BioMedInformatics

MDPI

Article

Anomaly Detection and Artificial Intelligence Identified the Pathogenic Role of Apoptosis and RELB Proto-Oncogene, NF-kB Subunit in Diffuse Large B-Cell Lymphoma

Joaquim Carreras [1,*] and Rifat Hamoudi [2,3,4,5]

[1] Department of Pathology, School of Medicine, Tokai University, 143 Shimokasuya, Isehara 259-1193, Japan
[2] Department of Clinical Sciences, College of Medicine, University of Sharjah, Sharjah P.O. Box 27272, United Arab Emirates; rhamoudi@sharjah.ac.ae
[3] Division of Surgery and Interventional Science, University College London, London NW3 2PF, UK
[4] ASPIRE Precision Medicine Research Institute Abu Dhabi, University of Sharjah, Sharjah P.O. Box 27272, United Arab Emirates
[5] BIMAI-Lab, Biomedically Informed Artificial Intelligence Laboratory, University of Sharjah, Sharjah P.O. Box 27272, United Arab Emirates
* Correspondence: joaquim.carreras@tokai.ac.jp; Tel.: +81-463-93-1121

Abstract: Background: Diffuse large B-cell lymphoma (DLBCL) is one of the most frequent lymphomas. DLBCL is phenotypically, genetically, and clinically heterogeneous. Aim: We aim to identify new prognostic markers. Methods: We performed anomaly detection analysis, other artificial intelligence techniques, and conventional statistics using gene expression data of 414 patients from the Lymphoma/Leukemia Molecular Profiling Project (GSE10846), and immunohistochemistry in 10 reactive tonsils and 30 DLBCL cases. Results: First, an unsupervised anomaly detection analysis pinpointed outliers (anomalies) in the series, and 12 genes were identified: *DPM2*, *TRAPPC1*, *HYAL2*, *TRIM35*, *NUDT18*, *TMEM219*, *CHCHD10*, *IGFBP7*, *LAMTOR2*, *ZNF688*, *UBL7*, and *RELB*, which belonged to the apoptosis, MAPK, MTOR, and NF-kB pathways. Second, these 12 genes were used to predict overall survival using machine learning, artificial neural networks, and conventional statistics. In a multivariate Cox regression analysis, high expressions of *HYAL2* and *UBL7* were correlated with poor overall survival, whereas *TRAPPC1*, *IGFBP7*, and *RELB* were correlated with good overall survival ($p < 0.01$). As a single marker and only in RCHOP-like treated cases, the prognostic value of *RELB* was confirmed using GSEA analysis and Kaplan–Meier with log-rank test and validated in the TCGA and GSE57611 datasets. Anomaly detection analysis was successfully tested in the GSE31312 and GSE117556 datasets. Using immunohistochemistry, RELB was positive in B-lymphocytes and macrophage/dendritic-like cells, and correlation with HLA DP-DR, SIRPA, CD85A (LILRB3), PD-L1, MARCO, and TOX was explored. Conclusions: Anomaly detection and other bioinformatic techniques successfully predicted the prognosis of DLBCL, and high *RELB* was associated with a favorable prognosis.

Keywords: anomaly detection; gene expression; diffuse large B-cell lymphoma; artificial intelligence; machine learning; artificial neural networks; prognosis; immuno-oncology; RELB; macrophages

Citation: Carreras, J.; Hamoudi, R. Anomaly Detection and Artificial Intelligence Identified the Pathogenic Role of Apoptosis and RELB Proto-Oncogene, NF-kB Subunit in Diffuse Large B-Cell Lymphoma. *BioMedInformatics* **2024**, *4*, 1480–1505. https://doi.org/10.3390/biomedinformatics4020081

Academic Editors: Ognjen Arandjelović and Burghardt Wittig

Received: 27 February 2024
Revised: 29 April 2024
Accepted: 31 May 2024
Published: 7 June 2024

1. Introduction

1.1. Clinicopathological Characteristics and Prognosis of Diffuse Large B-Cell Lymphoma

This study aimed to identify new prognostic markers of diffuse large B-cell lymphoma (DLBCL) using anomaly detection analysis. By identifying outlier cases, the genes associated with those unusual cases were identified, and their prognostic value was assessed.

The classification of hematologic malignancies integrates data from several sources, including pathologic characteristics, pathophysiology, treatment, and outcomes. The current classification is the World Health Organization (WHO) revised 4th edition (WHO4R) [1],

which has recently been updated into the International Consensus Classification 2022 (ICC2022) [1–4], and the proposed *5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms* (WHO5) [5]. In this classification, mature B-cell neoplasms are hematological cancers originating from lymphocytes with a lymphocyte subtype or cell lineage of B cells.

These neoplasms are classified according to several parameters, such as morphological characteristics, architectural distribution of the neoplastic cells, immunophenotypic markers, genetic alterations, and clinical features of the patients [6–13]. They are classified into different subtypes based in part on the postulated cell of origin.

DLBCL is one of the most frequent histological subtypes of hematological neoplasia, accounting for approximately 25–30% of non-Hodgkin lymphomas.

The incidence of DLBCL in the United States and the United Kingdom is approximately 7 cases per 100,000 people per year. In Europe, there are 5 cases per 100,000 people per year [14–16]. Interestingly, the incidence differs according to ethnicity. White Americans have a higher incidence than Blacks, Asians, and Native Americans [14,15,17].

The diagnostic criteria of DLBCL are heterogeneous and include several subtypes such as T cell/histiocyte-rich large B cell lymphoma, the primary DLBCL of the mediastinum, intravascular large B cell lymphoma, lymphomatoid granulomatosis, the primary DLBCL of the central nervous system, the primary cutaneous DLBCL leg type, DLBCL associated with chronic inflammation, and Epstein–Barr virus-positive (EBER)-positive DLBCL. In the WHO classification, other categories are included [1,2,5], which have features of overlap between DLBCL and other subtypes (Burkitt lymphoma), such as high-grade B-cell lymphoma with *MYC* and *BCL2* and/or *BCL6* rearrangements and high-grade B-cell lymphoma not otherwise specified [18–32].

DLBCL originates from mature B cells that have the histological appearance of centroblast or immunoblasts, which are two types of activated B cells. The histological appearance of DLBCL is variable because of the heterogeneity of the morphological characteristics of the neoplastic B lymphocytes and the tumor-immune microenvironment. This heterogeneity is shown in Figure 1.

Clinically, most patients present with a rapidly growing mass located in the lymph nodes or abdomen. In approximately 60% of cases, the disease will present as an advanced stage. Two subtypes of DLBCL have been identified on the basis of gene expression and the postulated cell-of-origin: the germinal center B cell type (GCB) and the activated B cell type (ABC) [2,33–35].

Predicting clinical evolution is currently performed using the International Prognostic Index (IPI) with its variants, the revised IPI, and the National Comprehensive Cancer Network (NCCN)-IPI [33–35]. The IPI uses as unfavorable predictors an age > 60 years, serum lactate dehydrogenase concentration above normal, ECOG performance status ≥ 2, Ann Arbor stage III or IV, and number of extranodal disease sites > 1 [34]. Gene expression profiling also stratifies patients into two prognostic groups, with the activated B cell type associated with a poorer prognosis. Integration with other genetic factors, such as the presence of *BCL2*, *MYC*, and *BCL6* translocations; copy number changes and LOH; and mutational profiling, has allowed the identification of different genetic subtypes (MCD, BN2, EZB, ST2, A53, and N1) [36]. Interestingly, these subtypes also showed different gene signatures, including malignant processes (proliferation signature and *MYC*, ribosomal proteins, glycolipid pathways), B cell differentiation, transcription factors (*IRF4*, *BCL6*, *OCT2*, and *TCF3*), oncogenic signaling (*NFKB*, *p53*, *NOTCH*, *PI3K*, and *JAK2*), and immune microenvironment (T follicular helper cells, CD4 T helper cells, CD8 cytotoxic T lymphocytes, regulatory T lymphocytes, natural killer cells, macrophages, dendritic cells, and fibrosis) [36].

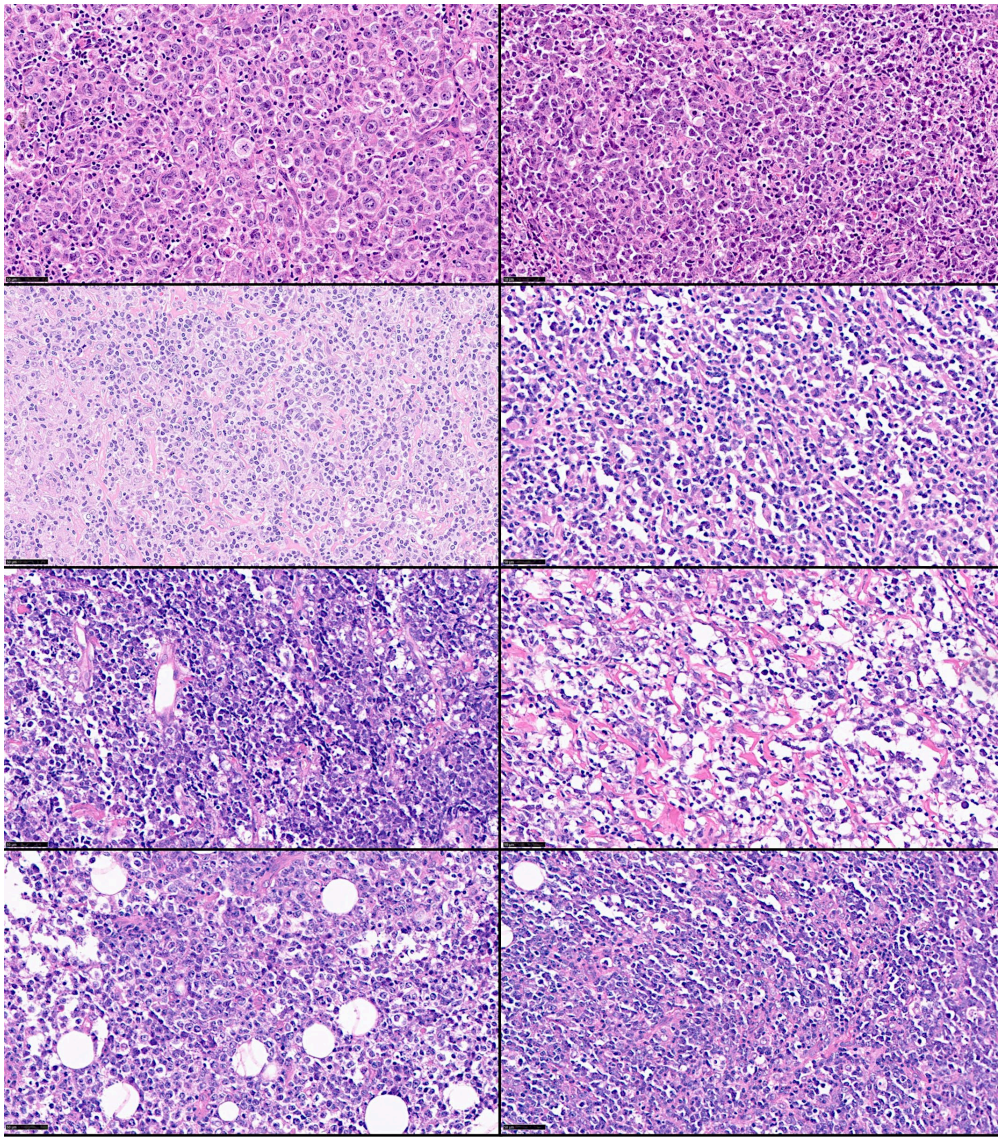

Figure 1. Histological heterogeneity of DLBCL. Despite the fact that DLBCL is a unique lymphoma subtype, its morphological characteristics are heterogeneous, including the neoplastic B lymphocytes and variable content of the tumor immune microenvironment. Hematoxylin and eosin stain (scale bar = 50 μm). The histological cases were retrieved from the lymphoma database of the Department of Pathology, Tokai University, School of Medicine.

1.2. Machine Learning and Anomaly Detection

1.2.1. Machine Learning

Machine learning can be defined as an analytic method that uses data and algorithms to emulate human learning and gradually improve accuracy [37]. It is a branch of artificial intelligence (AI) that uses statistical methods and algorithms to make classifications and predictions [37]. Neural networks are a subfield of machine learning, and deep learning is a subfield of neural networks [38].

A machine learning algorithm has three components: the decision process algorithms make predictions or classifications; the error function evaluates the prediction of the model; and the model optimization process adjusts the weights autonomously to improve the performance of the model [38].

There are three main types of learning: supervised, unsupervised, and reinforcement learning. Supervised learning uses labeled datasets to make classifications, predictions, and regression. Unsupervised learning uses unlabeled datasets to identify not readily apparent patterns and classify cases [39]. Reinforcement learning is an area of machine learning that handles sequential decision-making problems in a situation of uncertainty. Reinforcement learning learns to optimize sequential decisions by finding the best strategy [40].

Machine learning is an area of artificial intelligence that fits mathematical models to observed data. Machine learning can be broadly divided into supervised learning, unsupervised learning, and reinforcement learning (Figure 2). Deep neural networks contribute to each of these areas. The type of analysis to be performed depends on the type of data and the aim of the study [41].

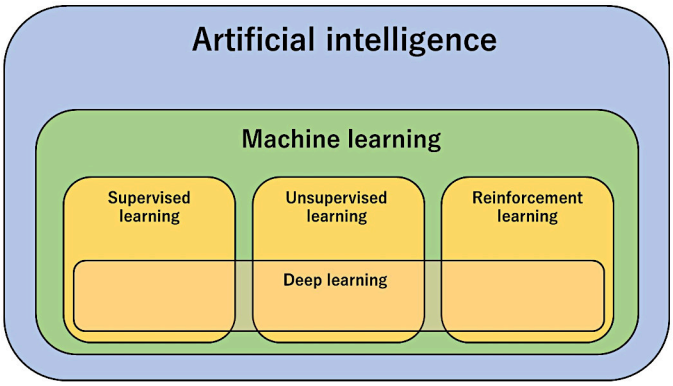

Figure 2. Types of artificial intelligence methods.

In this study, anomaly detection was used to identify anomalies (rare events) in the dataset. A model was constructed from the input data (gene expression) without corresponding labels (i.e., "no supervision"). Rather than learning a mapping from input to output, the goal is to describe or understand the structure of the data. Subsequently, supervised learning was used to predict the overall survival outcome (dead vs. alive).

Different types of machine learning methods, including supervised, unsupervised, and reinforcement learning, are shown in Figure 3.

1.2.2. Segmentation Analysis

Segmentation is the technique of splitting cases into different groups depending on their characteristics. There are several segmentation methods, such as K-Means, Kohonen, TwoSteps cluster, TwoStep-AS, and Anomaly detection.

K-Means is a type of clustering analysis that is unsupervised because there is no definition of the target variable (field). The dataset is clustered into different groups to search for patterns in the input data. Within a cluster, the cases are similar to each other, but the characteristics differ between clusters. From the data, the centers of the clusters are searched, and the cases are assigned to the most similar cluster based on the input variables. Of note, the order of the data may affect the clustering output [42]

Kohonen clustering analysis is also known as knet or self-organizing map (S.O.M). A type of neural network that performs unsupervised clustering. Within a group, the cases are similar and different from a different cluster. The basic unit of the neural network is the neuron. The network architecture organizes neurons into input and output layers. All input neurons connect to output neurons, and the connections have a weight (w), which is also known as strength. The output is a map of a two-dimensional grid in which the units have no connections [43,44].

A

Supervised learning

- **Linear Models**
  - Bayesian Regression
  - Elastic-Net
  - Generalized Linear Models
  - LARS Lasso
  - Lasso
  - Least Angle Regression
  - Logistic regression
  - Multi-task Elastic-Net
  - Multi-task Lasso
  - Ordinary Least Squares
  - Orthogonal Matching Pursuit (OMP)
  - Passive Aggressive Algorithms
  - Perceptron
  - Polynomial regression: extending linear models with basis functions
  - Quantile Regression
  - Ridge regression and classification
  - Robustness regression: outliers and modeling errors
  - Stochastic Gradient Descent - SGD
- **Linear and Quadratic Discriminant Analysis**
  - Linear Discriminant Analysis (LDA)
  - Quadratic Discriminant Analysis (QDA)
- **Kernel ridge regression**
- **Support Vector Machines**
  - SVC, NuSVC, LinearSVC
  - OneClassSVM
- **Stochastic Gradient Descent**
- **Nearest Neighbors**
  - Unsupervised Nearest Neighbors
    - Finding the Nearest Neighbors
    - KDTree and BallTree Classes
  - Nearest Neighbors Classification
  - Nearest Neighbors Regression
  - Nearest Neighbor Algorithms
    - Brute Force
    - K-D Tree
    - Ball Tree
  - Nearest Centroid Classifier
    - Nearest Shrunken Centroid
- **Gaussian Processes**
  - Gaussian Process Regression (GPR)
  - Gaussian Process Classification (GPC)
- **Cross decomposition**
  - PLSCanonical
  - PLSSVD
  - PLSRegression
  - Canonical Correlation Analysis
- **Naive Bayes**
  - Gaussian Naive Bayes
  - Multinomial Naive Bayes
  - Complement Naive Bayes
  - Bernoulli Naive Bayes
  - Categorical Naive Bayes
  - Out-of-core naive Bayes model fitting
- **Decision Trees**
  - Classification
  - Regression
  - Multi-output problems
  - Tree algorithms: ID3, C4.5, C5.0 and CART
  - Missing Values Support
  - Minimal Cost-Complexity Pruning
- **Ensembles: Gradient boosting, random forests, bagging, voting, stacking**
  - Gradient-boosted trees
  - Random forests and other randomized tree ensembles
  - Bagging meta-estimator
  - Voting Classifier
  - Voting Regressor
  - Stacked generalization
  - AdaBoost
- **Multiclass and multioutput algorithms**
  - Multiclass classification
  - Multilabel classification
  - Multiclass-multioutput classification
  - Multioutput regression
- **Feature selection**
  - Removing features with low variance
  - Univariate feature selection
  - Recursive feature elimination
  - Feature selection using SelectFromModel
  - Sequential Feature Selection
  - Feature selection as part of a pipeline
- **Semi-supervised learning**
  - Self Training
  - Label Propagation
- **Isotonic regression**
- **Probability calibration**
  - Calibration curves
  - Calibrating a classifier
- **Neural network models (supervised)**
  - Multilayer Perceptron

B

Unsupervised learning

- **Gaussian mixture models**
  - Gaussian Mixture
  - Variational Bayesian Gaussian Mixture
- **Manifold learning**
  - Introduction
  - Isomap
  - Locally Linear Embedding
  - Modified Locally Linear Embedding
  - Hessian Eigenmapping
  - Spectral Embedding
  - Local Tangent Space Alignment
  - Multi-dimensional Scaling (MDS)
  - t-distributed Stochastic Neighbor Embedding (t-SNE)
- **Clustering**
  - Overview of clustering methods
  - K-means
  - Affinity Propagation
  - Mean Shift
  - Spectral clustering
  - Hierarchical clustering
  - DBSCAN
  - HDBSCAN
  - OPTICS
  - BIRCH
  - Clustering performance evaluation
- **Biclustering**
  - Spectral Co-Clustering
  - Spectral Biclustering
  - Biclustering evaluation
- **Decomposing signals in components (matrix factorization problems)**
  - Principal component analysis (PCA)
  - Kernel Principal Component Analysis (kPCA)
  - Truncated singular value decomposition and latent semantic analysis
  - Dictionary Learning
  - Factor Analysis
  - Independent component analysis (ICA)
  - Non-negative matrix factorization (NMF or NNMF)
  - Latent Dirichlet Allocation (LDA)
- **Covariance estimation**
  - Empirical covariance
  - Shrunk Covariance
  - Sparse inverse covariance
  - Robust Covariance Estimation
- **Novelty and Outlier Detection**
  - Overview of outlier detection methods
  - Novelty Detection
  - Outlier Detection
  - Novelty detection with Local Outlier Factor
- **Density Estimation**
  - Density Estimation: Histograms
  - Kernel Density Estimation
- **Neural network models (unsupervised)**
  - Restricted Boltzmann machines

C

Reinforcement learning

- **Temporal difference**
- **Deep adversarial networks**
- **Q-learning**

Figure 3. Types of machine learning methods for predictive data analysis. In addition to anomaly detection analysis, there are many other types of machine learning that can be classified as supervised (**A**),

unsupervised (**B**), and reinforcement learning (**C**). Of note, this figure includes methods usually used in predictive data analysis, but it does not focus on deep learning and reinforcement learning (please refer to popular deep learning frameworks such as tensorflow, keras, and pytorch, for documentation).

An image of the K-Means cluster (left), Kohonen clustering analysis (middle), and anomaly detection (right) are shown in Figure 4.

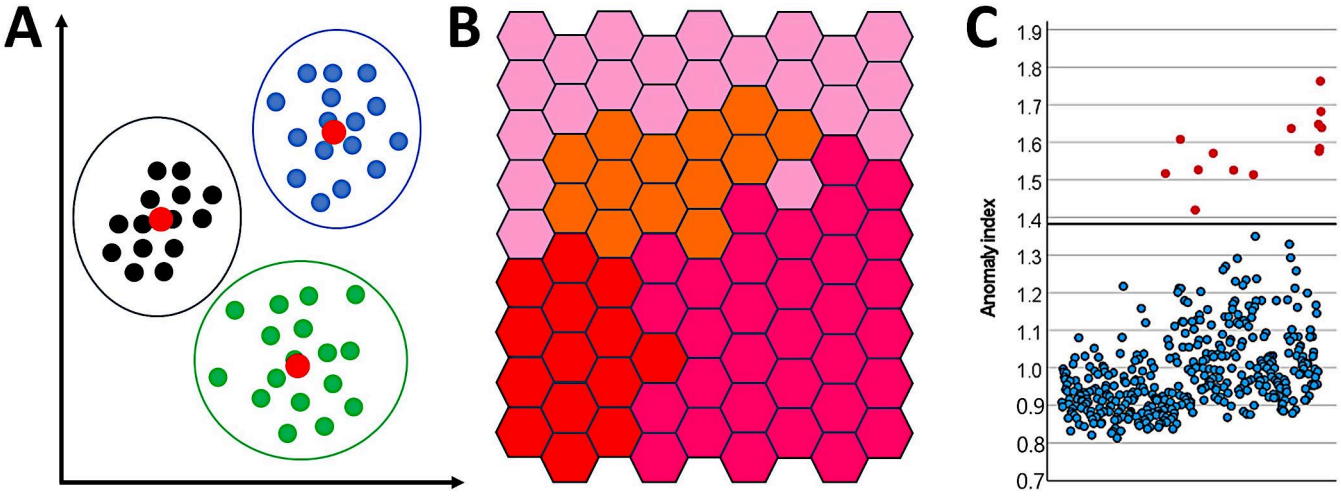

Figure 4. Segmentation analysis. This figure shows example images of the K-Means cluster (**A**), Kohonen clustering analysis (**B**), and anomaly detection (**C**).

The TwoStep cluster is also an unsupervised method. As in the K-Means and Kohonen methods, the cases are grouped in clusters with similar characteristics, whereas differences are observed between clusters. The method follows two steps. First, the raw input data are compressed into different subclusters. Second, a hierarchical clustering method joins the subclusters into larger clusters. Of note, this method is sensitive to the order of the training data. The TwoStep cluster has the advantage of handling mixed types of variables, can use large datasets, and can exclude outliers. However, it cannot handle missing data.

1.2.3. Anomaly Detection Analysis

An anomaly is a data point or collection of data that does not follow the same pattern or has the same structure as the rest of the data [45]. Anomaly detection is a machine learning method that identifies data points, events, and/or observations that deviate from a dataset's ordinary distribution [46]. In other words, anomaly detection is a technique that allows the identification of rare events that do not fit normal patterns. Examples of applications of this technique can be found in the following link: https://paperswithcode.com/task/anomaly-detection (accessed on 18 October 2023).

The anomaly detection procedure is designed to quickly detect unusual cases for data-auditing purposes in the exploratory data analysis step before any inferential data analysis. It searches for unusual cases and can be useful for detecting outliers within a large amount of data. The algorithm is designed for generic anomaly detection, which means that the definition of an anomalous case is not specific to any particular application [47]. Therefore, it can identify outliers even if they do not follow any known pattern. This method analyzes several variables to identify clusters that include cases with similar characteristics. Later, each record is compared with the others of the peer group to identify the anomalies. For each record, an anomaly index is assigned. The higher the anomaly index, the greater the deviation of a particular case from the average. An index above 2 is a good cutoff for identifying anomalies because it indicates a deviation twice the average. Of note, the identified cases should be assessed as suspected anomalies because, after close analysis,

they may turn out to be true outliers. The algorithm is divided into three stages: modeling, scoring, and reasoning.

2. Aim

This study aimed to identify new prognostic markers of DLBCL using anomaly detection analysis. By identifying outlier cases, the genes associated with those unusual cases were identified. Then, the prognostic value of the identified genes was evaluated in all cases of the series using other techniques, including several machine learning and artificial neural networks, and conventional biostatistics, such as Cox regression and Kaplan–Meier with log-rank test (Figure 5).

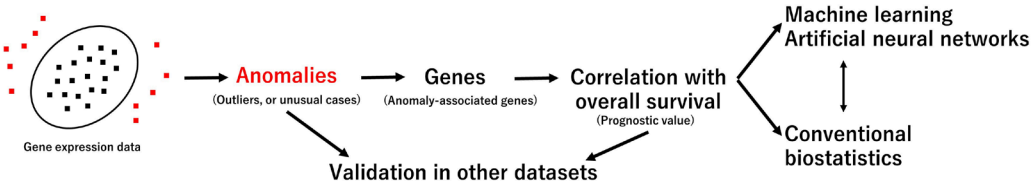

Figure 5. Aim and methodology. The discovery set was the Lymphoma/Leukemia Molecular Profiling Project (LLMPP) GSE10846 gene expression dataset (last update 25 March 2019) of 414 cases.

3. Materials and Methods

The gene expression of DLBCL is an important source of data for identifying prognostic markers. This study analyzed the gene expression of one of the most relevant DLBCL gene expression datasets of the Lymphoma/Leukemia Molecular Profiling Project (LLMPP). The dataset was GSE10846, which is a retrospective study of 414 DLBCL cases [48,49]. The last update of this dataset was 25 March 2019.

GSE10846 is a very well clinically characterized series of DLBCL. Despite being some years old, it serves the purpose of this research because we are looking for genes associated with the pathogenesis of DLBCL. Of note, to test the predictive value of one of the most relevant genes, only RCHOP-like cases were used.

A complete description of the clinicopathological characteristics of this series is presented in our previous publication that analyzed *CSF1R* expression [50]. In summary, 55% of the cases were male and aged > 60 years, NCCN-IPI was high–intermediate and high risk in 35.8%, the cell-of-origin molecular subtype was activated B cell type and unclassified in 45.8%, and the treatment was RCHOP-like in 56.3% of the cases.

The method used was anomaly detection analysis, which is a model designed to identify outliers in the gene expression data. This method is unsupervised. While traditional methods usually look into a few variables at the same time (one or two), the anomaly detection method can examine several fields (genes). The variables are analyzed to find clusters or peer groups that are similar. Each record can then be compared with others in its peer group to identify possible anomalies. The further away a case is from the typical center, the more likely it is to be abnormal. The anomaly detection algorithm is presented in the Zenodo repository [51].

The GSE10846 data were downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) public functional genomics data repository (https://www.ncbi.nlm.nih.gov/gds; accessed on 15 February 2024). The gene expression array used in this series was the GPL570, Affymetrix Human Genome U133 Plus 2.0 Array (HG-U133_Plus_2). The data were normalized and log2 transformed [50]. The series comprises 420 cases, 414 cases of DLBCL, and 6 cases of reactive lymphoid tissue. The series contains 20,684 genes. The gene expression values were collapsed to one value for each gene in the case of multiple probes using collapse to the maximum expression function [50]. The output identified case anomalies and the most relevant genes that contributed to them.

Further analysis consisted of several machine learning and artificial neural networks, as we recently published [52–57]. Finally, a conventional Cox regression for overall survival, backward conditional, was performed using the same set of genes to easily understand the prognostic value of these markers. Table 1 describes the basics of the machine learning and neural network analyses used in this study [58].

The immunohistochemical expression of RELB and other macrophage/dendritic cells-related markers was performed in 10 reactive tonsils and 30 cases of DLBL not-other-wise specified (NOS), including RELB, HLA DP-DR, SIRPA, CD85A, PD-L1, MARCO, and TOX. The primary antibodies were as follows: RELB (D7D7W, #10544, Cell Signaling Technology (CST)), HLA DP-DR (JS76, Spanish National Cancer Research Center (CNIO), Madrid, Spain), SIRPA (SIRPα/SHPS1, D6I3M, #13379, CST), CD85A (LILRB3, FRAS92B, CNIO), PD-L1 (E1J2J, #15165, CST), MARCO (MAKI373B, CNIO), and TOX (TOX1, NAN448B, CNIO). The immunohistochemistry was performed as previously described [52,53,55,57,59,60] using a Leica Bond-Max fully automated immunohistochemistry and in situ hybridization staining system (Leica Biosystems K.K., Tokyo, Japan). The slides were first visualized in an Olympus BX53 light microscope and later fully digitalized using a NanoZoomer S360 digital slide scanner (Hamamatsu whole slide imaging—WSI) and evaluated using the NDP.view2 image viewing software (version 2.9.29, U12388-01, Hamamatsu Photonics K.K., Hamamatsu, Japan).

Table 1. A brief description of the machine learning methods used in this study.

Model	Description
Anomaly detection	Method that quickly looks for unusual cases based on deviations from the norms of their cluster groups [51].
Bayesian Network	Creates a graphical model that shows variables (nodes) linked using arcs. Probabilistic independencies between nodes are displayed. The arcs do not necessarily represent cause and effect [52,53,55,61].
C5.0	Builds a decision tree. It splits the samples on the basis of the variable that provides more information and has more weight. Then, multiple splits are made based on other variables until the cases cannot be further divided. Finally, splits with few contributions to the model are removed. This model can only predict a categorical target [58,62].
C&R Tree	The classification and regression (C&R) tree is similar to the C5.0 method. All splits are binary [63].
CHAID	Chi-squared Automatic Interaction Detection (CHAID) creates decision trees using calculations based on the chi-square test. Crosstabulations between the input variables and the output are examined, and the variables are ranked according to their significance for selection in the tree model [64–68].
Discriminant	Creates a predictive model for group membership [69,70].
KNN Algorithm	Nearest Neighbor Analysis classifies cases based on their similarity to other cases. This method identifies the pattern of the data [71].
Logistic regression	Also known as nominal regression, it is a method that classifies records based on predictors in a manner similar to linear regression but with a categorical target variable.
LSVM	The data were classified on the basis of a linear support vector machine. This method is useful for large datasets with many variables [72,73].
Neural Network	Basic units, known as neurons, are organized into different layers. The input layer contains nodes with input variables (predictors). The output layer contains nodes with the target fields. Nodes are interconnected by different strengths (weights). The number of hidden layers defines the "deep" of the network. Using training, the weights are changed from random to optimized, and the network replicates the known outcomes [74–79].
Quest	Quick, Unbiased, Efficient Statistical (QUEST) tree creates a binary classification method. All splits are binary.

Table 1. *Cont.*

Model	Description
Random Forest	This is an implementation of the bagging algorithm. A collection of decision trees is used to make predictions [80–82].
Random Trees	It is based on the C&R methodology and uses recursive partitioning to split records into segments with similar outputs [83].
SVM	A support vector machine (SVM) is suitable when the dataset contains a very large number of predictors. It is a solid classification and regression technique that does not overfit the training data [84,85].
Tree-AS	This method creates a decision tree using CHAID or exhaustive CHAID, which is more time-consuming [52,53,57].
XGBoost Linear	Implementation of a gradient boosting algorithm with a linear model as the base [86].
XGBoost Tree	Implementation of a gradient boosting algorithm with a tree model as the base [87–94].

Additional descriptions of machine learning and neural network models are presented in the companion manuscript "Artificial Intelligence Analysis and Reverse Engineering of Molecular Subtypes of Diffuse Large B-Cell Lymphoma Using Gene Expression Data". *BioMedInformatics* 2024, 4, 295–320. https://doi.org/10.3390/biomedinformatics4010017 (accessed on 15 February 2024) [58].

All analyses were performed on a desktop equipped with an AMD Ryzen 9 5900X and NVIDIA GeForce RTX 3060 Ti GPU and 16 GB of RAM. Conventional statistics were calculated using IBM SPSS version 27.0.1.0 64-bit edition (IBM Corporation, Orchard Rd, Armonk, NY 10504, USA).

Anomaly detection analysis was also performed using other series to confirm that the method is applicable. The GSE31312 and GSE117556 datasets were used, which have 498 and 928 cases of DLBCL.

The gene expression analysis of *RELB* was also performed in TCGA (n = 267) and GSE57611 (n = 30).

4. Results

4.1. Anomaly Detection Analysis

The anomaly detection analysis using GSE10846 ranked the cases according to the anomaly index, which ranged from 0.813 to 1.763 (Supplementary Excel File). Of note, cases with anomaly index values of less than 1 or even 1.5 would not be considered anomalies. The distribution of anomaly index values is shown in Figure 6.

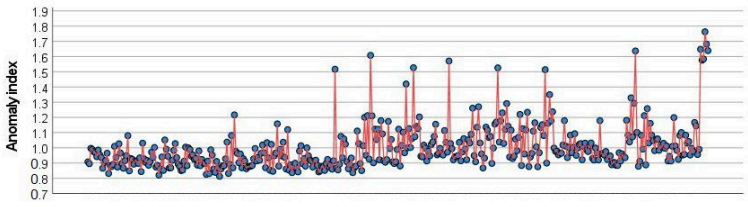

Figure 6. Anomaly index values. Anomaly detection analysis identifies outliners, or unusual cases, in the data. It records information on what normal behavior looks like and identifies outliers even if they do not conform to any known pattern. It is an unsupervised method that examines large numbers of variables to identify clusters or peer groups. Then, each record is compared to others in its peer group to identify possible anomalies. Each record (blue circle) is assigned an abnormality index. High index implies a higher average of the case than the average. In the setup, several options can be specified, such as the adjustment of coefficient, number of peer groups, noise level, and noise ratio.

The model also identified the 12 genes that contributed to anomaly detection: *DPM2*, *TRAPPC1*, *HYAL2*, *TRIM35*, *NUDT18*, *TMEM219*, *CHCHD10*, *IGFBP7*, *LAMTOR2*, *ZNF688*, *UBL7*, and *RELB* (Table 2).

Table 2. Genes identified in anomaly detection analysis using the GSE10846 series.

Gene	Name	Function
DPM2	Dolichyl-Phosphate Mannosyltransferase Subunit 2, Regulatory	Regulation of protein stability
TRAPPC1	Trafficking Protein Particle Complex Subunit 1	Endoplasmic reticulum-to-Golgi vesicle-mediated transport
HYAL2	Hyaluronidase 2	Positive regulation of the extrinsic apoptotic signaling pathway. Related to bladder cancer inflammation and tumor-associated myeloid cells [95]
TRIM35	Tripartite Motif Containing 35	Multiple biological processes, including cell death, glucose metabolism, and innate immune response. Correlation with high infiltration of NK cells in DLBCL [96], tumor suppressor in breast cancer [97], predicts survival in hepatocellular carcinoma, and is related to tumorigenesis
NUDT18	Nudix Hydrolase 18	Elimination of potentially toxic nucleotide metabolites
TMEM219	Transmembrane Protein 219	Apoptosis
CHCHD10	Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 10	Positive regulation of mitochondrial outer membrane permeabilization involved in the apoptotic signaling pathway
IGFBP7	Insulin-Like Growth Factor Binding Protein 7	Prostacyclin production and cell adhesion. Related to Epstein–Barr virus tumorigenesis, mantle cell lymphoma, and lung cancer [56,98,99]
LAMTOR2	Late Endosomal/Lysosomal Adaptor, MAPK, a nd MTOR Activator 2	Activation of mTORC1, with control of cell growth and related to the risk of breast cancer [100]
ZNF688	Zinc Finger Protein 688	Negative regulation of transcription by RNA polymerase II
UBL7	Ubiquitin Like 7	Ubiquitin-dependent protein catabolic process, cellular response to stress. Autoantibody signature in hepatocellular carcinoma [101]; necroptosis-related marker in stomach adenocarcinoma [102]
RELB	RELB Proto-Oncogene, NF-KB Subunit	NF-kappa-B is a pleiotropic transcription factor involved in many biological processes, such as inflammation, immunity, differentiation, cell growth, tumorigenesis, and apoptosis. Pathogenic marker of DLBCL [103,104]

Information based on GeneCards and UniProtKB/Swiss-Prot.

The anomaly detection methodology was also tested in another series of DLBCL, GSE31312. This is a series of 498 de novo adult DLBCL cases treated with RCHOP. Gene expression was performed using the Affymetrix HG-U133 Plus 2.0 platform. The last update was 3 August 2020. Anomaly detection classified the series into two peer groups of 315 and 183 cases. In the peer group of 183, the contribution was also of 12 genes, but different (*NACA4P*, *DAZAP2*, *RSP28*, *RPS7*, *TSPOAP1_AS1*, *MT_ND5*, *MIR142*, *MGC16275*, *SHOC2*, *CALM1*, *GLUL*, and *SIT29*). Therefore, the anomaly detection method can be applied to series other than GSE10846. Of note, because of the intrinsic heterogenicity of DLBCL, including the characteristics of unusual cases (anomalies), the two series provided different results. This is not a bad result. Other methods, such as artificial neural networks, can also provide different results in each analysis due to different factors, including the random number generator.

Anomaly detection was also performed using the GSE117556 dataset. This dataset belongs to a retrospective analysis of whole transcriptome data for 928 DLBCL patients from the REMoDLB clinical trial. The platform was an Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip [105]. RNA was extracted from formalin-fixed, paraffin-embedded (FFPE) biopsies. The method classified the series into two peer groups of 661 and 267 records. In the second peer group, the contribution was of 27 genes.

4.2. Prediction of Overall Survival Using Machine Learning and Artificial Neural Networks Based on 12 Genes

The 12 genes previously identified in the anomaly detection analysis were used as predictors (inputs) of the prognosis of patients with DLBCL in the GSE10846 series. The prognosis was defined by the outcome of overall survival (output variable, dead versus alive). Several machine learning models and artificial neural networks were tested, including the C5.0 decision tree, logistic regression, Bayesian network, discriminant analysis, KNN algorithm, LSVM, random trees, SVM, Tree-AS, XGBoost linear, SGBoost tree, CHAID tree, Quest tree, C&R tree, random forest, and neural network.

The models were ranked according to overall accuracy (%), and the best models were the XGBoost tree, random forest, and C5 tree (Table 3).

Of note, the analysis was performed in all cases, CHOP-like, and RCHOP-like cases.

Table 3. Prediction of overall survival outcome (dead vs. alive) using machine learning and artificial neural networks, based on 12 previously identified genes in anomaly detection analysis.

Model	No. of Genes	Overall Accuracy (%)
XGBoost Tree	12	99.8
Random Forest	12	98.6
Random Trees	12	93.9
C5	7	75.4
KNN Algorithm	12	73.4
CHAID	5	71.7
Neural Network	12	71.3
Logistic regression	12	71.0
LSVM	12	70.1
SVM	12	69.3
Discriminant	12	68.4
C&R Tree	12	68.4
Tree-AS	3	65.5
Quest	6	64.5
XGBoost Linear	12	60.2
Bayesian Network	12	0.0

The performance was assessed with the overall accuracy that is the percentage of records for which the outcome was correctly predicted. The formula is shown in Appendix B.

4.3. Cox Regression Analysis of Overall Survival Using the 12 Genes

The 12 genes were used as predictors of overall survival using conventional Cox regression analysis in the GSE10846 series. The method was backward conditional. In the last step (n = 8), only five genes retained significant values. In this model, *TRAPPC1*, *IGFBP7*, and *RELB* were associated with a favorable prognosis, and *HYAL2* and *UBL7* were associated with a poor prognosis (Table 4).

Table 4. Prediction of the overall survival using Cox regression analysis based on the 12 genes.

Gene	B	*p* Value	Hazard Risk	95% CI for HR	
				Lower	Upper
TRAPPC1	−0.391	0.023	0.676	0.483	0.946
HYAL2	0.757	0.000	2.133	1.461	3.113
IGFBP7	−0.683	0.000	0.505	0.400	0.637
UBL7	0.507	0.001	1.660	1.234	2.233
RELB	−0.361	0.003	0.697	0.549	0.885

Backward conditional method.

When *MYC* and *BCL2* were added to the equation, the Cox regression analysis only kept *MYC* as a significant predicted value (p value = 0.008, HR = 1.280, 95% CI 1.066–1.536), in addition to the other five genes that had similar values as in Table 4.

Similar results were found when NCCN-IPI was added to the equation with the five genes. NCCN-IPI was also significant, as were the other five genes (p value < 0.001, HR = 2.438, 95% CI = 1.713–3.469).

In this model, the molecular subtypes of GCB and ABC had no predictive value when combined with the five genes.

Finally, the prognostic value as a single variable of *RELB* was tested using survival analysis with Kaplan–Meier and log-rank tests. In the DLBCL cases treated with RCHOP-like cases, high *RELB* expression was associated with a favorable prognosis of the patients (Figure 7).

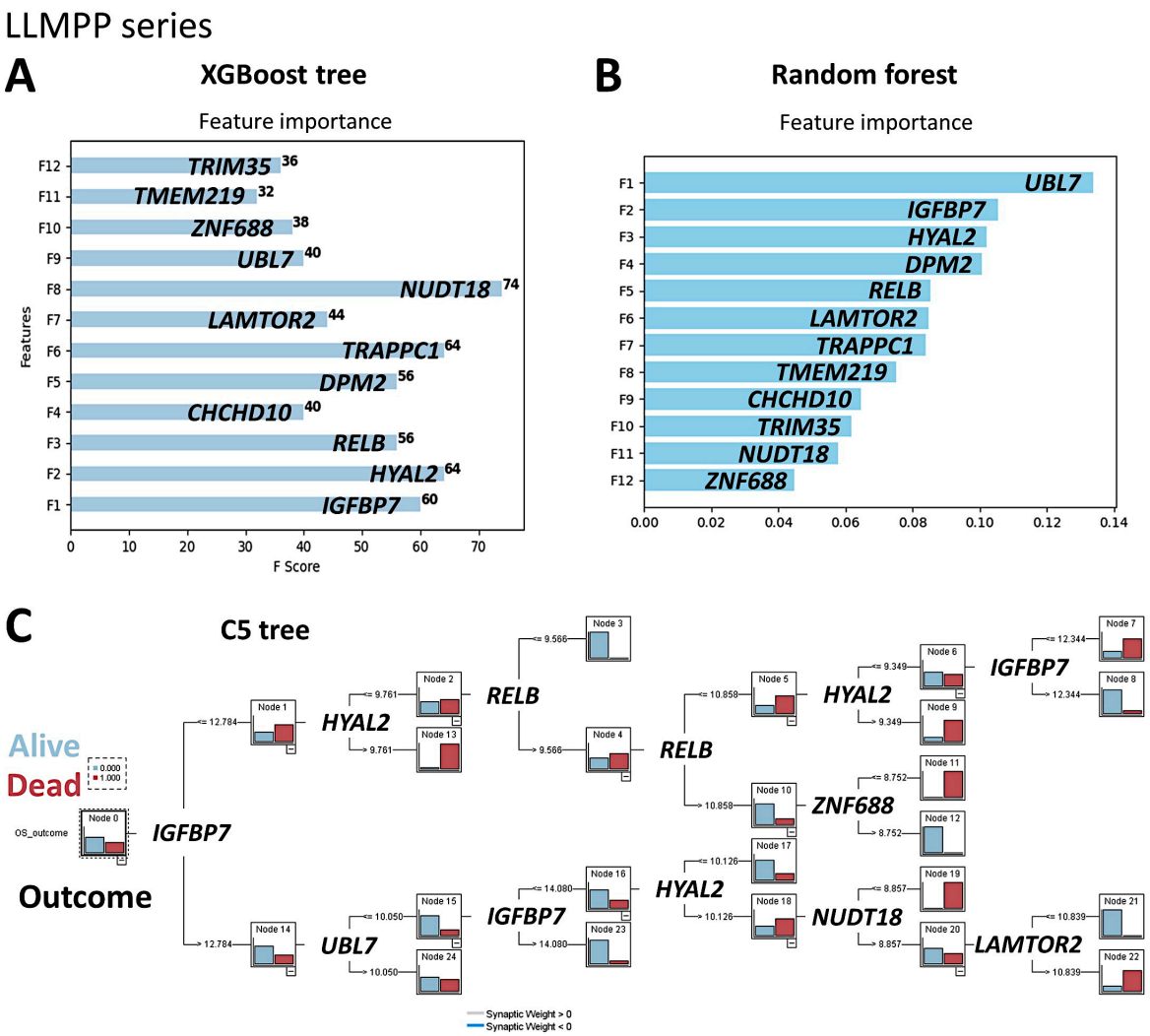

Figure 7. *Cont.*

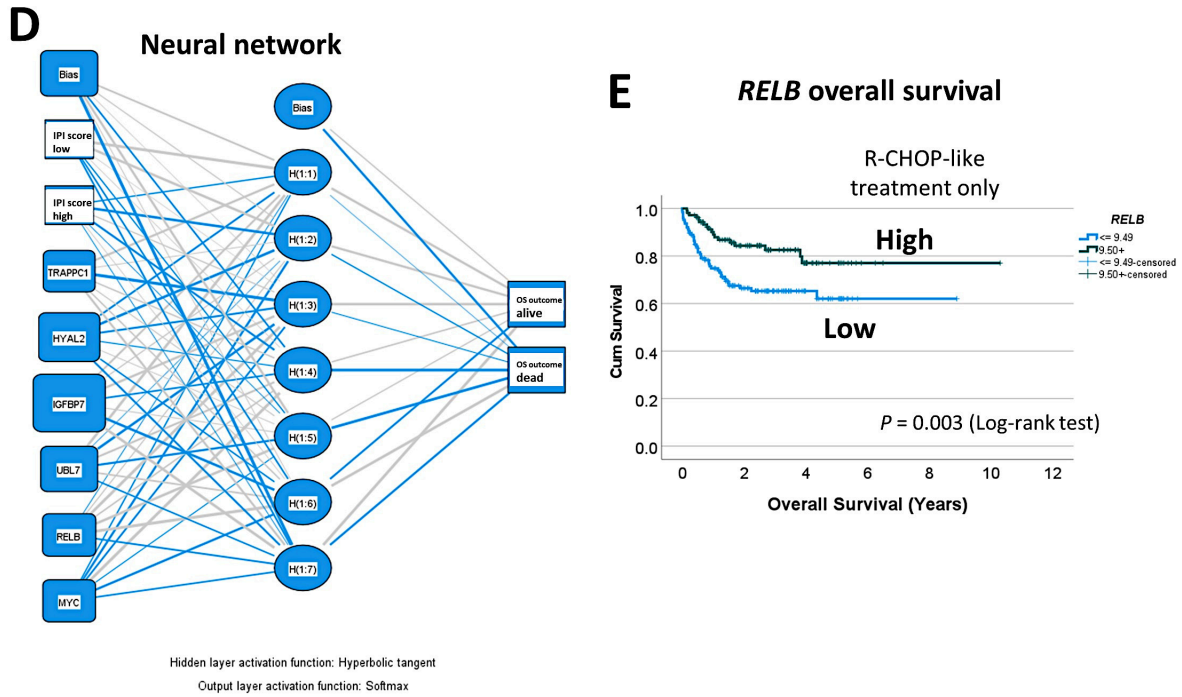

Figure 7. Machine learning and artificial neural networks using the LLMPP gene expression dataset. Abnormality detection analysis identified 12 genes. The prognostic value of these genes for overall survival was tested using several artificial intelligence analysis techniques. XGBoost tree (**A**), random forest (**B**), C5 tree (**C**), and neural network (**D**). Of note, the prognostic value of *RELB* was confirmed in the RCHOP-like cases of the LLMPP series using conventional overall survival analysis of Kaplan–Meier with log-rank tests (**E**). High gene expression of *RELB* was associated with favorable overall survival (**E**).

The prognostic value of *RELB* was evaluated in other series of patients. In TCGA and GSE57611, high RELB gene expression was associated with favorable overall survival (Hazard-risk 0.45 and 0.1645, respectively (p values 0.0018 and 0.0171) (Appendix A, Figure A1).

4.4. Validation of the Predictive Value of RELB for Overall Survival of Patients Using Gene Set Enrichment Analysis (RCHOP-Treated Cases)

The predictive value of *RELB* in DLBCL was assessed using GSEA analysis in the RCHOP-treated cases of the LLMPP series. Gene set enrichment analysis (GSEA) is a computational method that determines whether an a priori-defined set of genes shows statistically significant, concordant differences between two biological states (e.g., phenotypes) [106–108]. In this study, the phenotypes were the overall survival outcome as dead and alive. The priori set of genes was the *RELB* pathway. To define the RELB pathway, the STRING platform was used. STRING is a protein–protein interaction network and functional enrichment analysis [109,110]. A functional network association analysis was performed using *RELB* as the hub gene to design the *RELB* pathway (1st shell $\leq$ 20 interactions; 2nd shell $\leq$ 5 interactions; confidence as the meaning of network edges). The network had 26 nodes and 227 edges, with an average node degree of 17.5, an averaged local clustering coefficient of 0.865, and protein–protein interaction enrichment p value < 0.001 (Figure 8A).

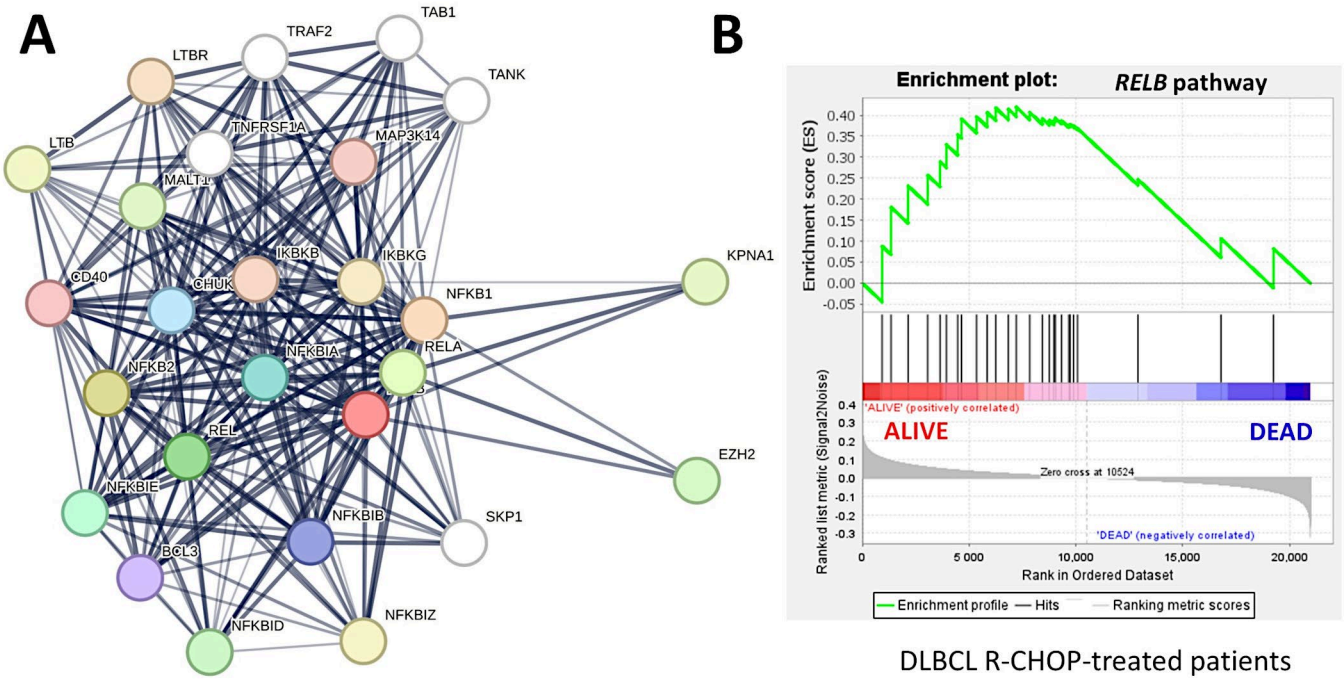

Figure 8. Protein−protein interaction analysis and gene set enrichment analysis (GSEA) of *RELB* gene and pathway. First, a functional network association analysis (protein−protein interaction network) focused on RELB created a pathway. Later, this RELB pathway was used in the GSEA analysis. The GSEA analysis confirmed the association of the RELB gene and pathway with a favorable overall survival of patients with DLBCL treated with R-CHOP therapy. Functional network association analysis (**A**), GSEA (**B**).

Later, the genes of the *RELB* network were used as a pathway for the GSEA analysis, and the results showed enrichment toward the alive phenotype (Figure 8B). Therefore, the *RELB* pathway was associated with a favorable overall survival of the DLBCL pathway, as identified in our previous analyses of machine learning and conventional biostatistics. In the core enrichment of the GSEA plot, 13 genes were identified, with *RELB* in the third position (Figure 8, Table 5).

Table 5. Gene set enrichment analysis (GSEA) using *RELB* network and pathway.

No.	Symbol	Title	Running Enrichment Score (ES)	Core Enrichment
1	*REL*	REL proto-oncogene, NF-kB subunit	0.0879	Yes
2	*LTB*	Lymphotoxin beta	0.1807	Yes
3	*RELB*	RELB proto-oncogene, NF-kB subunit	0.2316	Yes
4	*TRAF2*	TNF receptor-associated factor 2	0.2571	Yes
5	*NFKB2*	Nuclear factor kappa B subunit 2	0.2892	Yes
6	*CD40*	CD40 molecule	0.3301	Yes
7	*MALT1*	MALT1 paracaspase	0.3536	Yes
8	*NFKBID*	NFKB inhibitor delta	0.3914	Yes
9	*NFKBIA*	NFKB inhibitor alpha	0.3964	Yes
10	*RELA*	RELA proto-oncogene, NF-kB subunit	0.4062	Yes
11	*IKBKG*	Inhibitor of nuclear factor kappa B kinase regulatory subunit	0.4174	Yes
12	*BCL3*	BCL3 transcription coactivator	0.4145	Yes
13	*TAB1*	TGF-beta activated kinase 1 (MAP3K7) binding protein 1	0.4192	Yes
14	*TANK*	TRAF family member-associated NFKB activator	0.4068	No
15	*NFKBIB*	NFKB inhibitor beta	0.3919	No
16	*EZH2*	Enhancer of zeste 2 polycomb repressive complex 2 subunit	0.3875	No
17	*TNFRSF1A*	TNF receptor superfamily member 1A	0.3872	No
18	*NFKBIE*	NFKB inhibitor epsilon	0.3934	No
19	*IKBKB*	Inhibitor of nuclear factor kappa B kinase subunit beta	0.3868	No
20	*SKP1*	S-phase kinase-associated protein 1	0.3765	No
21	*CHUK*	Component of inhibitor of nuclear factor kappa B kinase complex	0.3782	No
22	*NFKB1*	Nuclear factor kappa B subunit 1	0.3747	No
23	*KPNA1*	Karyopherin subunit alpha 1	0.3685	No
24	*MAP3K14*	Mitogen-activated protein kinase kinase kinase 14	0.2468	No
25	*LTBR*	Lymphotoxin beta receptor	0.106	No
26	*NFKBIZ*	NFKB inhibitor zeta	0.082	No

This table shows the genes used in the GSEA analysis of Figure 8B.

4.5. Immunohistochemical Analysis of RELB and Immune Microenvironment

The histological protein expression of *RELB* was analyzed by immunohistochemistry in 10 reactive tonsils (i.e., reactive tissue control) and 30 cases of DLBCL NOS. In reactive tonsils, the expression of RELB was mainly located in the germinal centers of reactive follicles. There, two types of intensity were identified: strong in macrophages/dendritic cells and weak in the B lymphocytes. In DLBCL NOS, the expression was heterogeneous, and four patterns were identified: 0 (negative), 1+ (weak), 2+ (moderate), and 3+ (strong). In DLBCL, the positive cells were heterogeneous when the staining was moderate/strong, with a mixture of B-cell staining and macrophage/dendritic cell-like. Additional markers were included in the panel to investigate the expression of macrophage-related immune microenvironment markers, including HLA DP-DR, SIRPA, CD85A, PD-L1, MARCO, and TOX (TOX1). In summary, the expression of RELB partially correlated with macrophage/dendritic cell markers but was also present in the B-lymphocytes (Figures 9 and 10).

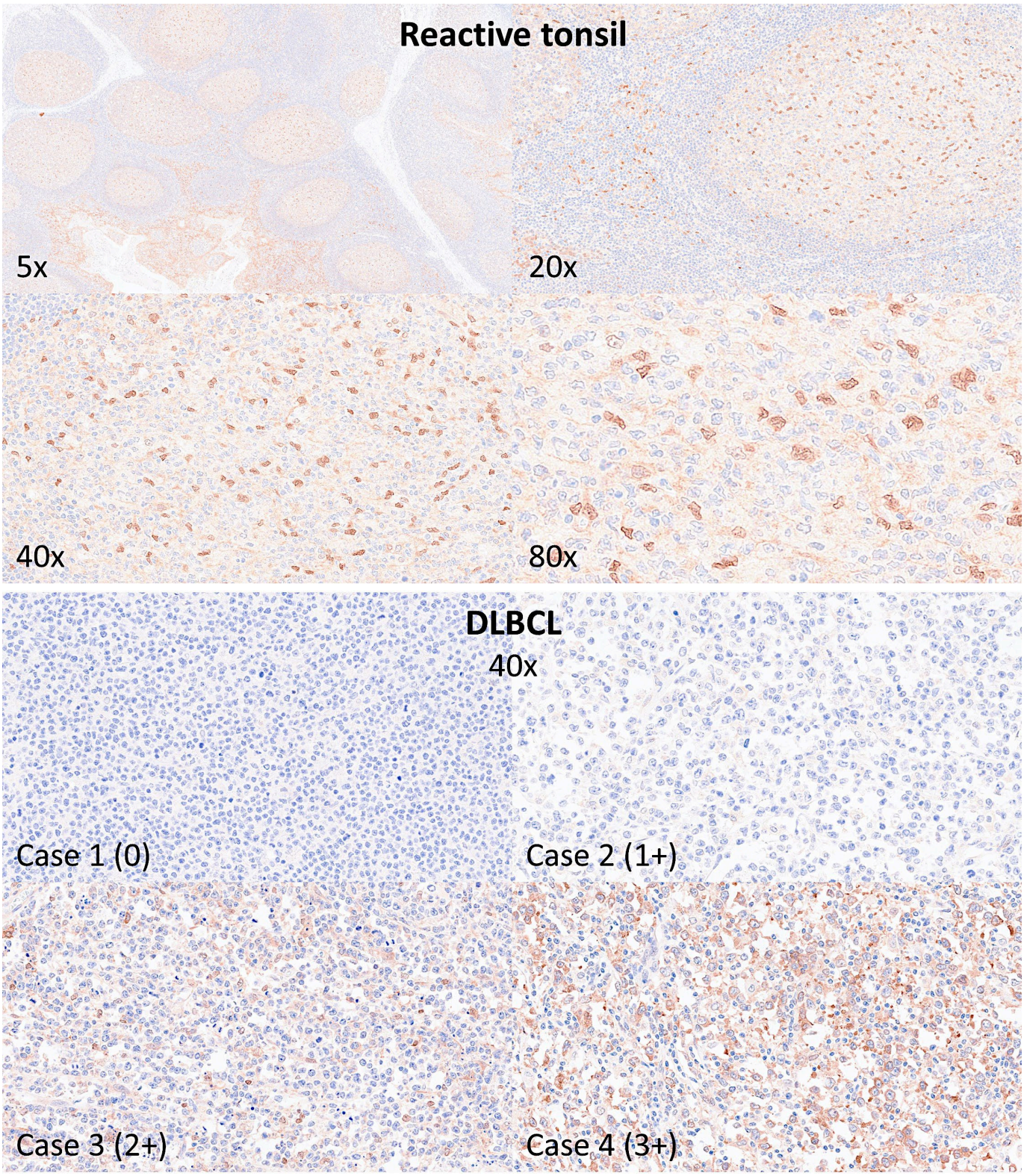

Figure 9. Immunohistochemical analysis of RELB in reactive tonsils and DLBCL. The protein expression of RELB was analyzed in 10 reactive tonsils (tissue control) and 30 cases of DLBCL not otherwise specified (NOS). In reactive tonsils, RELB expression was mainly present in the germinal centers of the follicles, with strong staining in macrophage/dendritic cells and weak in the B-lymphocytes. In DLBCL NOS, the staining was heterogeneous, ranging from 0 to 3+, and expressed by neoplastic B-lymphocytes and cells of the microenvironment.

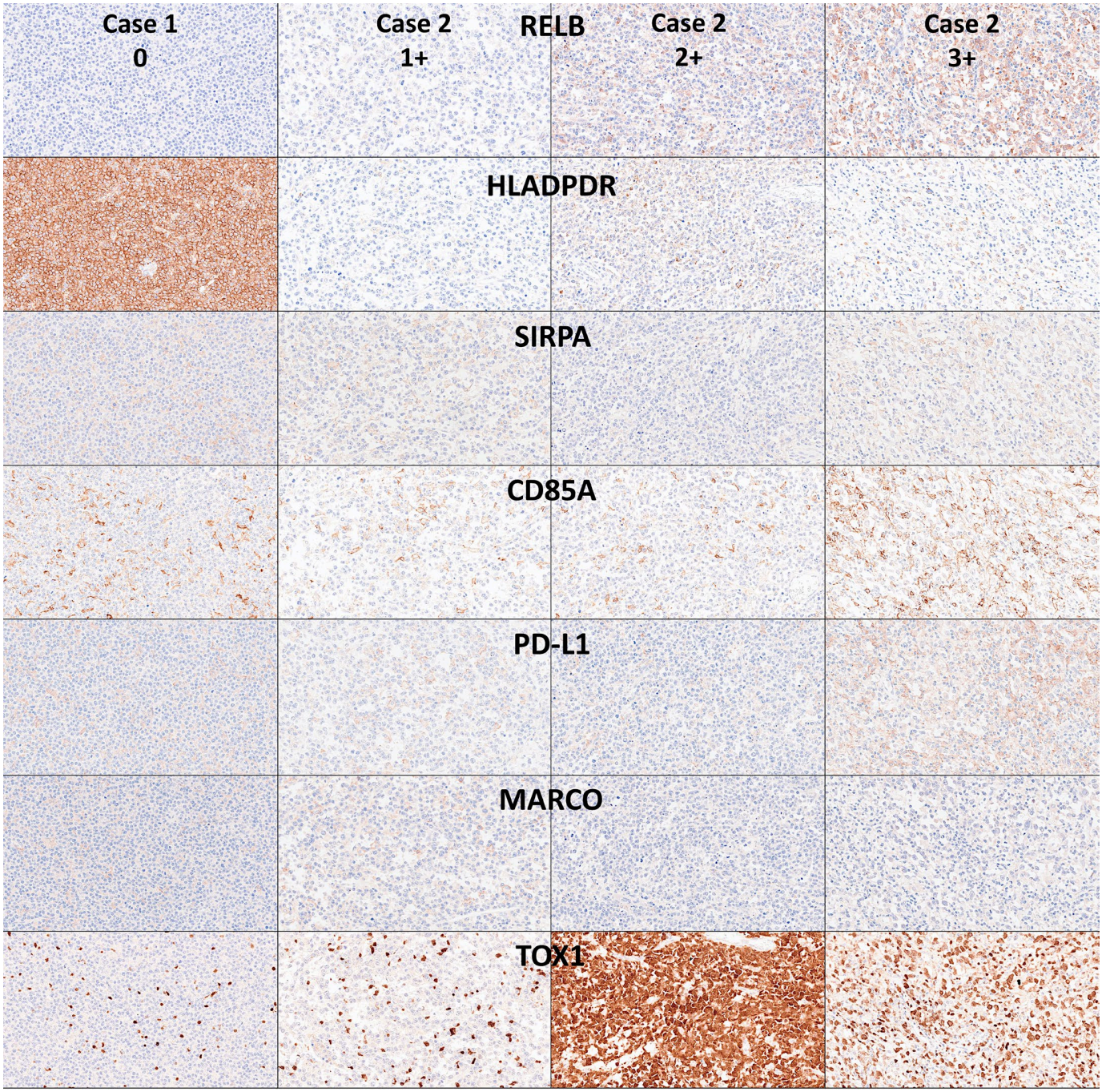

Figure 10. Immunohistochemical analysis of RELB in relationship with other immune microenvironment markers in DLBCL NOS. The expression of RELB in DLBCL was heterogeneous, with a pattern compatible with mixture of macrophage/dendritic cells and B-lymphocytes. Correlation with other macrophage-associated and immune microenvironment/immune checkpoint markers was performed using HLA DP-DR, SIRPA, CD85A, PD-L1, MARCO, and TOX (TOX1). Original magnification 400×.

5. Discussion

Diffuse large B-cell lymphoma (DLBCL) is one of the most frequent histological subtypes of non-Hodgkin lymphomas, accounting for approximately 20–30% of cases. DLBCL is a heterogeneous diagnostic category with heterogeneous morphological, genetic, and clinical characteristics. The current classification dates back to 2017 with the revised 4th edition [1], and several subtypes were defined, including T cell/histiocyte-rich large B cell

lymphoma, the primary mediastinal large B cell lymphoma, intravascular B cell lymphoma, the primary DLBCL of the central nervous system, the primary cutaneous DLBCL, leg type, and EBV-positive DLBCL not–otherwise–specified (NOS) [1]. An important subtype is high-grade B-cell lymphoma with *MYC* and *BCL2* and/or *BCL6* rearrangements, which in some cases had previously been called Burkitt-like lymphoma [1,2]. In this study, our diagnostic category was diffuse large b-cell lymphoma not otherwise specified.

The molecular pathogenesis of DLBCL includes a complex and multistage pathological mechanism that results in the proliferation of a germinal center or postgerminal center B cell clone. One of the best characterized molecular changes is the acquisition of rearrangements of *BCL6*, *BCL2*, and *MYC*.

The *MYC* proto-oncogene is a transcription factor that binds to DNA nonspecifically yet recognizes the 5′-CAC[GA]TG-3′ sequence [111,112]. MYC activates the transcription of several genes that have tumor-promoting functions [111,112]. In DLBCL, *MYC* gene rearrangement occurs in approximately 10% of cases [113], and in 80% of translocation-positive cases, the partner is the *IGH* locus. The presence of *MYC* rearrangement, copy-number gain (amplification), and/or overexpression is associated with poor prognosis [113–115]. Despite the importance of *MYC* in DLBCL pathogenesis, most cases are *MYC* rearrangement negative. In our series, the *REL* high group was characterized by a lower frequency of *MYC* translocation: *REL* high vs. low, 11.5% vs. 88.5% ($p = 0.009$).

Using a novel analysis approach, we identified 12 genes with prognostic value in DLBCL: *DPM2*, *TRAPPC1*, *HYAL2*, *TRIM35*, *NUDT18*, *TMEM219*, *CHCHD10*, *IGFBP7*, *LAMTOR2*, *ZNF688*, *UBL7*, and *RELB*. The functions and biological relevance of these genes are shown in Table 1. Most of these genes have multiple functions, but a proportion of them are related to the control of apoptosis, such as *HYAL2*, *TRIM35*, *TMEM219*, *CHCHD10*, and *UBL7*. In DLBCL, the dysregulation of the apoptosis pathway is an important pathogenic mechanism. In up to 30% of DLBCL cases, especially in the germinal center B cell-like subtype, there is BCL2 overexpression. BCL2 is an oncogene that inhibits apoptosis and leads to the enhanced survival of tumor cells [116].

We also identified a marker of the NF-kappa-B pathway, the RELB proto-oncogene, NF-KB subunit. NF-kappa-B is a pleiotropic transcription factor involved in many biological processes, such as inflammation, immunity, differentiation, cell growth, tumorigenesis, and apoptosis. It is a pathogenic marker of DLBCL [103,104]. We found that the high expression of *RELB* was associated with a favorable prognosis. Our results are consistent with previously reported data in DLBCL [117,118].

The work of Chi Young Ok et al. [118] is of special interest. This study analyzed a large cohort of 533 cases of de novo DLBCL, and the gene and protein expression of the five NF-KB pathway subunits (p50, p52, p65, RELB, and c-Rel) was assessed. All subunits were expressed by GCB and ABC DLBCL, but there were differences between the two subtypes of the cell of origin. The expression of p52/RELB was associated with improved OS and PFS. When cases were stratified into GCB and ABC, p52 or p52/RELB expression status was associated with better OS and PFS only within the GCB subtype.

NF-KB signaling is an important regulator of apoptosis. Several genetic alterations and other mechanisms activate the NF-KB pathway. The constitutive activation of the NF-KB pathway contributes to cancer development, progression, and therapy resistance [119]. NF-KB signaling is categorized as canonical or noncanonical.

The canonical pathway is activated by C-like receptors 4, the TNF receptor family, and the antigen receptors BCR and TCR, whereas the noncanonical pathway is activated by other receptors, such as BAFF-R, CD40, RANK, CD30, and LTβ-R [119]. The canonical pathway includes SYK, BTK, CARD1/MLAT1/BCL10, and RELA. Target genes are related to survival, anti-apoptosis, cell proliferation, inflammation, and innate immunity.

Conversely, the noncanonical leads to the activation of IKK, p100/RELB/P50. This pathway targets genes related to lymphoid organogenesis, adaptive immunity, anti-inflammatory properties, and B-cell maturation [119].

This study and the use of the anomaly detection technique have limitations. This study focused on the GSE10846 dataset. This is a series that was made public on 28 November 2008 and was last updated on 25 March 2019. Therefore, it is a relatively old series of DLBCL cases. This retrospective study included 181 clinical samples from CHOP-treated patients and 233 samples from Rituximab–CHOP–treated patients. The array used was Affymetrix U133 plus 2.0. Currently, there is newer technology to assess gene expression data, such as Clariom assays from Thermo Fisher Scientific and next-generation sequencing (RNA-Seq) to reveal the presence and quantity of RNA molecules in biological samples. Therefore, this study used a series with relatively old technology, and approximately half of the patients had received CHOP therapy. However, this series was created by the Lymphoma/Leukemia Molecular Profiling Project (LLMPP). It is very well annotated, and the clinicopathological characteristics of the samples are complete and reliable. The analysis was first performed using all 414 cases but was later repeated using only the R-CHOP cases. For example, Figure 2 shows the overall survival of patients based on *RELB* expression only in RCHOP-like cases, and the prognostic relevance of *RELB* was maintained.

The anomaly detection procedure searches for unusual cases based on deviations from the norms of their cluster groups [47]. This procedure allows the rapid detection of unusual cases during the exploratory data analysis step before any inferential data analysis [47]. However, this algorithm is designed for generic anomaly detection, and the definition of anomalous cases is not specific to any particular application [46,47].

The anomaly detection analysis using GSE10846 ranked the cases according to the anomaly index, which ranged from 0.813 to 1.763 (Supplementary Excel File). There is no definitive cutoff for selecting anomalous cases. Cases with anomaly index values less than 1 or even 1.5 would not be considered anomalies, but the selection cases should be tested by other techniques to confirm that they are true anomalous cases.

The results of the anomaly detection technique depend on the series of cases. This is a limitation because anomalous cases may have different clinicopathological characteristics and different gene expression profiles depending on the series, especially if the disease is heterogeneous, such as DLBCL. Anomaly detection was technically successful in the GSE31312 and GSE117556 datasets, but the genes identified were different. This is due to the heterogeneous profile of DLBCL and the characteristics of each series. This is not a bad result. However, we confirmed the relevance of *RELB* for predicting DLBCL not only in the GSE10846 but also in the TCGA and GSE57611 series.

The model identified 12 genes that contributed to anomaly detection in the GSE10846 series: *DPM2*, *TRAPPC1*, *HYAL2*, *TRIM35*, *NUDT18*, *TMEM219*, *CHCHD10*, *IGFBP7*, *LAMTOR2*, *ZNF688*, *UBL7*, and *RELB* (Table 1). The importance of these genes was validated using other machine learning techniques and conventional statistics.

When the 12 genes were used as predictors of overall survival using a conventional Cox regression analysis in the GSE10846 series, in the last step, only five genes retained a significant value. In this model, *TRAPPC1*, *IGFBP7*, and *RELB* were associated with a favorable prognosis, whereas *HYAL2* and *UBL7* were associated with a poor prognosis (Table 4). Similar results were found when NCCN-IPI was added to the equation with the five genes. NCCN-IPI was also significant, as were the other five genes (p value < 0.001, HR = 2.4). Therefore, despite its limitations, this bioinformatics approach provides useful information regarding the pathogenesis of DLBCL. Of note, further analysis will include the validation of *RELB* in individual series of cases.

Jintao Wu et al. recently identified *RELB* as a potential molecular biomarker for immunotherapy in human pan-cancer [120]. Using the Cancer Genome Atlas Program (TCGA) dataset, they found that RELB was detected in human cancers and that the expression was associated with the overall survival of the patients, with a favorable in some cases, such as glioblastoma multiforme and lung adenocarcinoma, and unfavorable in others, such as breast cancer. Interestingly, using gene set enrichment analysis, an association of *RELB* and the tumor immune microenvironment and immune checkpoint was identified [120]. This is a relevant result because immuno-oncology and immunotherapeutic therapies in DLBCL

include monoclonal anti-CD20 antibody (rituximab), monoclonal anti-PD-1 antibodies (nivolumab and pembrolizumab), monoclonal anti-PD-L1 antibodies (avelumab, durvalumab, and atezolizumab), and chimeric antigen receptor (CAR) T-cell therapy [121,122]. The role of *RELB* in the pathogenesis of DLBCL is complex [103,104,118,123–126]. Further analysis of the impact of *RELB* on the prognosis of DLBCL and their relationship with known and well stablished markers such as *MYC*, *BCL2*, and *BCL6* [12,127] is warranted.

6. Conclusions

In conclusion, using a statistical approach based on anomaly detection and artificial intelligence of gene expression data of DLBCL, we identified pathogenic markers related to apoptosis, MAPK and MTOR, and the NF-KB pathway. High expression of the *RELB* proto-oncogene is associated with a favorable prognosis of DLBCL.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biomedinformatics4020081/s1, Anomaly detection Excel File.

Author Contributions: Conceptualization, J.C.; formal analysis, J.C.; investigation, J.C. and R.H.; resources, J.C.; writing—original draft preparation, J.C.; writing—review and editing, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), grant number KAKEN 23K06454. Rifat Hamoudi is funded by ASPIRE, the technology program management pillar of Abu Dhabi's Advanced Technology Research Council (ATRC), via the ASPIRE Precision Medicine Research Institute Abu Dhabi (AS-PIREPMRIAD) award grant number VRI-20-10.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki, and was approved by the Institutional Review Board of TOKAI UNIVERSITY, SCHOOL OF MEDICINE (protocol code IRB14R-080 and IRB20-156).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: For original data, please contact joaquim.carreras@tokai-u.jp.

Acknowledgments: We thank all members of the Lymphoma/Leukemia Molecular Profiling Project for sharing the GSE10846 dataset and the authors of the GSE31312 dataset. We thank Giovanna Roncador, head of Monoclonal antibodies unit of Centro Nacional de Investigaciones Oncologicas (CNIO) (Spanish National Cancer Research Centre) for the primary antibodies.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

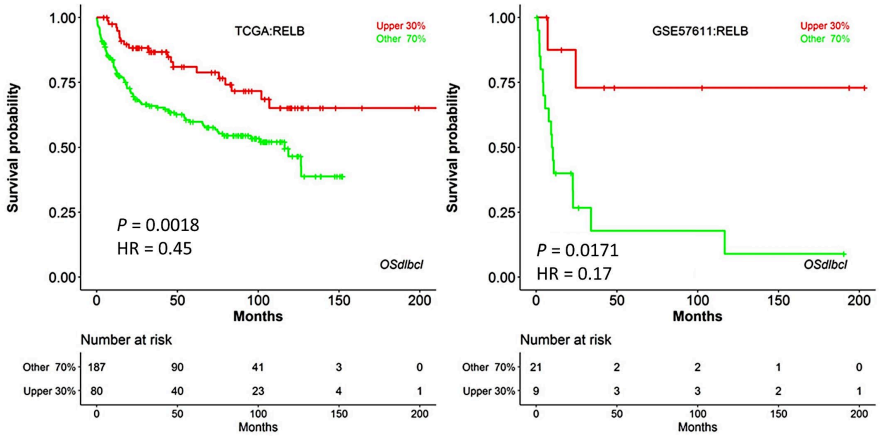

Figure A1. Validation of the association between *RELB* gene expression and overall survival in other series.

Appendix B

Overall accuracy is the percentage of records for which the outcome is correctly predicted.

The formula is as follows:

$$a = \frac{\sum_{i=1}^{n} m(i)}{n} \cdot 100\%,\ m(i) = \begin{cases} 1, & \text{if}(\hat{x_i} = x_i) \\ 0, & \text{otherwise} \end{cases}$$

where $\hat{x_i}$ is the predicted outcome value for record i and x_i is the observed value.

References

1. Swerdlow, S.H.; Campo, E.; Harris, N.L.; Jaffe, E.S.; Pileri, S.A.; Stein, H.; Thiele, J.; Vardiman, J.W. (Eds.) *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, 4th ed.; International Agency for Research on Cancer (IARC): Lyon, France, 2017.
2. Campo, E.; Jaffe, E.S.; Cook, J.R.; Quintanilla-Martinez, L.; Swerdlow, S.H.; Anderson, K.C.; Brousset, P.; Cerroni, L.; de Leval, L.; Dirnhofer, S.; et al. The International Consensus Classification of Mature Lymphoid Neoplasms: A report from the Clinical Advisory Committee. *Blood* **2022**, *140*, 1229–1253. [CrossRef]
3. Cazzola, M.; Sehn, L.H. Developing a classification of hematologic neoplasms in the era of precision medicine. *Blood* **2022**, *140*, 1193–1199. [CrossRef]
4. De Leval, L.; Alizadeh, A.A.; Bergsagel, P.L.; Campo, E.; Davies, A.; Dogan, A.; Fitzgibbon, J.; Horwitz, S.M.; Melnick, A.M.; Morice, W.G.; et al. Genomic profiling for clinical decision making in lymphoid neoplasms. *Blood* **2022**, *140*, 2193–2227. [CrossRef]
5. Alaggio, R.; Amador, C.; Anagnostopoulos, I.; Attygalle, A.D.; Araujo, I.B.O.; Berti, E.; Bhagat, G.; Borges, A.M.; Boyer, D.; Calaminici, M.; et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia* **2022**, *36*, 1720–1748. [CrossRef]
6. Quintanilla-Martinez, L.; Swerdlow, S.H.; Tousseyn, T.; Barrionuevo, C.; Nakamura, S.; Jaffe, E.S. New concepts in EBV-associated B, T, and NK cell lymphoproliferative disorders. *Virchows Arch.* **2023**, *482*, 227–244. [CrossRef]
7. Laurent, C.; Cook, J.R.; Yoshino, T.; Quintanilla-Martinez, L.; Jaffe, E.S. Follicular lymphoma and marginal zone lymphoma: How many diseases? *Virchows Arch.* **2023**, *482*, 149–162. [CrossRef]
8. Kurz, K.S.; Kalmbach, S.; Ott, M.; Staiger, A.M.; Ott, G.; Horn, H. Follicular Lymphoma in the 5th Edition of the WHO-Classification of Haematolymphoid Neoplasms-Updated Classification and New Biological Data. *Cancers* **2023**, *15*, 785. [CrossRef]
9. Gianelli, U.; Thiele, J.; Orazi, A.; Gangat, N.; Vannucchi, A.M.; Tefferi, A.; Kvasnicka, H.M. International Consensus Classification of myeloid and lymphoid neoplasms: Myeloproliferative neoplasms. *Virchows Arch.* **2023**, *482*, 53–68. [CrossRef]
10. De Leval, L.; Feldman, A.L.; Pileri, S.; Nakamura, S.; Gaulard, P. Extranodal T- and NK-cell lymphomas. *Virchows Arch.* **2023**, *482*, 245–264. [CrossRef]
11. Coupland, S.E.; Du, M.Q.; Ferry, J.A.; de Jong, D.; Khoury, J.D.; Leoncini, L.; Naresh, K.N.; Ott, G.; Siebert, R.; Xerri, L.; et al. The fifth edition of the WHO classification of mature B-cell neoplasms: Open questions for research. *J. Pathol.* **2024**, *262*, 255–270. [CrossRef]
12. Carreras, J.; Nakamura, N. Artificial Intelligence, Lymphoid Neoplasms, and Prediction of MYC, BCL2, and BCL6 Gene Expression Using a Pan-Cancer Panel in Diffuse Large B-Cell Lymphoma. *Hemato* **2024**, *5*, 119–143. [CrossRef]
13. Jaffe, E.S.; Carbone, A. B- and T-/NK-Cell Lymphomas in the 2022 International Consensus Classification of Mature Lymphoid Neoplasms and Comparison with the WHO Fifth Edition. *Hemato* **2024**, *5*, 157–170. [CrossRef]
14. Morton, L.M.; Wang, S.S.; Devesa, S.S.; Hartge, P.; Weisenburger, D.D.; Linet, M.S. Lymphoma incidence patterns by WHO subtype in the United States, 1992–2001. *Blood* **2006**, *107*, 265–276. [CrossRef]
15. Smith, A.; Howell, D.; Patmore, R.; Jack, A.; Roman, E. Incidence of haematological malignancy by sub-type: A report from the Haematological Malignancy Research Network. *Br. J. Cancer* **2011**, *105*, 1684–1692. [CrossRef]
16. Sant, M.; Allemani, C.; Tereanu, C.; De Angelis, R.; Capocaccia, R.; Visser, O.; Marcos-Gragera, R.; Maynadie, M.; Simonetti, A.; Lutz, J.M.; et al. Incidence of hematologic malignancies in Europe by morphologic subtype: Results of the HAEMACARE project. *Blood* **2010**, *116*, 3724–3734. [CrossRef]
17. Shirley, M.H.; Sayeed, S.; Barnes, I.; Finlayson, A.; Ali, R. Incidence of haematological malignancies by ethnic group in England, 2001–2007. *Br. J. Haematol.* **2013**, *163*, 465–477. [CrossRef]
18. Chadburn, A.; Gloghini, A.; Carbone, A. Classification of B-Cell Lymphomas and Immunodeficiency-Related Lymphoproliferations: What's New? *Hemato* **2023**, *4*, 26–41. [CrossRef]
19. De Leval, L.; Jaffe, E.S. Lymphoma Classification. *Cancer J.* **2020**, *26*, 176–185. [CrossRef]
20. Ricard, F.; Cheson, B.; Barrington, S.; Trotman, J.; Schmid, A.; Brueggenwerth, G.; Salles, G.; Schwartz, L.; Goldmacher, G.; Jarecha, R.; et al. Application of the Lugano Classification for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The PRoLoG Consensus Initiative (Part 1-Clinical). *J. Nucl. Med.* **2023**, *64*, 102–108. [CrossRef]
21. Hartmann, S.; Fend, F. Classification of Hodgkin lymphoma and related entities: News and open questions. *Pathologie* **2023**, *44*, 184–192. [CrossRef]
22. Shimkus, G.; Nonaka, T. Molecular classification and therapeutics in diffuse large B-cell lymphoma. *Front. Mol. Biosci.* **2023**, *10*, 1124360. [CrossRef]

23. Goodlad, J.R.; Cerroni, L.; Swerdlow, S.H. Recent advances in cutaneous lymphoma-implications for current and future classifications. *Virchows Arch.* **2023**, *482*, 281–298. [CrossRef]
24. King, R.L.; Hsi, E.D.; Chan, W.C.; Piris, M.A.; Cook, J.R.; Scott, D.W.; Swerdlow, S.H. Diagnostic approaches and future directions in Burkitt lymphoma and high-grade B-cell lymphoma. *Virchows Arch.* **2023**, *482*, 193–205. [CrossRef]
25. Kurz, K.S.; Ott, M.; Kalmbach, S.; Steinlein, S.; Kalla, C.; Horn, H.; Ott, G.; Staiger, A.M. Large B-Cell Lymphomas in the 5th Edition of the WHO-Classification of Haematolymphoid Neoplasms-Updated Classification and New Concepts. *Cancers* **2023**, *15*, 2285. [CrossRef]
26. Carreras, J. The pathobiology of follicular lymphoma. *J. Clin. Exp. Hematopathol.* **2023**, *63*, 152–163. [CrossRef]
27. Rosenwald, A.; Menter, T.; Dirnhofer, S. Classification of aggressive B-cell lymphomas: News and open questions. *Pathologie* **2023**, *44*, 166–172. [CrossRef]
28. Rodriguez-Pinilla, S.M.; Dojcinov, S.; Dotlic, S.; Gibson, S.E.; Hartmann, S.; Klimkowska, M.; Sabattini, E.; Tousseyn, T.A.; de Jong, D.; Hsi, E.D. Aggressive B-cell non-Hodgkin lymphomas: A report of the lymphoma workshop of the 20th meeting of the European Association for Haematopathology. *Virchows Arch.* **2024**, *484*, 15–29. [CrossRef]
29. Attygalle, A.D.; Chan, J.K.C.; Coupland, S.E.; Du, M.Q.; Ferry, J.A.; Jong, D.; Gratzinger, D.; Lim, M.S.; Naresh, K.N.; Nicolae, A.; et al. The 5th edition of the World Health Organization Classification of mature lymphoid and stromal tumors—An overview and update. *Leuk. Lymphoma* **2024**, *65*, 413–429. [CrossRef]
30. Arber, D.A.; Campo, E.; Jaffe, E.S. Advances in the Classification of Myeloid and Lymphoid Neoplasms. *Virchows Arch.* **2023**, *482*, 1–9. [CrossRef]
31. Song, J.Y.; Dirnhofer, S.; Piris, M.A.; Quintanilla-Martinez, L.; Pileri, S.; Campo, E. Diffuse large B-cell lymphomas, not otherwise specified, and emerging entities. *Virchows Arch.* **2023**, *482*, 179–192. [CrossRef]
32. Campo, E. The 2022 classifications of lymphoid neoplasms: Keynote. *Pathologie* **2023**, *44*, 121–127. [CrossRef]
33. Liu, Y.; Barta, S.K. Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. *Am. J. Hematol.* **2019**, *94*, 604–616. [CrossRef]
34. Ruppert, A.S.; Dixon, J.G.; Salles, G.; Wall, A.; Cunningham, D.; Poeschel, V.; Haioun, C.; Tilly, H.; Ghesquieres, H.; Ziepert, M.; et al. International prognostic indices in diffuse large B-cell lymphoma: A comparison of IPI, R-IPI, and NCCN-IPI. *Blood* **2020**, *135*, 2041–2048. [CrossRef]
35. Zhou, Z.; Sehn, L.H.; Rademaker, A.W.; Gordon, L.I.; Lacasce, A.S.; Crosby-Thompson, A.; Vanderplas, A.; Zelenetz, A.D.; Abel, G.A.; Rodriguez, M.A.; et al. An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. *Blood* **2014**, *123*, 837–842. [CrossRef]
36. Wright, G.W.; Huang, D.W.; Phelan, J.D.; Coulibaly, Z.A.; Roulland, S.; Young, R.M.; Wang, J.Q.; Schmitz, R.; Morin, R.D.; Tang, J.; et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. *Cancer Cell* **2020**, *37*, 551–568.e514. [CrossRef]
37. What Is Artificial Intelligence? IBM Topics Artificial-Intelligence. Available online: https://www.ibm.com/topics/artificial-intelligence (accessed on 22 January 2024).
38. Deep Learning vs. Machine Learning. IBM Topics Artificial-Intelligence. Available online: https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks (accessed on 22 January 2024).
39. What Is Unsupervised Learning? IBM Topics Unsupervised-Learning. Available online: https://www.ibm.com/topics/unsupervised-learning (accessed on 22 January 2024).
40. What Is Reinforcement Learning? IBM Developer. Available online: https://developer.ibm.com/learningpaths/get-started-automated-ai-for-decision-making-api/what-is-automated-ai-for-decision-making/ (accessed on 22 January 2024).
41. Prince, S.J.D. *Understanding Deep Learning*; MIT Press: Cambridge, MA, USA, 2023.
42. McLachlan, G.J.; Bean, R.W.; Ng, S.K. Clustering. *Methods Mol. Biol.* **2017**, *1526*, 345–362. [CrossRef]
43. Orsoni, M.; Giovagnoli, S.; Garofalo, S.; Magri, S.; Benvenuti, M.; Mazzoni, E.; Benassi, M. Preliminary evidence on machine learning approaches for clusterizing students' cognitive profile. *Heliyon* **2023**, *9*, e14506. [CrossRef]
44. Zampighi, L.M.; Kavanau, C.L.; Zampighi, G.A. The Kohonen self-organizing map: A tool for the clustering and alignment of single particles imaged using random conical tilt. *J. Struct. Biol.* **2004**, *146*, 368–380. [CrossRef]
45. RStudio. Anomaly Detection in R (DataCamp), Ch. 1—Statistical Outlier Detection. Available online: https://rpubs.com/michaelmallari/anomaly-detection-r (accessed on 22 January 2024).
46. Developer, I. Anomaly Detection. Available online: https://developer.ibm.com/apis/catalog/ai4industry--anomaly-detection-product/Introduction (accessed on 22 January 2024).
47. Corporation, I. *IBM Business Predictive Analytics, Algorithms Guide*; IBM Software Group 1994; IBM Corporation: Armonk, NY, USA, 2021.
48. Cardesa-Salzmann, T.M.; Colomo, L.; Gutierrez, G.; Chan, W.C.; Weisenburger, D.; Climent, F.; Gonzalez-Barca, E.; Mercadal, S.; Arenillas, L.; Serrano, S.; et al. High microvessel density determines a poor outcome in patients with diffuse large B-cell lymphoma treated with rituximab plus chemotherapy. *Haematologica* **2011**, *96*, 996–1001. [CrossRef]
49. Lenz, G.; Wright, G.; Dave, S.S.; Xiao, W.; Powell, J.; Zhao, H.; Xu, W.; Tan, B.; Goldschmidt, N.; Iqbal, J.; et al. Stromal gene signatures in large-B-cell lymphomas. *N. Engl. J. Med.* **2008**, *359*, 2313–2323. [CrossRef]

50. Carreras, J.; Kikuti, Y.Y.; Miyaoka, M.; Roncador, G.; Garcia, J.F.; Hiraiwa, S.; Tomita, S.; Ikoma, H.; Kondo, Y.; Ito, A.; et al. Integrative Statistics, Machine Learning and Artificial Intelligence Neural Network Analysis Correlated CSF1R with the Prognosis of Diffuse Large B-Cell Lymphoma. *Hemato* **2021**, *2*, 182–206. [CrossRef]
51. Carreras, J. Supplementary Data 2 (Version 2). Zenodo. 2024. Available online: https://zenodo.org/records/11058101 (accessed on 24 April 2024).
52. Carreras, J. Artificial Intelligence Analysis of Ulcerative Colitis Using an Autoimmune Discovery Transcriptomic Panel. *Healthcare* **2022**, *10*, 1476. [CrossRef]
53. Carreras, J. Artificial Intelligence Analysis of Celiac Disease Using an Autoimmune Discovery Transcriptomic Panel Highlighted Pathogenic Genes including BTLA. *Healthcare* **2022**, *10*, 1550. [CrossRef]
54. Carreras, J.; Hamoudi, R.; Nakamura, N. Artificial Intelligence Analysis of Gene Expression Data Predicted the Prognosis of Patients with Diffuse Large B-Cell Lymphoma. *Tokai J. Exp. Clin. Med.* **2020**, *45*, 37–48.
55. Carreras, J.; Hiraiwa, S.; Kikuti, Y.Y.; Miyaoka, M.; Tomita, S.; Ikoma, H.; Ito, A.; Kondo, Y.; Roncador, G.; Garcia, J.F.; et al. Artificial Neural Networks Predicted the Overall Survival and Molecular Subtypes of Diffuse Large B-Cell Lymphoma Using a Pancancer Immune-Oncology Panel. *Cancers* **2021**, *13*, 6384. [CrossRef]
56. Carreras, J.; Nakamura, N.; Hamoudi, R. Artificial Intelligence Analysis of Gene Expression Predicted the Overall Survival of Mantle Cell Lymphoma and a Large Pan-Cancer Series. *Healthcare* **2022**, *10*, 155. [CrossRef]
57. Carreras, J.; Roncador, G.; Hamoudi, R. Artificial Intelligence Predicted Overall Survival and Classified Mature B-Cell Neoplasms Based on Immuno-Oncology and Immune Checkpoint Panels. *Cancers* **2022**, *14*, 5318. [CrossRef]
58. Carreras, J.; Yukie Kikuti, Y.; Miyaoka, M.; Miyahara, S.; Roncador, G.; Hamoudi, R.; Nakamura, N. Artificial Intelligence Analysis and Reverse Engineering of Molecular Subtypes of Diffuse Large B-Cell Lymphoma Using Gene Expression Data. *BioMedInformatics* **2024**, *4*, 295–320. [CrossRef]
59. Carreras, J.; Kikuti, Y.Y.; Miyaoka, M.; Hiraiwa, S.; Tomita, S.; Ikoma, H.; Kondo, Y.; Ito, A.; Nagase, S.; Miura, H.; et al. Mutational Profile and Pathological Features of a Case of Interleukin-10 and RGS1-Positive Spindle Cell Variant Diffuse Large B-Cell Lymphoma. *Hematol. Rep.* **2023**, *15*, 188–200. [CrossRef]
60. Carreras, J.; Kikuti, Y.Y.; Hiraiwa, S.; Miyaoka, M.; Tomita, S.; Ikoma, H.; Ito, A.; Kondo, Y.; Itoh, J.; Roncador, G.; et al. High PTX3 expression is associated with a poor prognosis in diffuse large B-cell lymphoma. *Cancer Sci.* **2022**, *113*, 334–348. [CrossRef]
61. Li, Q.; Dou, M.; Zhang, J.; Jia, P.; Wang, X.; Lei, D.; Li, J.; Yang, W.; Yang, R.; Yang, C.; et al. A Bayesian network model to predict neoplastic risk for patients with gallbladder polyps larger than 10 mm based on preoperative ultrasound features. *Surg. Endosc.* **2023**, *37*, 5453–5463. [CrossRef]
62. C5.0 Node. Available online: https://www.ibm.com/us-en (accessed on 25 April 2024).
63. Asadi, F.; Salehnasab, C.; Ajori, L. Supervised Algorithms of Machine Learning for the Prediction of Cervical Cancer. *J. Biomed. Phys. Eng.* **2020**, *10*, 513–522. [CrossRef]
64. Bottel, L.; Brand, M.; Dieris-Hirche, J.; Pape, M.; Herpertz, S.; Te Wildt, B.T. Predictive power of the DSM-5 criteria for internet use disorder: A CHAID decision-tree analysis. *Front. Psychol.* **2023**, *14*, 1129769. [CrossRef]
65. Diaz-Perez, F.M.; Garcia-Gonzalez, C.G.; Fyall, A. The use of the CHAID algorithm for determining tourism segmentation: A purposeful outcome. *Heliyon* **2020**, *6*, e04256. [CrossRef]
66. Kaya, S.; Guven, G.S.; Aydan, S.; Toka, O. A comprehensive framework identifying readmission risk factors using the CHAID algorithm: A prospective cohort study. *Int. J. Qual. Health Care* **2018**, *30*, 366–374. [CrossRef]
67. Meydanlioglu, A.; Akcan, A.; Oncel, S.; Adibelli, D.; Cicek Gumus, E.; Sarvan, S.; Kavla, I. Prevalence of obesity and hypertension in children and determination of associated factors by CHAID analysis. *Arch. Pediatr.* **2022**, *29*, 30–35. [CrossRef]
68. Murphy, E.L.; Comiskey, C.M. Using chi-Squared Automatic Interaction Detection (CHAID) modelling to identify groups of methadone treatment clients experiencing significantly poorer treatment outcomes. *J. Subst. Abus. Treat.* **2013**, *45*, 343–349. [CrossRef]
69. Solberg, H.E. Discriminant analysis. *CRC Crit. Rev. Clin. Lab. Sci.* **1978**, *9*, 209–242. [CrossRef]
70. Chan, Y.H. Biostatistics 303. Discriminant analysis. *Singap. Med. J.* **2005**, *46*, 54–61, quiz 62.
71. Carreras, J. KNN Algorithms (Version 1). Zenodo. 2024. Available online: https://zenodo.org/records/11058452 (accessed on 24 April 2024).
72. Lu, J.; Chen, Q.; Li, D.; Zhang, W.; Xing, S.; Wang, J.; Zhang, X.; Liu, J.; Qing, Z.; Dai, Y.; et al. Reconfiguration of Dynamic Functional Connectivity States in Patients With Lifelong Premature Ejaculation. *Front. Neurosci.* **2021**, *15*, 721236. [CrossRef]
73. Arabi, E.M.; Ahmed, K.S.; Mohra, A.S. Advanced Diagnostic Technique for Alzheimer's Disease using MRI Top-Ranked Volume and Surface-based Features. *J. Biomed. Phys. Eng.* **2022**, *12*, 569–582. [CrossRef]
74. Ali, R.; Hussain, J.; Lee, S.W. Multilayer perceptron-based self-care early prediction of children with disabilities. *Digit. Health* **2023**, *9*, 20552076231184054. [CrossRef]
75. Ivanov, A.S.; Nikolaev, K.G.; Novikov, A.S.; Yurchenko, S.O.; Novoselov, K.S.; Andreeva, D.V.; Skorb, E.V. Programmable Soft-Matter Electronics. *J. Phys. Chem. Lett.* **2021**, *12*, 2017–2022. [CrossRef]
76. Majidzadeh Gorjani, O.; Byrtus, R.; Dohnal, J.; Bilik, P.; Koziorek, J.; Martinek, R. Human Activity Classification Using Multilayer Perceptron. *Sensors* **2021**, *21*, 6207. [CrossRef]
77. Lyu, J.; Shi, H.; Zhang, J.; Norvilitis, J. Prediction model for suicide based on back propagation neural network and multilayer perceptron. *Front. Neuroinform.* **2022**, *16*, 961588. [CrossRef]

78. Fujita, T.; Sato, A.; Narita, A.; Sone, T.; Iokawa, K.; Tsuchiya, K.; Yamane, K.; Yamamoto, Y.; Ohira, Y.; Otsuki, K. Use of a multilayer perceptron to create a prediction model for dressing independence in a small sample at a single facility. *J. Phys. Ther. Sci.* **2019**, *31*, 69–74. [CrossRef]
79. Radhakrishnan, S.; Nair, S.G.; Isaac, J. Multilayer perceptron neural network model development for mechanical ventilator parameters prediction by real time system learning. *Biomed. Signal Process. Control* **2022**, *71*, 103170. [CrossRef]
80. Rigatti, S.J. Random Forest. *J. Insur. Med.* **2017**, *47*, 31–39. [CrossRef]
81. Rhodes, J.S.; Cutler, A.; Moon, K.R. Geometry- and Accuracy-Preserving Random Forest Proximities. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10947–10959. [CrossRef]
82. Asadi, S.; Roshan, S.; Kattan, M.W. Random forest swarm optimization-based for heart diseases diagnosis. *J. Biomed. Inform.* **2021**, *115*, 103690. [CrossRef]
83. Elbeltagi, A.; Pande, C.B.; Kumar, M.; Tolche, A.D.; Singh, S.K.; Kumar, A.; Vishwakarma, D.K. Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models. *Environ. Sci. Pollut. Res. Int.* **2023**, *30*, 43183–43202. [CrossRef]
84. Mehta, S.D.; Sebro, R. Computer-Aided Detection of Incidental Lumbar Spine Fractures from Routine Dual-Energy X-Ray Absorptiometry (DEXA) Studies Using a Support Vector Machine (SVM) Classifier. *J. Digit. Imaging* **2020**, *33*, 204–210. [CrossRef]
85. Han, H.; Jiang, X. Overcome support vector machine diagnosis overfitting. *Cancer Inform.* **2014**, *13*, 145–158. [CrossRef]
86. Yehuda, B.; Rabinowich, A.; Link-Sourani, D.; Avisdris, N.; Ben-Zvi, O.; Specktor-Fadida, B.; Joskowicz, L.; Ben-Sira, L.; Miller, E.; Ben Bashat, D. Automatic Quantification of Normal Brain Gyrification Patterns and Changes in Fetuses with Polymicrogyria and Lissencephaly Based on MRI. *AJNR Am. J. Neuroradiol.* **2023**, *44*, 1432–1439. [CrossRef]
87. Raubitzek, S.; Neubauer, T. An Exploratory Study on the Complexity and Machine Learning Predictability of Stock Market Data. *Entropy* **2022**, *24*, 332. [CrossRef]
88. Thedinga, K.; Herwig, R. A gradient tree boosting and network propagation derived pan-cancer survival network of the tumor microenvironment. *iScience* **2022**, *25*, 103617. [CrossRef]
89. Thedinga, K.; Herwig, R. Gradient tree boosting and network propagation for the identification of pan-cancer survival networks. *STAR Protoc.* **2022**, *3*, 101353. [CrossRef]
90. Pfob, A.; Sidey-Gibbons, C.; Lee, H.B.; Tasoulis, M.K.; Koelbel, V.; Golatta, M.; Rauch, G.M.; Smith, B.D.; Valero, V.; Han, W.; et al. Identification of breast cancer patients with pathologic complete response in the breast after neoadjuvant systemic treatment by an intelligent vacuum-assisted biopsy. *Eur. J. Cancer* **2021**, *143*, 134–146. [CrossRef]
91. Nistal-Nuno, B. Machine learning applied to a Cardiac Surgery Recovery Unit and to a Coronary Care Unit for mortality prediction. *J. Clin. Monit. Comput.* **2022**, *36*, 751–763. [CrossRef]
92. Tran, T.; Le, U.; Shi, Y. An effective up-sampling approach for breast cancer prediction with imbalanced data: A machine learning model-based comparative analysis. *PLoS ONE* **2022**, *17*, e0269135. [CrossRef]
93. Pfob, A.; Mehrara, B.J.; Nelson, J.A.; Wilkins, E.G.; Pusic, A.L.; Sidey-Gibbons, C. Machine learning to predict individual patient-reported outcomes at 2-year follow-up for women undergoing cancer-related mastectomy and breast reconstruction (INSPiRED-001). *Breast* **2021**, *60*, 111–122. [CrossRef]
94. Janjua, H.; Barry, T.M.; Cousin-Peterson, E.; Kuo, P.C. Defining the relative contribution of health care environmental components to patient outcomes in the model of 30-day readmission after coronary artery bypass graft (CABG). *Surgery* **2021**, *169*, 557–566. [CrossRef]
95. Dominguez-Gutierrez, P.R.; Kwenda, E.P.; Donelan, W.; O'Malley, P.; Crispen, P.L.; Kusmartsev, S. Hyal2 Expression in Tumor-Associated Myeloid Cells Mediates Cancer-Related Inflammation in Bladder Cancer. *Cancer Res.* **2021**, *81*, 648–657. [CrossRef]
96. Tan, X.; Cao, F.; Tang, F.; Lu, C.; Yu, Q.; Feng, S.; Yang, Z.; Chen, S.; He, X.; He, J.; et al. Suppression of DLBCL Progression by the E3 Ligase Trim35 Is Mediated by CLOCK Degradation and NK Cell Infiltration. *J. Immunol. Res.* **2021**, *2021*, 9995869. [CrossRef]
97. Wang, R.; Huang, K.L.; Xing, L.X. TRIM35 functions as a novel tumor suppressor in breast cancer by inducing cell apoptosis through ubiquitination of PDK1. *Neoplasma* **2022**, *69*, 370–382. [CrossRef]
98. Dang, W.; Cao, P.; Yan, Q.; Yang, L.; Wang, Y.; Yang, J.; Xin, S.; Zhang, J.; Li, J.; Long, S.; et al. IGFBP7-AS1 is a p53-responsive long noncoding RNA downregulated by Epstein-Barr virus that contributes to viral tumorigenesis. *Cancer Lett.* **2021**, *523*, 135–147. [CrossRef]
99. Wu, S.G.; Chang, T.H.; Tsai, M.F.; Liu, Y.N.; Hsu, C.L.; Chang, Y.L.; Yu, C.J.; Shih, J.Y. IGFBP7 Drives Resistance to Epidermal Growth Factor Receptor Tyrosine Kinase Inhibition in Lung Cancer. *Cancers* **2019**, *11*, 36. [CrossRef]
100. De Araujo, M.E.; Erhart, G.; Buck, K.; Muller-Holzner, E.; Hubalek, M.; Fiegl, H.; Campa, D.; Canzian, F.; Eilber, U.; Chang-Claude, J.; et al. Polymorphisms in the gene regions of the adaptor complex LAMTOR2/LAMTOR3 and their association with breast cancer risk. *PLoS ONE* **2013**, *8*, e53768. [CrossRef]
101. Zhang, S.; Liu, Y.; Chen, J.; Shu, H.; Shen, S.; Li, Y.; Lu, X.; Cao, X.; Dong, L.; Shi, J.; et al. Autoantibody signature in hepatocellular carcinoma using seromics. *J. Hematol. Oncol.* **2020**, *13*, 85. [CrossRef]
102. Luo, L.; Li, L.; Liu, L.; Feng, Z.; Zeng, Q.; Shu, X.; Cao, Y.; Li, Z. A Necroptosis-Related lncRNA-Based Signature to Predict Prognosis and Probe Molecular Characteristics of Stomach Adenocarcinoma. *Front. Genet.* **2022**, *13*, 833928. [CrossRef]

103. Eluard, B.; Nuan-Aliman, S.; Faumont, N.; Collares, D.; Bordereaux, D.; Montagne, A.; Martins, I.; Cagnard, N.; Caly, M.; Taoui, O.; et al. The alternative RelB NF-kappaB subunit is a novel critical player in diffuse large B-cell lymphoma. *Blood* **2022**, *139*, 384–398. [CrossRef]
104. Nuan-Aliman, S.; Bordereaux, D.; Thieblemont, C.; Baud, V. The Alternative RelB NF-kB Subunit Exerts a Critical Survival Function upon Metabolic Stress in Diffuse Large B-Cell Lymphoma-Derived Cells. *Biomedicines* **2022**, *10*, 348. [CrossRef]
105. Sha, C.; Barrans, S.; Cucco, F.; Bentley, M.A.; Care, M.A.; Cummin, T.; Kennedy, H.; Thompson, J.S.; Uddin, R.; Worrillow, L.; et al. Molecular High-Grade B-Cell Lymphoma: Defining a Poor-Risk Group That Requires Different Approaches to Therapy. *J. Clin. Oncol.* **2019**, *37*, 202–212. [CrossRef]
106. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [CrossRef]
107. Mootha, V.K.; Lindgren, C.M.; Eriksson, K.F.; Subramanian, A.; Sihag, S.; Lehar, J.; Puigserver, P.; Carlsson, E.; Ridderstrale, M.; Laurila, E.; et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **2003**, *34*, 267–273. [CrossRef]
108. Broad Institute, Inc. Massachusetts Institute of Technology; Regents of the University of California. Gene Set Enrichment Analysis. Available online: https://www.gsea-msigdb.org/gsea/index.jsp (accessed on 23 April 2024).
109. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S.; et al. The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **2023**, *51*, D638–D646. [CrossRef]
110. SIB—Swiss Institute of BioInformatics; Novo Nordisk Foundation Center Protein Research; EMBL—European Molecular Biology Laboratory. STRING. Available online: https://string-db.org/ (accessed on 23 April 2024).
111. Kim, J.Y.; Cho, Y.E.; Park, J.H. The Nucleolar Protein GLTSCR2 Is an Upstream Negative Regulator of the Oncogenic Nucleophosmin-MYC Axis. *Am. J. Pathol.* **2015**, *185*, 2061–2068. [CrossRef]
112. Shi, Y.; Xu, X.; Zhang, Q.; Fu, G.; Mo, Z.; Wang, G.S.; Kishi, S.; Yang, X.L. tRNA synthetase counteracts c-Myc to develop functional vasculature. *eLife* **2014**, *3*, e02349. [CrossRef]
113. Barrans, S.; Crouch, S.; Smith, A.; Turner, K.; Owen, R.; Patmore, R.; Roman, E.; Jack, A. Rearrangement of MYC is associated with poor prognosis in patients with diffuse large B-cell lymphoma treated in the era of rituximab. *J. Clin. Oncol.* **2010**, *28*, 3360–3365. [CrossRef]
114. Kawasaki, C.; Ohshim, K.; Suzumiya, J.; Kanda, M.; Tsuchiya, T.; Tamura, K.; Kikuchi, M. Rearrangements of bcl-1, bcl-2, bcl-6, and c-myc in diffuse large B-cell lymphomas. *Leuk. Lymphoma* **2001**, *42*, 1099–1106. [CrossRef]
115. Stasik, C.J.; Nitta, H.; Zhang, W.; Mosher, C.H.; Cook, J.R.; Tubbs, R.R.; Unger, J.M.; Brooks, T.A.; Persky, D.O.; Wilkinson, S.T.; et al. Increased MYC gene copy number correlates with increased mRNA levels in diffuse large B-cell lymphoma. *Haematologica* **2010**, *95*, 597–603. [CrossRef]
116. Leveille, E.; Johnson, N.A. Genetic Events Inhibiting Apoptosis in Diffuse Large B Cell Lymphoma. *Cancers* **2021**, *13*, 2167. [CrossRef]
117. Odqvist, L.; Montes-Moreno, S.; Sanchez-Pacheco, R.E.; Young, K.H.; Martin-Sanchez, E.; Cereceda, L.; Sanchez-Verde, L.; Pajares, R.; Mollejo, M.; Fresno, M.F.; et al. NFkappaB expression is a feature of both activated B-cell-like and germinal center B-cell-like subtypes of diffuse large B-cell lymphoma. *Mod. Pathol.* **2014**, *27*, 1331–1337. [CrossRef]
118. Ok, C.Y.; Xu-Monette, Z.Y.; Li, L.; Manyam, G.C.; Montes-Moreno, S.; Tzankov, A.; Visco, C.; Dybkaer, K.; Routbort, M.J.; Zhang, L.; et al. Evaluation of NF-kappaB subunit expression and signaling pathway activation demonstrates that p52 expression confers better outcome in germinal center B-cell-like diffuse large B-cell lymphoma in association with CD30 and BCL2 functions. *Mod. Pathol.* **2015**, *28*, 1202–1213. [CrossRef]
119. Yu, L.; Li, L.; Medeiros, L.J.; Young, K.H. NF-kappaB signaling pathway and its potential as a target for therapy in lymphoid neoplasms. *Blood Rev.* **2017**, *31*, 77–92. [CrossRef]
120. Wu, J.; Yu, X.; Zhu, H.; Chen, P.; Liu, T.; Yin, R.; Qiang, Y.; Xu, L. RelB is a potential molecular biomarker for immunotherapy in human pan-cancer. *Front. Mol. Biosci.* **2023**, *10*, 1178446. [CrossRef]
121. Modi, D.; Potugari, B.; Uberti, J. Immunotherapy for Diffuse Large B-Cell Lymphoma: Current Landscape and Future Directions. *Cancers* **2021**, *13*, 5827. [CrossRef]
122. Zhang, J.; Medeiros, L.J.; Young, K.H. Cancer Immunotherapy in Diffuse Large B-Cell Lymphoma. *Front. Oncol.* **2018**, *8*, 351. [CrossRef]
123. Gasparini, C.; Celeghini, C.; Monasta, L.; Zauli, G. NF-kappaB pathways in hematological malignancies. *Cell. Mol. Life Sci.* **2014**, *71*, 2083–2102. [CrossRef]
124. Jayawant, E.; Pack, A.; Clark, H.; Kennedy, E.; Ghodke, A.; Jones, J.; Pepper, C.; Pepper, A.; Mitchell, S. NF-kappaB fingerprinting reveals heterogeneous NF-kappaB composition in diffuse large B-cell lymphoma. *Front. Oncol.* **2023**, *13*, 1181660. [CrossRef]
125. Lim, S.K.; Peng, C.C.; Low, S.; Vijay, V.; Budiman, A.; Phang, B.H.; Lim, J.Q.; Jeyasekharan, A.D.; Lim, S.T.; Ong, C.K.; et al. Sustained activation of non-canonical NF-kappaB signalling drives glycolytic reprogramming in doxorubicin-resistant DLBCL. *Leukemia* **2023**, *37*, 441–452. [CrossRef]

126. Oien, D.B.; Sharma, S.; Hattersley, M.M.; DuPont, M.; Criscione, S.W.; Prickett, L.; Goeppert, A.U.; Drew, L.; Yao, Y.; Zhang, J.; et al. BET inhibition targets ABC-DLBCL constitutive B-cell receptor signaling through PAX5. *Blood Adv.* **2023**, *7*, 5108–5121. [CrossRef]
127. Carreras, J.; Ikoma, H.; Kikuti, Y.Y.; Miyaoka, M.; Hiraiwa, S.; Tomita, S.; Kondo, Y.; Ito, A.; Nagase, S.; Miura, H.; et al. Mutational, immune microenvironment, and clinicopathological profiles of diffuse large B-cell lymphoma and follicular lymphoma with *BCL6* rearrangement. *Virchows Arch.* **2024**, *484*, 657–676. [CrossRef]