*Article*

# Systems Drug Discovery for Diffuse Large B Cell Lymphoma Based on Pathogenic Molecular Mechanism via Big Data Mining and Deep Learning Method

Shan-Ju Yeh [ID], Tsun-Yung Yeh and Bor-Sen Chen *

Laboratory of Automatic Control, Signal Processing and Systems Biology, Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan; m793281@gmail.com (S.-J.Y.); qpctotca@gmail.com (T.-Y.Y.)
* Correspondence: bschen@ee.nthu.edu.tw

**Abstract:** Diffuse large B cell lymphoma (DLBCL) is an aggressive heterogeneous disease. The most common subtypes of DLBCL include germinal center b-cell (GCB) type and activated b-cell (ABC) type. To learn more about the pathogenesis of two DLBCL subtypes (i.e., DLBCL ABC and DLBCL GCB), we firstly construct a candidate genome-wide genetic and epigenetic network (GWGEN) by big database mining. With the help of two DLBCL subtypes' genome-wide microarray data, we identify their real GWGENs via system identification and model order selection approaches. Afterword, the core GWGENs of two DLBCL subtypes could be extracted from real GWGENs by principal network projection (PNP) method. By comparing core signaling pathways and investigating pathogenic mechanisms, we are able to identify pathogenic biomarkers as drug targets for DLBCL ABC and DLBCL GCD, respectively. Furthermore, we do drug discovery considering drug-target interaction ability, drug regulation ability, and drug toxicity. Among them, a deep neural network (DNN)-based drug-target interaction (DTI) model is trained in advance to predict potential drug candidates holding higher probability to interact with identified biomarkers. Consequently, two drug combinations are proposed to alleviate DLBCL ABC and DLBCL GCB, respectively.

**Keywords:** diffuse large B cell lymphoma (DLBCL); deep neural network; drug discovery; drug combination

## 1. Introduction

Non-Hodgkin lymphoma (NHL), a lymphoid tissue malignancy, is one of the most prevalent cancers worldwide [1]. Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of NHL in western countries [2]. Meanwhile, it is a biologically heterogeneous and aggressive disease. The survival rate is usually less than one year for patients without treatment. Along with the thriving of DNA array technology, gene expression profiling studies have confirmed the existence of DLBCL subtypes involving in germinal center B cells (GCB) DLBCL and activated B cells (ABC) DLBCL. It represents lymphomas caused at different stages of lymphatic differentiation. Moreover, DLBCL GCB is a lymphocyte from the germinal center, therefore, it expresses some genes often observed in germinal center B cells including BCL6 and CD10 [3]. The main pathological feature of DLBCL ABC is the NFκB signaling pathway resulting in significant impacts on the cell proliferation and the regulation of apoptosis. It is noted that there is a large difference between DLBCL GCB and DLBCL ABC in terms of the clinical survival rate. The five-year survival rate of DLBCL GCB is about 60%, while the five-year survival rate of DLBCL ABC is about 35%. The pathogenesis of DLBCL in two subtypes is currently unknown.

The current standard therapy for DLBCL is R-CHOP, including five drugs, rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone. Among them, rituximab acts with CD20 to drive caspase-independent cell apoptosis death [4]. However, rituximab-induced hypogammaglobulinemia occurred [5]. For cyclophosphamide, it could target the

gene CD95 and trigger activation-induced cell death after activation [6]. One study indicates that patients treated with cyclophosphamide have a 4.5-fold increased risk of bladder cancer [7]. Doxorubicin, an anthracycline drug, has been implicated in cardiotoxicity. Its main mechanisms have something to do with DNA damage, membrane damage, oxidative stress, and the apoptosis pathways [8]. Targeting the p53 gene to participate in cell cycle arrest, DNA repair, or apoptosis [9], vincristine is widely used to treat malignant tumors; however, vocal cord paralysis caused by neurotoxicity has been found [10]. Prednisone, a glucocorticoid drug, inhibits NFκB and other inflammatory transcription factors, while the long-term steroid therapy may induce osteoporosis and liver cancer [11,12]. Instead of using rituximab, novel anti-CD20 agents (i.e., obinutuzumab and ofatumumab) were suggested for B-chronic lymphocytic leukemia and follicular lymphoma as well [13]. In addition, several innovative treatments for DLBCL have been approved by the U.S. Food and Drug Administration (FDA) including the anti-CD79b antibody drug conjugate polatuzumab vedotin (Pola) with bendamustine and rituximab (Pola-BR) [14]; the oral nuclear transport (XPO1) inhibitor selinexor [15]; and the combination of the anti-CD19 monoclonal antibody tafasitamab with the immunomodulatory agent lenalidomide [16,17]. Considering the different side effects of current treatments, drug combinations with multi-targets therapies toward DLBCL are worth studying.

It usually takes more than 12 years to develop a novel drug. The average cost of the drug development is about USD 2.6 billion [18]. There are few drugs that start from actual human testing that ever make it to marketing [19]. Due to huge demand for new anticancer drugs and various combinations of cell-target based screenings [20], drug repositioning based on computational methods has become popular in drug discovery. Drug-target interaction (DTI) prediction facilitates the process of drug discovery. It is the exploration of new drugs that interact with a particular target. The computational methods for DTI can be broadly classified into ligand-based approaches, docking approaches, and chemogenomic approaches [21]. The concept of ligand-based approaches is to predict the interactions based on the similarities between the protein ligands. However, without using sequencing information, it is hard to discover possible novel interactions due to the limitation of known ligands and protein families [22]. Utilizing 3D structures of proteins as well as drugs, docking approaches are based on the simulations to predict DTI [23–25], while these tasks would be challenged for certain membrane proteins, the 3D structures of which are unavailable. For chemogenomic approaches, it combines the chemical space of drugs and the genomics space of proteins into feature vectors to overcome the drawbacks of ligand-based and docking approaches. Chemogenomic approaches is suitable for machine learning (ML) methods for prediction of DTI [26]. In ML methods, the knowledge about drugs and proteins are represented by feature vectors that are used to train models for predicting the interactions between new drugs and/or new targets [27]. Furthermore, different learning-based models have been developed for DTI predictions, such as deep belief neural networks [28,29], convolutional neural networks [30,31], multilayer perceptrons [32,33], and graph neural network [34–36]. From the viewpoint of application, taking advantage of a chemogenomic approach, we trained a deep neural network (DNN)-based DTI prediction framework in advance for obtaining potential drug candidates toward the identified biomarkers.

In this study, we propose systems biology methods including systems modeling, system identification, system order detection scheme, and a principal network projection method to identify essential biomarkers as drug targets based on investigating pathogenic molecular mechanisms. Afterward, for identified biomarkers, we follow system drug design procedure taking drug design specifications into account, such as drug-target interaction ability, drug regulation ability, and drug toxicity to suggest potential drug combinations for DLBCL GCB and DLBCL ABC, respectively. The corresponding systems drug discovery flowchart is shown in Figure 1. It is noted that we build a DNN-based DTI model in advance for helping us obtain drug candidates, which have higher interaction probability toward the identified biomarkers (drug targets). Consequently, both famotidine and chlorzoxazone are regarded as common molecular drugs, which contribute to

inhibiting tumor metastasis, migration, and invasion for DLBCL ABC and DLBCL GCB. Furthermore, etoposide is designed specifically for cancer cell DNA damage of DLBCL ABC, and methotrexate is designed specifically for abnormal cell cycle of DLBCL GCB.
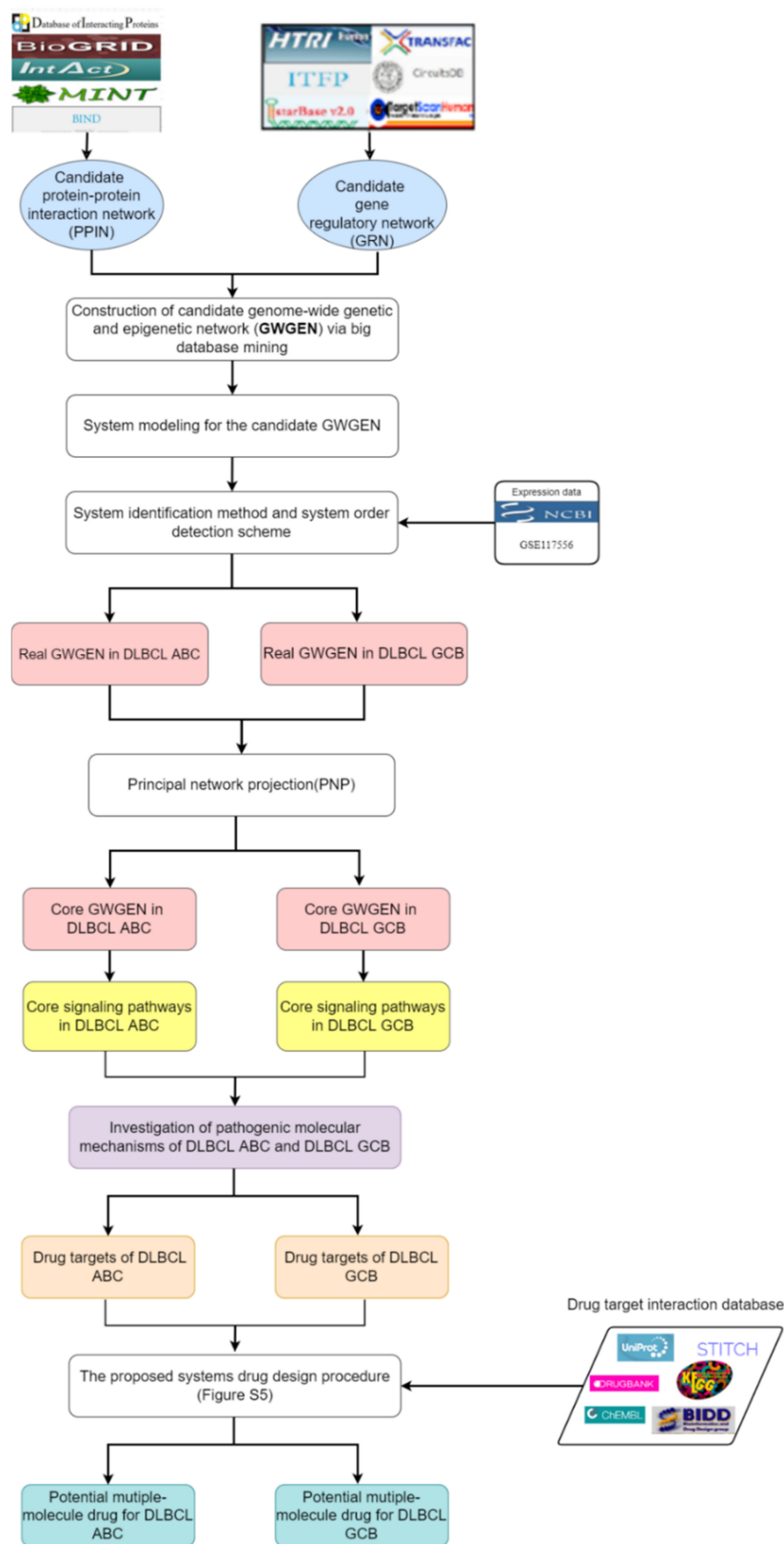


**Figure 1.** Flowchart of systems drug discovery based on systems biology approaches and drug design specifications.

## 2. Results

### 2.1. The Pathogenic Molecular Mechanisms in DLBCL ABC

From the core signaling pathways of DLBCL ABC in Figure 2, macrophage migration inhibitory factor (MIF) is found to be an important regulator of the innate immune system. MIF is classified as a pro-inflammatory cytokine [37]. MIF binds to CD74 on other immune cells to trigger an acute immune response [38]. Receptor CD74 (HLA class II histocompatibility antigen γ-chain receptor) in DLBCL ABC receives microenvironment factor MIF (macrophage migration inhibitory factor) to regulate TF STAT3 and MYC [39], respectively. The signaling transduction protein SRC, which was affected by phosphorylation, could transmit signals from CD74 to TF STAT3 in DLBCL ABC. Moreover, SRC was phosphorylated at the specific tyrosine residue by other tyrosine kinases, playing an important role in regulating embryonic development and cell growth [40]. Another signaling transduction protein, SORBS3, encodes a SH3 domain-containing adaptor protein. The presence of the SH3 domain is responsible for making the protein bind other cytoplasmic molecules, which are helpful for cytoskeletal organization, cell migration, gene expression and signaling. The constitutive activation of STAT3 signal promotes the growth, survival, angiogenesis and metastasis of tumor cells [41]; the overexpression of abnormally acetylated (activated) TF STAT3 can upregulate its target gene *HIF1A* [42], thereby promoting cellular functions, including cell proliferation as well as autophagy and inhibiting apoptosis [43]. At the same time, TF STAT3 would upregulate the target gene *ID2* [44] resulting in the promotion of cell cycle and epithelial-mesenchymal transition (EMT). Besides, it would upregulate the target gene *BCL2*, triggering the inhibition of autophagy and apoptosis. Upregulated by the acetylated STAT3, TF JDP2 is related to the inhibition of cell differentiation, cell cycle, and apoptosis. After being modified by the phosphorylation, the activated TF JDP2 would upregulate the DNA-methylated target gene *IL6*, which leads to promoting cell apoptosis and immune response against cancer [45].

Increased expression of SRC would trigger another core signaling pathway transmitting signals to TF FOXL1 through signaling transduction protein HIST1H2BA. FOXL1 plays an important role in regulating the expression of genes involved in cell metabolism, proliferation and differentiation. The overexpression of FOXL1 can upregulate miRNA MIR15A. The overexpression of MIR15A awakened by the upstream signals would inhibit the target genes *CCND1* and *ACTB* to promote their respective cellular functions. However, the total expression of *CCND1*, which is also activated by another TF NFκB1 and miRNA MIR497, is upregulated. Moreover, the target gene *CCND1* can promote the cell cycle progression and the target gene *ACTB* can promote the cell apoptosis and metastasis.

In the next core signaling pathway, after the ligand MIF combining with the receptor CD64, the signals are transmitted to TF MYC via the signaling transduction proteins BIN2, ATL2, and AR in DLBCL ABC. It is known that BIN2 related pathways are immune system [46]. Moreover, it can facilitate cell movement and migration through podosomes that interact with cell membrane and mediate cytoskeleton. Among this core signaling pathway, AR, an androgen receptor, is a DNA-binding transcription factor that regulates gene expression of *BCL2* [47]. It can regulate gene expression in eukaryotes and affect cell proliferation and differentiation. MYC is a proto-oncogene, which plays an important role in cell cycle progression, apoptosis and metastasis [48]. Overexpressed TF MYC will promote the upregulation of the target gene *BCL2*, further inhibiting cell autophagy and apoptosis, and promoting immune response.
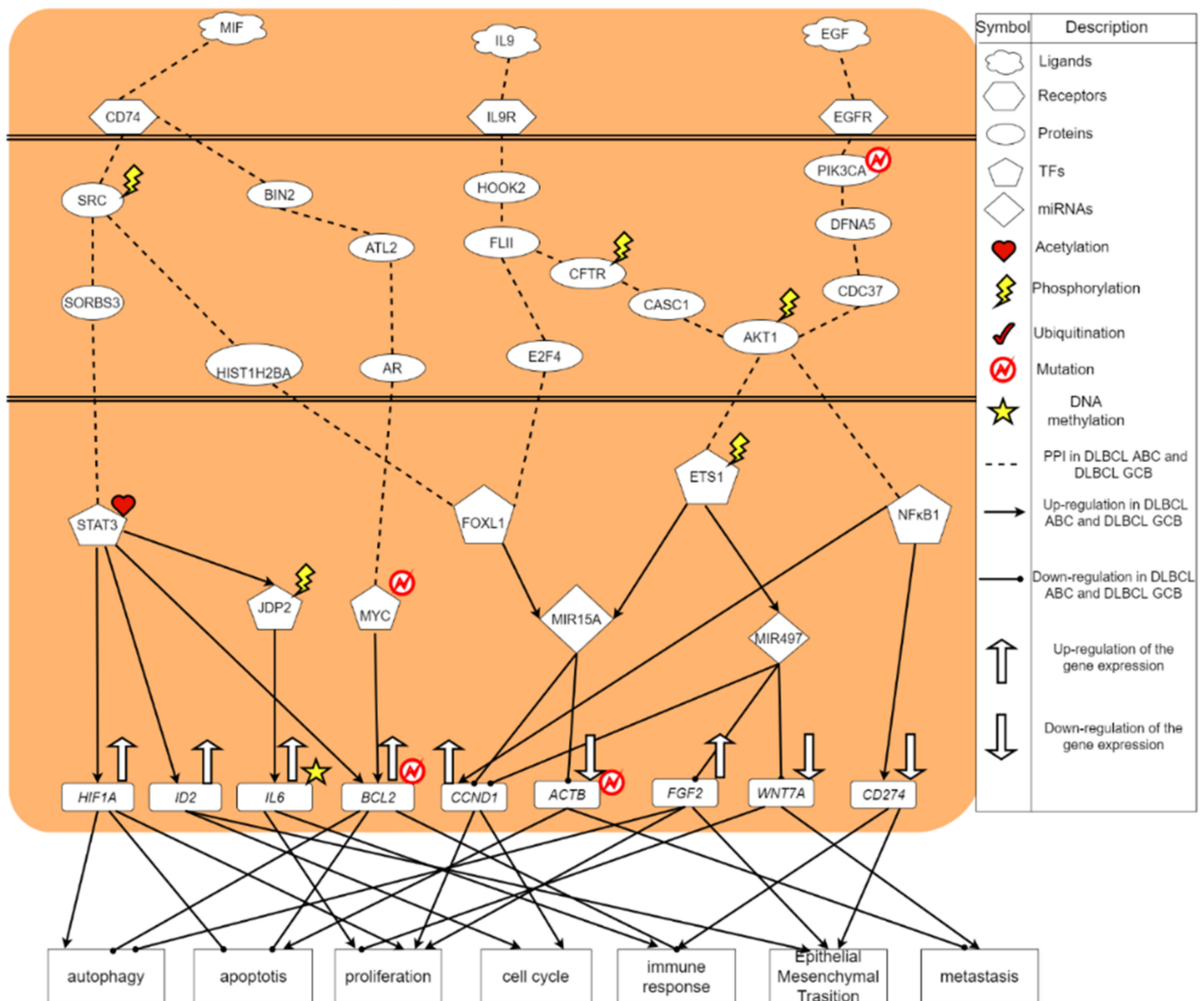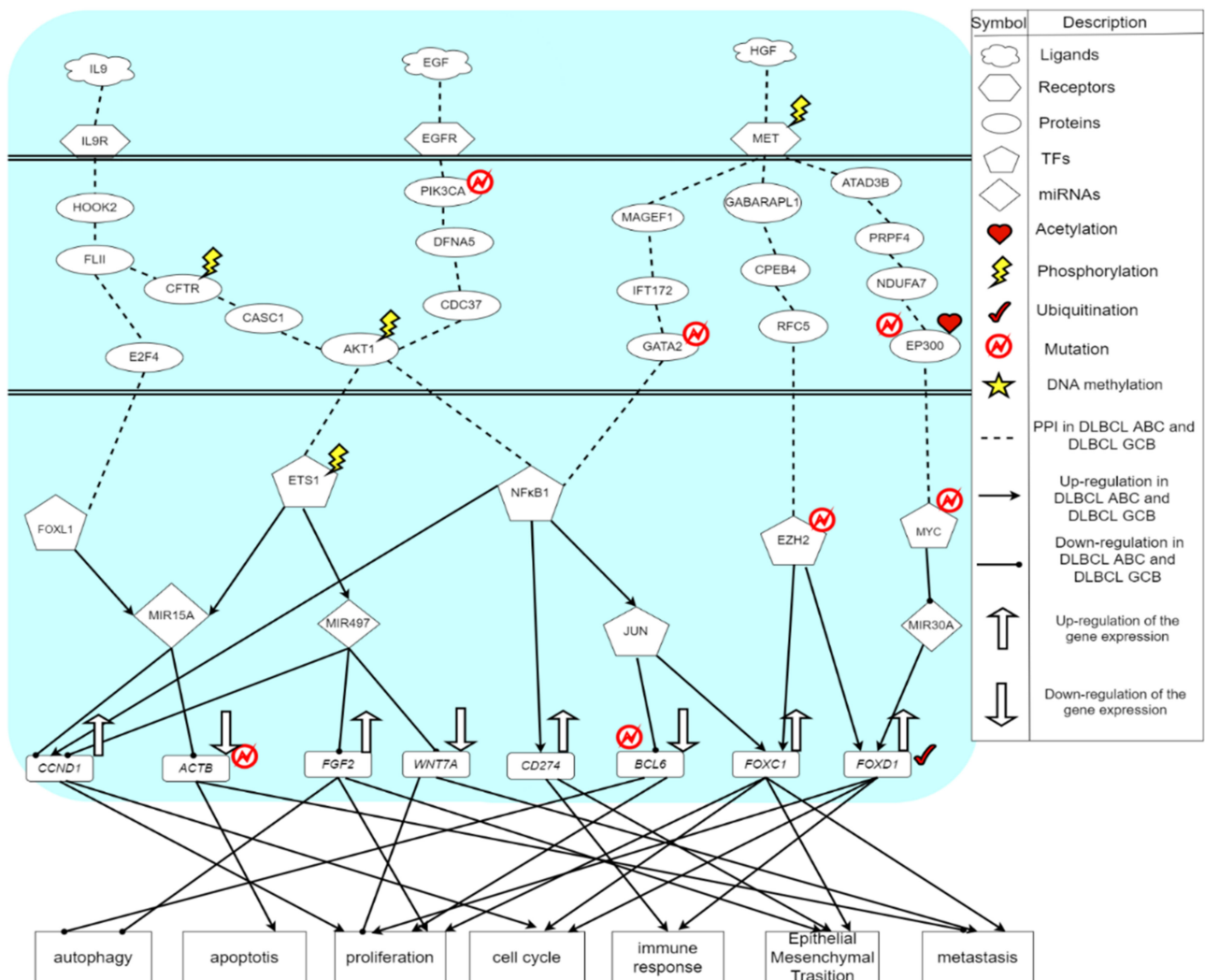
**Figure 2.** The core signaling pathways of DLBCL ABC. The black dotted line indicates protein–protein interactions in DLBCL ABC; the black arrow head of solid lines means activating cellular functions; the black circle head of solid lines means inhibiting cellular functions; the up arrow on the target gene indicates a high expression. The down arrow on the target gene indicates a low expression.

*2.2. The Carcinogenic Molecular Mechanism in DLBCL GCB*

The core signaling pathways of DLBCL GCB are shown in Figure 3. The microenvironment factor is a hepatocyte growth factor (HGF). HGF is secreted by mesenchymal cells and acts as a multifunctional cytokine on cells of primary epithelial origin [49]. Its ability to stimulate mitosis, cell movement and cytoplasmic matrix invasion makes it angiogenic, and plays a significant role in tumorigenesis and tissue regeneration [50]. The tyrosine kinase receptor MET receives the microenvironment factor HGF to regulate TF NFκB1, EZH2 and MYC, respectively. MET is an essential tyrosine kinase receptor for embryonic development, organ growth and wound healing. Through the signaling transduction proteins MAGEF1, IFT172 and GATA2, the mutated GATA2 protein will transmit the signal from MET to TF NFκB1. Among them, MAGEF1 can promote the degradation of proteasome and weaken the activity of some DNA repair and metabolic enzymes. In order to form cilia, IFT is necessary for the movement of other signaling proteins in the cilia [51]. Therefore, IFT172 plays a role in many different signaling pathways. IFT is considered to be a mediator of Hedgehog signaling and is one of the most important pathways in embryoge-

nesis. Furthermore, GATA2 plays an important role in regulating the transcription of genes related to the development and proliferation of hematopoietic and endocrine cells [52]. The mutation of GATA2 is associated with a variety of genetic and immune diseases, including myelodysplastic syndrome and acute myeloid leukemia [53]. The overexpressed NFκB1 can upregulate TF JUN. An improper activation of NFκB is related to many inflammatory diseases, while continuing to inhibit NFκB can cause abnormal immune cell development or delayed cell growth. This signal transduction event can lead to many biological processes such as inflammation, immunity, differentiation, cell growth, triggering growth, tumorigenesis and apoptosis. TF JUN was found to play an important role in cell proliferation [53]. In DLBCL GCB, TF JUN will downregulate the target gene *BCL6* and upregulate the target gene *FOXC1*, thereby correspondingly resulting in cell proliferation, autophagy, cell cycle, epithelial-mesenchymal transition (EMT) and cell metastasis.



**Figure 3.** The core signaling pathways of DLBCL GCB: The dotted black line indicates protein–protein interactions of DLBCL GCB; the black arrow head of solid lines means activating cellular functions; the black circle head of solid lines means inhibiting cellular functions; the up arrow on the target gene indicates an up-regulation. The down arrow on the target gene indicates a down-regulation.

In addition, after receiving the signal from phosphorylated MET, the signaling transduction proteins GABARAPL1, CPEB4, and RFC5 transmit the signal to TF EZH2. GABARAPL1 is a protein related to autophagy [54]. CPEB4 is related to the cell cycle progression promoting the growth and proliferation of tumors [55]. RFC5 is involved in DNA replication and repair. TF EZH2 is responsible for healthy embryo development through the epigenetic

maintenance of genes, which take charge of regulating development and differentiation [56]. The mutation or overexpression of EZH2 is associated with a variety of cancers [57]. Blocking the activity of EZH2 may slow down tumor growth. It is known that EZH2 has become a target for inhibition as it was upregulated in a variety of cancers [58]. In Figure 2, the abnormally activated TF EZH2 can upregulate target gene *FOXC1*, further promoting cellular functions, including cell proliferation, cell cycle, epithelial-mesenchymal transition (EMT) and metastasis [59]. In addition, EZH2 will also upregulate *FOXD1*, causing cell proliferation, cell cycle and immune response.

In the next core signaling pathway, after HGF binding to MET, the signal will be transmitted to TF MYC via the signaling transduction proteins ATAD3B, PRPF4, NDUFA7, and EP300, where ATAD3B is a protein related to immunity, PRPF4 is involved in pre-mrna splicing and modification [60], and NDUFA7-related pathways include respiration electronic transportation [61]. The mutant protein EP300 affected by acetylation can promote the transmission of upstream signals to downstream regulators. EP300 plays an important role in regulating cell growth and division, and promotes cell maturation and differentiation. One study indicates that EP300 protein is crucial for normal development of multicellular organisms before and after birth [62]. The expression of EP300 in DLBCL is significantly reduced. The mutations in EP300 usually remove or inactivate the histone acetyltransferase (HAT) coding domain of any gene [63]. In addition, study has pointed out that TF MYC was usually expressed constitutively in cancer [64], which led to an increased expression of many genes, some of which were involved in cell proliferation, in turn leading to cancer formation. The abnormally activated TF MYC will silence miRNA MIR30A [65]. Moreover, the low expression of MIR30A can negatively regulate the target gene *FOXD1*, thereby promoting cell proliferation, cell cycle and immune response.

Finally, from the results shown in Figures 1 and 2, although the cancer cells of DLBCL GCB have a stronger proliferative ability than the cancer cells of DLBCL ABC, we find that the ability of anti-apoptosis in DLBCL GCB is worse than DLBCL ABC. As a result, the cancer cells of DLBCL GCB are not conducive to spreading to other cell tissues. In other words, with stronger anti-apoptosis and anti-immune ability, DLBCL ABC will possess excessive cancer cell proliferation, which enhance the effect of metastasis and EMT. Therefore, DLBCL ABC has a higher mortality rate than GCB subtype.

### 2.3. The Common and Specific Carcinogenic Molecular Mechanism between DLBCL ABC and DLBCL GCB

In Figure 4, we have investigated the common and specific core signaling pathways between DLBCL ABC and DLBCL GCB. The microenvironment factor IL9 is a cell growth factor that can stimulate cell proliferation and prevent apoptosis [66]. Interleukin 9 receptor (IL9R) accepts IL9, a pleiotropic cytokine, belonging to the group of interleukins, through the signaling transduction proteins HOOK2, FLII, E2F4 in DLBCL ABC and DLBCL GCB, to upregulate TF FOXL1. In this core signaling pathway, the transduction signaling protein FLII plays a role in regulating cytoskeletal rearrangement involved in cell division and cell metastasis; E2F4 plays an important role in controlling the cell cycle and inhibiting tumor proteins. TF FOXL1 would regulate the expression of genes involved in cell metabolism, proliferation and differentiation [67].

The upregulated TF FOXL1 will overexpress MIR15A while MIR15A negatively regulates target genes *CCND1* and *ACTB*, respectively [68]. There have been studies showing that ACTB will mutate in DLBCL [69]. As for *CCND1*, it will cause cell proliferation and cell cycle progression [70]. After DNA replication, the replication chromosomes are separated into two independent cells. Generally speaking, the cell cycle can be divided into interphase (I) and mitosis (M). The stages of mitosis include prophase, prometaphase, metaphase, anaphase and telophase. The interphase (phase I) can usually be divided into the early stage of DNA synthesis (G1), the period of DNA synthesis (S) and the late stage of DNA synthesis (G2) [71]. The entire cell cycle can be expressed as: G1 phase → S phase → G2 phase → M phase. In G1 phase, the G1 checkpoint mechanism will prepare to ensure DNA

synthesis. Once the cell cycle checkpoint (Start or Restriction Point) is passed, the cell cycle is initiated and the process is irreversible, in which CCND1 is the most important cell cycle checkpoint [72]. DNA replication occurs in the S phase. During G2, the cells will prepare for mitosis, but in some cases the cells will jump out of the cell cycle and enter the so-called G0 phase. In G0 phase, the cells will leave the cycle and stop dividing. In fact, many cells in the human body are usually in the G0 phase, for example, nerve cells will never divide. The downregulation of ACTB will inhibit cell apoptosis and promote cell metastasis. Therefore, this core signaling pathway leads to the proliferation, the anti-apoptosis and the promotion of metastasis to exacerbate cancer progression in DLBCL patients. In addition, the receptor IL9R also transmits signals to TF ETS1 through signaling transduction proteins HOOK2, FLII, CFTR, CASC1, AKT1, where CFTR (cystic fibrosis transmembrane conductance regulator) is a membrane protein and chloride channel in vertebrates, which transport negatively charged particles called chloride ions into or out of the cell [73].
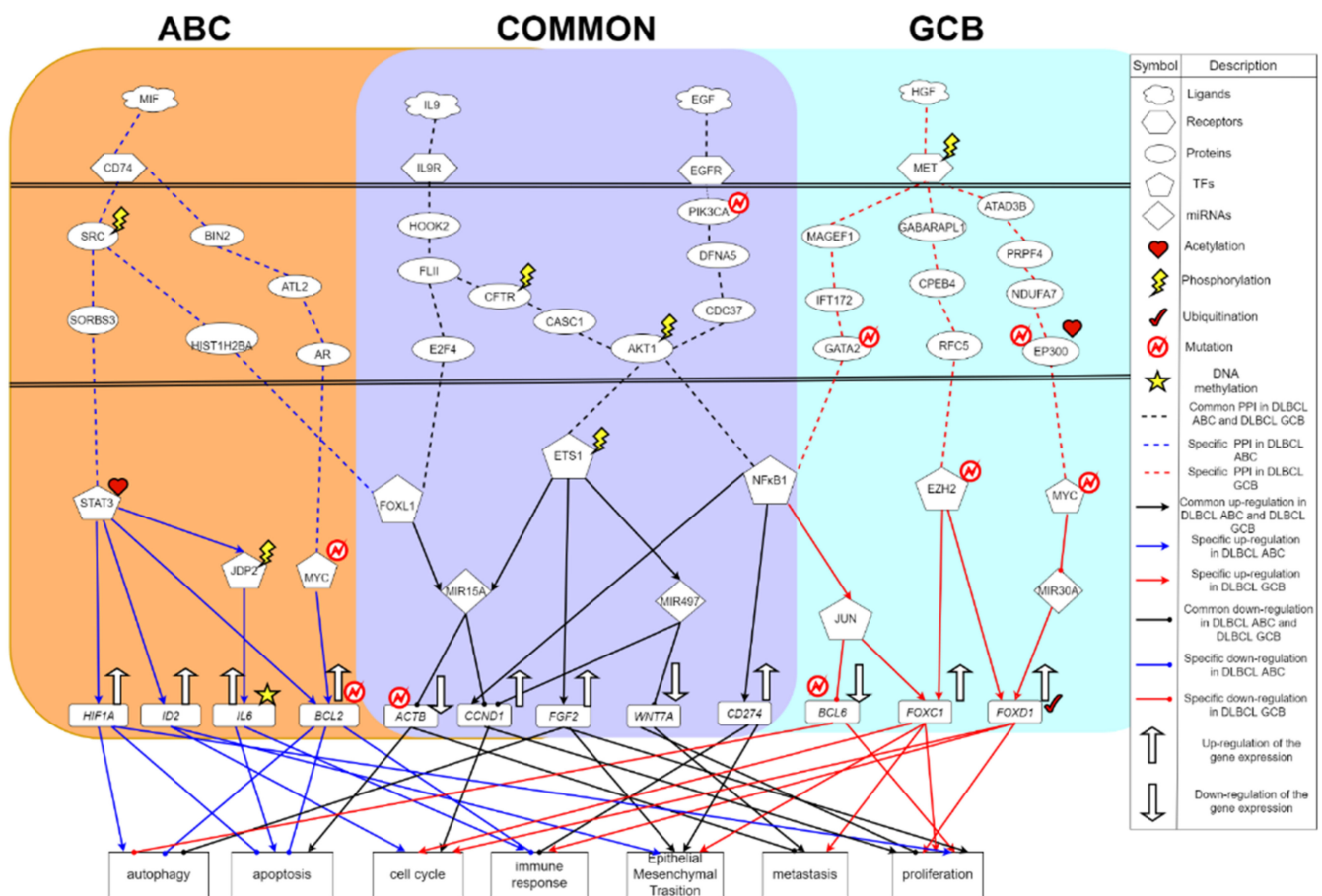


**Figure 4.** The common and specific core signaling pathways of DLBCL ABC and DLBCL GCB. This figure summarizes the genetic and epigenetic pathogenic molecular mechanisms of DLBCL ABC and DLBCL GCB. The core signaling pathways shown in the purple background are the common core signaling pathways of DLBCL ABC and DLBCL GCB. The blue line indicates specific core signaling pathways of DLBCL ABC; the red line indicates specific core signaling pathways of DLBCL GCB; the black line indicates common core signaling pathways of DLBCL ABC and DLBCL GCB; the arrow head of solid lines means activating cellular functions; the circle head of solid lines means inhibiting cellular functions. The up arrow on the target gene indicates an up-regulation. The down arrow on the target gene indicates a down-regulation.

For the next core signaling pathway in Figure 4, the epidermal growth factor receptor (EGFR) receives the microenvironmental factor EGF, and then the signal is transmitted

through the transduction proteins via the mutated PIK3CA, DFNA5, CDC37, and the phosphorylated AKT1 to TF ETS1, in which AKT1 is related to apoptosis [74]. It is found that AKT1 will phosphorylate AKT and inhibit apoptosis. ETS1 is a TF related to maintaining the proliferation of DLBCL and regulating the differentiation of germinal centers [75]. The overexpression of ETS1 will cause proliferation, survival and differentiation of lymphoma cells [76]. Modified by phosphorylation, the TF ETS1 will promote the expression of MIR15A and MIR497; meanwhile, ETS1 will upregulate the target gene FGF2. Furthermore, MIR497 will inhibit *CCND1* and *WNT7A*. It is known that *WNT7A* will promote cell proliferation and metastasis [77]. MIR15A will inhibit *CCND1* and *ACTB*. Although *CCND1* is inhibited by MIR15A and MIR497, *CCND1* is upregulated by another TF like NFκB1 as well. Hence, the total expression of *CCND1* in DLBCL is still upregulated, leading to subsequent cell proliferation and metastasis. In addition, *FGF2* can cause the inhibition of autophagy [78]. Autophagy is an orderly cell degradation and recycling process in all eukaryotes. There are generally three different forms of autophagy, including microautophagy, macroautophagy, and chaperone-mediated autophagy (CMA) [79]. One of their functions is to transport the cargo to the lysosome for degradation and recycling. *FGF2* will also promote cell proliferation and epithelial-mesenchymal transition (EMT). Therefore, this corresponding core signaling pathway will cause cell proliferation; simultaneously, autophagy and apoptosis will be inhibited, which further promote cancer cell metastasis and EMT in DLBCL. Moreover, the formation process of EMT will destroy the adhesion between normal cells improving the ability of migration and invasion for the cancer cells. It brings benefit to the cancer metastasis. This process not only accelerates the spread of cancer cells but also makes cancer cells spread intensely [79].

The microenvironment factor EGF (epidermal growth factor) plays an important role in regulating cell growth, proliferation and differentiation. After EGF binding to the receptor EGFR on the cell surface, the signals would be transmitted to TF NFκB1 through the signaling transduction proteins including the mutated PIK3CA, DFNA5, CDC37, and the phosphorylated AKT1. The role of PIK3CA is to promote the catalytic reaction of the message transmission [80]; and the mutation of PIK3CA will change the way of cells, regulating physiological responses to cause the formation of cancer. DFNA5 also has been found in other types of cancer such as stomach cancer, colorectal cancer, and breast cancer. Its characteristic is to induce apoptosis. Among this pathway, CDC37, a molecular chaperone protein, has a specific function in cell signaling transduction. It binds to a variety of kinases and regulates cyclin. Moreover, the TF NFκB1 will promote the upregulation of the target gene *CCND1*. Improper activation of NFκB is related to many inflammatory diseases, and continuous inhibition of NFκB will lead to poor development of immune cells or the delay of cell growth [81]. In summary, this signal transduction event is associated with many biological processes, such as inflammation, immunity, differentiation, and cell growth. It finally triggers cell growth, tumorigenesis and apoptosis. Besides, the upregulated *CCND1* further promotes cellular functions including cell proliferation and cell cycle progression. TF NFκB1 also upregulates the expression of target gene *CD274* (*PD-L1*), which in turn triggers cellular functions of EMT and immune responses [82].

### 2.4. Systems Drug Design Procedure Considering Drug-Target Interaction, Drug Regulation Ability, and Drug Toxicity

After investigating pathogenic molecular mechanisms, we identified two pools of essential biomarkers as drug targets for two subtypes of DLBCL shown in Table 1. The systems drug design procedure is in Figure S5. Firstly, we consider the drug-target interaction ability toward the identified biomarkers in terms of the application of DNN-based DTI model. Subsequently, filtered by drug regulation ability and drug toxicity, the number of predicted drug candidates would be narrowed down. For training DNN-based DTI model, there were 70% of the data as training set, including 10% of the data as validation set. The remaining 30% of the data were used as the testing set. To the architecture of DNN-based DTI model, it is a fully connected neural network consisting of one input

layer, four hidden layers and one output layer, of which four hidden layers have 512, 256, 128, 64 neurons, respectively. The dropout was added in each hidden layer for reducing overfitting. We used ReLU as the activation function for each hidden layer. In the output layer, we chose sigmoid to be the activation function for limiting the output value between zero and one. It is noted that the drugs with higher interaction probability (greater than 0.5) would be selected as drug candidates. Evaluating the robustness of hyperparameters including the number of nodes, dropout, and learning rate, we performed 10-fold cross validation. The corresponding 10-fold cross validation results could be found in Figure S6. The average accuracy of testing is 98.698% (standard deviation: 0.0659). Furthermore, we plot the receiver operating characteristic curve (ROC) in Figure S7. The area under ROC of the DNN-based DTI model is 0.99. Here, except for drug-target interaction, we regard drug regulation ability and drug toxicity as our drug design specifications as well. By referring to the connectivity map (CMap) [83], we could find the gene signatures after treating with more than 1300 compounds in numbers of cultivated cell lines. The goal here is to find the drugs owning the ability to reverse the abnormal gene expression. Meanwhile, according to the median lethal dose, which is looked up at DrugBank [84], we expect that the selected candidate small molecules could have less toxicity (Table S4). Consequently, we suggested famotidine, chlorzoxazone, and etoposide to be the potential multiple-molecule drug for alleviating DLBCL ABC (Table 2); famotidine, chlorzoxazone, and methotrexate as potential multiple-molecule drug for mitigating DLBCL GCB (Table 3).

**Table 1.** The biomarkers (drug targets) are identified for DLBCL ABC and DLBCL GCB.

| Cancer | Biomarkers (Drug Targets) |
|---|---|
| DLBCL ABC | FOXL1 NFκB1 AKT1 MYC STAT3 |
| DLBCL GCB | FOXL1 NFκB1 AKT1 MYC EZH2 |

**Table 2.** The multiple-molecule drug and the corresponding target proteins for DLBCL ABC.

| Targets \ Drugs | FOXL1 | NFκB1 | AKT1 | MYC | STAT3 |
|---|---|---|---|---|---|
| Famotidine | | O | O | | O |
| Chlorzoxazone | O | O | | | O |
| Etoposide | | O | | O | O |

O: The drug targets to its potential target proteins.

**Table 3.** The multiple-molecule drug and the corresponding target proteins for DLBCL GCB.

| Targets \ Drugs | FOXL1 | NFκB1 | AKT1 | MYC | EZH2 |
|---|---|---|---|---|---|
| Famotidine | | O | O | | O |
| Chlorzoxazone | O | O | | | O |
| Methotrexate | | | O | O | O |

O: The drug targets to its potential target proteins.

## 3. Discussion

Based on the core signaling pathways, we investigated the downstream carcinogenic pathogenesis and identified five significant biomarkers as drug targets for DLBCL ABC and DLBCL GCB, respectively (Table 1). Among these biomarkers, STAT3 and MYC can influence cancer cell survival and promote proliferation. The immune response of a human can be inhibited by NFκB1. Both AKT1 and EZH2 are associated with cancer metastasis and invasion, resulting in the deterioration of tumors. In contrast, FOXL1 can promote apoptosis and inhibit cancer cell metastasis. In order to reduce the ability of inhibiting apoptosis and promoting proliferation by HIF1A, to diminish the ability of promoting proliferation and

cell cycle by ID2, to decrease the ability of reducing apoptosis, and to reduce the ability of inhibiting autophagy and suppressing immunity by BCL2, STAT3 was selected as the biomarker to be inhibited. For enhancing the ability of promoting apoptosis and inhibiting metastasis caused by the target gene ACTB, FOXL1 was selected as a biomarker to be up-regulated. Moreover, to reduce the ability of inhibiting autophagy, promoting proliferation and EMT by FGF2, and to inhibit the ability of promoting cancer cell metastasis by WNT7A, AKT1 was selected as a biomarker to be inhibited. In addition, considering the significant impact of immune response on DLBCL ABC and GCB, NFκB1 was selected as the drug target to be inhibited, thereby reducing the ability of both suppressing immune response by CD274 and promoting cell cycle and proliferation by CCND1. For the purpose of reducing the ability of promoting proliferation, metastasis, and EMT by FOXC1 and reducing the proliferation caused by FOXD1, EZH2 was selected as a biomarker to be down-regulated.

Using immunotherapy against cancer gains a lot of attention in recent years. Here, we selected NFκB1 as a drug target to indirectly inhibit PD-L1 and block the related mechanisms having contribution to escape immunity. Among the proposed two multiple-molecule drugs, chlorzoxazone is a drug for treating muscle spasms [85]. It acts on the spinal cord by suppressing reflexes. One tumor related study has shown that it was used with other drugs to inhibit tumor growth, including tumor metastasis, migration, and invasion [86]. It is known that STAT3 and NFκB1 are significant activators of carcinogenic signal transduction. Chlorzoxazone can effectively reduce the expression of STAT3, NFκB1, and EZH2, and upregulate FOXL1, therefore, it might be an effective drug for DLBCL. In addition, famotidine could decrease the production of stomach acid. Its pharmacologic activity is used in the treatment of acid-related gastrointestinal conditions, including duodenal ulcer, esophageal adenocarcinoma and chronic gastroesophageal reflux disease in adults and children. Meanwhile, famotidine can inhibit the occurrence of cancer by inhibiting STAT3. Its drug targets include NFκB1, AKT1, and EZH2 as well. Moreover, etoposide has been studied to replace the DLBCL current standard treatment with R-CHOP, including rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone, among them, the disadvantage of doxorubicin is its cardiotoxicity [87]. Another study also mentioned that etoposide substituted could treat most of DLBCL patients who cannot receive anthracycline treatment [88]. Methotrexate, a chemotherapeutic drug and immunosuppressive agent, has been commonly applied in combination with other drugs for the treatment of breast cancer, leukemia, lung cancer, lymphoma, autoimmune disease, and ectopic pregnancy [89]. By down-regulating its target protein EZH2, methotrexate can treat cancers [89]. Note that dysregulation of EZH2 is closely related to oncogenesis of various tissue types. More and more evidences show that targeting EZH2 has great therapeutic potential in cancers [90]. Methotrexate can not only downregulate the expression of EZH2, but also interact with AKT1 and MYC. Hence, we suggest it as one of the small molecule drugs in our proposed drug combination.

Given the substantial costs and long development timeline of new drug discovery, the repurposing of old drugs to treat common and rare disease becomes an attractive proposition. In other words, drug repurposing is a strategy for identifying new uses for approved drugs that are outside the scope of known medical indication. In this study, by the proposed systems biology approaches and drug design specifications, we suggested multiple-molecule drugs (drug combinations) for DLBCL ABC and GCB, respectively. Those suggested small molecules are FDA approved. Although some studies have shown that etoposide and methotrexate were used in DLBCL, their combinations with famotidine and chlorzoxazone are still worth studying in the future in terms of synergistic and antagonistic effects. Leveraging computational biology methods, this study might provide new perspectives of understanding the pathogenic molecular mechanisms of DLBCL ABC and DLBCL GCB at a system level and give an alternative way to accelerate systems drug discovery for new therapeutics.

## 4. Materials and Methods

### 4.1. Overview of Systems Drug Discovery for DLBCL ABC and DLBCL GCB

The DLBCL microarray data is from the National Center for Biotechnology Information (NCBI) with accession number GSE117556. The corresponding platform is GPL14951. The dataset samples were divided into two subtypes, DLBCL ABC and DLBCL GCB, in which the ABC subtype has worse prognosis. There are 468 samples and 249 samples for DLBCL GCB and DLBCL ABC, respectively. The flowchart of systems drug discovery is in Figure 1. For identifying essential biomarkers as drug targets to alleviate DLBCL ABC and DLBCL GCB, we investigated the pathogenic molecular mechanisms based on the systems biology methods: (1) big database mining; (2) system modeling; (3) system identification and system order detection scheme; (4) principal network projection method.

Firstly, by big database mining, we constructed a candidate genome-wide genetic and epigenetic network (GWGEN), which is represented by a Boolean matrix (i.e., 0 or 1 if interaction is nonexistent or existent between two nodes). It is noted that both DLBCL ABC and DLBCL GCB shared the same candidate GWGEN. The candidate GWGEN consists of the candidate protein–protein interaction network (PPIN) and candidate gene regulatory network (GRN). For the candidate PPIN, we refer to the following database: DIP [91], IntAct [92], BioGRID [93], BIND [94], and MINT [95]. To the candidate GRN, we collect the pairs of transcription factors and target genes from ITFP [96] and HTRIdb [97]. Moreover, we look up databases including TargetScan [98], CircuitsDB [99], and StarBase2.0 [100] for the post-transcriptional regulations between miRNA, lncRNA and their target genes. After conducting system modeling for proteins, genes, miRNAs, and lncRNAs, we could evaluate system models' parameters by the system identification method with the help of microarray dataset in two subtypes of DLBCL. There might be false-positive interactions in the candidate GWGEN caused by various experimental conditions. Therefore, we performed a system order detection approach to prune these false-positive interactions for obtaining real GWGENs for DLBCL ABC and DLBCL GCB, respectively (Figures S1 and S2). The total number of nodes (i.e., transcription factors, receptors, proteins, miRNAs, lncRNAs) and their corresponding edges in the candidate GWGEN, real GWGEN of DLBCL ABC, and real GWGEN of DLBCL GCB are shown in Table S1. However, the real GWGENs were still too complicated to analyze. Applying principal network projection (PNP) method, from the real GWGENs, we could extract core GWGENs (Figures S3 and S4), which is comprised of the top 3000 nodes based on the descending order of projection value. The higher the projection value is, the more contribution provided by the node in the real GWGEN. Additionally, we did gene enrichment analyses by the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources version 6.8 based on the genes in core GWGENs (Tables S2 and S3). Projecting the corresponding core GWGENs in the annotation of Kyoto Encyclopedia of Genes and Genomes (KEGG), we could further investigate the common and specific pathogenic molecular mechanisms and identify essential biomarkers as drug targets.

For suggesting potential drug candidates toward these identified drug targets, we followed systems drug design procedure shown in Figure S5. The drug design specifications include drug-target interaction probability, drug regulation ability, and drug toxicity. To estimate drug-target interaction probability, we trained a DNN-based DTI model in advance. We regard the drugs having higher predicted probability as drug candidates. Subsequently, the number of those predicted drug candidates would be narrowed down by considering drug regulation ability and toxicity. Here, we aim to find drugs having the ability to reverse abnormal gene expression with low toxicity. More details will be discussed in the following sections.

### 4.2. Constructing the System Models in the GWGEN to Identify Real GWGEN of DLBCL GCB and DLBCL ABC

To investigate the molecular mechanisms of DLBCL GCB and DLBCL ABC, we constructed the interactive and regulatory models in the candidate GWGEN, including protein–

protein interactions, transcriptional regulations, miRNA regulations, and lncRNA regulations. For the candidate PPIN (PPIN), the *i*-th protein is described in the following equation:

$$p_i[n] = \sum_{\substack{k=1 \\ k \neq i}}^{Y_i} \alpha_{ik} p_i[n] p_k[n] + \phi_{i,PPIN} + \eta_{i,PPIN}[n] \text{ for } i = 1, \ldots, I, \text{ and } n = 1, \ldots N. \quad (1)$$

where $\alpha_{ik}$ is the interaction ability between the *i*-th protein and the *k*-th interactive protein; $p_i[n]$ denotes the expression level of the *i*-th protein for the sample *n*; $p_k[n]$ represents the expression level of the *k*-th protein for the *n* data sample; $Y_i$ indicates the total number of proteins interacting with the *i*-th protein and *I* denotes the total number of proteins in the candidate PPIN; *N* is the total number of data samples; $\phi_{i,PPIN}$ is the basal level of the *i*-th protein caused by some unknown interactions including phosphorylation and acetylation; $\eta_{i,PPIN}[n]$ represents the stochastic noise as a result of the modeling residue and measurement noise for the *n* data sample.

For the candidate gene regulatory network (GRN) in the candidate GWGEN, the systematic gene regulation model for the *q*-th gene of DLBCL cells to sample *n* can be governed by the following form:

$$g_q[n] = \sum_{\substack{j=1 \\ j \neq q}}^{J_q} A_{qj} z_j[n] + \sum_{w=1}^{W_q} B_{qw} x_w[n] - \sum_{h=1}^{H_q} C_{qh} d_h[n] g_q[n] + \phi_q + \eta_q[n]$$

$$\text{for } q = 1, \ldots, Q \text{ and } n = 1, \ldots N. \quad (2)$$

where $g_q[n]$ represents the expression level of the *q*-th gene; $J_q$ indicates the total number of TFs binding to the *q*-th gene; $W_q$ represents the total number of lncRNAs binding to the *q*-th gene; $H_q$ denotes the total number of miRNAs inhibiting the *q*-th gene; $A_{qj}$ denotes the transcription regulatory ability from the *j*-th TF to the *q*-th gene; $B_{qw}$ is the regulation ability from the *w*-th lncRNA to the *q*-th gene; $C_{qh} \geq 0$ represents the post-transcription regulatory ability, with which the *h*-th miRNA inhibits the *q*-th gene; $z_j[n]$, $x_w[n]$, and $d_h[n]$ indicate the expression of the *j*-th TF, the *w*-th lncRNA, and the *h*-th miRNA, respectively. *Q* is the total number of genes and *N* denotes the total number of data samples; $\phi_q$ represents the basal level of the *q*-th gene expression due to unknown regulations containing post-transcriptional modifications; $\eta_q[n]$ is the stochastic noise of the *q*-th gene for the data sample *n* caused by the model uncertainty and data noise. Furthermore, in the same way, the systematic models of the candidate lncRNA regulation network and the candidate miRNA regulation network can be referred to in the Supplementary Materials.

*4.3. Using the System Identification Method and System Order Detection Approach to Build Real GWGENs of DLBCL GCB and DLBCL ABC*

To estimate the unknown parameters for the PPI model in the candidate PPIN, we utilize a system identification method and system order detection approach on our systematic models with the help of genome-wide microarray data of patient samples. The PPI equation in Equation (1) could be rewritten as below:

$$p_i[n] = \begin{bmatrix} p_i[n]p_1[n] & \cdots & p_i[n]p_{Y_i}[n] & 1 \end{bmatrix} \times \begin{bmatrix} \alpha_{i1} \\ \vdots \\ \alpha_{iY_i} \\ \phi_i \end{bmatrix} + \eta_i[n] \quad (3)$$

$$= \xi_i[n] \cdot \varphi_{i,P} + \tau_i[n], \text{ for } i = 1, \ldots, I \text{ and } n = 1, \ldots, N.$$

where $\xi_i[n]$ determines the regression vector, which could be computed by the microar- ray data; $\varphi_{i,P}$ indicates the unknown parameter vector for the $i$-th protein. The Equation (3) of the $i$-th protein could be augmented for $N$ samples as below:

$$
\begin{bmatrix} p_i[1] \\ p_i[2] \\ \vdots \\ p_i[N] \end{bmatrix} = \begin{bmatrix} \xi_{i,P}[1] \\ \xi_{i,P}[2] \\ \vdots \\ \xi_{i,P}[N] \end{bmatrix} \cdot \varphi_{i,P} + \begin{bmatrix} \tau_i[1] \\ \tau_i[2] \\ \vdots \\ \tau_i[N] \end{bmatrix} \tag{4}
$$

Furthermore, the Equation (4) could be simplified represented as:

$$
P_i = \Xi_{i,P} \cdot \varphi_{i,P} + T_i \tag{5}
$$

Therefore, the unknown parameters in the vector $\varphi_{i,P}$ could be estimated by solving the least square estimation problem:

$$
\hat{\varphi}_{i,P} = \min_{\varphi_{i,P}} \frac{1}{2} \| \Xi_{i,P} \cdot \varphi_{i,P} - P_i \|_2^2 \tag{6}
$$

where $\hat{\varphi}_{q,G}$ is the estimated vector including the estimated interaction parameters for the $i$-th protein.

In the same way, the gene regulation model in Equation (2) could be rewritten as below:

$$
\begin{aligned}
g_q[n] &= \begin{bmatrix} z_1[n] & \cdots & z_{J_q}[n] & x_1[n] & \cdots & x_{W_q}[n] & g_q[n]d_1[n] & \cdots & g_q[n]d_{H_q}[n] & 1 \end{bmatrix} \times \begin{bmatrix} A_{q1} \\ \vdots \\ A_{qJ_q} \\ B_{q1} \\ \vdots \\ B_{qW_q} \\ -C_{q1} \\ \vdots \\ -C_{qH_q} \\ \phi_q \end{bmatrix} + \eta_q[n] \\
&= \xi_q[n] \cdot \varphi_{q,G} + \tau_q[n], \text{ for } q = 1, \ldots, Q \text{ and } n = 1, \ldots, N.
\end{aligned} \tag{7}
$$

where $\xi_q[n]$ indicates the regression vector, which could be obtained from the microarray data and $\varphi_{q,G}$ denotes the unknown parameters vector for the $q$-th gene. We could expand the Equation (7) for $N$ samples as shown below:

$$
\begin{bmatrix} g_q[1] \\ g_q[2] \\ \vdots \\ g_q[N] \end{bmatrix} = \begin{bmatrix} \xi_{q,G}[1] \\ \xi_{q,G}[2] \\ \vdots \\ \xi_{q,G}[N] \end{bmatrix} \cdot \varphi_{q,G} + \begin{bmatrix} \tau_q[1] \\ \tau_q[2] \\ \vdots \\ \tau_q[N] \end{bmatrix} \tag{8}
$$

Moreover, the Equation (8) could be simplified in the following form:

$$
G_q = \Xi_{q,G} \cdot \varphi_{q,G} + T_q \tag{9}
$$

Hence, by solving the following constrained linear least square estimation problem, we could have the estimated regulatory parameters in the vector $\varphi_{q,G}$.

$$\hat{\varphi}_{q,G} = \min_{\varphi_{q,G}} \frac{1}{2} \left\| \Xi_{q,G} \cdot \varphi_{q,G} - G_q \right\|_2^2$$

$$\text{subject to} \begin{bmatrix} 0 & \cdots & \cdots & 0 & 0 \cdots & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & & \vdots & \vdots \ddots & & \vdots & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots & \vdots \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & \cdots & 0 & 0 \cdots & \cdots & 0 & 0 \cdots & & 0 & 1 & 0 \end{bmatrix} \varphi_{q,G} \leq \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \qquad (10)$$

$$\underbrace{\qquad}_{J_q} \qquad \underbrace{\qquad}_{W_q} \qquad \underbrace{\qquad}_{H_q}$$

where $\hat{\varphi}_{q,G}$ is the estimated vector including estimated regulatory parameters in the Equation (2). Meanwhile, the miRNA repression parameters $C_{qh}$ are guaranteed to be positive (i.e., $C_{qh} \geq 0$) for $h = 1, \ldots, H_q$.

It is noted that there are many false-positive interactions in the candidate GWGEN as a result of various experimental conditions in different databases. Here, we applied a system order detection approach in Equations (5) and (9) to prune the false-positive interactions. According to the Akaike information criterion (AIC) theory [101], the smallest AIC value would lead to the most accurate model. In other words, the smaller the AIC value we get, the closer we detect to the real system order. The formulas of AIC for determining the system order of interactions among the $i$-th protein and the $q$-th gene are given as below:

$$AIC(Y_i) = \log(\hat{\rho}_{i,P}^2) + \frac{2(Y_i+1)}{N}$$
$$\text{where } \hat{\rho}_{i,P} = \sqrt{\frac{(P_i - (\Xi_{i,P} \cdot \hat{\varphi}_{i,P}))^T (P_i - (\Xi_{i,P} \cdot \hat{\varphi}_{i,P}))}{N}} \qquad (11)$$

$\hat{\rho}_{i,P}$ and $Y_i$ denote the estimated residual error and the number (system order) of PPIs with the $i$-th protein, respectively; $\hat{\varphi}_{i,P}$ denotes the estimated interaction parameters of the $i$-th protein by solving (6). Based on the AIC theory, the real system order $Y_i^*$ resulting in the smallest $AIC(Y_i^*)$.

$$AIC(J_q, W_q, H_q) = \log(\hat{\rho}_{q,G}^2) + \frac{2(\theta_{q,G}+1)}{N}$$
$$\text{where } \hat{\rho}_{q,G} = \sqrt{\frac{(G_q - (\Xi_{q,G} \cdot \hat{\varphi}_{q,G}))^T (G_q - (\Xi_{q,G} \cdot \hat{\varphi}_{q,G}))}{N}} \text{ and } \theta_{q,G} = J_q + W_q + H_q \qquad (12)$$

$\hat{\rho}_{i,P}$ and $\theta_{q,G}$ represent the estimated residual error and the number of regulations on the $q$-th gene, respectively; $\hat{\varphi}_{q,G}$ is the estimated parameter vector of the $q$-th gene obtained by solving (9). It is noted that the real system order $J_q^* + W_q^* + H_q^*$ lead to the smallest $AIC(J_q^* + W_q^* + H_q^*)$. For each protein, gene, miRNA, and lncRNA, we used forward and backward search to find the real system order by AIC. The unimportant interactions among the candidate GWGEN, which are out of the system order, would be removed via the system order detection approach. By doing so, we could find the real GWGENs of DLBCL GCB and DLBCL ABC, respectively. The system identification method and system order detection approach could be applied to the lncRNA and miRNA system models as well (Supplementary Materials).

### 4.4. Extracting the Core GWGENs from the Real GWGENs by Principal Network Projection (PNP) Method

Although we have pruned the false-positive interactions from the candidate GW-GEN by the system identification method and system order detection approach, the real GWGENs of DLBCL GCB and DLBCL ABC in Figures S1 and S2 are still too complex to investigate the common and specific pathogenic molecular mechanisms between DLBCL GCB and DLBCL ABC. Therefore, we utilize the PNP method to extract the core GWGENs from the real GWGENs of DLBCL GCB and DLBCL ABC. Before using the PNP method, we have to build a combined network matrix $Z$ as follows:

$$Z = \begin{bmatrix}
\hat{\alpha}_{11} & \cdots & \hat{\alpha}_{1k} & \cdots & \hat{\alpha}_{1K} & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\alpha}_{i1} & \cdots & \hat{\alpha}_{ik} & \cdots & \hat{\alpha}_{iK} & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\alpha}_{I1} & \cdots & \hat{\alpha}_{Ik} & \cdots & \hat{\alpha}_{IK} & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\
\hat{A}_{11} & \cdots & \hat{A}_{1j} & \cdots & \hat{A}_{1J} & \hat{B}_{11} & \cdots & \hat{B}_{1w} & \cdots & \hat{B}_{1W} & -\hat{C}_{11} & \cdots & -\hat{C}_{1h} & \cdots & -\hat{C}_{1H} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{A}_{q1} & \cdots & \hat{A}_{qj} & \cdots & \hat{A}_{qJ} & \hat{B}_{q1} & \cdots & \hat{B}_{qw} & \cdots & \hat{B}_{qW} & -\hat{C}_{q1} & \cdots & -\hat{C}_{qh} & \cdots & -\hat{C}_{qH} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{A}_{Q1} & \cdots & \hat{A}_{Qj} & \cdots & \hat{A}_{QJ} & \hat{B}_{Q1} & \cdots & \hat{B}_{Qw} & \cdots & \hat{B}_{QW} & -\hat{C}_{Q1} & \cdots & -\hat{C}_{Qh} & \cdots & -\hat{C}_{QH} \\
\hat{\beta}_{11} & \cdots & \hat{\beta}_{1j} & \cdots & \hat{\beta}_{1J} & \hat{\chi}_{11} & \cdots & \hat{\chi}_{1w} & \cdots & \hat{\chi}_{1W} & -\hat{\gamma}_{11} & \cdots & -\hat{\gamma}_{1h} & \cdots & -\hat{\gamma}_{1H} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\beta}_{v1} & \cdots & \hat{\beta}_{vj} & \cdots & \hat{\beta}_{vJ} & \hat{\chi}_{v1} & \cdots & \hat{\chi}_{vw} & \cdots & \hat{\chi}_{vW} & -\hat{\gamma}_{v1} & \cdots & -\hat{\gamma}_{vh} & \cdots & -\hat{\gamma}_{vH} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\beta}_{V1} & \cdots & \hat{\beta}_{Vj} & \cdots & \hat{\beta}_{VJ} & \hat{\chi}_{v1} & \cdots & \hat{\chi}_{Vw} & \cdots & \hat{\chi}_{VW} & -\hat{\gamma}_{V1} & \cdots & -\hat{\gamma}_{Vh} & \cdots & -\hat{\gamma}_{VH} \\
\hat{\sigma}_{11} & \cdots & \hat{\sigma}_{1j} & \cdots & \hat{\sigma}_{1J} & \hat{\delta}_{11} & \cdots & \hat{\delta}_{1w} & \cdots & \hat{\delta}_{1W} & -\hat{\omega}_{11} & \cdots & -\hat{\omega}_{1h} & \cdots & -\hat{\omega}_{1H} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\sigma}_{m1} & \cdots & \hat{\sigma}_{mj} & \cdots & \hat{\sigma}_{mJ} & \hat{\delta}_{m1} & \cdots & \hat{\delta}_{mw} & \cdots & \hat{\delta}_{mW} & -\hat{\omega}_{m1} & \cdots & -\hat{\omega}_{mh} & \cdots & -\hat{\omega}_{mH} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\hat{\sigma}_{M1} & \cdots & \hat{\sigma}_{Mj} & \cdots & \hat{\sigma}_{MJ} & \hat{\delta}_{M1} & \cdots & \hat{\delta}_{Mw} & \cdots & \hat{\delta}_{MW} & -\hat{\omega}_{M1} & \cdots & -\hat{\omega}_{Mh} & \cdots & -\hat{\omega}_{MH}
\end{bmatrix} \in \Re^{(I*+Q*+V*+M*)\times(J*+W*+H*)} \tag{13}$$

where the estimated parameters in (13) are obtained by solving the constrained linear least square estimation problem and conducting a system order detection approach based on AIC. The entry, which is pruned by AIC, would be padded with zero. The *i*-th row of $Z$ denotes the interaction and regulation parameters of the *i*-th node in the real GWGEN. The PNP method is based on the singular value decomposition of $Z$ shown as below:

$$Z = UKG^T \tag{14}$$

where $U \in \Re^{(I*+Q*+V*+M*)\times(I*+Q*+V*+M*)}$, $G \in \Re^{(J*+W*+H*)\times(J*+W*+H*)}$, $K = diag(k_1,\cdots,k_r,\cdots,k_{J*+W*+H*}) \in \Re^{(I*+Q*+V*+M*)\times(J*+W*+H*)}$ and $K$ denotes the diagonal matrix which consists of $J*+W*+H*$ singular values of $Z$ in descending order (i.e., $k_1 \geq \cdots \geq k_r \geq \cdots \geq k_{J*+W*+H*} \geq 0$). The normalization of singular values is defined as below:

$$E_r = \frac{k_r^2}{\sum\limits_{r=1}^{J*+W*+H*} k_r^2}, \quad \sum\limits_{i=1}^{J*+W*+H*} E_r = 1 \tag{15}$$

Here, we choose the top $R$ normalized singular values of combined network matrix $Z$ with the minimum $R$ to satisfy $\sum\limits_{r=1}^{R} E_r \geq 0.85$. It shows that we could use the top $R$ singular vectors to construct 85% network structure as principal network structure. Afterwards, we project each node in the real GWGEN (i.e., each row in $Z$) to the top $R$ singular vectors in $G^T$ as below:

$$F(t,a) = d_{t,:} \cdot r_{a,:}^T, \text{ for } t = 1,\ldots I^* + Q^* + V^* + M^*, \ a = 1,\ldots R. \tag{16}$$

where $d_{t,:}$ denotes the *t*-th row vector of $Z$; $r_{a,:}^T$ is the *a*-th singular vector of $G^T$. Subsequently, we compute the 2-norm projection value for each node in the following:

$$P(t) = \sqrt{\sum\limits_{a=1}^{R} F^2(t,a)}, \text{ for } t = 1,\ldots I^* + Q^* + V^* + M^*, \ a = 1,\ldots R. \tag{17}$$

where $P(t)$ denotes the 2-norm projection value of each $t$-th node in the real GWGEN on the top $R$ singular vectors. The greater a projection value is, the more significant the $t$-th node in the principal structure of the real GWGEN. If the projection value approaches zero, it means that the related $t$-th node is almost independent to the principal network structure. In other words, the greater the projection value of a node in real GWGEN is, the higher probability is that a node will be an important component in the principal network structure. Lastly, the core GWGENs of DLBCL ABC and DLBCL GCB could be extracted from the real GWGENs based on the top-rank 3000 projection values of the nodes. Moreover, the core GWGENs of DLBCL ABC and DLBCL GCB are shown in Figures S3 and S4.

*4.5. Deep Neural Netwok (DNN)-Based Drug-Target Interaction (DTI) Model for Multiple-Molecule Drug Design*

To train a DNN-based DTI model, the drug-target interaction dataset came from BindingDB [102]. We picked drugs that at least had four interactions. Hence, in the selected dataset, there are 80,291 known drug-target interactions between 38,015 drugs and 7292 proteins. In order to simply avoid a class imbalance issue, which would degrade the training performance or make the learning progress biased toward the majority class, we randomly chose the negative instance (unknown drug-target pair) in the same size as positive instance (known drug-target pair). We trained the model using 70% of the data containing 10% of the data as validation set. The remaining 30% of data were used as testing set. Delineating the drug-target pair in a numerical vector, we transformed them into a feature vector by PyBioMed Python package under python 2.7 environment [103]. The PyMolecule module in PyBioMed was responsible to transform the drug descriptors. The drug features include commonly used structural and physicochemical information. The PyProtein module in PyBioMed was applied to transform the target descriptors. The target features were computed based on the structural and physicochemical properties of proteins and peptide from amino acid sequence. The feature vector for each drug-target pair can be represented in the following form:

$$w_{drug-target} = [D, T] = [d_1, d_2, \cdots, d_X, t_1, t_2, \cdots, t_Y] \tag{18}$$

where $w_{drug-target}$ denotes a feature vector of drug-target pair, $X$ and $Y$ are the total number of drug features and target features, which are 363 and 996, respectively; $D$ and $T$ indicates the feature vector of the relevant drug-target pair; $d_X$ is the $x$-th drug feature and $t_Y$ is the $y$-th target feature. Since the drug and target features are measured in different scales, we performed normalization before training. Then, we applied principal component analysis (PCA) [104] to decrease the feature size from 1359 to 618. By doing so, we not only could remove noisy feature but also reduce memory consumption.

For the architecture of the DNN-based DTI model, the input layer contains 618 neurons, followed 512, 256, 128, and 64 neurons in the hidden layers, respectively. The output layer is with one neuron. The optimal hyperparameters were found based on 10-fold cross validation (Figure S6). Each layer of DNN-based DTI model could be simplified into a function as follows:

$$h_n = \sigma(w^T x_n + b) \tag{19}$$

where $x_n$ denotes the input of the $n$-th drug-target feature vector, $h_n$ indicates the output of each layer; $w$ is the weighting matrix; $b$ is the bias vector; $\sigma$ is the activation function, by which sigmoid activation function is used for the output layer and ReLU [105] is used for the hidden layer. We added dropout on each hidden layer for reducing overfitting. Meanwhile, the model training would be terminated once the model performance stopped to improve on the validation set by early stopping function. Moreover, we chose the binary cross-entropy to be the cost function:

$$C_n(w,b) = -\frac{1}{N}\sum_{n=1}^{N}\left(p_n\log(\hat{p}_n) + (1-p_n)\log(1-\hat{p}_n)\right)$$

$$L(w,b) = \frac{1}{N}\sum_{n=1}^{N}C_n(p_n,\hat{p}_n) \tag{20}$$

where $L(w,b)$ is the average of total loss; $p_n$ denotes the $n$-th true positive instance (1) or true negative instance (0) of drug-target binding; $\hat{p}_n$ denotes the $n$-th predicted probability of positive instance (1) or predicted probability of negative instance (0) of drug-target binding. For obtaining the optimal network parameter set $\phi^*$, the cost function is in the following:

$$\phi^* = \arg\min_{\phi} L(\phi) \tag{21}$$

The above equation could be achieved by the backpropagation algorithm [106]. The updated weight and bias parameters for the $j$-th epoch is shown as below:

$$\phi^j = \phi^{j-1} - \eta\nabla L(\phi^{j-1}),$$

$$\text{where } \nabla L(\phi^{j-1}) = \begin{bmatrix} \frac{\partial L(\phi^{j-1})}{\partial w_1} \\ \vdots \\ \frac{\partial L(\phi^{j-1})}{\partial w_h} \\ \frac{\partial L(\phi^{j-1})}{\partial b_1} \\ \vdots \\ \frac{\partial L(\phi^{j-1})}{\partial b_h} \end{bmatrix}. \tag{22}$$

where $\eta$ is the learning rate, which is 0.001; $\nabla L(\phi^{j-1})$ denotes the gradient of $L(\phi^{j-1})$.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23126732/s1.

**Author Contributions:** Conceptualization B.-S.C.; Methodology, S.-J.Y. and T.-Y.Y.; Software, T.-Y.Y.; Validation S.-J.Y. and T.-Y.Y.; Formal Analysis, S.-J.Y. and T.-Y.Y.; Investigation, T.-Y.Y.; Data Curation, T.-Y.Y.; Writing—Original Draft Preparation, S.-J.Y. and T.-Y.Y.; Writing—Review and Editing, S.-J.Y. and B.-S.C.; Visualization, T.-Y.Y.; Supervision, B.-S.C.; Funding Acquisition, B.-S.C. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The DLBCL microarray data is from GSE117556 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117556, accessed on 18 May 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Clarke, C.A.; Glaser, S.L.; Dorfman, R.F.; Bracci, P.M.; Eberle, E.; Holly, E.A.J.C.E.; Biomarkers, P. Expert review of non-Hodgkin's lymphomas in a population-based cancer registry: Reliability of diagnosis and subtype classifications. *Cancer Epidemiol. Biomark. Prev.* **2004**, *13*, 138–143. [CrossRef] [PubMed]
2. Prochazka, V.; Jarošová, M.; Prouzova, Z.; Nedomova, R.; Papajik, T.; Indrák, K. Immune escape mechanisms in diffuse large B-cell lymphoma. *Int. Sch. Res. Not.* **2012**, *2012*, 208903. [CrossRef]
3. Sehn, L.H.; Gascoyne, R.D. Diffuse large B-cell lymphoma: Optimizing outcome in the context of clinical and biologic heterogeneity. *Blood* **2015**, *125*, 22–32. [CrossRef] [PubMed]
4. Salles, G.; Barrett, M.; Foà, R.; Maurer, J.; O'Brien, S.; Valente, N.; Wenger, M.; Maloney, D.G. Rituximab in B-cell hematologic malignancies: A review of 20 years of clinical experience. *Adv. Ther.* **2017**, *34*, 2232–2273. [CrossRef] [PubMed]
5. Mok, C.C. Rituximab for the treatment of rheumatoid arthritis: An update. *Drug Des. Dev. Ther.* **2014**, *8*, 87. [CrossRef]

6.  Al-Homsi, A.S.; Roy, T.S.; Cole, K.; Feng, Y.; Duffner, U. Post-transplant high-dose cyclophosphamide for the prevention of graft-versus-host disease. *Biol. Blood Marrow Transplant.* **2015**, *21*, 604–611. [CrossRef] [PubMed]
7.  Travis, L.B.; Curtis, R.E.; Glimelius, B.; Holowaty, E.J.; Van Leeuwen, F.E.; Lynch, C.F.; Hagenbeek, A.; Stovall, M.; Banks, P.M.; Adami, J.; et al. Bladder and Kidney Cancer Following Cyclophosphamide Therapy for Non-Hodgkin's Lymphoma. *JNCI J. Natl. Cancer Inst.* **1995**, *87*, 524–531. [CrossRef] [PubMed]
8.  Thorn, C.F.; Oshiro, C.; Marsh, S.; Hernandez-Boussard, T.; McLeod, H.; Klein, T.E.; Altman, R.B. Doxorubicin pathways: Pharmacodynamics and adverse effects. *Pharm. Genom.* **2011**, *21*, 440. [CrossRef]
9.  Vayssade, M.; Faridoni-Laurens, L.; Bénard, J.; Ahomadegbe, J.-C. Expression of p53-family members and associated target molecules in breast cancer cell lines in response to vincristine treatment. *Biochem. Pharmacol.* **2002**, *63*, 1609–1617. [CrossRef]
10. Samoon, Z.; Shabbir-Moosajee, M. Vincristine-induced vocal cord palsy and successful re-treatment in a patient with diffuse large B cell lymphoma: A case report. *BMC Res. Notes* **2014**, *7*, 318. [CrossRef]
11. Kimberly, R.P. Mechanisms of action, dosage schedules, and side effects of steroid therapy. *Curr. Opin. Rheumatol.* **1991**, *3*, 373–379. [CrossRef]
12. Modlinski, R.; Fields, K.B. The effect of anabolic steroids on the gastrointestinal system, kidneys, and adrenal glands. *Curr. Sports Med. Rep.* **2006**, *5*, 104–109. [CrossRef]
13. Papageorgiou, S.G.; Thomopoulos, T.P.; Liaskas, A.; Vassilakopoulos, T.P. Monoclonal Antibodies in the Treatment of Diffuse Large B-Cell Lymphoma: Moving beyond Rituximab. *Cancers* **2022**, *14*, 1917. [CrossRef]
14. Sehn, L.H.; Herrera, A.F.; Flowers, C.R.; Kamdar, M.K.; McMillan, A.; Hertzberg, M.; Assouline, S.; Kim, T.M.; Kim, W.S.; Ozcan, M. Polatuzumab vedotin in relapsed or refractory diffuse large B-cell lymphoma. *J. Clin. Oncol.* **2020**, *38*, 155. [CrossRef]
15. Kalakonda, N.; Maerevoet, M.; Cavallo, F.; Follows, G.; Goy, A.; Vermaat, J.S.; Casasnovas, O.; Hamad, N.; Zijlstra, J.M.; Bakhshi, S. Selinexor in patients with relapsed or refractory diffuse large B-cell lymphoma (SADAL): A single-arm, multinational, multicentre, open-label, phase 2 trial. *Lancet Haematol.* **2020**, *7*, e511–e522. [CrossRef]
16. Salles, G.; Duell, J.; Barca, E.G.; Tournilhac, O.; Jurczak, W.; Liberati, A.M.; Nagy, Z.; Obr, A.; Gaidano, G.; André, M. Tafasitamab plus lenalidomide in relapsed or refractory diffuse large B-cell lymphoma (L-MIND): A multicentre, prospective, single-arm, phase 2 study. *Lancet Oncol.* **2020**, *21*, 978–988. [CrossRef]
17. Cheson, B.D.; Nowakowski, G.; Salles, G. Diffuse large B-cell lymphoma: New targets and novel therapies. *Blood Cancer J.* **2021**, *11*, 68. [CrossRef]
18. Mohs, R.C.; Greig, N.H. Drug discovery and development: Role of basic biological research. *Alzheimer's Dement. Transl. Res. Clin. Interv.* **2017**, *3*, 651–657. [CrossRef]
19. Takebe, T.; Imai, R.; Ono, S. The current status of drug discovery and development as originated in United States academia: The influence of industrial and academic collaboration on drug discovery and development. *Clin. Transl. Sci.* **2018**, *11*, 597–606. [CrossRef]
20. Bailón-Moscoso, N.; Romero-Benavides, J.C.; Ostrosky-Wegman, P. Development of anticancer drugs based on the hallmarks of tumor cells. *Tumor Biol.* **2014**, *35*, 3981–3995. [CrossRef]
21. Sachdev, K.; Gupta, M.K. A comprehensive review of feature based methods for drug target interaction prediction. *J. Biomed. Inform.* **2019**, *93*, 103159. [CrossRef]
22. Butina, D.; Segall, M.D.; Frankcombe, K. Predicting ADME properties in silico: Methods and models. *Drug Discov. Today* **2002**, *7*, S83–S88. [CrossRef]
23. Li, H.; Gao, Z.; Kang, L.; Zhang, H.; Yang, K.; Yu, K.; Luo, X.; Zhu, W.; Chen, K.; Shen, J. TarFisDock: A web server for identifying drug targets with docking approach. *Nucleic Acids Res.* **2006**, *34*, W219–W224. [CrossRef]
24. Cheng, A.C.; Coleman, R.G.; Smyth, K.T.; Cao, Q.; Soulard, P.; Caffrey, D.R.; Salzberg, A.C.; Huang, E.S. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **2007**, *25*, 71–75. [CrossRef]
25. Pujadas, G.; Vaque, M.; Ardevol, A.; Blade, C.; Salvado, M.; Blay, M.; Fernandez-Larrea, J.; Arola, L. Protein-ligand docking: A review of recent advances and future perspectives. *Curr. Pharm. Anal.* **2008**, *4*, 1–19. [CrossRef]
26. Bagherian, M.; Sabeti, E.; Wang, K.; Sartor, M.A.; Nikolovska-Coleska, Z.; Najarian, K. Machine learning approaches and databases for prediction of drug–target interaction: A survey paper. *Brief. Bioinform.* **2021**, *22*, 247–269. [CrossRef]
27. Nath, A.; Kumari, P.; Chaube, R. Prediction of human drug targets and their interactions using machine learning methods: Current and future perspectives. *Comput. Drug Discov. Des.* **2018**, *1762*, 21–30.
28. Wen, M.; Zhang, Z.; Niu, S.; Sha, H.; Yang, R.; Yun, Y.; Lu, H.J. Deep-learning-based drug–target interaction prediction. *J. Proteome Res.* **2017**, *16*, 1401–1409. [CrossRef]
29. Gao, K.Y.; Fokoue, A.; Luo, H.; Iyengar, A.; Dey, S.; Zhang, P. Interpretable drug target prediction using deep neural representation. In Proceedings of the IJCAI, Stockholm, Sweden, 13–19 July 2018; pp. 3371–3377.
30. Öztürk, H.; Özgür, A.; Ozkirimli, E.J.B. DeepDTA: Deep drug–target binding affinity prediction. *Bioinformatics* **2018**, *34*, i821–i829. [CrossRef]
31. Lee, I.; Keum, J.; Nam, H. DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences. *PLoS Comput. Biol.* **2019**, *15*, e1007129. [CrossRef]
32. You, J.; McLeod, R.D.; Hu, P. Predicting drug-target interaction network using deep learning model. *Comput. Biol. Chem.* **2019**, *80*, 90–101. [CrossRef] [PubMed]

33. Ezzat, A.; Wu, M.; Li, X.-L.; Kwoh, C.-K. Drug-target interaction prediction via class imbalance-aware ensemble learning. *BMC Bioinform.* **2016**, *17*, 509. [CrossRef] [PubMed]

34. Li, Y.; Qiao, G.; Wang, K.; Wang, G. Drug–target interaction predication via multi-channel graph neural networks. *Brief. Bioinform.* **2021**, *23*, bbab346. [CrossRef] [PubMed]

35. Cheng, Z.; Yan, C.; Wu, F.; Wang, J. Drug-target interaction prediction using multi-head self-attention and graph attention network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *1*, 33956632. [CrossRef]

36. Zhao, T.; Hu, Y.; Valsdottir, L.R.; Zang, T.; Peng, J. Identifying drug–target interactions based on graph convolutional network and deep neural network. *Brief. Bioinform.* **2020**, *22*, 2141–2150. [CrossRef]

37. Tillmann, S.; Bernhagen, J.; Noels, H. Arrest Functions of the MIF Ligand/Receptor Axes in Atherogenesis. *Front. Immunol.* **2013**, *4*, 115. [CrossRef]

38. Figueiredo, C.R.; Azevedo, R.A.; Mousdell, S.; Resende-Lara, P.T.; Ireland, L.; Santos, A.; Girola, N.; Cunha, R.L.O.R.; Schmid, M.C.; Polonelli, L.; et al. Blockade of MIF-CD74 Signalling on Macrophages and Dendritic Cells Restores the Antitumour Immune Response against Metastatic Melanoma. *Front. Immunol.* **2018**, *9*, 1132. [CrossRef]

39. Gil-Yarom, N.; Radomir, L.; Sever, L.; Kramer, M.P.; Lewinsky, H.; Bornstein, C.; Blecher-Gonen, R.; Barnett-Itzhaki, Z.; Mirkin, V.; Friedlander, G.; et al. CD74 is a novel transcription regulator. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 562–567. [CrossRef]

40. Roskoski, R. Src protein–tyrosine kinase structure and regulation. *Biochem. Biophys. Res. Commun.* **2004**, *324*, 1155–1164. [CrossRef]

41. Huang, X.; Meng, B.; Iqbal, J.; Ding, B.B.; Perry, A.M.; Cao, W.; Smith, L.M.; Bi, C.; Jiang, C.; Greiner, T.C.; et al. Activation of the STAT3 signaling pathway is associated with poor survival in diffuse large B-cell lymphoma treated with R-CHOP. *J. Clin. Oncol.* **2013**, *31*, 4520–4528. [CrossRef]

42. Pawlus, M.R.; Wang, L.; Hu, C.J. STAT3 and HIF1α cooperatively activate HIF1 target genes in MDA-MB-231 and RCC4 cells. *Oncogene* **2014**, *33*, 1670–1679. [CrossRef]

43. Carmeliet, P.; Dor, Y.; Herbert, J.-M.; Fukumura, D.; Brusselmans, K.; Dewerchin, M.; Neeman, M.; Bono, F.; Abramovitch, R.; Maxwell, P.; et al. Role of HIF-1α in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. *Nature* **1998**, *394*, 485–490. [CrossRef]

44. Wein, F.; Otto, T.; Lambertz, P.; Fandrey, J.; Hansmann, M.-L.; Küppers, R. Potential role of hypoxia in early stages of Hodgkin lymphoma pathogenesis. *Haematologica* **2015**, *100*, 1320–1326. [CrossRef]

45. Burger, R. Impact of interleukin-6 in hematological malignancies. *Transfus. Med. Hemotherapy* **2013**, *40*, 336–343. [CrossRef]

46. Sánchez-Barrena, M.J.; Vallis, Y.; Clatworthy, M.R.; Doherty, G.J.; Veprintsev, D.B.; Evans, P.R.; McMahon, H.T. Bin2 is a membrane sculpting N-BAR protein that influences leucocyte podosomes, motility and phagocytosis. *PLoS ONE* **2012**, *7*, e52401. [CrossRef]

47. Davey, R.A.; Grossmann, M. Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *Clin. Biochem. Rev.* **2016**, *37*, 3–15.

48. Nguyen, L.; Papenhausen, P.; Shao, H. The Role of c-MYC in B-Cell Lymphomas: Diagnostic and Molecular Aspects. *Genes* **2017**, *8*, 116. [CrossRef]

49. Tjin, E.P.M.; Groen, R.W.J.; Vogelzang, I.; Derksen, P.W.B.; Klok, M.D.; Meijer, H.P.; van Eeden, S.; Pals, S.T.; Spaargaren, M. Functional analysis of HGF/MET signaling and aberrant HGF-activator expression in diffuse large B-cell lymphoma. *Blood* **2006**, *107*, 760–768. [CrossRef]

50. Lam, B.Q.; Dai, L.; Qin, Z. The role of HGF/c-MET signaling pathway in lymphoma. *J. Hematol. Oncol.* **2016**, *9*, 135. [CrossRef]

51. Haycraft, C.J.; Banizs, B.; Aydin-Son, Y.; Zhang, Q.; Michaud, E.J.; Yoder, B.K. Gli2 and Gli3 localize to cilia and require the intraflagellar transport protein polaris for processing and function. *PLoS Genet.* **2005**, *1*, e53. [CrossRef]

52. Lentjes, M.H.F.M.; Niessen, H.E.C.; Akiyama, Y.; de Bruïne, A.P.; Melotte, V.; van Engeland, M. The emerging role of GATA transcription factors in development and disease. *Expert Rev. Mol. Med.* **2016**, *18*, e3. [CrossRef]

53. Crispino, J.D.; Horwitz, M.S. GATA factor mutations in hematologic disease. *Blood* **2017**, *129*, 2103–2110. [CrossRef]

54. Chakrama, F.Z.; Seguin-Py, S.; Le Grand, J.N.; Fraichard, A.; Delage-Mourroux, R.; Despouy, G.; Perez, V.; Jouvenot, M.; Boyer-Guittaut, M. GABARAPL1 (GEC1) associates with autophagic vesicles. *Autophagy* **2010**, *6*, 495–505. [CrossRef]

55. Zhong, X.; Xiao, Y.; Chen, C.; Wei, X.; Hu, C.; Ling, X.; Liu, X. MicroRNA-203-mediated posttranscriptional deregulation of CPEB4 contributes to colorectal cancer progression. *Biochem. Biophys. Res. Commun.* **2015**, *466*, 206–213. [CrossRef]

56. Béguelin, W.; Popovic, R.; Teater, M.; Jiang, Y.; Bunting, K.L.; Rosen, M.; Shen, H.; Yang, S.N.; Wang, L.; Ezponda, T.; et al. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell* **2013**, *23*, 677–692. [CrossRef]

57. Bisserier, M.; Wajapeyee, N. Mechanisms of resistance to EZH2 inhibitors in diffuse large B-cell lymphomas. *Blood* **2018**, *131*, 2125–2137. [CrossRef]

58. Kim, K.H.; Roberts, C.W. Targeting EZH2 in cancer. *Nat. Med.* **2016**, *22*, 128–134. [CrossRef]

59. Elian, F.A.; Yan, E.; Walter, M.A. FOXC1, the new player in the cancer sandbox. *Oncotarget* **2017**, *9*, 8165–8178. [CrossRef] [PubMed]

60. Park, S.; Han, S.-H.; Kim, H.-G.; Jeong, J.; Choi, M.; Kim, H.-Y.; Kim, M.-G.; Park, J.-K.; Han, J.E.; Cho, G.-J.; et al. Suppression of PRPF4 regulates pluripotency, proliferation, and differentiation in mouse embryonic stem cells. *Cell Biochem. Funct.* **2019**, *37*, 608–617. [CrossRef] [PubMed]

61. Song, H.; Dang, X.; He, Y.-Q.; Zhang, T.; Wang, H.-Y. Selection of housekeeping genes as internal controls for quantitative RT-PCR analysis of the veined rapa whelk (*Rapana venosa*). *PeerJ* **2017**, *5*, e3398. [CrossRef] [PubMed]

62. Yao, T.-P.; Oh, S.P.; Fuchs, M.; Zhou, N.-D.; Ch'ng, L.-E.; Newsome, D.; Bronson, R.T.; Li, E.; Livingston, D.M.; Eckner, R. Gene Dosage–Dependent Embryonic Development and Proliferation Defects in Mice Lacking the Transcriptional Integrator p300. *Cell* **1998**, *93*, 361–372. [CrossRef]

63. Pasqualucci, L.; Dominguez-Sola, D.; Chiarenza, A.; Fabbri, G.; Grunn, A.; Trifonov, V.; Kasper, L.H.; Lerach, S.; Tang, H.; Ma, J.; et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **2011**, *471*, 189–195. [CrossRef]

64. Dang, C.V. MYC on the Path to Cancer. *Cell* **2012**, *149*, 22–35. [CrossRef]

65. Ortega, M.; Bhatnagar, H.; Lin, A.P.; Wang, L.; Aster, J.C.; Sill, H.; Aguiar, R.C.T. A microRNA-mediated regulatory loop modulates NOTCH and MYC oncogenic signals in B- and T-cell malignancies. *Leukemia* **2015**, *29*, 968–976. [CrossRef]

66. Lv, X.; Feng, L.; Ge, X.; Lu, K.; Wang, X. Interleukin-9 promotes cell survival and drug resistance in diffuse large B-cell lymphoma. *J. Exp. Clin. Cancer Res.* **2016**, *35*, 106. [CrossRef]

67. Chen, A.; Zhong, L.; Lv, J. FOXL1 overexpression is associated with poor outcome in patients with glioma. *Oncol. Lett.* **2019**, *18*, 751–757. [CrossRef]

68. Ni, H.; Tong, R.; Zou, L.; Song, G.; Cho, W.C. MicroRNAs in diffuse large B-cell lymphoma. *Oncol. Lett.* **2016**, *11*, 1271–1280. [CrossRef]

69. Lohr, J.G.; Stojanov, P.; Lawrence, M.S.; Auclair, D.; Chapuy, B.; Sougnez, C.; Cruz-Gordillo, P.; Knoechel, B.; Asmann, Y.W.; Slager, S.L.; et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3879–3884. [CrossRef]

70. Vela-Chávez, T.; Adam, P.; Kremer, M.; Bink, K.; Bacon, C.M.; Menon, G.; Ferry, J.A.; Fend, F.; Jaffe, E.S.; Quintanilla-Martínez, L. Cyclin D1 positive diffuse large B-cell lymphoma is a post-germinal center-type lymphoma without alterations in the CCND1 gene locus. *Leuk Lymphoma* **2011**, *52*, 458–466. [CrossRef]

71. Vermeulen, K.; Van Bockstaele, D.R.; Berneman, Z.N. The cell cycle: A review of regulation, deregulation and therapeutic targets in cancer. *Cell Prolif.* **2003**, *36*, 131–149. [CrossRef]

72. Gennaro, V.J.; Stanek, T.J.; Peck, A.R.; Sun, Y.; Wang, F.; Qie, S.; Knudsen, K.E.; Rui, H.; Butt, T.; Diehl, J.A.; et al. Control of CCND1 ubiquitylation by the catalytic SAGA subunit USP22 is essential for cell cycle progression through G1 in cancer cells. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E9298. [CrossRef]

73. Hwang, T.-C.; Kirk, K.L. The CFTR ion channel: Gating, regulation, and anion permeation. *Cold Spring Harb. Perspect Med.* **2013**, *3*, a009498. [CrossRef]

74. Wang, J.; Xu-Monette, Z.Y.; Jabbar, K.J.; Shen, Q.; Manyam, G.C.; Tzankov, A.; Visco, C.; Wang, J.; Montes-Moreno, S.; Dybkær, K.; et al. AKT Hyperactivation and the Potential of AKT-Targeted Therapy in Diffuse Large B-Cell Lymphoma. *Am. J. Pathol.* **2017**, *187*, 1700–1716. [CrossRef]

75. Dittmer, J. The Biology of the Ets1 Proto-Oncogene. *Mol. Cancer* **2003**, *2*, 29. [CrossRef]

76. Bonetti, P.; Testoni, M.; Scandurra, M.; Ponzoni, M.; Piva, R.; Mensah, A.A.; Rinaldi, A.; Kwee, I.; Tibiletti, M.G.; Iqbal, J.; et al. Deregulation of ETS1 and FLI1 contributes to the pathogenesis of diffuse large B-cell lymphoma. *Blood* **2013**, *122*, 2233–2241. [CrossRef]

77. Ochoa-Hernández, A.B.; Ramos-Solano, M.; Meza-Canales, I.D.; García-Castro, B.; Rosales-Reynoso, M.A.; Rosales-Aviña, J.A.; Barrera-Chairez, E.; Ortíz-Lazareno, P.C.; Hernández-Flores, G.; Bravo-Cuellar, A.; et al. Peripheral T-lymphocytes express WNT7A and its restoration in leukemia-derived lymphoblasts inhibits cell proliferation. *BMC Cancer* **2012**, *12*, 60. [CrossRef]

78. Yuan, H.; Li, Z.-M.; Shao, J.; Ji, W.-X.; Xia, W.; Lu, S. FGF2/FGFR1 regulates autophagy in FGFR1-amplified non-small cell lung cancer cells. *J. Exp. Clin. Cancer Res.* **2017**, *36*, 72. [CrossRef]

79. Parzych, K.R.; Klionsky, D.J. An overview of autophagy: Morphology, mechanism, and regulation. *Antioxid. Redox Signal.* **2014**, *20*, 460–473. [CrossRef]

80. Cui, W.; Zheng, S.; Liu, Z.; Wang, W.; Cai, Y.; Bi, R.; Cao, B.; Zhou, X. PIK3CA expression in diffuse large B cell lymphoma tissue and the effect of its knockdown in vitro. *OncoTargets Ther.* **2017**, *10*, 2239–2247. [CrossRef]

81. Compagno, M.; Lim, W.K.; Grunn, A.; Nandula, S.V.; Brahmachary, M.; Shen, Q.; Bertoni, F.; Ponzoni, M.; Scandurra, M.; Califano, A.; et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **2009**, *459*, 717–721. [CrossRef]

82. Dong, P.; Xiong, Y.; Yue, J.; Hanley, S.J.B.; Watari, H. Tumor-Intrinsic PD-L1 Signaling in Cancer Initiation, Development and Treatment: Beyond Immune Evasion. *Front. Oncol.* **2018**, *8*, 386. [CrossRef] [PubMed]

83. Lamb, J. The Connectivity Map: A new tool for biomedical research. *Nat. Rev. Cancer* **2007**, *7*, 54–60. [CrossRef] [PubMed]

84. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [CrossRef] [PubMed]

85. Roy, P.; Ghosh, A. Mechanochemical cocrystallization to improve the physicochemical properties of chlorzoxazone. *CrystEngComm* **2020**, *22*, 4611–4620. [CrossRef]

86. Sogawa, C.; Eguchi, T.; Tran, M.T.; Ishige, M.; Trin, K.; Okusha, Y.; Taha, E.A.; Lu, Y.; Kawai, H.; Sogawa, N.; et al. Antiparkinson Drug Benztropine Suppresses Tumor Growth, Circulating Tumor Cells, and Metastasis by Acting on SLC6A3/DAT and Reducing STAT3. *Cancers* **2020**, *12*, 523. [CrossRef] [PubMed]

87. Prusila, R.E.I.; Peroja, P.; Jantunen, E.; Turpeenniemi-Hujanen, T.; Kuittinen, O.J.H.O. Treatment of diffuse large B-cell lymphoma in elderly patients: Replacing doxorubicin with either epirubicin or etoposide (VP-16). *Hematol. Oncol.* **2019**, *37*, 136–142. [CrossRef] [PubMed]

88. Moccia, A.A.; Schaff, K.; Hoskins, P.; Klasa, R.; Savage, K.J.; Shenkier, T.; Gascoyne, R.D.; Connors, J.M.; Sehn, L.H. R-CHOP with Etoposide Substituted for Doxorubicin (R-CEOP): Excellent Outcome in Diffuse Large B Cell Lymphoma for Patients with a Contraindication to Anthracyclines. *Blood* **2009**, *114*, 408. [CrossRef]

89. Huang, W.-Y.; Yang, P.-M.; Chang, Y.-F.; Marquez, V.E.; Chen, C.-C. Methotrexate induces apoptosis through p53/p21-dependent pathway and increases E-cadherin expression through downregulation of HDAC/EZH2. *Biochem. Pharmacol.* **2011**, *81*, 510–517. [CrossRef]

90. Han Li, C.; Chen, Y. Targeting EZH2 for cancer therapy: Progress and perspective. *Curr. Protein Pept. Sci.* **2015**, *16*, 559–570. [CrossRef]

91. Salwinski, L.; Miller, C.S.; Smith, A.J.; Pettit, F.K.; Bowie, J.U.; Eisenberg, D. The database of interacting proteins: 2004 update. *Nucleic Acids Res.* **2004**, *32*, D449–D451. [CrossRef]

92. Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N.H.; Chavali, G.; Chen, C.; Del-Toro, N. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **2014**, *42*, D358–D363. [CrossRef]

93. Chatr-Aryamontri, A.; Breitkreutz, B.-J.; Oughtred, R.; Boucher, L.; Heinicke, S.; Chen, D.; Stark, C.; Breitkreutz, A.; Kolas, N.; O'Donnell, L. The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* **2015**, *43*, D470–D478. [CrossRef]

94. Bader, G.D.; Betel, D.; Hogue, C.W. BIND: The biomolecular interaction network database. *Nucleic Acids Res.* **2003**, *31*, 248–250. [CrossRef]

95. Licata, L.; Briganti, L.; Peluso, D.; Perfetto, L.; Iannuccelli, M.; Galeota, E.; Sacco, F.; Palma, A.; Nardozza, A.P.; Santonico, E. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* **2012**, *40*, D857–D861. [CrossRef]

96. Zheng, G.; Tu, K.; Yang, Q.; Xiong, Y.; Wei, C.; Xie, L.; Zhu, Y.; Li, Y. ITFP: An integrated platform of mammalian transcription factors. *Bioinformatics* **2008**, *24*, 2416–2417. [CrossRef]

97. Bovolenta, L.A.; Acencio, M.L.; Lemke, N.J. HTRIdb: An open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genom.* **2012**, *13*, 405. [CrossRef]

98. Agarwal, V.; Bell, G.W.; Nam, J.-W.; Bartel, D.P. Predicting effective microRNA target sites in mammalian mRNAs. *elife* **2015**, *4*, e05005. [CrossRef]

99. Friard, O.; Re, A.; Taverna, D.; De Bortoli, M.; Corá, D. CircuitsDB: A database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinform.* **2010**, *11*, 435. [CrossRef]

100. Li, J.-H.; Liu, S.; Zhou, H.; Qu, L.-H.; Yang, J.-H. starBase v2. 0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **2014**, *42*, D92–D97. [CrossRef]

101. Sakamoto, Y.; Ishiguro, M.; Kitagawa, G.J.D. *Akaike Information Criterion Statistics*; Reidel, D., Ed.; Springer: Dordrecht, The Netherlands, 1986; Volume 81, p. 26853.

102. Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045–D1053. [CrossRef] [PubMed]

103. Dong, J.; Yao, Z.J.; Zhang, L.; Luo, F.; Lin, Q.; Lu, A.P.; Chen, A.F.; Cao, D.S. PyBioMed: A python library for various molecular representations of chemicals, proteins and DNAs and their interactions. *J. Cheminform.* **2018**, *10*, 16. [CrossRef] [PubMed]

104. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **2008**, *26*, 303–304. [CrossRef] [PubMed]

105. Nwankpa, C.; Ijomah, W.; Gachagan, A.; Marshall, S. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv* **2018**, arXiv:1811.03378.

106. Hecht-Nielsen, R. Theory of the backpropagation neural network. In *Neural Networks for Perception*; Elsevier: Amsterdam, The Netherlands, 1992; pp. 65–93.