Molecular and clinical diversity in primary central nervous system lymphoma

To cite this version:

HAL Id: hal-03870175
https://u-picardie.hal.science/hal-03870175
Submitted on 20 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Molecular and clinical diversity in primary central nervous system lymphoma


1Institut du Cerveau-Paris Brain Institute-ICM, Inserm, Sorbonne Université, CNRS, Paris, France; 2Department of Hematology and Medical Oncology, University Medical Center Göttingen, Göttingen; 3Department of Hematology, Oncology and Cancer Immunology, Campus Benjamin Franklin, Charité—Universitätsmedizin Berlin, Berlin; 4Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany; 5Department of Neuropathology, Groupe Hospitalier Pitié Salpêtrière, APHP, Paris; 6Department of Pathology, CHU de Bordeaux, Hôpital Pellegrin, Bordeaux; 7Department of Neurosurgery, Bordeaux University Hospital Center, Pellegrin Hospital, Bordeaux; 8EA 7435—IMOTION, University of Bordeaux, Bordeaux; 9Department of Pathology, PBH, CHU Angers, Angers; 10CRCINA, Université de Nantes—université d’Angers, Angers; 11Department of Hematology, CHU Angers, Angers; 12Department of Neurology, Hôpitaux Civils de Colmar, Colmar; 13Department of Neuropathology, Hôpitaux Civils de Colmar, Strasbourg; 14Department of Pathology, CHU de Toulouse, IUC-Oncopele, Toulouse; 15INSERM U1037, Cancer Research Center of Toulouse (CRCIT), Toulouse; 16Université Toulouse III Paul Sabatier, Toulouse; 17Department of Hematology, IUC Toulouse Oncopole, Toulouse; 18Neuroepidemiology Department, University Hospital Timone, Aix Marseille University, Marseille; 19Department of Neuro-oncology, CHU Timone, APHM, Marseille; 20Institut de Physiopathologie CNRS, INSERM, Aix-Marseille University, Marseille; 21Department of Biopathology, CHRU Nancy, CHRU/ICL, Bâtiment B86, Vandoeuvre-lès-Nancy; 22Department of Legal Medicine, CHRU Nancy, Vandoeuvre-lès-Nancy; 23INSERM U12556, University of Lorraine, Vandoeuvre-lès-Nancy; 24Centre de Recherche des Cordeliers, Université de Strasbourg; 25Department of Neurology, CHRU Nancy, Nancy; 26Department of Neuro-oncology, CHRU-Nancy, Université de Lorraine, Nancy; 27Department of Hematology, CHU Amiens-Picardie, Amiens; 28Department of Anatomopathology, Lariboisière Hospital, Assistance Publique—Hôpitaux de Paris, University of Paris, Paris; 29Pathology Department, Bicêtre University Hospital, Public Hospital Network of Paris, Le Kremlin Bicêtre; 30Department of Neurology, Université de Paris, AP-HP, Hôpital Saint Louis, Paris; 31Department of Hematology, Institut Bergonié Hospital, Bordeaux; 32Department of Neuroradiology, Sorbonne Université, Assistance Publique—Hôpitaux de Paris, Groupe Hospitalier Pitié-Salpêtrière-Charles Foix, Paris; 33Department of Neurology-2, Sorbonne Université, Assistance Publique—Hôpitaux de Paris, Groupe Hospitalier Pitié-Salpêtrière-Charles Foix, Paris; 34Department of Neurology, Perpignan Hospital, Perpignan; 35Department of Neurosurgery, Sorbonne Université, Assistance Publique—Hôpitaux de Paris, Groupe Hospitalier Pitié-Salpêtrière-Charles Foix, Paris; 36Department of Pathology; 37Hematology, APHP, Hôpital Pitié-Salpêtrière and Sorbonne University, Paris, France; 38Broad Institute of MIT and Harvard, Cambridge, USA; 39INSERM UMR_5138—Centre de Recherche des Cordeliers—Université Pierre et Marie Curie et Université Paris Descartes, Paris; 40Hematology Unit, Institut Curie, Saint-Cloud, France

Background: Primary central nervous system lymphoma (PCNSL) is a rare and distinct entity within diffuse large B-cell lymphoma presenting with variable response rates probably to underlying molecular heterogeneity. Patients and methods: To identify and characterize PCNSL heterogeneity and facilitate clinical translation, we carried out a comprehensive multi-omic analysis [whole-exome sequencing, RNA sequencing (RNA-seq), methylation sequencing, and clinical features] in a discovery cohort of 147 fresh-frozen (FF) immunocompetent PCNSLs and a validation cohort of formalin-embedded (FFPE) 93 PCNSLs with RNA-seq and clinico-radiological data. Results: Consensus clustering of multi-omic data uncovered concordant classification of four robust, non-overlapping, prognostically significant clusters (CS). The CS1 and CS2 groups presented an immune-cold hypermethylated profile but a distinct clinical behavior. The ‘immune-hot’ CS4 group, enriched with mutations increasing the Janus kinase (JAK)−signal transducer and activator of transcription (STAT) and nuclear factor-kB activity, had the most favorable clinical outcome, while the heterogeneous-immune CS3 group had the worse prognosis probably due to its association with meningeval infiltration and enriched HIST1H1E mutations. CS1 was characterized by high Polycomb repressive complex 2 activity and CDKN2A/B loss leading to higher proliferation activity. Integrated analysis on proposed targets suggests potential use of immune checkpoint inhibitors/JAK1 inhibitors for CS4, cyclin D-Cdk4,6 plus phosphoinositide 3-kinase (PI3K) inhibitors for CS1, lenalidomide/demethylating drugs for CS2, and enhancer of
INTRODUCTION
Primary central nervous system lymphoma (PCNSL) is a rare subtype of extranodal non-Hodgkin’s lymphoma within diffuse large B-cell lymphoma (DLBCL), but has a less favorable prognosis than its systemic counterpart and has been proved to be molecularly a different biological entity.1-6 The standard treatment relies on high-dose methotrexate (HD-MTX) regimen with or without consolidation and is associated with treatment resistance or relapses in up to 60% of the patients.7,8

Biologically, initial studies have found PCNSL to be at late B-cell germinal center (GC) exit stages and to have constitutive nuclear factor-kB (NF-κB) activity driven by mutations in genes of the B-cell receptor (BCR) pathway, of the toll-like receptor (TLR) pathway (MYD88), and CARD11.6,9-17 Recently, DLBCL has been divided into molecular subtypes. The integration of genome-wide data from multi-omic data revealed four molecular patterns in PCNSL and subtype-based targeted interventions.

Patients
A total of 147 biopsies from treatment-naive FF (discovery cohort) and 93 FFPE (validation cohort) tumor samples from immunocompetent EBV- PCNSL were recollected from different hospitals across France (see Supplementary Table S1, available at https://doi.org/10.1016/j.annonc.2022.11.002). None of the patients received corticosteroids for 1 week before the brain biopsy. All patients had a complete systemic evaluation to rule out secondary central nervous system DLBCL. Diagnoses and section selection for sequencing were established at the reference institution by specialized (neuro)pathologists. The tumor cell content for each sample was estimated to be at least 60% based on histomorphological evaluation. FFPE tissue sections available for all data (FF and FFPE cohorts) were used for immunophenotypic characterization with antibodies directed against CD20, CD10, BCL6, CD3, Ki67, and MUM1/IRF4. EBV-encoded small RNAs in situ hybridization was carried out as previously described.31 We obtained appropriate consent from relevant institutional review boards, which coordinated the consent process at each tissue-source site; written informed consent was obtained from all participants. The Pitié Salpêtrière Hospital ethics committee approved the study (Ile-de-France VI, No. DC-2009-957) and CNIL (DR-2013-279). All patients received HD-MTX regimens according to French National ‘Lymphome oculo-cérébral, LOC’ PCNSL network.7 Moreover, 19/134 (14.2%; FF cohort) and 24/93 (25.8%; FFPE cohort) received intensive chemotherapy with autologous stem cell rescue (IC-ASCR). Magnetic resonance images (MRI) of PCNSL at the time of first diagnosis and recurrence in sufficient quality were available for 90 patients within the validation cohort. We carried out, on the FF cohort, exome sequencing (n = 115), RNA sequencing (RNA-seq, n = 123), and DNA methylation profiling (n = 64). In the FFPE cohort, only RNA-seq was carried out (n = 93). Study design and sample size for each sequencing approach are displayed in Figure 1A.

Multi-omic data integration for PCNSL molecular subtyping
The multi-omic cohort comprised 85 FF samples from which we obtained six data types including mutations, CNA alterations, fusions, gene expression, immune cell proportions, TCR/BCR clonotypes, and fusion transcripts (Figure 1A). The most appropriate cluster number was obtained by clustering prediction index (CPI) and Gap-statistics analyses. High robust clustering using these data was obtained by consensus clustering resulting from 10 different multi-omics clustering algorithms (iClusterBayes, moCluster, CIMLR, IntNMF, ConsensusClustering, COCA, NEMO, PINSPlus, SNF, and LRA).32-42 Clustering performance was evaluated for each cluster number (n = 2-8) using the weighted silhouette width.33 Genomic data were deposited at the European
Multi-omic cohort was carried out using tumor tissue and matched peripheral blood samples for 98 PCNSL patients, while for 17 patients it was tumor-only. For the methyl-seq, we also included four control peripheral blood samples. Different types of information were available for each omic data, as seen inside each bar plot. The multi-omic cohort consisted of 85 samples that had WES, RNA-seq, and clinic data within the discovery fresh-frozen cohort (red dashed box). (B) The number of affected samples within the top driver genes (identified by dNdScv algorithm) in the cohort of 115 PCNSL patients is shown. Bar plots are filled according to mutation type (missense, nonsense, splice site, frameshift, multiht, or other) or CNA events (gain, deletion, or deep deletion). The frequency of affected samples within the cohort is annotated at the top of each bar plot for each driver gene. (C) The boxplot comparison of the proportion of mutations attributed to c-AID or SBS9 mutational processes when occurring at clonal times versus subclonal times is shown. P values were calculated by Wilcoxon matched-pairs signed rank test for paired data (per sample). (D) The GISTIC2.0-defined recurrent copy number focal deletions (blue, left) and gains (red, right) as mirror plots in DLBCL (n = 296 from Chapuy et al., 2018) and PCNSL (n = 108 from this study) are shown. Chromosome position is on the y-axis, and significance is on the x-axis. CNAs are labeled with their associated cytoband/arm followed in brackets by the frequency of the alteration (DLBCL / PCNSL). (E) The GISTIC2.0 results of significantly recurring (events with a q-value of < 0.1) focal amplifications (red) and deletions (blue) are shown, where the chromosomes are plotted in the x-axis and the GISTIC-scores (G-score) are plotted in the y-axis. Genes affected for each focal event are annotated followed by the percentage of altered samples (n = 108). Genes are underlined if they are driver genes or in bold if they are transcriptionally affected by the focal event (related to Supplementary Figure S8 and Table S4, available at https://doi.org/10.1016/j.annonc.2022.11.002).

c-AID, canonical activation-induced cytidine deaminase; CNA, copy number alteration; DLBCL, diffuse large B-cell lymphoma; EBV, Epstein–Barr virus-negative; FC, fold change; KPS, Karnofsky performance status; methyl-seq, methylation sequencing; MRI, magnetic resonance imaging; OS, overall survival; PCNSL, primary central nervous system lymphoma; PFS, progression-free survival; RNA-seq, RNA sequencing; SNP, single-nucleotide polymorphism; TME, tumor microenvironment; WES, whole-exome sequencing.

Figure 1. Study design and PCNSL mutational landscape. (A) The study design including the workflow and the data composition for each cohort is demonstrated. WES was carried out using tumor tissue and matched peripheral blood samples for 98 PCNSL patients, while for 17 patients it was tumor-only. For the methyl-seq, we also included four control peripheral blood samples. Different types of information were available for each omic data, as seen inside each bar plot. The multi-omic cohort consisted of 85 samples that had WES, RNA-seq, and clinic data within the discovery fresh-frozen cohort (red dashed box). (B) The number of affected samples within the top driver genes (identified by dNdScv algorithm) in the cohort of 115 PCNSL patients is shown. Bar plots are filled according to mutation type (missense, nonsense, splice site, frameshift, multiht, or other) or CNA events (gain, deletion, or deep deletion). The frequency of affected samples within the cohort is annotated at the top of each bar plot for each driver gene. (C) The boxplot comparison of the proportion of mutations attributed to c-AID or SBS9 mutational processes when occurring at clonal times versus subclonal times is shown. P values were calculated by Wilcoxon matched-pairs signed rank test for paired data (per sample). (D) The GISTIC2.0-defined recurrent copy number focal deletions (blue, left) and gains (red, right) as mirror plots in DLBCL (n = 296 from Chapuy et al., 2018) and PCNSL (n = 108 from this study) are shown. Chromosome position is on the y-axis, and significance is on the x-axis. CNAs are labeled with their associated cytoband/arm followed in brackets by the frequency of the alteration (DLBCL / PCNSL). (E) The GISTIC2.0 results of significantly recurring (events with a q-value of < 0.1) focal amplifications (red) and deletions (blue) are shown, where the chromosomes are plotted in the x-axis and the GISTIC-scores (G-score) are plotted in the y-axis. Genes affected for each focal event are annotated followed by the percentage of altered samples (n = 108). Genes are underlined if they are driver genes or in bold if they are transcriptionally affected by the focal event (related to Supplementary Figure S8 and Table S4, available at https://doi.org/10.1016/j.annonc.2022.11.002).

c-AID, canonical activation-induced cytidine deaminase; CNA, copy number alteration; DLBCL, diffuse large B-cell lymphoma; EBV, Epstein–Barr virus-negative; FC, fold change; KPS, Karnofsky performance status; methyl-seq, methylation sequencing; MRI, magnetic resonance imaging; OS, overall survival; PCNSL, primary central nervous system lymphoma; PFS, progression-free survival; RNA-seq, RNA sequencing; SNP, single-nucleotide polymorphism; TME, tumor microenvironment; WES, whole-exome sequencing.
Genome-phenome Archive (EGAD00001008706; http://www.ebi.ac.uk/ega/), which is hosted by the European Bioinformatics Institute (EBI). The RBrAlymp algorithm is publicly accessible at https://github.com/iS4i4S/PCNSL-RBrAlymp. All other materials are available upon request from the authors.

**Statistical analyses**

All statistical analyses were carried out using the R statistical programming environment (v4.0; The R Project for Statistical Computing, Vienna, Austria). Differences in proportions and binary/categorical variables were calculated from two-sample Z-tests or Fisher’s exact test. The Kruskal–Wallis test was used to test for a difference in distribution between three or more independent groups, and the Mann–Whitney U test was used for differences in distributions between two population groups unless otherwise noted. P values were corrected for multiple comparisons using the Benjamini–Hochberg method when applicable. Overall survival (OS) and progression-free survival (PFS) analysis were assessed using log-rank Kaplan–Meier curves and multivariate Cox proportional hazards regression modeling. See Supplementary Appendix for full details, available at https://doi.org/10.1016/j.annonc.2022.11.002.

**RESULTS**

**PCNSL mutational landscape**

We identified 32,544 mutations in the 115 PCNSL FF samples analyzed (85% tumor-normal pairs, median = 3.23 mutations/Mb; range = 0.02-85.49 mutations/Mb; Supplementary Table S2 and Figure S1, available at https://doi.org/10.1016/j.annonc.2022.11.002). We applied the dNdScv44 algorithm to identify driver mutations finding the hallmark mutations of PCNSL like MYD88 (64%), PIM1 (59%), PRDM1 (57%), GRHPR (50%), HLA-A/B/C (49%, 30%, and 13%), BTG2 (47%), CD79B (43%), CDKN2A (28%), ETV6 (26%), TBL1X1R1 (25%), KLHL14 (25%), CARD11 (22%), and HIST1H1E (18%) which are involved in BCR-TLR-mediated NF-kB signaling, antigen presentation, cell cycle, histone modification, and B-cell differentiation regulation1,12,18,21,45,46 (Figure 1B, Supplementary Table S3 and Figure S2A, available at https://doi.org/10.1016/j.annonc.2022.11.002). Moreover, we detected canonical activation-induced cytidine deaminase (c-AID) off-target mutations and found that they represent 7.9% (95% confidence interval (CI) 6.8% to 8.5%) of single-nucleotide variant mutations and fall within driver genes like PIM1 (47%), CD79B (10%), IRF4 (9%), and HIST1H1E (6%) (Supplementary Table S4 and Figure S2B, available at https://doi.org/10.1016/j.annonc.2022.11.002).

Next, we used the clonal allelic status of each mutation along with the mutational process that probably originated them to gain information about the relative timing of these processes. Interestingly, both c-AID and non-c-AID (COSMIC signature S589) mutations are significantly more active at clonal stages (P = 0.007 and P = 0.018, respectively), hence reflecting the importance of AID activity in early stages of PCNSL tumorigenesis (Figure 1C), as previously described.1,18,21,45,46

Regarding focal CNA, we identified significant recurrent amplifications in 18q21.33 (42%) and 19p13.13 (34%), and deletions in 6p21 (39%), 6q21 (65%), 6q27 (49%), and 9p21.3 (28%) which have a higher frequency than those observed by Chapuy et al. in systemic DLBCL (Figure 1D).1,18 Furthermore, we found recently described amplifications in 1q32.1 (33%, IL10) and 11q23.3 (26%, CD3G), and deletions in 6p25.3 (21%, IRF4), 22q11.22 (29%, GGTLC2), and 14q32.33 (84%) that produce significant expression changes in CD3G [fold change (FC) = 1.25], IRF4 (FC = −1.03), and GGTLC2 [FC = −1.76, false discovery rate (FDR) q-value <0.1], respectively (Figure 1E, Supplementary Table S5 and Figure S3, available at https://doi.org/10.1016/j.annonc.2022.11.002).

**Multi-omic data integration reveals PCNSL molecular subtypes with clinical outcome implications**

To address the PCNSL heterogeneity, we carried out a cluster of clusters analysis using six levels of information including messenger RNA (mRNA) expression, mutations, CNA, fusion transcripts, TCR/BCR clonotypes, and immune cell fractions available in the multi-omic cohort (Figure 2A, Supplementary Figure S4, available at https://doi.org/10.1016/j.annonc.2022.11.002). Using the optimal cluster number resulting from CPI, Gap-statistics, and weighted silhouette width, we identified four PCNSL subtypes (CS1-CS4) that display different clinical outcomes in OS (global log-rank P < 0.001, Figure 2B). Patients in CS4 had the longest OS (median 66.8 months; 95% CI 19.8-67.2 months) and lived significantly longer than those in both clusters CS2 (median 18 months; 95% CI 8.3-53.4 months; P = 0.024) and CS3 (median 13.8 months; 95% CI 6.1-16.7 months; P = 0.003), and slightly longer, but not significantly, than those in CS1 (median 26.2 months; 95% CI 13.3-63.9 months; P = 0.094). Additionally, these observations remained significant after adjusting by age and Karnofsky performance status (KPS) in Cox proportional hazards ratio multivariate models (Figure 2B). Interestingly, CS4 was independently associated with a better response when considering PFS in univariate and multivariate models (Figure 2C). Furthermore, we did not observe significant differences in the median number of predicted immunogenic neoantigens (P = 0.44), the fraction of c-AID mutations (P = 0.25), or the number of patients receiving IC-ASCR (P = 0.32, Supplementary Table S6, available at https://doi.org/10.1016/j.annonc.2022.11.002).

**Transcriptomic data correctly assign multi-omic defined PCNSL subtypes in FF and FFPE samples**

Given the difficulty of acquiring FF tissue and of analyzing and implementing multi-omic data into routine clinical practice, we sought to evaluate the use of only RNA expression, obtained from FFPE or FF tissue, to categorize patients into the four subtypes. To do so, we first identified, for each subtype, the top 100 unique up-regulated...
biomarkers by carrying out differential expression analysis of one group versus the others and then extracting the non-overlapping genes across subtypes (P adjusted < 0.05; Supplementary Table S7, available at https://doi.org/10.1016/j.annonc.2022.11.002); afterward we applied different prediction methods on the mRNA data (Supplementary Figures S5-S7, available at https://doi.org/10.1016/j.annonc.2022.11.002). We obtained a Cohen’s $\kappa$ coefficient of 0.90 ($P < 0.001$, Supplementary Figure S8A, available at https://doi.org/10.1016/j.annonc.2022.11.002) when evaluating the accuracy of correctly assigning patients from the multi-omic cohort. Additionally, when expanding to the FF-RNA complete set ($n = 123$) or when using the FFPE cohort ($n = 93$), we observed the same behaviors regarding clinical outcome (global $P < 0.001$) across molecular subtypes in both univariate and multivariate models using OS and PFS (Figure 2D, Supplementary Figure S8, available at https://doi.org/10.1016/j.annonc.2022.11.002).

Next, we evaluated the contribution of each of the six data types to outcome prediction models by using Harrell’s concordance index (C-index). A C-index of 0.60 (95% CI 0.56-0.65) in FF and 0.71 (95% CI 0.68-0.74) in FFPE was observed using KPS and age, which are the clinical features currently used in the Memorial Sloan Kettering Cancer Center prognostic score for PCNSL. When adding different data types to the multi-omic cohort modeling, we observed higher predictive power using mRNA expression compared to the others (C-index = 0.91 ± 0.02). We further validated these observations in the FFPE cohort obtaining a C-index of 0.83 (95% CI 0.80-0.85) and 0.93 (95% CI 0.91-0.95) when adding the mRNA level or the TME + RNA levels to the model, respectively (Figure 2D). Altogether, these results show that RNA-seq data alone from FFPE or FF tissue can be used to correctly identify PCNSL subgroups.

**Distinct genetic signatures within PCNSL subtypes and DLBCLs**

Next, we aimed to characterize the differences in genetic alterations across groups by carrying out inside-outside cluster Fisher tests (q-value ≤ 0.1). The C54 cluster...
presents 10 enriched events that included mutations in SOCS1, which is a negative regulator of the Janus kinase (JAK)—signal transducer and activator of transcription 3 (STAT3) pathway, MPEGL1, PIM2, and deletion of 17q25.1 involving GRB2 that indirectly regulates the NF-κB pathway.

We observed 43 events within the CS1 cluster including mutations involved in the NF-κB pathway (IRF4 via 6p25.3 deletion), B-cell differentiation (IRF4 via 6p25.3 deletion, TOX, and BCL6), proliferation via interruption of cell cycle arrest (CDKN2A/2B fusions and FOXO1), and B-cell lymphomagenesis (e.g. ET36, OSBPRL1). Patients within the CS3 cluster exhibit 12 events from which HIST1H1E arises as the top enriched, and has been proved to enhance self-renewal properties and disrupt chromatin architecture in B-cell lymphomas.1,2,18,45,49 The CS2 cluster did not present any genomic enriched events with respect to the other subtypes. Furthermore, most of these distinctive events arrived as early events (clonal) in tumorigenesis like IRF4 and BCL6 in CS1 (Figure 3A, Supplementary Table S8, available at https://doi.org/10.1016/j.jannonc.2022.11.002). Of note, most of these mutations were not observed in the clusters previously defined by Chapuy et al. or were more CS-specific (e.g. 9p11.2 del; Figure 3B).1 We also used the LymphGen algorithm to identify the correspondence of our samples to the genetic subtypes described by Wright et al.45

We observed a similar proportion of samples that were classified as either MCD (40%, 34/85) or other (39%, 33/85), while the remaining fell within either A53 (13%, 11/85) or BN2 (8%, 7/85). The MCD was mainly composed of CS1 enriched events with respect to the other subtypes, and CS2 samples (91%, 77/85) were activated B cells as per the cell-of-origin classification.

**B-cell differentiation stages, pathways, and TME distinctions between PCNSL molecular subtypes**

To determine whether or not these mutations could affect oncogenic pathways and B-cell differentiation programs transcriptionally, we recovered and analyzed the expression of different curated gene signatures initially proposed and used by Wright et al. and Kotlov et al.45,50 CS1 was characterized by the up-regulation of PI3K, glycolytic activity, and cell proliferation signatures, which is consistent with the observed mutations in the cell cycle arrest; additionally, it presented hyperactivation of the Polycomb repressive complex 2 (PRC2) which has been proven to inhibit major histocompatibility complex (MHC)-I expression, through histone methylation (Figure 4A, P < 0.05).51 Moreover, p53 activity was enriched in the CS2 subtype.52,63 Interestingly, even though all subtypes presented mutations within the NF-κB pathway, it was transcriptionally active only in clusters CS3 and CS4. Additionally, mitogen-activated protein kinase (MAPK) and JAK—STAT pathways were up-regulated in those clusters, respectively (Figure 4A).

Regarding B-cell differentiation programs, CS1 expresses a mixture of GC cells, which is consistent with both the observed MYC induction activity, 6p25.3-19q13.12 deletions, and BCL6 mutations (Figure 2C). On the other hand, the cluster CS4 presents an enrichment in terminally differentiated plasma cells that goes in line with BCL6 negative regulation, the absence of MYC induction, and BCL6 mutations. The most heterogeneous cluster was CS3, presenting features of both GC and mature B cells. Intriguingly, CS2 did not feature any B-cell stage enrichment but instead a lymphatic endothelial cell (LEC) gene signature (Figure 4A, Supplementary Figure S9, available at https://doi.org/10.1016/j.jannonc.2022.11.002). These different B-cell states, including plasmablasts/plasma cells, have been previously described in DLBCL and PCNSL using bulk/single-cell RNA-seq approaches.54-56 Additionally, Milpied et al. reported the co-existence of different B-cell states within a patient with follicular lymphoma.57 To further corroborate these observations, we analyzed BCR clonotypes and the immunoglobulin (Ig) heavy-chain variable (VH) and constant region expression across clusters. We observed across clusters the presence of BCR clones’ diversity and IgVH4-34 expression, which further supports the presence of self-antigen-dependent chronic active BCR signaling across clusters (Supplementary Figure S10, available at https://doi.org/10.1016/j.jannonc.2022.11.002).45,58

Then, we aimed to describe the TME differences between subtypes by using CIBERSORTx-derived immune deconvolution and Kotlov’s B-cell lymphoma-specific TME gene signatures.60 The CS1 cluster is immunologically ‘neutral’; meanwhile the CS2 cluster, which is immunologically depleted, exhibits expression of vascular endothelial cells, memory resting CD4+ T cells, monocytes, and activation of γ-aminobutyric acid synthesis, which has been recently linked to B cells that inhibit CD8+ T cells’ killer function and promote monocyte differentiation into anti-inflammatory macrophages.59 The CS4 cluster has a hot-inflammatory TME due to the presence of active CD8+ T cells and natural killer cells (high cytolytic activity score), and a high expression of MHC-I/II, and immune checkpoint molecules.50 Conversely, heterogeneity was again observed for the CS3 subtype, with only inactivated macrophages M0 being more significantly enriched (Figure 4A, Supplementary Figure S11, available at https://doi.org/10.1016/j.jannonc.2022.11.002). When analyzing TCR sequences, we found significantly higher T-cell clonotype diversity in CS4 versus others (P < 0.05, Supplementary Figure S12, available at https://doi.org/10.1016/j.jannonc.2022.11.002). Similar results were observed in the FFPE cohort (Supplementary Figure S13, available at https://doi.org/10.1016/j.jannonc.2022.11.002).

Moreover, immune staining revealed higher presence of both CD3+ and CD163+ cells in CS4, in line with the transcriptionally observed hot TME (Figure 4B). To evaluate the LEC signature, we stained PROX1, a homeobox gene master regulator (MR) in LEC differentiation that forms part of our transcriptionally defined LEC gene signature and that was used by Louveau et al. when describing LECs in the central nervous system,61,62 finding a higher number of PROX1+ cells in the CS2 subtype compared to the others. Finally, we
Figure 3. Distinct genetic signatures within PCNSL subtypes and systemic DLBCL. (A) The landmark genetic alterations are shown for each PCNSL subtype (boxed for each cluster) identified by a one-sided Fisher test (event within cluster versus outside cluster) and ranked by significance (FDR corrected q-value ≤ 0.1 selected, red
observed loss of repressive H3 lysine 23 trimethylation (H3K27me3− cells) in the CS3 subtype, a phenomenon that has been previously associated with B-cell lymphomas carrying HIST1H1B-E mutations.49

**Subtype-specific MRs**

To elucidate subtype-specific MRs, we determined the transcription factors (TFs) of each molecular PCNSL subtype using the genes that are differentially expressed within each cluster and a curated collection of TF targets.64 We found increased TF activity of TEAD1 (in CS1 and CS2); PRDM14 (CS1) which has roles as histone methyltransferase and leukemia initiator64; IRF4, SPIB (both targets of lenalidomide), MYC, and PROX1 (CS2); E2F1 and IRF3 (CS3); and STAT3/1 and NFκB1 (CS4) (Figure 5A, Supplementary Tables S9 and S10 and Figure S14, available at https://doi.org/10.1016/j.annonc.2022.11.002).

**Epigenetic attributes across PCNSL subtypes**

Firstly, we evaluated the DNA methylation—transcriptome correlation by considering all PCNSL samples (n = 64; i.e. not grouping by subtypes) and a total of 27 111 genes with available data, and found that only 12.4% were correlated with their promoter methylation (P adjusted < 0.05 and ρ < 0). Interesting genes included TERT, CD79A, and MGMT (ρ = −0.50, −0.59, and −0.62, respectively; P adjusted < 0.001, Supplementary Table S11A, available at https://doi.org/10.1016/j.annonc.2022.11.002), which goes in line with the literature.65 When grouping by molecular subtypes, we found CCL22, which is a chemoattractant for primary activated T lymphocytes,66 less expressed in the CS2 cluster and associated with gene hypermethylation (Supplementary Table S11, available at https://doi.org/10.1016/j.annonc.2022.11.002).

We proceeded to investigate epigenetic differences among subtypes. CS2 displayed higher hypermethylation globally, within promoters, and at chromosome ends (P = 0.006, P < 0.001, and P < 0.001; Figure 5B); however, the CS1 subtype presented higher methylation within CpG islands (P = 0.001, Supplementary Figures S15 and S16, available at https://doi.org/10.1016/j.annonc.2022.11.002). Interestingly, a hypermethylator phenotype has been previously associated with a depleted TME in systemic DLBCL.59

Gene Ontology analyses on differentially methylated promoters revealed B-cell differentiation programs to be hypomethylated in CS1 but hypermethylated in CS2 (Figure 5C, Supplementary Table S12, available at https://doi.org/10.1016/j.annonc.2022.11.002). We observed cell adhesion hypermethylation only in CS2, which was previously reported when comparing PCNSL versus DLBCL.67 Genomic region enrichment analysis on hypermethylated promoters identified strong enrichment of binding sites for the histone/chromatin proteins H3K27me3, and EZH2 in CS1, as expected from the high PR2 observed transcriptionally. On the other hand, the CS2 subtype exhibited enrichment for BCL11A, NF-kB, IRF4, and BCL6 (Figure 5D, Supplementary Table S13, available at https://doi.org/10.1016/j.annonc.2022.11.002). These findings together with the observed increased LEC gene signature and PROX1 immune staining support the idea of a phenotypic shift to more LEC-like in this PCNSL subtype.

**CS3 subtype is associated with meningeal infiltration to cerebrospinal fluid**

Here, we investigated if brain MRI analysis (n = 90, FFPE cohort) could provide more insights about the molecular subtypes. CS4 tumors arose more in the isthmus of the corpus callosum (7/34 versus 0/56, P < 0.001). Conversely, CS2/CS3 were more frequent in the brainstem (3/19 and 4/16 versus 1/55; P = 0.049 and P = 0.008, respectively), when compared to the other clusters. Strikingly, we found no association with tumor size or multiple lesions. However, meningeal infiltration of the cerebrospinal fluid (CSF) was only found within CS3 tumors (6/16 versus 0/74; P < 0.001; Figure 6A, Supplementary Tables S14 and S15, available at https://doi.org/10.1016/j.annonc.2022.11.002).

**Therapeutic associations of the PCNSL molecular subtypes**

We next wondered if the PCNSL molecular subtypes may exhibit greater benefit from specific targeted therapies. Even though the hallmark PCNSL alterations targeting My-T-BCR protein supercomplex, CD79A/B CRC subunits, and the CBM (CARD11-BCL10-MALT1) complex were relatively constant across subgroups, the NF-kB hyperactive group (CS4) presented more GRB2/LYN deletions and absence of PLCG2 mutations, which either represses the BCR complex or affects the CBM complex activation. Furthermore, NF-kB activity could not be explained by self-antigen-dependent chronic active BCR signaling up-regulation since IgVH4-34 expression was similar across groups.45,43,68 These observations suggest that CS4 and CS3 may be more sensitive to Bruton’s tyrosine kinase (BTK) inhibitors. The CS4 cluster also presented high Jak–STAT activity and mutated SOCS1 (JAK1 repressor), making it potentially responsive to JAK1 inhibitors.69,70 Moreover, the absence of PR2 activity and presence of MHC-I, only monoallelic deletions in HLA-A/B2M/CD58 (Supplementary Figure S17, available at https://doi.org/10.1016/j.annonc.2022.11.002), and checkpoint molecules’
expression indicate a potential use of immune checkpoint inhibitors (ICIs) for CS4. On the other hand, EZH2 inhibitors in combination with ICI could potentially increase MHC-I expression and immune detection in CS3. 71

Additionally, we observed a higher frequency of cases with genetic alterations involved in the cell cycle for CS1 (97%, P < 0.001, Supplementary Figures S18 and S19, available at https://doi.org/10.1016/j.annonc.2022.11.002); hence, cyclin D-Cdk4,6 plus PI3K inhibitors could be beneficial for CS1 patients.

Despite not presenting enriched genetic signatures, the CS2 cluster may be potentially susceptible to inhibition of the TFs IRF4 (e.g. lenalidomide), SPIB, and MEIS1 (e.g. MEISi-1), demethylating drugs, and/or inhibition of GAD67. 45,59

To further evaluate the proposed therapeutic targets, we downloaded the RNA-seq expression, the genome-wide CRISPR loss-of-function screens’ scores, and the drug screening response data from 20 DLBCL cell lines. 72-74

We applied the partition around medoids method 70 using our PCNSL RNA-seq data as training expression and found DLBCL cell lines KARPAS-422, A3-KAW, and SU-DHL-4 to be transcriptionally similar to PCNSL subtypes CS1, CS2, and CS3, respectively. Next, we investigated the cumulative gene dependency on different pathways measured by CRISPR screen Chronos score since it has been demonstrated to exhibit the lowest copy number and screen quality bias. 71 We found the CS3-related cell line (SU-DHL-4) to have higher dependency on proteins involved in BCR subunits and BCR-dependent NF-kB activation (Figure 6B), while the CS1-related one benefited from CDKN2A/B knockouts. The PRC2 elements were equally essential for both CS1 and CS3. Interestingly, the IRF4/SPIB knockouts were selectively essential for the CS2-related cell line which corresponds to the MR analysis’ results. When looking at the half maximal inhibitory concentration (IC50) of ibrutinib in these cell lines, we found the lowest value in the CS3-related cell line (IC50 = 1.71; Figure 6C), thus supporting our hypothesis about CS3 subtype sensitivity to BTK inhibitors. Moreover, the CS1-related cell line showed the lowest IC50 toward AMG-319 (IC50 = 10.19), a PI3K-mammalian target of rapamycin (mTOR) inhibitor.

**DISCUSSION**

Identifying groups of patients with shared biologic and prognostic markers is extremely challenging mainly due to high genetic, phenotypic, and TME heterogeneity. Thanks to a collaborative nationwide effort, here, we carried out a multi-omic analysis of a large cohort of immunocompetent EBV™ PCNSL. We identified four PCNSL molecular subtypes with specific oncogenic pathways, gene expression phenotypes, methylation profiles, TME, tumor location, outcome, and potential therapeutic targets. Moreover, our study gives plausible explanations to the PCNSL response heterogeneity based on finding that many previously PCNSL characteristic

---

**Figure 4. Phenotypic analysis of the multi-omic defined PCNSL subtypes.** (A) A heatmap with either gene signature activity (measured by GSVA) or immune cell proportions (CIBERSORTx deconvoluted) across molecular subtypes is shown. P values indicate higher expression of the colored group when compared against the others (Wilcoxon test, left side of plot). (B) The hematoxylin–eosin staining and IHC of CD20, CD3, CD163, PROX1, and H3K27me3 in samples belonging to the different PCNSL subtypes are shown. Scale bars are indicated in each figure.
features, based on MCD or C5 DLBCL subtypes, are cluster-specific (CS1-CS4) and not shared across all PCNSL tumors. For example, PCNSLs (MCD/C5 DLBCLs) are mainly characterized by mutations leading to constitutive NF-κB activation but this was only observed, transcriptionally, for CS4 and CS3; however, the outcome for these clusters is very different mainly due to meningeal infiltration to the CSF probability, TME, HIST1H1E mutations, and B-cell differentiation program differences. Hence, we propose different tailored treatments according to the pathway activation of each CS, suggesting, for example, that CS4 might be more likely to respond to ICI treatment, and BTK/JAK1 inhibitors. Our initial multi-omic analysis along with our DepMap/Genomics of Drug Sensitivity in Cancer (GDSC) analysis on DLBCL cell lines suggest the CS3 subtype to be particularly more susceptible to BTK inhibitors, while the CS2 subtype to lenalidomide, and the CS1 subtype to PI3K inhibitors (Figure 7). When analyzing the TME and epigenetic features across subtypes, we found similar results to those previously described by Kotlov et al. in DLBCLs where the hypermethylated subtypes (CS1/CS2) had a depleted TME.

On top of this and given the importance of routine clinical implementation, we propose RBraLymP (RNA-based Brain Lymphoma Profiler), which uses gene expression data from either FFPE or FF tissue, to identify the PCNSL molecular subtypes associated with multi-omic features. The RBraLymP algorithm is publicly accessible at https://github.com/i54i45/PCNSL-RBraLymP such that existing and new therapy efforts can be directed to the most appropriate patients.

Our study presents various limitations. Identification of subtypes was achieved using relevant features obtained from different static bulk-omics data, and hence does not respond to tumor intra-heterogeneity or single-cell’s cross-talk as other approaches like single-cell RNA-seq or spatial transcriptomics would potentially do. Another limitation would be the absence of a validation PCNSL cohort having...
In summary, our multi-omics analysis builds on the same data types as the discovery cohort; however, this is at the moment difficult given the rarity of the disease. Finally, our analysis on DLBCL cell lines would benefit from further validations using PCNSL-derived cell lines, organoids, or patient-derived xenograft models with drug sensitivity and RNA-seq data.

In summary, our multi-omics analysis builds on the current classification of DLBCL by the addition of the molecular heterogeneity within PCNSL that may inform on its pathogenesis. Our study discovered a link between genetic and neoplastic signaling pathways, pointing to potential treatment targets. Selecting treatment for PCNSL based on individual genetic changes is not desirable from the standpoint of precision medicine, as it is likely that combinations of genetic aberrations influence therapeutic response. The genetic subgroups we define could serve as a conceptual foundation for developing targeted therapeutic approaches for these poorly understood and with high mortality malignancies.

ACKNOWLEDGEMENTS

We thank Yannick Marie and Emeline Mundwiller (Paris Brain Institute), Lucile Armentoul, and Mira Ayadi (Carte d’Identité de Tumeurs, Ligue contre le Cancer) for the help with exome, RNA sequencing, and bisulfite DNA sequencing libraries’ preparation; Silvia Duran for helpful discussions; Figure 7 was created with biorender.com and all the clinicians, pathologists, researchers, and patients associated with the French National ‘Lymphome Oculo-Cérébral’ included in Rare cancers of the central nervous system, RENOCLIP-LOC Network, approved by the French National Institute of Cancer (INCa).

FUNDING

This work was supported by a grant from Investissements d’avenir and by the grant INCA-DGOS-Inserm_12560 of the Siric Curamus [grant number INCa-DGOS-Inserm_12560], the program ‘investissements d’avenir’ [grant number ANR-10-IAIHU-06], PRT-K/INCa grant LOC-model reference 2017-1-RT-04 (no grant number), DRCI de l’APHP (no grant number), CRC2013_105_R1/projet Tri LOC, BETPSY project, overseen by the French National Research Agency, as part of the second ‘Investissements d’Avenir’ program [grant number ANR-18-RHUS-0012], Foundation RAM active investments, ARTC foundation (no grant number), an unrestricted grant from Bristol Myers Squibb (BMS) [grant number RD06618], ICGex project (no grant number), and IDEATIOn project with an unrestricted grant from MSD Avenir (no grant number).
Figure 7. Pathogenetic and therapeutic implications of the PCNSL subtypes. Shown is a schematic representation summarizing the major molecular findings and proposed potential therapeutic targets. Contribution of each molecular subtype to the indicated alteration where color bar width indicates the prevalence of each subtype (Mutational alterations section). Asterisks indicate if a genetic alteration is enriched in any CS subtype (related to Figure 3A). Arrows indicate transcriptional gene signature activity where the height indicates the relative up- or negative regulation (according to Figure 4A).

DISCLOSURE

GA reports grants from Biogen, Novartis, Roche, Sanofi, Abbvie, Pfizer, and CSL Behring, outside the submitted work. AA reports research grant with an unrestricted grant from Bristol-Myers Squibb (BMS). All other authors have declared no conflicts of interest.

REFERENCES


