspectives. *Journal of Clinical Medicine*, 9(11):1–17, nov 2020.

[185] SOHAIB ARSHAD. Ornet, Digital Pathology – A step towards better Diagnostics and Treatment, https://ornet.eu/digital-pathology-a-step-towards-better-diagnostics-and-treatment/, Accessed on, 2022.

[186] B. Acs, M. Rantalainen, and J. Hartman. Artificial intelligence as the next step towards precision pathology. *Journal of Internal Medicine*, 288(1):62–81, jul 2020.

[187] Christof A. Bertram and Robert Klopfleisch. The Pathologist 2.0: An Update on Digital Pathology in Veterinary Medicine. *Veterinary Pathology*, 54(5):756–766, sep 2017.

[188] Liron Pantanowitz. Digital images and the future of digital pathology. *Journal of Pathology Informatics*, 1(1):15, jan 2010.

[189] Gloria Bueno, M. Milagro Fernández-Carrobles, Oscar Deniz, and Marcial García-Rojo. New Trends of Emerging Technologies in Digital Pathology. *Pathobiology*, 83(2-3):61–69, apr 2016.

[190] Navid Farahani and Liron Pantanowitz. Overview of Telepathology. *Surgical pathology clinics*, 8(2):223–231, jun 2015.

[191] Raouf E. Nakhleh. Has diagnostic (analytic) accuracy improved in anatomic pathology? Are we better today than we were 20 years ago? *Archives of pathology & laboratory medicine*, 139(6):716–718, jun 2015.

[192] Liron Pantanowitz, John H. Sinard, Walter H. Henricks, et al. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Archives of pathology & laboratory medicine*, 137(12):1710–1722, dec 2013.

[193] Pinky A. Bautista and Yukako Yagi. Digital simulation of staining in histopathology multispectral images: enhancement and linear transformation of spectral transmittance. *https://doi.org/10.1117/1.JBO.17.5.056013*, 17(5):056013, may 2012.

[194] Pinky A. Bautista and Yukako Yagi. Multispectral Enhancement towards Digital Staining. *Analytical Cellular Pathology (Amsterdam)*, 35(1):51, 2012.

[195] John K.C. Chan. The Wonderful Colors of the Hematoxylin–Eosin Stain in Diagnostic Surgical Pathology. *http://dx.doi.org/10.1177/1066896913517939*, 22(1):12–32, jan 2014.

[196] D. Wittekind. Traditional staining for routine diagnostic pathology including the role of tannic acid. 1. Value and limitations of the hematoxylin-eosin stain. *http://dx.doi.org/10.1080/10520290310001633725*, 78(5):261–270, oct 2009.

[197] Kassandra Larson, Hao H. Ho, Pushpa L. Anumolu, and T. Minsue Chen. Hematoxylin and eosin tissue stain in Mohs micrographic surgery: A review. *Dermatologic Surgery*, 37(8):1089–1099, aug 2011.

[198] E. K.W. Schulte. Standardization of biological dyes and stains: pitfalls and possibilities. *Histochemistry 1991 95:4*, 95(4):319–328, feb 1991.

[199] Larry E. Morrison, Mark R. Lefever, Heather N. Lewis, et al. Conventional histological and cytological staining with simultaneous immunohistochemistry enabled by invisible chromogens. *Laboratory Investigation 2021 102:5*, 102(5):545–553, dec 2021.

[200] Edward C. Stack, Chichung Wang, Kristin A. Roman, and Clifford C. Hoyt. Multiplexed immunohistochemistry, imaging, and quantitation: A review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. *Methods*, 70(1):46–58, nov 2014.

[201] Steve Lu, Julie E. Stein, David L. Rimm, et al. Comparison of Biomarker Modalities for Predicting Response to PD-1/PD-L1 Checkpoint Blockade: A Systematic Review and Meta-analysis. *JAMA Oncology*, 5(8):1195–1204, aug 2019.

[202] Giulia Mazzaschi, Denise Madeddu, Angela Falco, et al. Low PD-1 expression in cytotoxic CD8 þ tumor-Infiltrating lymphocytes confers an immune-privileged tissue microenvironment in NSCLC with a prognostic and predictive value. *Clinical Cancer Research*, 24(2):407–419, jan 2018.

[203] Daniela S. Thommen, Viktor H. Koelzer, Petra Herzig, et al. A transcriptionally and functionally distinct PD-1+ CD8+ T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nature Medicine 2018 24:7*, 24(7):994–1004, jun 2018.

[204] Paul C. Tumeh, Christina L. Harview, Jennifer H. Yearley, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*, 515(7528):568–571, nov 2014.

[205] Pierre Hainaut and Amelie Plymoth. Targeting the hallmarks of cancer: Towards a rational approach to next-generation cancer therapy. *Current Opinion in Oncology*, 25(1):50–51, jan 2013.

[206] Marion Le Rochais, Patrice Hemon, Jacques Olivier Pers, and Arnaud Uguen. Application of High-Throughput Imaging Mass Cytometry Hyperion in Cancer Research. *Frontiers in Immunology*, 13, mar 2022.

[207] H. Raza Ali, Hartland W. Jackson, Vito R.T. Zanotelli, et al. Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. *Nature Cancer 2020 1:2*, 1(2):163–175, feb 2020.

[208] Peter Bankhead, Maurice B. Loughrey, José A. Fernández, et al. QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, 7(1), dec 2017.

[209] Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell Detection with Star-convex Polygons. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11071 LNCS:265–273, jun 2018.

[210] Yongjun Xu, Xin Liu, Xin Cao, et al. Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, 2(4):100179, nov 2021.

[211] Steve J. Bickley, Ho Fai Chan, and Benno Torgler. Artificial intelligence in the field of economics. *Scientometrics*, 127(4):2055–2084, apr 2022.

[212] David Elliott and Eldon Soifer. AI Technologies, Privacy, and Security. *Frontiers in Artificial Intelligence*, 5:60, apr 2022.

[213] Iqbal H. Sarker. AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science 2022 3:2*, 3(2):1–20, feb 2022.

[214] R.E. Uhrig. Introduction to artificial neural networks. *Proceedings of IECON 95 - 21st Annual Conference on IEEE Industrial Electronics*, 1:33–37, 1995.

[215] Qiming Zhang, Haoyi Yu, Martina Barbiero, et al. Artificial neural networks enabled by nanophotonics. *Light: Science & Applications 2019 8:1*, 8(1):1–14, may 2019.

[216] Bin Ding, Huimin Qian, and Jun Zhou. Activation functions and their characteristics in deep neural networks. *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*, pages 1836–1841, jul 2018.

[217] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[218] Simon Haykin. *Neural networks and learning machines, 3/E*. Pearson Education India, 2009.

[219] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2323, 1998.

[220] Andrew Janowczyk and Anant Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of pathology informatics*, 7(1), jan 2016.

[221] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data 2021 8:1*, 8(1):1–74, mar 2021.

[222] Stephanie Robertson, Hossein Azizpour, Kevin Smith, and Johan Hartman. Digital image analysis in breast pathology-from image processing techniques to artificial intelligence. *Translational research : the journal of laboratory and clinical medicine*, 194:19–35, apr 2018.

[223] Kelly Elliott, Stephen McQuaid, Manuel Salto-Tellez, and Perry Maxwell. Immunohistochemistry should undergo robust validation equivalent to that of molecular diagnostics. *Journal of clinical pathology*, 68(10):766–770, oct 2015.

[224] Hye Yoon Chang, Chan Kwon Jung, Junwoo Isaac Woo, et al. Artificial Intelligence in Pathology. *Journal of pathology and translational medicine*, 53(1):1–12, 2019.

[225] Esther Abels, Liron Pantanowitz, Famke Aeffner, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *The Journal of Pathology*, 249(3):286, nov 2019.

[226] Guang Yang, Cora Sau, Wan Lai, et al. Efficient and Robust Cell Detection: A Structured Regression Approach. *Medical image analysis*, 344(6188):1173–1178, 2018.

[227] Weidi Xie, J Alison Noble, and Andrew Zisserman. Microscopy Cell Counting with Fully Convolutional Regression Networks. *MICCAI 1st Workshop on Deep Learning in Medical Image Analysis*, 2015.

[228] Joseph Paul Cohen, Geneviève Boucher, Craig A. Glastonbury, et al. Count-ception: Counting by Fully Convolutional Redundant Counting. *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, 2018-Janua:18–26, 2018.

[229] Yao Xue, Nilanjan Ray B, Judith Hugh, and Gilbert Bigras. Cell Counting by Regression Using Convolutional Neural Network. *Computer Vision–ECCV 2016 Workshops*, 9913:274–290, 2016.

[230] Yousef Al-Kofahi, Alla Zaltsman, Robert Graves, et al. A deep learning-based algorithm for 2-D cell segmentation in microscopy images. *BMC Bioinformatics*, 19(1):1–11, oct 2018.

[231] Andong Wang, Qi Zhang, Yang Han, et al. A novel deep learning-based 3D cell segmentation framework for future image-based disease detection. *Scientific Reports 2022 12:1*, 12(1):1–15, jan 2022.

[232] Ahmed Fakhry, Tao Zeng, and Shuiwang Ji. Residual Deconvolutional Networks for Brain Electron Microscopy Image Segmentation. *IEEE transactions on medical imaging*, 36(2):447–456, feb 2017.

[233] Reza Moradi Rad, Parvaneh Saeedi, Jason Au, and Jon Havelock. Blastomere cell counting and centroid localization in microscopic images of human embryo. *2018 IEEE 20th International Workshop on Multimedia Signal Processing, MMSP 2018*, pages 1–6, 2018.

[234] Ezgi Mercan, Sachin Mehta, Jamen Bartlett, et al. Assessment of Machine Learning of Breast Pathology Structures for Automated Differentiation of Breast Cancer and High-Risk Proliferative Lesions. *JAMA network open*, 2(8), 2019.

[235] Gabriele Campanella, Matthew G. Hanna, Luke Geneslaw, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, aug 2019.

[236] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem:770–778, dec 2016.

[237] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv*, sep 2014.

[238] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem:770–778, dec 2015.

[239] Babak Ehteshami Bejnordi, Maeve Mullooly, Ruth M. Pfeiffer, et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 31(10):1502–1512, oct 2018.

[240] Kunal Nagpal, Davis Foote, Yun Liu, et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ digital medicine*, 2(1), dec 2019.

[241] Sidra Nawaz, Andreas Heindl, Konrad Koelble, and Yinyin Yuan. Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. *Modern Pathology*, 28(6):766–777, 2015.

[242] Rikiya Yamashita, Jin Long, Atif Saleem, et al. Deep learning predicts postsurgical recurrence of hepatocellular carcinoma from digital histopathologic images. *Scientific Reports 2021 11:1*, 11(1):1–14, jan 2021.

[243] Pierre Courtiol, Charles Maussion, Matahi Moarii, et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nature Medicine 2019 25:10*, 25(10):1519–1525, oct 2019.

[244] Xiangxue Wang, Andrew Janowczyk, Yu Zhou, et al. Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital H&E images. *Scientific Reports 2017 7:1*, 7(1):1–10, oct 2017.

[245] Deanna Cross and James K. Burmester. The Promise of Molecular Profiling for Cancer Identification and Treatment. *Clinical Medicine and Research*, 2(3):147, 2004.

[246] Eoghan R. Malone, Marc Oliva, Peter J.B. Sabatini, et al. Molecular profiling for precision cancer therapies. *Genome Medicine 2020 12:1*, 12(1):1–19, jan 2020.

[247] Ricardo L.B. Costa and Brian J. Czerniecki. Clinical development of immunotherapies for HER2+ breast cancer: a review of HER2-directed monoclonal antibodies and beyond. *npj Breast Cancer 2020 6:1*, 6(1):1–11, mar 2020.

[248] Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature medicine*, 24(10):1559–1567, oct 2018.

[249] Deepak Anand, Kumar Yashashwi, Neeraj Kumar, et al. Weakly supervised learning on unannotated H&E-stained slides predicts BRAF mutation in thyroid cancer with high accuracy. *The Journal of Pathology*, 255(3):232–242, nov 2021.

[250] David La Barbera, António Polónia, Kevin Roitero, et al. Detection of HER2 from Haematoxylin-Eosin Slides Through a Cascade of Deep Learning Classifiers via Multi-Instance Learning. *Journal of Imaging 2020, Vol. 6, Page 82*, 6(9):82, aug 2020.

[251] Gil Shamai, Yoav Binenbaum, Ron Slossberg, et al. Artificial Intelligence Algorithms to Assess Hormonal Status From Tissue Microarrays in Patients With Breast Cancer. *JAMA Network Open*, 2(7):e197700–e197700, jul 2019.

[252] Hongming Xu, Sunho Park, Sung Hak Lee, and Tae Hyun Hwang. Using transfer learning on whole slide images to predict tumor mutational burden in bladder cancer patients. *bioRxiv*, page 554527, feb 2019.

[253] James A. Diao, Wan Fung Chui, Jason K. Wang, et al. Dense, high-resolution mapping of cells and tissues from pathology images for the interpretable prediction of molecular phenotypes in cancer. *bioRxiv*, page 2020.08.02.233197, aug 2020.

[254] Wenqi Lu, Simon Graham, Mohsin Bilal, et al. Capturing cellular topology in multi-gigapixel pathology images. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2020-June:1049–1058, jun 2020.

[255] Paul Gamble, Ronnachai Jaroensri, Hongwu Wang, et al. Determining breast cancer biomarker status and associated morphological features using deep learning. *Communications Medicine 2021 1:1*, 1(1):1–12, jul 2021.

[256] Sami Blom, Lassi Paavolainen, Dmitrii Bychkov, et al. Systems pathology by multiplexed immunohistochemistry and whole-slide digital image analysis. *Scientific Reports*, 7(1):15580, dec 2017.

[257] Jessica Kalra and Jennifer Baker. *Multiplex Immunohistochemistry for Mapping the Tumor Microenvironment*. Springer New York, 2017.

[258] Michele Bortolomeazzi, Lucia Montorsi, Damjan Temelkovski, et al. A simpli (single-cell identification from multiplexed images) approach for spatially resolved tissue phenotypingat single-cell resolution. *bioRxiv*, 2021.

[259] Weibo Yu, Elizabeth Rao, Curtis D Chin, et al. A preliminary study of deep-learning algorithm for analyzing multiplex immunofluorescence biomarkers in body fluid cytology specimens. *Acta Cytologica*, 65(4):348–353, 2021.

[260] Yinyin Yuan. Spatial Heterogeneity in the Tumor Microenvironment. *Cold Spring Harbor Perspectives in Medicine*, 6(8):a026583, aug 2016.

[261] Shan E.Ahmed Raza, Khalid Abduljabbar, Mariam Jamal-Hanjani, et al. Deconvolving convolutional neural network for cell detection. *Proceedings - International Symposium on Biomedical Imaging*, 2019-April:891–894, apr 2019.

[262] Clifford C Hoyt. Multiplex immunofluorescence and multispectral imaging: Forming the basis of a clinical test platform for immuno-oncology. *Frontiers in Molecular Biosciences*, 8, 2021.

[263] Jia-Ren Lin, Benjamin Izar, Shu Wang, et al. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-cycif and conventional optical microscopes. *Elife*, 7, 2018.

[264] Amélie Viratham Pulsawatdi, Stephanie G Craig, Victoria Bingham, et al. A robust multiplex immunofluorescence and digital pathology workflow for the characterisation of the tumour immune microenvironment. *Molecular oncology*, 14(10):2384–2402, 2020.

[265] Neofytos Dimitriou, Ognjen Arandjelović, and Peter D Caie. Deep learning for whole slide image analysis: an overview. *Frontiers in medicine*, page 264, 2019.

[266] Yeman Brhane Hagos, Priya Lakshmi Narayanan, Ayse U Akarca, et al. Concorde-net: Cell count regularized convolutional neural network for cell detection in multiplex immunohistochemistry images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 667–675. Springer, 2019.

[267] Katherine Sanchez, Isaac Kim, Brie Chun, et al. Multiplex immunofluorescence to measure dynamic changes in tumor-infiltrating lymphocytes and pd-l1 in early-stage breast cancer. *Breast Cancer Research*, 23(1):1–15, 2021.

[268] Chung-Wein Lee, Yan J Ren, Mathieu Marella, et al. Multiplex immunofluorescence staining and image analysis assay for diffuse large b cell lymphoma. *Journal of immunological methods*, 478:112714, 2020.

[269] Dragan Maric, Jahandar Jahanipour, Xiaoyang Rebecca Li, et al. Whole-brain tissue mapping toolkit using large-scale highly multiplexed immunofluorescence imaging and deep neural networks. *Nature communications*, 12(1):1–12, 2021.

[270] Parmida Ghahremani, Yanyun Li, Arie Kaufman, et al. Deepliif: Deep learning-inferred multiplex immunofluorescence for ihc image quantification. *bioRxiv*, 2021.

[271] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, et al. Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem:2818–2826, dec 2015.

[272] Shan E.Ahmed Raza, Linda Cheung, Muhammad Shaban, et al. Micro-Net: A unified model for segmentation of various objects in microscopy images. *Medical Image Analysis*, 52:160–173, feb 2019.

[273] Tushar Mungle, Suman Tewary, Indu Arun, et al. Automated characterization and counting of Ki-67 protein for breast cancer prognosis: A quantitative immunohistochemistry approach. *Computer methods and programs in biomedicine*, 139:149–161, feb 2017.

[274] Lu Chen, Ji Bao, Qiang Huang, and Huaiqiang Sun. A robust and automated cell counting method in quantification of digital breast cancer immunohistochemistry images. *Polish journal of pathology : official journal of the Polish Society of Pathologists*, 70(3):162–173, 2019.

[275] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.

[276] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, dec 2014.

[277] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, et al. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[278] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[279] Martín Abadi, Ashish Agarwal, Paul Barham, et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[280] Dylan Bannon, Erick Moen, Morgan Schwartz, et al. DeepCell Kiosk: scaling deep learning–enabled cellular image analysis with Kubernetes. *Nature Methods 2021 18:1*, 18(1):43–45, jan 2021.

[281] Giuseppe Gritti, Chiara Pavoni, and Alessandro Rambaldi. Is there a role for minimal residual disease monitoring in follicular lymphoma in the chemo-immunotherapy era? *Mediterranean Journal of Hematology and Infectious Diseases*, 9(1), 2017.

[282] Julia González-Rincón, Miriam Méndez, Sagrario Gómez, et al. Unraveling transformation of follicular lymphoma to diffuse large B-cell lymphoma. *PLOS ONE*, 14(2):e0212813, feb 2019.

[283] Corrado Tarella, Angela Gueli, Federica Delaini, et al. Rate of primary refractory disease in B and T-Cell non-Hodgkin's Lymphoma: Correlation with long-term survival. *PLoS ONE*, 9(9), sep 2014.

[284] Stefano Luminari, Martina Manni, Sara Galimberti, et al. Response-Adapted Postinduction Strategy in Patients With Advanced-Stage Follicular Lymphoma: The FOLL12 Study. *Journal of Clinical Oncology*, oct 2021.

[285] Laura Pasqualucci, Hossein Khiabanian, Marco Fangazio, et al. Genetics of Follicular Lymphoma Transformation. *Cell Reports*, 6(1):130–140, jan 2014.

[286] Björn Engelbrekt Wahlin, Mohit Aggarwal, Santiago Montes-Moreno, et al. A unifying microenvironment model in follicular lymphoma: Outcome is predicted by programmed death-1-positive, regulatory, cytotoxic, and helper T cells and macrophages. *Clinical Cancer Research*, 16(2):637–650, jan 2010.

[287] Anna Maria Tsakiroglou, Susan Astley, Manàs Dave, et al. Immune infiltrate diversity confers a good prognosis in follicular lymphoma. *Cancer immunology, immunotherapy : CII*, 70(12):3573–3585, dec 2021.

[288] Robert Kridel, Christian Steidl, and Randy D. Gascoyne. Tumor-Associated macrophages in diffuse large b-cell lymphoma. *Haematologica*, 100(2):143–145, feb 2015.

[289] Junxin Li, Wenji Sun, Priyanka B. Subrahmanyam, et al. NKT Cell Responses to B Cell Lymphoma. *Medical sciences : open access journal*, 2(2):82, apr 2014.

[290] Klanova M, Oestergaard MZ, Trněný M, et al. Prognostic Impact of Natural Killer Cell Count in Follicular Lymphoma and Diffuse Large B-cell Lymphoma Patients Treated with Immunochemotherapy. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 25(15):4632–4643, aug 2019.

[291] Juliette Ferrant, Faustine Lhomme, Simon Le Gallou, et al. Circulating Myeloid Regulatory Cells: Promising Biomarkers in B-Cell Lymphomas. *Frontiers in Immunology*, 0:3686, feb 2021.

[292] Metcalf RA, Monabati A, Vyas M, et al. Myeloid cell nuclear differentiation antigen is expressed in a subset of marginal zone lymphomas and is useful in the differential diagnosis with follicular lymphoma. *Human pathology*, 45(8):1730–1736, 2014.

[293] Anil V. Parwani. Next generation diagnostic pathology: Use of digital pathology and artificial intelligence tools to augment a pathological diagnosis, dec 2019.

[294] Charlotte Syrykh, Arnaud Abreu, Nadia Amara, et al. Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning. *npj Digital Medicine*, 3(1):1–8, dec 2020.

[295] Hiroaki Miyoshi, Kensaku Sato, Yoshinori Kabeya, et al. Deep learning shows the capability of high-level computer-aided diagnosis in malignant lymphoma. *Laboratory Investigation*, 100(10):1300–1310, oct 2020.

[296] Maher Albitar, Hong Zhang, Andre H. Goy, et al. Determining Clinical Course of Diffuse Large B-Cell Lymphoma Using Targeted Transcriptome and Machine Learning Algorithms. *Blood*, 138(Supplement 1):2395, nov 2021.

[297] Steven H. Swerdlow, Elias Campo, Stefano A. Pileri, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*, 127(20):2375–2390, may 2016.

[298] Andreas Heindl, Ivana Sestak, Kalnisha Naidoo, et al. Relevance of Spatial Heterogeneity of Immune Infiltration for Predicting Risk of Recurrence after Endocrine Therapy of ER+ Breast Cancer. *Journal of the National Cancer Institute*, 110(2), feb 2018.

[299] Carlo C. Maley, Konrad Koelble, Rachael Natrajan, et al. An ecological measure of immune-cancer colocalization as a prognostic factor for breast cancer. *Breast Cancer Research*, 17(1):1–13, 2015.

[300] Matthew J. Zdilla, Scott A. Hatfield, Kennedy A. McLean, et al. Circularity, solidity, axes of a best fit ellipse, aspect ratio, and roundness of the foramen ovale: a morphometric analysis with neurosurgical considerations. *The Journal of craniofacial surgery*, 27(1):222, jan 2016.

[301] Yang Mingqiang, Kpalma Kidiyo, Ronsin Joseph, et al. A Survey of Shape Feature Extraction Techniques. *Pattern Recognition Techniques, Technology and Applications*, nov 2008.

[302] Zhao Wang, Hsiang Chieh Lee, Osman O. Ahsen, et al. Computer-Aided Analysis of Gland-Like Subsurface Hyposcattering Structures in Barrett's Esophagus Using Optical Coherence Tomography. *Applied Sciences 2018, Vol. 8, Page 2420*, 8(12):2420, nov 2018.

[303] Andreas Heindl, Adnan Mujahid Khan, Daniel Nava Rodrigues, et al. Microenvironmental niche divergence shapes BRCA1-dysregulated ovarian cancer morphological plasticity. *Nature Communications*, 9(1):1–14, dec 2018.

[304] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, mar 2020.

[305] Cameron Davidson-Pilon, Jonas Kalderstam, Noah Jacobson, et al. lifelines: 0.25.10, mar 2021.

[306] Harald Steck, Balaji Krishnapuram, Cary Dehing-oberije, et al. On Ranking in Survival Analysis: Bounds on the Concordance Index. *Advances in Neural Information Processing Systems*, 20, 2007.

[307] Megha Mittal, Abhiney Puri, Rajat Nangia, and Alisha Sachdeva. Follicular lymphoma transforming into anaplastic diffuse large B-cell lymphoma of oral cavity: A case report with review of literature. *Journal of Oral and Maxillofacial Pathology : JOMFP*, 19(3):379, sep 2015.

[308] Aashi Jindal, Prashant Gupta, Jayadeva, and Debarka Sengupta. Discovery of rare cells from voluminous single cell expression data. *Nature Communications 2018 9:1*, 9(1):1–9, nov 2018.

[309] Lan Jiang, Huidong Chen, Luca Pinello, and Guo Cheng Yuan. GiniClust: Detecting rare cell types from single-cell gene expression data with Gini index. *Genome Biology*, 17(1):1–13, jul 2016.

[310] R. K. Gershon and K. Kondo. Cell interactions in the induction of tolerance: the role of thymic lymphocytes. *Immunology*, 18(5):723, may 1970.

[311] Kiniwa Y, Miyahara Y, Wang HY, et al. CD8+ Foxp3+ regulatory T cells mediate immunosuppression in prostate cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 13(23):6947–6958, dec 2007.

[312] Jianqing Hao, Helin Wang, Lai Song, et al. Infiltration of CD8+ FOXP3+ T cells, CD8+ T cells, and FOXP3+ T cells in non-small cell lung cancer microenvironment. *International journal of clinical and experimental pathology*, 13(5):880–888, 2020.

[313] Iradj Sobhani and Sabine Le Gouvello. Critical role for CD8+FoxP3+ regulatory T cells in colon immune response in humans. *Gut*, 58(6):743–744, jun 2009.

[314] Christian T. Mayer, Stefan Floess, Abdul Mannan Baru, et al. CD8+Foxp3+ T cells share developmental and phenotypic features with classical CD4+Foxp3+ regulatory T cells but lack potent suppressive activity. *European Journal of Immunology*, 41(3):716–725, 2011.

[315] Dung T. Le, Brian H. Ladle, Timothy Lee, et al. CD8+Foxp3+ tumor infiltrating lymphocytes accumulate in the context of an effective anti-tumor response. *International journal of cancer. Journal international du cancer*, 129(3):636, aug 2011.

[316] Vieyra-Lobato Mrtha R, Vela-Ojeda Jorge, Montiel-Cervantes Laura, et al. Description of CD8 + Regulatory T Lymphocytes and Their Specific Intervention in Graft-versus-Host and Infectious Diseases, Autoimmunity, and Cancer. *Journal of immunology research*, 2018, 2018.

[317] Berendsen Madeleine R, Stevens Wendy B C, van den Brand Michiel, et al. Molecular Genetics of Relapsed Diffuse Large B-Cell Lymphoma: Insight into Mechanisms of Therapy Resistance. *Cancers*, 12(12):1–26, dec 2020.

[318] Overgaard Nana H, Jung Ji-Won, Steptoe Raymond J, and Wells James W. CD4+/CD8+ double-positive T cells: more than just a developmental stage? *Journal of leukocyte biology*, 97(1):31–38, jan 2015.

[319] Juliette Desfrançois, Agnès Moreau-Aubry, Virginie Vignard, et al. Double positive CD4CD8 $\alpha\beta$ T cells: A new tumor-reactive population in human melanomas. *PLoS ONE*, 5(1), jan 2010.

[320] Rahemtullah Aliyah, Reichard Kaaren K., Preffer Frederic I., et al. A double-positive CD4+CD8+ T-cell population is commonly found in nodular lymphocyte predominant Hodgkin lymphoma. *American journal of clinical pathology*, 126(5):805–814, nov 2006.

[321] Spolski Rosanne, Li Peng, and Leonard Warren J. Biology and regulation of IL-2: from molecular mechanisms to human therapy. *Nature reviews. Immunology*, 18(10):648–659, oct 2018.

[322] Interleukin-2 (IL2) for Metastatic Melanoma - Melanoma Research Alliance, 2021.

[323] Franz O. Smith, Stephanie G. Downey, Jacob A. Klapper, et al. Treatment of Metastatic Melanoma Using Interleukin-2 Alone or in Conjunction with Vaccines. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 14(17):5610, sep 2008.

[324] Diwakar Davar, Fei Ding, Melissa Saul, et al. High-dose interleukin-2 (HD IL-2) for advanced melanoma: A single center experience from the University of Pittsburgh Cancer Institute. *Journal for ImmunoTherapy of Cancer*, 5(1):1–10, sep 2017.

[325] Emil Kumar, Lucy Pickard, and Jessica Okosun. Pathogenesis of follicular lymphoma: genetics to the microenvironment to clinical translation. *British Journal of Haematology*, 194(5):810–821, sep 2021.

[326] Heather Fairfield, Carolyne Falank, Lindsey Avery, and Michaela R. Reagan. Multiple myeloma in the marrow: pathogenesis and treatments. *Annals of the New York Academy of Sciences*, 1364(1):32, 2016.

[327] Christine Galustian, Brendan Meyer, Marie Christine Labarthe, et al. The anti-cancer agents lenalidomide and pomalidomide inhibit the proliferation and function of T regulatory cells. *Cancer immunology, immunotherapy : CII*, 58(7):1033–1045, jul 2009.

[328] Chitta S. Kasyapa, Taimur Sher, and Asher A. Chanan-Khan. Multiple myeloma and immunomodulation: Regulating the regulatory cells. *Leukemia and Lymphoma*, 53(7):1253–1254, jul 2012.

[329] T. Dosani, M. Carlsten, I. Maric, and O. Landgren. The cellular immune system in myelomagenesis: NK cells and T cells in the development of myeloma [corrected] and their uses in immunotherapies. *Blood cancer journal*, 5(4):e306, 2015.

[330] Mika Casey and Kyohei Nakamura. The Cancer-Immunity Cycle in Multiple Myeloma. *ImmunoTargets and Therapy*, 10:247, jul 2021.

[331] Christina Hadjiaggelidou and Eirini Katodritou. Regulatory T-Cells and Multiple Myeloma: Implications in Tumor Immune Biology and Treatment. *Journal of clinical medicine*, 10(19), oct 2021.

[332] Gang An, Chirag Acharya, Xiaoyan Feng, et al. Osteoclasts promote immune suppressive microenvironment in multiple myeloma: Therapeutic implication. *Blood*, 128(12):1590–1603, 2016.

[333] Nouf Alrasheed, Lydia Lee, Ehsan Ghorani, et al. Marrow-Infiltrating Regulatory T Cells Correlate with the Presence of Dysfunctional CD4 + PD-1 + Cells and Inferior Survival in Patients with Newly Diagnosed Multiple Myeloma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 26(13):3443–3454, jul 2020.

[334] Yeman Brhane Hagos, Priya Lakshmi Narayanan, Ayse U. Akarca, et al. ConCORDe-net: Cell count regularized convolutional neural network for cell detection in multiplex immunohistochemistry images. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, 11764 LNCS:667–675, 2019.

[335] Oscar E. Brück, Susanna E. Lallukka-Brück, Helena R. Hohtari, et al. Machine Learning of Bone Marrow Histopathology Identifies Genetic and Clinical Determinants in Patients with MDS. *Cancer discovery*, 2(3):238–249, may 2021.

[336] Ende Zhao, Huanbin Xu, Lin Wang, et al. Bone marrow and the control of immunity. *Cellular and Molecular Immunology*, 9(1):11, jan 2012.

[337] Darragh Duffy, Hélène Perrin, Valérie Abadie, et al. Neutrophils Transport Antigen from the Dermis to the Bone Marrow, Initiating a Source of Memory CD8+ T Cells. *Immunity*, 37(5):917–929, nov 2012.

[338] Nicola Maciocia, Ashutosh Wechalekar, and Kwee Yong. Monoclonal gammopathy of undetermined significance (MGUS) and smoldering myeloma (SMM): a practical guide to management. *Hematological Oncology*, 35(4):432–439, dec 2017.

[339] Antonio Palumbo, Hervé Avet-Loiseau, Stefania Oliva, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 33(26):2863–2869, 2015.

[340] Ramraj Chandradevan, Ahmed A. Aljudi, Bradley R. Drumheller, et al. Machine-based detection and classification for bone marrow aspirate differential counts: initial development focusing on nonneoplastic cells. *Laboratory Investigation 2019 100:1*, 100(1):98–109, sep 2019.

[341] Luise Fischer, Caroline Herkner, Reni Kitte, et al. Foxp3+ Regulatory T Cells in Bone and Hematopoietic Homeostasis. *Frontiers in Endocrinology*, 10:578, sep 2019.

[342] Miriam Shapiro-Shelef, Kuo I. Lin, David Savitsky, et al. Blimp-1 is required for maintenance of long-lived plasma cells in the bone marrow. *The Journal of experimental medicine*, 202(11):1471–1476, dec 2005.

[343] A. L. Shaffer, Kuo I. Lin, Tracy C. Kuo, et al. Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity*, 17(1):51–62, 2002.

[344] Wen Tan Hu, Jun Wen Liu, Hong Chun Liu, and Ting Ting Wang. Expression of Blimp-1 mRNA in Bone Marrow Mononuclear Cells in Multiple Myeloma Patients and Its Clinical Significance. *Zhongguo shi yan xue ye xue za zhi*, 24(5):1443–1448, oct 2016.

[345] Yeman Brhane Hagos, Catherine S.Y. Lecat, Dominic Patel, et al. Cell abundance aware deep learning for cell detection on highly imbalanced pathological data. *Proceedings - International Symposium on Biomedical Imaging*, 2021-April:1438–1442, feb 2021.

[346] Radhakrishna Achanta, Appu Shaji, Kevin Smith, et al. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2281, 2012.

[347] Konstantinos Zormpas-Petridis, Rosa Noguera, Daniela Kolarevic Ivankovic, et al. SuperHistopath: A Deep Learning Pipeline for Mapping Tumor Heterogeneity on Low-Resolution Whole-Slide Digital Histopathology Images. *Frontiers in oncology*, 10, jan 2021.

[348] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770—-778, 2016.

[349] Yeman Brhane Hagos, Catherine S.Y. Lecat, Dominic Patel, et al. Cell abundance aware deep learning for cell detection on highly imbalanced pathological data. *Proceedings - International Symposium on Biomedical Imaging*, 2021-April:1438–1442, apr 2021.

[350] Zi Jun Xu, Yu Gu, Cui Zhu Wang, et al. The M2 macrophage marker CD206: a novel prognostic indicator for acute myeloid leukemia. *OncoImmunology*, 9(1), jan 2020.

[351] Stéfan Van Der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, et al. Scikit-image: Image processing in python. *PeerJ*, 2014(1), 2014.

[352] Yuanhong Li, Ming Dong, and Jing Hua. A Gaussian Mixture Model to Detect Clusters Embedded in Feature Subspace. *Communications in Information and Systems*, 7(4):337–352, jan 2007.

[353] Fabian Pedregosa FABIANPEDREGOSA, Vincent Michel, Olivier Grisel OLIVIERGRISEL, et al. Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, Pedregosa FABIANPEDREGOSA, F., Michel, V., Grisel OLIVIERGRISEL, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., . *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[354] George J. Grevera. Distance Transform Algorithms And Their Implementation And Evaluation. *Deformable Models*, pages 33–60, 2007.

[355] Stavros Tsogkas and Sven Dickinson. AMAT: Medial Axis Transform for Natural Images. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-Octob:2727–2736, mar 2017.

[356] Philip J. Clark and Francis C. Evans. Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology*, 35(4):445–453, oct 1954.

[357] Yeman B. Hagos, Ayse U. Akarca, Alan Ramsay, et al. High inter-follicular spatial co-localization of CD8+FOXP3+ with CD4+CD8+ cells predicts favorable outcome in follicular lymphoma. *Hematological oncology*, 2022.

[358] Karl Francis and Bernhard O. Palsson. Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proceedings of the National Academy of Sciences of the United States of America*, 94(23):12258–12262, nov 1997.

[359] Irene M. Ghobrial, Alexandre Detappe, Kenneth C. Anderson, and David P. Steensma. The bone-marrow niche in MDS and MGUS: implications for AML and MM. *Nature Reviews Clinical Oncology 2018 15:4*, 15(4):219–233, jan 2018.

[360] Y. Shiozawa, A. M. Havens, K. J. Pienta, and R. S. Taichman. The bone marrow niche: habitat to hematopoietic and mesenchymal stem cells, and unwitting host to molecular parasites. *Leukemia 2008 22:5*, 22(5):941–950, feb 2008.

[361] Alvaro Gomariz, Patrick M. Helbling, Stephan Isringhausen, et al. Quantitative spatial analysis of haematopoiesis-regulating stromal cells in the bone marrow microenvironment by 3D microscopy. *Nature Communications 2018 9:1*, 9(1):1–15, jun 2018.

[362] Jacob J. Robino, Nathalie Pamir, Sara Rosario, et al. Spatial and biochemical interactions between bone marrow adipose tissue and hematopoietic stem and progenitor cells in rhesus macaques. *Bone*, 133, apr 2020.

[363] Alessandro Allegra, Alessandro Tonacci, Raffaele Sciaccotta, et al. Machine Learning and Deep Learning Applications in Multiple Myeloma Diagnosis, Prognosis, and Treatment Selection. *Cancers*, 14(3), feb 2022.

[364] Yan Xu, Zhipeng Jia, Liang Bo Wang, et al. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinformatics*, 18(1):1–17, may 2017.

[365] Thorsten Falk, Dominic Mai, Robert Bensch, et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods 2018 16:1*, 16(1):67–70, dec 2018.

[366] Michelle A. Lawson, Michelle M. McDonald, Natasa Kovacic, et al. Osteoclasts control reactivation of dormant myeloma cells by remodelling the endosteal niche. *Nature Communications 2015 6:1*, 6(1):1–15, dec 2015.

[367] Christian M. Schürch, Leo Rasche, Leonie Frauenfeld, et al. A review on tumor heterogeneity and evolution in multiple myeloma: pathological, radiological, molecular genetics, and clinical integration. *Virchows Archiv : an international journal of pathology*, 476(3):337–351, mar 2020.

[368] Carole H Sudre, Wenqi Li, Tom Vercauteren, et al. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10553 LNCS:240–248, jul 2017.

[369] Yawara Kawano, Oksana Zavidij, Jihye Park, et al. Blocking IFNAR1 inhibits multiple myeloma-driven Treg expansion and immunosuppression. *The Journal of clinical investigation*, 128(6):2487–2499, jun 2018.

[370] Oksana Zavidij, Nicholas J. Haradhvala, Tarek H. Mouhieddine, et al. Single-cell RNA sequencing reveals compromised immune microenvironment in precursor stages of multiple myeloma. *Nature Cancer 2020 1:5*, 1(5):493–506, apr 2020.

[371] Karthick Raja Muthu Raja, Lucie Rihova, Lenka Zahradova, et al. Increased T regulatory cells are associated with adverse clinical features and predict progression in multiple myeloma. *PloS one*, 7(10), oct 2012.

[372] Andreas Heindl, Ivana Sestak, Kalnisha Naidoo, et al. Relevance of Spatial Heterogeneity of Immune Infiltration for Predicting Risk of Recurrence After Endocrine Therapy of ER+ Breast Cancer. *Journal of the National Cancer Institute*, 110(2):166–175, feb 2018.

[373] Slavica Vuckovic, Christian E. Bryant, Ka Hei Aleks Lau, et al. Inverse relationship between oligoclonal expanded CD69 TTE and CD69+ TTE cells in bone marrow of multiple myeloma patients. *Blood Advances*, 4(19):4593–4604, oct 2020.

[374] Bénedith Oben, Guy Froyen, Kylee H. Maclachlan, et al. Whole-genome sequencing reveals progressive versus stable myeloma precursor conditions as two distinct entities. *Nature Communications 2021 12:1*, 12(1):1–11, mar 2021.

[375] Sunil S. Badve and Yesim Gökmen-Polar. Ductal carcinoma in situ of breast: update 2019. *Pathology*, 51(6):563, oct 2019.

[376] Harold J Burstein, Kornelia Polyak, Julia S Wong, et al. Ductal carcinoma in situ of the breast. *New England Journal of Medicine*, 350(14):1430–1441, 2004.

[377] Anosheh Afghahi, Erna Forgó, Aya A. Mitani, et al. Chromosomal copy number alterations for associations of ductal carcinoma in situ with invasive breast cancer. *Breast Cancer Research*, 17(1):1–13, aug 2015.

[378] Anna K. Casasent, Aislyn Schalck, Ruli Gao, et al. Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing. *Cell*, 172(1-2):205–217.e12, jan 2018.

[379] Angelo Fortunato, Diego Mallo, Shawn M. Rupp, et al. A new method to accurately identify single nucleotide variants using small FFPE breast samples. *Briefings in Bioinformatics*, 22(6):1–12, nov 2021.

[380] Nobuyuki Otsu. THRESHOLD SELECTION METHOD FROM GRAY-LEVEL HISTOGRAMS. *IEEE Trans Syst Man Cybern*, SMC-9(1):62–66, 1979.

[381] Quoc Dang Vu, Simon Graham, Tahsin Kurc, et al. Methods for segmentation and classification of digital microscopy tissue images. *Frontiers in Bioengineering and Biotechnology*, 7(APR):53, 2019.

[382] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5), 2001.

[383] Santanu Roy, Shubhajit Panda, and Mahesh Jangid. Modified Reinhard Algorithm for Color Normalization of Colorectal Cancer Histopathology Images. *European Signal Processing Conference*, 2021-Augus:1231–1235, 2021.

[384] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, et al. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018.

[385] Onur Can Koyun and Tülay Yildirim. Adversarial nuclei segmentation on h&e stained histopathology images. In *2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–5. IEEE, 2019.

[386] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[387] Qifeng Chen and Vladlen Koltun. Photographic image synthesis with cascaded refinement networks. In *Proceedings of the IEEE international conference on computer vision*, pages 1511–1520, 2017.

[388] Faranak Sobhani, Sathya Muralidhar, Azam Hamidinekoo, et al. Spatial interplay of tissue hypoxia and T-cell regulation in ductal carcinoma in situ. *npj Breast Cancer 2022 8:1*, 8(1):1–11, sep 2022.

[389] Adriana Knopfelmacher, Jana Fox, Yungtai Lo, et al. Correlation of histopathologic features of ductal carcinoma in situ of the breast with the oncotype DX DCIS score. *Modern Pathology 2015 28:9*, 28(9):1167–1173, jun 2015.

[390] Po Lin Chiu, Chun Hao Chang, Yu Ling Lin, et al. Rapid and Safe Isolation of Human Peripheral Blood B and T Lymphocytes through Spiral Microfluidic Channels. *Scientific Reports 2019 9:1*, 9(1):1–10, may 2019.

[391] Dirce M. Carraro, Eliana V. Elias, and Victor P. Andrade. Ductal carcinoma in situ of the breast: morphological and molecular features implicated in progression. *Bioscience Reports*, 34(1):19–28, 2014.

[392] Fei Fei Xu, Sai Fang Zheng, Cheng Xu, et al. Prognostic and predictive significance of tumor infiltrating lymphocytes for ductal carcinoma in situ. *oncoimmunology*, 10(1), 2021.

[393] Charles L Hitchcock. The future of telepathology for the developing world. *Archives of pathology & laboratory medicine*, 135(2):211–214, 2011.

[394] Pedro Farinha, Abdulwahab Al-Tourah, Karamjit Gill, et al. The architectural pattern of FOXP3-positive T cells in follicular lymphoma is an independent predictor of survival and histologic transformation. *Blood*, 115(2):289–295, jan 2010.

[395] Philippe Solal-Céligny, Pascal Roy, Philippe Colombat, et al. Follicular lymphoma international prognostic index. *Blood*, 104(5):1258–1265, sep 2004.

[396] Philippe Solal-Céligny, Pascal Roy, Philippe Colombat, et al. Follicular lymphoma international prognostic index. *Blood*, 104(5):1258–1265, sep 2004.

[397] Michael S Toss, Islam Miligy, Abdubaqi Al-Kawaz, et al. Prognostic significance of tumor-infiltrating lymphocytes in ductal carcinoma in situ of the breast. *Modern Pathology*, 31(8):1226–1236, 2018.

[398] Fei-Fei Xu, Sai-Fang Zheng, Cheng Xu, et al. Prognostic and predictive significance of tumor infiltrating lymphocytes for ductal carcinoma in situ. *Oncoimmunology*, 10(1):1875637, 2021.

[399] Giancarlo Pruneri, Matteo Lazzeroni, V Bagnardi, et al. The prevalence and clinical relevance of tumor-infiltrating lymphocytes (tils) in ductal carcinoma in situ of the breast. *Annals of Oncology*, 28(2):321–328, 2017.

[400] Shona Hendry, Jia-Min B Pang, David J Byrne, et al. Relationship of the breast ductal carcinoma in situ immune microenvironment with clinicopathological and genetic features. *Clinical Cancer Research*, 23(17):5210–5217, 2017.

[401] Tian Yu. The Pathologist, Examining Future Trends in Spatial Biology, https://thepathologist.com/diagnostics/future-trends-in-spatial-biology, Accessed on September 2022, 2022.

[402] Lyla Atta and Jean Fan. Computational challenges and opportunities in spatially resolved transcriptomic data analysis. *Nature Communications 2021 12:1*, 12(1):1–5, sep 2021.

[403] Sabrina M. Lewis, Marie Liesse Asselin-Labat, Quan Nguyen, et al. Spatial omics and multiplexed imaging to explore cancer biology. *Nature Methods 2021 18:9*, 18(9):997–1012, aug 2021.

[404] Giovanni Palla, David S. Fischer, Aviv Regev, and Fabian J. Theis. Spatial components of molecular tissue biology. *Nature Biotechnology 2022 40:3*, 40(3):308–318, feb 2022.

[405] Mark Bustoros, Romanos Sklavenitis-Pistofidis, Jihye Park, et al. Genomic Profiling of Smoldering Multiple Myeloma Identifies Patients at a High Risk of Disease Progression. *Journal of Clinical Oncology*, 38(21):2380, jul 2020.

S1

Supplementary data for Chapter 2

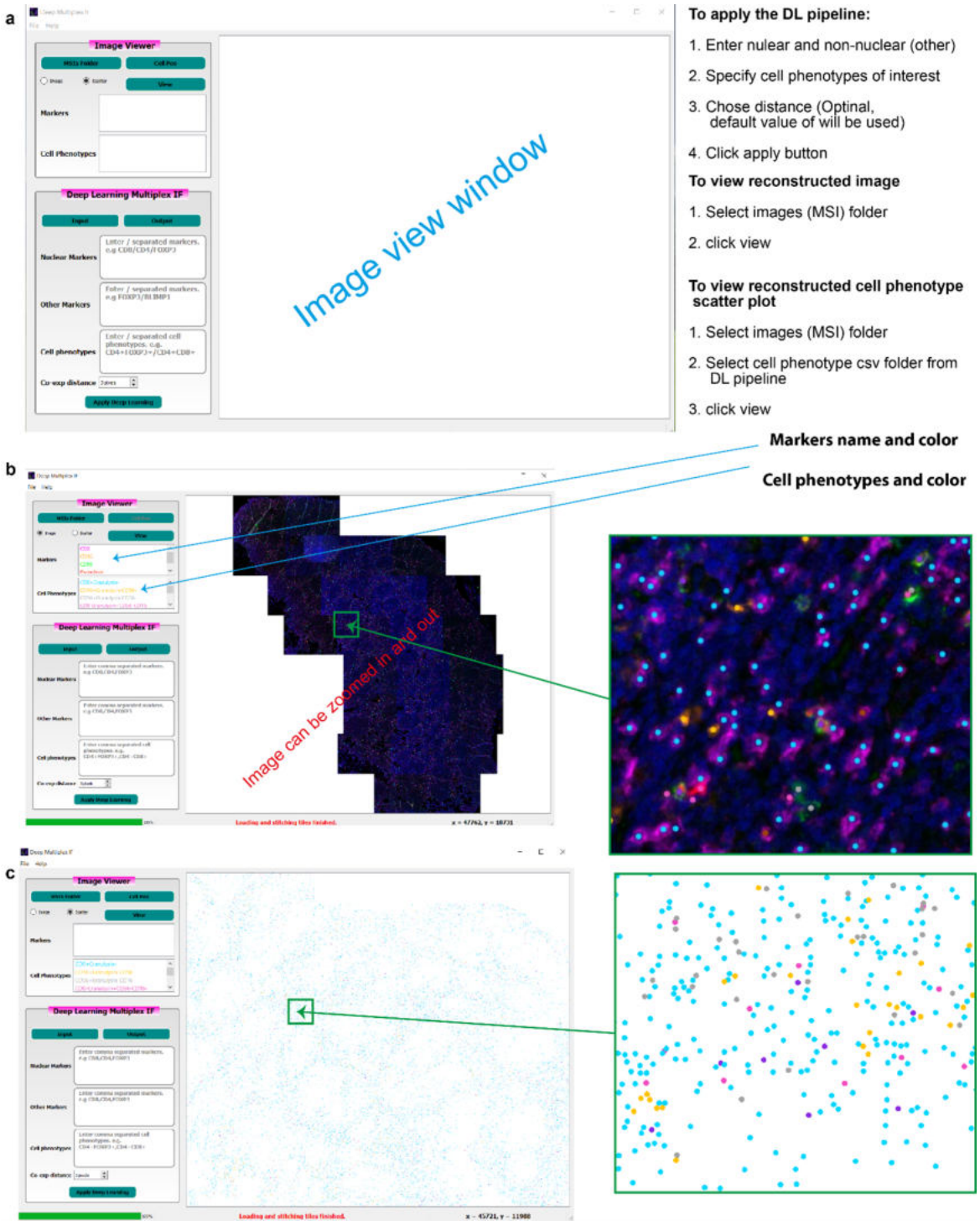


Figure S1.1: **DeepMIF Graphical user interface: A)** an Overview of the user interface. It has image viewer and deep learning-based image analysis section. **B)** sample reconstructed image. The IF images from Vectra 3 scanning system are tiles of around 4000x3000 pixels size. For visualization, the whole slide image will be reconstructed back and rendered using our image viewer. **C)** Scatter plot of cells detected by our dlearning-basedased cell phenotyping in multispectral images. The viewer is responsive to user zoom in/out, and sliding horizontal and vertical.

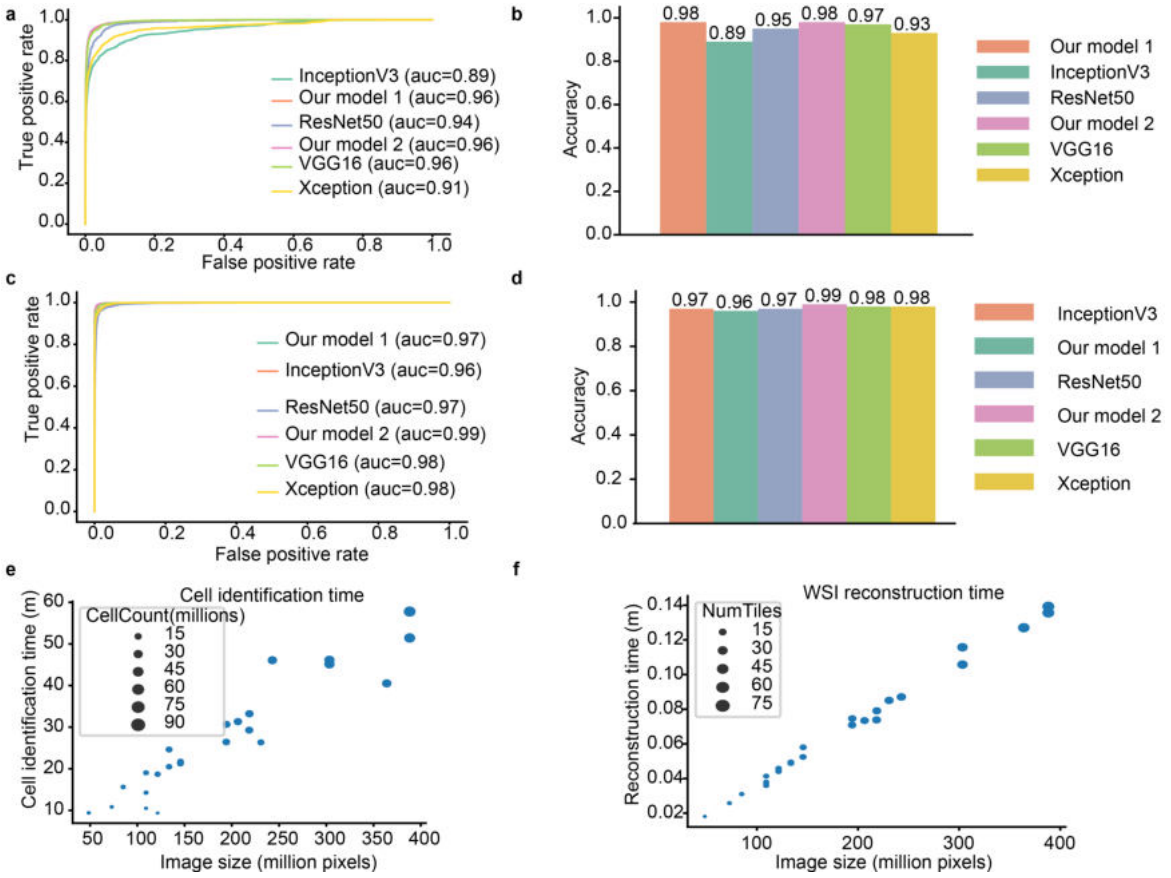


Figure S1.2: **Models and graphical user interface performance evaluation:** Cell classifiers' performance evaluation on independent data from immune T cell panel data using area under the curve (AUC) (**A**) and accuracy (**B**) and on independent data from natural killer T cell and macrophage panels using AUC (**C**) and accuracy (**D**). Speed of cell identification **E**) and image viewer reconstruction statistics (F).

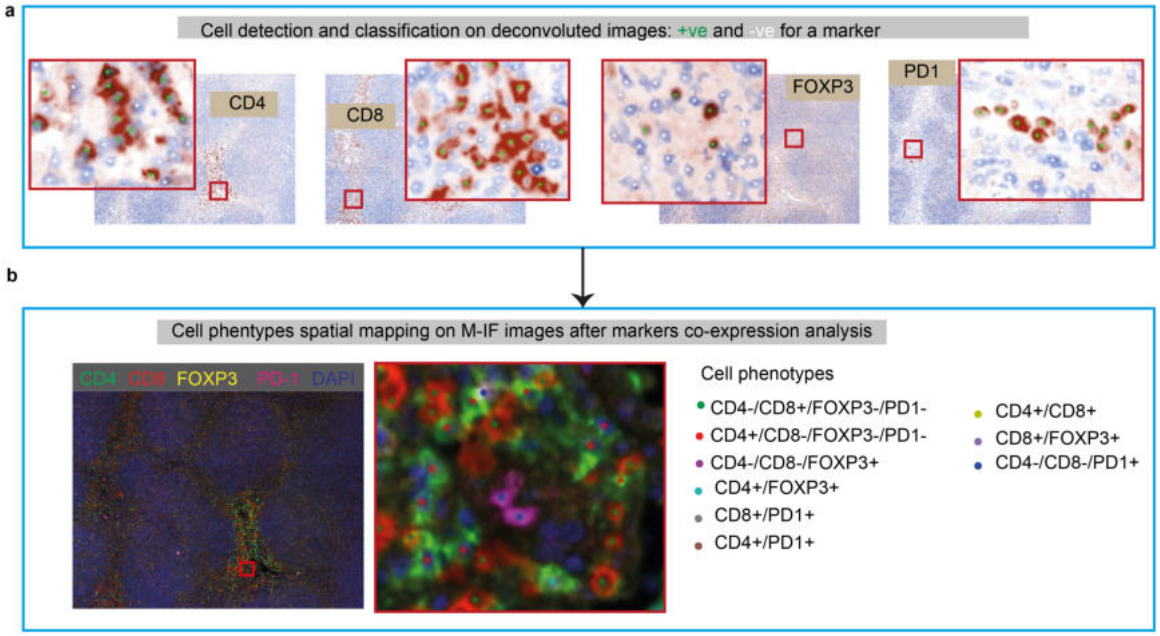


Figure S1.3: **Sample illustrative images of cell detection and classification:** Sample images showing cell detection results on deconvoluted images **A)** and on multispectral immunofluorescence image (**B**).

S2

Supplementary data for Chapter 3

Table S2.1: **Description of antibodies used in the follicular lymphoma study.** MM = Mouse Monoclonal; PR = Rabbit polyclonal; RM = Rabbit monoclonal; LM = Leica Microsystems Ltd., Newcastle-upon-Tyne, UK; AT=Agilent Technologies LDA UK Ltd. Cheshire, UK; AP = Abcam Plc. Cambridge, UK; SCB = Santa Cruz Biotechnology, Inc., Texas, U.S.A; *Dr G Roncador, CNIO, Madrid (Spain)

Molecule	**Antibody type**	**Clone name**	**Dilution**	**Source**	**Opal fluorophore**
Anti-CD4	MM	4B12	1:50	LM	Opal-620
Anti-CD8	MM	4B11	1:200	LM	Opal-650
Anti-CD11b	RM	EP1345Y	1:500	AP	Opal-520
Anti-CD14	MM	SP192	1:100	AP	Opal-570
Anti-CD15	MM	MMA	RTU	LM	Opal-650
Anti-CD16	MM	2H7	1:40	LM	Opal-570
Anti-CD56	MM	CD564	RTU	LM	Opal-520
Anti-CD68	MM	PGM1	1:100	AT	Opal-520
Anti-CD163	MM	10D6	1:200	LM	Opal-690
Anti-CD206	RP	-	1:250	AP	Opal-620
Anti-Granulysin	MM	F-9	1:300	SCB	Opal-690
Anti-Granzyme B	MM	11F1	1:80	LM	Opal-540
Anti-FOXP3	MM	236A/E3	1:2	*Gifted	Opal-570
Anti-PD1	MM	NAT 105/E3	1:350	AP	Opal-540
Anti-PDL1	RM	22C3	RTU	AT	Opal-570

Table S2.2: **Distribution of human annotation collected from deconvoluted images belonging to negative and positive cell classes.**

Deconvoluted image	Negative cells	Positive cells
Training (75%) and testing (25%) dataset from Immune T cells panel		
CD4	6 651	3 992
CD8	5 254	985
FOXP3	12 413	1 809
PD-1	7 527	1 696
Validation data from NK/T cells and macrophages panels		
CD16	425	107
CD206	594	241

Table S2.3: **Validation of MIF panels and our cell identification method using correlation between CD8+ cells density in the MIF panels.**. All correlation values were computed using a non-parametric Spearman correlation.

Panel names	Correlation (r)	P value
Myeloid cells panel vs. immune T cells panel	0.81	1.03×10^{-8}
Myeloid cells panel vs. NKT cells panel	0.72	2.53×10^{-6}
NKT cells panel vs. immune T cells panel	0.55	8.42×10^{-4}

Table S2.4: **Statistical significance p value (between relapsed and not relapsed cases) and RFS logrank p values for density of cells outside follicles**. For statistical comparisons among groups, a two-sided, nonparametric, unpaired, Wilcoxon signed-rank test was used, unless stated otherwise. To correct for multiple testing, we applied Benjamini-Hochberg (BH).

Cell Name	P value	RFS logrank p value
CD16+/Granulysin-CD56-	0.23016	0.695645
CD4-CD8-FOXP3+	0.233	0.281826
CD11B+CD14+	0.316867	0.058755
CD15+CD14+	0.316867	0.153487
CD11B+CD15+	0.316867	0.232623
CD8-CD11B+CD14-CD15-	0.360238	0.675889
CD11B-CD14-CD15+	0.469798	0.622003
CD4+FOXP3+	0.53056	0.833982
CD11B-CD14+CD15-	0.61019	0.866561
CD8+CD11B+	0.61019	0.253616
CD163+PDL1-	0.79584	0.293668
CD68-CD206+	0.79584	0.397422
CD68+CD206+	0.89485	0.405931
CD163+PDL1+	0.89485	0.684397
CD68+CD206-	0.89485	0.713578
CD163-PDL1+	0.89485	0.781852
CD4-CD8+FOXP3-	0.92479	0.456855
CD56+Granulysin+	0.92479	0.767382
CD8+Granulysin+	0.92479	0.807344
CD4+CD8+	0.92479	0.207722
CD8-Granulysin+CD56-CD16-	0.92479	0.757035
CD16+Granulysin+	0.92479	0.844875
CD16+CD56+	0.92479	0.177508
CD56+Granulysin-CD16-	0.92479	0.212977
CD4+CD8-FOXP3-	0.92479	0.952659

Table S2.5: **Statistical significance p value (between relapsed and not relapsed cases) and RFS logrank p values for density of cells inside follicles.** For statistical comparisons among groups, a two-sided, nonparametric, unpaired, Wilcoxon signed-rank test was used, unless stated otherwise. To correct for multiple testing, we applied Benjamini-Hochberg (BH).

Cell Name	P value	RFS logrank p value
CD8+FOXP3+	0.14226	0.076275
CD56+Granulysin+	0.325973	0.259931
CD16+Granulysin-CD56-	0.325973	0.550402
CD8-Granulysin+CD56-CD16-	0.325973	0.259931
CD16+CD56+	0.325973	0.310445
CD4-CD8-FOXP3+	0.326	0.084855
CD8-CD11B+CD14-CD15-	0.369798	0.766891
CD8+CD11B+	0.369798	0.763889
CD11B+CD15+	0.369798	0.763889
CD11B+CD14+	0.369798	0.550043
CD15+CD14+	0.369798	0.763889
CD11B-CD14-CD15+	0.369798	0.763889
CD8+Granulysin+	0.371392	0.310445
CD4-CD8+FOXP3-	0.39792	0.309384
CD4+CD8+	0.39792	0.529019
CD4+FOXP3+	0.425844	0.495265
CD68-CD206+	0.503355	0.944195
CD68+CD206+	0.503355	0.606061
CD163+PDL1-	0.503355	0.606061
CD68+CD206-	0.503355	0.944195
CD163-PDL1+	0.52662	0.944195
CD56+Granulysin-CD16-	0.651385	0.310445
CD163+PDL1+	0.74822	0.944195
CD16+Granulysin+	0.8061	0.550402
CD11B-CD14+CD15-	0.86506	0.766891
CD4+CD8-FOXP3-	0.89485	0.921643

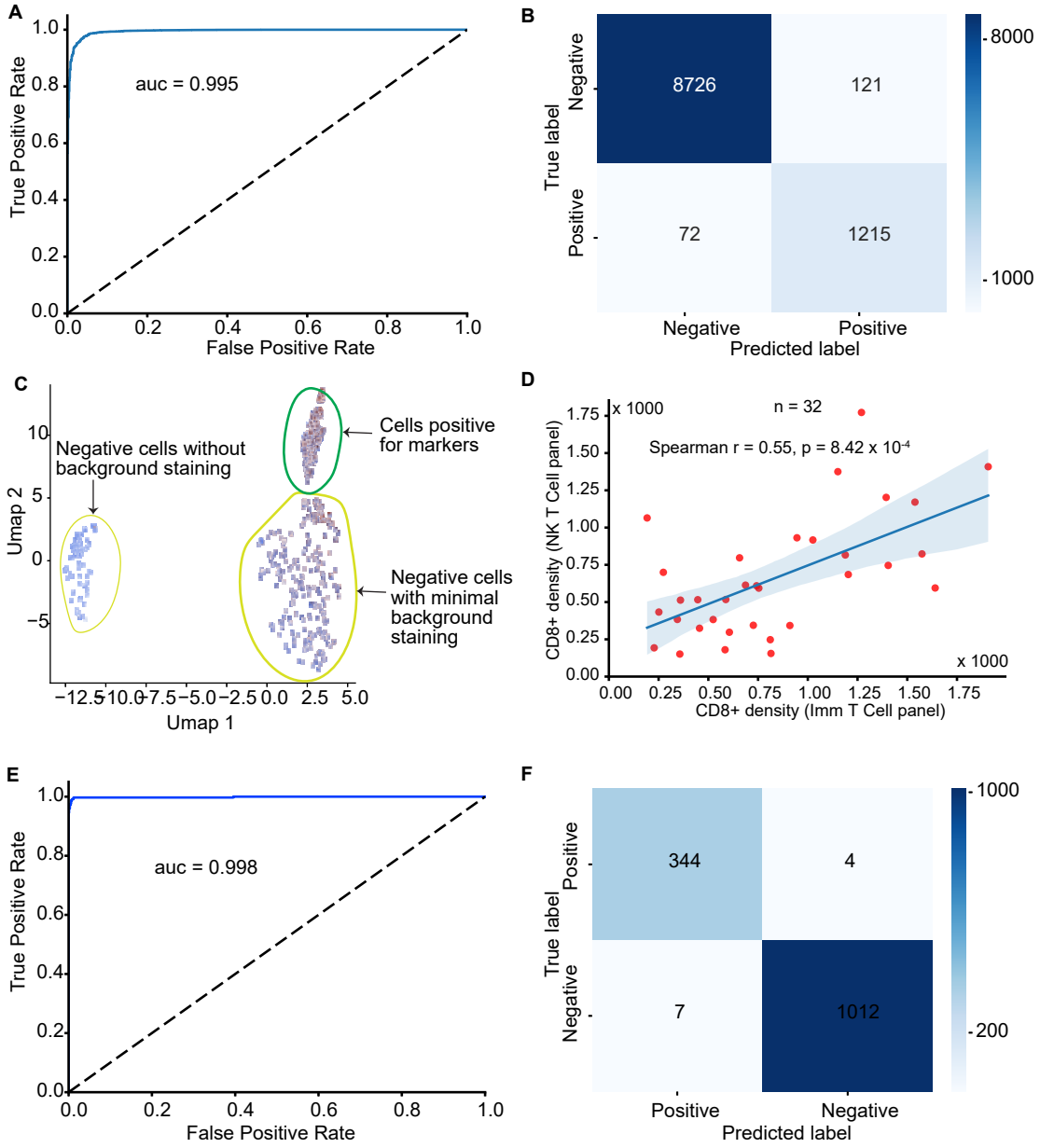


Figure S2.1: **Deep learning models performance evaluation. A,B)** Classifier performance evaluation using receiver operating characteristic curve (ROC) and area under the curve (AUC) (**A**) and confusion matrix (**B**) on a separately held test data. **C)** Image scatters plot showing two-dimensional mapping sampled testing data after UMAP dimensionality reduction of deep learning features. **D)** Spearman correlation of density of CD8+ cells. **E,F** Classifier performance evaluation using receiver ROC and AUC (**E**) and confusion matrix (**F**) on validation data collected from macrophages and NK/T cells panels.

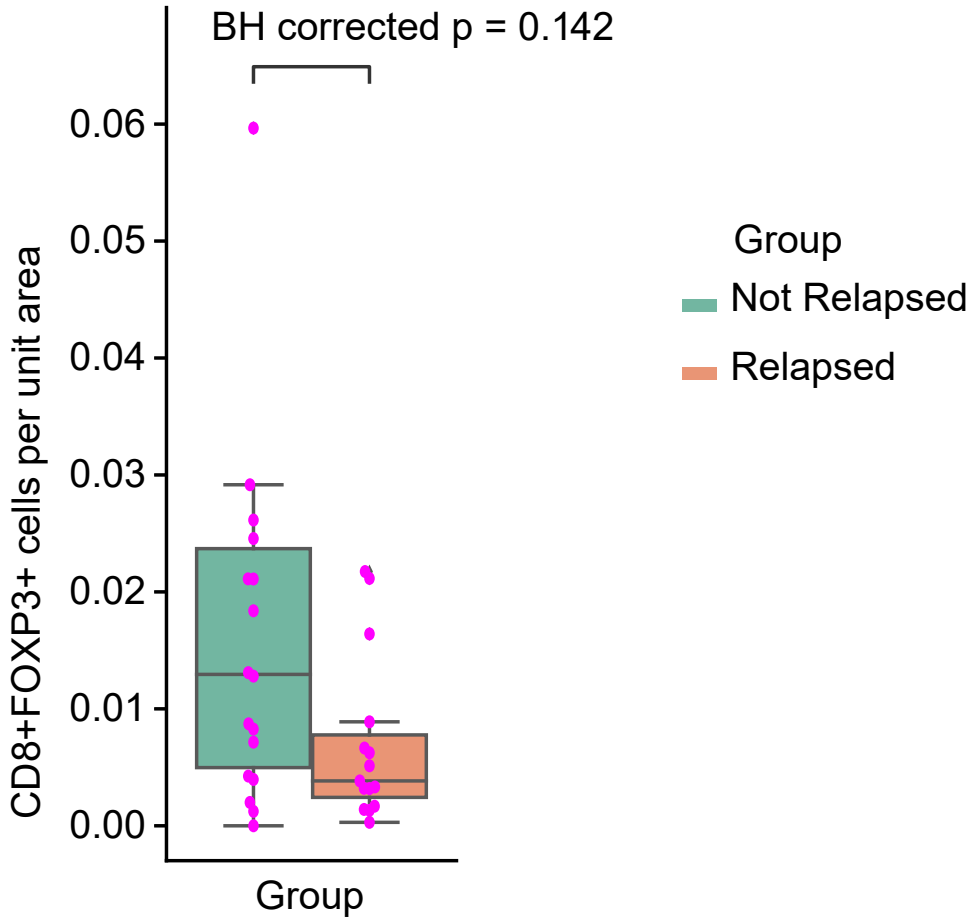


Figure S2.2: **Distribution density of CD8+FOXP3+ cells within follicles.** Cell density is measured per $1000\mu m^2$ between relapsed (n = 15) and not relapsed (n = 17) cases. For statistical comparisons among groups, a two-sided, nonparametric, unpaired, Wilcoxon signed-rank test was used. To correct for multiple testing, we applied Benjamini-Hochberg (BH). All p values displayed are after multiple testing correction.

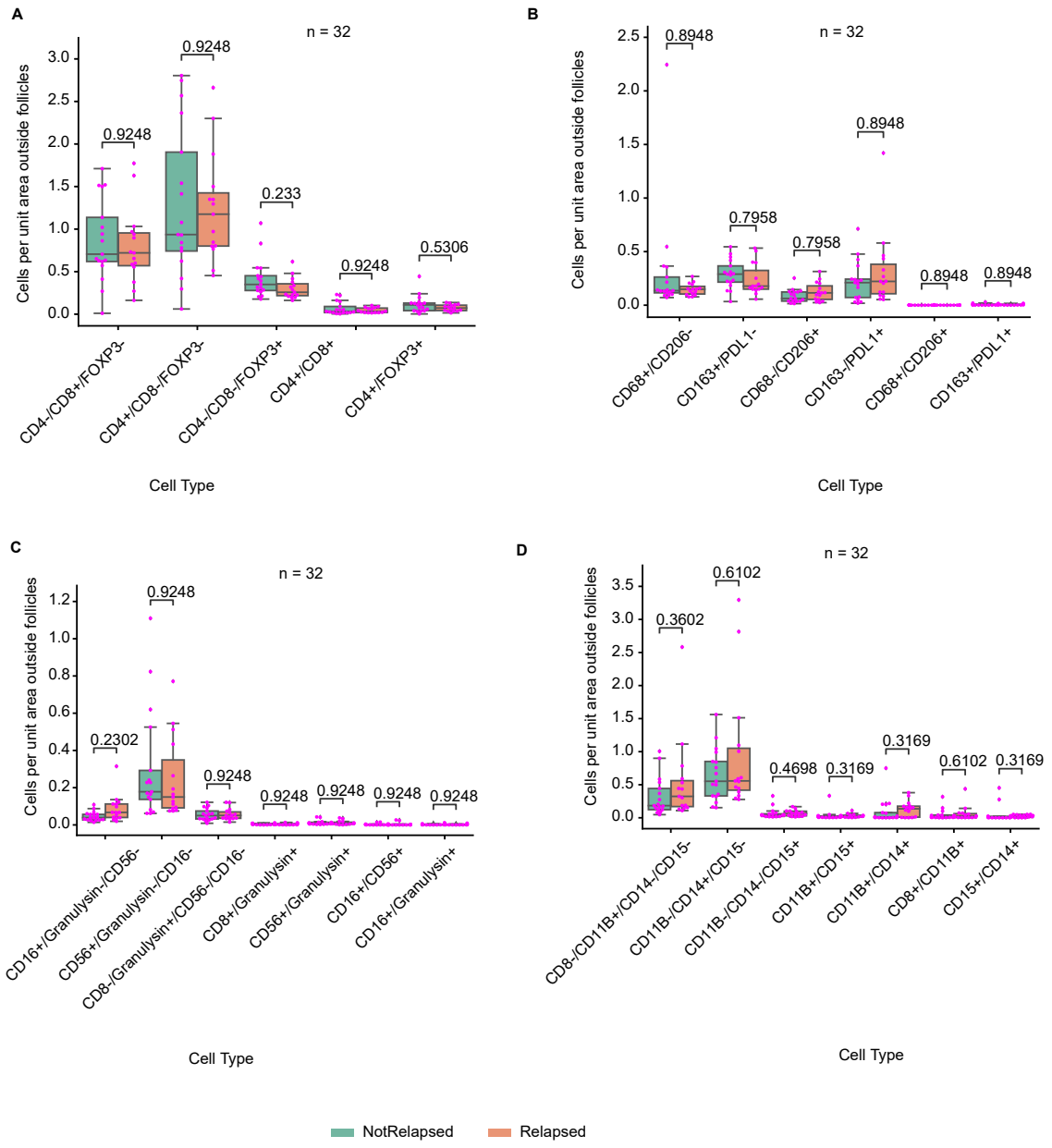


Figure S2.3: **Distribution cell phenotypes outside follicles. A-D** Boxplot showing difference in density of cells (cells/ 1000µm2) outside follicles between relapsed (n = 15) and not relapsed (n = 17) cases. For statistical comparisons among groups, a two-sided, nonparametric, unpaired, Wilcoxon signed-rank test was used. To correct for multiple testing, we applied Benjamini-Hochberg (BH). All p values displayed are after multiple testing correction.

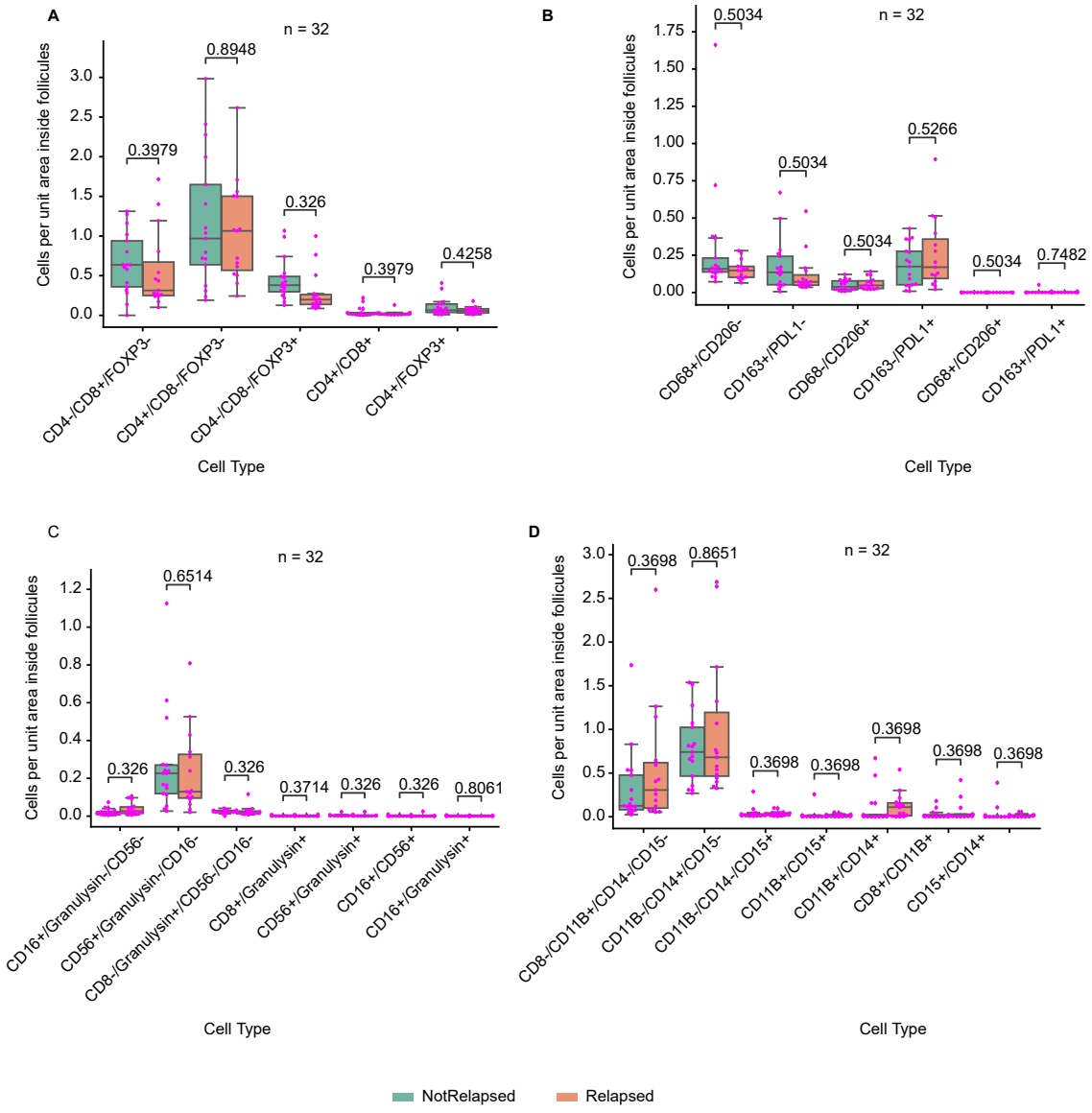


Figure S2.4: **Distribution cell phenotypes inside follicles. A-D** Boxplot showing difference in density of cells (cells/ 1000µm2) inside follicles between relapsed (n = 15) and not relapsed (n = 17) cases. For statistical comparisons among groups, a two-sided, nonparametric, unpaired, Wilcoxon signed-rank test was used. To correct for multiple testing, we applied Benjamini-Hochberg (BH). All p values displayed are after multiple testing correction.

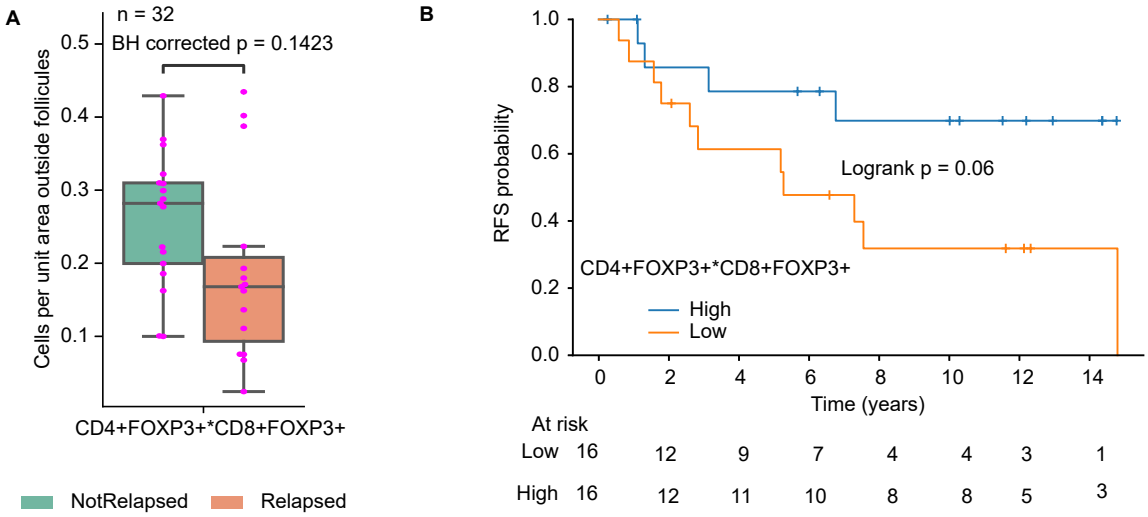


Figure S2.5: **Co-localization of CD8+FOXP3+ with CD4+FOXP3+ outside follicle. A)** Boxplot showing difference in co-localization of CD8+FOXP3+ with CD4+FOXP3+ cells outside follicels between relapsed (n = 15) and not relapsed (n = 17) cases. **B)** Kaplan-Meier curves illustrating RFS of patients dichotomized using median co-localization of CD8+FOXP3+ with CD4+FOXP3+ cells outside follicles.

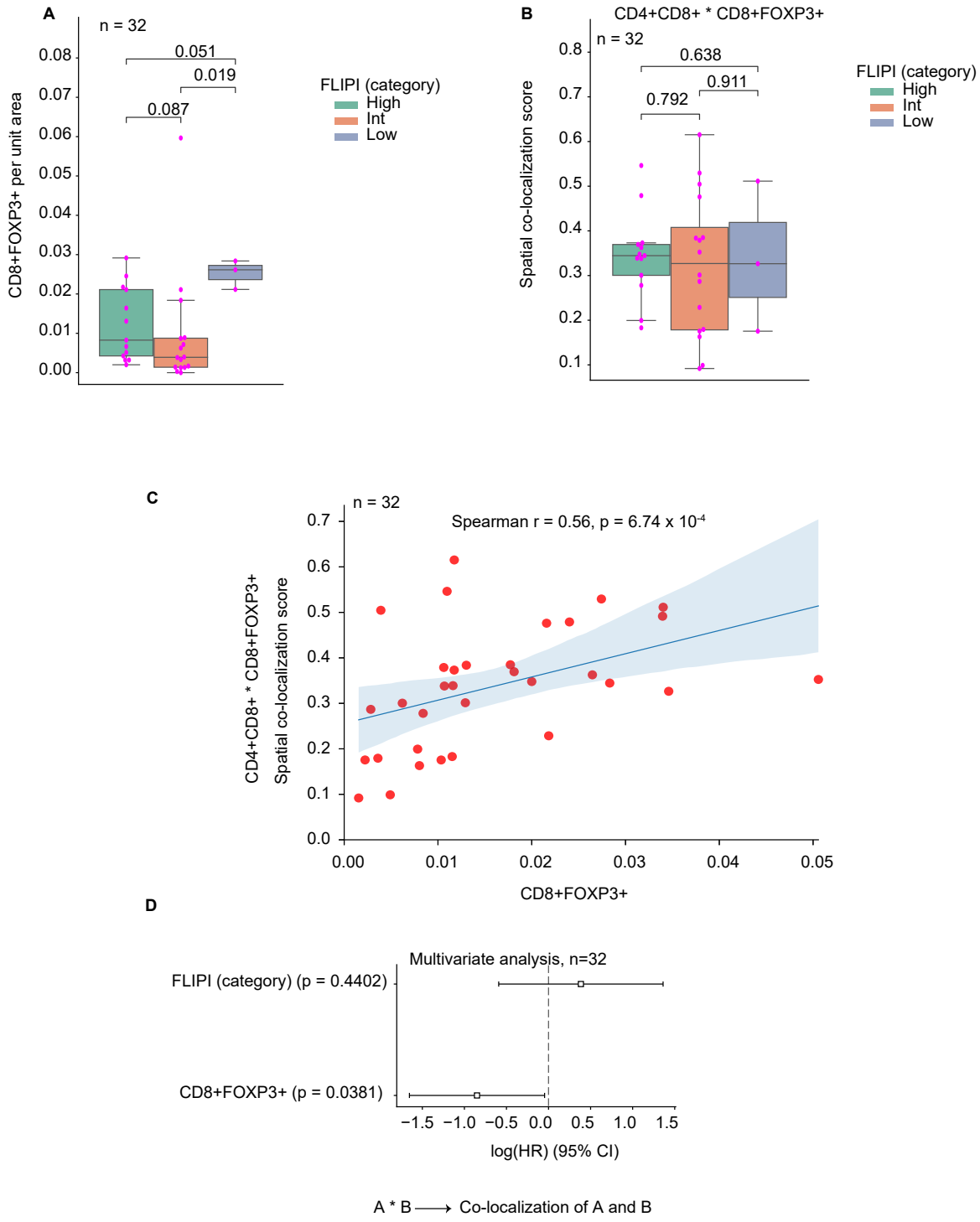


Figure S2.6: **Association of inter-follicular CD8+FOXP3+ cell density with other standard scores. A)** Boxplot showing difference interfollicular CD8+FOXP3+ cell density between the three FLIPI categories. **B)** Boxplot showing difference interfollicular spatial score between the three FLIPI categories. **C)** Correlation between spatial score and density values. For statistical comparisons among groups, a two-sided, nonparametric, unpaired, Wilcoxon signed-rank test was used, unless stated otherwise. **D)** Forest plots showing multivariate Cox regression analyses. Continuous values were used for the density parameter. Follicular lymphoma international prognostic index (FLIPI).

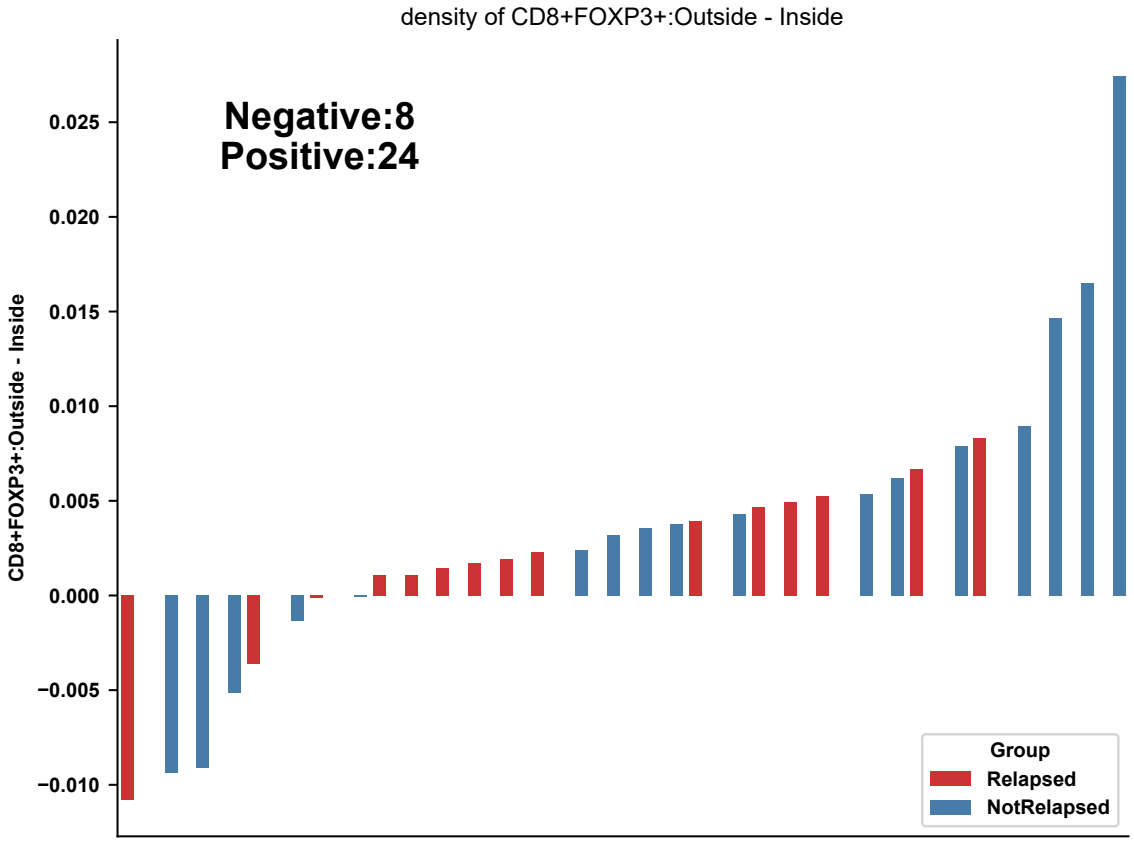


Figure S2.7: **Difference in density of CD8+FOXP3+ cells in the inter- and intra-follicular regions**. In 24 cases out of 32 cases, the density of CD8+FOXP3+ cells was higher in the inter-follicular region than in the intra-follicular region.

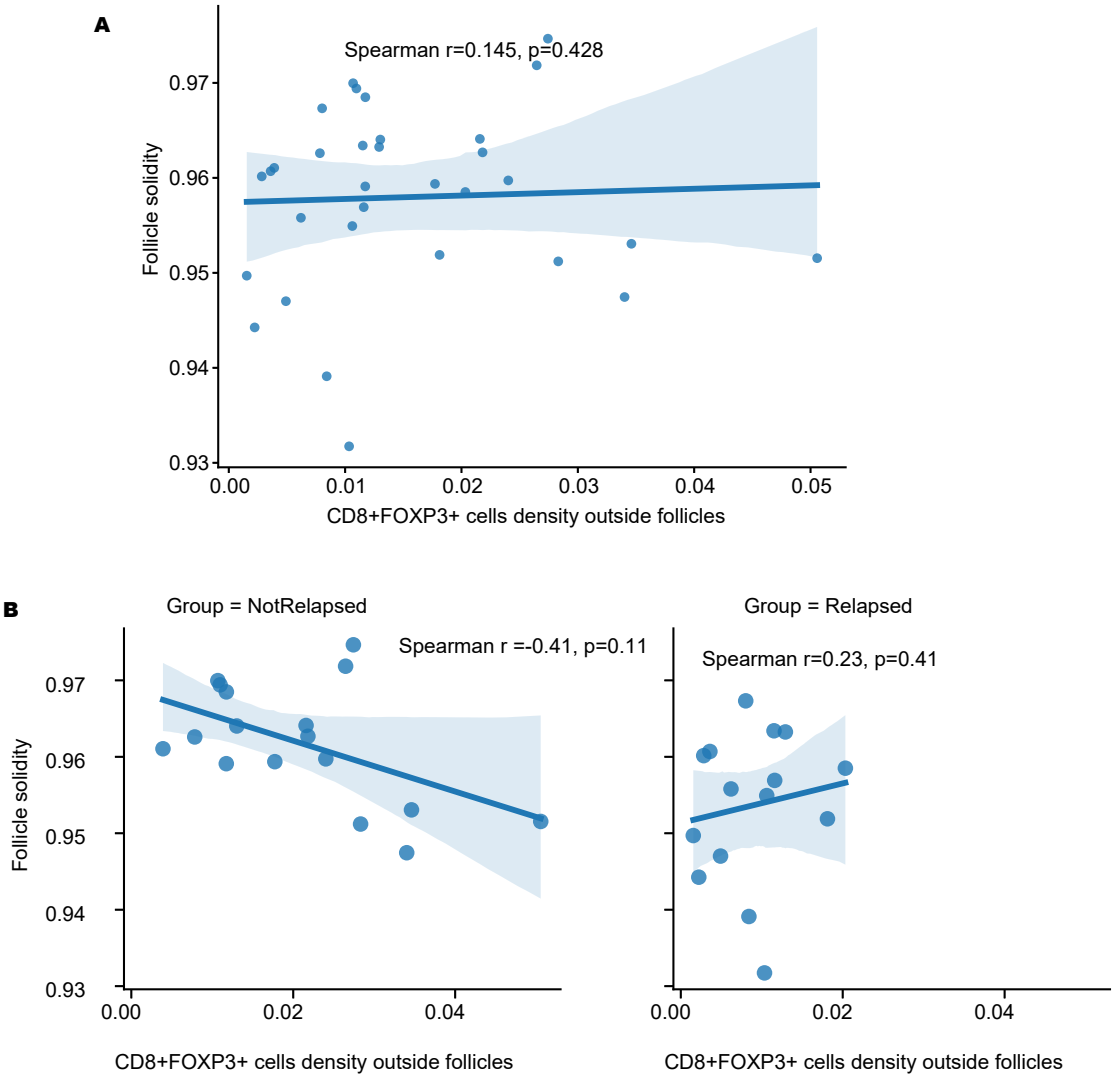


Figure S2.8: **Association of inter-follicular CD8+FOXP3+ cell density and neoplastic follicles morphology: A,B)** Correlation between CD8+FOXP3+ cell density and neoplastic follicles solidity for all patients and group-wise (B).

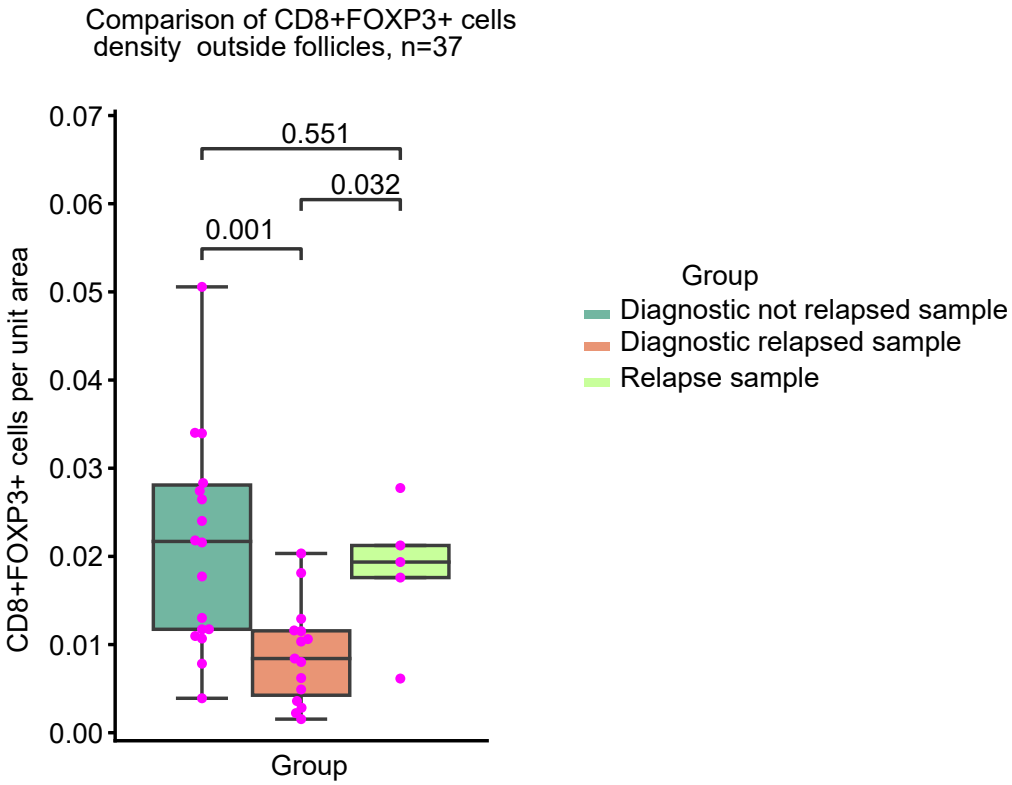


Figure S2.9: **Difference in inter-follicular CD8+FOXP3+ cell density between relapse and diagnostic samples:**

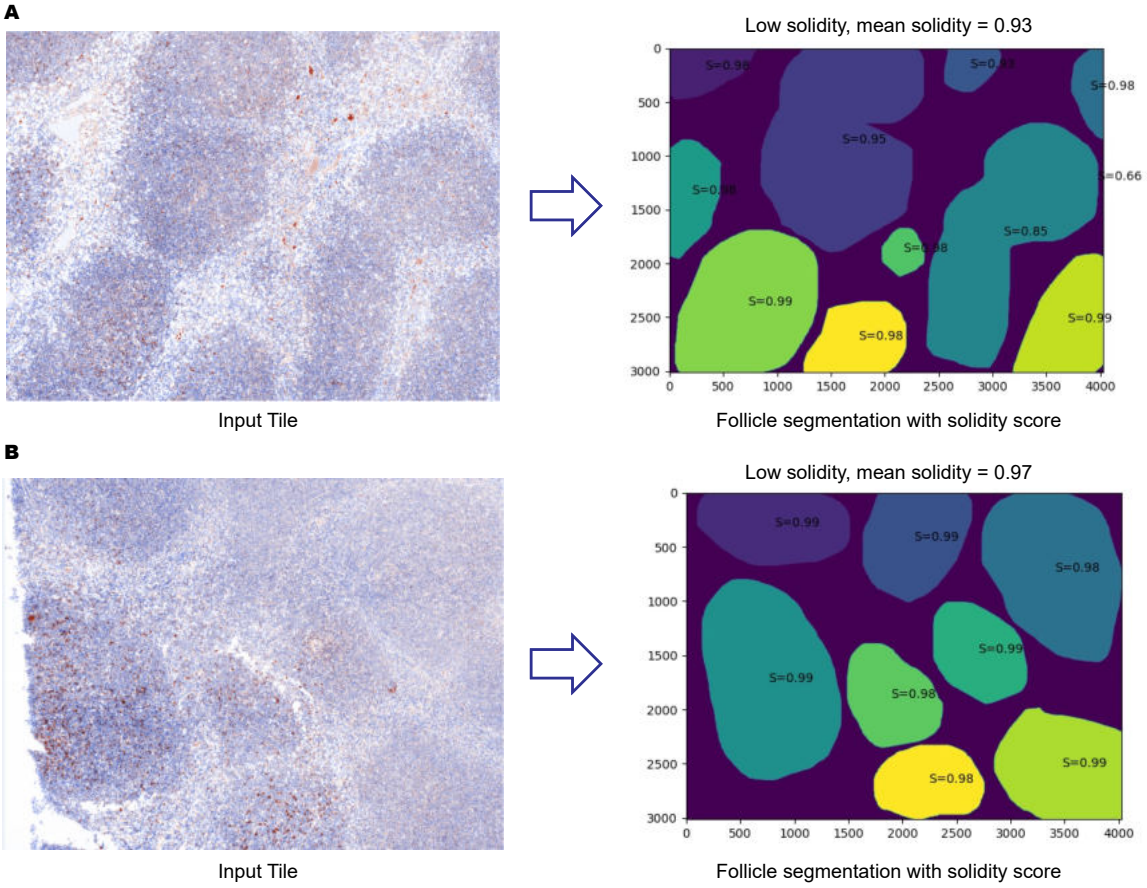


Figure S2.10: **Sample images showing tiles with high and low solidity scores of follicles**. Solidity (S) is a measure of shape irregularity. Its value ranges between 0 and 1. From 1 to 0 the level of irregularity increases. Colours represent neoplastic follicles label, it is not associated with solidity score.

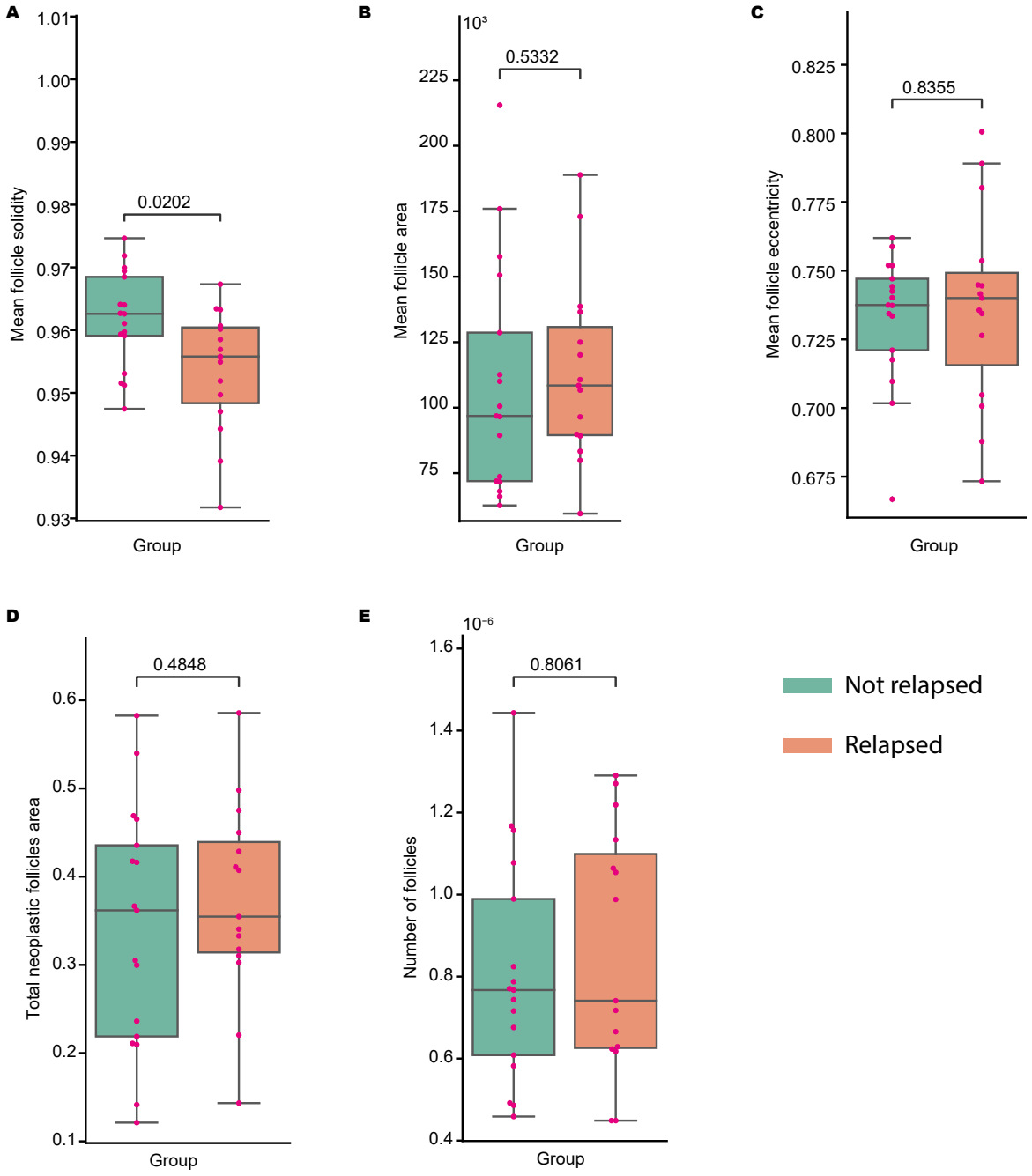


Figure S2.11: **Neoplastic follicles morphological features and follicular lymphoma prognosis. A-C)** Boxplots showing difference in neoplastic follicles solidity (A), average area (B), eccentricity (C) between relapsed (n = 15) and not relapsed (n = 17). **D-E)** Boxplots showing difference in total follicle area (D), number of follicles between relapsed (n = 15) and not relapsed (n = 17). These features were normalised by total amount of tissue in the slide. Area was measured in μm^2. For statistical comparisons among groups, a two-sided, nonparametric, unpaired, Wilcoxon signed-rank test was used.

S3

Supplementary data for Chapter 4

Table S3.1: **Antibodies used for Leica Bond slide stainer.** *RTU – ready-to-use from manufacturer

Antibody	Supplier	Species	Dilution	Control
BLIMP-1	CNIO	Mouse	1:4	Tonsil
CD4	Novocastra	Mouse	RTU*	Tonsil
CD8	Novocastra	Mouse	RTU	Tonsil
FOXP3	eBioscience	Rat	1:100	Tonsil

Table S3.2: **Staining protocols: CD4 / FOXP3 / CD8 panel.** ERP = Epitope retrieval protocol; AIT= Antibody incubation time (min); PPIT = Post primary incubation time (min); PIT = Polymer incubation time (min)

Antibody	Colour	Dilution	ERP	AIT	PPIT	PIT
CD4	Brown	RTU	ER2: 20	15	8	8
FOXP3	Blue	1:100	ER1: 10	15	8	8
CD8	Red	RTU		15	8	8

Table S3.3: **Staining protocols: CD4 / CD8 / BLIMP1 panel.** ERP = Epitope retrieval protocol; AIT= Antibody incubation time (min); PPIT = Post primary incubation time (min); PIT = Polymer incubation time (min)

Antibody	Colour	Dilution	ERP	AIT	PPIT	PIT
CD4	Brown	RTU	ER2: 20	15	8	8
BLIMP1	Blue	1:4		15	8	8
CD8	Red	RTU		15	8	8

Table S3.4: **Cell detection model Recall evaluation on detecting rare cell type, FOXP3+CD4+.**

Model	CellName	Recall
CONCORDe-Net	FOXP3+CD4+	0.5897
RatioWeight	FOXP3+CD4+	0.5983
ExpTpe2	FOXP3+CD4+	0.641
ExpType1	FOXP3+CD4+	0.6325
U-net	FOXP3+CD4+	0.547

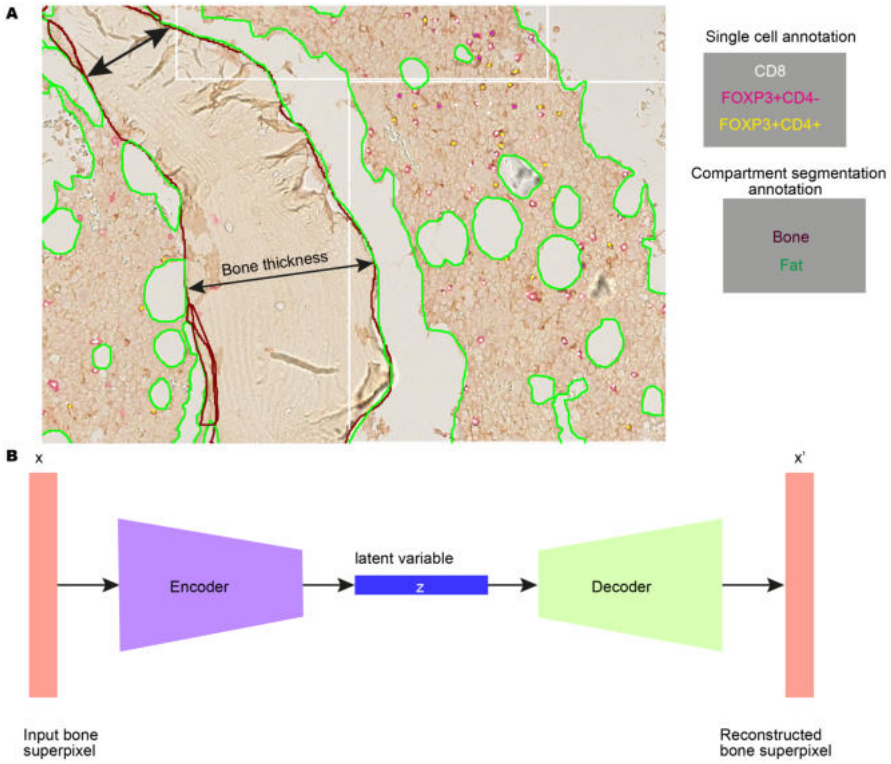


Figure S3.1: **Machine learning algorithms to understand bone physiology: A)** Sample image showing expert manual segmentation annotation used to train and validate deep learning models. **B)** Autoencoder architecture that learns low dimensional embedding of bone structure superpixels.

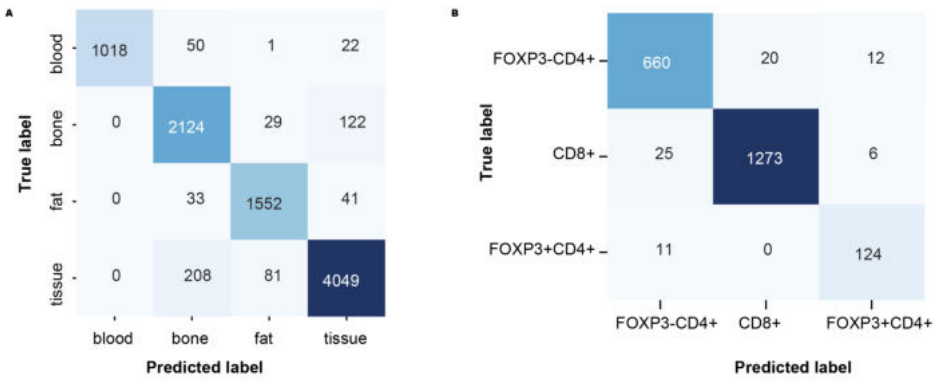


Figure S3.2: **Confusion matrix showing classification performance of MoSaicNet based compartment classification (A) and AwareNet single cell classifier (B) models.**

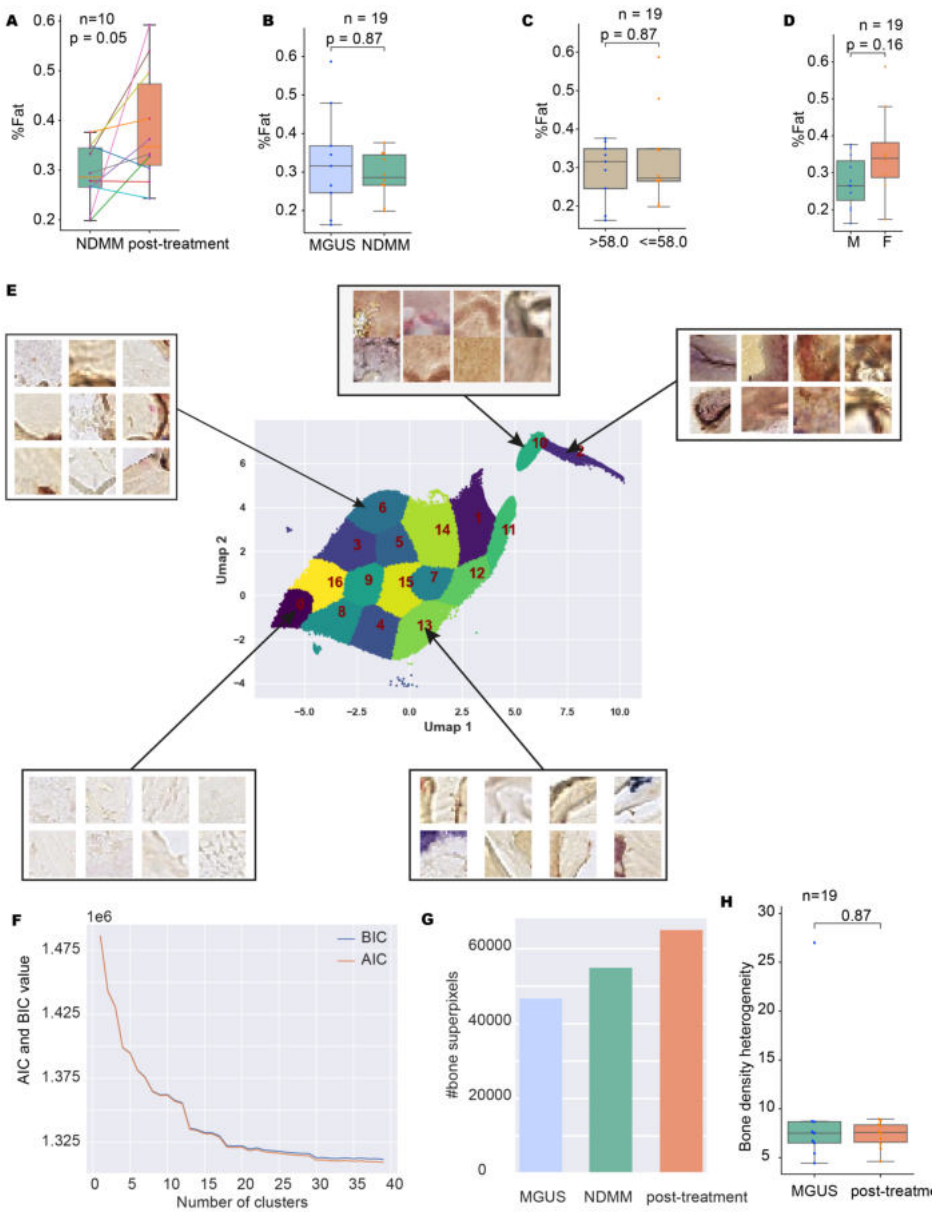


Figure S3.3: **Understanding bone physiology from BM tissue samples: A-D)** Boxplots showing the difference in %fat between NDMM and post-treatment **(A)**, between MGUS and NDMM **(B)**, and between different age groups split by median **(C)** and Gender groups **(D)**. **E)** A 2-dimensional mapping of superpixels using MoSaicNet learned 200-dimensional features after dimensionality reduction by Uniform Manifold Approximation and Projection (UMAP). To group similar bone superpixels, we applied Gaussian mixture clustering. The numbers indicate the clustered index. Sample images are displayed for some of the clusters. **F)** Finding the optimal number of clusters using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The optimal number of clusters was 19. **G)** Number of bone superpixels extracted from MGUS (9 patients), NDMM (10 patients) and post-treatment (10 patients) samples. **H)** Bone density heterogeneity difference between MGUS and post-treatment samples.

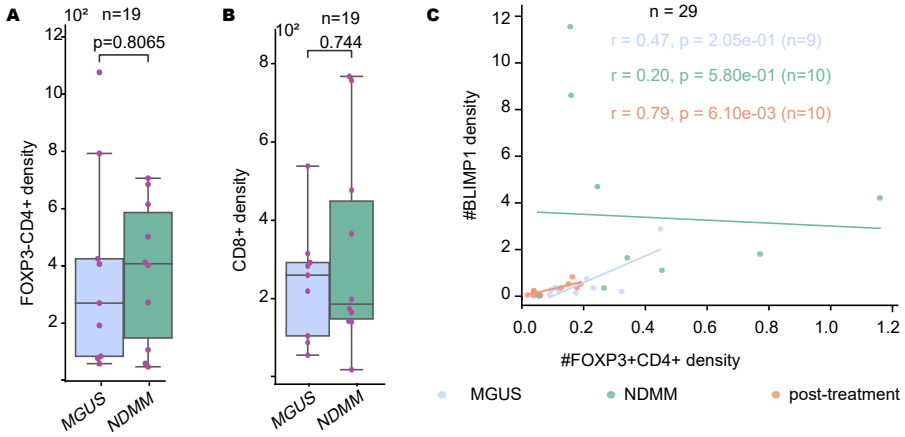


Figure S3.4: **Comparison of density of immune T cells and plasma cells in MGUS and NDMM samples: (A-B)** Boxplots showing the difference in density of FOXP3+CD4+ cells **(A)**, and BLIMP1+ cells **(B)** between MGUS samples and NDMM samples (n=19 samples). **C)** Correlation between the density of FOXP3+CD4+ and BLIMP1+ cells in the different patient groups. The cell density is presented per 1 mm^2 tissue area. Cell density is presented per 1 mm^2 tissue area.

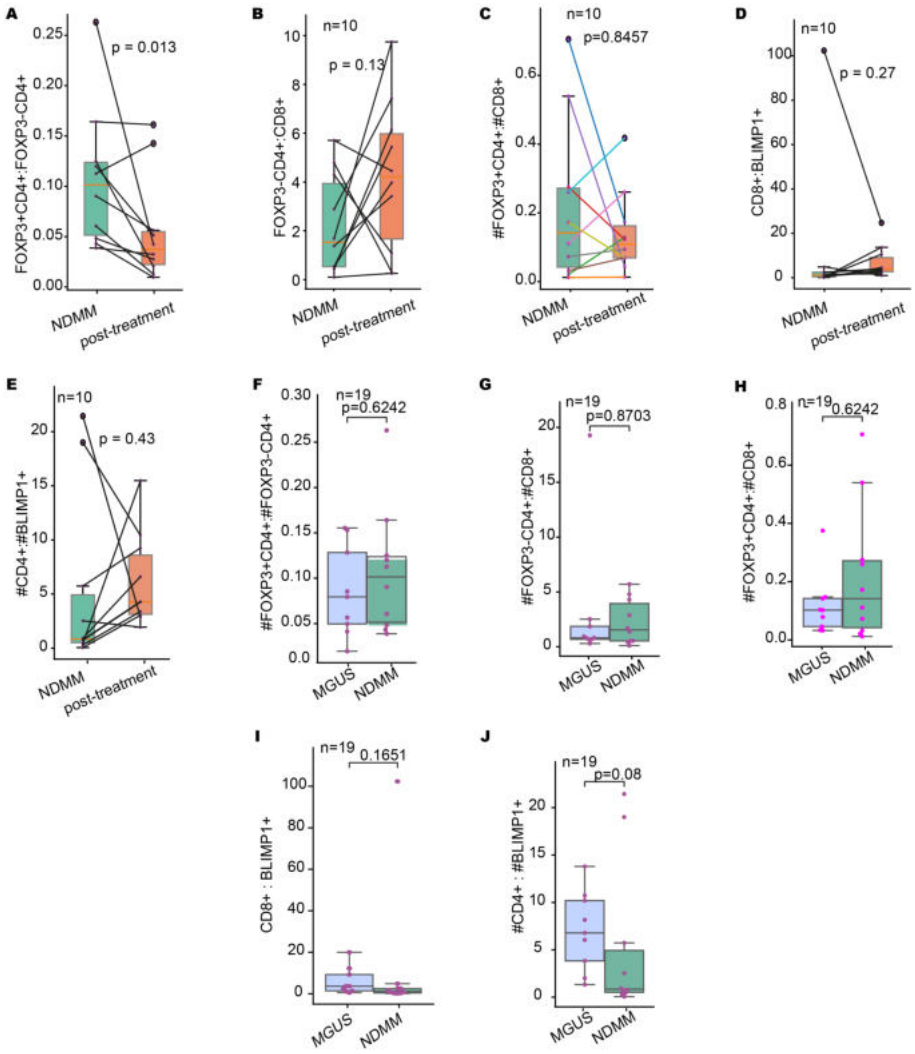


Figure S3.5: **Comparison of ratio of abundance of different cell types: A-B)** Boxplots showing the difference in FOXP3+CD4+:FOXP3-CD4+ ratio **(A)**, FOXP3-CD4+:CD8+ ratio (B), FOXP3+CD4+:CD8+ ratio (C), CD8+:BLIMP1+ ratio **(D)**, and CD4+:BLIMP1+ ratio **(E)** between paired NDMM and post-treatment samples. The outlier sample (Sample id = UH15-19506) in (D), contains high number of CD8+ cells but low density of BLIMP1+ cells. A region from this sample's tissue section is shown in Figure S3.6. **F-J)**Boxplots showing the difference in FOXP3+CD4+:FOXP3-CD4+ ratio **(F)**, FOXP3-CD4+:CD8+ ratio (G), FOXP3+CD4+:CD8+ ratio (H), CD8+:BLIMP1+ ratio **(I)**, and CD4+:BLIMP1+ ratio **(J)** between paired NDMM and MGUS samples.

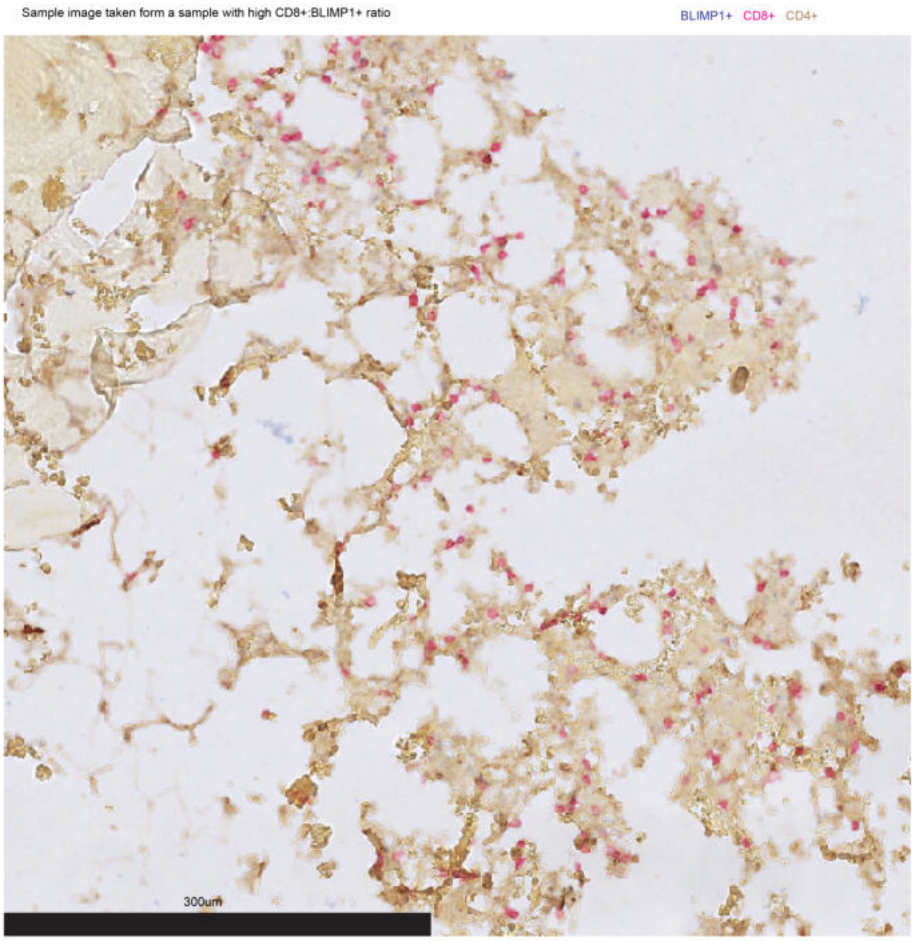


Figure S3.6: **Sample image with high number of CD8+ cells but low BLIMP1+ cells.**

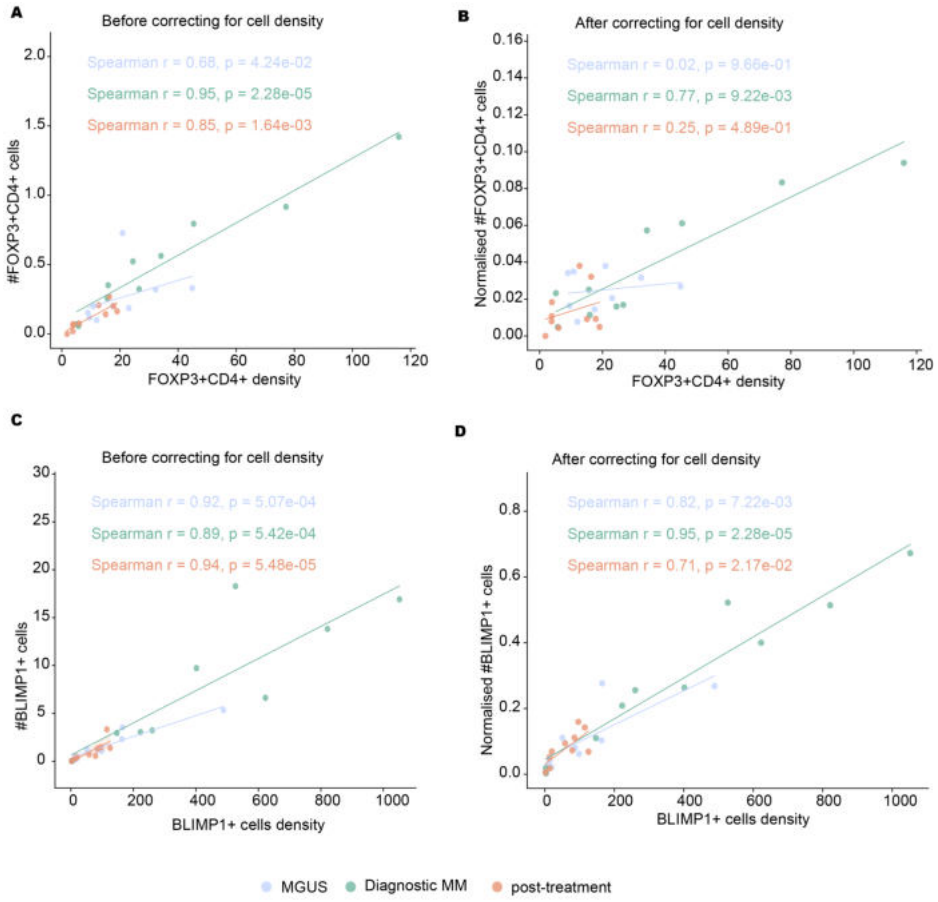


Figure S3.7: **Correlation between cell density and cells' spatial proximity scores:** Scatter plots showing the correlation between cells' spatial proximity scores before and after correcting for density. **A, B)** Correlation between the number of (#) FOXP3+CD4+ cells in proximity with CD8+ cells before (A) and after (B) correcting for FOXP3+CD4+ cell density. After correction, the correlation was reduced in all patient groups. **C, D)** Correlation between the number of (#) BLIMP1+ cells in proximity with CD8+ cells before (C) and after (D) correcting for BLIMP1+ cell density. After correction, the correlation was reduced in the MGUS and post-treatment groups, but not in the NDMM group. All cell spatial proximity scores were computed for $100\mu m$ distance.

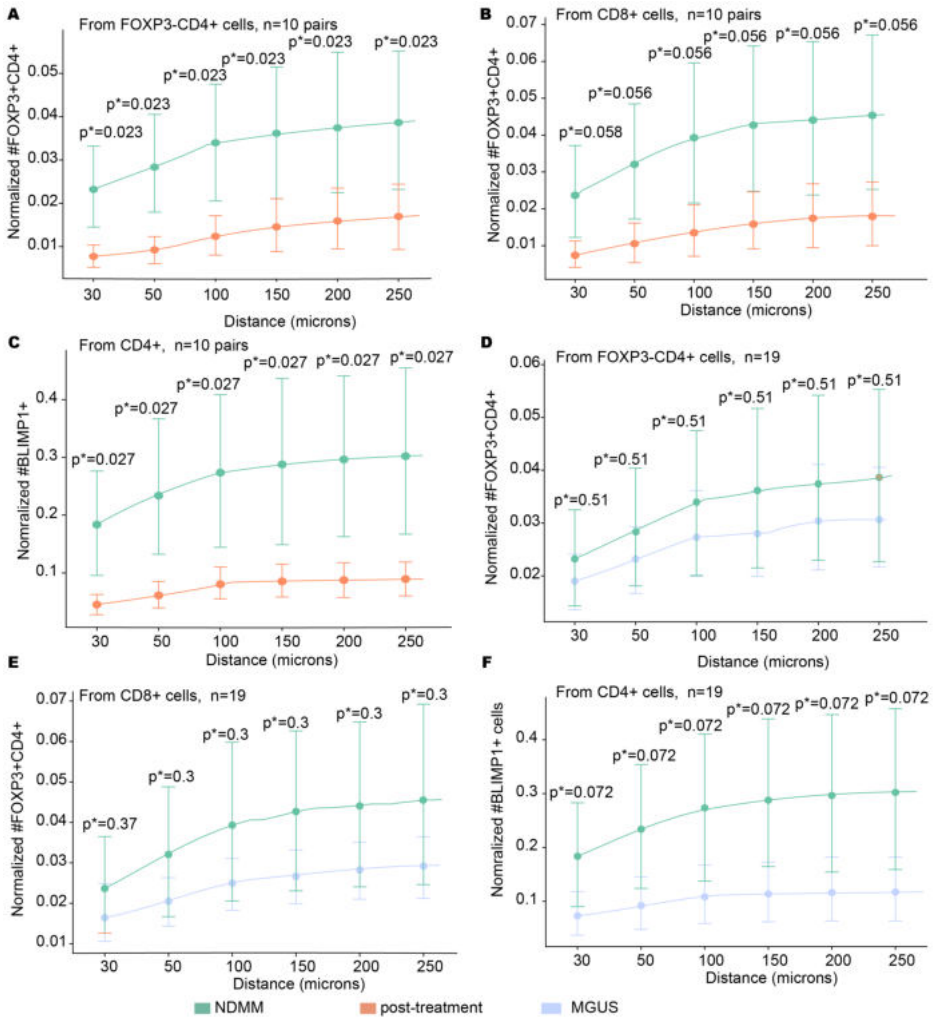


Figure S3.8: **Spatial proximity of immune cells to regulatory T cells and tumour cells: A-B)** Point plots showing the difference in the number of FOXP3+CD4+ cells within a distance (μm) from FOXP3-CD4+ cells (A) and CD8+ cells (B) between paired NDMM and post-treatment samples. **C)** Point plot showing the difference in the number of BLIMP1+ cells within a distance from CD4+ cells between MGUS and NDMM samples as a function of distance. **D-E)** Point plots showing the difference in the number of FOXP3+CD4+ cells within a distance (μm) from FOXP3-CD4+ (D) cells and CD8+ (E) cells among MGUS and NDMM samples. **F)** A point plot showing the difference in the number of BLIMP1+ cells within a distance from CD4+ cells between MGUS and NDMM samples. In the point plots, the points represent the mean and the bars are 95% confidence intervals, indicating uncertainty. The p* indicate p values after multiple testing correction using the Benjamini-Hochberg method.

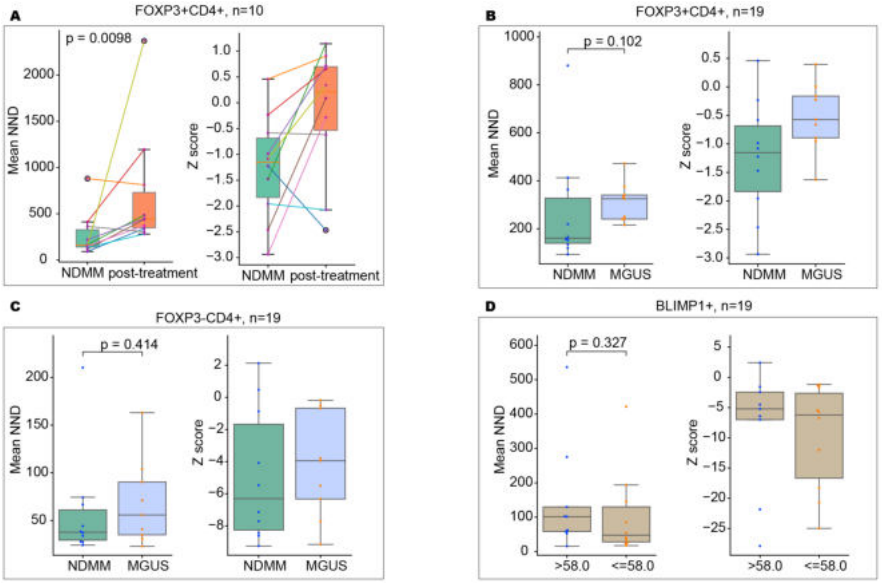


Figure S3.9: **Clustered or dispersed pattern of immune and tumour cells in a BM trephine sample: A-B)** Boxplots showing the difference in NND and Z score of FOXP3+CD4+ cells between NDMM and post-treatment **(A)**, FOXP3+CD4+ cells between NDMM and MGUS **(B)**, FOXP3-CD4+ cells between NDMM and MGUS **(C)**, and BLIMP1+ cells between age groups (median split) **(D)**. The Z score shows the significance of the difference between the NND distribution for a given cell type from a complete spatial random distribution and the observed NND (Methods).

S4

Supplementary Data for Chapter 5

Table S4.1: **Correlation between pathologists stromal TILs scores and cAI-TIL computed using different stromal boundary scales.** Correlation (r) and p values were computed using Spearman method.

Boundary (millimetres)	r value	p value
0.03	0.23	0.12
0.05	0.35	0.017
0.067	0.43	0.0029
0.08	0.45	0.001788976
0.1	0.53	1.45E-04
0.2	0.65	1.12E-06
0.3	0.66	5.99085E-07
0.4	0.66	6.28E-07
0.5	0.66	5.12E-07
0.6	0.67	3.28E-07
0.7	0.67	3.83E-07
0.8	0.66	7.075E-07

Table S4.2: **The concordance index (C-index) between observed time and predicted recurrence free survival using cAI-TIL feature.** Continuous values of cAI-TIL were used to measure C-index.

Stromal boundary (mm)	C-index
0.03	0.58
0.05	0.55
0.067	0.54
0.08	0.54
0.10	0.53
0.20	0.51
0.30	0.51
0.40	0.51
0.50	0.51
0.60	0.52
0.70	0.52
0.80	0.52

Table S4.3: **Patient level: DCIS morphology and mutation burden association for all cases:** Correlation (r) values were computed using Spearman method. Stromal cAI-TIL was computed using stromal boundary of $30\mu m$. n=34 patients.

Features	r value	p value
Common mutations vs cAI-TIL	0.113	0.526
Common mutations vs Duct Area	0.122	0.492
Common mutations vs Duct Solidity	0.041	0.817
Common mutations vs Num Ducts	0.11	0.537
Mutations divergence % vs cAI-TIL	-0.133	0.475
Mutations divergence % vs Duct Area	0.016	0.93
Mutations divergence % vs Duct Solidity	0.171	0.357
Mutations divergence % vs Num Ducts	-0.062	0.742
Total vs cAI-TIL	-0.006	0.974
Total vs Duct Area	0.386	0.024
Total vs Duct Solidity	0.179	0.312
Total vs Num Ducts	0.17	0.335

Table S4.4: **Patient level: DCIS morphology and mutation burden association for no-IBE, DCISr and IBCr cases:** Correlation (r) values were computed using Spearman method. Stromal cAI-TIL was computed using stromal boundary of 30μm. IBCr (n=14), DCISr (n=14),no-IBE (n=6).

Patient group	Features	r value	p value
DCISr	Common mutations vs cAI-TIL	-0.269	0.352
no-IBE	Common mutations vs cAI-TIL	0.257	0.623
IBCr	Common mutations vs cAI-TIL	0.382	0.178
DCISr	Common mutations vs Duct Area	-0.194	0.506
no-IBE	Common mutations vs Duct Area	-0.2	0.704
IBCr	Common mutations vs Duct Area	0.629	0.016
DCISr	Common mutations vs Duct Solidity	0.011	0.97
no-IBE	Common mutations vs Duct Solidity	0.2	0.704
IBCr	Common mutations vs Duct Solidity	0.205	0.481
DCISr	Common mutations vs Num Ducts	0.301	0.296
no-IBE	Common mutations vs Num Ducts	-0.314	0.544
IBCr	Common mutations vs Num Ducts	-0.033	0.91
DCISr	Mutations divergence % vs cAI-TIL	-0.039	0.904
no-IBE	Mutations divergence % vs cAI-TIL	-0.714	0.111
IBCr	Mutations divergence % vs cAI-TIL	-0.283	0.349
DCISr	Mutations divergence % vs Duct Area	0.128	0.691
no-IBE	Mutations divergence % vs Duct Area	0.314	0.544
IBCr	Mutations divergence % vs Duct Area	0.433	0.139
DCISr	Mutations divergence % vs Duct Solidity	0.637	0.026
no-IBE	Mutations divergence % vs Duct Solidity	0.257	0.623
IBCr	Mutations divergence % vs Duct Solidity	-0.008	0.978
DCISr	Mutations divergence % vs Num Ducts	-0.117	0.716
no-IBE	Mutations divergence % vs Num Ducts	0.086	0.872
IBCr	Mutations divergence % vs Num Ducts	-0.02	0.949
DCISr	Mutation burden vs cAI-TIL	0.033	0.911
no-IBE	Mutation burden vs cAI-TIL	0.143	0.787
IBCr	Mutation burden vs cAI-TIL	-0.154	0.599
DCISr	Mutation burden vs Duct Area	0.068	0.817
no-IBE	Mutation burden vs Duct Area	0.6	0.208
IBCr	Mutation burden vs Duct Area	0.748	0.002
DCISr	Mutation burden vs Duct Solidity	0.02	0.946
no-IBE	Mutation burden vs Duct Solidity	0.429	0.397
IBCr	Mutation burden vs Duct Solidity	0.176	0.547
DCISr	Mutation burden vs Num Ducts	0.218	0.454
no-IBE	Mutation burden vs Num Ducts	0.486	0.329
IBCr	Mutation burden vs Num Ducts	0.004	0.988

Table S4.5: **Univariate analysis of prognostic value of mutation burden and mutation divergence**. Continuous values of mutation burden and mutation divergence were used.

Feature	HR	95% CI	P value
Total	1.01	0.99-1.03	0.0518
Divergence	1.01	0.99-1.02	0.271

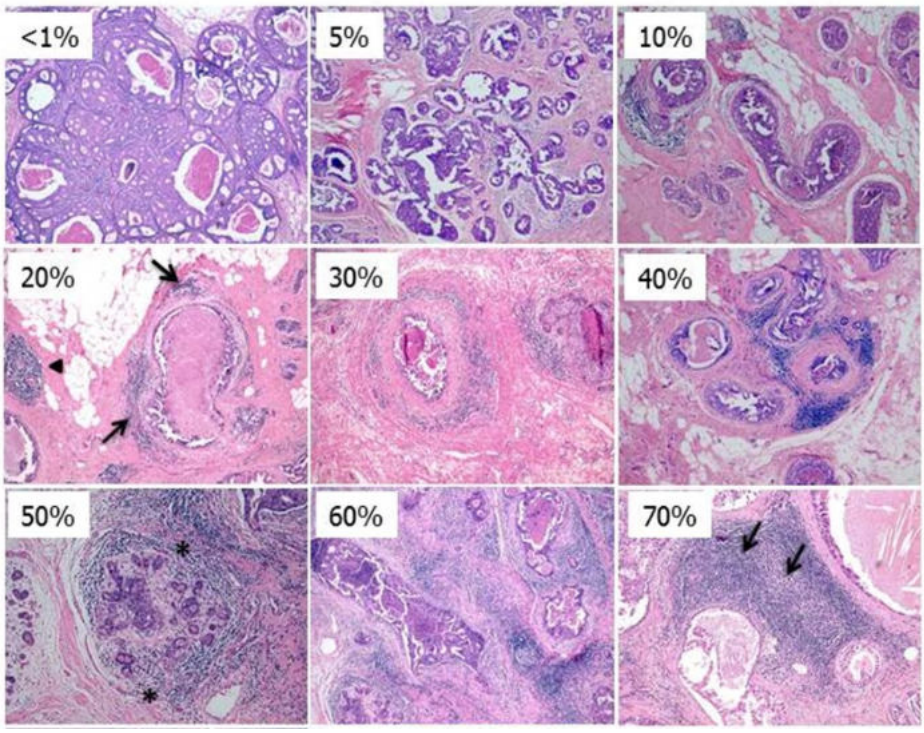


Figure S4.1: **Sample images showing how pathologists score TILs in DCIS.** Image credit: https://www.tilsinbreastcancer.org/

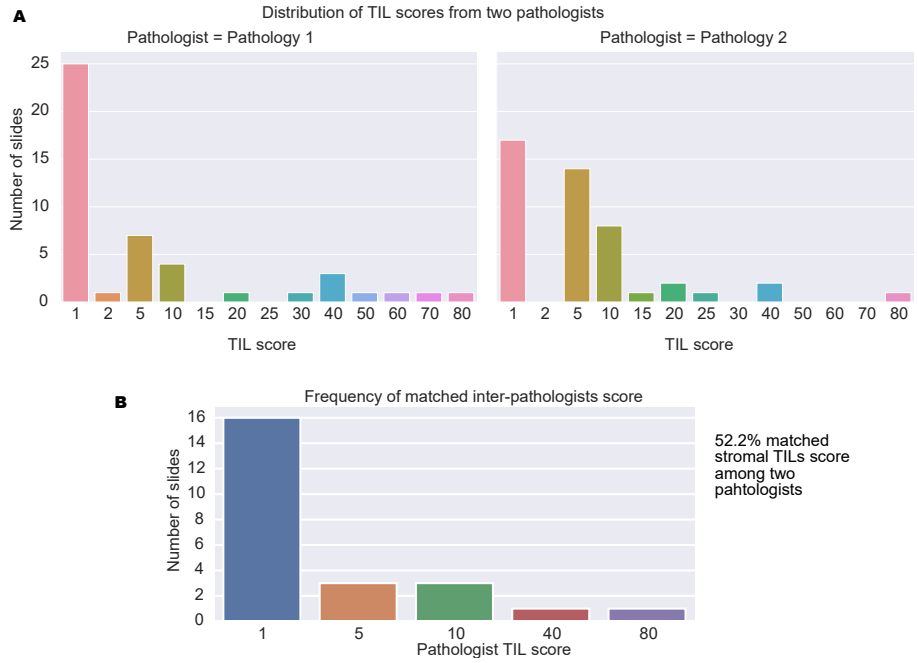


Figure S4.2: **Distribution of pathologists stromal TILs scoring: A)** Frequency plot of pathologists TILs scores. **B)** Frequency plot of matched stromal TILs scores.

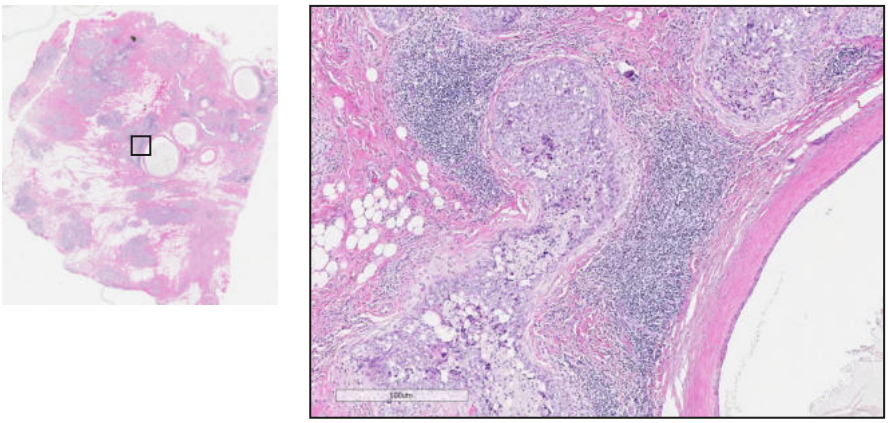


Figure S4.3: **Inter-pathologist stromal TILs scoring variation was mostly in high TILs slides.** The stromal TILs scores for this was 25% by pathologist 1 and 70% by pathologist 2.

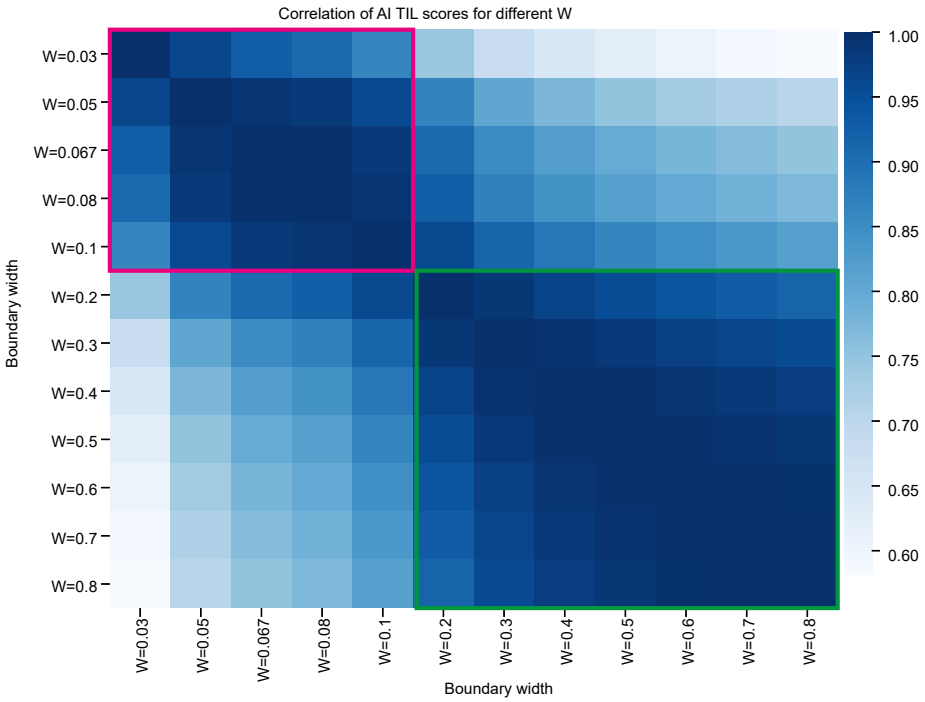


Figure S4.4: **Correlation between stromal AI-TIL computed using different stromal boundaries**

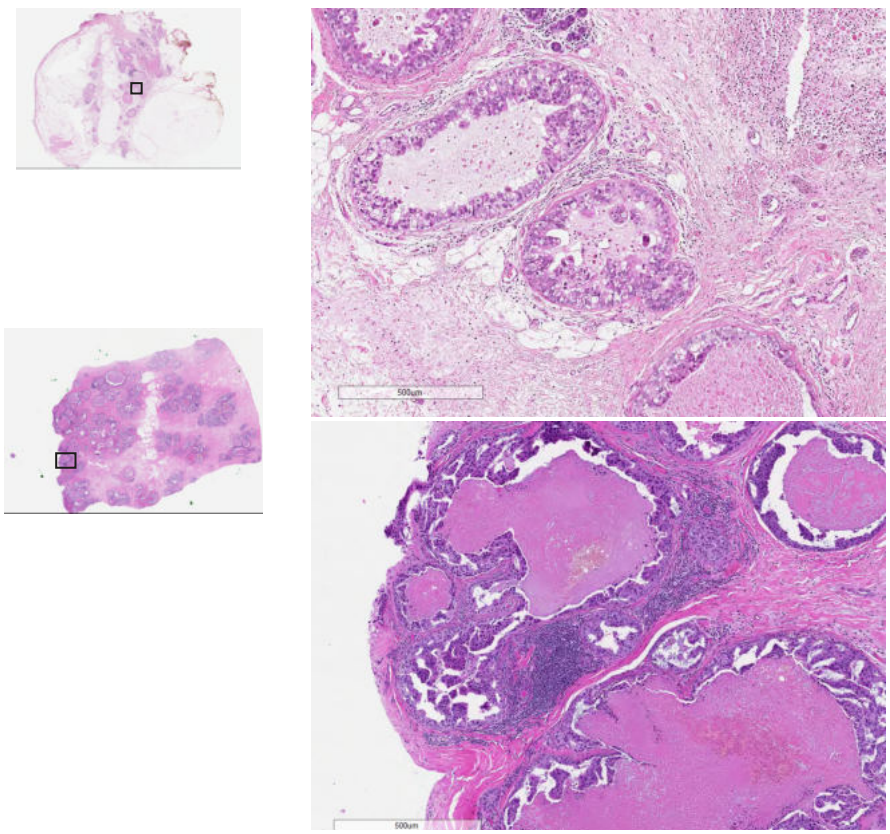


Figure S4.5: **Sample images with low inter-pathologist stromal TILs scoring variability but high disagreement between pathologists score and automated method.**

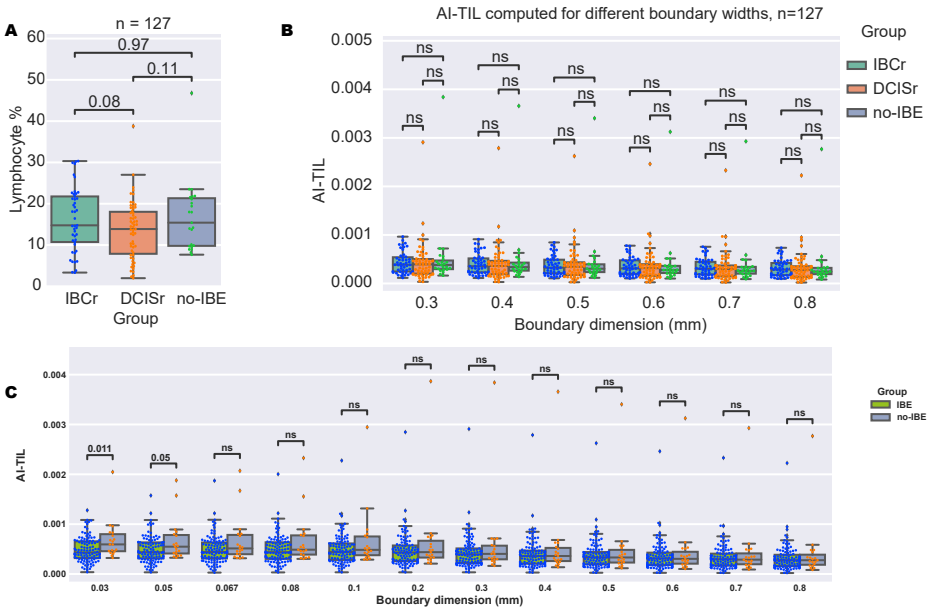


Figure S4.6: **Association between lymphocyte infiltration and recurrence in DCIS:A** Boxplots showing the difference in lymphocyte % (**A**) between patient groups. **B)** Boxplots showing the difference in DCIS AI-TIL between recurrence groups for various TILs scoring boundary. **B)** Boxplots showing the difference in DCIS AI-TIL between recurrence and non-recurrence patients for various TILs scoring boundary. Number of patients: recurrence groups (no-IBE (n=19), DCISr (n=61) and IBCr (n=47)); IBE is combination of DCISr and IBCr groups (n=108). The p-values were computed using a two-sided unpaired non-parametric Wilcoxon method considering $p<0.05$ significant. ER = estrogen receptor; +Ve = Positive; -Ve = Negative.

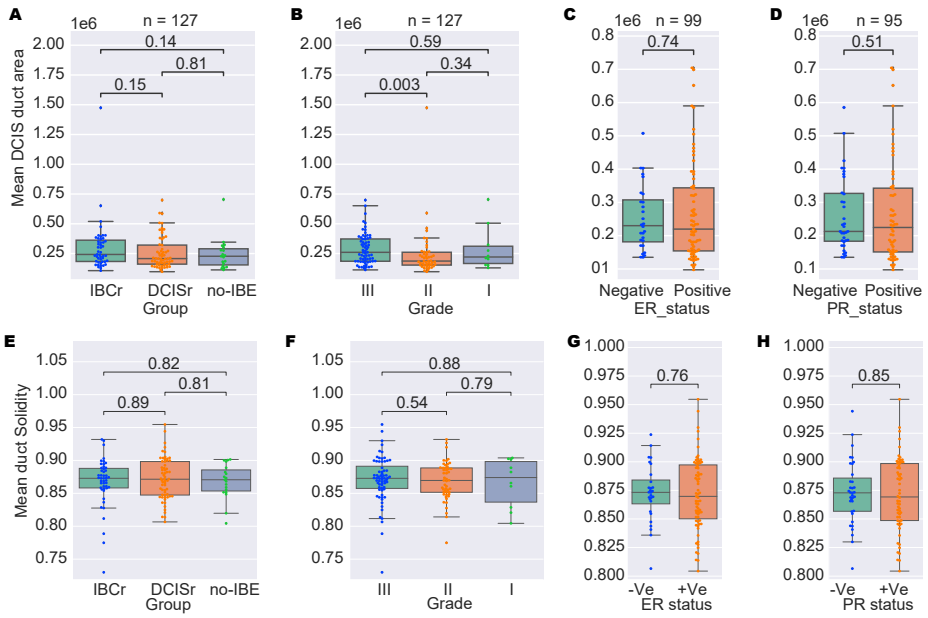


Figure S4.7: **Association of DCIS ducts morphology with recurrence status, grade, ER status, and PR status:** Boxplots showing difference in DCIS duct area (**A-D**) and DCIS duct solidity (**E-H**) between different patient groups. The label of y-axis is shared among all plots in a row. Number of patients: recurrence groups (no-IBE (n=19), DCISr (n=61) and IBCr (n=47)); Grade (**C**) (III (n=68), II (n=49), and I (n=10)); ER status (+ve (n=71) and -Ve (n=28), and PR status (+Ve (n=60) and -Ve (n=35)). ; PR = progesterone receptor; ER = estrogen receptor; +Ve = Positive; -Ve = Negative. The p-values were computed using a two-sided unpaired non-parametric Wilcoxon method considering $p<0.05$ significant.

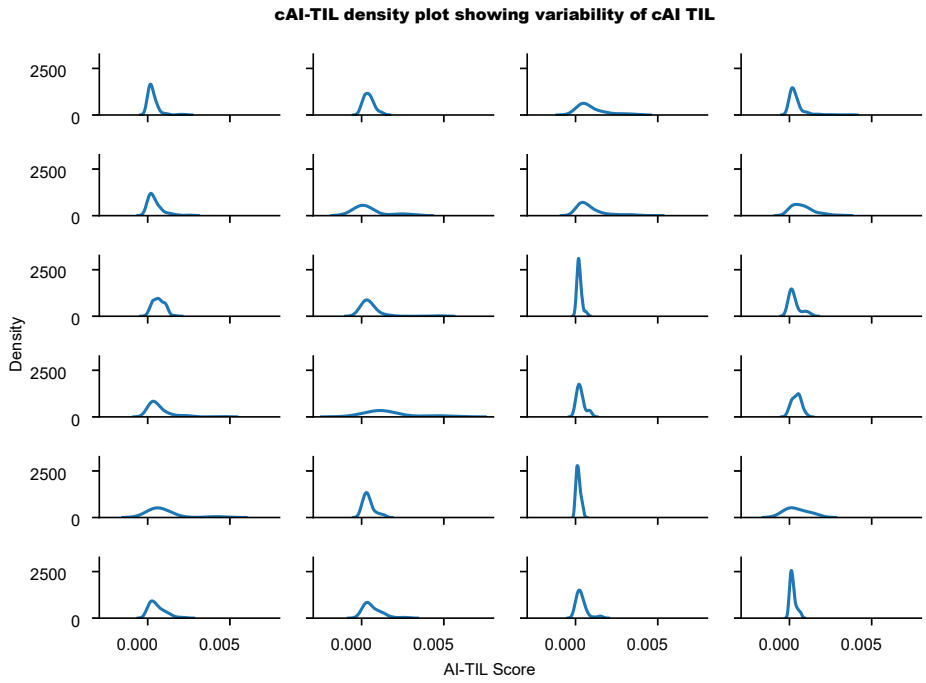


Figure S4.8: **Kernel density plot showing heterogeneity of cAI-TIL score within and between slides.** One plot per slide. cAI-TIL score was computed for every DCIS duct and density estimation was applied. cAI-TIL was computed using boundary width of 0.03*mm*. Wider bandwidth indicates higher cAI-TIL heterogeneity. The label of y-axis is shared among all plots in a row.

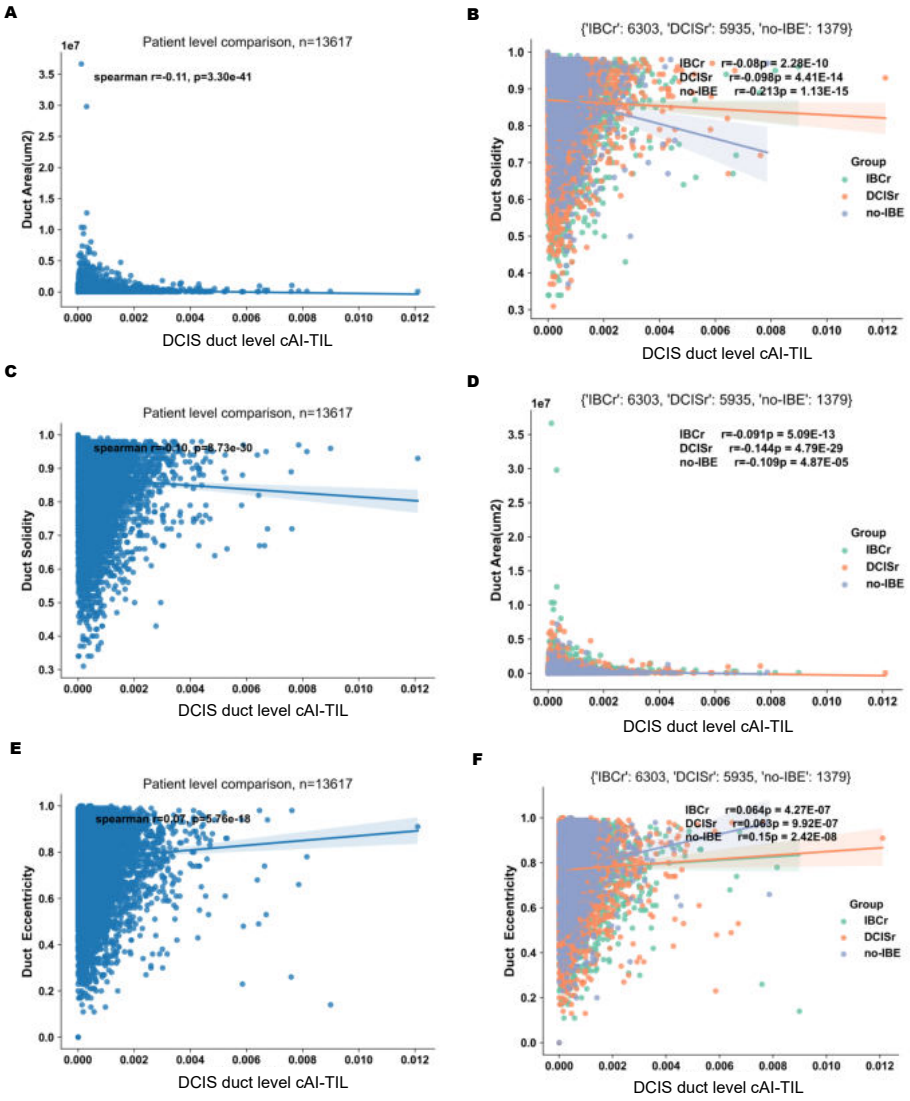


Figure S4.9: **DCIS ducts morphology and cAI-TIL infiltration at duct level:** Correlation between of DCIS duct area (**A,B**), solidity (**C,D**), and eccentricity (**E,F**) with stromal cAI-TIL computed using boundary width of 0.03*mm*. n represents number DCIS ducts. A point represents a DCIS duct. All correlations were computed using spearman method.

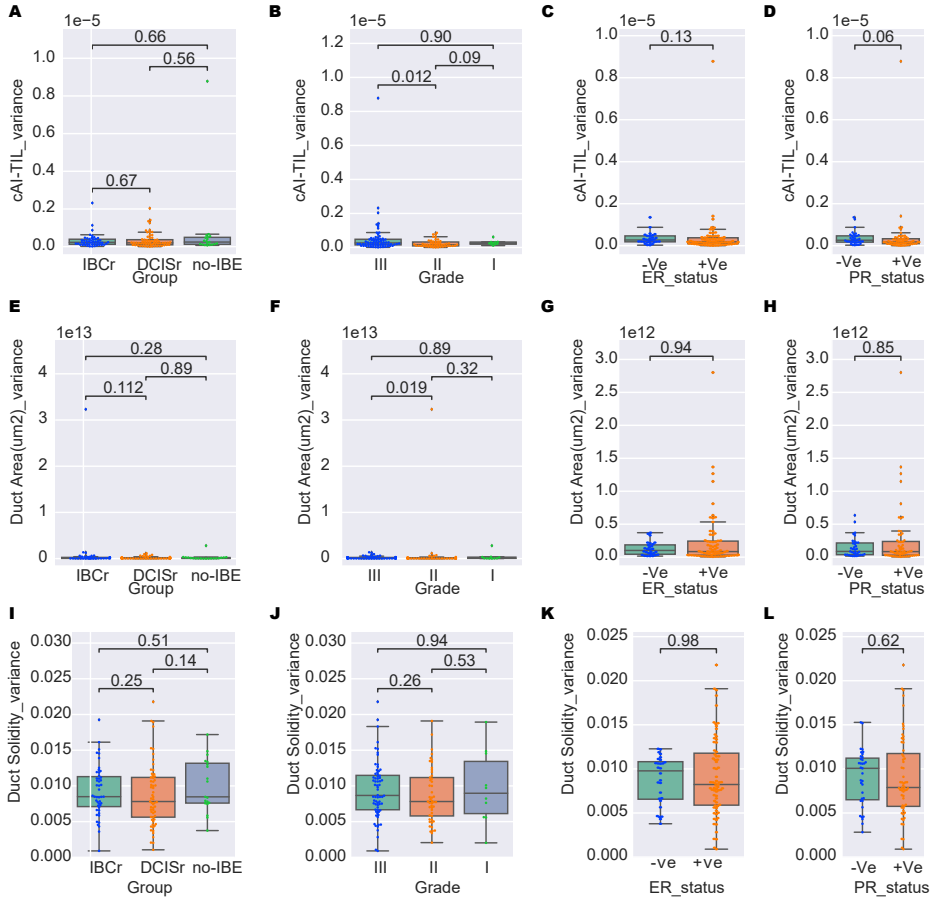


Figure S4.10: **Association between heterogeneity of DCIS morphology and cAI-TIL at duct DCIS duct level with clinical variables:** Boxplots showing difference in variance of cAI-TIL (**A-D**), DCIS duct area (**E-H**) and DCIS duct solidity (**I-L**) between different patient groups. The label of y-axis is shared among all plots in a row. Number of patients: recurrence groups (no-IBE (n=19), DCISr (n=61) and IBCr (n=47)); Grade (**C**) (III (n=68), II (n=49), and I (n=10)); ER status (+ve (n=71) and -Ve (n=28), and PR status (+Ve (n=60) and -Ve (n=35)). ns = not significant; PR = progesterone receptor; ER = estrogen receptor

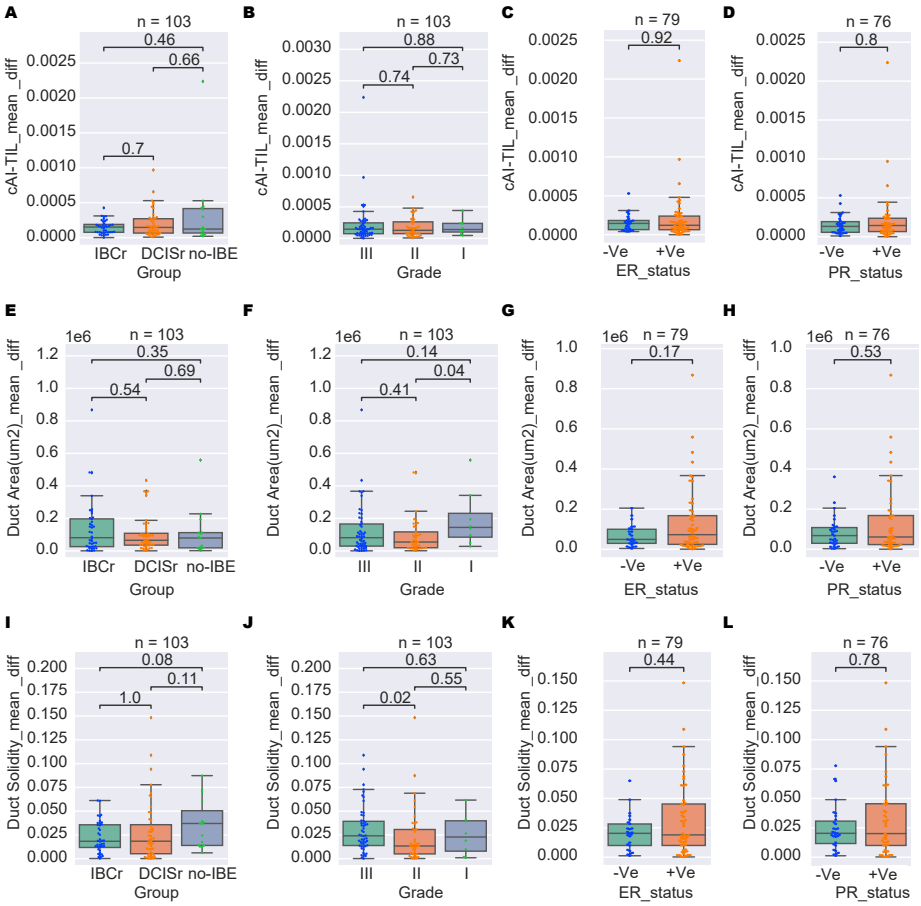


Figure S4.11: **Association between inter tumour block heterogeneity of DCIS morphology and cAI-TIL with clinical variables:** Boxplots showing difference in inter tumour block difference in cAI-TIL (**A-D**), DCIS duct area (**E-H**) and DCIS duct solidity (**I-L**) between different patient groups. The label of y-axis is shared among all plots in a row. Number of patients: PR = progesterone receptor; ER = estrogen receptor. Slide level features were computed as mean of DCIS level feature. Then, inter block heterogeneity was computed as absolute value of difference between the scores from the two blocks.

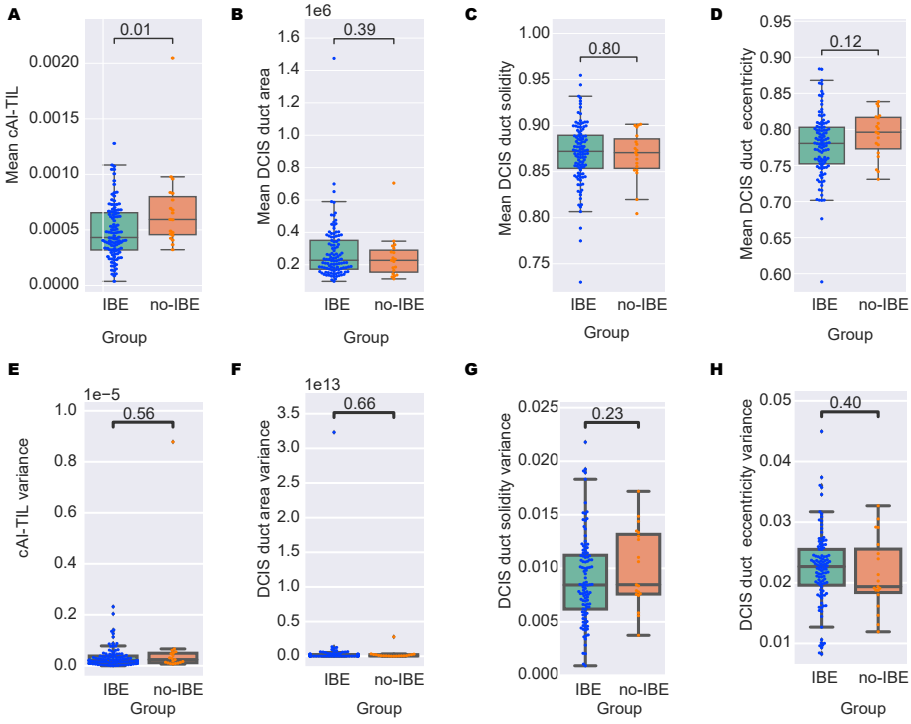


Figure S4.12: **Difference between patients with IBE and no-IBE in terms of DCIS ducts morphologic features and cAI-TIL heterogeneity:** cAI-TIL ($W = 0.03\mu m$), solidity, DCIS duct area, and eccentricity were computed at duct level. This allows estimating the mean and variance of these features at slide level or patient level. When a patient has multiple slides, mean aggregation was applied. Every point in the plots represents a patient. IBE (n=108), and no-IBE (n=19)

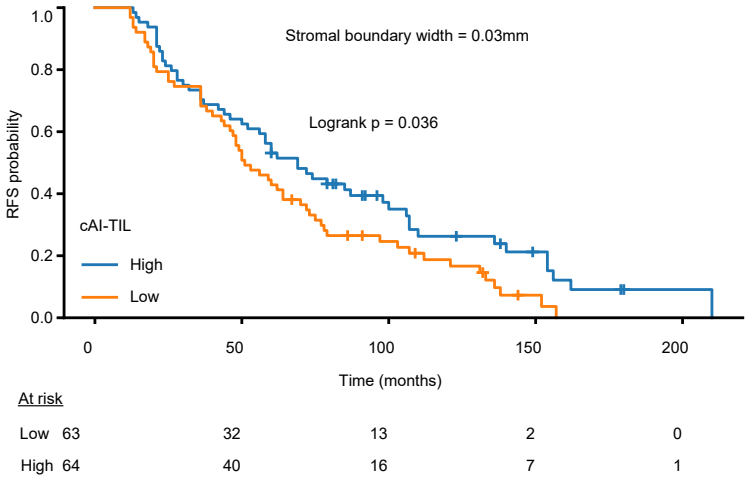


Figure S4.13: **B) Kaplan–Meier curves illustrating RFS of patients dichotomized using median cAI-TIL.** Median value of cAI-TIL was used to stratify the patients into high and low group.

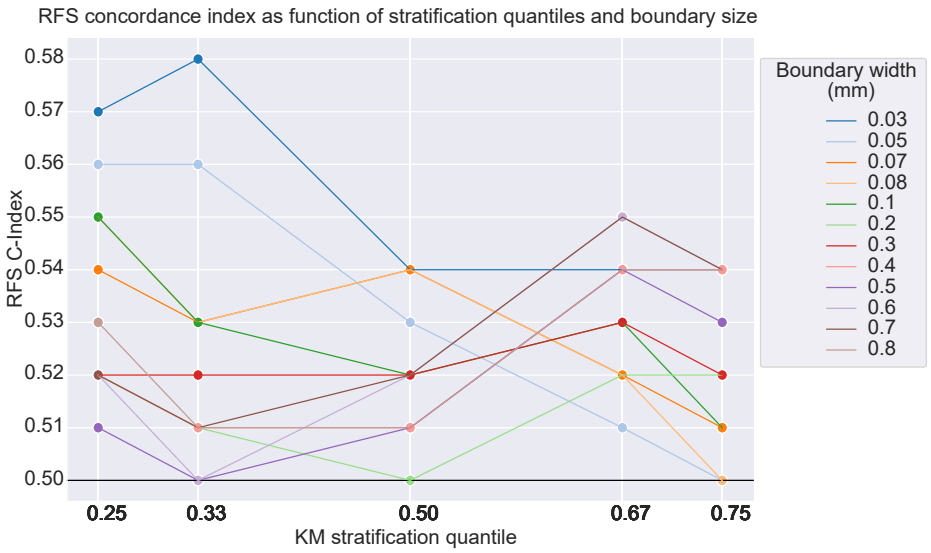


Figure S4.14: Evaluation of concordance index

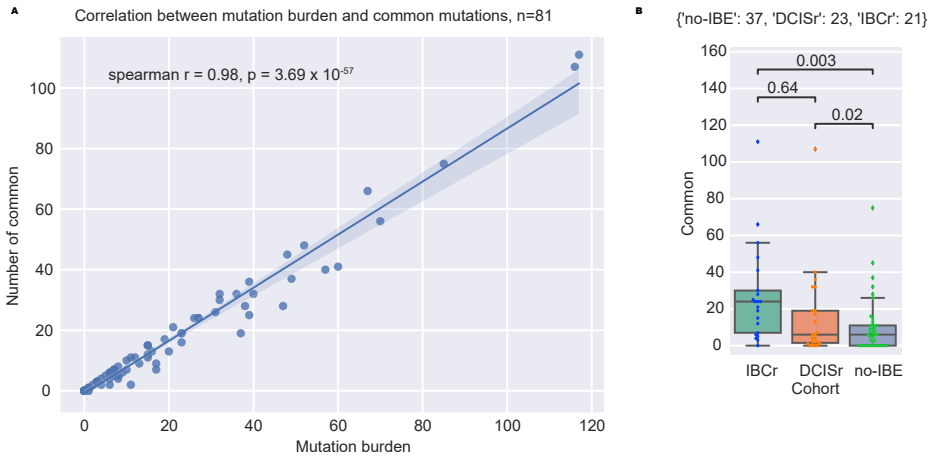


Figure S4.15: **Analysis of number of common mutations: A)** Correlation between number of common mutation and mutation burden. **B)** A boxplot showing difference in common mutations between recurrence groups.

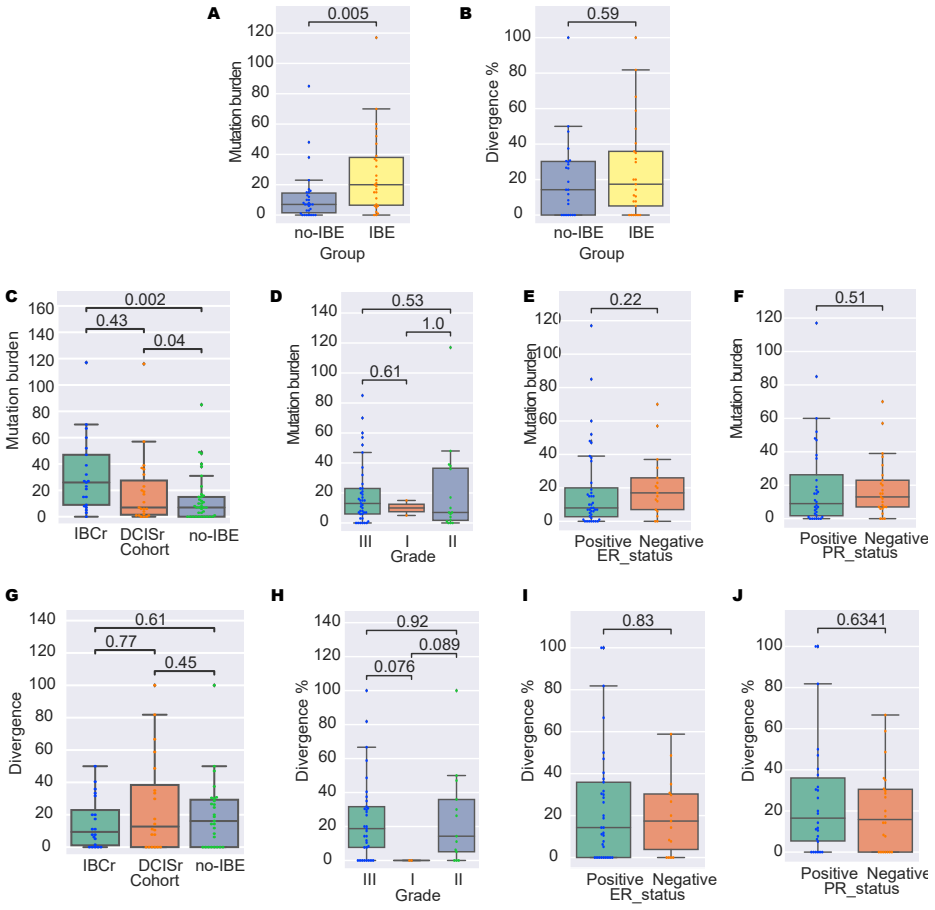


Figure S4.16: **Association of mutation data with recurrence, grade, ER status, and PR status: A,B** Boxplots showing difference in mutation burden (A) and mutation divergence % (B) between IBE and no-IBE cases. **C-J)** Boxplots showing difference in mutation burden (**C-F**) and mutation divergence %(**G-J**) between different patient groups. Number of patients: recurrence groups (no-IBE (n=37), DCISr (n=23) and IBCr (n=21)); Grade (**C**) (III (n=33), II (n=13), and I (n=2)); ER status (Positive (n=33) and Negative (n=15), and PR status (Positive (n=26) and Negative (n=22)). IBE group contains combined DCISr and IBCr patients (n=44). PR = progesterone receptor; ER = estrogen receptor

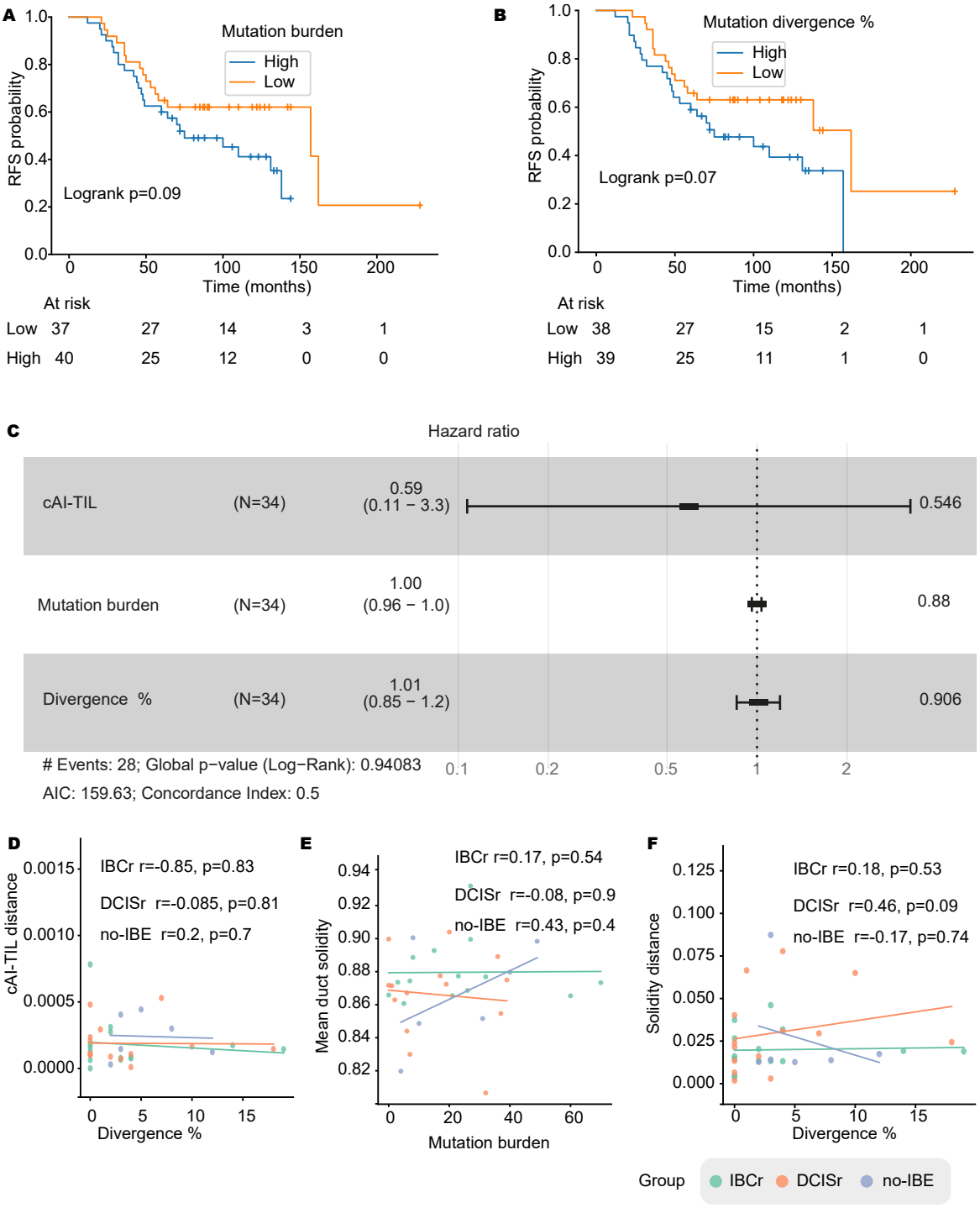


Figure S4.17: **Association of mutation features with histologic features and recurrence free survival A,B)** Kaplan–Meier curves illustrating RFS of patients dichotomized using median mutation burden (A) and mutation divergence % (B). **C)** Forest plots showing multivariate Cox regression analyses. Continuous values of mutation burden, cAI-TIL and mutation divergence % were used. **D-F)** Scatter plot showing correlation of cAI-TIL distance vs mutation divergence % (D, no-IBE (n=6), DCISr (n=12) and IBCr (n=13)), mean duct solidity vs mutation burden (E, no-IBE (n=6), DCISr (n=14) and IBCr (n=14)) and DCIS duct solidity distance vs mutation divergence (F, no-IBE (n=6), DCISr (n=12) and IBCr (n=13)).