

Mutational analysis of primary central nervous system lymphoma

Aurélie Bruno^{1,2,3,4}, Blandine Boisselier^{1,2,3,4,5}, Karim Labreche^{1,2,3,4}, Yannick Marie^{5,6}, Marc Polivka⁷, Anne Jouvét⁸, Clovis Adam⁹, Dominique Figarella-Branger¹⁰, Catherine Miquel¹¹, Sandrine Eimer¹², Caroline Houillier¹³, Carole Soussain¹⁴, Karima Mokhtari^{1,2,3,4,6}, Romain Daveau¹⁵ and Khê Hoang-Xuan^{1,2,3,4,13}

¹ Sorbonne Universités, UPMC Univ Paris 06, UM 75, ICM, F-75013 Paris, France

² Institut National de la Santé et de la Recherche Médicale, U1127, ICM, Paris, F-75013 Paris, France

³ Centre National de la Recherche Scientifique, UMR 7225, ICM, Paris, F-75013 Paris, France

⁴ ICM, Paris, 75013 France

⁵ Plateforme de Génotypage Séquençage, ICM, F-75013, Paris, France

⁶ Onconeurothèque, Groupe Hospitalier Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, Paris, France

⁷ Centre Hospitalier Universitaire Lariboisière, Assistance Publique-Hôpitaux de Paris, Service d'Anatomopathologie, Paris, France

⁸ Hospices Civils de Lyon, Hôpital Neurologique, Bron, France and Université Lyon 1, Institut National de la Santé et de la Recherche Médicale Unité 842, Lyon, France

⁹ Centre Hospitalier Universitaire Bicêtre, Assistance Publique-Hôpitaux de Paris, Service d'anatomopathologie, Bicêtre, France

¹⁰ Centre Hospitalier Universitaire La Timone, Assistance Publique-Hôpitaux de Marseille, Institut National de la Santé et de la Recherche Médicale Unité 911, Centre de Recherches en Oncologie biologique et Onco-pharmacologie, Université de la Méditerranée and Tumorothèque de l'Assistance Publique-Hôpitaux de Marseille (AC 2013-1786), Marseille, France

¹¹ Centre hospitalier Sainte Anne, Université Paris Descartes, Sorbonne Paris Cité, Paris, France

¹² Service de Pathologie, CRB Tumorothèque, Centre Hospitalier Universitaire Bordeaux, Bordeaux, France

¹³ Assistance Publique-Hôpitaux de Paris, Hôpital de la Pitié-Salpêtrière, Service de Neurologie 2-Mazarin, Paris, France; and the LOC network (INCa)

¹⁴ Hôpital René Huguenin, Institut Curie, Service d'Hématologie, Saint Cloud, France; and the LOC network (INCa)

¹⁵ Institut National de la Santé et de la Recherche Médicale Unité 830, Génétique et Biologie des Cancers, Institut Curie, Paris, France

Correspondence to: Khê Hoang-Xuan, *email:* khe.hoang-xuan@psl.aphp.fr

Keywords: Primary CNS lymphoma, exome sequencing, somatic mutations, NFKB, B cell differentiation

Received: April 21, 2014

Accepted: June 7, 2014

Published: June 8, 2014

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Little is known about the genomic basis of primary central nervous system lymphoma (PCNSL) tumorigenesis. To investigate the mutational profile of PCNSL, we analyzed nine paired tumor and germline DNA samples from PCNSL patients by high throughput exome sequencing. Eight genes of interest have been further investigated by focused resequencing in 28 additional PCNSL tumors to better estimate their incidence. Our study identified recurrent somatic mutations in 37 genes, some involved in key signaling pathways such as NFKB, B cell differentiation and cell cycle control. Focused resequencing in the larger cohort revealed high mutation rates for genes already described as mutated in PCNSL such as *MYD88* (38%), *CD79B* (30%), *PIM1* (22%) and *TBL1XR1* (19%) and for genes not previously reported to be involved in PCNSL tumorigenesis such as *ETV6* (16%), *IRF4* (14%), *IRF2BP2* (11%) and *EBF1* (11%). Of note, only 3 somatically acquired SNVs were annotated in the COSMIC database. Our results demonstrate a high genetic heterogeneity of PCNSL

and mutational pattern similarities with extracerebral diffuse large B cell lymphomas, particularly of the activated B-cell (ABC) subtype, suggesting shared underlying biological mechanisms. The present study provides new insights into the mutational profile of PCNSL and potential targets for therapeutic strategies.

INTRODUCTION

Primary central nervous system lymphoma (PCNSL) represents a rare subgroup of diffuse large B-cell lymphoma (DLBCL) that arises in the brain, eyes, meninges or spinal cord, accounting for up to 5% of primary malignant brain tumors and 1% of non-Hodgkin's lymphomas (NHL) in adults. Despite the application of intensive treatment including high-dose methotrexate based poly-chemotherapy with or without whole brain radiotherapy, the median overall survival ranges from 2 to 4 years with a poorer prognosis than extracerebral DLBCL [1]. The pathogenesis of PCNSL remains largely unclear, which is partly due to the rarity of the tumor tissue available for research studies. Transcriptomic studies have identified deregulated genes involved in the IL4/JAK/STAT6, cell adhesion-related, unfolded protein response (UPR) and apoptosis signaling pathways [2–5]. Copy number variation studies [4, 6–8] have revealed frequent chromosome losses affecting the 6q, 6p21.32 and 9p21 regions. However, the mutational landscape of PCNSL is still poorly known. A whole exome sequencing strategy has successfully identified pivotal gene mutations in several hematologic and brain malignancies [9, 10]. In a previous study, we have reported preliminary results based on four PCNSL cases investigated by this technique and identified recurrent mutations in *MYD88* and *TBLIXR1* [8]. Here, we have expanded our series and we present the results of nine paired germline and tumour samples, allowing for the identification of recurrent gene mutations that have not yet been reported in PCNSL. We confirmed the most relevant mutations and genes in a validation set of 28 PCNSL cases.

RESULTS

Mutational pattern of PCNSL revealed by whole exome sequencing

To investigate the mutational profile of PCNSL, we performed high throughput exome sequencing on 9 cases. DNA from case-matched blood was also sequenced to screen out germline polymorphisms. On average, 9.8×10^7 (8.1×10^7 - 1.4×10^8) 75-bp paired reads were sequenced per sample, 5.8×10^7 (3.8×10^7 - 8.3×10^7) of these were specifically positioned onto the human reference exome (as defined by the Agilent SureSelect 50 Mb probes) after the removal of both low-quality mapped reads and potential

PCR-derived duplicates (Supplementary Figure 1). This provided 76% (64-86) coverage over the targeted regions at a minimum depth of 20X (Supplementary Figure 2), wherein 82% of the bases were suitable for variant detection. Across the coding regions of the 9 matched tumor and germline pairs we investigated, we detected 17×10^3 (15×10^3 - 19×10^3) SNVs and 226 (176-263) indels. A total of 25×10^3 (20×10^3 - 32×10^3) SNVs and 21×10^2 (18×10^2 - 27×10^2) indels were also called outside of the targeted exons, but those primarily fell into neighboring introns and, in most cases, were already described as known polymorphisms. To assess the quality of our calls, we reviewed population-scale variant distributions from the 1000-genomes project and found no difference with either paired germline or PCNSL samples when considering all high-quality called SNVs and comparing the (i) transition to transversion rates, (ii) mutational spectrum and (iii) variant annotation (Supplementary Figure 3). Then, we focused on somatic SNVs identified in tumor DNA and not present in germline DNA (Figure 1). On average, we identified 220 (126-358) somatically acquired point mutations per sample and no hypermutated tumors were found. Among them, 62 (26-101) and 143 (89-231) were synonymous and non-synonymous, respectively. The non-synonymous to synonymous ratio was thus 2.4 (1.8-3.4), and there was a non-silent mutation rate of 2.9 (1.8-4.6) per Mb, the latter being lower than previously published estimations in DLBCL [11]. Half of those non-synonymous SNVs, i.e., 74 (49-111) were predicted as functionally deleterious in the dbNSFP database [12, 13]. Transitions accounted for 68% of somatic events (Figure 1B) similar to pattern observed in DLBCL [14, 15]. To confirm the depth at the somatic mutation sites, reads aligned at these genomic positions were visualized using IGV software (Broad Institute).

Identification of 37 genes recurrently affected by somatic non-synonymous mutations

Only SNVs located within coding regions were considered. After having removed germline variations, synonymous SNVs, indels and known polymorphisms, we identified 37 genes, harboring 142 somatically point mutations (Supplementary Table 1), that were mutated in at least 2 patients. Among these 142 mutations, 133 led to an amino acid exchange while the remaining nine led to the gain or loss of a stop codon. These 37 recurrently mutated genes were prioritized based on (i) the number of mutated tumors, (ii) the prediction of the functional impact

Table 1: Prioritization of the 37 genes of interest identified by whole exome sequencing. The present study identified 37 genes affected by non-synonymous somatic SNVs in at least 2 of the 9 patients of the discovery set. In this table are listed all these genes prioritized according to (i) number of mutated patients, (ii) functional impact prediction (FISM), (iii) number of mutations per gene. Functional impact prediction columns indicate the number of patients harboring at least one mutation for each FISM category (1 corresponds to the highest impact). Of importance, number of mutated patients and number of mutations per gene take into account all somatic mutations identified by exome sequencing before any attempt of validation.

Genes	Chromosome	Mutations	Patients	Functional prediction impact (FISM)						
				NA	≥0.5	≥0.6	≥0.7	≥0.8	≥0.9	=1
<i>PIMI</i>	6	32	8	0	8	8	7	6	6	5
<i>IGLL5</i>	22	12	6	6	0	0	0	0	0	0
<i>MYD88</i>	3	2	5	5	0	0	0	0	0	0
<i>TBL1XR1</i>	3	4	4	0	4	4	4	4	4	3
<i>CSMD3</i>	8	4	4	0	4	4	4	3	3	1
<i>CD79B</i>	17	3	3	0	2	2	2	2	2	1
<i>HIST1H2AC</i>	6	8	3	0	3	3	3	3	1	1
<i>ETV6</i>	12	5	3	0	3	3	2	2	1	1
<i>KLHL14</i>	18	7	2	0	2	2	2	2	2	2
<i>IRF4</i>	6	3	2	0	2	2	2	2	2	2
<i>PRKCD</i>	3	2	2	0	2	2	2	2	2	2
<i>ABCC8</i>	11	2	2	0	2	2	2	2	2	1
<i>ZFHX4</i>	8	2	2	0	2	2	2	2	2	1
<i>SALL3</i>	18	2	2	0	2	2	2	1	1	1
<i>IRF2BP2</i>	1	3	2	0	2	2	1	1	1	1
<i>CD37</i>	19	2	2	0	2	2	1	1	1	1
<i>OSBPL10</i>	3	7	2	0	2	2	2	2	2	0
<i>EBF1</i>	5	3	2	0	2	2	2	2	2	0
<i>DST</i>	6	2	2	0	2	2	2	2	1	0
<i>MIF4GD</i>	17	2	2	0	2	2	2	2	1	0
<i>HIST1H1D</i>	6	3	2	0	2	2	2	1	1	0
<i>BTG1</i>	12	2	2	0	2	2	2	1	1	0
<i>MEP1B</i>	18	2	2	0	2	2	2	1	1	0
<i>THBS4</i>	5	2	2	0	2	2	2	1	1	0
<i>ADAMTS5</i>	21	2	2	0	2	2	1	1	1	0
<i>HIST1H1E</i>	6	2	2	0	2	1	1	1	1	0
<i>MPEG1</i>	11	3	2	1	1	1	1	1	1	0
<i>OBSCN</i>	1	2	2	0	2	2	2	2	0	0
<i>C10orf71</i>	10	2	2	0	2	2	2	1	0	0
<i>HMCN1</i>	1	2	2	0	2	2	2	1	0	0
<i>MYH4</i>	17	2	2	0	2	2	1	1	0	0
<i>TBC1D4</i>	13	2	2	0	2	1	1	1	0	0
<i>SLC2A12</i>	6	2	2	0	2	2	1	0	0	0
<i>ETS1</i>	11	2	2	0	2	2	0	0	0	0
<i>MUC16</i>	19	2	2	2	0	0	0	0	0	0
<i>UNC80</i>	2	2	2	2	0	0	0	0	0	0
<i>ACTG1</i>	17	1	2	2	0	0	0	0	0	0

and (iii) the number of SNVs per gene (Table 1). Then, somatic mutations were verified by Sanger sequencing on tumor and germline DNA. For *PIMI* and *MYD88* genes, only “hot spot” mutations E226K and L265P, respectively, were validated. To better understand the biological processes that are potentially altered by somatic mutations, we used gene ontology [16] annotations for these 37 genes. This functional categorization highlighted the variability of the biological processes that are altered in PCNSL (Figure 2), including transcription (e.g., *ETV6*, *IRF2BP2*, *EBF1*, *IRF4*, *TBLIXR1*), cell cycle (e.g., *PIMI*, *BTGI*), nucleosome assembly (e.g., *HIST1H1D*, *HIST1H2AC*) and cell adhesion (e.g., *MUC16*, *ACTG1*). In terms of signaling pathways, we identified mutations in the genes involved in the NFKB, WNT and B-cell or T-cell receptor signaling pathways.

Analysis of 8 relevant genes in an independent series of 28 PCNSL

In order to specify their mutation frequency in PCNSL, we selected 8 genes for further investigation in an independent validation panel of PCNSL tumors (n=28). This selection was based both on high mutation rate in our discovery set and biological relevance. *PIMI*, *TBLIXR1*, *ETV6*, *IRF4*, *IRF2BP2* and *EBF1* were resequenced for their coding exons by pyrosequencing. We identified

133 variations, including 122 SNVs and 11 deletions. Among them, 39 variations were missense mutations (Supplementary Table 2), including 35 variations that were not previously described in the dbSNP database as known polymorphisms. For each missense SNV, functional impact was predicted using SIFT or Polyphen2 tools and identified 11 SNVs with putative damaging consequences predicted by both softwares. Twenty-five out of the 35 missense mutations were validated by Sanger sequencing and corresponded to 22 SNVs and 3 frameshift deletions. The somatic state of the validated mutations was verified with direct sequencing. Considering the whole cohort, including the discovery and the validation sets, somatic variations were found in 22% (8/37) of the PCNSL cases for *TBLIXR1*, 19% (7/37) for *PIMI*, 16% (6/37) for *ETV6*, 14% (5/37) for *IRF2BP2* and 11% (4/37) for *IRF4* and *EBF1* each (Fig 3A). Of note, 3 non-sense mutations affecting *ETV6* and *IRF2BP2* genes and 3 deletions leading to a frameshift in *TBLIXR1*, *ETV6* and *EBF1* were observed (Figure 3B). Somatic mutations on the hot spots L265P of *MYD88* and Y196 of *CD79B* were already referenced in the COSMIC database. One somatic mutation within the *PIMI* gene was also identified in this database (e.g., COSM220740) as reported in DLBCL cases [9, 14]. Four other somatic mutations identified within the *PIMI*, *ETV6* and *IRF4* genes in this study occur in the same codon as the alterations that are mainly reported in hematopoietic or lymphoid malignancies.

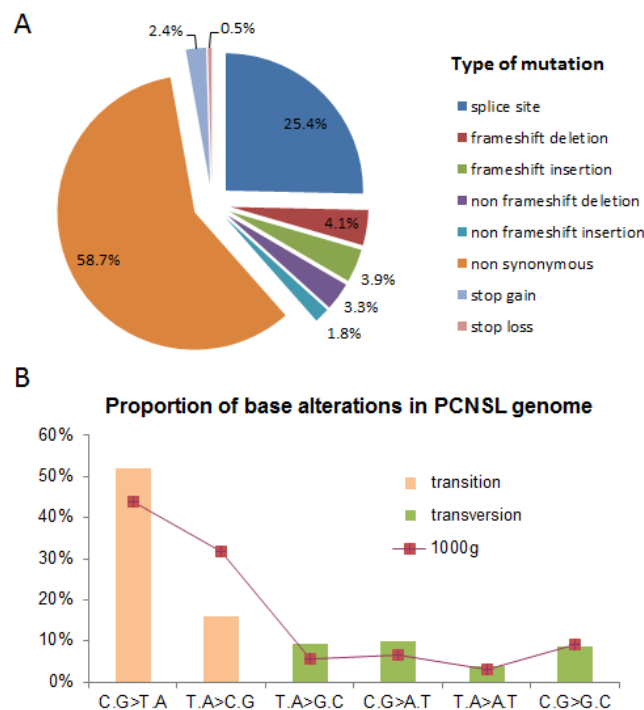


Figure 1: Mutation pattern of PCNSL samples investigated by whole exome sequencing. To address PCNSL mutational profile, whole exome sequencing was conducted in nine paired blood and tumor samples. (A) Pie chart represents relative distribution of somatically acquired mutations classified according to their type. (B) Histogram depicts the proportion of PCNSL somatic mutations in each mutational class of transitions and transversions compared with 1000-genomes project data (red line).

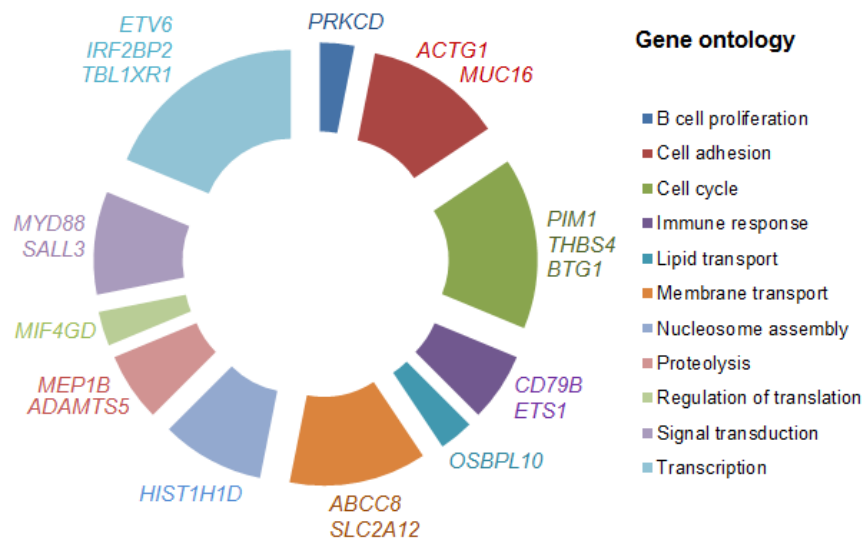


Figure 2: Gene ontology of PCNSL genes. Relative distribution of the 37 genes somatically mutated in PCNSL by gene ontology categories. The spans of the arcs indicate the relative numbers of genes annotated with respect to gene ontology terms. Representative genes in each category are shown next to each arc.

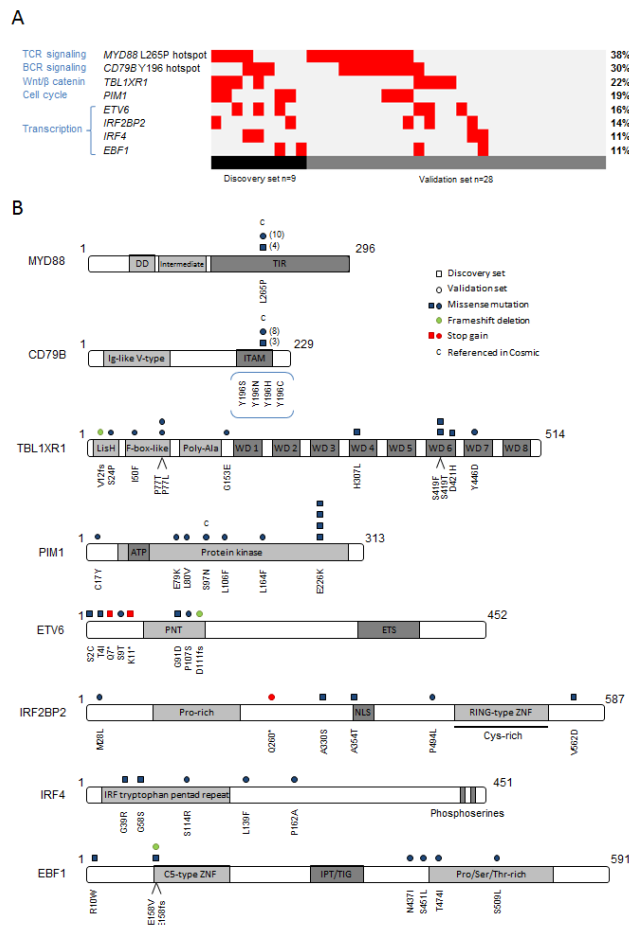


Figure 3: Investigation of 8 relevant genes recurrently affected by point mutations in PCNSL. Based on genes identified by whole exome sequencing, we selected 8 relevant genes to be sequenced in a larger cohort: *CD79B*, *EBF1*, *ETV6*, *IRF4*, *IRF2BP2*, *MYD88*, *PIM1* and *TBL1XR1*. (A) Repartition of validated mutations by gene within the whole population of 37 PCNSL cases. (B) Schematic representation of all validated mutations identified in the discovery (□) and the validation sets (○) with their position according to protein domains. Symbol color indicates mutation type. Number of □ or ○ indicates the number of mutated patients except for L265P *MYD88* and Y196 *CD79B* mutations.

Direct sequencing of *MYD88* and *CD79B* focused on the hot spot mutations identified in the discovery panel; the L265P mutation was found in 4/9 cases, and Y196 mutations were found in 3/9 cases. In the validation panel, 10 additional patients harbored the *MYD88* L265P mutation and 8 additional cases harbored *CD79B* Y196 mutations. Considering the whole population, *MYD88* L265P and *CD79B* Y196 mutations were identified in 38% (14/37) and 30% (11/37) of PCNSL tumors, respectively (Figure 3A), representing the most recurrently mutated genes in our series.

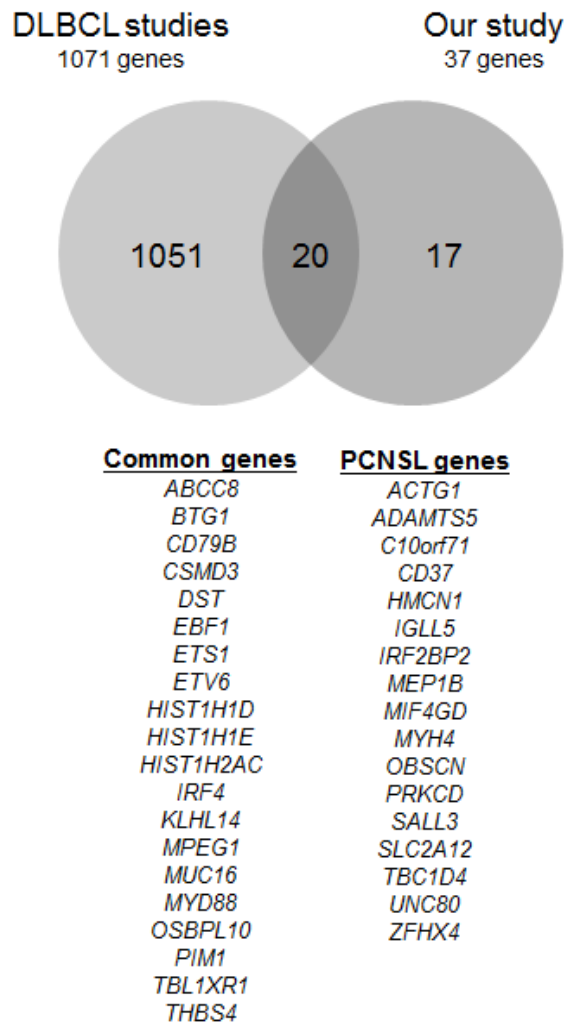


Figure 4: Overlaps in genes discovered in DLBCL studies and our 37 genes of interest. The Venn diagram depicts the comparison between gene mutations from the five DLBCL exomes studies and the present PCNSL study. The gene lists used were as follows: Lohr et al. (Table 1 in Ref. 11, n=72 genes), Pasqualucci et al. (Table S3 and Fig. S4 in Ref. 15, n=108 validated somatic genes), Zhang et al. (Table S3 in Ref. 14, n=322 genes), Morin et al. (in Ref. 9, n=315 known and confirmed somatic genes; Table S3 in Ref. 37, n=588 known and confirmed somatic genes).

DISCUSSION

The present study investigated the coding genomes of PCNSL in order to provide information on the mutational landscape of these tumors. We described an overview of the genes that are recurrently mutated in PCNSL, including (i) genes previously known to be mutated in PCNSL, such as *MYD88*, *CD79B*, *PIM1* and *TBL1XR1*; (ii) genes altered by somatic mutations in other B cell malignancies that have not yet been reported in PCNSL, such as *ETV6*, *IRF4* or *EBF1*; and (iii) genes that are altered in solid tumors, such as *IRF2BP2*. These results reveal the genetic heterogeneity of this disease and highlight the major signaling pathways that are deregulated in PCNSL.

In our series, genes coding for nuclear factor- κ B (NF κ B) pathway regulators (i.e., *MYD88*, *CD79B* and *TBL1XR1*) represented the most frequently altered genes. *MYD88* encodes a signaling adaptor protein that induces NF κ B and JAK/STAT3 pathway activation after the stimulation of the Toll-like and IL1/IL18 receptors as well as interferon β production [17, 18]. *CD79B* encodes a B-cell receptor (BCR) subunit that is essential for BCR signaling, leading to NF κ B activation [19]. We identified *MYD88* L265P and *CD79B* Y196 hot spot mutations in 38% and 30% of the PCNSL patients, respectively. We confirm and expand the results of Montesinos-Rongen et al [20, 21] who have recently investigated PCNSL for mutations in several genes involved in the BCR signaling cascade and reported a 36% (7/14) and 20% (5/25) mutation rate in *MYD88* and *CD79B*, respectively. These two hot-spot mutations have been described as oncogenic activating alterations leading to constitutive NF κ B activation in DLBCL [22, 23]. Additionally, we found a significant association ($p=0.0044$, Chi-square test) between the *MYD88* L265P and *CD79B* Y196 mutations, suggesting collaborative effects of the NF κ B activating pathways in PCNSL. The *TBL1XR1* gene, which encodes for a transcriptional regulator involved both in the Wnt/B catenin [24, 25] and NF κ B pathways [26], was mutated in 22% of our PCNSL cohort. The *TBL1XR1* mutation rate in our series and the recurrent deletions of 3q26.32 (*TBL1XR1* locus) reported in PCNSL [7], extracerebral DLBCL [15], and acute lymphoblastic leukemia [27, 28] suggest its potential role as a tumor suppressor. Taken together, mutations in *MYD88*, *CD79B* and *TBL1XR1* affected 54% (20/37) of our cohort, suggesting that NF κ B pathway deregulation is a driving mechanism in PCNSL tumorigenesis. Other genes, such as *CARD11* and *TNFAIP3*, which belong to this pathway are also reported to be mutated at lower rates in 16% and 3% of PCNSL, respectively [29].

A second set of alterations was detected in genes involved in B-cell proliferation and differentiation, such as *ETV6*, *EBF1*, *IRF4* and *ETS1*. To our knowledge these gene mutations have never been reported in

PCNSL. The *ETV6* tumor suppressor gene encodes an Ets family transcriptional repressor factor required for hematopoiesis [30] and largely described as a partner of gene translocation in lymphoid and myeloid hematopoietic tumors [31]. In our series, we found *ETV6* mutations in 16% of cases, including 2 cases with non-sense mutations. In line with this, several studies have reported heterozygous and homozygous deletions of 12p13.2 corresponding to the *ETV6* locus (15% in the present series) in PCNSL [6, 7]. *IRF4*, also known as *MUM1* encoding a lymphocyte-specific transcription factor [32], and *EBF1*, encoding an activator of transcription involved in lymphoid development [33], were found to be mutated in 11% of our cohort. We also reported, in our discovery set, 2 somatic mutations affecting *ETS1*, encoding another Ets family transcription factor involved in the negative regulation of plasmocytic differentiation [34]. A variety of *ETS1* alterations, including deletions [35] or gains [36] and somatic mutations [9, 37], have been reported in B cell malignancies. Finally, 11 tumors from our 37 samples (30%) harbored one or more mutation of genes involved in B cell proliferation and differentiation, supporting the role of B lymphoid development deregulation in PCNSL tumorigenesis.

A hallmark of oncogenesis is the alteration of genes controlling the cell cycle. We and others have previously identified *CDKN2A* homozygous deletions as a frequent alteration in PCNSL [4, 6–8] with an unfavorable impact on the prognosis [8]. In the present study, we found recurrent mutations in cell cycle regulator genes such as *PIMI* [38] (7/37; 19%), *IRF2BP2* [39] (5/37; 14%) and *BTG1* [40] (2/9). *PIMI* is a proto-oncogene that encodes a serine/threonine kinase and is known to be frequently targeted by somatic hypermutation in PCNSL [41]. Of note, 6 of the 7 seven mutations identified on *PIMI* in the present study were located on the protein kinase domain. A variety of inhibitors are currently under development for PIM family proteins [42]^(Tab2), rendering these proteins attractive targets for therapy [5]. *IRF2BP2* encodes a zinc finger protein that interacts with partners such as *TP53* and the oncogene *IRF2*. *IRF2BP2* acts as a repressor of *IRF2*, leading to the inhibition of interferon responsive gene expression and *NFAT1*, which is involved in the cell cycle. Recently, a novel fusion between *IRF2BP2* and the *CDX1* homeobox gene was described in a patient suffering from a mesenchymal chondrosarcoma [43]. Intriguingly, the patient also had PCNSL; unfortunately the brain tumor tissue was not investigated.

Our results revealed many similarities between genomic abnormalities of extracerebral DLBCL and PCNSL. Indeed, among the 37 genes of interest identified in this study, 20 have described mutations in DLBCL exome studies [9, 11, 14, 15, 37] (Figure 4). More specifically, mutations in the genes involved in the NFκB signaling pathway and in *PIMI*, as observed in PCNSL, are likely associated with the activated B-cell like (ABC)

subtype of DLBCL. In contrast, histone-modifying genes, such as *CREBBP*, *EZH2* and *MLL2*, which are recurrently altered in the germinal center B-cell like (GCB) subtype of DLBCL [9, 14, 15], were not found in our series. These observations are in agreement with previous studies showing that the PCNSL gene expression profile is more closely related to post-GCB and ABC cells than to GCB cells [2, 44].

The present study has several limitations. Even if the small number of cases analyzed is generally acceptable given the rarity of the disease and small amount of available tissue, it provides a limited power of analysis and we likely underestimate the PCNSL gene mutations. In addition, the sequencing methods used do not investigate noncoding portions of the genome. Altogether, this could explain the relatively low overlap with a recent study of the Mayo Clinic including 10 PCNSL investigated by whole exome sequencing (O'Neill BP *et al.*, 2013, ASH Annual Meeting Abstract). Alternatively, these results could also illustrate a high molecular heterogeneity within PCNSL as observed in extracerebral DLBCL exome studies [14]. However, our results contribute to the description of the PCNSL mutational landscape and provide insights into the prominent signaling pathways that are disrupted in PCNSL tumorigenesis. Genomic similarities with the ABC subtype of extracerebral DLBCL may open the possibility for parallels in therapeutic strategies of both lymphomas. For example, lenalidomide which induces IRF4 levels decrease [45], and ibrutinib which targets B-cell receptor signaling (Wilson WH *et al.*, 2012, ASH Annual Meeting Abstract) have shown promising results in extracerebral ABC-DLBCL. In this setting, they might also be attractive therapeutic strategies for PCNSL.

METHODS

PCNSL sample selection and patient characteristics

Thirty-seven PCNSL patients were selected for the present study. All tumors were classified as CD20+ DLBCL according to the WHO classification and demonstrated to contain at least 90% tumor cells based on morphology and immunohistochemistry. All the patients were newly diagnosed and immunocompetent. The participants provided written consent for sample collection and genetic analysis. This study was approved by the local ethical committee (CPPRB Pitié-Salpêtrière). Based on the high quality and sufficient levels of DNA, nine paired frozen tumor and blood tissues were selected to constitute the discovery set investigated by whole exome sequencing, and 28 tumor samples constituted the validation set investigated by direct sequencing. The sex ratio was 1.18 (male/female) and the median age at

diagnosis was 61 years, ranging from 17 to 83.

Isolation and quality assessment of DNA

Tumor DNA from 34 cryopreserved and 3 FFPE samples was extracted using the QIAamp DNA Mini Kit (Qiagen) and iPrep™ ChargeSwitch® Forensic Kit (Life Technologies), respectively, according to the manufacturer's instructions. A conventional saline method was used for the extraction of germline DNA from the blood samples. DNA was quantified using a NanoDrop spectrophotometer, and the quality was assessed on a 1% agarose gel.

Whole exome sequencing

Whole exome sequencing was possible for PCNSL patients with available paired frozen tumor and blood samples and with a minimal amount of 5 µg of tumor and germline DNA. Genomic DNA capture was performed using biotinylated oligonucleotides probes library (Human All Exon v2 – 46 Mb, Agilent) according to Agilent in-solution enrichment methodology (SureSelect Human All Exon Kits Version 2, Agilent). Sequence capture, enrichment and elution were performed according to manufacturer's instructions and protocols (SureSelect, Agilent). Massively parallel sequencing was realized on an Illumina GAIIX as paired-end 75 b reads.

Mapping and variant calling

Mapping of high-quality paired-end sequenced reads onto the GRCh37 build of the human reference genome was performed by Integragen using the Illumina ELAND 2 software tool. Raw alignments were first filtered for both low-quality mapped reads and assumed PCR duplicates with the SAMtools view (-q 20) and the Picard MarkDuplicates utilities, respectively [46]. The resulting filtered BAM files were subsequently confined to the genomic coordinates delineating the Agilent SureSelect 50-Mb probes using the intersectBed command of the BEDtools suite [47]. A commonly used combination of SAMtools mpileup and BCFtools view was then applied to the latter bounded alignments in order to call single nucleotide variations (SNVs) as well as short insertions and deletions (indels) within the targeted genomic regions. Mapping and coverage summary statistics were additionally obtained by an in-house post-processing of SAMtools idxstats and mpileup outputs.

Annotating called variants

Variant annotation was performed with the unpublished Genomic and Functional Annotation Pipeline

(GFAP) software, developed and routinely used at Institut Curie (<http://gfap.curie.fr/>). Briefly, GFAP consists of a set of tools that automatically: (i) retrieve and store suitable information from public variant databases such as 1000-genomes [48], dbSNP [49] or COSMIC [50, 51], (ii) match submitted variants against built-in databases and annotate them with respect to their genomic localization, (iii) assign an integrated functional impact prediction to non-synonymous variants (including stop-gains and losses) using dbNSFP database [12, 13] which compiles several tools such as SIFT [52] or Polyphen2 [53].

Validation set

Samples were selected based on the availability of tumor DNA. The validation set was investigated for known hotspot mutations by Sanger sequencing and for all exons of highly mutated genes by pyrosequencing. The tumor DNA was amplified using the primers listed in Supplementary Table 3. The amplification conditions were 94°C for 3 min followed by 45 cycles of 94°Cx15 sec, 60°Cx45 sec and 72°Cx1 min, with a final step at 72°C for 8 min. Exon 13 of *TBL1XR1* was amplified using Touch Down PCR with a gradient from 62 to 55°C during 6 cycles followed by 30 cycles at 55°C for primer annealing. The PCR products were purified according to the Agencourt® AMPure® XP PCR purification protocol (Beckman Coulter) with the Biomek® 3000 Automation Workstation.

Sanger sequencing

Sequencing reactions were performed in both orientations using the Big-Dye® Terminator Cycle Sequencing Ready Reaction (Perkin Elmer). The extension products were purified with the Agencourt® CleanSEQ® protocol according to the manufacturer's instructions (Beckman Coulter). The purified sequences were analyzed on an ABI Prism 3730 DNA Analyzer (Applied Biosystems). The forward and reverse sequences were visualized using Chromas Lite software.

Pyrosequencing

The universal tailed amplicon resequencing approach (454 Sequencing Technology, Roche) was used for coding exons sequencing. This system employs a second PCR, aiming MID (multiplex identifier) and 454 adaptors incorporation, an emulsion PCR according to the emPCR Amplification Method Manual Lib-A protocol (GS Junior Titanium Series, Roche), enrichment and pyrosequencing according to the Sequencing Method Manual (Roche). Sequences analysis was performed using CLC Genomics Workbench software.

ACKNOWLEDGMENTS

This work is part of the national program Cartes d'Identité des Tumeurs ® (CIT) <http://cit.ligue-cancer.net/> funded and developed by the Ligue nationale contre le cancer. This study benefited from the LOC study group network (réseau national de centres experts des lymphomes primitifs du SNC, INCa). The research leading to these results has received funding from the program "Investissements d'avenir" ANR-10-IAIHU-06 and from the Association pour la recherche sur les tumeurs cérébrales (ARTC). The biological resource centers of CHU Bordeaux and of Groupe Hospitalier Pitié Salpêtrière took part to samples collection.

Conflict of interest

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

A.B., B.B., K.L. R.D., Y.M., K.H-X. designed and performed research, analyzed data and wrote the paper. K.M., M.P., A.J., D.F-B, C.A., C.M., S.E., C.H., C.S. contributed to collect biological tissue and analyzed data.

REFERENCES

1. Ricard D, Idbaih A, Ducray F, Lahutte M, Hoang-Xuan K, Delattre J-Y. Primary brain tumours in adults. *Lancet*. 2012; 379: 1984-1996.
2. Montesinos-Rongen M, Brunn A, Bentink S, Basso K, Lim WK, Klapper W, Schaller C, Reifenberger G, Rubenstein J, Wiestler OD, Spang R, Dalla-Favera R, Siebert R, et al. Gene expression profiling suggests primary central nervous system lymphomas to be derived from a late germinal center B cell. *Leukemia*. 2008; 22: 400-405.
3. Tun HW, Personett D, Baskerville KA, Menke DM, Jaeckle KA, Kreinest P, Edenfield B, Zubair AC, O'Neill BP, Lai WR, Park PJ, McKinney M. Pathway analysis of primary central nervous system lymphoma. *Blood*. 2008; 111: 3200-3210.
4. Sung CO, Kim SC, Karnan S, Karube K, Shin HJ, Nam D-H, Suh Y-L, Kim S-H, Kim JY, Kim SJ, Kim WS, Seto M, Ko Y-H. Genomic profiling combined with gene expression profiling in primary central nervous system lymphoma. *Blood*. 2011; 117: 1291-1300.
5. Rubenstein JL, Fridlyand J, Shen A, Aldape K, Ginzinger D, Batchelor T, Treseler P, Berger M, McDermott M, Prados M, Karch J, Okada C, Hyun W, et al. Gene expression and angiotropism in primary CNS lymphoma. *Blood*. 2006; 107: 3716-3723.
6. Schwindt H, Vater I, Kreuz M, Montesinos-Rongen M, Brunn A, Richter J, Gesk S, Ammerpohl O, Wiestler OD, Hasenclever D, Deckert M, Siebert R. Chromosomal imbalances and partial uniparental disomies in primary central nervous system lymphoma. *Leukemia*. 2009; 23: 1875-1884.
7. Braggio E, McPhail ER, Macon W, Lopes MB, Schiff D, Law M, Fink S, Sprau D, Giannini C, Dogan A, Fonseca R, O'Neill BP. Primary central nervous system lymphomas: a validation study of array-based comparative genomic hybridization in formalin-fixed paraffin-embedded tumor specimens. *Clin Cancer Res*. 2011; 17: 4245-4253.
8. Gonzalez-Aguilar A, Idbaih A, Boisselier B, Habbita N, Rossetto M, Laurence A, Bruno A, Jouvet A, Polivka M, Adam C, Figarella-Branger D, Miquel C, Vital A, et al. Recurrent mutations of MYD88 and TBL1XR1 in primary central nervous system lymphomas. *Clin Cancer Res*. 2012; 18: 5203-5211.
9. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, Krzywinski M, Scott DW, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. 2011; 476: 298-303.
10. Parsons DW, Jones S, Zhang X, Lin JC-H, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu I-M, Gallia GL, Olivi A, McLendon R, Rasheed BA, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 2008; 321: 1807-1812.
11. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, Cruz-Gordillo P, Knoechel B, Asmann YW, Slager SL, Novak AJ, Dogan A, Ansell SM, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci USA*. 2012; 109: 3879-3884.
12. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat*. 2011; 32: 894-899.
13. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations. *Hum Mutat*. 2013; 34: 2393-2402.
14. Zhang J, Grubor V, Love CL, Banerjee A, Richards KL, Mieczkowski PA, Dunphy C, Choi W, Au WY, Srivastava G, Lugar PL, Rizzieri DA, Lagoo AS, et al. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc Natl Acad Sci USA*. 2013; 110: 1398-1403.
15. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, Wells VA, Grunn A, Messina M, Elliot O, Chan J, Bhagat G, Chadburn A, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet*. 2011; 43: 830-837.
16. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25: 25-29.

17. Iwasaki A, Medzhitov R. Regulation of adaptive immunity by the innate immune system. *Science*. 2010; 327: 291-295.
18. Ishii KJ, Akira S. Innate immune recognition of, and regulation by, DNA. *Trends Immunol*. 2006; 27: 525-532.
19. Dal Porto JM, Gauld SB, Merrell KT, Mills D, Pugh-Bernard AE, Cambier J. B cell antigen receptor signaling 101. *Mol Immunol*. 2004; 41: 599-613.
20. Montesinos-Rongen M, Godlewska E, Brunn A, Wiestler OD, Siebert R, Deckert M. Activating L265P mutations of the MYD88 gene are common in primary central nervous system lymphoma. *Acta Neuropathol*. 2011; 122: 791-792.
21. Montesinos-Rongen M, Schäfer E, Siebert R, Deckert M. Genes regulating the B cell receptor pathway are recurrently mutated in primary central nervous system lymphoma. *Acta Neuropathol*. 2012; 124: 905-906.
22. Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao W, Lim K-H, Kohlhammer H, Xu W, Yang Y, Zhao H, Shaffer AL, Romesser P, Wright G, et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature*. 2011; 470: 115-119.
23. Davis RE, Ngo VN, Lenz G, Tolar P, Young RM, Romesser PB, Kohlhammer H, Lamy L, Zhao H, Yang Y, Xu W, Shaffer AL, Wright G, et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature*. 2010; 463: 88-92.
24. Li J, Wang C-Y. TBL1-TBLR1 and beta-catenin recruit each other to Wnt target-gene promoter for transcription activation and oncogenesis. *Nat Cell Biol*. 2008; 10: 160-169.
25. Perissi V, Scafoglio C, Zhang J, Ohgi KA, Rose DW, Glass CK, Rosenfeld MG. TBL1 and TBLR1 phosphorylation on regulated gene promoters overcomes dual CtBP and NCoR/SMRT transcriptional repression checkpoints. *Mol Cell*. 2008; 29: 755-766.
26. Perissi V, Aggarwal A, Glass CK, Rose DW, Rosenfeld MG. A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. *Cell*. 2004; 116: 511-526.
27. Parker H, An Q, Barber K, Case M, Davies T, Konn Z, Stewart A, Wright S, Griffiths M, Ross FM, Moorman AV, Hall AG, Irving JA, et al. The complex genomic profile of ETV6-RUNX1 positive acute lymphoblastic leukemia highlights a recurrent deletion of TBL1XR1. *Genes Chromosomes Cancer*. 2008; 47: 1118-1125.
28. Zhang J, Mullighan CG, Harvey RC, Wu G, Chen X, Edmonson M, Buetow KH, Carroll WL, Chen I-M, Devidas M, Gerhard DS, Loh ML, Reaman GH, et al. Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Blood*. 2011; 118: 3080-3087.
29. Montesinos-Rongen M, Schmitz R, Brunn A, Gesk S, Richter J, Hong K, Wiestler OD, Siebert R, Küppers R, Deckert M. Mutations of CARD11 but not TNFAIP3 may activate the NF-kappaB pathway in primary CNS lymphoma. *Acta Neuropathol*. 2010; 120: 529-535.
30. Hock H, Meade E, Medeiros S, Schindler JW, Valk PJM, Fujiwara Y, Orkin SH. Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev*. 2004; 18: 2336-2341.
31. De Braekeleer E, Douet-Guilbert N, Morel F, Le Bris M-J, Basinko A, De Braekeleer M. ETV6 fusion genes in hematological malignancies: a review. *Leuk Res*. 2012; 36: 945-961.
32. De Silva NS, Simonetti G, Heise N, Klein U. The diverse roles of IRF4 in late germinal center B-cell differentiation. *Immunol Rev*. 2012; 247: 73-92.
33. Pongubala JMR, Northrup DL, Lancki DW, Medina KL, Treiber T, Bertolino E, Thomas M, Grosschedl R, Allman D, Singh H. Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5. *Nat Immunol*. 2008; 9: 203-215.
34. John SA, Clements JL, Russell LM, Garrett-Sinha LA. Ets-1 regulates plasma cell differentiation by interfering with the activity of the transcription factor Blimp-1. *J Biol Chem*. 2008; 283: 951-962.
35. Overbeck BM, Martin-Subero JI, Ammerpohl O, Klapper W, Siebert R, Giefing M. ETS1 encoding a transcription factor involved in B-cell differentiation is recurrently deleted and down-regulated in classical Hodgkin's lymphoma. *Haematologica*. 2012; 97: 1612-1614.
36. Flossbach L, Holzmann K, Mattfeldt T, Buck M, Lanz K, Held M, Möller P, Barth TFE. High-resolution genomic profiling reveals clonal evolution and competition in gastrointestinal marginal zone B-cell lymphoma and its large cell variant. *Int J Cancer*. 2013; 132: 116-127.
37. Morin RD, Mungall K, Pleasance E, Mungall AJ, Goya R, Huff R, Scott DW, Ding J, Roth A, Chiu R, Corbett RD, Chan FC, Mendez-Lago M, et al. Mutational and structural analysis of diffuse large B-cell lymphoma using whole genome sequencing. *Blood*. 2013; 122: 1256-1265.
38. Shirogane T, Fukada T, Muller JM, Shima DT, Hibi M, Hirano T. Synergistic roles for Pim-1 and c-Myc in STAT3-mediated cell cycle progression and antiapoptosis. *Immunity*. 1999; 11: 709-719.
39. Koeppel M, van Heeringen SJ, Smeenk L, Navis AC, Janssen-Megens EM, Lohrum M. The novel p53 target gene IRF2BP2 participates in cell survival during the p53 stress response. *Nucleic Acids Res*. 2009; 37: 322-335.
40. Rouault JP, Rimokh R, Tessa C, Paranhos G, Ffrench M, Duret L, Garoccio M, Germain D, Samarut J, Magaud JP. BTG1, a member of a new family of antiproliferative genes. *EMBO J*. 1992; 11: 1663-1670.
41. Montesinos-Rongen M, Van Roost D, Schaller C, Wiestler OD, Deckert M. Primary diffuse large B-cell lymphomas of the central nervous system are targeted by aberrant somatic hypermutation. *Blood*. 2004; 103: 1869-1875.
42. Narlik-Grassow M, Blanco-Aparicio C, Carnero A. The PIM family of serine/threonine kinases in cancer. *Med Res*

Rev. 2014; 34: 136-159.

43. Nyquist KB, Panagopoulos I, Thorsen J, Haugom L, Gorunova L, Bjerkehagen B, Fosså A, Guriby M, Nome T, Lothe RA, Skotheim RI, Heim S, Micci F. Whole-transcriptome sequencing identifies novel IRF2BP2-CDX1 fusion gene brought about by translocation t(1;5)(q42;q32) in mesenchymal chondrosarcoma. *PLoS ONE*. 2012; 7: 49705.
44. Camilleri-Broët S, Crinière E, Broët P, Delwail V, Mokhtari K, Moreau A, Kujas M, Raphaël M, Iraqi W, Sautès-Fridman C, Colombat P, Hoang-Xuan K, Martin A. A uniform activated B-cell-like immunophenotype might explain the poor prognosis of primary central nervous system lymphomas: analysis of 83 cases. *Blood*. 2006; 107: 190-196.
45. Yang Y, Shaffer AL 3rd, Emre NCT, Ceribelli M, Zhang M, Wright G, Xiao W, Powell J, Platig J, Kohlhammer H, Young RM, Zhao H, Yang Y, et al. Exploiting synthetic lethality for the therapy of ABC diffuse large B cell lymphoma. *Cancer Cell*. 2012; 21: 723-737.
46. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078-2079.
47. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26: 841-842.
48. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491: 56-65.
49. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001; 29: 308-311.
50. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet*. 2008; 10: 10-11.
51. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res*. 2011; 39: 945-950.
52. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009; 4: 1073-1081.
53. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7: 248-249.