A novel molecular classification of diffuse large B cell lymphoma based on Metabolism-related genes

ying duan

Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

Yanzhen Luo

Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

hong cen (✉ cenhong981@gxmu.edu.cn)

Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

<https://orcid.org/0000-0001-7233-5692>

Research article

Keywords: classification, Metabolism-related genes, Diffuse large B-cell lymphoma, Immune checkpoint, prognosis

Posted Date: December 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-132445/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose

About 30–40% of patients with diffuse large B-cell lymphoma (DLBCL) relapse or fail to respond to first-line treatment. The molecular heterogeneity is considered to be the main factor affecting the therapeutic response of DLBCL. The existing classification methods can not fully explain these heterogeneity, so we try to explain DLBCL heterogeneity by defining DLBCL subtypes from the perspective of metabolism.

Methods

In this study, we integrated five DLBCL data sets (GSE10846, GSE11318, GSE53786, GSE87371 and GSE23501) (n = 742) from geo database, screened 106 metabolic related genes (MAD > 0.5, Cox P < 0.001), and identified dlbc2 subclasses (nmftype1, nmftype2) by non-negative matrix factorization clustering (NMF).

Results

nmftype1 showed low metabolic activity ,while nmftype2 showed high metabolic activity. Compared with the two subtypes of immune infiltration, it was found that nmftype1 was mainly infiltrated by B cells, and nmftype2 was mainly infiltrated by T cells and macrophages, and the high expression of nmftype2 was more in immune checkpoint. The difference of metabolic subtype OS was statistically significant, and the overall survival (OS) of nmftype1 was worse than that of nmftype2. The combination of metabolic subtypes and ABCGCB subtypes can predict the prognosis of DLBCL patients better than the existing ABCGCB subtypes. Finally, 34 gene classifiers were identified. The consistency results were verified by GSE31312 (n = 470), and a new classification of DLBCL based on metabolic gene expression profile was established.

Conclusions

We have obtained a new DLBCL typing method, which has prognostic significance. It has a certain correlation with immune escape and can guide individualization application of immunotherapy and metabolic therapy.

Background

Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of non-Hodgkin lymphoma (NHL), accounting for 30% – 40% of all NHL cases[1, 2]. Most DLBCL patients can be cured with current first-line chemoimmunotherapy based on a combination of anthracyclines and anti-CD20 antibodies (such as R-CHOP). However, about 30–40% of patients still have relapse or ineffective first-line treatment[3].

Previous studies have shown that DLBCL is a malignant tumor with great heterogeneity in gene mutation, copy number (CN) change, structural variation, clinical presentation and prognosis, etc. [4–6], and molecular heterogeneity of DLBCLs is considered to be the main factor influencing the response to R-CHOP treatment. [7, 8] Elucidating the underlying molecular mechanism of DLBCL heterogeneity and discovering different subsets to subdivide DLBCL is essential for guiding clinical treatment and predicting prognosis.

DLBCL can be divided into germinal center B cell like (GCB) and activated B cell like (ABC) according to the cells of origin. About 10%-20% of the cases cannot be classified[9]. although the prognosis of ABC subtype has been confirmed to be worse than that of GCB subtype. subtypes[7], Heterogeneity still exists within GCB and ABC subtypes, and there are better and worse prognostic subgroups in each group[10]. Schmitz et al. Sequenced the whole exome of 372 genes, RNA sequence and gene copy number 574 DLBCL patients. and classified DLBCL patients into four subclasses: MCD (MYD88L265P and CD79B co-mutated), BN2 (BCL6 fusions or NOTCH2 mutated), N1 (NOTCH1 mutated), and EZB (EZH2 mutated or BCL2 translocated)[6]. However, this classification method has several drawbacks: 1. The sample size of subclass is small, especially N1 type, and the sample size is only one digit after stratification. 2. 55% of DLBCL specimens were not classified by this method. 3. Because the samples of ABC and unclassified subtypes were artificially enriched in this study, they could not represent the true subtype distribution. (Dubois and Jardin, 2018). Chapuy et al. conducted a comprehensive genetic analysis on 304 DLBCL samples and identified 5 DLBCL subclasses (C1-C5) with prominent genetic characteristics. However, this classification method requires in-depth genomic analysis, which certain overlaps with the subtypes obtained by Schmitz [3, 5]. Therefore, larger sample size and simpler classification methods and better heterogeneous subtype identification are needed to guide clinical practice[3].

Changes in metabolic activity enable cancer cells to acquire and maintain malignant characteristics[11, 12]. In 2005, Stefano Monti et al. Clustered gene expression profiles of 176 DLBCL samples to obtain three subsets of oxidative phosphorylation (OxPhos), B-cell receptor/proliferation, host response (HR). Genes involved in electron transport chain (ETC) complexes, OxPhos metabolism, and other mitochondrial functions were enriched in the OxPhos subtype. BCR/proliferation subtypes express higher levels of multiple components of the B cell receptor (BCR) signaling cascade and other B cell specific or essential transcription factors than the other two subtypes[13]. HR subset had T cell enrichment, and increased expression of complement and inflammatory mediators, etc[14]. These studies indicate that metabolism is associated with tumorigenesis and Cancer progression, and DLBCL has metabolic heterogeneity. These studies suggest the possibility of prospective classification of DLBCL from metabolism.

Methods

2.1 Patients and samples

Through gene expression database (GEO, <https://www.ncbi.nlm.nih.gov/geo/>)Using "DLBCL" and "GPL570" as key words, the eligible samples were retrieved. According to the inclusion criteria of GEO dataset :(1) The study species is human;(2) the patients were diagnosed with DLBCL ;(3) The sample were tissue specimen;(4) The platform file was GPL570;(5) The sample size was more than 50 cases;(6) there was mRNA expression data;(7) Existence of prognostic information;(8) The detection method is chip detection. five data sets of GSE23501[15], GSE53786[16], GSE10846[7], GSE87371[17], GSE11318 [18] are screened out(Table 1).

we annotate the data set by converting the probe ID to a gene symbol and extract the mRNA expression data of DLBCL sample in each data set. and got a total of 1024 sample data, The number of samples contributed by each dataset is shown in Table 1(Table 1). Using intersect function, we found that there were 114 identical samples in GSE53786 and GSE 10846, 41 identical samples in GSE 53786 and GSE 1318, and 163 identical samples in GSE 10846 and GSE 11318 (Figure S1). After removing the repeated samples, 747 samples were obtained, including 300 cases of GSE10846, 40 cases of GSE11318, 65 cases of GSE23501, 119 cases of GSE53786 and 223 cases of GSE87371. After removing the samples without survival data (including 3 cases of GSE11318 and 2 cases of GSE87371), the metadata set of 742 samples was obtained.

Table 1
the GEO dataset used in this study

GEO datasets	Total Sample Number of Data Sets	Number of DLBCL Samples Containing mRNA Expression Data	Original data processing	Country
GSE23501	69	65	normalization	The United States
GSE53786	119	119	normalization	The United States
GSE10846	420	414	normalization	The United States
GSE87371	223	223	normalization	The United States
GSE11318	406	203	normalization	The United States
(Supplementary note: Because the following steps need to combine these data for analysis, selecting the same platform data set for research can avoid more interference. The most commonly used platform for DLBCL samples in the GEO database is GPL570, Therefore, the data sets of the GPL570 platform files are selected for analysis)				

2.2. Identification of DLBCL subclasses

This study download the c2.cp.kegg.v7.0.symbols document from the The Molecular Signatures Database(MSigDB) (<https://www.gsea-msigdb.org/>). According to this table, genes related to metabolic pathway were extracted. we calculated the median absolute deviation (MAD) of each gene and used a univariate cox regression to evaluate the correlation between the expression of each gene and the overall survival (OS). We used $MAD > 0.5$ and $P < 0.001$ as criteria to screen important prognostic metabolic genes. $MAD > 0.5$ and $P < 0.001$ were used as standard screen metabolic genes with significant prognosis. The mRNA expression data of metabolic genes with significant prognosis were clustered unsupervised by NMR package [19]. The K value which the apparent correlation coefficient begins to decrease is selected as the optimal cluster number [20]. The same method was applied to the GSE31312 dataset [21]. GENE PATTERN's submap module(<https://cloud.genepattern.org/>) compare the two datasets to detect the consistency of the subsets. Then principal component analysis (PCA) was used to verify the distribution of the above metabolic gene mRNA expression data.

2.3 Characterization of DLBCL subclasses

Mf1type1 was set as the experimental group and mf2type2 as the control group. The difference of mRNA expression between the two subtypes was analyzed by limma package. The expression differential genes were obtained according to the standard of log fold change ($\log FC$) > 1.5 , $P < 0.05$. Then, the KEGG and go enrichment analysis of the differentially expressed genes were carried out by using the cluster profiler package, and the enrichment analysis results were obtained by selecting $p.adjust < 0.05$ as the standard.

2.4 Gene Set Enrichment Analysis(GSEA)

According to the results of limma package, the genes were sorted according to the size of $\log FC$. Taking the c2.cp.kegg.v7.0.symbols file downloaded from the GSEA official website as a reference, GSEA analysis was performed with GSEABase R package, and enrichment pathways were screened according to the criteria of $p.adjust < 0.05$. Because nm1type2 was taken as the control group, the positive value of enriched normalized enrichment score (NES) was regarded as the specific metabolic pathway of nm1type1, and the negative value of enriched NES was regarded as the specific metabolic pathway of nm1type2.

2.5 Evaluation of immune infiltration

CIBERSORT was used to calculate the infiltration score of immune cells in the sample. According to the threshold value of $P < 0.05$, the sample with credible infiltration results. Wilcoxon.test was used to compare the infiltration scores of immune cells in the two groups. Based on current drug inhibitors in clinical trials or drugs that have been approved for specific cancer types, 10 potential targeted immune checkpoint genes selected(CTLA-4/CCL2/ CD276/CD4/CXCR4/IL1A/IL6/LAG3/PDL1(CD274)/ TGFB1). The t test was used to compare the differences in the expression of immune checkpoint genes between the two metabolic subtypes[22].

2.6 Comparison of clinical characteristics

GSE10846 with complete clinical information was selected for clinical feature analysis. Chi-square test was used to compare the clinical characteristics of different subtypes.

2.7 Prognostic analysis of the two subtypes

The prognosis of the two subtypes was compared by log-rank test in each data set and combined data of five data sets. ABC subtype samples were extracted from the combined data of five data sets and this data were used to compare the prognosis of the two metabolic subtypes in the Patients with ABC subtype. The same procedure was performed in patients with GCB subtype. The same analysis was performed on GSE31312 to verify the correlation between metabolic subtypes and prognosis.

2.8 Comparison of metabolic subtypes and GEP subtypes

The distribution of GEP subtypes in the two metabolic subtypes was compared with ggstatsplot package. In the combined data of five data sets, the metabolic subtype prognostic model, GEP subtype prognostic model, and combined prognosis model were constructed by using the coxh function of survival package, The concordance index (c-index), The Akaike information criterion(AIC) value and log-rank χ^2 value of the three models were calculated, and the predictive ability of the three models for prognosis was compared through the three values. The log rank χ^2 value is larger, which indicates that the prediction ability of the model is better. C-index is used to quantify the prediction performance of the model, and AIC is used to compare the model fitting[23].

2.9 Classifier generation and performance verification

34 genes which play a major role in NMF clustering analysis were selected to establish a subtype prediction classifier. The 34- gene-Classifer was used to predict the metabolic subtypes of GSE31312 dataset. The consistency between the prediction results and those obtained by NMF before GSE 31312 was compared.

2.10 Statistical analysis methods:

In this article, R x64 3.5.2 version was used for calculation and statistical analysis. Unpaired t-test was used to compare two groups with normally distributed variables. wilcox.test was used to compare two groups with non-normally distributed variables. Contingency table variables were analyzed by Chi-square test or Fisher's exact test. The Kaplan-Meier method was used for survival analysis, and the log-rank test was used for prognostic comparison. A univariate Cox proportional hazards regression model was used to estimate the hazard ratios for univariate analyses. A two-tailed P value < 0.05 was statistically significant.

Results

3.1 Identify two metabolism-related subtypes according to metabolic genes

A flowchart was developed to systematically describe our study (Fig. 1A). The Canonical Pathways gene set from the KEGG database were downloaded from the GSEA website(<https://www.gsea-msigdb.org/>). According to this table, the key words of "metabolism" were used as screening conditions to extract the genes related to metabolic pathways, and 948 metabolism-related genes were obtained(Table S1). Among them, 883 genes were expressed in the five data sets. The expression data of these 883 metabolic genes were extracted from the 5 data sets, and The ComBat function of the sav package was used to remove the batch effect(Figure S2), and the merge function merges the data, The metadata was used for subsequent non-negative matrix decomposition (NMF) clustering.

Before NMF is executed, the filtering process is performed. MAD function of DescTools package was used to calculate MAD of each gene in the combined data of 5 data sets. 778 eligible genes were selected according to $MAD > 0.5$. Then, Cox regression was performed to evaluate the association between all candidate genes and overall survival (OS) by using r-package "survival". 106 metabolic genes with strong prognosis were obtained (Table S2). Finally, 106 genes with high variability ($MAD > 0.5$) and significant prognostic value ($P < 0.001$) were used for sample clustering.

Using the NMF function of the NMF package, according to the expression profiles of the 106 candidate genes mentioned above, the NMF clustering was performed on the combined sample set of five data sets. The K-value at which the cophenetic coefficient begins to decline was selected as the optimal decomposition point, and $k = 2$ was finally obtained as the optimal clustering number.[20] (Fig. 1B). The 34 genes that made major contributions to the NMF were obtained via the NMF package's extractFeatures function.("NNMT", "PTGDS", "HNMT", "ASAH1", "ENPP2", "DPYD", "OAT", "ALDH2", "DEGS1", "TD02", "KYNU", "PLA2G7", "GUCY1A1", "PAPSS2", "GUCY1B1", "PLPP3", "CMPK2", "PIK3CG", "PTGIS", "KMO", "PLPP1", "MTMR6", "AKR1C3", "PAPSS1", "PRPS2", "NMRK1", "NANS", "IMPDH2", "PAICS", "SRM", "PFKP", "GLO1", "POLD2", "PNP"). NMF analysis was performed again with the major contributing genes, and two subclasses were obtained through NMF, which were 469 cases of nmftype1 and 273 cases of nmftype2.(Figure S3) (Table S3) In order to verify the distribution of subclasses, we also performed PCA to reduce the dimension of features, and found that the subtypes are basically consistent with the two-dimensional PCA distribution pattern (Fig. 1C). Two subtypes were obtained by performing NMF typing on GSE31312(Figure S4). The submap of GENEPATTER was used to determine whether the subclasses identified in the two datasets were related. The results show that the nmftype1 and nmftype2 subclasses in the metadata set are highly correlated with the corresponding subclasses in GSE31312. (Fig. 1D) (Table S4)

3.2 Correlation of the DLBCL subclasses with metabolism-associated signatures

In this study, NMF classification was based on metabolism-related gene expression levels, for which we further studied the metabolic characteristics of the two subclasses. Nmftype1 was set as the

experimental group and nmftype 2 as the control group. The expression data of the two subtypes were compared by limma package, The gene logFC sequence was obtained by comparing the gene expression difference between the two subgroups. We downloaded C2.CP. Kegg. V7.0.symbols document from GSEA's official website as a reference and used GSEABase R package for GSEA analysis [24]. According to the standard of $\text{adjust.p} < 0.05$, 47 enrichment pathways were obtained, Among them, there are 6 pathways related to metabolism, If NES was positive, the pathway was considered as nmftype1 specific metabolic pathway. If NES was negative, the pathway was considered as nmftype2 specific metabolic pathway. It was found that there was a specific metabolic pathway in nmftype1, namely pyrimidine related metabolic pathway. There are 5 specific metabolic pathways in nmftype2, including the metabolic process of CYTOCHROME P450, ARACHIDONIC ACID, DRUG METABOLISM CYTOCHROME_P450, TRYPTOPHAN, etc. (Fig. 1E) (Table S5)

3.3 Transcriptome comparison of the two subtypes

The difference of mRNA expression between the two subtypes was analyzed according to limma package, 556 differentially expressed genes (482 down-regulated genes and 74 up-regulated genes) were obtained with the threshold of $\log\text{FC} > 1$ and $P < 0.05$ (Table S6). Up-regulated gene sets were regarded as nmftype1 specific gene set, and down-regulated gene sets were regarded as nmftype2 specific gene set. Then the ClusterProfiler package was used to analyze KEGG and GO enrichment of up-regulated and down-regulated genes respectively, and the enrichment analysis results were obtained using $P\text{Adjust} < 0.05$ as the screening criteria. The enrichment analysis results were obtained using $\text{adjust.P} < 0.05$ as the screening criteria (Figure S5). GO enriched nmftype1 was associated with genetic material. Nmftype2 is related to extracellular matrix. The results of KEGG enrichment showed that nmftype1 had no specific enrichment, and the first three specific enrichment pathways of nmftype2 were ECM-receptor interaction, Focal adhesion, PI3K-Akt signaling pathway. (Table S7)(Table S8)

3.4 Evaluation of immune infiltration

CIBERSORT was used to obtain the infiltration score of immune cells in the samples (Table S9). Samples with credible infiltration results were selected according to the threshold value of $P < 0.05$, and the number of samples changed from 742 to 737 (nmftype 1 samples changed from 469 to 466, nmftype 2 samples changed from 273 to 271). The corresponding heat map of immune cell infiltration was shown in the figure (Fig. 2A). Wilcox.test was used to compare the infiltration scores of each immune cell in the two groups, There were significant differences between the two groups in infiltration of multiple immune cells. B-cell infiltration of nmftype1 subtype was significantly higher than that of nmftype2 subtype, while T-cell and macrophage infiltration of nmftype2 subtype was higher than that of nmftype1 subtype (Fig. 2B). The relationship between subtypes and gene expression of 10 potential targeted immune checkpoints was further investigated (CTLA-4/CCL2/ CD276/CD4/CXCR4/IL1A/IL6/LAG3/

PDL1(CD274)/ TGFBI). These genes were selected based on current drug inhibitors in clinical trials or drugs that have been approved for specific cancer types. The results showed that there were differences among the 9 immune targets, among which 7 genes (CTLA-4/CCL2/ CD276/ CXCR4/ IL6/LAG3/PDL1 (CD274)) were highly expressed in nmftype2 compared with nmftype1 ($P < 0.05$). The 2 immune checkpoint genes of nmftype1 had higher expression levels than nmftype2. (Fig. 2C)

3.5. Relationship between subtypes and clinical characteristics

To further study the relationship between subtypes and clinical features. GSE10846 with complete clinical information was selected as the case set for clinical feature analysis (Table 2). Chi-square test was used to compare the clinical characteristics of different subtypes. It was found that Extranodalsites ($P = 0.040$) and GEP subtype($P < 0.001$) were correlated with the metabolic subtype in the GSE10846 dataset, but not with age($P = 0.779$) gender ($P = 0.536$), stage($P = 1$), ECOG($P = 0.372$) in the GSE10846 dataset. (Fig. 3A)

Table 2
Relationship between GSE10846 subtypes and clinical data

Characteristic	GSE10846	nmfsubtype1	nmfsubtype2	χ^2	P
Age					
≤60	188	113	75		
> 60	226	140	86	0.079	0.779
Sex					
Male	224	130	94		
Female	172	106	66	0.383	0.536
stage					
I-II	188	116	72		
I-II"	218	134	84	0	1
ECOG					
0-1	296	180	116		
≥ 2	93	62	31	0.798	0.372
Extranodalsites					
0-1	353	226	127		
≥ 2	30	13	17	4.201	0.040
subtype					
ABC	167	124	43		
GCB	183	95	88	17.66	<0.001

3.6 Comparison of prognostic characteristics of the two subtypes

significant prognostic difference was observed in combined data set (log-rank test $P < 0.0001$), the overall survival of nmfsubtype1 was worse than that of nmfsubtype2. And there were significant differences in the prognosis among the four data sets GSE10846(OS $P < 0.0001$) GSE11318(OS $P = 0.00085$) GSE53786(OS $P = 0.011$) GSE87371(OS $P < 0.016$). The combined samples were stratified by ABC and GCB, and the nmfsubtype1 suggested a worse prognosis than nmfsubtype2 in GCB samples ($P = 0.0049$), The same analysis was performed on GSE31312 data set, and consistent results were found in GSE31312 (overall data OS: $P = 0.00021$)(GCB data OS: $P < 0.0001$)(Fig. 3B)

3.7 Comparison of metabolic subtype classification and existing GEP subtype classification

Comparing the distribution of the two metabolic subtypes and GEP subtypes with ggstatplot package, it can be seen that ABC subtype accounts for 46%, GCB subtype accounts for 35% in nmftype1 subtype, ABC subtype accounts for 21% and GCB subtype accounts for 52% in nmftype2 subtype. The original unclassified subtype is also classified according to metabolic subtypes. (Fig. 4A). Three prognostic model indexes[23] were calculated by log-rank test and cox prognostic model. ABC and GCB subtypes: OS(Chisq = 45.6, p = 1e-11) cox prognostic model (C-index: 0.619 ± 0.018), AIC = 2352.155); nmftype1 and nmftype2 subtypes: OS (Chisq = 23.8, P = 1E-06) COX prognostic model (C-index: 0.574 ± 0.015 ,AIC = 3054.996); The two subtypes were combined to obtain four groups of ABC-nmftype1 and ABC-nmftype2, GCB-nmftype1 and GCB-nmftype2: OS (Chisq = 50.7, p = 6e-11)cox prognostic model(C-index: 0.637 ± 0.018 ,AIC = 2347.423). According to the results, the prediction ability of the nmftype subtype was weaker than that of the ABCGCB subtype. However, the new classification method combined with the existing subtype classification method had better prediction effect than the existing ABCGCB subtype. (Fig. 4B).

3.8Building a gene classifier

In order to build a simplified classifier for clinical use, 34 genes which contributed to NMF clustering were selected to construct the gene classifier. (Fig. 4C)(Table S10). This gene classifier was used to predict the metabolic subclass of GSE31312 and the predicted grouping results of this classifier were compared with the original NMF classification, (Table S11) We observed a consistency of 73.2% for nmftype1 subclass and 86.9% for nmftype2 subclass. The results indicate that the DLBCL classification can be determined repeatably with 34 gene signatures. (Fig. 4B).

Discussion

Although several DLBCL classifications based on gene expression and mutation have been proposed in recent years, there are still problems such as insufficient sample size, complexity, and inability to classify samples completely. In this study, five DLBCL data sets (GSE10846, GSE11318, GSE53786, GSE87371 and GSE23501) from the GEO database GPL570 platform were combined to obtain 742 DLBCL samples, and two metabolic subtypes, nmftype1 and nmftype2 were identified. The study found that nmftype2 had obvious metabolic characteristics, high expression of immune checkpoint gene, mainly infiltration of T cells and macrophages, and a good prognosis. The metabolic signal of nmftype1 was lower than that of nmftype2, most of the immune checkpoint genes were low in expression, and b-cell infiltration was predominant, indicating a worse prognosis. In general, this study classified DLBCL from the perspective of metabolism, and obtained two clusters with active and failing metabolic activities. Moreover, further analysis of metabolic subtypes found that DLBCL has different characteristics of immune infiltration and prognosis.

In 2005, Stefano Monti et al. obtained oxidative phosphorylation(OxPhos), B-cell receptor/proliferation(BRP) and host response (HR) three subclasses by clustering DLBCL gene expression profiles. [13] BCR-DLBCLs are characterized by glycolysis. OxPhos DLBCLs are dependent on

mitochondrial energy transfer and nutrient utilization. [25] In this study, two subtypes with different metabolic characteristics were also found. Nmftype1 showed low metabolic activity. There was only one pathway related to the metabolism of pyrimidine, and nmftype2 showed high metabolic activity. The abundance of metabolic characteristics indicates that nmftype2 patients are more likely to benefit from metabolic therapy. For example, RNA interference or drug interference with tigecycline (Tigecyl), an FDA approved inhibitor, has selective toxicity to OxPhos-DLBCL cell line tumors. [25] In 2019, Johanna chiche et al. Found that DLBCL with low GAPDH expression was metabolized by OXPHOS and depended on mTORC1 signal transduction and glutamine decomposition. Lymphoma with low GAPDH expression showed poor response to R-CHOP treatment and was sensitive to mitochondrial metabolic inhibitors. [26] These studies provide insights into the prediction of potential responses to metabolic therapy. The identification of two subtype specific metabolic pathways is expected to provide metabolic therapy for some metabolic processes, providing an alternative for chemotherapy resistant patients.

In complex tumor microenvironment, metabolic reprogramming is mainly guided by the serine/threonine kinase mTOR (mammalian target of rapamycin). [27] The enrichment of nmftype1 by GO was related to genetic material, Nmftype2 is associated with extracellular matrix. The results of KEGG enrichment showed that nmftype1 had no specific enrichment, and the first three specific enrichment pathways of nmftype2 were ECM-receptor interaction, Focal adhesion, PI3K-Akt signaling pathway. Focal adhesion signal inhibitor E7123 can induce caspase dependent cell death in DLBCL. [28] The efficacy of E7123 in the treatment of DLBCL has been verified in a mouse model of diffuse large B-cell lymphoma with central nervous system involvement. [29] ROR1 significantly promotes the tumorigenesis of DLBCL by regulating PI3K / Akt / mTOR signaling pathway. Targeting ROR1 may provide a promising strategy for the treatment of DLBCL. [30] In rare cases, patients with relapsed and refractory DLBCL have achieved complete response, which suggests that some B-lymphomas may be extremely sensitive to rapalogs, possibly because they show active mTORC1 signal transduction [27]. focal adhesion signal inhibitor E7123 and targeted ROR1 may be therapeutic strategies for nmftype2, but may not be effective for nmftype1. When PI3K and Akt inhibitors were used in cell lines and mouse models, the phosphatidylinositol-3-kinase (PI3K) α/δ (PI3K α/δ) Inhibitor AZD8835 showed significant effect in ABC DLBCL model, while azd5363, a protein kinase B (Akt) inhibitor AZD5363 induced apoptosis in PTEN deficient DLBCLs. [31] the enrichment of nmftype2 on PI3K Akt signaling pathway suggests that nmftype2 is more likely to benefit from the treatment of PI3K and Akt inhibitors. The discovery of two specific pathways of DLBCL metabolic subtypes can reveal the molecular pathogenesis and potential heterogeneity of the two DLBCL metabolic subtypes, and provide a reference for the development of new therapeutic schemes and the screening of clinically beneficial patients.

At present, many immunotherapies have shown encouraging results in patients with recurrent or refractory DLBCL [32]. Metabolic changes in tumor microenvironment (TME) can inhibit anti-tumor immunity such as immune cell infiltration by producing immunosuppressive metabolites. Metabolic disorders of cancer cells will further affect the expression of cell surface markers, thus interfering with immune monitoring [33]. the two metabolic subtypes showed different characteristics of immune infiltration, among which nmftype1 subtype showed significant B cell infiltration, and nmftype2 subtype T

cell and macrophage infiltration were significant. Insufficient infiltration of T cells and / or NK cells indicates low survival rate[34], which may be related to the poor prognosis of nmftype1. Seven potential immunotherapeutic targets or immune checkpoint gene of high metabolic nmftype2 subtype were higher than those of low metabolic nmftype1 (CTLA-4 / CCL2 / PD-1 (CD276) / CXCR4 / / IL6 / Lag3 / PDL1 (CD274)) ($P < 0.05$). The expression level of IL1A / TGFB in nmftype1 was higher than that of nmftype2 subtypes ($P < 0.05$). High expression of PD-L1 / PD-L2 in DLBCL cells is associated with good prognosis[35], which is consistent with our conclusion. Because tumor microenvironment promotes immune escape, the clinical benefits of immunosuppression are still small. [36] the combination of immunotherapy and antimetabolism therapy may change this situation. previous studies have shown that the targeted blocking effect of CD73 significantly enhances the therapeutic activity of anti PD-1 and anti CTLA-4 monoclonal antibodies, which confirms this conclusion. [37] the two metabolic subtypes have different potential targets for immunotherapy, which provides guidance for the combination of immunotherapy and antimetabolic drugs. Nmftype2 is characterized by high metabolism and high expression of immunotherapy target, which is more likely to benefit from immunotherapy and metabolic therapy. The high expression of IL1A / TGFB1 in nmftype1 may provide possibilities to improve the poor prognosis of nmftype1 subtypes.

In order to further guide the clinical practice, we also studied the relationship between the classification of the two subtypes and the clinical features and prognosis. Through the analysis of geo10846 data set, we found that the two subtypes were only correlated with Extranodalsites ($p = 0.040$) and ABCGCB subtype ($p < 0.001$). Considering the incomplete clinical feature collection, this problem can be further studied. The results showed that the classification of metabolic subtypes was closely related to prognosis (OS). Nmftype1 showed worse prognosis (OS). In the stratified analysis of ABC and GCB subtypes, nmf1 subtype still indicated worse prognosis (OS) in GCB samples ($P = 0.0049$), and was verified in GSE31312. Compared with the original GEP subtype distribution, this classification method can better define the unclassified subtypes. The model method for distinguishing metabolic subtypes is not a simple repetition of GCB and ABC classification, and has new clinical significance. Although the ability of metabolic subtypes to predict prognosis is weaker than ABCGCB subtypes, the combination of the metabolic model classification and the original ABCGCB classification has a better predictive effect on the prognosis than ABCGCB alone. GCB-nmftype2, GCB-nmftype1, ABC-nmftype2 and ABC-nmftype1 in turn indicate the possibility of worse prognosis. With this model, the existing subtypes that can not be classified can be defined, the heterogeneity of ABCGCB can be solved to a certain extent, and a better prognosis prediction model than the existing ABCGCB subtypes is established. In order to further simplify and guide the clinical application of the metabolic classification model, we constructed 34 gene classifiers to predict which metabolic subtype the samples belong to. By verifying the metadata set itself, it is found that the grouping result of the classifier is consistent with that of the original NMF classification method.

In the existing studies defining DLBCL subtypes, the sample size of this study is at the forefront [3]. This study provides a new insight into the heterogeneity of DLBCL from the perspective of metabolism. Two metabolic subtypes with different characteristics in transcription, metabolism, immunity and prognosis have been identified. We provide data on the role of immune landscape in DLBCL and how Metabolism-

Related DLBCL subtypes affect the cellular composition and immune function of tumor microenvironment. New knowledge may lead to promising treatments in the near future. A new prognostic model combining metabolic subtypes and ABCGCB subtypes has been proposed for the first time, and the prediction ability of this model for DLBCL is better than the existing ABCGCB subtypes. Finally, we studied the 34 gene simplified classifier to predict the metabolic subtype model, which provides the possibility for clinical application. This study provides insights into the potential response to metabolic therapy. By revealing the characteristics of metabolism, malignant tumor molecular subsets with unique biomarkers can be identified, which brings opportunities for new treatment opportunities. Through better understanding of metabonomic changes, these targets can be transformed into drug treatable tumor metabolites and developed anti metabolic therapy for high-risk DLBCL.

Conclusions

In conclusion, we classified DLBCL by metabolic genes and got a new typing method. This classification has prognostic significance and is helpful to guide the individualized application of immunotherapy and metabolic therapy. This study provides a new understanding of the heterogeneity of DLBCL from the perspective of metabolism.

Abbreviations

DLBCL Diffuse large B-cell lymphoma

NHL Non-Hodgkin's lymphoma

GEO Gene Expression Omnibus

GSE gene expression data series

OS overall survival

NHL Non-Hodgkin lymphoma

R-CHOP rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone

CN copy number

GCB Germinal center B-cell like

ABC activated B-cell

OxPhos oxidative phosphorylation

HR host response

BCR B cell receptor

RNA Ribonucleic acid

COO cell-of-origin

GEP gene expression profiles

NMF non-negative matrix factorization

GSEA Gene Set Enrichment Analysis

PCA Principal components analysis

LogFC Log Fold Change

KEGG Kyoto Encyclopedia of Genes and Genomes

GO Gene Ontology

NES normalized enrichment score

Cox cox proportional-hazards model

C-index the concordance index

AIC The Akaike information criterion

MAD median absolute deviation

GAPDH Glyceraldehyde-3-phosphate dehydrogenase

PI3K The phosphatidylinositol 3-kinase

ROR1 Receptor Tyrosine Kinase-like Orphan Receptor-1

ECM extracellular matrix

CYP cytochrome P450

mTOR mammalian target of rapamycin

TME tumor microenvironment

PI3K α/δ phosphatidylinositol-3-kinase α/δ

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and material: The datasets generated and analysed during the current study are available in the GEO repository(<https://www.ncbi.nlm.nih.gov/geo/>)

Competing interests: The authors declare that they have no competing interests.

Funding: This work has been supported by the Natural Science Foundation of Guangxi Province (2016GXNSFDA380029)

Code availability : R code to reproduce all analyses and results presented here.

Authors' contributions: YingDuan were responsible for data analysis, interpretation, mapping and drafting of manuscripts. Luo Yanzhen is responsible for screening and collating the data. Hongcen supervised the whole analysis and provided guidance and instructions. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge Natural Science Foundation of Guangxi Province for the grant to carry out research work (2016GXNSFDA380029).

References

1. Smith A, Howell D, Patmore R, Jack A, Roman E. Incidence of haematological malignancy by sub-type: a report from the Haematological Malignancy Research Network. *British journal of cancer*. 2011;105(11):1684-92.
2. Roman E, Smith AG. Epidemiology of lymphomas. *Histopathology*. 2011;58(1):4-14.
3. Dubois S, Jardin F. Novel molecular classifications of DLBCL. *Nat Rev Clin Oncol*. 2018;15(8):474-6.
4. Reddy A, Zhang J, Davis NS, Moffitt AB, Love CL, Waldrop A, et al. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell*. 2017;171(2):481-94.e15.
5. Chapuy B, Stewart C, Dunford AJ, Kim J, Kamburov A, Redd RA, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nature medicine*. 2018;24(5):679-90.
6. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *The New England journal of medicine*. 2018;378(15):1396-407.
7. Lenz G, Wright G, Dave SS, Xiao W, Powell J, Zhao H, et al. Stromal gene signatures in large-B-cell lymphomas. *The New England journal of medicine*. 2008;359(22):2313-23.

8. Lenz G, Staudt LM. Aggressive lymphomas. *The New England journal of medicine*. 2010;362(15):1417-29.
9. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403(6769):503-11.
10. Li S, Young KH, Medeiros LJ. Diffuse large B-cell lymphoma. *Pathology*. 2018;50(1):74-87.
11. Cantor JR, Sabatini DM. Cancer cell metabolism: one hallmark, many faces. *Cancer discovery*. 2012;2(10):881-98.
12. Pavlova NN, Thompson CB. The Emerging Hallmarks of Cancer Metabolism. *Cell metabolism*. 2016;23(1):27-47.
13. Caro P, Kishan AU, Norberg E, Stanley IA, Chapuy B, Ficarro SB, et al. Metabolic signatures uncover distinct targets in molecular subsets of diffuse large B cell lymphoma. *Cancer cell*. 2012;22(4):547-60.
14. Monti S, Savage KJ, Kutok JL, Feuerhake F, Kurtin P, Mihm M, et al. Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood*. 2005;105(5):1851-61.
15. Shaknovich R, Geng H, Johnson NA, Tsikitas L, Cerchietti L, Greally JM, et al. DNA methylation signatures define molecular subtypes of diffuse large B-cell lymphoma. *Blood*. 2010;116(20):e81-9.
16. Scott DW, Wright GW, Williams PM, Lih CJ, Walsh W, Jaffe ES, 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(8):1214-7.
17. Dubois S, Viailly PJ, Bohers E, Bertrand P, Ruminy P, Marchand V, et al. Biological and Clinical Relevance of Associated Genomic Alterations in MYD88 L265P and non-L265P-Mutated Diffuse Large B-Cell Lymphoma: Analysis of 361 Cases. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2017;23(9):2232-44.
18. Lenz G, Wright GW, Emre NC, Kohlhammer H, Dave SS, Davis RE, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(36):13520-5.
19. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC bioinformatics*. 2010;11:367.
20. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(12):4164-9.
21. Visco C, Li Y, Xu-Monette ZY, Miranda RN, Green TM, Li Y, et al. Comprehensive gene expression profiling and immunohistochemical studies support application of immunophenotypic algorithm for molecular subtype classification in diffuse large B-cell lymphoma: a report from the International DLBCL Rituximab-CHOP Consortium Program Study. *Leukemia*. 2012;26(9):2103-13.
22. Yang C, Huang X, Liu Z, Qin W, Wang C. Metabolism-associated molecular classification of hepatocellular carcinoma. *Molecular oncology*. 2020;14(4):896-913.

23. Weiss A, Chavez-MacGregor M, Lichtensztajn DY, Yi M, Tadros A, Hortobagyi GN, et al. Validation Study of the American Joint Committee on Cancer Eighth Edition Prognostic Stage Compared With the Anatomic Stage in Breast Cancer. *JAMA oncology*. 2018;4(2):203-9.
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(43):15545-50.
25. Norberg E, Lako A, Chen PH, Stanley IA, Zhou F, Ficarro SB, et al. Differential contribution of the mitochondrial translation pathway to the survival of diffuse large B-cell lymphoma subsets. *Cell death and differentiation*. 2017;24(2):251-62.
26. Chiche J, Reverso-Meinietti J, Mouchotte A, Rubio-Patiño C, Mhaidly R, Villa E, et al. GAPDH Expression Predicts the Response to R-CHOP, the Tumor Metabolic Status, and the Response of DLBCL Patients to Metabolic Inhibitors. *Cell metabolism*. 2019;29(6):1243-57.e10.
27. Ricci JE, Chiche J. Metabolic Reprogramming of Non-Hodgkin's B-Cell Lymphomas and Potential Therapeutic Strategies. *Frontiers in oncology*. 2018;8:556.
28. Bosch R, Dieguez-Gonzalez R, Céspedes MV, Parreño M, Pavón M, Grañena A, et al. A novel inhibitor of focal adhesion signaling induces caspase-independent cell death in diffuse large B-cell lymphoma. *Blood*. 2011;118(16):4411-20.
29. Bosch R, Moreno MJ, Dieguez-Gonzalez R, Céspedes MV, Gallardo A, Trias M, et al. A novel orally available inhibitor of focal adhesion signaling increases survival in a xenograft model of diffuse large B-cell lymphoma with central nervous system involvement. *Haematologica*. 2013;98(8):1242-9.
30. Mao Y, Xu L, Wang J, Zhang L, Hou N, Xu J, et al. ROR1 associates unfavorable prognosis and promotes lymphoma growth in DLBCL by affecting PI3K/Akt/mTOR signaling pathway. *BioFactors* (Oxford, England). 2019;45(3):416-26.
31. Erdmann T, Klener P, Lynch JT, Grau M, Vočková P, Molinsky J, et al. Sensitivity to PI3K and AKT inhibitors is mediated by divergent molecular mechanisms in subtypes of DLBCL. *Blood*. 2017;130(3):310-22.
32. Zhang J, Medeiros LJ, Young KH. Cancer Immunotherapy in Diffuse Large B-Cell Lymphoma. *Frontiers in oncology*. 2018;8:351.
33. Li X, Wenes M, Romero P, Huang SC, Fendt SM, Ho PC. Navigating metabolic pathways to enhance antitumour immunity and immunotherapy. *Nat Rev Clin Oncol*. 2019;16(7):425-41.
34. Xu-Monette ZY, Xiao M, Au Q, Padmanabhan R, Xu B, Hoe N, et al. Immune Profiling and Quantitative Analysis Decipher the Clinical Role of Immune-Checkpoint Expression in the Tumor Immune Microenvironment of DLBCL. *Cancer immunology research*. 2019;7(4):644-57.
35. Atanackovic D, Luetkens T. Biomarkers for checkpoint inhibition in hematologic malignancies. *Seminars in cancer biology*. 2018;52(Pt 2):198-206.
36. Renner K, Singer K, Koehl GE, Geissler EK, Peter K, Siska PJ, et al. Metabolic Hallmarks of Tumor and Immune Cells in the Tumor Microenvironment. *Frontiers in immunology*. 2017;8:248.

37. Allard B, Pommey S, Smyth MJ, Stagg J. Targeting CD73 enhances the antitumor activity of anti-PD-1 and anti-CTLA-4 mAbs. Clinical cancer research : an official journal of the American Association for Cancer Research. 2013;19(20):5626-35.

Figures

Figure 1

Identifying DLBCL subclasses using the NMF consensus cluster in the metadata set .(A) Research flow chart. (B) NMF clustering was performed using 106 metabolism-related prognostic genes. The Cophenetic correlation coefficient of K 2-6 is shown. (C) The two DLBCL subclasses were verified in PCA analysis. (D) The subclass correlation obtained by executing NMF in GSE31312 and the subclass correlation obtained by executing NMF in metadata dataset was compared.(E) Specific metabolic pathways were obtained by the analysis of two subclasses of GSEA

Figure 2

Immune characteristics of two subclasses in the metadata dataset. The violin figure of the abundance of immune distinguished by different subclasses, The statistical difference was compared through the wilcox.test, the P values are labeled above each the violin figure. (C) Expression levels of 10 immune checkpoint genes in two DLBCL subclasses (normalized count).Wilcox. test was used to compare the statistical differences. P values were marked with an asterisk above each boxplot (NS means no significance, *P < 0.05, **P < 0.01, ***P < 0.001).

GSE10846

GSE11318

GSE53786

GSE87371

overall survival

nmftype1

nmftype2

p < 0.0001

Number at risk

nmftype1

nmftype2

Time(month)

overall survival

nmftype1

nmftype2

p = 0.00085

Number at risk

nmftype1

nmftype2

Time(month)

overall survival

nmftype1

nmftype2

p = 0.011

Number at risk

nmftype1

nmftype2

Time(month)

overall survival

nmftype1

nmftype2

p = 0.016

Number at risk

nmftype1

nmftype2

Time(month)

overall survival

nmftype1

nmftype2

p = 0.41

Number at risk

nmftype1

nmftype2

Time(month)

overall survival

nmftype1

nmftype2

p < 0.0001

Number at risk

nmftype1

nmftype2

Time(month)

overall survival

nmftype1

nmftype2

p = 0.0049

Number at risk

nmftype1

nmftype2

Time(month)

overall survival

nmftype1

nmftype2

p = 0.66

Number at risk

nmftype1

nmftype2

Time(month)

overall survival

nmftype1

nmftype2

p = 0.00021

Number at risk

nmftype1

nmftype2

Time(month)

overall survival

nmftype1

nmftype2

p = 5e-04

Number at risk

nmftype1

nmftype2

Time(month)

overall survival

nmftype1

nmftype2

p = 0.3

Number at risk

nmftype1

nmftype2

Time(month)

Figure 3

(A) clinical features of the DLBCL subclass in the GSE10846 dataset (B) the differences in OS between the two subtypes (nmftype1, nmftype2) in the metadata set and five separate datasets(GSE23501 GSE53786 GSE10846 GSE87371 GSE11318)were verified by GSE31312, and the statistical significance of the differences was determined by logarithmic rank test.

Figure 4

(A) distribution of GEP subtypes in the two metabolic subtypes. The predictive ability of the three prognostic models(metabolic subtype, GEP subtypes, joint metabolic subtypes and GEP subtypes) for OS was compared. Chi-square value was calculated by log-rank test, and C-index and AIC were calculated by cox prognostic model. (C) Heatmap of gene classifier expression level of 34 and identified subtypes in GSE31312.(D) Concordance of DLBCL molecular subclass prediction between the 34-gene classifier and original prediction based on NMF in GSE31312.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig.S1.tif](#)
- [Fig.S2.tif](#)
- [Fig.S3.tif](#)
- [Fig.S4.tif](#)
- [Fig.S5.tif](#)
- [TableS1S11.xlsx](#)