Cerebrospinal Fluid Metabolic Markers Predict Prognosis behavior of Primary CNS Lymphoma with HD-MTX Based Chemotherapeutic Treatment

Liying Zhou1,*, Qing Li2,*, Jingshen Xu1, Shuaikang Wang1, Zhiqiang Song1,4, Xinyi Chen1, Yan Ma2, Zhigang Lin2, Bobin Chen2, He Huang1,3,5*

1. Shanghai Key Laboratory of Metabolic Remodeling and Health, Institute of Metabolism and Integrative Biology, Fudan University, Shanghai, 200438, China;
2. Department of Hematology, Huashan Hospital, Fudan University, Shanghai, 200438, China;
3. Shanghai Qi Zhi Institute, Shanghai, 200030, China;
4. School of Life Sciences, Inner Mongolia University, Hohhot Inner Mongolia, 010021, China;
5. Lead contact;

* These authors contribute equally

# Correspondence: he_huang@fudan.edu.cn (H.H.), and bbchen@fudan.edu.cn (B.C)

Corresponding author: He Huang

Mailing Add: 2005 Songhu Road, Shanghai City, China, 200438

Phone: 0086-130 5237 5779

Fax: 021-3124 2086

he_huang@fudan.edu.cn

© The Author(s) 2022. Published by Oxford University Press, the Society for Neuro-Oncology and the European Association of Neuro-Oncology. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Corresponding author 2: Bobin Chen

Mailing Add: 108 Luxiang Road, Huashan Hospital, Shanghai City, China, 200438

Phone: 0086-134 0218 6236

Fax: 021-5288 9999

bbchen@fudan.edu.cn
Abstract

**Background:** Primary central nervous system lymphoma (PCNSL) is a highly aggressive non-Hodgkin’s B-cell lymphoma which normally treated by a high-dose methotrexate (HD-MTX)-based chemotherapy. However, such treatment cannot always guarantee a good prognosis outcome while suffering several side effects. Thus, biomarkers or biomarker-based model that can predict PCNSL patient prognosis would be benefit.

**Methods:** We firstly collected 48 patients with PCNSL and applied HPLC-MS/MS based metabolomic analysis on such retrospective PCNSL patient samples. We then selected the highly dysregulated metabolites to build a logical regression model that can distinguish the survival time length by a scoring standard. Finally, we validated the logical regression model on a 33-patient prospective PCNSL cohort.

**Results:** Six metabolic features were selected from the cerebrospinal fluid (CSF) that can form a logical regression model to distinguish the patients with relatively good prognosis (Z score ≤ 0.06) from the discovery cohort. We applied the metabolic marker-based model onto a prospective recruited PCNSL patient cohort for a further validation, and the model preformed nicely on such validation cohort (AUC = 0.745).

**Conclusions:** We developed a logical regression model based on metabolic markers in CSF that can effectively predict PCNSL patient prognosis before the HD-MTX based chemotherapy treatments.

Key words: metabolomics, primary CNS lymphoma, high-dose methotrexate, prediction model.
Key points:

1. Cerebrospinal fluid metabolites are associated with prognostic outcome of PCNSL patients received HD-MTX based chemotherapy.

2. The prospective study validates the metabolic marker-based model for the prognostic outcome prediction before the HD-MTX based treatment.

3. Our results provide an applicable model that predict the prognosis of HD-MTX based treatment using a scoring standard.

Importance of the Study

Primary central nervous system lymphoma (PCNSL) is a highly aggressive non-Hodgkin’s B-cell lymphoma which accounts for 3-5% of all primary brain tumors. High-dose methotrexate (HD-MTX)-based chemotherapy serving as the first-line treatment causes strong adverse effects in patients. Although several clinical and basic researches have investigated the way to predict the prognostic outcome of PCNSL patients before receiving the HD-MTX based treatment, none of the biomarkers or prediction models were evaluated by a prospective PCNSL cohort due to the rare incidence rate of such disease. In this study, we applied metabolomics technique to discover a metabolic feature-based prediction model from a retrospective PCNSL cohort, and validated our model using a prospective PCNSL cohort. Our results indicated the prognosis of PCNSL patient with HD-MTX based treatment would associate with CSF metabolism disorders.
Introduction

Primary central nervous system lymphoma (PCNSL) is a highly aggressive non-Hodgkin’s B-cell lymphoma type with a median overall survival (OS) around four years after diagnosis\(^1\). It accounts for 3-5\% of all primary brain tumors\(^2\) and 4-6\% of all extranodal lymphomas\(^3\). More than 90\% of PCNSLs are diffuse large B-cell lymphomas (DLBCL), while the remaining ones are T-cell, Burkitt, lymphoblastic and marginal zone lymphomas\(^2\). The incidence of PCNSL has increased in recent years, particularly in patients over the age of 60 years, and is now at an overall rate of 0.5 per 100,000\(^3\).

There is no consensus on the optimal treatment regimen for PCNSL, but high-dose methotrexate-based chemotherapy is widely used. High-dose chemotherapy (HDC) followed by autologous stem cell transplantation (ASCT) is the most recent options for consolidation therapy\(^4\). Despite advances in initial treatment, up to half of patients relapse and 10\% to 15\% have primary refractory disease\(^5\). The median overall survival has improved significantly from 2.5 months to 26 months over the past few decades\(^6\), but the 5-year survival rate of PCNSL patients is still very low, especially in the elderly and those who cannot tolerate high-dose chemotherapy. Considering strong toxicity of MTX, the high-dose methotrexate (HD-MTX)-based chemotherapy readily results in adverse effects in patients, including hepatorenal complications and hematological toxicities\(^7\). Thus, finding of biomarkers and clinical indexes that can predict the prognostic outcome of HD-MTX-based treated patients would be beneficial.

Several clinical prognostic models were developed, including the International Extranodal Lymphoma Study Group (IELSG) score\(^8\), the Nottingham-Barcelona score\(^9\), and the Memorial Sloan-Kettering Cancer Center (MSKCC) classification of PCNSL\(^10\). Although a variety of prognostic biomarkers of
PCNSL have been reported by previous studies\textsuperscript{11–18}, it is difficult to comprehensively reflect the overall characteristics of prognosis with only a single or a few biomarkers. The anti-cancer target of MTX is dihydrofolate reductase (DHFR) which can convert dihydrofolate to tetrahydrofolate. The inhibition of DHFR also perturbs the DNA synthesis and one carbon metabolism\textsuperscript{19,20}. It is naturally to hypothesize that metabolite alterations may be associated with the HD-MTX-based chemotherapy outcomes, and metabolic molecular markers can be a type of promising biomarker to predict the efficacy of HD-MTX-based chemotherapy in PCNSL patients.

In this work, we first performed a retrospective cohort study to identify global changes in metabolites and screen biomarkers for prognosis prediction, followed by a prospective cohort study to validate our findings. By integrating clinical characteristics, therapeutic regimens and metabolic markers, we identified a molecular model with prognostic value that is helpful for risk stratification and management of PCNSL patients.

\textbf{Methods}

\textit{Patients}

The discovery cohort consisted of 48 patients with PCNSL diagnosed in 2018 to 2020 at our institution. Diagnosis of all participants were confirmed based on WHO classification of tumors of hematopoietic and lymphoid tissues. Patients would be excluded if they had systemic lymphoma and immunodeficiency disease. The validation cohort included 35 patients prospectively recruited from 2020-2021 with the same inclusion/exclusion criteria and sample processing protocol and followed up to July 2022. Two samples were removed due to clinical information inadequateness, leaving 33
patients. Informed consent was obtained from all patients for the use of their samples, in accordance with the guidelines of the respective Ethical Committees on Human Research approved at our institution.

All patients received an HD-MTX-based chemotherapy regimen. Each HD-MTX treatment was administered as a 3-hour infusion. Pre-hydration and alkalinization were initiated at least 72 h before MTX administration. Standard leucovorin rescue was initiated 24 h after the start of HD-MTX infusion. Every 3 weeks, 8 cycles were repeated until tumor progression or toxicity occurred. The clinical features of all patients were collected from the medical records, including age, gender, height, weight, performance status, time of diagnosis, surgical resection, biopsy type, lesion site, number of lesions, HIV status, serum lactate dehydrogenase (LDH) level, etc. Magnetic resonance imaging (MRI) was used to assess the location and quantity of lesions in all patients. The continuity of treatment was evaluated by contrast-enhanced MRI scans every cycle and by PET-CT after 3 cycles of treatment and after all therapeutic procedures. PFS was calculated from the date of diagnosis to the date of disease progression, the first relapse, death from any cause, or last follow-up.

Sample collection

All patients underwent biopsy before CSF sampling and were diagnosed pathologically. Each PCNSL patient underwent a lumbar puncture to examine cerebrospinal fluid (CSF) at baseline before chemotherapy. An additional 2 mL of CSF was obtained, then centrifuged for 10 min to remove cells and frozen at -80°C until further use.
**Metabolite extraction**

The metabolomic approach was adopted from a published method\textsuperscript{21}. Centrifuge at 14,000×g for 10 min at 4°C. Methanol: water (v:v, 80:20) was prechilled at -80°C overnight, and 4.5 mL was added to 50 microliters cerebrospinal fluid. Mix and incubate at -80°C for 30 min. Centrifuge at 4°C with 4,000×g for 10 min, and the supernatant was then collected in another 15 mL centrifuge tube. The 80% methanol extracted metabolites were then dried using a SpeedVac (LABCONCO Refrigerated CentriVap Concentrator) and stored at -80°C before MS analysis.

**Targeted metabolomics analysis**

The metabolomic approach was adopted from a published method\textsuperscript{22}. In general, samples were resuspended in 50 μL of water: acetonitrile (v:v, 50:50), and 5 μL was injected into a 6500 QTRAP triple-quadrupole MS (SCIEX) coupled to an HPLC system (Shimadzu). Metabolites were eluted via hydrophilic interaction chromatography (HILIC) by using a 4.6-mm i.d. × 10 cm Amide XBridge column (Waters) with a flow rate of 400 μL/min using buffer A (20 mM ammonium hydroxide/20 mM ammonium acetate (pH 9.2) at a 95:5 ratio with water: acetonitrile) and buffer B (acetonitrile). Gradients were run from 85% buffer B to 42% buffer B at 0–5 min; from 42% buffer B to 0% buffer B at 5–16 min; 0% buffer B was held from 16–24 min; from 0% buffer B to 85% buffer B at 24–25 min; and 85% buffer B was held for 7 min. All ions were acquired by 306 selected reaction monitoring transitions in a positive and negative mode switching fashion. Electrospray ionization (ESI) voltage was +4,900 and −4,500 V in positive or negative mode, respectively. Data were analyzed by MetaboAnalyst software (https://www.metaboanalyst.ca/).
Data processing and model

All the CSF metabolomics data were first processed with Log10 transformed and normalized using median value using MetaboAnalyst software\textsuperscript{23}. The normalized data were then corrected batch effect by an R package; sva::ComBat\textsuperscript{24,25}. For the feature selection, we compared the mostly different metabolites in patient CSF before the HD-MTX based chemotherapy treatment, and ranked the detected metabolites based on their fold change (FC) between GP and PP groups. Six features were selected due to their Log2FC > 0.5 or Log2FC < -0.5. For the prediction model, we applied a logistics regression model onto the discovery cohort and test such model on the validation cohort. Comparing the p-values of different models using survival analysis\textsuperscript{26}, we selected the best performing logistics regression model as our prediction model.

Statistical analyses

The patients’ baseline characteristics were summarized using descriptive statistics, and descriptive analyses were conducted for all variables. The Chi-square, Fisher’s exact test, Mann–Whitney tests were used for statistical analyses. Survival curves were plotted by the Kaplan–Meier method and analyzed by the log-rank test. All tests were two-sided, and p<0.05 was taken as statistically significant. All statistical analyses were performed using Statistical Package for Social Science, version 26.0 (IBM SPSS Statistics, Armonk, NY: IBM Corp.) and Graphpad Prism version 9.0.0 (Graphpad Software).
Results

Clinical characteristics of PCNSL patients

The basic characteristics of two set PCNSL patient cohorts are shown in Table 1, which is similar with that of the validation set. Among 48 PCNSL patients in the discovery set, 24 patients (50%) were males and 24 (50%) were females, and validation set were 19 males vs. 14 females. White blood cell (WBC) number and protein level in CSF were comparable between the two cohort sets.

The status of survival was evaluated before the end of routine follow up. Patients in discovery cohort were classified in good prognosis (GP) group or poor prognosis (PP) group based on their duration of progression-free survival (PFS) longer than 12 months or shorter than 6 months, respectively. There were also similarities of gender, age, weight and BMI between GP and PP groups (Table 2). Higher ECOG scores and number of lesions were in GP group. There were no difference of WBC and protein in CSF between the two groups. In general, the patients that classified to GP and PP group were only dependent on their prognosis outcome after HD-MTX-based chemotherapy treatments.

Metabolomic Analysis of retrospective PCNSL patient cohort

CSF was collected and processed by a targeted metabolomic analysis using a LC-MS based technique (Figure 1A). As all the retrospective PCNSL cohort patients were classified into either GP group or PP group based on their PFS lengths, we then preformed a survival analysis on these two groups. The Kaplan-Meier curves were shown in Figure 1B which indicated that PFS can distinct GP group from PP group significantly. The targeted metabolomic technique were used with profiles 155 metabolites.
in CSF sample from PCNSL patients, which includes amino acids, carbohydrates, organic acids, nucleic acids, vitamins, some lipids and alkaloids (Figure 1C). As we expected, the Partial Least Squares Discrimination Analysis (PLS-DA) of GP and PP group metabolome can cluster two distinct ellipses, which indicates metabolites were altered differently in CSF samples in these two groups (Figure 1D).

To find out the mostly different metabolites in patient CSF before the HD-MTX based chemotherapy treatment, we ranked the detected metabolites based on their fold change (FC) between GP and PP groups. We then selected 6 features that were highly dysregulated in GP group based on their fold change (Figure 1E). In these 6 features, dehydroascorbic acid and AMP were upregulated (Log2FC > 0.5) and 2-isopropylmalic acid, cholesterol, p-hydroxybenzoate and 4-pyridoxic acid were downregulated (Log2FC < -0.5) (Figure S1).

Logical regression model development for the prediction of PCNSL prognosis

To build up the prediction model that can provide a potential outcome of PCNSL patients after the HD-MTX based chemotherapy, we applied a logical regression model. The workflow of the model development is illustrated in Figure 2, we firstly processed data with Log10 transformed and normalized using median value and then corrected the batch effect between the discovery and validation cohorts using an empirical Bayes method. The six dysregulated features mentioned above were applied to the modeling buildup. Using a logistics regression based model, we constitute our model as:
Equation 1:

\[ Z_{\text{score}} = 0.9125 \times (4\text{PA}) + 0.7934 \times (PHBA) + 0.3153 \times (CHOL) + 0.5088 \times (21A) - 0.4737 \times (AMP) - 0.1061 \times (DHAA) \]

The 4PA, PHBA, CHOL, 2IA, and DHAA represented 4-pyridoxic acid, p-hydroxybenzoate, cholesterol, 2-isopropylmalic acid and dehydroascorbic acid, respectively. Next, the logistic regression model was then validated in the discovery cohort, which gave a good performance (Figure 3A and 3B, AUC = 0.789). The Z scores ranged from -3.38 to 2.52, with a high score associated with a poorer outcome and lower score indicated a better outcome. The optimal cut-off was a Z score of 0.06. As we expected, the prediction model performed well in term of the survival time length; the good prognosis group (Z ≤ 0.06) had a median survival time of 18.1 months (discovery cohort), whereas the poor prognosis group (Z > 0.06) had a median survival time of 2.2 months (discovery cohort). We also performed survival analyses on these six selected features individually, but none of their Kaplan-Meier curves showed ability to differentiate GP and PP group (Figure 3C-H).

Validation of prediction model on the validation PCNSL cohort

Although the logistic regression we built could nicely separate the patients with longer PFS from the discovery cohort, it was still not convinced that if such model can apply to a new PCNSL patient set. For such reason, we next recruited a new PCNSL patient cohort before their HD-MTX based treatments. As we mentioned before, this validation cohort was similar to the discovery cohort on gender, age, as well as the BMI. We firstly collected the CSF from the validation cohort patients and profiled their CSF metabolome. Once we received the clinical outcomes of these patients two years
later, we validated our prediction model by survival analysis. The Kaplan-Meier curve of the validation set can also be separated nicely (Figure 4A and B, p = 0.032, AUC = 0.745), which the patients with Z ≤ 0.06 had longer survival time (median 18.6 months) and Z > 0.06 had shorter survival time (median 2.4 months) as we expected. This result indicated that the logical regression model we built based on CSF metabolic features could nicely apply to the prediction of PCNSL patient prognosis outcomes with HD-MTX based treatments.

Discussion

As an aggressive lymphoma occurs in central nervous system, prognosis of PCNSL was generally inferior. Although HD-MTX based chemotherapies is the first line therapeutic regimens for PCNSL patients, most patients still suffered relatively short survival\textsuperscript{22,28}. Considering the several side effects including nephrotoxicity, hepatotoxicity, neurotoxicity, mucositis, and myelosuppression of such high dosage chemotherapy\textsuperscript{29}, a prediction of HD-MTX based chemotherapy prognoses before the treatment would benefit the PCNSL patients. Ryuya Yamanaka et. al. applied different factors including gene expression and cancer morphology to create multiple promising prediction models on retrospective cohort studies in last decade\textsuperscript{11,30}. Although the prediction models revealed a good ability to identify those patients who are unlikely benefit from the standard therapies, the models were not validated by a clinical prospective cohort. Retrospective study can provide lots of information in a relatively short period, but it has high bias risk and comparison limitation\textsuperscript{31}. In our study, it is the first time we recruited two sets of PCNSL patients for the prognosis predictive model development, which 48-patient discovery group was collected from 2018 to 2020 and 35 patients prospectively recruited from 2020 to 2021 as validation group. At the time we started this clinical study, we only obtained the prognosis results of the discovery group patients. Thus, the prediction model buildup was all based on the 48-patient discovery group retrospectively. Once we received the clinical outcomes of prospectively recruited PCNSL patients, we found our prediction model can indeed distinguish the
patients with poor HD-MTX treatment outcome (Figure 4A and B), which would be helpful for PCNSL patients to make a decision if a HD-MTX based chemotherapy is preferred.

Although most of the prediction model of PCNSL prognosis of HD-MTX based chemotherapies were based on gene expression levels, it was interesting to notice that the gene set used for prediction model is not strongly associated. Some protein markers such as interleukin 6 (IL-6) and myeloid differentiation primary response (88) (MYD88) were identified as PCNSL biomarkers to distinguish from other brain tumors, which are significantly higher in PCNSL CSF and serum samples. Thus, we also measured the levels of IL-6 and MyD88 in PCNSL patient samples. However, the levels of both IL-6 and MyD88 were not significantly different between good and poor prognosis patients in our cohorts (Figure S2A-B). Considering that patients we recruited in this study were all diagnosed as PCNSL, the IL-6 or MyD88 might not be a good feature to distinguish their HD-MTX based treatment outcomes. In addition, the MyD88 L265D mutation was believed as a prognostic biomarker in PCNSL, the sensitive genome measurement would be a good tool to differ their HD-MTX based treatment outcomes in the future. In addition, considering the anti-cancer target of MTX is a metabolic enzyme, DHFR, that can perturb one carbon metabolism, it is reasonable to hypothesize that outcome of HD-MTX based chemotherapies might associated with patient metabolic status. Recently, Feng-Xiang et. al. characterized PCNSL from four different types of brain tumors using 27 CSF metabolites and Jae et. al. could also diagnose PCNSL using CSF metabolites. These two studies implied that the metabolic status in PCNSL patients’ CSF could be various. In addition, Yasuo et. al. also observed an excessive glycolysis metabolism in methotrexate-resistant PCNSL-derived cells, which shed light on the clinical prognosis prediction studies using metabolomics approach. In our study, we applied a targeted metabolomic analysis approach that includes most compounds involved in central carbon metabolism (Figure 1C and Supplementary Table 1) to profile the metabolic alteration of PCNSL patient CSF sample. Since the metabolomic profiling was performed before the HD-MTX based chemotherapy treatments, the dysregulated metabolites in CSF were more likely associated with patient outcome for such treatment. To our surprise, the 6 dysregulated
metabolites (dehydroascorbic acid, AMP, 2-isopropylmalic acid, cholesterol, p-hydroxybenzoate and 4-pyridoxic acid) in CSF is not highly associated with glycolysis metabolism discovered in MTX-resistant cell\textsuperscript{49}. Such differences could due to the observation variances between clinical and \textit{in vitro} system. Of these changed CSF metabolites, 4-pyridoxic acid caught our attention due to its involvement of one carbon metabolism. DHFR is necessary for the change of DHF to THF, while THF can readily turned to N5,10 methylene THF with the help of SHMT and PLP (Figure S3A). The PLP can turn to other forms of vitamin B6 and eventually be metabolized to 4-pyridoxic acid\textsuperscript{50}. When the methotrexate is not working properly in central nervous system, the one carbon metabolism might be dysregulated in PCNSL tumor. The correlation between one carbon metabolism with lymphoma was studied. Multiple research groups were observed a predominated correlation between reduced risk of non-Hodgkin lymphoma (NHL) with higher dietary vitamin B6 and methionine intakes\textsuperscript{51–53}. Unhee Lim et. al. took a further look at the genetic alterations and found such one-carbon metabolism contribute to lymphomagenesis not only on nutrient dietary intake but also associates with genetic variation\textsuperscript{54}. In addition, such vitamin B6 metabolic dysregulation might associated with other types cancer that Lorenzo et. al. found that vitamin B6 sensitizes non-small cell lung cancer (NSCLC) and low pyridoxal kinase (PDXK) expression would associate with poor disease outcome\textsuperscript{55}. In our study, the accumulation of 4-pyridoxic acid in CSF (Figure S3B) might due to an efflux of vitamin B6. The vitamin B6 cannot perform its proper function in one carbon metabolism pathway but was flushed as an acidic form via CSF. For such reason, we hypothesized that the accumulation of the 4-pyridoxic acid in poor prognosis patient CSF might also result from the dysregulation of such one carbon metabolism.

Although we successfully developed a model that can predict PCNSL patient prognosis with HD-MTX based chemotherapy treatments, there are still several limitations in this study. Because the PCNSL patient recruitment was only happened in Huashan hospital that only Chinese population was considered in this prospective study, the results of our clinical study may have bias which need multicentral study for a further validation in the future. Moreover, due to the relatively low occurrence
rate of PCNSL, the low number of samples limited our further mechanistic studies. A long period
clinical study would be helpful in the future. It is unfortunately that the metabolites in serum could not
distinguish the outcome differences of PCNSL patients (Figure S4A and B), so that this predication
model was based on the metabolome changes in CSF not serum making the clinical test application
difficult. A more convenient CSF sample collection would improve the clinical practice to predict
PCNSL patient prognosis with HD-MTX based chemotherapy treatments.

Conclusions

In summary, we applied metabolomic technique to study the prognosis of PCNSL patient with HD-
MTX based chemotherapy in prospective cohort study. The metabolic markers can formulate a
prediction model to distinguish the poor prognosis patients using a scoring standard (Z score > 0.06)
before receiving the HD-MTX based chemotherapy. Such model shows a high potential for the
clinical application in the future, and the metabolic markers discovered in this study would provide a
direction for the PCNSL mechanistic research.
Acknowledgments

We thank Ke Qiao, Lanlan Zhang, Rongrong Cheng, and Hongfang Zhao for metabolomic and lipidomic data generation from Single Cell Quantitative Metabolomics and Lipidomics Core Facility of IMIB at Fudan University.

Fundings

This work was supported by National Key R&D Program of China (2020YFA0803800 and 2019YFA0801900), Natural Science Foundation of China (92057115) and Shanghai Sailing Program (20YF1402600) from H.H., and Shanghai Shenkang clinical science and technology innovation project (SHDC12020112) from C. B..

Institutional Review Board Statement

This study was approved by the Institutional Review Board of Huashan Hospital (protocol No.2021-777 and No.2022-767).

Contributors

Conceptualization, LZ, QL, JX, BC, HH; Investigation, LZ, QL, JX, HH; Analysis, LZ, QL, JX, HH; Writing, LZ, QL, JX, HH; Funding Acquisition, HH; Supervision, BC, HH. Dr. He Huang and Dr. Bobin Chen have accessed and verified the data. All authors confirm that they had full access to all...
the data in the study and accept responsibility to submit for publication. All the authors promise the accuracy and completeness of the data. All authors participated the revision and approved the final version of manuscript.

Data sharing statement

The source code for the prediction model is available on Github (https://github.com/JavaScripy/2022/tree/main/PCNSL). The original data of this study are available on request from the corresponding author Dr. He Huang (he_huang@fudan.edu.cn)

Declaration of interests

The authors declare no potential conflicts of interest.
References:


(18) Zhang, X.; Wu, Y.; Sun, X.; Cui, Q.; Bai, X.; Dong, G.; Gao, Z.; Wang, Y.; Gao, C.; Sun, S.; et al. The PI3K/AKT/MTOR Signaling Pathway Is Aberrantly Activated in Primary Central Nervous System Lymphoma and Correlated with a Poor Prognosis. *BMC Cancer* 2022, 22 (1), 190.


Figure legends

Figure 1. Metabolomic analysis of PCNSL patient CSF. (A) The workflow of metabolomic analysis; (B) Patients in discovery cohort were separated by their PFS length; (C) Metabolites detected in patient CSF covered most metabolite classes; (D) Metabolites detected in good prognosis patients’ CSF were different to the ones in poor prognosis patients; (E) Six metabolic features were selected with a higher difference between good and poor prognosis patients’ CSF.

Figure 2. The pipeline of logical regression model development. All the features were selected from discovery cohort, and then different indexes of the model were tested to locate a best one. The final model was validated using a prospectively collected cohort which survival analysis was performed.

Figure 3. Survival analyses of the model and individual metabolic features. (A) The model of Z score could distinguish good (Z \leq 0.06) and poor (Z > 0.06) prognosis PCNSL patients depending on their PFS length; (B) and receiver operating characteristic curve indicated a good classification power of Z score; while none of (C-H) the metabolic features could separate the GP and PP group individually.

Figure 4. Applying the prediction model on a prospectively recruited PCNSL cohort. (A) The Z score could also nicely distinguish the good (Z \leq 0.06) prognosis PCNSL patients from the validation PCNSL cohort, (B) and receiver operating characteristic curve indicated a good classification power of Z score.
Table 1. The clinical characteristics of discovery and validation PCNSL patient cohort

<table>
<thead>
<tr>
<th></th>
<th>Discovery cohort</th>
<th>Validation cohort</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=48</td>
<td>n=33</td>
<td></td>
</tr>
<tr>
<td>Gender, n(%)</td>
<td></td>
<td></td>
<td>0.502</td>
</tr>
<tr>
<td>Male</td>
<td>24(50%)</td>
<td>19(58%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>24(50%)</td>
<td>14(42%)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>56[50-66]</td>
<td>62[50.5-67]</td>
<td>0.600</td>
</tr>
<tr>
<td>Weight (Kg)</td>
<td>68.521</td>
<td>62.606</td>
<td>0.124</td>
</tr>
<tr>
<td>BMI (kg/m$^2$)</td>
<td>22.929</td>
<td>18.868</td>
<td>0.551</td>
</tr>
<tr>
<td>Elevated CSF WBC count</td>
<td>19/46</td>
<td>9</td>
<td>0.200</td>
</tr>
<tr>
<td>Elevated CSF protein level</td>
<td>33</td>
<td>26</td>
<td>0.187</td>
</tr>
</tbody>
</table>
Table 2. The demographics and clinical characteristics of discovery cohort

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All (n = 48)</th>
<th>Good prognosis (n = 25)</th>
<th>Poor prognosis (n = 23)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>0.773</td>
</tr>
<tr>
<td>Male</td>
<td>24</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>24</td>
<td>13</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Age, median [IQR]</td>
<td>56[50-66]</td>
<td>54[49.5-63]</td>
<td>61[51-68]</td>
<td>0.189</td>
</tr>
<tr>
<td>Weight, Mean (Kg)</td>
<td>68.5</td>
<td>71.1</td>
<td>65.6</td>
<td>0.542</td>
</tr>
<tr>
<td>BMI, Mean (kg/m^2)</td>
<td>22.929</td>
<td>25.889</td>
<td>19.712</td>
<td>0.380</td>
</tr>
<tr>
<td>ECOG score</td>
<td></td>
<td></td>
<td></td>
<td>0.046*</td>
</tr>
<tr>
<td>0-1</td>
<td>15</td>
<td>11</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>≥2</td>
<td>33</td>
<td>14</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Multiple lesions</td>
<td>18</td>
<td>6</td>
<td>12</td>
<td>0.044*</td>
</tr>
<tr>
<td>Involvement of deep structure</td>
<td>35</td>
<td>19</td>
<td>16</td>
<td>0.616</td>
</tr>
<tr>
<td>Elevated CSF WBC count</td>
<td>19/46</td>
<td>10/23</td>
<td>9/23</td>
<td>0.765</td>
</tr>
<tr>
<td>Elevated CSF protein level</td>
<td>33</td>
<td>15</td>
<td>18</td>
<td>0.173</td>
</tr>
<tr>
<td>HD-MTX</td>
<td></td>
<td></td>
<td></td>
<td>0.709</td>
</tr>
<tr>
<td>8g/m^2</td>
<td>30</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;8g/m²</td>
<td>Idarubicin</td>
<td>Rituximab</td>
<td>Dexamethasone</td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>------------</td>
<td>-----------</td>
<td>---------------</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td></td>
<td>0.613</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>33</td>
<td>18</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>15</td>
<td>7</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.945</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>45</td>
<td>24</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>48</td>
<td>25</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2

1. Discovery cohort
2. Data processing
3. Batch effect correlation
4. Feature selection
5. Modeling: Logistics regression
6. Validation cohort
7. Survival analysis
Figure 3

A. Z score in discovery cohort

- Survival probability
- Number at risk
  - Z≤0.06
  - Z>0.06
- p = 0.012

B. Sensitivity (True positive rate)

- AUC = 0.788
- 95% CI: 0.618-0.925

C. Dehydroascorbic acid

- Survival probability
- Number at risk
  - Low
  - High
  - p = 0.62

D. AMP

- Survival probability
- Number at risk
  - Low
  - High
  - p = 0.42

E. 2-Isopropylmalic acid

- Survival probability
- Number at risk
  - Low
  - High
  - p = 0.36

F. Cholesterol

- Survival probability
- Number at risk
  - Low
  - High
  - p = 0.07

G. p-Hydroxybenzoate

- Survival probability
- Number at risk
  - Low
  - High
  - p = 0.13

H. 4-Pyridoxic acid

- Survival probability
- Number at risk
  - Low
  - High
  - p = 0.15