

Final report to the NKFI grant No PD-115792

Primary central nervous system lymphoma (PCNSL) is a rare disease with a generally poor outcome and with very limited therapeutic choices. Therefore, PCNSL represents a significant clinical challenge with an urgent need of better biomarkers and novel therapeutic targets. In this project we proposed to investigate different aspects of PCNSL on a relatively large cohort of patients, using different molecular and immunohistochemical (IHC) approaches, to reveal the pathogenic and prognostic significance of (i) the activity of the PI3K/mTOR pathway, (ii) the cell of origin (COO) of PCNSL, (iii) the mutational profile of selected candidate genes, and (iv) microRNA (miR, miRNA) expression profiles.

1, Collection of primary and secondary central nervous system lymphoma cases

We collected formalin-fixed paraffin embedded (FFPE) blocks and relevant clinical data of 127 primary and 24 secondary CNS lymphoma cases from 3 centers (1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary; Department of Pathology, University of Pécs, Pécs, Hungary; and Division of Neuropathology, UCL Institute of Neurology, London, UK.). We also collected 51 cases of nodal diffuse large B-cell lymphoma (DLBCL) and 5 non-malignant tissues as control for the IHC and sequencing studies, respectively.

2, Nucleic acid isolation from FFPE tissues

We isolated DNA from 105 cases including PCNSL and SCNSL cases as well as control samples using the FFPE Tissue Kit (Qiagen, N.V., Venlo, Netherlands) following the manufacturer's instructions. The isolated DNA samples were stored at 4°C until further use.

We extracted RNA from 96 cases including PCNSL and SCNSL using the RecoverAll kit (Life Technologies/Ambion, Inc., Foster City, CA) according to the manufacturer's instructions. The isolated RNA samples were stored at -70°C until further use.

3, Results of the mTOR study

Dysregulation of the mTOR pathway has been recognized as a critical event in the development of different hematological malignancies including systemic DLBCL. Generally, more than half of the DLBCL cases show aberrantly activated mTOR pathway. mTOR represents a therapeutic target, and mTOR inhibitors alone or in combination with other agents have demonstrated promising results in relapsed or refractory DLBCL. Despite the promising results with mTOR inhibitors in DLBCL and the need for better understanding the pathomechanisms of PCNSL, there is very limited data on mTOR pathway activity in PCNSL.

3A, Immunohistochemical analysis of the PI3K/Akt/mTOR pathway.

The primary aim of this part of the study was to determine the mTOR pathway activity in PCNSL with an IHC approach. 31 cases of PCNSL and 51 cases of nodal DLBCL were included. The majority of both PCNSL (83.9%) and DLBCL cases (62.75%) showed p-S6 expression (Figure 1.). pmTOR was positive in 12.9% of PCNSL and in 54.9% of DLBCL cases. p-p70S6K1 was positive in 6.5% of PCNSL and in 31.4% of DLBCL cases. p-4EBP1 was positive in 12.9% of PCNSL and in 29.4% of DLBCL cases. In this cohort, only 25.8% of the PCNSL cases, while 66.7% of the DLBCL cases showed mTOR pathway activity by IHC as defined by our criteria (Figure 1.). The mTOR pathway activity was significantly less frequent in PCNSL compared with DLBCL ($p < 0.001$). Interestingly, among the mTOR pathway inactive cases p-S6 was positive in 78.3% of PCNSL and 35.3% of DLBCL cases, respectively (Figures 1 and 2). In line with these results, positivity of p-S6 was significantly associated with the activated mTOR pathway in DLBCL ($p = 0.005$), but not in PCNSL ($p = 0.198$).

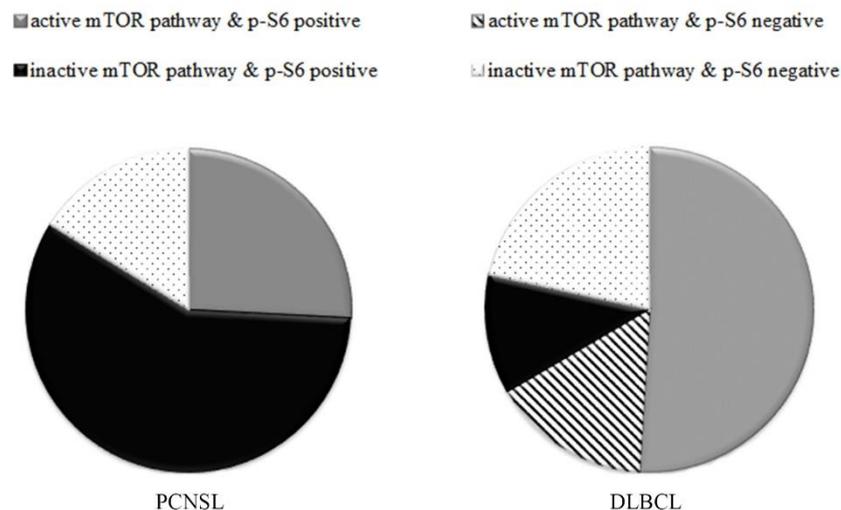


Figure 1. Distribution of mTOR signalling pathway activity and p-S6 expression. p-S6 positivity may represent mTOR pathway activity in DLBCL, however in PCNSL the number of p-S6 positive cases largely exceed the number of mTOR active cases.

After demonstrating that the number of p-S6 positive cases largely exceeded the number of cases with mTOR activity in PCNSL, we also aimed to identify the mTOR independent alternative pathways leading to S6 phosphorylation in PCNSL. Amongst the other kinases (p-RSK, p(T229)-p70S6K1 and PASK) potentially responsible for S6 phosphorylation, PASK proved to be positive in all cases of PCNSL and DLBCL (the latter were used as control) (Figure 2).

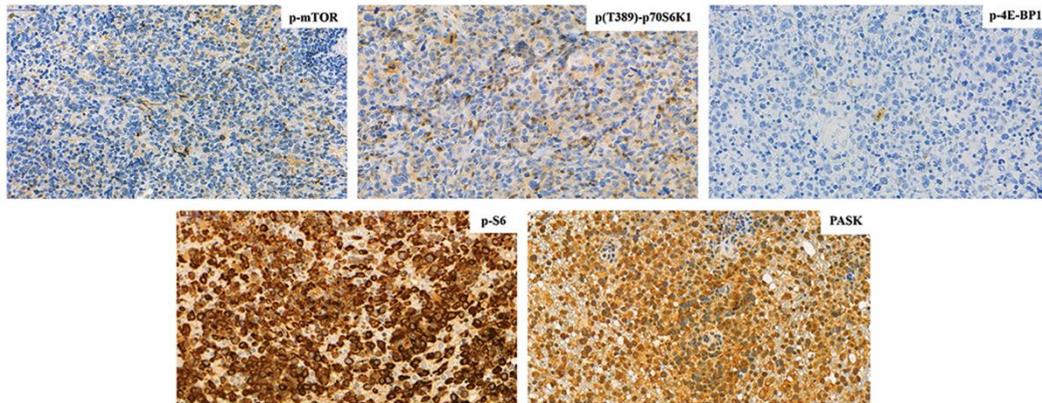


Figure 2. Representative immunohistochemistry images of the mTOR pathway activity analysis in PCNSL. The majority (58.1%; 18/31) of PCNSL cases display p-S6 positivity without mTOR pathway activity. All cases of PCNSL were PASK positive.

3B, Cell culture experiments

In order to prove that PASK could be responsible for S6 phosphorylation, diffuse large B-cell lymphoma cell line (MedB-1/BHD1) with known p-S6 expression due to mTOR activity was cultured and treated with mTOR inhibitor rapamycin, PASK inhibitor and both in combination, and p-S6 level was measured with Western blot and flow cytometry analyses.

As expected, rapamycin significantly reduced the level of p-S6. Inhibition of PASK also led to a significant downregulation of p-S6 (even in the presence of an active mTOR), suggesting that PASK contributes to the phosphorylation of S6 in DLBCL cells. The effect of the combined PASK inhibitor and rapamycin treatment was the most robust (Figure 3).

3C, Summary

In summary, we demonstrated that mTOR pathway is active only in around 25% of the cases. These patients, however, could benefit from mTOR-inhibitor therapy. We also showed, that the increased p-S6 expression cannot entirely be explained by mTOR pathway activity in PCNSL, and PASK may contribute to an mTOR independent S6 phosphorylation. We found frequent PASK expression in PCNSL as well as in nodal DLBCL, suggesting a potential role of PASK in the pathomechanism of these lymphomas. PASK may be a novel candidate for targeted therapy.

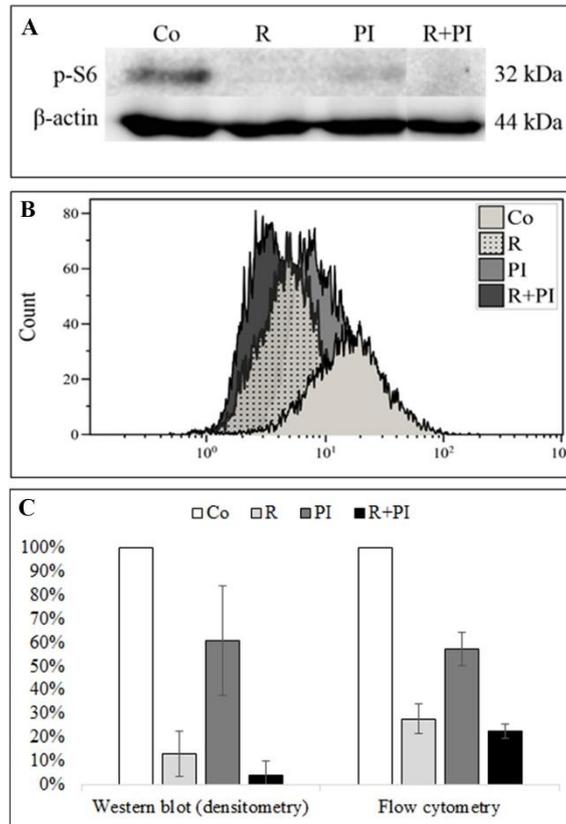


Figure 3. The effect of different treatments on p-S6 level. Inhibition of mTOR, PASK and both in BHD1 cell culture greatly reduced the level of p-S6 as confirmed by Western blot (A and C) and flow cytometry (B and C) analysis. The combined PASK inhibitor (PI) and rapamycin (R) treatment demonstrated the most robust effect.

3D, Publishing of the mTOR study

We published these results in a peer-reviewed journal:

- Marosvári D, Nagy N, Kriston C, Deák B, Hajdu M, Bödör C, Csala I, Bagó AG, Szállási Z, Sebestyén A, Reiniger L. Discrepancy Between Low Levels of mTOR Activity and High Levels of P-S6 in Primary Central Nervous System Lymphoma May Be Explained by PAS Domain-Containing Serine/Threonine-Protein Kinase-Mediated Phosphorylation. *J Neuropathol Exp Neurol.* 2018 Apr 1;77(4):268-273. doi: 10.1093/jnen/nlx121.

We also presented our results as oral presentations:

- Low mTOR activity in PCNSL (Alacsony mTOR aktivitás primer központi idegrendszeri lymphomákban, XXVIth Congress of the Hungarian Society Of Haematology And Transfusion, 19.05.2017, Szeged, Hungary)
- mTOR activity is reduced in PCNSL (Poster presentation at the Semmelweis Symposium, 17-18.11.2016, Budapest, Hungary)

4, Results of the cell of origin and deep next generation sequencing studies of PCNSL and SCNSL

The fundamental difference in biology between germinal center (GC) and activated B-cell (ABC) type DLBCL is reflected in different driver oncogenic pathways and mutation targets, which also translates in different efficacy of the novel targeted therapies between the 2 subgroups.

4A, Cell of origin (molecular subtype)

Considering the emerging novel therapies, precise assignment of patients into the molecular subgroups may well become of major clinical importance. The NanoString Lymphoma Subtyping Test (LST) assay (NanoString Technologies, Inc., Seattle, WA) was developed to establish an FFPE compatible, gene expression based test for the molecular subtyping of B-cell lymphomas and represents a more accurate technique compared with standard IHC algorithms.

The COO of 77 PCNSL and 17 SCNSL patients was determined with the NanoString LST assay on the nCounter Analysis System using the extracted RNA samples. These data were compared with the results obtained by the Hans IHC algorithm. Using the Hans algorithm, 95% of the PCNSL cases showed ABC (non-GC) phenotype, with 5% of the patients presenting with GC phenotype. In contrast, the LST-assay identified only 80.5% of the cases as ABC subtype, with 13% GC and 6.5% UC molecular subtypes, respectively. As for the SCNSL cases, 47% showed ABC and 53% showed GC phenotype with the Hans algorithm. The ratio was identical using the LST-assay, with 47% ABC and 53% GC subtypes. The subclassification obtained with the LST-assay showed discordant readouts in 16% of all analyzed cases as compared to the IHC results. 12.5% of the cases classified as ABC by the Hans algorithm showed a different readout when analyzed using the LST-assay, with 7 cases assigned to the GC group and 5 unclassified cases, while only one (1%) IHC-GC case was classified as ABC using the LST-assay. In the SCNSL group, only a single GC and a single ABC case did not match when comparing the readouts of the Hans algorithm and the LST assay. Taken together, using the LST-assay, a significantly lower proportion of cases (80.5% versus 95%, $p=0.0219$) displayed an ABC phenotype in PCNSL (Figure 4.).

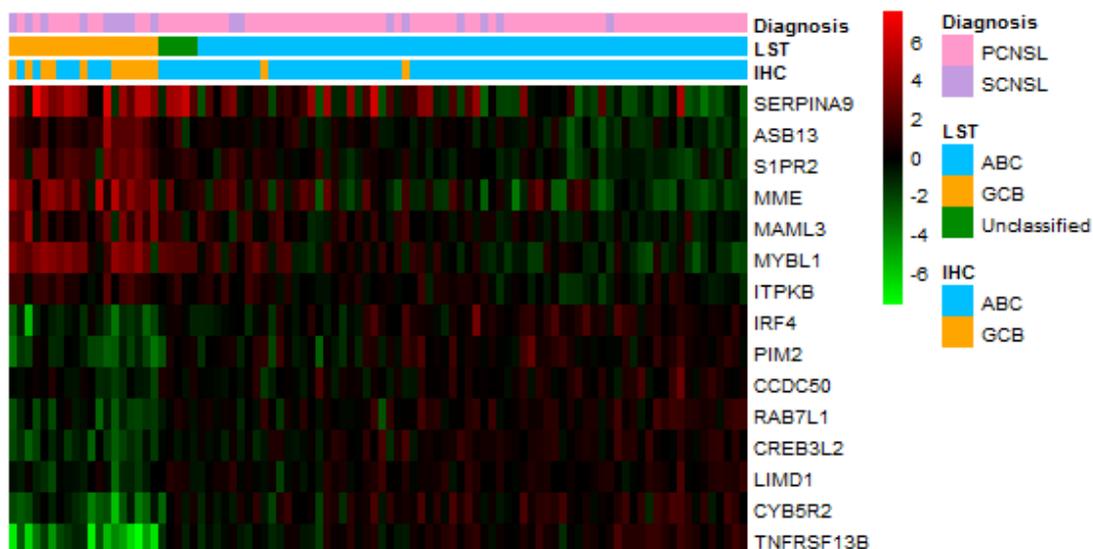


Figure 4. The NanoString LST readouts are illustrated in form of a gene expression heat map with the 15 target genes contributing to the model.

The survival of the patients with PCNSL and SCNSL did not show a significant difference ($p=0.1970$) (Figure 5A). Interestingly, the molecular subtype did not have an impact on the survival of patients, with survival showing no differences between the GC and ABC subtype neither in the entire cohort ($p=0.3981$) (Figure 5B) nor in the PCNSL cases ($p=0.8727$) (Figure 5C). These results are limited by the modest number of cases and the heterogeneous nature of the treatment regimens applied.

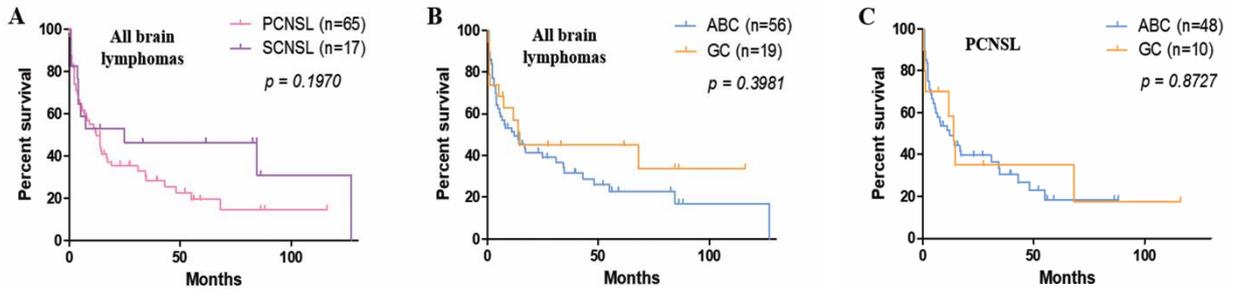


Figure 5. Effect of the primary or secondary nature (A) and molecular subtypes (B, C) on the survival of patients with primary and secondary central nervous system lymphomas. No survival difference was observed between PCNSL and SCNSL or cases with ABC or GC categories. Unclassified cases are not presented on the graphs B and C.

4B, Mutation profiles

In addition to the precise determination of the molecular subtypes, the identification of key genetic alterations is also becoming increasingly important for developing novel therapies and applying a personalized therapeutic approach in the treatment of PCNSL patients. Thus, in this study we performed complementary targeted mutation profiling on a subset of cases.

The genomic profiles of 64 PCNSL and 12 SCNSL cases were determined using ultra-deep targeted NGS of 14 recurrently mutated target genes, including *CARD11*, *CCND3*, *CD79B*, *CSMD2*, *CSMD3*, *IRF4*, *C-MYC*, *MYD88*, *PAX5*, *PIM1*, *PRDM1*, *PTPRD* and *TP53* using the TruSeq Custom Amplicon dual strand approach on a HiSeq4000 Instrument (Illumina). A subset of somatic variants with variant allele frequency of >20% was validated by bidirectional Sanger sequencing.

A total of 239 mutations were detected across the 76 brain lymphomas with the variant allele frequencies (VAF) ranging between 1.8 and 96.2% with a mean value of 41.4%. The vast majority (81%) of the mutations presented with a VAF higher than 20%. We detected a total of 210 somatic mutations in the 64 PCNSL cases across the 14 genes analyzed, with an average of 3.3 mutations per case ranging between 0 and 10. The most frequently mutated genes in the PCNSL cohort included MYD88 (66%), PIM1 (41%), KMT2D (31%) and PRMD1 (30%). The mutation frequencies in the remaining genes were as follows: C-MYC (19%), IRF4 (19%), CD79B (17%), TP53 (11%), CCND3 (9%), CARD11 (8%), PAX5 (3%), CSMD2 (3%) and CSMD3 (3%). No mutation was found in PTPRD gene. In the 12 SCNSL cases, a total of 29 somatic mutations were detected, with an average of 2.4 mutations per case ranging between 0 and 5. Individual cases harbored mutations in 1.8 genes on average, ranging between 0 and 4. PRDM1 (50%), followed by MYD88 (42%) and PIM1 (25%) were the most frequently mutated target genes in the SCNSL cohort. The mutations frequencies in the remaining genes were lower: KMT2D (17%), CD79B

(8%), IRF4 (8%), CCND3 (8%), C-MYC (8%), TP53 (8%) and PAX5 (8%). No mutation was found in CARD11, CSMD2, CSMD3 and PTPRD genes (Figures 6 and 7A).

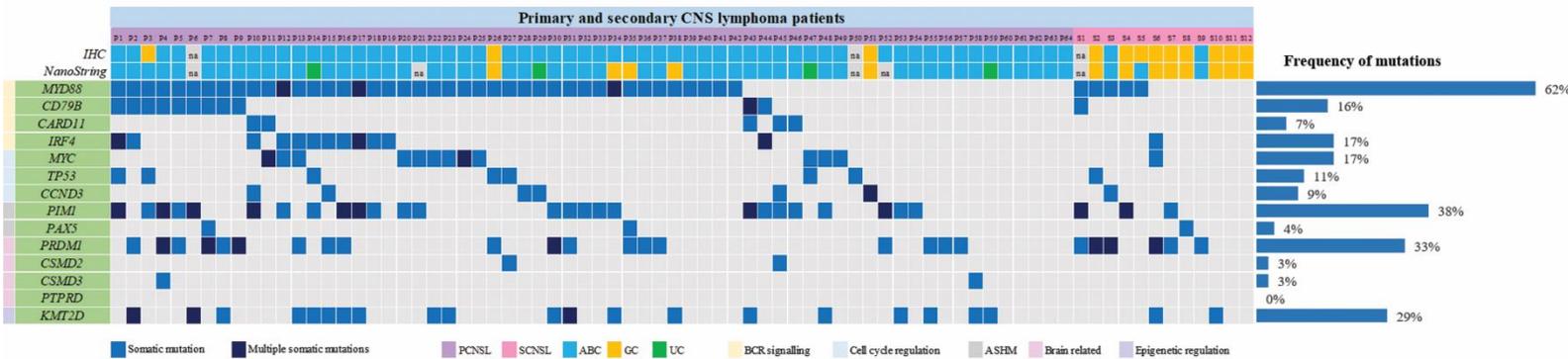


Figure 6. Illustrated are the mutation patterns of the 14 genes identified in 76 primary and secondary central nervous system lymphomas by next-generation sequencing, and molecular subtypes of 71 and 73 cases as defined by the NanoString LST-assay and the Hans algorithm using immunohistochemistry, respectively.

4C, Correlation of mutation profiles and COO

Considering all brain lymphomas, we observed an enrichment of MYD88 (67% vs 46%), PIM1 (39% vs 23%), IRF4 (20% vs 8%) and MYC (19% vs 8%) mutations in cases with ABC subtype, with CD79B, CARD11, CSMD2 and CSMD3 mutations being present exclusively in ABC cases (19%, 9%, 4% and 4% vs 0% for the four genes, respectively). On the other hand, mutations of TP53 (15% vs 6%) and PAX5 (15% vs 2%) appeared more frequent in GC cases. PRDM1, KMT2D and CCND3 genes showed similar mutational frequencies by comparing the GC and ABC cases (Figure 7B).

In PCNSL, enrichment of PIM1 mutations was observed (41% vs 20%) in cases with ABC subtype, with IRF4, CD79B, MYC, CARD11, CSMD2 and CSMD3 mutations being present exclusively in ABC cases (22%, 20%, 20%, 10%, 4% and 4% vs 0% for the six genes, respectively). In PCNSL cases with GC subtype, mutations of TP53 (20% vs 6%), PAX5 (20% vs 2%) and CCND3 (20% vs 8%) appeared more frequent compared to the ABC cases. MYD88, PRDM1 and KMT2D genes showed similar mutational frequencies across the two subtypes (Figure 7C). Despite the apparent enrichment of mutations, none of these differences reached statistical significance when compared between the GC and ABC groups.

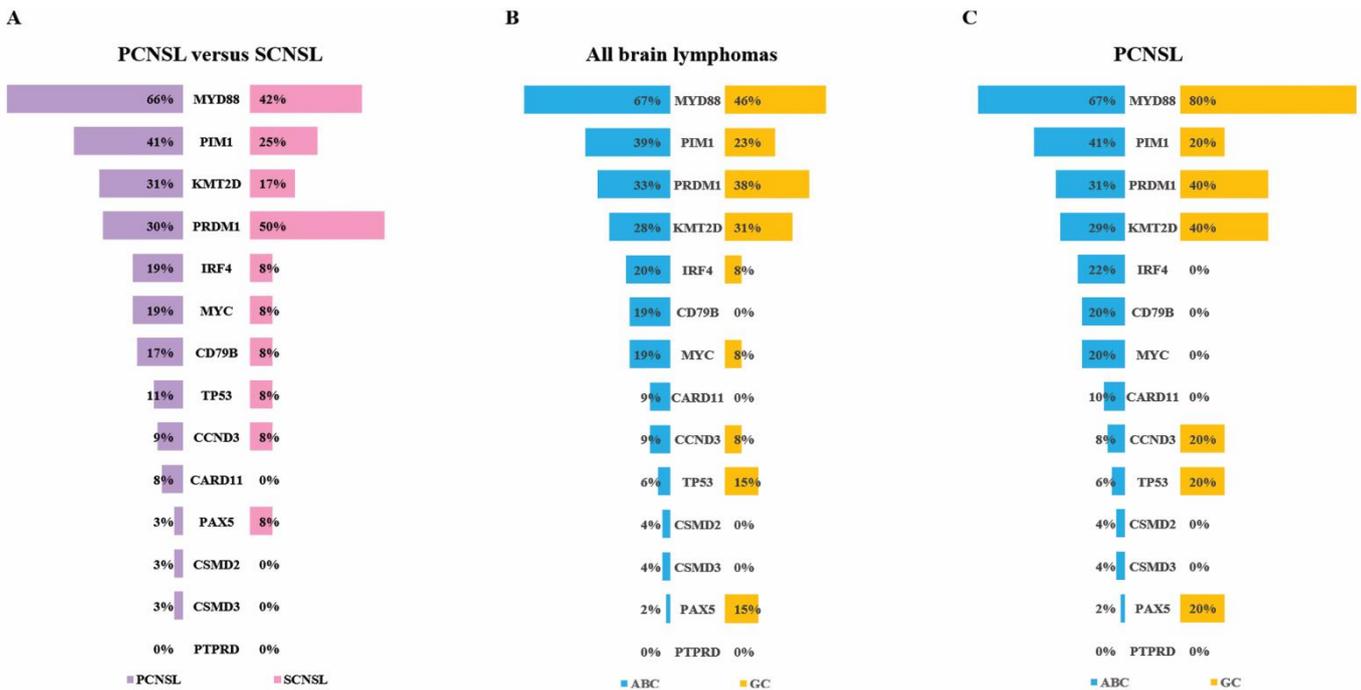


Figure 7. Comparison of mutation profiles between primary and secondary central nervous system lymphomas (A), all brain lymphomas of activated B-cell type (ABC) versus germinal center B-cell type (GC) (B), and primary brain lymphomas of ABC versus GC type (C).

4D, Correlation of mutation profiles and survival

Cases with *CD79B* mutation showed significantly worse survival compared to cases with a wild type *CD79B* gene, when considering all lymphomas ($p=0.0126$) (Figure 8A), all ABC subtype lymphomas ($p=0.0072$), PCNSL cases only ($p=0.0268$) (Figure 8B) and ABC subtype PCNSL cases ($p=0.0189$).

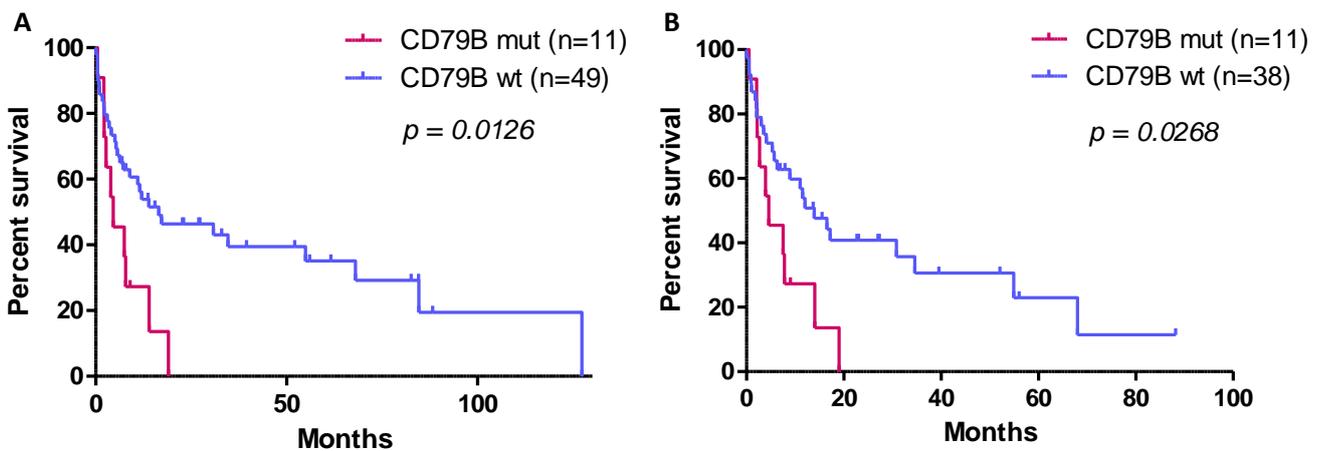


Figure 8. *CD79B* mutation is associated with worse overall survival in all central nervous system lymphomas (A) and in primary central nervous system lymphomas (B).

4E, Summary

In summary, we successfully applied for the first time the NanoString LST-assay for COO determination using FFPE derived RNA samples from a large cohort of patients with brain lymphoma. Interestingly, this molecular profiling revealed a higher proportion of cases with a GC subtype compared to the parallel IHC analysis (13% vs 5%), representing a considerably higher proportion as previously appreciated in the literature. We also determined the mutational profile of ABC and GC subgroups.

Considering our increasing knowledge on the genomic complexity of PCNSL and emergence of novel therapies with differential activity in the GC and ABC patient groups, precise assignment of the molecular subtypes using routinely available FFPE tissues and complementary mutation analysis of the actionable mutation targets will most likely support and drive personalized therapeutic decisions during the management of PCNSL patients. *CD79B* mutation status may be used as a prognostic marker in PCNSL.

4F, Publishing of the molecular subtypes and genomic profile study

We published these results in a peer-reviewed journal:

- Bödör C, Alpár D, Marosvári D, Galik B, Rajnai H, Bártai B, Nagy Á, Kajtár B, Burján A, Deák B, Schneider T, Alizadeh H, Matolcsy A, Brandner S, Storhoff J, Chen N, Liu M, Ghali N, Csala I, Bagó AG, Gyenesei A, Reiniger L. Molecular Subtypes and Genomic Profile of Primary Central Nervous System Lymphoma. *J Neuropathol Exp Neurol.* 2020 Feb 1;79(2):176-183. doi: 10.1093/jnen/nlz125.

We also presented our results as oral presentations:

- Cell of Origin and Genomic Profile of Primary Central Nervous System Lymphoma Determined Using the Nanostring LST Assay and Ultra Deep Targeted Next Generation Sequencing (Poster presentation at the 59th American Society of Hematology (ASH) Annual Meeting & Exposition, 9-12.12.2017, Atlanta, USA)
- Cell of Origin and Genomic Profile of Primary Central Nervous System Lymphoma Determined Using the Nanostring LST Assay and Ultra Deep Targeted Next Generation Sequencing (Poster presentation at the 23rd Congress of the European Hematology Association (EHA), 14-17.06.2018, Stockholm, Sweden)
- Cell of Origin and Genomic Profile of Primary Central Nervous System Lymphomas (Primer központi idegrendszeri lymphomák sejteredete és genomikai profilja, XXIst Malignant Lymphoma Congress of the Hungarian Society Of Haematology And Transfusion, 24-26.05.2018, Debrecen, Hungary)
- Determination of the COO of PCNSL by NanoString LST assay (A primer központi idegrendszeri lymphomák sejteredetének meghatározása NanoString LST assay alkalmazásával, XXVith Congress of the Hungarian Society Of Haematology And Transfusion, 19.05.2017, Szeged, Hungary)

5, MicroRNA profiling using the NanoString technique

The discovery of microRNAs (miRNAs, miRs) has opened a new field for unraveling and therapeutically targeting diseases. MicroRNAs provide a robust signal and can be stably extracted from FFPE samples, making them preferential as cancer biomarkers. They are also promising targets for molecular therapy in cancer, due to their role as oncomiRs or tumor-suppressors.

We performed expression profiling of 798 human miRNAs in 73 FFPE brain biopsy samples of primary (n=64) and secondary (n=9) CNS lymphomas using the Human v3 miRNA expression assay kit on the NanoString platform (nCounter miRNA Expression Assay), followed by complex bioinformatics analysis to reveal changing expression signatures. We aimed to identify potential novel biomarkers characterizing subgroups among brain lymphomas, as well as to examine differences based on their primary and secondary nature, molecular subtype, mutational patterns and survival.

We decided to use the NanoString technique instead of TaqMan low-density array, as this product provides a sensitive and reproducible method for detecting specific miRNAs within purified RNA isolated from FFPE samples, and the assay utilizes NanoString's nCounter platform to offer direct, digital counts of each miRNA without the use of reverse transcription or amplification. With this approach we significantly expanded the number of studied miRNAs compared to the originally planned 25 selected miRNAs.

5A, Differential miRNA expression between primary and secondary CNS lymphomas and molecular subtypes, and in association to mutation status

We identified 28 up- and 3 downregulated miRs in SCNSL compared with PCNSL (Table 1.). Amongst these, we can find miR-30c-5p, which has already been described to be significantly increased in CSF specimens of patients with SCNSL compared with PCNSL. In general, miR-30c-5p has a tumor-suppressive role in cancer pathogenesis, and shows low expression in various malignancies which is in line with our findings. Its significantly lower expression in PCNSL may contribute to their more aggressive behavior. Multiple studies have found higher expression of miR-21, a well-known oncomir, both in PCNSL and DLBCL cases compared to controls. In our study miR-21-5p generally showed high expression with a significant increase in SCNSL cases. We found 3 members of the miR-17-92 cluster (miR-19a-3p, miR-18a-5p and miR-106b-5p) to be upregulated in SCNSL compared with PCNSL, with a moderately high overall expression. This miR cluster has a strong oncogene activity in various malignancies including DLBCL.

Regarding the molecular subtypes, cases in the ABC group showed significantly higher expression of miR-155-5p, miR-222-3p and miR-522-3p and lower expression of miR-92a-3p compared with the GC group. In line with our results, higher expression of miR-155-5p and miR-222-3p has already been associated with the ABC subtype in DLBCL. We found that the ABC molecular subtype also correlated with higher miR-522-3p, miR-454-3p and miR-455-5p expression compared with the UC subgroup (Table 1.).

This is the first study demonstrating differentially expressed miRNAs in association with the mutational status of the PRDM1, C-MYC and CARD11 genes in CNS lymphomas (Table 1.).

5B, Validation of the NanoString results

It is widely known, that the different miRNA profiling platforms do not perform evenly. Nevertheless, we successfully validated the observed expression patterns of miR-148b-3p, miR-32-5p, miR-411-5p and miR-379-5p by ddPCR and/or RT-PCR methods.

Table 1. Differentially expressed miRs in CNS lymphoma cases

	microRNA	log2fc	p value	
SCNSL vs. PCNSL	hsa-miR-873-3p	-7,605495464	0,000876427	
	hsa-miR-6721-5p	-4,737186299	0,021128771	
	hsa-miR-509-5p	-2,172209427	0,016905487	
	hsa-miR-454-3p	2,752454143	0,013576217	
	hsa-miR-532-3p	2,922664981	0,017949539	
	hsa-miR-301a-3p	3,134127201	0,009885335	
	hsa-miR-320e	3,514487651	0,016284426	
	hsa-miR-374b-5p	3,534303786	0,010781895	
	hsa-miR-421	3,544794939	0,003828792	
	hsa-miR-30c-5p	3,551215934	0,006835022	
	hsa-miR-361-5p	3,677931406	0,014664301	
	hsa-miR-133a-3p	3,748778245	0,003628062	
	hsa-miR-551b-3p	3,751949004	0,005957892	
	hsa-miR-630	4,205357434	0,009337205	
	hsa-miR-1290	4,68380395	0,00353477	
	hsa-miR-574-5p	4,686394191	0,008364763	
	hsa-miR-107	4,710465086	0,021339641	
	hsa-miR-362-5p	4,865043246	3,61057E-05	
	hsa-miR-1246	4,92325355	0,022931902	
	hsa-miR-1915-3p	5,049033128	0,006481489	
	hsa-miR-148b-3p	5,209166621	0,016151903	
	hsa-miR-1285-5p	5,275968883	3,47504E-05	
	hsa-miR-21-5p	5,33362096	3,61814E-05	
	hsa-miR-7-5p	5,970379157	0,000770613	
	hsa-miR-423-5p	6,745393249	4,63583E-06	
	hsa-miR-32-5p	6,748178046	0,000144372	
	hsa-let-7e-5p	7,267979671	6,93184E-06	
	hsa-miR-19a-3p	7,327707072	3,5505E-05	
	hsa-miR-106b-5p	7,838756224	1,69779E-06	
	hsa-miR-18a-5p	8,588801883	2,88482E-08	
	hsa-miR-4516	8,739384027	0,00012008	
	GC vs. ABC	hsa-miR-222-3p	-3,819316989	8,4365E-05
		hsa-miR-522-3p	-2,344856677	0,000224632
hsa-miR-155-5p		-2,264963893	0,00010359	
hsa-miR-92a-3p		4,700026682	6,98953E-07	
UC vs. ABC	hsa-miR-522-3p	-3,05646362	1,85611E-05	
	hsa-miR-455-5p	-2,091607287	2,39961E-05	
	hsa-miR-454-3p	-2,118681159	4,22071E-05	
PRDM1 mut. vs. non-mut	hsa-miR-551a	-1,273201518	0,00604504	
	hsa-miR-107	1,416273171	0,003099768	
	hsa-miR-195-5p	1,53371272	0,002010757	
	hsa-miR-526b-5p	1,686516974	0,006075545	
	hsa-miR-376a-3p	1,874457116	0,008297192	
	hsa-miR-130a-3p	2,343294171	0,021347819	
	hsa-miR-548y	2,885455808	8,47191E-06	
	hsa-miR-30a-5p	3,610061286	0,000690782	
C-MYC mut. vs. non-mut	hsa-miR-411-5p	-2,501352047	5,27432E-06	
CARD11 mut. vs. non-mut	hsa-miR-379-5p	-3,547278055	2,3031E-05	

5C, Unsupervised clustering of miRNA expression data

We checked if the data grouped according to disease characteristics using principal component analysis (PCA). The PCA results did not indicate the presence of distinct groups based on disease subtype (primary or secondary) or molecular subtype (Figure 9). Additionally, the samples do not cluster based on RNA isolation group, place of origin, scanning date, degradation time, % tumor content, age or sex. As we could not define any specific grouping of samples based on this analysis, we investigated the miRNA expression patterns further, using additional clustering methods.

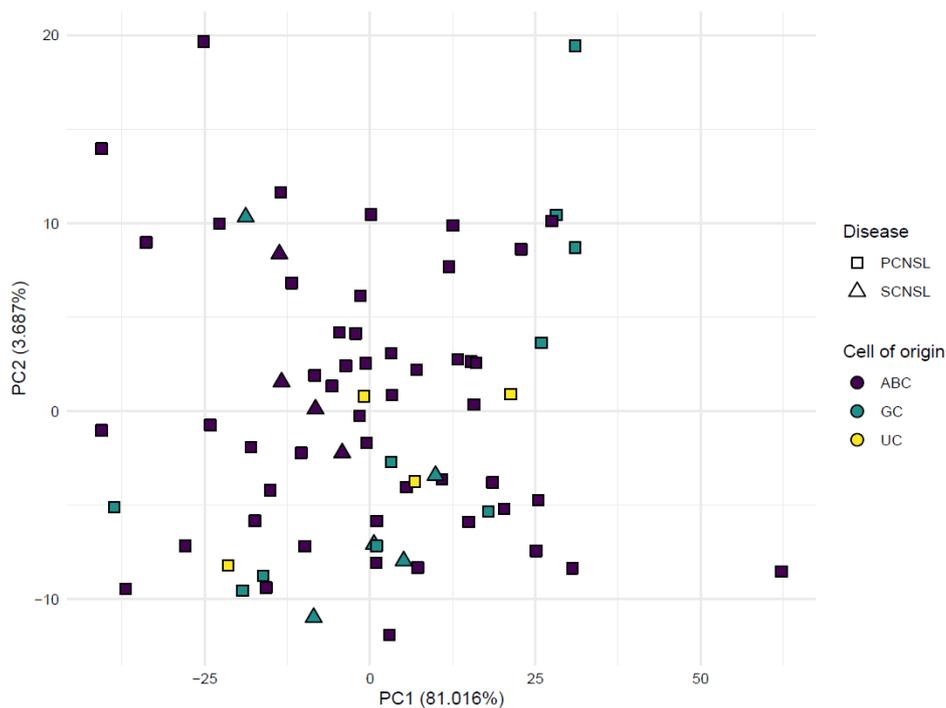


Figure 9. Principal component analysis of miRNA expression patterns of all samples, after removing potential batch effects from the normalized and voom transformed data. The shape of the points shows the primary or secondary disease category, while the color corresponds to the molecular subtype.

First, we did a binary clustering of the data (Figure 10), where we only considered the expressed/not expressed status of miRNAs. Based on the binary clustering, two sample groups become apparent. A small group, consisting of 8 samples are clearly distinct from the rest of the cohort. The binary expression patterns of this group are markedly different from the larger group (Table 2 and Figure 10).

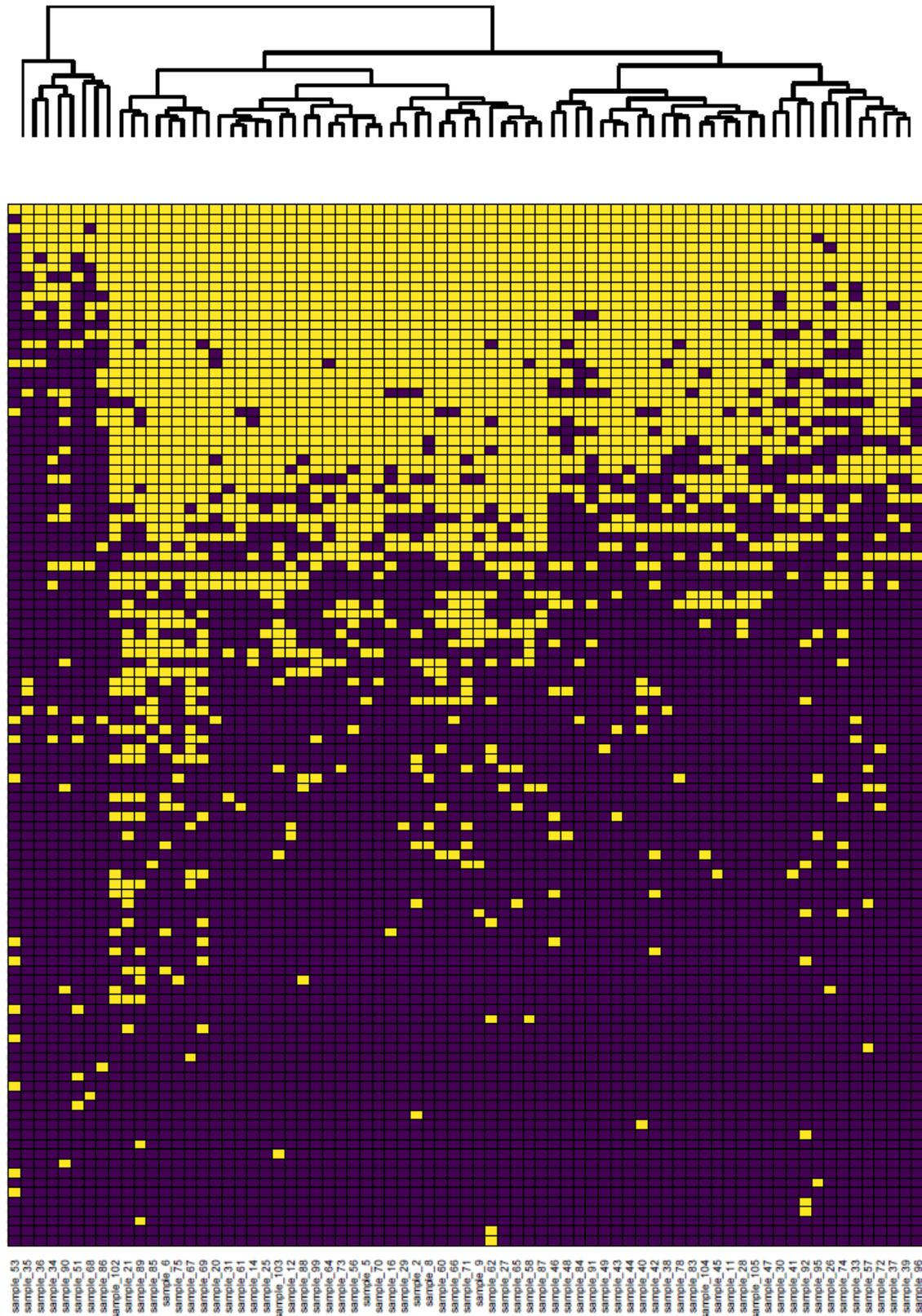


Figure 10. Unsupervised clustering of samples using binary (ON/OFF) miRNA expression patterns. The heatmap shows the binary expression pattern of miRNAs that were considered expressed in at least one sample. Columns correspond to samples, while rows correspond to a specific miRNA. Yellow tiles show expressed, purple tiles show non-expressed miRNAs. On top of the heatmap a hierarchical clustering tree shows the relationship between samples. miRNAs that were not expressed in any of the samples are not shown.

Table 2. Sample clusters. Summary of different clustering analysis methods, showing sample ids (sample_id) that were included in the unknown “small clusters” based on a specific method (Binary cluster PCNSL+SCNSL, Binary cluster PCNSL only, Kmeans cluster PCNSL+SCNSL, Kmeans cluster PCNSL only). The table shows the number of times a sample id was present in the small cluster during a specific analysis (Present in cluster) and the size of the cluster in a specific analysis (Cluster size).

sample_id	Binary cluster PCNSL + SCNSL	Binary cluster PCNSL only	Kmeans cluster PCNSL + SCNSL	Kmeans cluster PCNSL only	Present in clusters
sample_26	no	no	yes	yes	2
sample_30	no	no	yes	yes	2
sample_33	no	yes	yes	yes	3
sample_34	yes	yes	no	no	2
sample_35	yes	yes	no	no	2
sample_36	yes	yes	yes	yes	4
sample_41	no	no	yes	yes	2
sample_51	yes	yes	yes	yes	4
sample_53	yes	yes	yes	yes	4
sample_68	yes	yes	yes	yes	4
sample_84	no	no	yes	yes	2
sample_86	yes	yes	yes	yes	4
sample_90	yes	yes	no	no	2
Cluster size	8	9	10	10	

For example, the hsa-miR-93-5p and hsa-let-7d-5p miRNAs were not expressed in any samples from the small cluster, while the hsa-miR-181a-5p miRNA was expressed only in one sample. These miRNAs were expressed only in a small fraction of the large cluster. Considering these expression patterns, we carried out a systematic analysis and validated if the expression of specific miRNAs is significantly associated with the small and large groups using a Fisher test. Based on the test results, we found 19 miRNAs, whose expression patterns are associated with the clusters (Table 3).

Interestingly, we only found miRNAs, whose lack of expression was associated with the small cluster, and did not find any, whose presence of expression was associated with the same cluster. Based on these patterns, we initially thought that the small cluster might be the result of a particularly bad quality sample group or strong technical bias. However, based on the normalized, batch-corrected PCA analysis of the expression data, this does not seem to be true. The cluster remains even after correcting for all known technical biases, and also does not seem to be associated with age or sex.

Table 3. Small cluster associated miRNAs. List of miRNAs that show a significant association with the “small cluster”. The table shows the miRNA id (mirna_id), the number of samples in the two clusters with expressed/not expressed status (Big_exp, Small_exp, Big_noexp, Small_noexp), the Fisher-test p-value (fisher_p), odds ratio (fisher_odds) and FDR (fisher_fdr).

mirna_id	Big_exp	Small_exp	Big_noexp	Small_noexp	fisher_p	fisher_odds	fisher_fdr
hsa-miR-93-5p	61	0	4	8	3,682E-08	Inf	2,876E-05
hsa-miR-181a-5p	63	1	2	7	1,721E-07	164,12768	6,719E-05
hsa-let-7d-5p	56	0	9	8	1,808E-06	Inf	0,0004708
hsa-let-7g-5p	65	3	0	5	3,728E-06	Inf	0,0007279
hsa-miR-29a-3p	54	0	11	8	5,623E-06	Inf	0,0008783
hsa-let-7i-5p	59	1	6	7	7,755E-06	60,978791	0,0010095
hsa-miR-15a-5p	62	2	3	6	1,277E-05	52,781391	0,0014248
hsa-miR-25-3p	58	1	7	7	1,529E-05	52,196405	0,0014924
hsa-miR-126-3p	49	0	16	8	5,471E-05	Inf	0,0042732
hsa-miR-20a-5p+hsa-miR-20b-5p	56	1	9	7	4,947E-05	40,06057	0,0042732
hsa-miR-15b-5p	63	3	2	5	7,261E-05	44,949371	0,0051554
hsa-miR-222-3p	52	1	13	7	0,000315	26,391301	0,0205025
hsa-miR-106a-5p+hsa-miR-17-5p	43	0	22	8	0,0004354	Inf	0,0259381
hsa-miR-223-3p	51	1	14	7	0,000465	24,131636	0,0259381
hsa-let-7a-5p	63	4	2	4	0,0008769	27,961779	0,0370103
hsa-miR-191-5p	63	4	2	4	0,0008769	27,961779	0,0370103
hsa-miR-199a-3p+hsa-miR-199b-3p	41	0	24	8	0,0007825	Inf	0,0370103
hsa-miR-23a-3p	65	5	0	3	0,0009004	Inf	0,0370103
hsa-miR-29b-3p	65	5	0	3	0,0009004	Inf	0,0370103

After the binary clustering analysis, we repeated sample clustering using a k-means based method and instead of the binary expression data, we used normalized expression values (Figure 11). Based on visual inspection, we decided to use 4 clusters for the k-means algorithm, both for the sample and the miRNA level analysis. Interestingly, the second largest k-means sample cluster (SCluster1) consisting of 10 samples, partially overlaps with the “small” cluster defined in the binary expression analysis (Table 2). In summary, both the binary expression clustering and the k-means clustering defined a small set of samples with markedly distinct miRNA expression patterns from the rest.

After inspecting the miRNA k-means clusters, we noticed that the expression of miRNAs in the second largest cluster (MCluster1) in SCluster1 is low for most of the samples. We checked the overlap of MCluster1 miRNAs with the 19 miRNAs that had significantly different binary expression in the previous analysis. All of the 19 miRNAs were present in the new k-means based cluster. Therefore, MCluster1 contains those miRNAs whose expression pattern defines the small sample cluster seen in all of the above analysis.

Repeating both analyses using only PCNSL samples lead to similar results, with the only difference being that the “small” binary cluster contained an additional sample (sample no. 33).

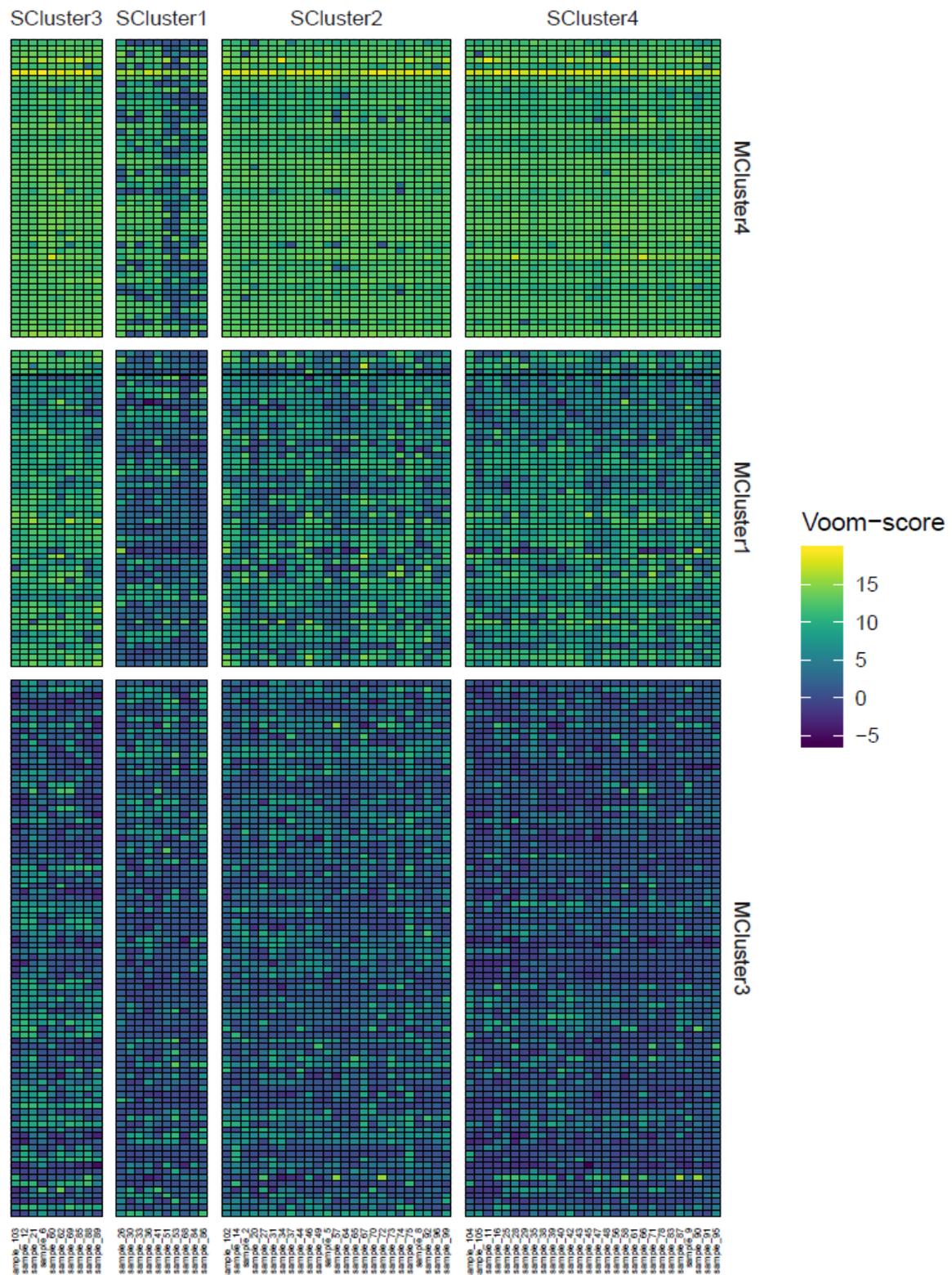


Figure 11. Unsupervised clustering of samples using normalized miRNA expression patterns. The heatmap shows the k-means clustering of miRNAs after normalization, voom transformation and batch effect removal. Columns correspond to samples, while rows correspond to a specific miRNA. Heatmap colors show expression intensity. We discarded the largest miRNA cluster (Mcluster2) as it contained mainly miRNAs with zero or very low expression.

5D, Pathway enrichment of differentially expressed miRNAs

Based on our data, pathway enrichment analysis revealed several downregulated pathways and gene sets in SCNSL compared with PCNSL. Additionally, PRDM1 mutation was also associated with the downregulation of several pathways (Figure 12.). Even though the evidence is circumstantial, as we did not directly measure the differential regulation of the genes comprising a pathway, but only their regulators, these pathways might be attractive targets for future drug development. Based on our results, NF- κ B and PI3K-MTOR-AKT pathways are generally more active in PCNSL compared to SCNSL and drugs available to target these pathways might be more effective for a selection of PCNSL cases. The unfolded protein response pathway was similarly upregulated in PCNSL cases. This pathway is considered as a general pro-survival mechanism for cancer cells, and small molecule inhibitors targeting the pathway are becoming available, suggesting a possible therapeutic target in PCNSL, as these lymphomas might be more sensitive to treatment. Considering that PCNSL is a rare and aggressive disease and the prognosis is poor, the pathways and molecular mechanisms analyzed in this study might be considered as novel drug targets.

Considering the pathway level changes in PRDM1 mutated samples, some drugs might be more effective in patients without PRDM1 mutations. The TGF-Beta signaling, the PI3K-MTOR-AKT, MYC target genes, G2M checkpoint genes and androgen response genes are all downregulated in samples with PRDM1 mutation, therefore the efficiency of their inhibitors might be decreased.

Pathway enrichment analysis shows similar pattern in the small distinct set of samples to PCNSL, with even more pronounced changes compared to the SCNSL vs PCNSL analysis. In addition, the WNT/beta-catenin pathway is also activated here. The small set of samples defined by our analysis might be a distinct PCNSL subgroup where patients might benefit from Wnt inhibitors.

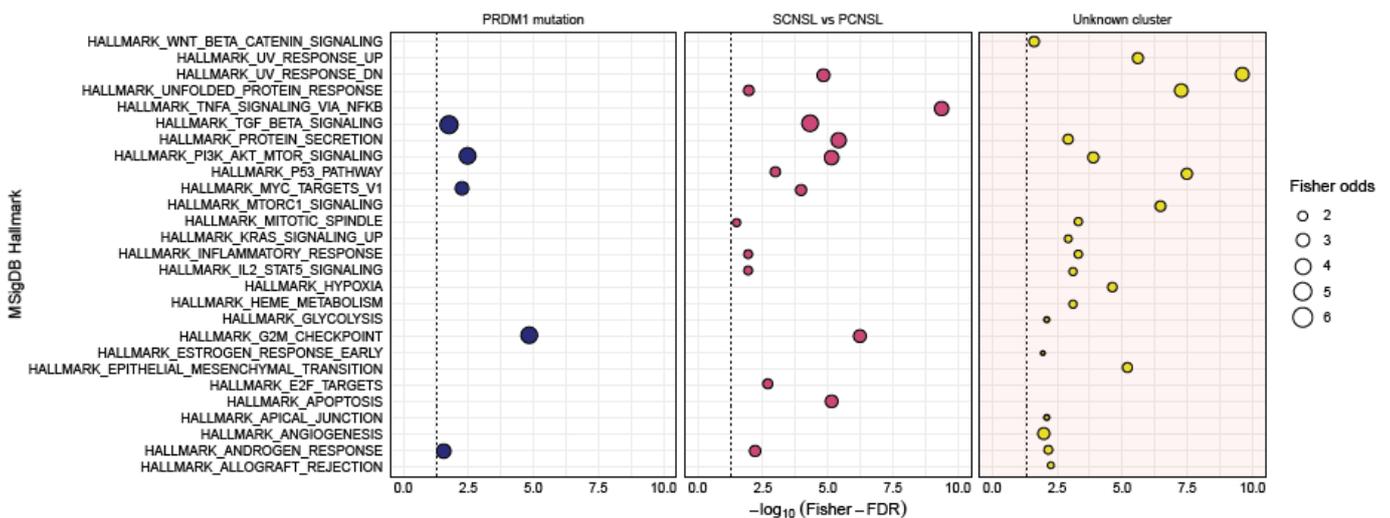


Figure 12. Gene set enrichment analysis. MSigDB pathways deregulated in specific comparisons, based on the differential expression patterns of miRNAs, from PRDM1 mutated and non-mutated or secondary and primary samples. Additionally, the figure shows the putatively deregulated pathways using the 19 miRNAs showing a significant association with the unknown “small cluster” based on the unsupervised clustering of expression data.

5E, miRNA expression profile association to survival characteristics

Survival analysis of all cases using the binary expression data showed miR-4488 to be significantly associated with a worse overall survival (Figure 13.), however, we did not find any correlation when analyzing PCNSL samples solely. It is important to highlight that these results are limited by the modest number of cases and the heterogeneous nature of the treatment regimens applied in this cohort.

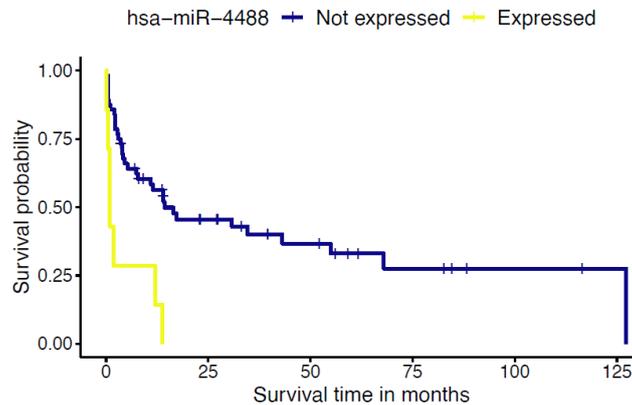


Figure 13. Survival curve of patients stratified based on the binary expression (ON/OFF) of hsa-miR-4488.

5F, Summary

In summary, our study identifies a novel CNS lymphoma subgroup defined by distinct miRNA expression patterns, describes putative subtype and cell-of-origin biomarkers, and proves the importance of specific miRNAs and pathways in CNS lymphoma pathogenesis. These results provide the basis for future research in the area of CNS lymphoma molecular diagnostics.

5G, Publishing of the miR expression signature study

We published our miR expression data in the Gene Expression Omnibus (GEO) database (NCBI). However, accession "GSE162956" is currently private and is scheduled to be released on Dec 31, 2021, or when the related manuscript has been accepted.

- <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE162956>

We prepared a manuscript entitled “*Distinct microRNA expression signatures of primary and secondary central nervous system lymphomas*” which is currently under review.

We also published a preprint version of our manuscript at medRxiv Pathology:

- <https://www.medrxiv.org/content/10.1101/2021.02.05.21249862v1>