

***Tumour microenvironment immune features predict outcomes in
Diffuse Large B-Cell Lymphoma***

Inaugural Dissertation

zur

Erlangung des Doktorgrades

Dr. nat. med.

der Medizinischen Fakultät und

der Mathematisch-Naturwissenschaftlichen Fakultät der

Universität zu Köln

vorgelegt von

Marta Pistone

aus Catania in Italien

Köln

2026

Betreuer/in: Prof. Dr. Reinhard Büttner

Referenten: Prof. Dr. Matthias Fischer
Prof. Dr. Achim Tresch

Datum der mündlichen Prüfung: 25.03.2026

Table of Contents

1. Zusammenfassung.....	6
2. Introduction	8
3. Aim of the study	22
4. Materials and Methods	23
4.1 Patient cohorts and inclusion criteria:.....	23
4.2 Data collection and sources	23
4.3 Sample processing and molecular profiling.....	24
4.4 DLBCL Proteogenotype classification and RNA-seq normalization	24
4.5 Table2. Data integration and sample annotation	25
4.6 Whole-Exome Sequencing (WES)	28
4.6.1 Evaluation and visualization of somatically acquired alterations	29
4.6.2 Table 3. Reagents and kits for DNA extraction and WES	31
4.6.3 Bulk-RNA sequencing (RNA-seq).....	32
4.6.4 RNA-seq sample processing and quantification	33
4.6.5 Table 4. Reagents and kits for RNA extraction and RNA-seq.....	33
4.6.6 RNA-Seq differential gene expression and signature gene analysis	34
4.6.7 Immune cell profiling via deconvolution of bulk transcriptomic data	34
4.6.8 Table 5. Bioinformatic tools	35
4.7 Multiplex Immunofluorescence (mIF) Staining	36
4.7.1 Table 6. Reagents used for mIF	37
4.7.2 Table 7. Antibodies	38
4.7.3 Multiplexed-imaging pre-processing and segmentation	39
4.8 Table 8. Abbreviations	40
5. Results	45
5.1 Overview of the DLBCL study cohort and multi-omic data integration. ...	46
5.2 Distribution of proteogenotypes in the cohort.....	46
5.2.1 Effect of the vote-threshold filtering on PG assignments	47
5.2.2 Survival analysis across PG subgroups.....	47
5.3 Spatial characterization of the tumour microenvironment using multiplex immunofluorescence.....	49
5.3.1 Representative mIF images across proteogenotypes	49
5.3.2 Deep-learning-based single-cell segmentation and reconstructed cellular maps	50

5.3.3 Global cell-type composition across proteogenotypes	50
5.3.4 Direct comparison of immune composition between PG4 and PG6	50
5.3.5 Tumour- centred spatial interactions in PG4 vs PG6.....	51
5.4 Differential gene expression between PG4 and other PGs.....	54
5.4.1 Enrichment of biological processes in PG4 vs others	54
5.4.2 Reduced immune-related transcriptional activity in PG4.....	55
5.4.3 Immune-related gene expression profiles distinguish PG4 from other PGs	55
5.5 Focused transcriptomic analysis of PG4 versus PG6	57
5.5.1 Differential gene expression reveals a strong immunological imbalance ..	57
5.5.2 Downregulated pathways in PG4 highlight impaired immune competence	58
5.5.3 Network analysis consolidates functional coherence of suppressed programs.....	58
5.4.4 Immune-response signature scores sharply separate PG4 and PG6	58
5.5.5 Heatmap of immune genes highlights the PG4-PG6 immune dichotomy .	59
5.6 Immune cell composition highlights distinct microenvironmental states in PG4	61
5.6.1 PG4 exhibits reduced abundance of multiple immune populations compared with other PGs	61
5.6.2 Direct comparison with PG6 confirms severe immune depletion in PG4 ..	62
5.6.3 Macrophage polarization reveals functional differences in PG4	62
5.6.4 Macrophage imbalance becomes significant in PG4 versus PG6.....	62
5.7 MYC target gene activation in PG4 compared with all other proteogenotypes	64
5.7.1 PG4 shows significantly elevated MYC target gene expression relative to PG other	64
5.7.2 Coordinated upregulation of MYC target genes in PG4	64
5.8 MYC target activation differentiates PG4 from the immune-enriched PG6 proteogenotype.....	66
5.8.1 Downregulated pathways in PG6 highlight functional consequences of MYC activation in PG4.....	66
5.8.2 PG4 exhibits markedly higher expression of HALLMARK MYC target genes	66
5.8.3 MYC target gene heatmap confirms coordinated upregulation in PG4	67
5.9 Somatic alterations across proteogenotypes	69

6. Discussion	71
6.1. Overview of the key findings and their clinical - biological relevance	71
6.2. Spatial and microenvironmental architecture of PG4 and PG6	74
6.3. Transcriptomic divergence between PG4 and PG6: Molecular determinants of prognostic polarization	76
6.4. MYC-driven biology and its role in shaping the PG4 phenotype	79
6.5. Genomic landscape of PG4 and PG6: Limited contribution of genetic alterations to proteogenotype identity	81
6.6. Integrated interpretation and clinical implications	84
6.7. Limitations and future directions	85
7. Conclusion.....	88
8. Bibliography.....	90
9. Erklärung	97

1. Zusammenfassung

Das diffus großzellige B-Zell-Lymphom (DLBCL) stellt den häufigsten Subtyp der Non-Hodgkin-Lymphome dar und zeichnet sich durch eine ausgeprägte biologische und klinische Heterogenität aus. Trotz bedeutender Fortschritte in der molekularen Charakterisierung zeigt ein erheblicher Anteil der Patientinnen und Patienten weiterhin ungünstige Verläufe, was die Notwendigkeit einer verbesserten prognostischen Stratifizierung und eines vertieften biologischen Verständnisses der Erkrankung unterstreicht. In der vorliegenden Arbeit analysierten wir eine Kohorte von 106 DLBCL-Patienten, aus deren formalinfixierten, paraffineingebetteten (FFPE) Tumourblöcken sowohl RNA als auch Gewebeschnitte gewonnen wurden. Aufbauend auf einem prognostischen Modell der Forschungsgruppe von Thomas Oellerich in Frankfurt, das DLBCL-Fälle in sieben prognostische Gruppen (PG1-PG7) klassifiziert, konzentrierten wir uns auf die Untersuchung der biologischen Unterschiede zwischen der Gruppe mit der ungünstigsten Prognose (PG4) und derjenigen mit der günstigsten Prognose (PG6). Zu diesem Zweck führten wir Bulk-RNA-Sequenzierung sowie multiplexe Immunfluoreszenzfärbungen (mIF) mit einem Panel von 20 Biomarkern durch, um eine integrierte molekulare und räumliche Charakterisierung der Tumormikroumgebung (TME) zu ermöglichen.

Die Transkriptomanalyse zeigte, dass immunbezogene Gene in PG4 im Vergleich zu PG6 signifikant herunterreguliert waren, was auf eine insgesamt unterdrückte Immunaktivität in der PG4-Untergruppe hindeutet. Die Pfadanalyse ergab eine verringerte Aktivierung immunologischer Signalwege, einschließlich der T-Zell-Rezeptor-Signalgebung, der Interferonantwort und der Antigenpräsentation. Die räumliche Analyse mittels mIF in Kombination mit maschinellem Lernen bestätigte, dass die TME von PG4-Fällen einen „immune-cold“-Phänotyp aufweist, charakterisiert durch eine reduzierte T-Zell-Infiltration, einen erhöhten Anteil M2-ähnlicher Makrophagen und eine verminderte Interaktion zwischen Tumour- und Immunzellen. Im Gegensatz dazu zeigten PG6-Proben einen „immune-hot“-Phänotyp mit einer ausgeprägten Präsenz zytotoxischer T-Zellen, einer aktiven Immun-Signalgebung und einer höheren Interaktionsrate zwischen Tumour- und Immunzellen.

Zusammenfassend zeigen unsere Ergebnisse eine deutliche immunologische Dichotomie zwischen den prognostischen Gruppen PG4 und PG6 im DLBCL und

verknüpfen transkriptionelle sowie räumliche Immunmerkmale mit dem klinischen Outcome. Die Befunde unterstützen das Konzept, dass ein immunsupprimiertes bzw. „immune-cold“-Mikromilieu zur schlechten Prognose der PG4-Patienten beiträgt, während eine „immune-hot“ und funktionell aktive TME den günstigen Verlauf in PG6 erklärt. Dieser integrierte molekulare und räumliche Ansatz liefert neue Einblicke in die Immunbiologie des DLBCL und hebt potenzielle Ansatzpunkte für immungerichtete therapeutische Strategien in Hochrisikogruppen hervor.

2. Introduction

Diffuse large B-cell lymphoma (DLBCL) is widely recognized as the most common subtype of non-Hodgkin's lymphoma, accounting for approximately 30-40% of all newly diagnosed cases worldwide (1,2). Despite being classified as a single histopathological entity, DLBCL encompasses a broad spectrum of biologically and clinically heterogeneous diseases, which differ markedly in their molecular features, clinical presentation, and response to therapy (3,4).

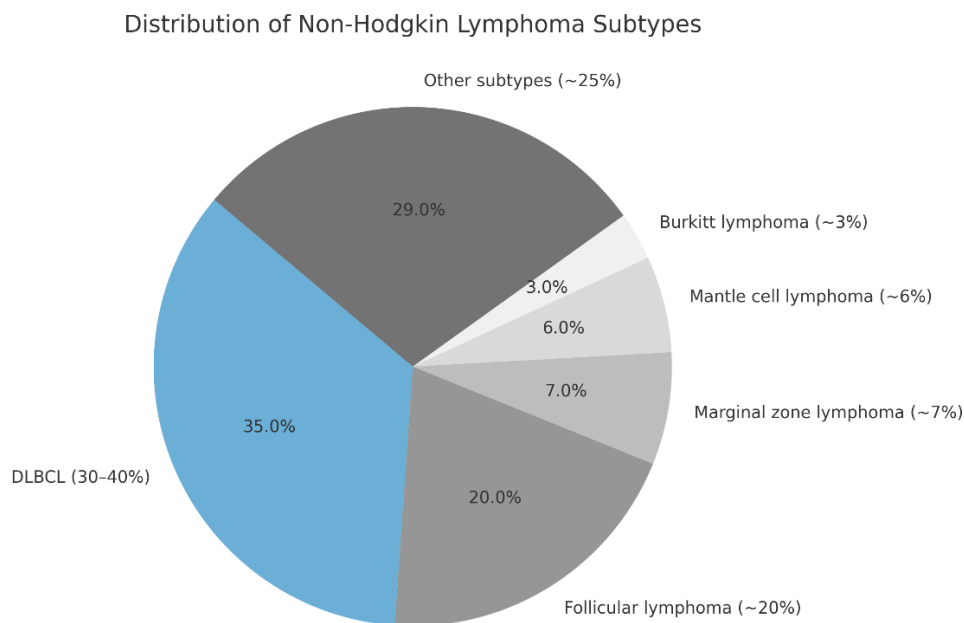


Figure 1. Distribution of non-Hodgkin lymphoma (NHL) subtypes. Diffuse large B-cell lymphoma (DLBCL) accounts for approximately 30-40% of all NHL cases, representing the most frequent subtype (2,5).

DLBCL is an aggressive disease that affects predominantly older adults, but can also be found in younger populations (2-4). The biological and clinical heterogeneity of DLBCL has posed significant challenges for diagnosis, disease prognosis, and treatment (2,6). While some patients experience durable remission after first-line therapy, a substantial proportion relapse or are refractory to available regimens, underscoring the urgent need for more precise molecular characterization and innovative therapeutic approaches (4,7). In the early 2000s, the introduction of rituximab combined with the CHOP chemotherapy backbone (cyclophosphamide, doxorubicin, vincristine, and prednisone), revolutionized the management of DLBCL

(2,8). The so-called R-CHOP regimen marked a treatment turning point, significantly improving survival rates compared to chemotherapy alone (2). However, despite these advances, around one third of patients fail to respond adequately or relapse within a few years. For these patients, prognosis remains poor, with median survival of less than one year(4).

Novel salvage therapies such as high-dose chemotherapy followed by autologous stem cell transplantation (ASCT) have been effective for chemo-sensitive relapsed disease, but not all patients are eligible due to age or comorbidities (9). Recently, immunotherapeutic approaches such as CAR-T cells have shown efficacy in relapsed/refractory settings, with long-term remissions observed in a subset of patients(7,10,11). This highlights the fact that DLBCL is not a single disease entity, but a constellation of biologically diverse malignancies that share morphological features at microscopic scale. Understanding this diversity has therefore become a major focus of modern lymphoma research.

Molecular profiling has played a pivotal role in dissecting the biology of DLBCL. The landmark study by Alizadeh et al. in 2000 introduced the concept of gene-expression profiling for lymphoma classification, distinguishing between the germinal centre B-cell like (GCB) and activated B-cell like (ABC) subtypes (6,12).

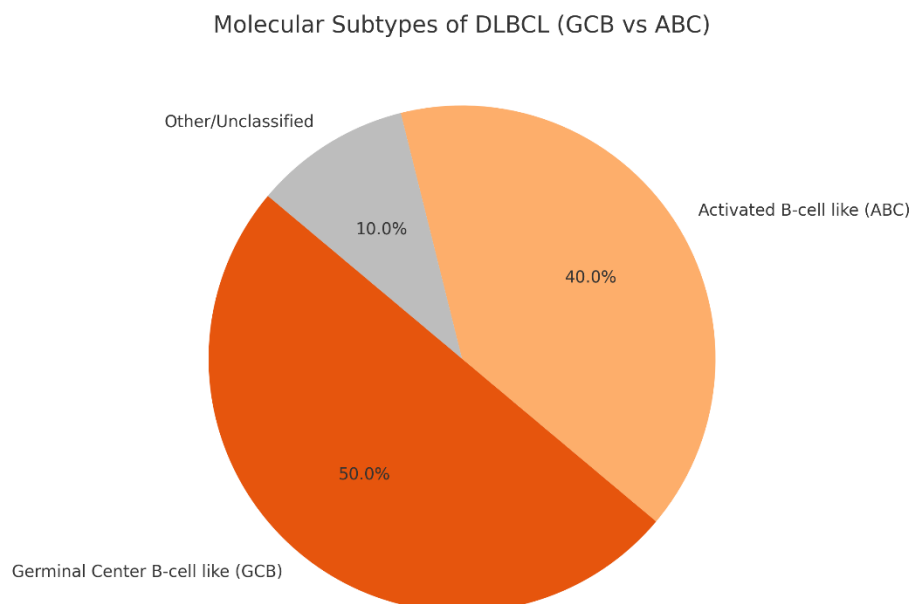


Figure 2. Evolution of molecular classification in DLBCL. Initial classification distinguished germinal centre B-cell like (GCB) and activated B-cell like (ABC) subtypes (6). More recent

genomic studies identified additional molecular subgroups(13), highlighting the biological heterogeneity of the disease.

These subtypes differ in their presumed cell of origin, signalling pathway activation, and clinical outcome. Patients with GCB-type disease tend to be better than those with ABC-type, who often display constitutive NF- κ B activation and a poorer prognosis (2,12). Recently, high-throughput sequencing studies have refined this model further, identifying additional molecular subgroups characterized by distinct genetic alterations, such as “double-hit” lymphomas with MYC and BCL2 or BCL6 rearrangements, and “double-expressor” lymphomas overexpressing MYC and BCL2 proteins (13,14).

Feature	GCB (Germinal Centre B-cell like)	ABC (Activated B-cell like)
Cell of origin	Germinal centre B cells	Post-germinal centre, plasmablastic B cells
Common genetic alterations	EZH2, BCL2, CREBBP, TNFRSF14	MYD88, CD79B, CARD11, PRDM1
Signalling pathways	BCL2-driven survival, PI3K/AKT	Chronic active BCR signalling, NF- κ B activation
Gene-expression profile	Resembles normal germinal centre B cells	Resembles activated peripheral blood B cells
Prognosis (with R-CHOP)	Generally favourable	Poorer outcome, higher relapse rate
Therapeutic implications	Sensitive to EZH2 inhibitors, BCL2-targeted therapy	Potential benefit from BTK inhibitors (e.g., ibrutinib)

Table 1. Key biological and clinical differences between germinal centre B-cell like (GCB) and activated B-cell like (ABC) subtypes of diffuse large B-cell lymphoma.(6,12,13) .

Building on this molecular framework, a recent study by Enßle *et al.* (under revision, 2025) proposed an integrated proteogenomic classification of DLBCL, which identified seven distinct proteogenotypes (PG1-PG7) that transcend cell-of-origin categories and genetic subtypes. Each proteogenotype (PG) is defined by characteristic transcriptional, proteomic, and microenvironmental features with prognostic relevance. Among these, PG4 emerged as a high-risk subgroup associated with inferior overall and progression-free survival, independent of the International Prognostic Index and cell-of-origin classification. In contrast, PG6 displayed an immune-rich phenotype and superior clinical outcomes. Mechanistically, PG4 tumours exhibit a dark zone-like B-cell phenotype, enhanced MYC and BCR/PI3K signalling, and augmented ribosomal and translational activity, driven by transcription factors such as TCF3/TCF4 and XBP1. These molecular programs converge to promote lymphoma cell proliferation and metabolic activation, while simultaneously fostering immune evasion. Indeed, both bulk and single-cell analyses revealed that PG4 is characterized by a depleted and exhausted T-cell compartment, particularly loss of CD4⁺ and cytotoxic CD8⁺ subsets, and enrichment of exhausted effector-memory T cells, defining an “immune-cold” microenvironment. In contrast, PG6 features an “immune-hot” landscape with high infiltration of activated T cells, macrophages, and antigen-presenting cells. These findings underscore the importance of integrating transcriptomic, proteomic, and spatial data to unravel disease heterogeneity in DLBCL. Building on this proteogenotypic framework, the analytical strategy of this thesis adopts a stepwise comparative approach to dissect the biological underpinnings of prognostic divergence in DLBCL. Because PG4 represents the poorest-outcome subgroup and PG6 the most immune-enriched and clinically favourable one, our analyses were designed to reflect both global and focused perspectives. First, PG4 was compared with the aggregated group of all remaining proteogenotypes (“PG other”) to obtain a broad overview of the transcriptional and immunological features that differentiate PG4 from the rest of the DLBCL landscape. This initial comparison captures the general molecular programs that define PG4 as a biologically adverse entity. Subsequently, we refined our investigation by concentrating specifically on the PG4 versus PG6 contrast, which represents the most biologically polarized comparison and thus offers the clearest insight into mechanisms underlying immune depletion and unfavourable prognosis (15). This two-tiered structure provides both breadth and

resolution in characterizing the interplay between MYC-driven tumour-intrinsic programs and microenvironmental immune activity across DLBCL. The identification of PG4 and PG6 as biologically and immunologically distinct entities provides a conceptual basis for microenvironment-driven prognostic stratification and rational immune-targeted therapeutic development.

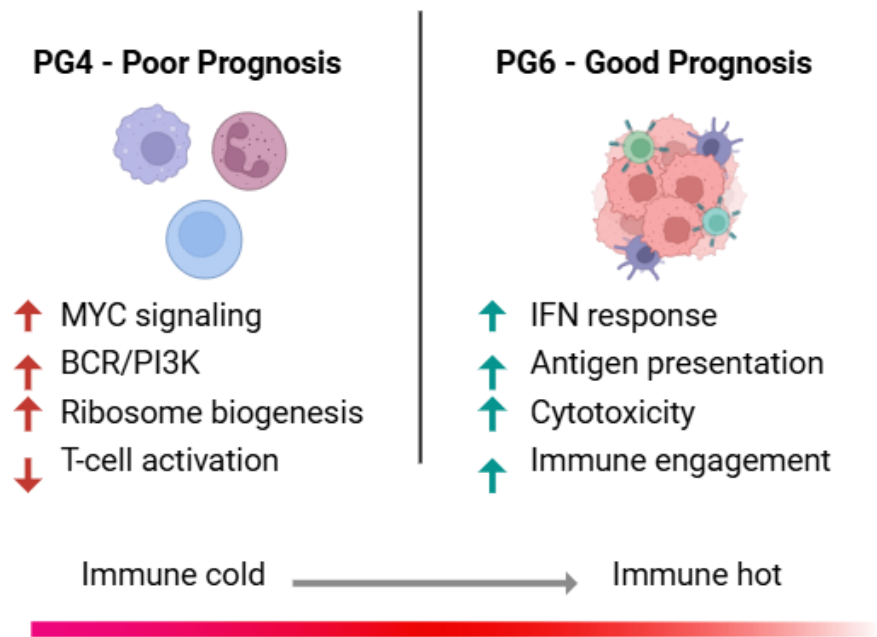


Figure 3. Schematic representation of the contrasting biological and immunological features of proteogenotypes PG4 and PG6 identified by Enßle et al. (manuscript in revision, 2025). PG4, associated with poor prognosis, displays a transcriptional program dominated by MYC and BCR/PI3K signalling, increased ribosome biogenesis, and reduced T-cell activation, resulting in an “immune-cold” microenvironment. In contrast, PG6, linked to favourable outcome, exhibits enhanced interferon response, antigen presentation, cytotoxicity, and immune engagement, consistent with an “immune-hot” phenotype. The gradient below illustrates the continuum of immune activity across the seven DLBCL proteogenotypes (PG1-PG7) (15) (BioRender).

The present thesis builds upon this framework to further explore, at both transcriptional and spatial levels, how the tumour microenvironment contributes to poor versus favourable prognostic phenotypes in DLBCL(15).

Despite progress in molecular classification, there remains a critical gap between biological insights and clinical application. Not all patients within a given molecular category behave uniformly, suggesting that additional factors shape disease course. Greater attention has turned to the tumour microenvironment (TME) as a determinant of lymphoma biology and therapy response(1,14). Far from being a passive background, the TME actively interacts with tumour cells, influencing proliferation, survival, immune evasion, and resistance to therapy (1,14). The TME of DLBCL includes immune cells such as T lymphocytes, macrophages, dendritic cells, and natural killer (NK) cells, as well as stromal components and the extracellular matrix (16). CD8+ cytotoxic T lymphocytes are central to anti-tumour immunity, yet in many DLBCL cases they are functionally exhausted, expressing inhibitory receptors and lacking effective cytolytic function (14). Regulatory T cells (Tregs) and tumour-associated macrophages (TAMs) contribute to an immunosuppressive milieu that favours tumour progression(1,16). Stromal cells secrete cytokines, chemokines, and growth factors that shape the tumour niche. Collectively, these interactions define an ecological system in which tumour cells and their environment co-evolve (1). Dendritic cells and NK cells contribute to immune regulation, while stromal fibroblasts remodel the extracellular matrix and secrete cytokines. Kotlov et al. (17) proposed a classification of DLBCL based on immune landscapes, showing that TME composition strongly correlates with prognosis. Steen et al. (14) further expanded this by identifying up to 44 cellular states using single-cell approaches, supporting an ecological model of lymphoma ecosystems.

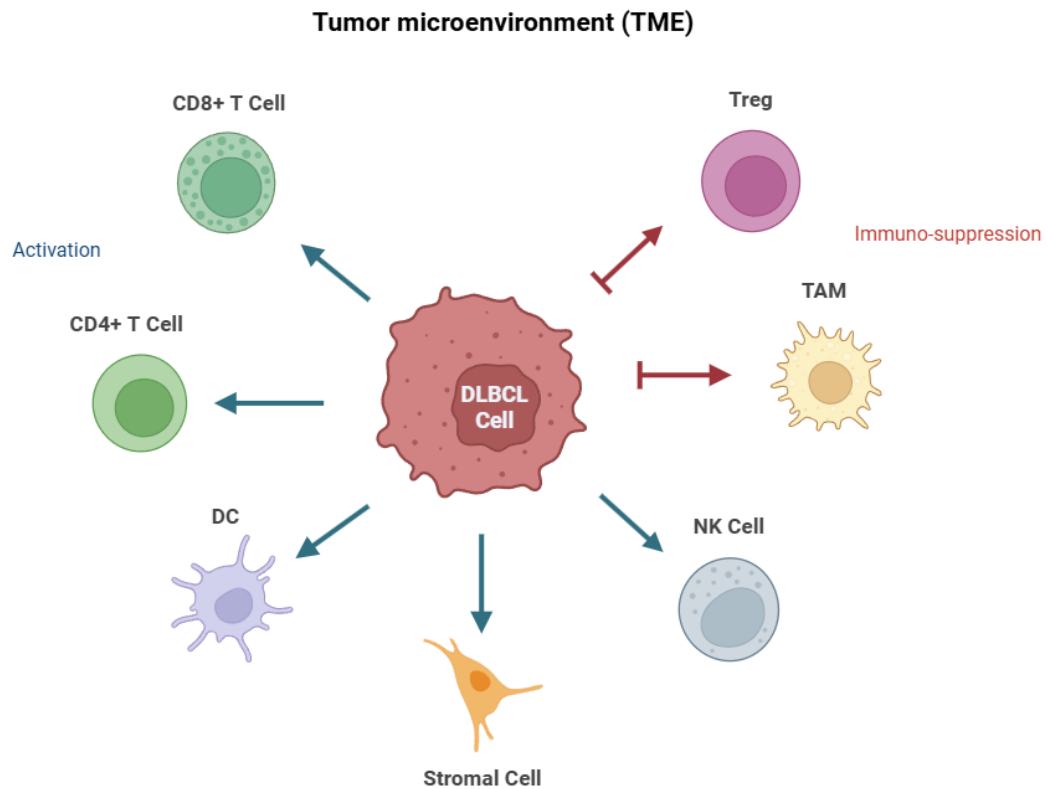


Figure 4. Schematic representation of the diffuse large B-cell lymphoma (DLBCL) tumour microenvironment. The ecosystem includes malignant B cells, immune populations (T cells, macrophages, dendritic cells (DC), NK cells), and stromal components. These interactions influence tumour growth, immune evasion, and response to therapy. (BioRender)

Recent single-cell RNA sequencing (scRNA-seq) studies have demonstrated remarkable diversity in both malignant and non-malignant cell populations within DLBCL, supporting the view that tumour ecosystems are as important as intrinsic tumour genetics (14). RNA sequencing (RNA-seq) has become the central technology in modern transcriptomics, offering high-resolution and quantitative analysis of gene expression(18). Bulk RNA-seq provides a global snapshot of transcript abundance across thousands of genes simultaneously (18). Compared to microarray technology, RNA-seq offers superior sensitivity, broader dynamic range, and the ability to detect novel transcripts, isoforms, and non-coding RNAs (18). In the context of DLBCL, bulk RNA-seq has enabled the identification of gene-expression signatures associated with prognosis, drug resistance, and pathway activation (6,13,14).

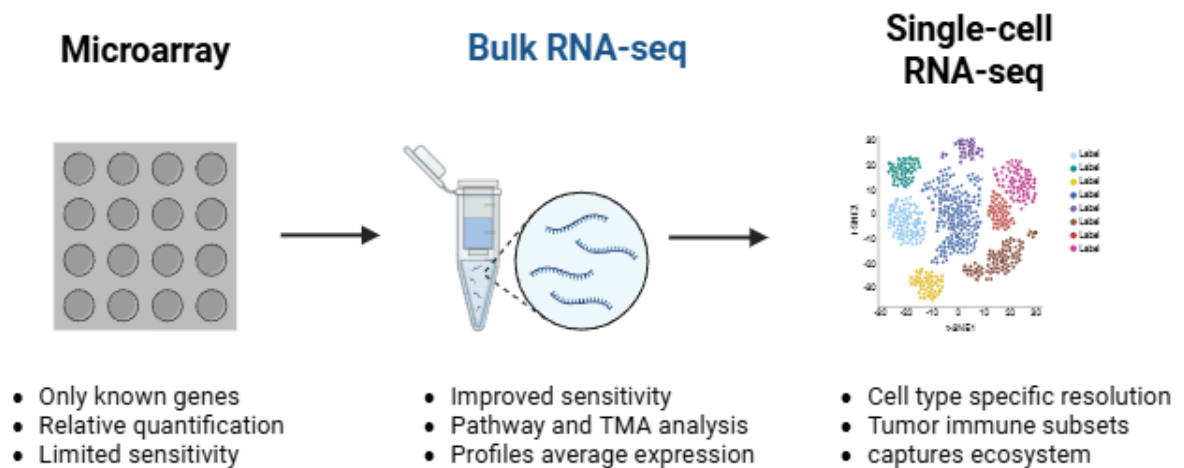


Figure 5. Evolution of transcriptomic technologies in lymphoma research. From early microarray studies to bulk and single-cell RNA sequencing, the resolution of transcriptomic profiling has progressively increased, allowing deeper insight into tumour heterogeneity and microenvironmental interactions. Adapted from Wang et al., 2009 (18) and Papalexi & Satija, 2018 (19) (Biorender).

Importantly, bulk RNA-seq data can also be leveraged to infer the composition of the TME using computational deconvolution algorithms such as CIBERSORT, xCell, and TIMER (20). These methods estimate the proportions of immune cell subsets present in bulk tissue samples, providing insights into how immune infiltration correlates with disease progression and patient outcome(20).

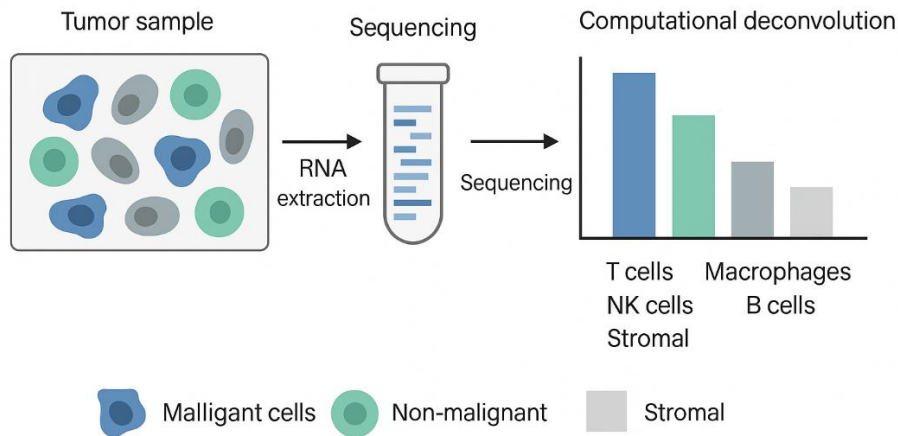


Figure 6. Computational deconvolution of bulk RNA-seq data. Transcriptomic data from tumour samples containing a mixture of malignant and non-malignant cells can be analysed using algorithms such as CIBERSORT, TIMER, or xCell to infer the relative abundance of immune and stromal cell subsets within the tumour microenvironment. Adapted from Newman et al., 2015(20) (BioRender).

Nevertheless, bulk RNA-seq has inherent limitations, because it represents an average signal across thousands or millions of cells, it masks cellular heterogeneity(14,19). Single-cell RNA-seq has emerged as a transformative technology to overcome this issue, capturing transcriptomic profiles at single-cell resolution. In DLBCL, scRNA-seq has revealed distinct malignant cell states and heterogeneous populations of infiltrating immune cells, including exhausted T cells and immunosuppressive macrophages (14,19). These findings underscore that understanding lymphoma requires not only molecular taxonomy of the tumour cells themselves, but also comprehensive mapping of the ecosystem in which they reside(14,17). Another critical aspect of modern oncology research is early detection and real-time monitoring of disease. Liquid biopsies, such as circulating tumour DNA (ctDNA) and circulating tumour cells (CTCs), are emerging as powerful tools for tracking tumour evolution (21). The ASCO Educational Book (2015) highlighted how ctDNA can serve as a biomarker for tumour burden, minimal residual disease, and treatment response (4). In DLBCL, ctDNA profiling has shown strong correlation with outcomes and provides a powerful complement to transcriptomic analyses. Rossi et

al. (8,21) emphasized that integrating ctDNA with tissue-based omics could enhance disease monitoring and guide therapy decisions. Such multimodal integration may offer a dynamic picture of disease evolution during treatment (4,21).

From a therapeutic perspective, the interplay between tumour cells and the TME has profound implications (1,14). Immunotherapies such as checkpoint inhibitors (anti-PD-1, anti-CTLA-4) and chimeric antigen receptor T cells (CAR-T) have shown promise in relapsed/refractory DLBCL, yet their success is often limited by microenvironmental barriers (14,22). CAR-T therapy targeting CD19 has transformed the management of refractory DLBCL, achieving durable responses in subsets of patients (7,11). However, resistance to CAR-T therapy is influenced by microenvironmental immunosuppression, antigen loss, and adaptive tumour mechanisms(7,11). For instance, the presence of suppressive macrophages and Tregs can blunt T cell responses, while stromal signalling pathways may foster resistance(14,16). Therefore, transcriptomic approaches that integrate tumour-intrinsic and microenvironmental features are essential for predicting immunotherapy outcomes and designing rational combinatorial strategies (1,4). Despite these advances, several challenges remain unresolved. First, most studies have investigated RNA-seq, TME, and liquid biopsy data in isolation, without integrating them into a unified framework (14,17,21). Second, transcriptional signatures often lack validation in large, independent cohorts, limiting their immediate clinical applicability(12,17). Third, the complexity of the TME demands sophisticated computational approaches to disentangle signals from malignant versus non-malignant cells(14,20). Emerging spatial transcriptomics and multiplex immunofluorescence approaches provide opportunities to address this gap (23). Multiplex immunofluorescence (mIF) has recently emerged as a powerful spatial profiling technology that enables simultaneous visualization and quantification of multiple biomarkers on a single formalin-fixed paraffin-embedded (FFPE) tissue section while preserving the spatial architecture of the tumour microenvironment (24,25).

Compared to conventional single-plex immunohistochemistry (IHC), mIF allows for high-dimensional characterization of immune, stromal, and tumour cell phenotypes within their native histologic context, which is essential for decoding the cellular ecosystems that shape DLBCL biology and therapeutic response(26,27). Technically, mIF workflows rely on iterative staining cycles or spectrally distinct

fluorophores, combined with multispectral imaging and computational unmixing to separate overlapping emission spectra. Tyramide Signal Amplification (TSA)-based approaches, such as the Akoya Opal platform, achieve high sensitivity even for low-abundance markers by covalently depositing fluorophores that remain bound after antibody stripping, allowing successive staining rounds on the same slide(24,27). Alternatively, cyclic immunofluorescence (cyclIF) and CODEX (Co-Detection by Indexing) use repeated rounds of staining, imaging, and fluorophore inactivation or barcode readout, enabling the detection of tens to hundreds of proteins per section with single-cell spatial resolution(25,28).

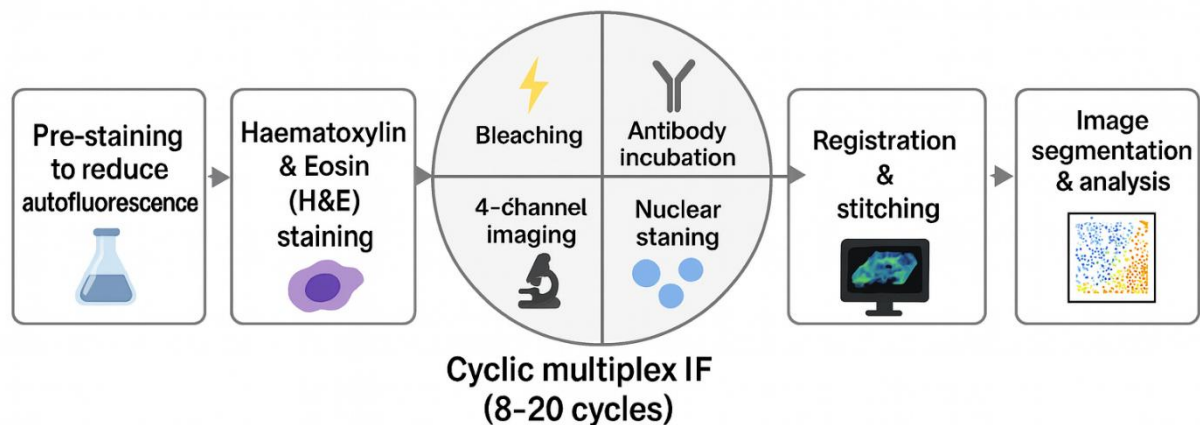


Figure 7. Schematic representation of the *t*-cyclIF workflow showing key steps, including pre-staining to reduce autofluorescence, haematoxylin and eosin (H&E) staining, cyclic antibody incubation and imaging (8-20 cycles), bleaching, and nuclear staining, followed by image registration, stitching, segmentation, and computational analysis (BioRender).

From a biological perspective, mIF provides crucial insight into the composition, activation status, and spatial organization of immune cell populations, including cytotoxic T cells, regulatory T cells, macrophages, and checkpoint molecules

expressing subsets(17,29). Recent studies have shown that spatial metrics such as immune-cell proximity to tumour cells, or spatial clustering of suppressive macrophages correlate with response to immunotherapy and overall survival in DLBCL and other lymphomas(30,31). Moreover, standardized multicentre validation studies (e.g., MITRE trial) have demonstrated that mIF can achieve inter-laboratory reproducibility comparable to quantitative assays such as ELISA, supporting its translational potential (32). Despite its advantages, mIF still faces challenges related to spectral overlap, autofluorescence, tissue degradation after multiple staining cycles, and computational complexity(25,27). However, continuous advances in image acquisition, fluorophore chemistry, and AI-driven image analysis pipelines are rapidly improving signal fidelity, segmentation accuracy, and analytical reproducibility (33-35). These developments position multiplex immunofluorescence as an indispensable tool for spatially resolved immuno-oncology and for integrating morphological and molecular layers of data in precision lymphoma research(17,35). The rationale of this thesis is therefore to employ a multimodal strategy combining bulk RNA-seq and immunofluorescence staining to dissect the DLBCL microenvironment and identify predictive markers of prognosis and therapy response.

This doctoral thesis is situated at the intersection of molecular profiling and tumour ecology. It seeks to characterize the transcriptomic landscape of DLBCL using bulk RNA-seq, with a particular focus on pathways implicated in tumour-microenvironment interactions. Furthermore, it aims to analyse the composition and functional diversity of the TME, leveraging both computational deconvolution of bulk data and insights from single-cell studies. Ultimately, this work aspires to propose an integrated model associating molecular and ecological features with clinical outcomes, thereby contributing to a more nuanced and personalized approach to lymphoma management.

Despite substantial advances in the genomic and transcriptomic profiling of diffuse large B-cell lymphoma (DLBCL), several critical knowledge gaps remain unresolved(2,13). Conventional cell-of-origin classification, while clinically relevant, fails to capture the full biological heterogeneity of the disease and provides limited insight into the tumour microenvironment interactions that shape therapeutic response(6,12). Although recent multi-omic efforts, including the proteogenotype (PG) framework, have improved the resolution of DLBCL stratification, the functional

implications of these subgroups particularly in terms of immune architecture and transcriptional regulation are still incompletely understood. One major limitation in current literature is the insufficient integration of immune-related pathways with tumour-intrinsic transcriptional programs(15). Several studies have highlighted the prognostic relevance of T-cell dysfunction, antigen-presentation impairment, and cytokine signalling disruption, yet a systematic characterization of how these immune programs interplay with proliferative and oncogenic networks across PG subgroups is still lacking. In particular, the mechanisms underlying the profound immune depletion observed in high-risk subgroups, such as PG4, have not been fully clarified, nor has the functional basis for the robust immune engagement associated with clinically favourable subgroups like PG6. Another unresolved gap concerns the degree to which transcriptomic signatures can be bridged to histological and spatially contextualized immune features. While bulk RNA-seq provides valuable information on pathway activation, it does not capture spatial relationships, cellular topology, or the distribution of effector and suppressor cells within the tumour microenvironment. Although technologies such as multiplex immunofluorescence (mIF) offer high-resolution spatial information, few studies have combined these approaches to deliver an integrated view of transcriptional and microenvironmental states across DLBCL proteogenotypes(14,15,17).

Finally, the biological drivers that underpin the extreme prognostic dichotomy between PG4 and PG6 remain insufficiently defined. Although PG4 has been associated with MYC activation and immune exclusion, and PG6 with immune enrichment and strong antigen-presentation programs, a comprehensive, multi-layered analysis directly comparing these two subgroups has not yet been systematically conducted. Understanding the molecular rationale behind such divergent biological states is essential both for refining risk stratification and for identifying potential therapeutic vulnerabilities.

These gaps underscore the need for an integrated multi-omic analysis that combines transcriptomic profiling, immune-cell deconvolution, MYC-target evaluation, and spatial interrogation of the tumour microenvironment. By addressing these unresolved questions, the present thesis aims to provide a more biologically informed framework for understanding DLBCL heterogeneity and its clinical implications(2,13-15).

Declaration on the use of AI-based tools

This doctoral thesis was written independently. AI-based tools, including DeepL and large language models, were used exclusively as support for writing, language editing, and grammatical correctness. All scientific content, analyses, interpretations, and conclusions are the sole responsibility of the author.

3. Aim of the study

The overarching aim of this doctoral project is to investigate the molecular and immunological determinants of prognostic heterogeneity in diffuse large B-cell lymphoma (DLBCL) using an integrated multi-omic approach. Building on the proteogenotype (PG) framework, the study is structured around the following specific objectives:

1. To characterize global transcriptomic differences between PG4 and all remaining proteogenotypes (PG other) in order to delineate the major molecular programs that distinguish high-risk tumours from the broader DLBCL landscape.
2. To conduct a focused comparative analysis between PG4 and PG6, the biologically most divergent subgroups, with the aim of identifying the transcriptional, immune, and oncogenic pathways that define their opposing phenotypic and clinical behaviours.
3. To quantify immune-cell composition across PG4, PG other, and PG6 through computational deconvolution of bulk RNA-seq data and to determine how immune states correlate with proteogenotypic classification.
4. To evaluate MYC target gene activation in PG4 compared with both PG other and PG6, and to define the extent to which MYC-driven transcriptional programs contribute to PG4 aggressiveness and immune exclusion.
5. To integrate transcriptomic signatures with spatial and cellular features derived from multiplex immunofluorescence (mIF), providing a multi-level characterization of tumour microenvironment interactions across proteogenotypes.

Through these aims, the study seeks to provide a mechanistic understanding of the molecular-immune axis that shapes DLBCL heterogeneity and to contribute to improved precision stratification of high-risk subgroups.

4. Materials and Methods

4.1 Patient cohorts and inclusion criteria:

This retrospective study included patients diagnosed with diffuse large B-cell lymphoma (DLBCL) between 2012 and 2024. Patients were identified from pathology archives and clinical databases of multiple contributing hospitals, including Johanniter Bonn, Universitätsklinikum Köln (UKK), and Malteser St. Hildegardis Hospital (Köln). Inclusion criteria required the availability of a diagnostic formalin-fixed paraffin-embedded (FFPE) tissue block, sufficient clinical annotation, and follow-up data for a minimum of 12 months, or until disease progression, relapse, or death. Patients with both primary and relapse samples were prioritized. Exclusion criteria included lack of clinical metadata or insufficient tissue material for further analysis. A cohort of matched-normal FFPE tissue samples was obtained for a total of 106 patients and an extra cohort of unmatched FFPE normal lymph nodes were collected for a total of 30 patients. Ethical approval was obtained from the institutional ethics committee. The study adhered to all applicable regulations, including participants' rights regarding sample destruction and data protection according to GDPR requirements.

4.2 Data collection and sources

Clinical and pathological data were manually curated from hospital information systems, pathology reports, and treatment protocols. The dataset included patient identifiers, biopsy block numbers for primary and relapse samples, treatment regimens (up to five lines), treatment intent (e.g., curative or palliative), response to therapy, and clinical outcomes including relapse, remission, and death. Stem cell transplantation (SCT) data and dates of each therapy initiation and completion were also recorded. All data were anonymized prior to analysis in compliance with data protection guidelines. Differences in sample numbers across RNA-seq, WES, and multiplex immunofluorescence analyses reflect data availability and modality-specific quality control and filtering procedures.

4.3 Sample processing and molecular profiling

To comprehensively characterize the biological features of aggressive B-cell lymphomas, we performed multi-modal molecular profiling on tumour tissue samples. Formalin-fixed paraffin-embedded (FFPE) blocks from diagnostic and, where available, relapse biopsies were subjected to whole-exome sequencing (WES), RNA sequencing (RNA-seq), and multiplex immunofluorescence (mIF) staining. These complementary approaches were used to assess gene expression patterns, somatic mutation landscapes, and the spatial organization of immune cell populations within the tumour microenvironment. Standardized protocols were applied for nucleic acid extraction, library preparation, sequencing, and downstream bioinformatic analysis to ensure reproducibility and data quality.

4.4 DLBCL Proteogenotype classification and RNA-seq normalization

DLBCL samples were classified according to the framework described by Enssle et al., *Pathogenesis of Diffuse Large B-cell Lymphoma Proteogenotypes* (Cell, 2024). In particular, the classification followed the definitions of the DLBCL proteogenotypes (PGs) established through integrated proteogenomic profiling. The reference model ("DLBCL PG classifier") developed in that study was used to assign samples to the corresponding proteogenotypic groups, as described by the authors and available at www.dlbcl-pg-classifier.uni-frankfurt.de. In our study, we focused exclusively on the comparison between PG4 and PG6 samples, which in our cohort included 29 and 9 cases, respectively. For transcriptomic analyses, we used the raw RNA-seq count data generated from our DLBCL samples. Normalization was performed using the Trimmed Mean of M-values (TMM) method implemented in the edgeR package (Bioconductor). This normalization accounts for library size and compositional differences between samples while preserving relative expression levels across genes. A threshold or vote-based filtering of 50/100 votes was applied after normalization, ensuring that all genes were retained for downstream statistical and exploratory analyses.

4.5 Table2. Data integration and sample annotation

Patient ID	Primary1 Relapse2	Hans Classifier	MYC Translocation	PG Classifier
DLBCL1	1	ABC	0	PG4
DLBCL2	1	ABC	0	PG4
DLBCL3	2	GCB	0	PG3
DLBCL4	1	GCB	0	PG4
DLBCL5	2	ABC	0	NA
DLBCL6	1	ABC	0	PG6
DLBCL7	1	ABC	0	PG6
DLBCL8	1	ABC	0	PG4
DLBCL9	1	ABC	0	PG5
DLBCL10	2	ABC	1	PG4
DLBCL11	2	ABC	0	NA
DLBCL12	1	ABC	0	PG6
DLBCL13	2	GCB	0	PG3
DLBCL14	2	ABC	0	PG4
DLBCL15	1	ABC	0	PG4
DLBCL16	1	ABC	0	PG6
DLBCL17	1	GCB	NA	PG3
DLBCL18	1	ABC	0	PG3
DLBCL19	1	ABC	0	PG5
DLBCL20	1	ABC	NA	PG3
DLBCL21	1	GCB	0	PG2
DLBCL22	2	ABC	0	PG6
DLBCL23	1	ABC	0	PG3
DLBCL24	2	GCB	0	PG3
DLBCL25	1	ABC	1	NA
DLBCL26	1	ABC	0	PG4
DLBCL27	1	GCB	0	NA

DLBCL28	1	ABC	0	PG3
DLBCL29	1	GCB	0	PG6
DLBCL30	2	ABC	0	PG4
DLBCL31	1	ABC	0	PG4
DLBCL32	1	GCB	0	PG4
DLBCL33	1	ABC	0	PG4
DLBCL34	1	ABC	0	PG6
DLBCL35	1	ABC	0	PG4
DLBCL36	1	GCB	0	PG5
DLBCL37	2	ABC	0	PG5
DLBCL38	1	GCB	0	PG3
DLBCL39	1	ABC	0	PG3
DLBCL40	1	ABC	0	PG6
DLBCL41	1	ABC	0	PG4
DLBCL42	1	ABC	0	PG2
DLBCL43	1	GCB	0	PG7
DLBCL44	1	GCB	0	NA
DLBCL45	1	ABC	0	PG3
DLBCL46	2	ABC	0	PG4
DLBCL47	1	ABC	0	NA
DLBCL48	1	ABC	0	PG7
DLBCL49	1	ABC	0	PG6
DLBCL51	1	GCB	0	PG5
DLBCL52	1	GCB	0	NA
DLBCL54	2	ABC	1	PG4
DLBCL55	1	ABC	0	PG4
DLBCL57	1	ABC	0	PG4
DLBCL58	1	ABC	0	PG5
DLBCL59	2	ABC	NA	NA
DLBCL60	1	GCB	NA	NA
DLBCL61	1	ABC	0	PG5
DLBCL62	2	GCB	0	PG5
DLBCL63	2	ABC	0	PG4

DLBCL64	1	ABC	0	NA
DLBCL66	1	ABC	0	PG5
DLBCL67	1	GCB	0	PG3
DLBCL68	1	ABC	0	PG4
DLBCL69	1	GCB	0	PG7
DLBCL70	1	GCB	0	PG7
DLBCL71	1	ABC	0	PG4
DLBCL72	2	GCB	0	PG3
DLBCL73	1	ABC	0	PG5
DLBCL74	1	GCB	0	PG5
DLBCL75	1	GCB	0	PG3
DLBCL77	1	GCB	0	PG4
DLBCL78	2	GCB	0	PG4
DLBCL79	1	GCB	0	PG5
DLBCL80	1	ABC	0	PG5
DLBCL81	2	GCB	0	NA
DLBCL82	1	ABC	0	PG5
DLBCL83	1	GCB	NA	PG4
DLBCL84	1	ABC	NA	PG4
DLBCL85	1	GCB	0	NA
DLBCL86	1	GCB	0	PG4
DLBCL87	1	GCB	0	PG4
DLBCL88	1	GCB	0	PG4
DLBCL89	1	ABC	0	NA
DLBCL90	1	ABC	NA	NA
DLBCL91	1	GCB	0	NA
DLBCL92	2	ABC	0	NA
DLBCL93	1	GCB	NA	NA
DLBCL94	1	GCB	0	NA
DLBCL95	1	GCB	NA	NA
DLBCL96	1	ABC	0	NA
DLBCL97	1	ABC	NA	PG5
DLBCL98	1	GCB	NA	NA

DLBCL99	1	GCB	0	PG4
DLBCL100	1	ABC	0	PG5
DLBCL101	1	GCB	0	PG4
DLBCL102	2	ABC	0	NA
DLBCL103	1	ABC	NA	NA
DLBCL104	1	ABC	0	NA
DLBCL105	1	ABC	NA	NA
DLBCL106	1	ABC	0	PG4

4.6 Whole-Exome Sequencing (WES)

Genomic DNA was isolated from FFPE tumour tissues using extraction kits optimized for degraded samples. Genomic DNA was isolated from FFPE tissue samples using the Maxwell® 16 FFPE Plus LEV DNA Purification Kit (Promega, Ref. AS1135). The automated extraction was performed on the Maxwell® 16 Instrument, ensuring standardized lysis, purification, and elution steps. After proteinase K digestion, DNA was purified via paramagnetic particle-based technology, which offers efficient removal of cross-linked proteins and inhibitors associated with formalin fixation. The purified DNA was eluted in nuclease-free water and quantified using Qubit Fluorometer. A minimum of 200 ng of genomic input DNA was preferred or 100 ng when sample material was limited.

Sequencing of the extracted DNA was performed by the Cologne Centre for Genomics (CCG, University of Cologne, Germany). DNA integrity and yield were monitored prior to downstream applications.

Libraries were sequenced on the Illumina NovaSeq 6000 platform with 2x100 bp paired-end reads. In total, 144 samples were processed, comprising 96 diffuse large B-cell lymphoma (DLBCL) tumours and 48 lymph node-derived normal tissues. Of the DLBCL samples, 48 had a matched germline material, while the other half lacked matched reference samples. The output of reads by the sequencing platform ranged from 78M to 210M.

Samples with insufficient sequencing depth were excluded from downstream analyses. Target regions were defined by intersecting the covered regions of the Agilent SureSelect XT HS2 Human All Exon V8 kit with a licensed database comprised of 55,084 clinically relevant amino acids with known somatic alterations

(Genomenon's Cancer Knowledgebase - CKB). This intersection performed with BCFtools v1.17 (36) resulted in a custom BED file representing the genomic intervals used for quality assessment. Using an in-house script together with GATK's DepthOfCoverage v4.4.0.0 (37), the proportion of these target regions covered at $\geq 70\times$ was calculated for each sample. Tumour samples in which $\geq 50\%$ of target amino acids failed to reach this depth threshold were removed from further analysis. Additionally, DLBCL tumours that did not receive more than 50 votes in the TMMLog2 analysis were excluded. Following these filtering steps, 45 tumour samples and 27 normal samples were retained, with normal samples included only when their corresponding tumour satisfied all quality thresholds.

4.6.1 Evaluation and visualization of somatically acquired alterations

WES data were processed on a high-performance computing (HPC) cluster using the community-curated nf-core/sarek workflow v3.2.2 (38,39). The pipeline was used for alignment of reads to the GATK GRCh38 reference genome, UMI-aware processing, and short-variant calling (single nucleotide variants, insertions, and deletions - SNVs and INDELs) with Mutect2 v4.4.0.0 (37). All default tools and parameters were executed with default settings, except for the removal of reads with a mapping quality below 30. A modified release from an in-house pipeline, TheSarekSnake (40), was integrated with nf-core/sarek to support pre- and post-processing steps and to enable downstream analyses of complex genomic features, including microsatellite instability (MSI) with MSIsensor-pro v1.3.0 (41) somatic copy number alterations (SCNAs), analysed with pureCN v2.12.0 (42,43), as well as calculation of tumour mutation burden (TMB).

Three distinctly formatted panels of normals (PON) were created using the lymph node-derived, FFPE-preserved normal tissues. For usage as a reference file with Mutect2, genomic coordinates of variants identified in at least 70% of the healthy samples were identified for removal of technical and fixation artifacts, additionally ensuring they were not listed either as known cancer hotspots (44) or as being established in the population based on gnomAD's release 2.1.1. This first PON included 33,350 variants. Secondly, following the developer's guidelines for MSIsensor-pro, a baseline file was constructed for determination of microsatellite instability status in tumour-only mode. The analysis targeted microsatellite regions covered by the exome enrichment kit, which included 33,683 microsatellite loci identified by the scan module from the tool. Lastly, the normal database for

determining SCNAs in unpaired mode was built using the recommendations from pureCN for compatibility with Mutect2 and with the reference GRCh38, which additionally includes the flags to genotype PON and germline sites, identifying known variants from dbSNP under "ID" and with the flag "DB" under the "INFO" field in the input files. The construction of the third PON required the use of the first panel built for the removal of artifacts.

Short-variant calls generated by Mutect2 were normalized and filtered with BCFtools v1.17. For matched tumour-normal pairs, variants were retained only if they passed the tools internal filters ("PASS" calls), had tumour allelic depth (AD) > 4, tumour depth (DP) ≥ 20 , normal DP ≥ 10 , a tumour variant allele frequency (VAF) for SNVs $\geq 5\%$ and $\geq 10\%$ for INDELs. For tumour-only samples, analogous filters were applied, except for additionally removing variants with VAF $\geq 40\%$ to avoid likely germline events. However, to mitigate potential false negatives by removing high allelic fraction calls, known cancer-hotspot positions were captured separately and flagged as low-confidence calls. The filtered and flagged calls were then merged and renormalized to generate the final VCFs.

Variant annotation of small mutations was conducted using Ensembl's VEP (version 110.0) against multiple databases, including ClinVar (January 2023 release), gnomAD v3.1.2, dbSNP (build 54), and COSMIC v97. Pathogenicity predictions are calculated with SIFT v6.2.1 and PolyPhen-2 (version 2.2.3). Separately, OncoKB (data version release 4.14 or newer) was accessed using OncoKB Annotator (release 3.4.0), as well as the release from 19.02.2024 from CKB. To exclude germline variants, variants were required not to be listed in gnomAD or at frequencies ≤ 0.01 . Additionally, they could not be classified as (likely) benign in ClinVar. Variants with a population frequency ≥ 0.01 were reported if they were classified in ClinVar with terms including the word "pathogenic" or if they were variants of unknown significance (VUS).

To quantify TMB, non-synonymous variants within coding regions covered by Agilent's capture kit were evaluated. To ensure accuracy and avoid overestimation, only high-confidence variants with tumour read depth ≥ 70 , variant allele frequency ≥ 0.05 , and population allele frequency ≤ 0.0001 in gnomAD were included.

For complex genomic alterations, the MSI status was determined in tumour-only mode using the baseline file for all samples. An MSI score equal to or greater than 5 was applied as the recommended cut-off for classification as MSI-H. Somatic copy

number alteration was assessed with pureCN's recommended guidelines in unpaired mode and with later releases of Mutect2. This involved lowering the minimum base quality to 20, employing the PSCBS segmentation algorithm, and retrieving reference files for known simple repeats from the UCSC repository.

For downstream visualization, short variants and copy-number events (amplifications and deletions) detected in the tumour samples were restricted to alterations with predicted functional consequences. Specifically, only small mutations annotated as (likely/predicted) gain-of-function, (likely/predicted) loss-of-function, resistance-associated, or otherwise classified as (likely) oncogenic by OncoKB or CKB were retained. For the SCNAs, focal amplifications with a copy number value (CN) ≥ 6 , non-focal amplifications with CN > 7 , and deletions with CN ≤ 1 were kept, following the developer's recommendations. These curated alterations were then aggregated across samples to construct multivariate heatmaps. Sample-level annotations, i.e. clinical subgrouping in PG Clusters, were added for interpretability; columns (samples) were sorted based on the clinical subgroups. Heatmaps were generated using the ComplexHeatmap R package v2.25 (45)

For each gene and alteration type with at least four alterations across the cohort, a 2x4 contingency table of alteration status (altered vs. not altered) across the PG clusters was constructed. This results in a set of 29 genes with small mutations/amplifications/deletions. Differences in alteration frequency across groups were assessed using Fisher's exact test (Fisher-Freeman-Halton extension for multi-group tables as implemented in R's `fisher.test`). Resulting p-values were corrected for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) procedure (FDR threshold > 0.1).

4.6.2 Table 3. Reagents and kits for DNA extraction and WES

Category	Reagent / Kit	Manufacturer	Notes
DNA extraction	Maxwell® 16 FFPE	Promega	Automated FFPE
	Plus LEV DNA Purification Kit		DNA extraction

Instrument	Maxwell® 16 Instrument	Promega	Automated purification
Enzyme	Proteinase K	Included	For FFPE digestion
DNA quantification	Qubit Fluorometer	Thermo Fisher	DNA yield assessment
WES capture kit	SureSelect XT HS2 Human All Exon V8	Agilent	Exome capture
Sequencing	NovaSeq 6000	Illumina	2×100 bp reads

4.6.3 Bulk-RNA sequencing (RNA-seq)

RNA was extracted from formalin-fixed paraffin-embedded (FFPE) tumour samples obtained at diagnosis and, when available, at relapse. Total RNA was extracted from formalin-fixed paraffin-embedded (FFPE) tissue sections using the High Pure FFPE RNA Kit (Roche Diagnostics, Ref. 06650775001), following the manufacturer's instructions. Briefly, tissue sections were deparaffinized and digested with proteinase K to facilitate lysis and release of nucleic acids. RNA purification was carried out through glass fiber column-based filtration, with multiple wash steps to remove contaminants and inhibitors commonly associated with FFPE-derived material. The eluted RNA was stored at -80°C until further use. RNA quality and concentration were assessed using Agilent TapeStation RNA ScreenTape assay and RNA integrity was evaluated when possible, taking into account the known degradation typical of FFPE samples.

The extracted RNA was then submitted to MACROGENE (Seoul, South Korea) for library preparation and RNA sequencing. Library preparation was performed following Illumina's protocol for low-input RNA derived from FFPE tissue. Sequencing was carried out on the Illumina NovaSeq 6000 S4 platform using 100 base pair paired-end reads, with a targeted output of approximately 1 gigabase (Gb) per sample. Raw reads underwent quality control, adapter trimming, and alignment to the reference genome using standardized bioinformatics workflows. Transcript abundance was quantified and used for downstream expression and clustering analyses.

4.6.4 RNA-seq sample processing and quantification

Total RNA extracted from lymphoma tumour samples was subjected to paired-end RNA sequencing and a total of 103 DLBCL samples and 31 normal lymph nodes were processed. As a result, raw RNA sequencing data in FASTQ format were obtained as 100bp reads from MACROGEN sequencing provider and amounted to 25M to 44M read pairs per sample. RNA-seq data were then processed on a high-performance computing (HPC) cluster using a community-curated Nextflow workflow for RNA-seq analysis, nf-core/rnaseq (version 3.12.0) (46). Quality control and adapter trimming were performed using FastQC (47) and Trim Galore (48), respectively. Samples with a read duplication rate exceeding 95%, as determined by Picard tools (49), were excluded from downstream analysis. Reads of the remaining samples were aligned to the UCSC hg38 human reference genome using STAR aligner (50), and gene-level quantification was carried out using Salmon (51), based on GENCODE GTF gene annotation file (hg38). All samples were processed in strand-specific mode. Gene count data were used for downstream analyses including differential gene expression and gene set enrichment analysis based on measured gene transcript abundance.

4.6.5 Table 4. Reagents and kits for RNA extraction and RNA-seq

Category	Reagent / Kit	Manufacturer	Notes
RNA extraction	High Pure FFPE RNA Kit	Roche Diagnostics	FFPE RNA purification
Deparaffinization	Xylene	Chemsolute	Standard FFPE dewaxing
Rehydration	Ethanol series	Merck	Standard
QC assay	RNA ScreenTape	Agilent	RNA integrity

Library prep	Illumina RNA-seq Library Prep (FFPE)	Illumina	Performed at Macrogen
Sequencing	NovaSeq 6000 S4	Illumina	100 bp PE reads

4.6.6 RNA-Seq differential gene expression and signature gene analysis

For biological interpretation, gene-level expression data were stratified by sample groups representing disease states, treatment responses, and molecular subtypes. Differential expression analysis between conditions (e.g., primary vs. relapse, or across clinical predictive groups such as PG4, PG6) was performed using DESeq2 (52) as part of the iDEP 2.0 framework (53). As such, raw count data were modelled using a negative binomial distribution, with dispersion estimates stabilized via empirical Bayes shrinkage. Adjusted p-values were calculated using the Benjamini-Hochberg procedure (54) to control the false discovery rate. Log2 fold changes were computed using maximum likelihood estimation.

Subsequently, curated immune and tumour microenvironment gene signatures, including CD8+ effector T cells, Tregs, Th1/Th2/Th17 subsets, M1/M2 macrophage polarization states, and exhaustion markers, were defined and used to extract subsets of differentially expressed genes for pairwise comparisons across conditions. Log2-fold changes and associated p-values for these gene classes and their component genes were visualized using bar plots, with statistical significance ($p < 0.05$) annotated per gene and per condition.

4.6.7 Immune cell profiling via deconvolution of bulk transcriptomic data

To assess the cellular composition of tumour-infiltrating immune cells across DLBCL subtypes, computational deconvolution was performed using the LM22 leukocyte gene signature matrix (20). This matrix, which captures transcriptional profiles of 22 human hematopoietic cell types, was applied to bulk RNA-seq data from the studied cohort to estimate cell type proportions in each sample. The deconvolution results

were obtained using CIBERSORTx (55), which fits a constrained regression model to predict relative abundances of immune cells based on the input expression matrix.

Sample-level clinical annotations were matched to the CIBERSORTx output using a metadata file. Only samples with intersecting identifiers in both the deconvolution output and metadata were retained to ensure valid comparisons. Subsequently, distributions of immune cell proportions were compared across selected categorical variables (e.g., primary vs. relapse status or PG subtypes). To visualize relative proportions across conditions, grouped boxplots were generated for each immune cell type using ggplot2 (43).

In addition to univariate comparisons, multivariate heatmaps were constructed to simultaneously visualize immune landscape patterns across all samples. Heatmaps were annotated by clinical variables and grouped accordingly. Cell types (rows) were sorted to aid interpretability, while columns (samples) were organized based on clinical subgroups but not clustered. Heatmap generation was performed using ComplexHeatmap (45).

The processed metadata and CIBERSORT output were later used to interpret sample-level immune heterogeneity, enabling integration of cell composition estimates with gene expression and clinical data across stratified disease cohorts.

4.6.8 Table 5. Bioinformatic tools

Category	Tool	Version / Reference	Notes
QC	FastQC	Andrews 2010	Quality control
Trimming	Trim Galore	Krueger 2015	Adapter trimming
Alignment	STAR aligner	Dobin et al. 2013	Alignment to hg38
Quantification	Salmon	Patro et al. 2017	Transcript quantification
Duplicates	Picard Tools	Broad Institute 2019	Duplicate rate

Pipeline	nf-core/rnaseq	v3.12.0	Workflow
DEG	DESeq2	Love et al. 2014	Differential expression
Deconvolution	CIBERSORT	Newman et al. 2019	Immune fractions

4.7 Multiplex Immunofluorescence (mIF) Staining

Multiplex immunofluorescence staining was employed to characterize the immune contexture of the tumour microenvironment. Formalin-fixed paraffin-embedded (FFPE) tissue sections were mounted on glass slides and incubated in a 60°C oven for 30 minutes to melt the paraffin. Slides were then dewaxed by two successive immersions in xylene (5 minutes each), followed by rehydration through a graded ethanol series: 100%, 90%, 70%, 50%, and 30% ethanol, each for 3 minutes. Rehydrated slides were then rinsed in 1× PBS (twice) to eliminate residual ethanol. Antigen retrieval was carried out by placing the slides in 500 mL of 10 mM citrate buffer (pH 6.0) and heating them at low power in a microwave or rice cooker until boiling. Slides were maintained at boiling temperature for 10 minutes, cooled to room temperature, and then washed three times with 1× PBS. Staining was performed using a cyclic immunofluorescence (cyclIF) strategy across six rounds. Each cycle involved incubation with 2-3 directly fluorophore-conjugated primary antibodies targeting specific immune or structural markers. For the first cycle, tissue sections were blocked with Odyssey Blocking Buffer (LI-COR Biosciences) for 30 minutes at room temperature in a humid chamber. In subsequent rounds, this step was omitted. Antibodies were diluted 1:100 in Odyssey Blocking Buffer and applied (250-500 µL/slide) overnight (~12 hours) at 4°C in a humidified chamber. After incubation, slides were washed four times in 1× PBS. Nuclei were counterstained with Hoechst 33342 (2 µg/mL) for 15 minutes at room temperature and washed again four times in 1× PBS. An overview of the antibody panel, targets, and performance evaluation is provided in Table 7. A total of 20 antibodies were tested across 6 rounds. The majority were fluorophore-conjugated rabbit or mouse

monoclonal antibodies, reactive against human antigens, targeting a variety of immune cell populations including T cells (CD3D, CD4, CD8a, FOXP3), B cells (CD20), NK cells (CD56), dendritic cells (CD11b), and macrophages (CD68, CD163, CD204). Structural and checkpoint markers such as Lamin A/C, Lamin B, CD31, PD-L1, LAG3, and TIM-3 were also included. Of the antibodies tested, most were validated as "established" based on expected signal distribution and intensity in control and experimental tissues. Some antibodies produced suboptimal results. Notably, anti-Cyclin D1 and CD56 showed poor staining in multiple controls, and anti-PD-L1 yielded weak signal with high background autofluorescence. Antibodies targeting LAG3 and TIM-3 failed to generate any detectable signal and were classified as "no signal", requiring further reassessment or substitution. Before imaging, slides were mounted with 1× PBS or 10% glycerol in PBS for sessions exceeding 30 minutes. Coverslips (24×60 mm, No. 1) were applied to prevent evaporation. After image acquisition, coverslips were removed by immersing slides in 1× PBS for ≥15 minutes in vertical staining jars, allowing detachment by gravity. After each imaging session, fluorophore inactivation was achieved by incubating the slides for 1 hour in a bleaching solution containing 4.5% hydrogen peroxide and 24 mM NaOH in 1× PBS, under white light at room temperature. Slides were then washed four times in 1× PBS and stored until the next staining cycle. Image acquisition was performed in collaboration with the Max Planck Institute for Biology of Ageing in Cologne, within the group led by Christian Kukat, Head of the FACS and Imaging Core Facility. The imaging system used was the Axioscan 7 slide scanner (Zeiss), which enabled high-throughput and high-resolution acquisition. A total of five imaging cycles were conducted on thirteen Tissue Microarray (TMA) slides. The scans were acquired at 20× magnification, using four fluorescence channels: DAPI ($\lambda_{\text{ex}} \approx 358 \text{ nm}$ / $\lambda_{\text{em}} \approx 461 \text{ nm}$), Alexa Fluor 488 (green; $\lambda_{\text{ex}} \approx 488 \text{ nm}$ / $\lambda_{\text{em}} \approx 519 \text{ nm}$), mCherry ($\lambda_{\text{ex}} \approx 587 \text{ nm}$ / $\lambda_{\text{em}} \approx 610 \text{ nm}$), and Alexa Fluor 647 (Cy5; $\lambda_{\text{ex}} \approx 649 \text{ nm}$ / $\lambda_{\text{em}} \approx 670 \text{ nm}$). This multiplexed approach enabled the detection of multiple markers across iterative staining and scanning rounds.

4.7.1 Table 6. Reagents used for mIF

Category	Reagent	Manufacturer	Notes
Dewaxing	Xylene	Chemsolute	Two immersions

Rehydration	Ethanol series	Merck	100→30%
Buffer	PBS (1×)	Gibco	Wash steps
Antigen retrieval	Citrate buffer pH 6.0	Lab-prepared, Merck, Carl Roth GmbH	Heated
Blocking buffer	Odyssey Blocking Buffer	LI-COR	First cycle
Nuclear stain	Hoechst 33342	Thermo Fisher	2 µg/mL

4.7.2 Table 7. Antibodies

Panel	Antibody	Species	Target / Cells	Fluorophore	Manufacturer	Catalogue number
1	CD8a	human	CD8+ T cells	Alexa-488	Thermo Fischer	53-0008-82
1	Lamin B	human	nuclear membrane	Alexa-647	abcam	ab194108
1	CD31	human	endothelial cells / vessels	Alexa-555	abcam	ab279331
2	CD11b	human	dendritic cells	Alexa-488	Thermo Fischer	53-0196-80
2	CD3D	human	T cells	Alexa-555	abcam	ab208514

2	CD20	human, monkey	B cells	eFluor 660	Thermo Fischer	50-0202- 82
3	CD16 3	human	macrophages	Alexa- 488	abcam	ab218293
3	CD68	human	macrophages	Alexa- 555	abcam	ab279323
3	CD20 4	human	macrophages	Alexa- 647	abcam	ab300100
4	CD4	human	CD4+ T cells	Alexa- 488	R&D Systems	FAB8165 G
4	FOXP 3	human, monkey	T reg cells	eFluor 570	Thermo Fischer	41-4777- 82
4	CD13 8	human	plasma cells, NK cells	Alexa- 647	BioLegen d	356523
5	Lamin A/C	human	nuclear membrane	Alexa- 488	abcam	ab185014
5	PD-L1	human	PD-L1	Alexa- 555	abcam	ab274896
5	CD56	human	NK cells	eFluor 660	Thermo Fischer	50-0565- 80

4.7.3 Multiplexed-imaging pre-processing and segmentation

Raw data consisted of 106 arrays of 15-channel, 16-bit multiplexed images of approximately 16000 x 16000 pixels (different sizes are due to the automatic stitching process of the individual tiles composing each image). Multiplexed Image

data was composed of four individual channels at different wavelengths (404 488, 555, 647 nm). The DAPI signal (405nm) across all cycles was used for registration and then discarded, with the exception of the signal from the first cycle. The autofluorescence and background signal was acquired from tissue free areas (e.g. image corners) by computing the average background intensity, which was then subtracted, improving the signal-to-noise ratio. The corrected images were then visually inspected for artifacts. Nuclear segmentation was performed by using Cellpose 2.0 (56) on the DAPI channel and the computed probability maps were used to define nuclear outlines. Cytoplasm was segmented by expansion of the nuclear object by 25 px or until the border of the next cell started. Cell and nuclear boundaries were used for computing mean and max intensity values from all the channels. Visualization of spatial marker expression was done by superimposing the normalized intensity values in the per cell nuclear centroid coordinates in the geometric space. Data was cleaned and processed with custom python code. The first step consisted in discarding the cells that, on a sample-by-sample basis, surpass the 95th percentile of mean target intensity. The expression of a specific target was then used to binarize a specific cell population via Otsu's algorithm, with the positive cells for that target being further analysed.

4.8 Table 8. Abbreviations

Abbreviation	Full Term
ABC	Activated B-cell-like
AID	Activation-induced cytidine deaminase
ASCT	Autologous stem cell transplantation
BCL2	B-cell lymphoma 2
BCL6	B-cell lymphoma 6
BCR	B-cell receptor
Bp	Base pair
CAR-T	Chimeric antigen receptor T cell

CCL	CC motif ligand
CD	Cluster of differentiation
CIBERSORT	Cell type Identification By Estimating Relative Subsets Of RNA Transcripts
CNV	Copy number variation
COO	Cell of origin
ctDNA	Circulating tumour DNA
CR	Complete remission
CTCs	Circulating tumour cells
cyclIF / CyCIF	Cyclic immunofluorescence
CXCL	C-X-C chemokine ligand
DAPI	4',6-diamidino-2-phenylindole
DC	Dendritic cell
DEG	Differentially expressed gene
DLBCL	Diffuse large B-cell lymphoma
DNA	Deoxyribonucleic acid
Dpi	Dots per inch
FFPE	Formalin-fixed, paraffin-embedded
FDR	False discovery rate
FISH	Fluorescence in situ hybridization
FOXP3	Forkhead box P3
FPKM	Fragments per kilobase of transcript per million reads
GCB	Germinal centre B-cell-like
Gb	Giga base
GEO	Gene Expression Omnibus

GO	Gene ontology
GSEA	Gene set enrichment analysis
GTF	Gene transfer format
GZMB	Granzyme B
H&E	Haematoxylin and eosin
HLA	Human leukocyte antigen
HSC	Hematopoietic stem cell
HSP	Heat-shock protein
HPC	High-performance computing
HR	Hazard ratio
IFIT	Interferon - induced proteins with tetratricopeptide repeats
IFN	Interferon
IF	Immunofluorescence
IHC	Immunohistochemistry
IPI	International Prognostic Index
KEGG	Kyoto Encyclopedia of Genes and Genomes
LAG3	Lymphocyte-activation gene 3
LCK	Tyrosine protein kinase
LN	Lymph node
mIF	Multiplex immunofluorescence
mIHC/IF	Multiplex immunohistochemistry / immunofluorescence
M0/M1/M2	Macrophage polarization states
MSI	Microsatellite instability
MRD	Minimal residual disease
mRNA	Messenger RNA

MTOR	Mechanistic target of rapamycin
MYC	Myelocytomatosis oncogene
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NK	Natural killer (cell)
Opal	Tyramide Signal Amplification system
PBMS	Peripheral blood mononuclear cell
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PD-1	Programmed cell death protein 1
PD-L1	Programmed death-ligand 1
PG	Proteogenotype
PI3K	Phosphoinositide 3-kinase
PRDM1	PR/Set Domain 1
PRF	Perforin
PRSS	Trypsin-1
QC	Quality control
R-CHOP	Rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
ROI	Region of interest
SCNA	Somatic copy-number alteration
scRNA-seq	Single-cell RNA sequencing
SCT	Stem cell transplantation
SNV	Single nucleotide variant

STAT	Signal transducer and activator of transcription
STING	Stimulator of interferon genes
TAM	Tumour-associated macrophage
TCR	T-cell receptor
TILs	Tumour-infiltrating lymphocytes
TIM3	T-cell immunoglobulin and mucin domain containing -3
TMA	Tissue microarray
TMB	Tumour mutation burden
TME	Tumour microenvironment
TPS	Tumour proportion score
Tregs	Regulatory T cells
TSA	Tyramide signal amplification
VAF	Variant allele frequency
WES	Whole-exome sequencing
WHO	World Health Organization

5. Results

The study cohort comprised 106 patients diagnosed with diffuse large B-cell lymphoma (DLBCL) between 2012 and 2024. All cases were histologically confirmed according to the 2017 WHO classification of hematopoietic and lymphoid neoplasms (3). Formalin-fixed, paraffin-embedded tissue samples were available for each case and constituted the core biological material for multi-omic analyses. The cohort included both *de novo* and relapsed/refractory cases, thus encompassing the clinical and molecular heterogeneity that typifies DLBCL. Detailed clinical information was retrieved for all patients, including demographic variables, disease stage, International Prognostic Index (IPI), treatment regimens, response to therapy, and follow-up data. As illustrated in Figure 1a, this dataset integrates complementary molecular and clinical layers, combining high-throughput sequencing, multiplexed imaging, and curated clinical annotations. FFPE-derived DNA, RNA, and protein material were subjected to whole-exome sequencing (WES), bulk RNA sequencing (RNA-seq), and multiplex immunofluorescence (mIF), providing a multidimensional molecular profile for each tumour. Genomic and transcriptomic analyses enabled the identification of somatic variants, gene expression programs, and pathway activities, while mIF preserved spatial information on immune and stromal components of the tumour microenvironment (TME). The integration of these data modalities with longitudinal clinical information established a comprehensive framework for molecular and clinical correlation analyses. This design allowed the study to address key biological and translational questions, such as the relationship between intrinsic tumour features and the immune contexture, or the molecular determinants of prognosis across proteogenotypic subgroups. Each sample was processed through standardized and quality-controlled workflows, ensuring reproducibility across data layers and comparability with external reference cohorts. The multi-omic integration depicted in Figure 8a reflects the central rationale of this thesis: to combine molecular, spatial, and clinical information to better understand the heterogeneity of DLBCL and its

microenvironment. By leveraging FFPE-based approaches, the study bridges clinical pathology with modern high-throughput technologies, generating a translationally relevant dataset capable of revealing the interplay between tumour-intrinsic alterations and immune ecosystem dynamics.

5.1 Overview of the DLBCL study cohort and multi-omic data integration.

The study included 106 DLBCL patients with available FFPE tissue material and comprehensive clinical data. Multi-layered profiling was performed from FFPE-derived DNA, RNA, and protein, encompassing whole-exome sequencing, RNA sequencing, and multiplex immunofluorescence imaging. These molecular and spatial data were integrated with clinical annotations to enable a multi-omic characterization of DLBCL biology and its association with clinical outcome.

5.2 Distribution of proteogenotypes in the cohort

Following the initial characterization of the cohort, we examined the distribution of the seven DLBCL proteogenotypes (PG1- PG7) across all samples with high-quality RNA-seq data. As shown in Figure 1b, the cohort displayed marked molecular heterogeneity, with PG4 emerging as the predominant subgroup (n = 31). This finding is particularly relevant given the biologically aggressive phenotype previously described for PG4, characterized by MYC- and BCR/PI3K- driven transcriptional programs, enhanced ribosomal biogenesis, and profound immune depletion. PG5 (n = 17) and PG3 (n = 14) were also well represented, providing a sufficient basis for comparative analyses. In contrast, PG6 (n = 9), PG7 (n = 4), and PG2 (n = 2) were less frequent, consistent with their lower prevalence and known variability in FFPE-derived transcriptomic classification. This distribution underscores the intrinsic heterogeneity of the cohort and motivates a focused comparison between the biologically adverse PG4 and the more immunologically active PG5/PG6 subgroups.

5.2.1 Effect of the vote-threshold filtering on PG assignments

To ensure robustness of the proteogenotype classification, we applied a vote-confidence threshold (<50 votes) to exclude samples with low assignment stability. After filtering, PG4 remained the most abundant subgroup (n = 26), confirming its strong molecular definition and high classification confidence (Figure 8c). PG5 retained all 17 cases, further supporting its stability within the dataset. PG3 decreased to 10 samples but remained sufficiently represented for exploratory analyses. PG6 decreased to 7 high-confidence samples, indicating classification consistency for this subgroup within FFPE-based RNA-seq. PG2 and PG7 were entirely removed, reflecting poor discriminative signal or degraded RNA quality. This filtering step refined the dataset to a more reliable subset, ensuring that subsequent analyses, particularly comparisons between PG4 and PG6 were based on stable classifications.

5.2.2 Survival analysis across PG subgroups

To investigate the clinical implications of PG4 in our cohort, we compared overall survival between PG4 and the pooled set of all other PGs using Kaplan-Meier analysis. Although the difference did not reach statistical significance (log-rank $p = 0.14$), the survival curves showed a clear trend toward inferior outcomes for PG4 patients (Figure 8d). The divergence between curves became most evident within the first 40-50 months of follow-up, a period typically associated with early disease progression and therapeutic resistance in high-risk DLBCL. While limited by sample size and heterogeneity of follow-up, this trend is nonetheless consistent with previously published evidence linking PG4 to unfavourable prognosis. These findings further reinforce the biological relevance of PG4 within this cohort and justify its selection as the principal high-risk subgroup for subsequent transcriptomic and spatial immunophenotypic analyses.

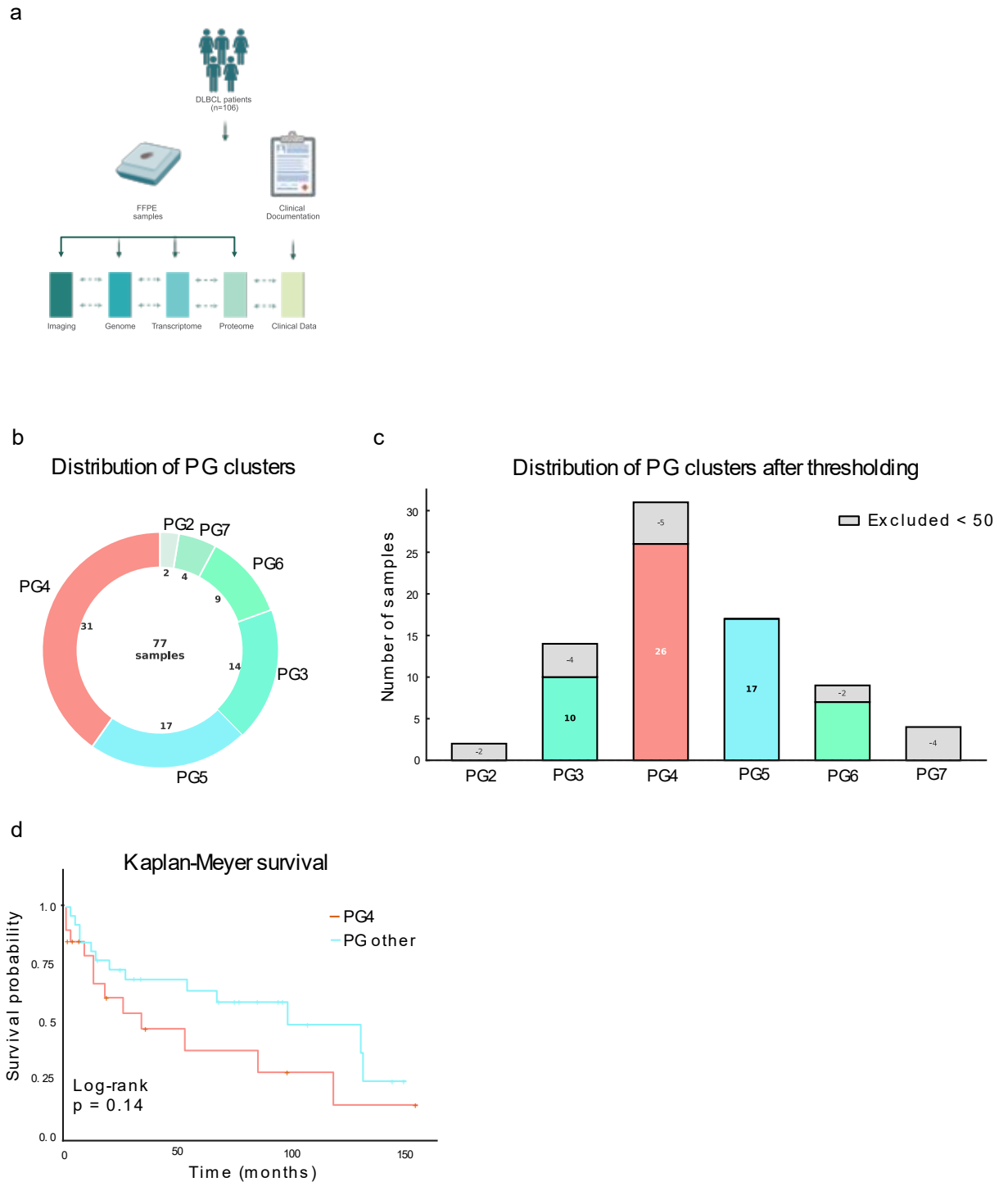


Figure 8. Multi-omic characterization and proteogenotype distribution in the DLBCL cohort.
 (a) Overview of the study cohort and multi-omic workflow.
 (b) Distribution of PG clusters across the 77 samples with high-quality RNA-seq data.
 (c) PG distribution after applying a <50-vote confidence threshold.
 (d) Kaplan-Meier overall survival comparing PG4 versus all other PGs.

5.3 Spatial characterization of the tumour microenvironment using multiplex immunofluorescence

Before engaging with transcriptomic analyses, we sought to obtain a spatial overview of the tumour microenvironment across proteogenotypes, with the aim of visually assessing baseline immune-cell infiltration and tissue architecture. Multiplex immunofluorescence (mIF) enables simultaneous detection of multiple immune and tumour-associated markers while preserving spatial context, offering a high-resolution perspective on cell composition and organization. Because PG4 and PG6 represent the biological extremes of immune depletion and immune activation, respectively, this analysis provides an essential morphological foundation for the transcriptomic contrasts described in the following sections.

5.3.1 Representative mIF images across proteogenotypes

The multiplex immunofluorescence images shown in Figure 9a provide a first qualitative assessment of the microenvironmental landscape across proteogenotypes. These images, acquired from representative TMA cores stained for DAPI, CD4, CD8, CD20, CD68 and CD163, reveal profound differences in both the density and spatial distribution of immune cells. PG4 samples are characterized by extensively tumour-dominated regions displaying minimal immune infiltration, with only a few scattered CD4⁺ and CD8⁺ T cells and virtually absent macrophage and B-cell staining. This spatial emptiness conveys a striking immune-desert phenotype. By contrast, PG6 samples exhibit a markedly different architecture: the tissue is densely populated by immune cells forming recognizable clusters of CD4⁺ helper and CD8⁺ cytotoxic T cells, as well as CD20⁺ B-cell aggregates and CD68⁺/CD163⁺ macrophage infiltrates. This arrangement suggests a highly interactive and active immune microenvironment. Intermediate phenotypes are visible in PG3 and PG5, which show variable levels of immune engagement and stromal architecture. Tonsil tissue included as a positive control highlights the expected richness of physiologically active lymphoid structures. This panel visually anchors the notion that PG4 and PG6 lie at opposite ends of the immunological spectrum.

5.3.2 Deep-learning-based single-cell segmentation and reconstructed cellular maps

Deep-learning-based segmentation enabled high-resolution single-cell reconstruction of each core, allowing detailed visualization of their cellular architecture. PG4 samples exhibited homogeneous tumour regions with sparse and isolated immune cells, highlighting both quantitative immune depletion and fragmentation of immune structure. Conversely, PG6 reconstructions demonstrated a highly diverse and intermixed landscape, where immune and tumour cells occupied shared spatial niches, suggestive of coordinated immune activity. PG3 and PG5 reconstructions showed intermediate degrees of immune -tumour intermixing.

5.3.3 Global cell-type composition across proteogenotypes

Cell-type proportions derived from the segmentation masks confirmed these observations. PG4 displayed the lowest abundance of nearly all immune populations, including helper and cytotoxic T cells, B cells, M1/M2 macrophages, and NK cells. PG6 showed the highest representation of these lineages, supporting the existence of an active and immunologically engaged microenvironment. The remaining PGs exhibited variable profiles, reflecting their biological diversity.

5.3.4 Direct comparison of immune composition between PG4 and PG6

By isolating PG4 and PG6 for direct comparison, Figure 9c sharply highlights the extreme dichotomy between immune-cold and immune-hot microenvironments. PG4 shows a dramatic reduction in all major immune subsets, including near-complete absence of macrophages and profoundly suppressed levels of CD4⁺ and CD8⁺ T cells. B-cell infiltration is also minimal, further emphasizing its immune-desert phenotype. PG6, however, shows strong infiltration by all these immune populations, with a balanced presence of effector T cells, B cells, and macrophage subsets. This comparison reinforces the understanding that PG4's adverse prognosis may be driven not merely by quantitative immune depletion but by the collapse of key effector circuits and antigen-presenting networks. The stark contrast between PG4 and PG6 serves as a biological premise for the subsequent

transcriptomic analyses demonstrating profound differences in immune-related gene expression.

5.3.5 Tumour- centred spatial interactions in PG4 vs PG6

To obtain a mechanistic understanding of how the tumour microenvironment is spatially organized across proteogenotypes, we quantified the interaction likelihood between tumour cells and all immune subsets, using the tumour compartment as the fixed reference population. This metric estimates how frequently a given immune cell type is found in close spatial proximity to tumour cells compared with random expectation, thereby capturing the degree to which each immune subset physically engages with, or is excluded from, tumour-rich regions. Because effective anti-tumour immunity depends critically on spatial crosstalk, particularly T-cell trafficking and macrophage-mediated antigen presentation, this analysis provides a high-resolution view of functional microenvironmental architecture.

Across the cohort, PG6 exhibited the highest tumour -immune interaction likelihoods, revealing a microenvironment in which macrophages and T cells consistently co-localize with malignant cells. Both M0/M1 and M2 macrophage subsets showed strong spatial adjacency to tumour regions, indicating preserved antigen-presenting and tissue-remodelling interfaces. Likewise, CD4⁺ helper T cells, CD8⁺ cytotoxic T cells, double-positive CD4⁺/CD8⁺ T cells, and T cells not otherwise specified (T-cell NOS) all demonstrated increased likelihoods of proximity to tumour cells. These patterns collectively portray PG6 as an immune-active and highly interactive ecosystem, in which effector cells, helper cells, and myeloid regulators are appropriately recruited to tumour sites and maintain a spatial configuration compatible with productive anti-tumour responses. Importantly, the engagement of CD8⁺ T cells in particular suggests preserved cytotoxic potential, while the balanced presence of macrophage subsets reflects a dynamic but functional immunological niche.

In stark contrast, PG4 displayed a profound collapse of tumour- immune spatial interactions, extending across nearly all measured cell types. Both M0/M1 and M2 macrophages showed sharply reduced likelihoods of being found near tumour cells, indicating that tumour regions in PG4 lack the antigen-presenting and phagocytic support traditionally provided by tissue-resident or infiltrating macrophages. This spatial disengagement suggests that macrophage recruitment, retention, or

positioning within tumour-rich areas is impaired, disrupting the myeloid- tumour interface that normally primes and sustains T-cell responses.

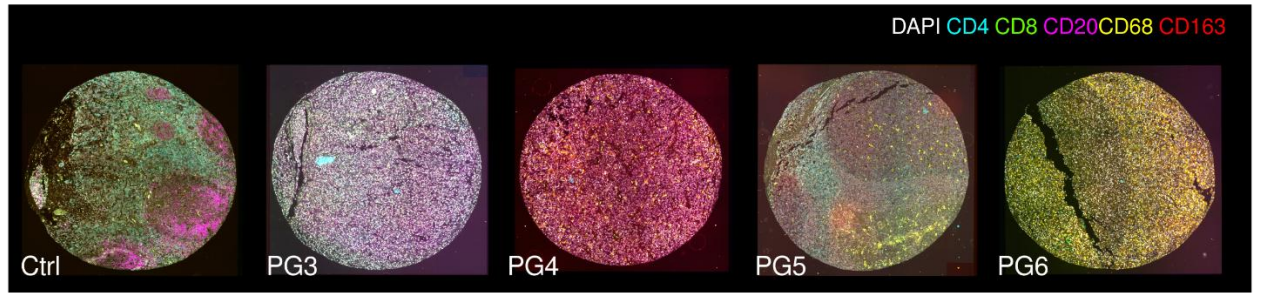
Even more striking was the behaviour of T-cell subsets. PG4 exhibited minimal or virtually absent tumour -T-cell interactions, regardless of T-cell phenotype. CD4⁺ helper T cells, CD8⁺ cytotoxic T cells, and double-positive T cells all demonstrated extremely low proximity to tumour cells, indicating that T cells either fail to infiltrate tumour regions or are actively excluded from them. This architectural segregation is consistent with a breakdown of chemokine-driven recruitment, stromal barrier formation, or functional T-cell paralysis, and aligns with the transcriptomic evidence of suppressed immune signalling pathways in PG4. Notably, even when rare T cells were present within PG4 tissues, they remained spatially distant from malignant cells, reinforcing the notion that the immune system is not merely depleted but structurally unable to engage with tumour targets.

Taken together, these findings reveal a fundamental architectural dichotomy between the immune-hot PG6 and the immune-cold PG4 proteogenotypes:

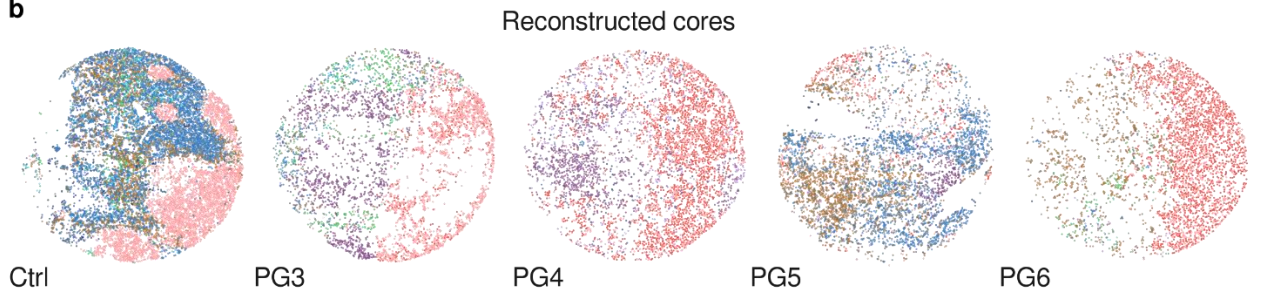
- PG6 tumour regions are deeply embedded within a multi-lineage immune network, with macrophages and T cells consistently interacting with malignant cells, supporting an immune-competent, surveillance-oriented microenvironment.
- PG4, by contrast, exhibits not only quantitative reductions in immune infiltration but a qualitative failure of spatial organization, wherein immune cells are physically segregated from tumour cells and thus unable to mount or sustain anti-tumour activity.

This collapse of tumour-directed spatial interactions in PG4 provides a mechanistic explanation for its profoundly immunosuppressed phenotype and unfavourable prognosis. It shows that immune insufficiency in PG4 is not solely a matter of cell number but reflects a deeper breakdown of the spatial and functional architecture required for effective anti-tumour immunity. The combination of depleted immune populations, disrupted tumour-myeloid interfaces, and absent tumour-T-cell engagement underscores the structural basis of immune failure in PG4 and highlights the spatially organized immune competence of PG6 as a potential driver of its favourable clinical outcome.

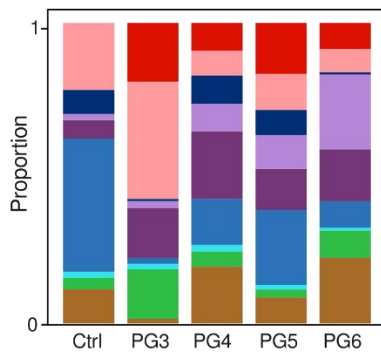
a



b



c



d

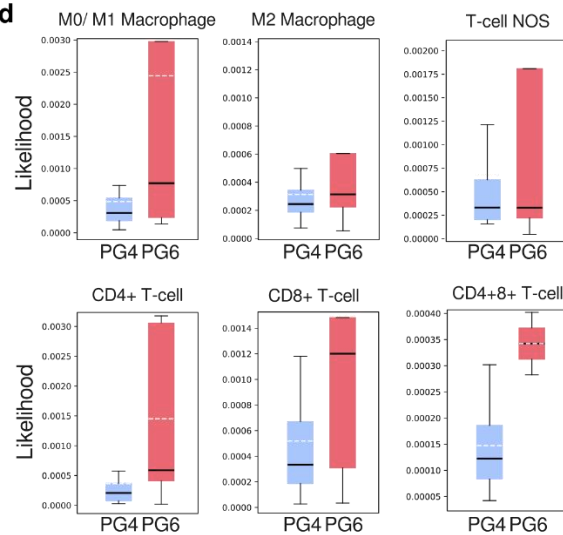


Figure 9. Spatial characterization of the DLBCL tumour microenvironment across proteogenotypes.

Representative multiplex immunofluorescence images showing DAPI, CD4, CD8, CD20, CD68 and CD163 staining across PGs.

(b) Single-cell segmentation maps illustrating the spatial distribution of tumour and immune populations.

(c) Global cell-type proportions across PGs derived from mIF segmentation.

Colour code indicates the following cell populations: tumour cells (red), B cells (light pink), endothelial cells (dark blue), M0/M1 macrophages (navy), M2 macrophages (purple), CD4⁺

T cells (light blue), CD4⁺/CD8⁺ T cells (cyan), CD8⁺ T cells (green), and T-cell NOS (brown).
(d) Tumour- centred spatial interactions in PG4 vs PG6

5.4 Differential gene expression between PG4 and other PGs

To investigate transcriptomic differences distinguishing PG4 from the remaining proteogenotypic subgroups, we performed a differential gene expression (DGE) analysis comparing PG4 versus all other PGs combined ("PG other"). As shown in Figure 9a, the volcano plot reveals a substantial set of significantly deregulated genes, with a clear asymmetry reflecting the distinct molecular programs characterizing PG4. A large number of genes were downregulated in PG4 (left side of the plot), consistent with the previously reported immune-depleted phenotype associated with this subgroup. Upregulated genes in PG4 were comparatively fewer and predominantly associated with proliferative or metabolic pathways that align with the established MYC- and BCR/PI3K-driven transcriptional profile of PG4. Collectively, the DGE landscape underscores a strong contrast between PG4 and the broader DLBCL population, driven largely by loss of immune-related transcriptional activity.

5.4.1 Enrichment of biological processes in PG4 vs others

To functionally interpret the transcriptional differences identified in the DGE analysis, we performed gene ontology (GO) over-representation analysis. As illustrated in Figure 10b, downregulated processes in PG4 mapped predominantly to developmental, differentiation, and adhesion-related pathways, several of which play essential roles in immune-cell maturation, cytoskeletal regulation, and intercellular communication. Notably, pathways linked to cellular development, morphogenesis, and cell-cell adhesion appeared strongly suppressed in PG4, suggesting impaired structural and immunologic organization within the tumour microenvironment. Conversely, upregulated pathways in PG4 were sparse and largely associated with cell-cycle progression, DNA replication, and biosynthetic activity findings in line with the hyperproliferative phenotype known to characterize PG4. The imbalance between suppressed immune/structural programs and

enhanced proliferative pathways is consistent with an aggressive tumour phenotype accompanied by profound microenvironmental dysfunction

5.4.2 Reduced immune-related transcriptional activity in PG4

Given the marked downregulation of immune and adhesion-related pathways, we next examined broad immune-related transcriptional signatures. As shown in Figure 10c, the GO:BP “Immune Response” meta-score was significantly lower in PG4 compared with the remaining PG groups. This finding is consistent with an immune-cold microenvironment, characterized by diminished expression of genes involved in T-cell activation, interferon signalling, antigen presentation, and leukocyte trafficking. This reduction in immune transcriptional competence supports the hypothesis that PG4 tumours may evade immune surveillance not only through tumour-intrinsic features but also through broad suppression of the immune ecosystem at the transcriptomic level.

5.4.3 Immune-related gene expression profiles distinguish PG4 from other PGs

To further interrogate immune transcriptional programs, we generated a heatmap of the top 100 GO:BP immune response genes differentially expressed between PG4 and PG other. The resulting pattern revealed a highly coherent separation of PG4 samples from all other proteogenotypes, with PG4 displaying a consistent downregulation across the majority of immune-associated genes examined. These genes included signatures related to cytotoxicity (e.g., granzymes, perforin), antigen presentation (e.g., HLA class II molecules), chemokine-mediated recruitment, and interferon-response pathways. The heatmap recapitulates the overarching trend observed in previous analyses: PG4 tumours exhibit a coordinated suppression of immune effector programs, supporting their classification as immunologically inert or “immune-cold.” In contrast, PG other samples composed predominantly of PG5 and other immunologically active proteogenotypes show higher expression of immune signalling genes, consistent with a more functionally engaged tumour microenvironment.

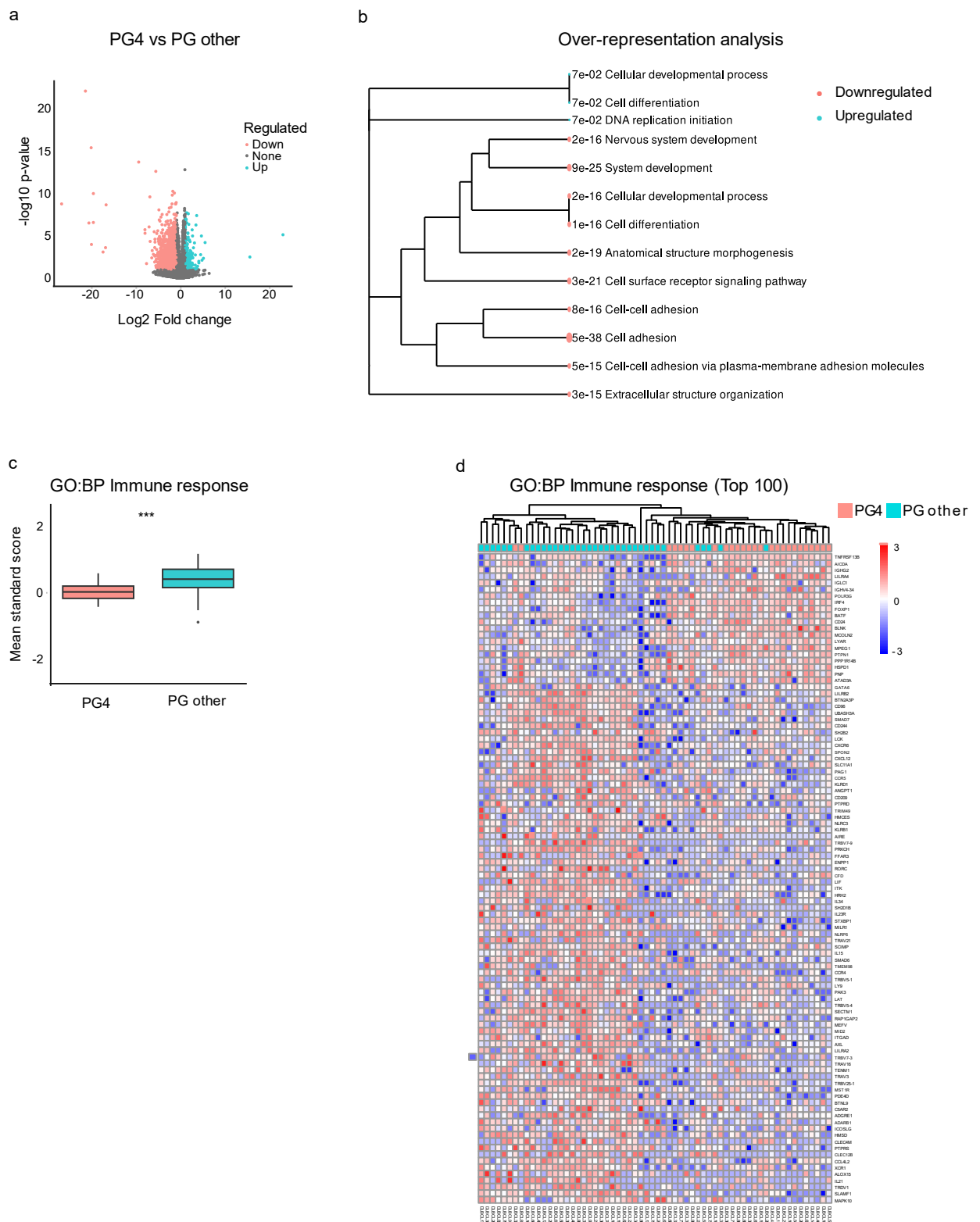


Figure 10. Differential transcriptomic features distinguishing PG4 from all other DLBCL proteogenotypes.

(a) Volcano plot displaying differentially expressed genes between PG4 and PG other, highlighting extensive downregulation of immune-related genes in PG4.

(b) GO over-representation analysis showing suppression of developmental, adhesion, and

immune-related pathways in PG4, and selective enrichment of proliferative programs.

(c) Boxplot of the GO:BP “Immune Response” signature score, significantly reduced in PG4.

(d) Heatmap of the top 100 immune-related GO:BP genes, illustrating the marked suppression of immune transcriptional programs in PG4 compared with PG other

5.5 Focused transcriptomic analysis of PG4 versus PG6

Given the broad transcriptional differences observed between PG4 and the remaining proteogenotypes, we subsequently focused on the biologically most divergent comparison within the cohort: PG4 versus PG6. These two subgroups represent opposite extremes of the DLBCL prognostic spectrum, with PG4 associated with poor outcome, heightened proliferative signalling, and an immune-depleted microenvironment, whereas PG6 exhibits a strongly immune-enriched, transcriptionally active, and clinically favourable phenotype. While the initial PG4 vs others analysis provided a global overview of PG4-specific transcriptional features, this grouping aggregates proteogenotypes with heterogeneous immune characteristics. Direct comparison with PG6 therefore enables a more refined and biologically interpretable characterization of the immune programs specifically suppressed in PG4. For this reason, we conducted a dedicated differential gene expression and pathway-enrichment analysis comparing these two proteotypic extremes.

5.5.1 Differential gene expression reveals a strong immunological imbalance

The transcriptomic comparison between PG4 and PG6 revealed a pronounced asymmetry, with PG4 exhibiting widespread downregulation of immune-related genes and PG6 showing robust upregulation of pathways associated with immune activation and response. As illustrated in Figure 11a, PG6 samples displayed elevated expression of chemokines, interferon-stimulated genes, cytotoxic mediators, and components of antigen-presentation machinery, consistent with an “immune-hot” phenotype. In contrast, PG4 showed coordinated suppression of T-cell and innate immunity related transcripts, consolidating its identity as a profoundly “immune-cold” proteogenotype. The breadth of this transcriptional divergence suggests deep biological separation between these two groups, extending beyond

differences in cellular infiltration to encompass fundamental impairment in immune signalling programs.

5.5.2 Downregulated pathways in PG4 highlight impaired immune competence

Gene ontology enrichment analysis of PG4-downregulated genes reinforced the marked immune deficiency in this subgroup. Figure 11b shows that nearly all significantly enriched pathways corresponded to core immune processes, including “immune response,” “immune system process,” “defense response,” and multiple layers of immune system regulation. Additional terms related to responses to external or biotic stimuli further underscore the lack of tumour-intrinsic capacity to engage with immune-derived cues. This pattern of pathway suppression aligns with the diminished activation of T-cell and antigen-presenting cell networks previously described for PG4 and suggests impaired ability to mount coordinated adaptive or innate immune responses.

5.5.3 Network analysis consolidates functional coherence of suppressed programs

To visualize the internal structure and interconnectedness of these pathways, we constructed a GO-based network representation. The resulting topology reveals a highly cohesive module of downregulated immune-related processes including immune response, defense response, immune system regulation, and interspecies interaction highlighting strong functional convergence. The clustering of these pathways indicates that PG4 is characterized not by isolated gene defects but by a systems-level breakdown of immune coordination. This network structure provides additional support for PG4’s classification as an immune-impaired entity.

5.4.4 Immune-response signature scores sharply separate PG4 and PG6

Consistent with the enrichment results, the GO:BP “Immune Response” signature score was significantly reduced in PG4 compared with PG6. PG6 samples exhibited uniformly high immune-response scores, while PG4 samples clustered tightly

around markedly diminished values. This difference reflects broad suppression of cytokine signalling, antigen presentation, chemotaxis, and cytotoxic effector functions in PG4. Such comprehensive transcriptional silencing reinforces the conclusion that PG4 tumours exist in a microenvironment unable to sustain or recruit effective anti-tumour immunity.

5.5.5 Heatmap of immune genes highlights the PG4-PG6 immune dichotomy

The heatmap of the top 100 immune-related GO:BP genes clearly recapitulated the immune dichotomy between PG4 and PG6. PG6 displayed strong and coordinated expression of immune effector molecules interferon-inducible transcripts, chemokines, HLA class II components, and cytotoxic mediators. PG4 samples, by contrast, showed broad and systematic repression across nearly all immune gene sets examined. This binary separation provides a transcriptomic foundation for the divergent clinical behaviour of these proteogenotypes and reinforces the notion that PG4 tumours lack the transcriptional machinery required for effective immune engagement.

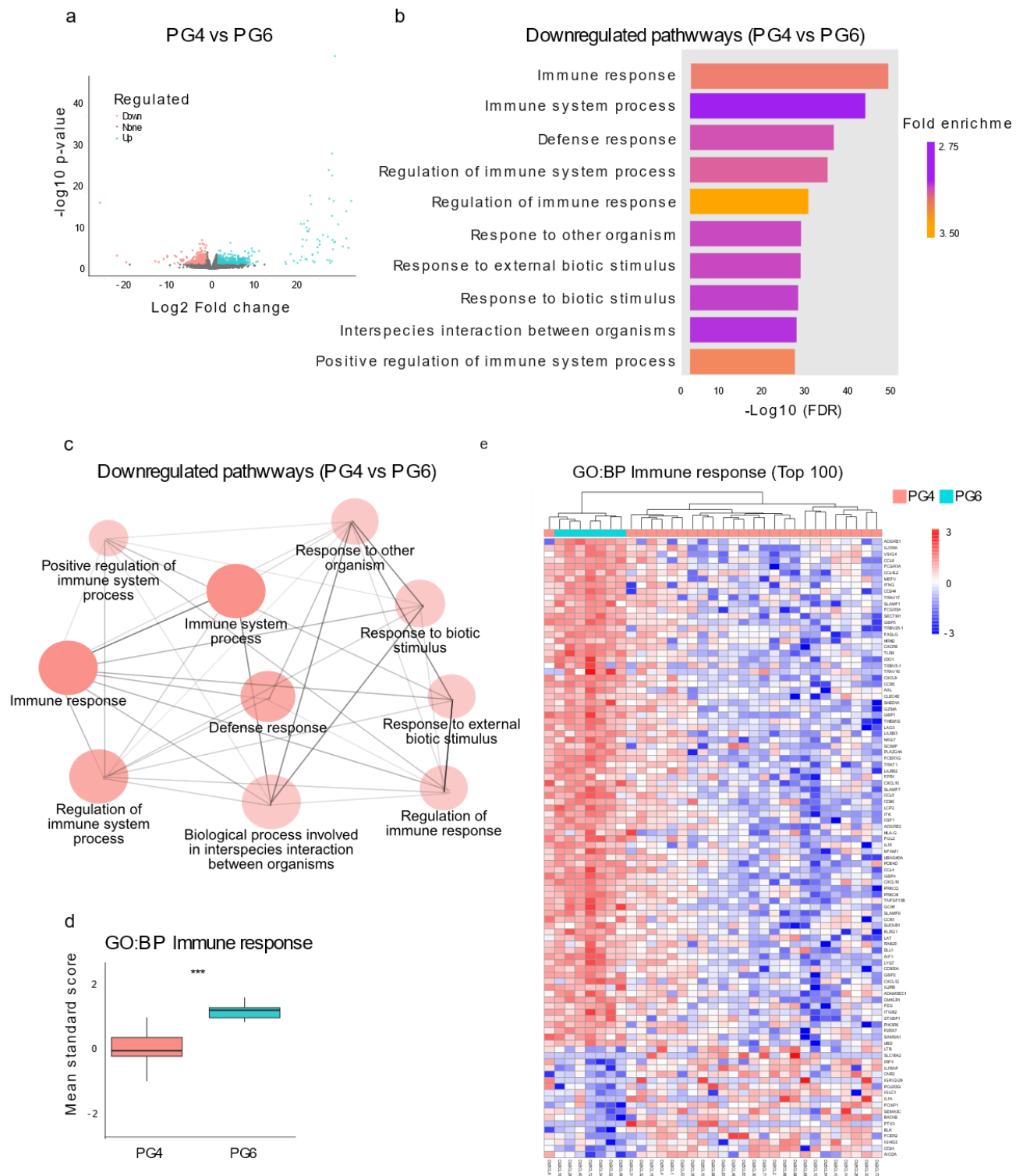


Figure 11. Transcriptomic differences between PG4 and PG6 reveal profound suppression of immune programs in PG4.

(a) Differential expression analysis showing widespread immune-related gene downregulation in PG4 and robust immune activation in PG6.

(b) GO-enrichment analysis of PG4-downregulated genes, highlighting suppression of core immune processes.

(c) Network of downregulated biological processes, illustrating functional coherence of

suppressed immune modules in PG4.

(d) GO:BP “Immune Response” signature scores, significantly lower in PG4 than PG6.

(e) Heatmap of the top 100 immune-related GO:BP genes, demonstrating the immune-inert phenotype of PG4 compared with the immune-active PG6

5.6 Immune cell composition highlights distinct microenvironmental states in PG4

To complement the transcriptomic analyses and to determine whether differences in gene expression translated into quantifiable variation in immune-cell infiltration, we performed computational deconvolution of bulk RNA-seq data using the LM22 leukocyte signature matrix. This allowed us to estimate the relative abundance of 22 immune cell populations across all samples and to compare PG4 with both PG other and PG6.

5.6.1 PG4 exhibits reduced abundance of multiple immune populations compared with other PGs

Figure 12a displays a heatmap of immune-cell proportions for PG4 versus PG other. Consistent with the transcriptomic signatures PG4 samples showed markedly reduced levels of several immune cell types, including CD8⁺ T cells, activated CD4⁺ memory T cells, M1 macrophages, dendritic cells, and NK cells. The suppression of these populations was evident across most PG4 samples, indicating a shared microenvironmental phenotype rather than isolated sample-specific variability. Conversely, PG other samples displayed a more heterogeneous immune composition, reflecting the broader biological diversity of the remaining proteogenotypes. Nonetheless, the overall pattern consistently showed higher relative immune infiltration in PG other compared with the uniformly depleted PG4 group. This analysis reinforces the notion that PG4 tumours exist in a globally immune-poor environment, which is congruent with the suppressed immune gene-expression programs identified earlier.

5.6.2 Direct comparison with PG6 confirms severe immune depletion in PG4

To further refine the interpretation of microenvironmental differences, we compared PG4 specifically to PG6, the subgroup previously identified as the most immune-active and clinically favourable. As shown in Figure 12b, PG6 samples exhibited substantially higher abundance of effector immune populations, including CD8⁺ T cells, activated CD4⁺ T cells, NK cells, M1 macrophages, and dendritic cells. In contrast, PG4 samples are consistently clustered with minimal immune infiltration, demonstrating a clear dichotomy between the two proteogenotypes. The PG4-PG6 comparison therefore reflects the most biologically polarized immune states in the dataset, providing further evidence that PG4's poor outcome may be driven, at least in part, by an inability to support an effective anti-tumour immune response.

5.6.3 Macrophage polarization reveals functional differences in PG4

Because macrophage subsets are key regulators of the tumour microenvironment, we examined differences in macrophage polarization states. Figure 12c shows the relative abundance of M1 (pro-inflammatory) and M2 (immunosuppressive) macrophages for PG4 versus PG other. Although these comparisons did not reach statistical significance, the trends were consistent with the general immune-depleted phenotype observed in PG4: both M1 and M2 macrophages were present at overall lower levels in PG4 samples. This suggests that PG4 does not simply shift macrophage polarization but exhibits a reduction in both inflammatory and suppressive macrophage subsets, reflecting global immune impairment rather than selective reprogramming.

5.6.4 Macrophage imbalance becomes significant in PG4 versus PG6

The contrast becomes more pronounced when PG4 is compared directly to PG6. PG6 exhibited significantly higher levels of M1 macrophages, consistent with activation of pro-inflammatory pathways and enhanced recruitment of antigen-presenting cells. M2 macrophages were also modestly increased in PG6 relative to PG4, suggesting that PG6 supports a more diverse and immunologically engaged macrophage milieu. These findings align with the spatial and transcriptomic observations from earlier figures, highlighting that PG6 hosts an active and

functionally competent immune microenvironment, while PG4 is characterized by profound macrophage depletion alongside suppressed T-cell responses.

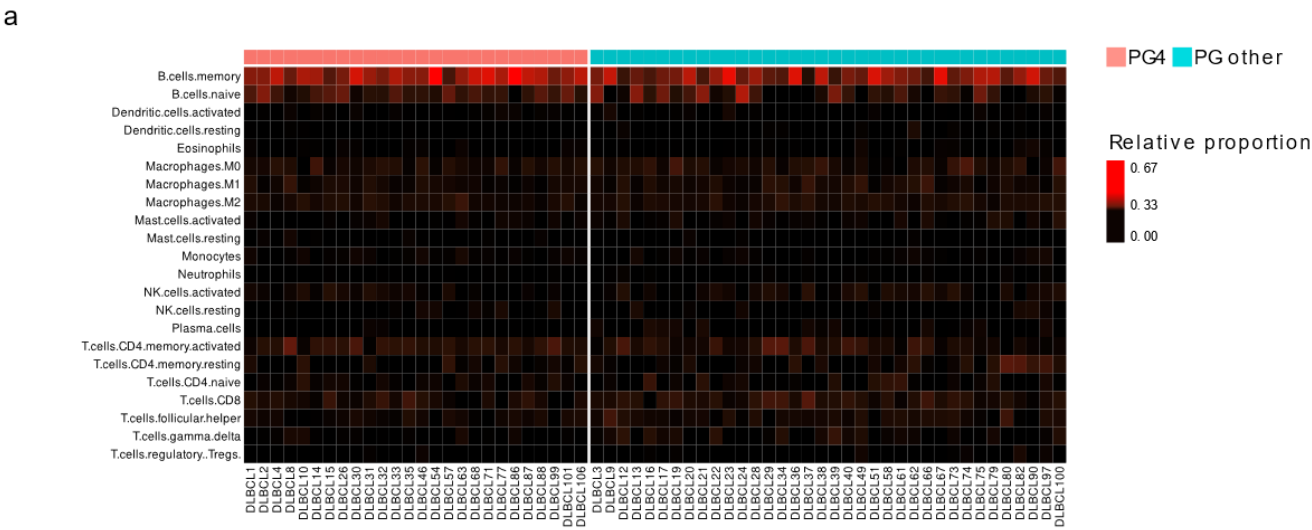
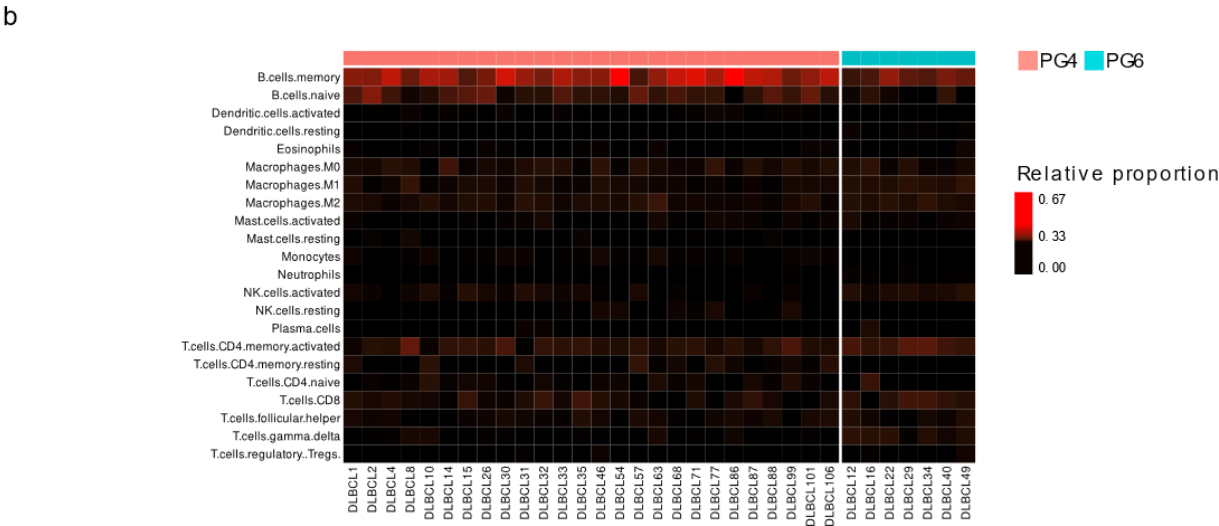
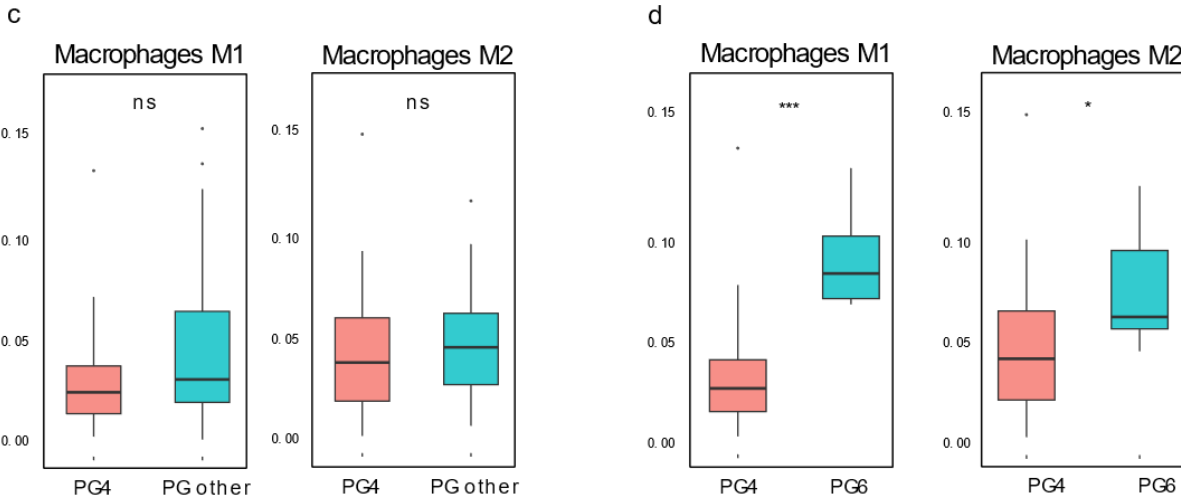


Figure 12. Immune cell deconvolution highlights major microenvironmental differences between proteogenotypes.

(a) Heatmap of CIBERSORT-derived immune cell proportions comparing PG4 versus PG other, showing reduced levels of T cells, NK cells, dendritic cells, and M1 macrophages in PG4.

(b) Heatmap of immune-cell fractions comparing PG4 and PG6, revealing the highest immune infiltration in PG6 and the most depleted immune microenvironment in PG4.

(c) Boxplots of M1 and M2 macrophages in PG4 versus PG other, showing overall reduced macrophage presence in PG4.

(d) Boxplots of M1 and M2 macrophages in PG4 versus PG6, highlighting significantly higher M1 levels in PG6.

5.7 MYC target gene activation in PG4 compared with all other proteogenotypes

To further investigate the proliferative programs that may contribute to the aggressive biology of PG4, we evaluated the expression of HALLMARK MYC TARGET genes in PG4 compared with all other proteogenotypes (PG other). MYC is a key oncogenic regulator associated with increased cellular growth, metabolic activation, and repression of immune processes. Its involvement in PG4 biology has been previously suggested and is explored here at the transcriptomic level.

5.7.1 PG4 shows significantly elevated MYC target gene expression relative to PG other

As shown in Figure 13a, PG4 exhibited a markedly higher mean standardized score for the HALLMARK MYC TARGETS (V1 union) gene-set compared with PG other, with strong statistical significance (Mann-Whitney U test $p = 7.9 \times 10^{-10}$). The distribution of scores indicates that elevated MYC target expression is a consistent feature across PG4 samples, while PG other showed substantially lower and more heterogeneous expression.

5.7.2 Coordinated upregulation of MYC target genes in PG4

The heatmap of the top 100 MYC target genes underscores this difference. PG4 samples displayed coherent, high-level expression of MYC-responsive genes

involved in ribosomal biogenesis, DNA replication, nucleotide metabolism, protein synthesis, and cell-cycle progression. In contrast, PG other samples showed markedly reduced activation of these genes, reflecting the broader diversity of proliferative states across the remaining proteogenotypes. The clear separation between PG4 and PG other highlights MYC target activation as a defining component of the PG4 transcriptional program.

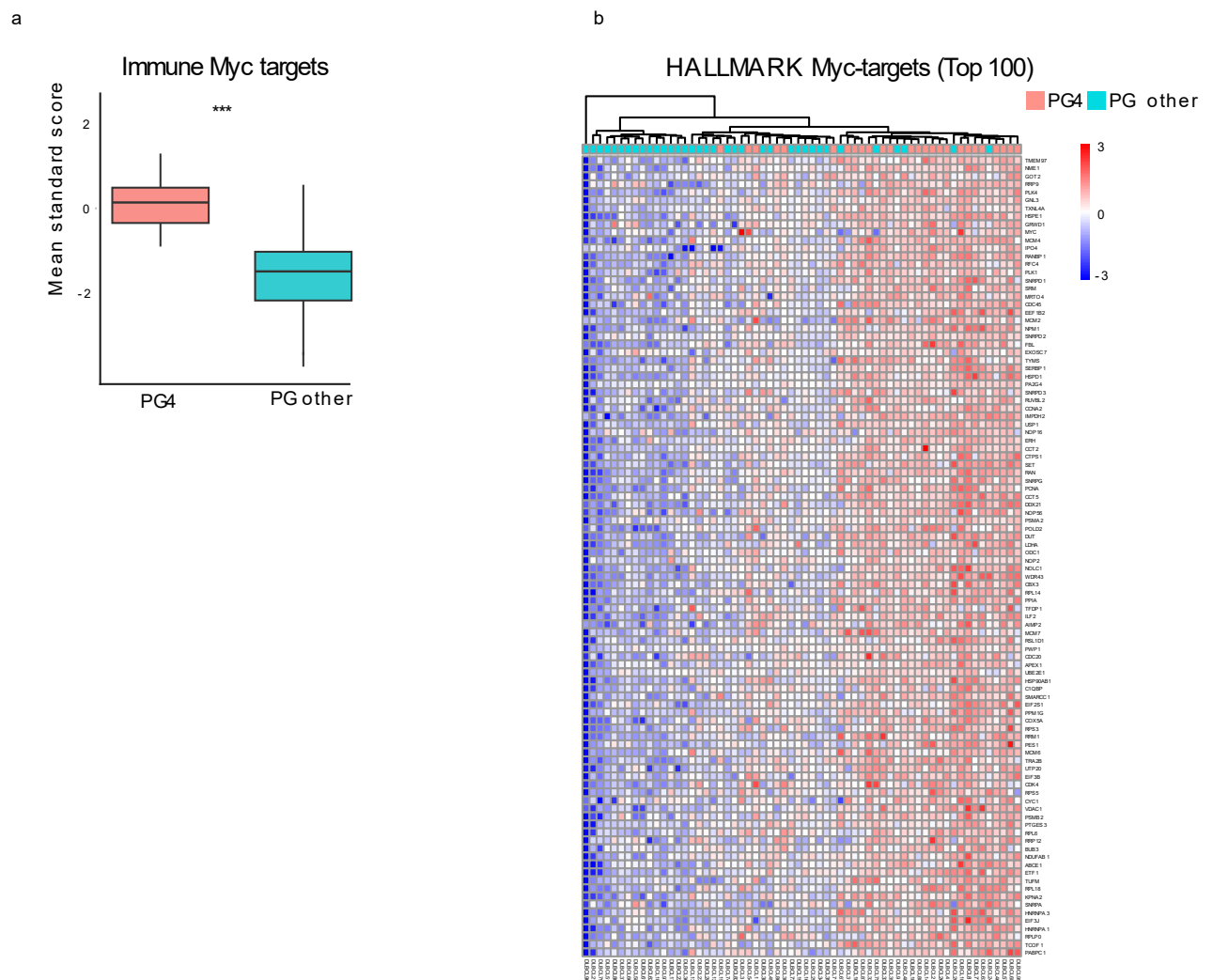


Figure 13. MYC target gene activation in PG4 compared with all other proteogenotypes. (a) Boxplot of the HALLMARK MYC TARGETS (V1 union) gene-set scores showing significantly higher MYC target expression in PG4 compared with PG other (Mann-Whitney U test $p = 7.9 \times 10^{-10}$). (b) Heatmap of the top 100 MYC target genes demonstrating broad upregulation of MYC-driven transcriptional programs in PG4 relative to the remaining proteogenotypes.

5.8 MYC target activation differentiates PG4 from the immune-enriched PG6 proteogenotype

To deepen the molecular characterization of the antagonistic phenotypes represented by PG4 and PG6, we assessed the expression of HALLMARK MYC TARGET genes, given the previously reported association of PG4 with proliferative and biosynthetic transcriptional programs. While the comparison of PG4 versus all other proteogenotypes provided a global overview of MYC target gene elevation in PG4, the direct contrast with PG6 offers a more refined perspective on MYC-driven biology in the context of the strongest immune-active subgroup.

5.8.1 Downregulated pathways in PG6 highlight functional consequences of MYC activation in PG4

The GO enrichment analysis revealed that pathways downregulated in PG6 relative to PG4 were strongly enriched for biological processes modulated directly or indirectly by MYC, including leukocyte activation, lymphocyte activation, defense response, and regulation of the immune system. These processes are known to be antagonistic to MYC-driven proliferative programs, suggesting that PG4 tumours may prioritize growth and biosynthetic activity at the expense of immune competence. This observation aligns with earlier findings of immune depletion in PG4 and supports a model in which MYC activation contributes mechanistically to the suppression of immune-related transcriptional networks.

5.8.2 PG4 exhibits markedly higher expression of HALLMARK MYC target genes

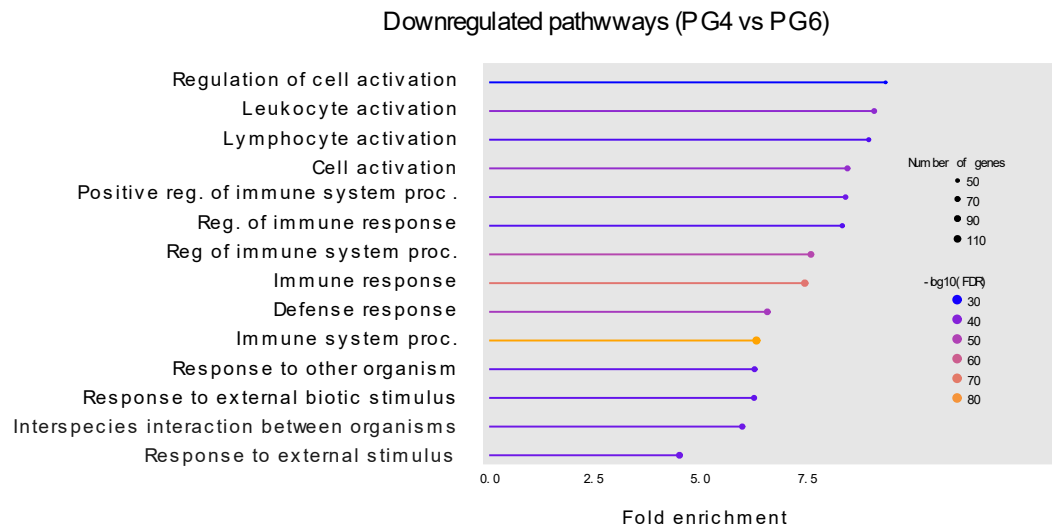
Consistent with proliferative program activation, PG4 samples displayed significantly higher expression of the HALLMARK MYC TARGET gene-set compared with PG6. The elevation of these MYC target genes including those involved in ribosomal biogenesis, metabolic activation, DNA replication, and protein synthesis reinforces the strong proliferative signature that characterizes PG4. In contrast, PG6 samples, which demonstrate broad immune activation and superior clinical outcomes, exhibited markedly reduced MYC target expression. The

reciprocal relationship between MYC activation and immune engagement is consistent with known MYC-mediated mechanisms of immune evasion.

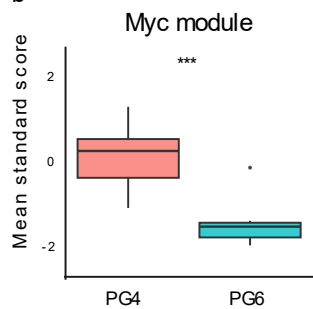
5.8.3 MYC target gene heatmap confirms coordinated upregulation in PG4

The heatmap of the top 100 HALLMARK MYC TARGET genes revealed a highly coherent transcriptional pattern in which PG4 samples displayed strong and widespread upregulation of MYC-regulated genes, whereas PG6 samples clustered separately with uniformly lower expression. This pattern indicates that MYC-driven transcriptional activation is a defining molecular feature of PG4 and provides a mechanistic contrast to the immune-enriched phenotype of PG6. The clean separation between the two proteogenotypes reflects the divergent biological programs underpinning their distinct clinical behaviours. Together, these findings support a model in which PG4 is characterized by dominant MYC target activation concurrent with broad immune suppression, whereas PG6 exhibits minimal MYC activation alongside heightened immune competence, providing a clear molecular rationale for their opposing prognostic profiles.

a



b



c

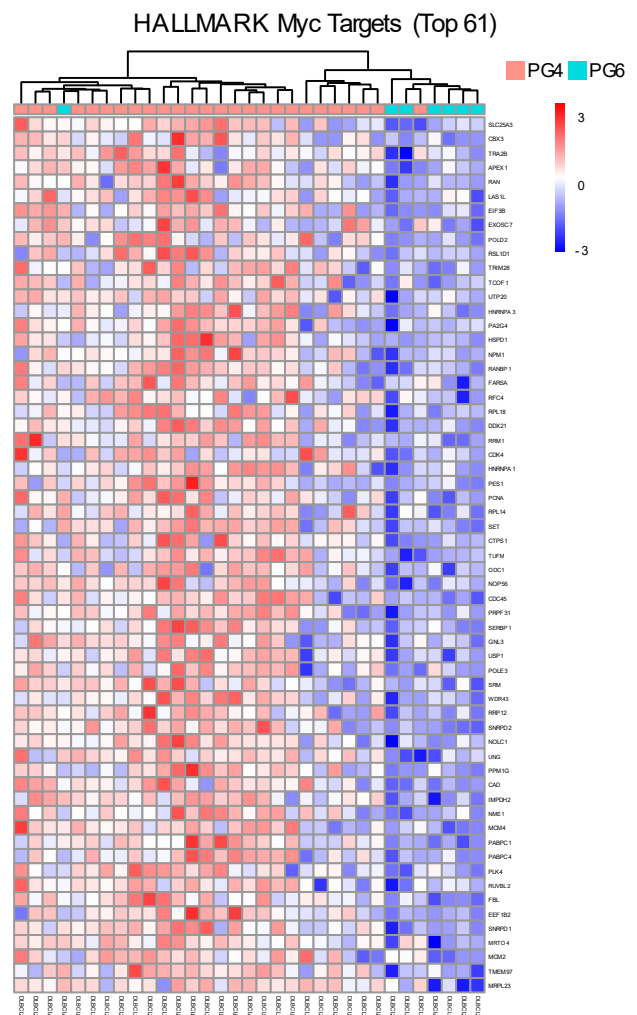


Figure 14. MYC target gene activation differentiates PG4 from the immune-rich PG6 proteogenotype.
 (a) GO enrichment analysis of pathways downregulated in PG6 relative to PG4, highlighting biological programs regulated by MYC and associated with cellular activation.

(b) Boxplot of the HALLMARK MYC TARGET gene-set scores showing significantly higher MYC target expression in PG4.
(c) Heatmap of the top 100 HALLMARK MYC TARGET genes, illustrating coordinated MYC-driven transcriptional activation in PG4 and reduced expression in PG6.

5.9 Somatic alterations across proteogenotypes

To complement the transcriptomic and microenvironmental analyses, we performed a comparative evaluation of somatic alterations across proteogenotypes using whole-exome sequencing (WES). A total of 45 samples passed stringent quality filtering and were included in the analysis of small mutations, copy-number alterations (CNAs) and predicted functional impact. Only alterations classified as oncogenic, likely oncogenic, or associated with gain- or loss-of-function were retained, ensuring that the analysis focused on biologically meaningful events.

Across the cohort, 29 genes met the threshold for inclusion in the multi-layered heatmap, comprising loci showing at least four functional alterations. The mutational patterns observed were heterogeneous, with most events occurring at low frequency and with no proteogenotype-specific clustering. Small mutations, amplifications and deletions were broadly distributed across PG groups without evident enrichment in any molecular category. Similarly, neither the mutation impact score (MIS) nor the tumour mutational burden (TMB) differed significantly between PG4, PG6 or the remaining PGs, indicating a generally quiet genomic landscape consistent with previous sequencing studies of DLBCL.

Despite the absence of proteogenotype-specific drivers, one observation emerged from the dataset: PRSS1 and PRSS2 displayed recurrent deletions predominantly in PG6. These genes encode pancreatic trypsinogen isoforms and have not been implicated in the pathogenesis of DLBCL or other B-cell malignancies. Their deletion in PG6 is therefore most consistent with a passenger event rather than a functionally relevant alteration, particularly given the lack of association with immune-related pathways, lymphoid signalling modules or antigen-presentation machinery. No additional recurrent CNAs or pathogenic mutations were specifically associated with PG6, and critically no genetic alterations emerged as candidates for explaining the severe immune depletion characteristic of PG4.

Taken together, the somatic mutation analysis reveals a genomic background largely devoid of PG-specific oncogenic events. The absence of structurally or

functionally convergent alterations across proteogenotypes supports the notion that the distinctive transcriptional identities and divergent immune phenotypes of PG4 and PG6 are not genetically driven. Instead, these results reinforce the central role of transcriptional programs, immune composition and spatial microenvironmental organization in shaping proteogenotypic behaviour, rather than recurrent somatic mutations or copy-number abnormalities.

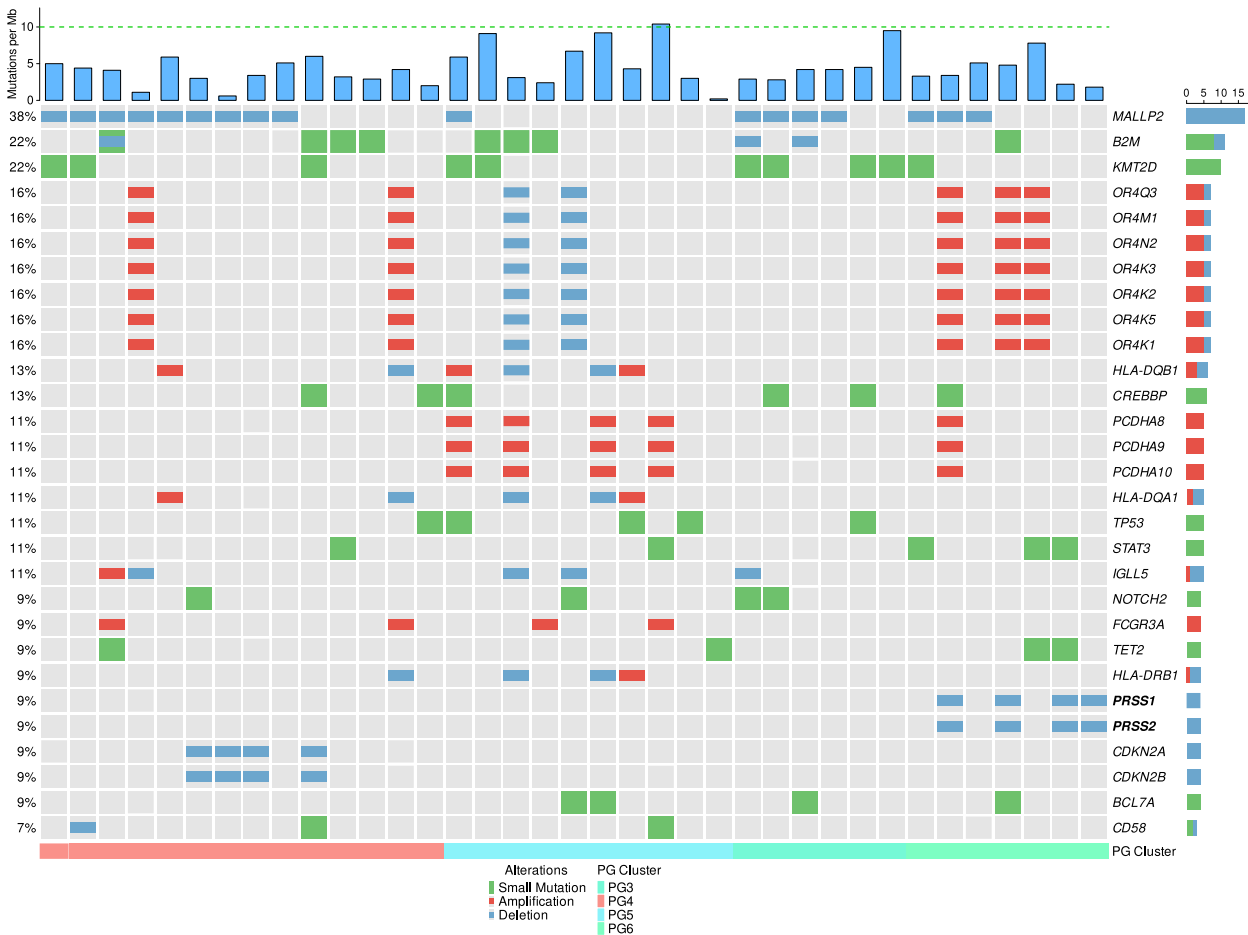


Figure 15. Somatic alterations across proteogenotypes. Heatmap showing small mutations, amplifications, deletions and predicted functional impact across 29 genes with ≥ 4 alterations in 45 DLBCL samples. Samples are grouped by proteogenotype, and genes are annotated for mutation type and functional consequence (gain-of-function, loss-of-function or oncogenic classification). PRSS1 and PRSS2 deletions appear predominantly in PG6, while no additional recurrent or PG-specific alterations are observed.

6. Discussion

6.1. Overview of the key findings and their clinical - biological relevance

Diffuse large B-cell lymphoma remains one of the most heterogeneous hematologic malignancies, both biologically and clinically, despite substantial advances in the molecular classification of the disease. Traditional stratification systems, most notably the germinal centre B-cell-like (GCB) and activated B-cell-like (ABC) cell-of-origin (COO) model and, more recently, the genetic subtypes described by Chapuy et al. and Schmitz et al. (10), have significantly contributed to understanding disease diversity but still fail to fully explain the profound variability in patient outcomes(57,58). In this context, the proteogenotype (PG) framework recently developed by Enßle et al. (13) (manuscript in revision) represents a major conceptual advance, as it captures the interplay between tumour-intrinsic transcriptional programs, immune microenvironment states, and proteomic features(17). Building upon this framework, the present thesis provides an integrated multi-omic and spatial analysis of DLBCL, with the specific aim of elucidating the biological distinction between the poorest-outcome proteogenotype (PG4) and the most immune-enriched and clinically favourable subgroup (PG6).

Across spatial, transcriptomic, and immunological layers, the results of this work converge to demonstrate a profound immunological and molecular dichotomy between PG4 and PG6. This divergence aligns with, but also substantially extends, previous findings by introducing a unified perspective in which immune exhaustion, microenvironmental collapse, MYC-driven proliferative programs, and spatial disorganization appear tightly interconnected(59,60). By integrating RNA sequencing (RNA-seq), multiplex immunofluorescence (mIF), computational deconvolution, and whole-exome sequencing (WES), the study provides compelling evidence that the poor prognosis associated with PG4 is not primarily driven by genetic alterations, but rather by synergistic tumour-intrinsic transcriptional programs and microenvironmental failure.

The first major observation emerging from this study is the widespread suppression of immune-related transcriptional activity in PG4, both in comparison to all other proteogenotypes (PG other) and, more starkly, to PG6. Differential gene expression analyses revealed that PG4 is characterized by downregulation of genes involved

in T-cell activation, antigen presentation, interferon signalling, and chemokine-mediated immune recruitment. Gene ontology (GO) analysis further showed enrichment of pathways associated with developmental processes, cell-cell adhesion, and immune system regulation among downregulated transcripts, supporting the notion that PG4 lacks not only immune effector function but also the basic structural and signalling architecture necessary for productive immune-tumour interactions. These findings are consistent with the immune-desert phenotype previously described for PG4, yet they provide new resolution by showing that immune suppression is not confined to isolated pathways but instead represents a systems-level collapse of the transcriptional programs sustaining immune competence(14).

Complementing transcriptomic evidence, spatial profiling through multiplex immunofluorescence confirmed an extremely immune-depleted and spatially disorganized microenvironment in PG4, in stark contrast to the highly immune-enriched architecture of PG6. Quantitative mIF data demonstrated that PG4 tissues display dramatic reductions in nearly all immune cell populations, including CD4⁺ helper T cells, CD8⁺ cytotoxic T cells, B cells, dendritic cells, NK cells, and both M1 and M2 macrophages, resulting in a microenvironment dominated by malignant B cells with minimal immune surveillance. Moreover, deep-learning based single-cell segmentation revealed that even when immune cells are present in PG4 samples, they remain spatially segregated from tumour clusters, suggesting impaired trafficking, exclusionary stromal barriers, or defective chemokine signalling(61,62). Spatial interaction likelihood analyses further confirmed the near absence of macrophage-T cell interactions in PG4, a feature that severely compromises antigen presentation and T-cell priming and likely contributes to functional immune paralysis. In contrast, PG6 samples exhibited dense immune infiltrates with preserved spatial organization, rich macrophage-T cell interfaces, and robust tumour-immune cell interactions, features characteristic of an “immune-hot” environment capable of exerting effective anti-tumour activity.

A second major theme uncovered in this work is the prominent activation of MYC target genes in PG4, which appears tightly coupled to its immune-cold phenotype. MYC-driven transcriptional programs, including ribosomal biogenesis, oxidative metabolism, DNA replication, and protein synthesis, were consistently upregulated in PG4 relative to both PG other and PG6. The HALLMARK MYC TARGET gene

sets showed sharp and coherent separation between PG4 and the other proteogenotypes, with PG4 samples displaying homogeneously elevated expression of MYC-responsive genes. This observation aligns with the dark zone - like phenotype previously described for PG4 and supports a model in which unrestricted MYC activity contributes not only to aggressive tumour proliferation but also to suppression of immune engagement, potentially through mechanisms such as metabolic competition, repression of antigen presentation pathways, or modulation of cytokine networks(63,64). The reciprocal pattern observed in PG6, characterized by low MYC target expression combined with strong immune activation, reinforces the biological antagonism between these two proteogenotypes and suggests that PG6 may represent a state in which immune surveillance successfully constrains tumour proliferation, thereby contributing to favourable clinical outcomes.

A third critical insight from this thesis is the absence of recurrent or proteogenotype-enriched genetic alterations that could explain the divergent phenotypes of PG4 and PG6. Whole-exome sequencing revealed a genomically quiet landscape, with no differences in small mutations, copy-number alterations, mutation impact scores, microsatellite instability, or tumour mutation burden across proteogenotypes. The only recurrent event identified was the deletion of PRSS1 and PRSS2 in PG6, which lacks known biological relevance in lymphoma and is most likely a passenger alteration. Importantly, no genomic features correlated with immune depletion, MYC activation, or proteogenotypic identity. This finding strongly supports the concept that PG4 and PG6 represent transcriptionally and microenvironmentally defined states, rather than genetically distinct entities, and underscores the need to study DLBCL biology beyond mutational profiling alone(17,65).

Taken together, these findings establish a robust multi-layered model for understanding the biology of high-risk versus favourable DLBCL proteogenotypes. PG4 emerges as a tumour-intrinsic, MYC-driven, immune-excluded state in which the collapse of cellular immunity and spatial immune architecture promotes unchecked malignant expansion. In contrast, PG6 represents a microenvironmentally competent, immune-enriched state in which coordinated innate and adaptive immune responses contribute to disease control. This dichotomy carries significant implications for prognostic assessment, biological stratification, and potential therapeutic targeting, especially in the context of

immunotherapies whose success depends on the presence of functional immune ecosystems.

The following sections of the Discussion will deepen this perspective by examining in detail:

- (i) the microenvironmental architecture revealed by mIF,
- (ii) the transcriptional landscape underlying the PG4 - PG6 contrast,
- (iii) the mechanistic contribution of MYC signatures,
- (iv) the absence of genomic drivers, and
- (v) the implications of this integrated multi-omic analysis for precision medicine in DLBCL.

6.2. Spatial and microenvironmental architecture of PG4 and PG6

A foundational step in understanding the biological divergence between PG4 and PG6 is the characterization of their tumour microenvironmental architecture. Spatial profiling through multiplex immunofluorescence (mIF), enhanced by deep learning-based cell segmentation, provides direct visualization of immune infiltration, cellular distribution, and interaction patterns(66–68). This analysis reveals striking microenvironmental contrasts between the two proteogenotypes and establishes the morphological and functional context needed to interpret the transcriptional mechanisms explored in the subsequent section.

Across all samples analysed, PG4 emerged as a prototypical immune-excluded proteogenotype(59). mIF images demonstrated large, dense tumour regions almost entirely devoid of immune cells. Both CD4⁺ and CD8⁺ T cells appeared sparse and scattered, often as isolated single cells without spatial organization or proximity to malignant B-cell clusters. Macrophages marked by CD68 and CD163 were nearly absent, and when present, they were typically confined to stromal or peripheral regions rather than integrated within the tumour mass. This cell-poor, architecturally “empty” landscape reflects a profound inability of the microenvironment to support immune recruitment, positioning, or coordinated responses. The absence of immune aggregates or structured niches suggests that immune infiltration is not merely reduced but actively impaired at multiple levels of tissue organization(69,70). In contrast, PG6 displayed an immune-enriched and spatially coordinated architecture, evident in both the density and organization of immune infiltrates. PG6 tissues contained abundant CD4⁺ and CD8⁺ T cells, as well as CD20⁺ B cells, which

were distributed throughout tumour areas rather than restricted to the periphery. Macrophages, including both M1- and M2-associated phenotypes, were readily detectable and often found in close proximity to T cells, forming micro-anatomical hubs reminiscent of tertiary lymphoid-like structures(70,71). These features signal an immunologically active environment capable of antigen presentation, T-cell priming, and effector engagement. The presence of immune clusters and coordinated multicellular structures in PG6 supports its role as a clinically favourable and immune-reactive proteogenotype.

Quantification of immune populations through single-cell segmentation confirmed these qualitative observations. PG4 contained the lowest levels across virtually all immune lineages such as helper and cytotoxic T cells, NK cells, dendritic cells, and macrophage subsets are indicating a global quantitative depletion rather than selective immune loss(60). PG6 showed significantly higher abundance across the same populations, reinforcing its immune-competent identity. Of particular relevance, PG4 showed reductions in both M1 and M2 macrophages, distinguishing it from immunosuppressive phenotypes typically characterized by M2 dominance. Instead, PG4 appears to lack the macrophage compartment altogether, depriving the tumour of essential antigen-presenting and cytokine-regulating functions.

The functional relevance of these spatial patterns becomes clearer when analysing cell-cell interactions. PG6 samples exhibited a high frequency of macrophage-T cell co-occurrence, indicative of active antigen presentation and immune activation circuits(62,72). T cells in PG6 were also observed in close spatial proximity to malignant B cells, reflecting successful infiltration into tumour nests and supporting cytotoxic engagement. PG4, by contrast, showed an almost complete collapse of these interactions. Macrophage-T cell interfaces were virtually absent, and T cells were spatially segregated from tumour cells, suggesting disrupted chemokine guidance, stromal barriers, or deficient antigen presentation. This functional disconnection provides a tissue-level explanation for the poor prognosis associated with PG4(29,68).

Importantly, the spatial findings align with and motivate the transcriptional analyses presented in the next section. The severe immune depletion and disorganization observed in PG4 raise a fundamental mechanistic question: why is this microenvironment unable to recruit or sustain immune cells? Conversely, the structured immune architecture of PG6 prompts an investigation into the

transcriptional programs that support immune activation, interferon signalling, and antigen presentation. To address this, the following section focuses on the transcriptional mechanisms that give rise to the spatial phenotypes described here, revealing how gene expression programs in PG4 and PG6 shape their respective immune landscapes. Taken together, these spatial findings describe two fundamentally distinct immune ecosystems and provide the basis for examining the transcriptomic landscape distinguishing PG4 from PG6.

6.3. Transcriptomic divergence between PG4 and PG6: Molecular determinants of prognostic polarization

A central objective of this study was to dissect the transcriptional programs that distinguish PG4, the poorest-outcome proteogenotype, from PG6, the most immune-enriched and clinically favourable subgroup. By performing layered differential expression analyses, first comparing PG4 to all other proteogenotypes and subsequently focusing specifically on PG4 versus PG6, the results reveal a deep and coherent transcriptional polarization that underpins the contrasting biological and clinical behaviour of these subtypes(17,65). Collectively, the findings demonstrate that PG4 is defined by the suppression of immune-related gene programs and the activation of proliferative, MYC-linked transcriptional circuits, whereas PG6 exhibits a reciprocal pattern characterized by robust immune engagement and attenuated proliferative signalling. This transcriptomic antagonism provides a molecular explanation for the divergent tumour microenvironment states and clinical outcomes observed between the two PGs.

The initial comparison of PG4 versus all remaining proteogenotypes (PG other) already revealed a broad landscape of transcriptional dysregulation in PG4, dominated by the downregulation of immune-related genes. However, this global comparison encompasses proteogenotypes with heterogeneous immunological profiles and therefore provides only a partial view of PG4's distinctive biology. By narrowing the analysis to the PG4 versus PG6 contrast, the two most biologically and clinically polarized subgroups, this thesis highlights the most extreme and biologically informative transcriptional divergence within the DLBCL proteogenotypic spectrum(14,60).

The volcano plot comparing PG4 and PG6 demonstrates a striking asymmetry in gene expression: hundreds of immune-related genes are substantially downregulated in PG4, whereas PG6 exhibits strong upregulation of transcripts associated with adaptive immunity, innate immune activation, antigen presentation, and inflammatory signalling. PG6 tumours display elevated expression of chemokines (e.g., CCL5, CXCL9, CXCL10), interferon-stimulated genes (e.g., IFIT family members, IRF1), and cytotoxic mediators such as granzyme B and perforin, signature components of an effective anti-tumour immune response(59,73). In contrast, PG4 exhibits a profound loss of these signals, mirroring the spatial immune desert observed in multiplex immunofluorescence. The directionality and magnitude of these changes are consistent across samples within each group, reflecting intrinsic and stable transcriptional states rather than stochastic or sample-specific fluctuations.

Gene ontology (GO) term enrichment analysis further supports these observations by demonstrating that the genes downregulated in PG4 are overwhelmingly concentrated in pathways essential for immune competence. GO categories such as “immune response”, “immune system process”, “leukocyte activation”, “regulation of immune system process”, and “defense response” are among the top suppressed biological processes in PG4 compared with PG6. These pathways encompass T-cell receptor signalling, antigen processing and presentation, interferon signalling, cytokine production, and chemotaxis, which are core components of adaptive and innate immunity. Of particular note is the downregulation of multiple HLA class II-related genes in PG4, which impairs antigen presentation to CD4⁺ T helper cells and compromises the orchestrated activation of cytotoxic responses(14,74). The breadth of suppressed immune pathways reveals that PG4 tumours do not simply exhibit reduced infiltration of immune effector cells; rather, they lack the transcriptional infrastructure required to initiate, sustain, or amplify anti-tumour immunity.

The network visualization of downregulated GO terms underscores the functional coherence of these suppressed pathways. Immune-related processes form a tightly interconnected module, indicating that transcriptional suppression in PG4 affects a unified set of immune functions rather than isolated biological events. Such modular suppression suggests a mechanistic foundation for the immune-desert phenotype of PG4, whereby multiple layers of immune regulation, ranging from chemokine

secretion to effector function, are simultaneously impaired(17). This systems-level collapse contrasts sharply with PG6, where immune pathways form a highly active and integrated network consistent with an immunologically “hot” environment.

Beyond broad enrichment of immune signatures, the heatmap of the top 100 immune-related GO:BP genes provides direct visual evidence of the binary transcriptional phenotypes separating PG4 and PG6. PG6 samples uniformly express high levels of genes involved in antigen presentation (HLA-DRA, HLA-DPA1, HLA-DPB1), T-cell activation (CD3D, CD3E, LCK), interferon signalling (STAT1, CXCL10, IFIT3), and cytotoxicity (GZMB, PRF1). In contrast, PG4 samples cluster tightly and display homogeneously low expression across these immune gene sets, illustrating a consistent and entrenched immune-inert transcriptional identity. The robustness of this clustering indicates that the PG4 phenotype is not the result of variable immune suppression at the level of individual genes, but rather reflects a coordinated silencing of entire immune programs, potentially mediated by upstream oncogenic or epigenetic regulators(60,68).

In parallel, PG4 demonstrates enrichment of transcriptional programs associated with proliferation, biosynthesis, and metabolic activation, features tightly linked to MYC-driven biology. Although a detailed discussion of MYC signalling is provided in the next section, it is important to highlight here that many of the genes upregulated in PG4 (e.g., those involved in ribosome biogenesis, nucleotide metabolism, DNA replication, and protein synthesis) belong to pathways known to be activated downstream of MYC and are inversely correlated with immune activation. This inverse relationship reinforces the notion that proliferative dominance in PG4 may contribute directly or indirectly to immune evasion, either through metabolic competition, suppression of antigen processing, or modulation of interferon pathways(63,64,75).

These transcriptional findings not only align with the spatial and immunophenotypic observations of this thesis but also integrate seamlessly with previously published single-cell studies demonstrating T-cell exhaustion and antigen presentation failure in high-risk DLBCL ecosystems. Steen et al. (11) reported that DLBCL tumours with suppressed HLA class II expression and diminished interferon signalling exhibit markedly reduced effector T-cell infiltration, precisely the molecular and spatial features observed in PG4 in the present study. Conversely, immune-enriched DLBCL subtypes characterized by robust interferon signalling and chemokine

expression, such as those described by Kotlov et al. (15), closely resemble the molecular architecture of PG6.

In summary, the transcriptomic differences between PG4 and PG6 reveal a deep molecular polarization that aligns with their opposing microenvironmental states and clinical outcomes. PG4 is defined by a transcriptional program of immune silence and proliferative dominance, reflecting an aggressive tumour-intrinsic phenotype combined with profound immune incompetence. PG6, on the other hand, embodies a biologically favourable state characterized by active immune engagement, high interferon signalling, and coordinated antigen presentation. This transcriptomic dichotomy not only explains the spatial and cellular findings described earlier but also offers mechanistic insights into why PG4 is resistant to immune-mediated tumour control, whereas PG6 is associated with superior prognosis(59,61). The strong transcriptional polarization observed between PG4 and PG6 suggests the involvement of upstream oncogenic regulators. Among these, MYC emerges as a central candidate due to its established role in sustaining proliferative programs and repressing antigen presentation and interferon signalling. We therefore next explore the contribution of MYC-driven biology to the PG4 phenotype.

6.4. MYC-driven biology and its role in shaping the PG4 phenotype

The transcriptomic contrasts described in Section 6.3 provide a compelling foundation for understanding the immune-rich state of PG6 and the immune-depleted architecture of PG4. However, a complete interpretation of these differences requires considering the tumour-intrinsic proliferative programs that characterize each proteogenotype. Among these, MYC-driven transcriptional activity emerges as a central axis distinguishing PG4 from the other DLBCL proteogenotypes and provides critical mechanistic insights into its uniquely aggressive behaviour(75,76).

Across multiple analyses, PG4 consistently displayed the highest expression of MYC target genes, reflecting a deeply entrenched proliferative state. Gene-set enrichment confirmed strong upregulation of both HALLMARK_MYC_TARGETS_V1 and V2, positioning PG4 as the most MYC-driven proteogenotype within the cohort(12,77). These signatures encompass fundamental cellular processes such as ribosomal biogenesis, nucleotide

metabolism, protein translation, and mitochondrial function, all essential components of the anabolic and replicative machinery underlying rapid tumour growth.

This MYC-dominated transcriptional program aligns with PG4's spatial phenotype. Tumours with sustained MYC activity frequently adopt an immune-poor architecture, a phenomenon described in several malignancies, including lymphomas, breast cancer, and hepatocellular carcinoma(63,64,78). Mechanistically, MYC activation has been shown to suppress antigen presentation through transcriptional repression of HLA-II machinery, reduce interferon responsiveness, alter metabolic landscapes to disadvantage immune cells, and reshape cytokine networks in ways that impair T-cell recruitment(63,74). These mechanisms offer a plausible biological explanation for why PG4 tumours, despite harbouring malignant B cells with high metabolic demand, show minimal immune infiltration and almost no immune-tumour cell interactions.

In line with these observations, PG4 demonstrated suppression of multiple pathways that are typically incompatible with high MYC activity. Interferon signalling, antigen presentation, and chemokine-mediated immune recruitment were all downregulated in PG4 compared to PG6. Such pathways are known to be negatively regulated by MYC either directly, through transcriptional repression, or indirectly, through epigenetic remodelling and metabolic competition. For example, MYC-driven increases in protein synthesis and oxidative phosphorylation create nutrient and oxygen landscapes that are unfavourable for cytotoxic T cells and dendritic cells, particularly in settings with limited metabolic plasticity(79,80).

PG4 also displayed features reminiscent of the dark-zone phenotype of germinal centre B cells, which is strongly associated with MYC activity. The dark zone is characterized by high proliferation, somatic hypermutation, and limited antigen presentation, biological states recapitulated in PG4's transcriptional signature(81,82). In this context, MYC activation may help maintain a proliferative, immune-silent tumour state that resists immune-mediated clearance. Conversely, PG6 shares transcriptional characteristics of the germinal centre light zone, which is more compatible with immune engagement, antigen presentation, and immune-driven selection.

Importantly, the high MYC activity in PG4 does not appear to be driven by recurrent genomic alterations, as whole-exome sequencing did not identify PG-specific

mutations or copy-number changes affecting MYC or its upstream regulators. This observation underscores that MYC activation in PG4 is likely regulatory and transcriptional, rather than mutational. It may arise from epigenetic programs, microenvironmental cues, or proteomic alterations characteristic of the proteogenotype. This finding is consistent with emerging evidence suggesting that MYC activation in DLBCL can be maintained through enhancer remodelling, chromatin accessibility changes, or dysregulated transcriptional complexes rather than genetic lesions alone(13,83).

The functional consequences of sustained MYC signalling in PG4 are multifaceted. At the cellular level, MYC enhances proliferation and plasma membrane remodelling, enabling rapid tumour expansion. At the microenvironmental level, MYC suppresses immune recruitment and antigen visibility, creating an ecosystem in which malignant B cells proliferate in relative isolation. At the clinical level, such MYC-driven immune exclusion likely contributes to resistance to both cytotoxic therapies and immunotherapies, consistent with the poor outcomes associated with PG4(84,85).

In contrast, PG6 exhibits markedly lower MYC target expression, suggesting a more balanced proliferative state that permits or even supports immune engagement. Tumours with attenuated MYC activity are more permissive to interferon signalling and antigen presentation, facilitating the immune-rich phenotype observed in PG6 and contributing to its superior prognosis(59,60).

Overall, the MYC-driven biology of PG4 provides a unifying mechanistic framework that integrates transcriptional suppression of immune pathways, spatial immune exclusion, and clinical aggressiveness. By sustaining a proliferative program that is intrinsically antagonistic to immune activation, MYC appears to be a key driver of PG4's immune-cold ecosystem and its associated poor clinical outcomes.

6.5. Genomic landscape of PG4 and PG6: Limited contribution of genetic alterations to proteogenotype identity

While transcriptomic and spatial analyses reveal profound differences between PG4 and PG6, the whole-exome sequencing (WES) component of this study demonstrates that these proteogenotypes do not differ at the level of recurrent genomic alterations. This finding is crucial, as it indicates that the aggressive,

immune-cold phenotype of PG4 and the immune-enriched, clinically favourable phenotype of PG6 arise predominantly from regulatory and microenvironmental programs, rather than from distinct mutational architectures. The genomic analysis therefore provides an important counterpoint to the transcriptional and spatial findings, highlighting that proteogenotype identity is not genomically encoded(12,13).

Across all samples, WES revealed a relatively quiet mutational landscape, consistent with what is often observed in unselected DLBCL cohorts(65). When comparing PG4 and PG6 across a wide range of genomic features including single nucleotide variants, small insertions and deletions, copy-number variations (CNVs), and mutation impact scores, no significant differences emerged. Neither the overall number of mutations per sample nor the distribution of predicted deleterious variants differed between the two proteogenotypes. This suggests that the biological divergence of PG4 and PG6 cannot be explained by differential mutational burden or by enrichment for damaging variants in key oncogenic pathways.

Similarly, tumour mutation burden (TMB) did not vary meaningfully between PGs. This is notable given the established relationship between TMB and immunogenicity in some solid cancers; however, in DLBCL, TMB has historically shown limited predictive power for microenvironmental features or treatment responses(86). Consistent with this, the shared low-to-intermediate TMB across PGs supports the conclusion that neither PG4's immune exclusion nor PG6's immune activation is driven by increased neoantigen load.

Analysis of microsatellite instability (MSI) further reinforced this point, with no evidence of MSI-high status in either PG4 or PG6. MSI-associated hypermutation, which can dramatically reshape tumour-immune interactions in certain malignancies, does not appear relevant to the biology of these proteogenotypes(12). Copy-number analysis likewise did not reveal PG-specific genomic aberrations. Most observed CNVs were sporadic or shared across multiple PGs without proteogenotype specificity. The single recurrent event identified is a deletion encompassing PRSS1 and PRSS2 in PG6 samples, which is not known to have biological relevance in lymphomagenesis and is most likely a benign passenger alteration. Importantly, no CNVs involving known regulators of B-cell identity, antigen presentation, interferon signalling, or MYC regulation were enriched in either PG4

or PG6. This further supports the conclusion that the immune and proliferative phenotypes of these proteogenotypes are not encoded at the genomic level(17,65). Perhaps most striking is the absence of genomic alterations that could account for the high MYC-target activity observed in PG4. Neither MYC rearrangements nor amplifications were enriched in PG4 compared with other PGs. This indicates that PG4's MYC-driven transcriptional program arises not from genetic changes to the MYC locus itself, but from non-genetic mechanisms such as enhancer activity, chromatin remodelling, or dysregulated upstream signalling pathways. These mechanisms are increasingly recognized as key drivers of MYC activation in lymphomas and further emphasize the importance of regulatory rather than mutational events in shaping PG4 biology(13,83,87).

Collectively, the absence of PG-specific genomic alterations underscores the central message of this study: the divergence between PG4 and PG6 is not genetically hardwired. Instead, their distinct phenotypes emerge from differences in transcriptional programming, microenvironmental composition, spatial architecture, and protein-level regulation. This conclusion aligns with recent large-scale lymphoma studies showing that transcriptional and microenvironmental classifications often outperform genomic subtyping in capturing clinically relevant heterogeneity(17).

From a clinical perspective, this finding has meaningful implications. Therapies that target genomic lesions may have limited utility in distinguishing or treating PG4 and PG6, whereas approaches that modulate the immune microenvironment or disrupt regulatory programs such as epigenetic therapies, interferon pathway activators, or metabolic modulators, may hold greater promise. Moreover, the lack of genome-encoded differences reinforces the value of proteogenomic profiling and spatial biology in capturing features that are invisible to mutational analysis alone.

In summary, the genomic analysis demonstrates that the fundamental differences between PG4 and PG6 do not stem from unique genetic alterations. Rather, they arise from transcriptional, spatial, and immunological programs that operate independently of the mutational landscape. These findings position PG4 and PG6 as strong examples of how regulatory states, rather than genetic lesions, can dominate tumour behaviour and clinical outcome in DLBCL.

6.6. Integrated interpretation and clinical implications

The multi-layered analyses performed in this study, including spatial profiling, transcriptomics, MYC-driven pathway evaluation and whole-exome sequencing, collectively reveal a coherent biological framework that explains the divergent clinical behaviours of PG4 and PG6. Although each analytical modality provides unique insights, their convergence offers a comprehensive understanding of how tumour-intrinsic programs and microenvironmental states interact to shape disease aggressiveness in DLBCL, consistent with emerging integrative models of lymphoma biology(14,17).

A central insight emerging from this integrated perspective is that the PG4 phenotype cannot be attributed to any single molecular event, but instead reflects a multilayered interplay between transcriptional repression of immune programs, MYC-driven proliferation, and structural collapse of the tumour immune ecosystem. Spatial imaging reveals that PG4 tumours are characterized by extreme immune depletion and lack of functional immune-tumour cell interfaces(59,61). These tissue-level observations would be insufficient to explain the phenotype without transcriptomic data, which demonstrate a coordinated downregulation of interferon signalling, chemokine expression, antigen presentation, and T-cell activation pathways in PG4. This indicates that immune exclusion is not simply due to absence of recruitment, but arises from a deeper failure to activate or sustain immune signalling at the transcriptional level(14).

The MYC-driven biology of PG4 provides an upstream mechanistic explanation. MYC is known to reprogram cellular metabolism, repress interferon responses, and inhibit antigen presentation pathways. Its strong and selective activation in PG4 not only sustains tumour proliferation but also generates a metabolic and transcriptional environment fundamentally inhospitable to cytotoxic lymphocytes and antigen-presenting cells. This MYC-driven immune suppression is consistent with reports in other lymphomas and solid tumours, where MYC hyperactivation correlates with immune evasion and resistance to immune-mediated clearance(63,64,78).

Importantly, WES analyses demonstrate that these features are not genetically encoded. The absence of recurrent mutations, copy-number alterations, or mutational signatures specific to PG4 or PG6 indicates that proteogenotype identity arises from non-genetic regulatory programs rather than genomic lesions. This is profoundly relevant, as it suggests that therapeutic strategies aimed solely at

targeting mutations are unlikely to shift a tumour from an immune-cold PG4 state into an immune-competent PG6-like state. Instead, approaches that modulate epigenetic landscapes, chromatin accessibility, interferon signalling, or metabolic environments may be more effective for reprogramming PG4 biology(13,65).

Clinically, the integrated findings of this study underscore a key principle: immune competence, not mutational burden, is the primary determinant of outcome within the DLBCL proteogenotype framework. PG6, with its immune-enriched architecture and strong interferon-driven activation, aligns closely with biologically favourable ecosystems described in other large-scale analyses of DLBCL and is associated with improved prognosis(17,59). By contrast, PG4's immune-depleted, MYC-driven phenotype resembles high-risk ecosystems characterized by exhaustion, antigen presentation failure, and immune paralysis(14,61).

These insights suggest several translational implications. First, immune-modulating therapies, such as interferon agonists, STING agonists, or epigenetic agents capable of restoring antigen presentation, may hold promise for PG4 patients by reactivating dormant immune programs(64,84). Second, metabolic modulators or MYC-targeting approaches may help disrupt the proliferative immune-suppressive axis that defines PG4 biology(79,80). Third, the profound differences between PG4 and PG6 emphasize the potential utility of proteogenotype-based stratification in future clinical trials, as it captures immune and transcriptional states that cannot be identified through genomic profiling alone(17).

Taken together, the integration of all omic layers reveals PG4 and PG6 not as genetically distinct diseases, but as regulatory immune ecosystems, with MYC activity, immune activation, and spatial architecture acting as the central axes of clinical behaviour. These findings refine our understanding of DLBCL heterogeneity and highlight novel biological routes through which aggressive disease may emerge independently of genetic diversity(59,85).

6.7. Limitations and future directions

Despite the strengths of this multi-omic and spatial analysis, several limitations should be acknowledged to contextualize the findings and guide future work. These limitations arise from cohort size, methodological constraints, and the inherent

complexity of inferring biological mechanisms from static tissue snapshots and bulk profiling approaches(67,88).

A first limitation relates to the sample size available for spatial imaging and WES. While the number of PG4 and PG6 cases included allowed us to characterize robust and recurring biological patterns, expanding the dataset would increase statistical power and permit deeper exploration of intra-proteogenotype heterogeneity. Particularly for PG4, the degree to which MYC-driven biology and immune exclusion co-occur across broader patient cohorts warrants validation in larger, independent datasets(14,17).

Second, although mIF provides high-quality spatial resolution, it is limited by the number of markers that can be visualized simultaneously. Increasing the multiplexing depth or applying spatial transcriptomics technologies would enable finer characterization of immune-cell states, cell - cell communication, and stromal architecture. Such approaches could clarify whether additional microenvironmental components such as fibroblasts, endothelial cells, or myeloid-derived suppressor cells contribute to the PG4 immune-excluded phenotype(29,89).

A related limitation concerns the use of bulk RNA sequencing, which captures average gene expression across heterogeneous cell populations. Although computational deconvolution and spatial profiling mitigate this issue, they cannot fully resolve rare cell states or subtle transcriptional programs that may influence immune activation or suppression. Future studies integrating bulk transcriptomics with single-cell RNA sequencing would be highly informative, enabling dissection of tumour- immune interactions at cellular resolution and clarifying whether PG4 tumours harbour distinct malignant B-cell subsets with unique regulatory states(88,90).

Another limitation is the reliance on inferred functional pathways from transcriptional signatures. While gene-set enrichment analyses provide valuable insights, they do not directly measure protein activity, signalling cascades, or metabolic states. Integrating phosphor-proteomics, metabolomics, or chromatin accessibility data could help confirm the presence of MYC-driven metabolic reprogramming or identify epigenetic mechanisms responsible for the immune-silent state of PG4(79,91).

The WES analysis, while comprehensive, is also limited by its focus on coding regions. Regulatory elements, enhancers, and non-coding RNAs which are increasingly recognized as key modulators of MYC activation and immune signalling

are not captured by this method. Expanding future analyses to include whole-genome sequencing or ATAC-seq could uncover regulatory DNA elements that contribute to proteogenotype-specific biology(87,92).

Finally, the clinical implications of PG4 and PG6 remain to be fully elucidated in the context of modern immunotherapies, such as CAR-T cells and bispecific antibodies. Whether the immune-cold architecture of PG4 predicts poor response to immunotherapy or whether it can be overcome by targeted immune activation, remains an important open question. Prospective studies evaluating proteogenotype-based responses to immunotherapies could directly translate the findings of this thesis into clinical practice(10,11).

In summary, while this work provides a detailed multi-omic characterization of PG4 and PG6, future studies integrating deeper spatial resolution, single-cell approaches, and functional assays will be essential for fully understanding the mechanisms driving proteogenotype identity and for translating these insights into therapeutic strategies. These efforts will not only address current limitations but also have the potential to refine DLBCL classification and improve patient outcomes(59).

7. Conclusion

Diffuse large B-cell lymphoma remains a clinically and biologically heterogeneous disease in which traditional cell-of-origin classifications and genomic subtyping only partially capture the complexity underlying patient outcomes. This thesis contributes to a deeper understanding of DLBCL by integrating spatial, transcriptomic, proteogenomic, and genomic data within the recently proposed proteogenotype framework. Through this multi-layered analysis, the work delineates the mechanisms driving the profound biological and clinical divergence between the poorest-outcome proteogenotype (PG4) and the most immune-enriched and favourable subgroup (PG6).

Across all analytical layers examined, PG4 consistently emerged as a tumour state defined by immune failure, MYC-driven proliferative dominance, and spatial disorganization. Transcriptomic profiling revealed that PG4 harbours a coordinated suppression of immune-related gene programs, affecting interferon signalling, chemokine production, antigen presentation, and T-cell activation. Spatial imaging further demonstrated that these transcriptional defects translate into a stark depletion of nearly all immune lineages and a collapse of immune -tumour cell interactions, resulting in a profoundly immune-excluded microenvironment. Simultaneously, PG4 is characterized by strong activation of MYC target genes and metabolic programs associated with rapid cellular proliferation, suggesting a mechanistic link between oncogenic transcriptional activity and immune silencing. In contrast, PG6 displayed a highly immune-competent phenotype characterized by robust interferon signalling, abundant chemokine expression, strong antigen presentation, and well-organized spatial immune architecture. Immune cells in PG6 infiltrated tumour regions effectively, formed structured multicellular hubs, and maintained functional engagement with malignant B cells features that collectively support efficient immune surveillance. This immune-rich ecosystem provides a biological foundation for the superior clinical outcomes associated with PG6.

Importantly, whole-exome sequencing revealed no significant differences in mutational burden, copy-number alterations, or genomic instability between PG4 and PG6. The absence of PG-specific genetic lesions indicates that proteogenotype identity, and the associated clinical behaviour, emerge primarily from regulatory and microenvironmental programs rather than genomic alterations. This finding shifts

the interpretative focus away from mutational drivers and toward transcriptional, proteomic, and spatial determinants of disease biology.

By integrating all omic layers, this thesis demonstrates that the fundamental distinction between PG4 and PG6 lies not in their genomes, but in their ecosystems. PG6 represents an immune-activated, microenvironmentally cooperative, biologically favourable state, whereas PG4 embodies a MYC-driven, immune-suppressed state lacking the molecular and structural prerequisites for immune-mediated tumour control. These findings refine the current understanding of DLBCL heterogeneity and highlight the necessity of evaluating tumour biology through a multi-dimensional lens that extends beyond traditional genomics.

The implications of this work are substantial for both research and clinical practice. First, immune competence rather than mutational burden, emerges as a dominant determinant of prognosis across proteogenotypes. Second, PG4 may benefit from therapeutic approaches aimed at restoring antigen presentation, enhancing interferon signalling, or modulating MYC-driven metabolic states. Third, the PG framework provides a valuable stratification system for future clinical trials by capturing immune and regulatory phenotypes not detectable by genomics alone. Finally, the study underscores the importance of multi-omic and spatial approaches in future lymphoma research, as they reveal key biological dimensions that remain invisible to single-modality analyses.

In summary, this thesis demonstrates that DLBCL outcomes are shaped by the interplay between tumour-intrinsic programs and microenvironmental architecture. PG4 and PG6 exemplify the extremes of this spectrum, providing mechanistic insights into how proliferative pressure, immune suppression, and spatial organization converge to drive disease behaviour. By elucidating these multi-layered biological principles, this work advances the conceptual and practical understanding of DLBCL and lays the foundation for improved classification, risk stratification, and targeted therapeutic strategies.

8. Bibliography

1. Takahara T, Nakamura S, Tsuzuki T, Satou A. The Immunology of DLBCL. *Cancers*. 2023 Jan 29;15(3):835.
2. Sehn L, Salles G. Diffuse large B-cell lymphoma. *N Engl J Med*. 2021;384(9):842–58.
3. Swerdlow S, Campo E, Harris N, others. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. IARC; 2017.
4. Srivastava S, Koay E, Borowsky A, De Marzo A, Ghosh S, Wagner P, et al. Cancer overdiagnosis: a biological challenge and clinical dilemma. *ASCO Educ Book*. 2015;35:e449–53.
5. Weltgesundheitsorganisation. WHO classification of tumours of haematopoietic and lymphoid tissues. Revised 4th edition. Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, et al., editors. Lyon: International Agency for Research on Cancer; 2017. 585 p. (World Health Organization classification of tumours).
6. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000 Feb;403(6769):503–11.
7. Locke F, Ghobadi A, Jacobson C, others. Long-term safety and activity of axicabtagene ciloleucel in refractory large B-cell lymphoma (ZUMA-1). *Lancet Oncol*. 2019;20(1):31–42.
8. Coiffier B, Lepage E, Briere J, others. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346(4):235–42.
9. Philip P, Chauvin F, Armitage J, Bron D, Hagenbeek A, Biron P, et al. Parental high-dose therapy and autologous bone marrow transplantation after failure of conventional chemotherapy in aggressive non-Hodgkin lymphoma. *N Engl J Med*. 1995;333(23):1540–5.
10. Schuster S, Bishop M, Tam C, others. Tisagenlecleucel in adult relapsed or refractory diffuse large B-cell lymphoma. *N Engl J Med*. 2019;380(1):45–56.
11. Neelapu S, Locke F, Bartlett N, others. Axicabtagene ciloleucel CAR T-cell therapy in refractory large B-cell lymphoma. *N Engl J Med*. 2017;377(26):2531–44.
12. Schmitz R, Wright G, Huang D, others. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med*. 2018;378(15):1396–407.

13. Chapuy B, Stewart C, Dunford A, others. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med*. 2018;24(5):679–90.
14. Steen C, Luca B, Esfahani M, others. The landscape of tumor cell states and ecosystems in diffuse large B-cell lymphoma. *Cancer Cell*. 2021;39(10):1422–1437.e10.
15. Enßle JC, Häupl B, Qoku A, Wang B, Wright GW, Barrans S, et al. Pathogenesis of diffuse large B-cell lymphoma proteogenotypes. 2025;
16. Keane C, Vari F, Hertzberg M, others. Tumor infiltrating T cells in DLBCL: clinical and biological significance. *Blood*. 2015;126(19):2193–202.
17. Kotlov N, Bagaev A, Revuelta M, others. Clinical and biological subtypes of DLBCL defined by gene expression-based immune landscapes. *Cancer Discov*. 2021;11(6):1468–89.
18. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57–63.
19. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol*. 2018;18(1):35–45.
20. Newman A, Liu C, Green M, others. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7.
21. Rossi D, Di Rocco A, Spina V, others. Liquid biopsy in lymphoid malignancies. *Blood*. 2021;137(9):1113–26.
22. Ansell S, Minnema M, Johnson P, others. Nivolumab for relapsed/refractory DLBCL. *J Clin Oncol*. 2019;37(6):481–9.
23. Salas M, Bejarano L, Gut I, others. Integration of spatial and multi-omics data in cancer. *Nat Cancer*. 2022;3:1410–24.
24. Gerdes M, et al. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded tissue. *Nat Methods*. 2013;10(8):868–74.
25. Lin J, Izar B, Wang S, et al. Highly multiplexed imaging of human tissues and tumors using CODEX. *Nat Protoc*. 2023;18(1):76–106.
26. Tan W, et al. Multiplex immunohistochemistry/immunofluorescence (mIHC/IF) for immune profiling in cancer research and clinical practice. *Nat Rev Clin Oncol*. 2024;21(2):107–25.
27. Stack E, et al. Multiplexed immunofluorescence and digital image analysis for clinical biomarker studies. *Front Oncol*. 2023;13:1203154.
28. Gut G, Wörsdörfer P, Dey S. Advanced multiplexed imaging and cyclic immunofluorescence in spatial omics. *Nat Methods*. 2025;22(1):45–58.

29. Keren L, Bosse M, Thompson S. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell*. 2018;174(6):1373–87.
30. Yu J, et al. Spatial organization of immune microenvironment predicts prognosis and therapy response in diffuse large B-cell lymphoma. *Blood Cancer J*. 2024;14(3):81.
31. Li Z, et al. Immune spatial architecture as a biomarker in lymphoma immunotherapy. *J Immunother Cancer*. 2024;12(5):e009352.
32. Danaher P, et al. Multisite reproducibility of multiplex immunofluorescence assays: results from the MITRE study. *J Immunother Cancer*. 2023;11(2):e006875.
33. Zhang T, et al. Artificial intelligence-based analysis of multiplex immunofluorescence images enhances predictive accuracy in lymphoma. *Lab Invest*. 2024;104(9):1012–25.
34. Ng K, et al. Standardization of multiplex immunofluorescence workflows for translational research. *Cancers Basel*. 2025;17(2):215.
35. Schueffler PJ, Fuchs TJ. Artificial intelligence in digital pathology: unlocking the potential of multiplex imaging. *Nat Rev Cancer*. 2021;21(9):509–21.
36. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2).
37. Van der Auwera GA, O'Connor BD. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media; 2020.
38. Hanssen F, Garcia MU, Folkersen L, Pedersen AS, Lescai F, Jodoin S, et al. Scalable and efficient DNA sequencing analysis on different compute infrastructures aiding variant discovery. *NAR Genomics Bioinforma*. 2024;6(2).
39. Garcia M, Juhos S, Larsson M, others. Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. *F1000Research*. 2020;9:63.
40. Wiggeshoff J, Welter V, Rehker J, others. Abstracts. *Molecular Oncology*. 2025.
41. Jia P, Yang X, Guo L, Liu B, Lin J, Liang H, et al. MSIsensor-Pro: Fast, Accurate, and Matched-Normal-Sample-Free Detection of Microsatellite Instability. *Genomics Proteomics Bioinformatics*. 2020;18(1):65–71.
42. Riester M, Singh A, Brannon A, others. PureCN: Copy number calling and SNV classification using targeted short read sequencing. *Source Code Biol Med*. 2016;11:13.
43. Oh S, Geistlinger L, Ramos M, others. Reliable analysis of clinical tumor-only whole exome sequencing data. *JCO Clin Cancer Inform*. 2020;

44. Chang X. Accelerating discovery of functional mutant alleles in cancer. *Cancer Discov.* 2017;
45. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32(18):2847–9.
46. Ewels P, Peltzer A, Fillinger S, others. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol.* 2020;38(3):276–8.
47. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Internet]. 2010. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
48. Krueger F. Trim Galore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files [Internet]. 2015. Available from: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
49. Broad Institute. Picard toolkit [Internet]. 2019. Available from: <http://broadinstitute.github.io/picard/>
50. Dobin A, Davis CA, Schlesinger F, others. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21.
51. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017;14(4):417–9.
52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
53. Ge SX, Son EW, Yao R. iDEP: An integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics.* 2018;19(1):534.
54. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol.* 1995;57(1):289–300.
55. Newman AM, Steen CB, Liu CL, others. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol.* 2019;37(7):773–82.
56. Stringer C, Wang T, Michaelos M, others. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods.* 2021;18:100–6.
57. Scott DW, Gascoyne RD. The biology of diffuse large B-cell lymphoma. *Blood.* 2014;123:171–81.
58. Wright GW, others. A probabilistic classification tool for genetic subtypes of diffuse large B cell lymphoma. *Nat Rev Clin Oncol.* 2020;

59. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours. *Nat Rev Cancer*. 2019;19:197–212.
60. Thorsson V, others. The immune landscape of cancer. *Immunity*. 2018;48:812–30.
61. Joyce JA, Fearon DT. T cell exclusion, immune privilege, and the tumor microenvironment. *Science*. 2015;348.
62. Carstens JL, others. Spatial computation of intratumoral T cells correlates with survival of patients with pancreatic cancer. *Nat Commun*. 2017;8.
63. Casey SC, others. MYC regulates the antitumor immune response through CD47 and PD-L1. *Cell*. 2016;166:404–17.
64. Topper MJ, others. Epigenetic therapy ties MYC depletion to reversing immune evasion and treating lung cancer. *Nat Med*. 2017;23:1119–27.
65. Reddy A, others. Genetic and functional drivers of diffuse large B cell lymphoma. *Cell*. 2017;171:481–94.
66. Giesen C, Wang HAO, Schapiro D, others. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods*. 2014;11:417–22.
67. Bodenmiller B. Multiplexed epitope-based tissue imaging for discovery and healthcare applications. *Nat Methods*. 2016;13:997–1008.
68. Sch"urch CM, Bhate SS, Barlow GL, others. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell*. 2020;182:1341–59.
69. Dieu-Nosjean MC, Goc J, Giraldo NA, others. Tertiary lymphoid structures in cancer and beyond. *Nat Rev Cancer*. 2014;14:367–82.
70. Cabrita R, Lauss M, Sanna A, others. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature*. 2020;577:561–5.
71. Helmink BA, Reddy ST, Gao J, others. B cells and tertiary lymphoid structures promote immunotherapy response. *Nature*. 2020;577:549–55.
72. Cassetta L, Pollard JW. Targeting macrophages: therapeutic approaches in cancer. *Nat Rev Drug Discov*. 2018;17:887–904.
73. Ayers M, Lunceford J, Nebozhyn M, others. IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest*. 2017;127:2930–40.
74. Steidl C, Shah SP, Woolcock BW, others. MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers. *N Engl J Med*. 2010;362:209–20.
75. Dang CV. MYC on the path to cancer. *Cell*. 2012;149:22–35.

76. Meyer N, Penn LZ. Reflecting on 25 years with MYC. *Nat Rev Cancer*. 2008;8:976–90.
77. Subramanian A, Tamayo P, Mootha VK, others. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102:15545–50.
78. Kortlever RM, Sodir NM, Wilson CH, others. MYC cooperates with Ras by programming inflammation and immune suppression. *Cell*. 2017;171:1301–15.
79. Buck MD, Sowell RT, Kaech SM, Pearce EL. Metabolic instruction of immunity. *Cell*. 2017;169:570–86.
80. Chang CH, Qiu J, O'Sullivan D, others. Metabolic competition in the tumor microenvironment is a driver of cancer progression. *Cell*. 2015;162:1229–41.
81. Victora GD, Nussenzweig MC. Germinal center dynamics revealed by multiphoton microscopy. *Cell*. 2012;149:1120–34.
82. Calado DP, Sasaki Y, Godinho SA, others. The cell-cycle regulator c-Myc is essential for the formation and maintenance of germinal centers. *Nat Immunol*. 2012;13:1092–100.
83. Lovén J, Hoke HA, Lin CY, others. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013;153:320–34.
84. Spranger S, Bao R, Gajewski TF. Mechanism of tumor rejection with doublets of CTLA-4, PD-1 blockade involves T cell recruitment. *Nature*. 2015;523:231–5.
85. Topalian SL, Taube JM, Anders RA, Pardoll DM. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat Rev Cancer*. 2016;16:275–87.
86. Chan TJ, Lenz G, Gascoyne RD, others. Tumor mutation burden is not associated with immune infiltration or survival in diffuse large B-cell lymphoma. *Blood*. 2019;134:207–19.
87. Hnisz D, Abraham BJ, Lee TI, others. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155:934–47.
88. Stuart T, Satija R. Comprehensive integration of single-cell data. *Nat Rev Genet*. 2019;20:257–72.
89. Ståhl PL, Salmén F, Vickovic S, others. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*. 2016;353:78–82.
90. Tirosh I, Izar B, Prakadan SM, others. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016;352:189–96.
91. Mertins P, Mani DR, Ruggles KV, others. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*. 2016;534:55–62.

92. Corces MR, Granja JM, Shams S, others. The chromatin accessibility landscape of primary human cancers. *Nat Genet.* 2018;50:1211–24.

10. Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Univ.- Prof. Dr. med. Reinhard Büttner betreut worden.

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Veränderungen, dem Promotionsausschuss unverzüglich mitzuteilen.

17.04.2026

Datum Unterschrift

Marta Pistone