UNIVERSITY OF URBINO CARLO BO

Department of Biomolecular Sciences

Ph.D. Course in Biomolecular and Health Sciences

XXXVI cycle

FLOW CYTOMETRIC AND ARTIFICIAL INTELLIGENCE APPROACH TO DIAGNOSTIC MARKERS FOR B-CELL LYMPHOMAS

SSD: BIO/17

Coordinator: Prof. Marco Bruno Luigi Rocchi

Supervisor: Prof. Stefano Papa

Co-Supervisor: Dr. Massimo Geuna

Ph.D. student: Dr. Elena Casanova

ACADEMIC YEAR 2022-2023

In loving memory of Letizia,

my travelling companion in Biology studies

and now my guardian angel.

INDEX

ABSTRACT

Mature B-cell lymphoid neoplasms are a wide and highly heterogeneous group of malignancies. Different morphology, phenotype, genotype, aggressiveness and response to therapy characterize each of these entities and individual cases within the same entity. Immunophenotypic characterization is a key element for their classification in order to guide to the correct therapeutic plan. Although the phenotypic study is routinely performed by immunohistochemistry (IHC), we verified that the use of flow cytometry (FC) could bring several advantages. Indeed, immunophenotype reveals the expression of genes (revealed by IHC) through the detection of cell surface and/or intracellular proteins (identified by FC) that are the final expression of these genes. Consequently, cytometric analysis can provide a wide range of information comparable in value to gene sequencing. Applied to samples from various sources (peripheral blood, bone marrow aspirates, lymph node and tissue biopsies, fine needle aspiration samples, and pleural and peritoneal effusions) from patients affected by the most common mature B-cell Non-Hodgkin lymphomas (B-NHL), FC allows the quantitative expression of multiple markers to be evaluated by analyzing millions of cells simultaneously and easily defining clonal populations. This results in the ability to almost perfectly isolate each neoplastic subpopulation and study it in its uniqueness. This feature distinguishes FC analysis from IHC, which, by presenting the global appearance of the sample being examined, preserves the architecture of the tissue while revealing all mixed clones. In addition, like IHC, FC can provide information about intracellular antigens. However, FC can be widely applied to fine needle aspiration (FNA) specimens and is more sensitive and specific than IHC. In addition, FC can be applied to peripheral blood (PB) samples of leukemia stage lymphoma. In any case, FC provides faster and less biased results than IHC because the data are expressed quantitatively.

However, the large amount of data generated by FC is very complex to analyze as a whole. In particular, it is difficult to correlate each marker with the precise diagnosis of a disease. Artificial intelligence (AI) can help by using sophisticated software and computing systems to compare and stratify all this data in a short time. In this way, AI provides more manageable data with a two-dimensional representation that is easier to interpret.

In a previous study, which, according to the authors, we have decided to reproduce in some parts for the sake of clarity, since it prepares the data presented in this work, we applied machine learning (ML) algorithms to a large dataset of B-NHL immunophenotypes to generate a robust and clinically applicable prediction system. This system would also allow us to overcome the time-consuming, optimize the use of antibodies and standardize a clinically applicable predictive system by establishing a panel of antibodies to be systematically used in a multiparametric immunophenotypic analysis of samples performed with a high-complexity flow cytometer..

We then applied additional intracellular markers to a homogeneous case series of 615 tissue samples whose diagnoses, all confirmed by histologic analysis, were grouped into 8 major categories of B-NHL patients. The Predictive Power Score (ppscore) method allowed us to assess the impact of each marker in defining each lymphoma category. Considering that a ppscore greater than 0.22 (the baseline score) is statistically significant for discriminating diagnostic categories, we surprisingly noticed the discriminatory power of intracellular

markers not commonly used in a multiparametric immunophenotypic approach to lymphoma diagnosis, such as IRF4 and Bcl6.The role of these markers was validated by combining each one with all the others in a classification tree, resulting in a structural relationship tree that separates the entire database into quasi-homogeneous groups of lymphomas. Finally, the Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction technique divides the 8 categories of lymphomas into clusters. Interestingly, the application of UMAP to the 10 markers evaluated as having the greatest impact on diagnosis results in a greater separation of clusters, but does not recognize different lymphoma entities. Exploring the expression of all markers used in the phenotypic panel, where all antibodies contribute to the UMAP representation independently of their statistical significance, the 8 groups of B-NHL result clearly defined and separated into clusters.

In conclusion, it is conceivable that the implementation of artificial intelligence applied to multiparametric flow cytometry could significantly contribute to an optimal diagnostic process where histopathological examination still remains the gold standard.

It remains to be investigated whether the use of these methods with a large number of markers can also predict categories of molecular or genetic alterations, which would be useful for a better classification even for therapeutic purposes of B-NHL.

INTRODUCTION

1.1 Impact of B-cell non-Hodgkin lymphoma on global health

B-cell Non-Hodgkin lymphomas (B-NHL) are one of the major health problems worldwide. The age-standardized rate (ASR) of B-NHL incidence showed an increasing trend during the last 30 years, in both sexes and in most geographic regions, even if globally, the incidence of different types of B-NHL varies considerably across the world. However, death and disability-adjusted life years (DALYs) caused by B-NHL showed decreasing trends globally, but this category of neoplasms remains a substantial challenge (Cai et al, 2021). The increase in overall survival rates of B-NHL was largely ascribed to the progress in several treatment studies and the application of the relevant research achievements, which included monoclonal antibody (mAbs), small molecule inhibitors, and the targeted chimeric antigen receptor T cells (CAR-T).

1.2 Mature B cell lymphoid neoplasms heterogeneity in WHO classifications

Mature B-cell lymphoid neoplasms encompass a heterogeneous and extensive group of malignancies arising from B cells at various stages of differentiation and maturation. These neoplasms can exhibit significant variability in their morphologic, phenotypic, and genotypic features, as well as differences in their clinical behavior and response to treatment. The heterogeneity of these lymphoid neoplasms makes accurate diagnosis and classification, crucial for determining the appropriate treatment strategy. Various techniques, such as immunohistochemistry, flow cytometry, and genetic testing, are employed to identify specific markers and genetic alterations associated with each subtype. Due to the distinct biological characteristics and clinical behavior of each entity, treatment approaches can differ significantly. Some lymphomas may be more responsive to chemotherapy, while others may benefit from targeted therapies, immunotherapy, or stem cell transplantation. As research advances, more tailored and precise treatments continue to emerge for these diverse entities, improving overall patient outcomes.

The World Health Organization (WHO) classification of lymphoid tumors has provided a global reference for the diagnosis of lymphoid neoplasms since its 3rd edition in 2001, which was based on the Revised European American Lymphoma (R.E.A.L) Classification, developed by the International Lymphoma Study Group (ILSG) in the early 1990s (Jaffe et al, 2001). The definitions established in successive editions of the WHO classifications have been adopted not only in clinical, as in basic and translational research, but have also been incorporated into the International Classification of Diseases (ICD) codes, thus serving as a global reference for epidemiological monitoring in all national and international health policy organizations.

From the 4th revised edition (Swerdlow et al, 2017; Jiang et al, 2017), it is possible to deduce what the differentiation between the most common types of mature B-cell lymphoid neoplasms may be. The major categories of B-NHL are listed below:

- **Diffuse Large B-cell Lymphoma (DLBCL)** is the most common type of B-NHL, and it is characterized by a diverse morphology and clinical behavior. DLBCL can be further classified into distinct molecular subtypes, each with its own prognosis and response to therapy. Despite recent therapeutic advances (Chiappella et al, 2017), up to 50% of patients relapse after standard chemoimmunotherapy R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine and prednisolone). Predictive biologic and gene expression markers remain undefined, but their identification is becoming increasingly urgent, as exciting research results on new therapeutic agents lead to hope for new monoclonal antibodies conjugated to cytotoxic drugs and bispecific antibodies that may lead to promising results for relapsing/refractory patients (Papageorgiou et al, 2022). Some of the characteristic markers used to classify DLBCL subtypes include:

    - Cell of origin markers:
        - Germinal center B-cell (GCB) subtype: CD10, BCL6 and MUM1/IRF4 (also known as MUM1/CD40).
        - Activated B cell (ABC) subtype: CD10 negative, BCL6 negative and MUM1/IRF4 positive.
        - MYC, BCL2 and BCL6 are often evaluated for rearrangements or gene amplifications:
            - Double-hit lymphoma: presence of concomitant rearrangements or gene amplifications of MYC and BCL2 or MYC and BCL6.
            - Triple-hit lymphoma: presence of concomitant rearrangements or gene amplifications of MYC, BCL2 and BCL6.
    - CD5 expression: CD5-positive DLBCL is associated with a worse prognosis.
    - EBV (Epstein-Barr virus) status: Presence of EBV in tumor cells, which may influence prognosis.

- **Follicular lymphoma (FL)** is the second most common type of B-NHL in western countries, it is typically indolent, slow-growing, and has a characteristic nodular growth pattern. It is associated with specific genetic abnormalities: these include the t(14;18) translocation, which is the hallmark of FL, and leads to overexpression of BCL2, which contributes to lymphoma cell survival by preventing apoptosis. Also, the cytomorphology is characteristic: the tumor is composed of follicle center cells, usually a mixture of centrocytes (small cells with cleaved nuclei) and centroblasts (large noncleaved nuclei). Different markers are commonly used for the diagnosis and classification of FL:

    - CD20 molecule is universally expressed by normal B cells in all stages of development, from the pre-B cell up to the mature plasma cell as well as by most B cell malignancies (Staschenko et al, 1980), and it is almost always expressed on FL cells. CD20-targeted therapies, such as Rituximab or Obinutuzumab (Freeman et al, 2018), have been highly effective in treating

and have become an integral part of the treatment approach in combination with chemotherapy for this disease.
    - CD10 is commonly found on the surface of normal germinal center B-cells and it is also expressed in FL cells, thus it is helpful in distinguishing FL from other types of lymphoma.
    - CD5 and CD23 are usually negative in FL and can help differentiate it from other types of B-NHL, such as chronic lymphocytic leukemia (CLL).
    - BCL6 appears overexpressed in FL: this dysregulation causes the uncontrolled growth of abnormal B cells and hinders the normal process of apoptosis. Understanding the role of the BCL6 gene in FL has been essential for the development of targeted therapies that aim to inhibit or block the activity of this gene.

  Common initial treatments are bendamustine (Treanda) plus rituximab and R-CHOP.

- **Chronic Lymphocytic Leukemia (CLL) and Small Lymphocytic Lymphoma (SLL)** are essentially the same disease, but they differ in their clinical presentation. CLL primarily presents with an increased number of abnormal B cells in the blood and bone marrow, whereas SLL presents with lymph node enlargement. The diagnosis of CLL requires $\geq 5 \times 10^3/\mu L$ of circulating monoclonal B lymphocytes with a CLL immunophenotype in the peripheral blood (PB). The term SLL refers to cases with a circulating CLL cell count $< 5 \times 10^3/\mu L$ and recognized lymph node, splenic, or extramedullary involvement. It is important to consider that CLL and SLL are complex diseases with various subtypes and clinical presentations. Evaluation of factors such as clinical symptoms, blood counts, and imaging contribute to accurate diagnosis, risk stratification, and treatment planning. Treatment strategies for CLL/SLL may include watchful waiting, targeted therapies (e.g. monoclonal antibodies and/or BCR inhibitors), chemotherapy (i.e. fludarabine, cyclophosphamide, bendamustine), and sometimes stem cell transplantation, depending on the specific case and disease progression. There are specific characteristic markers associated with these conditions, which are used to aid in diagnosis and management:

    - CD5 is a cell surface marker typically found on normal mature T cells, but is also expressed on abnormal lymphocytes in CLL/SLL, distinguishing it from other types of lymphoma.
    - CD19 is a cell surface protein that is typically present on B cells, including CLL/SLL cells. However, in CLL/SLL, the expression of CD19 is usually maintained, and this characteristic can be exploited for targeted therapies using monoclonal antibodies directed against CD19, such as Tafasitamab (Boxhammer et al, 2019).
    - CD20 is an important therapeutic target in CLL/SLL (Freeman et al, 2018), nevertheless its expression level is usually low. Some instance may account for a certain degree of heterogeneity in CD20 expression:

        - Clonal Heterogeneity: CLL/SLL is a heterogeneous disease, meaning that cancer cells within an individual patient can differ in their characteristics. Some cells may express higher levels of CD20, while others may have lower or even undetectable levels (Tam et al, 2008).
        - Clonal Evolution: over time, CLL/SLL cells can undergo changes and acquire new genetic mutations, leading to variations in CD20 expression levels (Zhang et al, 2014).
        - Microenvironment Interactions: the interaction between CLL/SLL cells and the surrounding microenvironment can affect the expression of CD20 (Dubois et al, 2023).
    - CD23 expression is important in CLL/SLL as it aids in the accurate diagnosis and differentiation between these two diseases and other B-NHL subtypes, which can have implications for treatment decisions and prognosis.
        - Differentiating CLL/SLL from Mantle Cell Lymphoma (MCL): MCL and CLL/SLL share CD5 expression, CD23 is usually absent in MCL, while it is almost always present in CLL/SLL, rarely dimly positive and in the most of cases brightly positive. However, with dimly positive expression, interpretation should be cautious (Gong et al, 2001; Barna et al, 2008; Yoshino et al, 2020)
        - Therapeutic considerations: in the context of targeted therapies, certain treatments, such as monoclonal antibodies like lumiliximab (Schnaiter et al, 2010), target CD23 and can be used as part of the treatment regimen for CLL.
    - Immunoglobulin heavy chain variable region (IgVH): Genetic analysis of the immunoglobulin heavy chain variable region gene is essential for the diagnosis and prognosis of CLL/SLL. The mutational status of immunoglobulin genes can influence disease progression and patient outcome (Lee et al, 2021).
    - ZAP-70 (Zeta-Associated Protein 70) is a key signaling molecule in lymphocytes: its expression is associated with more aggressive forms of CLL/SLL (Schroers et al, 2005).
    - CD38 is also a marker associated with a more advanced and aggressive form of CLL/SLL that predicts poorer outcomes (Schroers et al, 2005).
    - Disruption of the TP53 gene, either by deletion at chromosome 17p13.1 (del17p) or mutations, is the most important prognostic/predictive biomarker in CLL (Bomben et al, 2023).

- **Mantle Cell Lymphoma (MCL)** is characterized by the overexpression of cyclin D1 due to the t(11;14) translocation, which involves the fusion of the CCND1 (cyclin D1) gene on chromosome 11 and the immunoglobulin heavy chain gene on chromosome 14. As a result of this translocation, cyclin D1 is overproduced, leading to uncontrolled cell growth and proliferation. Thus MCL typically presents with an aggressive clinical course, it tends to grow and spread rapidly. It often involves multiple lymph nodes, bone marrow, and other organs at the time of diagnosis.

However its incidence is relatively rare, accounting for approximately 6% of all B-NHL. MCL is more commonly diagnosed in older adults, with a median age of around 60 years, and it is also more prevalent in men than women. Treatment for MCL usually involves a combination of chemotherapy, targeted therapy, immunotherapy, and stem cell transplantation, depending on the stage and aggressiveness of the disease (Kumar et al, 2022). Despite initial responses to treatment, MCL often relapses after remission. The prognosis for MCL varies depending on several factors, including the patient's age, overall health status, the stage of the disease at diagnosis, and specific biological features of the lymphoma cells. The international prognostic index (IPI) and the Mantle Cell Lymphoma International Prognostic Index (MIPI) are some of the tools used to assess prognosis. Flow cytometry is an essential diagnostic tool used in the evaluation of Mantle Cell Lymphoma (MCL). The analysis of specific markers expressed on the surface of cells helps identify and differentiate MCL from other lymphomas. Some of the key markers are used in flow cytometry to diagnose MCL:

- Almost all cases of MCL show overexpression of Cyclin D1.
- The classic immunophenotype is strongly positive for pan-B cell antigens CD5, CD19, CD43, positive for FMC7, and negative for CD10, CD23, and Bcl-6 (Inamdar et al, 2016).
- Also CD20, CD22 and CD79b are expressed and they are essential in confirming the B-cell lineage of the lymphoma.
- CD38 and CD200 can aid in differentiating the usual aggressive MCL variants ($CD38^+/CD200^-$) and rare, non-nodal cases with monoclonal asymptomatic lymphocytosis, cyclin D1–positive (MALD1), which are $CD38^-/CD200^+$ (Espinet et al, 2014).

Burkitt lymphoma (BL) is a highly aggressive type of B-cell cancer that can affect both children and adults. Research on BL has uncovered the first recurrent chromosomal abnormality in cancer, known as t(8;14)(q24;q32), and subsequently revealed the pivotal role of the MYC gene and Epstein-Barr virus (EBV) in the development of this malignancy. Most patients with BL are treated with chemoimmunotherapy, such as R-CHOP, R-CODOX-M (rituximab, cyclophosphamide, vincristine, doxorubicin, and methotrexate), R-IVAC (rituximab, ifosfamide, etoposide, cytarabine), DA-EPOCH-R (rituximab with dose adjusted etoposide, prednisolone, vincristine, cyclophosphamide and doxorubicin). However, patients with relapsed or refractory disease usually die of lymphoma (López et al, 2022). BL usually exhibits a characteristic immunophenotype:

- BL cells are positive for pan-B-cell antigens CD19, CD22 and CD79a, with a monoclonal light chain expression of kappa or lambda.
- CD10, CD38, CD43, CD71 and Bcl-6 are also expressed.
- The blast markers CD34 and TdT are negative.
- SPF are very high (>30%) as is the proliferation marker Ki-67 (100%)

- **Marginal Zone Lymphomas (MZL)** can be subdivided into three categories: extranodal (MALT), nodal (NMZL), and splenic marginal zone lymphomas (SMZL). The frequency varies by category: MALT lymphoma is approximately 7.5% of B-NHL, NMZL is less than 2%, and SMZL is less than 1%. They are often associated with chronic inflammatory conditions and can behave indolently or aggressively depending on the subtype. Common initial treatments are bendamustine (Treanda) plus rituximab and R-CHOP. In cases without adequate morphology, immunophenotypic characterization by flow cytometry (FC) relies heavily on exclusion of other low-grade lymphomas (Arcaini et al, 2003; Kost et al, 2008):

    - MZL expresses typical pan-B-cell antigens CD19, CD20, and CD79a.
    - CD5, CD10, CD23, and CD43 are not expressed.
    - Expression of CD11c is highly associated with SMZL.
    - Levels of CD19 expression in conjunction with CD11c and CD38 expression can distinguish MZL from CD10 negative FL.

- **Lymphoplasmacytic lymphoma (LPL)** is a rare type of low-grade or indolent B-NHL, composed of small lymphocytes, plasmacytoid lymphocytes, and plasma cells that typically involve the bone marrow. The 2008 World Health Organization (WHO) guidelines for the classification of blood disorders define Waldenström's macroglobulinemia (WM) as a subset of LPL that has a measurable presence of monoclonal immunoglobulin (Ig) M gammopathy, with bone marrow involvement by LPL. The abnormal presence of monoclonal IgM can lead to an increase in blood viscosity in up to 30% of patients. Symptoms can vary considerably among individual patients, and many patients are asymptomatic at diagnosis (Naderi et Yang, 2013). The typical immunophenotype of LPL demonstrates:

    - expression of CD19, CD20, CD22, FMC7, BCL2, CD38, and CD79a with monotypic surface light chain.
    - CD5, CD10, and CD23 are usually absent (Konoplev et al, 2005). However, up to 20% of cases may express CD5, CD10, or CD23 (Hunter et al, 2005).
    - Immunoglobulin light-chain restriction can usually be demonstrated in both the small lymphocytes and plasma cells on tissue sections.
    - The plasma cells in LPL have been reported to have a unique immunophenotype, when compared with plasma cells found in marginal-zone lymphoma or plasma cell myeloma, with coexpression of PAX5 and CD45 with CD19, respectively (Morice et al, 2009).

  Treatments for LPL generally include watchful waiting; combinations of chemotherapy drugs that may be used include DRC (dexamethasone, rituximab and cyclophosphamide), BRD (bortezomib and rituximab, with or without dexamethasone), CVP (cyclophosphamide, vincristine and prednisone, with or without rituximab), thalidomide with rituximab; targeted therapy drugs used alone or

in combination with rituximab, bortezomib and ibrutinib; stem cell transplant; radiotherapy.

- **High-grade B-cell Lymphoma (HGBL)** is a group of aggressive lymphomas with multiple subtypes, which can present significant challenges in diagnosis and treatment. This category includes tumors with Burkitt-like or blastoid morphology that do not have double-hit cytogenetics and that cannot be classified as other well-defined lymphoma subtypes. Most of them have germinal center B-cell phenotype, and up to 45% carry a single-hit MYC rearrangement, but otherwise, they have no unifying immunophenotypic or cytogenetic characteristics (Olszewski et al, 2022). HGBL includes also Double-hit Lymphoma (DHL) and Triple-hit Lymphoma (THL), which are aggressive subtypes of DLBCL with concurrent rearrangements of MYC and BCL2 and/or BCL6 genes (Kim et al, 2020), and encompasses other aggressive B-cell lymphomas with similar genetic alterations, leading to poor prognosis and resistance to standard treatments. Treatment approaches for high-grade B-NHL typically involve intensive chemotherapy regimens, sometimes combined with targeted therapies and immunotherapy (most often R-CHOP), depending on the subtype and individual patient factors. Novel therapies, including chimeric antigen receptor (CAR) T-cell therapy, have shown promising results in relapsed or refractory cases (Denlinger et al, 2022). There are currently no established immunophenotypic criteria for the examination of lymphomas with MYC rearrangements. Immunophenotyping can be performed by standard methods using a panel of monoclonal antibodies against CD20, CD38, CD10, CD43, CD27, FMC-7, Ki-67, bcl-2, CD79b, CD23, CD22, and surface immunoglobulins (IgM, IgD, IgG, and IgA) to identify antigen expression in the tumor cell population (Tsagarakis et al, 2020).

The last edition is WHO-HAEM5, the drafting of which began in 2018 and was completed in 2022. It applies a hierarchical system for the classification of hematologic malignancies, organizing diseases in order of increasing levels of specification: category (e.g., mature B-cell), family/class (e.g., large B-cell lymphomas), entity/type (e.g., diffuse large B-cell lymphoma, not otherwise specified), and subtype (e.g, diffuse large B-cell lymphoma, not otherwise specified, germinal center B-cell-like) (Alaggio et al, 2022).

1.3 Immunophenotypic characterization of mature B cell lymphoid neoplasms

Immunophenotypic characterization is a key element for B-NHL classification, in order to direct to the correct therapeutic plan. Although the phenotypic study is routinely performed by immunohistochemistry (IHC), we verified that the use of flow cytometry (FC) could bring several advantages (van Oostrum et al, 2019). Indeed, the immunophenotype reveals expression of genes (unveiled by IHC) through detection of cell surface and/or intracellular proteins (identified by FC) which are the final expression of these genes. Consequently, cytometric analysis can provide a wide range of information, comparable in value to gene sequencing. Applied to samples derived from different sources (peripheral blood, bone

marrow aspirates, lymph node and tissue biopsies, fine needle aspiration samples and pleural and peritoneal effusions) of patients affected by the most common mature B-NHL, FC allows to evaluate the quantitative expression of several markers by analyzing millions of cells simultaneously, and easily defining clonal populations. This results in the ability to isolate almost perfectly each neoplastic subpopulation, examining it in its uniqueness. This feature differentiates FC analysis from IHC, which, by presenting the global appearance of the examined sample, preserves the architecture of the tissue while showing all the mixed clones. In addition, like IHC, FC can provide information on intracellular antigens. However, FC can be widely used in fine needle aspiration (FNA) samples, being more sensitive and specific than IHC (Barrena et al, 2011; Demurtas et al, 2010). Moreover, FC can be applied on peripheral blood (PB) samples of leukemia-stage lymphomas. In any case, FC provides faster and less biased results than IHC because the data are expressed quantitatively.

In the diagnostic process of certain diseases, it is already known the crucial role of FC, as in the demonstration of the typical CLL immunophenotype. CLL typically exhibits a characteristic immunophenotype, expressing CD5, CD19, CD20 ^{dim}, CD22 ^{dim}, CD23 , CD43 ^{bright}, CD45 ^{dim}, CD79b $^{dim/neg}$, CD81 ^{dim}, CD200, and surface monoclonal immunoglobulin (Ig) ^{dim}, but is negative for CD10, CD103, and CD123 and other T-cell and myeloid antigens. The primary distinction to consider when diagnosing CLL/SLL involves distinguishing it from conditions such as monoclonal B-cell lymphocytosis (MBL) and MCL. MBL is the presence of $<5\times10^3/\mu L$ of circulating monoclonal B lymphocytes in the absence of associated lymphadenopathy, organomegaly, extramedullary involvement, or other features of B-cell lymphoproliferative disorders. CLL-type MBL (75% of MBL cases) has an immunophenotype identical to that of CLL, but the expression of CD200 and CD23 in combination with CD20, CD22, CD45, CD79b and CD81 ^{bright} and CD43 differentiates CLL from MCL. Rarely, CLL cases may show an atypical immunophenotype, such as the absence of CD5 and CD23, normal intensity of CD20, CD22, and CD79b, aberrant expression of cell antigens, and atypical immunoreactivity (Salem et al, 2019).

But definitively, FC is a rapid and cost-effective technique for evaluating the expression of many lymphoid markers in all mature B-cell neoplasms, including DLBCL, the most common B-NHL. In a recent study, a gene expression-based risk score was constructed based on the expression levels of BCL2, BCL6, CD11c and LAIR1 to predict the outcome of patients with DLBCL (Devin et al, 2019).

The typical immunophenotype of FL, identified by FC, by analyzing a lymph node FNA sample, such as a bone marrow aspirate (BM), is $CD19^{+\,(dim)}$, $CD20^{+\,(bright)}$, $CD10^{+\,(uniform)}$, $CD5^{neg}$, $CD23^{neg}$, $CD200^{\,neg}$, $CD11c^{\,neg}$, with surface expression of restricted immunoglobulin κ-chain or λ-chain (Khanlari et al, 2022). Evidence on FC of monotypic expression of light chains, uniform expression of CD10, and decreased intensity of CD19, CD20, and CD38 are other features that support the diagnosis of FL over reactive follicular hyperplasia (Mantei et al, 2009).

The high diagnostic accuracy and efficacy of FC in BL was also demonstrated. CD16/CD56 expression without $CD38^{higher}$ and the absence of CD16/CD56 with $CD38^{higher}$ expression proved to be reliable, fast and cost-effective methods for the diagnosis of 11q aberration and MYC rearrangements, respectively, in $CD10^{+}$ aggressive lymphomas (Rymkiewicz et al, 2018).

However, it has been shown that FC immunophenotyping can play an important role in the detection of aggressive MYC rearranged B-cell lymphomas (Tsagarakis et al, 2020). However, the detection of MYC rearranged B-cell lymphomas may pose a diagnostic challenge for flow cytometry specialists. It has been proposed that the expression of $CD38^{bright}$ and $CD45^{low}$ is highly specific for MYC-arranged HGBLs compared with non-MYC-arranged DLBCLs, with a combined sensitivity of 67% and specificity of approximately 100%. This information could guide the efforts of the flow cytometrist in the second diagnostic step, namely the distinction between BL, DHL, and MYC-DLBCL. Previous studies have shown that there is significant overlap between BL, DHL, and MYC-DLBCL in terms of expression of CD10, CD19, CD20, CD38, and CD45 (Mandelker et al., 2014). Furthermore, decreased staining of CD20 and/or bright staining of CD38 have been described in association with LDL (Alsuwaidan et al., 2019; Wu et al., 2010). However, they cannot be used as biomarkers for early diagnosis of DHL because of their moderate sensitivity.

Ultimately, the correct subclassification of B-cell lymphomas requires the integration of histologic and IHC findings, as well as the results of FC immunophenotyping, cytogenetic analysis, and molecular studies. For example, in the context of MCL, molecular measurement of minimal residual disease (MRD) is typically performed on peripheral blood or bone marrow using methods such as high-sensitivity FC, allele-specific oligonucleotide polymerase chain reaction (PCR), or next-generation sequencing (NGS)-based IGH clonal rearrangement assay (Sethi et al, 2021). In other words, the integration of ancillary studies and biopsy results from other disease sites, in addition to careful histologic review and an expanded antibody panel, are essential to obtain a correct final classification.

1.4 Cluster analysis to get important biological informations

The large amount of data obtained by FC is very complex to analyze as a whole. In particular, it is difficult to correlate each marker with the precise diagnosis of a disease. To examine an archive containing several thousand of different data, cluster analysis can be very useful: it is indeed possible to see how markers are related to each other, thus obtaining important biological information and a rough classification of different cases.

Gene expression profiling studies have been successfully used to identify molecules for use as potential prognosticators. In analogy to gene expression profiling, an original method adapted to oncohematology diagnostics, useful for identify the immunophenotypic signature of subgroups of CLL with different prognosis, called “surface antigen expression profiling”,

has been proposed previously. According to this method, surface marker expression data can be successfully analyzed with data mining tools identical to those used in gene expression profiling studies, including unsupervised and supervised algorithms, with the goal of identifying the immunophenotypic signature of CLL subpopulations with different prognoses. By expanding the panel of markers, some of them (e.g., TCL1, CCR7, FCRL2, FCRL3, and CD150) revealed to have surprising prognostic relevance in CLL (Zucchetto et al, 2011).

It is possible to imagine applying this model to a larger database built by collecting a large number of phenotypes and thoroughly analyzing samples from patients with different oncohematological diseases. Results show that cases related to specific, different diseases (e.g. CLL, Hairy Cell Leukemia, FL, BL) cluster with specific markers, that characterize them and are co-expressed. Then it becomes interesting to investigate why certain markers are expressed in the same cluster and why cases with different diagnoses are found in the same clusters. Could there be a common origin in the form of a normal counterpart followed by different evolutionary paths? The resulting classification highlights biological similarities. It may be of interest to study whether these similarities are reflected in the clinic, in terms of aggressiveness and response to treatment, and whether they can be used to provide prognostic and therapeutic guidance.

1.5 Machine Learning: ability to classify

The satisfactory use of cluster analysis suggests proceeding down this path, using more sophisticated systems. To analyze a copious amount of data obtained by FC, another help comes from artificial intelligence (AI), which through the use of very complex software and computational systems can compare and stratify all this data in a short time. In this way, AI provides more simply manageable data with a two-dimensional representation that is easier to interpret.

In this field, machine learning (ML) allows computers to learn from their own experience. As human beings, ML algorithms involve a period of "training", in which they learn a predictive model, based on supervised analysis, and a period of "validation", in which this model is applied to cases never seen before (Taye 2023). It is known that several tools have already been implemented in the medical field, for the interpretation of radiological (Kohli et al, 2017; Wang et al, 2012), histopathological (Li et al, 2019; Bayramoglu et al, 2016) and fundus oculi (Xiao et al, 2017; Maji et al, 2015) images, as well as the prediction of clinical outcomes based on electronic medical records (Miotto et al, 2016). Since FC produces a large amount of data and its interpretation requires high analytical skills, it naturally represents the ideal field of application for ML algorithms: consequently, ML tools have already been applied to different stages of FC, from data pre-processing (O'Neill et al, 2013; Rahim et al, 2018) to detection of disease or minimal residual disease (MRD) (Angeletti et al, 2018; Ko et al, 2018).

1.6 Interesting implementations in AI for a more refined classification

Dimensionality reduction methods ranked highly, such as generic t-distributed stochastic neighbor embedding (t-SNE) and its initial Matlab-based implementation for FC data viSNE (Cheung et al, 2021). t-SNE is a nonlinear dimensionality reduction technique particularly effective in visualizing complex and nonlinear relationships within the data. It aims to map high-dimensional data points to a lower-dimensional space in such a way that similar data points in the original space are represented close together in the reduced space. Conversely, dissimilar data points are represented farther apart. The primary differences between t-SNE and ViSNE lie in their specific applications and the datasets they are best suited for. While t-SNE is a general-purpose technique applicable to various high-dimensional datasets, ViSNE is tailored for single-cell data analysis. Both t-SNE and ViSNE have been widely used in the field of bioinformatics and single-cell genomics to reveal the underlying cellular heterogeneity and identify distinct cell populations within a sample.

Software with graphical user interfaces also ranked highly, including PhenoGraph, SPADE1, FlowSOM, and Citrus, with unsupervised learning methods outnumbering supervised learning methods, and algorithm type popularity spread across K-Means, hierarchical, density-based, model-based, and other classes of clustering algorithms.

PhenoGraph is an unsupervised clustering algorithm used to analyze high-dimensional single-cell data, such as data obtained by FC and mass cytometry (CyTOF) (Levine et al, 2015). It is widely used for its ability to identify cell populations based on their phenotypic profiles. PhenoGraph constructs indeed a graph representation of the high-dimensional data, where each cell represents a node and the edges between nodes are weighted based on the similarity between cell phenotypes. The underlying assumption is that cells belonging to the same population have similar phenotypes and are therefore more likely to be connected in the graph.

SPADE1 (Spanning-tree Progression Analysis of Density-normalized Events) is another popular unsupervised algorithm for FC data analysis. SPADE1 constructs a density-based spanning tree to identify cell populations, thus it has been widely adopted in the field of single-cell analysis, because it provides valuable insights into cellular heterogeneity, differentiation, and cell-cell interactions.

FlowSOM is an algorithm used for unsupervised clustering and visualization of high-dimensional single-cell data, primarily in the context of FC (Van Gassen et al, 2015). FlowSOM uses self-organizing maps (SOMs), a type of artificial neural network, for dimensionality reduction and visualization of high-dimensional data. SOMs are particularly useful for preserving the topological relationships between cells, making them suitable for analyzing complex and heterogeneous datasets.

Citrus is a method used for finding statistically significant associations between cellular phenotypes and experimental outcomes, such as disease status or treatment response

(Polikowsky et al, 2019). Citrus first performs dimensionality reduction using t-SNE or principal component analysis (PCA). This step reduces the high-dimensional FC data to a lower-dimensional space, making it more computationally tractable and facilitating the identification of phenotypic signatures. Then Citrus uses a random forest classifier to find features (fluorescent markers) that significantly stratify cells based on the experimental variable of interest. It evaluates the ability of each marker to discriminate between different experimental groups, considering both the purity of cell populations and the strength of association with the experimental variable. Thus Citrus ranks the markers based on their importance in stratifying the data. It selects the top-ranking markers to create a stratification signature, which is a set of markers defining the subpopulations of cells that are associated with the experimental outcome.

In our project, we applied the Predictive Power Score (PPScore) method and the Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction technique.

PPScore is a Python library used for feature selection in machine learning and data analysis. It evaluates the predictive power of each feature in a data set by measuring its relationship with the target variable. It can be used to identify the most informative features, which can be useful for building more accurate predictive models or reducing dimensionality.

UMAP is a dimensionality reduction technique used in the field of machine learning and data analysis to visualize and cluster high-dimensional data. It is an alternative to t-SNE, known for its efficiency in preserving global and local data structures during the dimensionality reduction process. Indeed, UMAP has gained significant attention, particularly in the context of single-cell data analysis, including applications in FC, mass cytometry and single-cell RNA sequencing (scRNA-seq) data analysis (Becht et al, 2018) for some of the key advantages:

- Preserving Global and Local Structures: UMAP aims to preserve both global and local structures of the data. It efficiently maintains the relationships between data points at different scales, allowing for a more accurate representation of complex data structures, such as clusters, gradients, and neighborhoods.
- Non-Linearity: UMAP can handle non-linear relationships in the data, making it suitable for analyzing high-dimensional datasets with complex and non-linear relationships between variables.
- Scalability: UMAP is computationally efficient and scalable to large datasets. It can handle millions of data points with relatively low memory requirements, which is especially crucial in the analysis of high-throughput single-cell data.
- Flexibility: UMAP allows users to control the balance between preserving local and global structures through the "n_neighbors" parameter, providing flexibility in the visualization process. Adjusting this parameter enables users to emphasize different aspects of the data according to their specific research needs.
- Robustness: UMAP is relatively robust to various input data types and noise, making it applicable to a wide range of data domains, including continuous, categorical, and mixed data.

- Visualization: UMAP produces visually appealing and informative two-dimensional embeddings, making it easier for researchers to explore and interpret complex datasets. These embeddings can provide insights into the underlying biological or functional characteristics of cell populations or other data points.
- Interpretability: UMAP provides a set of parameters that allow users to control the trade-off between optimizing global and local structures. This enables researchers to fine-tune the visualization according to their domain knowledge and research questions.
- Open-Source Implementation: UMAP is available as an open-source Python library, making it easily accessible and usable for the broader research community.

It's interesting to note that unsupervised learning methods are dominant in this context. Unsupervised learning refers to machine learning techniques where the algorithm identifies patterns and structures in the data without explicit guidance or labeled examples. In contrast, supervised learning requires labeled data for training.

In addition, the popularity of clustering algorithms seems to be diverse, with various types being used. Some of the common types mentioned are K-Means, hierarchical clustering, density-based clustering, model-based clustering, and other unspecified classes of algorithms. This indicates that different methods are preferred depending on the specific dataset and research goals.

AIM OF THESIS

The research project that constituted my PhD thesis work focuses on an in-depth study of the phenotypic characteristics of different types B-NHL.

We know that lymphoid neoplasms are a very heterogeneous group of malignancies, which respond differently to specific treatments, and immunophenotypic characterization is a key element for their classification, in order to direct to the correct therapeutic plan. Their classification thus requires skillful evaluation by expert hematopathologists, but the risk of error remains higher in these tumors than in many other areas of pathology.

We also know that FC, applied to samples derived from different sources of patients affected by the most common mature B-NHL, allows to evaluate at the same time the quantitative expression of several markers, easily defining clonal populations.

My purpose is to identify which of the markers used over the past 20 years seemed most useful and specific for characterizing (and classifying) these types of lymphomas, analyzing cases received and characterized for immunophenotype at the Laboratory of Immunopathology of A.O. Ordine Mauriziano in Turin.

Therefore, I firstly focused my attention on a wide series of B-NHL, obtained from both blood and non-blood samples, all with a diagnosis confirmed by pathologist and/or hematologist, to analyze by means of AI the expression of several markers. Secondly, to better evaluate the efficacy in the diagnostic process of intracellular markers, identified in the WHO classification of lymphoid tumors for their capacity to differentiate these pathologies, we consider a homogeneous case series constituted by only tissue samples, whose diagnoses, all confirmed by histologic analysis, were grouped into 8 major categories of B-NHL patients.

Then, data obtained were evaluated with different statistical analysis systems, supplied by AI, to select the best approach able to identify the markers of greatest interest and most effective to characterize homogeneous groups of lymphomas.

This work aims to be functional to set up an antibody panel for systematic use in a multiparametric immunophenotypic analysis of samples, performed with a high complexity flow cytometer.

MATERIALS AND METHODS

3.1 Sample collection

Six hundreds eighteen tissue samples (lymph nodes, biopsies, FNA), collected in our laboratory since 2009 to 2022, and all referred to the 8 major categories of B-NHL, were analyzed in this study, including 193 DLBCL, 193 FL, 92 MZL (57 NMZL,12 SMZL, 23 MALT), 67 CLL, 27 MCL,16 LPL, 10 BL, 20 HGBL. All diagnoses were established according to the 2017 WHO criteria and confirmed by histologic analysis.

3.2 Flow cytometry

All cases were extensively characterized for immunophenotype by multiparametric FC. A cell suspension of tissue samples was obtained by mechanical disaggregation, immediately counted and evaluated for cell viability in the hemacytometer. If viability was less than 70% and/or the total number of viable cells was greater than 10 x 10^6, mononucleated cells were then purified by density gradient; otherwise, death cells were excluded from analysis using 7AAD (Beckman Coulter) staining. Cells were then washed twice with RPMI 1640 + FBS 10% and resuspended in the same medium at 5–20 x 10^6/ml. Fifty to 100 ml of samples were directly added to a monoclonal antibody mixture, incubated 15–30 min at room temperature, washed twice with PBS/BSA 0.5%, resuspended and immediately acquired. For detection of intracellular markers, after surface staining with backbone markers, cells were fixed, permeabilized and stained with appropriate amount of monoclonal antibody using FoxP3 Staining Buffer Set (Miltenyi Biotec), according to manufacturer's instructions. This study is a 14-year retrospective study: over this period, the instruments and reagents used have been changed several times, adapting to technological evolution. The oldest flow cytometric sample acquisitions (2009–2010) were performed by a CyAn ADP (Dako) equipped with 3 solid state lasers (488 nm, 405 nm and 642 nm) and 8 colors; then, from 2010 to 2019 samples were acquired with a 3-laser and 10-color Navios (Beckman Coulter) and finally, since 2020 with a 3-laser and 13-color DxFlex (Beckman Coulter). Analyses were performed by Kaluza 1.3 software (Beckman Coulter). Similarly, most of the antibodies used were initially only available with FITC, PE, PerCP and APC fluorochromes; over time, monoclonal antibodies conjugated with Pacific Blue, APC-Alexa700, APC-Alexa750, PE-Cy5 / PE-Cy5.5 and PE-Cy7 have been added. **Table 1** lists the antibodies used since 2009. CD45, CD 19, CD20, CD5, FMC7, CD9, CD11c, CD22, CD23, CD24, CD25, CD31, CD38, CD43, CD44, CD79b, CD81, CD103, CD123 were provided by Beckman Coulter; Kappa and Lambda F(ab)2 polyclonal antibodies were provided by Agilent-Dako; CD6, CD21, CD72, CD74, CD200, CXCR3, Bcl-2, MIB-1, IRF-4 were provided by Miltenyi Biotec GmbH; CD10 and Bcl-6 were provided by BD Biosciences. The expression of each marker was evaluated on monoclonal B cells only using a gating strategy where CD19 was combined with one or more antibodies to obtain a purity of clonal restricted cells greater than 95% **(Figure 1)**. Essentially, an initial set of test tubes, now condensed into a single tube with 12 antibodies and 8 colors, is used to detect and identify the monoclonal cell population, i.e., the backbone markers which guarantee a purity of the monoclonal population greater than

95% in 97% of cases, and greater than 90% in the remaining samples. The percentage of positive cells was calculated by comparing unstained vs. stained cells in each sample.

Antigen	Clone	Clonality	Isotype	Conjugated	Dilution
CD45	J33	Ms mAb	IgG1	Krome-Orange	1:33
CD19	J3-119	Ms mAb	IgG1	PC7/PB	1:33
CD20	B9E9	Ms mAb	IgG2a	FITC/PC5.5	1:20
CD5	BL1a	Ms mAb	IgG2a	PC5.5	1:33
Kappa	polyclonal	Rb poAb	F(ab')2	FITC/APC	1:20
Lambda	polyclonal	Rb poAb	F(ab')2	PE	1:33
FMC7	FMC7	Ms mAb	IgM	FITC	1:20
CD6	M-T411	Ms mAb	IgG1k	APC	1:20
CD9	ALB6	Ms mAb	IgG1	APC-Alexa Fluor750	1:33
CD10	HI10a	Ms mAb	IgG1k	BV421	1:33
CD11c	BU15	Ms mAb	IgG1	PE	1:20
CD21	HB5	Ms mAb	IgG1	APC	1:20
CD22	SJ10.1H11	Ms mAb	IgG1	APC	1:20
CD23	HD50	Ms mAb	IgG2b	APC	1:20
CD24	ALB9	Ms mAb	IgG1	APC-Alexa Fluor750	1:33
CD25	1HT44H3	Ms mAb	IgG2a	APC	1:10
CD31	1F11	Ms mAb	IgG1	PE	1:10
CD38	T16	Ms mAb	IgG1	APC-Alexa Fluor750	1:33
CD43	DTF1	Ms mAb	IgG1	APC-Alexa Fluor750	1:33
CD44	J.173	Ms mAb	IgG1	FITC	1:10
CD72	REA231	REA mAb	IgG1	FITC	1:20
CD74	5-329	Ms mAb	IgG1k	FITC	1:20
CD79b	CB3-1	Ms mAb	IgG1	PE	1:10
CD81	JS64	Ms mAb	IgG2a	PE	1:20
CD103	2G5	Ms mAb	IgG2a	FITC	1:100
CD123	SSDCLY107D	Ms mAb	IgG1	APC	1:20
CD200	OX-104	Ms mAb	IgG1	FITC	1:100
CXCR3	REA232	REA mAb	IgG1	PE	1:10
Bcl-2	*REA872*	*REA mAb*	*IgG1*	*FITC*	*1:50*
Bcl-6	*K112-91*	*Ms mAb*	*IgG1k*	*PE*	*1:50*
MIB-1	*REA183*	*REA mAb*	*IgG1*	*FITC*	*1:100*
IRF-4	*REA201*	*REA mAb*	*IgG1*	*APC*	*1:20*

Table 1. *Flow-cytometric antibodies selection. Recommended antibodies dilution each 1×10^6 cells. Bcl-2, Bcl-6, MIB-1 and IRF-4 are transcription factors and require intracellular staining. mAb: monoclonal Antibody; poAB: polyclonal Antibody; Ms: Mouse; Gt: Goat; REA: Recombinant Human Antibody.*

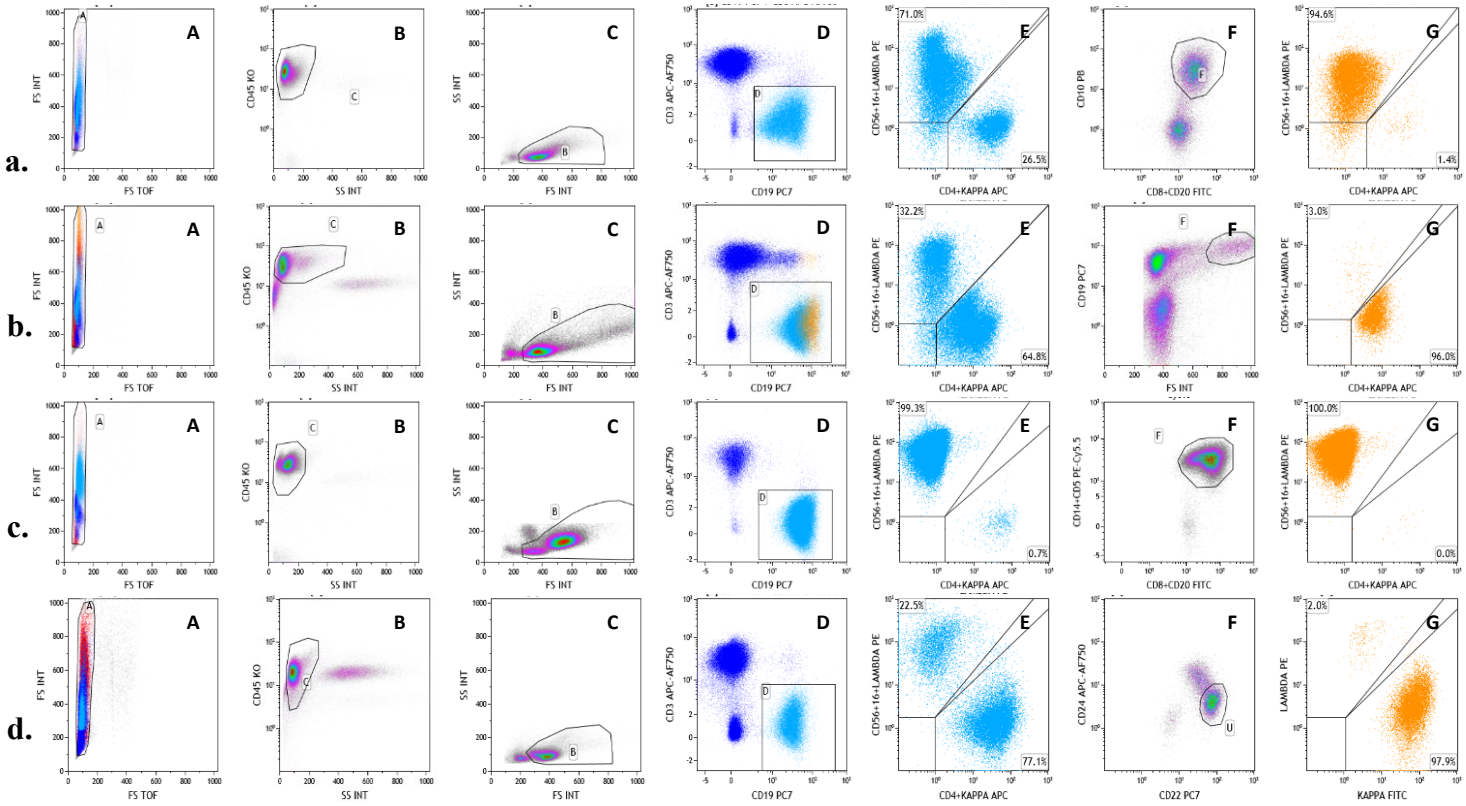


Figure 1. *Flow cytometric gating strategy in different sample types/lymphoma entities. The sequence of gate is aimed to obtain the greatest purity of neoplastic cells as identified by means of the clonal restriction of immunoglobulin light chains (kappa or lambda). In all cases the first four plots (A, B, C and D) follow an identical scheme: plots A are FS TOF (Forward Scatter Time Of Flight) vs FS INT (Forward Scatter Intensity) and the gate A is drawn to exclude doublets. Plots B, activated on gate A, are SS INT (Side Scatter Intensity) vs CD45 and the gate C is drawn around CD45 positive cells with low SS, to exclude granulocytes (CD45+/SS^{high}) and unlysed/ghost erythrocytes (CD45-/SS^{low}). Plots C, activated on the Boolean gate [A AND C], are FS INT vs SS INT were all the cells identified with the previous gates are further gated (gate B) to exclude debris (low FS and low SS). Plots D, activated on the Boolean gate [A AND C AND B], are CD19 (total B cells) vs CD3 (total T cells) and the gate D is drawn to select all CD19 positive cells.*

Figure 1a. *In a LN of a FL patient, plot E (gated on [A AND C AND B AND D]) shows the kappa/lambda distribution (with a ratio of 1/2.73) on the total CD19+ B cells. Plot F (gated on [A AND C AND B AND D]) shows the presence of a population of CD20+high, CD10+ cells that are gated in region F. The plot G E (gated on [A AND C AND B AND D AND F]) shows the expression of kappa and lambda light chains on CD19+, CD20+high, CD10+ cells: 94.6% of B cells (virtually 98.5%) are clonally restricted for lambda light chain.*

Figure 1b. *On a FNAB referred to a DLBCL patient, plot E (gated on [A AND C AND B AND D]) shows the kappa/lambda distribution (with a ratio of 2/1) on the total CD19+ B cells. Plot F (gated on [A AND C AND B AND D]) shows the presence of a minor population of CD19+ B cells with a high FS (Large B cells, gated on F). The plot G (gated on [A AND C*

AND B AND D AND F]) shows the expression of kappa and lambda light chains on the CD19+ large B cells that are clonally restricted for kappa light chain (96% of expression at low intensity).

Figure 1c. *On a FNAB referred to a MCL patient, plot E (gated on [A AND C AND B AND D]) shows the kappa/lambda distribution on the total CD19+ B cells with less than 1% of kappa positive cells. Plot F (gated on [A AND C AND B AND D]) shows that almost all CD19+ cells co-express CD5 and CD20, both at high fluorescence intensity, gated on F. The plot G (gated on [A AND C AND B AND D AND F]) shows the expression of kappa and lambda light chains on the CD19+, CD20+, CD5+ B cells that are clonally restricted for lambda light chain at 100% with high fluorescence intensity.*

Figure 1d. *On a Spleen excised from a SL patient, plot E (gated on [A AND C AND B AND D]) shows the kappa/lambda distribution on the total CD19+ B cells with a kappa:lambda ratio of 3.4:1. Plot F (gated on [A AND C AND B AND D]) shows the expression of CD22 and CD24 on the total B cells, a discrete population of CD22+high/CD24+low cells is clearly identified and gated on F. The plot G (gated on [A AND C AND B AND D AND F]) shows the expression of kappa and lambda light chains on the CD19+, CD22+high, CD24+low B cells that are clonally restricted for kappa light chain with a purity of 97.9%.*

3.3 Machine Learning

Each predictive model consists of a single classification tree, a supervised learning technique introduced by Breiman (Breiman et al, 1984). A classification tree is a rooted and directed acyclic graph (DAG) defined as $\Gamma = (\Phi \cup \Delta \cup \Theta \cup E)$. The root node $\phi \in \Phi$ with an in-degree 0 represents the most effective classification rule. The terminal nodes $\theta \in \Theta$ with an out-degree 0 represent the class to which a generic input sample is assigned. Each decision node $d \in D$ implements a threshold (T_h) comparison with an FC marker. The result of such comparison fires one of the two outgoing edges $e_f, e_t \in E$, enabling the connection to the nodes at the lower levels: if a marker $m_k \leq T_H$ then the true branch et is active, otherwise the false branch e_f is active. Depending on the values assumed by the FC markers of a given sample, only one root-to-leaf path is activated. Classification trees were obtained with a script developed in Python, using the open-source libraries Pandas (Mc Kinney, 2011) for data preprocessing, and SciKit learn (Pedregosa et al, 2011) for training and validation. For each classification tree, the training algorithm was applied to a subset of the dataset containing the 75% of the samples, while the validation was performed on the remaining 25% of the population. Since the training and validation samples were randomly drawn from the entire dataset, and since the available labels are significantly unbalanced, we ensured that all classes were represented in both the training and validation sets by using stratified random selection (Neyman, 1992). The settings for each classification tree were established through a comprehensive grid search analysis. This involved a thorough exploration of various configurations for classification trees based on empirical assessments. Specifically, the grid search analysis for each classification tree included the following parameters:

- Maximum depth, with values spanning from 6 to 12.
- Minimum number of samples required for a split, with values ranging from 2 to 10.
- Minimum number of samples required for a leaf node, with values in the range of 1 to 10.
- Maximum number of features criterion (options: sqrt, log2, or unconstrained).

Grid search was combined with the k-fold cross-validation (Bengio et al, 2004), where k is set according to the number of unique labels. It is important to note that k-fold was performed only on samples from the training set. The best performing model was selected based on its highest accuracy. It was then evaluated using the validation dataset. The training algorithm relies on the Gini impurity (Breiman, 1996) or on the information gain (Quinlan, 1986) to quantify the amount of entropy of each possible partition.

3.4 PPScore Calculation

The PPScore quantifies the agreement between predicted values and actual observations, providing insights into the model's performance. To calculate the PPScore, we followed the approach proposed by Florian Wetschoreck (Wetschoreck, 2020). Leveraging the Predictive Power Score, we evaluated the predictive performance of each individual marker against all lymphoma categories. Initially, we entered into an Excel sheet the cases of B-NHL patients whose tissue samples we had received since 2009, reporting the measured value for each marker used. During this period, the flow cytometric data were obtained from different machines and different reagents settings. As the fresh samples were no longer available to repeat immunostainings, we tried to make the data collected over the years repeatable by excluding parameters such as the MFI or the choice of cutoff, which may vary by operator, different instruments, batches and fluorochromes of antibodies. Thus, we decided to express only the percentage of positive cells using the internal control of unstained cells as negative. Furthermore, the expression of more than 50 different antigens has been evaluated in the course of all these years. However, these markers were used with different frequencies, so it was necessary to understand which ones were useful to our study. Then we applied the PPS technique to the entire biopsy database, replacing missing values with the mode and eliminating markers with more than 55% missing values, regardless their diagnosis. These predictions were generated using the PPScore Python package (https://github.com/8080labs/ppscore/). The dataset from Excel was exported to a Pandas DataFrame and processed in Python. This metric quantifies the similarity between predicted and observed values within each bin.. The PPScore is calculated by analyzing the conditional mutual information between two features in a dataset. It measures the predictive relationship between the features by considering both the strength and the shape of the relationship. The calculation involves estimating the information shared by the features, taking into account the conditional distributions and probabilities. The PPScore algorithm captures both linear and non-linear relationships, providing a comprehensive measure of predictive power. The

resulting score ranges from 0 to 1, with 0 indicating no predictive power and 1 indicating perfect predictive power.

3.5 Uniform Manifold Approximation and Projection (UMAP)

We chose the UMAP algorithm because it is competitive with t-SNE for visualization quality, and arguably preserves more of the global structure with superior run time performance. Furthermore, UMAP has no computational restrictions on embedding dimension, making it viable as a general purpose dimension reduction technique for machine learning (McInnes et al, 2018). In our study, we used UMAP to visualize the relationships between the various markers used (x) and the diagnosis (y) related to the phenotypes included in our case series. We applied the UMAP algorithm using the UMAP-learn in Python and evaluated a set of different values for the n_neighbors parameter which determines the number of nearest neighbors used to construct the fuzzy topological representation, influencing the balance between local and global structure preservation in the resulting low-dimensional representation. At the same time, we kept the min_dist parameter, which controls the minimum distance between points in the low-dimensional representation, close to 0 in all experiments. UMAP transformed the high-dimensional data into a lower-dimensional space, capturing the intrinsic relationships and similarities among samples. The resulting embeddings were used for downstream analysis and visualization.
The UMAP embeddings were visualized using the matplotlib Python package (https://ieeexplore.ieee.org/document/4160265). We color-coded the points in the UMAP plot to represent the different clusters representing the 8 major classes of B-NHL.

RESULTS

4.1 FC could help in classifying lymphoid neoplasms.

In a previous study (Gaidano et al, 2020), we demonstrated how FC could significantly help to obtain a precise classification of B-NHL, sustaining the traditional approach performed by IHC. Moreover, the dataset comprised 1465 immunophenotypic profiles of clonal B-NHL, deriving from tissue biopsies, liquor, effusions, PB or BM samples, collected from 2003 to 2019. The immunophenotypic panel included more than 50 antibodies. However surface markers with a 50% or higher missing-value rate were completely discarded from the database (**Figure 2**).

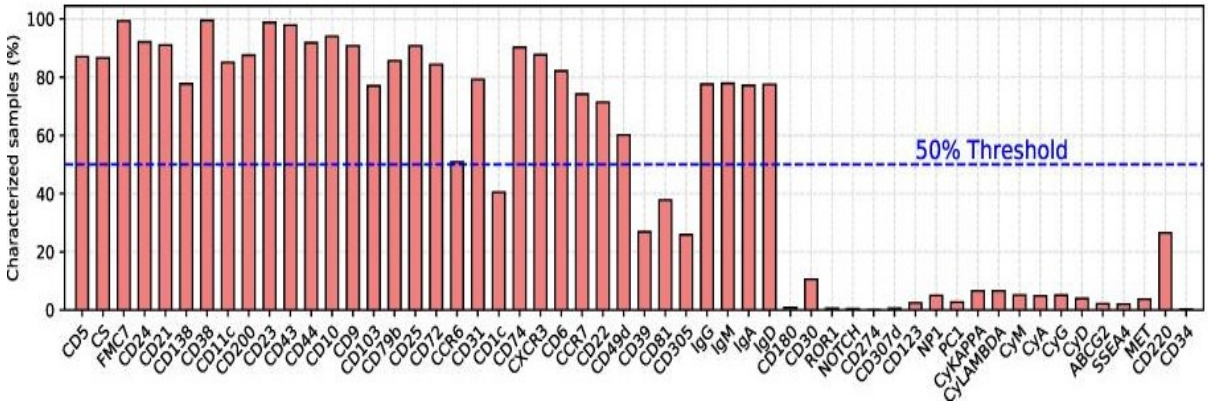


Figure 2. List of surface markers associated with the percentage of samples characterized for each marker. *The blue dotted line indicates the threshold (50%) above which markers were considered.*

To deal with missing values in the remaining markers, we replaced them with the mean values specific to each class. Then we used a ML algorithm to automatically select the most important and representative set of markers.

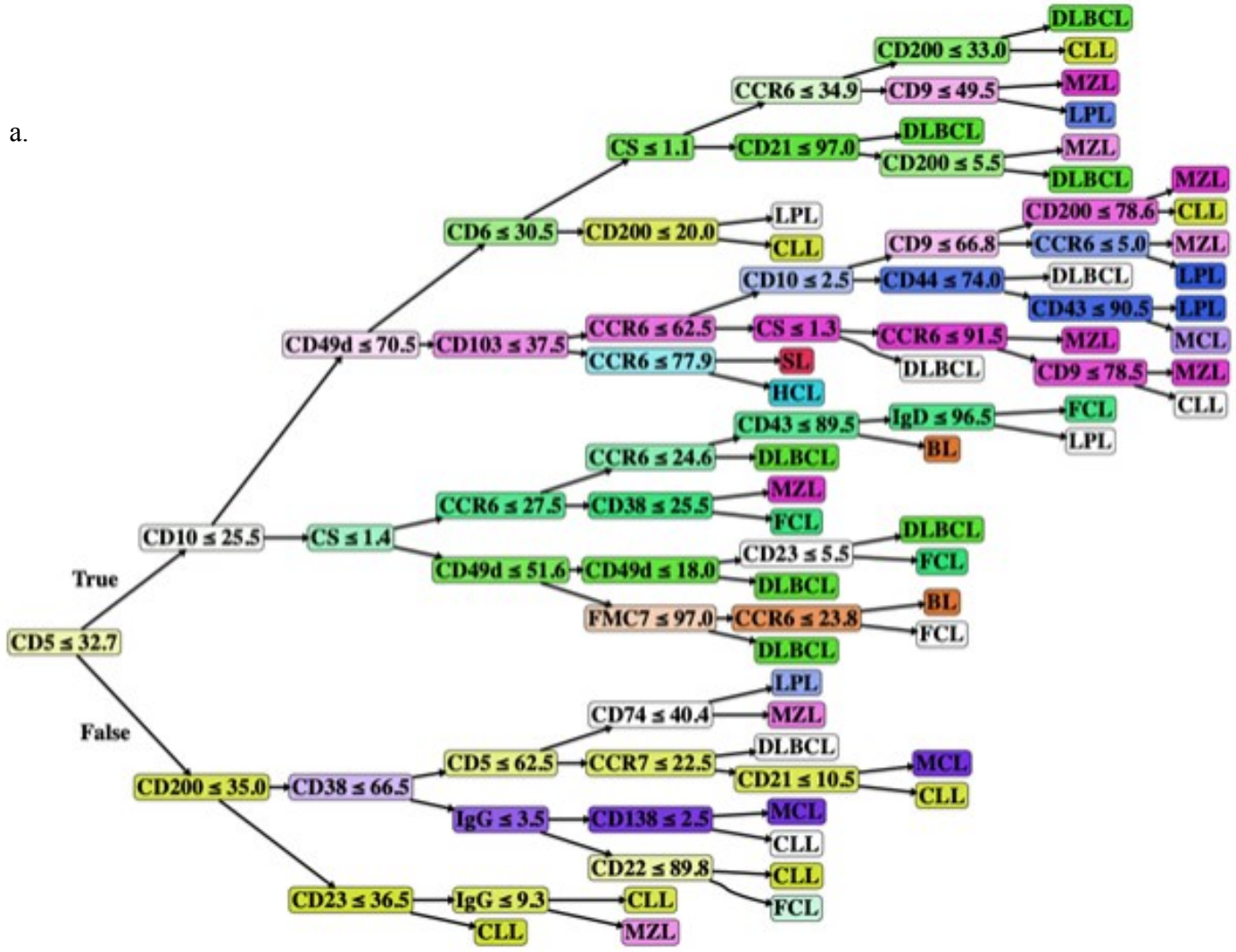
4.2 Classification trees in ML allow to define homogeneous groups of neoplasms, characterized by the expression of specific surface markers.

To validate the predictive models, the first step was to divide the entire dataset into two parts:

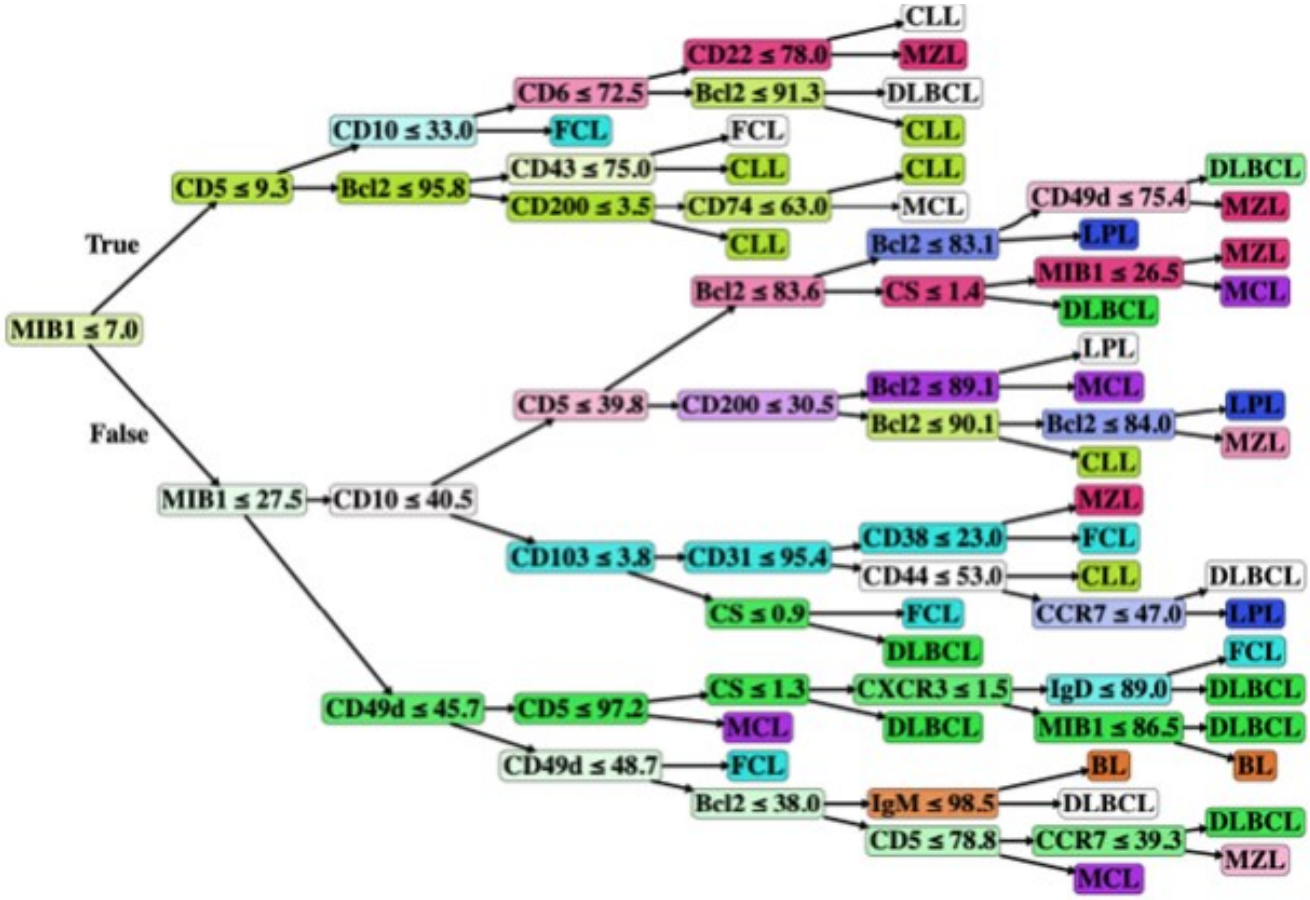
- Training Set (75%): This portion of the data is used to build the classification tree models. In machine learning, this is where the model learns patterns and relationships in the data.
- Validation Set (25%): This part of the data is used to evaluate the performance of the predictive models. It helps assess how well the models generalize to unseen data.

Then classification trees were employed to classify instances into different B-NHL classes. Classification trees consist of three key components: a starting point known as the root node, which corresponds to the most influential marker in terms of distinguishing characteristics, intermediary decision nodes featuring two branches, and terminal nodes called leaf nodes. These leaf nodes not only signify the B-NHL classification but also provide an indication of the likelihood that the assigned classification is accurate (**Figure 3***)*.

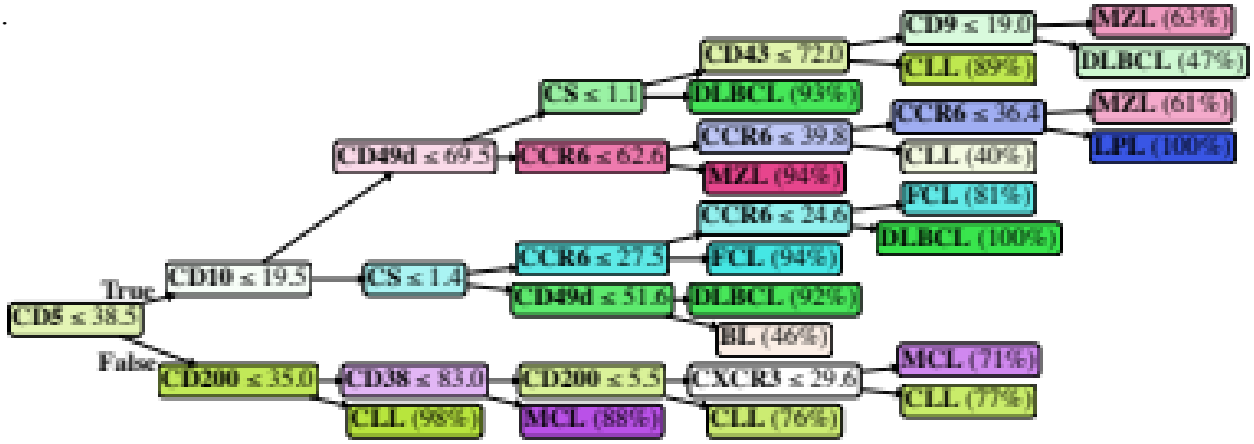
a.



b.


MIB1 ≤ 7.0
True
False
CD5 ≤ 9.3
MIB1 ≤ 27.5
CD10 ≤ 33.0
Bcl2 ≤ 95.8
CD10 ≤ 40.5
CD49d ≤ 45.7
CD6 ≤ 72.5
CD43 ≤ 75.0
CD200 ≤ 3.5
CD5 ≤ 39.8
CD103 ≤ 3.8
CD5 ≤ 97.2
CD49d ≤ 48.7
CD22 ≤ 78.0
Bcl2 ≤ 91.3
FCL
CLL
MZL
DLBCL
CD74 ≤ 63.0
MCL
Bcl2 ≤ 83.6
CD200 ≤ 30.5
CD31 ≤ 95.4
CS ≤ 0.9
CS ≤ 1.3
Bcl2 ≤ 38.0
Bcl2 ≤ 83.1
CS ≤ 1.4
Bcl2 ≤ 89.1
Bcl2 ≤ 90.1
CD38 ≤ 23.0
CD44 ≤ 53.0
CXCR3 ≤ 1.5
IgM ≤ 98.5
CD5 ≤ 78.8
CD49d ≤ 75.4
MIB1 ≤ 26.5
Bcl2 ≤ 84.0
CCR7 ≤ 47.0
IgD ≤ 89.0
MIB1 ≤ 86.5
CCR7 ≤ 39.3
LPL
BL

c.


CD5 ≤ 38.5
True
False
CD10 ≤ 19.5
CD200 ≤ 35.0
CD49d ≤ 69.5
CS ≤ 1.4
CD38 ≤ 83.0
CCR6 ≤ 62.6
CCR6 ≤ 27.5
CD200 ≤ 5.5
CD49d ≤ 51.6
CS ≤ 1.1
CD43 ≤ 72.0
CCR6 ≤ 39.8
CCR6 ≤ 24.6
CXCR3 ≤ 29.6
CD9 ≤ 19.0
CCR6 ≤ 36.4
CLL (98%)
MCL (88%)
DLBCL (93%)
MZL (94%)
FCL (94%)
DLBCL (92%)
BL (46%)
CLL (76%)
CLL (89%)
CLL (40%)
FCL (81%)
DLBCL (100%)
MCL (71%)
CLL (77%)
MZL (63%)
DLBCL (47%)
MZL (61%)
LPL (100%)

d.

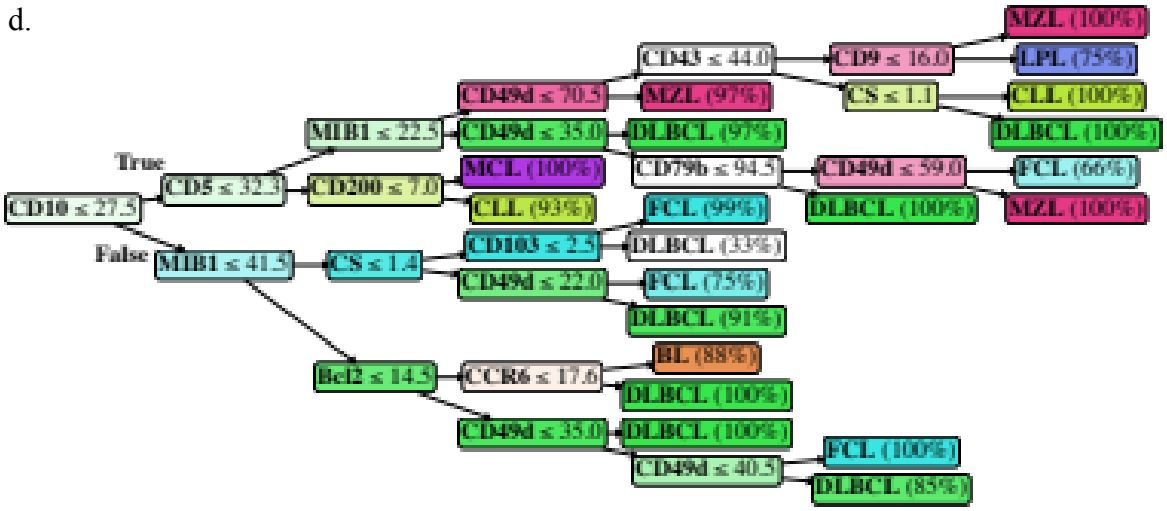


Figure 3. *Classification trees of Model I (a), Model II (b), Model III (c), and Model IV (d). Values reported inside decision nodes represent threshold values. Percentage values reported in leaf nodes represent the prediction probability, i.e., how likely the achieved class is the correct one.*

Furthermore, given the unique nature of intracellular markers, not usually employed in immunophenotypic analysis performed by FC, we opted to create several predictive models, both with and without MIB1 and Bcl2. This approach allowed us to address the inherent issue of missing data in retrospective studies, yielding the most meaningful findings while avoiding any potential distortions. This approach began by ML techniques to the complete training dataset while excluding MIB1 and Bcl2 (referred to as Model I, **Figure 3a**). Subsequently, we introduced MIB1 and Bcl2 but omitted HCL and SL from the analysis due to inconsistent evaluation of these two markers in these LNH subclasses (referred to as Model II, **Figure 3b**). To assess the specific contributions of MIB1 and Bcl2, the same dataset used for Model II was reanalyzed without including MIB1 and Bcl2, resulting in Model III (**Figure 3c**). Finally, to demonstrate that the significance of MIB1 and Bcl2 was not an artifact caused by the data filling method, we focused exclusively on non-blood samples, where these two markers were prevalent (comprising 74.81% of the samples for Bcl2 and 96.35% for MIB1). In this model, known as Model IV (**Figure 3d**), HCL and SL cases were excluded, due to the limited availability of non-blood samples from these subclasses (only 2 samples).

The most important markers in each model represent the root and the first branches of each tree; this reflects their ability to clearly discriminate between different classes (**Figure 4**). Some markers, such as CD5 and CD10, recur in all 4 models because of their fundamental role in leading to the diagnosis. Interestingly, it is absolutely important to introduce the use of intracytoplasmic markers, such as MIB1, which proves to be a powerful discriminator when applied to the phenotypic study of tissue biopsies (Model II and Model IV).

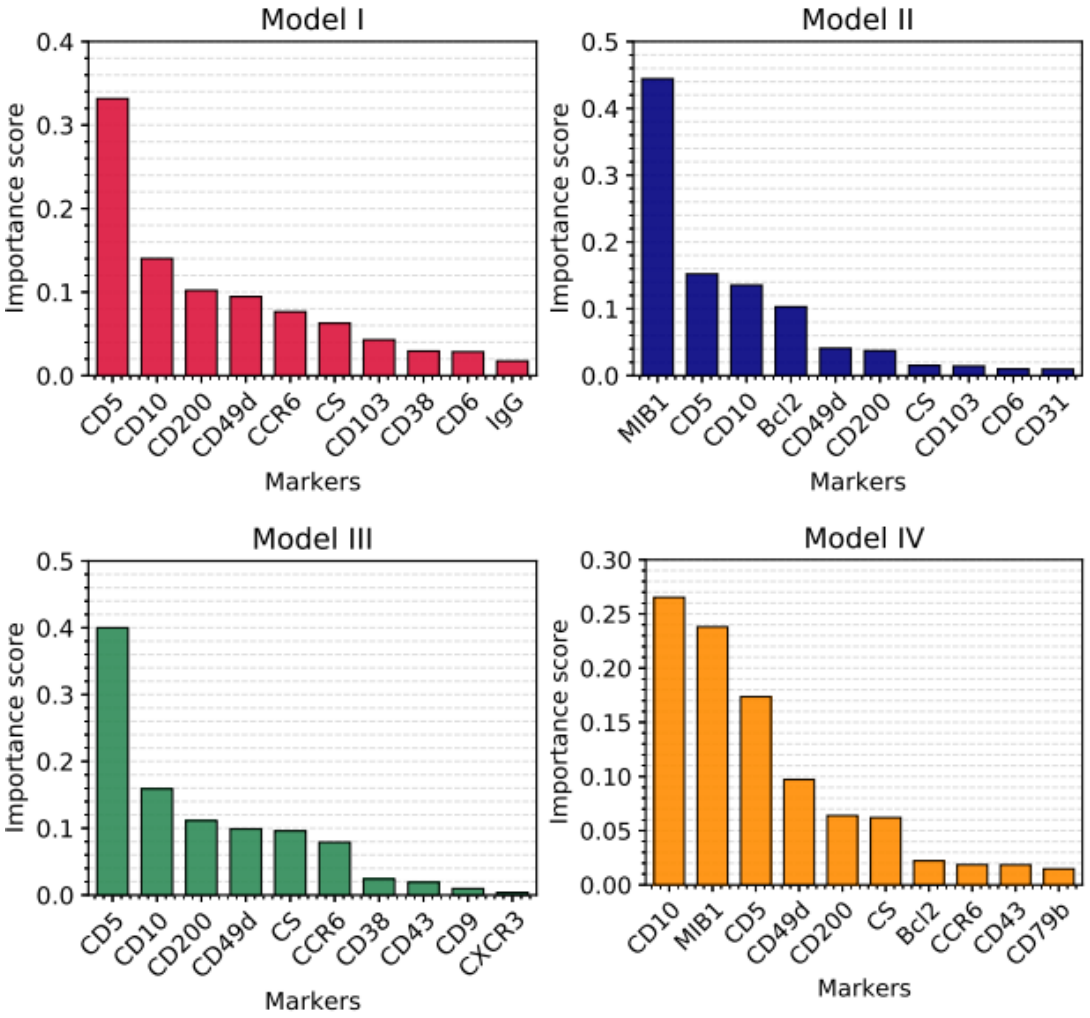


Figure 4. *Ten most representative markers characterize each model defined by the classification trees.*

4.3 Predictive models could categorize B-NHL in up to nine of the most common pathological entities with optimal accuracies.

Each individual tree created was tested on a separate validation set, consisting of samples that were not part of the initial training set. In this context, we used confusion matrices as a means to visualize the performance of each model. These matrices provide a visual representation of how predicted classifications compare to actual sample classifications. Specifically, the true class to which each sample belongs is shown on the ordinate axis, while the predicted class is displayed on the abscissa axis. In an ideal scenario where a classifier achieves 100% accuracy, all samples would align perfectly along the diagonal in the confusion matrix. **Figures 5a** to **5d** present the confusion matrices generated by the Model I, Model II, Model III, and Model IV classification trees, respectively.

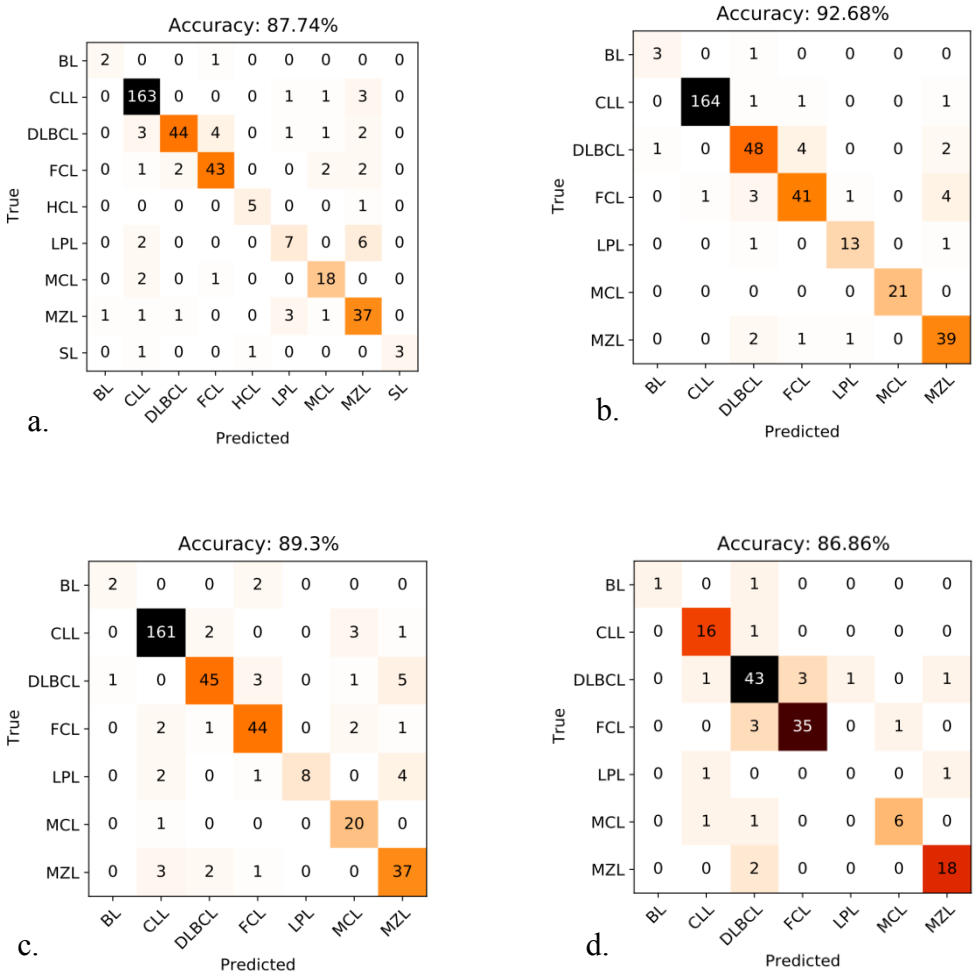


Figure 5. *Confusion matrices and corresponding overall accuracy of algorithms deriving from Model I (Fig.5a), II (Fig.5b), III (Fig.5c) and IV (Fig.5d)*

The accuracy, which is calculated as the proportion of correctly predicted samples divided by the total number of samples, obtained from the confusion matrix of Model I, is 87.74%, for

Model II is 92.68%, for Model III is 89.3%, and for Model IV is 87.59%. It is important to underline that the obtained training accuracy for each model is respectively 85.97%, 91.27%, 86.67%, and 87.35%. This means that there is no overfitting, i.e., the predictive models are neither too rigid nor calibrated on the training set, but they are rather capable of generalizing classification rules on new data quite effectively.

4.4 FC can contribute significantly to the study of B-NHLs performed on tissue samples.

In order to study more thoroughly the diagnostic and, if possible, prognostic value of the markers used in the phenotypic characterization of the samples analyzed, we selected, from the numerous cases received between 2009 and 2022, the most complete and significant phenotypes related to those patients who had received a diagnosis of B-NHL after undergoing biopsy collection. The reports were validated by IHC, so we simultaneously confirmed that FC can be an equally reliable tool in oncohematologic diagnostics performed on tissue. At the time of initial diagnosis, the mean age of the 618 patients, 318 men and 300 women, was 66 years. However, the number of cases attributed to each gender and the average age of the patients varied according to the type of B-NHL diagnosed, as shown in **Table 2**. It is important to point out that the number of subjects reported for each category of B-NHL does not always reflect the real frequency of diagnosed cases: in fact, for some types of neoplasia (e.g., CLL), the analyses are mainly performed on PB.

diagnosis	tot cases	average age at time of diagnosis	cases F	average age F	cases M	average age M
DLBCL	**193**	**66**	**96**	**68**	**97**	**63**
FL	**193**	**64**	**102**	**63**	**91**	**65**
MZL	**92**	**67**	**41**	**70**	**51**	**68**
CLL	**67**	**71**	**31**	**72**	**36**	**69**
MCL	**27**	**64**	**13**	**65**	**14**	**63**
LPL	**16**	**69**	**7**	**71**	**9**	**68**
HGBL	**20**	**67**	**9**	**73**	**11**	**62**
BL	**10**	**48**	**1**	**70**	**9**	**46**
	618		**300**		**318**	

Table 2. *Patient characteristics in selected cases.*

Clonal populations, previously identified by mean of "backbone markers" (CD19, CD20, or any combination giving on gated cells a percentage of clonally restricted cell greater than 95%), were characterized with the large panel of antibodies described above (see "Materials and Methods"). In this study, we included routinely some cytoplasmic (cy) and nuclear (nu)

markers such as cyCD79a, cyBCL-2, cyZAP70, nuMIB-1(Ki-67) and more recently nuIRF4/MUM1, nuBCL-6 and nuMNDA.

4.5 PPScore is useful to evaluate the impact of each marker to define every B-NHL category.

We evaluated the predictive power of each marker across all lymphoma categories using the PPScore. The PPScore method allowed us to evaluate the impact of each individual marker (x), including the diagnosis, on all other markers (y). The closer the value obtained is to 1, the higher the correlation. (**Table 3**).

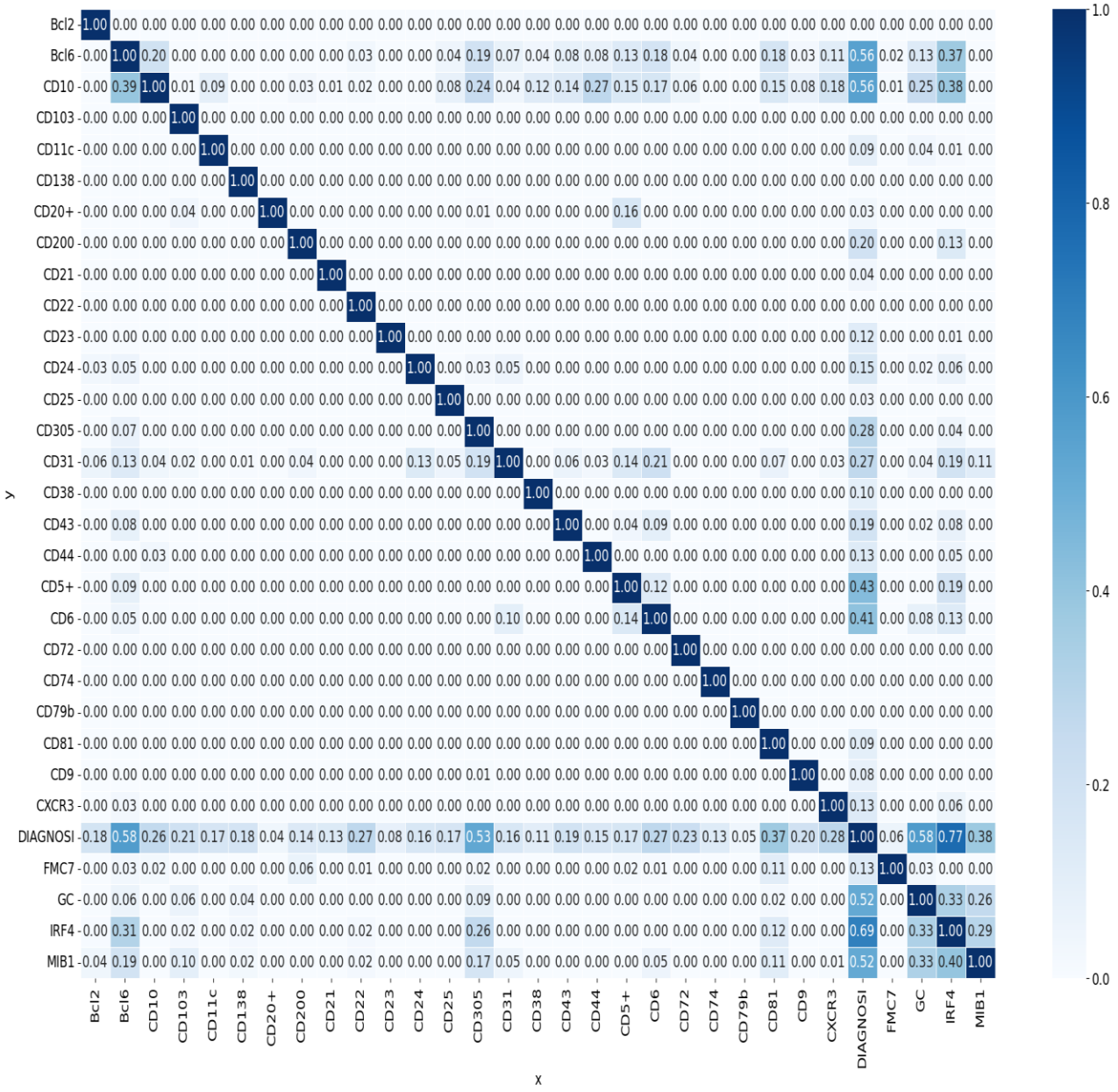


Table 3. *PPScore calculation considering the diagnostic impact of each single marker belonging to the entire antibody panel*

PPScore values ≥ 0.22 (the baseline score) are statistically significant to discriminate diagnostic categories. The results obtained demonstrate how the use of surface and intracellular markers allows us to define the major categories of B-NHL with a high degree of accuracy. Ten or less markers seem to be sufficient to achieve an adequate classification capability. (**Table 4**).

x	y	ppscore	case	is_valid_score	metric	baseline_score	model_score	model
IRF4	DIAGNOSI	0.772383	classification	True	weighted F1	0.234146	0.825679	DecisionTreeClassifier()
GC	DIAGNOSI	0.582396	classification	True	weighted F1	0.234146	0.680176	DecisionTreeClassifier()
Bcl6	DIAGNOSI	0.580251	classification	True	weighted F1	0.234146	0.678534	DecisionTreeClassifier()
CD305	DIAGNOSI	0.528243	classification	True	weighted F1	0.234146	0.638703	DecisionTreeClassifier()
MIB1	DIAGNOSI	0.376755	classification	True	weighted F1	0.234146	0.522685	DecisionTreeClassifier()
CD81	DIAGNOSI	0.374881	classification	True	weighted F1	0.234146	0.521250	DecisionTreeClassifier()
CXCR3	DIAGNOSI	0.279717	classification	True	weighted F1	0.234146	0.448369	DecisionTreeClassifier()
CD6	DIAGNOSI	0.272696	classification	True	weighted F1	0.234146	0.442992	DecisionTreeClassifier()
CD22	DIAGNOSI	0.271921	classification	True	weighted F1	0.234146	0.442398	DecisionTreeClassifier()
CD10	DIAGNOSI	0.260594	classification	True	weighted F1	0.234146	0.433723	DecisionTreeClassifier()

Table 4. *Markers of greatest prognostic value according to PPScore calculation*

We further validated the role of these markers by combining them in a classification tree. Here, each marker is analyzed in combination with all the others leading to a structural relationship tree that separates the entire database in quasi-homogeneous groups of lymphomas (**Figure 6**).

The distribution of most significant phenotypic markers in each category of B-NHL was obtained from statistical analysis in Python (**Figure 7**).

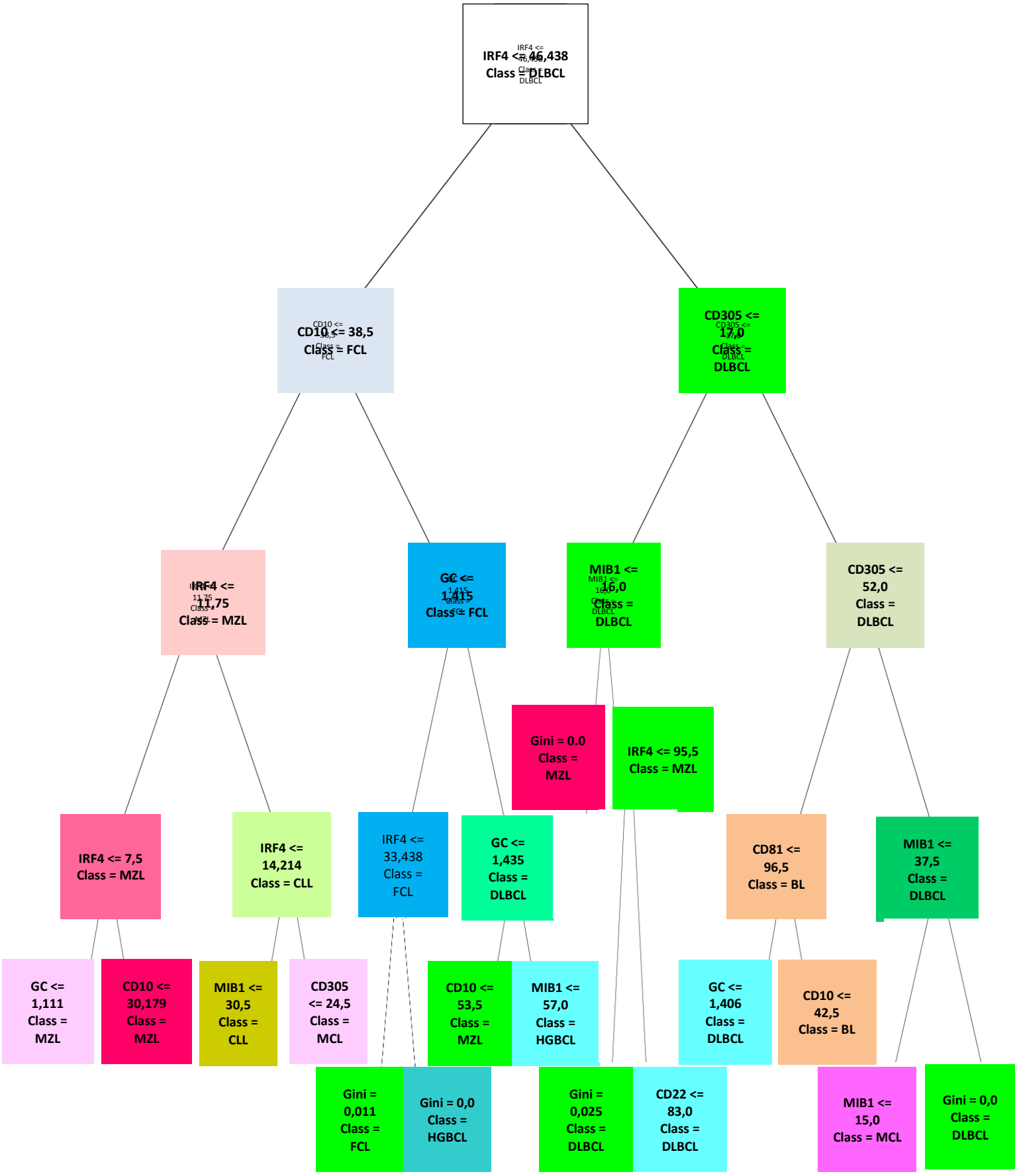


Figure 6. *Classification tree based on 10 most predictive markers.*

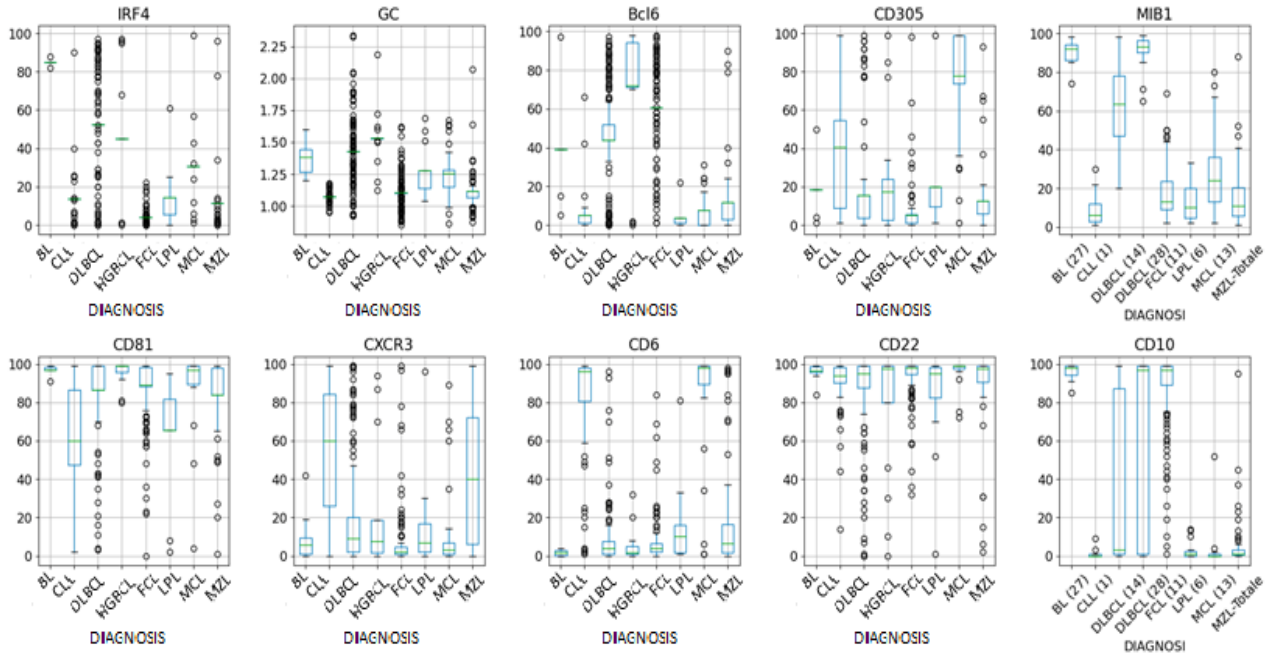


Figure 7. *Box plot representation of clusters obtained from statistical analysis in Python*

Observing the distribution of the clusters in the box plots, it is interesting to note that most of the markers have a distribution that is strongly correlated with the different types of B-NHL, and that high expression (or lack of expression) of some of them is characteristic of some NHLs.

4.6 UMAP separates B-NHL categories in clusters with a high degree of accuracy.

The UMAP dimensionality reduction technique is employed to group the 8 lymphoma categories into distinct clusters. Notably, when UMAP is applied to the 10 markers that are identified as having the most significant impact on diagnosis (as shown in **Figure 8**), it leads to a more pronounced separation of clusters. However, it does not differentiate between various lymphoma entities, as indicated by the arrow. In contrast, by exploring the expression patterns of all markers included in the phenotypic panel (as illustrated in **Figure 9**), where each antibody contributes to the UMAP representation regardless of its statistical significance, a clear and distinct separation of the 8 B-NHL groups into clusters becomes evident. Thus, we can conclude that PPScore analysis unveiled 10 markers which, among all, strongly correlated with the diagnosis. Nevertheless, a greater number of markers, combining intracellular with unconventional markers (CD305, CD81), increases the ability of UMAP to separate different entities.

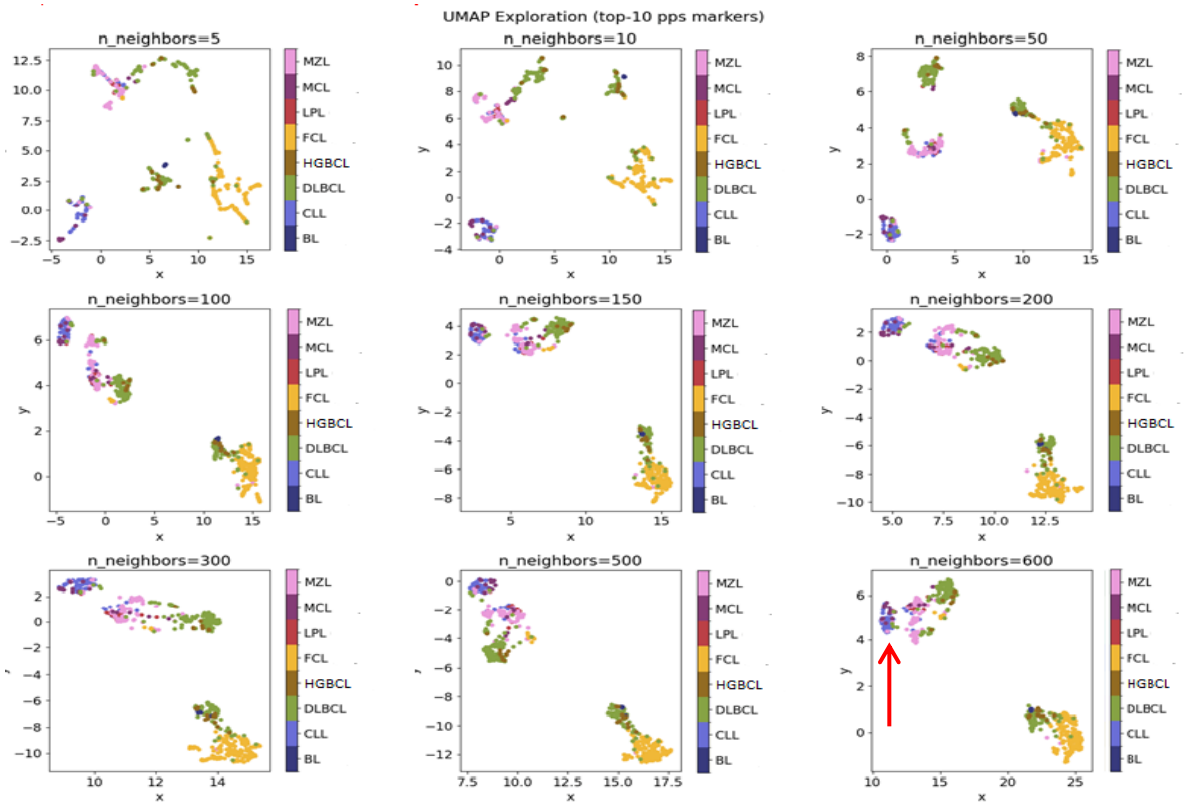


Figure 8. *UMAP exploration on best predictor markers.*

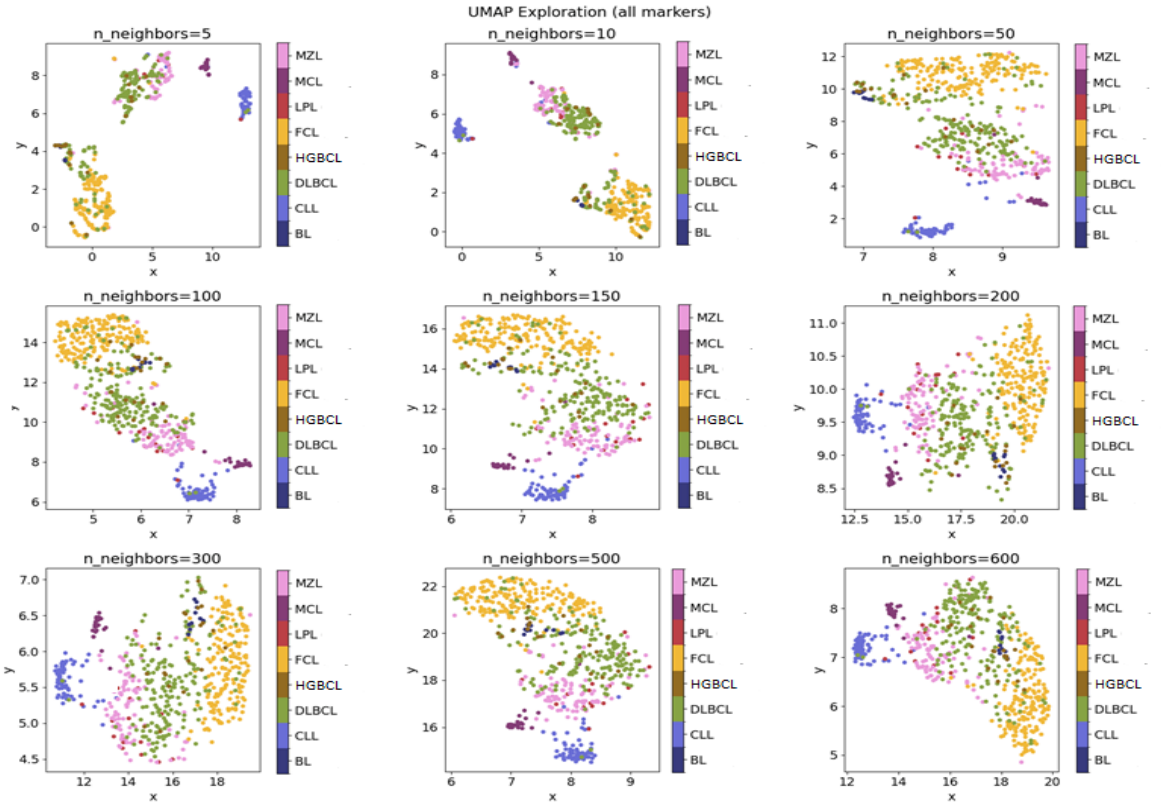


Figure 9. *UMAP exploration on all markers.*

DISCUSSION

It seems clear that the diagnosis of B-NHLs to date is highly variable and complex and, as recommended by the WHO classification of hematopoietic and lymphoid tumors, can only be properly achieved by careful integration of clinical, morphologic, immunophenotypic, molecular, and cytogenetic data. Although guidelines recommend that diagnosis should be made by excisional lymph node biopsy and that histopathologic evaluation of tissue architecture should guide subsequent immunohistochemical and biomolecular studies (Campo et al, 2011; Cheson et al, 2014), the surgical procedure is not readily applicable to patients with deep neoplastic lesions (e.g., retrosternal or retroperitoneal) or to frail or debilitated patients in whom invasive surgery may cause clinical complications with severe physical and psychological consequences. Core needle biopsy (CNB), which is performed under ultrasound guidance with a 16-gauge needle in cases where surgery is not indicated or in cases of disease recurrence, could be an alternative to excisional biopsy for the diagnosis of B-NHL. It should also be noted that while CNB limits the preservation of tissue architecture but allows IHC investigations, fine-needle aspirate biopsy (FNAB) does not allow the preservation of tissue architecture nor the application of IHC. However, several studies have investigated and demonstrated its validity when combined with supporting techniques such as FC, the results of which allow to achieve a high degree of diagnostic accuracy to be achieved (Amador Ortiz et al, 2011; Hu et al, 2013). Indeed, as mentioned above, FC provides a much greater amount of information than IHC. In addition, FC allows near perfect isolation of neoplastic populations as if they were a single entity, whereas in IHC the various cellular components appear mixed. However, the large amount of data obtained is difficult to analyze as a whole. Above all, it is particularly complex to correlate the diagnosis of a given type of B-NHL with the expression of each marker, or with the co-expression of multiple markers. Help comes from AI, which, through sophisticated reduction systems, quickly compares and stratifies all the elements obtained. Thus we obtained information comfortably manageable, through more easily interpretable 2D graphs. The potential of AI, when applied to a diverse set of cases involving mature B-NHL, has proven surprising. Previously, ML allowed the identification of distinct groups of tumors based on the expression of surface markers. By including intracellular markers in the phenotyping of tissue samples from patients with B-NHL, we achieved a more comprehensive analysis with the goal of developing an AI-based classification system that would determine which markers are most effective in distinguishing different groups of lymphomas.

First of all, we decide to express immunophenotype data using the percentage of positive cells for each marker used, regardless of the fluorescence intensity. The percentage is the only parameter insensitive to the change of instrument and or reagent, that were obvious in the long period of time in which the data were collected. Moreover, it is extremely easy, and fits with our intent to create an algorithm which could support decisions for future samples, possibly in different laboratories.
Afterwards, we had to decide which AI technique was more suitable with our purpose. After several trials with different techniques, we decided to use trees as classifiers: in fact, more advanced AI techniques such as convolutional neural networks did not improve significantly the accuracy; moreover, they are much more complex and heavyweight and, most importantly, did not allow us to understand how the algorithm was working. On the other

hand, a classification tree is a simple but extremely effective technique which clearly shows us the way it arrives to classify samples, allowing to compare the behavior of the algorithm to the human one, eventually confirming or suggesting improvements to our routine clinical practice.
A theoretical disadvantage of classification trees is that they define net cut-off values for markers: it is difficult to attribute any real meaning to these thresholds, and these values could potentially vary from one laboratory to another one, even if our database spans form 2003 to 2019, using different markers, different instruments and different operators thus recapitulating, at least partially, a multicenter experience. However, the fact that trees use thresholds and that a single marker is used along the tree with different thresholds in order to discriminate different groups of B-NHL has indeed a biological meaning: it confirms that a marker is not just "positive" or "negative", but rather the percentage of positive or negative cells is important to discriminate between different entities or group of entities. Moreover, it tells us that it could be incorrect to use fixed thresholds to define a marker as positive or negative because

- there could be different thresholds for the same marker, indicating different B-NHL
- thresholds should vary from one marker to another one.

In other words, the AI uses a threshold of a specific marker, defined as a percentage of positive cells, to split a population in two main categories (for example people with clear or dark eyes) and then it may use the same marker with a different threshold to split one (or both) category in two subcategories (for example people with green or blue eyes among the clear eyes category). Then the use of another marker allows AI to distinguish further categories (for example the eyes shape).
Finally, we had to deal with missing values: as our database spans from 2003 to 2019, different markers have been used according to the best knowledge and techniques available at that time. In addition, the use of markers is a reflection of daily clinical practice, so some markers may be under-used in the setting of suspicion of a particular disease. This aspect is particularly important for MIB1 and Bcl2, as these two markers are mostly utilized on samples other than peripheral blood or bone marrow aspirates, and for HCL and SL classes, where we had virtually only blood samples. This issue led us to generate multiple models, as described in the Results section.
These analyses allowed us not only to evaluate our data in the most honest way, but also to compare the performances with and without MIB1 and Bcl2, which are two markers that have been used more and more in our laboratory and are a source of satisfaction.

In Model I, the whole dataset was analyzed, but MIB1 and Bcl2 were removed. The I tree is characterized by a good overall accuracy (87,74%) and a very good per class accuracy, ranging from 94,28 to 99,46%. However, the best indicators are probably the positive predictive value (PPV) and negative predictive value (NPV) that range, respectively, from 58,33 to 100% and from 96,56 to 99,73%. Merging information from the confusion matrix and the percentage of accuracy, sensitivity and specificity, we can say that Model I, overall, has high specificity and NPV, while sensitivity and PPV are suboptimal for some B-NHL

subclasses: the algorithm seems to confuse LPL with MZL, which is quite expected, considering that these 2 subclasses are both indolent lymphomas with largely overlapping features. The best recognized class in terms of sensitivity is CLL, probably thanks to the very high number of available samples the algorithm could learn from.
Analyzing the tree, we can appreciate that it recapitulates the classical FC approach to B-NHL classification: CD5 is the root that divides the right part of the tree (CD5+ve), where there are mainly MCL and CLL, distinguished by CD200, from the left part (CD5-ve). The CD5-ve section of the tree is divided by CD10 in a middle part (CD5-ve CD10+ve), mainly occupied by DLBCL, FCL and BL, and in a left part (CD5-ve CD10-ve), mainly identifying MZL, SL, HCL and LPL. However, there are some considerations worthies to be noted: as mentioned above, CD5 and CD10, together with many other markers, are recurrent into the tree with different thresholds. In this sense, it would be more correct to say CD5> or < 32.704% (at least in this model), rather than CD5+ve or -ve. The same concept applies to all markers.

Figure 4 clearly shows that, apart from CD5, CD10 and CD200, one of the most important parameters to classify B-NHL is cell size, a variable underused as classifier in FC (Ichinohasama et al, 1997; Manocha et al, 2003), despite its common use as a gate (to overcome the problem of the intrinsic variability, due to different set up and different instruments, of the Forward Scatter as a measure of the cell volume we calculate for each case the ratio between the forward scatter of neoplastic B cells and residual T lymphocytes).
Some markers which are generally used in a particular setting can have a role in classifying other B-NHL: CD200 and CD23, for example, which are basically used to distinguish CLL from MCL, seem to be differentially expressed in other B-NHL, such as DLBCL or MZL.

In Model II, MIB1 and Bcl2 were included, but we had to remove HCL and SL from the analyses because samples from these two subclasses were not consistently tested for these two markers. The introduction of MIB1 and Bcl2 in the analysis significantly enhances the overall accuracy (92,68%), with major improvements in sensitivity, specificity, PPV and NPV of every subclass, which are all extremely high.
MIB1 is the most important marker in this algorithm (**Figure 3B**) and it represents the root of the tree. MIB1 identifies Ki-67, i.e., a nuclear antigen which is expressed by proliferating cells. Ki-67 is known to be differentially expressed in B-NHL, being highly expressed in aggressive B-NHL and to have a strong prognostic impact.
Model II surely outperforms Model I, but the starting dataset is not the same: to evaluate the net benefit of MIB1 and Bcl2, we have to compare Model II with Model III where MIB1 and Bcl2 were removed from the analysis. Although the benefit in terms of overall accuracy seems to be modest (92,68% in Model II vs 89,3% in Model III), we can appreciate that in the subclasses' analyses the sensitivity and PPV drop when MIB1 and Bcl2 are removed. In summary, Model II outperforms Model III, whose tree and characteristics are very similar to Model I.
Finally, to demonstrate that the importance of MIB1 and Bcl2 was not an artefact due to the filling methodology utilized, we analyzed non-blood samples only, which are for the most part characterized with these 2 markers (Model IV). This resulted in a significant reduction of

available samples, from 1465 to 548, however, the overall accuracy is still 86.86%, with very good data on sensitivity and PPV; more importantly, MIB1 is the second most important marker and Bcl2 is still between the most powerful and discriminatory markers, just like in Model II (**Figure 3D**). Considering the different starting datasets (Model II is much more enriched in CLL compared to Model IV), these two models have very similar trees.
This clearly demonstrates that the importance of MIB1 and Bcl2 is real, excluding any artefactual effect.
Given the data obtained on tissue samples, we conducted further analysis on a larger group of tissue samples, obtained between 2009 and 2022, in which additional intracellular markers, including BCL6, IRF4/MUM1, and MNDA, were added to the panel.
The criteria for selecting the markers to be used in the analysis, the method of filling in missing data, and the expression of the data as a percentage of positive cells for the specific marker, were the same as those used in the previously described analysis.
On the contrary, we decided to follow an alternative analytical strategy. Instead of grouping the cases in four different dataset (Models I to IV), we maintained a single dataset of cases and used three different analytical approaches.
The first method we applied was PPSscore analysis. This is an analytical method used for feature selection in machine learning that can identify those features (in this case, markers) that have the greatest predictive power in relation to a target variable (in this case, diagnosis).
The results of this approach were particularly interesting: among the top five markers with the greatest predictive power, three were intracellular markers (IRF4, which was found to be the main one, BCL6 and MIB1) one was cell size (GC) and finally an unconventional surface marker (CD305).
It is interesting to note that IRF4 and BCL6 are among the main markers used in IHC for defining the two main subcategories of DLBCL (follicular center type and activated type), as well as for differentiating different types of both indolent (e.g., FL) and aggressive (e.g., HGBL) lymphomas. The predictive power of MIB1 has been discussed previously, and it is therefore predictable that it will also be present in this analytical model.
The cell size parameter has also been previously discussed. It is not surprising that this parameter is among those with the greatest predictive power, infact cell size represent a fundamental criterion in histological analysis for the classification of lymphomas.
The presence among the best predictive markers of CD305 is less easily explained both due to its scarce application in clinical FC and the relatively limited literature (van Dongen et al., 2012). CD305 is considered a very useful marker in the diagnosis of HCL (Garnache Ottou et al., 2014), but is poorly described in other types of lymphoma. In our series, in which HCLs are not present, CD305 is highly expressed both in MCL and in a subset of DLBCL, furthermore it has a very wide distribution pattern in SLL/CLL (**Figure 7**).
Further studies are needed to better understand the precise role of CD305 in the context of B-NHL, nevertheless, it remains to underline the ability of AI models to identify unsuspected feature with clinical impact.
As a second analytical approach we once again employed the tree classification model. The markers identified by the PPS score were again both as the root node (IRF4) and among the main intermediate nodes. Unlike the PPS score, in the tree classification model the CD10

takes on a more significant role and is placed immediately below the root node in the classification system. The differences between the two models are however relatively limited and, also in this model, CD305 plays a fundamental role in differentiating the different types of B-NHL.
Finally, the third analytical approach was the analysis with the dimensional reduction model UMAP. Unlike the other two analytical systems, UMAP, rather than identifying markers with an intrinsic predictive value, creates maps in which the weight of each marker weighed in relation to that of all the other markers generates clusters.
From our analysis conducted both using only the ten markers identified by the PPS score and using all the markers, it clearly emerges that through flow cytometric analysis it is possible to identify the main groups of lymphomas with high accuracy. Furthermore, the positioning of certain lymphoma groups in a specific space of the UMAP map is not random, but rather determined by the intrinsic characteristics of that type of lymphoma.
As an example of this, we can observe how the distribution of DLBCLs can be defined in two main groups, one "close" to FLs and one "close" to MZLs, probably recapitulating the heterogeneity of DLBCLs. Finally, the "distance" between SLL/CLL and FL appreciable in all maps can be interpreted as the "distance" in terms of the normal counterpart of the two types of lymphoma.
In conclusion, our results demonstrate that the combination of surface and intracellular markers allows accurate categorization of the major B-NHL groups, with only ten markers proving sufficient for effective classification. Furthermore, the use of a broader set of markers, including intracellular and unconventional markers, improved the ability of UMAP to discriminate between different entities.

Ultimately, this study suggests that the integration of AI with multiparametric FC could significantly improve the diagnostic process in B-NHL, complementing traditional histopathologic examinations. Further research is needed to determine whether this approach can predict categories of malignancies with specific molecular or genetic alterations, which may have implications for improved classification and therapeutic decision-making in B-NHL.

With this study, we hope to have started a big job, aimed at helping to achieve a more accurate and standardized method in the diagnosis and search for cures of the most common types of B-NHL. Nevertheless, it must be emphasized that although AI holds great promise in B-NHL diagnosis, there are several pitfalls and challenges that need to be addressed:

- AI models heavily rely on data quality. Biased or incomplete datasets can lead to prejudiced or inaccurate results. Ensuring that the training data is representative of diverse patient populations is essential to avoid biased outcomes.
- AI algorithms need rigorous clinical validation to prove their effectiveness and safety. This process can be time-consuming and expensive, and results may not always align with initial expectations.
- AI models can sometimes become overly tailored to the training data, resulting in poor generalization to new and unseen cases. Proper model validation and evaluation are essential to prevent overfitting.

- Many AI algorithms provide results without clear explanations.

In some cases, there may be a scarcity of data for rare subtypes or conditions of B-NHL. AI models may struggle to provide reliable diagnoses in such scenarios.

CONCLUSIONS

AI has the potential to revolutionize the diagnosis and management of B-NHL in several perspectives.

AI Algorithms can analyze vast amounts of patient data, including medical images, genomic data, and clinical records, with incredible precision. This is certainly a useful strategy for researchers to discover new insights into B-NHL biology and potential therapeutic strategies.

Meanwhile, powered predictive models can identify patterns and risk factors associated with B-NHL. This could aid in the early detection of the disease, allowing for prompt intervention and potentially better treatment outcomes. Furthermore, this can speed up the diagnostic process and reduce the chances of human error.

Not to mention, the ability to reduce large amounts of data to easily interpretable patterns enables the integration of data from multiple sources. This holistic view of patient data can provide a more complete understanding of the disease and its progression. Anyway, exploring the correlation of specific markers with prognosis, AI-driven monitoring systems can track patients' responses to treatment in real-time.

Furthermore, unraveling the specific relationship between certain markers and precise diagnoses, AI could also help create personalized treatment plans based on a patient's specific subtype of B-NHL, genetic makeup, and response to therapy. This personalized approach has the potential to result in treatments that are both more efficient and associated with fewer side effects. Therefore, AI can accelerate drug discovery by identifying potential therapeutic targets and predicting the effectiveness of existing drugs for B-NHL treatment. This could lead to the development of new and more targeted therapies.

Definitively, by streamlining diagnostic processes and reducing the need for extensive manual labor, AI has the potential to lower healthcare costs associated with B-NHL diagnosis and treatment.

However, it's essential to consider the challenges of implementing AI in B-NHL diagnosis, such as data privacy, regulatory approvals, and the need for ongoing validation and training of AI models. Additionally, AI should complement rather than replace human expertise, with healthcare professionals playing a critical role in interpreting AI-generated insights and making clinical decisions.

REFERENCES

- Alaggio R., Amador C., Anagnostopoulos I. et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia* 36:1720–1748 (2022).
- Alsuwaidan A., Pirruccello E., Jaso J. *et al.* Bright CD38 expression by flow cytometric analysis is a biomarker for double/triple hit lymphomas with a moderate sensitivity and high specificity. *Cytometry B* 96(5), 368–374 (2019).
- Amador-Ortiz C., Chen .L, Hassan A. *et al.* Combined core needle biop- sy and fine-needle aspiration with ancillary studies correlate highly with traditional techniques in the diagnosis of nodal-based lym- phoma. *Am J Clin Pathol.* 135:516-243(2011).
- Angeletti C. A method for the interpretation of flow cytometry data using genetic algorithms. *J. Pathol. Inform.* 9:16 (2018).
- Arcaini L., Paulli M., Boveri E. *et al*. Marginal zone-related neoplasms of splenic and nodal origin. *Haematologica* 88:80– 93 (2003).
- Barna G., Reiniger L., Tátrai P., Kopper L., Matolcsy A. The cut-off levels of CD23 expression in the differential diagnosis of MCL and CLL. *Hematol. Oncol.* 26(3):167–170 (2008).
- Barrena S., Almeida J., Del Carmen García-Macias M. *et al.* Flow cytometry immunophenotyping of fine-needle aspiration specimens: Utility in the diagnosis and classification of non-Hodgkin lymphomas. *Histopathology* 58:906–918 (2011).
- Bayramoglu N., Kannala J., Heikkilä J. Deep learning for magnification independent breast cancer histopathology image classification. *In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico*, 2440–2445 (2016).
- Becht E., McInnes L., Healy J. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37:38–44 (2019).
- Bengio Y., Grandvalet Y. No unbiased estimator of the variance of k-fold cross-validation. *J. Mach. Learn. Res.* 5: 1089–1105 (2004).
- Bomben R., Rossi F.M., Vit F. *et al.* Clinical impact of TP53 disruption in chronic lymphocytic leukemia patients treated with ibrutinib: a campus CLL study*Leukemia* 37:914–918 (2023)
- Boxhammer R., Striebel F., Baumgartner R., Endell J. Expression of CD19 Antigen on Chronic Lymphocytic Leukemia Cells after Tafasitamab (Anti-CD19) Treatment: Phase I Trial Data. *Blood* 134(1):5061 (2019).
- Breiman L. Some properties of splitting criteria. *Mach. Learn.* 24: 41–47 (1996).
- Breiman L., Friedman, J., Stone C.J., Olshen, R.A. Classification and Regression Trees. *CRC Press* (1984).
- Cai W., Zeng Q., Zhang X., Ruan W. Trends Analysis of Non-Hodgkin Lymphoma at the National, Regional, and Global Level, 1990–2019: Results From the Global Burden of Disease Study 2019. *Front. Med.* 8:738693 (2021).
- Campo E., Swerdlow S.H., Harris N.L. *et al*. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood* 117:5019-32 (2011).

- Cheson B.D., Fisher R.I., Barrington S.F. *et al*. Recommendations for ini- tial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol.* 32:3059-68 (2014).
- Cheung M., Campbell J.J., Whitby L. *et al.* Current trends in flow cytometry automated data analysis software. *Cytometry A* 99(10):1007-1021 (2021).
- Chiappella A., Santambrogio E, Castellino A. *et al.* Integrating novel drugs to chemoimmunotherapy in diffuse large B-cell lymphoma. *Expert Rev. Hematol.* 10(8):697-705 (2017).
- Demurtas A., Accinelli G., Pacchioni D *et al.* Utility of flow cytometry immunophenotyping in fine-needle aspirate cytologic diagnosis of non-Hodgkin lymphoma: A series of 252 cases and review of the literature. *Appl. Immunohistochem. Mol. Morphol.* 18:311–322 (2010).
- Denlinger N, Bond D, Jaglowski S. CAR T-cell therapy for B-cell lymphoma. *Curr Probl Cancer.* 46(1):100826 (2022).
- Devin J., Kassambara A., Bruyer A. *et al.* Phenotypic Characterization of Diffuse Large B-Cell Lymphoma Cells and Prognostic Impact. *J. Clin. Med.* 8(7):1074 (2019).
- Dubois K., Tannoury M., Bauvois B., *et al.* Extracellular Vesicles in Chronic Lymphocytic Leukemia: Tumor Microenvironment Messengers as a Basis for New Targeted Therapies? *Cancers* 15(8):2307 (2023).
- Espinet B., Ferrer A., Bellosillo B. *et al*. Distinction between asymptomatic monoclonal B-cell lymphocytosis with cyclin D1 overexpression and mantle cell lymphoma: from molecular profiling to flow cytometry. *Clin. Cancer Res*. 15;20(4):1007-1019 (2014).
- Freeman, C.L. and Sehn, L.H. A tale of two antibodies: obinutuzumab versus rituximab. *Br. J. Haematol*. 182:29-45 (2018).
- Gaidano V., Tenace V., Santoro N. *et al*. A Clinically Applicable Approach to the Classification of B-Cell Non-Hodgkin Lymphomas with Flow Cytometry and Machine Learning. *Cancers* 12:1684-1701 (2020).
- Garnache Ottou F., Chandesris M.-O., Lhermitte L. *et al.* Peripheral blood 8 colour flow cytometry monitoring of hairy cell leukaemia allows detection of high-risk patients. *Br. J. Haematol.* 166:50-59 (2014)
- Gong J.Z., Lagoo A.S., Peters D. *et al.* Value of CD23 Determination by Flow Cytometry in Differentiating Mantle Cell Lymphoma From Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma, *Am. J. Clin. Pathol.*116(6): 893–897 (2001)
- Hu Q., Naushad H., Xie Q. *et al.* Needle-core biopsy in the pathologic diagnosis of malignant lymphoma showing high reproducibility among pathologists. *Am J Clin Pathol.* 140: 238-47 (2013).
- Hunter Z.R., Branagan A.R., Manning R. *et al.* CD5, CD10, and CD23 expression in Waldenstrom's macroglobulinemia. *Clin. Lymphoma* 5(4):246-9 (2005).

- Ichinohasama R., DeCoteau J.F., Myers J. *et al*. Three-color flow cytometry in the diagnosis of malignant lymphoma based on the comparative cell morphology of lymphoma cells and reactive lymphocytes. *Leukemia*. 11(11):1891-903 (1997).
- Inamdar AA, Goy A, Ayoub NM, Attia C, Oton L, Taruvai V, Costales M, Lin YT, Pecora A, Suh KS. Mantle cell lymphoma in the era of precision medicine-diagnosis, biomarkers and therapeutic agents. *Oncotarget* 7(30):48692-48731 (2016).
- Jaffe E.S., Harris N., Stein H., Vardiman J.W. World Health Organization classification of Tumours. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. 3rd ed. Lyon: IARC (2001).
- Jiang M., Bennaniband N.N., Feldman A.L. Lymphoma classification update: B-cell non-Hodgkin lymphomas. *Expert Rev. Hematol.* 10(5):405–415 (2017).
- Khanlari M., Chapman J.R. Follicular lymphoma: updates for pathologists. *J Pathol Transl Med.* 56(1):1-15 (2022).
- Kim H., Kim H.J., Kim S.H. Diagnostic Approach for Double-Hit and Triple-Hit Lymphoma Based on Immunophenotypic and Cytogenetic Characteristics of Bone Marrow Specimens. *Ann Lab Med.* 40(5):361-369 (2020).
- Ko B.S., Wang Y.F., Li J.L. *et al.* Clinically validated machine learning algorithm for detecting residual diseases with multicolor flow cytometry analysis in acute myeloid leukemia and myelodysplastic syndrome. *EBioMedicine* 37:91–100 (2018).
- Kohli M., Prevedello L.M., Filice R.W., Geis J.R. Implementing machine learning in radiology practice and research. *Am. J. Roentgenol*. 208:754–760 (2017).
- Konoplev S., Medeiros L.J., Bueso-Ramos C.E. *et al.* Immunophenotypic profile of lymphoplasmacytic lymphoma/Waldenström macroglobulinemia. *Am J Clin Pathol.* 124(3):414–420 (2005).
- Kost C.B., Holden J.T., Mann K.P. Marginal zone B-Cell lymphoma: A retrospective immunophenotypic analysis. *Cytometry B* 74B:282–286 (2008).
- Kumar A., Eyre T.A., Lewis K.L. *et al.* New Directions for Mantle Cell Lymphoma in 2022 American Society of Clinical *Oncology Educational Book 2022* 42:614-628 (2022).
- Lee L. W., Ching T., Howie B. *et al.* Assessment of Immunoglobulin Heavy Chain Variable Region (IGHV) Mutation Status with a Next-Generation Sequencing (NGS) Immunosequencing Assay for Measurable Residual Disease (MRD). *Blood* 138(1):4676-4677 (2021).
- Levine J.H., Simonds E.F., Bendall S.C, *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell.* 162(1):184-97 (2015).
- Li C., Xue D., Hu Z. *et al*. A Survey for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks. *In Proceedings of the International Conference on Information Technologies in Biomedicine, Da Nang, Vietnam,* 222–233 (2019).
- López C., Burkhardt B., Chan J.K.C. *et al.* Burkitt lymphoma. *Nat. Rev. Dis. Primers* 8,78 (2022).
- Maji D., Santara A., Ghosh S. *et al*. Deep neural network and random forest hybrid architecture for learning to detect retinal vessels in fundus images. *In Proceedings of*

the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milano, Italy, 3029–3032 (2015).

- Mandelker D. L., Dorfman D. M., Li B. *et al.* Antigen expression patterns of MYC-rearranged versus non-MYC-rearranged B-cell lymphomas by flow cytometry. *Leukemia & Lymphoma*, 55(11), 2592–2596 (2014).
- Manocha S., Matrai Z., Osthoff M. *et al.* Correlation between cell size and CD38 expression in chronic lymphocytic leukaemia. *Leuk Lymphoma* 44(5):797-800 (2003).
- Mantei K., Wood B.L. Flow cytometric evaluation of CD38 expression assists in distinguishing follicular hyperplasia from follicular lymphoma. *Cytometry B* 76:315–20 (2009).
- McInnes L., Healy J., Melville J. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. *arXiv:1802.03426* (2020) .
- McKinney W. Pandas: A foundational Python library for data analysis and statistics. *Python High Perform.Sci. Comput.* 14:1–9 (2011).
- Miotto R., Li L., Kidd B.A., Dudley J.T. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Sci. Rep*. 6:1–10 (2016).
- Morice W..G, Chen D., Kurtin P.J. *et al*. Novel immunophenotypic features of marrow lymphoplasmacytic lymphoma and correlation with Waldenström's macroglobulinemia. *Mod Pathol*. 22(6):807-16 (2009).
- Naderi N, Yang DT. Lymphoplasmacytic lymphoma and Waldenström macroglobulinemia. *Arch. Pathol. Lab. Med*. 137(4):580-585 (2013).
- Neyman, J. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Breakthroughs in Statistics. Springer*; 123–150 (1992).
- O'Neill K., Aghaeepour N., Špidlen J., Brinkman R. Flow cytometry bioinformatics. *PLoS Comput. Biol.* 9(12): e1003365. (2013)
- Olszewski A.J., Kurt H., Evens A.M. Defining and treating high-grade B-cell lymphoma, NOS. *Blood* 140(9): 943–954 (2022).
- Papageorgiou S.G., Thomopoulos T.P., Liaskas A., Vassilakopoulos T.P. Monoclonal Antibodies in the Treatment of Diffuse Large B-Cell Lymphoma: Moving beyond Rituximab. *Cancers* 14(8):1917 (2022).
- Pedregosa F., Varoquaux G., Gramfort A. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12: 2825–2830 (2011).
- Polikowsky H.G., Drake K.A. Supervised Machine Learning with CITRUS for Single Cell Biomarker Discovery. *Methods Mol. Biol.* 1989:309-332 (2019).
- Quinlan J.R. Induction of decision trees. *Mach. Learn.* 1: 81–106 (1986).
- Rahim A., Meskas J., Drissler S. *et al*. High throughput automated analysis of big flow cytometry data. *Methods* 134:164–176 (2018).
- Rymkiewicz G., Grygalewicz B., Chechlinska M. *et al.* A comprehensive flow-cytometry-based immunophenotypic characterization of Burkitt-like lymphoma with 11q aberration. *Mod Pathol.* 31(5):732-743 (2018).

- Salem DA, Stetler-Stevenson M. Clinical Flow-Cytometric Testing in Chronic Lymphocytic Leukemia. *Methods Mol Biol.* 2032:311-321 (2019).
- Schnaiter A., Stilgenbauer S. Refractory chronic lymphocytic leukemia: new therapeutic strategies. *Oncotarget.* 1(7):472-82 (2010).
- Schroers R., Griesinger F., Trümper L. *et al.* Combined analysis of ZAP-70 and CD38 expression as a predictor of disease progression in B-cell chronic lymphocytic leukemia. *Leukemia* 19(5):750-758 (2005).
- Sethi S., Epstein-Peterson Z., Kumar A. and Ho C. Current Knowledge in Genetics, Molecular Diagnostic Tools, and Treatments for Mantle Cell Lymphomas. *Front. Oncol.* 11:739441 (2021).
- Stashenko P., Nadler L.M., Hardy R., Schlossman S.F. Characterization of a human B lymphocyte-specific antigen. *J. Immunol.*125, 1678– 1685 (1980).
- Swerdlow S.H., Campo E., Harris N.L. *et al.* World Health Organization classification of Tumours of Haematopoietic and Lymphoid Tissues. Revised 4th ed. Lyon: IARC (2017).
- Tam C.S., Otero-Palacios J., Abruzzo L.V. *et al.* Chronic lymphocytic leukaemia CD20 expression is dependent on the genetic subtype: a study of quantitative flow cytometry and fluorescent in-situ hybridization in 510 patients. *Br. J. Haematol.* 141(1), 36-40 (2008).
- Taye M.M. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers* 12:91-116 (2023).
- Tsagarakis N.J., Papadhimitriou S.I., Pavlidis D. *et al.* Contribution of immunophenotype to the investigation and differential diagnosis of Burkitt lymphoma, double-hit high-grade B-cell lymphoma, and single-hit *MYC*-rearranged diffuse large B-cell lymphoma. *Cytometry B* 98(5):412-420 (2020)
- van Gassen S., Callebaut B., Van Helden M.J. *et al.* FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry A* 87(7):636-45(2015).
- van Oostrum M., Müller M., Klein F. *et al*. Classification of mouse B cell types using surfaceome proteotype maps. *Nat. Commun.* 10:5734 (2019).
- van Dongen J.J., Lhermitte L., Bottcher S. *et al.* EuroFlow antibody panels for standardized *n*-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia* 26(9): 1908–1975 (2012).
- Wang S., Summers R.M. Machine learning and radiology. *Med. Image Anal.* 16:933–951 (2012).
- Wetschoreck F. RIP correlation. Introducing the Predictive Power Score. *https://towardsdatascience.com/rip-correlation-introducing-the-predictive-power-score-3d90808b9598* (2020).
- Wu D., Wood B.L., Dorer R., Fromm J. R. "Double-hit" mature B-cell lymphomas show a common immunophenotype by flow cytometry that includes decreased CD20 expression. *Am J Clin Pathol.* 134(2), 258–265 (2010).
- Xiao D., Yu S., Vignarajan J. *et al*. Retinal hemorrhage detection by rule-based and machine learning approach. In *Proceedings of the 2017 39th Annual International*

Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 660–663 (2017).

- Yoshino T., Tanaka T., Sato Y. Differential diagnosis of chronic lymphocytic leukemia/small lymphocytic lymphoma and other indolent lymphomas, including mantle cell lymphoma. *J Clin Exp Hematop*. 60(4):124-129 (2020).
- Zhang S., Kipps T.J. The pathogenesis of chronic lymphocytic leukemia. *Annu Rev Pathol.* 9:103-18 (2014).
- Zucchetto A., Cattarossi I., Nanni P. *et al.* Cluster analysis of immunophenotypic data: the example of chronic lymphocytic leukemia. *Immunol. Lett.* 134(2):137-144 (2011).

ACKNOWLEDGEMENTS

At the end of this challenging yet exciting period of study, I would like to thank all the people who have supported me and made it possible for me to achieve such an important goal.

First of all, I would like to thank the many people associated with the University of Urbino, especially my supervisor Stefano Papa, who believed in me and gave me the opportunity to enter the Ph.D. program in Biomolecular and Health Sciences; Professor Barbara Canonico, Professor Elena Barbieri and the Director Professor Marco Rocchi, who were always available to advise me along the way; the secretarial staff and the staff of the International College, who made me feel at home even though I was far from home.

Thinking back to my time in Urbino, I would like to address affectionate thoughts to Chiara Barattini, a formidable companion of study and adventure, and to my friends Rosa Fioravante and Elisa Ippoliti, who, despite their humanistic studies, never missed a single one of my reports.

Meanwhile, at home, I could count on the support of my husband Claudio, patient and understanding, who never let me lack his material, moral and logistical support; my children Luca and Annalisa, proud of my choice, but never as proud as I am of them; and my parents and sister, who followed my progress in my studies from afar with trepidation. I will never forget my father's offer, at the beginning of my doctoral journey, to show himself as a living example of the importance of research in the field of hematology, having been struck by B-NHL twenty years ago, which he managed to overcome brilliantly.

But as far as the conception and realization of my whole research project is concerned, I would like to thank the Laboratory of Immunopathology of the Mauriziano Hospital in Turin, where I made my first steps in the field of FC so many years ago, and where I was welcomed again with open arms to continue this work. Working with us were hematologist Valentina Gaidano, for case collection, engineer Valerio Tenace, for data processing with AI, and let's not forget the valuable advice of expert researcher Francesca De Bacco.

Above all, from a scientific and professional point of view, I would be a nobody if I had not had the good fortune to meet my mentor, my co-tutor Massimo Geuna, to whom I owe the greatest thanks. From the very beginning of my career, which he made possible by offering me my first fellowship, he appreciated my sincere enthusiasm for research and guided me by sharing his deep knowledge of cytometry, his rigor in performing analyses, his curiosity to go beyond what appears, and his unwavering will to reach new heights. He taught me and then followed me, corrected me, supported me, encouraged me, advised me, and helped me all these years. I am honored to introduce myself as one of his students. It has always been my dream to work in the field of oncology research, ever since high school: if I have been able to land there, and if I have come this far, I can say that much of the credit is due to him.