

Supplementary File

Title: A CSF ctDNA-based molecular panel for minimally invasive diagnosis of primary CNS lymphoma

Author: Haoyi Li, Huizhi Wang, Shaoping Shen, Bing Li, Jinhao Zhang, Hao Duan, Ming Li, Xinze Lv, Gehong Dong, Leilei Yuan, Zairan Wang, Mingxiao Li, Linxiao Li, Tengfei Zhu, Qian Liu, Lin Sang, Dabiao Zhou, Ling Qin, Song Lin, Yonggao Mou, Zheng Zhao, Fangang Meng, Feng Li, Xiaohui Ren

-Supplementary Tables	Page
Table S1. Clinical overview of patients in whom tissue biopsy was not initially performed	2-3
Table S2. List of 112 hotspot mutation probes specific to systemic DLBCL	4
Table S3. Data screening process for PCNSL (pcnsl_msk_2024)	5
Table S4. Data screening process for grade 4 glioma cases (glioma_msk_2018)	6
Table S5. Sequencing platform characteristics	7-8
Table S6. Gene-level mutation detection rates in the TIS-based training set and the external validation set after cross-platform harmonization	9-11
Table S7. Gene-level mutation detection rates across the CSF-based validation cohorts/set after cross-platform harmonization	12-14
Table S8. Comparison of gene-level mutation frequencies between PCNSL and grade 4 glioma in the TIS-based training set	15-16
Table S9. Comparative overview of liquid biopsy-based molecular diagnostics for PCNSL across two studies	17
-Supplementary Figures	
Figure S1. Unsupervised clustering analysis differentiating PCNSL and Grade 4 glioma based on genetic mutation profiles detected by 112 hotspot mutation probes	18
Figure S2. Distributions of gene-level mutation detection rates in PCNSL, grade 4 glioma and non-PCNSL across analytical datasets	19
Figure S3. Gene-level mutation detection rates in PCNSL vs. grade 4 glioma	20
Figure S4. Reliability assessment of three machine learning (ML)-generated panels in differentiating patients with PCNSL from grade 4 glioma in the TIS-based training set and external validation set	21
Figure S5. Concordance of mutational profiles between tissue and liquid specimen-informed sequencing in the BJTTH cohort	22-23
Figure S6. Positive detection rates of CSF ctDNA across the three validation sets.	24
Figure S7. Feasibility of three machine learning (ML)-generated panels for diagnosing PCNSL using CSF ctDNA profiling in CSF-based MSKCC validation set	25
Figure S8 Mutation sites of the eight genes included in the N8-RF-PCNSL	26

-Supplementary Tables**Table S1. Clinical overview of patients in whom tissue biopsy was not initially performed**

ID	Age	Gender	CSF IL-10 > 5 pg/mL	Molecular Diagnosis by N8-RF-PCNSL	Combined Diagnosis*	Reasons for Biopsy Infeasible [#]	Subsequent Management
1	79	M	Y	PCNSL	PCNSL	(1) & (2)	CR at 3 months (chemo-free therapy)
2	70	F	Y	PCNSL	PCNSL	(2)	CR at 3 months (chemo-free therapy)
3	76	M	N	PCNSL	PCNSL	(1)	CR at 3 months (chemo-free therapy)
4	75	M	N	PCNSL	PCNSL	(1) & (2)	PR at 3 months (chemo-free therapy)
5	73	M	N	PCNSL	PCNSL	(1)	PR at 3 months (chemo-free therapy)
6	55	M	Y	PCNSL	PCNSL	(1)	CR at 3 months (chemo-free therapy)
7	75	F	Y	PCNSL	PCNSL	(1) & (5)	CR at 3 months (chemo-free therapy)
8	67	F	Y	PCNSL	PCNSL	(3)	CR at 3 months (chemo-free therapy)
9	80	M	Y	PCNSL	PCNSL	(1) & (3)	CR at 3 months (chemo-free therapy)
10	75	F	Y	PCNSL	PCNSL	(1)	CR at 3 months (chemo-free therapy)
11	70	M	Y	PCNSL	PCNSL	(1) & (2)	CR at 3 months (chemo-free therapy)
12	69	F	Y	PCNSL	PCNSL	(4)	CR at 3 months (chemo-free therapy)
13	75	F	Y	PCNSL	PCNSL	(1)	CR at 3 months (chemo-free therapy)
14	72	F	Y	PCNSL	PCNSL	(3)	CR at 3 months (chemo-free therapy)
15	74	F	Y	PCNSL	PCNSL	(2)	CR at 3 months (chemo-free therapy)
16	76	M	N	PCNSL	PCNSL	(1) & (2)	CR at 3 months (chemo-free therapy)
17	83	M	N	PCNSL	PCNSL	(1)	CR at 3 months (chemo-free therapy)
18	81	F	Y	PCNSL	PCNSL	(1) & (5)	CR at 3 months (chemo-free therapy)
19	79	M	Y	PCNSL	PCNSL	(1)	CR at 3 months (chemo-free therapy)
20	67	M	Y	PCNSL	PCNSL	(3)	CR at 3 months (chemo-free therapy)
21	58	M	Y	PCNSL	PCNSL	(1) & (2)	CR at 3 months (chemo-free therapy)
22	62	M	Y	PCNSL	PCNSL	(3)	CR at 3 months (chemo-free therapy)
23	75	M	Y	PCNSL	PCNSL	(1) & (3)	Death before treatment
24	58	M	Y	PCNSL	PCNSL	(5)	PR at 3 months (chemo-free therapy)
25	52	F	Y	PCNSL	PCNSL	(3)	CR at 3 months (chemo-free therapy)
26	60	M	Y	PCNSL	PCNSL	(4)	PR at 3 months (chemo-free therapy)
27	73	M	N	Non-PCNSL	Non-PCNSL	(4)	Clinically diagnosed demyelination without pathological confirmation
28	61	M	N	Non-PCNSL	Non-PCNSL	(1) & (2)	Death before pathological confirmation
29	61	F	N	Non-PCNSL	Non-PCNSL	(5)	Clinically diagnosed PCNSL without pathological confirmation
30	62	F	N	Non-PCNSL	Non-PCNSL	(5)	Clinically diagnosed cancers of unknown primary without pathological confirmation
31	68	M	N	Non-PCNSL	Non-PCNSL	(3) & (5)	Clinically diagnosed GBM without pathological confirmation

32	48	F	N	Non-PCNSL	Non-PCNSL	(5)	Clinically diagnosed cerebral infarction without pathological confirmation
33	75	M	N	Non-PCNSL	Non-PCNSL	(1) & (2)	Clinically diagnosed GBM without pathological confirmation
34	36	F	N	Non-PCNSL	Non-PCNSL	(3)	Anaplastic astrocytoma confirmed by biopsy at an outside institution
35	55	F	N	Non-PCNSL	Non-PCNSL	(2)	GBM confirmed by biopsy at an outside institution
36	76	F	N	Non-PCNSL	Non-PCNSL	(1) & (4)	PCNSL confirmed by biopsy after corticosteroid-associated recurrence
37	38	M	N	Non-PCNSL	Non-PCNSL	(4)	PCNSL confirmed by biopsy after corticosteroid-associated recurrence
38	59	M	N	Non-PCNSL	Non-PCNSL	(2)	GBM confirmed by biopsy at our institution
39	61	M	Y	Non-PCNSL	Non-PCNSL	(2)	PCNSL confirmed by biopsy at our institution
40	68	F	N	Non-PCNSL	Non-PCNSL	(2)	Surgically confirmed astrocytoma (CNS WHO Grade 3) at an outside institution
41	70	M	N	Non-PCNSL	Non-PCNSL	(1)	GBM confirmed by biopsy at our institution
42	41	M	N	Non-PCNSL	Non-PCNSL	(3) & (5)	Surgically confirmed Craniopharyngioma at our institution

*Combined diagnosis: N8-RF-PCNSL (CSF ctDNA) plus CSF IL-10 testing. As elevated IL-10 lacks disease specificity, definitive molecular classification as PCNSL relied primarily on N8-RF-PCNSL.

#Reasons for biopsy infeasible: (1) High anesthesia risk (advanced age, severe cardiopulmonary disease, or major systemic comorbidities); (2) Long-term antithrombotic or antiplatelet therapy; (3) Deep-seated lesion with high hemorrhagic risk; (4) Pre-biopsy corticosteroid exposure; (5) Patient refusal of invasive procedure

Chemo-free therapy: Zanubrutinib (BTKi) plus Lenalidomide (IMiD)

BTKi: Bruton's tyrosine kinase inhibitor, CR: complete remission, CSF: cerebrospinal fluid, F: female,

GBM: glioblastoma, IL-10: interleukin-10, IMiD: Immunomodulatory drug, M: male, N: No, PCNSL:

primary central nervous system lymphoma, PR: partial remission, Y: yes

Table S2. List of 112 hotspot mutation probes specific to systemic DLBCL

<i>ALK</i> NM_004304.4	<i>ARID1A</i> NM_006015.4	<i>ARID1B</i> NM_020732.3	<i>ARID2</i> NM_152641.2	<i>ASXL3</i> NM_030632.2
<i>ATG5</i> NM_001286106.1	<i>ATM</i> NM_000051.3	<i>B2M</i> NM_004048.2	<i>BCL2</i> NM_000633.2	<i>BCL6</i> NM_001130845.1
<i>BCOR</i> NM_001123383.1	<i>BCORL1</i> NM_021946.4	<i>BIRC3</i> NM_001165.4	<i>BRAF</i> NM_004333.4	<i>BTK</i> NM_000061.2
<i>CARD11</i> NM_032415.5	<i>CCND1</i> NM_053056.2	<i>CCND2</i> NM_001759.3	<i>CCND3</i> NM_001760.4	<i>CD28</i> NM_006139.3
<i>CD58</i> NM_001779.2	<i>CD79A</i> NM_001783.3	<i>CD79B</i> NM_000626.3	<i>CDKN2A</i> NM_000077.4	<i>CDKN2B</i> NM_004936.3
<i>CHD8</i> NM_001170629.1	<i>CIITA</i> NM_000246.3	<i>CREBBP</i> NM_004380.2	<i>CRYBG1</i> NM_001624.3	<i>CTLA4</i> NM_005214.4
<i>CTNNB1</i> NM_001904.3	<i>CXCR4</i> NM_003467.2	<i>DDX3X</i> NM_001356.4	<i>DNMT3A</i> NM_022552.4	<i>DNMT3B</i> NM_006892.3
<i>DTX1</i> NM_004416.2	<i>DUSP22</i> NM_001286555.1	<i>EP300</i> NM_001429.3	<i>EZH2</i> NM_004456.4	<i>FAS</i> NM_000043.5
<i>FOXO1</i> NM_002015.3	<i>FOXO3</i> NM_001455.3	<i>FYN</i> NM_002037.5	<i>GATA3</i> NM_001002295.1	<i>GNA13</i> NM_006572.5
<i>ID3</i> NM_002167.4	<i>IDH2</i> NM_002168.3	<i>IGHD</i> NA	<i>IGHJ</i> NA	<i>IGHV</i> NA
<i>IRF4</i> NM_002460.3	<i>ITK</i> NM_005546.3	<i>ITPKB</i> NM_002221.3	<i>JAK1</i> NM_001320923.1	<i>JAK3</i> NM_000215.3
<i>KDM6A</i> NM_001291415.1	<i>KIR2DL4</i> NM_002255.5	<i>KIR3DL2</i> NM_006737.3	<i>KIT</i> NM_000222.2	<i>KLHL6</i> NM_130446.2
<i>KLRC1</i> NM_002259.4	<i>KLRC2</i> NM_002260.3	<i>KLRK1</i> NM_007360.3	<i>KMT2A</i> NM_001197104.1	<i>KMT2C</i> NM_170606.2
<i>KMT2D</i> NM_003482.3	<i>KRAS</i> NM_033360.3	<i>MAP2K1</i> NM_002755.3	<i>MAP3K14</i> NM_003954.4	<i>MEF2B</i> NM_001145785.1
<i>MET</i> NM_000245.3	<i>MFHAS1</i> NM_004225.2	<i>MGA</i> NM_001164273.1	<i>MTOR</i> NM_004958.3	<i>MYC</i> NM_002467.4
<i>MYD88</i> NM_002468.4	<i>NF1</i> NM_000267.3	<i>NOTCH1</i> NM_017617.4	<i>NOTCH2</i> NM_024408.3	<i>NRAS</i> NM_002524.4
<i>NSD2</i> NM_001042424.2	<i>PDGFRA</i> NM_006206.4	<i>PIK3CA</i> NM_006218.3	<i>PIM1</i> NM_001243186.1	<i>PRDM1</i> NM_001198.3
<i>PTEN</i> NM_000314.6	<i>RHOA</i> NM_001664.3	<i>SETD2</i> NM_014159.6	<i>SF3B1</i> NM_012433.3	<i>SGK1</i> NM_001143676.1
<i>SOCS1</i> NM_003745.1	<i>SPEN</i> NM_015001.2	<i>SPI1</i> NM_001080547.1	<i>STAT3</i> NM_139276.2	<i>STAT5B</i> NM_012448.3
<i>STAT6</i> NM_001178078.1	<i>STK11</i> NM_000455.4	<i>SYK</i> NM_001174167.2	<i>TBX21</i> NM_013351.1	<i>TCF3</i> NM_003200.3
<i>TET2</i> NM_001127208.2	<i>TNFAIP3</i> NM_001270507.1	<i>TNFRSF14</i> NM_003820.3	<i>TP53</i> NM_000546.5	<i>TP63</i> NM_003722.4
<i>TP73</i> NM_005427.3	<i>TRAF2</i> NM_021138.3	<i>TRAF3</i> NM_145725.2	<i>TSC1</i> NM_000368.4	<i>TSC2</i> NM_000548.4
<i>XPO1</i> NM_003400.3	<i>ZAP70</i> NM_001079.3			

Note: NM represents the transcript number used to annotate (RefSeq).

DLBCL: diffuse large B-cell lymphoma

Table S3. Data screening process for PCNSL (pcnsl_msk_2024)

	Screening Details	Patient ID
	Log in to the website:	
Step 1	https://www.cbioportal.org/study/summary?id=pcnsl_msk_2024	
Step 2	A total of 44 patients with 39 TIS and 47 CSF samples were identified.	
Step 3	Fifteen patients* (14 CSFs and 14 TISs) were excluded because they were pathologically confirmed SCNSL (“SCNSL” option under the “Cancer Subtype” module). Consequently, data from 29 patients with PCNSL were eligible for further analysis.	*14-184_2 (Tumor, CSF1), 14-184_8 (Tumor, CSF1), 14-184_9 (Tumor), 14-184_13 (Tumor), 14-184_14 (Tumor, CSF1), 14-184_15 (Tumor), 14-184_18 (Tumor, CSF1), 14-184_22 (Tumor), 14-184_24 (CSF1), 14-184_26 (Tumor, CSF1), 14-184_27 (Tumor, CSF1), 14-184_31 (Tumor, CSF1), 14-184_34 (Tumor, CSF1, CSF2), 14-184_36 (Tumor, CSF1, CSF2), 14-184_39 (Tumor, CSF1, CSF2)
Step 4	Among these 29 PCNSL patients with 58 samples (33 CSFs and 25 TISs), 10 patients* had duplicate CSF ctDNA sequencing records. The second records, which were associated with treatment monitoring, were removed. Finally, data from 29 patients with 23 CSFs and 25 TISs were retained for analysis.	*14-184_45_CSF2, 14-184_42_CSF2, 14-184_33_CSF2, 14-184_20_CSF2, 14-184_40_CSF2, 14-184_37_CSF2, 14-184_38_CSF2, 14-184_29_CSF2, 14-184_32_CSF2, 14-184_43_CSF2

CSF: cerebrospinal Fluid, ctDNA: circulating tumor DNA, PCNSL: primary central nervous system lymphoma, SCNSL: secondary central nervous system lymphoma, TIS: tumor tissue

Table S4. Data screening process for grade 4 glioma cases (glioma_msk_2018)

	Screening Details	Patient ID
Step 1	Log in to the website: http://www.cbioportal.org/study?id=glioma_msk_2018	
Step 2	A total of 42 patients with 39 TIS and 47 CSF samples were identified.	
Step 3	By selecting the “Glioblastoma Multiforme” option under the “Cancer Type Detailed” module, 28 patients pathologically confirmed as grade 4 glioma with 26 TIS and 33 CSF samples were retrieved.	
Step 4	Among these, 11 patients had 13 recurrent TIS samples* (“Recurrence” option under “Sample Type”), which were excluded. After filtering, 28 patients with 33 CSF and 13 TIS samples were retained.	*Patient-21-T, Patient-07-T3, Patient-20-T, Patient-11-T, Patient-18-T, Patient-17-T, Patient-28-T3, Patient-07-T2, Patient-07-T4, Patient-34-T, Patient-33-T, Patient-24-T, Patient-36-T2
Step 5	Of these 28 patients, 4* had CSF samples obtained via ventriculoperitoneal shunt and 1# had a duplicate CSF sample. These samples were also excluded, resulting in a final dataset of 28 patients with 28 CSF and 13 TIS samples for analysis.	* Patient-15-CSF-VP, Patient-17-CSF-VP, Patient-21-CSF-VP, Patient-34-CSF-VP # Patient-07-CSF2

CSF: cerebrospinal Fluid, TIS: tumor tissue

Table S5. Sequencing platform characteristics

	BJTTH/BJFTH	BJTTH (CGGA)	MSKCC_PCNSL	MSKCC_grade 4 glioma
Sequenced sample types	BJTTH: TIS, CSF and PLA BJFTH: CSF	TIS	TIS and CSF	TIS and CSF
Assay/Platform	Custom lymphoma-focused 112-gene hybrid-capture panel	Whole-exome sequencing	MSK-HemePACT	MSK-IMPACT 341/410/468
Capture scope	Approximately 314 kb; coding regions of 112 lymphoma-associated genes	Exome-wide protein-coding regions	585 cancer-related genes enriched for hematologic malignancies	TIS: Coding exons of 341, 410 or 468 cancer-associated genes CSF: Coding exons of 410 cancer-associated genes and selected introns
Sequencing method	Hybrid-capture targeted sequencing	Illumina HiSeq, paired-end sequencing	Hybrid-capture targeted sequencing	Hybrid-capture targeted sequencing
Reference genome	hg19	hg19	hg19	hg19
Normal comparator	Matched WBC	Matched blood	Matched normal	Matched normal in 73/85 source cases; pooled normals otherwise
Reported sequencing depth	Average depth 1000× for TIS/CSF; Average depth 10000× for PLA	Average depth 100× for TIS	N/A	Broader MSK series mean unique depth 718× for TIS
Variant-detection threshold	TIS/CSF: SNV ≥ 8 reads (MAF $\geq 1.0\%$), indel ≥ 5 reads (MAF $\geq 0.5\%$) PLA: SNV ≥ 8 (MAF $\geq 0.1\%$), indel ≥ 2 (MAF $\geq 0.05\%$)	SAVI2-derived empirical Bayesian threshold	Independent call: hotspot TD ≥ 20 , VAD ≥ 8 , VAF $\geq 2\%$; novel TD ≥ 20 , VAD ≥ 10 , VAF $\geq 5\%$ Secondary concordance search: VAD ≥ 2 and VAF $\geq 1\%$	CSF: Independent call: hotspot TD ≥ 20 , VAD ≥ 8 , VAF $\geq 2\%$; novel TD ≥ 20 , VAD ≥ 10 , VAF $\geq 5\%$ Secondary matched-pair search: VAD ≥ 2 and VAF $\geq 1\%$ TIS: Hotspot: DP ≥ 20 , mutant reads ≥ 8 , VAF $\geq 2\%$; non-hotspot: DP ≥ 20 , mutant reads ≥ 10 , VAF $\geq 5\%$

Aligner	BWA 0.7.10	BWA-MEM 0.7.12	Source MSK pipeline	MSK-IMPACT pipeline
Variant calling	GATK 3.2 and VarScan 2.4.3	SAVI2; SAMtools/bcftools	Source MSK pipeline	Source MSK-IMPACT pipeline

BCFtools: Binary Call Format tools, BJFTH: Beijing Fengtai Hospital, BJTTH: Beijing Tiantan Hospital, BWA: Burrows–Wheeler Aligner, CGGA: Chinese Glioma Genome Atlas, CSF: Cerebrospinal fluid, DP: Sequencing depth, GATK: Genome Analysis Toolkit, indel: Insertion/deletion, kb: Kilobase, MAF: Mutant allele frequency, MEM: Maximal exact matches, MSK: Memorial Sloan Kettering, MSKCC: Memorial Sloan Kettering Cancer Center, N/A: Not available, PCNSL: Primary central nervous system lymphoma, PLA: Plasma, SAMtools: Sequence Alignment/Map tools, SAVI2: Statistical Algorithm for Variant Frequency Identification 2, SNV: Single-nucleotide variant, TD: Total depth, TIS: Tumor tissue, VAD: Variant allele depth, VAF: Variant allele frequency, WBC: White blood cell.

Table S6. Gene-level mutation detection rates in the TIS-based training set and the external validation set after cross-platform harmonization

	TIS-based training set		TIS-based external validation set	
	PCNSL (n = 64)	Grade 4 glioma (n = 54)	PCNSL (n = 25)	Grade 4 glioma (n = 13)
<i>MYD88</i>	73.4%	0.0%	76.0%	0.0%
<i>PIM1</i>	68.8%	0.0%	68.0%	7.7%
<i>CD79B</i>	46.9%	1.9%	44.0%	0.0%
<i>DTX1</i>	31.3%	1.9%	48.0%	0.0%
<i>GNA13</i>	23.4%	0.0%	20.0%	0.0%
<i>IRF4</i>	23.4%	0.0%	20.0%	0.0%
<i>KMT2D</i>	18.8%	0.0%	40.0%	7.7%
<i>CARD11</i>	15.6%	1.9%	24.0%	7.7%
<i>ITPKB</i>	15.6%	1.9%	0.0%	0.0%
<i>PRDM1</i>	12.5%	0.0%	12.0%	0.0%
<i>TP53</i>	12.5%	48.1%	12.0%	46.2%
<i>CD58</i>	12.5%	0.0%	8.0%	0.0%
<i>EP300</i>	12.5%	0.0%	8.0%	7.7%
<i>B2M</i>	9.4%	0.0%	16.0%	0.0%
<i>SOCS1</i>	9.4%	0.0%	16.0%	7.7%
<i>BCL2</i>	9.4%	0.0%	4.0%	0.0%
<i>PTEN</i>	9.4%	16.7%	4.0%	46.2%
<i>CCND3</i>	7.8%	0.0%	12.0%	0.0%
<i>CREBBP</i>	6.3%	0.0%	16.0%	7.7%
<i>CIITA</i>	6.3%	1.9%	12.0%	0.0%
<i>KMT2C</i>	6.3%	0.0%	12.0%	7.7%
<i>XPO1</i>	6.3%	0.0%	12.0%	0.0%
<i>KLHL6</i>	6.3%	0.0%	8.0%	0.0%
<i>TNFAIP3</i>	6.3%	0.0%	4.0%	0.0%
<i>BCL6</i>	6.3%	1.9%	0.0%	0.0%
<i>MEF2B</i>	4.7%	0.0%	8.0%	0.0%
<i>FOXO1</i>	4.7%	1.9%	4.0%	0.0%
<i>NOTCH2</i>	4.7%	1.9%	4.0%	0.0%
<i>SGK1</i>	4.7%	0.0%	0.0%	0.0%
<i>CXCR4</i>	3.1%	0.0%	16.0%	7.7%
<i>BCOR</i>	3.1%	1.9%	12.0%	7.7%
<i>MYC</i>	3.1%	0.0%	4.0%	0.0%
<i>CD28</i>	3.1%	0.0%	0.0%	0.0%
<i>CD79A</i>	3.1%	0.0%	0.0%	0.0%
<i>CDKN2A</i>	3.1%	1.9%	0.0%	0.0%
<i>TET2</i>	3.1%	0.0%	0.0%	7.7%
<i>DDX3X</i>	1.6%	0.0%	12.0%	0.0%

<i>ATM</i>	1.6%	0.0%	4.0%	0.0%
<i>EZH2</i>	1.6%	0.0%	4.0%	0.0%
<i>NF1</i>	1.6%	18.5%	4.0%	15.4%
<i>ALK</i>	1.6%	0.0%	0.0%	7.7%
<i>ARID2</i>	1.6%	1.9%	0.0%	15.4%
<i>BIRC3</i>	1.6%	0.0%	0.0%	0.0%
<i>FAS</i>	1.6%	0.0%	0.0%	0.0%
<i>GATA3</i>	1.6%	1.9%	0.0%	0.0%
<i>ID3</i>	1.6%	0.0%	0.0%	0.0%
<i>MFHAS1</i>	1.6%	0.0%	0.0%	0.0%
<i>RHOA</i>	1.6%	0.0%	0.0%	0.0%
<i>SF3B1</i>	1.6%	0.0%	0.0%	7.7%
<i>TRAF2</i>	1.6%	0.0%	0.0%	0.0%
<i>ZAP70</i>	1.6%	0.0%	0.0%	0.0%
<i>ARID1B</i>	0.0%	1.9%	12.0%	7.7%
<i>CCND2</i>	0.0%	0.0%	8.0%	7.7%
<i>PDGFRA</i>	0.0%	7.4%	8.0%	7.7%
<i>SETD2</i>	0.0%	3.7%	8.0%	0.0%
<i>DNMT3A</i>	0.0%	0.0%	4.0%	0.0%
<i>DNMT3B</i>	0.0%	0.0%	4.0%	0.0%
<i>IDH2</i>	0.0%	0.0%	4.0%	0.0%
<i>KDM6A</i>	0.0%	0.0%	4.0%	0.0%
<i>KIT</i>	0.0%	1.9%	4.0%	7.7%
<i>NOTCH1</i>	0.0%	3.7%	4.0%	0.0%
<i>TCF3</i>	0.0%	0.0%	4.0%	0.0%
<i>TP63</i>	0.0%	0.0%	4.0%	0.0%
<i>ARID1A</i>	0.0%	0.0%	0.0%	7.7%
<i>ASXL3</i>	0.0%	0.0%	0.0%	0.0%
<i>ATG5</i>	0.0%	0.0%	0.0%	0.0%
<i>BCORL1</i>	0.0%	5.6%	0.0%	0.0%
<i>BRAF</i>	0.0%	0.0%	0.0%	0.0%
<i>BTK</i>	0.0%	0.0%	0.0%	0.0%
<i>CCND1</i>	0.0%	0.0%	0.0%	0.0%
<i>CDKN2B</i>	0.0%	0.0%	0.0%	0.0%
<i>CHD8</i>	0.0%	0.0%	0.0%	0.0%
<i>CRYBG1</i>	0.0%	0.0%	0.0%	0.0%
<i>CTLA4</i>	0.0%	0.0%	0.0%	0.0%
<i>CTNNB1</i>	0.0%	0.0%	0.0%	15.4%
<i>DUSP22</i>	0.0%	0.0%	0.0%	0.0%
<i>FOXO3</i>	0.0%	0.0%	0.0%	0.0%
<i>FYN</i>	0.0%	0.0%	0.0%	7.7%
<i>IGHD</i>	0.0%	0.0%	0.0%	0.0%
<i>IGHJ</i>	0.0%	0.0%	0.0%	0.0%

<i>IGHV</i>	0.0%	0.0%	0.0%	0.0%
<i>ITK</i>	0.0%	1.9%	0.0%	0.0%
<i>JAK1</i>	0.0%	0.0%	0.0%	7.7%
<i>JAK3</i>	0.0%	1.9%	0.0%	0.0%
<i>KIR2DL4</i>	0.0%	0.0%	0.0%	0.0%
<i>KIR3DL2</i>	0.0%	1.9%	0.0%	0.0%
<i>KLRC1</i>	0.0%	0.0%	0.0%	0.0%
<i>KLRC2</i>	0.0%	0.0%	0.0%	0.0%
<i>KLRK1</i>	0.0%	0.0%	0.0%	0.0%
<i>KMT2A</i>	0.0%	3.7%	0.0%	7.7%
<i>KRAS</i>	0.0%	0.0%	0.0%	0.0%
<i>MAP2K1</i>	0.0%	0.0%	0.0%	7.7%
<i>MAP3K14</i>	0.0%	0.0%	0.0%	0.0%
<i>MET</i>	0.0%	0.0%	0.0%	7.7%
<i>MGA</i>	0.0%	0.0%	0.0%	7.7%
<i>MTOR</i>	0.0%	0.0%	0.0%	15.4%
<i>NRAS</i>	0.0%	0.0%	0.0%	0.0%
<i>NSD2</i>	0.0%	0.0%	0.0%	0.0%
<i>PIK3CA</i>	0.0%	9.3%	0.0%	30.8%
<i>SPEN</i>	0.0%	0.0%	0.0%	7.7%
<i>SPI1</i>	0.0%	0.0%	0.0%	0.0%
<i>STAT3</i>	0.0%	0.0%	0.0%	7.7%
<i>STAT5B</i>	0.0%	1.9%	0.0%	0.0%
<i>STAT6</i>	0.0%	0.0%	0.0%	0.0%
<i>STK11</i>	0.0%	0.0%	0.0%	7.7%
<i>SYK</i>	0.0%	0.0%	0.0%	7.7%
<i>TBX21</i>	0.0%	0.0%	0.0%	0.0%
<i>TNFRSF14</i>	0.0%	0.0%	0.0%	0.0%
<i>TP73</i>	0.0%	0.0%	0.0%	0.0%
<i>TRAF3</i>	0.0%	0.0%	0.0%	0.0%
<i>TSC1</i>	0.0%	3.7%	0.0%	7.7%
<i>TSC2</i>	0.0%	0.0%	0.0%	7.7%

n: Number of samples, PCNSL: Primary central nervous system lymphoma, TIS: Tumor tissue

Table S7. Gene-level mutation detection rates across the CSF-based validation cohorts/set after cross-platform harmonization

	CSF-based BJTH		CSF-based BJFTH		CSF-based MSKCC	
	validation cohort		validation cohort		validation set	
	PCNSL (n = 73)	Non-PCNSL (n = 15)	PCNSL (n = 27)	Non-PCNSL (n = 14)	PCNSL (n = 23)	Grade 4 glioma (n = 28)
<i>MYD88</i>	68.5%	0.0%	66.7%	0.0%	30.4%	3.6%
<i>PIM1</i>	58.9%	0.0%	70.4%	0.0%	26.1%	7.1%
<i>CD79B</i>	41.1%	0.0%	51.9%	0.0%	8.7%	0.0%
<i>DTX1</i>	26.0%	0.0%	22.2%	0.0%	17.4%	0.0%
<i>KMT2D</i>	19.2%	6.7%	22.2%	7.1%	13.0%	10.7%
<i>GNA13</i>	19.2%	0.0%	7.4%	0.0%	8.7%	0.0%
<i>PRDM1</i>	17.8%	0.0%	14.8%	0.0%	8.7%	7.1%
<i>IRF4</i>	17.8%	0.0%	11.1%	0.0%	0.0%	3.6%
<i>TP53</i>	16.4%	13.3%	18.5%	21.4%	4.3%	57.1%
<i>CARD11</i>	15.1%	0.0%	29.6%	0.0%	8.7%	7.1%
<i>ITPKB</i>	13.7%	0.0%	25.9%	0.0%	0.0%	0.0%
<i>CD58</i>	11.0%	0.0%	14.8%	0.0%	4.3%	0.0%
<i>EP300</i>	11.0%	0.0%	3.7%	0.0%	0.0%	10.7%
<i>B2M</i>	9.6%	0.0%	14.8%	0.0%	8.7%	0.0%
<i>SOCS1</i>	9.6%	0.0%	11.1%	0.0%	8.7%	3.6%
<i>BCL2</i>	9.6%	0.0%	7.4%	0.0%	0.0%	0.0%
<i>PTEN</i>	9.6%	0.0%	0.0%	7.1%	0.0%	35.7%
<i>CCND3</i>	8.2%	0.0%	3.7%	0.0%	4.3%	3.6%
<i>TNFAIP3</i>	8.2%	0.0%	11.1%	0.0%	4.3%	0.0%
<i>KLHL6</i>	8.2%	0.0%	11.1%	0.0%	0.0%	0.0%
<i>CREBBP</i>	6.8%	0.0%	3.7%	0.0%	8.7%	14.3%
<i>CIITA</i>	5.5%	0.0%	14.8%	0.0%	0.0%	0.0%
<i>MYC</i>	5.5%	0.0%	0.0%	0.0%	0.0%	3.6%
<i>KMT2C</i>	4.1%	0.0%	0.0%	0.0%	8.7%	14.3%
<i>MEF2B</i>	4.1%	0.0%	3.7%	0.0%	8.7%	0.0%
<i>BCOR</i>	4.1%	0.0%	7.4%	7.1%	4.3%	10.7%
<i>NOTCH2</i>	4.1%	0.0%	3.7%	0.0%	4.3%	7.1%
<i>XPO1</i>	4.1%	0.0%	3.7%	0.0%	4.3%	0.0%
<i>TRAF2</i>	4.1%	0.0%	3.7%	0.0%	0.0%	3.6%
<i>CXCR4</i>	2.7%	0.0%	0.0%	0.0%	8.7%	7.1%
<i>CDKN2A</i>	2.7%	0.0%	14.8%	0.0%	4.3%	3.6%
<i>TP63</i>	2.7%	0.0%	3.7%	0.0%	4.3%	3.6%
<i>ARID2</i>	2.7%	0.0%	0.0%	0.0%	0.0%	21.4%
<i>CD28</i>	2.7%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>CD79A</i>	2.7%	0.0%	3.7%	0.0%	0.0%	0.0%
<i>TET2</i>	2.7%	0.0%	0.0%	0.0%	0.0%	7.1%

<i>ATM</i>	1.4%	6.7%	3.7%	7.1%	4.3%	7.1%
<i>DDX3X</i>	1.4%	0.0%	7.4%	0.0%	4.3%	0.0%
<i>DNMT3A</i>	1.4%	0.0%	0.0%	7.1%	4.3%	3.6%
<i>FOXO1</i>	1.4%	0.0%	3.7%	0.0%	4.3%	3.6%
<i>BCL6</i>	1.4%	0.0%	3.7%	0.0%	0.0%	3.6%
<i>EZH2</i>	1.4%	0.0%	0.0%	7.1%	0.0%	0.0%
<i>FAS</i>	1.4%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>KDM6A</i>	1.4%	0.0%	3.7%	0.0%	0.0%	3.6%
<i>MFHAS1</i>	1.4%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>NF1</i>	1.4%	0.0%	0.0%	7.1%	0.0%	14.3%
<i>SF3B1</i>	1.4%	0.0%	3.7%	0.0%	0.0%	3.6%
<i>SGK1</i>	1.4%	0.0%	3.7%	0.0%	0.0%	0.0%
<i>ZAP70</i>	1.4%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>ARID1A</i>	0.0%	0.0%	0.0%	0.0%	4.3%	10.7%
<i>ARID1B</i>	0.0%	0.0%	0.0%	0.0%	4.3%	10.7%
<i>BCORL1</i>	0.0%	0.0%	0.0%	7.1%	4.3%	0.0%
<i>KIT</i>	0.0%	0.0%	0.0%	0.0%	4.3%	3.6%
<i>NOTCH1</i>	0.0%	0.0%	3.7%	0.0%	4.3%	7.1%
<i>PDGFRA</i>	0.0%	13.3%	0.0%	21.4%	4.3%	14.3%
<i>SETD2</i>	0.0%	0.0%	0.0%	7.1%	4.3%	3.6%
<i>ALK</i>	0.0%	0.0%	0.0%	0.0%	0.0%	10.7%
<i>ASXL3</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>ATG5</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>BIRC3</i>	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%
<i>BRAF</i>	0.0%	6.7%	0.0%	0.0%	0.0%	7.1%
<i>BTK</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>CCND1</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>CCND2</i>	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%
<i>CDKN2B</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>CHD8</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>CRYBG1</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>CTLA4</i>	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%
<i>CTNNB1</i>	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%
<i>DNMT3B</i>	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%
<i>DUSP22</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>FOXO3</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>FYN</i>	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%
<i>GATA3</i>	0.0%	0.0%	0.0%	0.0%	0.0%	7.1%
<i>ID3</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>IDH2</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>IGHD</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>IGHJ</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>IGHV</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

<i>ITK</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>JAK1</i>	0.0%	0.0%	0.0%	0.0%	0.0%	7.1%
<i>JAK3</i>	0.0%	6.7%	3.7%	0.0%	0.0%	10.7%
<i>KIR2DL4</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>KIR3DL2</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>KLRC1</i>	0.0%	0.0%	0.0%	7.1%	0.0%	0.0%
<i>KLRC2</i>	0.0%	0.0%	3.7%	0.0%	0.0%	0.0%
<i>KLRK1</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>KMT2A</i>	0.0%	0.0%	0.0%	0.0%	0.0%	10.7%
<i>KRAS</i>	0.0%	13.3%	0.0%	0.0%	0.0%	3.6%
<i>MAP2K1</i>	0.0%	0.0%	0.0%	0.0%	0.0%	7.1%
<i>MAP3K14</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>MET</i>	0.0%	0.0%	0.0%	0.0%	0.0%	10.7%
<i>MGA</i>	0.0%	0.0%	0.0%	0.0%	0.0%	10.7%
<i>MTOR</i>	0.0%	0.0%	0.0%	7.1%	0.0%	10.7%
<i>NRAS</i>	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%
<i>NSD2</i>	0.0%	0.0%	3.7%	0.0%	0.0%	0.0%
<i>PIK3CA</i>	0.0%	0.0%	0.0%	0.0%	0.0%	21.4%
<i>RHOA</i>	0.0%	0.0%	3.7%	0.0%	0.0%	0.0%
<i>SPEN</i>	0.0%	0.0%	0.0%	0.0%	0.0%	10.7%
<i>SPI1</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>STAT3</i>	0.0%	0.0%	3.7%	0.0%	0.0%	3.6%
<i>STAT5B</i>	0.0%	6.7%	0.0%	0.0%	0.0%	0.0%
<i>STAT6</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>STK11</i>	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%
<i>SYK</i>	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%
<i>TBX21</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>TCF3</i>	0.0%	0.0%	0.0%	0.0%	0.0%	7.1%
<i>TNFRSF14</i>	0.0%	0.0%	3.7%	0.0%	0.0%	0.0%
<i>TP73</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>TRAF3</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>TSC1</i>	0.0%	0.0%	0.0%	0.0%	0.0%	10.7%
<i>TSC2</i>	0.0%	0.0%	3.7%	0.0%	0.0%	7.1%

BJTTH: Beijing Tiantan Hospital, BJFTH: Beijing Fengtai Hospital, CSF: Cerebrospinal fluid, MSKCC: Memorial Sloan Kettering Cancer Center, n: Number of samples, PCNSL: Primary central nervous system lymphoma

Table S8. Comparison of gene-level mutation frequencies between PCNSL and grade 4 glioma in the TIS-based training set

Gene	PCNSL, n/N (%)	Grade 4 glioma, n/N (%)	Fisher's exact <i>P</i>	FDR-adjusted <i>P</i>
<i>MYD88</i>	47/64 (73.4)	0/54 (0.0)	1.12E-18	7.17E-17
<i>PIM1</i>	44/64 (68.8)	0/54 (0.0)	3.81E-17	1.22E-15
<i>CD79B</i>	30/64 (46.9)	1/54 (1.9)	6.33E-09	1.35E-07
<i>DTX1</i>	20/64 (31.3)	1/54 (1.9)	1.57E-05	2.50E-04
<i>TP53</i>	8/64 (12.5)	26/54 (48.2)	3.05E-05	3.91E-04
<i>GNA13</i>	15/64 (23.4)	0/54 (0.0)	4.65E-05	4.25E-04
<i>IRF4</i>	15/64 (23.4)	0/54 (0.0)	4.65E-05	4.25E-04
<i>KMT2D</i>	12/64 (18.8)	0/54 (0.0)	4.25E-04	3.40E-03
<i>NF1</i>	1/64 (1.6)	10/54 (18.5)	2.48E-03	1.76E-02
<i>CD58</i>	8/64 (12.5)	0/54 (0.0)	7.47E-03	3.99E-02
<i>EP300</i>	8/64 (12.5)	0/54 (0.0)	7.47E-03	3.99E-02
<i>PRDM1</i>	8/64 (12.5)	0/54 (0.0)	7.47E-03	3.99E-02
<i>CARD11</i>	10/64 (15.6)	1/54 (1.9)	1.10E-02	5.04E-02
<i>ITPKB</i>	10/64 (15.6)	1/54 (1.9)	1.10E-02	5.04E-02
<i>PIK3CA</i>	0/64 (0.00)	5/54 (9.3)	1.81E-02	7.71E-02
<i>B2M</i>	6/64 (9.4)	0/54 (0.0)	3.06E-02	1.09E-01
<i>BCL2</i>	6/64 (9.4)	0/54 (0.0)	3.06E-02	1.09E-01
<i>SOCS1</i>	6/64 (9.4)	0/54 (0.0)	3.06E-02	1.09E-01
<i>PDGFRA</i>	0/64 (0.0)	4/54 (7.4)	4.12E-02	1.39E-01
<i>CCND3</i>	5/64 (7.8)	0/54 (0.0)	6.17E-02	1.97E-01
<i>BCORL1</i>	0/64 (0.0)	3/54 (5.6)	9.29E-02	2.83E-01
<i>CREBBP</i>	4/64 (6.3)	0/54 (0.0)	1.24E-01	3.05E-01
<i>KLHL6</i>	4/64 (6.3)	0/54 (0.0)	1.24E-01	3.05E-01
<i>KMT2C</i>	4/64 (6.3)	0/54 (0.0)	1.24E-01	3.05E-01
<i>TNFAIP3</i>	4/64 (6.3)	0/54 (0.0)	1.24E-01	3.05E-01
<i>XPO1</i>	4/64 (6.3)	0/54 (0.0)	1.24E-01	3.05E-01
<i>KMT2A</i>	0/64 (0.0)	2/54 (3.7)	2.07E-01	4.42E-01
<i>NOTCH1</i>	0/64 (0.0)	2/54 (3.7)	2.07E-01	4.42E-01
<i>SETD2</i>	0/64 (0.0)	2/54 (3.7)	2.07E-01	4.42E-01
<i>TSC1</i>	0/64 (0.0)	2/54 (3.7)	2.07E-01	4.42E-01
<i>MEF2B</i>	3/64 (4.7)	0/54 (0.0)	2.49E-01	4.98E-01
<i>SGK1</i>	3/64 (4.7)	0/54 (0.0)	2.49E-01	4.98E-01
<i>PTEN</i>	6/64 (9.4)	9/54 (16.7)	2.76E-01	5.34E-01
<i>BCL6</i>	4/64 (6.3)	1/54 (1.9)	3.73E-01	6.83E-01
<i>CIITA</i>	4/64 (6.3)	1/54 (1.9)	3.73E-01	6.83E-01
<i>ARID1B</i>	0/64 (0.0)	1/54 (1.9)	4.58E-01	6.95E-01
<i>ITK</i>	0/64 (0.0)	1/54 (1.9)	4.58E-01	6.95E-01
<i>JAK3</i>	0/64 (0.0)	1/54 (1.9)	4.58E-01	6.95E-01
<i>KIR3DL2</i>	0/64 (0.0)	1/54 (1.9)	4.58E-01	6.95E-01

<i>KIT</i>	0/64 (0.0)	1/54 (1.9)	4.58E-01	6.95E-01
<i>STAT5B</i>	0/64 (0.0)	1/54 (1.9)	4.58E-01	6.95E-01
<i>CD28</i>	2/64 (3.1)	0/54 (0.0)	4.99E-01	6.95E-01
<i>CD79A</i>	2/64 (3.1)	0/54 (0.0)	4.99E-01	6.95E-01
<i>CXCR4</i>	2/64 (3.1)	0/54 (0.0)	4.99E-01	6.95E-01
<i>MYC</i>	2/64 (3.1)	0/54 (0.0)	4.99E-01	6.95E-01
<i>TET2</i>	2/64 (3.1)	0/54 (0.0)	4.99E-01	6.95E-01
<i>FOXO1</i>	3/64 (4.7)	1/54 (1.9)	6.24E-01	8.32E-01
<i>NOTCH2</i>	3/64 (4.7)	1/54 (1.9)	6.24E-01	8.32E-01
<i>BCOR</i>	2/64 (3.1)	1/54 (1.9)	1.00E+00	1.00E+00
<i>CDKN2A</i>	2/64 (3.1)	1/54 (1.9)	1.00E+00	1.00E+00
<i>ALK</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00
<i>ARID2</i>	1/64 (1.6)	1/54 (1.9)	1.00E+00	1.00E+00
<i>ATM</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00
<i>BIRC3</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00
<i>DDX3X</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00
<i>EZH2</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00
<i>FAS</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00
<i>GATA3</i>	1/64 (1.6)	1/54 (1.9)	1.00E+00	1.00E+00
<i>ID3</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00
<i>MFHAS1</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00
<i>RHOA</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00
<i>SF3B1</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00
<i>TRAF2</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00
<i>ZAP70</i>	1/64 (1.6)	0/54 (0.0)	1.00E+00	1.00E+00

Bold type indicates statistical significance at $P < 0.05$.

FDR: False discovery rate, n: Number of mutation-positive samples, N: Total number of samples,
PCNSL: Primary central nervous system lymphoma, TIS: Tumor tissue

Table S9. Comparative overview of liquid biopsy-based molecular diagnostics for PCNSL across two studies

	Mutter et al.	Li et al.
Sample Type	TIS, CSF and PLA	TIS, CSF and PLA
Liquid Sample Size	CNSL_CSF: 46 CNSL_PLA: 52 Non-CNSL_CSF: 16 Non-CNSL_PLA: 16	PCNSL_CSF: 123 PCNSL_PLA: 76 Non-PCNSL_CSF: 57 PCNSL_TIS-CSF-PLA: 61 triplets Suspected PCNSL_CSF: 42
Sequencing Method	CAPP-Seq	Targeted panel-Seq
Panel Composition	26 features: -Total mutations -MYD88/CD79B gene -Cluster assignment (×6) -Top 3 cluster-defining region counts	N8-RF-PCNSL: <i>MYD88</i> , <i>PIM1</i> , <i>CD79B</i> , <i>GNA13</i> , <i>IRF4</i> , <i>DTX1</i> , <i>KMT2D</i> , <i>B2M</i>
Diagnostic Sensitivity	CSF: 59% PLA: 25%	CSF (BJTTH): 79% CSF (BJFTH): 89% CSF (MSKCC): 43%
Diagnostic Specificity	CSF: 100% PLA: 100%	CSF: 100% CSF (BJFTH): 100% CSF (MSKCC): 89%
Diagnostic Sensitivity in <i>MYD88</i>^{wt} Patients	N/A	CSF: 35% CSF (BJFTH): 67% CSF (MSKCC): 19%
Diagnostic Specificity in <i>MYD88</i>^{wt} Patients	N/A	CSF: 100% CSF (BJFTH): 100% CSF (MSKCC): 93%
Biopsy-infeasible real-world cohort	N/A	Over 60% of patients were molecularly diagnosed as PCNSL, who subsequently demonstrated favorable radiographic responses.

BJFTH: Beijing Fengtai Hospital, BJTTH: Beijing Tiantan Hospital, CNSL: Central nervous system lymphoma, CSF: Cerebrospinal fluid, MSKCC: Memorial Sloan Kettering Cancer Center, N/A: Not applicable, PCNSL: Primary central nervous system lymphoma, PLA: Plasma, RF: Random forest, TIS: Tumor tissue, wt: Wild type.

-Supplementary Figures

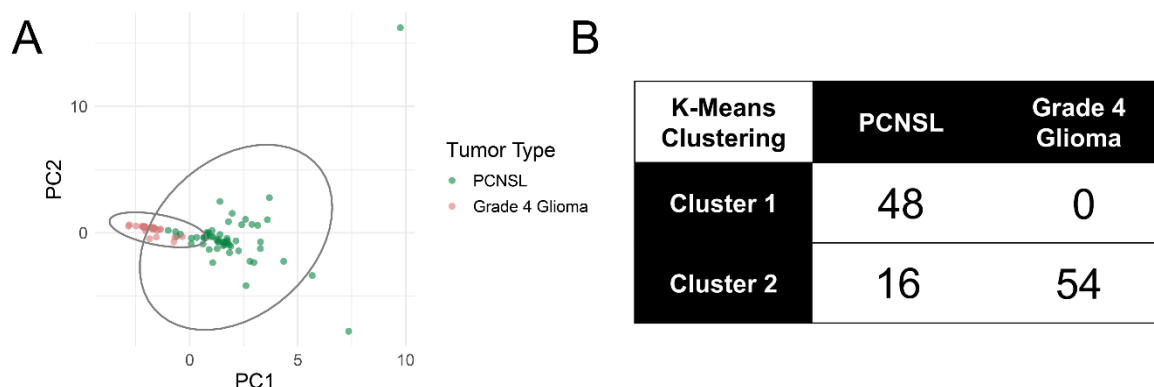


Figure S1. Unsupervised clustering analysis differentiating patients with PCNSL from grade 4 glioma based on genetic mutation profiles detected by 112 hotspot mutation probes. (A) PCA plot illustrating the distribution of PCNSL (green dots) and grade 4 glioma (red dots) based on the two principal components (PC1 and PC2). Each dot represents an individual tumor sample. Ellipses represent the 95% confidence intervals for each tumor type, highlighting the degree of separation and clustering patterns within each tumor type. (B) A contingency table presenting the result of K-means clustering applied to distinguish between PCNSL and grade 4 glioma. The clustering assignments are compared to the actual tumor types, allowing evaluation of the concordance between unsupervised clustering and pathological diagnosis.

PCA: principal component analysis, PCNSL: primary central nervous system lymphoma

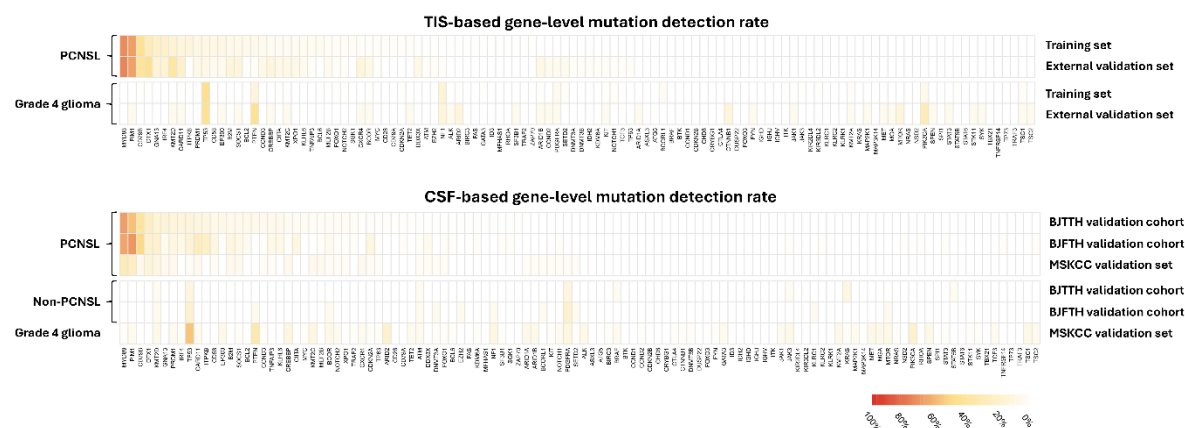


Figure S2. Distributions of gene-level mutation detection rates in PCNSL, grade 4 glioma and non-PCNSL across analytical datasets.

BJFTH: Beijing Fengtai Hospital, BJTTH: Beijing Tiantan Hospital, CSF: cerebrospinal fluid, MSKCC: Memorial Sloan Kettering Cancer Center, PCNSL: primary central nervous system lymphoma, TIS: tumor tissue

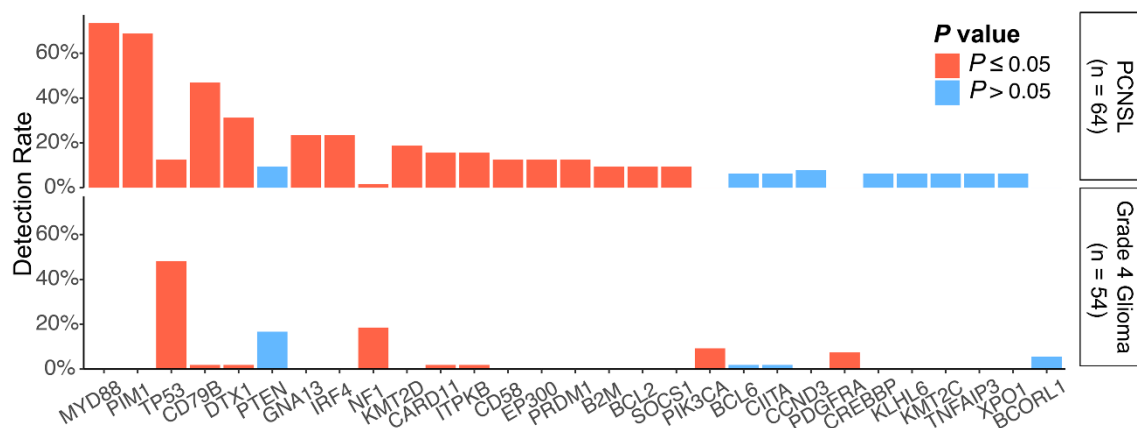


Figure S3. Gene-level mutation detection rates in PCNSL vs. grade 4 glioma. Bar plot illustrating the detection rates of genes with differential mutation frequencies between PCNSL and grade 4 glioma. Genes with statistics of $P \leq 0.05$ in mutation rates between the two groups (i.e., DMGs) are shown in red, whereas genes without significant differences are shown in blue.

DMGs: differentially mutated genes, PCNSL: primary central nervous system lymphoma, vs.: versus

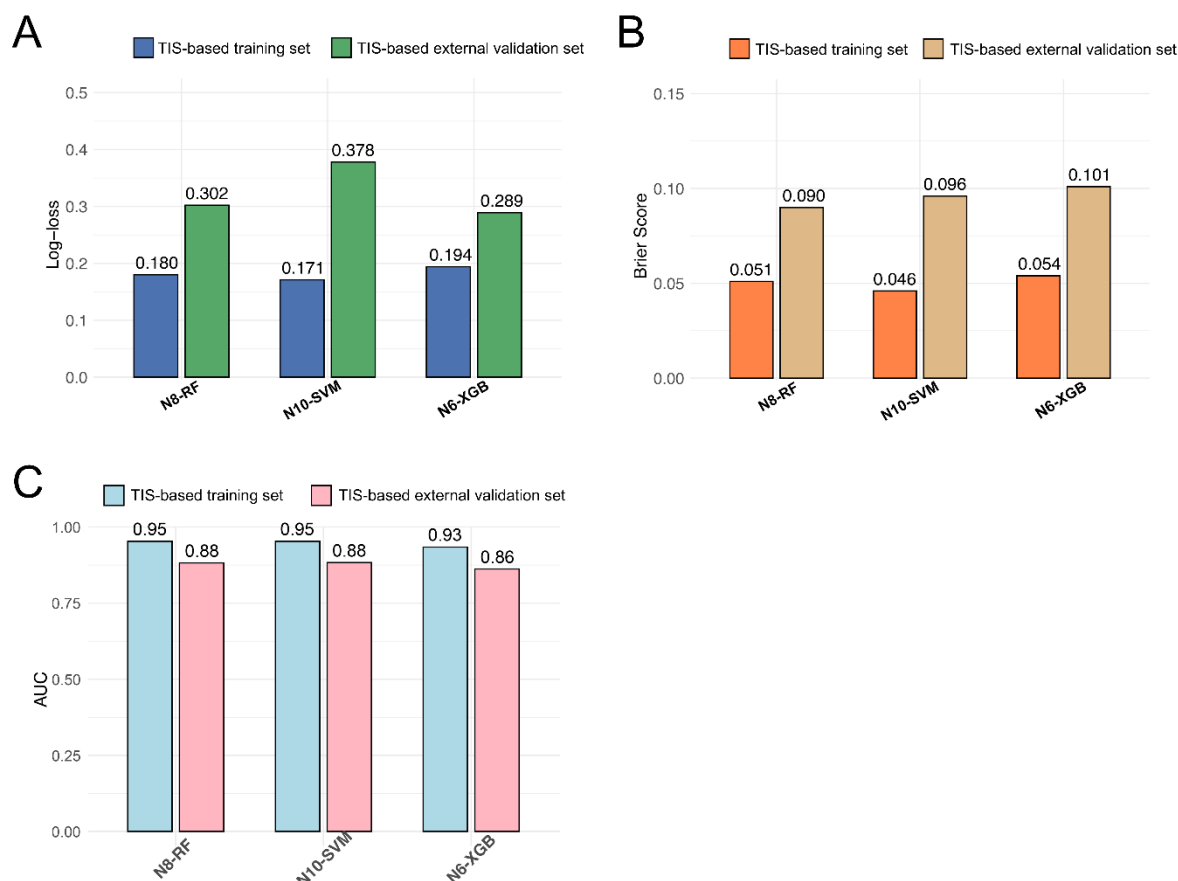


Figure S4. Reliability assessment of three ML-generated panels in differentiating patients with PCNSL from grade 4 glioma in the TIS-based training set and external validation set. (A) Bar plot showing the Log-loss values of the three ML-generated panels in both the TIS-based training set (blue bars) and the TIS-based external validation set (green bars). Log-loss imposes heavy penalties on confident but incorrect predictions, making it particularly sensitive to poorly calibrated or extreme probability estimates. Lower values indicate better predictive performance. (B) Bar plot showing the Brier scores of the three ML-generated panels in both the TIS-based training set (orange bars) and the TIS-based external validation set (brown bars). Brier score reflects the mean squared difference between predicted probabilities and observed outcomes. Brier score is strongly associated with probability calibration. Its values range from 0 to 1, with lower scores indicating better calibrated and more accurate probability estimates. (C) Bar plot comparing the AUC values of the three ML-generated panels between the TIS-based training set (blue bars) and the TIS-based external validation set (pink bars). AUC: 0.95 vs. 0.88 for N8-RF, $P = 0.21$; AUC: 0.95 vs. 0.88 for N10-SVM, $P = 0.26$; AUC: 0.93 vs. 0.86 for N6-XGB, $P = 0.21$.

AUC: area under the receiver operating characteristic curve, ML: machine learning, RF: random forest, SVM: support vector machine, TIS: tumor tissue, vs.: versus, XGB: XGboost

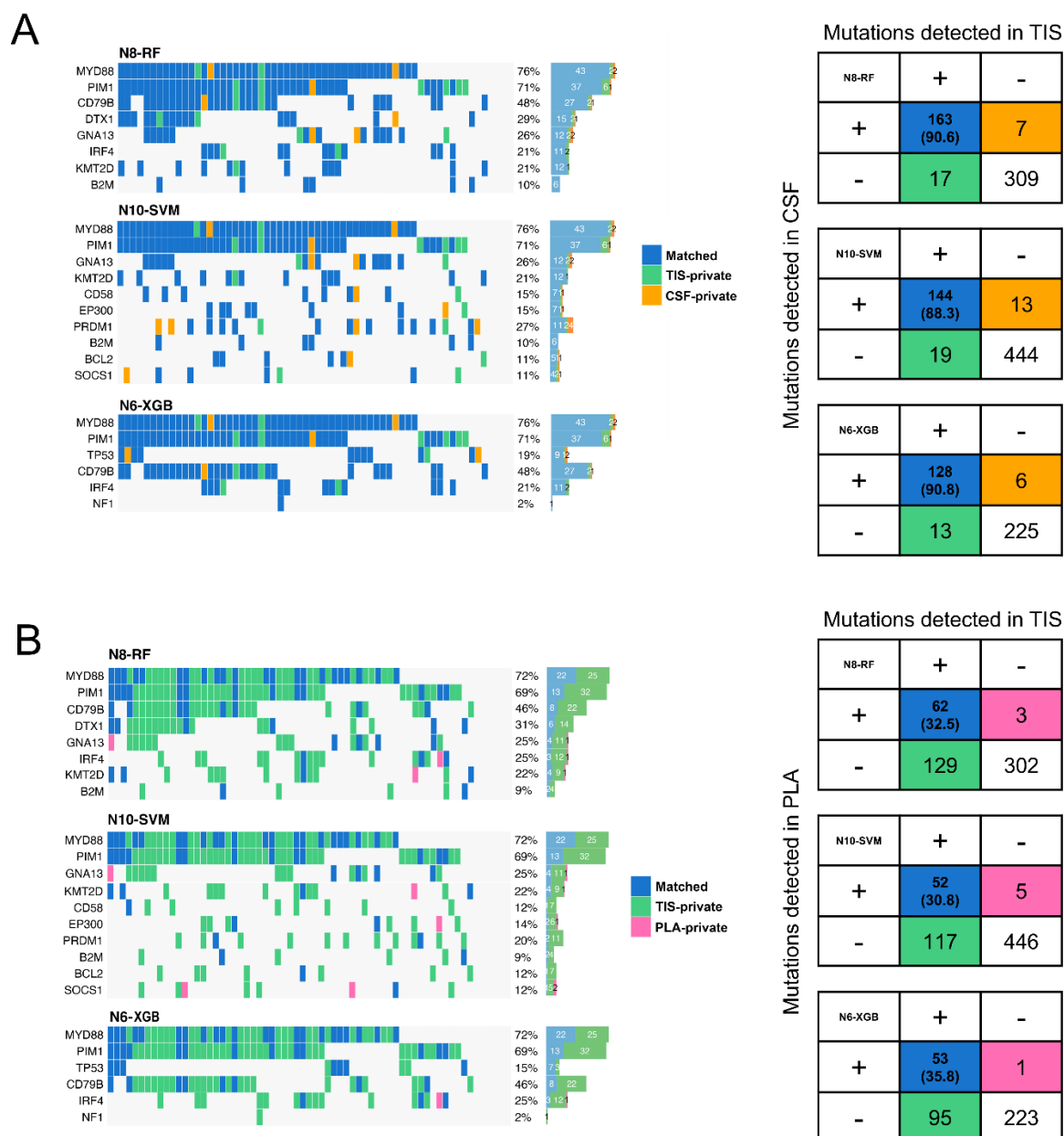


Figure S5. Concordance of mutational profiles between tissue and liquid specimen-informed sequencing in the BJTTH cohort. (A-B) waterfall plots illustrating genetic biomarker concordance for the N8-RF-PCNSL, N10-SVM-PCNSL and N6-XGB-PCNSL between TIS and paired liquid samples. Blue blocks represent the mutated genes simultaneously detected in the TIS and the corresponding CSF/PLA. Green and orange/pink blocks represent the mutated genes uniquely identified in TIS or CSF/PLA, respectively. Contingency tables display mutation detection status across TIS and CSF/PLA pairs, with the horizontal axis representing TIS results and the vertical axis representing CSF/PLA results; "+" denotes a positive detection, and "-" denotes a negative detection. Sensitivity (%) was inferred based on gene mutations detected in TIS, which were either identified in CSF or PLA ctDNA

CSF: cerebrospinal fluid, PLA: plasma, RF: random forest, SVM: support vector machine, TIS: tumor tissue, XGB: XGboost

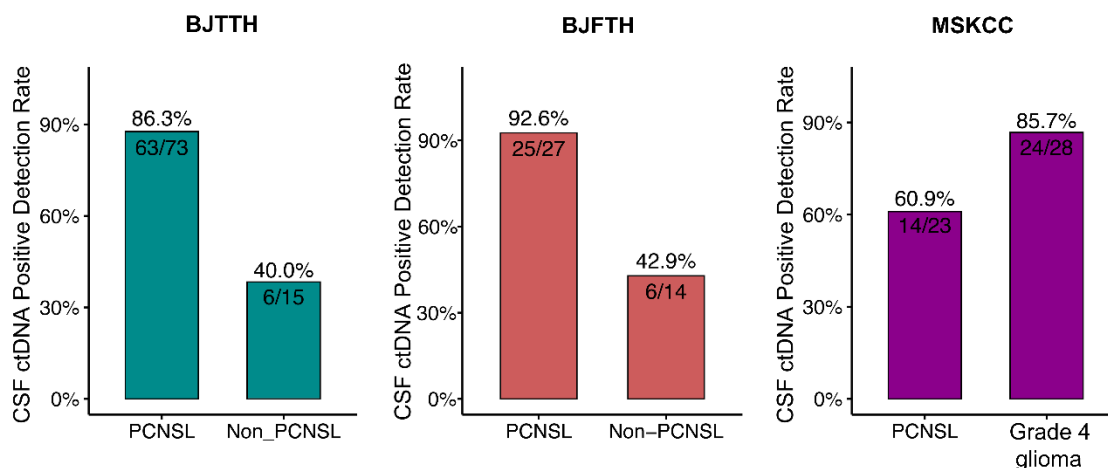


Figure S6. Positive detection rates of CSF ctDNA across the three validation sets. CSF ctDNA was detected using targeted sequencing data from the 112-hotspot-probe panel applied in the BJTTH and BJFTH cohorts and the MSKCC-specific panels used in the publicly available MSKCC dataset.

BJFTH: Beijing Fengtai Hospital, BJTTH: Beijing Tiantan Hospital, CSF ctDNA: cerebrospinal fluid circulating tumor DNA, MSKCC: Memorial Sloan Kettering Cancer Center, PCNSL: primary central nervous system lymphoma

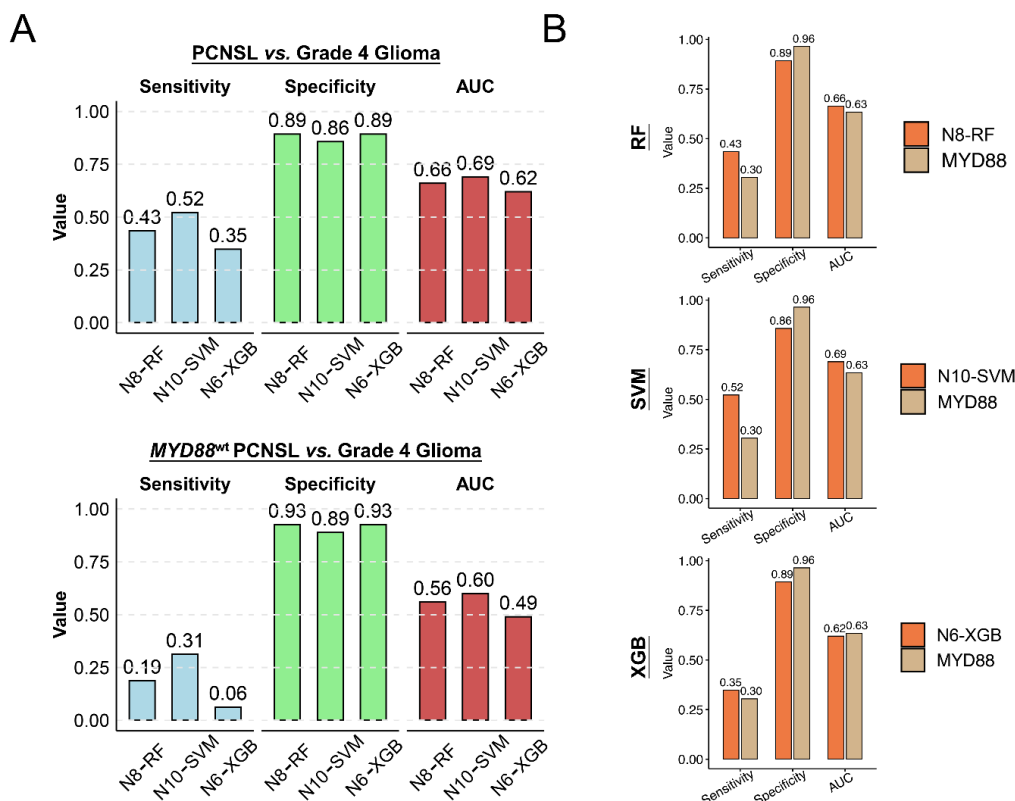


Figure S7. Feasibility of three ML-generated panels for diagnosing PCNSL using CSF ctDNA profiling in CSF-based MSKCC validation set. (A) Diagnostic performance of ML-generated panels in differentiating patients with PCNSL/*MYD88*^{wt} PCNSL from grade 4 glioma through CSF ctDNA detection. Bar charts display and compare sensitivity, specificity and AUCs across the three established panels. MSKCC: AUC_{RF} = 0.66 (95% CI: 0.55-0.78), AUC_{SVM} = 0.69 (95% CI: 0.57-0.81), AUC_{XGB} = 0.62 (95% CI: 0.51-0.74); MSKCC (*MYD88*^{wt} subgroup): AUC_{RF} = 0.56 (95% CI: 0.45-0.67), AUC_{SVM} = 0.60 (95% CI: 0.47-0.73), AUC_{XGB} = 0.49 (95% CI: 0.42-0.57). (B) Comparative diagnostic performance of the N8-RF, N10-SVM and N6-XGB panels vs. *MYD88*^{mut} mono-assay, respectively, in differentiating patients with PCNSL from grade 4 glioma through CSF ctDNA detection.

AUC: area under the receiver operating characteristic curve, CI: confidence interval, CSF ctDNA: cerebrospinal fluid circulating tumor DNA, ML: machine learning, MSKCC: Memorial Sloan Kettering Cancer Center, PCNSL: primary central nervous system lymphoma, RF: random forest, SVM: support vector machine, vs.: versus, wt: wild type, XGB: XGboost

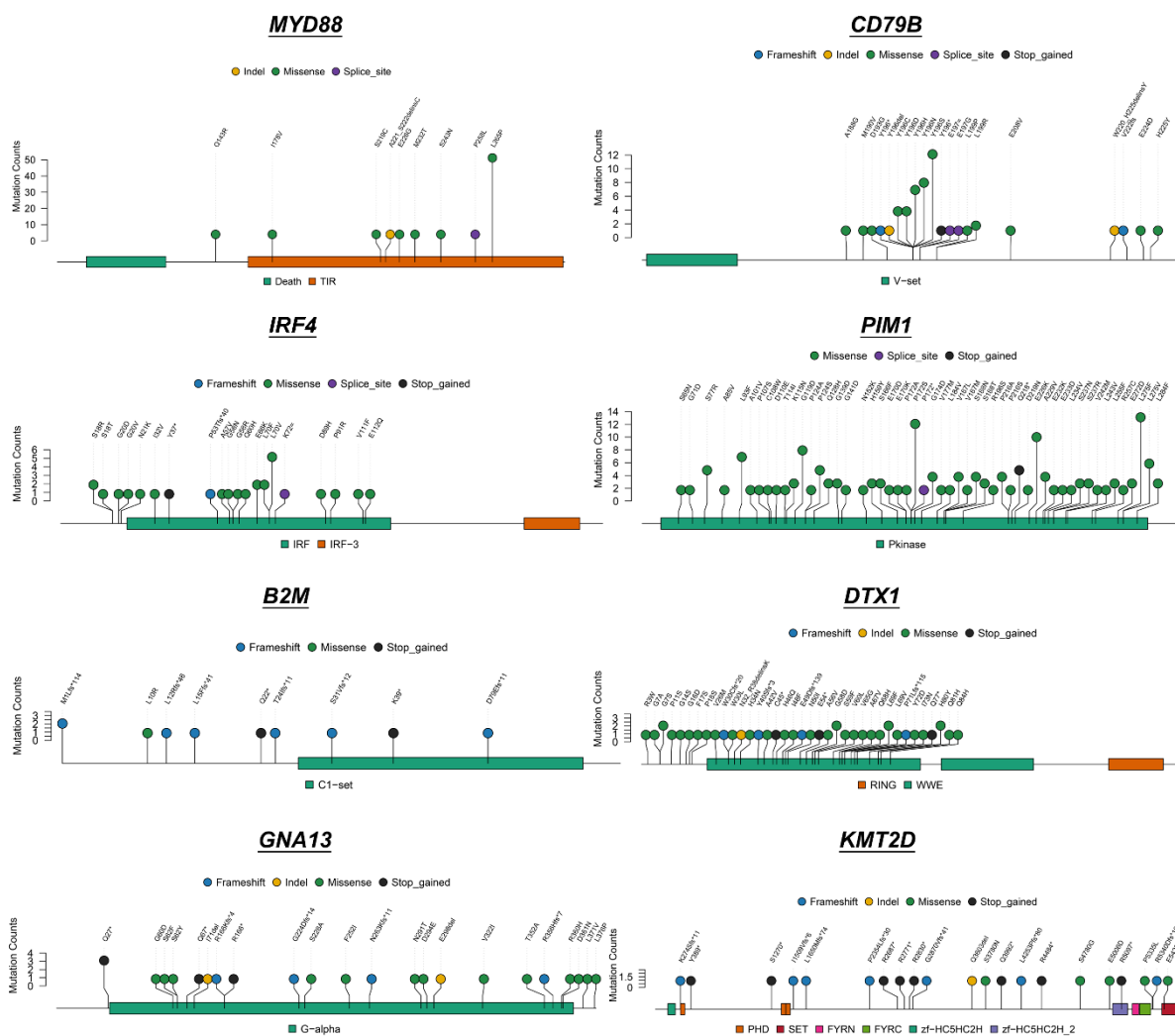


Figure S8. Mutational loci of the eight genes included in the N8-RF-PCNSL. Mutational loci of the eight genes included in the panel were presented using both the CSF-based BJTTH and BJFTH validation cohorts. Two genes, *MYD88* and *CD79B*, had significantly recurrent mutational hotspots. Specifically, *MYD88* mutations predominantly occurred at the L265P (52/70, 74.3%) site; and *CD79B* at the Y196S (12/47, 25.5%), Y196N (8/47, 17.0%) and Y196H (7/47, 14.9%) sites. The remaining genes displayed dispersed mutational loci without specific hotspots. Each mutation type is represented by a color, and the mutation count of each variant is indicated by the height of the corresponding stem.

BJFTH: Beijing Fengtai Hospital, BJTTH: Beijing Tiantan Hospital, CSF: cerebrospinal fluid, PCNSL: primary central nervous system lymphoma, RF: random forest