Genomic Landscape and Molecular Subtypes of Primary Central Nervous System Lymphoma

Shengjie Li[1,2,3,4,5,6*#], Danhui Li[7#], Zuguang Xia[8,9#], Jianing Wu[6], Jun Ren[6], Yingzhu Li[6], Jiazhen Cao[10], Ying Sun[11], Liyang Zhang[12], Hongwei Ye[12], Chengxun Li[13*], Wenjun Cao[6*], Xingtao Zhou[5*], Ying Mao[1,2,3,4,14*]

1. Department of Neurosurgery, Huashan Hospital, Department of Clinical Laboratory, Eye & ENT Hospital, Shanghai Medical College, Fudan University, Shanghai, China

2. National Center for Neurological Disorders, Shanghai, China

3. Shanghai Key Laboratory of Brain Function Restoration and Neural Regeneration, Shanghai, China

4. Neurosurgical Institute of Fudan University, Shanghai, China

5. Eye Institute and Department of Ophthalmology, Eye & ENT Hospital, Fudan University, Shanghai, China

6. Department of Clinical Laboratory, Eye & ENT Hospital, Shanghai Medical College, Fudan University, Shanghai, China

7. Department of Pathology, RenJi Hospital, School of Medicine, Shanghai JiaoTong University, Shanghai, 200127, China

8. Department of Medical Oncology, Fudan University Shanghai Cancer Center, Fudan University, Shanghai, China

9. Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China

10. Department of Clinical Laboratory, Huashan Hospital, Fudan University, Shanghai, China

11. GenomiCare Biotechnology (Shanghai) Co. Ltd., Shanghai, China

12. Guangzhou KingMed Center for Clinical Laboratory Co. Ltd., Guangzhou, China

13.Institutes of Biomedical Sciences, Fudan University, Shanghai, 200032, China

14. Lead contact

#Shengjie Li, Danhui Li, and Zuguang Xia contributed equally to this work.

Corresponding Authors:

Wenjun Cao, Department of Clinical Laboratory, Eye & ENT Hospital, Fudan University, No. 83 Fenyang Road, Shanghai, China; Electronic address: wgkjyk@aliyun.com.

Chengxun Li, Institutes of Biomedical Sciences, Fudan University, Shanghai, 200032, China Electronic address: licx21@m.fudan.edu.cn

Xingtao, Zhou, Eye Institute and Department of Ophthalmology, Eye & ENT Hospital, Fudan University, Shanghai, 200031, China. Electronic address: doctzhouxingtao@163.com.

Ying Mao, Department of Neurosurgery, Huashan Hospital, Fudan University, No. 12 Wulumuqi Road, Shanghai, 200040, China; Electronic address: maoyin@fudan.edu.cn

Abstract

Primary central nervous system lymphoma (PCNSL) is a rare and aggressive brain tumor with a poor prognosis and is almost exclusively diffuse large B-cell lymphoma (DLBCL). Its genetic characteristics and molecular subtypes in Chinese patients remain poorly understood, which in turn makes developing effective new therapies challenging. By enrolling 176 newly diagnosed Chinese PCNSLs at five tertiary care centers with extensive follow-up, we performed a genomic study aimed at expanding the genomic landscape and developing new molecular subtypes. We first confirmed that the molecular subtyping of DLBCL, as previously published, is not applicable to Chinese PCNSLs. We then identified (n = 58) and validated (n = 82) three prominent genetic subtypes related to different clinical and molecular features of PCNSL, and further confirmed by an independent external Chinese PCNSL cohort (n = 36). We called these BMIs (from the co-occurrence of mutations in two genes among BTG1, MYD88, and IRF4), which are associated with favorable outcomes; E3s (so-called EP300 mutations), which are associated with unfavorable outcomes; and UCs (unclassified, without characteristic mutations). Importantly, EP300 was mutated in more Asians (16.98%) than in Western PCNSLs (< 4.53%), resulting in unfavorable outcomes independent of the specific mutation site. Our analysis comprehensively reveals the genomic landscape of Chinese PCNSL and emphasizes the clinical value of molecular classification for improving precision medicine strategies.

Keywords: Primary central nervous system lymphoma; Genomic landscape; Molecular subtypes; EP300; Precision medicine

Introduction

Primary central nervous system lymphoma (PCNSL), a rare subtype of extranodal non-Hodgkin lymphoma, accounts for >60% of PCNSL cases and has an unfavorable prognosis even when the patient receives standard treatment, which is based on a high-dose methotrexate (HD-MTX) regimen[1,2]. PCNSL is a unique lymphoma that differs from other tumors and is characterized as a genetically heterogeneous disorder marked by a variety of low-frequency mutations, somatic copy number alterations, and structural variants[3–6]. Thus, combining standard immunochemotherapy with promising novel agents that target specific pathways via different molecular clusters may improve patient prognosis[7,8].

Pathologically, nearly 95% of PCNSLs are diffuse large B-cell lymphomas (DLBCLs). Recently, DLBCL has been categorized into various molecular clusters on the basis of genomic sequencing and RNA sequencing, including quartet categorization by *L.M. Staudt*'s team[9], quintet categorization by *M.A. Shipp*'s team[10], *L.M. Staudt*'s team's refined septet method[11], and the LymphPlex classification proposed by *W.L. Zhao*'s team[12]. However, PCNSL has been proven to be a biological entity that is molecularly distinct from DLBCL[13], so these molecular clusters of DLBCL may not be applicable to PCNSL. Notably, the integration of genome-wide data from multiomic studies by *A. Alentorn* et al.[14] showed four molecular patterns of PCNSL mainly use gene expression data. These patterns have a distinctive prognostic impact, providing a basis for future clinical stratification and subtype-based targeted interventions. Owing to the molecular heterogeneity between various ethnic groups of PCNSL patients[3,15], the applicability of the proposed classifications for PCNSL to other populations requires further confirmation. No effective biomarker or molecular classification scheme exists to tailor therapies for individual PCNSL patients. Furthermore, the issue of PCNSL molecular heterogeneity in Chinese patients has not been adequately addressed, primarily because the data have come from small, single-center studies of Chinese PCNSL patients[16,17].

To address these issues, we conducted whole-exome sequencing (WES) on specimens obtained from 140 patients with newly diagnosed PCNSL who were seen at one of three tertiary care cancer centers. These patients were divided into a discovery cohort and a

validation cohort. Among the 140 patients, 94.29% received treatment with an HD-MTX-based chemotherapy regimen followed by consolidation therapy (detailed in Table S1), and all 140 patients underwent long-term follow-up. Additionally, WES/Whole Genome Sequencing (WGS) data were obtained from 36 Chinese PCNSL patients at two other tertiary care cancer centers[17], serving as an independent external cohort. This representative and clinically annotated PCNSL cohort was utilized to comprehensively detect mutations. Importantly, to facilitate routine clinical implementation, we identified three molecular subtypes composed only of single-nucleotide variations (SNVs); these SNVs were associated with different outcomes, and this was confirmed in the validation cohort and in an independent external Chinese PCNSL cohort (n=36)[17]. Our results comprehensively reveal the genomic landscape of PCNSLs in Chinese patients and provide a clinically actionable PCNSL classification system, thereby improving precision medicine strategies.

METHODS

Sample collection and clinicopathological features of PCNSL patients

In this study, a total of 140 newly diagnosed PCNSL patients were retrospectively enrolled from January 2010 to December 2020 from three independent medical centers, namely, Huashan Hospital of Fudan University, Renji Hospital of Shanghai Jiao Tong University, and Shanghai Cancer Center of Fudan University. The study comprised two cohorts: the discovery cohort, which included 58 patients from Huashan Hospital of Fudan University, Shanghai; and the validation cohort (n = 82), which consisted of 25 patients from Huashan Hospital of Fudan University, 51 patients from Renji Hospital of Shanghai Jiao Tong University, and 6 patients from Shanghai Cancer Center of Fudan University. In the discovery cohort, paired blood samples were also obtained for analysis in conjunction with each tumor sample. In the validation cohort, only unpaired FFPE tumor tissue samples were obtained for analysis.

Additionally, WES/WGS data of 36 patients newly diagnosed with PCNSL from Fujian Cancer Hospital and The First Affiliated Hospital of Fujian Medical University between February 2012 and October 2020 were included in this study, as an independent external Chinese PCNSL cohort. The general clinical characteristics of these 36 PCNSL patients and

119 the sequencing protocol have been previously described[17].

120 None of the patients had received steroid treatment for PCNSL, and PCNSL tumors were
121 diagnosed on the basis of the World Health Organization criteria[13]. All patients underwent a
122 2-deoxy-2[F-18] fluoro-D-glucose positron emission tomography/computed tomography
123 (FDG PET/CT) scan and bone marrow aspiration to exclude systemic tumor manifestation
124 on the basis of the guidelines of the International PCNSL Collaborative Group[18] and the
125 European Association of Neuro-Oncology[19]. PCNSL patients for whom ocular involvement
126 was present were excluded. Each tumor tissue sample was independently reviewed by at
127 least 2 pathologists to confirm that the tumor sample was histologically consistent with the
128 PCNSL. All patients underwent long-term follow-up, which ran to December 30, 2022.

129 **Ethics statement**

130 This study was conducted with the approval of the Institutional Review Boards at Huashan
131 Hospital of Fudan University (Approval No. 2022-529), Shanghai Cancer Center of Fudan
132 University (Approval No. 1612167-18), Renji Hospital of Shanghai Jiao Tong University
133 (Approval No. LY2024-112-C), and Fujian Cancer Hospital (Approval no. SQ 2020-106-01).
134 All participants provided written informed consent before participating in the clinical study and
135 before the collection of tumor tissues. The research was carried out in strict adherence to the
136 ethical principles of the Declaration of Helsinki and conformed to international norms of good
137 clinical practice.

138 **Treatment Regimens**

139 All PCNSL patients were treated with HD-MTX-based combination immunochemotherapy,
140 followed by consolidation therapy (detailed in Table S1). The following chemotherapy
141 protocols were used for induction: HD-MTX combined with IDA, HD-MTX combined with R,
142 HD-MTX combined with IDA and R, or a combination of BTK inhibitors. Whole-brain radiation
143 therapy or stem cell transplantation is a form of consolidated therapy. The detailed therapeutic
144 schedule was the same as that previously described[20].

145 **Sample processing**

146 The collection and processing of samples for this study were carried out by the Department
147 of Pathology or Department of Neurosurgery at Huashan Hospital of Fudan University, Renji

Hospital of Shanghai Jiao Tong University, and Shanghai Cancer Center of Fudan University. In the discovery cohort, surplus tumor tissues and paired blood samples obtained during surgical procedures were collected. To guarantee the quality of these samples, each specimen was promptly and accurately labeled with relevant patient information within 30 minutes of collection and then immediately snap-frozen in liquid nitrogen. These samples were stored at -80°C until being shipped to GenomiCare Biotechnology (Shanghai, China). In the validation cohort, the residual tumor tissue specimens were subjected to FFPE. These FFPE samples were transported in a dry ice container to Sinotech Genomics Technologies in Shanghai, China, accompanied by a time and temperature tracker to ensure proper monitoring during shipment. Upon arrival, the samples were promptly stored in liquid nitrogen to maintain their integrity until further processing.

DNA extraction

For frozen fresh tissue and matched peripheral blood samples, total genomic DNA (gDNA) was extracted via the Maxwell RSC Blood DNA Kit (AS1400, Promega) on a Maxwell RSC system (AS4500, Promega) following the manufacturer's instructions. For FFPE tissue, total gDNA was extracted via the QIAamp DNA FFPE Tissue Kit (56404, Qiagen) following the manufacturer's instructions. The integrity and concentration of the gDNA were determined via the Qsep100 System (BIOptic, China) and a Qubit 3.0 fluorometer (Thermo Fisher Scientific). The OD_{260} was measured with a NanoDrop One (Thermo Fisher Scientific).

WES library preparation and sequencing

For frozen tissue and matched peripheral blood samples, exome DNA was extracted via the SureSelect Human All Exon V7 Kit (5991-9039EN, Agilent). The library was subsequently prepared via the SureSelectXT Low Input Target Enrichment and Library Preparation system (G9703-90000, Agilent). For the FFPE sample, the exome DNA was captured via the SureSelect Human All Exon V8 Kit (5191-6874, Agilent) and prepared into a library via the SureSelect XT HS2 DNA Reagent Kit (G9983A, Agilent). The library was validated via the use of an Agilent 2100 Bioanalyzer and a Qubit 3.0 fluorometer. Paired-end 150 bp read sequencing was performed on an Illumina NovaSeq 6000. Image analysis and base calling were performed via onboard RTA3 software (Illumina).

177 **Somatic mutation calling**

178 Quality control was conducted via FastQC software. The cleaned and trimmed FASTQ files
179 were aligned to the UCSC human reference genome (hg19) via the Burrows–Wheeler Aligner
180 (BWA) with default parameters. For each paired blood sample, SNVs and short
181 insertions/deletions (INDELs) were identified via Sentieon TNseq[21] with default parameters.
182 The identified somatic mutations were removed if they did not satisfy any one of the following
183 criteria: a variant allele frequency (VAF) of at least 0.05, support from a minimum of three
184 reads, annotation by the Variant Effect Predictor (VEP) package[22], and transformation into a
185 mutation annotation format (MAF) file for further analysis with maftools[23].

186 When a paired normal sample (FFPE sample) was not available, a panel of normal (PON)
187 files was used as a paired control to call the SNVs and InDels. A panel of normal files was
188 created using the reads from clinical blood samples that showed no evidence of tumor
189 contamination from 100 different individuals collected by Genomicare Biotechnology
190 (Shanghai). The reads were retained when they passed the standard WES QC, as described
191 in the WES subsection. The reads for reference alleles and alternative alleles were gathered
192 for each candidate site in all 100 samples, except for false positives, which were those
193 exhibiting a candidate allele frequency exceeding 0.05 in more than 5 samples.

194 We applied an additional PON mask for all the candidate somatic SNV and INDEL sites to
195 exclude germline mutations, as previously reported[10,24]. In general, in addition to the standard
196 retention rules, mutations were retained if they were known driver alterations highlighted as
197 being biologically significant in the COSMIC database. Mutations with a tumor mutation
198 frequency ≥20% were also retained. Other germline mutations were retained if their existence
199 ratio in the Exome Aggregation Consortium (ExAC) exac_all was < 0.00005, their existence
200 ratio in the Asian population exac_eas was < 0.0003, and their existence ratio in previous
201 GenomiCare samples germline_gc was < 0.0006. Retained mutations were considered
202 somatic mutations.

203 Thus, the PON file and PON mask limit possible recurrent artifacts of sequencing and the
204 presence of germline mutations in somatic mutations that are falsely selected owing to a lack
205 of paired blood data.

206 **Somatic copy number alteration calling and profiling**

207 Using the method described in the ExomeCNV package[25], we implemented a normalized
208 depth–coverage ratio method to identify CNVs in paired samples. To correct for five potential
209 biases (the size of exonic regions, batch effects, both the quantity and quality of sequencing
210 data, local GC content, and genomic mappability) that affect the raw read counts, we utilized
211 a standard normal distribution model. Genes with a haploid copy number ≥3 or ≤1.2 were
212 defined as amplified or deleted, respectively, and a minimum tumor content (purity) of 20%
213 was needed.

214 **Significance analysis of recurrent somatic copy number alterations**

215 Sequenza[26] was used to generate the segment files. The segment files were used as the
216 input for GISTIC2.0[27] to detect recurrent arm-level and focal peaks in copy number alterations
217 via the parameters -js 40 -conf 0.99 -ta 0.2 -td 0.2 -brlen 0.8, with other parameters set to the
218 defaults.

219 **Cancer-driver gene mutations filtering**

220 After the aforementioned steps of calling SNVs and CNVs, we further filtered the resulting
221 mutated genes by intersecting them with a designated group of genes recognized as cancer-
222 related genes. This group of genes was sourced from two prominent public cancer gene
223 databases. Part 1 of the group was obtained from the OncoKB curated cancer gene list
224 (https://www.oncokb.org/cancerGenes)[28]. These genes are classified as cancer genes by
225 OncoKB on the basis of their inclusion in various sequencing panels, the Sanger Cancer
226 Gene Census, or the criteria established by *Vogelstein* et al.[29]. Part 2 was derived from the
227 ranked CIViC gene candidates table (https://github.com/griffithlab/civic-
228 server/blob/master/public/downloads/RankedCivicGeneCandidates.tsv)[30]. The final list of
229 cancer-related genes consisted of all genes in Part 1 and genes in Part 2 only when its
230 'panel_count' value (in the CIViC tsv file listed above) was ≥ 2.

231 **Bioinformatic analysis**

232 The mutational signature classification was based on all single-base substitution (SBS)
233 signatures of the COSMIC mutational signatures[31]. TMB was defined as the aggregate count
234 of somatic nonsynonymous mutations, including SNVs or INDELs, within the tumor exome

for each patient[32]. This count was then divided by the overall size of the targeted regions (35 for WES), resulting in the TMB, expressed in counts per megabase (counts/Mb). The MATH score[33] was calculated on the basis of the width of the variant allele frequency (VAF) distribution, utilizing maftools for the analysis[23]. The pathway map was generated via maftools as previously described[34].

Classification model for identifying molecular types

CoxNet survival analyses and least absolute shrinkage and selection operator (LASSO) regressions were performed via the scikit-learn package in Python (version 3.9.7). K–M survival curves (log-rank test) were used for OS analysis via the survival and survminer packages in R (version 4.1.2).

(1) Selection of 24 initial features in the discovery cohort

Molecular features derived from SNVs were selected to develop subsequent classification models.

First, to identify genes with high variability for use in clustering, those with low mutation frequencies were filtered out. This process led to the selection of genes with a mutation frequency greater than 5% for clustering purposes. A total of 150 genes were screened for further analysis.

In the second phase, OS was analyzed as the dependent variable, with the aforementioned 150 genes serving as independent variables. The feature selection steps were as follows: (1) a comprehensive approach involving CoxNet survival analyses and LASSO regressions was employed to identify characteristic genes; (2) the importance score of each feature in the model was calculated and ranked in descending order; (3) features that ranked in the top 50% were recorded as potential features once; (4) processes 1-3 were repeated 500 times, and all potential features were recorded; and (5) among the feature sets obtained in the above process, those recorded more than 250 times were selected.

Ultimately, 24 feature genes, namely, EP300, KMT2D, MPEG1, TUSC3, MYD88, PTCH1, DST, ZNF521, BTG1, NOTCH2, PRDM1, ADGRL3, CDH11, CREBBP, ATRX, TCL1A, GNA13, ETV6, IRF4, HIST2H3D, CD79B, GRM8, MYH11, and SPTA1, were selected.

(2) Differential feature selection

To determine the most suitable features among the initial 24 for the classification model, Kaplan–Meier analyses were conducted. Features qualifying for further analysis met the following criteria: a log-rank test P value of less than 0.2 and any group with more than 5 patients. Ultimately, 6 DEGs, namely, EP300, CREBBP, DST, MYD88, BTG1, and IRF4, were identified.

(3) Model training utilizing random combinations of 6 differential features

To adhere to clinical practice guidelines for predicting patient outcomes and implementing individualized management strategies for PCNSL patients, patients with PCNSL were categorized into three subgroups (those with good, poor, and intermediate prognoses) in the present study. Six differential features were ultimately selected for building a classification model to classify patients into the three subtypes. These features were randomly combined to form three types, yielding a total of 288 combinations. To identify the most appropriate classification model among these combinations, Kaplan–Meier analyses were performed. Models were chosen on the basis of a log-rank test P value of less than 0.05 and subgroup sample sizes greater than 8. Ultimately, seven candidate molecular subtype combinations met these criteria: (molecular markers shown in the format of subtype 1 vs. subtype 2 vs. subtype 3): ① EP300 (predominant) vs. MYD88/BTG1/IRF4≥2 vs. others; ② EP300 vs. MYD88/BTG1/IRF4≥2 (predominant) vs. others; ③ BTG1 (predominant) vs. CREBP/DST≥1 vs. others; ④ BTG1 vs. CREBP/DST≥1 (predominant) vs. others; ⑤ BTG1/IRF4≥1 vs. EP300/CREBP/DST≥1 vs. others; ⑥ BTG1/IRF4≥1 vs. EP300/DST≥1 vs. others; and ⑦ CREBP/DST≥1 vs. MYD88/BTG1/IRF4≥2 vs. others.

(4) Validation of candidate molecular subtypes

The WES data and OS data of 82 patients in the validation cohort were included to validate the candidate molecular subtypes. The mutation frequencies of EP300, CREBBP, DST, MYD88, BTG1, and IRF4 in the validation cohort were greater than 5%. Kaplan–Meier survival analysis was used to validate the seven candidate molecular subtype combinations. The performance of [EP300 (predominant) vs. MYD88/BTG1/IRF4≥2 vs. others] (3-way P = 0. 0202) and [EP300 vs. MYD88/BTG1/IRF4≥2 vs. others (predominant)] (3-way P = 0. 0346) was better than those of the other five molecular subtypes (3-way P > 0.05 for both).

Additionally, compared with the [EP300 vs. MYD88/BTG1/IRF4≥2 vs. others (predominant)] molecular subtype, the Kaplan–Meier survival curve of [EP300 (predominant) vs. MYD88/BTG1/IRF4≥2 vs. others] showed no crossover. Finally, the performance of [EP300 (predominant) vs. MYD88/BTG1/IRF4≥2 vs. others] was the best, so we selected it as the molecular subtype of Chinese PCNSL patients. In other words, PCNSL patients with both EP300 and BMI (BTG1, IRF4, and MYD88) variants were categorized into the E3 subtype.

Hematoxylin & eosin and immunohistochemical staining

PCNSL tumor tissue was fixed with 10% paraformaldehyde and embedded in paraffin. The paraffin-embedded tissues were cut into 4-μm-thick sections. These sections were then subjected to staining with hematoxylin and eosin (H&E) or for immunohistochemistry. For H&E staining, the sections were stained with hematoxylin and eosin following standard protocols (ab245880, Abcam). Immunohistochemistry was performed for ki-67 (ab15580, Abcam), CD2 (ab314761, Abcam), CD5 (ab75877, Abcam), CD10 (ab256494, Abcam), CD19 (ab134114, Abcam), CD20 (ab78237, Abcam), CD79a (ab79414, Abcam), BCL2 (MA5-11757, Thermo Scientific), BCL6 (PA5-14259, Thermo Scientific), MUM1 (ab247079, Abcam), c-Myc (MA1-980, Thermo Scientific), and p53 (MA5-14516, Thermo Scientific) via an automated Leica BOND-III staining system. H&E and immunohistochemistry images were obtained with a KFBIO scanner (KF-PRO-005-EX, Zhejiang, China) and visualized via KFBIO SlideViewer software. Each staining was independently reviewed by at least 2 pathologists.

In situ hybridization of fluorescence

PCNSL tumor tissue was fixed with 10% paraformaldehyde and embedded in paraffin. The paraffin-embedded tissues were cut into 4-μm-thick sections. Fluorescence in situ hybridization was performed in accordance with standard protocols via the following commercial probes: MYC break-apart, BCL-2 break-apart, and BCL-6 break-apart probes (LBP Medicine Science & Technology Co., Ltd., Guangzhou, China). Staining was independently evaluated by two hematopathologists, and discrepancies were resolved by another hematopathologist.

HIV, HBV, and EBV detection

Antibody and antigen levels of HIV were measured via an Elecscy HIV combi PT assay kit

322 (Roche, Germany) with a Cobas e 601 analyzer (Roche). Hepatitis B virus (HBV)-specific
323 PCR was performed via an HBV virus nucleic acid test kit (Sansure Biotech, China) with an
324 Applied Biosystems 7500 Real-Time PCR System (Thermo Scientific). EBV-encoded RNA in
325 situ hybridization (EBER-ISH) was performed to assess the presence of Epstein-Barr virus
326 (EBV) in tumor tissue. The testing utilized a fluorescein isothiocyanate -coupled specific
327 peptide nucleic acid probe (Roche Diagnostics, Mannheim, Germany).

328 **Statistical analysis**

329 All the statistical analyses were performed via R version 4.1.2 and GraphPad Prism version
330 9. The specific statistical analyses employed for each figure are detailed in the respective
331 legends. We used PASS software to calculate the sample size for our study, with a
332 significance level of 0.05 and a power of 0.80. The observed group proportions were 0.6, 0.2,
333 and 0.2, whereas the expected proportions were 0.33 for each group. The sample size
334 required to meet these criteria was 25 participants. Consequently, the sample size utilized in
335 this study is deemed adequate. Whole-exome sequencing data were processed via nf-
336 core/sarek v3.4.2[35] of the nf-core collection of workflows[36], which uses reproducible software
337 environments from the Bioconda[37] and Biocontainers[38] projects. The pipeline was executed
338 with Nextflow v24.04.2[39].
339

340 **Results**

341 **PCNSL genomic landscape**

342 Using WES, we detected mutations in 140 patients who were newly diagnosed with PCNSL
343 at one of three cancer centers; in 59% of cases, blood samples were lacking. Because EBV-
344 positive PCNSL is recognized as an immunodeficiency-associated PCNSL in the current
345 WHO CNS5 classification[40,41], all participants in the study who tested negative for EBV were
346 included. Additionally, DLBCL patients with HBV and HIV infection may have different
347 mutational profiles[42,43]; thus, all participants in the study who tested negative for HIV and HBV
348 were included. The key demographic and clinical characteristics of the patients are
349 summarized in **Figure S1** and **Table S1**. The study design and flow chart are displayed in
350 **Figure 1A**. After filtering, we identified 115,938 genetic variation events in the 140 PCNSL

351 samples analyzed (median = 19.06 mutations/Mb, range = 0.24–91.2 mutations/Mb; median
352 number of variants = 949.5/sample) (**Figure S2A, Figure S3A**).
353 We applied a filter to the 140 PCNSLs on the basis of the genes curated in the CIViC and
354 OncoKB databases to identify candidate cancer genes with hallmark mutations in PCNSL.
355 Some of the mutated genes were PIM1 (in 62.1% of PCNSLs), MYD88 (55.0%), KMT2D
356 (48.6%), CD79B (45.7%), BTG2 (38.6%), LRP1B (37.9%), FAT4 (37.1%), RNF213 (35%),
357 FAT1 (34.3%), ZFHX3 (33.6%), DTX1 (31.4%), KMT2C (31.4%), SETD1B (31.4%), MPEG1
358 (29.3%), CREBBP (27.9%), ANKRD11 (27.1%), KAT6B (27.1%), BTG1 (25%), and PTEN
359 (25%) (**Figure 1B**), which are involved in chromatin histone modification, BCR-TLR-mediated
360 NF-κB signaling, immune signaling, the cell cycle, PI3 kinase signaling, MAPK signaling, and
361 Wnt/β-catenin signaling (**Figure S2B**)[3,5,9–11,44–51]. In this study, we also identified several
362 other pathways in PCNSL, such as pathways related to genome integrity, RTK signaling, RNA
363 abundance, TGFB signaling, TOR signaling, and apoptosis (**Figure S2B**), many of which
364 have defined roles in other cancers[34]. The mutually exclusive or cooccurring relationships of
365 these genes are shown in **Figure S3D**. Survival associated with candidate cancer genes with
366 mutation frequencies greater than 15% plus the NOTCH2 and EZH2 genes[9] is shown in
367 **Figure S4**.
368 In the search for focal copy number alterations (CNAs) in the 140 PCNSLs, we detected
369 significant recurrent amplifications at chromosomal locations 1q21.1, 4q16.3, 11p15.5,
370 11q12.2, 17q12, 19p13.3, 22q11.1, Yq11.221, and Yq12. We also detected deletions at 11q11,
371 11q12.1, 14q32.33, 17q11.2, 18p11.21, 19p12, 19q13.42, and 22q11.21 (**Figure 1C**). The
372 arm-level CNAs among all 140 PCNSL samples are displayed in **Figure S2C**. The pattern of
373 somatic mutations caused by the different mutational processes in the genome, termed
374 *mutational signatures*, was calculated and compared to the well-established signatures in
375 COSMIC[52]. The mutational signature contributions are shown in **Figure S2D**. The COSMIC
376 5, COSMIC 45, and COSMIC 1 signatures contributed the most to the PCNSL genome
377 (**Figure S2E**). These factors are associated with aging, tobacco smoking, NER deficiency,
378 the regulation of oxidative DNA damage repair, and the spontaneous deamination of 5-
379 methylcytosine. Taken together, these results establish a comprehensive genomic landscape

380 of PCNSL in Chinese individuals.

381 **Comparison of the genomic landscapes of the discovery and validation cohorts**

382 Owing to differences in sample type between the discovery and validation cohorts (fresh
383 tumor [discovery cohort] vs. formalin-fixed paraffin-embedded [FFPE] [validation cohort]) and
384 the presence or absence of matched samples (paired [discovery cohort] vs. tumor-only
385 [validation cohort]), we applied further filtering and recalibration steps to reduce false-positive
386 calls in the validation cohort (for details, see the Methods subsection on somatic mutation
387 calling) and compared the genomic landscapes of the discovery and validation cohorts. The
388 key demographic and clinical characteristics of the patients in the discovery and validation
389 cohorts are summarized in **Table S2**. We identified 12,222 genetic variation events (median
390 = 3.84 mutations/Mb, median variants = 159/sample, **Figure S3B, Figure S5A**) in the
391 analyzed discovery cohort of PCNSL samples and 103,716 genetic variation events (median
392 = 6.85 mutations/Mb, median variants = 1597/sample, **Figure S3C, Figure S5B**) in the
393 analyzed validation cohort of PCNSL samples. The genetic variation profiles of the discovery
394 and validation cohort of PCNSL samples were similar, as revealed by principal component
395 analysis (PCA) (**Figure 2A**), but the mutation rate was higher in the validation cohort of
396 PCNSL samples (**Figure 2B**).

397 PIM1, MYD88, CD79B, and KMT2D were the most frequently mutated candidate cancer
398 genes of PCNSL. PIM1 (70.7%), MYD88 (63.8%), CD79B (51.7%), and KMT2D (39.7%)
399 (**Figure S5C**) were identified in the discovery cohort, and PIM1 (56.1%), MYD88 (48.8%),
400 CD79B (41.5%), and KMT2D (54.6%) were also observed at similar mutation frequencies in
401 the validation cohort (**Figure S5D**). However, several of the less frequently mutated candidate
402 cancer genes in PCNSL whose mutation frequencies were significantly different when the
403 discovery and validation cohorts were compared included ZFHX3, TRRAP, SZT2, RNF213,
404 RANBP2, PRKDC, NUMA1, NOTCH1, NOTCH2, KMT2C, KMT2B, KAT6B, FLT3, FAT4,
405 FAT1, and ANKRD11 (all $P < 0.05$) (**Figure 2C**). The arm-level CNAs between the discovery
406 and validation cohort of PCNSL samples were mostly similar and are displayed in **Figure 2D**.
407 Significant differences in both recurrent amplifications (1q, 8q, and 12p) and deletions (6p,
408 8p, 19p, and 19q) were observed. The results of significant recurrent focal amplifications and

deletions were visualized via GISTIC 2.0 in the discovery (**Figure S5E**) and validation (**Figure S5F**) cohorts.

The tumor mutational burden (TMB) describes the number of mutations in one tumor sample and is often related to the prognosis of patients receiving clinical treatments[53]. The mutant-allele tumor heterogeneity (MATH) score reflects intratumor heterogeneity[33]. Microsatellite instability (MSI) arises from the impaired function of DNA mismatch repair proteins, and its molecular characteristics are regarded as significant genetic markers[54]. Although the TMB (**Figure 2E**), MATH score (**Figure 2F**), MSI (**Figure 2G**), and the variant allele frequency (VAF) (**Figure 2H**) were greater in the validation cohort than in the discovery cohort ($P < 0.001$), the mean values were relatively close. The 3 signatures extracted from the discovery samples presented cosine similarities of 86.7%, 93.1% and 94.4% to COSMIC signatures 6, 45, and 10a, respectively (**Figure S5G**), whereas those extracted from the validation samples presented similarities of 80.4%, 70.9%, and 77.9% to COSMIC signatures 5, 1, and 23, respectively (**Figure S5H**). Taken together, these results suggest that the genetic variation profiles of the discovery and validation PCNSL samples exhibit certain discrepancies.

The unique genomic landscape of PCNSLs in Chinese patients

To date, only single-center studies with small sample sizes have reported heterogeneity in the genomic landscape among different populations of Chinese PCNSL patients[15]. Thus, in this large, multicenter study, we further compared the genetic landscape of our cohort with previously published PCNSL/DLBCL data. The frequencies of recurrent genetic alterations in PCNSL patients were compared with those in DLBCL patients from the United States, Finland, Denmark, Israel, the United Kingdom, Hong Kong, and India[49]. Mirror bar plots (**Figure 3A**) revealed that a set of genes, such as PIM1 (62.1% vs. 16.88%), MYD88 (55% vs. 18.01%), IRF4 (15% vs. 3.98%), and EP300 (19.29% vs. 5.97%), had significantly higher mutation rates in Chinese PCNSL patients than in DLBCL patients, suggesting genetic diversity of the cancer genome between diseases. Compared with those in PCNSLs in French patients [14] (**Figure 3B**), the mutation frequencies of EP300 (19.29% vs. 3.48%) and KMT2D (48.6% vs. 22.61%) were significantly higher in the PCNSLs of Chinese patients. Compared with those

438 in the PCNSLs[3] of Japanese patients (**Figure 3C**), the mutation frequencies of MYD88 (55%
439 vs. 85.6%) and BTG2 (38.6% vs. 87.8%) were significantly lower in the PCNSLs of Chinese
440 patients. These data revealed a unique mutational landscape of Chinese PCNSL patients in
441 terms of the frequency of SNVs. Next, the frequencies of recurrent SNAs with the GISTIC2.0
442 program in Chinese PCNSL patients were compared with those in DLBCL patients. At the
443 focal level of SNAs, Chinese PCNSL samples had significant levels of 11q12.4, 18p11.21,
444 19q13.42, 22q11.1, and 22q11.21 deletions along with amplifications in arms 1q21.1, 4p16.3,
445 11q12.2, and 17q12 (**Figure 3D**). Compared with PCNSLs from Japanese patients, those
446 from Chinese patients also had significant levels of deletions in arms 11q12.1, 18p11.21,
447 19q13.42, 22q11.1, and 22q11.21 along with amplifications in arms 1q21.1, 4p16.3, 11q12.2,
448 and 17q12 (**Figure 3E**). These results suggest genetic diversity in the PCNSL genome
449 between races.

450 Owing to the genetic diversity of the PCNSL genome across races and the genetic diversity
451 between the PCNSL and DLBCL cancer genomes, the existing molecular subtyping methods
452 for PCNSL[14] and DLBCL[9–12,55,56] may not be applicable to Chinese PCNSL patients. We
453 assessed the prognostic value of these published molecular subtypes for overall survival (OS)
454 and progression-free survival (PFS) in our Chinese PCNSL cohort. Actionable PCNSL
455 classification can perfectly differentiate prognoses, thereby improving precision medicine
456 strategies. According to cell-of-origin (COO) molecular subtyping[55,56], there were no
457 significant differences in OS (P = 0.42) or PFS (P = 0.67) between the germinal center type
458 (GCB) and non-GCB subtypes (**Figure S6A**). Similarly, according to LymphPlex molecular
459 subtyping[12], no significant differences in OS (P = 0.68) or PFS (P = 0.71) were observed
460 between the BN2-like, MCD-like, EZB-like, N1-like, TP53, and other subtypes (**Figure S6B**).
461 We further categorized our data on the basis of the DLBCL subtyping (C0-C5) proposed by
462 *M. A. Shipp* et al.[10] Within our cohort, the C4 subtype was associated with shorter OS (P =
463 0.009, P = 0.013) and PFS (P = 0.096, P = 0.047) than the C0 and C5 subtypes, respectively
464 (**Figure S6C**). However, the overall efficacy of this DLBCL subtyping system in predicting
465 outcomes was less than satisfactory (OS, P = 0.074). The data were also analyzed via the
466 five subtypes (BN2, EZB, MCD, N1, and others) [9] and seven subtypes (A53, BN2, EZB, MCD,

467 N1, ST2, and others)[11] of DLBCL proposed by *Louis M Staudt* et al. There were no significant
468 differences in OS ($P > 0.05$) or PFS ($P > 0.05$) between BN2, N1, and other categories within
469 the 5-subtype model (**Figure S6D**). Overall, the seven-subtype model (**Figure S6E**) had
470 inferior performance, with no significant differences observed in OS ($P = 0.61$) or PFS ($P =$
471 0.47). With respect to the four subtypes (CS1-4) of PCNSL proposed by A Alentorn [14], our
472 data revealed that the OS ($P = 0.12$) and PFS ($P = 0.54$) of patients with the CS3 subtype
473 were shorter than those of patients with the other three subtypes (**Figure S6F**), but the
474 differences were not significant. Furthermore, the double-hit gene expression signature
475 defines a distinct subgroup of germinal center B-cell-like DLBCL, which is associated with a
476 poor prognosis[13]. However, in this study, there were no significant differences in OS ($P > 0.05$)
477 or PFS ($P > 0.05$) between the double- or triple-hit PCNSL patients and the other PCNSL
478 patients, as shown in **Figure S7** (A: all; B: discovery cohort; C: validation cohort). Taken
479 together, these results suggest that the molecular subtyping of PCNSL and DLBCL, as
480 previously published, is not applicable to Chinese PCNSL patients.

481 **PCNSL molecular subtypes with clinical outcome implications**

482 To establish a molecular classification system for PCNSL, we initially carried out consensus
483 clustering on 58 patients (discovery cohort) who received immunochemotherapy (HD-MTX-
484 based chemotherapy regimen followed by consolidation therapy) via WES data to generate
485 cluster assignments (refer to Methods - Classification model to identify molecular types).
486 WES data from 82 patients (validation cohort) who received the same treatment regimen
487 were used to confirm the molecular clusters. Additionally, WES/WGS sequencing data from
488 36 newly diagnosed PCNSL patients at Fujian Cancer Hospital and The First Affiliated
489 Hospital of Fujian Medical University were included as an independent external Chinese
490 PCNSL cohort to further validate the molecular clusters[17].

491 The molecular classification scheme, presented in **Figure 4A**, **Table S3** and **Figure S8**,
492 was determined independently of clinical information and finalized before the analysis of
493 clinical data, allowing us to analyze the relationships between genetic subtypes and survival
494 in this entire cohort. Finally, EP300, BTG1, MYD88, and IRF4 were included in the molecular
495 subtyping of Chinese PCNSLs.

All nonsynonymous mutations in EP300, BTG1, MYD88, and IRF4 were visualized within the functional domains of the encoded proteins (**Figure S9**). As previously reported[10,57,58], MYD88 (L265P) was the most prevalent mutation, while EP300, BTG1, and IRF4 all harbored scattered mutations. We evaluated the prognostic significance of these four identified genes for OS and PFS. The survival outcomes of patients with EP300 mutations were significantly poorer than those of patients with wild-type EP300, and patients harboring mutations in either BTG1 (OS, P = 0.017; PFS, P = 0.12), MYD88 (OS, P = 0.016; PFS, P = 0.069), or IRF4 (OS, P = 0.045; PFS, P = 0.018) had favorable OS (**Figure S10A**) and PFS (**Figure S10B**) relative to individuals with their wild-type counterparts.

In the discovery cohort, we identified three subtypes of PCNSL: BMI (n = 16, 27.59%), E3 (n = 9, 15.52%), and UC (n = 33, 56.90%). The E3 subtype was defined as those with mutations in the EP300 gene. The PCNSLs that carried mutations in at least two of the three genes, BTG1, MYD88, and IRF4, were characterized as the BMI subtype. The other PCNSLs that did not fit into either of these categories were classified under the UC subtype (Unclassified). The three subtypes differed significantly in OS ($P < 0.001$) and PFS ($P = 0.043$), the BMI subtype ($P < 0.001$, $P = 0.023$) had much more favorable outcomes than the E3 and UC subtypes did, and the E3 subtype ($P < 0.001$, $P = 0.037$) had far worse outcomes than the BMI and UC subtypes did (**Figure 4B**). These differences between the three subtypes of PCNSL were detected in the validation cohort (OS, $P = 0.020$; PFS, $P = 0.048$; n = 82, tumor-only sample), and patients with E3 had significantly shorter survival times than those with BMI (OS, $P = 0.007$) or UC (OS, $P = 0.043$) (**Figure 4C**). Similar results were observed in the entire cohort (n = 140, OS, $P < 0.0001$; PFS, $P = 0.0004$; **Figure S10C**): 26 patients were classified as BMI (18.57%), 27 as E3 (19.29%), and 87 as UC (62.14%). Furthermore, the raw WES or WGS data, along with follow-up data from an independent cohort of 36 Chinese patients with PCNSL[17], were obtained. The three identified subtypes significantly differed in OS (P = 0.0004; **Figure 4D**). Specifically, the BMI subgroup (P = 0.0004, P = 0.005) had notably more favorable outcomes than the E3 and UC subgroups did. Conversely, the E3 subtype (P = 0.0004, P = 0.0231) had considerably worse outcomes than either the BMI or UC subgroup did. These results further validate the robustness of our

525 molecular classification.

526 BTK inhibitors have been shown to improve the prognosis of PCNSL patients by blocking
527 B-cell receptor signaling pathways, which influence the growth and survival of B cells.
528 Consequently, we further analyzed whether the use of BTK inhibitors affects the robustness
529 of our classification. There were no significant differences in OS ($P > 0.05$, **Figure S11A**) or
530 PFS ($P > 0.05$, **Figure S11B**) between patients who received BTK inhibitors ($n = 8$) and those
531 who did not receive BTK inhibitors ($n = 132$). Additionally, after initial treatment options (HD-
532 MTX combined with idarubicin (IDA), HD-MTX combined with rituximab (R), HD-MTX
533 combined with IDA and R, or a combination of BTK inhibitors) and consolidated therapy
534 (WBRT, stem cell transplant) were adjusted, multivariable Cox regression analysis revealed
535 that the use of different treatment options did not affect the robustness of our classification
536 (**Table S4**).

537 For the E3 subtype, the mean mutation prevalence in the PCNSLs of Asian patients[3,16,17]
538 (16.98%) was significantly higher than that in the PCNSLs of Western patients[4,5,14,59–61]
539 (4.53%) (**Figure 4E**). Patients with the E3 subtype had significantly poorer survival, which
540 was not affected by the mutation site (OS, $P = 0.82$; PFS, $P = 0.69$; **Figure 4F**). Univariate
541 Cox regression analysis revealed that PCNSL molecular subtypes are associated with clinical
542 outcomes (**Table S5**), which was also confirmed by multivariate Cox regression hazard ratio
543 analysis after adjusting for important confounders (age, IELSG, MSKCC, LDH levels, deep
544 brain location, and CSF protein levels) (**Figure 4G)**. These findings further confirmed that
545 PCNSL molecular subtypes are associated with clinical outcomes. The key demographic and
546 clinical characteristics of the three MS patients are summarized in **Table S6**.

547 We also constructed a Sankey diagram to visualize the correspondence between our
548 samples and the previously described genetic subtypes of PCNSL and DLBCL (**Figure S12**).
549 The Chinese PCNSL subtypes were significantly distinct from those previously described.
550 Taken together, these results indicate that three unique molecular subtypes of Chinese
551 PCNSL are associated with clinical outcomes.

552 **Genomic landscape and tumor microenvironment across PCNSL subtypes**

553 We further aimed to determine whether these three molecular subtypes of Chinese PCNSL

554 patients could affect the mutational landscape, oncogenic pathways and tumor
555 microenvironment. Intergroup mutation distribution analysis was performed to identify unique
556 and shared mutations. A total of 6,102 (median variants per sample = 196.5), 41,548 (median
557 variants per sample = 1719), and 73,440 mutated genes (median variants per sample = 918)
558 were identified in the BMI, E3, and UC subtypes, respectively, with 518 mutated genes shared
559 among the three subtypes (**Figures 5A** and **S13A**). Key candidate cancer genes in the 140
560 PCNSL patients, grouped by these three subtypes, are presented in **Figure 5B**. The mutation
561 frequencies of EP300, MYD88, IRF4, FAT4, KMT2C, KMT2A, MPEG1, PRKDC, SPEN,
562 CARD11, RANBP2, NUMA1, NSD1, PASK, USP6, PTPN13, SLX4, ADGRA2, TET1, and
563 NOTCH2 were significantly different (P< 0.05) among these three subtypes. Arm-level CNAs
564 among the samples of the three subtypes are displayed in **Figure 5C**, showing significant
565 differences in recurrent deletions (in 6q and 12p) but not in amplifications (**Figure 5D**). The
566 results of significant recurrent focal amplifications and deletions were visualized by GISTIC
567 2.0 in the BMI, E3, and UC subtypes (**Figure S13B**). There was no significant difference ($P >$
568 0.05) in the MATH (**Figure 5E**) or MSI (**Figure 5F**) scores among the three subtypes, but the
569 TMB was greater in the E3 subgroup than in the BMI ($P < 0.0001$) or UC subgroup ($P < 0.001$)
570 (**Figure 5G**). Furthermore, the VAF was lower in the E3 subgroup than in the BMI ($P < 0.0001$)
571 or UC subgroup ($P < 0.001$) (**Figure 5H**).

572 The contributions of mutational signatures across the three subtypes are illustrated in
573 **Figure S13C**. The three signatures extracted from the E3 subtype samples displayed cosine
574 similarities of 82.0%, 93.8%, and 95.2% to the COSMIC signatures 6, 10a, and 45,
575 respectively (**Figure S13D**). In contrast, those extracted from BMI subtype samples
576 presented similarities of 77.3%, 81.7%, and 58.3% to COSMIC signatures 5, 84, and 87,
577 respectively (**Figure S13D**). The signatures from the UC subtype samples were 80.8%,
578 70.3%, and 91.7% and similar to the COSMIC signatures 5, 1, and 45 (**Figure S13D**),
579 respectively.

580 Hematoxylin–eosin (H&E) and immunohistochemical (IHC) staining of CD2, CD5, CD10,
581 CD19, CD20, CD79a, Ki-67, BCL2, BCL6, MUM1, c-Myc, and p53 from samples of different
582 PCNSL subtypes revealed significant malignant progression in patients with the E3 subtype

583 of PCNSL, as shown in **Figure S14A** and **Table S7**. This is evidenced by the formation of
584 poorly differentiated and aggressively growing tumors (Ki-67), which contrasts with the other
585 two subtypes. Immunostaining revealed a greater presence of CD2+, CD5+, CD10+, CD19+,
586 CD20+, and CD79a+ cells in the E3 subtype. Additionally, a greater presence of BCL2+,
587 BCL6+, MUM1+, and c-Myc+ cells, which are markers associated with the diagnosis of B-cell
588 lymphomas[62], was observed in the E3 subtype.

589 To explore potential therapeutic strategies for the genetic subtypes of PCNSL, we
590 examined groups of genetic aberrations that target oncogenic signaling pathways (**Figure**
591 **S14B**). Genetic events affecting NF-κB regulators that negatively regulate the stability of NF-
592 κB-dependent mRNAs were detected in 85.2% and 84.6% of the E3 and BMI subtype
593 samples but not in the UC subtype samples ($P < 0.05$). These findings suggest that the E3
594 and BMI subtypes may be more responsive to BTK inhibitors[63]. The PI3 kinase pathway, a
595 pathway that can indirectly activate NF-κB, is genetically altered in 66.7% of E3 patients[64].
596 We also noted a higher frequency of genetic alterations in chromatin histone modifiers (100%),
597 RTK signaling (77.8%), protein homeostasis/ubiquitination (66.7%), cell cycle pathways
598 (77.8%), genome integrity (82.5%), Wnt/B-catenin signaling (59.3%), the chromatin SWI/SNF
599 complex (70.4%), and splicing (37%) in E3 patients. Therefore, a combination of cyclin D-
600 Cdk4,6 and PI3 kinase inhibitors might be beneficial for patients with the E3 subtype.

601 Taken together, these results suggest that the three molecular subtypes of PCNSL in
602 Chinese patients each have a unique genomic landscape, tumor microenvironment, and
603 oncogenic pathways.

604 **Mutational landscape and oncogenic pathways in PCNSLs with different sites of onset**

605 Next, we aimed to determine whether the site of PCNSL onset in Chinese patients
606 influences the mutational landscape and oncogenic pathways. **Figure 6A** shows that the
607 areas most often affected by PCNSL were the basal ganglia, brainstem, cerebellum, corpus
608 callosum, frontal lobe, occipital lobe, parietal lobe, temporal lobe, thalamus, and ventricles,
609 as well as combinations of multiple sites. The most frequently involved sites of PCNSL were
610 the frontal lobe (22.14%), temporal lobe (15.71%), and basal ganglia (8.57%).

611 First, we found that there was no significant difference in the proportion of different disease

612 sites among the three subtypes (chi-square test, corpus callosum, P = 0.887; basal ganglia,
613 P=0.645; thalamus, P= 0.317; brainstem, P= 1.000; ventricle, P= 0.631; front lobe, P= 0.576;
614 multiple, P= 0.949; temporal lobe, P= 0.214; cerebellum, P= 0.831; occipital lobe, P= 0.778;
615 parietal lobe, P= 0.887). BMI-related tumors were slightly more prevalent in the temporal lobe
616 (7 out of 26 patients vs. 15 out of 114 patients, chi-square test P = 0.082) than other subtypes
617 were. Conversely, E3 events were likely to be more frequent in the thalamus (2 out of 27
618 cases vs. 2 out of 113 cases, chi-square test P = 0.113). Larger studies are needed to confirm
619 this observation.

620 Next, we explored whether the tumor sites differed between the discovery and validation
621 cohorts. As shown in **Figure S15A**, there was no significant difference in the proportions of
622 different disease sites between the discovery and validation cohorts (chi-square test P =
623 0.610). Furthermore, the proportions of deep brain lesions in the discovery and validation
624 cohorts were not significantly different (P = 0.478).

625 Sixteen mutated genes were represented among all the sites (**Figure 6B**). Key candidate
626 cancer genes in the 140 PCNSL patients, grouped by site of disease onset, are presented in
627 **Figure 6C**. Mutations in the PIM1, CD79B, and MYD88 genes were relatively common across
628 all PCNSL onset sites. The frontal lobe had the highest mutation frequencies of five genes:
629 PIM1 (51.61%), MYD88 (51.61%), and ZFHX3 (48.39%). In the temporal lobe, the most
630 prevalent mutations were in PIM1 (68.18%), SC1. In the basal ganglia, the genes with the
631 highest mutation frequencies were PIM1 (66.67%), MYD88 (50%), ZFHX3 (41.67%), and
632 DTX1 (41.67%). The mutation frequencies of LRP1B and KMT2D were significantly different
633 (P< 0.05) among these sites. Arm-level CNAs among the samples of these sites are displayed
634 in **Figure 6D**, and significant differences in both recurrent amplifications (2q, 10q, and 10p)
635 and deletions (9p) were observed. The difference in the MATH score was significant (P <
636 0.05) (**Figure 6E**) across these sites, but there was no significant difference (P > 0.05) in the
637 TMB (**Figure 6F**) or MSI (**Figure 6G**) score between several sites.

638 Next, we further divided the tumor sites into deep brain and shallow brain tumor subgroups.
639 A total of 854 (81%) mutated genes were represented between the deep brain and shallow
640 brain tumor subgroups (**Figure S15B**). The mutational landscape between deep brain and

641 shallow brain tumors is shown in **Figure S15C**. The mutation frequencies of LRP1B, FAT1,
642 KMT2B, CIC, PTPRB, MAP3KI, ARID1B, and CPS1 in the deep brain and shallow brain
643 tumors were significantly different (P< 0.05). Arm-level CNAs between deep brain and
644 shallow brain tumors are displayed in **Figure S15D**, and no significant differences (P >0.05)
645 in either recurrent amplification or deletion were observed. There was no significant difference
646 (*P* > 0.05) in the MATH (**Figure S15E**), TMB (**Figure S15F**), or MSI (**Figure S15G**) score
647 between deep brain and shallow brain tumors.

648 We further conducted pathway enrichment analysis (**Figure 6H**) and revealed significant
649 involvement of NF-κB signaling in the tumors of patients with onset in the parietal lobe and
650 brainstem, each with an involvement rate of 100%. We also noted a significant difference (P
651 <0.05) in the frequency of genetic alterations in transcription factors, other chromatin, immune
652 signaling, NF-κB signaling, chromatin histone modifiers, RTK signaling, PI3K signaling, RNA
653 abundance, the chromatin SWI/SNF complex, protein homeostasis/ubiquitination, and the
654 cell cycle among these sites. Furthermore, we noted a significant difference (P <0.05) in the
655 frequency of genetic alterations in the chromatin SWI/SNF complex but not in the other
656 pathways (**Figure S15H**). Taken together, these results indicate that the mutational
657 landscape in Chinese patients with PCNSL varies between sites of onset.

658

659 **Discussion**

660 Owing to the genetic, phenotypic, and tumor microenvironment heterogeneity of PCNSL,
661 identifying classification and prognostic biomarkers for patients is extremely challenging,
662 which in turn makes developing effective new therapies challenging. Here, we performed a
663 study of 176 Chinese PCNSL patients that was adequately powered to expand the genomic
664 landscape and explore its clinical significance utilizing an unprecedented sample size and a
665 multicenter approach. We defined recurrent mutations, CNAs, and associated cancer
666 candidate genes in Chinese PCNSL patients and compared these comprehensive genetic
667 signatures with those of PCNSL patients and systemic DLBCL patients of other races,
668 revealing that genetic heterogeneity is a defining feature of PCNSL. Our results highlight the
669 complexity of PCNSLs, which have a median of 19.06 mutations/Mb and a median of 949.5

variants per sample.

Although the genetic variation profiles of the PCNSL discovery and validation samples were similar, notable differences were evident in the mutation frequencies of certain genes (e.g., NOTCH1, NOTCH2, and KMT2C), the TMB, the MSI, and the MATH score. These discrepancies might stem from variations in specimen types (fresh tissue versus FFPE) and the absence of paired samples in the validation cohort[65]. Furthermore, genetic heterogeneity between patient populations[15] and biological differences between PCNSL and DLBCL[13] may be the main reasons why the published classifications for DLBCL and PCNSL are not applicable to Chinese PCNSL patients. Currently, the classification frameworks for systemic DLBCL offer valuable insights into the underlying biology based on molecular mechanisms. However, it is important to note that biological insights do not necessarily correlate with clinical outcomes, especially given that treatment paradigms may evolve over time.

More importantly, in our study, we identified three robust molecular subtypes of PCNSL in Chinese patients: BMI, E3, and UC. Distinct clinical outcomes, activated cellular pathways, and alterations in the genetic landscape distinguish the three subtypes and could be utilized to personalize therapy for patients with different subtypes. For the E3 subtype, the mean mutation rate was significantly higher in the PCNSLs of Chinese and Japanese patients (16.88%) than in those of Western PCNSL patients (4.53%). This difference may be attributed to variations in age, lifestyle, and EBV infection rates[17,41,66]. Patients with the E3 subtype had significantly poorer survival outcomes, a finding that is consistent with reported findings in DLBCL[67]. EP300 and CREBBP are two closely related members of the KAT3 family of histone acetyltransferases that function similarly[68]. However, in this study, the survival patients with EP300 mutations was significantly poorer, whereas CREBBP mutations did not contribute to an inferior prognosis (**Figure S4**). This discrepancy might stem from the distinct roles of EP300 and CREBBP in PCNSL. Specifically, in a recent study, EP300, but not CREBBP, was reported to play an essential role in supporting the viability of classical Hodgkin's lymphoma by directly modulating the expression of the oncogenic MYC/IRF4 network, surface receptor CD30, immunoregulatory cytokine interleukin 10, and immune checkpoint protein PD-L1[69]. The molecular mechanisms through which EP300 mutations facilitate the progression of

PCNSL are still not understood. We plan to explore these mechanisms in a forthcoming study.

PCNSLs that carried mutations in at least two of three genes, namely, BTG1, MYD88, and IRF4, were characterized as the BMI subtype with favorable survival outcomes. The prognostic significance of MYD88 mutations in PCNSL patients remains inconclusive. While several studies have reported no effect on OS[70–73], only two have reported an unfavorable outcome[74,75]. In line with our findings, *Olimpia* et al.[76] and *Claudio* et al.[77] have indicated that the survival time is significantly longer in patients with PCNSLs harboring MYD88 mutations. This discrepancy may be attributed to heterogeneity in genetics, phenotype, and the tumor microenvironment across various patient populations; variances in sample size; and the adjustment of confounding variables, including age and treatment methods. In this large study, we found that Chinese PCNSL patients harboring MYD88 mutations have longer survival times. However, further research is needed to verify these findings in diverse populations. The associations of IRF4 and BTG1 mutations with clinical outcomes in PCNSL patients have not been previously reported. This results of this study revealed that mutations in both IRF4 and BTG1 were associated with favorable outcomes. IRF4 has been identified as a driver oncogene that transcriptionally regulates downstream target genes, such as MYC, and coordinates the transcriptional program with NF-κB references[78–81]. Numerous in vitro and clinical studies have highlighted the abnormal overexpression and oncogenic roles of IRF4 in various mature lymphoid neoplasms[81–84]. Therefore, we hypothesized that the reduced expression and/or decreased transcriptional activity of IRF4 resulting from such mutations may explain the association of IRF4 mutations with favorable survival outcomes in PCNSL patients. BTG1 mutation has been associated with extranodal dissemination and unfavorable outcomes in B-cell lymphoma patients[85–87]. However, in this study, BTG1 mutation was linked to favorable outcomes in PCNSL patients. In a substantial discovery cohort consisting of 533 patients with B-cell precursor acute lymphoblastic leukemia, *Blanca* and colleagues[88] reported that deletions of BTG1 alone did not affect prognosis. The complete remission rate among acute myeloid leukemia patients with low BTG1 expression was significantly greater than that among patients with high BTG1 expression[89]. The molecular mechanisms through which mutations in BTG1 decelerate the progression of PCNSL warrant further research.

This study has several limitations. First, the proportion of unclassified PCNSLs (UC subtypes) is high, possibly due to the lack of multiomics data needed for classification. However, the three-class classification based on WES proposed in this study has strong clinical practicality and translational value, as it relies solely on the four genes identified through WES. In other words, clinicians can predict patient outcomes and further perform individualized management of PCNSLs by detecting mutations in these four genes without the need for additional multiomics testing. Second, the three identified molecular subtypes can be used to predict patient outcomes. However, it remains challenging to determine whether these three molecular subtypes can still be used to accurately distinguish biological subtypes without additional RNA-seq data layers.

In summary, by performing genomic sequencing on tumor specimens from 176 Chinese PCNSL patients, we identified three molecular subtypes that can be used to predict patient outcomes. These genetic signatures associated with PCNSL outcomes illuminate the path toward individualized management of PCNSL patients and the discovery of novel treatment strategies for this challenging disease.

Data and code availability

The raw sequence data have been deposited in the Genome Sequence Archive[90] in the National Genomics Data Center[91], China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human HRA006122), which is publicly accessible at https://ngdc.cncb.ac.cn/gsa-human. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

The raw sequencing data of the DLBCL WES data are accessible from the TCGA database and were reanalyzed via pipelines and filter settings that were identical to those used in the present study. The raw WES data of the Japanese PCNSL patients were obtained from the Japanese Genotype–Phenotype Archive (JGA, http://trace.ddbj.nig.ac.jp/jga) and reanalyzed via pipelines and filtering settings that were identical to those used in the present study, which is hosted by DDBJ under the accession number JGAS00000000021[3]. The raw sequencing WES/WGS data of an external independent Chinese PCNSL cohort (n = 36)[17] were obtained

from the China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences and reanalyzed via pipelines and filtering settings that were identical to those used in the present study, under the accession number GSA-Human HRA002475.

The frequencies of recurrent genetic alterations in DLBCLs[49] and French PCNSLs[14] are accessible from the manuscript or supplemental information and were not reanalyzed via identical pipelines and filtering settings as those used in the present study.

Acknowledgment: This study was funded by the National Natural Science Foundation of China (82302582), the Shanghai Municipal Health Commission Project (20224Y0317), the Shanghai Science and Technology Commission (17430750200), the Join Breakthrough Project for New Frontier Technologies of Shanghai Hospital Development Center (SHDC12016120), and the Youth Medical Talents–Clinical Laboratory Practitioner Program (2022-65).

Competing Interests: The authors declare no potential conflicts of interest.

Author contributions: SJ. L, CX. L, DH. L, WJ. C, ZG. X, XT. Z, and Y. M conceived and designed the project. SJ. L, DH. L, WJ. C, JZ. C, J. R, and ZG. X collected the clinical samples. SJ. L, JN. W, JZ. C, LY. Z, HW. Y, and Y. S analyzed the WES data and performed bioinformatic analyses. SJ. L, DH. L, J. R, LY. Z, HW. Y, and JN. W integrated the sequencing data, drew the display items. SJ. L, DH. L, and CX. L wrote the manuscript. XT. Z and Y. M oversaw the ethical guidelines and data regulation. WJ. C and Y. M supervised the project. All of the authors contributed to the final version of the paper.

References

1. Chen, T., Liu, Y., Wang, Y., Chang, Q., Wu, J., Wang, Z., Geng, D., Yu, J.-T., Li, Y., Li, X.-Q., et al. (2022). Evidence-based expert consensus on the management of primary central nervous system lymphoma in China. J Hematol Oncol *15*, 136. https://doi.org/10.1186/s13045-022-01356-7.

2. Ferreri, A.J.M., Calimeri, T., Cwynarski, K., Dietrich, J., Grommes, C., Hoang-Xuan, K., Hu, L.S., Illerhaus, G., Nayak, L., Ponzoni, M., et al. (2023). Primary central nervous system lymphoma. Nat. Rev. Dis. Primers *9*, 29. https://doi.org/10.1038/s41572-023-00439-0.

3. Fukumura, K., Kawazu, M., Kojima, S., Ueno, T., Sai, E., Soda, M., Ueda, H., Yasuda, T., Yamaguchi, H., Lee, J., et al. (2016). Genomic characterization of primary central nervous system lymphoma. Acta Neuropathol. (Berl.) *131*, 865–875. https://doi.org/10.1007/s00401-016-1536-2.

4. Braggio, E., Van Wier, S., Ojha, J., McPhail, E., Asmann, Y.W., Egan, J., Da Silva, J.A., Schiff, D., Lopes, M.B., Decker, P.A., et al. (2015). Genome-Wide Analysis Uncovers Novel Recurrent Alterations in Primary Central Nervous System Lymphomas. Clin. Cancer Res. *21*, 3986–3994. https://doi.org/10.1158/1078-0432.CCR-14-2116.

5. Radke, J., Ishaque, N., Koll, R., Gu, Z., Schumann, E., Sieverling, L., Uhrig, S., Hübschmann, D., Toprak, U.H., López, C., et al. (2022). The genomic and transcriptional landscape of primary central nervous system lymphoma. Nat. Commun. *13*, 2558. https://doi.org/10.1038/s41467-022-30050-y.

6. Vater, I., Montesinos-Rongen, M., Schlesner, M., Haake, A., Purschke, F., Sprute, R., Mettenmeyer, N., Nazzal, I., Nagel, I., Gutwein, J., et al. (2015). The mutational pattern of primary lymphoma of the central nervous system determined by whole-exome sequencing. Leukemia *29*, 677–685. https://doi.org/10.1038/leu.2014.264.

7. Rosenwald, A., Wright, G., Chan, W.C., Connors, J.M., Campo, E., Fisher, R.I., Gascoyne, R.D., Muller-Hermelink, H.K., Smeland, E.B., Giltnane, J.M., et al. (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N. Engl. J. Med. *346*, 1937–1947. https://doi.org/10.1056/NEJMoa012914.

8. Wilson, W.H., Young, R.M., Schmitz, R., Yang, Y., Pittaluga, S., Wright, G., Lih, C.-J., Williams, P.M., Shaffer, A.L., Gerecitano, J., et al. (2015). Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. Nat. Med. *21*, 922–926. https://doi.org/10.1038/nm.3884.

9. Schmitz, R., Wright, G.W., Huang, D.W., Johnson, C.A., Phelan, J.D., Wang, J.Q., Roulland, S., Kasbekar, M., Young, R.M., Shaffer, A.L., et al. (2018). Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. N. Engl. J. Med. *378*, 1396–1407. https://doi.org/10.1056/NEJMoa1801445.

10. Chapuy, B., Stewart, C., Dunford, A.J., Kim, J., Kamburov, A., Redd, R.A., Lawrence, M.S., Roemer, M.G.M., Li, A.J., Ziepert, M., et al. (2018). Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat. Med. *24*, 679–690. https://doi.org/10.1038/s41591-018-0016-8.

11. Wright, G.W., Huang, D.W., Phelan, J.D., Coulibaly, Z.A., Roulland, S., Young, R.M., Wang, J.Q., Schmitz, R., Morin, R.D., Tang, J., et al. (2020). A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. Cancer Cell *37*, 551-568.e14. https://doi.org/10.1016/j.ccell.2020.03.015.

12. Shen, R., Fu, D., Dong, L., Zhang, M.-C., Shi, Q., Shi, Z.-Y., Cheng, S., Wang, L., Xu, P.-P., and Zhao, W.-L. (2023). Simplified algorithm for genetic subtyping in diffuse large B-cell lymphoma. Signal Transduct. Target. Ther. *8*, 145. https://doi.org/10.1038/s41392-023-01358-y.

13. Swerdlow, S.H., Campo, E., Pileri, S.A., Harris, N.L., Stein, H., Siebert, R., Advani, R., Ghielmini, M., Salles,

824 G.A., Zelenetz, A.D., et al. (2016). The 2016 revision of the World Health Organization classification of
825 lymphoid neoplasms. Blood *127*, 2375–2390. https://doi.org/10.1182/blood-2016-01-643569.

826 14. Hernández-Verdin, I., Kirasic, E., Wienand, K., Mokhtari, K., Eimer, S., Loiseau, H., Rousseau, A., Paillassa,
827 J., Ahle, G., Lerintiu, F., et al. (2023). Molecular and clinical diversity in primary central nervous system
828 lymphoma. Ann. Oncol. *34*, 186–199. https://doi.org/10.1016/j.annonc.2022.11.002.

829 15. Yuan, X., Yu, T., Zhao, J., Jiang, H., Hao, Y., Lei, W., Liang, Y., Li, B., and Qian, W. (2023). Analysis of the
830 genomic landscape of primary central nervous system lymphoma using whole-genome sequencing in
831 Chinese patients. Front. Med. *17*, 889–906. https://doi.org/10.1007/s11684-023-0994-x.

832 16. He, X., Fan, X., Shan, Y., Ji, X., Su, L., and Wang, Y. (2023). Analysis of genomic alterations in primary central
833 nervous system lymphoma. Medicine (Baltimore) *102*, e34931.
834 https://doi.org/10.1097/MD.0000000000034931.

835 17. Zhu, Q., Wang, J., Zhang, W., Zhu, W., Wu, Z., Chen, Y., Chen, M., Zheng, L., Tang, J., Zhang, S., et al. (2022).
836 Whole-Genome/Exome Sequencing Uncovers Mutations and Copy Number Variations in Primary Diffuse
837 Large B-Cell Lymphoma of the Central Nervous System. Front Genet *13*, 878618.
838 https://doi.org/10.3389/fgene.2022.878618.

839 18. Abrey, L.E., Batchelor, T.T., Ferreri, A.J.M., Gospodarowicz, M., Pulczynski, E.J., Zucca, E., Smith, J.R., Korfel,
840 A., Soussain, C., DeAngelis, L.M., et al. (2005). Report of an international workshop to standardize baseline
841 evaluation and response criteria for primary CNS lymphoma. J Clin Oncol *23*, 5034–5043.
842 https://doi.org/10.1200/JCO.2005.13.524.

843 19. Hoang-Xuan, K., Bessell, E., Bromberg, J., Hottinger, A.F., Preusser, M., Rudà, R., Schlegel, U., Siegal, T.,
844 Soussain, C., Abacioglu, U., et al. (2015). Diagnosis and treatment of primary CNS lymphoma in
845 immunocompetent patients: guidelines from the European Association for Neuro-Oncology. Lancet
846 Oncol *16*, e322-332. https://doi.org/10.1016/S1470-2045(15)00076-5.

847 20. Li, Q., Ma, J., Ma, Y., Lin, Z., Kang, H., and Chen, B. (2021). Improvement of outcomes of an escalated high-
848 dose methotrexate-based regimen for patients with newly diagnosed primary central nervous system
849 lymphoma: a real-world cohort study. Cancer Manag Res *13*, 6115–6122.
850 https://doi.org/10.2147/CMAR.S322467.

851 21. Freed, D., Aldana, R., Weber, J.A., and Edwards, J.S. (2017). The Sentieon Genomics Tools–A fast and
852 accurate solution to variant calling from next-generation sequence data. BioRxiv, 115717.

853 22. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F.
854 (2016). The Ensembl Variant Effect Predictor. Genome Biol *17*, 122. https://doi.org/10.1186/s13059-016-
855 0974-4.

856 23. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and
857 comprehensive analysis of somatic variants in cancer. Genome Res *28*, 1747–1756.
858 https://doi.org/10.1101/gr.239244.118.

859 24. Wang, Y., Bae, T., Thorpe, J., Sherman, M.A., Jones, A.G., Cho, S., Daily, K., Dou, Y., Ganz, J., Galor, A., et al.

(2021). Comprehensive identification of somatic nucleotide variants in human brain tissue. Genome Biol *22*, 92. https://doi.org/10.1186/s13059-021-02285-3.

25. Sathirapongsasuti, J.F., Lee, H., Horst, B.A.J., Brunner, G., Cochran, A.J., Binder, S., Quackenbush, J., and Nelson, S.F. (2011). Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinformatics *27*, 2648–2654. https://doi.org/10.1093/bioinformatics/btr462.

26. Favero, F., Joshi, T., Marquard, A.M., Birkbak, N.J., Krzystanek, M., Li, Q., Szallasi, Z., and Eklund, A.C. (2015). Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Ann Oncol *26*, 64–70. https://doi.org/10.1093/annonc/mdu479.

27. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol *12*, R41. https://doi.org/10.1186/gb-2011-12-4-r41.

28. Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol *2017*, PO.17.00011. https://doi.org/10.1200/PO.17.00011.

29. Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558. https://doi.org/10.1126/science.1235122.

30. Griffith, M., Spies, N.C., Krysiak, K., McMichael, J.F., Coffman, A.C., Danos, A.M., Ainscough, B.J., Ramirez, C.A., Rieke, D.T., Kujan, L., et al. (2017). CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet *49*, 170–174. https://doi.org/10.1038/ng.3774.

31. Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature *578*, 94–101. https://doi.org/10.1038/s41586-020-1943-3.

32. Chalmers, Z.R., Connelly, C.F., Fabrizio, D., Gay, L., Ali, S.M., Ennis, R., Schrock, A., Campbell, B., Shlien, A., Chmielecki, J., et al. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med *9*, 34. https://doi.org/10.1186/s13073-017-0424-2.

33. Mroz, E.A., and Rocco, J.W. (2013). MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. Oral Oncol *49*, 211–215. https://doi.org/10.1016/j.oraloncology.2012.09.007.

34. Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadoy, S., Liu, D.L., Kantheti, H.S., Saghafinia, S., et al. (2018). Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell *173*, 321-337.e10. https://doi.org/10.1016/j.cell.2018.03.035.

35. Garcia, M., Juhos, S., Larsson, M., Olason, P.I., Martin, M., Eisfeldt, J., DiLorenzo, S., Sandgren, J., Díaz De Ståhl, T., Ewels, P., et al. (2020). Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. F1000Res *9*, 63. https://doi.org/10.12688/f1000research.16665.2.

36. Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and

895 Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nat
896 Biotechnol *38*, 276–278. https://doi.org/10.1038/s41587-020-0439-x.

897 37. Grüning, B., Dale, R., Sjödin, A., Chapman, B.A., Rowe, J., Tomkins-Tinch, C.H., Valieris, R., Köster, J., and
898 Bioconda Team (2018). Bioconda: sustainable and comprehensive software distribution for the life
899 sciences. Nat Methods *15*, 475–476. https://doi.org/10.1038/s41592-018-0046-7.

900 38. da Veiga Leprevost, F., Grüning, B.A., Alves Aflitos, S., Röst, H.L., Uszkoreit, J., Barsnes, H., Vaudel, M.,
901 Moreno, P., Gatto, L., Weber, J., et al. (2017). BioContainers: an open-source and community-driven
902 framework for software standardization. Bioinformatics *33*, 2580–2582.
903 https://doi.org/10.1093/bioinformatics/btx192.

904 39. P, D.T., M, C., Ew, F., Pp, B., E, P., and C, N. (2017). Nextflow enables reproducible computational workflows.
905 Nature biotechnology *35*. https://doi.org/10.1038/nbt.3820.

906 40. Gandhi, M.K., Hoang, T., Law, S.C., Brosda, S., O'Rourke, K., Tobin, J.W.D., Vari, F., Murigneux, V., Fink, L.,
907 Gunawardana, J., et al. (2021). EBV-associated primary CNS lymphoma occurring after
908 immunosuppression is a distinct immunobiological entity. Blood *137*, 1468–1477.
909 https://doi.org/10.1182/blood.2020008520.

910 41. Kaulen, L.D., Denisova, E., Hinz, F., Hai, L., Friedel, D., Henegariu, O., Hoffmann, D.C., Ito, J., Kourtesakis,
911 A., Lehnert, P., et al. (2023). Integrated genetic analyses of immunodeficiency-associated Epstein-Barr
912 virus- (EBV) positive primary CNS lymphomas. Acta Neuropathol *146*, 499–514.
913 https://doi.org/10.1007/s00401-023-02613-w.

914 42. Ren, W., Ye, X., Su, H., Li, W., Liu, D., Pirmoradian, M., Wang, X., Zhang, B., Zhang, Q., Chen, L., et al. (2018).
915 Genetic landscape of hepatitis B virus-associated diffuse large B-cell lymphoma. Blood *131*, 2670–2681.
916 https://doi.org/10.1182/blood-2017-11-817601.

917 43. Berhan, A., Bayleyegn, B., and Getaneh, Z. (2022). HIV/AIDS Associated Lymphoma: Review. Blood
918 Lymphat Cancer *12*, 31–45. https://doi.org/10.2147/BLCTT.S361320.

919 44. Pasqualucci, L., Trifonov, V., Fabbri, G., Ma, J., Rossi, D., Chiarenza, A., Wells, V.A., Grunn, A., Messina, M.,
920 Elliot, O., et al. (2011). Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet *43*,
921 830–837. https://doi.org/10.1038/ng.892.

922 45. Morin, R.D., Mendez-Lago, M., Mungall, A.J., Goya, R., Mungall, K.L., Corbett, R.D., Johnson, N.A., Severson,
923 T.M., Chiu, R., Field, M., et al. (2011). Frequent mutation of histone-modifying genes in non-Hodgkin
924 lymphoma. Nature *476*, 298–303. https://doi.org/10.1038/nature10351.

925 46. Lohr, J.G., Stojanov, P., Lawrence, M.S., Auclair, D., Chapuy, B., Sougnez, C., Cruz-Gordillo, P., Knoechel,
926 B., Asmann, Y.W., Slager, S.L., et al. (2012). Discovery and prioritization of somatic mutations in diffuse
927 large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci U S A *109*, 3879–3884.
928 https://doi.org/10.1073/pnas.1121343109.

929 47. Morin, R.D., Mungall, K., Pleasance, E., Mungall, A.J., Goya, R., Huff, R.D., Scott, D.W., Ding, J., Roth, A.,
930 Chiu, R., et al. (2013). Mutational and structural analysis of diffuse large B-cell lymphoma using whole-

931 genome sequencing. Blood *122*, 1256–1265. https://doi.org/10.1182/blood-2013-02-483727.

932 48. de Miranda, N.F.C.C., Georgiou, K., Chen, L., Wu, C., Gao, Z., Zaravinos, A., Lisboa, S., Enblad, G., Teixeira,
933 M.R., Zeng, Y., et al. (2014). Exome sequencing reveals novel mutation targets in diffuse large B-cell
934 lymphomas derived from Chinese patients. Blood *124*, 2544–2553. https://doi.org/10.1182/blood-2013-
935 12-546309.

936 49. Reddy, A., Zhang, J., Davis, N.S., Moffitt, A.B., Love, C.L., Waldrop, A., Leppa, S., Pasanen, A., Meriranta, L.,
937 Karjalainen-Lindsberg, M.-L., et al. (2017). Genetic and Functional Drivers of Diffuse Large B Cell
938 Lymphoma. Cell *171*, 481-494.e15. https://doi.org/10.1016/j.cell.2017.09.027.

939 50. Bruno, A., Boisselier, B., Labreche, K., Marie, Y., Polivka, M., Jouvet, A., Adam, C., Figarella-Branger, D.,
940 Miquel, C., Eimer, S., et al. (2014). Mutational analysis of primary central nervous system lymphoma.
941 Oncotarget *5*, 5065–5075. https://doi.org/10.18632/oncotarget.2080.

942 51. Chapuy, B., Roemer, M.G.M., Stewart, C., Tan, Y., Abo, R.P., Zhang, L., Dunford, A.J., Meredith, D.M.,
943 Thorner, A.R., Jordanova, E.S., et al. (2016). Targetable genetic features of primary testicular and primary
944 central nervous system lymphomas. Blood *127*, 869–881. https://doi.org/10.1182/blood-2015-10-
945 673236.

946 52. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli,
947 N., Borg, A., Børresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer.
948 Nature *500*, 415–421. https://doi.org/10.1038/nature12477.

949 53. Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J., Lee, W., Yuan, J., Wong, P.,
950 Ho, T.S., et al. (2015). Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade
951 in non-small cell lung cancer. Science *348*, 124–128. https://doi.org/10.1126/science.aaa1348.

952 54. Yamamoto, H., Watanabe, Y., Arai, H., Umemoto, K., Tateishi, K., and Sunakawa, Y. (2024). Microsatellite
953 instability: A 2024 update. Cancer Sci *115*, 1738–1748. https://doi.org/10.1111/cas.16160.

954 55. Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T.,
955 Yu, X., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.
956 Nature *403*, 503–511. https://doi.org/10.1038/35000501.

957 56. Hans, C.P., Weisenburger, D.D., Greiner, T.C., Gascoyne, R.D., Delabie, J., Ott, G., Müller-Hermelink, H.K.,
958 Campo, E., Braziel, R.M., Jaffe, E.S., et al. (2004). Confirmation of the molecular classification of diffuse
959 large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood *103*, 275–282.
960 https://doi.org/10.1182/blood-2003-05-1545.

961 57. Phelan, J.D., Young, R.M., Webster, D.E., Roulland, S., Wright, G.W., Kasbekar, M., Shaffer, A.L., Ceribelli,
962 M., Wang, J.Q., Schmitz, R., et al. (2018). A multiprotein supercomplex controlling oncogenic signalling in
963 lymphoma. Nature *560*, 387–391. https://doi.org/10.1038/s41586-018-0290-0.

964 58. Ngo, V.N., Young, R.M., Schmitz, R., Jhavar, S., Xiao, W., Lim, K.-H., Kohlhammer, H., Xu, W., Yang, Y., Zhao,
965 H., et al. (2011). Oncogenically active MYD88 mutations in human lymphoma. Nature *470*, 115–119.
966 https://doi.org/10.1038/nature09671.

967 59. Geng, H., Mo, S.S., Chen, L., Ballapuram, A., Tsang, M., Lu, M., Rauschecker, A.M., Wen, K.W., Devine, W.P.,
968 Solomon, D.A., et al. (2024). Identification of Genomic Biomarkers of Disease Progression and Survival in
969 Primary CNS Lymphoma. Blood Adv, bloodadvances.2024014460.
970 https://doi.org/10.1182/bloodadvances.2024014460.

971 60. Zorofchian, S., El-Achi, H., Yan, Y., Esquenazi, Y., and Ballester, L.Y. (2018). Characterization of genomic
972 alterations in primary central nervous system lymphomas. J Neurooncol *140*, 509–517.
973 https://doi.org/10.1007/s11060-018-2990-6.

974 61. Barakat, M., Albitar, M., Whitney, R., and Abdulhaq, H. (2021). Diversity of genetic alterations of primary
975 central nervous system lymphoma in Hispanic versus non-Hispanic patients. Cancer Treat Res Commun
976 *27*, 100310. https://doi.org/10.1016/j.ctarc.2021.100310.

977 62. van Imhoff, G.W., Boerma, E.-J.G., van der Holt, B., Schuuring, E., Verdonck, L.F., Kluin-Nelemans, H.C.,
978 and Kluin, P.M. (2006). Prognostic impact of germinal center-associated proteins and chromosomal
979 breakpoints in poor-risk diffuse large B-cell lymphoma. J Clin Oncol *24*, 4135–4142.
980 https://doi.org/10.1200/JCO.2006.05.5897.

981 63. Iwasaki, H., Takeuchi, O., Teraguchi, S., Matsushita, K., Uehata, T., Kuniyoshi, K., Satoh, T., Saitoh, T.,
982 Matsushita, M., Standley, D.M., et al. (2011). The IκB kinase complex regulates the stability of cytokine-
983 encoding mRNA induced by TLR-IL-1R by controlling degradation of regnase-1. Nat Immunol *12*, 1167–
984 1175. https://doi.org/10.1038/ni.2137.

985 64. Kloo, B., Nagel, D., Pfeifer, M., Grau, M., Düwel, M., Vincendeau, M., Dörken, B., Lenz, P., Lenz, G., and
986 Krappmann, D. (2011). Critical role of PI3K signaling for NF-kappaB-dependent survival in a subset of
987 activated B-cell-like diffuse large B-cell lymphoma cells. Proc Natl Acad Sci U S A *108*, 272–277.
988 https://doi.org/10.1073/pnas.1008969108.

989 65. Spencer, D.H., Sehn, J.K., Abel, H.J., Watson, M.A., Pfeifer, J.D., and Duncavage, E.J. (2013). Comparison of
990 clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens.
991 J Mol Diagn *15*, 623–633. https://doi.org/10.1016/j.jmoldx.2013.05.004.

992 66. Yokoyama, A., Kakiuchi, N., Yoshizato, T., Nannya, Y., Suzuki, H., Takeuchi, Y., Shiozawa, Y., Sato, Y., Aoki,
993 K., Kim, S.K., et al. (2019). Age-related remodelling of oesophageal epithelia by mutated cancer drivers.
994 Nature *565*, 312–317. https://doi.org/10.1038/s41586-018-0811-x.

995 67. Huang, Y.-H., Cai, K., Xu, P.-P., Wang, L., Huang, C.-X., Fang, Y., Cheng, S., Sun, X.-J., Liu, F., Huang, J.-Y.,
996 et al. (2021). CREBBP/EP300 mutations promoted tumor progression in diffuse large B-cell lymphoma
997 through altering tumor-associated macrophage polarization via FBXW7-NOTCH-CCL2/CSF1 axis. Signal
998 Transduct Target Ther *6*, 10. https://doi.org/10.1038/s41392-020-00437-8.

999 68. Meyer, S.N., Scuoppo, C., Vlasevska, S., Bal, E., Holmes, A.B., Holloman, M., Garcia-Ibanez, L., Nataraj, S.,
1000 Duval, R., Vantrimpont, T., et al. (2019). Unique and Shared Epigenetic Programs of the CREBBP and EP300
1001 Acetyltransferases in Germinal Center B Cells Reveal Targetable Dependencies in Lymphoma. Immunity
1002 *51*, 535-547.e9. https://doi.org/10.1016/j.immuni.2019.08.006.

1003 69. Wei, W., Song, Z., Chiba, M., Wu, W., Jeong, S., Zhang, J.-P., Kadin, M.E., Nakagawa, M., and Yang, Y.
1004 (2023). Analysis and therapeutic targeting of the EP300 and CREBBP acetyltransferases in anaplastic large
1005 cell lymphoma and Hodgkin lymphoma. Leukemia *37*, 396–407. https://doi.org/10.1038/s41375-022-
1006 01774-z.

1007 70. Nayyar, N., White, M.D., Gill, C.M., Lastrapes, M., Bertalan, M., Kaplan, A., D'Andrea, M.R., Bihun, I., Kaneb,
1008 A., Dietrich, J., et al. (2019). MYD88 L265P mutation and CDKN2A loss are early mutational events in
1009 primary central nervous system diffuse large B-cell lymphomas. Blood Adv *3*, 375–383.
1010 https://doi.org/10.1182/bloodadvances.2018027672.

1011 71. Zhou, Y., Liu, W., Xu, Z., Zhu, H., Xiao, D., Su, W., Zeng, R., Feng, Y., Duan, Y., Zhou, J., et al. (2018). Analysis
1012 of Genomic Alteration in Primary Central Nervous System Lymphoma and the Expression of Some Related
1013 Genes. Neoplasia *20*, 1059–1069. https://doi.org/10.1016/j.neo.2018.08.012.

1014 72. Zheng, M., Perry, A.M., Bierman, P., Loberiza, F., Nasr, M.R., Szwajcer, D., Del Bigio, M.R., Smith, L.M.,
1015 Zhang, W., and Greiner, T.C. (2017). Frequency of MYD88 and CD79B mutations, and MGMT methylation
1016 in primary central nervous system diffuse large B-cell lymphoma. Neuropathology *37*, 509–516.
1017 https://doi.org/10.1111/neup.12405.

1018 73. Yamada, S., Ishida, Y., Matsuno, A., and Yamazaki, K. (2015). Primary diffuse large B-cell lymphomas of
1019 central nervous system exhibit remarkably high prevalence of oncogenic MYD88 and CD79B mutations.
1020 Leuk Lymphoma *56*, 2141–2145. https://doi.org/10.3109/10428194.2014.979413.

1021 74. Takano, S., Hattori, K., Ishikawa, E., Narita, Y., Iwadate, Y., Yamaguchi, F., Nagane, M., Akimoto, J., Oka, H.,
1022 Tanaka, S., et al. (2018). MyD88 Mutation in Elderly Predicts Poor Prognosis in Primary Central Nervous
1023 System Lymphoma: Multi-Institutional Analysis. World Neurosurg *112*, e69–e73.
1024 https://doi.org/10.1016/j.wneu.2017.12.028.

1025 75. Hattori, K., Sakata-Yanagimoto, M., Okoshi, Y., Goshima, Y., Yanagimoto, S., Nakamoto-Matsubara, R.,
1026 Sato, T., Noguchi, M., Takano, S., Ishikawa, E., et al. (2017). MYD88 (L265P) mutation is associated with an
1027 unfavourable outcome of primary central nervous system lymphoma. Br J Haematol *177*, 492–494.
1028 https://doi.org/10.1111/bjh.14080.

1029 76. Curran, O.E., Poon, M.T.C., Gilroy, L., Torgersen, A., Smith, C., and Al-Qsous, W. (2021). MYD88 L265P
1030 mutation in primary central nervous system lymphoma is associated with better survival: A single-center
1031 experience. Neurooncol Adv *3*, vdab090. https://doi.org/10.1093/noajnl/vdab090.

1032 77. Agostinelli, C., Morandi, L., Righi, S., Cirillo, L., Iommi, M., Tonon, C., Mazzatenta, D., Zoli, M., Rossi, M.,
1033 Bagnato, G., et al. (2023). Genomic Profiling of Primary Diffuse Large B-Cell Lymphoma of the Central
1034 Nervous System Suggests Novel Potential Therapeutic Targets. Mod Pathol *36*, 100323.
1035 https://doi.org/10.1016/j.modpat.2023.100323.

1036 78. Wong, R.W.J., Tan, T.K., Amanda, S., Ngoc, P.C.T., Leong, W.Z., Tan, S.H., Asamitsu, K., Hibi, Y., Ueda, R.,
1037 Okamoto, T., et al. (2020). Feed-forward regulatory loop driven by IRF4 and NF-κB in adult T-cell
1038 leukemia/lymphoma. Blood *135*, 934–947. https://doi.org/10.1182/blood.2019002639.

1039 79. Tsuboi, K., Iida, S., Inagaki, H., Kato, M., Hayami, Y., Hanamura, I., Miura, K., Harada, S., Kikuchi, M., Komatsu,
1040 H., et al. (2000). MUM1/IRF4 expression as a frequent event in mature lymphoid malignancies. Leukemia
1041 *14*, 449–456. https://doi.org/10.1038/sj.leu.2401696.

1042 80. Nakagawa, M., Shaffer, A.L., Ceribelli, M., Zhang, M., Wright, G., Huang, D.W., Xiao, W., Powell, J., Petrus,
1043 M.N., Yang, Y., et al. (2018). Targeting the HTLV-I-Regulated BATF3/IRF4 Transcriptional Network in Adult
1044 T Cell Leukemia/Lymphoma. Cancer Cell *34*, 286-297.e10. https://doi.org/10.1016/j.ccell.2018.06.014.

1045 81. Weilemann, A., Grau, M., Erdmann, T., Merkel, O., Sobhiafshar, U., Anagnostopoulos, I., Hummel, M.,
1046 Siegert, A., Hayford, C., Madle, H., et al. (2015). Essential role of IRF4 and MYC signaling for survival of
1047 anaplastic large cell lymphoma. Blood *125*, 124–132. https://doi.org/10.1182/blood-2014-08-594507.

1048 82. Shaffer, A.L., Emre, N.C.T., Lamy, L., Ngo, V.N., Wright, G., Xiao, W., Powell, J., Dave, S., Yu, X., Zhao, H., et
1049 al. (2008). IRF4 addiction in multiple myeloma. Nature *454*, 226–231. https://doi.org/10.1038/nature07064.

1050 83. Amanda, S., Tan, T.K., Ong, J.Z.L., Theardy, M.S., Wong, R.W.J., Huang, X.Z., Ali, M.Z., Li, Y., Gong, Z., Inagaki,
1051 H., et al. (2022). IRF4 drives clonal evolution and lineage choice in a zebrafish model of T-cell lymphoma.
1052 Nat Commun *13*, 2420. https://doi.org/10.1038/s41467-022-30053-9.

1053 84. Yang, Y., Shaffer, A.L., Emre, N.C.T., Ceribelli, M., Zhang, M., Wright, G., Xiao, W., Powell, J., Platig, J.,
1054 Kohlhammer, H., et al. (2012). Exploiting synthetic lethality for the therapy of ABC diffuse large B cell
1055 lymphoma. Cancer Cell *21*, 723–737. https://doi.org/10.1016/j.ccr.2012.05.024.

1056 85. Delage, L., Lambert, M., Bardel, É., Kundlacz, C., Chartoire, D., Conchon, A., Peugnet, A.-L., Gorka, L.,
1057 Auberger, P., Jacquel, A., et al. (2023). BTG1 inactivation drives lymphomagenesis and promotes
1058 lymphoma dissemination through activation of BCAR1. Blood *141*, 1209–1220.
1059 https://doi.org/10.1182/blood.2022016943.

1060 86. Mlynarczyk, C., Teater, M., Pae, J., Chin, C.R., Wang, L., Arulraj, T., Barisic, D., Papin, A., Hoehn, K.B., Kots,
1061 E., et al. (2023). BTG1 mutation yields supercompetitive B cells primed for malignant transformation.
1062 Science *379*, eabj7412. https://doi.org/10.1126/science.abj7412.

1063 87. Tijchon, E., van Emst, L., Yuniati, L., van Ingen Schenau, D., Havinga, J., Rouault, J.-P., Hoogerbrugge, P.M.,
1064 van Leeuwen, F.N., and Scheijen, B. (2016). Tumor suppressors BTG1 and BTG2 regulate early mouse B-
1065 cell development. Haematologica *101*, e272-276. https://doi.org/10.3324/haematol.2015.139675.

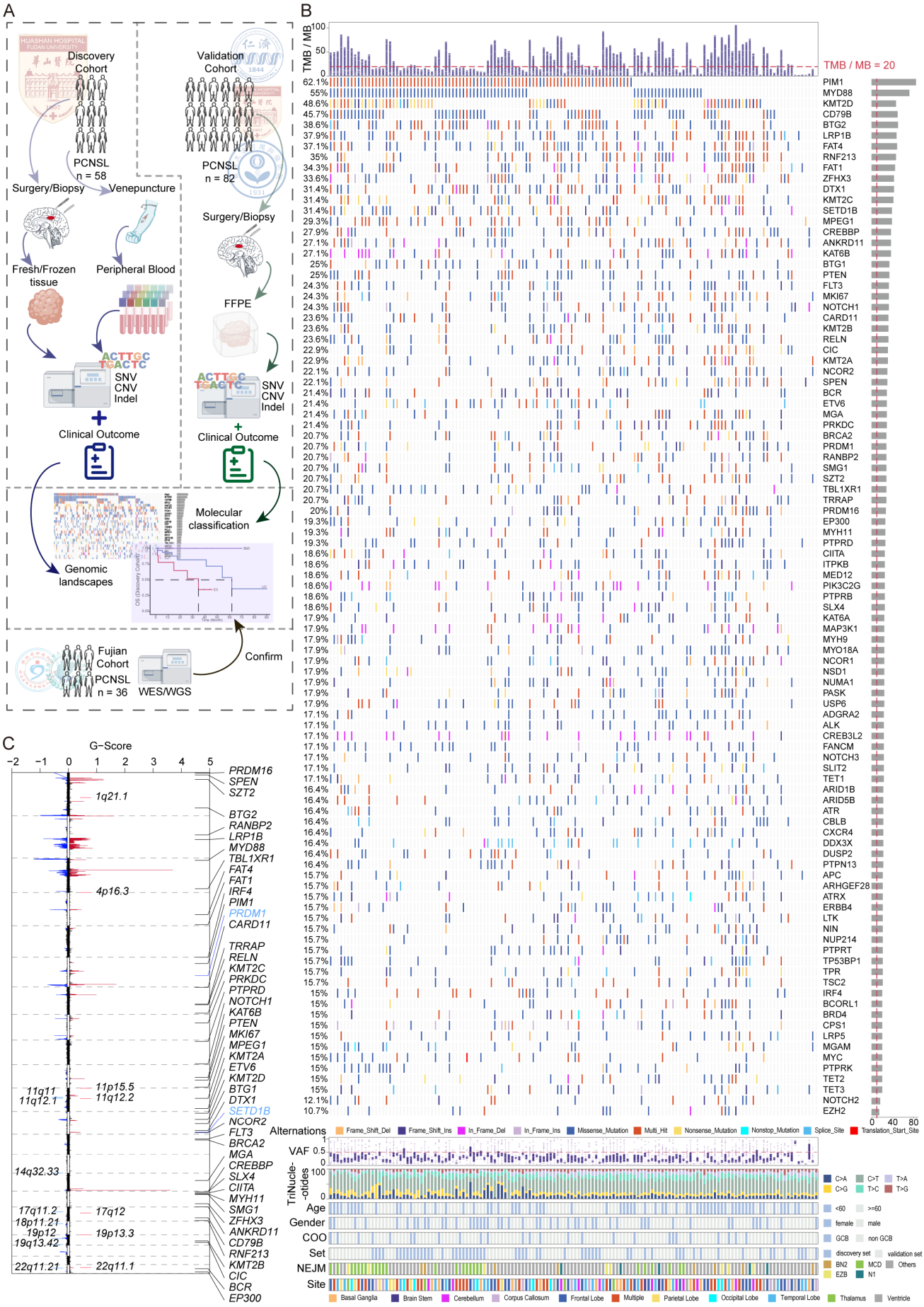
1066 88. Scheijen, B., Boer, J.M., Marke, R., Tijchon, E., van Ingen Schenau, D., Waanders, E., van Emst, L., van der
1067 Meer, L.T., Pieters, R., Escherich, G., et al. (2017). Tumor suppressors BTG1 and IKZF1 cooperate during
1068 mouse leukemia development and increase relapse risk in B-cell precursor acute lymphoblastic leukemia
1069 patients. Haematologica *102*, 541–551. https://doi.org/10.3324/haematol.2016.153023.

1070 89. Li, Y., Mao, X., Li, M., Li, L., Tong, X., and Huang, L. (2024). The predictive value of BTG1 for the response
1071 of newly diagnosed acute myeloid leukemia to decitabine. Clin Epigenetics *16*, 16.
1072 https://doi.org/10.1186/s13148-024-01627-9.

1073 90. Chen, T., Chen, X., Zhang, S., Zhu, J., Tang, B., Wang, A., Dong, L., Zhang, Z., Yu, C., Sun, Y., et al. (2021).
1074 The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. Genomics

Proteomics Bioinformatics *19*, 578–583. https://doi.org/10.1016/j.gpb.2021.08.001.

91. CNCB-NGDC Members and Partners (2022). Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2022. Nucleic Acids Res *50*, D27–D38. https://doi.org/10.1093/nar/gkab951.

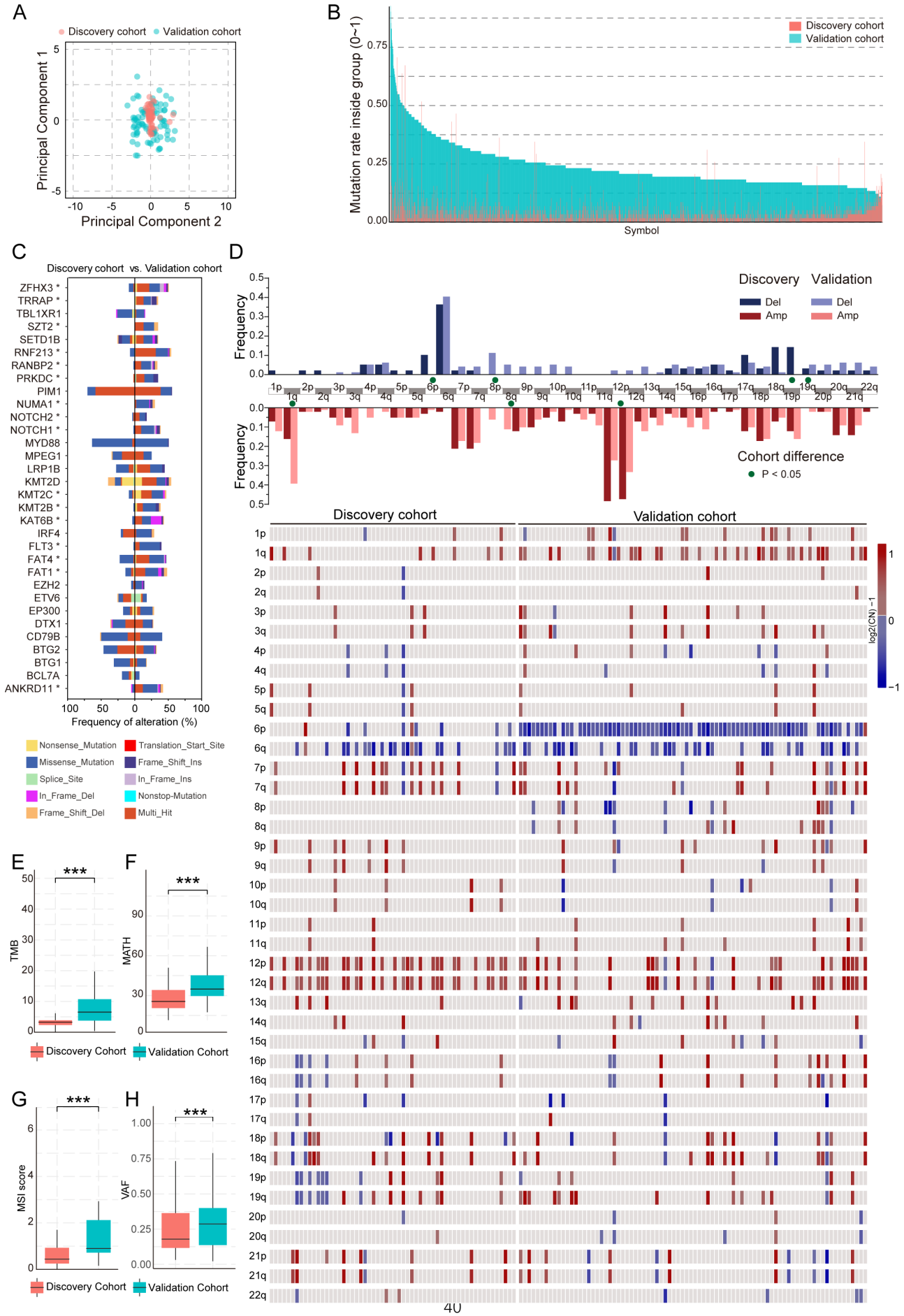

A
Discovery Cohort
PCNSL n = 58
Surgery/Biopsy
Venepuncture
Fresh/Frozen tissue
Peripheral Blood
SNV CNV Indel
Clinical Outcome
Validation Cohort
PCNSL n = 82
Surgery/Biopsy
FFPE
SNV CNV Indel
Clinical Outcome
Molecular classification
Genomic landscapes
Fujian Cohort
PCNSL n = 36
WES/WGS
Confirm
B
TMB / MB
TMB / MB = 20
Alternations
Frame_Shift_Del
Frame_Shift_Ins
In_Frame_Del
In_Frame_Ins
Missense_Mutation
Multi_Hit
Nonsense_Mutation
Nonstop_Mutation
Splice_Site
Translation_Start_Site
VAF
TriNucle-otides
Age
Gender
COO
Set
NEJM
Site
C
G-Score
1q21.1
4p16.3
11q11
11q12.1
11p15.5
11q12.2
14q32.33
17q11.2
18p11.21
19p12
19q13.42
17q12
19p13.3
22q11.21
22q11.1
PRDM16
SPEN
SZT2
BTG2
RANBP2
LRP1B
MYD88
TBL1XR1
FAT4
FAT1
IRF4
PIM1
PRDM1
CARD11
TRRAP
RELN
KMT2C
PRKDC
PTPRD
NOTCH1
KAT6B
PTEN
MKI67
MPEG1
KMT2A
ETV6
KMT2D
BTG1
DTX1
SETD1B
NCOR2
FLT3
BRCA2
MGA
CREBBP
SLX4
CIITA
MYH11
SMG1
ZFHX3
ANKRD11
CD79B
RNF213
KMT2B
CIC
BCR
EP300

1082 **Figure 1. Study design and the PCNSL mutational landscape in Chinese patients**

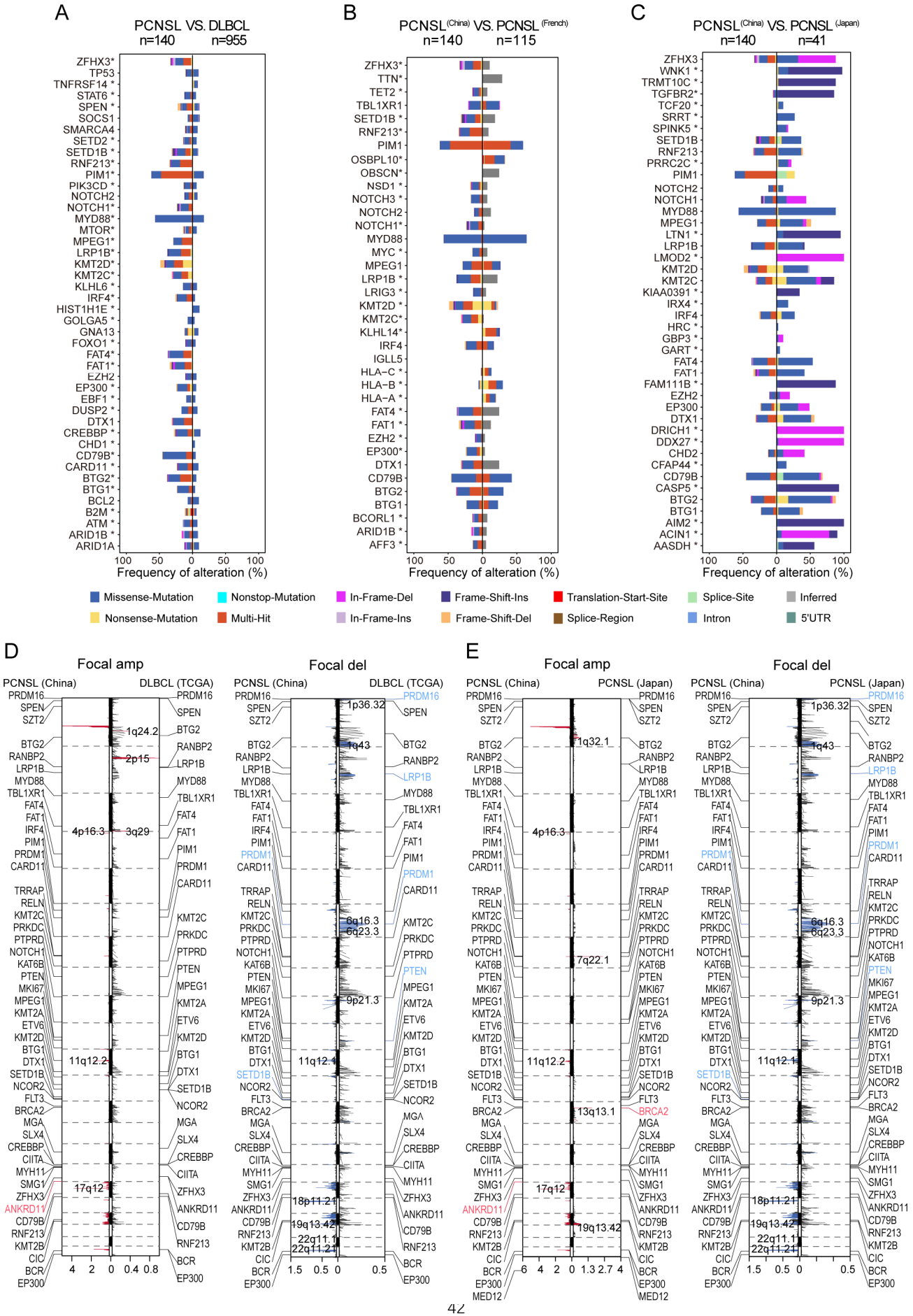
1083 A: Study design, including workflow and data composition for the discovery cohort and
1084 validation cohort.

1085 B: Number and frequency of recurrent mutations. Gene–sample matrix of recurrently mutated
1086 genes ranked by mutation frequency (n = 140). The total mutation density across the cohort
1087 is displayed at the top, with the variant allele fraction and trinucleotides at the bottom.

1088 C: GISTIC2.0 results of significant recurrent focal amplifications (red) and deletions (blue).
1089 Genes affected by each focal event are annotated (n = 140). X-axis: plot of chromosomes;
1090 Y-axis: G score.


A
Discovery cohort
Validation cohort
Principal Component 1
Principal Component 2
B
Discovery cohort
Validation cohort
Mutation rate inside group (0~1)
Symbol
C
Discovery cohort vs. Validation cohort
ZFHX3 *
TRRAP *
TBL1XR1
SZT2 *
SETD1B
RNF213 *
RANBP2 *
PRKDC *
PIM1
NUMA1 *
NOTCH2 *
NOTCH1 *
MYD88
MPEG1
LRP1B
KMT2D
KMT2C *
KMT2B *
KAT6B *
IRF4
FLT3 *
FAT4 *
FAT1 *
EZH2
ETV6
EP300
DTX1
CD79B
BTG2
BTG1
BCL7A
ANKRD11 *
Frequency of alteration (%)
Nonsense_Mutation
Missense_Mutation
Splice_Site
In_Frame_Del
Frame_Shift_Del
Translation_Start_Site
Frame_Shift_Ins
In_Frame_Ins
Nonstop-Mutation
Multi_Hit
D
Discovery
Validation
Del
Amp
Frequency
Cohort difference
P < 0.05
Discovery cohort
Validation cohort
log2(CN)
E
TMB
Discovery Cohort
Validation Cohort
F
MATH
G
MSI score
H
VAF

1092 **Figure 2. Comparison of mutation profiles between the discovery cohort and validation**
1093 **cohort of PCNSL patients**
1094 A: Principal component analysis of whole-exome sequencing (WES) data between the
1095 discovery cohort and validation cohort of PCNSL patients
1096 B: Mutation rate of WES data between the discovery cohort and validation cohort of PCNSL
1097 patients
1098 C: A mirror bar plot showing the frequencies of genetic alterations in the discovery cohort of
1099 PCNSL patients compared with those in the validation cohort.
1100 D: Arm-level copy number alterations of the discovery cohort and validation cohort samples
1101 are displayed. The frequencies of amplifications and deletions were compared.
1102 E: Comparison of the tumor mutational burden (TMB) between the discovery cohort and the
1103 validation cohort of PCNSL patients
1104 F: The levels of mutant-allele tumor heterogeneity (MATH) in the discovery cohort and the
1105 validation cohort of PCNSL patients were compared.
1106 G: The levels of microsatellite instability (MSI) in the discovery cohort and the validation
1107 cohort of PCNSL patients were compared.
1108 H: The variant allele frequency (VAF) in the discovery cohort and the validation cohort of
1109 PCNSL patients were compared.
1110 The discovery cohort included blood-paired fresh tumor tissue samples (n = 58), and the
1111 validation cohort included unpaired FFPE tumor tissue samples (n = 82). Independent-
1112 samples t tests and chi-squared tests were used. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.


A
PCNSL VS. DLBCL
n=140 n=955
B
PCNSL(China) VS. PCNSL(French)
n=140 n=115
C
PCNSL(China) VS. PCNSL(Japan)
n=140 n=41
Frequency of alteration (%)
Missense-Mutation
Nonstop-Mutation
In-Frame-Del
Frame-Shift-Ins
Translation-Start-Site
Splice-Site
Inferred
Nonsense-Mutation
Multi-Hit
In-Frame-Ins
Frame-Shift-Del
Splice-Region
Intron
5'UTR
D
Focal amp
Focal del
PCNSL (China)
DLBCL (TCGA)
E
PCNSL (Japan)

1114 **Figure 3. Mutation profiles of Chinese PCNSLs, DLBCLs, and PCNSLs of other races**

1115 A: A mirror bar plot showing the frequencies of genetic alterations between Chinese PCNSL
1116 patients (n = 140) and DLBCL patients (n = 955).

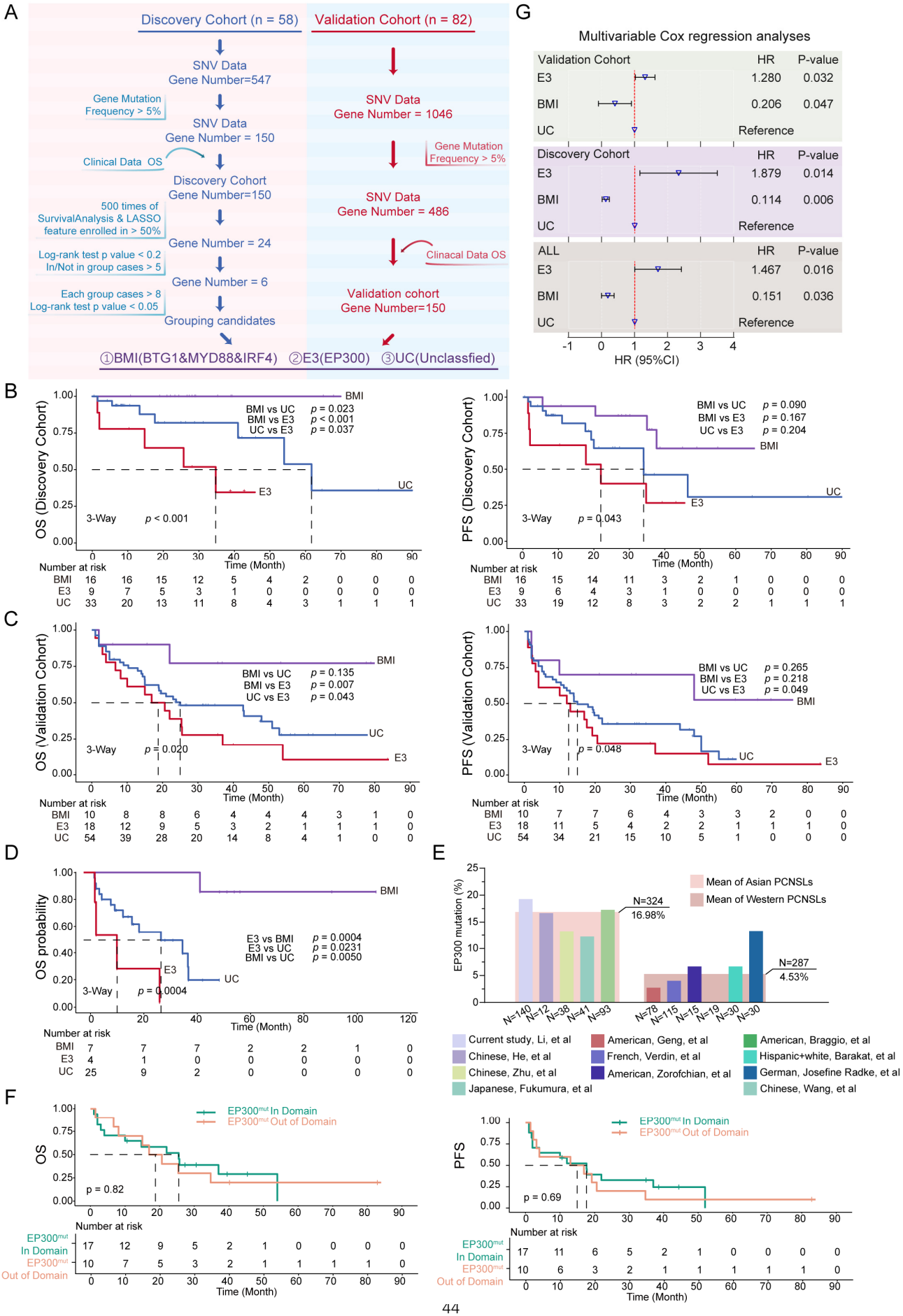
1117 B: A mirror bar plot showing the frequencies of genetic alterations between Chinese PCNSL
1118 patients (n = 140) and French PCNSL patients (n = 115).

1119 C: A mirror bar plot showing the frequencies of genetic alterations between Chinese PCNSL
1120 patients (n = 140) and Japanese PCNSL patients (n = 41).

1121 D: GISTIC-defined recurrent copy number focal deletions (blue) and gains (red) as mirror
1122 plots in DLBCL (TCGA, n = 46) and Chinese PCNSLs (n = 140) are shown. X-axis: plot of
1123 chromosomes; Y-axis: G score.

1124 E: The GISTIC2.0-defined recurrent copy number focal deletions (blue) and gains (red) in
1125 Japanese PCNSL patients (n = 41) and Chinese PCNSL patients (n = 140) are shown as
1126 mirror plots. Y-axis: plot of chromosomes; X-axis: G score.

1127 The chi-squared test was used. *$P < 0.05$. In the figure legend of A-C, 'Inferred' means that
1128 the mutation types were not directly provided but were calculated from the supplemental
1129 tables of the corresponding paper; thus, the results are 'Inferred', and the *in facto* mutation
1130 number will not be less than the inferred value.


A
Discovery Cohort (n = 58)
Validation Cohort (n = 82)
SNV Data
Gene Number=547
Gene Mutation Frequency > 5%
SNV Data
Gene Number = 150
Clinical Data OS
Discovery Cohort
Gene Number=150
500 times of SurvivalAnalysis & LASSO feature enrolled in > 50%
Gene Number = 24
Log-rank test p value < 0.2 In/Not in group cases > 5
Gene Number = 6
Each group cases > 8 Log-rank test p value < 0.05
Grouping candidates
SNV Data
Gene Number = 1046
Gene Mutation Frequency > 5%
SNV Data
Gene Number = 486
Clinacal Data OS
Validation cohort
Gene Number=150
①BMI(BTG1&MYD88&IRF4) ②E3(EP300) ③UC(Unclassfied)
G
Multivariable Cox regression analyses
Validation Cohort HR P-value
E3 1.280 0.032
BMI 0.206 0.047
UC Reference
Discovery Cohort HR P-value
E3 1.879 0.014
BMI 0.114 0.006
UC Reference
ALL HR P-value
E3 1.467 0.016
BMI 0.151 0.036
UC Reference
HR (95%CI)
B
OS (Discovery Cohort)
BMI vs UC p = 0.023
BMI vs E3 p < 0.001
UC vs E3 p = 0.037
3-Way p < 0.001
Time (Month)
Number at risk
BMI 16 16 15 12 5 4 2 0 0 0
E3 9 7 5 3 1 0 0 0 0 0
UC 33 20 13 11 8 4 3 1 1 1
PFS (Discovery Cohort)
BMI vs UC p = 0.090
BMI vs E3 p = 0.167
UC vs E3 p = 0.204
3-Way p = 0.043
Number at risk
BMI 16 15 14 11 3 2 1 0 0 0
E3 9 6 4 3 1 0 0 0 0 0
UC 33 19 12 8 3 2 2 1 1 1
C
OS (Validation Cohort)
BMI vs UC p = 0.135
BMI vs E3 p = 0.007
UC vs E3 p = 0.043
3-Way p = 0.020
Number at risk
BMI 10 8 8 6 4 4 4 3 1 0
E3 18 12 9 5 3 2 1 1 1 0
UC 54 39 28 20 14 8 4 1 0 0
PFS (Validation Cohort)
BMI vs UC p = 0.265
BMI vs E3 p = 0.218
UC vs E3 p = 0.049
3-Way p = 0.048
Number at risk
BMI 10 7 7 6 4 3 3 2 0 0
E3 18 11 5 4 2 2 1 1 1 0
UC 54 34 21 15 10 5 1 0 0 0
D
OS probability
E3 vs BMI p = 0.0004
E3 vs UC p = 0.0231
BMI vs UC p = 0.0050
3-Way p = 0.0004
Number at risk
BMI 7 7 7 2 2 1 0
E3 4 1 0 0 0 0 0
UC 25 9 2 0 0 0 0
E
EP300 mutation (%)
Mean of Asian PCNSLs
Mean of Western PCNSLs
N=324
16.98%
N=287
4.53%
N=140 N=12 N=38 N=41 N=93
N=78 N=115 N=15 N=19 N=30 N=30
Current study, Li, et al
Chinese, He, et al
Chinese, Zhu, et al
Japanese, Fukumura, et al
American, Geng, et al
French, Verdin, et al
American, Zorofchian, et al
American, Braggio, et al
Hispanic+white, Barakat, et al
German, Josefine Radke, et al
Chinese, Wang, et al
F
OS
EP300mut In Domain
EP300mut Out of Domain
p = 0.82
Number at risk
EP300mut In Domain 17 12 9 5 2 1 0 0 0 0
EP300mut Out of Domain 10 7 5 3 2 1 1 1 1 0
Time (Month)
PFS
EP300mut In Domain
EP300mut Out of Domain
p = 0.69
Number at risk
EP300mut In Domain 17 11 6 5 2 1 0 0 0 0
EP300mut Out of Domain 10 6 3 2 1 1 1 1 1 0
Time (Month)

1131

1132 **Figure 4. Integrated genetic drivers reveal PCNSL molecular subtypes with clinical**
1133 **outcome implications**

1134 A: Schematic of the typing strategies used to identify PCNSL molecular subtypes in the
1135 discovery cohort (n = 58) and in the validation cohort (n = 82).

1136 B: Kaplan–Meier estimates of overall survival (OS) and progression-free survival (PFS)
1137 among patients belonging to each molecular subtype in the discovery cohort.

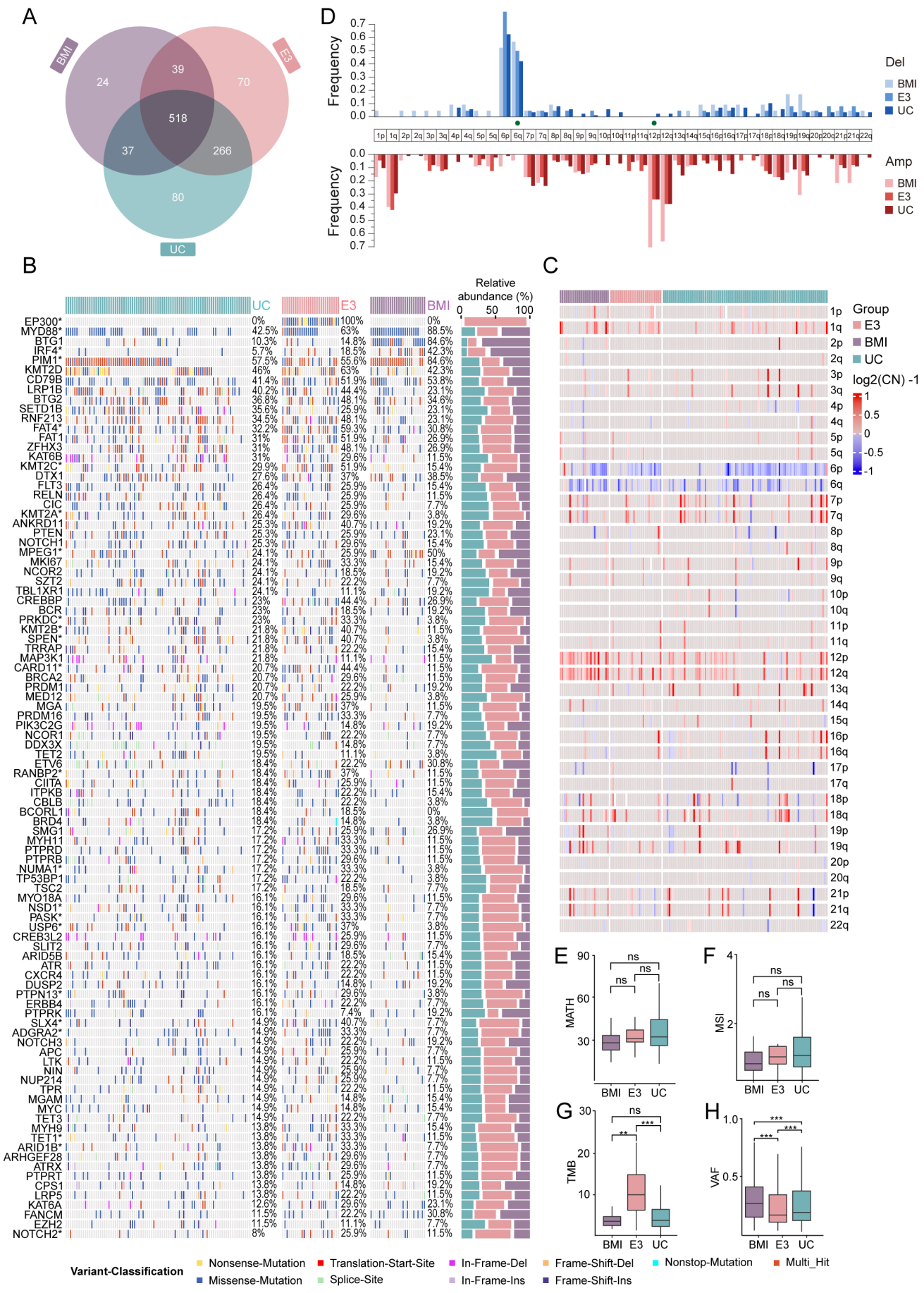
1138 C: Kaplan–Meier estimates of overall survival (OS) and progression-free survival (PFS)
1139 among patients belonging to each molecular subtype in the validation cohort.

1140 D: Kaplan–Meier estimates of overall survival (OS) among patients belonging to each
1141 molecular subtype in an independent external cohort (n=36).

1142 E: The prevalence of EP300 mutations in PCNSL patients of different races.

1143 F: Kaplan–Meier analyses for comparisons of the overall survival (OS) and progression-free
1144 survival (PFS) of patients with mutations within and those with mutations outside the EP300
1145 domain.

1146 G: Hazard ratio estimates of overall survival in PCNSL patients (n = 140).


A
BMI
E3
UC
24
39
70
518
37
266
80
D
Frequency
Del
BMI
E3
UC
Amp
B
UC
E3
BMI
Relative abundance (%)
C
Group
E3
BMI
UC
log2(CN)
E
MATH
F
MSI
G
TMB
H
VAF
ns
BMI E3 UC
Variant-Classification
Nonsense-Mutation
Missense-Mutation
Translation-Start-Site
Splice-Site
In-Frame-Del
In-Frame-Ins
Frame-Shift-Del
Frame-Shift-Ins
Nonstop-Mutation
Multi_Hit

1148 **Figure 5. Mutation profiles across three molecular subtypes of PCNSL**

1149 A: Venn diagram showing unique and shared mutations among the three identified PCNSL
1150 subtypes.

1151 B: Number and frequency of recurrent mutations and the gene–sample matrix of recurrently
1152 mutated genes among the three PCNSL subtypes. The relative abundance across the
1153 molecular subtypes is displayed on the right.

1154 C: Arm-level copy number alterations among the three molecular subtype samples are
1155 displayed.

1156 D: The frequencies of amplifications and deletions were compared among the three
1157 molecular subtype samples.

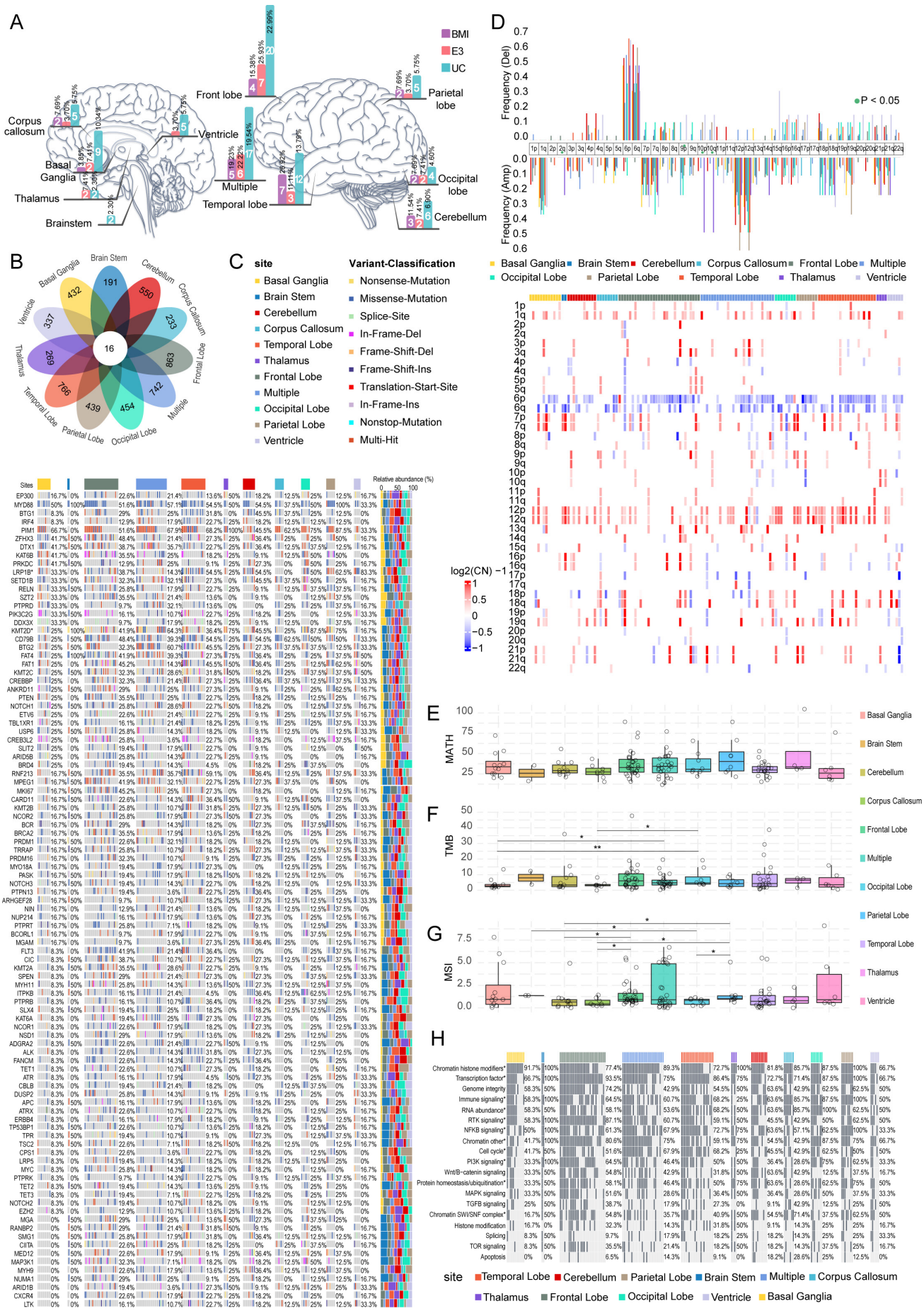
1158 E: The levels of mutant-allele tumor heterogeneity (MATH) were compared among the three
1159 molecular subtype samples.

1160 F: The levels of microsatellite instability (MSI) were compared among the three MS samples.

1161 G: The tumor mutational burden (TMB) was compared among the three MS samples.

1162 H: The levels of variant allele frequency (VAF) were compared among the three MS samples.

1163 Independent-samples t tests and chi-squared tests were used. *$P < 0.05$; **$P < 0.01$; ***$P <$
1164 0.001; ns $P > 0.05$.


A
BMI
E3
UC
Front lobe
Parietal lobe
Corpus callosum
Ventricle
Basal Ganglia
Thalamus
Brainstem
Multiple
Temporal lobe
Occipital lobe
Cerebellum
B
Basal Ganglia
Brain Stem
Cerebellum
Corpus Callosum
Frontal Lobe
Multiple
Occipital Lobe
Parietal Lobe
Temporal Lobe
Thalamus
Ventricle
16
C
site
Basal Ganglia
Brain Stem
Cerebellum
Corpus Callosum
Temporal Lobe
Thalamus
Frontal Lobe
Multiple
Occipital Lobe
Parietal Lobe
Ventricle
Variant-Classification
Nonsense-Mutation
Missense-Mutation
Splice-Site
In-Frame-Del
Frame-Shift-Del
Frame-Shift-Ins
Translation-Start-Site
In-Frame-Ins
Nonstop-Mutation
Multi-Hit
Relative abundance (%)
D
Frequency (Del)
Frequency (Amp)
P < 0.05
log2(CN)
E
MATH
F
TMB
G
MSI
H
Chromatin histone modifiers*
Transcription factor*
Genome integrity
Immune signaling*
RNA abundance*
RTK signaling*
NFKB signaling*
Chromatin other*
Cell cycle*
PI3K signaling*
Wnt/B-catenin signaling
Protein homeostasis/ubiquitination*
MAPK signaling
TGFB signaling
Chromatin SWI/SNF complex*
Histone modification
Splicing
TOR signaling
Apoptosis
site
Temporal Lobe
Cerebellum
Parietal Lobe
Brain Stem
Multiple
Corpus Callosum
Thalamus
Frontal Lobe
Occipital Lobe
Ventricle
Basal Ganglia

1166 **Figure 6. Mutational profiles across different sites of onset in PCNSL**

1167 A: The number and percentage of PCNSL patients with various molecular subtypes at
1168 different disease onset sites.

1169 B: Venn diagram of unique and shared mutations among PCNSLs with different sites of onset.

1170 C: The number and frequency of recurrent mutations, along with a gene–sample matrix of
1171 recurrently mutated genes in PCNSL patients at different sites of onset, are presented. The
1172 relative abundance across the molecular subtypes is displayed on the right.

1173 D: Arm-level copy number alterations across different sites of onset are displayed. The
1174 frequencies of amplifications and deletions were compared.

1175 E: The levels of mutant-allele tumor heterogeneity (MATH) were compared among the
1176 samples representing different sites of onset.

1177 F: Comparisons of the tumor mutational burden (TMB) among the different sites of onset
1178 samples were performed.

1179 G: The levels of microsatellite instability (MSI) were compared among the samples
1180 representing different sites of onset.

1181 H: Pathways affected by oncogenes in PCNSL patients with different sites of onset.

1182 The chi-squared test and one-way ANOVA were used. *$P < 0.05$.

1183