International Journal of
Molecular Sciences

MDPI

Article

Extracellular Vesicles in Diffuse Large B Cell Lymphoma: Characterization and Diagnostic Potential

Rune Matthiesen [1,*], Paula Gameiro [2], Andreia Henriques [1], Cristian Bodo [3], Maria Carolina Strano Moraes [3], Bruno Costa-Silva [3], José Cabeçadas [2], Maria Gomes da Silva [2], Hans Christian Beck [4] and Ana Sofia Carvalho [1,*]

[1] Computational and Experimental Biology Group, NOVA Medical School-Research, Faculdade de Ciências Médicas, Universidade NOVA de Lisboa, Campo dos Mártires da Pátria, 130, 1169-056 Lisbon, Portugal
[2] Instituto Português de Oncologia, Departament of Hematology, 1099-213 Lisbon, Portugal
[3] Champalimaud Physiology and Cancer Programme, Champalimaud Foundation, 1400-038 Lisbon, Portugal
[4] Centre for Clinical Proteomics, Department of Clinical Biochemistry, Odense University Hospital, DK-5000 Odense, Denmark
* Correspondence: rune.matthiesen@nms.unl.pt (R.M.); ana.carvalho@nms.unl.pt (A.S.C.)

Abstract: Diffuse large B cell lymphoma (DLBCL) is an aggressive B cell lymphoma characterized by a heterogeneous behavior and in need of more accurate biological characterization monitoring and prognostic tools. Extracellular vesicles are secreted by all cell types and are currently established to some extent as representatives of the cell of origin. The present study characterized and evaluated the diagnostic and prognostic potential of plasma extracellular vesicles (EVs) proteome in DLBCL by using state-of-the-art mass spectrometry. The EV proteome is strongly affected by DLBCL status, with multiple proteins uniquely identified in the plasma of DLBCL. A proof-of-concept classifier resulted in highly accurate classification with a sensitivity and specificity of 1 when tested on the holdout test data set. On the other hand, no proteins were identified to correlate with non-germinal center B-cell like (non-GCB) or GCB subtypes to a significant degree after correction for multiple testing. However, functional analysis suggested that antigen binding is regulated when comparing non-GCB and GCB. Survival analysis based on protein quantitative values and clinical parameters identified multiple EV proteins as significantly correlated to survival. In conclusion, the plasma extracellular vesicle proteome identifies DLBCL cancer patients from healthy donors and contains potential EV protein markers for prediction of survival.

Keywords: diffuse large B cell lymphoma; extracellular vesicles; cancer; diagnostic; liquid biopsy; plasma

check for updates

Citation: Matthiesen, R.; Gameiro, P.; Henriques, A.; Bodo, C.; Moraes, M.C.S.; Costa-Silva, B.; Cabeçadas, J.; Gomes da Silva, M.; Beck, H.C.; Carvalho, A.S. Extracellular Vesicles in Diffuse Large B Cell Lymphoma: Characterization and Diagnostic Potential. *Int. J. Mol. Sci.* **2022**, *23*, 13327. https://doi.org/10.3390/ijms232113327

Academic Editor: Aamir Ahmad

Received: 1 October 2022
Accepted: 29 October 2022
Published: 1 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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

B-cell lymphomas develop from abnormal B-lymphocytes and account for 85% of all non-Hodgkin's lymphomas (NHLs). Diffuse large B-cell lymphoma (DLBCL) is the most common form of lymphoma, accounting for up to 40% of all newly diagnosed cases of NHL [1]. As a heterogeneous entity, DLBCL is also characterized by a variety of morphologic variants, diverse molecular abnormalities, and variable clinical behaviors and therapeutic responses [2,3]. It can arise de novo or as a transformation from a low grade B-cell malignancy, such as follicular lymphoma or chronic lymphocytic leukemia. The disease demonstrates remarkable progression-free survival and overall survival in clinically and biologically defined subgroups of patients. Unfortunately, approximately 40% of patients will relapse or develop refractory disease. Simple clinical characteristics, such as age, lactate dehydrogenase levels, the number and sites of metastases involvement, the stage, and performance status can be scored in a manner that stratifies R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone) treated DLBCL patients in

groups with varying survival [4]. Gene expression profiling (GEP) studies of DLBCL revealed two main sub-types: germinal center B-cell like (GCB) and activated B-cell like (ABC). In addition, a rarer primary mediastinal B-cell type was identified. Recently, deeper molecular analysis of large cohorts of DLBCL newly diagnosed biopsies revealed different molecular subtypes that were associated with responses to immunochemotherapy [5,6]. Non-GCB tumors, defined by immunohistochemistry algorithms, include the unclassifiable and the ABC DLBCL subtype; the latter is associated with poor treatment outcomes. The gold-standard method for DLBCL gene expression sub-typing is the Affymetrix method based on frozen tissue (FT) [7–9]. This severely limits the ability to utilize this assay in routine clinical practice, as fresh tissue is not routinely collected during patient diagnosis. The Lymph2Cx assay, a digital 20-gene expression based test (NanoString) for cell of origin (COO) assignment requiring formalin-fixed paraffin-embedded tissue (FFPET), has demonstrated very high rates of accurate assignment when tested against the gold-standard FT method [10]. Despite good segregation between ABC and GCB subtypes, both methods rely on RNA, which is susceptible to degradation by chemical treatments such as those used in FFPET preparation and requires fresh tissue as a source. COO assignment in standard practice applying immunohistochemical methods (BCL6/CD10/MUM1) in commonly available FFPET is less precise but relatively inexpensive. Attempts to develop multiple immunohistochemistry (IHC)-based algorithms such as Hans, Tally, Choi and Visco-Young suffered from lower than desired agreement rates with GEP classifications, and required staining of four to eight sections of limited biopsy material [1,11].

Success of targeted therapies for DLBCL relies on minimally invasive, robust strategies that segregate patients into molecular groups with high reliability. Liquid biopsy is a minimally invasive method that potentially provides information on circulating materials such as cell-free DNA (cfDNA), circulating tumor cells (CTCs), and extracellular vesicles (EVs) to detect molecular alterations that are indicative of cancer progression.

Using PCR technology, circulating tumor cells in the blood have been reported to be associated with recurrence in patients with indolent lymphoma and mantle cell lymphoma. Unlike indolent lymphomas, DLBCLs are associated with a relatively low level of circulating tumour cells. Next-generation sequencing (NGS) applied to cell-free tumour DNA in the serum of patients with DLBCL detected clonal immunoglobulin gene sequences, which correlated with poor survival [12].

Since the discovery of EVs, including exosomes, in human body fluids, compelling research studies have been published exploiting the potential of EVs as a source of stable biomarkers. EVs are cell-derived vesicles composed of lipid bilayers, constituting a heterogeneous vesicle population of different sizes, biogenesis and origins. These structures shed from all cell types, retaining part of the donor cell signature as well as specific EV-related factors [13]. EVs have been described as being involved in cell-cell communication in various physiological processes, such as antitumoral immune responses, and conferring resistance to therapy [14]. Additionally, EVs functionally operate at a systemic level and contain circulating targets for diagnosis, prognosis, and potential treatment targets. Given that EVs have been recognized as a promising source of diagnostic biomarkers, various attempts to establish methods to detect disease-specific circulating EVs have been reported. Recently, the proteome of tumor-derived EVs in cancer and non-cancer patients' plasma has been shown to discriminate cancer from non-cancer and to identify proteins with potential to serve as prognostic biomarkers for melanoma [15]. A pan cancer signature of plasma EVs that extended to several types of cancer, allowing early-stage detection, treatment response, and potentially for diagnosing tumors of unknown primary origin, was proposed [16]. EVs present the advantage that the biomolecules are encapsulated in a lipid bilayer, preventing attack by free nucleases and proteases. In addition, for mass spectrometry-based proteomics, EVs serve as an efficient method to remove abundant proteins. EVs are also readily available from a standard blood sample enabling longitudinal monitoring. In this study, we explored the proteome of plasma EVs from newly diagnosed DLBCL patients and age-matched healthy donors (HD) by state-of-the-art high-throughput

mass spectrometry. The results obtained can be explored in the future for accurately identify minimal residual disease (MRD) as well as predicting patient survival using EV proteomics. MRD monitoring could help identify early relapses or treatment failure.

2. Results

The present study aimed at comparing the proteome of vesicles isolated from the plasma of DLBCL patients and healthy donors as well as between DLBCL patients. Patients' characteristics are summarized in Table 1.

Table 1. Patients demographics.

Characteristics	*N*	% or Mean (Range)
Sex		
Male/Female	16/16	50%
Age (years)	32	67.2 (30–94)
LDH [1]		
≤250	19	180 (100–241)
>250	13	680 (262–2537)
IPI score [2]		
Low risk = 0/1	10	31.30%
low-intermediate risk = 2	6	18.80%
high-intermediate risk = 3	12	37.5%
high risk = 4/5	4	12.50%
Ann Arbor stage		
I	7	21.90%
II	5	15.70%
III	2	6.30%
IV	18	56.30%
DLBCL subtype [4]		
GCB type	21	65.60%
Non-GCB type	10	31.30%
ND [3]	1	3.10%

[1] LDH: Lactate dehydrogenase; [2] IPI: International prognostic indices; [3] ND: Non determined. [4] Subtyping based on Hans algorithm.

2.1. Patient Characteristics

This study included 32 patients diagnosed with DLBCL (Table 1) and fifteen age-matched healthy donors. Twenty-one patients were classified as GCB, ten as non-GCB and one case was unclassified by Hans immunohistochemistry algorithm [17]. Twenty-six patients presented a complete response, while the remaining presented partial responses, progressive disease, or had died due to disease progression.

2.2. Plasma EVs of DLBCL Patients and HDs

The present study's primary aim is to compare the proteome of extracellular vesicles isolated from the plasma of DLBCL patients and HDs. For mass spectrometry analysis, we have isolated extracellular vesicles from 32 newly diagnosed DLBCL patients and 15 aged matched HDs using sequential centrifugation, as previously described [18]. The particles have been analyzed by nanoparticle tracking analysis (Figure 1a,b). EVs plasma were characterized by different size subpopulations enriched in the 100 nm range. However, the comparison of particles size distribution between cancer and HDs showed a more heterogeneous population for lymphoma patients compared to healthy donors (Figure 1c,d).

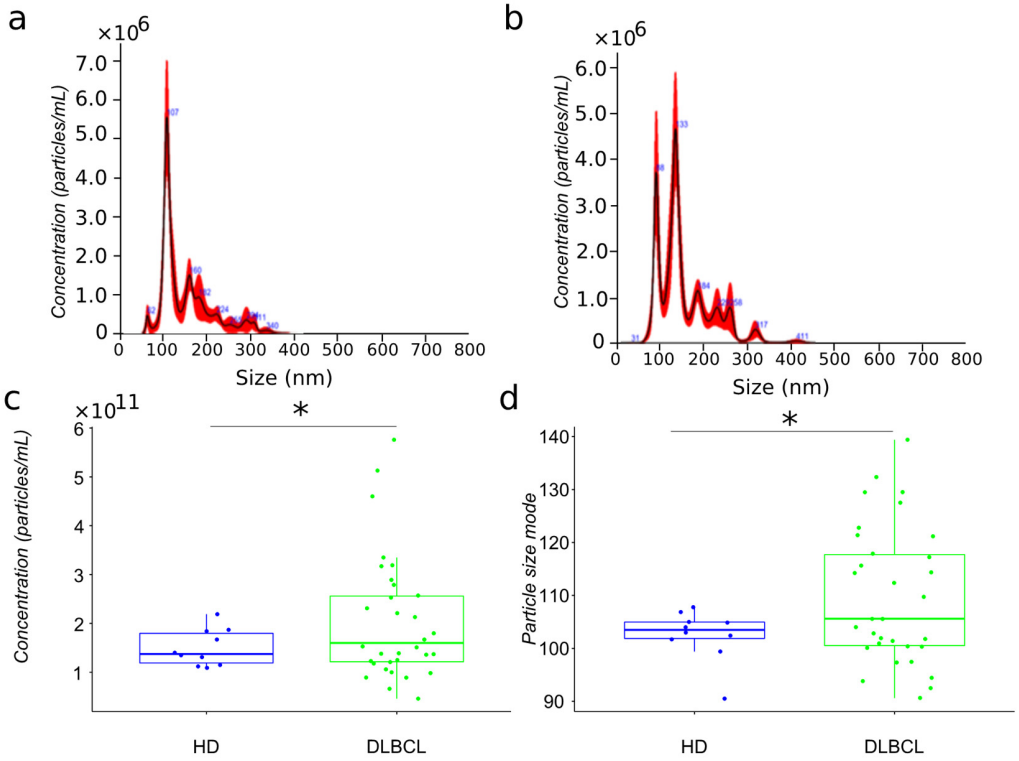


Figure 1. Biophysical characterization by nanoparticle tracking analysis of plasma EV samples from HD and DLBCL groups. Particle size distribution in representative samples of HD and DLBCL groups ((**a**,**b**), respectively). Distribution of estimated particle concentration and particle size mode between DLBCL patients and HDs ((**c**,**d**), respectively). * means significance.

The particle size of plasma EVs samples from patients showed a significant increase compared to those from HDs. Notably, the size range of particles in the EV preparations from DLBCL patients varied between 90–140 nm while with the exception of a single individual, particles from HDs varied between 100–110 nm (Figure 1). Subgrouping according to prognosis (survival and deceased patients) showed no significant differences (data not shown).

2.3. EV Protein Markers

We previously proposed an EV quality control software tool based on mass spectrometry data [14,19]. The tool compares the expression of EV markers frequently reported in the literature and the top ten most abundant exosome markers from a publicly available EV database (ExoCarta) [20]. This tool was further developed in the present study to also include markers for large micro vesicles (RGAP1, MIC60) and furthermore highlight widely accepted exosome markers such as syntenin (SDCBP), ALIX (PDCD6IP) and CD63 (Figure 2). Two alternative isolation methods of EV enrichment were used in the reference data for EVs from cell lines: differential ultra-centrifugation [16] and PEG-based precipitation [21]. All markers are present in the plasma EVs, although at a slightly lower intensity compared to EVs isolated by similar methods from cell line conditioned media and bronchoalveolar lavage (BAL). EV isolation from plasma is notoriously difficult and known to contain platelet EVs [22]. Markers indicating contamination from other subcellular fractions such as the endoplasmic reticulum (CANX), Golgi apparatus (GOLGA2), mitochondria (BCL2), and nucleus (NUP98) were absent or in lower range compared to previous studies. Microvesicles markers were not observed by any of the methods and sample sources included (Figure 2).

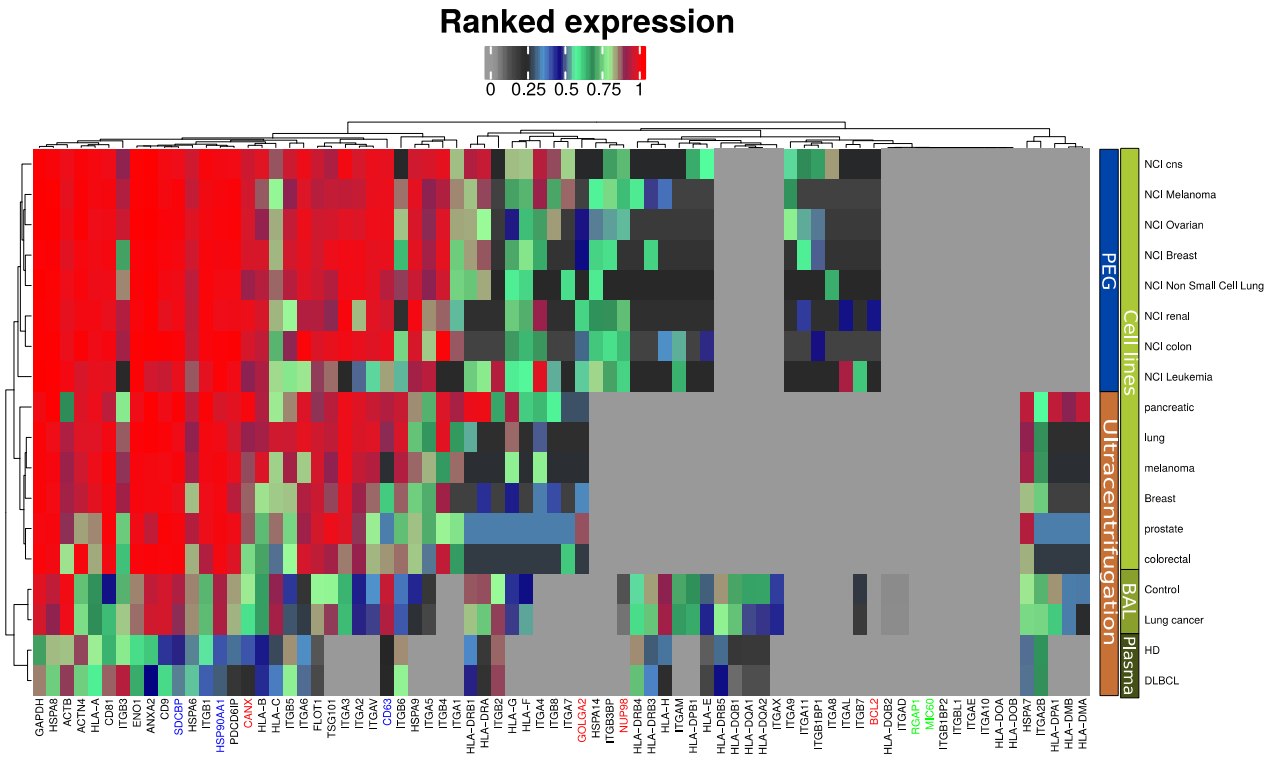


Figure 2. MS-based quantitative comparison between enriched plasma EVs, BAL EVs and cell line isolated EVs of frequently reported exosome protein markers and the ten most abundant exosome markers from a publicly available EV database (ExoCarta). Blue indicates generally accepted exosome markers. Green indicates micro vesicle markers. Red-labeled proteins indicate non-EV proteins.

2.4. Qualitative Comparison of EV Proteome from DLBCL Patients and HDs

Plasma EV proteomes from DLBCL patients and HDs were analyzed by mass spectrometry. Across a total of 98 liquid chromatography tandem mass spectrometry (LC-MS/MS) runs, 2595 protein isoforms were identified. The protein isoforms were mapped to their corresponding genes, resulting in a final number of 1197 proteins. Figure 3a depicts the number of proteins identified factored on cancer status. The DLBCL group displayed a high number of identified proteins, although there were considerably fewer samples from the HD group compared to the DLBCL group. The DLBCL group also displayed more unique proteins compared to the HD group (Figure 3b). The unique proteins for the DLBCL and HD groups are listed in Table S1. According to KEGG pathway analysis (Figure 3c), proteins identified in DLBCL were more enriched in proteasome, infection related functions, antigen presentation, and glycolysis and gluconeogenesis functions than those from the HD group. Complement and coagulation cascades and systemic lupus erythematosus were found significantly enriched in proteins identified from both the DLBCL and the HD group. We used Gene Ontology (GO) (http://www.geneontology.org/, accessed on 15 August 2022) to perform enrichment analysis based on the Gene Ontology–Biological Process (GO-BP), Cellular Component (GO-CC) and Molecular Function (MF) annotations. Figures A1–A3 represent the gene ontology functional enrichment analysis of all identified proteins from the DLBCL and the HD group. This enrichment analysis indicated that the EV plasma proteome of patients and HDs were particularly enriched in GO-CC associated with extracellular exosomes with 548 (DLBCL) and 362 (HD) proteins classified in this category, confirming the quality of the EVs isolation method used (Figure A1). EV plasma proteome included the described protein cargo with tetraspanins membrane organizers, CD9, CD81, CD82, CD63, intracellular trafficking components such as RAB, GTPases, annexins and chaperone molecules from the HSP90 and HSP70 protein families (Supplementary Table S1). As shown in Figure A2, the molecular functions associated with the proteome of

all identified proteins in plasma EVs were threonine-type endopeptidase activity, structural constituent of cytoskeleton, GTPase activity (mainly DLBCL) and antigen binding (mainly HD). The most significant enrichment in GO-BP associated with innate immune response and platelet activation, with 79 (HD)/105 (DLBCL) and 45 (HD)/67 (DLBCL) proteins identified in each category, respectively (Figure A3). Additionally for GO-BP, negative regulation of endopeptidase activity was mainly identified in HDs (Figure A3). GTPase activity, antigen processing related functions, and ubiquitin ligase activity in cell cycle were mainly enriched in samples from DLBCL patients (Figure A3).

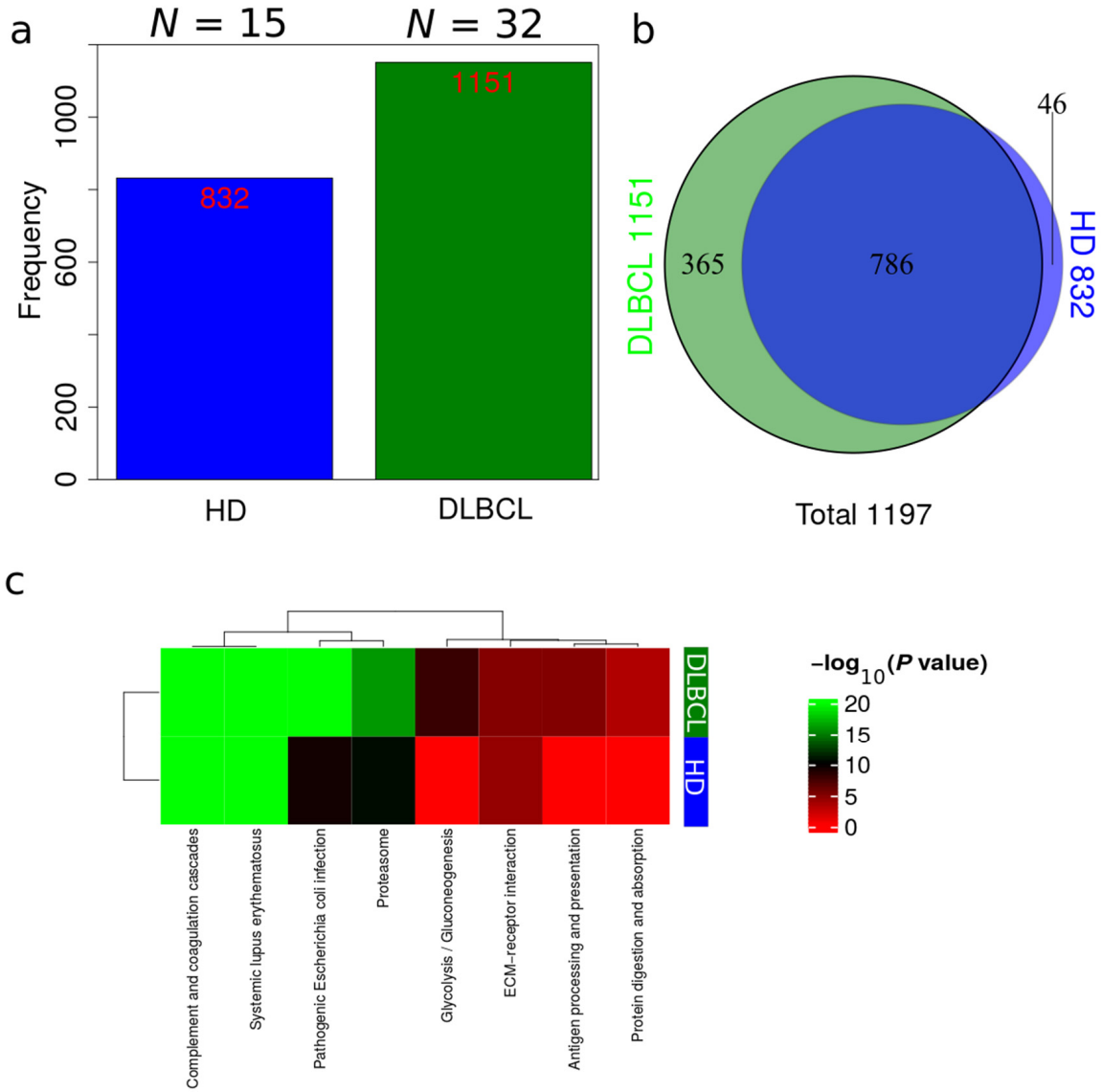


Figure 3. Overview of identifications in DLBCL patients and HDs. (**a**) Total number of identifications from DLBCL patients and HDs. (**b**) Venn diagram comparing identifications in DLBCL patients and HDs. (**c**) Summary of functional KEGG pathway analysis that are either enriched in HDs or DLBCL patients based on all identified proteins from DLBCL and HD groups.

To evaluate if the protein richness is indeed higher in the DLBCL group, a subsampling of 100 permutations was drawn, assuming a different number of samples. The total number of non-redundant proteins for each sample number was counted, and standard error was calculated for each draw (Figure 4). This modeling provides a view of protein richness given the measurement strategy in the two sample groups. The estimated protein richness

at the plateau by Jack1 method was 973 (45), and it was 1291 (53) for the HD and DLBCL groups, respectively (Figure 4b, gene collapsed proteins).

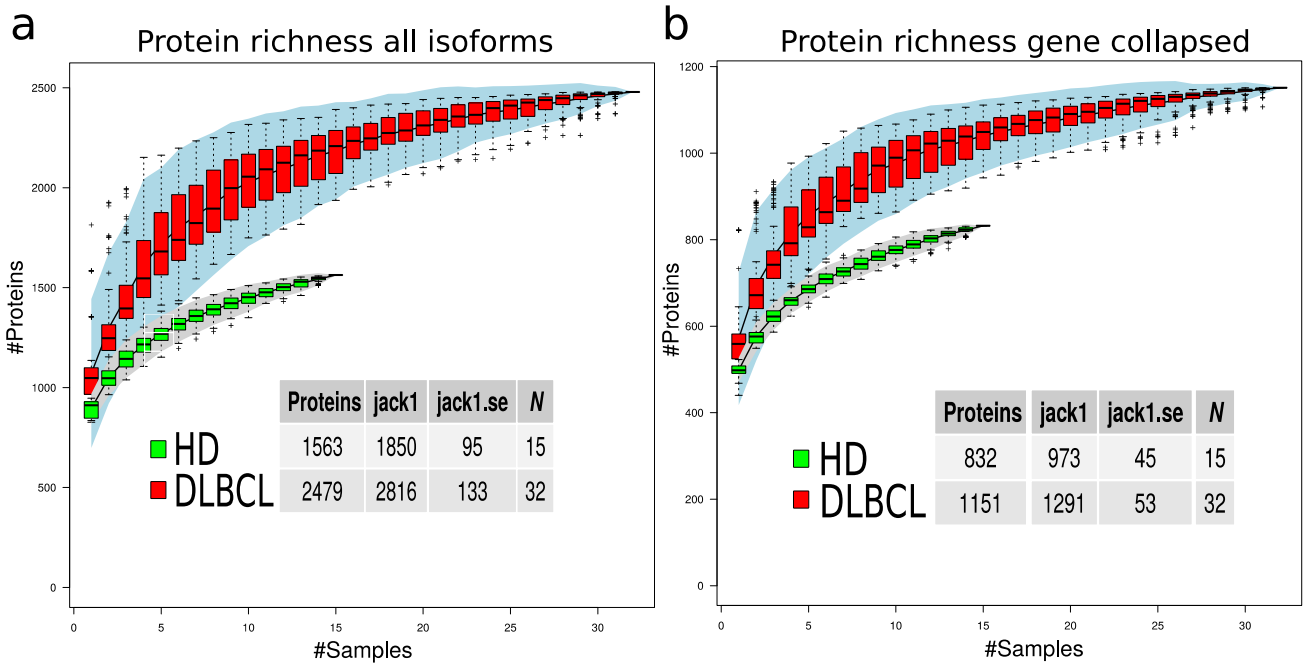


Figure 4. Protein richness in DLBCL and HD groups. (**a**) Protein richness curves for DLBCL (red boxes) and HD group (green boxes) when all protein isoforms are used. (**b**) Protein richness curves for DLBCL (red boxes) and HD group (green boxes) when gene-collapsed proteins were used. The lower tables indicate the estimated number of proteins in the two sample groups using different estimation methods and the corresponding standard error.

The statistical difference in number of protein identifications was estimated based on 50 random subsamples of 10 HD and 10 DLBCL sample groups (Figure A4) and the result indicates a highly significant difference in the number of proteins identified in the two sample groups. Overall, our analysis suggests that DLBCL samples contain a more diverse proteome and the variance among the number of identified proteins in the DLBCL group is higher than in the HD group (e.g., variance depicted in Figures 4 and A4).

2.5. Quantitative Analysis of Differential Regulated EV Proteome

The label-free quantitative proteomics data were obtained in two batches. To correct for batch effects, MS data was analyzed separately for principal component analysis (Figure A5) and considered in subsequent statistical analysis using the limma R package. In total, 1193 (DLBCL versus HD) and 1183 (non-GCB versus GCB) proteins were identified and quantified after collapsing into the encoded genes and removing low expression proteins. A statistical comparison of quantitative proteomics data between DLBCL and HD group (Table S2a) and between non-GCB and GCB (Table S3a) revealed significant differences in 264/402 expressed proteins from DLBCL versus HD (adj. p-value < 0.05/p-value < 0.05). For non-GCB and GCB 51 differentially expressed proteins (DEPs) were identified (p-value < 0.05).

Volcano plots showing the comparison between DLBCL and HD, and between non-GCB and GCB proteomes, are depicted in Figure 5a,b, respectively. Non-GCB versus GCB resulted in less DEP in relation to DLBCL versus HD comparison (Figure 5a,b). Given that platelets are considered a major contaminant in EV preparations, potential contaminant proteins from erythrocytes, platelets and the blood coagulation system extracted from Geyer et al. [23] are highlighted in red, orange and blue in the figure, respectively. Despite the fact that platelet annotated proteins are present in the plasma EV samples, minimal overlap of potential contaminant proteins with significant DEPs was observed

(Figure 5a,b). Immunoglobulins and guanine nucleotide-binding protein family members were up-regulated between DLBCL and HD (Figure 5a). The main immunoglobulin proteins are characterized by the specific rearrangements in the immunoglobulin kappa and lambda light chains such as IGKV3-20, IGKV3D-11 and IGLV1-40. The G proteins GNA12/GNA13 as well as GNAI1/3 were the main signal transducers overexpressed in DLBCL, while SERPINF1, VNN1, CLHC1 and HLA-A were downregulated in the EV plasma of DLBCL. In the GCB group, plasma EVs are enriched in histone proteins such as histone H2B as well as in heat shock protein TRAP1 and the glycosyltransferase ST3GAL6 (Figure 5b).

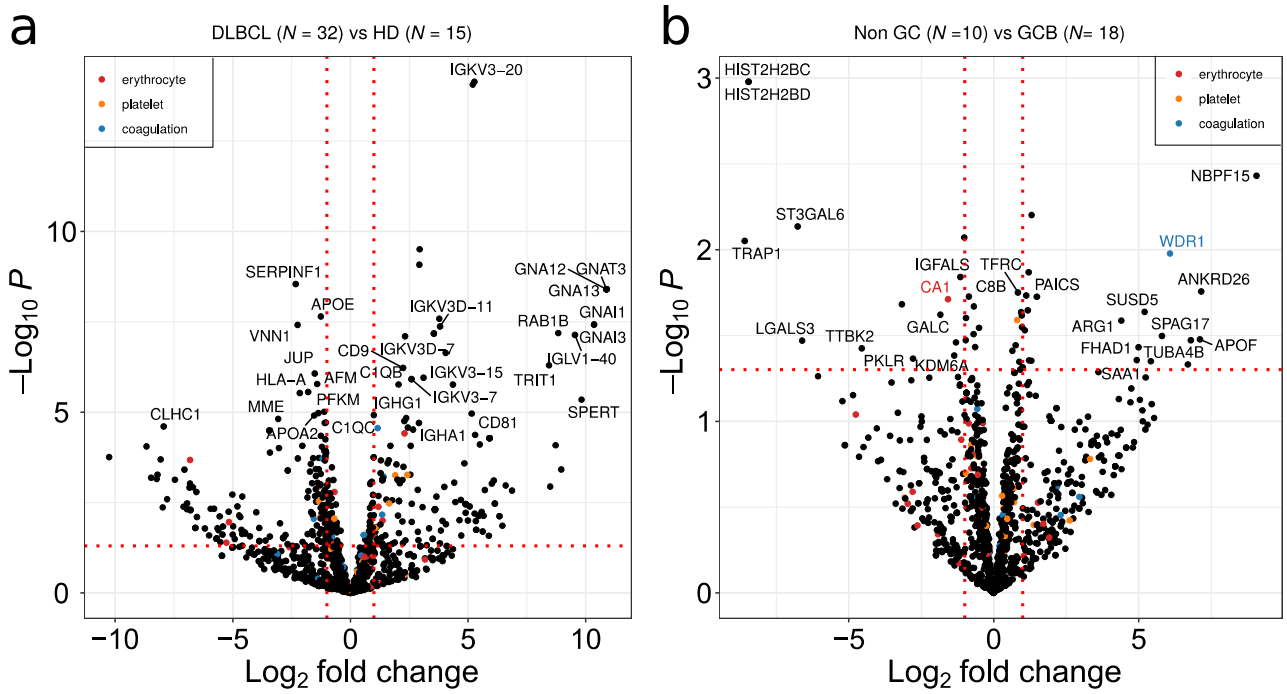


Figure 5. Volcano plot depicting deregulated proteins when comparing (**a**) DLBCL versus HD and (**b**) non-GCB versus GCB. A *p*-value threshold of 0.05 is indicated by a horizontal red dotted line. The $\log_2$ fold up or down regulation is represented by the red vertical line. Potential contaminants from erythrocyte, platelet and coagulation are indicated by red, orange and blue dots, respectively.

The results of the functional enrichment analysis of differentially expressed proteins are depicted in Figure 6 and summarized in Tables S2 and S3. Functional analysis of DEP for the comparison of DLBCL versus HD showed proteins involved in lupus related systemic inflammation (Figure 6a), platelet activation (Figure 6b), regulation of glycoprotein metabolic processes (Figure 6b), and proteins involved in GTPase activity (Figure 6b,c). Deregulated platelet activation (Figure 6b and Table S2) proteins were mainly up-regulated in DLBCL. Note that platelet contaminant markers described by Geyer et al. [22] are distinct from the functional category of platelet activation. In addition, within the GTPase activity category, there were significantly up-regulated and down regulated proteins in DLBCL (Figure 6c). The cellular location of deregulated proteins included the extracellular space, blood microparticles, and the nucleosome and complement complex (Figure 6d).

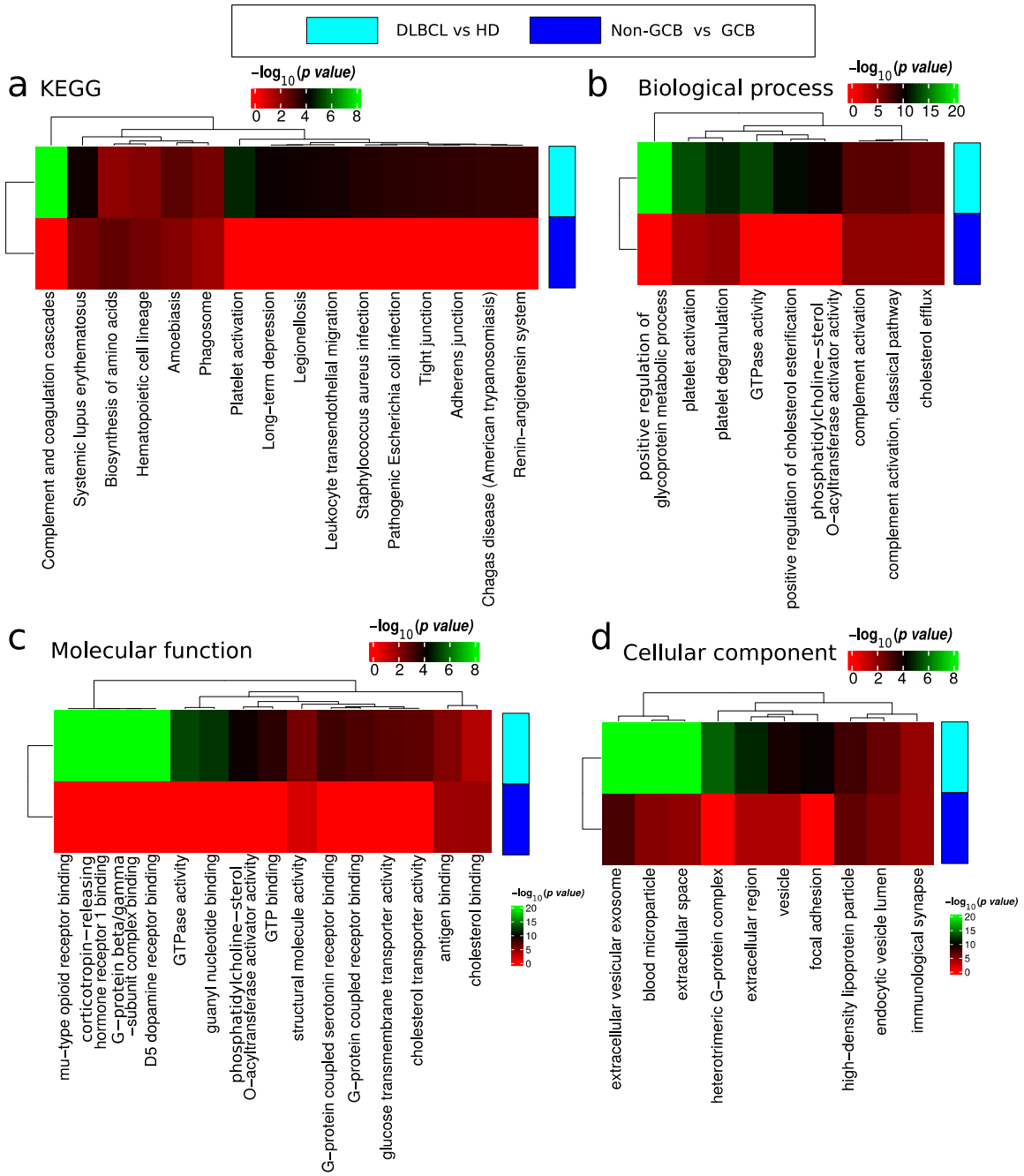


Figure 6. Functional enrichment analysis of differentially expressed proteins between DLBCL and HD and non-GCB and GCB. Analysis of (**a**) KEGG pathway, (**b**) biological process, (**c**) molecular function and (**d**) cellular component gene categories.

2.6. DLBCL vs. HD Classification

A partial least square (PLS) regression model was built, given the large number of significantly regulated proteins when comparing DLBCL and HD cases. A 66% split, ensuring a balanced training set, was applied to divide the data into training and test sets (Figure 7a). The PLS model tested on the test data displayed excellent performance based on receiving operating characteristics (ROC), with an area under the curve (AUC) of 1 (Figure 7b). The confusion matrix revealed zero misclassifications out of 15 cases (Figure 7d). The model was also built without enforcing a balanced training set, which again resulted in perfect classification without any misclassifications. Caution must, however, be highlighted due to the low number of test samples available. For a larger test set, some errors would

be expected. However, given the large proteome differences between EVs from DLBCL patients and HD, we expect that this classification based on the plasma EV proteome will present high accuracy in practice. The most important protein features for the model are represented in Figure 7c. Immunoglobulin proteins appeared with the highest importance score for the model closely followed by G protein GNAI3. This model is highlighted as a proof of concept to demonstrate the feasibility of classifications based on plasma EVs. However, it must be optimized on a larger dataset, which may change the optimal proteins and classification algorithm.

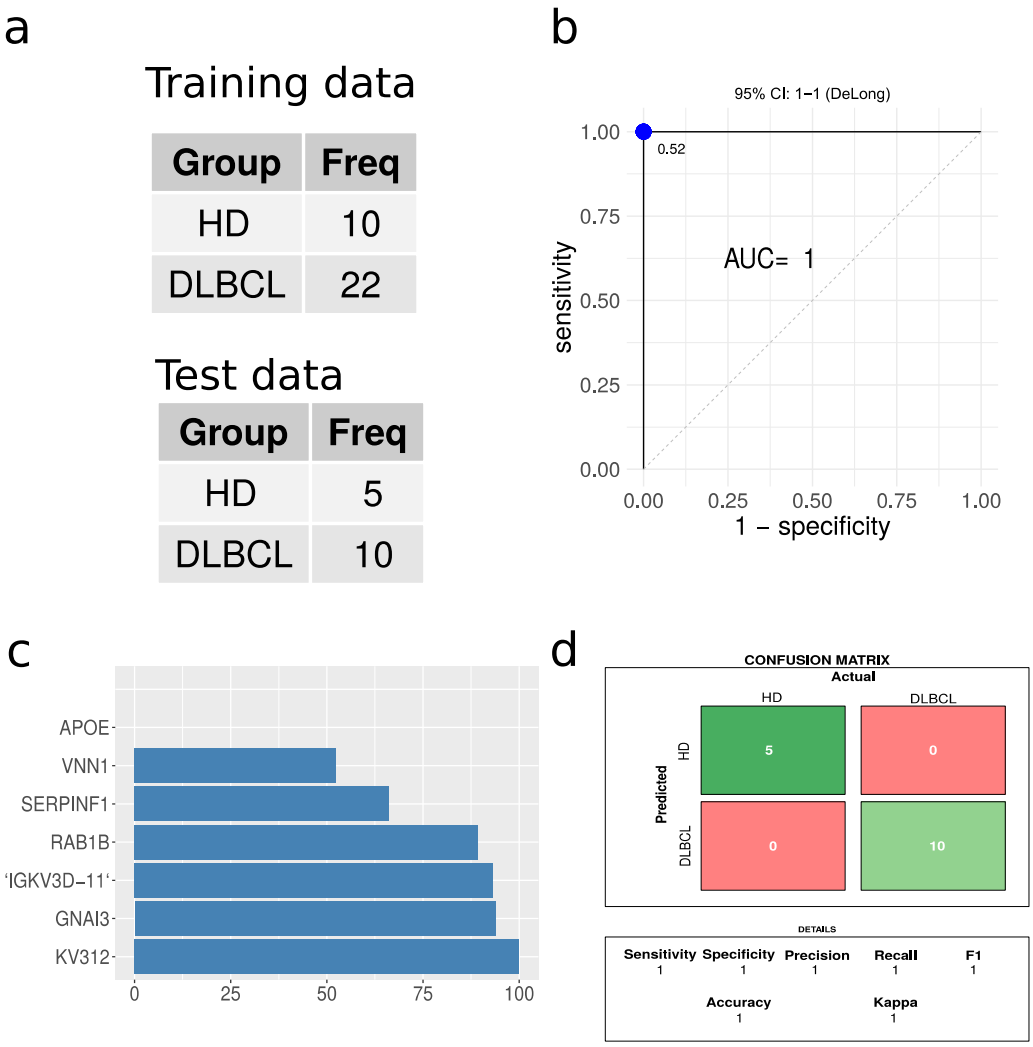


Figure 7. Performance of partial least squares regression analysis. (**a**) A summary of the training and test data sets used. (**b**) Receiver operating characteristics for the PLS regression model. The blue dot indicates the threshold based on Youden's J statistic. (**c**) The top ten protein features in terms of importance. (**d**) The confusion matrix and the performance parameters associated with it.

2.7. Survival Analysis

Univariate Cox proportional hazards regression models were fitted to the observed survival, proteomics and clinical data for all 32 DLBCL patients in the cohort. The proportional hazard (PH) assumption was assessed using Schoenfeld residuals [24]. The Cox–Mantel log-rank tests for the top proteins (p-value < 0.005 and Cox–Mantel log-rank tests < 0.01) are listed in Table 2. None of the measured clinical variables correlated significantly with survival when the above thresholds were applied. In addition, the total number of proteins identified in each case was also not correlated with survival. Figure 8 depicts the Kaplan–

Meier plots for immunoglobulin lambda constant 1 (IGLC1), immunoglobulin lambda like polypeptide 5 (IGLL5), proteasome subunit beta type-2 (PSMB2) and coronin-1A (CORO1a), which are among the most significant predictors for survival.

Table 2. Cox–Mantel log-rank tests and *p*-value estimates from Cox proportional hazards regression models.

Protein	*p*	LogRank	Protein	*p*	LogRank
IGLC1	3.31×10^{-4}	1.21×10^{-5}	POTEE	2.40×10^{-3}	3.07×10^{-4}
IGLL5	3.32×10^{-4}	1.99×10^{-5}	POTEI	2.60×10^{-3}	3.28×10^{-4}
IGLC6	3.58×10^{-4}	3.95×10^{-5}	POTEF	2.65×10^{-3}	3.74×10^{-4}
PSMB2	5.18×10^{-4}	1.98×10^{-5}	IGLC3	2.73×10^{-3}	7.07×10^{-4}
CORO1A	8.72×10^{-4}	2.84×10^{-6}	PSMA2	2.85×10^{-3}	9.08×10^{-4}
PSMA1	1.01×10^{-3}	4.16×10^{-6}	IGHV3_16	3.42×10^{-3}	4.72×10^{-4}
ARPC3	1.30×10^{-3}	1.76×10^{-5}	TUBA3C	3.46×10^{-3}	1.40×10^{-4}
HLA_E	1.42×10^{-3}	2.05×10^{-4}	ACTBL2	3.63×10^{-3}	1.31×10^{-5}
PSMB4	1.50×10^{-3}	1.38×10^{-4}	FCHO2	3.70×10^{-3}	3.57×10^{-4}
PON1	1.88×10^{-3}	1.01×10^{-4}	TUBA4A	4.05×10^{-3}	1.81×10^{-4}
ACTC1	1.98×10^{-3}	4.90×10^{-5}	SERPING1	4.06×10^{-3}	1.53×10^{-3}
ACTA1	1.98×10^{-3}	4.90×10^{-5}	TUBA1C	4.10×10^{-3}	9.24×10^{-5}
ACTG1	2.03×10^{-3}	3.94×10^{-5}	TUBA1A	4.17×10^{-3}	9.37×10^{-5}
ACTB	2.12×10^{-3}	6.31×10^{-5}	ENO3	4.32×10^{-3}	1.12×10^{-3}
ACTA2	2.32×10^{-3}	7.35×10^{-5}	TUBA1B	4.98×10^{-3}	1.51×10^{-4}
ACTG2	2.32×10^{-3}	7.35×10^{-5}			

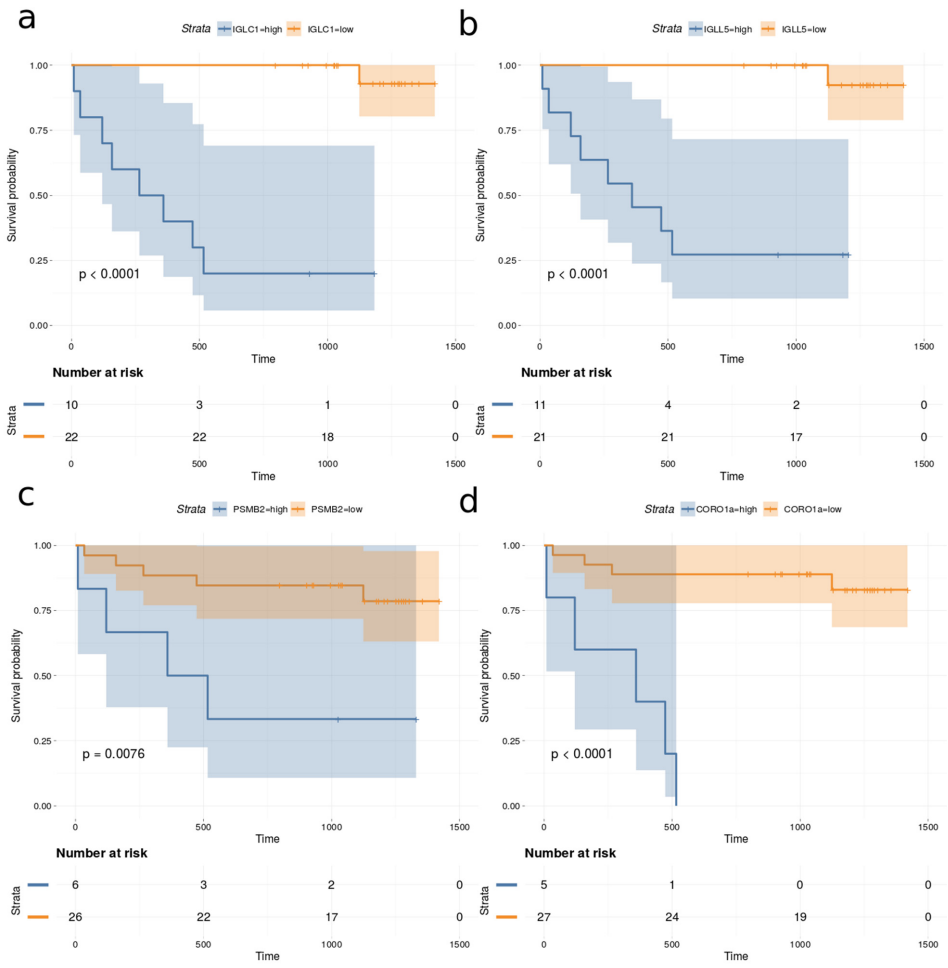


Figure 8. Kaplan–Meier plots based on dividing high and low expression of (**a**) IGLC1, (**b**) IGLL5, (**c**) PSMB2 and (**d**) CORO1a.

3. Discussion

Blood circulating EVs can reflect changes occurring in different tissues and organs. The abundance and population diversity of EVs in plasma is still a matter of debate. The quantified particle concentration in the EVs plasma fraction from DLBCL patients and HDs showed that cancer patient-derived plasma contained a significantly higher number of particles compared to that from HDs and a broader range of particle size mode (Figure 1). This has also been observed by others reporting a clinical significance for other types of cancers, such as, for example, glioblastoma [25]. To access EV plasma composition, we have presented here a quality assessment tool using the most reported proteins publicly available in EV database (ExoCarta) [20,26,27] based on MS-based proteomics studies of EVs isolated from the conditioned cell medium (CCM) by different isolation methods. The marker expression in the current study is then compared with marker expression from previous large EV studies. EVs from cell lines, body fluids such as bronchoalveolar lavage fluid, and plasma studies were included. DLBCL-plasma EVs were enriched in tetraspanins CD9, CD63 and CD81, and enriched in classical EV markers generated via the endocytic pathway (Figure 2). Proteins involved in EV biogenesis [13] such as PDCD6IP, FLOT1 and TSG101 were also identified, although at a lower expression level compared with the tetraspanins proteins. When compared to CCM-EVs, the protein level of EV markers was higher in CCM-EVs compared with plasma EVs (Figure 2). Isolation of plasma EVs is normally restricted to a few milliliters of sample compared with an almost infinite amount of CCM potentially available. On the other hand, the diversity of EV subpopulations and other particles in plasma surpasses that of CCM, which normally originates from a single cell line. Nevertheless, when the ranked expression of tetraspanins and PDCD6IP, FLOT1 and TSG101 is compared between CCM and plasma derived EV samples, our study suggests that the proteins identified are enriched in extracellular vesicles of the endocytic pathway. However, the presence of vesicles generated via the secretory pathway in plasma cannot be discarded. Indeed, GO functional enrichment analysis of cellular components of identified proteins among DLBCL and HD groups are equally enriched in cellular component categories such as extracellular region, extracellular vesicular exosome, extracellular space, proteasome complex and blood microparticle (Figure A1). Nevertheless, in cancer patients, proteins in cellular component categories related to vesicle and focal adhesion are more enriched compared to those from HD (Figure A1). In our study, CD9 and CD81 were 2.2- and 5.2 $\log_{2}$- fold significantly up regulated (adj. p-value < 0.001) between the DLBCL and HD group, which suggests that the increased particle concentration in plasma comes from vesicles with CD9 and CD81 markers. A significant increase in CD9-CD63 EVs was also observed in the plasma of colorectal cancer patients compared to age-matched HDs [28]. Blood can, in principle, contain a variety of EVs released from many different cell types, but studies demonstrate that the main source of EVs in HD plasma originates from platelets (primarily) and erythrocytes [29]. On the other hand, tumour cells secrete an increased number of EVs, which have been described to contain molecules capable of transferring oncogenic traits to the neighboring cells or to disseminate to distant organs. In our study we observed that the proteome of DLBCL plasma EV contain unique proteins which can derive from the tumor itself or from cells involved in the systemic response (Table S2). The accumulation of exosomes carrying tissue factor (TF) in plasma has been linked to venous thrombolytic events, which are common in hematologic malignancies [30].

TF can interact with components of the coagulation cascade, and the large numbers of pro-coagulant vesicles in the plasma of cancer patients is considered a prognostic factor for dismal outcome [30]. Furthermore, it should be noted that, as proposed by other authors, immune cells probably secrete most EVs in lymph nodes or locally at the required site and, thus, only low levels of these particles would be present in the circulation [31]. These examples illustrate the variety of studies that are being carried out in the cancer field and the promising potential of EVs. Nevertheless, they also highlight the long path ahead.

With regard to plasma EVs, our study demonstrated that it reflected a pathological condition since 1151 proteins have been identified in EV plasma from DLBCL patients compared to 832 in HD EV plasma samples from HD. Even though the sampling between DLBCL and HD groups was unbalanced, estimation of protein richness verified a highly significant difference in the number of proteins identified in the two sample groups (Figure 3). This significant higher proteome complexity in malignant samples is in line with our previous findings from bronchoalveolar lavage samples [19,32]. Regarding protein quantification, 264/402 proteins were significantly regulated (adj p-value < 0.05/p-value < 0.05) between DLBCL patients and HDs. Among the significantly regulated proteins were some related to innate immune response. More specifically, these included immunoglobulins and proteins binding to G protein–coupled receptors such as guanine nucleotide-binding proteins, GNA12/13 (Figures 5a and 6). B-cell lymphomas are characterized by somatic hypermutation (SHM) events in immunoglobulin genes that lead to abnormal SHM; IGV SHM may enhance the B-cell receptor (BCR) affinity and B-cell survival, suggesting unfavorable prognostic effects [33]. Young et al. demonstrated that DLBCL tumor cells bear rearranged IgVH segments that are specific for ABC subtype tumors, and its viability depends on autoantigens [34]. Recently, we have performed a comprehensive proteomics profiling of EVs isolated from a panel of DLBCL cell lines, including GCB and ABC-like subtypes. In fact, the B-cell receptor pathway was significantly enriched in EVs secreted from DLBCL-ABC subtype cells [14]. G proteins are signaling transducers, and although they don't mediate directly signaling cascades triggered by the antigen activation of immune cells, G proteins might have important modulatory roles. For example, Gi protein signaling inhibition enhanced the capacity of splenocytes to produce IL-12 in response to both microbial and nonmicrobial stimuli.

G-proteins are up-regulated in GCB cells, with *Gna13* transcripts appearing more abundantly [35]. EVs have been implicated in several functions, such as in adaptive immune response to eliminate tumor cells [31]. The H2B protein family was regulated between non-GCB and GCB groups, and its detection in plasma has been associated with neutrophil extracellular traps (NETs) as well as vesicles [13]. Interestingly, early reports on NETs describe the antimicrobial activity of histones [36].

Among the proteins differentially regulated between DLBCL and HD groups, when DLBCL subtypes were considered, there was a clear enrichment for proteins in the Systemic Lupus erythematosus (SLE) KEGG pathway (Figure 6). SLE is an autoimmune disease in which immune complexes comprised of autoantibodies and self-antigens are deposited, particularly in the renal glomeruli, and mediate a systemic inflammatory response by activating complement or via Fcgamma receptor (FcgR)-mediated neutrophil and macrophage activation. In fact, the risk of SLE in hematological cancers, especially in non-Hodgkin's lymphoma, is increased approximately four-fold [37], and this has been postulated to be due to both immunological abnormalities and drug exposures as well as viral origin. Pentraxin 3 (PTX3) was downregulated in the EV fraction plasma of DLBCL patients. PTX3 can be induced by different soluble factors such as proteins from the interferon family, interleukins and tumour necrosis factors [38]. The role of PTX3 in cancer is not clear, and its contribution as a tumour promoter or suppressor is being equally reported. In multiple myeloma (MM), PTX3 affects in vitro and in vivo fibroblast growth factor 2-mediated MM angiogenesis, promoting cytotoxicity in MM cells by inhibiting the cross-talk between plasma cells, endothelial cells and fibroblasts [39]. On the other hand, it has been demonstrated in PTX3 deficient mice that tumorigenesis is promoted and is mediated by complement-dependent tumor-promoting inflammation [40].

Single cell analysis using CyTOF demonstrated that DLBCL tumor cells are phenotypically unique to each patient, with a small number of samples displaying distinct sub-clonal structures, often distinguished by the differential expression of immune-related proteins such as MHC-II rather than GCB vs. ABC [41]. However, this might be a result of the limited number of samples. However, several studies [42] have described poor agreement with non-GCB and GCB classification for prognosis, suggesting that in DLBCL, systemic compo-

nents have the potential to contribute to refining the subclassification of this most prevalent type of NHL. In fact, we identified multiple protein markers associated with survival, such as IGLC1, IGLL5, PSMB2 and CORO1a (Table 2). All of these potential protein markers have previously been associated with poor prognosis in other cancers [43–46]. It is possible that IGLC1 and IGLL5 stems from co-isolated immune complexes rather than EVs or may also be associated with the membrane of the EVs. PSMB2 is part of the proteasome complex, which is typically upregulated in many cancers. Unfortunately, inhibition of the proteasome with molecules such as bortezomib [47] and lenalidomide [48] in relapsed/refractory DLBCL has demonstrated modest responses. However, PSMB2 is not targeted by bortezomib and lenalidomide and may constitute a potential target. CORO1a plays a role in recurrent viral infection and is also associated with a poor survival rate, having a positive correlation with lymph node metastasis in gastric cancer patients [49]. Additionally, the present results demonstrate excellent separation for DLBCL patients compared to HDs, suggesting that EV proteins or a subset thereof may be useful for diagnosis or monitoring risk of recurrence. The model presented here is considered a proof-of-concept and should be tested on a larger cohort and compared with other clinical parameters already proven to have diagnostic value. Finally, this study opens a new research line aimed at understanding diversity in the EV proteome of DLBCL patients and its impact on therapeutic response.

4. Materials and Methods

4.1. Plasma of Cancer Patients and Healthy Donors

We have collected blood from DLBCL patients at diagnosis, prior to treatment, and of age-matched HDs, and separated the blood into blood cells, platelets and plasma. Venous blood samples were obtained from patients with DLBCL that were receiving treatment at the Portuguese Institute of Oncology of Lisbon Francisco Gentil, as well as from healthy volunteers. Blood samples were collected at diagnosis at a single center between December 2017 and January 2020. All patients were treated with R-CHOP therapy or related regimens (R DAEPOCH, RCEOP), with the exception of one patient who received palliative care. On average, the treatment period lasted between three to five months, with a median follow-up of 1 year. All patients received R-CHOP or R-CHOP-like chemotherapy with curative intent. At the time of diagnosis, only one patient presented with central nervous system involvement. Peripheral blood was collected from newly diagnosed DLBCL patients prior to treatment and from age matched healthy volunteers into EDTA lavendertop tubes. Collection tubes were centrifuged at 200× *g* for 10 min at RT. The top layer was centrifuged at 1000× *g* for 10 min to collect the plasma. Plasma was aliquoted into 2 mL vials and stored at −80 °C until further use.

4.2. Isolation of Extracellular Vesicles from Human Plasma

As previously described [18], frozen plasma specimens were thawed and centrifuged at 3000× *g* for 20 min at 4 °C and then at 12,000× g for 60 min at 4 °C. Clarified plasma was ultracentrifuged in an Optima TM L-80XP ultracentrifuge (Beckman Coulter, Brea, CA, USA) at 100,000× *g* at 4 °C for 120 min with a Type 50.2 Ti rotor (k-factor: 157.7) to pellet EVs. The supernatant was carefully removed, and crude EV-containing pellets were resuspended in ice-cold PBS, followed by floatation on a sucrose cushion (30%, D_2O) (Sigma-Aldrich, Saint Louis, MO, USA) for 60 min at 100,000× *g* at 4 °C to remove non-EV protein complexes. After washing, by pelleting the EVs collected in the sucrose cushion for 16 h at 100,000× *g* at 10 °C, EVs were resuspended in phosphate-buffered saline (Sigma-Aldrich, Saint Louis, MO, USA) and subjected to nanoparticle tracking analysis using NanoSight NS300 equipment (Malvern Instruments, Inc., Westborough, MA, USA).

4.3. Protein Measurements

Protein concentrations in isolated exosome fractions were measured using a bicinchoninic acid (BCA) protein assay kit (Pierce Biotechnology, Rockford, IL, USA) according to the manufacturer's instructions.

4.4. Peptide Sample Preparation

Samples containing a minimum of 20 µg of total proteins were further processed by the filter-aided sample preparation (FASP) method. In short, protein solutions containing SDS and DTT were loaded onto filtering columns (Millipore, Billerica, MA, USA) and washed exhaustively with 8M urea (GE, Healthcare, Marlborough, MA, USA) in HEPES buffer (Sigma-Aldrich, Saint Louis, MO, USA) as previously described [14]. Proteins were reduced with DTT (BioRad, Hercules, California, USA) and alkylated with IAA (GE, Healthcare, Marlborough, MA, USA). Before trypsin digestion overnight at 37 °C, proteins were equilibrated with ammonium bicarbonate buffer (Sigma-Aldrich, Saint Louis, MO, USA). Protein digestion was performed by overnight digestion with sequencing grade trypsin (Promega, Madison, WI, USA).

4.5. Mass Spectrometry Analysis

As previously described [14], samples were analyzed by mass spectrometry-based proteomics using nano-LC-MSMS equipment (Dionex RSLCnano 3000) coupled to an Exploris 480 Orbitrap mass spectrometer (Thermo Scientific, Hemel Hempstead, UK). In brief, samples were loaded onto a custom-made fused capillary pre-column (2 cm length, 360 µm OD, 75 µm ID, flowrate 5 µL per minute for 6 min) packed with ReproSil Pur C18 5.0 µm resin (Dr. Maisch, Ammerbuch-Entringen, Germany), and separated using a capillary column (25 cm length, 360 µm outer diameter, 75 µm inner diameter) packed with ReproSil Pur C18 1.9-µm resin (Dr. Maisch, Ammerbuch-Entringen, Germany) at a flow of 250 nL per minute. A 56 min linear gradient from 89% A (0.1% formic acid) to 32% B (0.1% formic acid in 80% acetonitrile) was applied. Mass spectra were acquired in positive ion mode in a data-dependent manner by switching between one Orbitrap survey MS scan (mass range *m/z* 350 to *m/z* 1200) followed by the sequential isolation and higher-energy collision dissociation (HCD) fragmentation and Orbitrap detection of fragment ions of the most intense ions with a cycle time of 2 s between each MS scan. MS and MSMS settings: maximum injection times were set to "Auto", normalized collision energy was 30%, ion selection threshold for MSMS analysis was 10,000 counts, and dynamic exclusion of sequenced ions was set to 30 s.

4.6. Database Search

The obtained data from the 98 LC-MS runs of 15 HD and 32 DLBCL EV samples were searched using VEMS [50] and MaxQuant [51] (all samples were run with two replicas except one DLBCL and one HD sample that were run as triplicates). The MSMS spectra were searched against a standard human proteome database from UniProt (3AUP000005640). Permuted protein sequences, where arginine and lysine were not permuted, were included in the database for VEMS. Trypsin cleavage allowing a maximum of four missed cleavages was used. Carbamidomethyl cysteine was included as fixed modification. Methionine oxidation, lysine and N-terminal protein acetylation, serine, threonine and tyrosine phosphorylation, ubiquitin diglycine and "LRGG" remnant on lysine were included as variable modifications. A 5 ppm mass accuracy was specified for precursor ions and 0.01 *m/z* for fragment ions. The false discovery rate (FDR) for protein identification was set at 1% for peptide and protein identifications. No restriction was applied for minimal peptide length for VEMS search. Identified proteins were divided into evidence groups as defined by Matthiesen et al. [52].

4.7. Statistical Analysis and Machine Learning

Statistical analysis of identified proteins was performed in the R statistical programming language. Protein richness in sample groups given the experimental protocol was estimated using the vegan R package [53]. vegan was also used to plot protein accumulation curves.

Quantitative data from MaxQuant and VEMS were analyzed in the R statistical programming language version 4.04 (The R Foundation, Vienna, Austria). Protein label free

quantitation (LFQ) and protein spectral counts from the two programs were preprocessed by removing common MS contaminants followed by a $\log_2(x + 1)$ transformation and removing common MS contaminants. Proteins with an average expression below 5 on the $\log_2$ scale were filtered out. Protein LFQ values were subjected to statistical analysis utilizing R package limma [54], where the contrast between DLBCL versus HD and non-GCB versus GCB were specified (Tables S1a and S2a).

Correction for multiple testing was applied using the method of Benjamini & Hochberg [55]. A volcano plot was plotted with ggplot software (The R Foundation, Vienna, Austria) followed by annotating data points with potential contaminant proteins from erythrocyte, platelet and coagulation extracted from Geyer et al. [23]. The R package pROC was applied for plotting receiver operator characteristic (ROC) and calculating the Youden's J statistic [56].

The partial least squares (PLS) regression models were constructed using the R package caret [57]. The data were $\log_2$ transformed, centered, and scaled. The training and test data (N = 47) included only data with a definitive diagnosis and no missing data. The data were split to ensure a balanced training set, with 66% of the data used for model training and the remaining 33% used to test the reported performance. Data splitting was constructed to ensure a balanced training set. Highly correlated features with correlation above 0.9 and zero variance features were removed. Before training the PLS classifier, a univariate feature selection based on *p*-values calculated by the limma R package in the training data set were applied. The presented model was restricted to the top 20 features. The training model was optimized using 10-fold repeated cross validation and accuracy as the performance metric.

4.8. Functional Enrichment Analysis

Functional enrichment based on the hypergeometric probability test was performed as described previously in R [58,59]. Functional enrichment was based on extracting all functional categories for which at least one of the samples showed a significant enrichment based on the hypergeometric probability test. For these functional categories, the matching proteins' gene names and numbers of proteins matching the functional categories were extracted and the estimated *p* values were $-\log_{10}$ transformed and plotted as heatmaps. Functional enrichment was performed for all identified proteins in each sample group and deregulated proteins when comparing sample groups. Cellular component (CC), biological process (BP), molecular function (MF) and KEGG functional annotation were considered in the analysis. In the case of KEGG enrichment for identified proteins, a filter was applied to only maintain the functional categories with at least 50% unique proteins identified compared to all functional categories with higher significance. The network of how all the top functional categories significantly enriched for KEGG overlap in terms of identified proteins is depicted in Figure A6.

4.9. Survival Analysis

Univariate Cox proportional hazards regression models were fitted using the R package survival [60] and the parameter ties were set to "breslow". Kaplan-Meier plots were constructed with the R package RTCGA [61].

5. Conclusions

The results described in this study constitute a proof of concept for the application of proteins EVs as a set of biomarkers for the detection of DLBCL in patient plasma. We demonstrated that the plasma EV fraction was enriched in EV and blood microparticles and could effectively identify DLBCL patients. Classification of DLBCL subtypes based on EVs proteome appeared to be correlated with immune system components such as immunoglobulins proteins or G-protein, but further research on larger patient cohorts is required to confirm this assertion. Moreover, we acknowledge that deeper characterization of plasma EV fraction would improve the sensitivity and specificity of biomarkers. In summary, our data indicate that the protein cargo of plasma EVs from DLBCL patients that reflects the pathological condition might serve as a liquid tumour biopsy upon further

validation, and it has the potential to become a proxy of clinical prognosis. Single parameter, multi-parameter and omics approaches can facilitate the identification of reliable cancer biomarkers in EVs, but further work is still required to validate the proposed candidates.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms232113327/s1.

Author Contributions: Conceptualization, R.M., A.S.C., M.G.d.S., J.C.; methodology, R.M., M.C.S.M., B.C.-S., A.S.C., P.G.; formal analysis, A.S.C. and R.M.; investigation A.S.C., P.G., C.B., M.C.S.M., B.C.-S., J.C., M.G.d.S., H.C.B. and R.M.; resources, A.S.C. and R.M.; writing—original draft preparation, A.S.C. and R.M.; writing—review and editing, R.M., P.G., A.H., C.B., M.C.S.M., B.C.-S., J.C., M.G.d.S., H.C.B., A.S.C.; visualization, A.S.C. and R.M.; project administration, A.S.C. and R.M.; funding acquisition, A.S.C. and R.M. All authors have read and agreed to the published version of the manuscript.

Funding: R.M. is supported by Fundação para a Ciência e a Tecnologia (CEEC position, 2019–2025 investigator). This article is a result of the projects (iNOVA4Health—UIDB/04462/2020), supported by Lisboa Portugal Regional Operational Programme (Lisboa2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). This work is also funded by FEDER funds through the COMPETE 2020 Programme and National Funds through FCT—Portuguese Foundation for Science and Technology under the projects number PTDC/BTM-TEC/30087/2017 and PTDC/BTM-TEC/30088/2017.

Institutional Review Board Statement: This study was approved by the Commission for Ethics and Health of the Instituto Português de Oncologia Francisco Gentil with the title "Diffuse Large B-cell lymphoma blood-based proteome signatures" with the reference UIC/1174 with favorable approval at 11 October 2018 where the samples have been collected. The study was also approved by the Ethical Commission of NMS | FCM-UNL (CEFCM) with the title Diffuse Large B-cell Lymphoma blood-based proteome signatures with the reference Nº01/2019/CEFCM) on 26 March 2019, where the samples have been analyzed.

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

Data Availability Statement: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [62] partner repository with the dataset identifier PXD036735.

Conflicts of Interest: The authors declare that they have no conflict of interest.

Appendix A

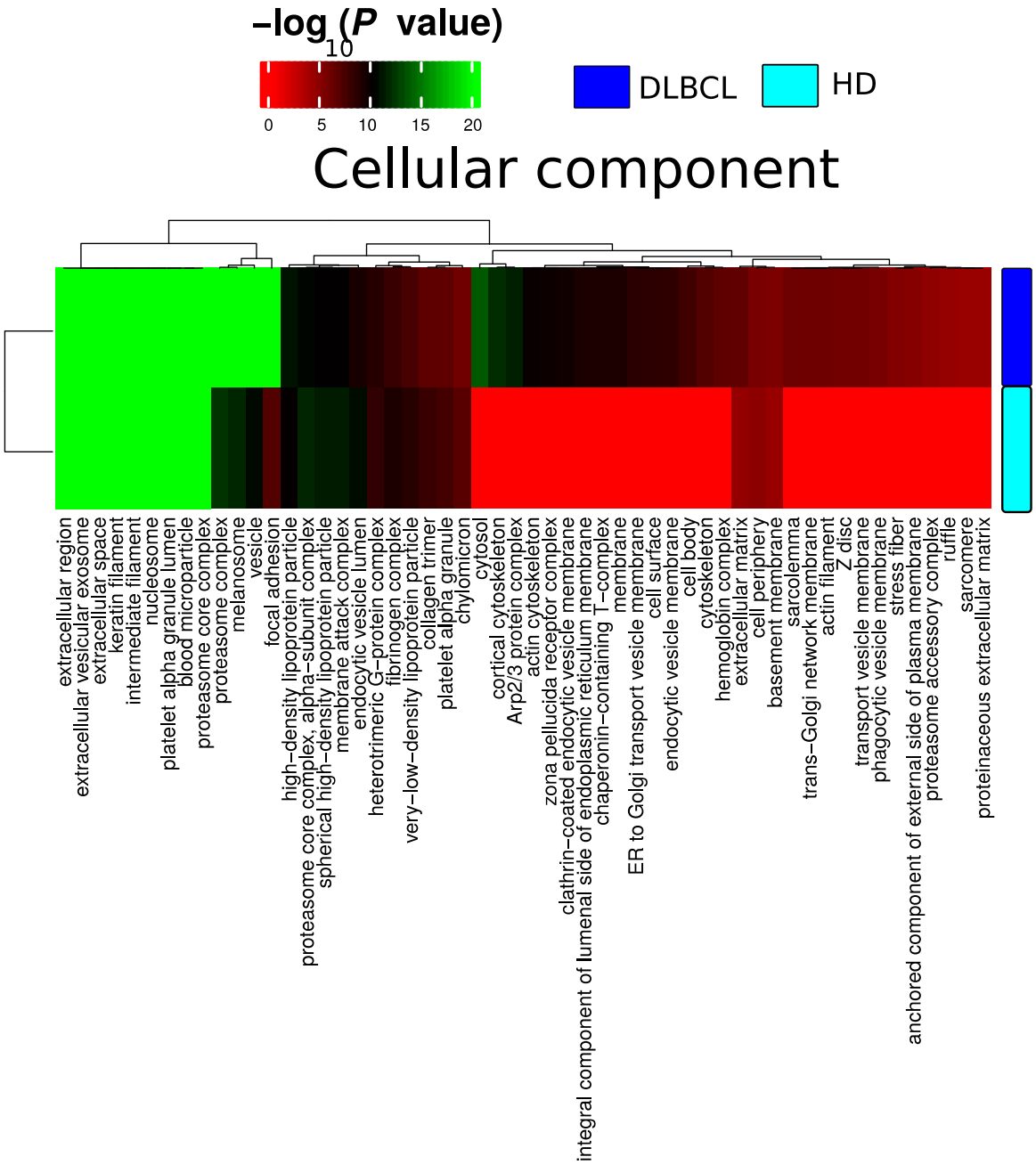


Figure A1. Summary of functional categories that are either enriched in HD or DLBCL based on functional enrichment in all identified protein from DLBCL and HD when compared with the cellular component.

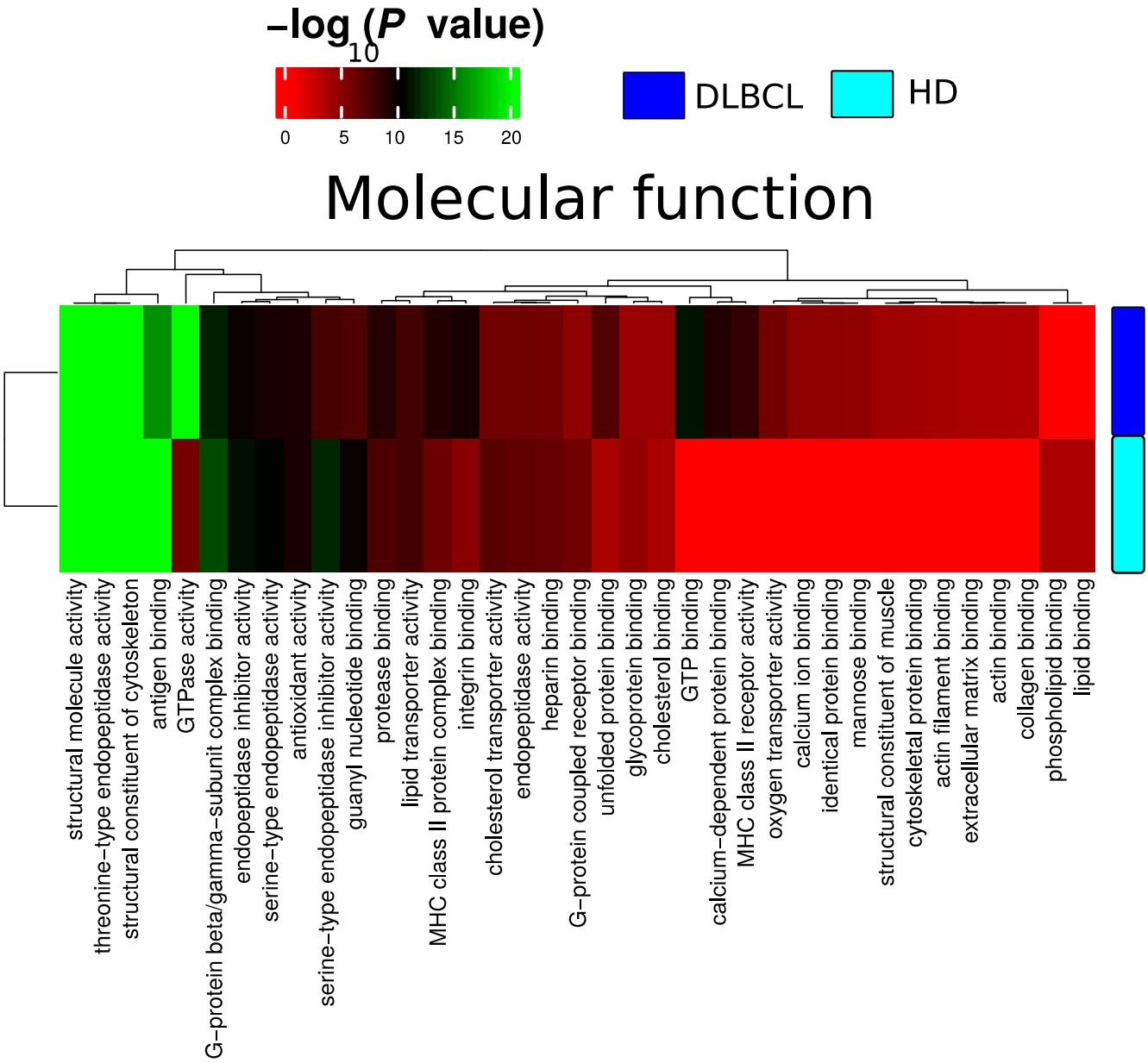


Figure A2. Summary of functional categories that are either enriched in HD or DLBCL based on functional enrichment in all identified protein from DLBCL and HD when compared with molecular function.

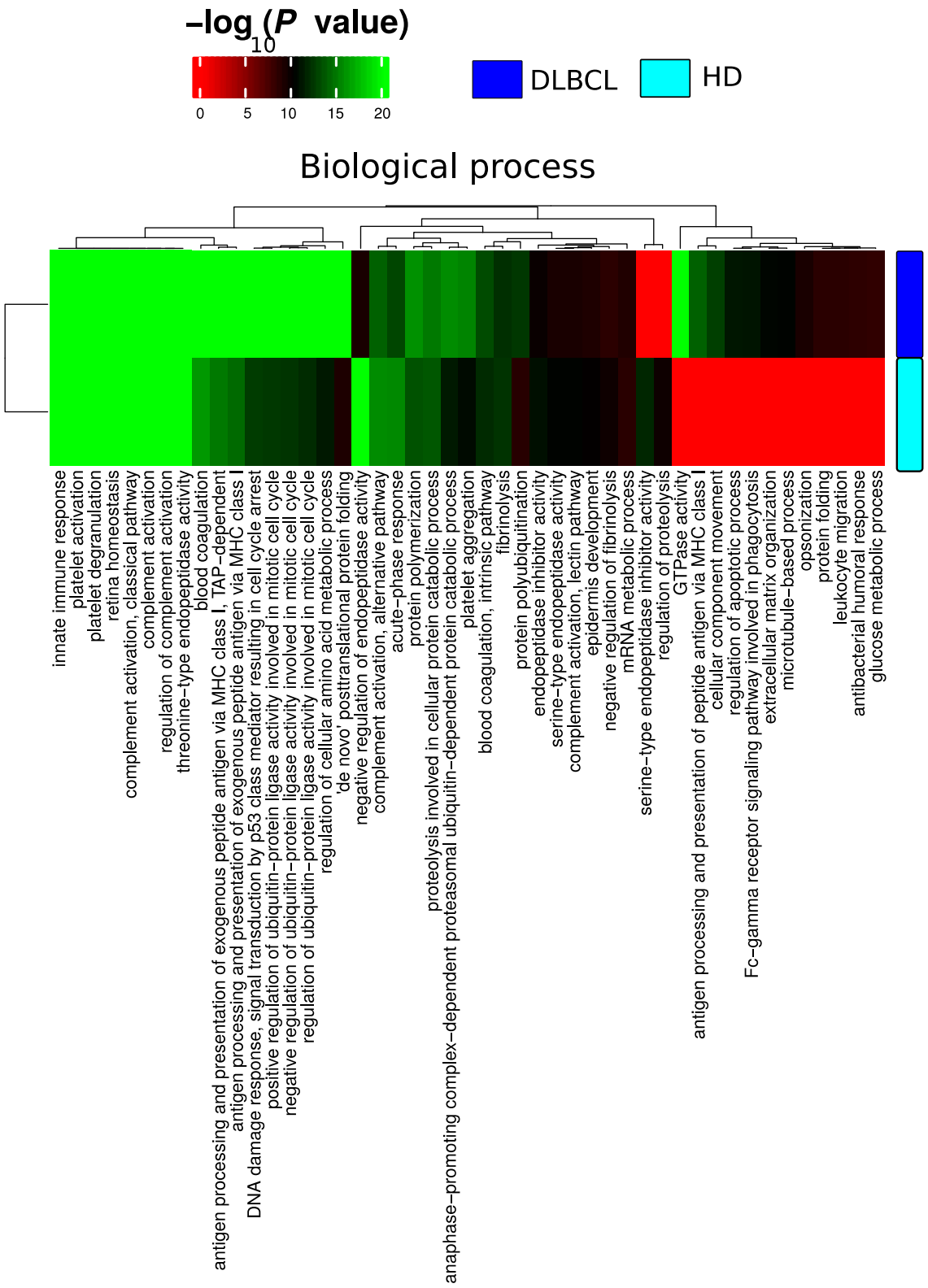


Figure A3. Summary of functional categories that are either enriched in HD or DLBCL based on functional enrichment in all identified protein from DLBCL and HD when compared with the biological process.

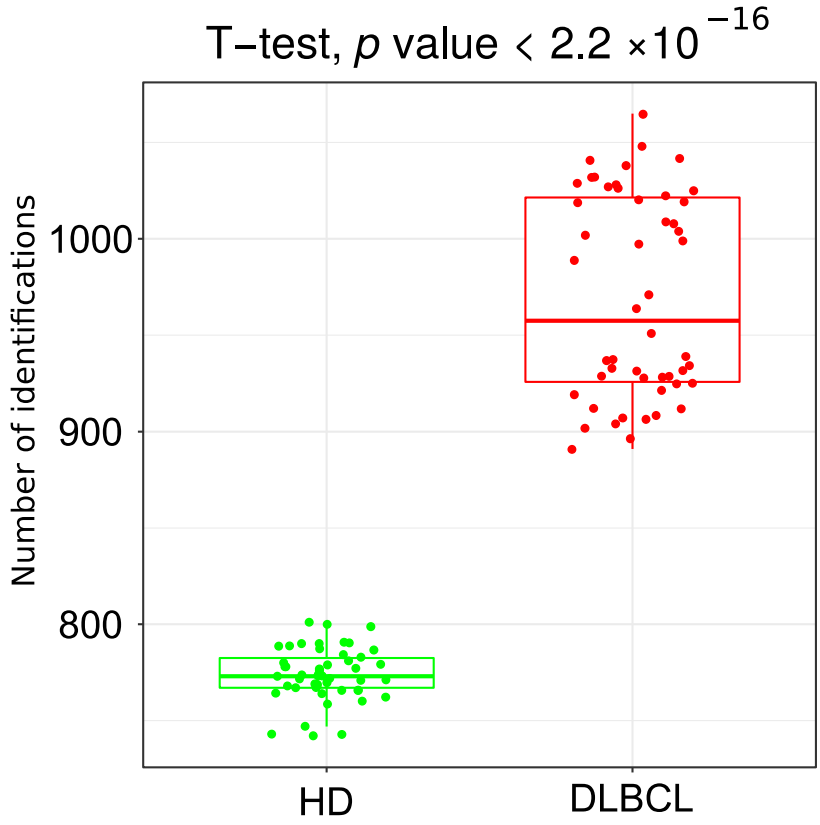


Figure A4. Boxplot of total number of gene collapsed proteins identified from 50 random subsamples of 10 DLBCL and 10 HD samples.

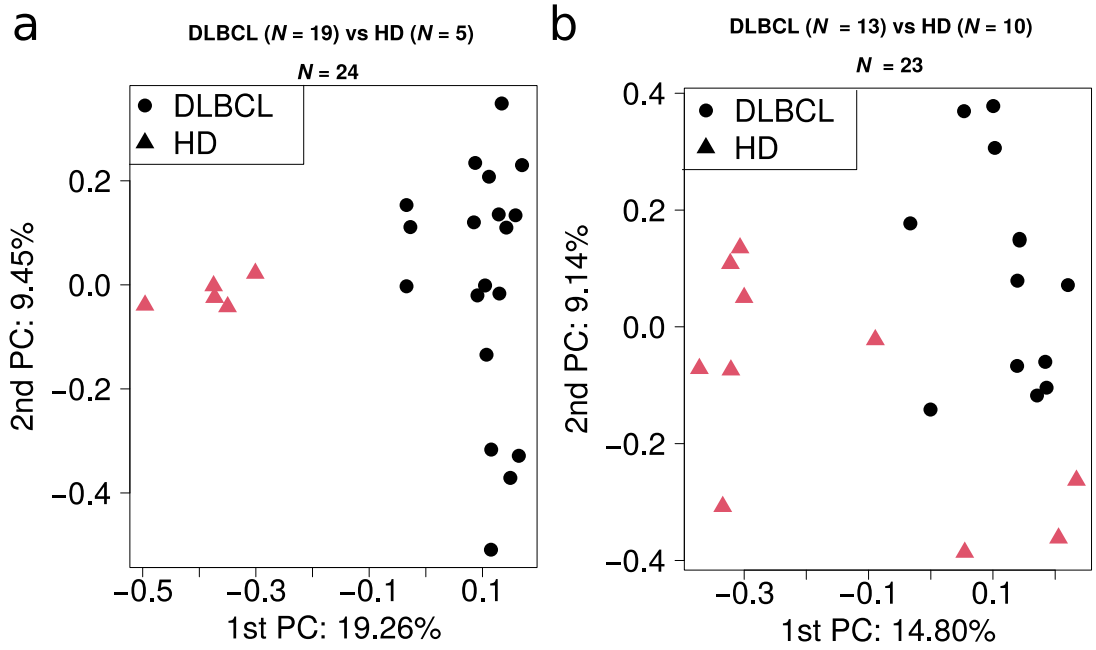


Figure A5. Principal component analysis of all quantitative values obtained from (**a**) batch 1 and (**b**) batch 2.

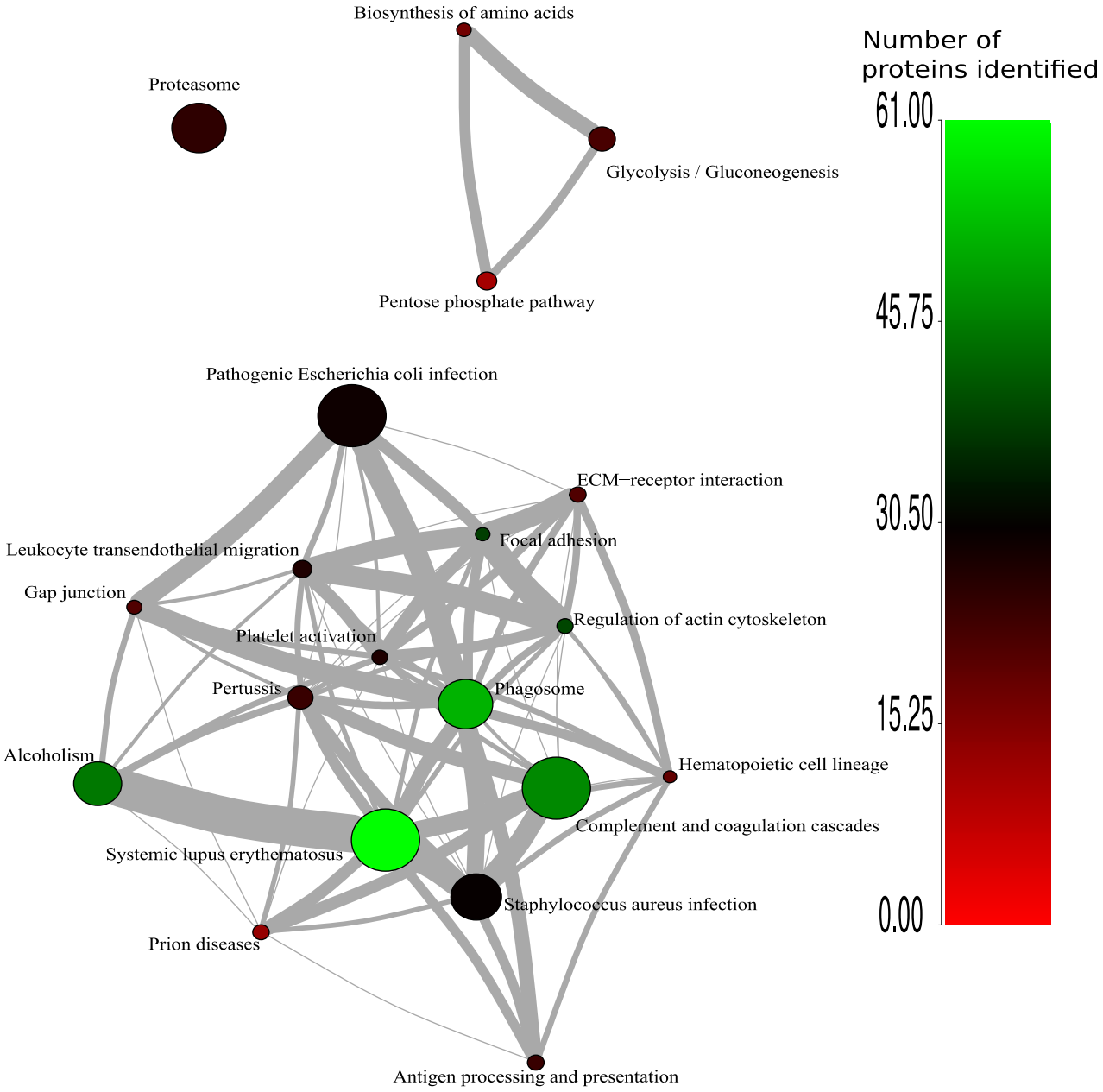


Figure A6. Analysis of overlapping proteins between enriched KEGG pathways for DLBCL. The color coding of the nodes indicates the number of proteins regulated for each category. The size of the nodes reflects $-\log_{10}$ of enrichment *p*-value. The thickness of the edges reflects the number of shared regulated proteins between the KEGG pathways (ranging from one to 31 proteins).

References

1. Swerdlow, S.H.; Campo, E.; Pileri, S.A.; Harris, N.L.; Stein, H.; Siebert, R.; Advani, R.; Ghielmini, M.; Salles, G.A.; Zelenetz, A.D.; et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **2016**, *127*, 2375–2390. [CrossRef]
2. Habermann, T.M. New developments in the management of diffuse large B-cell lymphoma. *Hematology* **2012**, *17* (Suppl. 1), S93–S97. [CrossRef] [PubMed]
3. Xie, Y.; Pittaluga, S.; Jaffe, E.S. The histological classification of diffuse large B-cell lymphomas. *Semin. Hematol.* **2015**, *52*, 57–66. [CrossRef] [PubMed]
4. 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]

5. 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] [PubMed]
6. 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]
7. Alizadeh, A.A.; Eisen, M.B.; Davis, R.E.; Ma, C.; 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]
8. Lenz, G.; Wright, G.W.; Emre, N.C.; 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]
9. Rosenwald, A.; Wright, G.; Chan, W.C.; Connors, J.M.; Campo, E.; Fisher, R.I.; Gascoyne, R.D.; Muller-Hermelink, H.K.; Smeland, E.B.; Giltnane, J.M.; et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **2002**, *346*, 1937–1947. [CrossRef]
10. 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]
11. Gutierrez-Garcia, G.; Cardesa-Salzmann, T.; Climent, F.; Gonzalez-Barca, E.; Mercadal, S.; Mate, J.L.; Sancho, J.M.; Arenillas, L.; Serrano, S.; Escoda, L.; et al. Gene-expression profiling and not immunophenotypic algorithms predicts prognosis in patients with diffuse large B-cell lymphoma treated with immunochemotherapy. *Blood* **2011**, *117*, 4836–4843. [CrossRef] [PubMed]
12. Roschewski, M.; Dunleavy, K.; Pittaluga, S.; Moorhead, M.; Pepin, F.; Kong, K.; Shovlin, M.; Jaffe, E.S.; Staudt, L.M.; Lai, C.; et al. Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: A correlative biomarker study. *Lancet Oncol.* **2015**, *16*, 541–549. [CrossRef]
13. van Niel, G.; D'Angelo, G.; Raposo, G. Shedding light on the cell biology of extracellular vesicles. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 213–228. [CrossRef]
14. Carvalho, A.S.; Baeta, H.; Henriques, A.F.A.; Ejtehadifar, M.; Tranfield, E.M.; Sousa, A.L.; Farinho, A.; Silva, B.C.; Cabecadas, J.; Gameiro, P.; et al. Proteomic Landscape of Extracellular Vesicles for Diffuse Large B-Cell Lymphoma Subtyping. *Int. J. Mol. Sci.* **2021**, *22*, 1004. [CrossRef] [PubMed]
15. Pietrowska, M.; Zebrowska, A.; Gawin, M.; Marczak, L.; Sharma, P.; Mondal, S.; Mika, J.; Polanska, J.; Ferrone, S.; Kirkwood, J.M.; et al. Proteomic profile of melanoma cell-derived small extracellular vesicles in patients' plasma: A potential correlate of melanoma progression. *J. Extracell. Vesicles* **2021**, *10*, e12063. [CrossRef] [PubMed]
16. Hoshino, A.; Kim, H.S.; Bojmar, L.; Gyan, K.E.; Cioffi, M.; Hernandez, J.; Zambirinis, C.P.; Rodrigues, G.; Molina, H.; Heissel, S.; et al. Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers. *Cell* **2020**, *182*, 1044–1061.e1018. [CrossRef] [PubMed]
17. Hans, C.P.; Weisenburger, D.D.; Greiner, T.C.; Gascoyne, R.D.; Delabie, J.; Ott, G.; Muller-Hermelink, H.K.; Campo, E.; Braziel, R.M.; Jaffe, E.S.; et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* **2004**, *103*, 275–282. [CrossRef] [PubMed]
18. Elzanowska, J.; Berrocal, L.; Garcia-Pelaez, B.; Vives-Usano, M.; Sebo, B.P.; Maia, J.; Batista, S.; Teppo, J.; Varjosalo, M.; Moraes, M.C.S.; et al. Defining Optimal Conditions for Tumor Extracellular Vesicle DNA Extraction for Mutation Profiling. *Cancers* **2022**, *14*, 3258. [CrossRef]
19. Carvalho, A.S.; Moraes, M.C.S.; Hyun Na, C.; Fierro-Monti, I.; Henriques, A.; Zahedi, S.; Bodo, C.; Tranfield, E.M.; Sousa, A.L.; Farinho, A.; et al. Is the Proteome of Bronchoalveolar Lavage Extracellular Vesicles a Marker of Advanced Lung Cancer? *Cancers* **2020**, *12*, 3450. [CrossRef]
20. Mathivanan, S.; Simpson, R.J. ExoCarta: A compendium of exosomal proteins and RNA. *Proteomics* **2009**, *9*, 4997–5000. [CrossRef]
21. Hurwitz, S.N.; Rider, M.A.; Bundy, J.L.; Liu, X.; Singh, R.K.; Meckes, D.G., Jr. Proteomic profiling of NCI-60 extracellular vesicles uncovers common protein cargo and cancer type-specific biomarkers. *Oncotarget* **2016**, *7*, 86999–87015. [CrossRef] [PubMed]
22. Palviainen, M.; Saraswat, M.; Varga, Z.; Kitka, D.; Neuvonen, M.; Puhka, M.; Joenvaara, S.; Renkonen, R.; Nieuwland, R.; Takatalo, M.; et al. Extracellular vesicles from human plasma and serum are carriers of extravesicular cargo-Implications for biomarker discovery. *PLoS ONE* **2020**, *15*, e0236439. [CrossRef] [PubMed]
23. Geyer, P.E.; Voytik, E.; Treit, P.V.; Doll, S.; Kleinhempel, A.; Niu, L.; Muller, J.B.; Buchholtz, M.L.; Bader, J.M.; Teupser, D.; et al. Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol. Med.* **2019**, *11*, e10427. [CrossRef] [PubMed]
24. Klein, J.P.; van Houwelingen, H.C.; Ibrahim, J.G.; Scheike, T.H. *Handbook of Survival Analysis*; Chapman & Hall: London, UK, 2016.
25. Osti, D.; Del Bene, M.; Rappa, G.; Santos, M.; Matafora, V.; Richichi, C.; Faletti, S.; Beznoussenko, G.V.; Mironov, A.; Bachi, A.; et al. Clinical Significance of Extracellular Vesicles in Plasma from Glioblastoma Patients. *Clin. Cancer Res.* **2019**, *25*, 266–276. [CrossRef] [PubMed]
26. Kalra, H.; Simpson, R.J.; Ji, H.; Aikawa, E.; Altevogt, P.; Askenase, P.; Bond, V.C.; Borras, F.E.; Breakefield, X.; Budnik, V.; et al. Vesiclepedia: A compendium for extracellular vesicles with continuous community annotation. *PLoS Biol.* **2012**, *10*, e1001450. [CrossRef] [PubMed]

27. Kim, D.K.; Lee, J.; Kim, S.R.; Choi, D.S.; Yoon, Y.J.; Kim, J.H.; Go, G.; Nhung, D.; Hong, K.; Jang, S.C.; et al. EVpedia: A community web portal for extracellular vesicles research. *Bioinformatics* **2015**, *31*, 933–939. [CrossRef]
28. Wei, P.; Wu, F.; Kang, B.; Sun, X.; Heskia, F.; Pachot, A.; Liang, J.; Li, D. Plasma extracellular vesicles detected by Single Molecule array technology as a liquid biopsy for colorectal cancer. *J. Extracell. Vesicles* **2020**, *9*, 1809765. [CrossRef]
29. Arraud, N.; Linares, R.; Tan, S.; Gounou, C.; Pasquet, J.M.; Mornet, S.; Brisson, A.R. Extracellular vesicles from blood plasma: Determination of their morphology, size, phenotype and concentration. *J. Thromb. Haemost.* **2014**, *12*, 614–627. [CrossRef]
30. van Es, N.; Hisada, Y.; Di Nisio, M.; Cesarman, G.; Kleinjan, A.; Mahe, I.; Otten, H.M.; Kamphuisen, P.W.; Berckmans, R.J.; Buller, H.R.; et al. Extracellular vesicles exposing tissue factor for the prediction of venous thromboembolism in patients with cancer: A prospective cohort study. *Thromb. Res.* **2018**, *166*, 54–59. [CrossRef]
31. Lindenbergh, M.F.S.; Stoorvogel, W. Antigen Presentation by Extracellular Vesicles from Professional Antigen-Presenting Cells. *Annu. Rev. Immunol.* **2018**, *36*, 435–459. [CrossRef]
32. Carvalho, A.S.; Cuco, C.M.; Lavareda, C.; Miguel, F.; Ventura, M.; Almeida, S.; Pinto, P.; de Abreu, T.T.; Rodrigues, L.V.; Seixas, S.; et al. Bronchoalveolar Lavage Proteomics in Patients with Suspected Lung Cancer. *Sci. Rep.* **2017**, *7*, 42190. [CrossRef] [PubMed]
33. Xu-Monette, Z.Y.; Li, J.; Xia, Y.; Crossley, B.; Bremel, R.D.; Miao, Y.; Xiao, M.; Snyder, T.; Manyam, G.C.; Tan, X.; et al. Immunoglobulin somatic hypermutation has clinical impact in DLBCL and potential implications for immune checkpoint blockade and neoantigen-based immunotherapies. *J. Immunother. Cancer* **2019**, *7*, 272. [CrossRef] [PubMed]
34. Young, R.M.; Wu, T.; Schmitz, R.; Dawood, M.; Xiao, W.; Phelan, J.D.; Xu, W.; Menard, L.; Meffre, E.; Chan, W.C.; et al. Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 13447–13454. [CrossRef] [PubMed]
35. Muppidi, J.R.; Schmitz, R.; Green, J.A.; Xiao, W.; Larsen, A.B.; Braun, S.E.; An, J.; Xu, Y.; Rosenwald, A.; Ott, G.; et al. Loss of signalling via Galpha13 in germinal centre B-cell-derived lymphoma. *Nature* **2014**, *516*, 254–258. [CrossRef] [PubMed]
36. Brinkmann, V.; Reichard, U.; Goosmann, C.; Fauler, B.; Uhlemann, Y.; Weiss, D.S.; Weinrauch, Y.; Zychlinsky, A. Neutrophil extracellular traps kill bacteria. *Science* **2004**, *303*, 1532–1535. [CrossRef]
37. Tessier-Cloutier, B.; Twa, D.D.; Baecklund, E.; Gascoyne, R.; Johnson, N.A.; Backlin, C.; Kamen, D.L.; Clarke, A.E.; Ramsey-Goldman, R.; Lee, J.L.; et al. Cell of origin in diffuse large B-cell lymphoma in systemic lupus erythematosus: Molecular and clinical factors associated with survival. *Lupus Sci. Med.* **2019**, *6*, e000324. [CrossRef]
38. Giacomini, A.; Ghedini, G.C.; Presta, M.; Ronca, R. Long pentraxin 3: A novel multifaceted player in cancer. *Biochim. Biophys. Acta Rev. Cancer* **2018**, *1869*, 53–63. [CrossRef]
39. Basile, A.; Moschetta, M.; Ditonno, P.; Ria, R.; Marech, I.; De Luisi, A.; Berardi, S.; Frassanito, M.A.; Angelucci, E.; Derudas, D.; et al. Pentraxin 3 (PTX3) inhibits plasma cell/stromal cell cross-talk in the bone marrow of multiple myeloma patients. *J. Pathol.* **2013**, *229*, 87–98. [CrossRef]
40. Bonavita, E.; Gentile, S.; Rubino, M.; Maina, V.; Papait, R.; Kunderfranco, P.; Greco, C.; Feruglio, F.; Molgora, M.; Laface, I.; et al. PTX3 is an extrinsic oncosuppressor regulating complement-dependent inflammation in cancer. *Cell* **2015**, *160*, 700–714. [CrossRef]
41. Nissen, M.D.; Kusakabe, M.; Wang, X.; Simkin, G.; Gracias, D.; Tyshchenko, K.; Hill, A.; Meskas, J.; Hung, S.; Chavez, E.A.; et al. Single Cell Phenotypic Profiling of 27 DLBCL Cases Reveals Marked Intertumoral and Intratumoral Heterogeneity. *Cytom. A* **2020**, *97*, 620–629. [CrossRef]
42. Read, J.A.; Koff, J.L.; Nastoupil, L.J.; Williams, J.N.; Cohen, J.B.; Flowers, C.R. Evaluating cell-of-origin subtype methods for predicting diffuse large B-cell lymphoma survival: A meta-analysis of gene expression profiling and immunohistochemistry algorithms. *Clin. Lymphoma Myeloma Leuk.* **2014**, *14*, 460–467.e462. [CrossRef] [PubMed]
43. Xia, Z.N.; Wang, X.Y.; Cai, L.C.; Jian, W.G.; Zhang, C. IGLL5 is correlated with tumor-infiltrating immune cells in clear cell renal cell carcinoma. *FEBS Open Bio* **2021**, *11*, 898–910. [CrossRef] [PubMed]
44. Liu, J.; Mi, J.; Liu, S.; Chen, H.; Jiang, L. PSMB5 overexpression is correlated with tumor proliferation and poor prognosis in hepatocellular carcinoma. *FEBS Open Bio* **2022**. [CrossRef]
45. Cheng, X.; Wang, X.; Wu, Z.; Tan, S.; Zhu, T.; Ding, K. CORO1C expression is associated with poor survival rates in gastric cancer and promotes metastasis in vitro. *FEBS Open Bio* **2019**, *9*, 1097–1108. [CrossRef]
46. Wang, X.; Qian, H.; Zhang, S. Discovery of significant pathways in breast cancer metastasis via module extraction and comparison. *IET Syst. Biol.* **2014**, *8*, 47–55. [CrossRef] [PubMed]
47. Lin, Z.; Chen, X.; Li, Z.; Zhou, Y.; Fang, Z.; Luo, Y.; Zhao, J.; Xu, B. The role of bortezomib in newly diagnosed diffuse large B cell lymphoma: A meta-analysis. *Ann. Hematol.* **2018**, *97*, 2137–2144. [CrossRef]
48. Rodgers, T.D.; Baran, A.; Reagan, P.M.; Casulo, C.; Zent, C.; Evans, A.; Burack, R.; Williams, A.M.; Friedberg, J.W.; Barr, P.M. Efficacy of lenalidomide in high-risk diffuse large B-cell lymphoma. *Br. J. Haematol.* **2020**, *188*, e33–e36. [CrossRef]
49. Yee, C.S.; Massaad, M.J.; Bainter, W.; Ohsumi, T.K.; Foger, N.; Chan, A.C.; Akarsu, N.A.; Aytekin, C.; Ayvaz, D.C.; Tezcan, I.; et al. Recurrent viral infections associated with a homozygous CORO1A mutation that disrupts oligomerization and cytoskeletal association. *J. Allergy Clin. Immunol.* **2016**, *137*, 879–888.e872. [CrossRef]
50. Carvalho, A.S.; Ribeiro, H.; Voabil, P.; Penque, D.; Jensen, O.N.; Molina, H.; Matthiesen, R. Global mass spectrometry and transcriptomics array based drug profiling provides novel insight into glucosamine induced endoplasmic reticulum stress. *Mol. Cell. Proteom.* **2014**, *13*, 3294–3307. [CrossRef]

51. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [CrossRef]
52. Matthiesen, R.; Prieto, G.; Amorim, A.; Aloria, K.; Fullaondo, A.; Carvalho, A.S.; Arizmendi, J.M. SIR: Deterministic protein inference from peptides assigned to MS data. *J. Proteom.* **2012**, *75*, 4176–4183. [CrossRef] [PubMed]
53. Oksanen, J. Vegan Community Ecology Package Version 2.6-2 April 2022. The Comprehensive R Archive Network. Available online: http://cran.r-project.org (accessed on 15 August 2022).
54. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **2004**, *3*, 3. [CrossRef] [PubMed]
55. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [CrossRef]
56. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.C.; Muller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [CrossRef] [PubMed]
57. Kuhn, M. Caret: Classification and Regression Training. R Package Version 6.0-93, 2022. The Comprehensive R Archive Network. Available online: http://cran.r-project.org (accessed on 15 August 2022).
58. Carvalho, A.S.; Molina, H.; Matthiesen, R. New insights into functional regulation in MS-based drug profiling. *Sci. Rep.* **2016**, *6*, 18826. [CrossRef]
59. Hackenberg, M.; Matthiesen, R. Annotation-Modules: A tool for finding significant combinations of multisource annotations for gene lists. *Bioinformatics* **2008**, *24*, 1386–1393. [CrossRef]
60. Therneau, T.M.; Grambsch, P.M. *Modeling Survival Data: Extending the Cox Model*; Springer: New York, NY, USA, 2000.
61. Kosinski, M.; Biecek, P. RTCGA: The Cancer Genome Atlas Data Integration. R Package Version 1.26.0, 2022. Available online: https://rtcga.github.io/RTCGA (accessed on 15 August 2022).
62. Deutsch, E.W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J.J.; Kundu, D.J.; Garcia-Seisdedos, D.; Jarnuczak, A.F.; Hewapathirana, S.; Pullman, B.S.; et al. The ProteomeXchange consortium in 2020: Enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* **2020**, *48*, D1145–D1152. [CrossRef]