CrossMark

ORIGINAL PAPER

# Genomic characterization of primary central nervous system lymphoma

Kazutaka Fukumura[1] · Masahito Kawazu[2] · Shinya Kojima[1] · Toshihide Ueno[1] ·
Eirin Sai[2] · Manabu Soda[1] · Hiroki Ueda[3] · Takahiko Yasuda[1] · Hiroyuki Yamaguchi[1] ·
Jeunghun Lee[4,6] · Yukiko Shishido-Hara[5] · Atsushi Sasaki[7] · Mitsuaki Shirahata[8] ·
Kazuhiko Mishima[8] · Koichi Ichimura[9] · Akitake Mukasa[10] · Yoshitaka Narita[11] ·
Nobuhito Saito[10] · Hiroyuki Aburatani[3] · Ryo Nishikawa[8] · Motoo Nagane[4] ·
Hiroyuki Mano[1,12]

**Abstract** Primary central nervous system lymphoma (PCNSL) is a rare malignancy confined to the central nervous system (CNS), and majority of PCNSL is pathologically classified as diffuse large B-cell lymphoma (DLBCL). We have now performed whole-exome sequencing for 41 tumor tissues of DLBCL-type PCNSL and paired normal specimens and also RNA-sequencing for 30 tumors, revealing a very high frequency of nonsynonymous somatic mutations in *PIM1* (100 %), *BTG2* (92.7 %), and *MYD88* (85.4 %). Many genes in the NF-κB pathway are concurrently mutated within the same tumors. Further, focal deletion or somatic mutations in the *HLA* genes are associated with poor prognosis. Copy number amplification and overexpression of genes at chromosome 7q35 were both found to predict short progression-free survival as well. Oncogenic mutations in *GRB2* were also detected, the effects of which in cultured cells were attenuated by inhibitors of the downstream kinases MAP2K1 and MAP2K2. Individuals with tumors positive for *MYD88* mutations also harbored the same mutations at a low frequency in peripheral blood mononuclear cells, suggesting that *MYD88* mutation-positive precancerous cells originate outside of the CNS and develop into lymphoma after additional genetic hits that confer adaptation to the CNS environment.

**Keywords** Primary central nervous system lymphoma · Oncogene · Genomics · MYD88

✉ Hiroyuki Mano
hmano@m.u-tokyo.ac.jp

1 Department of Cellular Signaling, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

2 Department of Medical Genomics, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan

3 Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 153-8904, Japan

4 Department of Neurosurgery, School of Medicine, Kyorin University Faculty of Medicine, Tokyo 181-8611, Japan

5 Department of Pathology, School of Medicine, Kyorin University Faculty of Medicine, Tokyo 181-8611, Japan

6 Department of Neurosurgery, Sassa General Hospital, Tokyo 188-0011, Japan

7 Department of Pathology, Saitama International Medical Center, Saitama Medical University, Saitama 350-1298, Japan

8 Department of Neuro-Oncology/Neurosurgery, Saitama International Medical Center, Saitama Medical University, Saitama 350-1298, Japan

9 Division of Brain Tumor Translational Research, National Cancer Center Research Institute, Tokyo 104-0045, Japan

10 Department of Neurosurgery, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan

11 Department of Neurosurgery and Neuro-Oncology, National Cancer Center Hospital, Tokyo 104-0045, Japan

12 Strategic Basic Research Program, Japan Science and Technology Agency, Saitama 332-0012, Japan

## Introduction

Primary central nervous system lymphoma (PCNSL) is a rare form of non-Hodgkin lymphoma that arises in and is confined to the central nervous system (CNS), accounting for 2 to 3 % of all CNS neoplasms [1, 13]. Incidence of PCNSL is recently increasing in the elderly [29]. Majority of PCNSL in immunocompetent individuals is pathologically classified as diffuse large B-cell lymphoma (DLBCL), whereas other types of lymphoma such as T-cell lymphoma are frequently observed in immunocompromised patients.

Standard chemotherapeutic regimens for systemic DLBCL show little efficacy in PCNSL, likely because of inefficient drug delivery across the blood–brain barrier [13, 20]. High-dose methotrexate with or without whole-brain irradiation is the most effective treatment for PCNSL, but it may be associated with neurotoxicity, and eventual relapse of lymphoma is frequently observed. The prognosis of PCNSL is thus poor with a median overall survival of 1 to 4 years, and it becomes even worse for immunocompromised cases [1, 25, 26].

The rarity of the disease and the difficulty of obtaining intracranial specimens have hindered understanding of the pathophysiology of PCNSL. The observed overexpression of *BCL6* and aberrant somatic hypermutation (aSHM) of many genes, together with expression of immunoglobulin M at the cell surface, have suggested that PCNSL cells may be arrested at the stage of terminal B-cell differentiation [6]. Recent chromosomal copy number analysis has successfully identified recurrent copy number variations in PCNSL, and exome sequencing has revealed frequent mutations in *MYD88*, *CD79B*, *ODZ4*, or *TBL1XR1* genes [3, 8, 28]. The landscape of genomic alterations in PCNSL has remained elusive, however, as a result of the small cohort size or lack of paired normal samples.

Here we conducted whole-exome sequencing on 41 DLBCL-type PCNSL tumors as well as paired normal samples, and further carried out RNA-sequencing (RNA-seq) on 30 tumor specimens. From the large dataset, we have successfully extracted the target genes of aSHM and the candidate driver genes in the carcinogenesis for PCNSL.

## Materials and methods

### Clinical specimens

Surgically resected tumor and paired peripheral blood mononuclear cell (PBMNC) specimens from PCNSL patients were analyzed with written informed consent. This project was approved by the institutional ethics committees of The University of Tokyo, Kyorin University, Saitama Medical University, National Cancer Center Research Institute, and National Cancer Center Hospital.

### Whole-exome sequencing

Genomic DNA was isolated from each sample and subjected to enrichment of exonic fragments with the use of a SureSelect Human All Exon Kit (Agilent). Massively parallel sequencing of the isolated fragments was performed with the HiSeq2000 platform (Illumina) according to the paired-end option. From the large data sets, we selected only sequence reads with a $Q$ value of $\geq 20$ at each base, and we further extracted unique reads that were subsequently mapped to the reference human genome sequence (hg19) with the bowtie 2 algorithm (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). Mismatches were discarded if (1) a given read contained $\geq 3$ independent mismatches, (2) they were already present in the "1000 genomes" database (http://www.1000genomes.org) or in normal human genome alterations of our in-house database, or (3) they were supported by only one strand of the genome. Somatic mutations were called by MuTect (http://www.broadinstitute.org/cancer/cga/mutect) and SomaticIndelDetector (http://www.broadinstitute.org/cancer/cga/node/87). Gene mutations were annotated by SnpEff (http://snpeff.sourceforge.net).

For the validation of the HiSeq2000 data, genomic regions corresponding to a subset of mutations were amplified with the primers designed with Ion AmpliSeq Designer (https://www.ampliseq.com/browse.action) and subjected to library preparation with Ion AmpliSeq Library Kit 2.0 followed by nucleotide sequencing with another next-generation sequencer (NGS), Ion PGM (both from Thermo Fisher Scientific).

### RNA-seq

Complementary DNAs were prepared from tumor tissue with the use of a NEBNext Ultra Directional RNA Library Prep Kit (New England BioLabs) and were subjected to NGS sequencing for 100 bp from both ends. The expression level of each gene was calculated with the use of the DESeq2 algorithm (http://bioconductor.org/packages/release/bioc/html/DESeq2.html) with the VST transformation, and gene fusions were detected by the deFuse pipeline (https://bitbucket.org/dranew/defuse). Gene sets for GSEA analysis were obtained from the GSEA website (http://www.broadinstitute.org/gsea/index.jsp). Differentially expressed genes between *MYD88* mutation-positive and -negative tumors were isolated by the Exact test in the R package (http://rpackages.ianhowson.com/bioc/edgeR/man/exactTest.html).

### Functional studies of cancer-related genes

Wild-type cDNA was obtained by polymerase chain reaction (PCR), and mutant forms were then generated with the

use of a QuickChange site-directed mutagenesis kit (Agilent). All cDNAs were verified by Sanger sequencing and then ligated into the pMXS retroviral vector (Cell Biolabs). The recombinant vectors were introduced together with an ecotropic packaging plasmid (Takara Bio) into HEK293T cells (American Type Culture Collection) by transfection to obtain infectious virus particles. For the focus formation assay, 3T3 cells (American Type Culture Collection) were infected with ecotropic recombinant retroviruses and cultured for 2 to 3 weeks in Dulbecco's modified Eagle's medium-F12 supplemented with 5 % calf serum (both from Invitrogen). Inhibitors of MAP2K1/2 were obtained from Selleck Chemicals.

For the analysis of IKBKB, a full length cDNA for IKBKB or IKBKB(V203I) was inserted into the pcDNA3 expression vector (Life Technologies), and the resulting plasmids were introduced into HEK293 cells by transfection. The cells were lysed for analysis at 12 h after transfection.

Antibodies for immunoblot analysis included sc-8014 for IKBKB and sc-6216 for LMNB1/2 (both from Santa Cruz Biotechnology) as well as #9242 for NFKBIA, #9246 for phosphorylated NFKBIA, #8242 for RELA, #3036 for phosphorylated RELA, and #4970 for ACTB (all from Cell Signaling Technology).

## Digital PCR analysis

Genomic DNA (25 ng per sample) was applied to the QX200 system (Bio-Rad) and subjected to PCR. Primers for PCR (forward and reverse, respectively) included 5′-CCTTGGCTTGCAGGT-3′ and 5′-TCTTTCTTCATTG CCTTGT-3′ for *MYD88*, 5′-CGGAGGGTCAGGGG-3′ and 5′-TCCATCACATTGCCACT-3′ for *MTMR8*, 5′-TCGG TGTGCTAGGTATG-3′ and 5′-GAGGTGAGACAAGGA GAG-3′ for *COL4A6*, 5′-AACATTTTGCAGTGCTG A-3′ and 5′-CTGAAGGATAGTTTCACCTG-3′ for *BEND2*, 5′-TGGTCAGAGAAGGAATAATG-3′ and 5′-ATAAC ATGCCTTAGGAGGG-3′ for *OGT*, 5′-GCCAATGTCCCC AATG-3′ and 5′-TTTCCGCCTCCCGA-3′ for *ARSE*, 5′-AGGATTGATATTAAAGGTAATTAAAC-3′ and 5′-ATG TTAAGCAACCAGTCTTA-3′ for *TBC1D8B*, and 5′-GAG AGCACCAAACTGAAG-3′ and 5′-CCACGAAGGGAAG GAA-3′ for *CCR9*. Taqman probes used to differentiate point mutations were 5′-TGGGGATCAGTCGCTT-3′ and 5′-TGGGGATCGGTCGC-3′ for *MYD88*, 5′-AAGTTA CCCATCACTAGCC-3′ and 5′-TCACGAGCCTGGGT T-3′ for *MTMR8*, 5′-TGGGATTCTTGACTGAACA-3′ and 5′-TGGGATTCTTGACTTAACATC-3′ for *COL4A6*, 5′-ACAAGAAATCTCAACCATGTT-3′ and 5′-ACAAGA AATCTCAAGCATGTT-3′ for *BEND2*, 5′-AAGGGAAA AAAAAAAGATTGGG-3′ and 5′-AAAGGGGAAAAAA AAAGATTGG-3′ for *OGT*, 5′-TCGTCCGCCATCAGA-3′

and 5′-AGGTCGTCCACCATCA-3′ for *ARSE*, 5′-ATAA AAAGTTTATTTCATCTAGGGAA-3′ and 5′-ACAAT AAAAAATTTATTTCATCTAGGG-3′ for *TBC1D8B*, and 5′-TGACCCTGAAGGTCATTC-3′ and 5′-TGACCCTGA AGTTCATTCT-3′ for *CCR9* (for wild-type and mutant sequences, respectively).

For measuring the relative abundance of mRNA for *CXCL2*, *CXCL3*, *MYC* or *NFKBIA* to that of *GAPDH*, RNA was subjected to real-time reverse transcription (RT)-PCR analysis with the Applied Biosystems 7500 platform. Primers used for PCR were 5′-CC ACACTCAAGAATGGGCAGAAAG-3′ and 5′-TCCT TCAGGAACAGCCACCAATAA-3′ for *CXCL2*, 5′-GCCAC ACTCAAGAATGGGAAGAAA-3′ and 5′-ATTTTCA GCTCTGGTAAGGGCAGG-3′ for *CXCL3*, 5′-GGCA AAAGGTCAGAGTCTGGATCA-3′ and 5′-GCGTAGTTG TGCTGATGTGTGGAG-3′ for *MYC*, 5′-CCATCATCCA TGAAGAAAGGCAC-3′ and 5′-ATCACAGCCAAGTGG AGTGGAGTC-3′ for *NFKBIA*, and 5′-GAATTTGGCT ACAGCAACAGGGTG-3′ and 5′-TTCAAGGGGTCTACA TGGCAACTG-3′ for *GAPDH*.

## Hierarchical clustering

Information on somatic mutations for 40 systemic DLBCL cases was obtained from a previous study [22]. Our PCNSL samples and the DLBCL samples were clustered according to the mutated gene profile with the use of Ward's method in the R package.

## Results

### Mutation burden of the PCNSL genome

We obtained surgically resected lymphoma tissue as well as paired PBMNCs from 41 individuals with PCNSL. Pathological analysis of the tumors revealed that they were all classified as DLBCL. Therefore, according to the current WHO classification, the tumors are primary DLBCL of the central nervous system. All but two cases were treatment-naïve, and none of the patients were positive for Epstein–Barr virus infection (Supplementary Table 1). Also, human immunodeficiency virus infection was negative in all cases but nine who were not tested. Immunohistochemical staining of the lymphoma specimens with antibodies to CD10/BCL6/MUM1 revealed 7 cases (#1–#7) to be of the germinal center B-cell-like (GCB) subtype and 32 cases (#8–#39) to be of the non-GCB subtype [9]. As expected, progression-free survival (PFS) for treatment-naïve individuals with non-GCB tumors was significantly shorter than that for those with GCB tumors ($P = 0.0266$, log-rank test) (Supplementary Fig. 1).

Genomic DNA isolated from tumors and PBMNCs was subjected to enrichment of exonic fragments and then to analysis with HiSeq2000 (Illumina). Sequence reads were obtained such that each base was examined at a mean depth of 158× and 81×, covering >98 % of the target regions, for tumor tissue and PBMNCs, respectively. Of 17,385 somatic mutations identified in the PCNSL specimens, 10,765 (62 %) were nonsynonymous substitutions, which was ~2.5 times the number of synonymous mutations (Fig. 1a). A total of 87 mutations were picked up and independently validated with an NGS based on a different platform (Ion PGM), yielding a concordance of 100 % (Supplementary Table 2). The tumor content of each PCNSL sample was calculated from the exome data with the use of the Karkinos computational pipeline [10], yielding a mean value of 69.0 % (range of 20.0 to 98.0 %) (Supplementary Table 1).

The mean number of nonsynonymous substitutions was 3.2 per megabase (Mb), similar to the value previously determined for systemic DLBCL [17, 22]. The mutation burden varied substantially among the specimens, however (Fig. 1b). The number of insertions/deletions (indels), for instance, was exceptionally high in tumor #6 at 2.2/Mb (mean value for all tumors was 0.2/Mb), whereas tumor #13 was enriched in nonsynonymous substitutions (48.1/Mb). These outlier cases harbored mutations in *POLE* or in genes related to the mismatch-repair system. The number of somatic mutations per sample did not differ significantly between GCB and non-GCB subtypes (Supplementary Fig. 2).

The predominant nucleotide substitution was a C>T transition (57.8 %), especially in the context of the sequence GCG (Supplementary Fig. 3). Patients #40 or #41 had received prior whole-brain irradiation or stereotactic
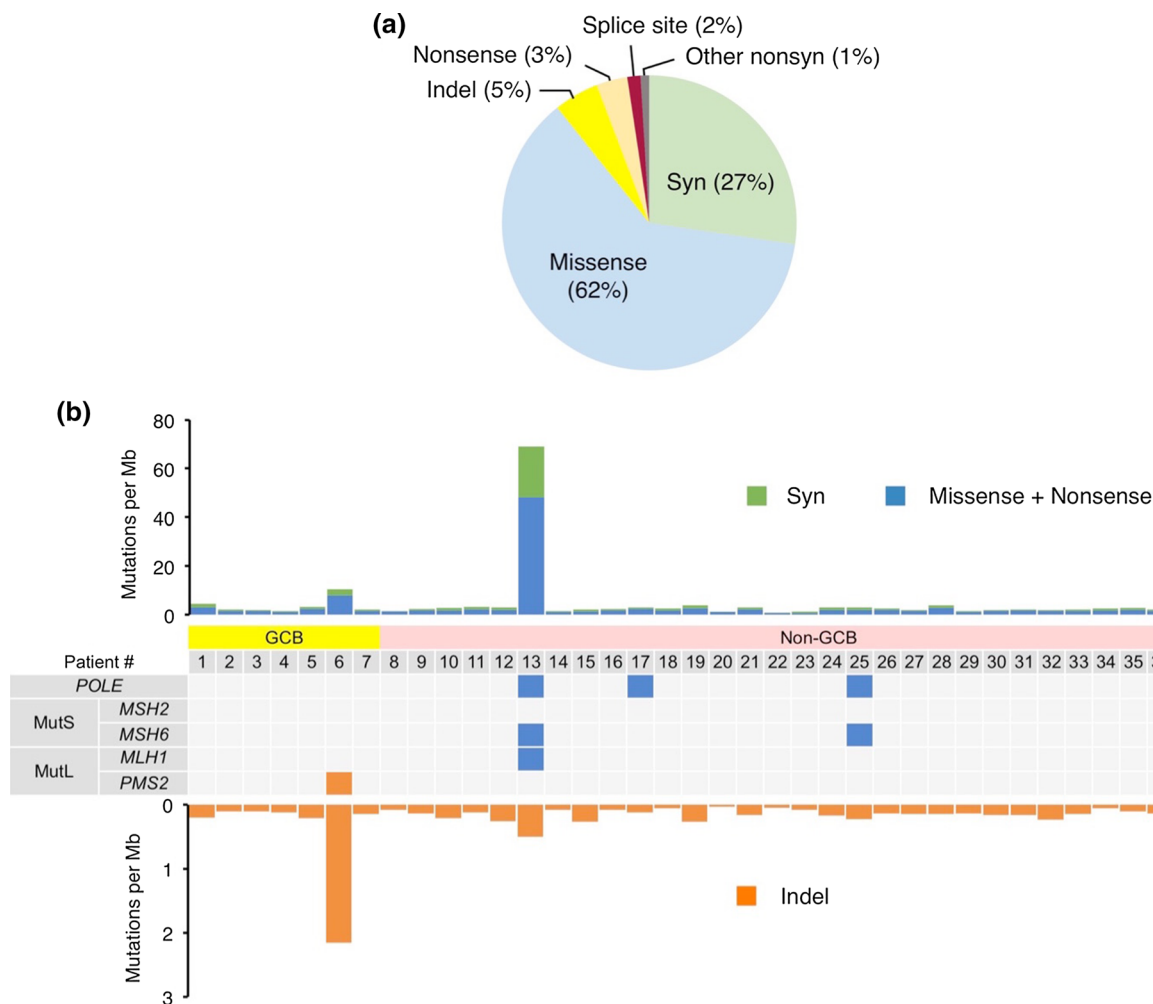


**Fig. 1** Mutation load of PCNSL. **a** Pie chart showing the percentages of different types of somatic mutation in PCNSL. *Syn* synonymous, *nonsyn* nonsynonymous. **b** Frequency of synonymous or nonsynonymous substitutions and of indels in each PCNSL specimen. Patient number, Hans's classification (GCB, non-GCB), and the presence of mutations in *POLE* or in genes for the MutS (*MSH2*, *MSH6*) or MutL (*MLH1*, *PMS2*) complexes of the mismatch-repair system are indicated in the *middle* panel. *ND* not determined

radiosurgery, respectively, and tumor #41 manifested an increased frequency of A>G transitions, consistent with radiation-induced nucleotide mutations (Supplementary Fig. 3).

## Profile of mutated genes in PCNSL

As in DLBCL, multiple mutations of many genes were observed in individual PCNSL tumors, suggestive of the operation of aSHM mediated by activation-induced cytidine deaminase. From our dataset, we predicted aSHM targets using the SHM indicator algorithm of Khodabakhshi et al. that takes into account (1) the mutation enrichment in the SHM hotspot motif "WRCY" where W indicates A or T, R indicates A or G, and Y indicates C or T, (2) the mutation ratio of C:G to A:T sites, and (3) the ratio of transition to transversion mutations [12]. We thus identified a total of seventy-six candidate genes of aSHM targets (SHM indicator <0.05) (Fig. 2, Supplementary Table 3). Strikingly, single or multiple somatic mutations in *PIM1* were identified in all 41 specimens irrespective of GCB/non-GCB subtypes, which is in contrast to the modest mutation rate (10–30 %) of *PIM1* in systemic DLBCL [12]. Similarly, *BTG2* was hypermutated in nearly all samples (92.7 %). On the other hand, frequent targets of aSHM in systemic DLBCL such as *RHOH* and *BCL6* were not affected in PCNSL (Supplementary Table 3).

MutSigCV analysis [16] identified 18 candidates for driver genes of carcinogenesis with a *q* value of <0.1, with three of these genes also being targets for aSHM (Supplementary Table 4). We additionally selected six genes with a *P* value of <0.05 that were mutated in ≥5 specimens as driver candidates (Supplementary Table 4). Somatic nonsynonymous mutations found in these 21 SHM-unrelated potential driver genes are shown in Fig. 3. Most of these genes were previously shown to harbor mutations in PCNSL and systemic DLBCL [17, 22, 31], but some of them are novel to PCNSL. Interestingly, none of the mutations in these 21 genes were significantly enriched in either the GCB or non-GCB subgroup.

Somatic mutations in *MYD88* were found frequently (75.6 %) with the p.L265P substitution being predominant (Fig. 3, Supplementary Fig. 4) [2, 14, 24], although such mutations were previously reported only in 10 to 20 % of systemic DLBCL (exclusively in non-GCB subtype) [17, 31]. Interestingly, genes in the NF-κB pathway are highly enriched in both the aSHM targets and the driver candidates, such as *PIM1*, *BTG2*, *CD44*, *XBP1*, *CD79B*, *MYD88*, and *NFKBIE*. Thus, the NF-κB pathway is often affected at multiple sites within the same tumors in PCNSL.

Further, mutual exclusivity of mutations in Fig. 3 was apparent for genes related to the immune reaction, to the cell cycle, or to epigenetic regulation (Supplementary Fig. 5). Of note, mutations within these three gene groups were not related to any clinicopathological features such
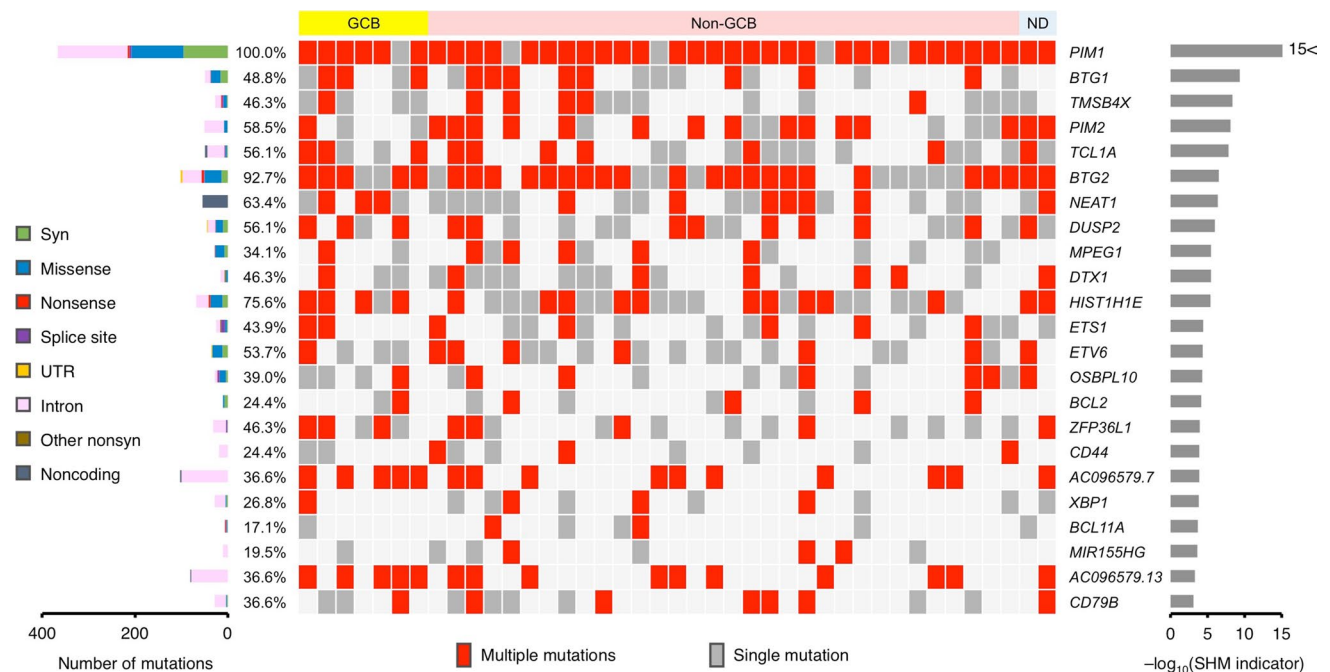


**Fig. 2** Targets of aberrant SHM in PCNSL. Twenty-three genes with the lowest SHM indicator (*right*) are shown with their mutation status in each patient. The histogram at the *left* shows the total number of mutations in each gene color-coded according to the mutation type.

Percentages indicate the proportion of tumors with mutations in the specified gene. Patients are sorted as in Fig. 1b. *UTR* untranslated region
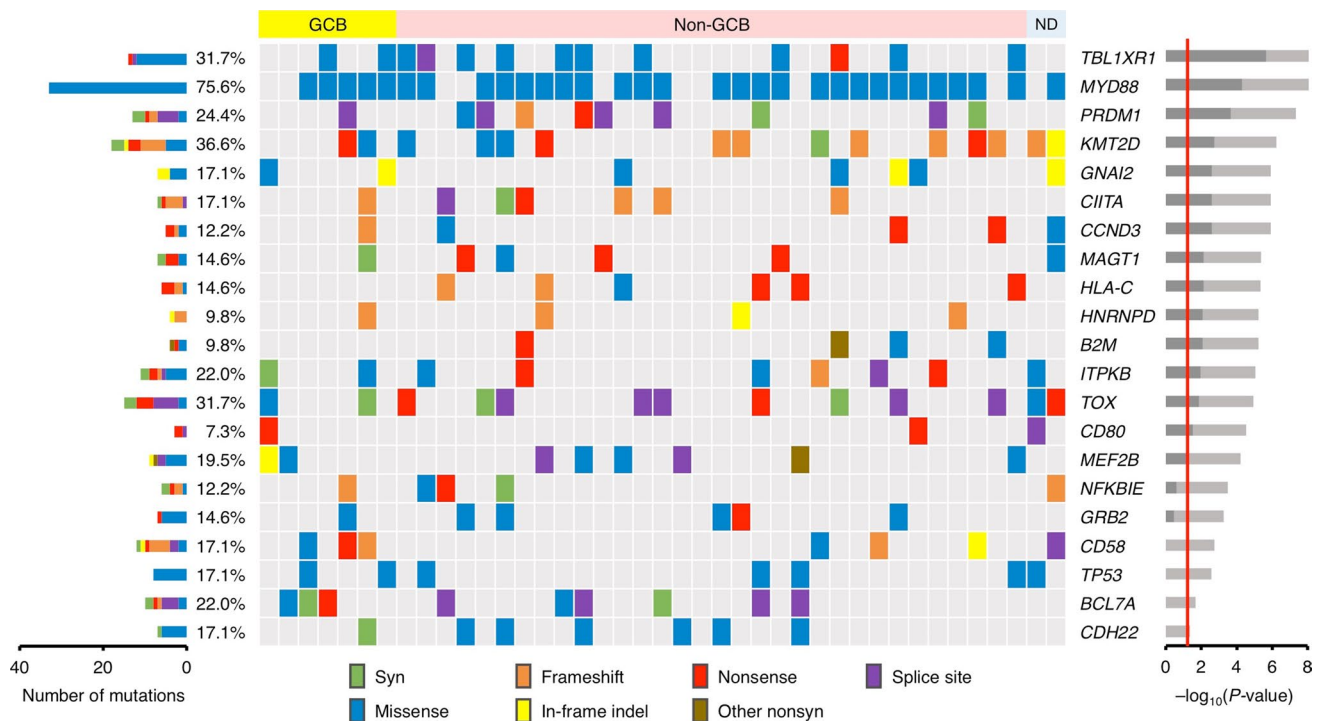
**Fig. 3** Candidates of driver genes in PCNSL. Mutation status of each patient is shown color-coded by the type of mutation for 21 candidate driver genes (SHM targets are not included). The *q* value and *P* value for the MutSigCV calculation are shown for each gene as *dark* and *light gray bars*, respectively (*right* panel). The histogram at the *left* and percentages are as in Fig. 2. Patients are sorted as in Fig. 1b

as age, sex, GCB/non-GCB subtypes or tumor recurrence (data not shown).

We then examined the relation of mutations of the driver genes to PFS, finding that alterations in *HLA-C* were associated with a shorter PFS compared with wild-type *HLA-C* ($P = 0.0262$, log-rank test) (Fig. 4a). Chromosome copy number for the tumors was inferred from the exome data, revealing a frequent gain in 1q, 7, 12, and 18q and a frequent loss of 6q (Fig. 4b). A focal deletion of 6p21-22 corresponding to the *HLA* locus was observed, consistent with previous data [8]. This latter finding, together with the linkage of *HLA-C* mutations to early relapse of PCNSL, suggests that impairment of HLA-dependent immune surveillance may play an important role in PCNSL development. Uniparental disomy of 6p and 9p was also frequently detected in PCNSL.

In our cohort, chromosome copy number alterations involving *TP53* was not associated with a shorter PFS ($P = 0.392$, log-rank test). The relationship between chromosome copy number changes with somatic mutations in the driver genes (Fig. 3) is demonstrated in Supplementary Fig. 6. Notably, copy number loss was frequently observed for *PRDM1* in addition to *HLA-C*, which is in line with the proposed tumor suppressive function of PRDM1 in lymphoma [18].

Of our PCNSL specimens, RNA was available for 30 samples. Complementary DNAs were prepared from

polyadenylated RNAs of each tumor, and analyzed with HiSeq2000, yielding a total of $29.1 \pm 5.5$ (mean $\pm$ SD) gigabases per tumor. Using this expression profile, we asked whether expression of certain gene sets was associated with disease progression. Gene set enrichment analysis (GSEA) [27] revealed that expression of the gene sets for primary immunodeficiency, nuclease activity, or deoxyribonuclease activity was significantly increased in the cohort without disease progression, whereas that of the gene set located at chromosome 7q35 (CHR7Q35) was significantly increased in the progression-positive cohort (Fig. 4c, Supplementary Fig. 7). Interestingly, the 7q35 locus manifested copy number amplification (Fig. 4b), suggesting that activation of the genes at this locus directly contributes to PCNSL relapse. We also extracted a gene set that is differentially expressed between *MYD88* mutation-positive and -negative tumors (Supplementary Table 5).

In addition, a search for gene fusions in the RNA-seq data identified only one in-frame fusion—between *INO80* and *NUSAP1* in tumor #11 (Supplementary Fig. 8)—which was verified by cDNA sequencing.

## Transformation-associated genes

GRB2 is an adapter protein that binds to tyrosine kinases and other docking proteins through its Src homology 2
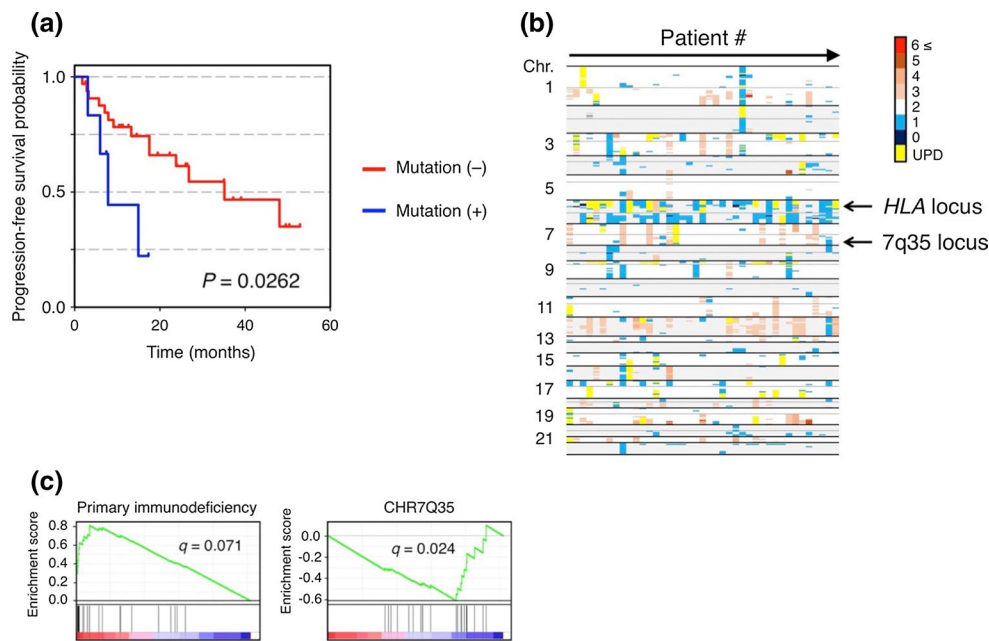
**Fig. 4** Chromosome copy number alterations and gene expression profiles in PCNSL. **a** Kaplan–Meier analysis of PFS probability for the treatment-naïve patients with or without *HLA-C* mutations. The *P* value for comparison of the two groups was determined by the log-rank test. **b** Chromosome copy number analysis of PCNSL. Copy number status is color-coded for chromosomes (Chr.) 1 to 22 (*top* to *bottom*) for patients with PCNSL (sorted from *left* to *right*). *HLA* and chromosome 7q35 loci are indicated. *UPD* uniparental disomy. **c** GSEA revealed "primary immunodeficiency" and "CHR7Q35″ gene sets to be significantly overexpressed in patients without or with disease progression, respectively. The false discovery rate (*q*) for each analysis is shown

(SH2) domain and transduces growth signals through the RAS-MAPK pathway [7]. We detected six missense and one nonsense mutation of *GRB2* in six patients (Fig. 5a, Supplementary Fig. 9). GRB2(V140G) showed oncogenic activity in a 3T3 cell transformation assay, with three other amino acid substitutions (L17R, L148R, A163T) also bestowing weak transforming ability (Supplementary Fig. 10). Given that GRB2 transduces growth signals in part through the RAS-MAPK pathway, we investigated whether inhibition of MAP2K1/2 activity might have therapeutic potential. Malignant transformation of 3T3 cells expressing GRB2(V140G) was attenuated by the MAP2K1/2 inhibitors, trametinib and selumetinib, in a concentration-dependent manner (Fig. 5b).

We also identified additional recurrent nonsynonymous mutations (Supplementary Table 6) in the exome data. A somatic p.V203I substitution in IKBKB, for instance, was newly discovered in two patients (Fig. 5c, Supplementary Fig. 11). We found that the ability of IKBKB(V203I) to phosphorylate NFKBIA (IκBα) was enhanced compared with that of the wild-type kinase. The V203I substitution of IKBKB results in enhanced degradation of NFKBIA and consequent activation of downstream canonical NF-κB signaling (Fig. 5d). While IKBKB(V203I) fails to transform 3T3 cells (data not shown), it could induce the transcription of NF-κB targets such as *CXCL2*, *CXCL3*, *MYC*

and *NFKBIA* (Supplementary Fig. 12). These data again confirm the importance of the NF-κB activity in the development of PCNSL.

## MYD88(L265P)-positive pre-lymphoma clones

Among the tumors apparently negative for *MYD88* mutations, manual inspection of the read data revealed mutations in an additional four cases (#2, #11, #18, and #23) (Supplementary Table 7), all of which were confirmed by Sanger sequencing, increasing the mutation frequency to 85.4 %. The frequency of *MYD88* mutations did not differ significantly between GCB- and non-GCB-subtypes (*P* = 0.98, Fisher's exact test). For some tumors (#11 and #23), computational pipelines failed to detect the *MYD88* somatic mutations because of its low frequency (15.1 % in #11 and 9.7 % in #23) which may be related to a low tumor content (~36 % in both cases) (Supplementary Table 1).

In contrast, somatic mutations of *MYD88* were not called in tumors #2 and #18, because NGS reads supporting the same mutations were also found in paired PBMNCs. Careful inspection of *MYD88* reads in all specimens revealed that NGS reads corresponding to MYD88(L265P) can be found in PBMNCs of patients #2, #11, #13, #18, and #33, all of whose tumor specimens were positive for MYD88(L265P) (Supplementary Table 7). To confirm the
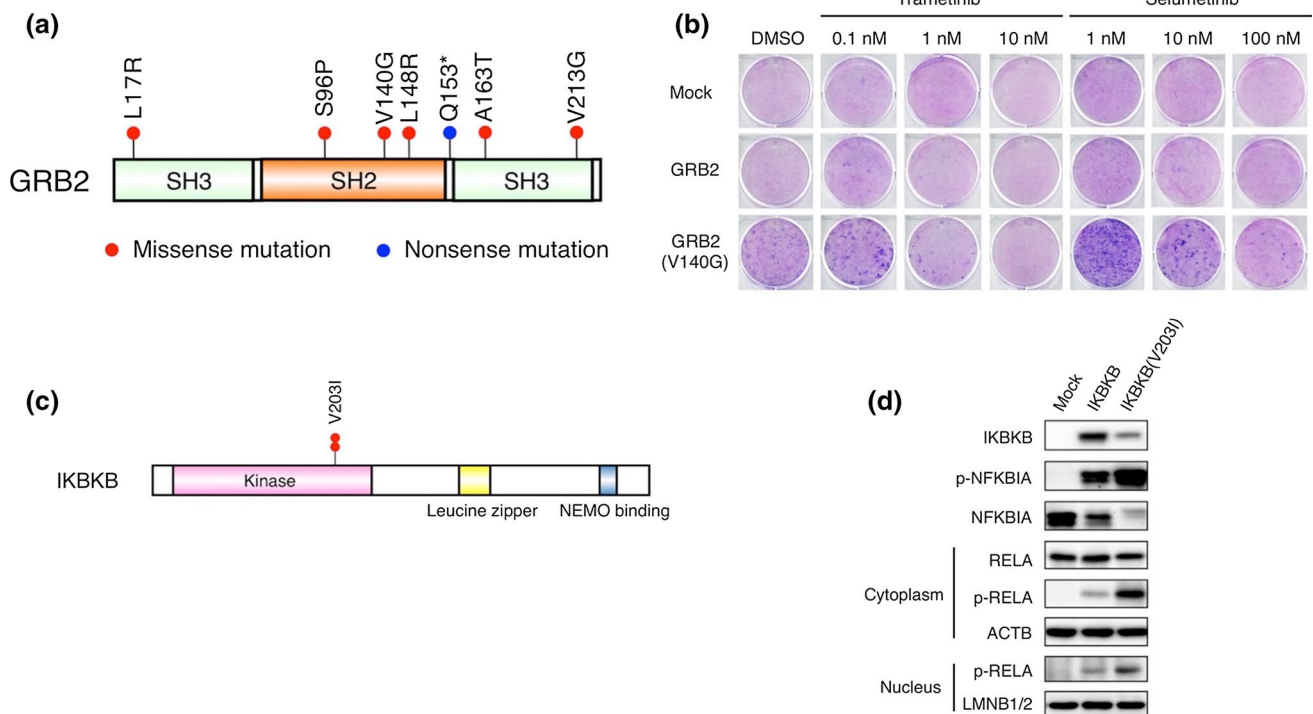
**Fig. 5** Transformation-associated mutations. **a** Domain organization of GRB2 showing the identified mutations. **b** 3T3 cells infected with retroviruses encoding GRB2 or GRB2(V140G) or with the empty retrovirus (Mock) were cultured for 2 weeks with the indicated concentrations of trametinib or selumetinib or with dimethyl sulfoxide (DMSO) vehicle beginning 2 days after infection. The cells were then stained with the Giemsa solution. **c** Domain organization of

IKBKB showing a recurrent missense mutation (p.V203I) in PCNSL. **d** HEK293 cells transfected with an expression vector for IKBKB or IKBKB(V203I) (or with the empty vector) were lysed and subjected to immunoblot analysis with antibodies to IKBKB or to total or phosphorylated (p-) forms of NFKBIA. Cytoplasmic and nuclear fractions prepared from the total cell lysates were also probed with antibodies to total or phosphorylated forms of RELA, to ACTB, or to LMNB1/2

mutation in PBMNCs, we amplified a genomic region corresponding to L265 of MYD88 by PCR from PBMNCs of patients #2, #11, and #18 and then ligated the amplification products into a cloning vector. Sanger sequencing of the resultant plasmids identified the L265P substitution in a small number of plasmids for each patient (Supplementary Fig. 13).

To detect the MYD88 mutation in PBMNCs with high sensitivity, we subjected genomic DNA of PBMNCs isolated from all patients whose tumors harbored MYD88(L265P) to digital PCR analysis. Unexpectedly, the mutant DNA was detected not only in PBMNCs of the five cases identified by the NGS data but also in those of an additional five cases (Supplementary Fig. 14, Supplementary Table 7) at a frequency above the sensitivity of our digital PCR assay (Supplementary Fig. 15).

Further, we asked whether gene mutations detected at a higher frequency in the tumors than that of *MYD88* were detectable by digital PCR in PBMNCs (Table 1). In NGS read data of case #2, for instance, tumor cells harbored *MYD88* and *MTMR8* mutations at a frequency of 65.7 and 80.6 %, respectively, whereas the corresponding values

for PBMNCs were 2.3 and 0.0 %. Digital PCR analysis detected the *MYD88* and *MTMR8* mutations in PBMNCs with a frequency of $1.39 \pm 0.26$ % (mean $\pm$ SD) and $0.02 \pm 0.03$ %, respectively. Likewise, the *MYD88* mutation was detected in PBMNCs of cases #4, #13, and #18, whereas none of the six gene mutations detected in tumors at a higher frequency than *MYD88* were detected in the PBMNCs by digital PCR analysis.

## Discussion

We have here characterized DLBCL-type PCNSL by whole-exome sequencing and RNA-seq analyses. With regard to genomic alterations, DLBCL-type PCNSL appears to be a relatively uniform disease, with a high mutation rate for *PIM1*, *MYD88*, and *BTG2* and uniform negativity for *BCL2* rearrangement (Supplementary Table 1). Furthermore, unsupervised clustering analysis of tumors according to gene mutation profile revealed that PCNSL forms a branch distinct from that of systemic DLBCL (Supplementary Fig. 16). These data suggest that

**Table 1** Digital PCR analysis of PBMNCs

| Patient | Mutant gene | Position | Nucleotide change | NGS data | | | | | | Digital PCR data |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Tumor | | | PBMNCs | | | Mutation ratio (%) in PBMNCs (mean ± SD) |
| | | | | Total reads | Mutant reads | Mutation ratio (%) | Total reads | Mutant reads | Mutation ratio (%) | |
| #2 | MYD88 | 3:38,182,641 | T>C | 347 | 228 | 65.7 | 172 | 4 | 2.3 | 1.39 ± 0.26 |
| | MTMR8 | X:63,576,174 | A>C | 268 | 216 | 80.6 | 68 | 0 | 0.0 | 0.02 ± 0.03 |
| #4 | MYD88 | 3:38,182,641 | T>C | 370 | 122 | 33.0 | 228 | 0 | 0.0 | 0.20 ± 0.12 |
| | COL4A6 | X:107,423,920 | C>A | 120 | 111 | 92.5 | 42 | 0 | 0.0 | 0.00 ± 0.00 |
| | BEND2 | X:18,209,292 | C>G | 49 | 41 | 83.7 | 23 | 0 | 0.0 | 0.00 ± 0.00 |
| #13 | MYD88 | 3:38,182,641 | T>C | 340 | 184 | 54.1 | 157 | 1 | 0.6 | 0.13 ± 0.04 |
| | OGT | X:70,767,686 | T>C | 85 | 58 | 68.2 | 45 | 0 | 0.0 | 0.00 ± 0.00 |
| | ARSE | X:2,876,366 | G>A | 84 | 57 | 67.9 | 60 | 0 | 0.0 | 0.00 ± 0.00 |
| #18 | MYD88 | 3:38,182,641 | T>C | 447 | 147 | 32.9 | 196 | 4 | 2.0 | 1.50 ± 0.19 |
| | TBC1D8B | X:106,065,558 | G>A | 48 | 23 | 47.9 | 26 | 0 | 0.0 | 0.00 ± 0.00 |
| | CCR9 | 3:45,942,935 | G>T | 500 | 180 | 36.0 | 258 | 0 | 0.0 | 0.00 ± 0.00 |

PCNSL is a distinct clinical entity irrespective of Hans's classification [9], which is in line with previous reports [4, 21], but this conclusion requires confirmation with a larger number of specimens.

*PIM1* was originally identified as a preferential proviral integration site of Moloney murine leukemia virus in T-cell lymphoma, and its high expression is presumed to contribute to carcinogenesis in hematological malignancies as well as solid tumors [23]. *PIM1* is, thus, believed to be a proto-oncogene. *PIM1* encodes a serine/threonine kinase, and its enzymatic inhibitor was shown effective against preclinical models of leukemia [11]. In our PCNSL analyses, however, *PIM1* is a target of aSHM in all tumors and often contains multiple nonsynonymous mutations within single tumors (Fig. 2). Some of the mutations are nonsense, and, hence, *PIM1* mutations are likely to be loss-of-function in this disorder. Indeed, some PIM1 mutants generated by aSHM in DLBCL were demonstrated to possess a decreased catalytic activity compared to that of the wild type [15]. Whether PIM1 acts for or against lymphomagenesis (or even possesses a kinase-independent function) in DLBCL/PCNSL warrants further examination.

The ABC (activated B-cell-like) subtype of DLBCL was shown dependent on a constitutive B-cell-receptor (BCR) signaling for cell viability [5]. Blockade of BCR signals with an inhibitor of a downstream BTK kinase (ibrutinib) has shown clinical efficacy against ABC-type DLBCL, especially that with double mutations of *CD79B* and *MYD88* [30]. In our PCNSL cohort, among the 35 lymphomas positive for *MYD88* mutations, 22 tumors (53.7 % of all cases) also carry concomitant *CD79B* alterations. BTK inhibitors that can efficiently penetrate blood–brain barrier

would be, thus, promising targeted drugs for about a half of PCNSL patients. Of note, while nonsynonymous mutations in *CD79B* or *MYD88* were reported to be highly enriched in non-GCB subtype of systemic DLBCL compared to GCB subtype [14], we observed no such subtype-preference for the mutations of the two genes in PCNSL ($P > 0.6$, Fisher's exact test).

In addition, we detected novel potentially oncogenic mutations of *GRB2* in ~14 % of PCNSL tumors. Importantly, blockade of the downstream kinases MAP2K1/2 with chemical agents reversed the malignant phenotype conferred by these mutations, opening up the possibility of a novel molecularly targeted therapy for this intractable disorder based on MAP2K1/2 inhibitors.

Whether PCNSL initially arises inside or outside of the CNS remained unclear. The fact that systemic dissemination of PCNSL is uncommon may indicate that the tumor develops from lymphocytes within the CNS. However, normal somatic hypermutation of lymphocytes requires the microenvironment provided by the germinal center of secondary lymphoid organs, suggesting that the initial transformation event in PCNSL may take place in such organs outside of the CNS.

We detected DNA corresponding to the MYD88(L265P) mutation, albeit at a low frequency, in PBMNCs of 10 cases (28.6 % of those with tumors positive for the mutant DNA), suggestive of clonal expansion of B cells harboring the mutant gene in peripheral blood. Such cells were unlikely to be tumor cells that had infiltrated into the systemic circulation from the CNS, given that gene mutations with a higher frequency than that for MYD88(L265P) in tumors were not detected in PBMNCs (Table 1). Likewise, due to

the same reason, it would not be plausible to assume that the *MYD88* mutation we observed was derived from circulating tumor-derived DNA or exosomes. Our results are consistent with the previous observation that tumor-associated rearrangements of immunoglobulin variable genes were clonally present in PBMNCs of PCNSL patients and that tumor cells and PBMNCs with the same rearrangement acquire additional mutations independently [19].

These data suggest that *MYD88* mutation-positive "pre-lymphoma" cells first appear outside of the CNS and circulate in peripheral blood and possibly in lymphatic vessels. They then enter the CNS and accumulate additional genetic or epigenetic alterations that provide a growth advantage in this environment. Definitive conclusion of the presence of pre-lymphoma cells in peripheral blood would be obtained by genomic analyses on sorted single cells of peripheral blood in PCNSL patients. Our observation that (1) the *MYD88* mutation was present in PBMNCs from one-quarter of *MYD88* mutation-positive PCNSL patients and (2) PCNSL carries *MYD88* mutations in ~86 % of cases together suggests that activation of MYD88 may be one of the initial and important genetic alteration of lymphocytes that supports their survival in the CNS.

**Compliance with ethical standards**

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments.

# References

1. Batchelor T, Loeffler JS (2006) Primary CNS lymphoma. J Clin Oncol 24:1281–1288. doi:10.1200/JCO.2005.04.8819

2. Bonzheim I, Giese S, Deuter C, Susskind D, Zierhut M, Waizel M, Szurman P, Federmann B, Schmidt J, Quintanilla-Martinez L et al (2015) High frequency of MYD88 mutations in vitreoretinal B-cell lymphoma: a valuable tool to improve diagnostic yield of vitreous aspirates. Blood 126:76–79. doi:10.1182/blood-2015-01-620518

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

4. Camilleri-Broet S, Criniere E, Broet P, Delwail V, Mokhtari K, Moreau A, Kujas M, Raphael M, Iraqi W, Sautes-Fridman C et al (2006) A uniform activated B-cell-like immunophenotype might explain the poor prognosis of primary central nervous system lymphomas: analysis of 83 cases. Blood 107:190–196. doi:10.1182/blood-2005-03-1024

5. Davis RE, Ngo VN, Lenz G, Tolar P, Young RM, Romesser PB, Kohlhammer H, Lamy L, Zhao H, Yang Y et al (2010) Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. Nature 463:88–92. doi:10.1038/nature08638

6. Deckert M, Montesinos-Rongen M, Brunn A, Siebert R (2014) Systems biology of primary CNS lymphoma: from genetic aberrations to modeling in mice. Acta Neuropathol 127:175–188. doi:10.1007/s00401-013-1202-x

7. Downward J (1994) The GRB2/Sem-5 adaptor protein. FEBS Lett 338:113–117

8. Gonzalez-Aguilar A, Idbaih A, Boisselier B, Habbita N, Rossetto M, Laurenge A, Bruno A, Jouvet A, Polivka M, Adam C et al (2012) Recurrent mutations of MYD88 and TBL1XR1 in primary central nervous system lymphomas. Clin Cancer Res 18:5203–5211. doi:10.1158/1078-0432.CCR-12-0845

9. Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, Muller-Hermelink HK, Campo E, Braziel RM, Jaffe ES et al (2004) Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood 103:275–282. doi:10.1182/blood-2003-05-1545

10. Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A, Yamamoto S, Tatsuno K, Katoh H, Watanabe Y et al (2014) Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. Nat Genet 46:583–587. doi:10.1038/ng.2984

11. Keeton EK, McEachern K, Dillman KS, Palakurthi S, Cao Y, Grondine MR, Kaur S, Wang S, Chen Y, Wu A et al (2014) AZD1208, a potent and selective pan-Pim kinase inhibitor, demonstrates efficacy in preclinical models of acute myeloid leukemia. Blood 123:905–913. doi:10.1182/blood-2013-04-495366

12. Khodabakhshi AH, Morin RD, Fejes AP, Mungall AJ, Mungall KL, Bolger-Munro M, Johnson NA, Connors JM, Gascoyne RD, Marra MA et al (2012) Recurrent targets of aberrant somatic hypermutation in lymphoma. Oncotarget 3:1308–1319. doi:10.18632/oncotarget.653

13. Korfel A, Schlegel U (2013) Diagnosis and treatment of primary CNS lymphoma. Nat Rev Neurol 9:317–327. doi:10.1038/nrneurol.2013.83

14. Kraan W, Horlings HM, van Keimpema M, Schilder-Tol EJ, Oud ME, Scheepstra C, Kluin PM, Kersten MJ, Spaargaren M, Pals ST (2013) High prevalence of oncogenic MYD88 and CD79B mutations in diffuse large B-cell lymphomas presenting at immune-privileged sites. Blood Cancer J 3:e139. doi:10.1038/bcj.2013.28

15. Kumar A, Mandiyan V, Suzuki Y, Zhang C, Rice J, Tsai J, Artis DR, Ibrahim P, Bremer R (2005) Crystal structures of proto-oncogene kinase Pim1: a target of aberrant somatic hypermutations in diffuse large cell lymphoma. J Mol Biol 348:183–193. doi:10.1016/j.jmb.2005.02.039

16. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA et al (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499:214–218. doi:10.1038/nature12213

17. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, Cruz-Gordillo P, Knoechel B, Asmann YW, Slager SL et al (2012) Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by

whole-exome sequencing. Poc Natl Acad Sci USA 109:3879–3884. doi:10.1073/Pnas.1121343109

18. Mandelbaum J, Bhagat G, Tang H, Mo T, Brahmachary M, Shen Q, Chadburn A, Rajewsky K, Tarakhovsky A, Pasqualucci L et al (2010) BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. Cancer Cell 18:568–579. doi:10.1016/j.ccr.2010.10.030

19. McCann KJ, Ashton-Key M, Smith K, Stevenson FK, Ottensmeier CH (2009) Primary central nervous system lymphoma: tumor-related clones exist in the blood and bone marrow with evidence for separate development. Blood 113:4677–4680. doi:10.1182/blood-2008-09-179366

20. Mead GM, Bleehen NM, Gregor A, Bullimore J, Shirley D, Rampling RP, Trevor J, Glaser MG, Lantos P, Ironside JW et al (2000) A medical research council randomized trial in patients with primary cerebral non-Hodgkin lymphoma: cerebral radiotherapy with and without cyclophosphamide, doxorubicin, vincristine, and prednisone chemotherapy. Cancer 89:1359–1370

21. Montesinos-Rongen M, Brunn A, Bentink S, Basso K, Lim WK, Klapper W, Schaller C, Reifenberger G, Rubenstein J, Wiestler OD et al (2008) Gene expression profiling suggests primary central nervous system lymphomas to be derived from a late germinal center B cell. Leukemia 22:400–405. doi:10.1038/sj.leu.2405019

22. Morin RD, Mungall K, Pleasance E, Mungall AJ, Goya R, Huff RD, Scott DW, Ding JR, Roth A, Chiu R et al (2013) Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. Blood 122:1256–1265. doi:10.1182/Blood-2013-02-483727

23. Narlik-Grassow M, Blanco-Aparicio C, Carnero A (2014) The PIM family of serine/threonine kinases in cancer. Med Res Rev 34:136–159. doi:10.1002/med.21284

24. Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao W, Lim KH, Kohlhammer H, Xu W, Yang Y, Zhao H et al (2011) Oncogenically active MYD88 mutations in human lymphoma. Nature 470:115–119. doi:10.1038/nature09671

25. Norden AD, Drappatz J, Wen PY, Claus EB (2011) Survival among patients with primary central nervous system lymphoma, 1973–2004. J Neurooncol 101:487–493. doi:10.1007/s11060-010-0269-7

26. Ricard D, Idbaih A, Ducray F, Lahutte M, Hoang-Xuan K, Delattre JY (2012) Primary brain tumours in adults. Lancet 379:1984–1996. doi:10.1016/S0140-6736(11)61346-9

27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102:15545–15550. doi:10.1073/pnas.0506580102

28. 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. doi:10.1038/leu.2014.264

29. Villano JL, Koshy M, Shaikh H, Dolecek TA, McCarthy BJ (2011) Age, gender, and racial differences in incidence and survival in primary CNS lymphoma. Br J Cancer 105:1414–1418. doi:10.1038/bjc.2011.357

30. Wilson WH, Young RM, Schmitz R, Yang Y, Pittaluga S, Wright G, Lih CJ, Williams PM, Shaffer AL, Gerecitano J et al (2015) Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. Nat Med 21:922–926. doi:10.1038/nm.3884

31. Zhang J, Grubor V, Love CL, Banerjee A, Richards KL, Mieczkowski PA, Dunphy C, Choi W, Au WY, Srivastava G et al (2013) Genetic heterogeneity of diffuse large B-cell lymphoma. Proc Natl Acad Sci USA 110:1398–1403. doi:10.1073/pnas.1205299110