



Article

Development and Validation of a Novel Four Gene-Pairs Signature for Predicting Prognosis in DLBCL Patients

Atsushi Tanabe ^{1,*} , Jerry Ndzinu ^{2,3} and Hiroeki Sahara ²

¹ Laboratory of Highly-Advanced Veterinary Medical Technology, Veterinary Teaching Hospital, Azabu University, 1-17-71 Fuchinobe Chuo-ku, Sagami-hara 252-5201, Kanagawa, Japan

² Laboratory of Biology, Azabu University School of Veterinary Medicine, 1-17-71 Fuchinobe Chuo-ku, Sagami-hara 252-5201, Kanagawa, Japan; jerry@azabu-u.ac.jp (J.N.); sahara@azabu-u.ac.jp (H.S.)

³ Department of Research and Development (R&D), Malignant Tumor Treatment Technologies, Inc., 130-42 Nagasone, Kita-ku, Sakai 591-8025, Osaka, Japan

* Correspondence: at-tanabe@azabu-u.ac.jp; Tel.: +81-42-769-1842

Abstract: Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of non-Hodgkin's lymphoma. Because individual clinical outcomes of DLBCL in response to standard therapy differ widely, new treatment strategies are being investigated to improve therapeutic efficacy. In this study, we identified a novel signature for stratification of DLBCL useful for prognosis prediction and treatment selection. First, 408 prognostic gene sets were selected from approximately 2500 DLBCL samples in public databases, from which four gene-pair signatures consisting of seven prognostic genes were identified by Cox regression analysis. Then, the risk score was calculated based on these gene-pairs and we validated the risk score as a prognostic predictor for DLBCL patient outcomes. This risk score demonstrated independent predictive performance even when combined with other clinical parameters and molecular subtypes. Evaluating external DLBCL cohorts, we demonstrated that the risk-scoring model based the four gene-pair signatures leads to stable predictive performance, compared with nine existing predictive models. Finally, high-risk DLBCL showed high resistance to DNA damage caused by anticancer drugs, suggesting that this characteristic is responsible for the unfavorable prognosis of high-risk DLBCL patients. These results provide a novel index for classifying the biological characteristics of DLBCL and clearly indicate the importance of genetic analyses in the treatment of DLBCL.

Keywords: diffuse large B-cell lymphoma; public database; gene signature; prognosis; drug sensitivity



Citation: Tanabe, A.; Ndzinu, J.; Sahara, H. Development and Validation of a Novel Four Gene-Pairs Signature for Predicting Prognosis in DLBCL Patients. *Int. J. Mol. Sci.* **2024**, *25*, 12807. <https://doi.org/10.3390/ijms252312807>

Academic Editor: Alessandro Poggi

Received: 21 October 2024

Revised: 22 November 2024

Accepted: 25 November 2024

Published: 28 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Non-Hodgkin's lymphoma (NHL) refers to a group of blood cancers that arise from lymphoid tissue, primarily lymph nodes. NHL is a highly heterogeneous disease composed of various subtypes exhibiting diverse clinical and molecular features. Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of NHL and has its own highly heterogeneous biology. Standard primary treatment for DLBCL is R-CHOP, a combination of rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone. Although addition of the anti-CD20 antibody rituximab to standard therapy has greatly improved DLBCL outcomes, approximately 40% of patients may relapse or become refractory to therapy [1,2]. Various salvage therapies for relapsed/refractory DLBCL are currently being developed and investigated, but few patients achieve long-term remission [2,3].

The International Prognostic Index (IPI), which classifies DLBCL patients into different risk groups based on five factors (age, serum LDH, performance status, stage, and number of extranodal lesions), is the most prevalent prognostic predictor for patients with DLBCL, but the IPI alone does not describe the biological characteristics of DLBCL [4,5]. Since the late 1990s, various classification models using gene-expression profiling as a predictor have been developed in an effort to clarify the biological heterogeneity of DLBCL.

The most representative and clinically utilized classification model is cell of origin (COO) subtyping, which classifies DLBCL into two groups: germinal-center B-cell (GCB) type and activated B-cell (ABC) type [6]. In general, the prognosis of ABC-DLBCL is worse than that of GCB-DLBCL. Classification models that focus on identifying patterns of chromosomal translocations and specific gene mutations have also been developed in response to recent advancements in next-generation sequencing techniques [7,8]. In addition to determining the genetic characteristics of lymphoma cells, classification models have also been developed to characterize components of the lymphoma microenvironment, such as stromal cells and immune cells [9,10]. However, various improvements must be instituted in order to render these prognostic predictors useful in the patient care context, such as improving their accuracy and reproducibility, expanding their versatility, and reducing the clinical testing cost.

In the present study, we conducted a comprehensive analysis of gene-expression profiles in public databases to identify novel gene signatures for classifying the biological characteristics and prognosis of DLBCL, using a univariable Cox proportional hazard model to extract genes relevant to prognosis of DLBCL. A risk-scoring model we developed showed a superior predictive stability compared to other existing risk-scoring models. In addition, this risk-scoring model was found to be clearly associated with chemotherapy resistance of DLBCL.

2. Results

2.1. The Development of Four Gene-Pair Signature Model

To identify novel prognostic gene signatures for DLBCL, we examined eight public mRNA-expression datasets obtained from two public microarray platforms (five datasets from Affymetrix, three datasets from ILMN) (Figure 1). To correct for batch effects between datasets, datasets from the same platform were merged into DLBCL training dataset 1 and 2 using pyComBat (Figure S1A,B). After extracting data for patients treated with CHOP or R-CHOP from the integrated data, univariable Cox regression analysis was performed to identify prognostic genes associated with overall survival of DLBCL patients. Finally, 408 prognostic genes were identified based on overlap between the two training datasets (Figure 2A,B and Table S1). To examine the relationships between the 408 newly identified prognostic genes and existing prognostic genes for DLBCL, a literature review was conducted. When 408 prognostic genes were matched to 4575 DLBCL-related genes in signatureDB (<http://lymphochip.nih.gov/signaturedb/>, accessed on 13 September 2024) [11], 242 genes overlapped (Table S1). In addition, 48 genes overlapped with the 777 genes in the BLYM-777 gene panel, which is based on more than 45 DLBCL studies [12].

Next, these 408 prognostic genes were combined for a total of 41,616 gene-pairs. To screen for gene-pairs associated with the prognosis of DLBCL patients, we performed univariate Cox regression analysis again. A total of four gene-pairs were common to training datasets (Figure 2C–K, and Table S2). Detailed information on the seven prognostic genes in the four gene-pairs is provided in Table 1. Finally, a risk score was calculated by directly summing the binary scores of the four gene-pairs.

Table 1. Seven prognostic genes included in four gene-pairs.

Gene Symbol	Gene Name	Entrez ID	Probe ID	
			HG-U133_Plus_2	Illumina HumanHT-12 V4.0
NEK6	NIMA related kinase 6	10783	223158_s_at	ILMN_1660871
RARRES2	retinoic acid receptor responder 2	5919	209496_at	ILMN_1810844
CHRNA1	cholinergic receptor nicotinic alpha 1 subunit	1134	206633_at	ILMN_2361768
URI1(C19orf2)	URI1 prefoldin like chaperone	8725	214173_x_at	ILMN_2406892
NIPA2	NIPA magnesium transporter 2	81614	212129_at	ILMN_1720344
ZNF22	zinc finger protein 22	7570	218005_at	ILMN_1798533
WDR12	WD repeat domain 12	55759	218512_at	ILMN_1770692

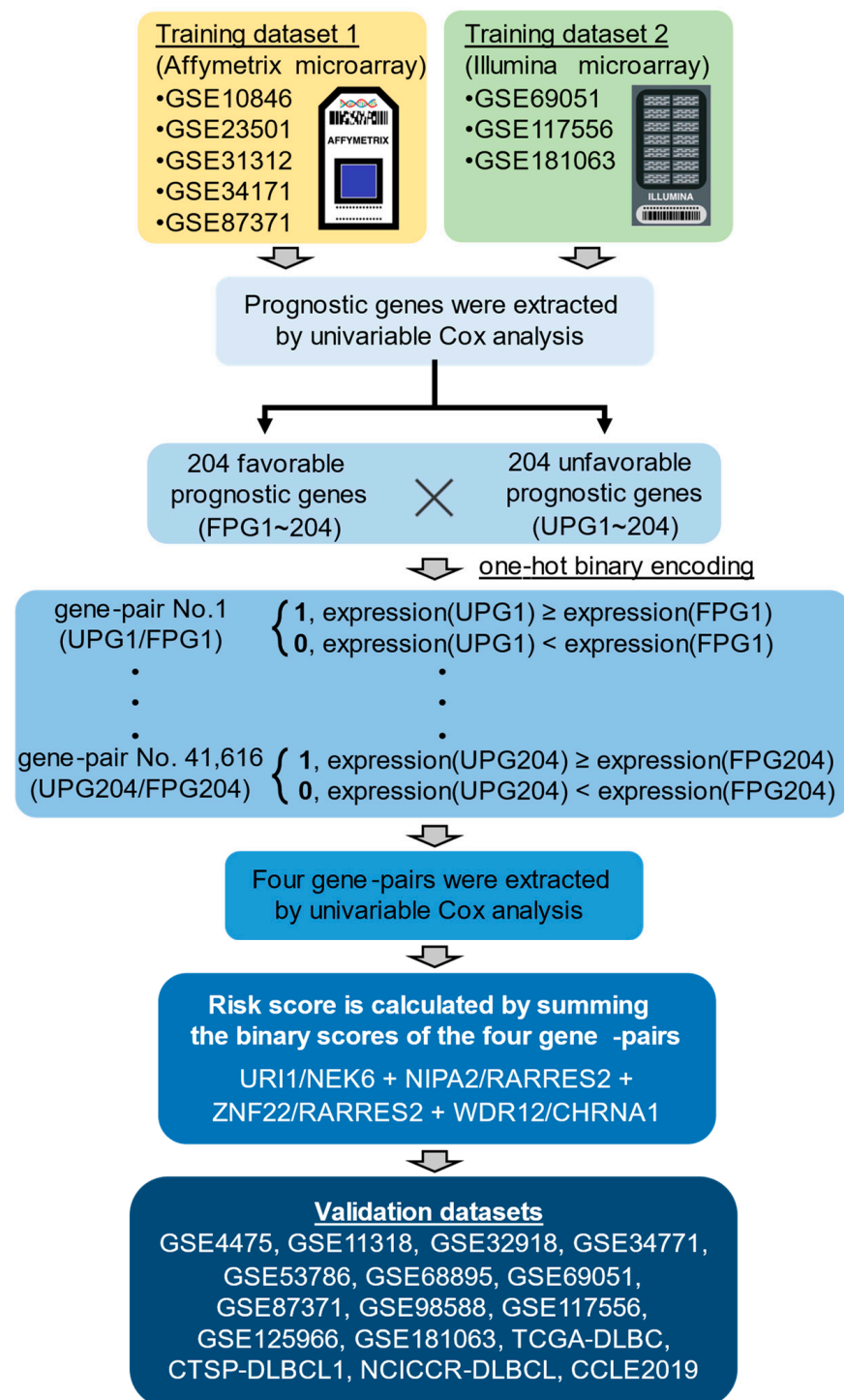


Figure 1. Workflow from training data integration to development of risk-scoring model. The training dataset 1 includes five Affymetrix microarray datasets and the training dataset 2 includes three Illumina microarray datasets were used for identification of prognostic genes. The risk-scoring model based on four prognostic gene-pairs was developed in the manner shown in Section 4.

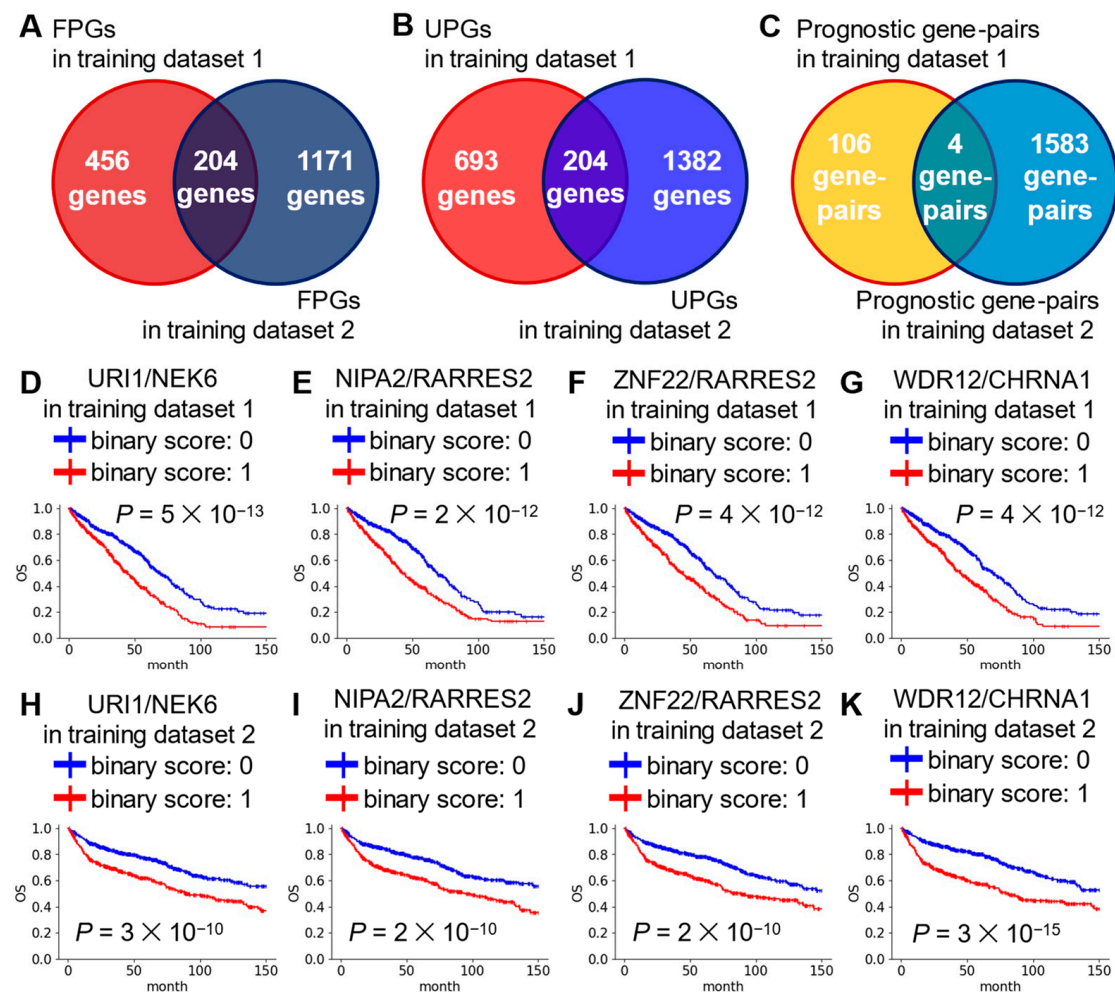


Figure 2. Screening of 408 prognostic genes and prognostic four gene-pairs. Venn diagrams summarizing the overlap of FPGs (A) or UPGs (B) or prognostic gene-pairs (C) between two training datasets. (D–K) KM survival curves showing OS in binary score groups of four gene-pairs. p values obtained using log-rank test.

2.2. Assessing the Performance of the Risk Score as a Prognostic Biomarker

To investigate predictive performance of the risk score, DLBCL patients in training datasets were divided into five groups based on the risk score. The KM survival curves showed that the OS was clearly reduced in the high-risk score group (Figure 3A,B). To confirm predictive performance of the risk score, we analyzed additional DLBCL datasets. Due to the lack of information on prognostic genes, the risk scores in GSE4475, GSE32918, and GSE68895 datasets were calculated using three, two, and two gene-pairs, respectively (Figure S2F,G,I). The KM survival curves showed that DLBCL patients in high-risk group had an unfavorable prognosis than those in the low-risk groups (Figures 3C,D and S2A–J). In addition, many patients with primary central nervous system lymphoma with unfavorable prognosis showed high-risk scores (Figure S3). These results suggested that the risk score based on the four gene-pairs is a useful biomarker for classifying prognostic differences among DLBCL patients.

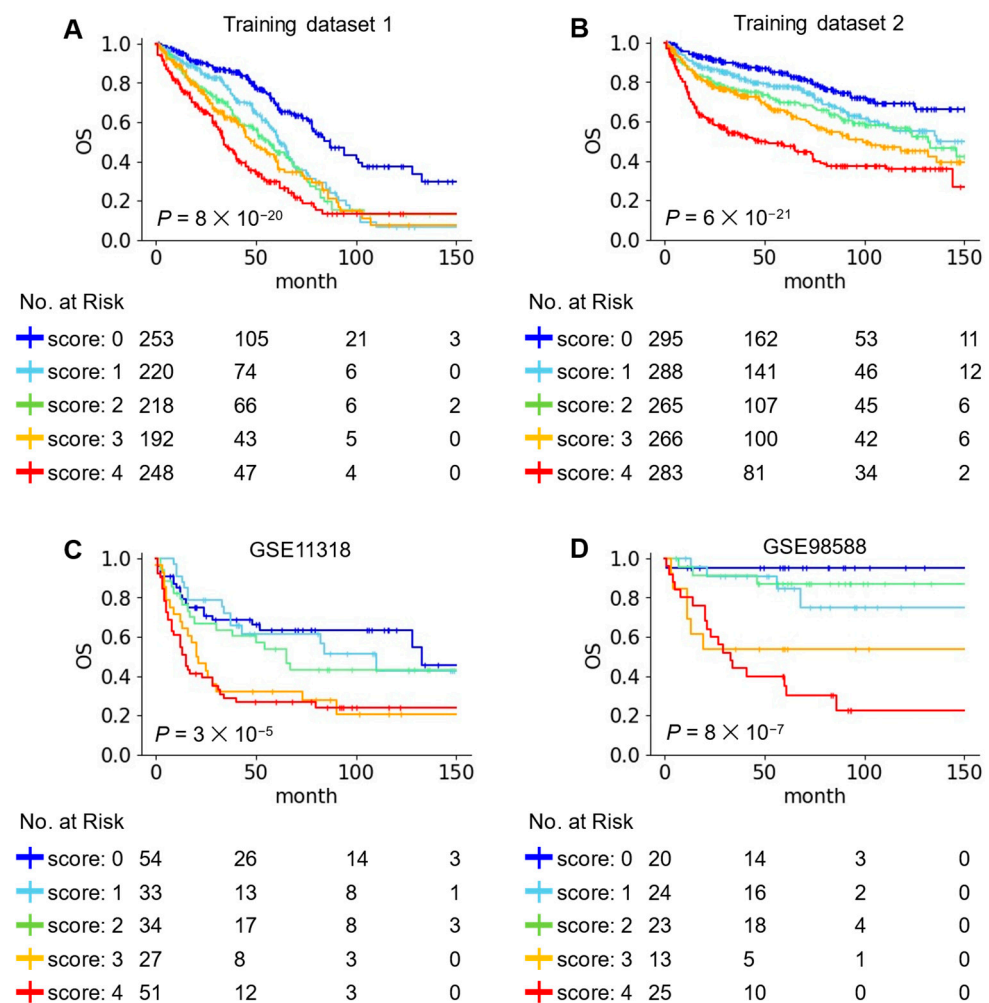


Figure 3. The predictive performance of the risk score in the training and validation datasets. (A–D) KM survival curves showing OS of five risk groups. p values obtained using log-rank test.

2.3. Predictive Performance of the Risk Score in Conditions Stratified by Clinical Characteristics

To investigate the relationship between the risk score and clinical characteristics of DLBCL, we examined predictive performances of the risk score in DLBCL patients with different clinical characteristics. In this analysis, all of DLBCL patients in GSE117556 and GSE181063 datasets were integrated into one dataset. First, we compared the predictive performance of the risk score in DLBCL patients in Low, Intermediate, and High IPI-scores (Figure 4A–D). Compared to the Low-IPI score group, patients in the High-IPI score group tended to have a higher percentage of high-risk scores. In all three groups stratified by IPI-scores, DLBCL patient with high-risk score exhibited significantly worse outcomes (Figure 4B–D). Retrospective analysis has divided DLBCL patients in GSE117556 and GSE181063 into four molecular subtypes: GCB, ABC, and unclassified (UNC), which are common COO subtypes, and molecular high-grade (MHG), which has a worse prognosis (Figure 4E). Notably, the risk score was also independent prognostic predictor in different molecular subtypes (Figure 4F–I). DLBCL harboring rearrangements in *Myc* and *Bcl2* or *Bcl6*, referred to as double-hit lymphoma, exhibits a distinctly worse prognosis (Figure 4J). DLBCL patients with high-risk score had a worse prognosis in both the no *Myc* rearrangement group and the double-hit group. (Figure 4K,L). These results indicated that the risk score remained a useful prognostic factor even after stratification by various clinical characteristics.

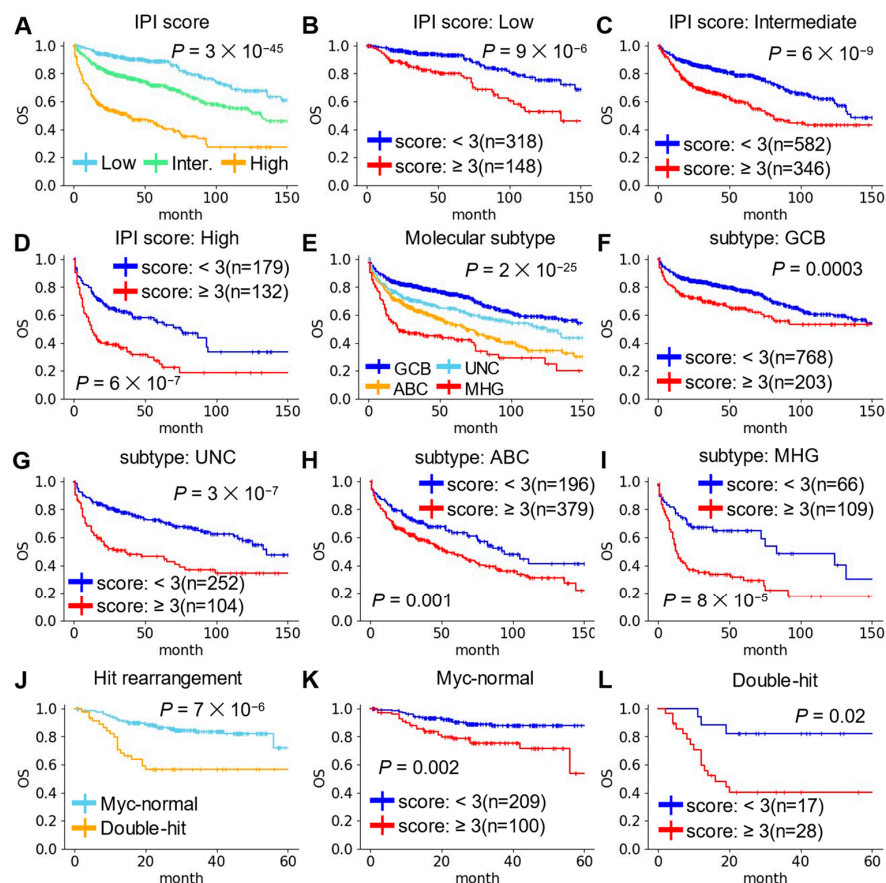


Figure 4. Survival analysis of the risk score in different clinical characteristics. (A) KM survival curves showing OS of patient with low-, intermediate-, and high-IPI-score. (B–D) KM survival curves showing OS of low- and high-risk patients in the three different IPI-score groups. (E) KM survival curves showing OS of patients with GCB, UNC, ABC, and MHG molecular subtypes. (F–I) KM survival curves showing OS of low- and high-risk patients in the four different molecular subtypes. (J) KM survival curves showing OS of patients with Myc-normal and double-hit DLBCL. (K,L) KM survival curves showing OS of low- and high-risk patients in the Myc-normal or double-hit DLBCL.

2.4. Comparison Predictive Performance of the Four Gene-Pair Signatures Against Other Signatures

To verify robustness of the four gene-pair signatures, we compared predictive performance against nine existing gene signatures [13–21]. The gene symbols and regression coefficients for the nine signatures are listed in Table S3. Compared to the distribution of risk scores for the four gene-pair signatures, the distribution of risk scores for the other signatures was highly variable across DLBCL datasets (Figure S4). To evaluate the predictive accuracy of the signatures, we calculated their area under the curve (AUC) scores in six external validation datasets and one internal test datasets using time-dependent ROC curve analyses (Figure 5, Table S4). The four gene-pair signatures showed the best AUC score in GSE98588 dataset (Figure 5, Table S4). When averaging AUC scores at 1 to 3 years, the four gene-pair signatures showed consistently accurate predictive performances (Figure 5, Table S4). Some of the other signatures showed relatively high prognostic accuracy in certain datasets, especially those in which they were trained, but many of them showed reduced accuracy in other datasets. For instance, the Xiong’s signature showed a high predictive accuracy in GSE11318, where it was trained, but it showed worse accuracy in other datasets (Figure 5, Table S4). When the accuracy of the AUC score was evaluated in rank order format, the four gene-pair signature and Ren’s signature tied for first place (Table 2). These results showed that the risk-scoring models based on the four gene-pair signature provides a robust prognostic predictor for DLBCL patients.

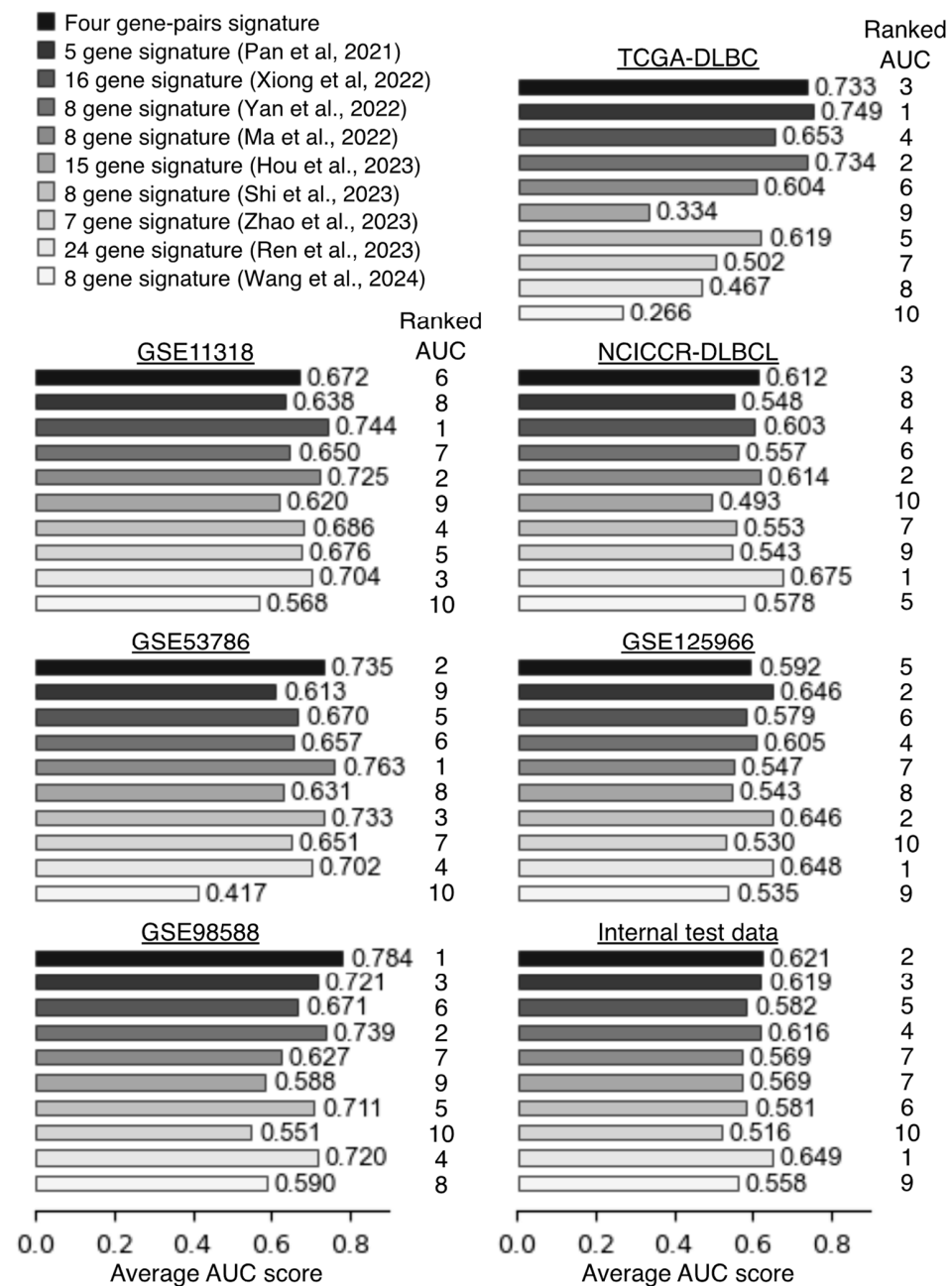


Figure 5. Comparison of AUC score between four gene-pair signature and nine other gene signatures. The numbers on the right side of the bar graph indicate the average AUC score at the 1-, 2-, and 3-year time points. The “Ranked AUC” is a relative ranking in predictive accuracy among 10 gene signatures [13–21].

Table 2. Comparison of ranked AUC scores between four gene-pair signature and other gene signatures.

Study	Ranked AUC Scores							Mean
	GSE11318	GSE53786	GSE98588	TCGA-DLBC	NCICCR-DLBCL	GSE125966	Internal Test Data	
Four gene-pairs	6	2	1	3	3	5	2	3.14
Pan et al., 2021 [13]	8	9	3	1	8	2	3	4.86
Xiong et al., 2022 [14]	1	5	6	4	4	6	5	4.43
Yan et al., 2022 [15]	7	6	2	2	6	4	4	4.43
Ma et al. 2022 [16]	2	1	7	6	2	7	7	4.57

Table 2. Cont.

Study	Ranked AUC Scores						Internal Test Data	Mean
	GSE11318	GSE53786	GSE98588	TCGA-DLBC	NCICCR-DLBC	GSE125966		
Hou et al., 2023 [17]	9	8	9	9	10	8	7	8.57
Shi et al., 2023 [18]	4	3	5	5	7	2	6	4.57
Zhao et al., 2023 [19]	5	7	10	7	9	10	10	8.29
Ren et al., 2023 [20]	3	4	4	8	1	1	1	3.14
Wang et al., 2024 [21]	10	10	8	10	5	9	9	8.71

2.5. Differential Signaling Pathways in Low- and High-Risk DLBCL

To explore differences in signaling pathways between low- and high-risk DLBCL, GSVA was performed. We assessed differences in the activity of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in the low- and high-risk groups. The GSVA scores of 85 pathways was significantly different between low- and high-risk DLBCL samples in training datasets (Figure 6, Table S5). Pathways involved in cell-cell interactions with stromal and immune cells and drug metabolism pathways were relatively activated in low-risk DLBCL. By contrast, cell cycle, energy synthesis, and DNA repair pathways were activated in high-risk DLBCL. The GSVA results suggested that differences in these signaling pathways might alter resistance to chemotherapy and had a critical impact on the prognosis of DLBCL patients.

2.6. High-Risk DLBCL Cell Lines Are Resistant to Chemotherapy

To examine the drug sensitivity of DLBCL associated with risk score, IC50 values of anti-cancer drugs and gene-expression data for 17 DLBCL cell lines were obtained from the Cancer Cell Line Encyclopedia and subjected to risk scoring analysis (Figure 7A). Comparative evaluation of drug sensitivity of low- and high-risk DLBCL cell lines revealed that high-risk DLBCL cell lines were significantly more resistant to 12 drugs and tended to be resistant to 14 drugs (Table 3). These drugs included doxorubicin, cisplatin, etoposide, and gemcitabine, all of which play a central role in chemotherapy of DLBCL (Figure 7B–D, Table 3). By contrast, there was no clear difference in susceptibility between the low- and high-risk groups for cytarabine and methotrexate (Figure 7E,F). Unfortunately, the comparative analysis did not find any drugs that were specifically effective against high-risk DLBCL cell lines (Table 3). DLBCL cell lines were also divided into low- and high-risk groups based on the Ren's signature, which has high predictive accuracy, and then comparative evaluation of drug sensitivity was performed. However, when the cell lines were grouped by the Ren's signature, the low-risk DLBCL cell lines were more resistant to various drugs, including cisplatin, gemcitabine, and cytarabine (Table S6). These results suggested that the risk score based on the four gene-pair signature is strongly correlated with the chemotherapy resistance of DLBCL.

Table 3. Comparison of IC50 values between low- and high-risk DLBCL cell lines.

Drug	Description	LN_IC50_Mean		p Value
		Risk Score: <3	Risk Score: ≥3	
OSU-03012	PDK1 (PDPK1)	1.053	2.057	0.006
GSK650394	SGK2, SGK3	2.713	4.277	0.018
Cisplatin	DNA crosslinker	2.343	4.116	0.024
Doxorubicin	Anthracycline	−2.945	−1.070	0.027
Mitomycin-C	DNA crosslinker	−1.648	0.197	0.031
Bosutinib	SRC, ABL, TEC	0.887	2.463	0.035
SN-38	TOP1	−5.205	−2.933	0.035
XMD15-27	CAMK2	3.414	4.391	0.040
BX795	TBK1, PDK1 (PDPK1), IKK, AURKB, AURKC	0.973	2.161	0.042

Table 3. Cont.

Drug	Description	LN_IC50_Mean		p Value
		Risk Score: <3	Risk Score: ≥3	
Lestaurtinib	FLT3, JAK2, NTRK1, NTRK2, NTRK3	−1.929	−0.734	0.048
Ponatinib	ABL, PDGFRA, VEGFR2, FGFR1, SRC, TIE2, ...	−1.324	0.860	0.045
ZM447439	AURKA, AURKB	1.345	2.756	0.048
Vorinostat	HDAC inhibitor Class I, IIa, IIb, IV	−0.091	0.715	0.057
BAY-61-3606	SYK	0.728	2.431	0.057
S-Trityl-L-cysteine	KIF11	0.043	1.467	0.066
HG6-64-1	BRAF	−0.219	1.562	0.069
Pelitinib	EGFR	−0.033	1.315	0.078
Gemcitabine	Pyrimidine antimetabolite	−3.887	−0.723	0.080
Tozasertib	AURKA, AURKB, AURKC, others	0.007	2.448	0.083
Tipifarnib	Farnesyl-transferase (FNTA)	0.711	2.413	0.083
ObatoclaxMesylate	BCL2, BCL-XL, BCL-W, MCL1	−2.383	−1.090	0.085
Dasatinib	ABL, SRC, Ephrins, PDGFR, KIT	−1.112	1.782	0.091
Etoposide	TOP2	0.163	2.373	0.094
Piperlongumine	Induces reactive oxygen species	1.503	2.614	0.097
CGP-60474	CDK1,CDK2,CDK5,CDK7,CDK9, PKC	−3.244	−2.015	0.097
Midostaurin	PKC, PPK, FLT1, c-FGR, others	−0.626	0.194	0.099

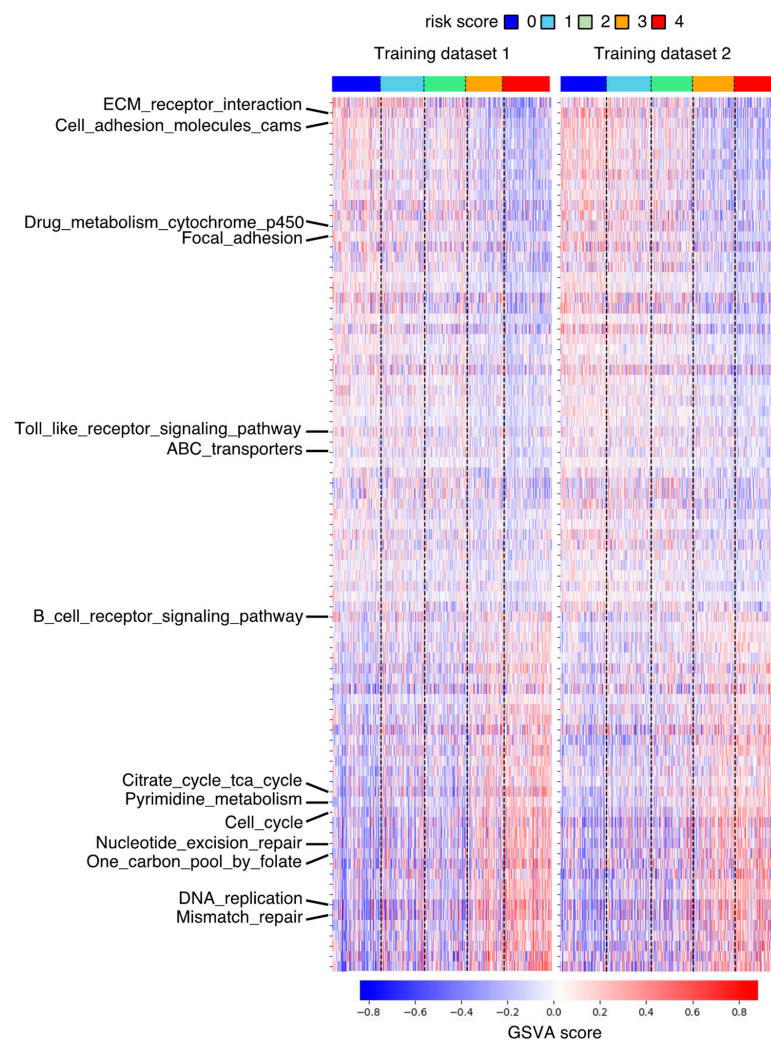


Figure 6. Differential signaling pathways in low- and high-risk DLBCL. DLBCL patients in two training datasets were stratified by risk score. GSEA scores of 85 KEGG pathways in each DLBCL patients in two training datasets were illustrated by heat map.

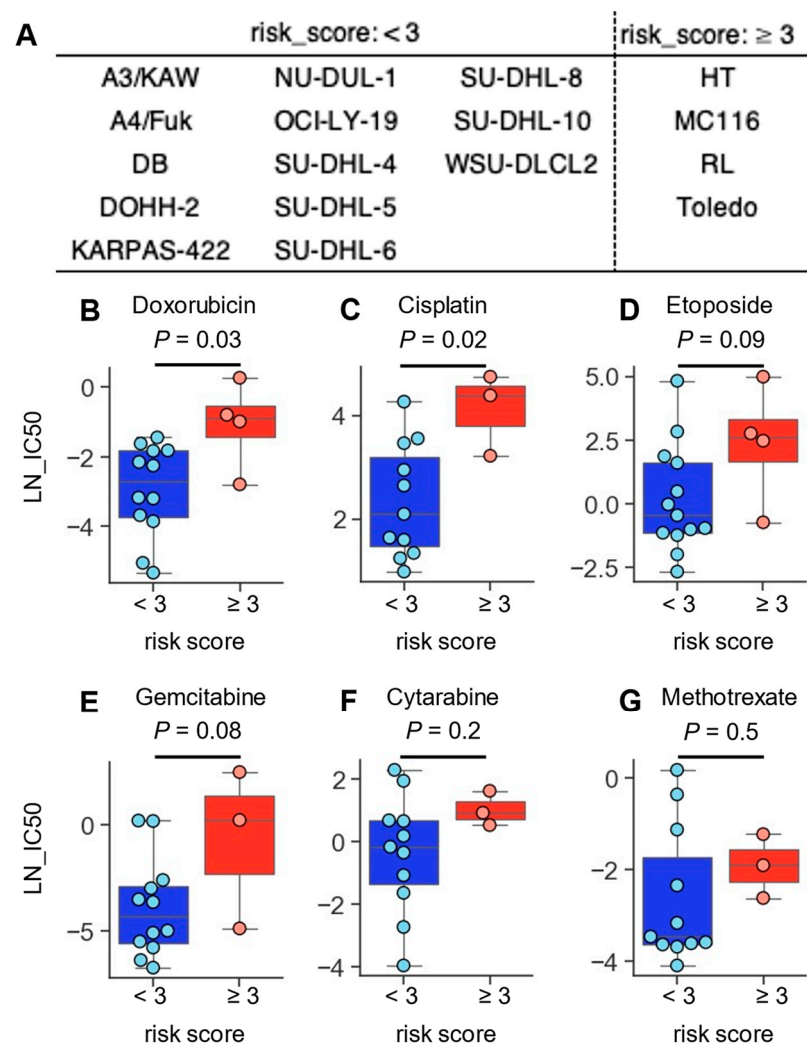


Figure 7. Differential drug sensitivity in low- and high-risk DLBCL cell lines. (A) 17 DLBCL cell lines were stratified into low- and high-risk groups by risk score (B–G) Boxplots showing log-normalized (LN) IC50 values of anti-cancer drugs in low- (blue) and high-risk (red) DLBCL cell lines. Each circle represents LN_IC50 values in an individual cell line. p values obtained using Welch t test.

3. Discussion

Since the late 1990s, advances in genetic analysis technologies have led to an increase in research on the genetic characteristics of various carcinomas, including DLBCL [6–8]. More recent studies have used a variety of methods to take advantage of cancer transcriptome data available in public databases [9,10]. In this study, we comprehensively analyzed DLBCL transcriptome data to extract new gene signatures and classify the prognostic characteristics of DLBCL.

Among the DLBCL prognostic genes we identified, 204 were found to be associated with a favorable prognosis. By contrast, 204 genes associated with an unfavorable prognosis were also identified, including many genes involved in lymphoma growth and survival, such as Myc and Bcl2 [7,8]. From the vast combination of those prognostic genes, a four gene-pairs model consisting of seven signature genes was developed. To investigate prognostic values and biological roles of seven signature genes in DLBCL, we searched signatureDB using those signature genes as search terms. NEK6 gene has been reported to be a marker gene for GCB-type DLBCL [22]. RARRES2 gene was reported by Lenz et al. as a Stromal-1 signature gene associated with a favorable prognosis [23]. No officially published reports on the CHRNA1 gene were found in signatureDB. NEK6, URI1, NIPA2, ZNF22, and WDR12 were all included in the EcoTyper signature genes [10]. NEK6 was

a signature gene of B-cell state 1 (S1) associated with a favorable prognosis, while URI1, ZNF22, NIPA2, and WDR12 were all signature genes of B-cell state 4/5 (S4/S5) associated with an unfavorable prognosis [10]. Thus, it is very likely that each of these signature genes is an important marker associated with prognosis of DLBCL, but further studies are needed to reveal the biological roles of these genes in DLBCL development and survival.

We developed a risk-scoring model based on the four gene-pairs with reference to a predictive model for acute myeloid leukemia published by Kong et al. in 2022 [24]. In the predictive model of Kong et al., the risk score was calculated after multiplying the binary scores of the gene-pairs by the regression coefficient, whereas in our risk-scoring model, the risk score was calculated as it is in the binary scores. Since the final risk score has five levels from 0 to 4, it is easier to determine the low-risk and high-risk threshold than in other risk-scoring models. The risk-scoring model based on the four gene-pairs showed stable accuracy as a prognostic predictor for DLBCL patients in training and validation datasets. The risk-scoring model was developed based on gene-expression data from DLBCL patients receiving standard R-CHOP/CHOP therapy, but this model also showed high predictive accuracy in a group of patients not receiving standard chemotherapy. Furthermore, it was shown to be an independent prognostic predictor even when combined with clinicopathologic predictors such as IPI-score, COO subtype, and myc-rearrangement.

To confirm the robustness of our risk-scoring model, we compared AUC scores in external validation datasets with other existing predictive models. The results showed that our risk-scoring model outperformed almost all other predictive models in terms of predictive stability. Among the other nine models, Ren's signature model was found to be an excellent predictive model, showing equal or more stable predictive accuracy than our risk-scoring model in all six datasets except TCGA-DLBC. However, the fact that Ren's signature consists of 24 signature genes is a point that makes it inferior to other predictive models when considering the operational cost of measuring gene expression.

The GSVA results showed that cell-cell interactions with stromal and immune cells were activated in low-risk DLBCL. Kotlov and colleagues described four DLBCL microenvironments based on analyses of transcriptome data from more than 4000 DLBCL cases [9]. Of the four microenvironments, "Depleted-DLBCL", characterized almost exclusively by B-lymphoma cells and very few other normal stromal and immune cells, had the most unfavorable prognosis. The results of the analysis by Kotlov et al. [9] strongly suggest that interactions between DLBCL and normal stromal and immune cells have a major effect on DLBCL survival and proliferation. The GSVA results also showed that several drug metabolic pathways were activated in low-risk DLBCL. Cyclophosphamide in R-CHOP therapy is a prodrug, which is metabolized in the liver to the active form [25,26]. The activation of drug metabolic pathways in low-risk DLBCL suggested that the metabolism of cyclophosphamide within the tumor might affect sensitivity to R-CHOP therapy. In contrast, the GSVA results showed that DNA repair pathways were activated in high-risk DLBCL. The activation of these pathways suggested that high-risk DLBCL are highly resistant to DNA damage caused by cyclophosphamide and doxorubicin.

To examine the association between risk score and DLBCL resistance to chemotherapy, drug resistance in low- and high-risk DLBCL cell lines was compared. In this study, the resistance to approximately 260 drugs in the CCLE dataset were compared between the two groups, and high-risk DLBCL cell lines were resistant to 26 drugs, equivalent to one-tenth of the total drugs. Especially, since doxorubicin and cisplatin are important drugs in first-line and second-line therapy for DLBCL, resistance to them might be an important factor in the unfavorable prognosis of DLBCL patients with high-risk scores. In addition, the result that high-risk DLBCL cell lines were resistant to drugs that cause DNA damage was consistent with the GSVA results. By contrast, the risk score based on Ren's signature did not correlate with drug resistance of DLBCL cell lines. The 24 signature genes in Ren's signature include marker genes for T-cells and dendritic cells, such as CD3E, CD28, and LY75. Thus, Ren's signature might be a risk-scoring model that reflects the state of

the DLBCL microenvironment, suggesting that the model does not work properly in cell culture systems where lymphoma cells are present alone.

One of the major clinical implications of the development of biomarkers and predictive models is to provide useful information for patient care. For example, a reliable biomarker that can be used to guide the choice of a treatment modality such as chimeric antigen receptor T cell therapy, which is highly effective but expensive and carries the risk of serious side effects, would be of great benefit to both health care workers and patients [27]. In the treatment of DLBCL, several common predictors, such as IPI score and COO subtype, help provide reliable prognostic information, but they are not often reflected in treatment modalities. On the other hand, recently developed clustering models based on the LymphGen algorithm have been shown to provide a decision index for predicting the efficacy of specific molecular-targeted drugs [7]. Our survival analysis and drug screening test showed that high-risk DLBCL was highly resistant to various chemotherapies, including R-CHOP. However, it was suggested that the sensitivity to methotrexate was almost the same in low-risk and high-risk DLBCL. Therefore, prophylactic high-dose methotrexate therapy for DLBCL patients with high-risk scores may improve prognosis [28]. In the second-line therapy for high-risk DLBCL, the usage of new molecular-targeted drugs and immunotherapies should be considered more aggressively than general treatment regimens, including cisplatin and etoposide.

Our study has important limitations. The first limitation is that the predictive performance of the risk-scoring model was demonstrated only in retrospective studies using publicly available gene-expression datasets and not in prospective studies using an original dataset. In addition, the datasets used in this study was DLBCL patient data from approximately 10 to 20 years ago. With advances in diagnostic techniques and treatment methods, the outcome of DLBCL has gradually improved since 2003, when R-CHOP therapy became the standard of care [29]. Therefore, prospective studies are required to confirm whether the risk-scoring model can adequately predict the prognosis of new patients. The second limitation is that the predictive performance of the risk-scoring model was examined using only microarray and total RNAseq-derived gene-expression data. When envisioning clinical applications of risk-scoring models, the cost of the work required for clinical testing and the financial cost to the patients are important issues to consider. Therefore, it is necessary to consider the predictive performance of alternative methods such as real-time PCR and NanoString technology, which have relatively low operational costs, instead of microarray and RNAseq, which have high operational costs. Especially, NanoString technologies is often used in translational research for clinical applications of classification models originally derived from microarray data, such as COO subtyping [12]. In any case, further clinical studies are needed to overcome the limitations of the research at this stage.

4. Materials and Methods

4.1. Datasets

A total of 19 microarray and RNA-seq datasets [23,30–47] were obtained from the NCBI Gene Expression Omnibus (GEO) [48] and the Genomic Data Commons Data Portal [49]. IC50 values of anti-cancer drugs and RNA-seq data in Cancer Cell Line Encyclopedia (CCLE) datasets were obtained from the cBioPortal for Cancer Genomics [50,51]. Details of the datasets used in this study are summarized in Table 4.

Table 4. The datasets used in this study.

Accession	Data Type	Platform	No. Samples	Ref
GSE68895	Microarray	Hu6800	77	[30]
GSE4475	Microarray	HG-U133A	127	[31]
TCGA-DLBC	RNA-seq	Illumina HiSeq 2000	47	[32]
GSE11318	Microarray	HG-U133_Plus_2	200	[33]
GSE10846	Microarray	HG-U133_Plus_2	414	[23]
GSE23501	Microarray	HG-U133_Plus_2	68	[34]

Table 4. *Cont.*

Accession	Data Type	Platform	No. Samples	Ref
GSE31312	Microarray	HG-U133_Plus_2	470	[35]
GSE34771	Microarray	HG-U133_Plus_2	34	[36]
GSE32918	Microarray	Illumina HumanRef-8 v3.0	172	[37]
GSE34171	Microarray	HG-U133_Plus_2	68	[38]
GSE53786	Microarray	HG-U133_Plus_2	119	[39]
GSE69051	Microarray	Illumina HumanHT-12 V4.0	117	[40]
GSE87371	Microarray	HG-U133_Plus_2	223	[41]
NCICCR-DLBCL	RNA-seq	Illumina HiSeq 2500	234	[42]
GSE98588	Microarray	HG-U133_Plus_2	105	[43]
CTSP-DLBCL1	RNA-seq	Illumina HiSeq 2500	18	[44]
GSE117556	Microarray	Illumina HumanHT-12 V4.0	928	[45]
GSE125966	RNA-seq	Illumina HiSeq 2500	553	[46]
GSE181063	Microarray	Illumina HumanHT-12 V4.0	1149	[47]
CCLE 2019	RNA-seq	Illumina HiSeq 2000	17	[50]

4.2. Identification of Prognostic Genes Using a Univariable Cox Proportional Hazard Model

Multiple independent microarray datasets were integrated into two training datasets using “pyComBat” (<https://epigenelabs.github.io/pyComBat/#using-pycombat>, accessed on 16 December 2023) [52] to eliminate batch effects (Figure S1A,B). After integration, DLBCL patients treated with regimen other than CHOP/R-CHOP were removed from the training datasets and designated as internal test data to evaluate the performance of the predictive models. Next, the Python package “lifelines” (v0.28.0) was used to perform univariate Cox regression analysis of the gene expression profiles and clinical information of the training datasets. Patients were divided into a high group (top 50%) or low group (bottom 50%) based on the median expression level for each gene. Hazard ratios (HRs) were calculated by fitting Cox proportional hazard models for the high and low groups. The threshold *p* value for the log-rank and Wilcoxon tests was set at <0.01, and genes with a HR < 1.0 were defined as favorable prognostic genes (FPGs), whereas genes with HR > 1.0 were defined as unfavorable prognostic genes (UPGs). Prognostic genes that overlapped in both training datasets were finally selected (Table S1).

4.3. Selection of Four Gene-Pairs and Calculation of the Risk Score

Data matrices of 204 FPGs and 204 UPGs were extracted from the training datasets and normalized using QuantileTransformer (output_distribution = ‘uniform’) in Python package “scikit-learn” (v1.5.1). Next, the UPGs were paired with the FPGs into a total of 41,616 gene-pairs. These pairs were binary transformed into scores of 1, if expression of the UPG was greater or equal to that of the FPG, or 0, if less. Univariate Cox regression analysis was performed using the binary scores of the 41,616 gene-pairs and clinical information from training datasets in the Python package “lifelines”. The gene-pairs with HR values > 1.8 were extracted from training datasets (Table S2). Then, four gene-pairs consisting of seven genes that overlapped in both training datasets were selected (Table 1). Finally, a risk score was calculated from the binary scores of the four gene-pairs. Where data for all prognostic genes were not available, the risk score was calculated using only the data of the genes registered in the dataset.

Survival analysis was performed using the Python package “lifelines”. Kaplan-Meier survival curves were generated to compare the clinical outcomes between patients stratified by risk score. Log-rank *p* values < 0.05 were considered to indicate statistical significance.

4.4. Time-Dependent ROC Curves and GSVA Analysis

Time-dependent receiver operating characteristic (ROC) curves and gene set variation analysis (GSVA) were performed by the R libraries “survivalROC” and “GSVA” in an Rstudio environment (R version 4.2.0; R Foundation) [53,54].

4.5. Data Analysis Environment and Statistical Analyses

Statistical analyses and visualization of data in the present study were primarily conducted using Python (version 3.9; Python Software Foundation) in a JupyterLab environment [55,56]. Two-tailed paired t-test was used for the comparison of two paired groups. Welch's t-test was used for the comparison of two unpaired groups. *p* values < 0.05 were considered to be statistically significant.

5. Conclusions

Taken together, we developed a risk-scoring model based on four gene-pair signatures using gene-expression profiles of DLBCL from public databases. Prognosis prediction by the risk-scoring model was highly reproducible, and its predictive accuracy was superior to many other existing models. The risk-scoring model also reflects the resistance of DLBCL to chemotherapy, which may also be useful in making treatment decisions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms252312807/s1>.

Author Contributions: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing—original draft preparation, visualization, A.T.; writing—review and editing, J.N. and H.S.; supervision, project administration, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in the TCGA Research Network: <https://www.cancer.gov/tcga> (accessed on 22 January 2024), the GEO database: <https://www.ncbi.nlm.nih.gov/geo/> (accessed on 16 December 2023), and the cBioPortal for Cancer Genomics: <https://www.cbioportal.org/> (accessed on 11 November 2023).

Conflicts of Interest: A.T. belongs to the Laboratory of Highly-Advanced Veterinary Medical Technology, an endowment course, supported by Malignant Tumor Treatment Technologies, Inc. and receives a salary. J.N. is an employee of Malignant Tumor Treatment Technologies, Inc. and receives a salary. No disclosures were reported by H.S.

References

1. Patriarca, A.; Gaidano, G. Investigational drugs for the treatment of diffuse large B-cell lymphoma. *Expert Opin. Investig. Drugs* **2021**, *30*, 25–38. [CrossRef] [PubMed]
2. Coiffier, B.; Sarkozy, C. Diffuse large B-cell lymphoma: R-CHOP failure-what to do? *Hematol. Am. Soc. Hematol. Educ. Progr.* **2016**, *2016*, 366–378. [CrossRef] [PubMed]
3. Maurer, M.J.; Ghesquière, H.; Jais, J.P.; Witzig, T.E.; Haioun, C.; Thompson, C.A.; Delarue, R.; Micallef, I.N.; Peyrade, F.; Macon, W.R.; et al. Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large B-cell lymphoma treated with immunochemotherapy. *J. Clin. Oncol.* **2014**, *32*, 1066–1073. [CrossRef] [PubMed]
4. International Non-Hodgkin's Lymphoma Prognostic Factors Project A predictive model for aggressive non-Hodgkin's lymphoma. *N. Engl. J. Med.* **1993**, *329*, 987–994. [CrossRef] [PubMed]
5. Jelacic, J.; Juul-Jensen, K.; Bukumiric, Z.; Roost Clausen, M.; Ludvigsen Al-Mashhadi, A.; Pedersen, R.S.; Poulsen, C.B.; Brown, P.; El-Galaly, T.C.; Stauffer Larsen, T. Prognostic indices in diffuse large B-cell lymphoma: A population-based comparison and validation study of multiple models. *Blood Cancer J.* **2023**, *13*, 1–12. [CrossRef] [PubMed]
6. Alizadeh, A.A.; Elsen, M.B.; Davis, R.E.; Ma, C.L.; Lossos, I.S.; Rosenwald, A.; Boldrick, J.C.; Sabet, H.; Tran, T.; Yu, X.; et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **2000**, *403*, 503–511. [CrossRef] [PubMed]
7. 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.e14. [CrossRef] [PubMed]
8. Sánchez-Beato, M.; Méndez, M.; Guirado, M.; Pedrosa, L.; Sequero, S.; Yanguas-Casás, N.; de la Cruz-Merino, L.; Gálvez, L.; Llanos, M.; García, J.F.; et al. A genetic profiling guideline to support diagnosis and clinical management of lymphomas. *Clin. Transl. Oncol.* **2024**, *26*, 1043–1062. [CrossRef] [PubMed]

9. Kotlov, N.; Bagaev, A.; Revuelta, M.V.; Phillip, J.M.; Cacciapuoti, M.T.; Antysheva, Z.; Svelkolkin, V.; Tikhonova, E.; Mihecheva, N.; Kuzkina, N.; et al. Clinical and biological subtypes of b-cell lymphoma revealed by microenvironmental signatures. *Cancer Discov.* **2021**, *11*, 1468–1489. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Steen, C.B.; Luca, B.A.; Esfahani, M.S.; Azizi, A.; Sworder, B.J.; Nabet, B.Y.; Kurtz, D.M.; Liu, C.L.; Khameneh, F.; Advani, R.H.; et al. The landscape of tumor cell states and ecosystems in diffuse large B cell lymphoma. *Cancer Cell* **2021**, *39*, 1422–1437.e10. [\[CrossRef\]](#)
11. Shaffer, A.L.; Wright, G.; Yang, L.; Powell, J.; Ngo, V.; Lamy, L.; Lam, L.T.; Davis, R.E.; Staudt, L.M. A library of gene expression signatures to illuminate normal and pathological lymphoid biology. *Immunol. Rev.* **2006**, *210*, 67–85. [\[CrossRef\]](#) [\[PubMed\]](#)
12. De Groot, F.A.; De Groen, R.A.L.; van den Berg, A.; Jansen, P.M.; Lam, K.H.; Mutsaers, P.G.N.J.; Van Noesel, C.J.M.; Chamuleau, M.E.D.; Stevens, W.B.C.; Plaça, J.R.; et al. Biological and Clinical Implications of Gene-Expression Profiling in Diffuse Large B-Cell Lymphoma: A Proposal for a Targeted BLYM-777 Consortium Panel as Part of a Multilayered Analytical Approach. *Cancers* **2022**, *14*, 1857. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Pan, M.; Yang, P.; Wang, F.; Luo, X.; Li, B.; Ding, Y.; Lu, H.; Dong, Y.; Zhang, W.; Xiu, B.; et al. Whole Transcriptome Data Analysis Reveals Prognostic Signature Genes for Overall Survival Prediction in Diffuse Large B Cell Lymphoma. *Front. Genet.* **2021**, *12*, 1–11. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Xiong, D.; Li, M.; Zeng, C. Construction and validation of a risk scoring model for diffuse large B-cell lymphoma based on ferroptosis-related genes and its association with immune infiltration. *Transl. Oncol.* **2022**, *16*, 101314. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Yan, J.; Yuan, W.; Zhang, J.; Li, L.; Zhang, L.; Zhang, X.; Zhang, M. Identification and Validation of a Prognostic Prediction Model in Diffuse Large B-Cell Lymphoma. *Front. Endocrinol.* **2022**, *13*, 1–12. [\[CrossRef\]](#)
16. Ma, J.; Wang, W.; Ma, J.; Xu, Z. A Novel and Validated 8-Pyoptosis-Related Genes Based Risk Prediction Model for Diffuse Large B Cell Lymphoma. *Biomolecules* **2022**, *12*, 1835. [\[CrossRef\]](#)
17. Hou, J.; Guo, P.; Lu, Y.; Jin, X.; Liang, K.; Zhao, N.; Xue, S.; Zhou, C.; Wang, G.; Zhu, X.; et al. A prognostic 15-gene model based on differentially expressed genes among metabolic subtypes in diffuse large B-cell lymphoma. *Pathol. Oncol. Res.* **2023**, *29*, 1–13. [\[CrossRef\]](#)
18. Shi, H.; Pan, Y.; Xiang, G.; Wang, M.; Huang, Y.; He, L.; Wang, J.; Fang, Q.; Li, L.; Liu, Z. A novel NET-related gene signature for predicting DLBCL prognosis. *J. Transl. Med.* **2023**, *21*, 1–18. [\[CrossRef\]](#)
19. Zhao, Z.; Shen, X.; Zhao, S.; Wang, J.; Tian, Y.; Wang, X.; Tang, B. A novel telomere-related genes model for predicting prognosis and treatment responsiveness in diffuse large B-cell lymphoma. *Aging* **2023**, *15*, 12927–12951. [\[CrossRef\]](#)
20. Ren, W.; Wan, H.; Own, S.A.; Berglund, M.; Wang, X.; Yang, M.; Li, X.; Liu, D.; Ye, X.; Sonnev, K.; et al. Genetic and transcriptomic analyses of diffuse large B-cell lymphoma patients with poor outcomes within two years of diagnosis. *Leukemia* **2024**, *38*, 610–620. [\[CrossRef\]](#)
21. Wang, Y.; Tsukamoto, Y.; Hori, M.; Iha, H. Disulfidptosis: A Novel Prognostic Criterion and Potential Treatment Strategy for Diffuse Large B-Cell Lymphoma (DLBCL). *Int. J. Mol. Sci.* **2024**, *25*, 7156. [\[CrossRef\]](#)
22. Wright, G.; Tan, B.; Rosenwald, A.; Hurt, E.H.; Wiestner, A.; Staudt, L.M. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9991–9996. [\[CrossRef\]](#)
23. 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. [\[CrossRef\]](#)
24. Kong, W.; He, L.; Zhu, J.; Brück, O.; Porkka, K.; Heckman, C.A.; Zhu, S.; Aittokallio, T. An immunity and pyroptosis gene-pair signature predicts overall survival in acute myeloid leukemia. *Leukemia* **2022**, *36*, 2384–2395. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Chang, T.K.H.; Yu, L.; Maurel, P.; Waxman, D.J. Enhanced cyclophosphamide and ifosfamide activation in primary human hepatocyte cultures: Response to cytochrome P-450 inducers and autoinduction by oxazaphosphorines. *Cancer Res.* **1997**, *57*, 1946–1954.
26. Hedrich, W.D.; Xiao, J.; Heyward, S.; Zhang, Y.; Zhang, J.; Baer, M.R.; Hassan, H.E.; Wang, H. Activation of the constitutive androstane receptor increases the therapeutic index of CHOP in lymphoma treatment. *Mol. Cancer Ther.* **2016**, *15*, 392–401. [\[CrossRef\]](#)
27. Levstek, L.; Janžič, L.; Ihan, A.; Kopitar, A.N. Biomarkers for prediction of CAR T therapy outcomes: Current and future perspectives. *Front. Immunol.* **2024**, *15*, 1–19. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Tolley, E.R.; Lewinter, C.; Pedersen, L.M.; Nielsen, T.H. Efficacy of intravenous high-dose methotrexate in preventing relapse to the central nervous system in R-CHOP(-like)-treated, high-risk, diffuse large B-cell lymphoma patients and its effect on mortality: A systematic review and meta-analysis. *Haematologica* **2024**, *109*, 3327–3337. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Durmaz, M.; Visser, O.; Posthuma, E.F.M.; Brouwer, R.E.; Issa, D.E.; de Jong, D.; Lam, K.H.; Blijlevens, N.M.A.; Zijlstra, J.M.; Chamuleau, M.E.D.; et al. Time trends in primary therapy and relative survival of diffuse large B-cell lymphoma by stage: A nationwide, population-based study in the Netherlands, 1989–2018. *Blood Cancer J.* **2022**, *12*, 38. [\[CrossRef\]](#)
30. Shipp, M.A.; Ross, K.N.; Tamayo, P.; Weng, A.P.; Kutok, J.L.; Aguiar, R.C.T.; Gaasenbeek, M.; Angelo, M.; Reich, M.; Pinkus, G.S.; et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.* **2002**, *8*, 68–74. [\[CrossRef\]](#)
31. Hummel, M.; Bentink, S.; Berger, H.; Klapper, W.; Wessendorf, S.; Barth, T.F.E.; Bernd, H.; Cogliatti, S.B.; Dierlamm, J.; Feller, A.C.; et al. A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. *N. Engl. J. Med.* **2006**, *354*, 2419–2430. [\[CrossRef\]](#) [\[PubMed\]](#)

32. Collins, F.S.; Barker, A.D. Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci. Am.* **2007**, *296*, 50–57. [\[CrossRef\]](#)
33. Lenz, G.; Wright, G.W.; Emre, N.C.T.; Kohlhammer, H.; Dave, S.S.; Davis, R.E.; Carty, S.; Lam, L.T.; Shaffer, A.L.; Xiao, W.; et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13520–13525. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Shaknovich, R.; Geng, H.; Johnson, N.A.; Tsikitas, L.; Cerchiatti, L.; Grealley, J.M.; Gascoyne, R.D.; Elemento, O.; Melnick, A. DNA methylation signatures define molecular subtypes of diffuse large B-cell lymphoma. *Blood* **2010**, *116*, 81–89. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Visco, C.; Li, Y.; Xu-Monette, Z.Y.; Miranda, R.N.; Green, T.M.; Li, Y.; Tzankov, A.; Wen, W.; Liu, W.; Kahl, B.S.; et al. Comprehensive gene expression profiling and immunohistochemical studies support application of immunophenotypic algorithm for molecular subtype classification in diffuse large B-cell lymphoma: A report from the International DLBCL Rituximab-CHOP Consortium. *Leukemia* **2012**, *26*, 2103–2113. [\[CrossRef\]](#)
36. Kawaguchi, A.; Iwade, Y.; Komohara, Y.; Sano, M.; Kajiwara, K.; Yajima, N.; Tsuchiya, N.; Homma, J.; Aoki, H.; Kobayashi, T.; et al. Gene expression signature-based prognostic risk score in patients with primary central nervous system lymphoma. *Clin. Cancer Res.* **2012**, *18*, 5672–5681. [\[CrossRef\]](#)
37. Barrans, S.L.; Crouch, S.; Care, M.A.; Worrillow, L.; Smith, A.; Patmore, R.; Westhead, D.R.; Tooze, R.; Roman, E.; Jack, A.S. Whole genome expression profiling based on paraffin embedded tissue can be used to classify diffuse large B-cell lymphoma and predict clinical outcome. *Br. J. Haematol.* **2012**, *159*, 441–453. [\[CrossRef\]](#)
38. Monti, S.; Chapuy, B.; Takeyama, K.; Rodig, S.J.; Hao, Y.; Yeda, K.T.; Inguilizian, H.; Mermel, C.; Currie, T.; Dogan, A.; et al. Integrative Analysis Reveals an Outcome-Associated and Targetable Pattern of p53 and Cell Cycle Deregulation in Diffuse Large B Cell Lymphoma. *Cancer Cell* **2012**, *22*, 359–372. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Scott, D.W.; Wright, G.W.; Williams, P.M.; Lih, C.J.; Walsh, W.; Jaffe, E.S.; Rosenwald, A.; Campo, E.; Chan, W.C.; Connors, J.M.; et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood* **2014**, *123*, 1214–1217. [\[CrossRef\]](#)
40. Sha, C.; Barrans, S.; Care, M.A.; Cunningham, D.; Tooze, R.M.; Jack, A.; Westhead, D.R. Transferring genomics to the clinic: Distinguishing Burkitt and diffuse large B cell lymphomas. *Genome Med.* **2015**, *7*, 64. [\[CrossRef\]](#)
41. Dubois, S.; Viailly, P.J.; Bohers, E.; Bertrand, P.; Ruminy, P.; Marchand, V.; Maingonnat, C.; Mareschal, S.; Picquenot, J.M.; Penther, D.; et al. Biological and clinical relevance of associated genomic alterations in MYD88 L265P and non-L265P-mutated diffuse large B-cell lymphoma: Analysis of 361 cases. *Clin. Cancer Res.* **2017**, *23*, 2232–2244. [\[CrossRef\]](#)
42. Schmitz, R.; Wright, G.W.; Huang, D.W.; Johnson, C.A.; Phelan, J.D.; Wang, J.Q.; Roulland, S.; Kasbekar, M.; Young, R.M.; Shaffer, A.L.; et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *N. Engl. J. Med.* **2018**, *378*, 1396–1407. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Chapuy, B.; Stewart, C.; Dunford, A.J.; Kim, J.; Kamburov, A.; Redd, R.A.; Lawrence, M.S.; Roemer, M.G.M.; Li, A.J.; Ziepert, M.; et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat. Med.* **2018**, *24*, 679–690. [\[CrossRef\]](#)
44. Phelan, J.D.; Young, R.M.; Webster, D.E.; Roulland, S.; Wright, G.W.; Kasbekar, M.; Shaffer, A.L.; Ceribelli, M.; Wang, J.Q.; Schmitz, R.; et al. A multiprotein supercomplex controlling oncogenic signalling in lymphoma. *Nature* **2018**, *560*, 387–391. [\[CrossRef\]](#) [\[PubMed\]](#)
45. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
46. McCord, R.; Bolen, C.R.; Koeppen, H.; Kadel, E.E.; Oestergaard, M.Z.; Nielsen, T.; Sehn, L.H.; Venstrom, J.M. PD-L1 and tumor-associated macrophages in de novo DLBCL. *Blood Adv.* **2019**, *3*, 531–540. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Lacy, S.E.; Barrans, S.L.; Beer, P.A.; Painter, D.; Smith, A.G.; Roman, E.; Cooke, S.L.; Ruiz, C.; Glover, P.; Van Hoppe, S.J.L.; et al. Targeted sequencing in DLBCL, molecular subtypes, and outcomes: A Haematological Malignancy Research Network report. *Blood* **2020**, *135*, 1759–1771. [\[CrossRef\]](#)
48. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2013**, *41*, 991–995. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Jensen, M.A.; Ferretti, V.; Grossman, R.L.; Staudt, L.M. The NCI Genomic Data Commons as an engine for precision medicine. *Blood* **2017**, *130*, 453–459. [\[CrossRef\]](#)
50. Ghandi, M.; Huang, F.W.; Jané-Valbuena, J.; Kryukov, G.V.; Lo, C.C.; McDonald, E.R.; Barretina, J.; Gelfand, E.T.; Bielski, C.M.; Li, H.; et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **2019**, *569*, 503–508. [\[CrossRef\]](#)
51. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal Complementary Data Sources and Analysis Options. *Sci. Signal.* **2013**, *6*, 1–20. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Behdenna, A.; Colange, M.; Haziza, J.; Gema, A.; Appé, G.; Azencott, C.A.; Nordor, A. pyComBat, a Python tool for batch effects correction in high-throughput molecular data using empirical Bayes methods. *BMC Bioinform.* **2023**, *24*, 1–9. [\[CrossRef\]](#) [\[PubMed\]](#)
53. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2013. Available online: <https://www.r-project.org/> (accessed on 4 May 2022).

54. RStudio Team. *RStudio: Integrated Development for R*; RStudio, PBC: Boston, MA, USA, 2020; Available online: <http://www.rstudio.com/> (accessed on 6 June 2023).
55. VanRossum, G.; Drake, F.L. *The Python Language Reference*; Python Software Foundation: Amsterdam, The Netherlands, 2010.
56. Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Willing, C. Jupyter Notebooks—A publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas, Proceedings of the 20th International Conference on Electronic Publishing, Göttingen, Germany, 7–9 June 2016*; IOS Press: Amsterdam, The Netherlands, 2016; pp. 87–90.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.