LUND UNIVERSITY

The immune microenvironment in mantle cell lymphoma

Targeted liquid and spatial proteomic analyses

Lokhande, Lavanya

2023

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Lokhande, L. (2023). *The immune microenvironment in mantle cell lymphoma: Targeted liquid and spatial proteomic analyses*. [Doctoral Thesis (compilation), Department of Immunotechnology]. Department of Immunotechnology, Lund University.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

The immune microenvironment in mantle cell lymphoma

Targeted liquid and spatial proteomic analyses

LAVANYA LOKHANDE
DEPT OF IMMUNOTECHNOLOGY | FACULTY OF ENGINEERING | LUND UNIVERSITY

Printed by Media-Tryck, Lund 2023 NORDIC SWAN ECOLABEL 3041 0903

Lund University
Faculty of Engineering
Department of Immunotechnology

ISBN 978-91-8039-546-5

The immune microenvironment in mantle cell lymphoma

The immune microenvironment in mantle cell lymphoma

Targeted liquid and spatial proteomic analyses

Lavanya Lokhande

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Engineering at Lund University to be publicly defended on 2nd of March 2023, at 09.00 in Horsalen, Medicon Village, Scheelevägen 2, 223 87 Lund

Faculty opponent
Karen Dybkær, Professor,
Department of Clinical Medicine, Aalborg University Hospital

Organization LUND UNIVERSITY Department of Immunotechnology Medicon Village (building 406) SE-223 87 Lund Sweden	**Document name** Doctoral Dissertation **Date of issue** 2nd March 2023
Author(s) Lavanya Lokhande	Sponsoring organization

Title and subtitle: The immune microenvironment in mantle cell lymphoma
Targeted liquid and spatial proteomic analyses

Abstract

The complex interplay of the tumour and immune cells affects tumour growth, progression, and response to treatment. Restoration of effective immune response forms the basis of onco-immunology, which further enabled the development of immunotherapy. In the era of precision medicine, pin-pointing patient biological heterogeneity especially in relation to patient-specific immune microenvironment is a necessity for the discovery of novel biomarkers and for development of patient stratification tools for targeted therapeutics. Mantle cell lymphoma (MCL) is a rare and aggressive subtype of B-cell lymphoma with poor survival and high relapse rates. Previous investigations of MCL have largely focused on the tumour itself and explorations of the immune microenvironment have been limited. This thesis and the included five papers, investigates multiple aspects of the immune microenvironment with respect to proteomic analysis performed on tissue and liquid biopsies of diagnostic and relapsed/refractory (R/R) MCL cohorts. Analyses based on liquid biopsies (serum) in particular are relevant for aggressive cases such as in relapse, where invasive procedures for extracting tissues is not recommended. Thus, paper I-II probes the possibility of using serum for treatment and outcome-associated biomarker discovery in R/R MCL, using a targeted affinity-based protein microarray platform quantifying immune-regulatory and tumor-secretory proteins in sera. Analysis performed in paper I using pre-treatment samples, identifies 11-plex biomarker signature (RIS – relapsed immune signature) associated with overall survival. Further integration of RIS with mantle cell lymphoma international prognostic index (MIPI) led to the development of $MIPI_{ris}$ index for the stratification of R/R MCL into three risk groups. Moreover, longitudinal analysis can be important in understanding how patient respond to treatment and this can further guide therapeutic interventions. Thus, paper II is a follow-up study wherein longitudinal analyses was performed on paired samples collected at pre-treatment (baseline) and after three months of chemo-immunotherapy (on-treatment). We show how genetic aberrations can influence systemic profiles and thus integrating genetic information can be crucial for treatment selection. Furthermore, we observe that the inter-patient heterogeneity associated with absolute values can be circumvented by using velocity of change to capture general changes over time in groups of patients. Thus, using velocity of change in serum proteins between pre- and on-treatment samples identified response biomarkers associated with minimal residual disease and progression. While exploratory analysis using high dimensional omics-based data can be important for accelerating discovery, translating such information for clinical utility is a necessity. Thus, in paper III, we show how serum quantification can be used complementary tissue-identified prognostic biomarkers and this can enable faster clinical implementation. Presence of CD163+ M2-like macrophages has shown to be associated with poor outcome in MCL tissues. We show that higher expression of sCD163 levels in sera quantified using ELISA, is also associated with poor outcome in diagnostic and relapsed MCL. Furthermore, we suggest a cut-off for sCD163 levels that can be used for clinical utility. Further exploration of the dynamic interplay of tumour-immune microenvironment is now possible using spatial resolved omics for tissue-based analysis. Thus, in paper IV and V, we analyse cell-type specific proteomic data collected from tumour and immune cells using GeoMx™ digital spatial profiler. In paper IV, we show that presence as well as spatial localization of CD163+ macrophage with respect to tumour regions impacts macrophage phenotypic profiles. Further modulation in the profile of surrounding tumour and T-cells is observed when macrophages are present in the vicinity. Based on this analysis, we suggest MAPK pathway as a potential therapeutic target in tumours with CD163+ macrophages. Immune composition can be defined not just by the type of cells, but also with respect to frequency and spatial localization and this is explored in paper V with respect to T-cell subtypes. Thus, in paper V, we optimized a workflow of multiplexed immunofluorescence image segmentation that allowed us to extract cell metrics for four subtypes of CD3+ T-cells. Using this data, we show that higher infiltration of T-cells is associated with a positive outcome in MCL. Moreover, by combining image derived metrics to cell specific spatial omics data, we were able to identify immunosuppressive microenvironment associated with highly infiltrated tumours and suggests new potential targets of immunotherapy with respect to IDO1, GITR and STING. In conclusion, this thesis explores systemic and tumor-associated immune microenvironment in MCL, for defining patient heterogeneity, developing methods of patient stratification and for identifying novel and actionable biomarkers.

Key words Immune microenvironment, serum analysis, spatial omics, image analysis, proteomics		
Classification system and/or index terms (if any)		
Supplementary bibliographical information		**Language** English
ISSN and key title		**ISBN** 978-91-8039-546-5 (print) 978-91-8039-547-2 (electronic)
Recipient's notes	**Number of pages** 94	Price
	Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature Date 2023-01-18

The immune microenvironment in mantle cell lymphoma

Targeted liquid and spatial proteomic analyses

Lavanya Lokhande

LUND
UNIVERSITY

Coverphoto made by author using Midjourney and Photoshop

Copyright pp. 1-94 Lavanya Lokhande
Paper 1 © BMC, part of Springer Nature
Paper 2 © Wiley online library
Paper 3 © by the Authors (Manuscript unpublished)
Paper 4 © by the Authors (Manuscript unpublished)
Paper 5 © by the Authors (Manuscript unpublished)

Faculty of Engineering
Department of Immunotechnology

ISBN 978-91-8039-546-5 (print)
ISBN 978-91-8039-547-2 (electronic)

Printed in Sweden by Media-Tryck, Lund University, Lund 2023

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se
MADE IN SWEDEN

To my family

"Success is like a shadow, don't catch it. Walk your way with it, and it will follow you automatically. Remember shadow follows you when walk towards brightness"

- Dr A. P. J. Abdul Kalam

Table of Contents

Papers included in this thesis

Paper I – Lokhande L, Kuci Emruli V, Kolstad A, Kolstad A, Hutchings M, Räty R, Jerkeman M, Ek S. Immune-related protein signature in serum stratify relapsed mantle cell lymphoma patients based on risk. BMC Cancer 20, 1202 (2020).

Paper II – Lokhande L, Kuci Emruli V, Eskelund CW, Kolstad A, Hutchings M, Räty R, Utoft Niemann C, Grønbæk K, Jerkeman M, Ek S. Serum proteome modulations upon treatment provides biological insight on response to treatment in relapsed mantle cell lymphoma. Cancer Reports. 2022; 5(7):e1524.

Paper III – Nikkarinen A, Lokhande L, Amini RM, Porwit A, Molin D, Enblad G, Jerkeman M, Weibull C, Hollander P, Ek S, Glimelius I. Soluble CD163 in serum predicts outcome in patients with mantle cell lymphoma both in chemoimmunotherapy treated and targeted therapy treated patients (*Manuscript*)

Paper IV – Rodrigues JM, Lokhande L, Gerdtsson AS, Nikkarinen A, Hollander P, Porwit A, Glimelius I, Jerkeman M, Ek S. Impact of spatial localization and presence of macrophages on immune suppression in mantle cell lymphoma (*Manuscript*)

Paper V – Lokhande L, Rodrigues JM, Nilsson D, Zelco A, Gerdtsson AS, Jerkeman M, Ek S. Combined deep-learning assisted image and spatial proteomic analyses reveals immune-regulatory targets on T-cell subsets in the mantle cell lymphoma microenvironment (*Manuscript*)

Author contribution

Paper I: Involved in the design of the study and performed all the practical lab work as well computational data analysis and pipeline development. Major parts of the manuscript were drafted by me.

Paper II: Involved in the design of the study and performed all the practical lab work as well computational data analysis and pipeline development. Major parts of the manuscript were drafted by me.

Paper III: Involved in the planning of the study, performed all the wet-lab work with respect to the assay. Also involved in the parts of data analysis and responsible for the analysis done on defining the clinically relevant cut-off. Actively participated in the correction of the manuscript.

Paper IV: Involved in the planning of the study and ROI selection. I was the main responsible for data pre-processing and developing the underlying computational workflows used for data analysis. Actively participated in the correction of the manuscript.

Paper V: Designed the study and was involved in ROI selection and image acquisition. Main responsible for outlining and evaluation of the pipeline optimization, for data pre-processing and combined data analysis. Major parts of the manuscript were drafted by me.

Abbreviations

AI - Artificial intelligence
ASCT - Autologous stem cell transplantation
BE - Batch effects
BECA - Batch effects correction algorithms
BF - Boundary F1
BLISS - B-cell lymphomas in Southern Sweden
BTK - Bruton's tyrosine kinase
CHOP - Cyclophosphamide, doxorubicin, vincristine, and prednisone
CNN - Convolutional Neural Network
DL – Deep learning
DSC - Dice Similarity Coefficient
DSP - Digital Spatial Profiler
ECOG - Eastern Cooperative Oncology Group
ELISA - Enzyme-linked immunosorbent assay
FL - Follicular lymphoma
GLOBOCAN - Global cancer observatory
HGP - Human Genome Project
HL - Hodgkin lymphoma
IF - Immunofluorescence
IGHV - Immunoglobulin heavy chain variable region
IHC - Immunohistochemistry
IMC - Imaging mass cytometry
IO - immuno-oncology
IoU - Intersection-Over-Union
LDH - Lactate dehydrogenase
LMM - Linear mixed models
MCL – Mantle cell lymphoma
mIF - Multiplexed immunofluorescence
MIPI - Mantle cell lymphoma prognostic index

$MIPI_{ris}$ - Mantle Cell Lymphoma International Prognostic Index – Relapsed Immune Signature
ML - Machine learning
MRD - Minimal residual disease
NCCN - National Comprehensive Cancer Network
NHL - Non-Hodgkin lymphoma
OPLS - Supervised orthogonal partial least square
OS – Overall Survival
PCA - Principal component analysis
PFS – Progression Free survival
PVCA - Principal variance component analysis
R/R MCL – Relapsed/Refractory Mantle cell lymphoma
RIS - Relapsed-immune signature
RLE - Relative log expression
ROI – Region of Interest
RUV - Removed unwanted variation
ScFv – Single-chain variable fragment
SVA - Surrogate variable analysis
SVM - Support vector machines
TIME - Tumour-immune microenvironment
TMA – Tissue microarray
TME - Tumour microenvironment
TP53 – Tumor protein 53
TTP - Time-to-progression
U-CAN - The Uppsala-Umeå Comprehensive Cancer Consortium
U-NET – U-type convolutional Neural Network
VIOLA - Vital Freezing of Lymphoma Cells
WBC – White blood cell
WHO-HAEM5 - World Health Organization Classification of Haematolymphoid Tumours

Chapter 1: Introduction

1.1 The shift in therapeutics – from "one size fits all" to personalized medicine

Cancer remains to be the leading cause of death worldwide with an estimated 10 million deaths and 19.3 million new cases in 2020[1]. Current projection suggests that nearly one out of three people during their lifetime would be affected by this malignant disease. Thus, developing more effective therapeutic strategies to treat cancer is a must.

Therapeutics in oncology has developed from the previous "one size fits all" approach to targeted patient treatment paving way to the era of personalized medicine[2] (Figure 1). The traditional approach of treatment decision making, relied on phenotypic outlook of the disease coupled with clinical experience[2]. This later evolved into Evidence Based Medicine (EBM) which is defined as "*conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients*"[3]. While this process had seen some success with its average treatment approach based on population dynamics, it inherently negated patient heterogeneity which influences the treatment outcome[4]. Thus, it has become abundantly clear that such an approach was not generally applicable, considering the observed variation in patient outcomes.

Furthermore, the choices for conventional first line therapeutics were initially restricted to surgical interventions, radiation therapy and chemotherapy. Technological advancements, such as genome sequencing, biomolecule expression profiling etc. have expanded the onco-therapeutic options. Now, several new, potentially curative approaches are being tested and even replacing first-line cancer therapy. Such treatment options include immunotherapy, cell-based therapy, CAR-T-cell therapy, cancer vaccines and gene therapy. Additionally, new technologies have augmented our biological understanding of cancer and the intricate mechanisms that lead to pathogenesis and disease progression. Therefore, such studies have highlighted patient heterogeneity and provided information on how the underlying biological processes can affect the overall disease progression and survival. Moreover, the development of new companion diagnostic and prognostic tools have helped in identifying subgroups of patients that require alternative therapeutic interventions, even though the underlying disease might be classified as

same. All-in-all, the need for therapies that are tailor-made to the specific individual and focuses on their intrinsic biology, have gradually become the future of onco-therapeutics. This has paved way to research exploring companion diagnostic and prognostic tools that can enable precision medicine efforts.

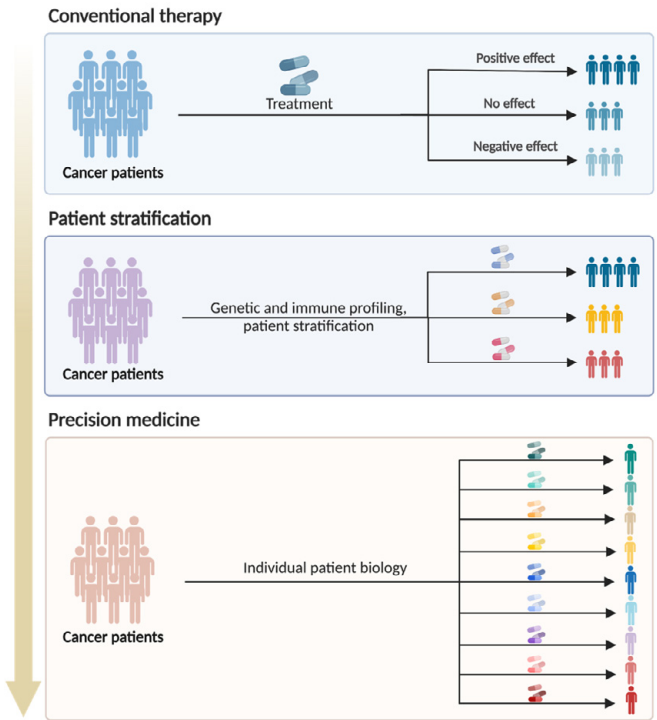

Figure 1: The movement from conventional therapy to precision medicine. Patients respond with great heterogeneity to conventional cancer treatment with some having beneficial response, some no response and sub-groups exhibiting adverse reactions to the therapy. Such variation is associated with heterogeneity in tumour biology and associated microenvironment. Thus, by using information from clinicopathology, and molecular assays (e.g., genetic make-up), patients can be stratified into subgroups with specialized treatment with respect to their respective biology. Further subgrouping and even tailored therapeutics for individuals is possible by comprehensively understanding patient's intrinsic make-up and such an approach is called precision or personalized medicine. This is a continuous goal for cancer therapeutics to ensure that each patient is positively benefited when treatment regimens are optimized. Figure was created using Biorender.

1.2 The rise of onco-immunology and impact on precision medicine

Originally, cancer research and therapeutics was focused only on cancer cells, and malignant growth was considered as a genetic disease[5]. However, in the last two decades, the systemic involvement in cancer development particularly with respect to the immune system and the contribution of the microenvironment, has been

intensively studied[5]. In 1950s, the concept of "immune surveillance of tumours" i.e., tumour-specific antigens released by the oncogenic cells and generating an immune response, was proposed, and validated over the years[6,7]. As proof of this, the presence of immune cells within tumour have shown to be associated with longer survival in many cancer subtypes[8–10].

However, cancer immunoediting theory was later proposed in light of the dual role of the immune system in both tumour inhibition and promotion[11–13]. The theory proposed a continuous process occurring in the tumour microenvironment, that could be highlighted as the three "E's" of cancer immunoediting; an initial step of cancer elimination, followed by immune-mediated equilibrium, and finally, immune escape/evasion of the cancer cells[11–13]. The importance of the immune system can be further highlighted by the new additions to the previous "hallmarks" of cancer. The original review by Hanahan and Weinberg published in 2000, initially proposed six cancer characteristics (evading apoptosis, insensitivity to anti-growth signals, self-sufficiency with respect to growth signals, tissue invasion/metastasis, angiogenesis, replicative immortality)[14]. In the 2011 update, four more hallmarks were proposed, of which immune evasion and metabolic reprogramming have been sufficiently validated by the 2022 update and have now been established as hallmarks complementing the original set[15,16]. The cancer-immunity cycle further explains that constant exchange of information and interaction between the tumour and surrounding immune cells[17]. Restoration of anti-tumour immune response by priming immune cells, especially the T cells, for cancer cell elimination is of particular interest and form the basis of onco-immunology[13]. Immunotherapy has led to tremendous improvement in patient outcome, as is evidenced by the implementation of Rituximab, a monoclonal anti-CD20 antibody in non-Hodgkin's lymphoma[18]. The complex and dynamic interaction of the immune and tumour cells contributes to the heterogeneity of the tumour-immune microenvironment (TIME) and stratification of patients with different immune compositions is a necessity for therapeutic personalization with respect to onco-immunology[13].

1.3 Technological progress that propelled the research in onco-immunology

Precision medicine relies on multi-modal information from various sources such as for example patient biometrics, clinical pathology, disease symptoms and phenotypes, socio-environmental factors, and multidimensional biological data[2]. The initial development of genomic sequencing was a crucial step in identifying genetic aberrations that drive tumorigenesis and progression, providing insights into some biological mechanisms and enabling identification of tools for clinical implementation[19]. However, it became clear that genetics alone would not suffice to

improve patient stratification and prognostication in all cancer sub-types, and therefore, focus on other molecular layers became critical[19]. Similarly, technologies have also moved from one biomolecule measurement to high-throughput high-dimensional measurements leading to development of "omics" and biomarker signatures. The term "omics" is broadly attributed to a "*collective study of molecular characterization and quantification of biological molecules from various subdomains of molecular biology using high-throughput technology*"[19]. Thus, from one-molecule single biological layer, we have moved to complex high-dimensional multi-layer biological exploration.

1.3.1 Challenges associated with clinical sampling and biological discovery

The two most common clinical samples collected from patients is solid biopsy tissue from sites of cancer presentation and liquid biopsies in the form of whole blood. There are advantages and disadvantages with both sampling methods. The disease genetics/phenotype and interaction with neighbouring tissue can be studied by using solid tumour tissue. Several experimental methodologies have been developed that allow tissue-based profiling for insights into the tumour microenvironment (TME). A popular method is flow cytometry that enables identification of rare cell types and is a widespread technique for immune cell profiling. Furthermore, during the era of sequencing, bulk genomics and single-cell sequencing platforms were developed that allowed high-throughput omics profiling. Such technologies involve disrupting the tissue; thus, the spatial information is lost. To retain the spatial biology, immunohistochemistry (IHC) or immunofluorescence (IF) were often used. But these techniques were limited in the number of simultaneous measurements that could be made. Additional factors to consider was the amount of available biological material, which is a limiting step since repeated sampling using invasive biopsy procedures is often not feasible or recommended. Now, we have moved into the era of spatial-omics wherein omics profiling can be done without disrupting the tissue. Such technologies especially enable the study of the TIME with much higher resolution than what was previously possible, allowing for investigation of the structural organization of the tumour, including immune cells and other components.

Tissue-based sampling is done for diagnostic reasons and is seldom recommended upon relapse or progression or in highly aggressive tumour types. Thus, to minimize invasive procedures, novel methods for continuous monitoring of the disease during and after treatment needs to be developed. In such cases, using liquid biopsies for extraction of systemic biological insight have become pivotal as sample collection is less invasive, and multiple sampling is a possibility which allows for longitudinal treatment monitoring. Additionally, the humoral component plays a major role in

the immune response, thus measuring immune-regulatory or tumour-secretory biomolecules can provide essential diagnostic or prognostic information[20].

1.3.2 The dawn of "Big Data" in translational cancer research

Technological innovations in relation to high-throughput profiling via omics using different biological samples have expanded the biological data and inference. Moreover, combined liquid and solid biopsies can provide more cohesive spatio-temporal insights into the tumour biology, further guiding precision oncology. The breakthroughs in omics technologies for both these sample types enable collection of high-dimensional information by using only a fraction of the total clinical sample. The quantified expression datasets along with the clinical features, form the basis of big data in clinical cohorts. However, to identify relevant features from such unstructured, complex, and large volumes of data require new processes for analysis. Thus, the application of artificial intelligence (AI), machine- and deep-learning (ML and DL) approaches for data mining in biology became prevalent[21]. As new modules are implemented for data mining, new biological insights are made, that can be further developed into predictive and prognostic models for clinical decision making, thus, aiding precision medicine. Hence, studies that focus on large datasets or "Big Data" that combine omics with patient metadata, are becoming front-runners to identify patterns that could help in demarcating risk groups that would be require separate clinical interventions[19].

1.4 Thesis overview

The field of precision medicine is relatively young and major strides are being made globally for evaluation of the TIME to improve the use of immunotherapeutic and stratification of patients. This thesis is centralized on a particular subtype of haematological malignancy called mantle cell lymphoma (MCL) with the aim of enhancing biological understanding and develop methods for patient stratification and prognostication. The outcome of treatment in MCL, an aggressive sub-type of B-cell lymphoma, is governed by both genetic aberrations, but also MCL immune microenvironment. Thus, this thesis and the included five papers are focused on exploring the immune microenvironment in MCL with respect to biomarker discovery and patient risk stratification (Figure 2). To achieve this, we have used technologies for biomolecule profiling in serum (non-cellular component of blood) and as well in solid tumour tissue using spatially resolved omics. Downstream bioinformatic workflows have guided selection of prognostically relevant biomarkers and in providing additional insights into MCL-related immune response in the TIME.

Paper I-III explore the humoral component of this lymphoproliferative disease, specifically serum analysis using targeted proteomics. Whilst paper IV-V explore tissue-based spatially-guided proteomics collected from tumour and specific sub-sets of immune cells (Figure 2).

In paper I-II, an antibody-based protein microarray technology (IMMRay™) using affinity proteomics was used for high through-put serum profiling. This assay profiled 371 single-chain variable fragment (scFv) antibodies against 158 unique immuno-regulatory and tumour-secretory serum proteins. In paper III, targeted serum protein expression was measured using sandwich enzyme-linked immunosorbent assay (ELISA). For spatial omics analysis in paper IV-V, GeoMX™ (Nanostring Inc) spatial omics platform was used to profile 63 informative proteins (plus 3 housekeeping proteins, 3 negative controls) collected from multiple cell types as identified by multiplexed immunofluorescence (mIF) staining of MCL tissues (section 3.3.2).

Multiple MCL cohorts with respect diagnosis or at relapse/during treatment, were assessed in the five studies. Paper I and II is centred around relapsed/refractory (R/R) patients belonging to a clinical trial cohort (MCL6-Philemon trial)[22], as biopsies are rarely taken from patients with aggressive relapse. These patients were homogenously treated with a combination of rituximab, ibrutinib and lenalidomide for 12 months, followed by maintenance with ibrutinib and rituximab until disease progression. Paper II was a longitudinal follow-up of paper I, wherein the previous samples collected at baseline or pre-treatment were compared to samples collected after 12 weeks of treatment (on-treatment)[23]. Paper III explored serum samples collected from both diagnostic and relapsed MCL patients from the MCL6-Philemon cohort as well as heterogenous cohorts collected from biobanks of Vital Freezing of Lymphoma Cells (VIOLA), the Uppsala-Umeå Comprehensive Cancer Consortium (U-CAN)[24]. Four patients from U-CAN were at the relapse stage and further combined with the Philemon samples for analysis in this paper. Paper IV-V is based on diagnostic MCL tissues collected from a population-based cohort of lymphomas (B-cell lymphomas in Southern Sweden, BLISS). Cores collected from this archival formalin-fixed paraffin-embedded (FFPE) tissues were transferred into tissue microarray (TMA) blocks, which was then sectioned, stained, and used for spatial omics data collection (elaborated further in section 3.3.2).

The main objectives for all study were identifying patterns of biological variation in relation to clinicopathological features such as patient outcome based on overall survival. This enabled identification of novel biomarkers and development of patient stratification tools. Specific aims of each paper have been elucidated further below, in detail.

The aim of paper I, was to investigate whether serum-based biomarkers could be identified that were correlated with overall survival in R/R MCL and whether new methods of patient stratification in relation to patient outcome and progression could be developed[25]. This was a necessity since, methods of risk stratification such as the MCL prognostic index (MIPI), have previously been developed in diagnostic MCL. Moreover, MIPI combines information only on clinical features and newer models that integrate biological information have also been suggested only in diagnostic cohorts (section 2.3.1). Therefore, prognosis at relapse is a lacking in MCL. Thus, developing prognostic tools with biological information based on blood/serum is a clinical need for R/R MCL. By using bioinformatic workflows, we were able to identify serum-based biomarkers in relation to overall survival which led to the development of a novel biomarker signature. We further combine it with MIPI to create a new prognostic index in R/R MCL that could be used for patient risk stratification.

The aim of paper II was to study how longitudinal profiles based on serum proteomics vary with respect to treatment. Thus, we assessed if any serum proteins were differentially regulated at the two timepoints collected (stated above) and the validity of using serum expression velocity i.e., change in serum protein expression between on-treatment and pre-treatment. Such an analysis can be important not just for evaluation treatment response, but also the temporal variation in patient biology. Our analysis revealed differential response in serum protein in relation to underlying genetic aberrations. We also show that using serum expression velocity of change, rather than absolute expression values at the two timepoints, can account for inter patient heterogeneity at baseline. Additionally, we were able to identify biomarkers in relation to minimal residual disease (MRD) and time to progression by categorizing patients into early and late progressors.

The aim of paper III was to evaluate the abundance and prognostic role of soluble CD163 (sCD163), a cell surface marker for M2-like macrophages. The presence of CD163+ macrophages in tissue has been shown to be correlated with poor outcome in primary MCL[26,27]. Thus, the goal was to assess whether tissue-based information is translatable to serum-based measurement, and whether it could be used instead as a complementary tool. This would also accelerate clinical implementation of CD163 as prognostic indicator. Our analysis revealed that the measure of sCD163 in sera was concurrent with tissue-based information and can be used as a surrogate biomarker for both diagnostic and relapsed MCL. Moreover, the study further suggests a robust cut-off for sCD163 levels for tentative clinical use at diagnosis as well as relapse.

While CD163+ macrophage presence is associated to poor outcome, their impact and the mechanistic modulation of the TIME is not clear. Thus, the aim of paper IV was to explore the role of CD163+ in altering the TIME. Paper IV investigates

whether the spatial localization of macrophages close or more distant to tumour cells affects the cell type specific phenotype and characteristics. In addition, the impact of the presence of macrophages in the MCL microenvironment on tumour and CD3+ T-cells was also investigated. Using spatially-guided proteome GeoMX™ data collected from CD20+ tumour cells, M2-like CD163+ macrophages, and CD3+ bulk T-cells, we show spatial localization in tissues can alter phenotypic characteristics of the cell type of interest, which confirms the necessity of using spatially resolved omics technologies for tissue-based analysis. Additionally, we demonstrate how the presence of a specific immune subset within tumour can alter the phenotypic profile of MCL tumour cells and T-cells in the adjoining neighbourhood. Thus, this analysis provided some context into the crosstalk between CD163 and CD20+ tumour cells and its contribution to immunosuppression.

The aim of paper V was to understand how composition and infiltration levels of different T-cells subtypes, are associated with overall survival and phenotypic variation in diagnostic MCL. Image analysis was used to extract cell metrics from mIF images, for defining infiltration based on cell frequency. Thus, this paper is partly technical wherein segmentation models (Cellpose and Stardist) were compared with respect to manually annotated data and a workflow for mIF image analysis was optimized. For biologically evaluation, we then combined the image derived metrics with GeoMx™ proteomics, to investigate modulations of the TIME in relation to cell frequency. In this study, four T-cells subsets were analysed by combinatorial expression of CD3, CD8 and CD57. The combination of CD8 and CD3 differentiates between cytotoxic and helper T-cells and CD57 differentiates between active and terminally differentiated T-cells[28–32]. Thus, following subtypes were defined– $T_{C,57-}$ (CD57- cytotoxic T cells), $T_{C,57+}$ (CD57+ cytotoxic T cells), $T_{H,57-}$ (CD57- Helper T cells), $T_{H,57+}$ (CD57+ Helper T cells). By using such a combined analysis, we show that T-cell infiltration is associated with overall survival in MCL. Varying infiltration levels can modulate tumour microenvironment with respect to T-cells and CD20+ tumour cells and were able to characterize immunosuppressive microenvironment with respect to level of T-cell infiltration. The complementary use of the two methodologies for investigating the TIME reveals that the frequencies and characteristics of single cell types cannot be studied as separate units, as the functionality of both tumour and immune cells adapt based on the composition of the TIME.

In conclusion, this thesis is centred around investigation of immune microenvironment in MCL via application of high-throughput technologies, using both liquid and lymphoid tissue biopsies. The overall aim was development of improved companion prognostic/predictive tools, models of risk stratification, and exploration of the TIME in MCL. The key findings and the scientific overview of

the five papers that comprise this thesis is elaborated in detail chapter five (section 5.1).

Chapter two focuses on MCL broadly, introducing the current known biological landscape, prognostics, and treatment options available. Chapter three covers the development of high-throughput technologies that propelled omics-exploration in biological sciences with a focus on liquid proteomics and spatially-guided omics and their applicability for biomarker discovery and companion diagnostic/prognostic tools. Chapter four describes the bioinformatic modules and the challenges associated with data pre-processing and data processing using ML and DL algorithms for evaluation of complex big data generated from clinical cohorts, with emphasis on methodologies used in the five included papers. Chapter five discusses scientific summaries, as stated above. It also provides an outlook towards the future use of information from the tissue and serum-based MCL microenvironment, based on the novel data generated from paper I-V.

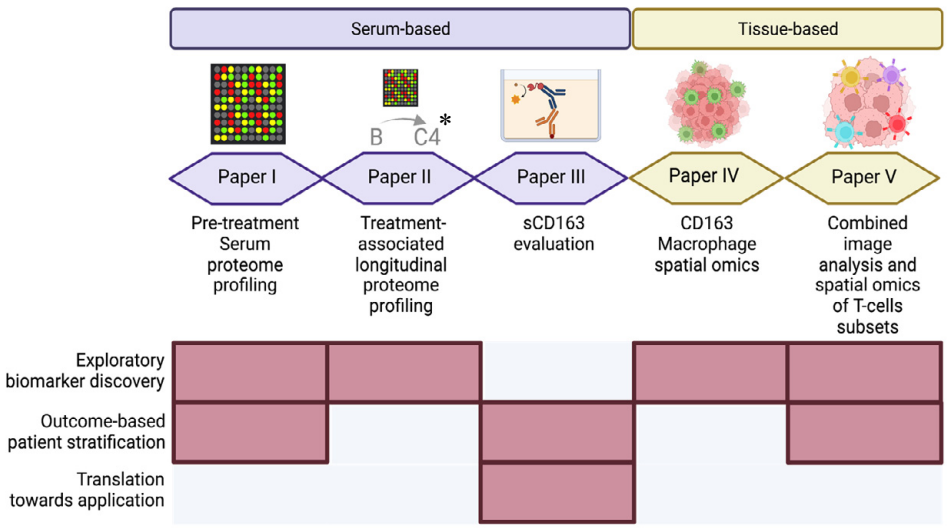

Figure 2: Overview of the five papers included in this thesis. Three of the five papers are based on serum protein analysis and the remaining two papers are based on tissue-based exploration using spatial omics. The goal of these papers can largely be attributed to exploratory biomarker discovery, outcome-based patient stratification using identified biomarkers and potentially translation into application for clinical use. Figure created using Biorender. *B = Baseline/Pre-treatment samples, C4=Samples collected on-treatment at cycle 4 after 12 weeks of treatment.

Chapter 2: Mantle Cell Lymphoma

2.1 Lymphomas – lymphoproliferative disease

Lymphomas include a complex group of haematological malignancies characterized by malignant growth of B, T and NK lymphocytes. The underlying complex mechanisms of lymphomagenesis can primarily be attributed to genetic aberrations and dysregulation in cell differentiation, growth, and death. B and T cells are particularly susceptible to malignant transformation, as their development includes complex genetic rearrangements, required for the development of the diverse immune repertoire as well maturation process that require migration to different tissue sites[33,34]. This is especially true for B cells, which undergo somatic hypermutation and immunoglobulin class switching and are thus, prone to accumulate oncogenic genetic alterations. Consequently, most lymphomas (~90%) are of B-cell origin[35,36].

Broadly, lymphoma can be categorized as Hodgkin (HL) and non-Hodgkin B- and T-cell lymphoma (NHL)[37]. However, differences in destabilizing genetic alterations, overall genomic landscape, phenotypic characterization, and clinical manifestation have further led to the recognition of more than 70 subtypes of haematolymphoid tumours, as reported in the recent update by the World Health Organization (WHO-HAEM5)[38]. Lymphomas are systemically treated, even in cases with early diagnosis showing localized tissue presentation.

Approximately, ~10 million deaths as a consequence of cancer were reported by the GLOBOCAN (Global cancer observatory) in 2020, and 2.8% of those deaths were caused by lymphomas[1]. This corresponds to 600,000 new cases of lymphomas, yearly[1]. There is also a geographical bias in the distribution of some subtypes which indicates that environmental and lifestyle factors, viral and genetic make-up, race, family history; may potentially influence lymphomagenesis[35,39–42]. For example, follicular lymphoma (FL) is more common in western countries, in comparison to Eastern Asian countries that see higher incidences of diffuse aggressive B-cell, peripheral and extra-nodal NK or T cell lymphoma[35,43].

Most B-cell lymphomas manifest as lymphadenopathy[44]. Subtypes of B-cell lymphomas have been initially characterised by phenotypic features corresponding to normal B cells in that localization, as MCL. However, technological

advancements particularly with sequencing and imaging, have led to sub-classification and identification of more specific subtypes based on genetical, immunological and molecular markers[43]. These subtypes exhibit a wide range of diverse indolent to aggressive behaviour that leads to varying clinical outcomes[44]. Some examples of indolent B-cell lymphoma are follicular lymphoma (FL), marginal zone lymphoma, small-cell lymphocytic lymphoma. More aggressive subtypes include diffuse-large B-cell lymphoma (DLBCL), mantle cell lymphoma and Burkitt lymphoma[45]. DLBCL is the most common aggressive subtype accounting for nearly 30-40% of B-cell lymphomas[44,46].

2.2 Mantle cell lymphoma

Mantle Cell Lymphoma (MCL) is an aggressive subtype of B-cell lymphoma with a median survival of 5-7 years[47]. It accounts for an average of 6% of lymphomas incidences in Western countries with a yearly incidence rate of 1-2/100,000 people[48–51]. However, it has been shown that the MCL incidences have been increasing with nearly 3,320 new cases reported in the United States in 2016[52]. The median age at diagnosis is 70 years with a striking 3:1 male pre-dominance[49,50,53].

The first description of the disease that later would be named MCL, was made in 1982 by Weisenburger *et al.*[54,55]. In 1994, it was confirmed as a separate subtype of lymphoma and the term "mantle cell lymphoma" was coined, referring to the location of the malignant B-cells in the mantle zone around the germinal centre of a lymphoid follicle[55,56]. The main site of presentation are lymph nodes and bone marrow, but secondary sites include Waldayer's ring, the central nervous system, and the gastro-intestinal tract[55–58].

In comparison to other lymphomas, MCL is associated with poor survival, historically. However, with developments in therapeutic care, the 5-year overall survival has improved from ~68% to ~82%[59]. Such improvement in survival rates was primarily seen due to implementation of immunotherapy, Rituximab, a monoclonal anti-CD20 antibody in 1997[18,60,61]. The clinical course is still highly heterogenous with several non-responsive cases already seen in front-line treatment. Patients that do respond to initial treatment have high-risk of aggressive relapses, with no fixed standard of care at this stage[62].

MCL patients most often show disseminated disease even at diagnosis with widespread bone marrow involvement, lymphadenopathy, and splenomegaly[63]. The clinical course of MCL is highly variable and depends on several key factors including age, as younger and fit patients have a better outcome primarily due to less age-related comorbidities and show less treatment-related complications. Even

patients who initially benefit from front line treatment, often have aggressive relapses with shorter OS and sometimes with chemoresistance[64]. Therefore, continuous efforts have been made to develop new treatment strategies that can improve the OS and prognostic factors that can help in monitoring disease progression.

2.2.1 Pathogenesis and mutational landscape

MCL pathogenesis is primarily characterised by t(11;14)(q13;q32) chromosomal translocation during V(D)J recombination of Ig gene in the immature bone marrow B-cell[65,66]. Such translocation can also occur in mature B-cells via activation-induced cytidine deaminase (AID)[67]. This translocation is in proximity to the Ig heavy chain sequence causing an upregulation in *CCND1* gene which subsequently leads to an overexpression of cyclin D1, a proto-oncogene involved in the dysregulation of various cancers[66,68,69]. Cyclin D1 is an important player in cell cycle transition from G1 to S phase as well as a transcriptional modulator[66,68].

Initially, it was considered that the above-mentioned specific translocation and overexpression of cyclin D1 were the primary hallmark of MCL, differentiating it from other B-cell NHL. However, a small subset of MCL cases (<1%) lack cyclin D1 overexpression but can alternatively have upregulated cyclin D2 and D3[70]. Furthermore, overexpression of SOX11, a neural transcription factor, was seen in the majority (90%) of MCL cases and especially with respect to cyclin D1-negative MCL[71–73]. Functionally, overexpression of SOX11 causes suppression in BCL6 transcription and increase in BCR signalling and PAX5[74–76]. However, the prognostic impact of SOX11 has been controversial and is dependent on the MCL inclusion criteria, as patients with indolent and leukemic MCL often lack SOX11[73,77–83].

One of the most important prognostic markers in MCL is genetic mutation or deletion of *TP53*[84]. But the mutational landscape of MCL is heterogenous with mutations reported in *ATM*, *NOTCH1*, *KMT2D*, *NOTCH2*, *UBR5*, *BIRC3, DNAH5, SF3B1, CHK2* etc.[55,85–88]. Additionally, MCL tumour cells typically express B-cell antigens (CD19 and CD20) and T-cell associated antigen CD5 as well as FMC7 and BCL2 but are commonly negative for CD10 and BCL6[58]. Several signalling pathways have been identified that are crucial to pathogenesis of MCL such as PI3K/AKT and NF-κB signalling pathways[89–91]. Epigenetic dysregulation may enhance proliferation in MCL and are frequently caused by underlying mutations in key genes such as *KMT2D*, *NSD2, KMT2C* or *UBR5*[89].

2.2.2 Histological and morphological variants of MCL

MCL exhibits a spectrum of histological and morphological variants. When localized to secondary lymphoid structures such as lymph nodes, MCL grow in either a nodular or diffused pattern. Based on the cytology of cells, they can further be classified as classical, blastoid or pleomorphic variant. Patients may present with non-nodal leukemic disease, although blood involvement is commonly found during progression of localized disease. Nodal MCL is defined by SOX11+ and IGHV non- or minimally mutated B cells. Indolent non-nodal leukemic variant, seen in 10-20% MCL patients, is characterized by IGHV-hypermutated and SOX11-negative tumour cells[50,55].

Classical or conventional MCL is a nodal subset and is typically exhibited in the lymph node with some extra-nodal presentation[55]. Accumulation and augmentation of secondary mutations, pathway abnormalities and functional dysregulation, give rise to even more aggressive blastoid (10%) or pleomorphic (5%) forms[92,93]. The blastoid variant is associated with high-proliferation index (Ki-67), *TP53* mutations, high c-myc expression and diffuse growth pattern[92,94]. Although, classical subtype, which is the most frequent subtype, can transform into blastoid morphology, most patients with blastoid MCL morphology present with the aggressive variant at diagnosis[56,92]. Pleomorphic MCL are composed of larger cells resembling diffuse large B-cell lymphomas, and frequently show tetraploidy[56,92].

2.3 Mantle cell lymphoma Prognostics

2.3.1 Mantle cell lymphoma prognostic index (MIPI) and variants

The need for prognostic guidance in MCL led to the development of the MCL international prognostic index (MIPI), in 2008, consolidating age, ECOG performance status, leukocyte count and serum lactate dehydrogenase LDH levels in a risk score that stratifies primary MCL patients in low-, intermediate- and high-risk groups[95]. In 2016, a study of several combined MCL cohorts (n=958), the five-year OS with respect to MIPI was reported as 83%, 63% and 34%, in low-, intermediate- and high- risk groups, respectively[96]. The MIPI-b was later developed by addition of Ki-67%, measuring proliferation, to MIPI[97].

MIPI score, and stratification was developed using diagnostic MCL cohorts and has since been validated in multiple cohorts, including different treatment regimens. However there have been controversial reports of its applicability in different cohorts. For example, the study by Shah *et al.*, 2008, reported MIPI being non-prognostic in

their homogenously treated cohort with rituximab plus fractionated cyclophosphamide, vincristine, doxorubicin, and dexamethasone (R-hyper-CVAD) [98].

Hence, to address biological heterogeneity, several variations of MIPI have been explored by different studies where the original score has been improved by addition of biological markers with prognostic value. Some examples are MIPI with p53 and SOX11 expression[82], MIPI-g with addition of *KMT2D* and *TP53* aberration[99], MIPI-B-miR by adding microRNA-18b-expression[100], R-MIPI including bone marrow involvement and serum albumin[101,102], and MIPI/MIPI-c improved with *TP53* and *WHSC1*[103], and revised MIPI stratification into five categories by addition of B-symptoms and ECOG[104]. These variants however lack cohort validation and have thus not been explored in clinical use.

Moreover, prognostic indices specific for R/R MCL have not been investigated and there is a general lack of knowledge for risk stratification in R/R MCL. In paper I, we show that MIPI stratification as well as *TP53* mutational status and proliferation index was not prognostic for R/R cohort and suggested an improved index (MIPI-relapsed immune signature)[25].

2.3.2 Clinicopathological prognostic features

Prognostic factors that are currently in clinical use include age, IGHV mutation, proliferation rate (Ki67), morphological subtype and *TP53* mutational status[50]. Blastoid and pleomorphic variants are associated with worse survival in comparison to classical and leukemic non-nodal variants[92]. Bulky disease and tumour burden are also predictive markers for survival[64,105]. Minimal residual disease (MRD) negativity after treatment has shown to be associated with long time to progression. A recent meta-analysis, showed that MRD-negative patients have improved progression free survival (PFS) and overall survival (OS) in comparison to MRD-positive patients, indicating that MRD could be a useful factor for predicting relapse[106]. In paper II, we confirm that positive MRD status is associated with increased serum levels of TGF-β1 and can be used as a surrogate marker for R/R MCL.

2.3.3 Molecular biomarkers with prognostic impact

Several mutations have shown to have a mutational impact such as *TP53* mutation or deletion, *NOTCH1* mutation, *KMT2D* mutation, MYC translocation or overexpression, high Ki-67 proliferation index in patients is associated with worse survival[83,87,99,107–109]. Of these, only *TP53* and Ki-67 are currently used in clinics.

To improve the biological understanding and patient stratification for optimized treatment - a plethora of studies have been conducted with the aim to identify molecular sub-groups of MCL. Most studies have been focused on oncogenic features of the tumour cells themselves, while more recent studies also focus on the tissue and blood-based microenvironment.

*2.3.3.1 **Blood-based prognostic factors***

Liquid biopsies may hold value for more non-invasive repetitive sampling during treatment or at relapse. Although serum-based analysis has been limited in MCL, a few studies have identified prognostic factors, tentatively useful for treatment monitoring. Such factors include IL-2Rα, IL-8, MIP-1β and Beta-2-Microglobulin[110,111]. Recently, it was shown that detectable presence of circulating tumour DNA (ctDNA) in serum correlated with poor outcome, and that molecular relapse often preceded clinical relapse when circulating tumour (ctDNA) was used for monitoring[112,113]. To address the need of developing serum-based prognostic tools, paper I-III explore serum-based soluble proteins in primary (paper III) and R/R MCL (paper I-III).

*2.3.3.2 **Tissue-based and histology related prognostic factors***

Positivity for p53, SOX11 and Pax5 protein were identified as potential subtypes of high-risk MCL and associated with negative PFS and OS[82,114,115]. Other non-validated histology markers include Bcl-2 Interacting Mediator of cell death (BIM) (higher expression being associated with higher OS)[116] and T-cell leukemia/lymphoma protein 1 (TCL1) (an oncoprotein with low expression associated with short OS)[117].

Patients with complex karyotypes have shown to be associated with inferior outcomes[118–120]. Controversial reports have also been suggested for using image derived metrics ([(18)F] fluorodeoxyglucose positron-emission-tomography (FDG-PET)) as a prognostic tool to assess disease staging and treatment-associated response[121–125].

*2.3.3.3 **Tumour-immune microenvironment related prognostic factors***

Therapeutically, targeting the immune system for tumour clearance has been of relevance in the last decade. This was highlighted by a survey conducted between 2017 and 2019, that reported 91% (2,030 to 3,876) increase of the number of active immuno-oncology (IO) drugs in development over two years[126]. Furthermore, immunotherapy, a treatment methodology that deals with enabling the immune system to fight cancer has become the frontrunner for therapeutic discovery. As elucidated in chapter 1.2, the composition of the TIME and the crosstalk between

the tumour and immune cells contribute to the patient's biological heterogeneity and thus, studying the TIME for precision prognostics is essential.

Immune cell-based studies are also limited in MCL with very few explorations of the TIME. However, some studies have reported immune cell counts and ratios being prognostic. For example, the presence of CD163+ macrophages in tissues were shown to be correlated with inferior outcome by us and others[26,27]. Nygren *et al.*, 2014, reported that total CD3, as well as CD8 and CD4 T-cells were higher in indolent MCL in comparison to aggressive histology and that CD4/CD8 T cell ratio was independently associated to survival, thus, suggesting its use for prognosis[127]. Similar observations were reported by Zhang *et al.*, 2016, wherein they showed that high monocyte count, low level of CD4+ T cells, and low CD4/CD8 T cell ratio were associated with unfavourable outcome[128]. Low absolute NK cell count and high ratio of FOXP3+ Treg cells to CD4 were also a predictor of inferior OS[129,130]. Higher absolute monocyte count (AMC) was shown to be correlated with poor survival, in contrast to CD68+ and CD163+ expression having no impact in this study cohort[131]. The prognostic impact of elevated AMC was validated with several studies[132–135]. Sahin *et al*., 2019, also found neutrophil/lymphocyte and platelet/lymphocyte ratio associated with increased risk of progression, thus suggesting that simple complete blood count (CBC) counts could be used for MCL prognosis[134]. It also important to note that in many models, age has not been accounted for and we know that age negatively affects anti-tumour immune response[136].

2.3.3.4 ***Prognostic indices and molecular signatures***

New prognostics indices and signatures have also been developed independent of MIPI. In Lv *et al.*, 2022, nomogram and immune-related prognostic signature (IRPI) was developed using CD4+ T cell count < 26.7%, CD8+>44.2%, beta-2-Microglobulin levels, platelet count and B symptoms which were all independent predictors of OS[137]. RNA-sequencing of circular RNAs was used to develop a 40-plex circSCORE that was prognostically significant to predict time-to-progression (TTP) and lymphoma-specific survival (LSS)[138]. A six gene signature (*AKT3, BCL2, BTK, CD79B, PIK3CD*, and *SYK*) for B-cell receptor pathway was found to be independent predictor in MCL[139]. The MCL35 gene signature of 17-genes was developed using quantification of RNA expression in MCL patients and further validated in randomized trails from the European MCL network[140,141]. This signature with or without MIPI/MIPI-c was prognostically significant for both OS and PFS; and identified a subset of patients with high-risk after the intensive first line of treatment[142]. The prognostic capability of MCL35 and outperforming simplified MIPI in R/R MCL was also recently validated[143].

2.4 Treatment strategies of MCL

Research advances and subsequent adaptation of treatment has revolutionized outcome for patients afflicted by MCL. Goy, 2021 has summarized three key reasons that have led to improved survival and low progression; 1) front line treatment as dose-intensive therapy followed by autologous stem cell transplantation (ASCT), 2) novel therapies that were developed in the last decade, 3) identification of prognostic factors that helped in understanding the clinical and biological heterogeneity in MCL[64]. However, response to treatment is highly variable and at relapse MCL has a dismal outcome as further described below.

2.4.1 Current treatment landscape in MCL

Traditionally, the choice of MCL treatment selection has been based on age, the ability to tolerate intensive therapies, and underlying comorbidities. Primary treatment was initially based on chemotherapy cocktail CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisone). Chemotherapeutic approaches have major side effects causing suffering and severe immune suppression for the patients; therefore, efforts have been made to move to chemotherapy-free regimens with particular focus on immunotherapy. In early 2000s, rituximab, a monoclonal anti-CD20 antibody, was introduced, which led to major improvement in OS. Thus, CHOP was exchanged for R-CHOP (Rituximab combined with CHOP) in frontline treatment. The current treatment standard for young and fit primary MCL patients, which was defined post the MCL Younger trial, now consists of high-dose cytarabine-containing regimen with rituximab followed by consolidation with ASCT if necessary, and rituximab maintenance[144–146].

In elderly, unfit patients that are ineligible for intensive cytarabine combination therapy, rituximab combined with less intensive chemo regimens are suggested. This most commonly includes combinations of rituximab with bendamustine and CHOP[64,147]. The introduction of Bruton's tyrosine kinase (BTK)-inhibitor ibrutinib in 2012 and the subsequent variants such acalabrutinib, zanubrutinib, are also being added to standard regimen[147,148]. Primary results from phase III SHINE trial, which investigates the combination of ibrutinib with BR (bendamustine and rituximab) in elderly unfit patients showed significant improvement in PFS in comparison to standard BR[149]. Currently, the National Comprehensive Cancer Network (NCCN) currently describes 17 possible therapeutic options for diagnostic and relapsed MCL[64].

2.4.1.1 Treatment alternatives of R/R MCL

BTK inhibitors have shown promising results in R/R patients[148]. Other options are proteosome inhibitor bortezomib, immunomodulator lenalidomide and BCL2 inhibitor venetoclax[150]. Rituximab is also a primary choice in combination with other mentioned drugs for R/R MCL. The MCL6-Philemon cohort used in paper I-III is a homogenously treated cohort where R/R patients were given a monthly dose combination of rituximab, ibrutinib and lenalidomide for a year, followed by maintenance of ibrutinib and rituximab until disease progression.

2.4.2 Alternative therapies

The current standard of care is still dependent on age, comorbidities, and ability to cope with intensive regimen. New drug possibilities are mostly tested in high-risk or R/R patients where current treatment alternatives have failed. Several new drugs, either stand-alone or in combination with previous options, are in different phases of clinical development/testing. These include PI3K inhibitors (Idelalisib, Copanlisib etc.), mTOR inhibitors (Eg: temsirolimus), CDK 4/6 inhibitors, antibody-drug conjugates - a new class of chemo-immunotherapy (wherein tumour-specific antibody selectively binds to specific antigens which is followed by the release of cytotoxic agents, Eg – polatuzumab vedotin), monoclonal bi-specific T-cell engagers (BiTEs have two binding sites for different antigens), epigenetic agents (vorinostat, cladribine, tazemetostat)[151]. Particularly, checkpoint inhibitors against primary MCL are of particular interest as they would be able to contribute to the goal of a chemo-free first line treatment[91]. Additionally, covalent BTKi such as ibrutinib and acalabrutinib have adverse events due to off-target inhibitions, therefore, non-covalent BTKi are under investigation[152]. One example is Pirtobrutinib, which is being tested against covalent BTKi's in a phase-III trial for heavily pre-treated MCL high-risk patients[150].

Chimeric Antigen Receptor (CAR)-T cell therapy in MCL is of particular interest. In a phase-2 ZUMA-2 trial with KTE-X19 (anti-CD19 CAR-T) in R/R MCL with previous BTKi intervention showed an outcome of 93% patients achieving ORR and 67% receiving complete response. Post one year, 57% of 60 patients were in remission[153]. With such positive outcome, FDA gave fast-track approval for KTE-X19 for treatment of R/R MCL in US in 2020[150]. In a comparative study between the current standard of care with KTE-X19 in Europe for 288 R/R MCL suggested that significant improvement in survival is possible with CAR-T cell therapy[154].

2.4.3 Risk-adapted therapy - A potential future for MCL

To move into risk adapted therapy, prognostic factors need to be applied at primary diagnostic setting. At this moment several suggestions for possible risk-based treatment selection have been suggested mostly based on age, *TP53* mutational status, and MIPI/MIPI-c risk groups. Although, there are not many trials or clinical evidence that suggest that such strategies would be successful, new trials are being designed based on certain risk inclusions. For example, the TARMAC trial in which R/R patients with *TP53* mutations and no response to previous standard therapy, are being treated with a combination of CD19 CAR-T (Tisagenlecleucel) and ibrutinib (NCT04234061)[155]. The ECOG-ACRIN EA4151 (NCT03267433), a phase-III randomized trial (under recruitment) for untreated MCL patients wherein patients are given standard chemo-immunotherapy. Patients who have achieved complete remission and tested MRD-negative then undergo ASCT plus rituximab maintenance. This is compared with the remaining group that only received rituximab maintenance[156].

Although we are still in the early phase to integrate risk-adaptation for clinical decision making and far from personalized medicine, the field is moving in the right direction. It is certain that high-risk patients, including elderly/frail and mutational high-risk, have poor outcome. Therefore, such groups require separate, tailored interventions. The efficacy of new treatment options and current trials for R/R that are under investigation. Especially CAR-T therapy would provide more solutions for high-risk patients. Furthermore, the simultaneous development of superior methods for patient stratification using biological information at diagnosis is essential. Such companion prognostic tools must be financially feasible and clinically applicable.

Chapter 3: Biomarker discovery using serum proteomics and spatial-omics

3.1 Evolution of bio-data science

The field of life sciences saw a huge revolution when sequencing technologies were invented in the 1970s. The technology break-through led to the decoding of the human genome by the early 2000s in the Human Genome Project (HGP, 1990-2003)[157]. Sequencing the human genome was pivotal for deciphering key functional aspects of cell and disease regulation. This was especially true for cancer, which is a disease primarily seen as product of genetic aberrations that alter the normal cell function. However, it soon became clear that the genetic make-up alone, is not sufficient to explain the complex and intricate biological mechanisms that govern malignant diseases[158,159].

Therefore, complementary companion tools were developed to study different molecular layers, which beyond genetics, include the transcriptome, proteome, and the epigenome. Additional advancements in methodologies allowed simultaneous measurements of multiple biomolecules from systems of interest. Thus, "omics" technologies were created that provide a snapshot of biology through high-throughput data collection[160]. As the influence of various molecular layers became clear, omics technologies expanded further from genomics (study of genes) into transcriptomics (study of gene transcripts i.e., the mRNA), proteomics (study of proteins), lipidomics (study of lipids), epigenomics (study of DNA modifications that alter gene activity) and metabolomics (study of low molecular weight metabolites)[160,161]. Proteomics can be particularly clinically vital as they represent the functional translation of the genome[162]. Together, these layers provide extensive disease-related biological information that have transformed medical science, thus becoming essential for the development of precision medicine.

Omics technologies have reformed clinical discovery wherein a common aim has been to correlate high-throughput biological data to clinical metadata of interest particularly disease progression and patient outcome[160]. Such discoveries have been essential for developing biomarkers ("biological markers") that provide critical

insights for patients. Additionally, such analysis has led to the development of omics-based tests that in many cases have been established for clinical use[160]. An example is the Oncotype DX 21-gene assay (Genomic Health, Inc., Redwood City, CA, USA) for prediction of breast cancer recurrence, that in parallel aids in treatment stratification by identifying patients likely to benefit from chemotherapy[163].

Furthermore, omics-technologies have enabled the capture of large datasets from single experiments, thus, catapulting the use of data science for multi-parametric analysis of biological complex data, along-with biostatistics. There are several challenges that are associated with such analysis, some of which have been explored in chapter 4, which is focused on computational strategies for exploring omics data. Here, the following sub-chapters will focus on biomarker discovery for clinical decision making and newly developed omics technologies particularly for proteome exploration of liquid biopsy and tissue-based spatial-omics.

3.2 Biomarker Discovery

The National Institutes of Health (NIH) defined the term "biomarker" as "*a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention*"[164]. In 2015, a collaborative approach by the FDA and NIH helped in establishing an active focused working group – the Biomarkers, Endpoints, and other Tools (BEST) resource that provides an updated glossary, to ensure clear terminologies are maintained[165]. In oncology, biomarkers define a large category of biomolecules that show altered behaviour in cancer patients in comparison to the healthy population. The utilization of such biomarkers enables the development of companion diagnostic and prognostic tools and is an essential step for translating biological insight to medical technology for improving clinical care[166].

3.2.1 Functional classes of biomarkers

Biomarkers can be broadly grouped into several non-exclusive categories – diagnostic/screening, prognostic, predictive, risk/predisposition, response, and monitoring[161,167]. Diagnostic biomarkers refer to biomolecules that distinguish the type of patient disease or disorder from similar conditions or healthy individuals. It also includes markers that provide additional classification on the cancer sub-types[167]. Cancer type was first classified based on the initial detection of affected organ. However, when the number of subtypes and their genetic heterogeneity were

revealed, molecular and imaging-based classification for clinical diagnosis have been further refined[167]. An example of a diagnostic biomarker would be the use of prostate-specific antigen (PSA) for the screening of prostate cancer in elderly men, which is implemented in Sweden[168,169].

A predisposition or risk biomarker indicates the risk of developing a disease or disorder[161,167]. For example, the most prevalent example of these are genetic mutations such as BReast CAncer genes 1 and 2 (BRCA1/2) mutations that predispose carriers to an increased risk of developing breast and ovarian cancer[170–172]. Response or surveillance biomarkers relate to factors which change with respect to treatment and are used to assess treatment effect. Presently, radiographic imaging such as PET or CT have been used to assess tumour burden as response to treatment. However, circulating tumour nucleic acids (ctDNA) and circulating microRNAs have also been suggested as a measure of tumour burden surveillance[173]. In general, a lot of emphasis has been on identifying liquid biopsy-based detection of biomolecules and this is due to that blood collection is an easier, less invasive sampling method which allows for sequential sampling during treatment as previously mentioned[174].

A prognostic biomarker highlights the possible disease course and likelihood of a clinical event (survival, progression etc.)[161,167]. In comparison, a predictive biomarker differentiates patients as responder's vs non-responders to a treatment. For example, Xie *et al.*, 2019, highlighted the need of differentiating the role of PD-1/PD-L1 in prognosis and prediction for lymphoma[175]. Several studies have highlighted the importance of PD-1/PD-L1 expression in tumour for lymphoma patient outcome[175]. Furthermore, the soluble quantification of PD-L1 was suggested as a potential predictive biomarker dissociating patients that might respond differently to immune checkpoint therapy in DLBCL, PTCL and NKTL[175].

Measurements that allow monitoring of disease can encompass many of the other above-mentioned types of biomarkers. However, the concept uniquely describes the possibility to follow patients over time without invasive procedures. The potential use of circulating tumour DNA (ctDNA) has been suggested as a useful marker for monitoring of disease and progression in various cancers and has been validated in a clinical trial recently in MCL[176,177]. It is important to note that a biomarker can be classified in multiple categories. For example, PSA is as an example of diagnostic and monitoring biomarker for prostate cancer. BRCA is a risk biomarker, but also used as a companion diagnostic/predictive biomarker. As we have advanced in technologies, several high throughput omics platforms for targeting diagnostic, prognostic or predictive biomarkers have been developed and the use of data-driven science is now in spotlight.

3.2.2 From bench to bedside – the roadmap of biomarker discovery

There are several steps in the process of biomarker discovery to clinical application which can be broadly labelled as analytic and clinical validity, and clinical utility[178]. The first step of analytical validity involves using experimental evidence from clinical cohorts[179]. Several factors can determine the data quality, including sample-related features such as consistency in sample collection method, sample processing, storage etc., and assay-related factors such as the sensitivity, specificity, and robustness of detection. Data reproducibility is a huge challenge, and several factors can affect this. For example, if the technological platform has not been validated, or the strategy for downstream analysis of the generated data has no consensus[179]. Hence, analytical validation directly refers to performance of the assay with respect to reliability and accuracy of the biomarker measurement[180,181].

Clinical validity is directly related to the ability of a potential biomarker to stratify patients into different categories that show altered response to an intervention[179]. Additionally, experimental reproducibility is a must and is shown by validating initial finding in multiple cohorts of samples. Independent validation in a separate sample set is a necessity, so that the discriminatory performance is not exaggerated[179]. Moreover, global validation implies that the baseline cohort similarity is maintained in validation cohorts[182]. The lack of validation and biased interpretation is a major reason for failure of many biomarkers that do not reach clinical application. Biases seen in any of the features mentioned above must be accounted for as they might cause batch effects (section 4.3.1) and influence the downstream biomarker discovery workflow. For example, in a follow-up SELDI-TOF MS study for validation of serum protein profiling for detection of prostate cancer, the investigators reported biases in sample collection time, between the disease vs normal control samples that affected the initial results[183]. In 2017, a meta-analysis review of 200 reports for ovarian cancer biomarkers highlighted that nearly 70% of publications had misrepresented or overinterpreted the study findings which led to exaggerated biomarker performance, thus biasing the overall results[184].

The third and final step is clinical utility which refers to the application of a potential biomarker for targeting clinically relevant questions. Prognostic and response biomarkers are of particular interest as they directly relate to patient outcome and impact the therapeutic decision making. A biomarker must answer a relevant clinical question and undergone several experimental validations, demonstrating high specificity and sensitivity for classification. Moreover, assays or biomarkers must be easily applicable with clear and established guidelines on usage, that is consistent globally. Improper reporting can lead to incorrect treatment decisions, which can have catastrophic effects on patients, as reviewed previously[179].

3.2.3 Single vs multi-parametric vs multi-omics biomarker signatures

Traditional approaches for biomarker discovery have relied on identifying single analytes to be used as biomarkers. But multiplexed technologies that simultaneously detect multiple biomolecules, have paved way for next generation of multiparametric diagnostics[185]. Biomarker signatures are defined set of biomolecules that when measured together allow individual biomarkers to vary, but more robustly separate groups of samples based on the question posed[186]. Therefore, the process deriving biologically relevant signatures have a structured workflow, wherein multiplexed technologies are applied to clinical cohorts followed by high-dimensional data analysis for discovery. For this purpose, machine learning methods are particularly applicable and can reduce the number of variables to a set of essential markers[186]. In several cases, such signatures have shown to be more accurate in explaining the heterogeneity in patient responses and in developing better patient stratification methods as they have higher discriminatory power[185]. In 2009, the OVA1®, a five protein (CA125, apolipoprotein A1 (ApoA-1), beta-2 microglobulin (B2M), transferrin (TF), and pre-albumin) serum-based signature received approval for clinical use from FDA[187]. This was followed by the second-generation Overa® model which replaces B2M and prealbumin with human epididymis protein 4 (HE4) and follicle stimulating hormone (FSH)[187]. The two scoring models are often used consecutively along with clinical assessment to detect the increasing malignancy risk for ovarian cancer.

Multiplexed technologies for identifying biomarker signatures for different cancers have been developed for several biomolecular layers. Sparano *et al.,* 2015 validated the clinical utility of the 21-gene signature OncotypeX for breast cancer[188]. Several publications have uncovered prognostic signatures that are yet to be validated or clinically implemented[189–191]. In MCL, as mentioned above, the MCL35 assay which is based on the gene expression panel of 17 informative genes, is a prognostic signature panel for MCL which has been validated in several cohorts, but not yet clinically established[140–143].

As the shift from single markers to signatures intensifies, methods are being developed that allow integration of different molecular layers into the same model capturing variation in pathophysiology of complex diseases, thus highlighting the importance of multi-omics signatures in precision medicine[192–195]. Several reports have explored the cancer biology with multi-omics approaches that are prognostically significant[196–199]. However, such data integration requires complex analysis. To our knowledge, no such combined multi-layer molecular signatures have been implemented in the clinic.

3.3 High through-put omics technologies

As stated previously, the two main types of patient samples routinely collected and stored in clinics are tissue biopsies and blood/serum/plasma. The technologies used for the assessment of these types of clinical samples are distinct, as they take advantage of the different tissue properties and the information that can be retrieved. In this thesis all papers, except paper III, deal with the use of high throughput assays to investigate the immunoproteomics and transcriptomics. The expression datasets were acquired along with the associated clinical metadata, which was used to identify molecular signatures associated with patient outcome and disease progression.

3.3.1 Immune-regulatory biomarkers using serum proteomics

Liquid biopsies refer to body effluents such as blood/serum/plasma for biological analysis. They have become important due to advantages associated with repeated sampling, as acquiring patient tissue implies invasive surgical intervention which are costly for the health care system, and cause suffering for the patient. Therefore, identifying blood-based biomarkers carries great value to monitor disease progression and effect of treatment. Human blood can provide information on systemic changes associated with disease, containing information in the form of serum proteins, circulating tumour cells (CTCs), cell free nucleic acids, platelets, and exosomes, whose investigation can highlight pathophysiological features[174]. Therefore, blood/serum-based analysis are particularly useful for monitoring disease progression and treatment response. One of the best and validated examples of this, is perhaps the prostate-specific antigen (PSA) and associated kinetics (PSA velocity (PSAV)) for monitoring prostate cancer[200–203].

Targeted serum proteomics can additionally provide insights into the cancer associated systemic immune-responses and thus be helpful in identifying immune-related biomarkers as altered expression profile can be a result of heterogenous disease states. Particularly in lymphoma, which is a disease that affects the cells of the immune system; exploring the serum immunoproteomics can provide novel insights not captured by other types of analyses.

Technological advancements have led to development of technologies for studying soluble proteins, such as 2D-PAGE, 2D-DIGE, SELDI-ToF-MS, iTRAQ etc.[162]. All methodologies have advantages and disadvantages, however global proteome assays such as mass-spectrometry deal with issues relating to high abundant proteins, which often mask lower abundant and rare molecules. On the other hand, affinity methods were developed in parallel and use binding reagents against target protein antigens, which allow for selective profiling. Typically, these technologies

are based on the use of antibodies as selective binders and are referred to as immunoassays. One of the earliest methods with high selectivity and specificity for protein profiling was ELISA[204]. In paper III, ELISA was used for evaluating the prognostic implication of macrophage associated with soluble CD163 levels in serum.

Traditional immunoassays were limited to single marker analysis, but targeted multiplexed assays have been developed as well. Multiplexed ELISA such as suspension based Luminex™, Cytometric Bead Arrays, Bio–PlexPro™ and planar arrays such as the Mesoscale Discovery Technology Platform (MSD®), Q–Plex™ have been developed[205]. Planar assays further propelled research into development of chip-based antibody-based microarrays for proteome detection. One such method is called IMMray®, which was used for the evaluation of the sera for profiling immune-regulatory and tumour-secretory proteins. This method was used in paper I and II and has been explained in detail below.

3.3.1.1 Antibody-based microarray – IMMray®

In 2002, an in-house (within the Department of Immunotechnology, Lund university) developed affinity-based protein antibody microarray technology was first described[206]. Since then, this technology has been updated both in terms of technique as well as its associated bioinformatic processes and is now commercialized by Immunovia AB. The technology is based on printing small volumes (picolitre scale) of hundreds of recombinant single chain variable fragment (scFv) antibodies in a matrix pattern on a slide. The antibodies are specific to target serum protein antigens and have been generated from a phage display library and are specifically targeting immune-regulatory and tumour-secretory proteins, including cytokines, chemokines, complement components, adhesion molecules, inflammatory and signalling molecules.

By comparing the profile in large cohorts of cases and control samples, clinically relevant biomarker signatures may be identified. Using this technology, various diseases have been explored such as breast cancer[207], pancreatic ductal adenocarcinoma (PDAC)[208–212], prostate cancer[213,214], DLBCL[215,216], pancreatitis[217], systemic lupus erythematosus[214,218,219]. Now, the IMMray® PanCan-D test which measures an 8-plex serum protein signature has been commercialized which can identify stage I and II PDAC based on serum proteome with high specificity and sensitivity[208]. This demonstrates how multiplex blood protein signatures can be condensed, validated, and translated for clinical implementation through combined academic and industrial efforts. In paper I and II, 158 unique proteins are quantified by measurement of multiple epitopes (total 371 single chain variable fragment (scFv)) using the IMMray® assay. The aim was to develop prognostic and responsive biomarker signatures.

3.3.2 Proteomics using spatial omics

The tumour microenvironment, with its own compilation of cell populations, is changing during tumour evolution and can be resembled to evolving ecosystems[220,221]. The application of ecological principles can often be made with respect to species richness, metabolic competition, the presence of cancer ecotones at the boundary of the TME[220]. Therefore, decoding the tumour ecosystem and milieu is pivotal for unravelling the biology. While tumour immunotherapy has been of pivotal interest in lymphoma treatment, there is still considerable diversity in treatment response. It has been suggested that this may be due to the heterogeneity seen in the TIME and its associated variation during different stages of the disease[222,223]. The shift in the tumour-immune profile can be indicative of neoplastic progression, thus deeper analysis of the complex TIME can reveal biomarkers indicative of progression or specific TIME-related patient subgroups. The interplay between the tumour and different immune cells, which define the TIME landscape, is important in understanding immune response to tumour and how the crosstalk can affect tumour progression.

Previously, the methodologies that could be used to study cellular interactions have had major limitations. An established traditional approach is to use flow cytometry for immune cell profiling, but this involves disrupting and disintegrating the tissue and thereby foregoing the spatial architecture[224]. Flow cytometry is also limited in its multiplex possibilities, which was later overcome by omics methods such as bulk or single cell sequencing and mass cytometry etc.[225]. In contrast, immuno-histochemistry (IHC) or immunofluorescent imaging retains the spatial information but are limited to low-plex analysis. Recent technological development has managed to combine several important features of previous platforms. Thus, spatially resolved omics-based profiling provides a more global view of the biology compared to bulk and single-cell sequencing, while keeping the tissues intact and thus retaining the spatial morphology. Such investigations expand the scope of onco-immunology by providing key insights with respect to spatial mechanism of cancer evolution, sub-clonal formation, metastatic process, cell-cell (tumour-cell and tumour-tumour) interactions, immune activation, immune evasion strategies, and therapeutic resistance[220,226,227]. Several platforms now exist for such deep profiling of the tissue such as the Visium Spatial Gene Expression by 10X Genomics, DBit-seq (deterministic barcoding in tissue for spatial omics sequencing), in-situ sequencing, co-detection by indexing (CODEX), Multiplexed error-robust fluorescence in situ hybridization (MERFISH) etc.[220,226,227]. Spatially resolved transcriptomics was awarded the nature method of the year in 2020[228]. In paper IV and V, we explored the immune response in MCL using GeoMx™-digital spatial profiler (DSP), an advanced multiplexed-IF based imaging and preparative (protein and mRNA probes) platform commercialized by Nanostring Technologies Inc.[229,230].

3.3.2.1 GeoMx™ digital spatial profiler (DSP)

The GeoMx™ technology is based on in-situ capturing wherein the sequence that is profiled is the barcode of antibody/oligo and not the target molecule itself. This platform uses up to three (total four with the nuclear marker) fluorescent antibodies to identify phenotypic markers that can be used to choose the cells of interest and guide the region of interest (ROI) selection. Barcoded antibodies/oligos coupled with a photocleavable linker are added and allowed to bind to the target molecules. Using millions of programmable micromirrors, UV light is directed to the specific segment of interest with high resolution and cleaves off the barcodes corresponding to the probes bound in that area. The released barcodes are aspirated and placed in a microtiter well. Barcodes from each segment are collected separately and provide a quantitative estimate of the amount of target molecules in each selected type of segment/cell type. Thus, quantitative information of hundreds of proteins or thousands of transcripts can be collected in parallel. This strategy has already proven to be successful in several key studies. In non-small cell lung cancer (NSCLC), it has been shown that presence of tumour associated CD163+ macrophages was associated with resistance to immunotherapy and outcome, wherein GeoMx™ was used to define differential tumour profiles in responder's vs non-responders [231]. Another GeoMx™ proteomics study on NSCLC identified CD44 expression in the tumour compartment as a novel predictor of PFS, specifically associated with the sensitivity to PD-1 axis inhibition[232]. GeoMx™ has also been combined with other omics technologies, for example in a study based on single cell sequencing and GeoMx™, which explored cell populations in spatial context that were associated with tumour progression in basal cell carcinoma[233]. These exhibit a small selection of studied done using one spatial omics technology. However, as stated above several spatial omics technologies are available commercially or are under-development. Thus, spatial omics is becoming as essential tool in the era of precision medicine.

In MCL, the complex interplay of the tumorigenic B-cells with the other immune cells (T/NK/macrophages/dendritic cells) remains to be characterized in detail which is one of the aims of this thesis. Using GeoMx™, we have studies, CD20+ B-cells, CD163+ macrophages, bulk immune CD3+ cells and sub-population of T-cells by combining CD3, CD8 and CD57 as morphology markers. We have further profiled the proteome in these cell subsets to study the effect of spatial localization and T-cell composition on phenotypic profile of these cell subsets and modulation of the tumour-immune microenvironment.

Chapter 4: Computational strategies for omics analysis

To extract patterns and information from large datasets that profile multiple biomolecules, advances with respect to computational approaches were necessary to deal with the bulk data. Inspiration taken from data science and applying them in life sciences has been an integral part to drive data-driven research forward. However, biological data are subject to different challenges that need to be accounted for, while designing solutions for extracting relevant information. Platform-associated challenges must also be considered in planning analytical workflows. In this thesis, two main datasets were analysed - the serum protein microarray and the spatial omics data. A huge part of the work involved in this thesis was focused on analysing these different data structures and dealing with technical challenges. Some of the challenges and the measures taken to solve those issues as well as the development of computational workflows, specifically used for extracting biologically relevant information, have been defined in greater detail in this chapter.

4.1 Biological data exploration using concepts of data science

In the last decade, the sheer volume of biological data generated globally have expanded with a reported 60 petabytes of genomics data generated in 2020 alone[234]. In a projection for the year of 2025, it has been estimated that between 100 million to 2 billion human genomes would be sequenced, exceeding the data acquired by three other main generators namely astronomy, YouTube and Twitter[235]. Biological data is growing at an exponential rate, therefore updating methodologies to analyse, store and manage such information is critical. Several databases already exist that have been specifically made to storage information for future assessments. Some examples in relation to cancer is The Cancer Genome Atlas (TCGA) (multi-omics data from 33 cancer subtypes), the Gene Expression Omnibus (GEO), International Cancer Genome Consortium (ICGC), Catalogue of Somatic Mutations in Cancer (COSMIC) etc.[236–239].

Clinical metadata can be defined as a compilation of information from several sources such as disease-related clinical data of patients (disease, treatment, outcome etc.), demographic information (gender, age etc.), family history, phenotypic symptoms, comorbidities, experimental analysis such as histopathology, DNA/RNA sequencing, images (PET, MRI etc)[240]. Such data can be both structured (organized in tabular format) which is a combination of categorical, continuous, and ordinal values or unstructured (non-tabular format like free text, narrative notes, images etc.)[240]. Data from clinical trials can be particularly more varied in terms of disease stage, multiple field sites, different clinical settings, time of medication, drug dosage, longitudinal assessment etc., can further add to the heterogeneity[234]. Therefore, it becomes critical to find new ways to analyse such data, but also manage, store, and share sensitive information.

Data science is a term used for extracting relevant features and predictions from big data by using principles of data mining, predictive analysis, and statistics[241]. The use of artificial Intelligence, based on ML, NLP (natural language processing) and DL have been pivotal for pattern recognition and have led to identification of clinically relevant biomarkers and development of intricate patient stratification models for several diseases. In omics analysis, methodologies developed for data science have been applied for identifying patterns and extracting clinically or biologically relevant features.

In our investigations with microarray proteome data collected from clinical trial cohort (paper I-II) and spatial-omics data generated from population-based cohort (paper VI-V), various technical challenges associated with data structures and bioinformatic workflows have been addressed and solved and have been highlighted in the sections below.

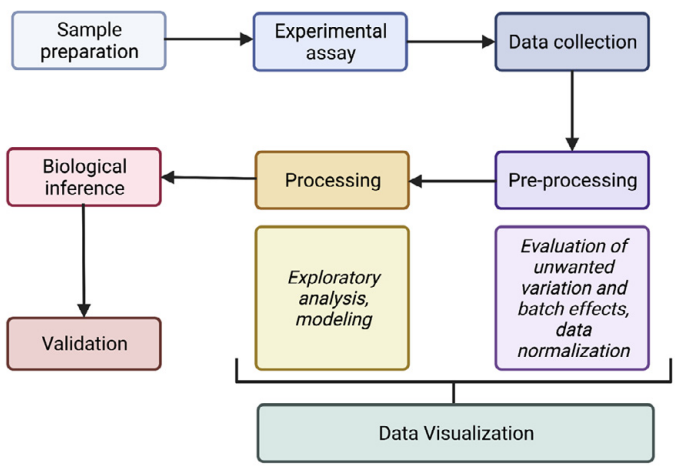

Figure 3: Basic workflow of exploratory data analysis. Samples collected are prepared as per the requirement of the experiment assay wherein features of interest are measured. Collected data is then initially pre-processed to assess for unwanted variations and batch effects as well as data normalization. Removal of redundant features through dimensionality reduction and imputation of missing information is also be performed at this step. Normalized data undergoes exploratory analysis based on the question and objective of the study. For example, applying regression models for extracting relevant features associated with clinical label. The final output or model as generated from this step is inferred to check for biological relevance. Validation of the output is required to determine the robustness of the generated model. Figure created using Biorender.

4.2 Challenges associated with pre-processing in complex biological big data

A typical workflow for exploration of omics datasets consists of several steps (Figure 3). Prior to data analysis, this includes sample collection and selection, sample pre-processing (if required) and the specific technical workflow. Once data is collected, the data analysis phase starts with data pre-processing which includes measuring unwanted variation, identify batch effects, imputing missing data, filtering of redundant or poor-quality data to reduce the dimensions, scaling, and normalization. This is followed by exploratory data mining for discovery and model development. The analytical strategies that define the overall computational pipeline is heavily dependent on factors including replicates, data/patient dependencies, technical factors, and data format/size.

Due to the large variability in biological data, it is essential to ensure that the data set is curated so down-stream analyses are free from batch-effects related to material/reagents, operator related variation, site-of collection/hospital etc.[242]. Data pre-processing refers to the application of a set of techniques to prepare the data prior to data exploration[243]. In the following chapters we have addressed problems that were tackled during the pre-processing of the data with particular emphasis on rationale that motivated our selection.

Most often, the analyses aim to identify variation in relation to a clinical feature, such as patient outcome. However, technical variation due to a variety of factors cause challenges with data handling. It is thus of importance that the same technical variation is found throughout the study cohort, without association to the parameter of interest. Clustering and classification visualization methods such as principal component analysis (PCA), principal variance component analysis (PVCA), OPLS (supervised orthogonal partial least square), density plots, box plots, relative log expression (RLE) plots were used to identify statistically relevant sources of unwanted variation. Large categorical distribution of unwanted variation caused by non-biological factors, that is statistically significant is called as “batch effects” (BE).

4.2.1 Unwanted variation caused by differences in sample collection, handling, and processing

All biological assays, including high-throughput methods can generate unwanted variation in out-put data. For example, a study showed variable sample temperature, multiple experimental batches, and instruments as a source of bias in metabolomics data[244]. In our large multi-centre studies, where samples were collected at different hospitals across time, could have been a source of error. We investigated all potential sources and especially, multiple field sites in paper I-II and archival tissue age collected during 2000-2014 in paper IV-V. However, when exploring for batch effects, no biases were seen in the output datasets after pre-processing and normalization, indicating that the developed pipelines can compensate for such technical biases.

Large effects based on assays being run on multiple days or different instruments being used etc., are major sources of batch effects (BE). In paper I-II, the experiments were spread over three days, and we identified it as a source of BE in our datasets. In paper VI-V, with respect to spatial-omics, the three different tissue microarray slides and multiple scans did not seem to cause batch effects in the expression datasets. The method to compensate for the BE is discussed in section 4.2.2.

Minute effects such as pipetting errors or handling issues, might be difficult to tackle. However, different batches of reagents could be a major source of error. In our studies, reagents batches were maintained such that no statistically relevant unwanted variation could be seen, expect in paper I-II, wherein two different batches of microarray slides were used. However, since the distribution of the batches was biased with respect to distribution of scan days, that unwanted variation that led to a BE caused by slide batch was eventually compensated when corrections on scan days was performed.

4.2.2 Compensating for batch effects and data normalization

Batch effects are inevitable in high-/low- throughput biological assays[245]. Batch effects generate biases in downstream data analysis by masking biological variation and can lead to over-or under-estimated model developments. Therefore, removal of BE is essential, and this is performed using normalization or BE-correction algorithms (BECA)[245].

Several BECA methodologies are available, but they can be dependent on the data type used to design these algorithms[246]. The use of mean scaling and zero-centering in linear models is type of batch correction[245]. Some examples are surrogate variable

analysis (SVA)[247] and batch mean-centering (BMC)[248] both initially developed for gene expression datasets, removed unwanted variation (RUV) for sequencing dataset[249].

ComBat, which was particularly developed for BE correction in small sample sizes, remains to be one of most popular BECA for microarray analysis due to observed higher performance[245,250]. Similarly, ComBat-seq was developed for BE removal in RNA-sequencing studies[251]. In paper I-II, the experiments were spread over three days, and we identified it as a source of BE in our datasets. Also, the slide batch was identified as a BE. Therefore, ComBat normalization was used. In paper VI-V, with respect to spatial-omics expression datasets, no major contributor to batch effect could be identified in all datasets, in-spite of multiple tissue microarrays and different scans sets. However, since the signal for each cell type was collected over heterogenous number of cells in an area for each core, it was imperative that the data was scaled prior to normalization. Therefore, the measured spatial area was used a scaling factor. Furthermore, after much assessment of linear (scaling by positive of negative controls) and non-linear normalization methods (assessed by NormalyzerDE), the non-linear normalization method of cyclic-loess (locally weighted smoothing) was selected[252].

4.2.3 Dealing with missing data

Missing data could imply lack of information on two fronts - clinical metadata or omics-acquired expression data. Missing data in high-throughput assays is frequent, therefore imputation methods have been developed[253,254]. Imputation is an artificial substitution of missing information based on a reasonable simulated estimate. One of the most basic methods is by using central statistics (mean, median and mode) for imputation for numerical parameters along with fixed value approach (where a constant value substitutes the missing datapoints). However, frequentist approach can also be used for nominal and ordinal parameters.

In paper I-II, missing data was caused by removal of expression datapoints that failed quality control. For example, each scFv antibody was measured in triplicates. Thus, outliers with more than 15% coefficient of variance (CV) were removed, which generated a few missing datapoints. This was dealt with by using bagged trees imputation method, a method often used for this platform[255]. No missing datapoints were imputed in the remaining studies with respect to ELISA or DSP-generated spatial-omics datasets.

4.2.4 Curse of dimensionality - The classic “n” vs “p” problem

In ML, “n” refers population size, and “p” refers to number of variables/features. In medical omics data, “n” would refer to the sample size of the cohort being tested, and “p” would refer to the number of molecular features (protein/transcripts etc.) being targeted. As the number of features grow, the projection into high dimensional space increases in volume which implies that data labels become sparse, thus causing a statistical issue[256]. Zhang *et al*., 2019, have shown that increasing the sample size greatly affects the robustness of identified biomarkers[257,258]. The small “n” and large “p” is known as the curse of dimensionality, which can be an issue in clinical-omics datasets[259]. Additionally, biological data deal with a high degree of multicollinearity which refers to independent “p” features being functionally correlated. A small sample size with high multicollinearity can lead to overfitting biases and lead to over-optimistic model performance[260–263].

The ratio of “n” vs “p” determines the type of analytical method that is further selected. Increasing sample sizes in clinical cohorts such that “p” <<< “n” is a way of mitigating such statistical issues, but this is not always feasible. The amount of “p” features is technologically dependent. For example, protein microarray data or multiplexed ELISA usually deal with features in the range of hundreds, in contrast to bulk or single-cell sequencing data would deal with feature numberings in the range of thousands. Thus, the required “n” would be different for these technologies. Additionally, there are limitations on “n” itself, which can be disease dependent. For example, MCL is rare disease with an annual incidence of one case per 200,000[264]. Hence, reaching a cohort size in the range hundreds is difficult, especially for relapsed MCL. High dimensional issues are often seen in bioimaging datasets[265]. There are now ML methods that specifically target such issues pertaining to small sample sizes and multicollinearity (discussed below). While, increasing “n” may not be a feasible solution in several clinical scenarios, reducing “p” dimension is possible, which is usually performed through dimensionality reduction or feature selection, or by condensing highly correlated values to a simpler design[259,263,266]. The proteomic dataset generated (n=44, p=356) in paper I and II, suffer from the curse of dimensionality. To compensate for this effect, false-discovery rate adjusted q-values as well as parallel independent analysis pipelines were used for biomarker discovery.

4.3 Data mining in biomarker discovery

4.3.1 The basics of machine learning in exploring tabular and structured data

Data-driven life sciences approach has now become a frontrunner for biological exploration particularly with respect to accelerating healthcare research. The power of ML and DL applied to pattern recognition challenges have transformed exploratory research in biology. It is hypothesized that AI will accelerate development of tools for improved clinical diagnosis, prediction, and prognosis. For example, McKinney *et al*., 2020, used an AI-tool based on DL to accurately measure breast cancer in its early stages using images[267]. In terminology, ML is a subgroup of AI that focuses on using mathematical algorithms for pattern prediction and discovery[260,268]. DL is subgroup of ML that uses multi-layered neural networks and is often used for image-based analysis models[268]. Although these fields are vast and highly multidisciplinary, this sub-chapter will focus on the methodologies applied in papers included in this thesis.

A typical ML-workflow deals with both labelled and unlabelled data formats which lead to the two main branches of learning that are often used in medical data - supervised and unsupervised[269]. Supervised methods are used to uncover features associated with target outcome[269]. They can broadly be divided into classification algorithms such as support vector machines (SVM), decision trees, random forests etc.; and regression algorithms like linear, logistic, and elastic net etc.[269]. Unsupervised methods can be categorized in dimensional reduction methods like principal component analysis (PCA) and clustering methods such as hierarchical clustering, k-means clustering etc.[269]. Unsupervised methods allow to identify natural grouping in datasets based on included features. These methodologies are used in tandem both for pre-processing when identifying batch effects and data processing for exploratory analyses. In paper I, the identification of different experimental days as batch effect was done by using both PCA combined with variance filtering followed by multi-group comparison and OPLS (supervised orthogonal partial least squares) and further adjusted using ComBat.

When dealing with the curse of dimensionality, there are two alternative paradigms i.e., either using i) models for feature selection, neglecting the correlation, and ii) using penalized models such as Elastic-net/LASSO/ridge regression which account for multicollinearity in the data[259]. Both elastic-net and LASSO can be used for feature selection, but ridge can only be used as a prediction tool[259]. These methods have been successfully used in defining prognostic biomarker signatures[190,270–273]. In DLBCL, a 7-plex prognostic signature was developed using Cox- and LASSO-regression[274]. In MCL, Lv *et al.*, 2022, developed the nomogram and immune-

related prognostic index (IRPI) to predict the overall survival (OS), by using a combined strategy of cox regression and LASSO[137].

Survival analysis aims to predict the time to an event (outcome, progression) with respect to censored data. Cox proportional hazards model is one of the most common approaches designed for small datasets[275], which has been used in our outcome-based analysis in all studies. Analyses workflows have been developed in parallel with the technical workflow for the IMMray™ platform. Their discovery pipeline is a combined strategy of backward elimination for feature selection followed by support vector machine for model training and the model is further validated on the test set using the leave-one-out cross validation[207]. This method was successfully used to identify a 21-protein signature for prediction of distant recurrence in primary breast cancer[207]. In paper I, parallel analytical strategies were applied wherein Cox regression was first used to select outcome-associated features (Figure 4). This was followed by backward elimination as also previously described in Carlsson *et al.*, 2011[207]. A second, independent and parallel strategy of elastic net regression was used to select feature associated with outcome. Such a parallel combined strategy was relevant in reducing redundant features, as an independent validation cohort was lacking. The strategy partly compensates for the challenges associated with the curse of dimensionality. Through this method, eleven proteins associated with outcome was identified.

The analytical challenges with spatial-omics datasets were different compared to the datasets generated by the IMMRay™ platform. The first challenge was dealing with repeated sampling measures. This was solved by using mixed effect models that account for correlations observed between multiple measures[276]. The total variation is partitioned into fixed and random effects. A generalized linear mixed models (GLMM) expands linear mixed models (LMM) to accounts for other types of response variable[277], and was applied in paper IV-V. Other higher order extensions of LMM have been made such as the glmmLasso that extends LASSO regression in GLMM setting, multi-kernel penalized linear mixed model with adaptive lasso (MKpLMM)[278–280].

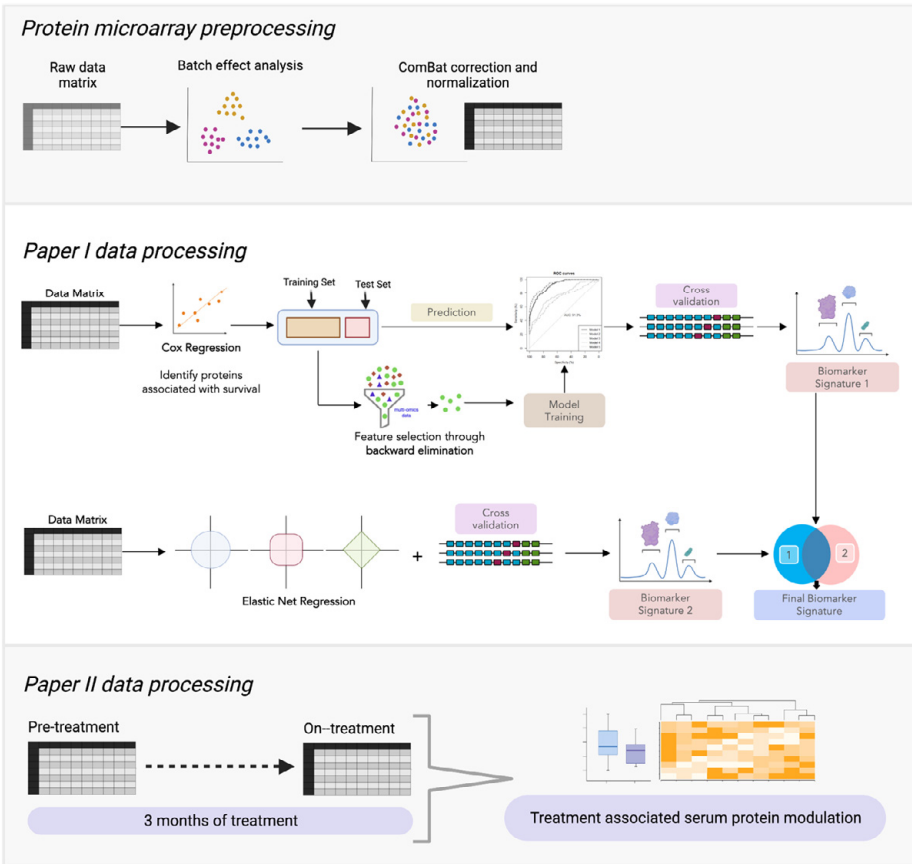

Figure 4: Data analysis workflow for paper I and II. Generated serum-based protein microarray data was pre-processed to evaluate any tentative batch effect, and ComBat was applied for batch correction and normalization. To extract survival-associated biomarkers, two parallel pipelines were used in paper I, first cox regression followed by backward elimination and secondly, elastic net regression. The final biomarker signature of eleven proteins was identified using the consensus from the two parallel strategies and used to develop a new stratification index. Further longitudinal profiling was done in paper II, to identify treatment associated biomarkers. The figure was created using Biorender.

4.3.2 Deep Learning for unstructured image analysis–complementary method to spatial-omics workflows

Imaging in terms of radiology, histology, immunofluorescence etc., often collected as part routine diagnosis and monitoring, reflect a huge volume of data collected in the healthcare industry. Significant improvement in image capture resolution of medical imaging methods has created a demand of automated image processing design for precise analysis of anatomical data that could refine and accelerate discovery of biological features (such as disease diagnosis)[281]. Traditional

computational approaches of analysing medical image include segmentation using contour based, watershed, intensity thresholding etc.[282]. However, AI-driven DL segmentation which uses convolutional neural network (CNN) to train the segmentation and classification model, have shown to have better performance than previous methods. AI based automated image segmentation to extract spatially significant features, have shown to outperform clinical estimates for prognosis and diagnosis[267,283–286].

The combined analyses of spatial metrics and expression data have the potential to advance prognostic models. For example, Zhu *et al.* 2021, developed the SpatioImageOmics (SIO) pipeline that integrates data from imaging mass cytometry (IMC), spatial-transcriptomics, and DL[287]. SIO extracts image features that when integrated with transcriptomics data identified features correlated to survival in advanced stage high-grade serous ovarian cancer (HGSC)[287]. McKinney *et al*., 2020, an AI system designed to assess mammography images for early detection of breast cancer, surpassed predictions made by radiologists[267].

Although multiple imaging formats exist, ranging from histology staining to radiography imaging, different pipelines have been developed that are suitable for specific formats. For example, available tools include DeepCell[288], StarDist[289,290], Cellpose[291], ilastik[292], NucleAlzer etc.[293,294]. In paper V, mIF image analysis was explored as one of the outputs of GeoMx™ were the fluorescent images of the tissue microarray (TMA) cores and ROI. Therefore, we selected specific models that have shown good outputs in fluorescent images and thus StarDist and Cellpose pre-trained and re-trained models for image segmentation were assessed. They are among the top state-of-the art U-NET architectures for such image types[291].

Model performance was measured using metrics in terms of true/false positive and negatives which define accuracy, prediction, and recall[281]. A primary comparison was by cell counts between DL segmentation output and manually annotated, since cell count was used downstream to define T-cell frequencies. Additionally, segmentation mask was compared using the Boundary F1 (BF) score, also called as Dice Similarity Coefficient (DSC) and Intersection-Over-Union (IoU). BF score is harmonic mean of precision and recall, is often used a measure of matching boundaries between predicted and ground truth segmentation[281]. IoU is the ratio of overlapping area to the union of area to compare between two models and in our case, this was used to compare the predicted and ground truth segmentation.

Based on the comparison of multiple workflows using different strategies, we selected a retrained Cellpose model, that performed an initial segmentation on Syto13 staining, followed by artificial expansion of the mask to the surface marker and classification using random forest for identifying different cell subsets. Thus, image derived metrics provided true estimations of cell-specific distribution as well

as cellular coordinates to measure cell-to-cell distances. In paper V, T-cell frequency as derived from image analysis, was further integrated as additional features with the spatial expression dataset collected from GeoMx™, to evaluate questions such as whether different T-cell cell ratios can impact the expression profile of other cell types and/or be correlated to patient survival.

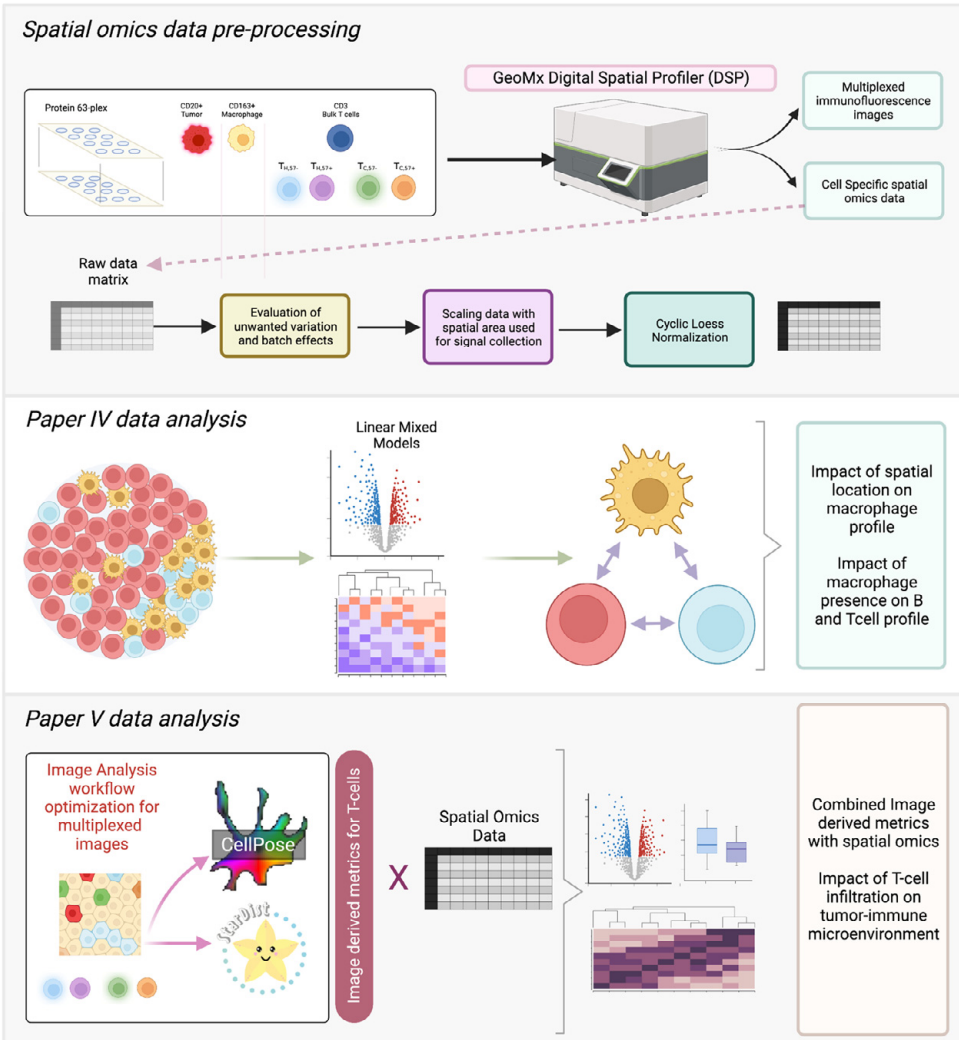

Figure 5: Data analysis workflow for paper IV and V. Spatial proteomics data (~63 informative immune-related proteins) were collected from MCL cells, bulk CD3 cells, M2-like macrophages and four T-cell subtypes using GeoMx DSP. Data was scaled based on total segmented cell area and normalized using non-linear cyclic loess normalization. In paper IV, MCL cells, bulk CD3 cells and M2-like macrophages were assessed with respect to spatial localization and variation in phenotypic profile. In paper V, multiple image analyses workflows were evaluated to extract cell-derived metrics from multiplexed immunofluorescence images. Sub-type specific cell frequencies were determined and used to assess the impact of T-cell infiltration on the tumour microenvironment. The figure was created using Biorender.

Chapter 5: Scientific conclusion and future outlook

5.1 Scientific conclusion

In the era of precision medicine, improvement of MCL patient outcomes is dependent on identifying features that can be used for patient stratification. Such features should allow identification of patients likely to respond to a particular treatment regimen. A major player in this scenario is the immune system, which has not been extensively studied in MCL.

As elaborated in chapter 1, section 1.4, the five papers that are part of this thesis, focus on various aspects of the immune system, both with respect to liquid biopsy for systemic biomarker discovery and lymphoid-tissue based exploration of the tumour-immune microenvironment. Technological advancements and complementary bioinformatic processes that allow parallel investigation of high dimensional data based on low number of biological samples, have become pivotal in driving high-throughput research. This data-driven process has been utilised in four out of five papers. Together, these methodologies have provided better molecular insights in diagnostic and R/R MCL immune microenvironment. The overall aims of each paper have been elucidated previously in section 1.4.

For the analysis of R/R MCL, serum proteome profiling was performed in paper I-II primarily because R/R MCL is a systemic disease and biopsies of secondary lymphoid organs rarely are taken after diagnosis is set. Thus, liquid biopsy is the preferred method of patient sampling. Moreover, very few clinical cohorts exist for R/R MCL and therefore biological exploration of R/R MCL has been limited. Particularly, studies that focus on determining prognostic features are rare.

Through high-throughput serum proteomics quantifying 158 immune regulatory and tumour-secretory proteins, we identified biomarkers in relation to overall survival in this R/R MCL cohort. This was facilitated by the bioinformatic methodologies previously elaboration in chapter 4.3.1 (Figure 4). Parallel strategies for feature selection were applied to remove redundant and false positive selections, since sample size was small, and no validation cohort was available. Although, several strategies for feature selection exist, but methodologies were selected based

on models popularly used for survival-based association also used to build MIPI (Cox model), methods previously applied for this data (backward elimination) and models that work well for small dataset with multicollinear features (elastic net regression).

Our results identified eleven prognostic proteins that were mostly chemokines, cytokines or signalling adaptor molecules, which were combined to form the relapsed-immune signature (RIS) and further integrated with MIPI to develop the $MIPI_{ris}$ score (Mantle Cell Lymphoma International Prognostic Index – Relapsed Immune Signature). This score was significant for both outcome and progression prediction and re-stratified the patients into three risk groups. Post the publication of this study, the outcome information was updated. The new 5-year follow-up time was even more significant for patient stratification using $MIPI_{ris}$ (Figure 6). This result is critical as it suggests how systemic information can be adapted for prognostic indication in MCL, circumventing sampling limitation in aggressive cases.

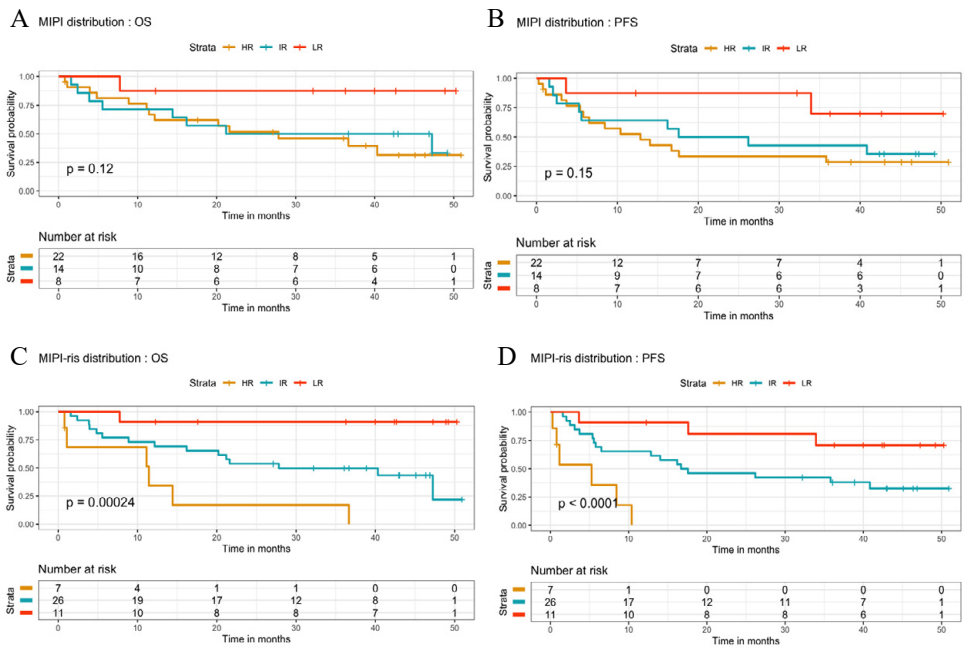

Figure 6: MIPI and $MIPI_{ris}$ risk stratification based on updated follow-up times. Kaplan-Meier curves of MIPI (A,B) and MIPIris (C,D) distribution for overall survival (OS) (A,C) and progression-free survival (PFS) (B,D) using the 5-year follow-up time updated in 2020.

Paper II explored longitudinal profiling of pre- and on-treatment samples (12 weeks of treatment) in the assessment of responsive biomarkers to understand if beneficial response can be measured early in the treatment of patients. To our knowledge, this

had not previously been performed in MCL. Differential expression profile was analysed using paired two-group comparison and three proteins were identified (PARP1, APLF, and GOLPH6) that showed decreased expression on-treatment and were shown to be associated with genomic alterations of *TP53* and *ATM*. Both these mutations have shown to be critical and related to poor outcome in MCL[108]. This is an example of response-associated biomarkers. This result was crucial as it elucidated how different genetic makeup can alter systemic responses to treatment, thus illuminating the need of adding genetic heterogeneity for further stratification in clinical cohorts.

In paper II, we also showed that using velocity of change (δ_{C4}) of individual protein levels comparing pre- and on-treatment was critical to account for variation and scale the data. We confirm that MRD status can be measured in sera using TGF-β as a surrogate and time-to-progression is associated with change in BTK expression. The use of MRD-guided treatment selection has been proposed in MCL and in general for lymphoma, particularly in relation to rituximab chemotherapy[295]. MRD status has been considered crucial in MCL both as a diagnostic and predictive tool and is usually measured by tissue analysis of bone marrow to find tumour infiltrates[296,297]. MRD measurement using current methodologies based on PCR, fails to provide information for approximately 15% of the patient's due technical failure[298]. While high-throughput sequencing has been suggested as alternative to provide more accurate readout, such a method can be more expensive and time-consuming[298]. Thus, establishing a serum-based monitoring biomarker using TGF-β, could prove to be useful. A continuous study measuring the expression of TGF-β and MRD status, from diagnosis, during treatment until relapse, is therefore needed to establish such a measure.

In both paper I-II, the lack of validation is a major drawback. However, the MCL7-Valeria clinical trial (conducted by the Nordic lymphoma group) is currently ongoing wherein R/R MCL patients are being treated with combined venetoclax, lenalidomide and rituximab[299]. Thus, the robustness of $MIPI_{ris}$ can be potentially tested in future using this group as a validation cohort.

Serum based discovery can crucial particularly considering clinical usability and ease of application. The value of serum-based analysis is further manifested in paper III, wherein we show how serum analysis is complementary to tissue-identified prognostic biomarkers. We report that high serum expression of sCD163, as measured by ELISA, was associated with poor survival at diagnosis and relapse and provide a potential clinically relevant cut-off (~2960 ng/ml) which can be used for both diagnosis and R/R MCL. The value of sCD163 in relation to outcome was more robust at diagnosis compared to time of relapse.

In paper IV, the crosstalk between CD163+ M2-like macrophages and surrounding tumour and T-cells was evaluated. M2-like macrophages are associated with immunosuppressive microenvironment that support tumor development. We show that that based on spatial localization (tumor-rich vs tumor-sparse) with respect to the CD20+ tumour cells, the proteome profile of CD163+ M2-like macrophages is modulated. For example, VISTA and STING were enriched in CD163+ and CD3+ segments of tumor-sparse areas. Thus, we conclude that the modulation of the immune response by M2 macrophages can be varied based on spatial localization. To understand how presence of CD163+ macrophage modulates the profile of tumor cells, differential profile of tumor cells with or without macrophage were evaluated. Immune checkpoint molecules PDL1 and PDL2 were downregulated in CD163+ TMEs. Tumours in CD163+ TME show increased expression of MAPK signalling pathways. Increased MAPK is associated with downregulation of T-cell proliferation and activation. Profile of T-cells suggested decreased antigen presentation due to expression of high IDO1 on tumours and low B2M and HLA-DR in T-cells in CD163+TME. Moreover, increased angiogenesis and neovascularization is observed in CD163+ TME based on high expression of SMA and CD34. Based on our analysis, we propose that TMEs with CD163+ macrophages can potentially be targeted for MAPK inhibition, for complementary treatment. Overall, this study enabled better comprehension of the crosstalk of CD163 with the surrounding TIME.

Paper V exhibits how important information can be derived from images and in this case, mIF images collected from GeoMx™. To enable us to extract correct metrics, a technical assessment of various combinatorial pipelines using StarDist and Cellpose methodologies was performed. These models were primarily selected as have been developed and previously used for analysis of IF images. Based on our analysis, we show that Cellpose image segmentation is superior for complex tissue segmentation. An optimized workflow was developed based retrained Cellpose nuclei segmentation followed by a random forest classifier. This workflow can be applied to mIF images given that cell circularity is closer to one.

Optimized workflow was used to extract features of interest from the four T-cell subtypes for TMA cores as well smaller ROI in tumour-rich or tumour-sparse regions. Our results show that cores containing tumour-sparse must be investigated carefully, as they are histologically heterogeneous and thus, evaluation must be limited to tumour-infiltrating T-cells particularly for assessing prognosis. Cox regression and Kaplan Meier curve analyses applied to T-cell metrics show that higher frequency of $T_{H,57-}$ and CD3 T-cells is associated with improved survival. Additionally, combined analysis with spatial omics data suggests that highly infiltrated tumours are associated with more immunosuppressive environment by the dual expression of IDO1 and STING on the T-cells and the negative correlation of these markers with GITR expression. In relation, low infiltration is associated

with a more immune active status by the enrichment of cytolytic markers on $T_{C,57-}$. Simultaneous modulations can be seen on the adjacent CD20+ tumour cells. We additionally observe that increased frequency of CD57+ T-cells is associated with distinct proteome profile of CD57- T-cells. These analyses also reveal the possibility of other neighbouring cell types such as the presence of myeloid lineage cells and memory T-cells in the vicinity of CD57- T-cells in high infiltrated tumours. Taken together, this manuscript represents how deep-learning assisted image segmentation can be used complementary to spatial omics and provide deeper insights in cellular mechanisms of the tumour-microenvironment.

The infiltration levels of total CD3 T-cells and CD57- helper T-cells was prognostically relevant in dichotomizing patients in outcome groups in response to immuno-chemotherapy in paper V. We propose that measuring infiltration of T-cells could be used at diagnosis as an additional measure of risk and suggesting that highly infiltrated tumour would benefit more from immunotherapy. We additionally propose that the type of immunotherapy would be dependent on neighbouring tumour microenvironment modulation as shown by combined spatial image analysis. Thus, we recommend further investigation into the interplay of GITR (checkpoint), IDO1 and STING (checkpoint) axis, which has been previously understudied in MCL and suggest them as additional immunotherapeutic targets with dependency on T-cell infiltration.

The five papers that are part of this thesis are based on evaluation of the immune microenvironment in MCL being investigated both in sera and tissue from clinical trial cohorts. This work has deepened our understanding of the systemic regulation and provided better insights into the MCL tumour-immune microenvironment. On a technical level, new and upcoming omics technologies have been used on patient cohorts to build this current story. The datasets thus generated from these different platforms, led to additional challenges with data processing which propelled computational pipeline development that resulted in the biological conclusions as elaborated in the papers/manuscripts. Further, complementation with image analysis has only highlighted how integration of different technologies can be used parallelly to provide a more holistic insight.

5.2 Future outlook

The future of MCL research is highly exciting and is moving towards more translational and interdisciplinary applications. Technological innovations would be paramount in driving this forward, particularly in generating huge amounts of data with limited patient material. As shown by our study, serum-based analysis could

be crucial for this purpose, especially for biological inferences from aggressive and relapsed cases.

Development of companion diagnostic/prognostic tools based on biological features that can stratify patients with high accuracy, but also enable treatment selection, is the future of MCL therapy. MCL subtyping based on genomic heterogeneity combined with variation in immune response can to some extent can be pivotal for the development of such methods. This is evidenced by the fact that longitudinal profiling had suggested how the mutational landscape can modulate systemic immune microenvironment. While risk stratification is crucial, it is necessary to identify indicators for the purpose of treatment selection. As we have shown with spatial omics, certain phenotypic profiles could potentially benefit from specialized immunotherapy based on T-cell infiltration and tumor profiles, such as for example IDO1 and MAPK inhibitors. Thus, more research that drive development of indicators that can advance treatment selection, as well as validation of previously proposed features is necessary in MCL.

The importance of spatial localization in tissues is highlighted in our work, and this has been shown recently by several other studies[300,301]. Currently, we have only profiled spatial proteins, but quantification of cell intrinsic transcriptome is also possible with GeoMx™. We are now exploring a spatial transcriptomic data (~1800 plex) that will further highlight the functional variation in MCL TIME associated with spatial localization and abundance of different cell types.

Development of comprehensive models is further possible by integrating multi-modal information. This is particularly necessary for patient stratification as integrated models which contains extensive insights into patient biology, would drive the development of precision prognostics. This can be achieved by applying high-throughput omics-based profiling in multiple molecular layers and further consolidating this information using data integration methods. Many bioinformatic models for data integration, from early to late-stage consolidation, have been developed[302], with several published studies highlighting the importance of integrated models. For example, a recent report used artificial neural network trained on multi-omics signature from breast cancer patients, for outcome predictions as well as drug response prediction[303]. Moreover, a multi-omics-based tool specific for cancer subtyping was recently proposed that was superior in performance to conventional approaches[304]. Data integration has been our particular interest, and we are now in the process of integrating GeoMx proteomic and transcriptomics spatially resolved data to build such models. We believe that this will be pivotal for defining patient-specific treatment for optimal response in MCL.

Furthermore, GeoMx is limited to compounded cell-specific information extracted from a spatial area. Currently, the technology only allows profiling of cell-types on

three surface markers with nuclei staining. Therefore, the extent of cell type deconvolution is limited. Recently, several new methods that perform single-cell spatial analysis are being developed, an example being CosMx™ by Nanostring[305]. Such methods would allow for further profiling especially for rare cell subtypes. Such an investigation along with clinicopathological information would benefit MCL in pin-pointing patient-specific biology and further enable disease subtyping.

The application of machine/deep learning and artificial intelligence to high-throughput data, would further enhance our understanding of MCL. Particularly DL assisted image analysis can propel our understanding of the MCL tumour, especially in relation to high multiplexed images by using IMC can be particularly crucial in defining cellular complex compositions. By integrating composition data with functional datasets (such the GeoMx™ data), the association between TIME composition and microenvironment functionality can be assessed. Moreover, research into image segmentation models using patient derived tissues, will accelerate development of DL and AI applied digital pathology tools for translational into clinical utility[306]. As we have shown in paper V, image analysis can be used as complementary method to extract composition of the time. The association of T-cell infiltration as assessed from image derived metrics in relation to outcome and its impact on MCL needs to be further validated especially in a homogenous treated cohort. We are now in the process of performing such an analysis using additional cohorts.

Conclusion

Research in MCL has made significant progress in the last decade, particularly in understanding the underlying biology and in the development of new treatment options. However, there is much to be investigated, especially in relation to the immune microenvironment, comprehensive knowledge of which is still lacking. The studies included in this thesis were a small step in that direction. Additionally, tools to guide treatment selection as well as development of targeted therapeutics and immunotherapy is a must, to further improve survival rates and quality of life of MCL patients. Precision therapy is now being made possible by application of new technologies and interdisciplinary research that integrate several scientific fields. This was the focus of this thesis as well and the included research papers have unlocked several new potential avenues for MCL research.

Popular Science Summary

In the last three decades, the drastic evolution in cancer care has resulted in major improvement in median survival for patients. Despite novel more effective drugs, cancer is still one of leading causes of death worldwide. It has become clear that the high variability seen in patients' response to treatment is due to the vast inherent biological variation. Society puts hope to precision medicine efforts, where more precise ways to target cancer are used by directing treatment towards the individual patients' unique molecular tumour footprint to further improve outcome.

To aid this process, development of tools that that can measure biological molecules, i.e., biomarkers, that is relevant for how patients respond to treatment is needed. Biomarkers can be of different molecular origin such as genes, mRNAs, or proteins. Variation in biomarkers, often expressed by cancer cells, can be used to measure, and define disease status, patient response to treatment or predict/prognosticate outcome.

The immune system plays a critical role in preventing and controlling cancer. However, in fully developed cancer, the immune system is often repressed and contributes to the aggressiveness of disease. Thus, identifying biomarkers that can measure the activity of the immune system is important to fully take advantage of immune-directed therapies that recently has been developed.

In this thesis, I describe our translational studies on immune regulation in an aggressive haematological malignancy named mantle cell lymphoma (MCL). MCL patients who often are elderly and with a male predominance, have overall short survival and high risk of relapse of disease. We investigated tissue biopsies and whole blood/serum (non-cellular component of blood), which are the patient material often clinically collected. Studying these materials provide different insights into tumour-immune regulation. Tissues allows us to define the heterogenous cellular landscape and biology that governs the course of the disease. Our focus was to understand how spatial composition of tumour-fighting immune cells such as macrophages and T-cells impacts MCL biology. For that purpose, we used a newly developed high-throughput spatially resolved omics technology that retains the spatial architecture and provides higher-resolution biological insights. We discovered that spatial localization of macrophages in relation to tumour cells and composition/frequency of T cells alters the phenotypic expression of immune-

related proteins. This in turns impacts the tumour expression thus highlighting the biological heterogeneity of MCL caused by factors in the tissue microenvironment. The insights into cellular crosstalk, will be further explored to identify potential targets for immunotherapy.

While tissue-based analysis traditionally is the go-to methods for investigations of cancer, blood-based evaluation can be pivotal for gaining insights of systemic immune regulation. Furthermore, collecting patient blood is a cost effective, less invasive method that allows repeated sampling. This is particularly suitable for frail patients such as relapsed patients where collection of tissue biopsy may not be recommended. Therefore, we investigated the systemic immune regulation by large scale profiling of serum proteins, with the objective of identifying clinically relevant biomarkers which can predict patient response to treatment, but as well as identify patterns of biological variation that can further categorize patient subgroups into risk groups that could potentially be used for therapeutic decision making. With this goal in mind, we have defined a new prognostic index using immune-related serum proteins that additionally allowed us to define molecular subgroups associated with relapse.

Clinical implementation of novel biomarkers is a major need to successfully implement precision medicine efforts where companion diagnostic tools are needed to identify patients likely to respond to a specific therapy. Thus, the development of assays that are feasible to handle and cost effective, is a necessity. In their current state, the above-mentioned technologies are great for discovery, but not so much for clinical implementation. However, there are ways of circumventing this issue. For example, presence of specific macrophages in tissue have shown to be related to outcome in this lymphoma. By targeted proteomic quantification, we were able to show the soluble form of the surface protein CD163 in sera could easily be used as alternative and implemented in clinics. We also defined clinically relevant cut-offs that differentiate patients into high and low risk outcome groups based on measured protein levels. This is an example of how serum expression could be complementary to tissue-based knowledge and how discovery could pave way to clinical application.

All-in-all, this work explores multiple aspects of the disease with particular emphasis towards immune responses. In a holistic approach, we advance from disease assessment using high-throughput technologies, to complex data analysis required for discovery, identify relevant biomarkers, and explore the tumour-immune microenvironment. This work has contributed and further supplemented the biological knowledge on this cancer subtype.

Acknowledgements

There was a time in my life when I seriously debated if I wanted to do a PhD. I had heard all the rumours you know!! But sometimes life intervenes, and I found myself in Lund, joining a wonderful PhD program surrounded by all these incredible people. It was perhaps one of the best decisions I made in my life, and I have always been grateful for this opportunity.

Doing a PhD is tough, full of trials, unnecessary risks, unexpected failures, sleepless nights, and comatose days. But it is also fun and exciting when you are surrounded by the right people who are there to support and help you, comfort you when you cry, listen to you vent and join you in being maniacally happy when something goes right. This thesis would not have been possible by the contributions made, both professionally and personally, by so many people. I am truly grateful to have met you all and am thankful that you were part of this journey with me.

No one deserves the credit more than my awesome supervisor, **Sara**. You inspire me in so many ways, as a scientist, as a mentor, as a role model, as a leader and dare I say it, as a friend. Thank you for selecting me as your *not 17th* student. I consider myself one of those few lucky PhDs that can say that they absolutely adore their supervisor (which I have to the annoyance of many) and could not have selected a better one. Thank you, truly, for the way you have supported me during my PhD, enabling me to always speak my mind, listening to my gazillion ideas, for always being available to help me even on weekends and odd hours, for being an amazing scientist from whom I learnt so much and for the Sauron's eye when a deadline is due. I grew so much in the last five years, which would not have been possible without your guidance. Your belief in me and all the encouragement for my ideas and my dreams, kept me moving forward. More than anything, I admire and respect you as a leader and a mentor, who knows how to keep your team together and motivated, even when things get rough. All the things that I learnt from you, have just made me a better person and a better scientist. Thank you so much for everything.

I am grateful to all my in incredible co-supervisors, that were a pivotal part of this journey. To **Mats**, thank you so much for your feedbacks and your clinical support. With the infinite clinical trials and for the absolute lightning speed replies to the most absurd questions, you were an important part of this journey. The clinical

aspect of biological research is something that I learnt from you, and it has also helped me shape my plans for the future. To **Anna**, I want to thank you for your brilliant insights, all the interesting and fruitful conversations, the guidance that you gave me, for your positivity and your encouragements. I have learnt so much from you and I am truly grateful for that. To **Venera**, thank you for the first introduction to the world high-dimensional data analysis. I started this PhD with the serum analysis under your guidance and then it just propelled me completely to bio data-science. Thank you for your constant smiles and your support in this thesis.

A special thank you for **Jana**, for all the support in the CanFaster program, for the amazing parties and after works at the department. And also, for your cheerfulness and the laughter's all around the fika room. I also want to thank **Maria K** for the assistance in the CanFaster program. To **Corinna** and **Lisa**, for being my introduction to the first lab work I did when I started my PhD. Thank you for your patience and your guidance and for all the things you taught me. To **Lina**, thank you for the keeping the spatial omics facility alive, for being there to answer all my questions and for lunch-time conversations, when I was regular at the department.

To all the members of the **CancerTarget** group, old and new, thank you for the constant support, for being there in general, for the interesting discussions and the group fikas. It has been really nice to witness the growth of our group and be surrounded by so many competent people. And also, thank you bearing with me when I present in our JC's, I know I speak a lot.

Thank you, **Cornelia**, for being so patient with me for the past five years, replying to all my random questions and helping me with all things financial and of regulatory nature. A very special thanks to the administration team in our department, both past and present members, – Cornelia, **Andreas** and **Ann-Sofie**, for keeping your doors open and for always responding to my one-minute questions.

To the **rest of the department**, both past and present members, thank you for making the department super welcoming when I first joined, for all the support I received overall during the various times as well as during crises and for keeping the general atmosphere pleasant. Five years ago, was my first time moving so far away from home, and the kindness, with the friendliness that I received from everyone, helped me a lot with my home sickness. And also, thank you for all the conversations around the table, during lunch, fika and Tuesday breakfasts. In addition, thank you **Magdalena** for the help during the Immunotechnology and Biopharma courses. To **Fredrik**, for the computation supports, and for saving my laptop from being infected by my stupidity. To **Jakob**, for setting impossibly high standards of a data scientist and of course, for your help when I was just starting to learn. Your work was an inspiration to me. To **Paul** and **Louella**, for all the bioinformatic support and general guidance into analysis. You helped me solve so

many random questions on analyses. Also, to **Mattis**, it was fun working with you for the few months you were at the department.

To all the **CanFaster's** that have been part of this journey with me. Many of you have finished and many remain. But I know each and every one on you has a bright future ahead. It was nice to be part of such a diverse team of people with different backgrounds and scientific expertise.

To all the co-authors of the various papers and manuscripts, thank you for all scientific support and feedbacks. A special thanks to **Anna N.** for the quick replies and discussion on paper III. Hoping for a quick acceptance of that paper. And to **Ingrid**, for the conversations with respect to the project as well as during the various conferences. Also, a special thanks to the **European Mantle Cell Lymphoma network** and **Nordic Lymphoma group** for the different conferences that kept teaching me so much.

I also want to the thank **Medicon Village** and **Lund University** for all the resources. And **Media Tryck**, the help in formatting and printing of this thesis.

For all my friends -

There are some people in life that no matter how much one tries, words would fail to convey the gratitude. To me it was the priceless **friendships** that I made here, people that were by my side when I needed them, shared my laughter as well as my pain. And I don't think I would be standing here with a completed thesis, without them and for all that they did for me, including saving me when I forgot my department ID so many ridiculous times, saving me when I locked myself in random rooms, getting me printouts when I am too lazy, being my party pals and my travel buddies. They gave me a sense of belonging and a community. It is my privilege that I came to know all of you and that we were a part of each other's journey.

To **Aastha**, you are one of the strongest, kindest, and bravest people that I know. You are an incredible person and the perseverance that you have is an inspiration, but truthfully an unattainable feat for me. Thank you for being so generous, supportive, and just caring so much for me. I am honestly, so grateful for your constant smiles, even when you are dealing with so many issues and your general cheery disposition. And for being my food, laughter, and dance partner in crime at all the random places. Please do not travel all around the world, keep a few places for us to travel together. Also, thank you for being an absolute genius in the cell lab and helping me with all of that. To **Sergio**, dear *husband* (Sorry Will, he was mine first ;)), we joined the department together, shared the same office and the rest was

history. You were the first friend I made here, and I am so grateful that I came to know you. Thank you for being my laughter, for all our inside jokes, bearing with me at the office, all the random brainstorming sessions and for solving coding issues when I am about to throw my laptop from the 4th floor. You are just a wonderful person (minus the cold freakiness part), with the best humour and an amazing chef (I am still waiting for my bread by the way). Triumvirate shall rule the world! Also, **Will**, so glad that you are a part of this journey. It has been amazing knowing you and being your friend. Thank you for having an amazing food palate and for all the Mui Gong trips. Let's do a food tourism trip to Asia together.

Joana, I don't think I could have asked for better companion and a co-PhD under Sara. You are a person that I know I can trust, and I am truly grateful for that. You inspire me with your efficiency and discipline, which is absolutely amazing. Thank you for the great trips, all the cell lab help, all the conferences, for the completely unnecessary McDonalds trip (you know what I am talking about, we take it to the grave) and all the pizzas (sometimes calzones for you) that we shared. You are an amazing scientist and I know your future is amazing because you will make it so. Dear **May**, I start with the cheesy eggy ramen, that alone supported me so many nights when I was writing this thesis. Also, the introduction to the K-drama world, you did not realise when you created a monster. Now it is my background noise (when I barely understand it, haha). I am so grateful to have you in my life because I know how much truly you care for me. Thank you for all the dinners and food you made for me, Leve bageri and the Friday donuts and all the packages that you got for me when I was sick. And just for being there helping at all times. Especially for all the GeoMx support and for always answering my questions (often the same ones). **Shuvolina**, thank you for giving me a taste of home, whenever I desperately craved it. I have been home sick so many times, and your food was such a comfort. Thank you for just being all round supportive, for asking about my well-being and taking care of me like a Didi. Sorry that the ghee was so late!

Many wonderful master students, who became cherished friends, were part of this journey. I hope that I was able to help you all as you have helped me, and that I was able to guide you in your projects as you required. I apologize for any mistakes that I may have made along the way as I learnt to be a mentor myself. You were amazing and I see a bright future for you all. Thank you for being part of my journey. To **Kajsa**, you were my first master student, so I may have made many mistakes. But you taught me so much and it was amazing to have you at the department. I miss our random scrolling and discussion over dresses when the analysis and writing got to us. To **Alexandra**, who was pivotal in initiating the microfluidics project, I cherish our conversations over the vacuum pumps and confocal rooms and that sushi night. I miss your cheery disposition and teaching me about berries. You are an astounding person and I wish you well. To **Linn** and **Ellinor**, I miss the avenger's movie nights with pizzas at the department as well as the super funny farewell party

(Hellllooooooo). To **Christine**, playing with paraffin and figuring out so many new things together was so much fun. To **Aura**, thank you for all the amazing computational discussions on image analysis and for the OPIS workflow. Your combined knowledge of biology and computation is impressive. To **Daniel**, the image analysis would be lost without you. Thank you for all your help, for teaching me so much about image segmentation and for the amazing dinners and random trips to the Asian supermarket. Once you are done with you master's thesis, please take a long vacation. To **Elias**, I did not supervise you personally, but I really enjoyed our chats on data exploration and analysis. It was really nice having you at the department.

To **Comida**, thank you for making me a part of you group of amazing people, for all the trips, the dinners, the art nights, and the beers. Also, a joint thanks for making me do the Nordic sauna+cold ocean dip process. I never, in my life, thought that I would ever do it.

To **Kat**, you taught me one of the most important thing, and that is the art of conversation (you know what I am talking about), something I struggled with a lot initially. Just the way you are, makes it so easy for someone to talk to. I know I can come to you for anything and the advice you would give would be spot-on. I love your witty sense of humour and the time we spend together, just talking and gossiping, is precious. Thank you for being there for me at all times and offering to help. To **Maria**, you were the one who added me to "Be creative", which was absolutely amazing because it gave me a community of people who became my cherished friends. I truly thank you for that. Your kind heart towards everyone and your compassion, are some of your many best qualities. I appreciate all the gestures of help and kindness you have shown towards me, over the years. I look forward to many more movie nights (with less thought process involved). **Alex**, for me you will always be "the Viking", the one who helped me drill holes (albeit with partial success) in my house walls, taught me what is shotgunning a beer, and MTG. You are always there to listen to my worries and give me wonderful advice that I am truly very thankful for. I am glad to have found a friend that shares a love for mangas, animes, superhero movies, and basically all fantasy related things as well as all the coding and scientific discussion. I have never met someone with such a wide knowledge of things as you do, and I have always been impressed by your intellect. **Laura**, you are fearless. And that's it. No, seriously, I love it. Your strength and charisma are awe-inspiring, and I absolutely love that about you. Thank you for choosing me as your neighbour, for all the random dinners, for the entrepreneurial spirit that we share, for all the fun, laughter and talks that we have had. I look forward to our future in Mumbai, the top floor office, and a penthouse view and with hopefully the company we build. **Ana**, we miss you a lot here. You have to be one of the most cheerful person that I know, always looking at the positives and being super happy all the times. Your love for elephants is

unparalleled! Wishing you, an amazing postdoc in Boston. **Jordi**, thank you for being there always, for being so nice in general and ready for a talk anytime. I loved the time we all spent together in India. Hopefully I will meet you two soon. **Martino**, you can easily change your career if you want, you can be a director or an editor (whichever you prefer). The movies you make, and your creative insights are just incredible. And also, your pasta's (food in general) is so delicious. I have tried and failed to recreate your recipes, so I will keep asking you to make me delicious food. **Matilde**, you are an artist, and that passion is something we share. Thank you for always being there and for your kindness. I am truly blessed to have a friend who is as compassionate and thoughtful as you are. Let's go for more art events together, like the kroki workshop. **Mo**, I see you as a tree, tall, quiet, grounded, strong and super supportive. Thank you for listening to me and for encouraging me when I have been worried. Your thoughtful approach to life is an inspiration to me. **Myriam**, I love your impressions, and how light-hearted you are. The way you easily smile and laugh, makes everyone feel at ease instantly. I admire that quality about you, a lot. I am grateful that I have come to know you. **Sid**, glad to have found someone who can eat and appreciate spicy food. You are the best spider that I have had the opportunity to kill. Your acting abilities are spot-on. To **Marta**, thank you so much for all these years of friendship, for your all warmth and care. You are such a passionate artist and I love all our art nights. Wishing for many more of them in the near future. **Oscar**, thank you for all the weekend afternoons at family's house, for being so considerate about me at all times and for being my pizza buddy. I am truly grateful to have you in my life. **Martina**, sadly we didn't get spend a lot of time together before you had to leave, but you are an amazing person and I hope we continue to build our friendship. To **Jessica**, my partner in crime when it comes to food indulgences. Thank you for just being such an amazing person, who loves shopping and eating as much as I do. Thank you for all the shopping trips especially when I really needed them and the encouragement that you have given me over the years. To **Chang**, for just being so thoughtful and caring. Thank you so much for being there for me and for all your support at all times. You are too sweet!!! Your friendship means a lot. To **Roberta**, thank you for being so attentive and for always lending me an ear. To **Marija**, I cherish our memories from Belgrade and the amazing time we shared together. To the many "**Be creatives/festives**" – **Martina, Eliska, Tiago C, Julie, Nadja, Albert, Veronica, Carletto, Kritika, Tiago B, Eva, Adrien, Fabio, Andreas B., Isak, Marcus, Ugne, Scott, Nic**, for being a super supportive family and for all the amazing parties, Thursday beers, birthday celebrations and the trips. Each and every one of you is an amazing person and has played an important role in my life. I want to thank you all for inviting me into your group and for being there for me at different moments of my life here in Sweden. Thank you so much for everything. I am truly grateful for having you all in my life and I hope that we continue to build our friendship.

Ole Ole Ole Ole Ole, the time we shared in that house together was priceless. We formed a strong bond, that I know will remain forever. We just click! Thank you for being there, for all the dinners we planned together, the brunches that you woke me up for, the wild song nights, the gossips, for being one of the artsy people that I could connect creatively with and being just as crazy as me. **Peter**, thank you for being the unofficial roommate, for the delicious sake and sharing my love of photography. We need to plan more photography trips!! You are an absolutely amazing person.

To **Sofia S**, you are so sweet, and we miss you dearly! It was so nice to have you at the department. Thank you for the impromptu dinners, the gossip, the ultra-funny stories that made me gasp for air and all the new places we tried together. To senpai (**Fabricio**) and **Mayra**, I look forward to more barbecues and the absolutely smashing birthday cakes. Senpai your cooking is amazing. You both are so sweet and lovely, and I am so glad to have met you. To **Rafsan**, we met in ELLIT and became friends over pool and beers. Thank you so much for the conversations on deep learning, imputation, autoencoders and for the absolutely unnecessary ikea trip.

To **Girish** and **Vaishali** and, they were my home away from home. They invited me, a stranger and welcomed me to Sweden with open arms. Words would fail me to convey my gratitude for their immense kindness and sincere thoughtfulness. I thank you both so much, for everything you have done for me, for inviting me to the various festivals and delicious home-cooked meals.

I also want to take this opportunity to thank **IISER, Pune**, for training me as a scientist. To all my teachers and professors who have guided me so far. A special thanks to **Chetana** Ma'am and **Urmila** for their guidance during my master thesis. And to the countless friends I have made along the way, during my school years and at IISER, I would not be the person I am today without your support at different times of my life.

Lastly, none of this would have been possible without the love and support of my family. I have loved every moment of my five years, but I have missed my family terribly. Knowing that they are there cherishing and cheering me has kept me going strong. I owe my life to them and there no words that can convey the depth of my feelings and my love for them. This thesis and my doctorate are dedicated solely to them. To **Baba**, who has been my rock and constant supporter, who has always been on my side, encouraging and comforting me. To **Ma**, who has been my inspiration and role model throughout my life, who I resemble the most in character and personality. To **Yudhu**, my beloved brother, who always surprises me with his thoughtfulness and his incredible wit, I hope I have been the good sister that you deserve. To **Mau**, who has been my second mother and is a close confidant, who no

doubt spoils me (I love it). To **Sanjay uncle**, who has always treated me and taken care of me like his daughter. To **Ayan**, my beloved brother number two, who will probably never grow up in my eyes, whose depth of insight often amazes me. To my late **grandparents**, who would have loved to see me finish this journey. I wish you here now, with me. I love you all so much. Thank you, from my bottom of my heart for all that you have done for me in this life.

References

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249 (2021).
2. Akhoon, N. Precision Medicine: A New Paradigm in Therapeutics. *Int J Prev Med* **12**, (2021).
3. Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B. & Richardson, W. S. Evidence based medicine: what it is and what it isn't. *BMJ : British Medical Journal* **312**, 71 (1996).
4. Blackstone, E. H. Precision Medicine Versus Evidence-Based Medicine: Individual Treatment Effect Versus Average Treatment Effect. *Circulation* **140**, 1236–1238 (2019).
5. Finn, O. J. Immuno-oncology: understanding the function and dysfunction of the immune system in cancer. *Annals of Oncology* **23**, viii6 (2012).
6. Burnet, M., Walter, F. R. S. & Hall, E. Cancer—A Biological Approach. *Br Med J* **1**, 779–786 (1957).
7. Ribatti, D. The concept of immune surveillance against tumors: The first theories. *Oncotarget* **8**, 7175 (2017).
8. Muppa, P. *et al.* Immune Cell Infiltration May Be a Key Determinant of Long-Term Survival in Small Cell Lung Cancer. *J Thorac Oncol* **14**, 1286–1295 (2019).
9. Huang, L. *et al.* Correlation of tumor-infiltrating immune cells of melanoma with overall survival by immunogenomic analysis. *Cancer Med* **9**, 8444 (2020).
10. Qin, Y. *et al.* Tumor-infiltrating B cells as a favorable prognostic biomarker in breast cancer: a systematic review and meta-analysis. *Cancer Cell Int* **21**, (2021).
11. Shankaran, V. *et al.* IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410**, 1107–1111 (2001).
12. Fridman, W. H. From Cancer Immune Surveillance to Cancer Immunoediting: Birth of Modern Immuno-Oncology. *The Journal of Immunology* **201**, 825–826 (2018).
13. Gubin, M. M. & Vesely, M. D. Cancer Immunoediting in the Era of Immuno-oncology. *Clin Cancer Res* **28**, 3917–3928 (2022).
14. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
15. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**, 31–46 (2022).

16. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
17. Chen, D. S. & Mellman, I. Oncology Meets Immunology: The Cancer-Immunity Cycle. *Immunity* **39**, 1–10 (2013).
18. Mohammed, R., Milne, A., Kayani, K. & Ojha, U. How the discovery of rituximab impacted the treatment of B-cell non-Hodgkin's lymphomas. *J Blood Med* **10**, 71 (2019).
19. Naithani, N., Sinha, S., Misra, P., Vasudevan, B. & Sahu, R. Precision medicine: Concept and tools. *Med J Armed Forces India* **77**, 249 (2021).
20. Shishido, S. N., Varahan, S., Yuan, K., Li, X. & Fleming, S. D. Humoral innate immune response and disease. *Clin Immunol* **144**, 142 (2012).
21. Wang, F. & Preininger, A. AI in Health: State of the Art, Challenges, and Future Directions. *Yearb Med Inform* **28**, 16 (2019).
22. Jerkeman, M. *et al.* Ibrutinib-Lenalidomide-Rituximab in Patients with Relapsed/Refractory Mantle Cell Lymphoma: Final Results from the Nordic Lymphoma Group MCL6 (PHILEMON) Phase II Trial. *Blood* **136**, 36–36 (2020).
23. Lokhande, L. *et al.* Serum proteome modulations upon treatment provides biological insight on response to treatment in relapsed mantle cell lymphoma. *Cancer Rep* **5**, e1524 (2021).
24. Glimelius, B. *et al.* U-CAN: a prospective longitudinal collection of biomaterials and clinical information from adult cancer patients in Sweden. *Acta Oncol (Madr)* **57**, 187–194 (2018).
25. Lokhande, L. *et al.* Immune-related protein signature in serum stratify relapsed mantle cell lymphoma patients based on risk. *BMC Cancer* **20**, (2020).
26. Rodrigues, J. M. *et al.* Infiltration of CD163-, PD-L1- and FoxP3-positive cells adversely affects outcome in patients with mantle cell lymphoma independent of established risk factors. *Br J Haematol* **193**, 520–531 (2021).
27. Li, P. *et al.* High Counts of CD68+ and CD163+ Macrophages in Mantle Cell Lymphoma Are Associated With Inferior Prognosis. *Front Oncol* **11**, (2021).
28. Brenchley, J. M. *et al.* Expression of CD57 defines replicative senescence and antigen-induced apoptotic death of CD8+ T cells. *Blood* **101**, 2711–2720 (2003).
29. Iida Masaki *et al.* Increase of Peripheral Blood CD57+T-Cells in Patients with Oral Squamous Cell Carcinoma. *Anticancer Res* **34**, 5729–5734 (2014).
30. Focosi, D., Bestagno, M., Burrone, O. & Petrini, M. CD57+ T lymphocytes and functional immune deficiency. *J Leukoc Biol* **87**, 107–116 (2010).
31. Andersson, E., Ohlin, M., Borrebaeck, C. A. K. & Carlsson, R. CD4+CD57+ T cells derived from peripheral blood do not support immunoglobulin production by B cells. *Cell Immunol* **163**, 245–253 (1995).
32. Hu, G. & Wang, S. Prognostic role of tumor-infiltrating CD57-positive lymphocytes in solid tumors: a meta-analysis. *Oncotarget* **9**, 8111 (2018).
33. Nogai, H., Dörken, B. & Lenz, G. Pathogenesis of non-Hodgkin's lymphoma. *Journal of Clinical Oncology* **29**, 1803–1811 (2011).

34. Shaffer, A. L., Young, R. M. & Staudt, L. M. Pathogenesis of Human B Cell Lymphomas. *Annu Rev Immunol* **30**, 565 (2012).
35. Armitage, J. O., Gascoyne, R. D., Lunning, M. A. & Cavalli, F. Non-Hodgkin lymphoma. *The Lancet* **390**, 298–310 (2017).
36. Armitage, J. O., Gale, R. P. & Jaffe, E. S. Lymphoma Nomenclature - What's in a name? *Br J Haematol* **197**, 539–543 (2022).
37. Aisenberg, A. C. Historical review of lymphomas. *Br J Haematol* **109**, 466–476 (2000).
38. Alaggio, R. *et al.* The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia 2022 36:7* **36**, 1720–1748 (2022).
39. Chiu, B. C. H. & Smith, S. M. Toward a global understanding of lymphoma: Epidemiologic clues from the second most populous country. *Leuk Lymphoma* **54**, 901–902 (2013).
40. Morton, L. M. *et al.* Etiologic Heterogeneity Among Non-Hodgkin Lymphoma Subtypes: The InterLymph Non-Hodgkin Lymphoma Subtypes Project. *JNCI Monographs* **2014**, 130–144 (2014).
41. Cerhan, J. R. & Slager, S. L. Familial predisposition and genetic risk factors for lymphoma. *Blood* **126**, 2265–2273 (2015).
42. Boffetta, P. I. Epidemiology of adult non-Hodgkin lymphoma. *Annals of Oncology* **22**, iv27–iv31 (2011).
43. Müller, A. M. S., Ihorst, G., Mertelsmann, R. & Engelhardt, M. Epidemiology of non-Hodgkin's lymphoma (NHL): Trends, geographic distribution, and etiology. *Ann Hematol* **84**, 1–12 (2005).
44. Bowzyk Al-Naeeb, A., Ajithkumar, T., Behan, S. & Hodson, D. J. Non-Hodgkin lymphoma. *BMJ* **362**, (2018).
45. Jaffe, E. S. & Pittaluga, S. Aggressive B-Cell Lymphomas: A Review of New and Old Entities in the WHO Classification. *Hematology Am Soc Hematol Educ Program* **2011**, 506 (2011).
46. Singh, R. *et al.* Non-Hodgkin's lymphoma: A review. *J Family Med Prim Care* **9**, 1834 (2020).
47. Vose, J. M. Mantle cell lymphoma: 2017 update on diagnosis, risk-stratification, and clinical management. *Am J Hematol* **92**, 806–813 (2017).
48. Thandra, K. C. *et al.* Epidemiology of Non-Hodgkin's Lymphoma. *Medical Sciences* **9**, 267–275 (2021).
49. Armitage, J. O. & Longo, D. L. Mantle-Cell Lymphoma. *N Engl J Med* **386**, 2495–2506 (2022).
50. Jain, P. & Wang, M. Mantle cell lymphoma: 2019 update on the diagnosis, pathogenesis, prognostication, and management. *Am J Hematol* **94**, 710–725 (2019).
51. Silkenstedt, E., Linton, K. & Dreyling, M. Mantle cell lymphoma – advances in molecular biology, prognostication and treatment approaches. *Br J Haematol* **195**, 162–173 (2021).

52. Teras, L. R. *et al.* 2016 US lymphoid malignancy statistics by World Health Organization subtypes. *CA Cancer J Clin* **66**, 443–459 (2016).
53. Wang, Y. & Ma, S. Risk Factors for Etiology and Prognosis of Mantle Cell Lymphoma. *Expert Rev Hematol* **7**, 233 (2014).
54. Weisenburger, D. D., Kim, H. & Rappaport, H. Mantle-Zone Lymphoma: A Follicular Variant of Intermediate L ymphocytic Lymphoma. doi:10.1002/1097-0142.
55. Pileri, S. A. & Falini, B. Mantle cell lymphoma. *Haematologica* **94**, 1488 (2009).
56. Jares, P. & Campo, E. Advances in the understanding of mantle cell lymphoma. *Br J Haematol* **142**, 149–165 (2008).
57. Samaha, H. *et al.* Mantle cell lymphoma: a retrospective study of 121 cases. *Leukemia 1998 12:8* **12**, 1281–1287 (1998).
58. Liu, Z. *et al.* CD5-Mantle Cell Lymphoma. *Am J Clin Pathol* **118**, 216–224 (2002).
59. Tarockoff, M., Gonzalez, T., Ivanov, S. & Sandoval-Sus, J. Mantle Cell Lymphoma: the Role of Risk-Adapted Therapy and Treatment of Relapsed Disease. *Curr Oncol Rep* **24**, 1313–1326 (2022).
60. Edwin, N. C. & Kahl, B. Evolving treatment strategies in mantle cell lymphoma. *Best Pract Res Clin Haematol* **31**, 270–278 (2018).
61. Salles, G. *et al.* Rituximab in B-Cell Hematologic Malignancies: A Review of 20 Years of Clinical Experience. *Adv Ther* **34**, 2232 (2017).
62. Eskelund, C. W. *et al.* Detailed Long-Term Follow-Up of Patients Who Relapsed After the Nordic Mantle Cell Lymphoma Trials: MCL2 and MCL3. *Hemasphere* **5**, 1 (2021).
63. Sandoval-Sus, J. D., Sotomayor, E. M. & Shah, B. D. Mantle Cell Lymphoma: Contemporary Diagnostic and Treatment Perspectives in the Age of Personalized Medicine. *Hematol Oncol Stem Cell Ther* **10**, 99–115 (2017).
64. Goy, A. Exploiting gene mutations and biomarkers to guide treatment recommendations in mantle cell lymphoma. *https://doi.org/10.1080/17474086.2021.1950529* **14**, 927–943 (2021).
65. Le, T., Chott, A., Krankenanstaltenverbund, W. & Pott, C. *Templated Nucleotide Addition and Immunoglobulin JH-Gene Utilization in t(11;14) Junctions: Implications for the Mechanism of Translocation and the Origin of Mantle Cell Lymphoma1 Article in Cancer Research*. https://www.researchgate.net/publication/12087740 (2001).
66. Li, J. Y. *et al.* Detection of Translocation t(11;14)(q13;q32) in Mantle Cell Lymphoma by Fluorescence in Situ Hybridization. *Am J Pathol* **154**, 1449 (1999).
67. Nadeu, F. *et al.* Genomic and epigenomic insights into the origin, pathogenesis, and clinical behavior of mantle cell lymphoma subtypes. *Blood* **136**, 1419 (2020).
68. Qie, S. & Diehl, J. A. Cyclin D1, Cancer Progression and Opportunities in Cancer Treatment. *J Mol Med (Berl)* **94**, 1313 (2016).
69. Alao, J. P. The regulation of cyclin D1 degradation: Roles in cancer development and the potential for therapeutic invention. *Mol Cancer* **6**, 1–16 (2007).

70. Sethi, S., Epstein-Peterson, Z., Kumar, A. & Ho, C. Current Knowledge in Genetics, Molecular Diagnostic Tools, and Treatments for Mantle Cell Lymphomas. *Front Oncol* **11**, 4835 (2021).
71. Narurkar, R., Alkayem, M. & Liu, D. SOX11 is a biomarker for cyclin D1-negative mantle cell lymphoma. *Biomark Res* **4**, 1–3 (2016).
72. Mozos, A. *et al.* SOX11 expression is highly specific for mantle cell lymphoma and identifies the cyclin D1-negative subtype. *Haematologica* **94**, 1555 (2009).
73. Ek, S., Dictor, M., Jerkeman, M., Jirström, K. & Borrebaeck, C. A. K. Nuclear expression of the non–B-cell lineage Sox11 transcription factor identifies mantle cell lymphoma. *Blood* **111**, 800–805 (2008).
74. Kuo, P. Y. *et al.* SOX11 augments BCR signaling to drive MCL-like tumor development. *Blood* **131**, 2247–2255 (2018).
75. Palomero, J. *et al.* SOX11 defines two different subtypes of mantle cell lymphoma through transcriptional regulation of BCL6. *Leukemia 2016 30:7* **30**, 1596–1599 (2015).
76. Vegliante, M. C. *et al.* SOX11 regulates PAX5 expression and blocks terminal B-cell differentiation in aggressive mantle cell lymphoma. *Blood* **121**, 2175–2185 (2013).
77. Wang, X. *et al.* The subcellular Sox11 distribution pattern identifies subsets of mantle cell lymphoma: correlation to overall survival. *Br J Haematol* **143**, 248–252 (2008).
78. Nygren, L. *et al.* Prognostic role of SOX11 in a population-based cohort of mantle cell lymphoma. *Blood* **119**, 4215–4223 (2012).
79. Fernàndez, V. *et al.* Genomic and gene expression profiling defines indolent forms of mantle cell lymphoma. *Cancer Res* **70**, 1408–1418 (2010).
80. Navarro, A. *et al.* Molecular subsets of mantle cell lymphoma defined by the IGHV mutational status and SOX11 expression have distinct biologic and clinical features. *Cancer Res* **72**, 5307–5316 (2012).
81. Ribera-Cortada, I. *et al.* Plasma cell and terminal B-cell differentiation in mantle cell lymphoma mainly occur in the SOX11-negative subtype. *Mod Pathol* **28**, 1435–1447 (2015).
82. Nordström, L. *et al.* SOX11 and TP53 add prognostic information to MIPI in a homogenously treated cohort of mantle cell lymphoma - a Nordic Lymphoma Group study. *Br J Haematol* **166**, 98–108 (2014).
83. Aukema, S. M. *et al.* Expression of TP53 is associated with the outcome of MCL independent of MIPI and Ki-67 in trials of the European MCL Network. *Blood* **131**, 417–420 (2018).
84. Sakhdari, A. *et al.* TP53 mutations are common in mantle cell lymphoma, including the indolent leukemic non-nodal variant. *Ann Diagn Pathol* **41**, 38–42 (2019).
85. Hill, H. A. *et al.* Genetic mutations and features of mantle cell lymphoma: a systematic review and meta-analysis. *Blood Adv* **4**, 2927 (2020).

86. Rodrigues, J. M., Porwit, A., Hassan, M., Ek, S. & Jerkeman, M. Targeted genomic investigations in a population-based cohort of mantle cell lymphoma reveal novel clinically relevant targets. *Leuk Lymphoma* **62**, 2637–2647 (2021).
87. Kridel, R. *et al.* Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. *Blood* **119**, 1963–1971 (2012).
88. Camacho, E. *et al.* ATM gene inactivation in mantle cell lymphoma mainly occurs by truncating mutations and missense mutations involving the phosphatidylinositol-3 kinase domain and is associated with increasing numbers of chromosomal imbalances. *Blood* **99**, 238–244 (2002).
89. Navarro, A., Beà, S., Jares, P. & Campo, E. Molecular pathogenesis of Mantle Cell Lymphoma. *Hematol Oncol Clin North Am* **34**, 795 (2020).
90. Bernard, S. *et al.* Inhibitors of BCR signalling interrupt the survival signal mediated by the micro-environment in mantle cell lymphoma. *Int J Cancer* **136**, 2761–2774 (2015).
91. Saleh, K. *et al.* Tumor Microenvironment and Immunotherapy-Based Approaches in Mantle Cell Lymphoma. *Cancers (Basel)* **14**, (2022).
92. Dreyling, M., Klapper, W. & Rule, S. Blastoid and pleomorphic mantle cell lymphoma: still a diagnostic and therapeutic challenge! *Blood* **132**, 2722–2729 (2018).
93. Jain, P. & Wang, M. Blastoid Mantle Cell Lymphoma. *Hematol Oncol Clin North Am* **34**, 941–956 (2020).
94. Tiemann, M. *et al.* Histopathology, cell proliferation indices and clinical outcome in 304 patients with mantle cell lymphoma (MCL): a clinicopathological study from the European MCL Network. *Br J Haematol* **131**, 29–38 (2005).
95. Hoster, E. *et al.* A new prognostic index (MIPI) for patients with advanced-stage mantle cell lymphoma. *Blood* **111**, 558–565 (2008).
96. Hoster, E. *et al.* Confirmation of the mantle-cell lymphoma International Prognostic Index in randomized trials of the European Mantle-Cell Lymphoma Network. *J Clin Oncol* **32**, 1338–1346 (2014).
97. Hoster, E. *et al.* Prognostic Value of Ki-67 Index, Cytology, and Growth Pattern in Mantle-Cell Lymphoma: Results From Randomized Trials of the European Mantle Cell Lymphoma Network. *J Clin Oncol* **34**, 1386–1394 (2016).
98. Shah, J. J., Fayad, L. & Romaguera, J. Mantle Cell International Prognostic Index (MIPI) not prognostic after R-hyper-CVAD. *Blood* **112**, 2583–2583 (2008).
99. Ferrero, S. *et al.* KMT2D mutations and TP53 disruptions are poor prognostic biomarkers in mantle cell lymphoma receiving high-dose therapy: a FIL study. *Haematologica* haematol.2018.214056 (2019) doi:10.3324/haematol.2018.214056.
100. Husby, S. *et al.* miR-18b overexpression identifies mantle cell lymphoma patients with poor outcome and improves the MIPI-B prognosticator. *Blood* **125**, 2669–2677 (2015).
101. Chihara, D. *et al.* Prognostic model for mantle cell lymphoma in the rituximab era: a nationwide study in Japan. *Br J Haematol* **170**, 657–668 (2015).

102. Villegas Da Ros, C. *et al.* Validation of R-MIPI and prognostic value of immunoglobulin light chain restriction in mantle cell lymphoma. *Br J Haematol* **177**, 816–818 (2017).
103. Yang, P. *et al.* Genomic landscape and prognostic analysis of mantle cell lymphoma. *Cancer Gene Therapy 2018 25:5* **25**, 129–140 (2018).
104. van de Schans, S. A. M., Janssen-Heijnen, M. L. G., Nijzie, M. R., Steyerberg, E. W. & van Spronsen, D. J. Validation, revision and extension of the Mantle Cell Lymphoma International Prognostic Index in a population-based setting. *Haematologica* **95**, 1503 (2010).
105. Kumar, A., Eyre, T. A., Lewis, K. L., Thompson, M. C. & Cheah, C. Y. New Directions for Mantle Cell Lymphoma in 2022. *Am Soc Clin Oncol Educ Book* **42**, 1–15 (2022).
106. Wu, X. *et al.* Association of minimal residual disease levels with clinical outcomes in patients with mantle cell lymphoma: A meta-analysis. *Leuk Res* **108**, 106605 (2021).
107. Eskelund, C. W. *et al.* TP53 mutations identify younger mantle cell lymphoma patients who do not benefit from intensive chemoimmunotherapy. *Blood* **130**, 1903–1910 (2017).
108. Mareckova, A. *et al.* ATM and TP53 mutations show mutual exclusivity but distinct clinical impact in mantle cell lymphoma patients. *https://doi.org/10.1080/10428194.2018.1542144* **60**, 1420–1428 (2019).
109. Choe, J. Y. *et al.* MYC overexpression correlates with MYC amplification or translocation, and is associated with poor prognosis in mantle cell lymphoma. *Histopathology* **68**, 442–449 (2016).
110. Yoo, C. *et al.* Serum beta-2 microglobulin as a prognostic biomarker in patients with mantle cell lymphoma. *Hematol Oncol* **34**, 22–27 (2016).
111. Sonbol, M. B. *et al.* Elevated Soluble IL-2Rα, IL-8, and MIP-1β Levels are Associated with Inferior Outcome and are Independent of MIPI Score in Patients with Mantle Cell Lymphoma. *Am J Hematol* **89**, E223 (2014).
112. Lakhotia, R. *et al.* Circulating tumor DNA predicts therapeutic outcome in mantle cell lymphoma. *Blood Adv* **6**, 2667–2680 (2022).
113. Roschewski, M. J. *et al.* Circulating tumor DNA to predict timing of relapse in mantle cell lymphoma. *https://doi.org/10.1200/JCO.2018.36.15_suppl.7576* **36**, 7576–7576 (2018).
114. Jing, C., Zheng, Y., Feng, Y., Cao, X. & Xu, C. Prognostic significance of p53, Sox11, and Pax5 co-expression in mantle cell lymphoma. *Sci Rep* **11**, 11896 (2021).
115. Rodrigues, J. M. *et al.* p53 is associated with high-risk and pinpoints TP53 missense mutations in mantle cell lymphoma. *Br J Haematol* **191**, 796–805 (2020).
116. Wang, J. D. *et al.* Proapoptotic protein BIM as a novel prognostic marker in mantle cell lymphoma. *Hum Pathol* **93**, 54 (2019).
117. Shin, S. J. *et al.* TCL1 expression predicts overall survival in patients with mantle cell lymphoma. *Eur J Haematol* **95**, 583–594 (2015).

118. Greenwell, I. B. *et al.* Complex karyotype in mantle cell lymphoma predicts inferior survival and poor response to intensive induction therapy. *Cancer* **124**, 2306 (2018).
119. Cohen, J. B. *et al.* Complex Karyotype Is Associated With Aggressive Disease and Shortened Progression-Free Survival in Patients With Newly Diagnosed Mantle Cell Lymphoma. *Clin Lymphoma Myeloma Leuk* **15**, 278-285.e1 (2015).
120. Sarkozy, C. *et al.* Complex karyotype in mantle cell lymphoma is a strong prognostic factor for the time to treatment and overall survival, independent of the MCL international prognostic index. *Genes Chromosomes Cancer* **53**, 106–116 (2014).
121. Karam, M. *et al.* FDG positron emission tomography/computed tomography scan may identify mantle cell lymphoma patients with unusually favorable outcome. *Nucl Med Commun* **30**, 770–778 (2009).
122. Bailly, C. *et al.* Prognostic value of FDG-PET in patients with mantle cell lymphoma: results from the LyMa-PET Project. *Haematologica* **105**, e33 (2020).
123. Bodet-Milin, C. *et al.* Prognostic impact of 18F-fluoro-deoxyglucose positron emission tomography in untreated mantle cell lymphoma: a retrospective study from the GOELAMS group. *Eur J Nucl Med Mol Imaging* **37**, 1633–1642 (2010).
124. Kedmi, M. *et al.* Is there a role for therapy response assessment with 2-[fluorine-18] fluoro-2-deoxy-D-glucose-positron emission tomography/computed tomography in mantle cell lymphoma? *Leuk Lymphoma* **55**, 2484–2489 (2014).
125. Hosein, P. J. *et al.* Utility of positron emission tomography scans in mantle cell lymphoma. *Am J Hematol* **86**, 841–845 (2011).
126. Yu, J. X., Hubbard-Lucey, V. M. & Tang, J. Immuno-oncology drug development goes global. *Nat Rev Drug Discov* **18**, 899–901 (2019).
127. Nygren, L. *et al.* T-cell levels are prognostic in mantle cell lymphoma. *Clin Cancer Res* **20**, 6096–6104 (2014).
128. Zhang, X. Y. *et al.* Negative prognostic impact of low absolute CD4+ T cell counts in peripheral blood in mantle cell lymphoma. *Cancer Sci* **107**, 1471–1476 (2016).
129. Zhou, X. H. *et al.* Low absolute NK cell counts in peripheral blood are associated with inferior survival in patients with mantle cell lymphoma. *Cancer Biomarkers* **24**, 439–447 (2019).
130. Balsas, P. *et al.* SOX11, CD70, and Treg cells configure the tumor-immune microenvironment of aggressive mantle cell lymphoma. *Blood* **138**, 2202–2215 (2021).
131. Koh, Y. W. *et al.* Absolute monocyte count predicts overall survival in mantle cell lymphomas: correlation with tumour-associated macrophages. *Hematol Oncol* **32**, 178–186 (2014).
132. George, A. *et al.* Prognostic impact of monocyte count at presentation in mantle cell lymphoma. *Br J Haematol* **164**, 890–893 (2014).
133. Porrata, L. F., Ristow, K. & Markovic, S. N. Absolute monocyte count at diagnosis and survival in mantle cell lymphoma. *Br J Haematol* **163**, 545–547 (2013).

134. Haydaroglu Sahin, H. Can the prognosis of mantle cell lymphoma be predicted by simple CBC counts? *Medicine* **98**, e16180 (2019).
135. Aprile von Hohenstaufen, K. *et al.* Prognostic impact of monocyte count at presentation in mantle cell lymphoma. *Br J Haematol* **162**, 465–473 (2013).
136. Drijvers, J. M., Sharpe, A. H. & Haigis, M. C. The effects of age and systemic metabolism on anti-tumor T cell responses. *Elife* **9**, 1–29 (2020).
137. Lv, H. *et al.* A novel clinical immune-related prognostic model predicts the overall survival of mantle cell lymphoma. *Hematol Oncol* **40**, 343–355 (2022).
138. Dahl, M. *et al.* Expression patterns and prognostic potential of circular RNAs in mantle cell lymphoma: a study of younger patients from the MCL2 and MCL3 clinical trials. *Leukemia 2021 36:1* **36**, 177–188 (2021).
139. Bomben, R. *et al.* A B-cell receptor-related gene signature predicts survival in mantle cell lymphoma: results from the Fondazione Italiana Linfomi MCL-0208 trial. *Haematologica* **103**, 849–856 (2018).
140. Scott, D. W. *et al.* New Molecular Assay for the Proliferation Signature in Mantle Cell Lymphoma Applicable to Formalin-Fixed Paraffin-Embedded Biopsies. *J Clin Oncol* **35**, 1668–1677 (2017).
141. Rauert-Wunderlich, H. *et al.* Validation of the MCL35 gene expression proliferation assay in randomized trials of the European Mantle Cell Lymphoma Network. *Br J Haematol* **184**, 616–624 (2019).
142. Holte, H. *et al.* The MCL35 gene expression proliferation assay predicts high-risk MCL patients in a Norwegian cohort of younger patients given intensive first line therapy. *Br J Haematol* **183**, 225–234 (2018).
143. Freeman, C. L. *et al.* Molecular determinants of outcomes in relapsed or refractory mantle cell lymphoma treated with ibrutinib or temsirolimus in the MCL3001 (RAY) trial. *Leukemia* **36**, 2479–2487 (2022).
144. Hermine, O. *et al.* Addition of high-dose cytarabine to immunochemotherapy before autologous stem-cell transplantation in patients aged 65 years or younger with mantle cell lymphoma (MCL Younger): a randomised, open-label, phase 3 trial of the European Mantle Cell Lymphoma Network. *Lancet* **388**, 565–575 (2016).
145. Eskelund, C. W. *et al.* 15-year follow-up of the Second Nordic Mantle Cell Lymphoma trial (MCL2): prolonged remissions without survival plateau. *Br J Haematol* **175**, 410–418 (2016).
146. Geisler, C. H. *et al.* Nordic MCL2 trial update: six-year follow-up after intensive immunochemotherapy for untreated mantle cell lymphoma followed by BEAM or BEAC + autologous stem-cell support: still very long survival but late relapses do occur. *Br J Haematol* **158**, 355–362 (2012).
147. Alnassfan, T. *et al.* Mantle cell lymphoma treatment options for elderly/unfit patients: A systematic review. *EJHaem* **3**, 276–290 (2022).
148. Rule, S. *et al.* Ibrutinib for the treatment of relapsed/refractory mantle cell lymphoma: extended 3.5-year follow up from a pooled analysis. *Haematologica* **104**, e211 (2019).

149. Wang, M. *et al.* Primary results from the double-blind, placebo-controlled, phase III SHINE study of ibrutinib in combination with bendamustine-rituximab (BR) and R maintenance as a first-line treatment for older patients with mantle cell lymphoma (MCL). *https://doi.org/10.1200/JCO.2022.40.17_suppl.LBA7502* **40**, LBA7502–LBA7502 (2022).
150. Al-Mansour, M. Treatment Landscape of Relapsed/Refractory Mantle Cell Lymphoma: An Updated Review. *Clin Lymphoma Myeloma Leuk* (2022) doi:10.1016/J.CLML.2022.07.017.
151. Pu, J. J., Savani, M., Huang, N. & Epner, E. M. Mantle cell lymphoma management trends and novel agents: where are we going? *Ther Adv Hematol* **13**, (2022).
152. Lewis, K. L. & Cheah, C. Y. Non-Covalent BTK Inhibitors—The New BTKids on the Block for B-Cell Malignancies. *J Pers Med* **11**, (2021).
153. Wang, M. *et al.* KTE-X19 CAR T-Cell Therapy in Relapsed or Refractory Mantle-Cell Lymphoma. *N Engl J Med* **382**, 1331 (2020).
154. Petersohn, S. *et al.* Cost-effectiveness analysis of KTE-X19 CAR T therapy versus real-world standard of care in patients with relapsed/refractory mantle cell lymphoma post BTKi in England. *J Med Econ* **25**, 730–740 (2022).
155. Minson, A. *et al.* A Phase II, Open-Label, Single Arm Trial to Assess the Efficacy and Safety of the Combination of Tisagenlecleucel and Ibrutinib in Mantle Cell Lymphoma (TARMAC). *Blood* **136**, 34–35 (2020).
156. Hammons, L. & Fenske, T. S. Treatment of Mantle Cell Lymphoma in the Frontline Setting: Are We Ready for a Risk-Adapted Approach? *J Pers Med* **12**, (2022).
157. Moraes, F. & Góes, A. A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. *Biochem Mol Biol Educ* **44**, 215–223 (2016).
158. Yates, J. R. Recent technical advances in proteomics. *F1000Res* **8**, (2019).
159. Pandey, A. & Mann, M. Proteomics to study genes and genomes. *Nature 2000 405:6788* **405**, 837–846 (2000).
160. Micheel, C. M. *et al.* Omics-Based Clinical Discovery: Science, Technology, and Applications. (2012).
161. Quezada, H., Guzmán-Ortiz, A. L., Díaz-Sánchez, H., Valle-Rios, R. & Aguirre-Hernández, J. Omics-based biomarkers: current status and potential use in the clinic. *Boletín Médico Del Hospital Infantil de México (English Edition)* **74**, 219–226 (2017).
162. Zhang, A. H., Sun, H., Yan, G. L., Han, Y. & Wang, X. J. Serum Proteomics in Biomedical Research: A Systematic Review. *Applied Biochemistry and Biotechnology 2013 170:4* **170**, 774–786 (2013).
163. McVeigh, T. P. & Kerin, M. J. Clinical use of the Oncotype DX genomic test to guide treatment decisions for patients with invasive breast cancer. *Breast Cancer: Targets and Therapy* **9**, 393–400 (2017).

164. Atkinson, A. J. *et al.* Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin Pharmacol Ther* **69**, 89–95 (2001).
165. Silver Spring (MD): Food and Drug Administration (US); Bethesda (MD): National Institutes of Health (US). BEST (Biomarkers, EndpointS, and other Tools) Resource. *BEST (Biomarkers , EndpointS , and other Tools) Resource* (2016).
166. Robb, M. A., McInnes, P. M. & Califf, R. M. Biomarkers and Surrogate Endpoints: Developing Common Terminology and Definitions. *JAMA* **315**, 1107–1108 (2016).
167. Califf, R. M. Biomarker definitions and their applications. *Exp Biol Med* **243**, 213 (2018).
168. Duffy, M. J. Biomarkers for prostate cancer: Prostate-specific antigen and beyond. *Clin Chem Lab Med* **58**, 326–339 (2020).
169. Enblad, A. P. *et al.* PSA testing patterns in a large Swedish cohort before the implementation of organized PSA testing. *Scand J Urol* **54**, 376–381 (2020).
170. Casaubon, J. T., Kashyap, S. & Regan, J.-P. BRCA 1 and 2. *StatPearls* (2022).
171. Effery, J. *et al.* The Risk of Cancer Associated with Specific Mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *https://doi.org/10.1056/NEJM199705153362001* **336**, 1401–1408 (1997).
172. Thorlacius, S. *et al.* Population-based study of risk of breast cancer in carriers of BRCA2 mutation. *Lancet* **352**, 1337–1339 (1998).
173. Lee, E. Y. & Kulkarni, R. P. Circulating biomarkers predictive of tumor response to cancer immunotherapy. *Expert Rev Mol Diagn* **19**, 895 (2019).
174. Poulet, G., Massias, J. & Taly, V. Liquid Biopsy: General Concepts. *Acta Cytol* **63**, 449–455 (2019).
175. Xie, M., Huang, X., Ye, X. & Qian, W. Prognostic and clinicopathological significance of PD-1/PD-L1 expression in the tumor microenvironment and neoplastic cells for lymphoma. *Int Immunopharmacol* **77**, 105999 (2019).
176. Calapre, L., Warburton, L., Millward, M., Ziman, M. & Gray, E. S. Circulating tumour DNA (ctDNA) as a liquid biopsy for melanoma. *Cancer Lett* **404**, 62–69 (2017).
177. de Kock, R. *et al.* Circulating biomarkers for monitoring therapy response and detection of disease progression in lung cancer patients. *Cancer Treat Res Commun* **28**, (2021).
178. Teutsch, S. M. *et al.* The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. *Genetics in Medicine* **11**, 3 (2009).
179. Henry, N. L. & Hayes, D. F. Cancer biomarkers. *Mol Oncol* **6**, 140 (2012).
180. Ou, F. S., Michiels, S., Shyr, Y., Adjei, A. A. & Oberg, A. L. Biomarker Discovery and Validation: Statistical Considerations. *Journal of Thoracic Oncology* **16**, 537–545 (2021).
181. Hayes, D. F. Biomarker validation and testing. *Mol Oncol* **9**, 960 (2015).

182. Ransohoff, D. F. & Gourlay, M. L. Sources of Bias in Specimens for Research About Molecular Markers for Cancer. *Journal of Clinical Oncology* **28**, 698 (2010).
183. McLerran, D. *et al.* SELDI-TOF MS whole serum proteomic profiling with IMAC surface does not reliably detect prostate cancer. *Clin Chem* **54**, 53–60 (2008).
184. Ghannad, M., Olsen, M., Boutron, I. & Bossuyt, P. M. A systematic review finds that spin or interpretation bias is abundant in evaluations of ovarian cancer biomarkers. *J Clin Epidemiol* **116**, 9–17 (2019).
185. Borrebaeck, C. A. K. Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer. *Nature Reviews Cancer 2017 17:3* **17**, 199–204 (2017).
186. Lagani, V., Kortas, G. & Tsamardinos, I. BIOMARKER SIGNATURE IDENTIFICATION IN "OMICS" DATA WITH MULTI-CLASS OUTCOME. *Comput Struct Biotechnol J* **6**, e201303004 (2013).
187. Liberto, J. M. *et al.* Current and Emerging Methods for Ovarian Cancer Screening and Diagnostics: A Comprehensive Review. *Cancers (Basel)* **14**, (2022).
188. Sparano, J. A. *et al.* Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *New England Journal of Medicine* **373**, 2005–2014 (2015).
189. Cheong, J. H. *et al.* Development and validation of a prognostic and predictive 32-gene signature for gastric cancer. *Nature Communications 2022 13:1* **13**, 1–9 (2022).
190. Lai, J., Xu, T. & Yang, H. Protein-based prognostic signature for predicting the survival and immunotherapeutic efficiency of endometrial carcinoma. *BMC Cancer* **22**, 1–17 (2022).
191. Chung, L. *et al.* Novel serum protein biomarker panel revealed by mass spectrometry and its prognostic value in breast cancer. *Breast Cancer Research* **16**, 1–12 (2014).
192. Cantafio, M. E. G. *et al.* From Single Level Analysis to Multi-Omics Integrative Approaches: A Powerful Strategy towards the Precision Oncology. *High Throughput* **7**, (2018).
193. Olivier, M., Asmis, R., Hawkins, G. A., Howard, T. D. & Cox, L. A. The Need for Multi-Omics Biomarker Signatures in Precision Medicine. *Int J Mol Sci* **20**, (2019).
194. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biol* **18**, (2017).
195. Correa-Aguila, R., Alonso-Pupo, N. & Hernández-Rodríguez, E. W. Multi-omics data integration approaches for precision oncology. *Mol Omics* **18**, 469–479 (2022).
196. Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).
197. Zhang, H. *et al.* Integrated proteogenomic characterization of human high grade serous ovarian cancer. *Cell* **166**, 755 (2016).

198. Li, L. *et al.* Integrated Omic analysis of lung cancer reveals metabolism proteome signatures with prognostic impact. *Nature Communications 2014 5:1* **5**, 1–12 (2014).
199. Boufraqech, M. & Nilubol, N. Multi-omics Signatures and Translational Potential to Improve Thyroid Cancer Patient Outcome. *Cancers (Basel)* **11**, (2019).
200. Javaeed, A., Ghauri, S. K., Ibrahim, A. & Doheim, M. F. Prostate-specific antigen velocity in diagnosis and prognosis of prostate cancer - a systematic review. *Oncol Rev* **14**, 64–71 (2020).
201. Nelson, T. J. *et al.* Association of Prostate-Specific Antigen Velocity With Clinical Progression Among African American and Non-Hispanic White Men Treated for Low-Risk Prostate Cancer With Active Surveillance. *JAMA Netw Open* **4**, e219452–e219452 (2021).
202. Patel, H. D. *et al.* Prostate specific antigen velocity risk count predicts biopsy reclassification for men with very low risk prostate cancer. *J Urol* **191**, 629–637 (2014).
203. Flores-Fraile, M. C. *et al.* The Association between Prostate-Specific Antigen Velocity (PSAV), Value and Acceleration, and of the Free PSA/Total PSA Index or Ratio, with Prostate Conditions. *J Clin Med* **9**, 1–14 (2020).
204. Hüttenhain, R., Malmström, J., Picotti, P. & Aebersold, R. Perspectives of targeted mass spectrometry for protein biomarker verification. *Curr Opin Chem Biol* **13**, 518 (2009).
205. Tighe, P. J., Ryder, R. R., Todd, I. & Fairclough, L. C. ELISA in the multiplex era: Potentials and pitfalls. *Proteomics Clin Appl* **9**, 406–422 (2015).
206. Steinhauer, C. *et al.* Improved affinity coupling for antibody microarrays: Engineering of double-(His)6-tagged single framework recombinant antibody fragments. *Proteomics* **6**, 4227–4234 (2006).
207. Carlsson, A. *et al.* Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases. *Proc Natl Acad Sci U S A* **108**, 14252–14257 (2011).
208. Brand, R. E. *et al.* Detection of Early-Stage Pancreatic Ductal Adenocarcinoma from Blood Samples: Results of a Multiplex Biomarker Signature Validation Study. *Clin Transl Gastroenterol* **13**, E00468 (2022).
209. Wingren, C. *et al.* Identification of Serum Biomarker Signatures Associated with Pancreatic Cancer. *Cancer Res* **72**, 2481–2490 (2012).
210. Gerdtsson, A. S. *et al.* A Multicenter Trial Defining a Serum Protein Signature Associated with Pancreatic Ductal Adenocarcinoma. *Int J Proteomics* **2015**, 1–10 (2015).
211. Ingvarsson, J. *et al.* Detection of pancreatic cancer using antibody microarray-based serum protein profiling. *Proteomics* **8**, 2211–2219 (2008).
212. Mellby, L. D. *et al.* Serum Biomarker Signature-Based Liquid Biopsy for Diagnosis of Early-Stage Pancreatic Cancer. *J Clin Oncol* **36**, 2887–2894 (2018).
213. Kuci Emruli, V. *et al.* Identification of a serum biomarker signature associated with metastatic prostate cancer. *Proteomics Clin Appl* **15**, e2000025 (2021).

214. Nordström, M. *et al.* Identification of plasma protein profiles associated with risk groups of prostate cancer patients. *Proteomics Clin Appl* **8**, 951–962 (2014).
215. Pauly, F. *et al.* Plasma immunoprofiling of patients with high-risk diffuse large B-cell lymphoma: a Nordic Lymphoma Group study. *Blood Cancer J* **6**, e501 (2016).
216. Pauly, F. *et al.* Identification of B-cell lymphoma subsets by plasma protein profiling using recombinant antibody microarrays. *Leuk Res* **38**, 682–690 (2014).
217. Sandström, A. *et al.* Serum proteome profiling of pancreatitis using recombinant antibody microarrays reveals disease-associated biomarker signatures. *Proteomics Clin Appl* **6**, 486–496 (2012).
218. Carlsson, A. *et al.* Serum Protein Profiling of Systemic Lupus Erythematosus and Systemic Sclerosis Using Recombinant Antibody Microarrays. *Mol Cell Proteomics* **10**, (2011).
219. Borrebaeck, C. A. K., Sturfelt, G. & Wingren, C. Recombinant antibody microarray for profiling the serum proteome of SLE. *Methods Mol Biol* **1134**, 67–78 (2014).
220. Wu, Y., Cheng, Y., Wang, X., Fan, J. & Gao, Q. Spatial omics: Navigating to the golden era of cancer research. *Clin Transl Med* **12**, (2022).
221. Reynolds, B. A., Oli, M. W. & Oli, M. K. Eco-oncology: Applying ecological principles to understand and manage cancer. *Ecol Evol* **10**, 8538–8553 (2020).
222. Thorsson, V. *et al.* The Immune Landscape of Cancer. *Immunity* **48**, 812 (2018).
223. Zhou, C., Liu, Q., Xiang, Y., Gou, X. & Li, W. Role of the tumor immune microenvironment in tumor immunotherapy (Review). *Oncol Lett* **23**, 1–7 (2022).
224. Lyons, Y. A., Wu, S. Y., Overwijk, W. W., Baggerly, K. A. & Sood, A. K. Immune cell profiling in cancer: molecular approaches to cell-specific identification. *NPJ Precis Oncol* **1**, 26 (2017).
225. Chuah, S. & Chew, V. High-dimensional immune-profiling in cancer: implications for immunotherapy. *J Immunother Cancer* **8**, e000363 (2020).
226. Lewis, S. M. *et al.* Spatial omics and multiplexed imaging to explore cancer biology. *Nature Methods 2021 18:9* **18**, 997–1012 (2021).
227. Wang, N., Li, X., Wang, R. & Ding, Z. Spatial transcriptomics and proteomics technologies for deconvoluting the tumor microenvironment. *Biotechnol J* **16**, 2100041 (2021).
228. Method of the Year 2020: spatially resolved transcriptomics. *Nature Methods 2020 18:1* **18**, 1–1 (2021).
229. Van, T. M. & Blank, C. U. A user's perspective on GeoMxTM digital spatial profiling. *Immuno-Oncology Technology* **1**, 11–18 (2019).
230. Merritt, C. R. *et al.* Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nature Biotechnology 2020 38:5* **38**, 586–599 (2020).
231. Larroquette, M. *et al.* Spatial transcriptomics of macrophage infiltration in non-small cell lung cancer reveals determinants of sensitivity and resistance to anti-PD1/PD-L1 antibodies. *J Immunother Cancer* **10**, e003890 (2022).

232. Moutafi, M. K. *et al.* Spatially resolved proteomic profiling identifies tumor cell CD44 as a biomarker associated with sensitivity to PD-1 axis blockade in advanced non-small-cell lung cancer. *J Immunother Cancer* **10**, (2022).
233. Yerly, L. *et al.* Integrated multi-omics reveals cellular and molecular interactions governing the invasive niche of basal cell carcinoma. *Nat Commun* **13**, (2022).
234. Jing, Y., Yang, J., Johnson, D. B., Moslehi, J. J. & Han, L. Harnessing big data to characterize immune-related adverse events. *Nature Reviews Clinical Oncology 2022 19:4* **19**, 269–280 (2022).
235. Stephens, Z. D. *et al.* Big Data: Astronomical or Genomical? *PLoS Biol* **13**, e1002195 (2015).
236. Clough, E. & Barrett, T. The Gene Expression Omnibus database. *Methods Mol Biol* **1418**, 93 (2016).
237. Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat Genet* **45**, 1113 (2013).
238. Zhang, J. *et al.* International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. *Database (Oxford)* **2011**, (2011).
239. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941–D947 (2019).
240. Willems, S. M. *et al.* The potential use of big data in oncology. *Oral Oncol* **98**, 8–12 (2019).
241. Fessele, K. L. The Rise of Big Data in Oncology. *Semin Oncol Nurs* **34**, 168–176 (2018).
242. Marx, V. The big challenges of big data. *Nature 2013 498:7453* **498**, 255–260 (2013).
243. García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M. & Herrera, F. Big data preprocessing: methods and prospects. *Big Data Analytics 2016 1:1* **1**, 1–22 (2016).
244. de Livera, A. M. *et al.* Statistical methods for handling unwanted variation in metabolomics data. *Anal Chem* **87**, 3606 (2015).
245. Goh, W. W. bin, Wang, W. & Wong, L. Why Batch Effects Matter in Omics Data, and How to Avoid Them. *Trends Biotechnol* **35**, 498–507 (2017).
246. Zhou, L., Chi-Hau Sue, A. & bin Goh, W. W. Examining the practical limits of batch effect-correction algorithms: When should you care about batch effects? *Journal of Genetics and Genomics* **46**, 433–443 (2019).
247. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* **3**, 1724–1735 (2007).
248. Sims, A. H. *et al.* The removal of multiplicative, systematic bias allows integration of breast cancer gene expression datasets – improving meta-analysis and prediction of prognosis. *BMC Medical Genomics 2008 1:1* **1**, 1–14 (2008).
249. Molania, R. *et al.* Removing unwanted variation from large-scale RNA sequencing data with PRPS. *Nature Biotechnology 2022* 1–14 (2022) doi:10.1038/s41587-022-01440-w.

250. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
251. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* **2**, (2020).
252. Willforss, J., Chawade, A. & Levander, F. NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis. *J Proteome Res* **18**, 732–740 (2019).
253. Aittokallio, T. Dealing with missing values in large-scale studies: microarray data imputation and beyond. *Brief Bioinform* **11**, 253–264 (2010).
254. Song, M. *et al.* A Review of Integrative Imputation for Multi-Omics Datasets. *Front Genet* **11**, 1215 (2020).
255. Kuhn, M., Johnson, K. & Modeling, P. Data Pre-processing. *Applied Predictive Modeling* 27–59 (2013) doi:10.1007/978-1-4614-6849-3_3.
256. Eckmann, J. P. & Tlusty, T. Dimensional reduction in complex living systems: Where, why, and how. *BioEssays* **43**, 2100062 (2021).
257. Zhang, Y. *et al.* Post hoc power analysis: is it an informative and meaningful analysis? *Gen Psychiatr* **32**, e100069 (2019).
258. Kim, S. Y. Effects of sample size on robustness and prediction accuracy of a prognostic gene signature. *BMC Bioinformatics* **10**, 1–10 (2009).
259. Kirpich, A. *et al.* Variable selection in omics data: A practical evaluation of small sample sizes. *PLoS One* **13**, (2018).
260. Xu, C. & Jackson, S. A. Machine learning and complex biological data. *Genome Biol* **20**, 1–4 (2019).
261. Pate, A., Emsley, R., Sperrin, M., Martin, G. P. & van Staa, T. Impact of sample size on the stability of risk scores from clinical prediction models: a case study in cardiovascular disease. *Diagnostic and Prognostic Research 2020 4:1* **4**, 1–12 (2020).
262. Berisha, V. *et al.* Digital medicine and the curse of dimensionality. *NPJ Digit Med* **4**, (2021).
263. Mirza, B. *et al.* Machine Learning and Integrative Analysis of Biomedical Big Data. *Genes (Basel)* **10**, (2019).
264. Lynch, D. T., Koya, S., Acharya, U. & Kumar, A. Mantle Cell Lymphoma. *StatPearls* (2022).
265. Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLoS One* **14**, e0224365 (2019).
266. Phan, J. H., Quo, C. F., Cheng, C. & Wang, M. D. Multiscale integration of -omic, imaging, and clinical data in biomedical informatics. *IEEE Rev Biomed Eng* **5**, 74–87 (2012).
267. McKinney, S. M. *et al.* International evaluation of an AI system for breast cancer screening. *Nature 2020 577:7788* **577**, 89–94 (2020).
268. Tran, K. A. *et al.* Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med* **13**, (2021).

269. Sanchez-Pinto, L. N., Luo, Y. & Churpek, M. M. Big Data and Data Science in Critical Care. *Chest* **154**, 1239 (2018).
270. Shan, Z. *et al.* Proteomic profiling reveals a signature for optimizing prognostic prediction in Colon Cancer. *J Cancer* **12**, 2199 (2021).
271. Das, J., Gayvert, K. M., Bunea, F., Wegkamp, M. H. & Yu, H. ENCAPP: elastic-net-based prognosis prediction and biomarker discovery for human cancers. *BMC Genomics* **16**, (2015).
272. Fukushima, A., Sugimoto, M., Hiwa, S. & Hiroyasu, T. Elastic net-based prediction of IFN-β treatment response of patients with multiple sclerosis using time series microarray gene expression profiles. *Scientific Reports 2019 9:1* **9**, 1–11 (2019).
273. Issa, I. I. *et al.* A bendamustine resistance gene signature in diffuse large B-cell lymphoma and multiple myeloma. *Cancer Drug Resistance* **4**, 208–222 (2021).
274. Kuang, Z., Li, X., Liu, R., Chen, S. & Tu, J. Comprehensive Characterization of Cachexia-Inducing Factors in Diffuse Large B-Cell Lymphoma Reveals a Molecular Subtype and a Prognosis-Related Signature. *Front Cell Dev Biol* **9**, (2021).
275. Cox, D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)* **34**, 187–202 (1972).
276. Oberg, A. L. & Mahoney, D. W. Linear mixed effects models. *Methods Mol Biol* **404**, 213–234 (2007).
277. Bono, R., Alarcón, R. & Blanca, M. J. Report Quality of Generalized Linear Mixed Models in Psychology: A Systematic Review. *Front Psychol* **12**, 1345 (2021).
278. Li, J., Lu, Q. & Wen, Y. Multi-kernel linear mixed model with adaptive lasso for prediction analysis on high-dimensional multi-omics data. *Bioinformatics* **36**, 1785 (2020).
279. Schelldorfer, J., Bühlmann, P. & de Geer, S. van. Estimation for High-Dimensional Linear Mixed-Effects Models Using ℓ1-Penalization. *Scandinavian Journal of Statistics* **38**, 197–214 (2011).
280. Therneau, T. Mixed Eects Cox Models. (2022).
281. Rizwan I Haque, I. & Neubert, J. Deep learning approaches to biomedical image segmentation. *Inform Med Unlocked* **18**, 100297 (2020).
282. Kowal, M., Żejmo, M., Skobel, M., Korbicz, J. & Monczak, R. Cell Nuclei Segmentation in Cytological Images Using Convolutional Neural Network and Seeded Watershed Algorithm. *J Digit Imaging* **33**, 231 (2020).
283. Thakur, N., Yoon, H. & Chong, Y. Current Trends of Artificial Intelligence for Colorectal Cancer Pathology Image Analysis: A Systematic Review. *Cancers (Basel)* **12**, 1–19 (2020).
284. Ladd, A. M. & Diehl, D. L. Artificial intelligence for early detection of pancreatic adenocarcinoma: The future is promising. *World J Gastroenterol* **27**, 1283 (2021).

285. Ström, P. *et al.* Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol* **21**, 222–232 (2020).
286. Ehteshami Bejnordi, B. *et al.* Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Mod Pathol* **31**, 1502–1512 (2018).
287. Zhu, Y. *et al.* SIO: A Spatioimageomics Pipeline to Identify Prognostic Biomarkers Associated with the Ovarian Tumor Microenvironment. *Cancers (Basel)* **13**, (2021).
288. van Valen, D. A. *et al.* Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments. *PLoS Comput Biol* **12**, (2016).
289. Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. Star-convex polyhedra for 3D object detection and segmentation in microscopy. *Proceedings - 2020 IEEE Winter Conference on Applications of Computer Vision, WACV 2020* 3655–3662 (2020) doi:10.1109/WACV45572.2020.9093435.
290. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. 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 (2018).
291. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* doi:10.1038/s41592-020-01018-x.
292. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat Methods* **16**, 1226–1232 (2019).
293. Caicedo, J. C. *et al.* Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nat Methods* **16**, 1247–1253 (2019).
294. Hollandi, R. *et al.* nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer. *Cell Syst* **10**, 453-458.e6 (2020).
295. Blombery, P. & Cheah, C. Y. Predicting the future in MCL with MRD. *Blood* **140**, 1332–1333 (2022).
296. Hoster, E. & Pott, C. Minimal residual disease in mantle cell lymphoma: insights into biology and impact on treatment. *Hematology: the American Society of Hematology Education Program* **2016**, 437 (2016).
297. Galimberti, S. *et al.* The Minimal Residual Disease in Non-Hodgkin's Lymphomas: From the Laboratory to the Clinical Practice. *Front Oncol* **9**, 528 (2019).
298. Ladetto, M., Tavarozzi, R. & Pott, C. Minimal residual disease (MRD) in mantle cell lymphoma. *Ann Lymphoma* **4**, 4–4 (2020).
299. Jerkeman, M. *et al.* Venetoclax, Lenalidomide and Rituximab for Patients with Relapsed or Refractory Mantle Cell Lymphoma - Data from the Nordic Lymphoma Group NLG-MCL7 (VALERIA) Phase I Trial: Stopping Treatment in Molecular Remission Is Feasible. *Blood* **136**, 15 (2020).

300. van Dam, S., Baars, M. J. D. & Vercoulen, Y. Multiplex Tissue Imaging: Spatial Revelations in the Tumor Microenvironment. *Cancers 2022, Vol. 14, Page 3170* **14**, 3170 (2022).
301. Fu, T. *et al.* Spatial architecture of the immune microenvironment orchestrates tumor immunity and therapeutic response. *Journal of Hematology & Oncology 2021 14:1* **14**, 1–25 (2021).
302. Cai, Z., Poulos, R. C., Liu, J. & Zhong, Q. Machine learning for multi-omics data integration in cancer. *iScience* **25**, 103798 (2022).
303. Malik, V., Kalakoti, Y. & Sundar, D. Deep learning assisted multi-omics integration for survival and drug-response prediction in breast cancer. *BMC Genomics* **22**, 1–11 (2021).
304. Yang, Y., Tian, S., Qiu, Y., Zhao, P. & Zou, Q. MDICC: novel method for multi-omics data integration and cancer subtype identification. *Brief Bioinform* **23**, (2022).
305. He, S. *et al.* High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nature Biotechnology 2022 40:12* **40**, 1794–1806 (2022).
306. Baxi, V., Edwards, R., Montalto, M. & Saha, S. Digital pathology and artificial intelligence in translational medicine and clinical practice. *Modern Pathology 2021 35:1* **35**, 23–32 (2021).