1 Distinct microRNA expression signatures of primary and secondary central nervous 2 system lymphomas

3 Endre Sebestyén¹, Ákos Nagy², Dóra Marosvári², Hajnalka Rajnai¹, Béla Kajtár³, Beáta
4 Deák⁴, András Matolcsy^{1,2,5}, Sebastian Brandner⁶, James Storhoff⁷, Ning Chen⁷, Attila G.
5 Bagó⁸, Csaba Bödör², Lilla Reiniger^{1,9}

6

7 1, 1st Department of Pathology and Experimental Cancer Research, Semmelweis University,
8 Budapest, Hungary

9 2, MTA-SE Momentum Molecular Oncohematology Research Group, 1st Department of
10 Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary

11 3, Department of Pathology, University of Pécs, Clinical Centre, Pécs, Hungary

12 4, Department of Medical Oncology and Haematology, National Institute of Oncology,
13 Budapest, Hungary

14 5, Department of Laboratory Medicine, Division of Pathology, Karolinska Institutet,
15 Karolinska University Hospital, 171 77 Stockholm, Sweden

16 6, Division of Neuropathology, The National Hospital for Neurology and Neurosurgery,
17 University College London Hospitals NHS Foundation Trust and Department of
18 Neurodegenerative Disease, UCL Queen Square Institute of Neurology, London, WC1N 3BG
19 United Kingdom

20 7, NanoString Technologies, Seattle, WA, USA

21 8, Department of Neurooncology, National Institute of Clinical Neurosciences, Budapest,
22 Hungary

23 9, SE-NAP Brain Metastasis Research Group, 2nd Department of Pathology, Semmelweis
24 University, Budapest, Hungary

25

26 **Correspondence:** Lilla Reiniger: reiniger.lilla@med.semmelweis-univ.hu

27 1st Department of Pathology and Experimental Cancer Research, Semmelweis University,
28 Üllői út 26, Budapest, Hungary, H-1085.

29 Tel.: + 36 1 459 1500/4435, Fax: + 36 1 317 1074

30

31 **Keywords:** microRNA, PCNSL, brain lymphoma, molecular subtype

32

33 **Author contributions:** Endre Sebestyén analyzed all NanoString data, interpreted the results
34 and wrote the paper, Ákos Nagy did the RT-PCR and ddPCR validation, Dóra Marosvári
35 isolated the RNA samples, Hajnalka Rajnai revised histology, Beáta Deák contributed
36 samples and clinical data, András Matolcsy revised histology, Sebastian Brandner contributed
37 samples and clinical data, James Storhoff and Ning Chen coordinated the NanoString studies,
38 Ning Chen performed NanoString data QC and analysis, Attila G. Bagó contributed samples
39 and clinical data, Csaba Bödör designed and coordinated the study, Lilla Reiniger designed
40 and coordinated the study, contributed samples, interpreted the results and wrote the paper.
41 All authors read and approved the final version of the manuscript.

42

43

Funding: This work was funded by the Hungarian Science Foundation (OTKA-PD115792 to Lilla Reiniger) and Hungarian National Research, Development and Innovation Office (NKFIH) (KH17-126718 to Csaba Bödör, NVKP_16-1-2016-0004 to András Matolcsy and FK-132666 to Endre Sebestyén). Furthermore, the study was supported by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary within the framework of the Molecular Biology thematic programme of the Semmelweis University to Csaba Bödör, the Hungarian Brain Research Program (2017-1.2.1-NKP-2017-00002 to Lilla Reiniger), Semmelweis University Science and Innovation Fund (STIA_18_KF to Endre Sebestyén and STIA_21_KF to Csaba Bödör), Elixir Hungary (to Csaba Bödör) and the Research Grant of the University of Pécs Medical School (KA-2019-32 to Béla Kajtár).

The UK Brain Archive Information Network (BRAIN UK) is funded by the Medical Research Council and Brain Tumour Research. Sebastian Brandner was partly supported by the National Institute for Health Research Biomedical Research Centre's funding scheme to UCLH.

Acknowledgements: We thank Rachel Bradshaw from NanoString Technologies for the microRNA profiling, Mingdong Liu from NanoString Technologies for help in NanoString data QC and analysis, and Zoltán Szállási for critical reading of the manuscript.

62

63

64 Abstract

65 Central nervous system (CNS) lymphoma is a rare and aggressive non-Hodgkin lymphoma
66 that might arise in the CNS (primary CNS lymphoma, PCNSL) or disseminates from a
67 systemic lymphoma to the CNS (secondary CNS lymphoma, SCNSL). Dysregulated
68 expression of microRNAs (miRNAs) is associated with various pathological processes and
69 miRNA expression patterns may have diagnostic, prognostic and therapeutic implications.
70 However, miRNA expression is understudied in CNS lymphomas. Here, we performed
71 expression analysis of 798 miRNAs in 73 CNS lymphoma samples using the NanoString
72 platform, followed by a detailed statistical analysis to identify potential novel biomarkers
73 characterizing subgroups and to examine differences based on their primary and secondary
74 nature, molecular subtype, mutational patterns and survival. We describe the general
75 expression patterns of miRNAs across CNS lymphoma samples and identified 31
76 differentially expressed miRNAs between primary and secondary groups. Additionally, we
77 identified 7 more miRNAs associated with a molecular subtype and 25 associated with
78 mutation status. Using unsupervised clustering methods, we defined a small but distinct
79 primary CNS lymphoma subgroup, with characteristically different expression patterns
80 compared to the rest of the cases. Finally, we identified differentially regulated pathways in
81 the above comparisons and assessed the utility of miRNA expression patterns in predicting
82 survival. Our study identifies a novel CNS lymphoma subgroup defined by distinct miRNAs,
83 proves the importance of specific miRNAs and pathways in their pathogenesis, and provides
84 the basis for future research.

85

86

87

88

89 **Background**

90 Central nervous system (CNS) lymphoma is a rare and aggressive non-Hodgkin lymphoma
91 that either arises in the CNS structures (primary CNS lymphoma, PCNSL) or disseminates
92 from a systemic lymphoma to the CNS (secondary CNS lymphoma, SCNSL). Histologically,
93 it predominantly manifests as a diffuse large B-cell lymphoma (DLBCL). Effective treatment
94 of CNS lymphomas remains a significant challenge as their molecular pathogenesis is not
95 well understood [1-4].

96 Both from a therapeutic and prognostic point of view, it is becoming increasingly important to
97 precisely define the molecular subtype of DLBCLs into germinal center B-cell (GC) type,
98 activated B-cell (ABC) type or “unclassified” (UC) cases, as described by Alizadeh *et al.* [5].
99 Patients in the ABC-type DLBCL group show an inferior outcome [5, 6] compared to the
100 other types. The fundamental difference in biology including oncogenic pathways and
101 mutation targets between GC- and ABC-type DLBCLs is also reflected in the different
102 efficacy of novel targeted therapies between these subgroups [7, 8]. A more precise, gene
103 expression-based molecular subtype assignment can be achieved from formalin-fixed paraffin
104 embedded (FFPE) tissue using the NanoString Lymphoma Subtyping Test (LST) assay
105 (NanoString Technologies, Inc., Seattle, USA) compared to the standard
106 immunohistochemical (IHC) methods. The NanoString assay also demonstrates a better
107 concordance with the gold-standard Affymetrix approach [9].

108 The discovery of microRNAs (miRNAs, miRs) has opened a new field for unraveling and
109 therapeutically targeting diseases. These small non-coding RNAs regulate diverse biological
110 processes through post-transcriptional gene expression modulation. Based on their seed
111 sequence, miRNAs bind to multiple target mRNAs, thereby promoting their degradation or
112 inhibiting translation [10-12]. Dysregulated expression of miRNAs is associated with a

myriad of pathological processes including hematological malignancies [13, 14], and distinct miRNA expression patterns may also have diagnostic, prognostic and therapeutic implications [15-19]. As miRNAs remain relatively well preserved in archival FFPE tissue specimens, they are readily available as a valuable source of information in cancer tissues [20-22]. For the quantification of miRNA transcripts in FFPE samples, the NanoString nCounter technology is a preferable choice over quantitative reverse transcription polymerase chain reaction (RT-PCR) [23, 24], with a high reproducibility similar to other platforms [25, 26]. The NanoString assay also performs well in relative quantification studies [27].

MiRNA expression of PCNSL has been studied using different methods such as RT-PCR [28-38], microarray [36, 39], *in situ* hybridization [37, 38], next-generation sequencing (NGS) [35] and NanoString [38] technologies on various tissue types including brain biopsy specimens (FFPE [32, 37, 39] or fresh [36]), cerebrospinal fluid (CSF) [28-31, 33, 38] and peripheral blood [35], serum [34] or plasma [33]. It has been shown and further confirmed that the combined detection of miR-21, miR-19b and miR-92a in CSF allowed a reliable diagnosis of PCNSL. Moreover, these miRNAs emerge as promising tools in treatment monitoring and follow-up [28, 29], similarly to U2 small nuclear RNA fragments [30]. In addition, plasma miR-21 may serve as a diagnostic [33], and serum miR-21 both as a diagnostic and prognostic marker for PCNSL [34]. Other miRNAs with prognostic value in PCNSL include miR-151a-5p and miR-151b, together with 10 additional miRs [35]. CSF levels of miR-21 may have potential as a predictor of chemotherapeutic effect [33]. Measuring miR-30c in the CSF can differentiate between PCNSL and SCNSL, as elevated levels of miR-30c have pathobiological significance in SCNSL [31]. A recent microarray study found a couple of miRNAs with positive or negative prognostic significance in PCNSL [36]. It has also been demonstrated, that PCNSL shows different miRNA expression profiles compared with nodal or testicular DLBCL [32, 37, 39].

In this study, we performed expression profiling of 798 human miRNAs in 73 FFPE brain biopsy samples of primary and secondary CNS lymphomas using the NanoString platform, followed by a 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.

Methods

Sample collection and patient information

FFPE brain biopsy specimens of 64 patients with PCNSL and 9 patients with SCNSL were analyzed in this study. Tissue samples were obtained from three centers: (i) 1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary; (ii) Department of Pathology, University of Pécs, Pécs, Hungary and (iii) Division of Neuropathology, The National Hospital for Neurology and Neurosurgery, University College London Hospitals, United Kingdom, through the UK Brain Archive Information Network (BRAIN UK). Permissions to use the archived tissue have been obtained from the Local Ethical Committee (TUKEB-1552012) and from BRAIN UK (Ref.: 16/018), and the study was conducted in accordance with the Declaration of Helsinki.

Information on the molecular subtypes determined by the Research Use Only version of the NanoString LST-assay (NanoString Technologies, Inc., Seattle, USA) and on the mutational status identified by ultra-deep NGS of 14 target genes as described previously [40] are summarized in Supplementary file 1, together with clinical and survival data. Survival data was available in 54 PCNSL and 9 SCNSL cases. The PCNSL cases consisted of 78.1% (50/64) ABC, 15.6% (10/64) GC and 6.3% (4/64) UC molecular subtypes. Among the SCNSL cases, 44.4% (4/9) were classified as ABC- and 55.6% (5/9) as GC-subtypes (Figure 1A).

MicroRNA profiling using the NanoString platform

RNA isolation from 64 PCNSL and 9 SCNSL samples was performed using the RecoverAll™ kit (Life Technologies/Ambion, Inc, Foster City USA) according to the manufacturer's instructions. Approximately 100 ng of total RNA from each sample was analyzed using the Human v3 miRNA expression assay kit, following manufacturer's

instructions (NanoString Technologies, Inc., Seattle, USA). The Human v3 miRNA assay covers 98% of miRNA sequences found in miRBase v22, including 798 probes that recognize human miRNAs and 29 assay control probes. Raw data were pulled from digital analyzer and imported into nSolver4.0 (NanoStringtm) for data quality check. Data sets that passed all QC checks (including imaging quality, binding density, positive control linearity and limit of detection) were exported as a .csv file for downstream statistical analysis. Raw miRNA counts for all samples are available in the GSE162956 dataset at NCBI GEO.

NanoString data normalization

Raw NanoString read counts and clinical information was imported into the R statistical environment (version 3.5.1). We discarded human miRNAs with extremely low expression (read count ≤ 1 in more than 60 samples) before normalization, resulting in 781 human miRNAs, besides the control probes and housekeeping genes. Data normalization was done with the NanoStringNorm package (version 1.2.1) [41] using the following options: CodeCount = “geo.mean”, Background = “mean.2sd”, SampleContent = “housekeeping.geo.mean”, OtherNorm = “quantile”. We used the ath, cel, osa and NEG miRNA classes as negative controls, the POS class as positive controls, the *ACTB*, *B2M*, *GAPDH*, *RPL19* and *RPLP0* genes as housekeeping genes, and everything else as endogenous miRNAs. During the initial data assessment, we considered those miRNAs expressed, that had a larger normalized expression value than the lowest positive control (Addt 2). Additionally, we calculated the minimum, maximum and average expression of all miRNAs across all samples, together with the number of samples where a miRNA was expressed across all samples using the normalized expression values. Finally, we calculated the minimum, maximum, average and median values for the PCNSL and SCNSL samples

separately. Based on the median expression values we ranked the miRNAs and categorized them into 15 separate expression groups.

Differential expression analysis

We used the limma package (version 3.38.3) for differential expression [42]. We included the primary or secondary category, the molecular subtype, origin of sample (institute or department), degradation time and NanoString scan dates in the linear model as covariates during differential expression analysis. We converted the normalized read counts with the voom [43] function of limma, fitted the linear model with the lmFit function and calculated p-values with the eBayes function [44]. The raw p-values were false discovery rate (FDR) adjusted using the Benjamini-Hochberg method. We considered miRNAs differentially expressed with FDR adjusted p-value < 0.05 and absolute log2 fold change between conditions > 1 (Supplementary file 3).

We carried out a number of comparisons between different sample groups. First, we calculated differentially expressed miRNAs between the primary and secondary samples. Additionally, we calculated differentially expressed miRNAs between the different molecular subtypes, within the primary samples. This included the GC vs ABC, UC vs ABC and GC vs UC comparisons. Finally, we calculated differences between the different sample groups stratified by mutation status, using the *CARD11*, *CCND3*, *CD79B*, *CSMD2*, *CSMD3*, *IRF4*, *KMT2D*, *MYC*, *MYD88*, *PAX5*, *PIM1*, *PRDM1* and *TP53* genes (Supplementary file 3).

Principal component analysis and unsupervised clustering

During the principal component analysis, we used the normalized and voom transformed data, additionally removing the potential batch effects of the institute, scan date and degradation time with the removeBatchEffect function of the limma package. We calculated the principal

components with the prcomp R function with the scale. parameter set to TRUE, and plotted the first and second principal components.

During the binary hierarchical clustering of the samples, we used the voom transformed, normalized read counts of all endogenous miRNAs, after removing the potential batch effects of the institute, scan date and degradation time with the removeBatchEffect function of the limma package. First, we binarized the expression values [45] with a cut-off of 12. All miRNAs below this threshold were considered not expressed, while the rest was considered expressed. Using the expressed/not expressed status of the miRNAs across all samples, we calculated sample distances using the dist function from R, with the binary method, and did a hierarchical clustering using the hclust function, with the ward.D2 method. We manually defined a “small” and a “big” cluster of samples (Supplementary file 4), and carried out a Fisher-test, using a 2x2 contingency table (expressed/not expressed and small/big cluster) to check for miRNAs whose expression is associated with the small and big clusters (Supplementary file 5). The Fisher-test p-values were FDR adjusted using the Benjamini-Hochberg method and miRNAs with FDR adjusted p-value < 0.05 were considered associated with the cluster type. We repeated the binary hierarchical clustering using only the primary samples without changing any other parameter.

Additionally, we performed a k-means clustering on the data, where we also used the voom transformed, normalized read counts of all endogenous miRNAs after removing the potential batch effects of the institute, scan date and degradation time with the removeBatchEffect function of the limma package. We used kmeans function of R, with the iter.max = 1000 parameter and setting the cluster number to 4. We clustered both the samples and the miRNAs. During data visualization we dropped the largest miRNA cluster as it mainly contained miRNAs that were not expressed in most samples or had a very low expression

level. We repeated the k-means clustering using only the primary samples without changing any other parameter.

miRNA pathway enrichment analysis

We used the miRNet database (accessed and downloaded on 2019-08-29) to collect putative targets of the differentially expressed miRNAs [46]. We downloaded all gene and lncRNA targets of these miRNAs from bone-marrow or brain tissue-based experiments. We considered a gene up- or downregulated if at least two of its putative regulators from miRNet were differentially expressed in the opposite direction in a specific tissue type. For example, the AGO1 gene was considered downregulated in the secondary vs primary comparison, as 7 of its regulatory miRNAs defined in miRNet were upregulated in the secondary vs primary comparison based on the NanoString analysis. Based on this filtering criteria, only bone marrow-based mRNA interactions were considered for further downstream analysis. Additionally, we downloaded version 7 of the MSigDB gene sets (H, C2 – C7) [47] and did a Fisher-test using a 2x2 contingency table as follows: genes were either part of a gene collection or not, and based on the miRNet analysis they were putatively differentially expressed or not. The Fisher-test p-values were FDR adjusted using the Benjamini-Hochberg method and gene collections with FDR adjusted p-value < 0.05 were considered enriched in a specific comparison (secondary vs primary, etc) (Supplementary file 6).

Survival analysis

Survival data was available for 63 patients with the overall survival in months, last follow-up date, and survival status. We used only the 54 primary CNS lymphoma cases during this analysis, and stratified patients according to the binary expression status of the investigated miRNAs, or the big/small cluster classification. We used the survival package (version 2.44-

1.1) [48] to create Surv objects in R and calculated log-rank test based p-values using the survminer (version 0.4.6) [49] package. The log-rank p-values were FDR adjusted using the Benjamini-Hochberg method (Supplementary file 7).

Validation of NanoString miRNA expression data

10 miRNAs with significant differential expression among subgroups were selected (Supplementary file 8 and 9) based on NanoString expression data. Additionally, 3 miRNAs with stable expression in the cohort were used as endogenous control.

Validation with droplet digital PCR

The selected samples were individually reverse-transcribed with the TaqMan™ MicroRNA Reverse Transcription Kit (Applied Biosystems, USA) following manufacturer's instructions. Hundred ng miRNA from each selected sample was reverse-transcribed with specific TaqMan™ miRNA assays (Applied Biosystems, USA) in ProFlex Thermal Cyclers (Applied Biosystems, USA) using the following parameter values: 16°C for 30 min, 42°C for 30 and 85°C for 5 min. cDNAs were further diluted with 50 ul nuclease-free water to obtain a final volume of 77 ul. 10 ul diluted cDNA sample were used in the subsequent droplet digital PCR (ddPCR) reactions with 1 ul of the specific TaqMan™ miRNA assays and 11 ul of ddPCR Supermix for Probes (No dUTP) (Bio-Rad Laboratories, USA). Following droplet generation, target miRNAs were amplified in a C1000 Touch Thermal Cycler (Bio-Rad Laboratories, USA) with parameters as follows: 10 min at 95°C for enzyme activation, followed by 40 cycles of denaturation at 94°C for 30 s and 1 min annealing/extension at 55°C, enzyme deactivation step set at 98°C for 10 min, and a final hold step at 4°C for an infinite time period. Results were analyzed using the QuantaSoft software (version 1.7; Bio-Rad, USA). All ddPCR reactions were performed with the detection of adequate events (>10000 droplets per sample). Results were determined in copy number per ul.

294 *Validation with quantitative RT-PCR*

295 Further NanoString miRNA expression data validation was completed with quantitative RT-
 296 PCR. Reverse Transcription was performed as described above with the exception that each
 297 miRNA target was individually reverse-transcribed with an endogenous control miRNA. Five
 298 ul of the cDNA samples were amplified in duplicates using 1.5 ul of the target and
 299 endogenous control specific TaqMan™ miRNA assays with 7.7 ul 2x TaqMan puffer
 300 (Applied Biosystems, USA) and 10.5 ul nuclease-free water on a Quantstudio 3 Real-Time
 301 PCR system (Applied Biosystems, USA). Ct values for sample target and endogenous
 302 miRNAs were determined as the average Ct value of the duplicates. Δ -Ct values were
 303 obtained extracting Ct values of endogenous control miRNAs from the Ct values of target
 304 miRNAs, which were subsequently used for statistical analysis.

305

306

Results

General miRNA expression patterns in CNS lymphomas

The NanoString panel contained 798 endogenous, 5 housekeeping genes (*ACTB*, *B2M*, *GAPDH*, *RPL19*, *RPLP0*), 9 positive and 16 negative controls. Based on the normalized expression values, almost half of the investigated miRNAs (381) were not expressed in any sample, and 54 were expressed only in one. On the other hand, 12 miRNAs were expressed in 72 and 2 miRNAs in all 73 samples (Supplementary file 2).

We compared the median expression of endogenous miRNAs between PCNSL and SCNSL samples (Figure 1B). miRNAs with a non-zero median expression in both groups were highly correlated (Spearman correlation 0.89). We found 676 and 661 miRNAs in PCNSL and SCNSL respectively, that had rank in the bottom 10 % based on their median expression. On the other hand, hsa-miR-4454+hsa-miR-7975 is an extreme outlier, with a larger median expression than most of the positive controls. We noticed that the hsa-miR-4454 expression detected by NanoString might be confounded by the expression of tRNA^{His}, as the miRNA is identical to the 3' end of this tRNA [50].

Differential miRNA expression between primary and secondary CNS lymphomas and molecular subtypes

We analyzed differential expression patterns of the miRNAs between PCNSL and SCNSL cases and different molecular subtypes using the limma R package, after removing known confounding factors. With an absolute log₂ fold change > 1 and FDR value < 0.05 cut-off, we found 31, 4 and 3 differentially expressed miRNAs between PCNSL and SCNSL cases, GC and ABC subtypes, and UC and ABC subtypes, respectively (Figure 2A, Table 1 and Supplementary file 3). Twenty-eight miRNAs showed increased and three showed decreased

expression in SCNSL compared to PCNSL. Three miRNAs showed lower and one miRNA higher expression in GC compared to ABC cases. All three differentially expressed miRNAs showed lower levels in UC compared to ABC subtypes (Supplementary file 3).

miRNA expression profile association to mutation status

We assessed whether miRNA expression profiles were associated with the mutation status of specific genes. We found 8 differentially expressed miRNAs associated with the mutation status of the *PRDM1* gene, and one miRNA associated with the mutation status of *MYC* and *CARD11* genes (Figure 2B, Table 1 and Supplementary file 3).

We also identified 8, 5 and 2 differentially expressed miRNAs in cases with *PAX5*, *CSMD3* and *CSMD2* mutations. However, there were only two samples harboring mutation in each of these genes, therefore, the results must be considered with caution (Supplementary file 3).

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 3). Additionally, the samples do not cluster based on RNA isolation group, place of origin, scanning date, degradation time, % tumor content, age or sex (Supplementary Figure 1 – 8). 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.

First, we did a binary clustering of the data (Figure 4), 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 (Supplementary file 4). The binary expression patterns of this group are markedly different from the larger group. 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 (Supplementary file 5). 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 (Supplementary Figures 1-8) 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.

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 5). 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 (Supplementary file 4). 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) (Supplementary Figures 9 and 10).

Pathway enrichment of differentially expressed miRNAs

After the differential expression analysis, we asked if specific pathways might be up- or downregulated in a specific comparison. Assuming that the differential regulation of a specific miRNA will lead to the differential regulation of its target mRNAs in the opposite direction, we carried out a pathway enrichment analysis, based on validated miRNA-mRNA interactions from the miRNET database. Due to the low number of differentially expressed miRNAs, we were able to do this analysis only in the case of SCNSL vs PCNSL and *PRDM1* mutated vs non-mutated comparisons. Focusing on the hallmarks (H) gene sets of MSigDB, we found several pathways that are putatively downregulated in these cases (Figure 6 and Supplementary file 6). The G2M checkpoint, PI3K-AKT-MTOR signaling, TGF-beta signaling pathways, MYC target and androgen response genes were downregulated in both comparisons besides a number of other pathways or gene sets. Additionally, in the SCNSL vs PCNSL comparison TNF-alpha signaling, apoptosis and UV response genes, the P53 pathway and E2F target genes were also downregulated.

We carried out a similar analysis using the 19 miRNAs defined during the binary clustering, that are not expressed in our “small cluster”. However, in this case, as the miRNAs are not expressed in the small sample cluster, the associated enriched pathways are putatively upregulated in the same samples. Pathways include the unfolded protein response, connected to cellular stress, the P53 and MTORC1 signaling pathways, the epithelial to mesenchymal transition or the hypoxia gene sets and UV response genes.

Additional, more detailed results for the SCNSL vs PCNSL, *PRDMI* mutated vs non-mutated and “small cluster” analysis using the C2 – C7 gene sets are available in Supplementary file 6.

miRNA expression profile association to survival characteristics

Finally, we asked whether the expression profile of specific miRNAs is associated to patient survival. We stratified patients according to the binary expression status of miRNAs and using the log-rank test, calculated the significance of overall survival (OS) differences between groups (Supplementary file 7) including both PCNSL and SCNSL samples. The only miRNA showing difference was hsa-miR-4488, with an FDR of 0.022 (Supplementary Figure 11 and 12) where expression of the miRNA was associated with worse survival. Median survival for patients with “Expressed” or “Not expressed” status of hsa-miR-4488 was one and 14 months, respectively. Repeating the analysis with only PCNSL samples (Supplementary file 7), we found no significant differences between groups after FDR correction. hsa-miR-18a-5p had the lowest FDR value (0.12). Of note, the survival analysis is hampered by the relatively low number of cases and the heterogeneous nature of the treatment regimens applied in this cohort.

Validation of NanoString results

Based on the differential expression results, we selected a number of miRNAs for additional validation. We used quantitative RT-PCR and ddPCR for different miRNAs. (Supplementary file 8). Using the RT-PCR Δ -Ct values and the ddPCR counts, we correlated the validation results with the original and normalized NanoString read counts (Supplementary file 9). RT-PCR Δ -Ct values are expected to demonstrate a high negative correlation, while ddPCR counts are expected to show a high positive correlation with NanoString read counts or normalized expression values. We validated hsa-miR-411-5p and hsa-miR-32-5p using RT-PCR, where the correlation of Δ -Ct values with the normalized read counts was ≤ -0.6 . Additionally, we validated hsa-let7g-5p, hsa-miR-191-5p and hsa-miR-379-5p using ddPCR where the correlation of ddPCR read counts with the NanoString read counts was ≥ 0.6 , besides hsa-miR-411-5p where the correlation of normalized read count with ddPCR counts was 0.67 and the correlation of the original NanoString read count with the ddPCR count was 0.47.

Discussion

CNS lymphomas represent a considerable clinical challenge, as their molecular pathogenesis is poorly explored. CNS lymphomas are difficult to investigate as they are rare, and usually small biopsies are obtained for diagnostic purposes, which may not be sufficient for further analyses [3, 4] in many cases. Moreover, the majority of these biopsies are archived as FFPE tissue blocks, which suffer from degradation of macromolecules [51-56]. However, miRNAs provide a robust signal and can be stably extracted from FFPE samples [20-22], making them ideal cancer biomarkers. They are also promising targets for molecular therapy in cancer, due to their role as oncomiRs or tumor-suppressors [57-59]. The NanoString nCounter technology allows direct quantitation of hundreds of miRNA transcripts from FFPE tissues with outstanding performance, therefore it may be a preferable choice over other transcriptomic methods [23-26].

Previous expression studies of PCNSL revealed various miRNAs with a potential diagnostic [28, 29, 31, 33, 34], prognostic [34-36] or predictive [33] value. These datasets were generated by different methods, and used diverse patient-derived sample types and control tissues [28-39], thus being difficult to synthesize and integrate. Moreover, most studies examined only a few miRNAs [28-30, 33, 34, 37] on a limited number of CNS lymphoma cases [28, 29, 32-39]. There is only a single study in the literature discussing the miRNA expression differences between PCNSL and SCNSL [31]. The authors demonstrated in CSF samples that miR-16, miR-30b, miR-30c, miR-191 and miR-204 were upregulated and miR-222 was downregulated in SCNSL, with miR-30c showing the largest expression difference [31].

In this study, we performed expression profiling of 798 human miRNAs in a large number of CNS lymphoma cases. We compared the miRNA expression patterns of FFPE brain biopsy

specimens of primary and secondary CNS lymphomas to minimize the tissue bias that may well be presented with the examination of systemic (nodal) DLBCL. We used the NanoString platform, which is a reliable method for miRNA expression analysis of FFPE tissues [23-27].

We identified 28 up- and 3 downregulated miRNAs in SCNSL compared with PCNSL (Table 1). Out of these, 23 up- and all 3 downregulated miRNAs are completely novel and until now, were not described in the context of CNS lymphomas. Reassuringly, the remaining miRNAs described in the literature have similar expression patterns to our results. Amongst the already described miRNAs, we found miR-30c-5p, that is significantly increased in CSF samples of patients with SCNSL compared with PCNSL [31]. In general, miR-30c-5p has a tumor-suppressive role in cancer pathogenesis, and shows low expression in various malignancies (reviewed in [60]) which is in line with our findings (median expression: 11.18 and 0; rank: 137 and 147 in SCNSL and PCNSL samples, respectively) (Supplementary file 2). Its significant lower expression in PCNSL may contribute to their more aggressive behavior.

Multiple studies have found higher expression of miR-21, a well-known oncomiR [61], both in PCNSL and DLBCL cases compared to controls [16, 28, 29, 33, 34, 62-66]. In our study, miR-21-5p generally showed high expression (median expression: 3278.55 and 2926.61; rank: 16 and 17 in SCNSL and PCNSL samples, respectively) with a significant increase in SCNSL cases. Higher expression of miR-21 has also been associated with worse overall survival in DLBCL patients [67]. 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 (median expression: 606.62 and 58.34, 1088.28 and 383.75, and 763.30 and 263.06; rank: 70 and 94, 52 and 72, and 62 and 75, respectively in SCNSL and PCNSL samples). This miRNA cluster has a strong oncogene activity in various malignancies including DLBCL [68-71]. Higher expression of miR-18a was found to be associated with a shorter overall survival in DLBCL [72]. Previous studies demonstrated

other members of the miR-17-92 cluster (miR-17-5p and miR-20a) to be upregulated in PCNSL compared with nodal DLBCL [32, 37]. Moreover, high levels of miR-19b-1 and miR-92a-1 were detected in the CSF of PCNSL patients [28, 29].

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 [16, 19, 64, 65, 71, 73-76] and miR-222-3p [18, 71, 73] 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.

This is the first study demonstrating differentially expressed miRNAs in association with the mutational status of the *PRDM1*, *MYC* and *CARD11* genes in CNS lymphomas. According to the literature, the only miRNA that has already been connected to any of these genes is miR-30a-5p, which directly targets *PRDM1* and modulates the WNT/beta-catenin pathway [77]. However, this association is not connected to the mutational status of *PRDM1* itself.

It is widely known, that the different miRNA profiling platforms do not perform consistently [24-27]. 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.

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. 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. The constitutive activation of NF- κ B was already described in the literature [78-80] for DLBCL, apart from the activation of the PI3K-MTOR-AKT [79-81].

Based on our results, these pathways are generally more active in PCNSL compared to SCNSL and drugs [82] 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 [83], 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. Additional upregulated pathways in the PCNSL vs SCNSL comparison include protein secretion, the p53 pathway, MYC target genes, the G2M checkpoint, E2F transcription factor target genes, apoptosis genes, and genes related to androgen response. Considering that PCNSL is a rare and aggressive disease and the prognosis is poor [84], 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 [80, 82, 85].

Intriguingly, principal component analysis (PCA) showed no clustering of cases according to either disease characteristics or subtypes even after accounting for the various batch effects and biases. Subsequent binary clustering of the cases according to miRNA expression (expressed or not expressed) revealed a small group of 8 samples clearly separating from the rest of the cohort. Additional sample clustering using a k-means based method with normalized expression values defined a sample cluster (SCluster1) consisting of 10 samples. Five samples from SCluster1 also overlapped with the small cluster defined in the binary expression analysis. Taken together, unsupervised clustering methods defined a small set of

samples with markedly distinct miRNA expression patterns. Interestingly, the lack of expression of 19 miRNAs was found to be associated with the small cluster in the binary clustering analysis. Moreover, all of these 19 miRNAs were part of a miRNA k-means cluster (MCluster1) in SCluster1. Pathway enrichment analysis shows similar pattern in this small distinct set of samples to PCNSL, with even more pronounced changes compared to the SCNSL vs PCNSL analysis. The WNT/beta-catenin pathway activated here, was described as activated in PCNSL [86], and the small set of samples defined here might be a distinct PCNSL subgroup where the pathway can be efficiently targeted with Wnt inhibitors [87]. Based on these results, this is a well-defined sample group within the PCNSL cases contributing significantly to the distinct expression patterns between PCNSL and SCNSL.

Survival analysis of all cases using the binary expression data showed miR-4488 to be significantly associated with a worse overall survival, however, we did not find any association 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.

Conclusions

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 lymphomas.

560 **Declarations**

561 *Ethics approval and consent to participate*

562 Permissions to use the archived tissue have been obtained from the Local Ethical Committee
563 (TUKEB-1552012) and from BRAIN UK (Ref.: 16/018), and the study was conducted in
564 accordance with the Declaration of Helsinki.

565 *Consent for publication*

566 Not applicable.

567 *Availability of data and materials*

568 The datasets supporting the conclusions of this article are included within the article, its
569 additional files and at the NCBI GEO database with accession number GSE162956.

570 *Competing interests*

571 The authors declare the following competing interest. James Storhoff and Ning Chen were
572 employees of NanoString Technologies when the original NanoString analyses were carried
573 out. James Storhoff is currently an employee of Veracyte. Ning Chen is currently an
574 employee of Adaptive Biotechnologies.

575

576

References

1. Camilleri-Broet S, Criniere E, Broet P, Delwail V, Mokhtari K, Moreau A, Kujas M, Raphael M, Iraqi W, Sautes-Fridman C *et al*: **A uniform activated B-cell-like immunophenotype might explain the poor prognosis of primary central nervous system lymphomas: analysis of 83 cases.** *Blood* 2006, **107**(1):190-196.
2. Liu J, Wang Y, Liu Y, Liu Z, Cui Q, Ji N, Sun S, Wang B, Wang Y, Sun X *et al*: **Immunohistochemical profile and prognostic significance in primary central nervous system lymphoma: Analysis of 89 cases.** *Oncol Lett* 2017, **14**(5):5505-5512.
3. Grommes C, Rubenstein JL, DeAngelis LM, Ferreri AJM, Batchelor TT: **Comprehensive Approach to Diagnosis and Treatment of Newly Diagnosed Primary CNS Lymphoma.** *Neuro Oncol* 2018.
4. Hilal T: **Primary central nervous system lymphoma: Consensus, controversies, and future directions.** *ADVANCES IN CELL AND GENE THERAPY*, n/a(n/a):e82.
5. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X *et al*: **Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.** *Nature* 2000, **403**(6769):503-511.
6. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM: **A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma.** *Proc Natl Acad Sci U S A* 2003, **100**(17):9991-9996.
7. Karmali R, Gordon LI: **Molecular Subtyping in Diffuse Large B Cell Lymphoma: Closer to an Approach of Precision Therapy.** *Curr Treat Options Oncol* 2017, **18**(2):11.
8. Sujobert P, Salles G, Bachy E: **Molecular Classification of Diffuse Large B-cell Lymphoma: What Is Clinically Relevant?** *Hematol Oncol Clin North Am* 2016, **30**(6):1163-1177.
9. Scott DW, Wright GW, Williams PM, Lih CJ, Walsh W, Jaffe ES, Rosenwald A, Campo E, Chan WC, Connors JM *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):1214-1217.
10. Bartel DP: **MicroRNAs: target recognition and regulatory functions.** *Cell* 2009, **136**(2):215-233.
11. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB: **Prediction of mammalian microRNA targets.** *Cell* 2003, **115**(7):787-798.
12. Brennecke J, Stark A, Russell RB, Cohen SM: **Principles of microRNA-target recognition.** *PLoS Biol* 2005, **3**(3):e85.
13. Di Marco M, Ramassone A, Pagotto S, Anastasiadou E, Veronese A, Visone R: **MicroRNAs in Autoimmunity and Hematological Malignancies.** *Int J Mol Sci* 2018, **19**(10).
14. Lawrie CH: **MicroRNAs in hematological malignancies.** *Blood Rev* 2013, **27**(3):143-154.
15. Seto AG, Beatty X, Lynch JM, Hermreck M, Tetzlaff M, Duvic M, Jackson AL: **Cobomarsen, an oligonucleotide inhibitor of miR-155, co-ordinately regulates multiple survival pathways to reduce cellular proliferation and survival in cutaneous T-cell lymphoma.** *Br J Haematol* 2018, **183**(3):428-444.
16. Caramuta S, Lee L, Ozata DM, Akcakaya P, Georgii-Hemming P, Xie H, Amini RM, Lawrie CH, Enblad G, Larsson C *et al*: **Role of microRNAs and microRNA machinery in the pathogenesis of diffuse large B-cell lymphoma.** *Blood Cancer J* 2013, **3**:e152.
17. Roehle A, Hoefig KP, Repsilber D, Thorns C, Ziepert M, Wesche KO, Thiery M, Loeffler M, Klapper W, Pfreundschuh M *et al*: **MicroRNA signatures characterize diffuse large B-cell lymphomas and follicular lymphomas.** *Br J Haematol* 2008, **142**(5):732-744.
18. Montes-Moreno S, Martinez N, Sanchez-Espirdion B, Diaz Uriarte R, Rodriguez ME, Saez A, Montalban C, Gomez G, Pisano DG, Garcia JF *et al*: **miRNA expression in diffuse large B-cell lymphoma treated with chemoimmunotherapy.** *Blood* 2011, **118**(4):1034-1040.

- 626 19. Iqbal J, Shen Y, Huang X, Liu Y, Wake L, Liu C, Deffenbacher K, Lachel CM, Wang C, Rohr J *et al*: **Global microRNA expression profiling uncovers molecular markers for classification and**
- 627 **prognosis in aggressive B-cell lymphoma.** *Blood* 2015, **125**(7):1137-1145.
- 628
- 629 20. Khan J, Lieberman JA, Lockwood CM: **Variability in, variability out: best practice**
- 630 **recommendations to standardize pre-analytical variables in the detection of circulating and**
- 631 **tissue microRNAs.** *Clin Chem Lab Med* 2017, **55**(5):608-621.
- 632 21. Liu A, Tetzlaff MT, Vanbelle P, Elder D, Feldman M, Tobias JW, Sepulveda AR, Xu X: **MicroRNA**
- 633 **expression profiling outperforms mRNA expression profiling in formalin-fixed paraffin-**
- 634 **embedded tissues.** *Int J Clin Exp Pathol* 2009, **2**(6):519-527.
- 635 22. Hall JS, Taylor J, Valentine HR, Irlam JJ, Eustace A, Hoskin PJ, Miller CJ, West CM: **Enhanced**
- 636 **stability of microRNA expression facilitates classification of FFPE tumour samples**
- 637 **exhibiting near total mRNA degradation.** *Br J Cancer* 2012, **107**(4):684-694.
- 638 23. Eastel JM, Lam KW, Lee NL, Lok WY, Tsang AHF, Pei XM, Chan AKC, Cho WCS, Wong SCC: **Application of NanoString technologies in companion diagnostic development.** *Expert Rev*
- 639 *Mol Diagn* 2019, **19**(7):591-598.
- 640
- 641 24. Leichter AL, Purcell RV, Sullivan MJ, Eccles MR, Chatterjee A: **Multi-platform microRNA**
- 642 **profiling of hepatoblastoma patients using formalin fixed paraffin embedded archival**
- 643 **samples.** *Gigascience* 2015, **4**:54.
- 644 25. Kolbert CP, Feddersen RM, Rakhshan F, Grill DE, Simon G, Middha S, Jang JS, Simon V, Schultz
- 645 DA, Zschunke M *et al*: **Multi-platform analysis of microRNA expression measurements in**
- 646 **RNA from fresh frozen and FFPE tissues.** *PLoS One* 2013, **8**(1):e52517.
- 647 26. Chatterjee A, Leichter AL, Fan V, Tsai P, Purcell RV, Sullivan MJ, Eccles MR: **A cross**
- 648 **comparison of technologies for the detection of microRNAs in clinical FFPE samples of**
- 649 **hepatoblastoma patients.** *Sci Rep* 2015, **5**:10438.
- 650 27. Godoy PM, Barczak AJ, DeHoff P, Srinivasan S, Etheridge A, Galas D, Das S, Erle DJ, Laurent
- 651 LC: **Comparison of Reproducibility, Accuracy, Sensitivity, and Specificity of miRNA**
- 652 **Quantification Platforms.** *Cell Rep* 2019, **29**(12):4212-4222 e4215.
- 653 28. Baraniskin A, Kuhnenn J, Schlegel U, Chan A, Deckert M, Gold R, Maghnouj A, Zollner H,
- 654 Reinacher-Schick A, Schmiegel W *et al*: **Identification of microRNAs in the cerebrospinal**
- 655 **fluid as marker for primary diffuse large B-cell lymphoma of the central nervous system.**
- 656 *Blood* 2011, **117**(11):3140-3146.
- 657 29. Baraniskin A, Kuhnenn J, Schlegel U, Maghnouj A, Zollner H, Schmiegel W, Hahn S, Schroers
- 658 R: **Identification of microRNAs in the cerebrospinal fluid as biomarker for the diagnosis of**
- 659 **glioma.** *Neuro Oncol* 2012, **14**(1):29-33.
- 660 30. Baraniskin A, Zaslavska E, Nopel-Dunnebacke S, Ahle G, Seidel S, Schlegel U, Schmiegel W,
- 661 Hahn S, Schroers R: **Circulating U2 small nuclear RNA fragments as a novel diagnostic**
- 662 **biomarker for primary central nervous system lymphoma.** *Neuro Oncol* 2016, **18**(3):361-
- 663 367.
- 664 31. Baraniskin A, Chomiak M, Ahle G, Gress T, Buchholz M, Turewicz M, Eisenacher M, Margold
- 665 M, Schlegel U, Schmiegel W *et al*: **MicroRNA-30c as a novel diagnostic biomarker for**
- 666 **primary and secondary B-cell lymphoma of the CNS.** *J Neurooncol* 2018, **137**(3):463-468.
- 667 32. Fischer L, Hummel M, Korfel A, Lenze D, Joehrens K, Thiel E: **Differential micro-RNA**
- 668 **expression in primary CNS and nodal diffuse large B-cell lymphomas.** *Neuro Oncol* 2011,
- 669 **13**(10):1090-1098.
- 670 33. Yang K, Wang S, Cheng Y, Tian Y, Hou J: **Role of miRNA-21 in the diagnosis and prediction of**
- 671 **treatment efficacy of primary central nervous system lymphoma.** *Oncol Lett* 2019,
- 672 **17**(3):3475-3481.
- 673 34. Mao X, Sun Y, Tang J: **Serum miR-21 is a diagnostic and prognostic marker of primary**
- 674 **central nervous system lymphoma.** *Neurol Sci* 2014, **35**(2):233-238.
- 675 35. Roth P, Keller A, Hoheisel JD, Codo P, Bauer AS, Backes C, Leidinger P, Meese E, Thiel E, Korfel
- 676 A *et al*: **Differentially regulated miRNAs as prognostic biomarkers in the blood of primary**
- 677 **CNS lymphoma patients.** *Eur J Cancer* 2015, **51**(3):382-390.

- 678 36. Takashima Y, Kawaguchi A, Iwadata Y, Hondoh H, Fukai J, Kajiwarra K, Hayano A, Yamanaka R:
679 **MicroRNA signature constituted of miR-30d, miR-93, and miR-181b is a promising**
680 **prognostic marker in primary central nervous system lymphoma.** *PLoS One* 2019,
681 **14(1):e0210400.**
- 682 37. Robertus JL, Harms G, Blokzijl T, Booman M, de Jong D, van Imhoff G, Rosati S, Schuurin E,
683 Kluin P, van den Berg A: **Specific expression of miR-17-5p and miR-127 in testicular and**
684 **central nervous system diffuse large B-cell lymphoma.** *Mod Pathol* 2009, **22(4):547-555.**
- 685 38. Drusco A, Bottoni A, Lagana A, Acunzo M, Fassan M, Cascione L, Antenucci A, Kumchala P,
686 Vicentini C, Gardiman MP *et al*: **A differentially expressed set of microRNAs in cerebro-**
687 **spinal fluid (CSF) can diagnose CNS malignancies.** *Oncotarget* 2015, **6(25):20829-20839.**
- 688 39. Zheng J, Xu J, Ma S, Sun X, Geng M, Wang L: **Clinicopathological study of gene**
689 **rearrangement and microRNA expression of primary central nervous system diffuse large**
690 **B-cell lymphomas.** *Int J Clin Exp Pathol* 2013, **6(10):2048-2055.**
- 691 40. Bodor C, Alpar D, Marosvari D, Galik B, Rajnai H, Batai B, Nagy A, Kajtar B, Burjan A, Deak B *et*
692 *al*: **Molecular Subtypes and Genomic Profile of Primary Central Nervous System**
693 **Lymphoma.** *J Neuropathol Exp Neurol* 2020, **79(2):176-183.**
- 694 41. Waggott DM: **NanoStringNorm: Normalize NanoString miRNA and mRNA Data.**
695 <https://CRAN.R-project.org/package=NanoStringNorm>, Accessed 7 August 2020.
- 696 42. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential**
697 **expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015,
698 **43(7):e47.**
- 699 43. Law CW, Chen Y, Shi W, Smyth GK: **voom: Precision weights unlock linear model analysis**
700 **tools for RNA-seq read counts.** *Genome Biol* 2014, **15(2):R29.**
- 701 44. Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK: **Robust Hyperparameter Estimation**
702 **Protects against Hypervariable Genes and Improves Power to Detect Differential**
703 **Expression.** *Ann Appl Stat* 2016, **10(2):946-963.**
- 704 45. Kaiser S, Santamaria R, Khamiakova T, Sill M, Theron R, Quintales L, Leisch F, De Troyer E:
705 **biclust: BiCluster Algorithms.** <https://CRAN.R-project.org/package=biclust>, Accessed 7
706 **August 2020.**
- 707 46. Fan Y, Xia J: **miRNet-Functional Analysis and Visual Exploration of miRNA-Target**
708 **Interactions in a Network Context.** *Methods Mol Biol* 2018, **1819:215-233.**
- 709 47. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A,
710 Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: a knowledge-based**
711 **approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005,
712 **102(43):15545-15550.**
- 713 48. Therneau TM, Lumley T, Atkinson E, Crowson C: **survival: Survival Analysis.** [https://CRAN.R-](https://CRAN.R-project.org/package=survival)
714 [project.org/package=survival](https://CRAN.R-project.org/package=survival), Accessed 7 August 2020.
- 715 49. Kassambara A, Kosinski M, Biecek P, Scheipl F: **survminer: Drawing Survival Curves using**
716 **'ggplot2'.** <https://CRAN.R-project.org/package=survminer>, Accessed 7 August 2020.
- 717 50. Reinsborough CW, Ipas H, Abell NS, Nottingham RM, Yao J, Devanathan SK, Shelton SB,
718 Lambowitz AM, Xhemalce B: **BCDIN3D regulates tRNAHis 3' fragment processing.** *PLoS*
719 *Genet* 2019, **15(7):e1008273.**
- 720 51. Dedhia P, Tarale S, Dhongde G, Khadapkar R, Das B: **Evaluation of DNA extraction methods**
721 **and real time PCR optimization on formalin-fixed paraffin-embedded tissues.** *Asian Pac J*
722 *Cancer Prev* 2007, **8(1):55-59.**
- 723 52. Ribeiro-Silva A, Zhang H, Jeffrey SS: **RNA extraction from ten year old formalin-fixed**
724 **paraffin-embedded breast cancer samples: a comparison of column purification and**
725 **magnetic bead-based technologies.** *BMC Mol Biol* 2007, **8:118.**
- 726 53. Klopffleisch R, Weiss AT, Gruber AD: **Excavation of a buried treasure--DNA, mRNA, miRNA**
727 **and protein analysis in formalin fixed, paraffin embedded tissues.** *Histol Histopathol* 2011,
728 **26(6):797-810.**

- 729 54. Masuda N, Ohnishi T, Kawamoto S, Monden M, Okubo K: **Analysis of chemical modification**
730 **of RNA from formalin-fixed samples and optimization of molecular biology applications for**
731 **such samples.** *Nucleic Acids Res* 1999, **27**(22):4436-4443.
- 732 55. Kokkat TJ, Patel MS, McGarvey D, LiVolsi VA, Baloch ZW: **Archived formalin-fixed paraffin-**
733 **embedded (FFPE) blocks: A valuable underexploited resource for extraction of DNA, RNA,**
734 **and protein.** *Biopreserv Biobank* 2013, **11**(2):101-106.
- 735 56. Giusti L, Angeloni C, Lucacchini A: **Update on proteomic studies of formalin-fixed paraffin-**
736 **embedded tissues.** *Expert Rev Proteomics* 2019, **16**(6):513-520.
- 737 57. Zheng B, Xi Z, Liu R, Yin W, Sui Z, Ren B, Miller H, Gong Q, Liu C: **The Function of MicroRNAs**
738 **in B-Cell Development, Lymphoma, and Their Potential in Clinical Practice.** *Front Immunol*
739 2018, **9**:936.
- 740 58. Esquela-Kerscher A, Slack FJ: **Oncomirs - microRNAs with a role in cancer.** *Nat Rev Cancer*
741 2006, **6**(4):259-269.
- 742 59. Rupaimoole R, Slack FJ: **MicroRNA therapeutics: towards a new era for the management of**
743 **cancer and other diseases.** *Nat Rev Drug Discov* 2017, **16**(3):203-222.
- 744 60. Han W, Cui H, Liang J, Su X: **Role of MicroRNA-30c in cancer progression.** *J Cancer* 2020,
745 **11**(9):2593-2601.
- 746 61. Selcuklu SD, Donoghue MT, Spillane C: **miR-21 as a key regulator of oncogenic processes.**
747 *Biochem Soc Trans* 2009, **37**(Pt 4):918-925.
- 748 62. Liu K, Du J, Ruan L: **MicroRNA-21 regulates the viability and apoptosis of diffuse large B-cell**
749 **lymphoma cells by upregulating B cell lymphoma-2.** *Exp Ther Med* 2017, **14**(5):4489-4496.
- 750 63. Tamaddon G, Geramizadeh B, Karimi MH, Mowla SJ, Abroun S: **miR-4284 and miR-4484 as**
751 **Putative Biomarkers for Diffuse Large B-Cell Lymphoma.** *Iran J Med Sci* 2016, **41**(4):334-339.
- 752 64. Go H, Jang JY, Kim PJ, Kim YG, Nam SJ, Paik JH, Kim TM, Heo DS, Kim CW, Jeon YK: **MicroRNA-**
753 **21 plays an oncogenic role by targeting FOXO1 and activating the PI3K/AKT pathway in**
754 **diffuse large B-cell lymphoma.** *Oncotarget* 2015, **6**(17):15035-15049.
- 755 65. Lawrie CH, Soneji S, Marafioti T, Cooper CD, Palazzo S, Paterson JC, Cattani H, Enver T, Mager
756 R, Boulwood J *et al*: **MicroRNA expression distinguishes between germinal center B cell-**
757 **like and activated B cell-like subtypes of diffuse large B cell lymphoma.** *Int J Cancer* 2007,
758 **121**(5):1156-1161.
- 759 66. Song J, Shao Q, Li C, Liu H, Li J, Wang Y, Song W, Li L, Wang G, Shao Z *et al*: **Effects of**
760 **microRNA-21 on apoptosis by regulating the expression of PTEN in diffuse large B-cell**
761 **lymphoma.** *Medicine (Baltimore)* 2017, **96**(39):e7952.
- 762 67. Li J, Fu R, Yang L, Tu W: **miR-21 expression predicts prognosis in diffuse large B-cell**
763 **lymphoma.** *Int J Clin Exp Pathol* 2015, **8**(11):15019-15024.
- 764 68. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-
765 Cardo C, Lowe SW, Hannon GJ *et al*: **A microRNA polycistron as a potential human**
766 **oncogene.** *Nature* 2005, **435**(7043):828-833.
- 767 69. Tagawa H, Seto M: **A microRNA cluster as a target of genomic amplification in malignant**
768 **lymphoma.** *Leukemia* 2005, **19**(11):2013-2016.
- 769 70. Psathas JN, Doonan PJ, Raman P, Freedman BD, Minn AJ, Thomas-Tikhonenko A: **The Myc-**
770 **miR-17-92 axis amplifies B-cell receptor signaling via inhibition of ITIM proteins: a novel**
771 **lymphomagenic feed-forward loop.** *Blood* 2013, **122**(26):4220-4229.
- 772 71. Lawrie CH, Chi J, Taylor S, Tramonti D, Ballabio E, Palazzo S, Saunders NJ, Pezzella F,
773 Boulwood J, Wainscoat JS *et al*: **Expression of microRNAs in diffuse large B cell lymphoma**
774 **is associated with immunophenotype, survival and transformation from follicular**
775 **lymphoma.** *J Cell Mol Med* 2009, **13**(7):1248-1260.
- 776 72. Alencar AJ, Malumbres R, Kozloski GA, Advani R, Talreja N, Chinichian S, Briones J, Natkunam
777 Y, Sehn LH, Gascoyne RD *et al*: **MicroRNAs are independent predictors of outcome in diffuse**
778 **large B-cell lymphoma patients treated with R-CHOP.** *Clin Cancer Res* 2011, **17**(12):4125-
779 4135.

780 73. Lim EL, Trinh DL, Scott DW, Chu A, Krzywinski M, Zhao Y, Robertson AG, Mungall AJ, Schein J,
781 Boyle M *et al*: **Comprehensive miRNA sequence analysis reveals survival differences in**
782 **diffuse large B-cell lymphoma patients.** *Genome Biol* 2015, **16**:18.

783 74. Zhong H, Xu L, Zhong JH, Xiao F, Liu Q, Huang HH, Chen FY: **Clinical and prognostic**
784 **significance of miR-155 and miR-146a expression levels in formalin-fixed/paraffin-**
785 **embedded tissue of patients with diffuse large B-cell lymphoma.** *Exp Ther Med* 2012,
786 **3**(5):763-770.

787 75. Eis PS, Tam W, Sun L, Chadburn A, Li Z, Gomez MF, Lund E, Dahlberg JE: **Accumulation of**
788 **miR-155 and BIC RNA in human B cell lymphomas.** *Proc Natl Acad Sci U S A* 2005,
789 **102**(10):3627-3632.

790 76. Huang X, Shen Y, Liu M, Bi C, Jiang C, Iqbal J, McKeithan TW, Chan WC, Ding SJ, Fu K:
791 **Quantitative proteomics reveals that miR-155 regulates the PI3K-AKT pathway in diffuse**
792 **large B-cell lymphoma.** *Am J Pathol* 2012, **181**(1):26-33.

793 77. Wang X, Wang K, Han L, Zhang A, Shi Z, Zhang K, Zhang H, Yang S, Pu P, Shen C *et al*: **PRDM1**
794 **is directly targeted by miR-30a-5p and modulates the Wnt/beta-catenin pathway in a**
795 **Dkk1-dependent manner during glioma growth.** *Cancer Lett* 2013, **331**(2):211-219.

796 78. Pasqualucci L, Dalla-Favera R: **SnapShot: diffuse large B cell lymphoma.** *Cancer Cell* 2014,
797 **25**(1):132-132 e131.

798 79. Sehn LH, Gascoyne RD: **Diffuse large B-cell lymphoma: optimizing outcome in the context of**
799 **clinical and biologic heterogeneity.** *Blood* 2015, **125**(1):22-32.

800 80. Vermaat JS, Pals ST, Younes A, Dreyling M, Federico M, Aurer I, Radford J, Kersten MJ, Eha
801 Lymphoma Group aSWGotEHA: **Precision medicine in diffuse large B-cell lymphoma: hitting**
802 **the target.** *Haematologica* 2015, **100**(8):989-993.

803 81. Takashima Y, Sasaki Y, Hayano A, Homma J, Fukai J, Iwadate Y, Kajiwara K, Ishizawa S,
804 Hondoh H, Tokino T *et al*: **Target amplicon exome-sequencing identifies promising diagnosis**
805 **and prognostic markers involved in RTK-RAS and PI3K-AKT signaling as central**
806 **oncopathways in primary central nervous system lymphoma.** *Oncotarget* 2018,
807 **9**(44):27471-27486.

808 82. Mendez JS, Grommes C: **Treatment of Primary Central Nervous System Lymphoma: From**
809 **Chemotherapy to Small Molecules.** *Am Soc Clin Oncol Educ Book* 2018, **38**:604-615.

810 83. Madden E, Logue SE, Healy SJ, Manie S, Samali A: **The role of the unfolded protein response**
811 **in cancer progression: From oncogenesis to chemoresistance.** *Biol Cell* 2019, **111**(1):1-17.

812 84. Ferreri AJM: **Secondary CNS lymphoma: the poisoned needle in the haystack.** *Ann Oncol*
813 2017, **28**(10):2335-2337.

814 85. Grommes C, Pastore A, Palaskas N, Tang SS, Campos C, Schartz D, Codega P, Nichol D, Clark
815 O, Hsieh WY *et al*: **Ibrutinib Unmasks Critical Role of Bruton Tyrosine Kinase in Primary CNS**
816 **Lymphoma.** *Cancer Discov* 2017, **7**(9):1018-1029.

817 86. Vallee A, Lecarpentier Y, Vallee JN: **Hypothesis of Opposite Interplay Between the Canonical**
818 **WNT/beta-catenin Pathway and PPAR Gamma in Primary Central Nervous System**
819 **Lymphomas.** *Curr Issues Mol Biol* 2019, **31**:1-20.

820 87. Jung YS, Park JI: **Wnt signaling in cancer: therapeutic targeting of Wnt signaling beyond**
821 **beta-catenin and the destruction complex.** *Exp Mol Med* 2020, **52**(2):183-191.

Figure captions

Figure 1. Sample characteristics and general expression patterns. **a)** Total number of primary and secondary central nervous system lymphoma samples categorized by molecular subtypes. **b)** Median normalized expression of miRNAs in PCNSL (x-axis) and SCNSL (y-axis) samples. Both axes are $\log_{10}$ based, and the hexagon color scale shows the number of miRNAs falling into a particular median expression range. As can be seen on the plot (bright yellow hexagon in the bottom left corner), a large number of miRNAs had a 0 or near 0 expression in both sample types. The Spearman correlation of PCNSL and SCNSL values is 0.89.

Figure 2. Differential expression analysis between sample groups. **a)** Volcano-plot of differential expression results comparing secondary and primary samples, the GC and ABC or UC and ABC sample groups. The x-axis shows the $\log_2$ fold change of a specific miRNA, while the y-axis shows the $-\log_{10}$ transformed FDR corrected p-value. **b)** Volcano-plot of differential expression results comparing mutated and non-mutated samples for a specific gene, where the color of the dots correlates with the number of mutated samples. As in panel **a)** x-axis shows the $\log_2$ fold change of a specific miRNA, while the y-axis shows the $-\log_{10}$ transformed FDR corrected p-value.

Figure 3. 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.

Figure 4. 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.

Figure 5. 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 removed the largest miRNA cluster (Mcluster2) from the visualization as it contained mainly miRNAs with zero or very low expression.

Figure 6. 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. The x axis shows the $-\log_{10}$ transformed, FDR corrected Fisher-test p-values, the y axis lists the deregulated MSigDB cancer hallmarks, and the dot size is proportional to the Fisher-test odds value. The dot colors correspond to the three different categories investigated (*PRDM1* mutated vs wild-type, SCNSL vs PCNSL and the unknown small cluster). The light red background of the “Unknown small cluster” facet indicates that pathways in this analysis are upregulated, while pathways in the other two comparisons are downregulated.

870 **Table 1.**

<i>Condition</i>	<i>microRNA</i>	<i>log2 fc</i>	<i>FDR</i>	<i>References</i>
SCNSL vs. PCNSL	hsa-miR-873-3p	-7,61	0,0342768	-
	hsa-miR-6721-5p	-4,74	0,0466017	-
	hsa-miR-509-5p	-2,17	0,0379788	-
	hsa-miR-454-3p	2,75	0,0342768	-
	hsa-miR-532-3p	2,92	0,0401021	-
	hsa-miR-301a-3p	3,13	0,0342768	-
	hsa-miR-320e	3,51	0,0366852	-
	hsa-miR-374b-5p	3,53	0,0342768	-
	hsa-miR-421	3,54	0,0342768	-
	hsa-miR-30c-5p	3,55	0,0342768	[31, 32]
	hsa-miR-361-5p	3,68	0,0342768	-
	hsa-miR-133a-3p	3,75	0,0342768	-
	hsa-miR-551b-3p	3,75	0,0342768	-
	hsa-miR-630	4,21	0,0342768	-
	hsa-miR-1290	4,68	0,0342768	-
	hsa-miR-574-5p	4,69	0,0342768	-
	hsa-miR-107	4,71	0,0469009	-
	hsa-miR-362-5p	4,87	0,0036679	-
	hsa-miR-1246	4,92	0,0499940	-
	hsa-miR-1915-3p	5,05	0,0342768	-
	hsa-miR-148b-3p	5,21	0,0364880	-
	hsa-miR-1285-5p	5,28	0,0036679	-
	hsa-miR-21-5p	5,33	0,0036679	[16, 28, 29, 33, 34, 62-66]
	hsa-miR-7-5p	5,97	0,0342768	-
	hsa-miR-423-5p	6,75	0,0012532	-
	hsa-miR-32-5p	6,75	0,0117086	-
	hsa-let-7e-5p	7,27	0,0014054	-
	hsa-miR-19a-3p	7,33	0,0036679	[68-71]
	hsa-miR-106b-5p	7,84	0,0006885	[68-71]
	hsa-miR-18a-5p	8,59	0,0000234	[68-72]
	hsa-miR-4516	8,74	0,0108205	-
GC vs. ABC	hsa-miR-222-3p	-3,82	0,0280039	[18, 71, 73]
	hsa-miR-522-3p	-2,34	0,0455441	-
UC vs. ABC	hsa-miR-155-5p	-2,26	0,0280039	[16, 19, 64, 65, 71, 73-76]
	hsa-miR-92a-3p	4,70	0,0005669	-
	hsa-miR-522-3p	-3,06	0,0097304	-
	hsa-miR-455-5p	-2,09	0,0097304	-
	hsa-miR-454-3p	-2,12	0,0114100	-
PRDMI mut vs. nm	hsa-miR-551a	-1,27	0,0427852	-
	hsa-miR-107	1,42	0,0427852	-
	hsa-miR-195-5p	1,53	0,0427852	-
	hsa-miR-526b-5p	1,69	0,0427852	-
	hsa-miR-376a-3p	1,87	0,0427852	-
	hsa-miR-130a-3p	2,34	0,0428542	-
	hsa-miR-548y	2,89	0,0068707	-
	hsa-miR-30a-5p	3,61	0,0427852	[77]
	hsa-miR-411-5p	-2,50	0,0042775	-
	hsa-miR-379-5p	-3,55	0,0186781	-
MYC mut vs. nm				
CARD11 mut vs. nm				

871

872 **Table captions**

873 **Table 1.** Differentially expressed miRNAs between primary and secondary central nervous
874 system lymphomas, molecular subtypes and in association to mutation status. The columns
875 show the tested condition, the log₂ fold-change, the FDR and the literature references if
876 known. Abbreviations: ABC, activated B-cell type; GC, germinal center B-cell type; mut,
877 mutated; nm, non-mutated; PCNSL, primary central nervous system lymphoma; SCNSL,
878 secondary central nervous system lymphoma; UC, unclassified; vs., versus.

879

sample_53
sample_36
sample_35
sample_34
sample_34
sample_51
sample_68
sample_102
sample_21
sample_89
sample_85
sample_6
sample_76
sample_67
sample_69
sample_20
sample_11
sample_61
sample_14
sample_25
sample_103
sample_12
sample_38
sample_99
sample_64
sample_73
sample_56
sample_70
sample_16
sample_29
sample_2
sample_2
sample_60
sample_66
sample_71
sample_9
sample_12
sample_27
sample_65
sample_58
sample_87
sample_46
sample_46
sample_84
sample_91
sample_49
sample_43
sample_43
sample_40
sample_42
sample_38
sample_78
sample_104
sample_45
sample_11
sample_28
sample_16
sample_47
sample_30
sample_41
sample_32
sample_36
sample_29
sample_74
sample_33
sample_57
sample_32
sample_39
sample_96

SCluster3

SCluster1

SCluster2

SCluster4

MCluster4

MCluster1

MCluster3

Voom-score