

A genome-wide association study identifies susceptibility loci for primary central nervous system lymphoma at 6p25.3 and 3p22.1: a LOC network study.

Karim Labreche^{1,2}, Mailys Daniau^{2,3}, Amit Sud¹, Philip J. Law¹, Louis Royer-Perron^{2,4}, Amy Holroyd¹, Peter Broderick¹, Molly Went¹, Marion Benazra^{2,3}, Guido Ahle⁵, Pierre Soubeyran^{6,7}, Luc Taillandier⁸, Olivier L. Chinot^{9,10}, Olivier Casasnovas¹¹, Jacques-Olivier Bay¹², Fabrice Jardin¹³, Lucie Oberic¹⁴, Michel Fabbro¹⁵, Gandhi Damaj¹⁶, Annie Brion¹⁷, Karima Mokhtari^{2,18,19}, Cathy Philippe²⁰, Marc Sanson^{2,4,19}, Caroline Houillier², Carole Soussain²¹, Khê Hoang-Xuan^{2,4,*}, Richard S. Houlston^{1,*}, Agusti Alentorn^{2,4,*}, LOC Network[†].

¹ Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, Surrey SM2 5NG; UK

² Inserm, U 1127, ICM, F-75013 Paris, France; CNRS, UMR 7225, ICM, F-75013 Paris, France; Institut du Cerveau et de la Moelle épinière ICM, Paris 75013, France; Sorbonne Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France.

³ Institut du Cerveau et de la Moelle épinière, Plateforme iGenSeq, 47 Boulevard de l'Hôpital, 75013 Paris, France.

⁴ AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Service de neurologie 2-Mazarin, 75013 Paris, France. Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France

⁵ Department of Neurology, Hôpitaux Civils de Colmar, 68024, Colmar Cedex, France.

⁶ Department of Medical Oncology, Institut Bergonié, Bordeaux, F-33000, France

⁷ U1218 INSERM Research Unit, Bordeaux, F-33000, France.

⁸ Neuro-oncology Department, Nancy University Hospital and CRAN UMR 7039 CNRS, SBS BEAM Department, Nancy University, Vandoeuvre-lès-Nancy, France.

⁹Department of pathology and Neuropathology, Hôpital de la Timone, Aix-Marseille Univ, AP-HM, Marseille, 13005, France.

¹⁰AMU, CRO2, 13005, Marseille, 13005, France.

¹¹Department of Hematology, Dijon University Hospital, Dijon, 2100, Dijon, France.

¹²Department of Hematology, Clermont-Ferrand University Hospital, Clermont-Ferrand, 63003, France.

¹³Department of Hematology, Cancer Center Henri Henri Becquerel Center, 76000 Rouen, France and INSERM U1245, Cancer Center Henri Becquerel, Institute of Research and Innovation in Biomedicine, University of Normandy, Rouen, 76000, France.

¹⁴Department of Hematology, IUCT – Oncopole, 31100 Toulouse – France.

¹⁵Institut du Cancer Val d’Aurelle 34298 Montpellier Cedex 5 – France

¹⁶Department of Hematology, University Hospital of Caen, Caen, 14033, France.

¹⁷Department of Hematology, CHRU Besançon, Besançon, 25030, France

¹⁸AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Department of Neuropathology Raymond Escourolle, Paris, F-75013, France

¹⁹OncoNeuroTek, Institut du Cerveau et de la Moelle épinière, ICM, Paris, F-75013, France.

²⁰Neurospin Centre CEA, Saclay 91191, Gif sur Yvette, France.

²¹Department of Hematology, Hôpital René Huguenin, Institut Curie, 92210, Saint-Cloud, France

* These authors jointly supervised this work

¥ Corresponding author: Agusti Alentorn; Tel: 00 33 1 42 16 41 60; email: agusti.alentorn@aphp.fr

Inserm, U 1127, ICM, F-75013 Paris, France; CNRS, UMR 7225, ICM, F-75013 Paris, France; Institut du Cerveau et de la Moelle épinière ICM, Paris 75013, France; Sorbonne Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France. 47, bd de l’hôpital 75013 PARIS - France

Word count : 5,419 words

ABSTRACT

Background: Primary central nervous system lymphoma (PCNSL) is a rare form of extra-nodal non-Hodgkin lymphoma. PCNSL is a distinct subtype of non-Hodgkin lymphoma, with over 95% of tumors belonging to the diffuse large B-cell lymphoma (DLBCL) group. We have conducted a genome-wide association study (GWAS) on immunocompetent patients to address the possibility that common genetic variants influence the risk of developing PCNSL

Methods: We performed a meta-analysis of two new genome-wide association studies of PCNSL totaling 475 cases and 1,134 controls of European ancestry. To increase genomic resolution, we imputed >10 million single-nucleotide polymorphisms (SNPs) using the 1000 Genomes Project combined with UK10K as reference. In addition we performed a transcription factor binding disruption analysis and investigated the patterns of local chromatin patterns by capture Hi-C data.

Results: We identified independent risk loci at 3p22.1 (rs41289586, *ANO10*, $P = 2.17 \times 10^{-8}$) and 6p25.3 near *EXOC2* (rs116446171, $P = 1.95 \times 10^{-13}$). In contrast the lack of an association between rs41289586 and DLBCL, suggests distinct germline predisposition to PCNSL and DLBCL. We found looping chromatin interactions between non-coding regions at 6p25.3 (rs11646171) with the *IRF4* promoter and at 8q24.21 (rs13254990) with the MYC promoter, both genes with strong relevance to B-cell tumorigenesis.

Conclusion: To our knowledge this is the first study providing insight into the genetic predisposition to PCNSL. Our findings represent an important step in defining the contribution of common genetic variation to the risk of developing PCNSL.

Importance of the study

Primary CNS lymphomas (PCNSL) are a rare type of diffuse large B-cell lymphoma (DLBCL). Molecular studies in PCNSL patients have revealed similar patterns of molecular characteristics as those in nodal DLBCL. However, it is unknown whether there is a genetic predisposition to PCNSL. We performed a meta-analysis of two new genome-wide association studies of PCNSL analyzing the genotype of 475 patients and over 1000 healthy subjects patients to identify common genetic variants influencing the risk of PCNSL. We have identified independent risk loci associated with PCNSL. This finding advances our understanding of the genetic basis of PCNSL development.

INTRODUCTION

Primary diffuse large B-cell lymphoma of the central nervous system (PCNSL) is a rare tumor that accounts for $\leq 1\%$ of all lymphomas, and approximately 2% of all primary CNS tumors. The WHO classification of tumors of hematopoietic and lymphoid tissues recognizes PCNSL as a distinct subtype of non-Hodgkin lymphoma (NHL), with over 95% of tumors having comparative histology to the diffuse large B-cell lymphoma (DLBCL).

Immunocompromised individuals are considered most at risk of PCNSL, however, the incidence of the PCNSL is increasing in the immunocompetent populations and now represent the vast majority of patients. The disease typically follows an aggressive course and despite advances in the treatment of PCNSL is still associated with very high mortality.

Although PCNSL is strongly linked to Epstein-Barr virus (EBV) infection in immunocompromised patients, its detection is virtually absent in PCNSL from immunocompetent patients and little else is known about its etiology or risk factors in the population. To address the possibility that common genetic variants influence the risk of developing PCNSL, we have conducted a genome-wide association study (GWAS) on immunocompetent patients. Specifically, we performed a meta-analysis of two new GWAS of PCNSL and identify independent SNPs at 3p22.1 and 6p25.3 associated with risk.

MATERIALS AND METHODS

Subjects and ethics

This study was based on two primary GWAS datasets: (1) GWAS-1 comprised 346 immunocompetent HIV negative patients (184 male; median age 68 years) with PCNSL ascertained through the Service de Neurologie Mazarin, Groupe Hospitalier Pitié-Salpêtrière Paris and the Lymphome oculo-cerebral network (LOC) between 2008-2017, which serves all of France. For controls we made use of Illumina HumanHap 660 data on 788 individuals from the SU.VI.MAX (SUpplementation en VItamines et MinerauxAntioXydants) study healthy subjects (women aged 35–60 years; men aged 45–60 years). (2) GWAS-2 comprised 129 immunocompetent HIV negative patients (76 male; median age 69 years) with primary DLBCL CNS tumors ascertained through the Service de Neurologie Mazarin, Groupe Hospitalier Pitié-Salpêtrière Paris and LOC 2001-2007. For controls, we made use of second series of Illumina HumanHap 660 data generated on 346 individuals from the SU.VI.MAX. Collection of patient samples and associated clinico-pathological information was undertaken with written informed consent and ethical review board approval in accordance with the tenets of the declaration of Helsinki. The diagnosis of PCNSL (ICD-10 C83.3; WHO 9690/3) was established in accordance with WHO guidelines and all patient samples were obtained at first diagnosis.

Genotyping and quality control

Constitutional DNA was extracted from venous blood samples using QIAamp DNA Blood Mini Kit (Qiagen) (OncoNeuroTek, Paris) and quantified using Caliper LabchipGX and Nanodrop. Cases were genotyped using the Infinium OmniExpress-24 v1.2 BeadChip array according to the manufacturer's recommendations (Illumina Inc, San Diego, CA, USA). Standard quality control measures were applied to the GWAS. Specifically, individuals with low call rate (<99%) as well as all individuals with non-European ancestry (using the HapMap version 2 CEU, JPT/CHB and YRI

populations as a reference) were excluded. SNPs with a call rate <90% were excluded as were those with a minor allele frequency (MAF) < 0.01 or displaying significant deviation from Hardy-Weinberg equilibrium (*i.e.* $P < 10^{-6}$). GWAS data were imputed to >10 million SNPs with IMPUTE2 v2.3 software using a merged reference panel consisting of data from 1000 Genomes Project (phase 1 integrated release 3, March 2012) and UK10K. Genotypes were aligned to the positive strand in both imputation and genotyping. Imputation was conducted separately for each GWAS, and in each, the data were pruned to a common set of SNPs between cases and controls before imputation. Poorly imputed SNPs defined by an information measure <0.80 were excluded. Tests of association between imputed SNPs and *P*-values were calculated using logistic regression under an additive genetic model in SNPTESTv2.5. The adequacy of the case-control matching and possibility of differential genotyping of cases and controls were evaluated using Quantile-Quantile (Q-Q) plots of test statistics (**Supplementary Fig. 1**). The fidelity of rs116446171, rs41289586, and rs13254990 and rs10806525 imputation was confirmed by direct genotyping either by sequencing or by KasPar allele-specific PCR (**Supplementary Table 1**).

HLA imputation and analysis

To examine if specific coding variants within HLA genes contributed to association signals, we imputed the classical HLA alleles (A, B, C, DQA1, DQB1, DRB1) and coding variants across the HLA region (chr6:29–34 Mb) using SNP2HLA - <http://www.broadinstitute.org/mpg/snp2hla/>. Imputation was based on a reference panel from the Type 1 Diabetes Genetics Consortium (T1DGC), which comprises genotype data from 5,225 individuals of European descent typed for HLA-A, B, C, DRB1, DQA1, DQB1, DPB1, DPA1 4-digit alleles. A total of 8,961 classical HLA alleles (two- and four-digit resolution) and 1,873 AA markers including 580 AA positions that were ‘multi-allelic’, were successfully imputed (info score >0.8 for variant). Multi-allelic markers were analyzed as binary markers and a meta-analysis was conducted where we tested SNPs, HLA alleles and AAs across the HLA region for association with PCNSL using SNPTESTv2.5.

Meta-analysis

Meta-analyses were performed using the fixed-effects inverse-variance method based on the β estimates and standard errors from each study using META v1.6. Cochran's Q -statistic to test for heterogeneity, and the I^2 statistic to quantify the proportion of the total variation due to heterogeneity were calculated.

eQTL analysis

To examine the relationship between SNP genotype and gene expression we carried out Summary-data-based Mendelian Randomization (SMR) analysis as per Zhu *et al.*, 2016 (<http://cnsgenomics.com/software/smr/index.html>). We used publicly available lymphoblastoid cell line data from the Genotype-Tissue Expression (GTEx) (<http://www.gtexportal.org>) v6p release and MuTHR. Briefly, GWAS summary statistics files were generated from the meta-analysis. Reference files were generated from merging 1000 genomes phase 3 and UK10K (ALSPAC and TwinsUK) vcfs. Results from the SMR test for each of the five risk loci are reported in **Supplementary Data 1**. As previously advocated only probes with at least one expression quantity trait loci (eQTL) P -value of $< 5.0 \times 10^{-8}$ were considered for SMR analysis. We set a threshold for the SMR test of $PSMR < 7.57 \times 10^{-4}$ and $PSMR < 2.5 \times 10^{-3}$ corresponding to a Bonferroni correction for 66 tests (66 probes with a top eQTL $P < 5.0 \times 10^{-8}$ across the 5 loci and two LCL eQTL dataset) and 20 tests (20 probes with a top eQTL $P < 5.0 \times 10^{-8}$ across the 5 loci and Muther eQTL dataset) respectively.

Functional annotation

Novel risk SNPs and their proxies (*i.e.* $r^2 > 0.2$ in the 1000 Genomes EUR reference panel) were annotated for putative functional effect based upon histone mark ChIP-Seq data for H3K27ac, H3K4Me1 and H3K27Me3 from GM12878 (LCL) and primary B-cells. We searched for overlap

with “super-enhancer” regions as defined by Hnisz *et al*, restricting the analysis to the GM12878 cell line and CD19⁺ B-cells. The novel risk SNPs and their proxies ($r^2 > 0.2$ as above) were intersected with regions of accessible chromatin in CLL cells, as defined by Rendeiro *et al*, which were used as a surrogate for likely sites of transcription factor (TF) binding. SNPs falling within accessible sites (n=47) were taken forward to TF binding motif analysis and were also annotated for genomic evolutionary rate profiling (GERP) score as well as bound TFs based on ENCODE project ChIP-Seq data.

Transcription factor binding disruption analysis

To examine enrichment in specific TF binding across risk loci, we adapted the variant set enrichment method of Cowper-Salari *et al*. Briefly, for each risk locus, a region of strong LD (defined as $r^2 > 0.8$ and $D' > 0.8$) was determined, and SNPs within were termed the associated variant set (AVS). TF ChIP-Seq uniform peak data were obtained from ENCODE for the GM12878 cell line, which included data for 82 TFs. For each of these marks, the overlap of the SNPs in the AVS and the binding sites was determined to produce a mapping tally. A null distribution was produced by randomly selecting SNPs with the same characteristics as the risk-associated SNPs, and the null mapping tally calculated. This process was repeated 10,000 times, and approximate *P*-values were calculated as the proportion of permutations where the null mapping tally was greater or equal to the AVS mapping tally. An enrichment score was calculated by normalizing the tallies to the median of the null distribution. Thus, the enrichment score is the number of s.d.'s of the AVS mapping tally from the mean of the null distribution tallies.

RESULTS

Association analysis

After quality control, the two GWAS provided SNP genotypes on a total of 475 cases and 1,134 controls (**Supplementary Fig. 1 and 2 - Supplementary Tables 2 and 3**). To increase genomic resolution, we imputed >10 million SNPs using the 1000 Genomes Project combined with UK10K as reference. Q-Q plots for SNPs with MAF >0.5% post imputation showed only minimal evidence of over-dispersion (λ values for both GWAS = 1.00; **Supplementary Fig. 3**). Meta-analyzing test results from the two GWAS, we derived joint odds ratios (OR) per-allele and 95% confidence intervals (CI) under a fixed-effects model for each SNP and associated *P*-values.

Genome-wide significant associations ($P < 5 \times 10^{-8}$) were shown for loci at 3p22.1 (rs41289586, $P = 2.17 \times 10^{-8}$) and 6p25.3 (rs116446171, $P = 1.95 \times 10^{-13}$) (**Fig. 1, Table 1**). Conditional analysis of GWAS data showed no evidence for additional independent signals at either of the two risk loci.

Following on from this we examined whether other reported risk loci for DLBCL influenced PCNSL risk. Respective association *P*-values for the 6p21.22-HLA (human leukocyte antigen) (rs2523607) and 2p23.3 (rs79480871) risk SNPs were 0.023 and 0.14 (**Supplementary Table 4**).

HLA alleles

Variation at HLA has been linked to risk of DLBCL and a number of other B-cell tumors. The strongest SNP association at 6p21 (HLA) for PCNSL was provided by rs2395192 ($P = 1.81 \times 10^{-7}$), which maps between *HLA-DRA* and *HLA-DRB5* (**Supplementary Fig.5, Table 1**). To obtain additional insight into plausible functional variants within the HLA region, we imputed the classical HLA alleles and amino acid residues using SNP2HLA. No imputed HLA alleles or amino acid positions reached genome-wide significance (**Supplementary Fig. 5**). The strongest coding

changes within the HLA region were observed for the HLA class II alleles DRB1 Ser11Pro (AA_DRB1_11_32660115_SP, $P = 3.35 \times 10^{-6}$) and presence of the haplotype SRG (DRB1_13_32660109_SRG, $P = 3.35 \times 10^{-6}$) (**Supplementary Table 5**).

Functional annotation of risk loci

To gain insight into the biological basis underlying associations at 6p25.3 and other promising risk loci, we first evaluated each of the risk SNPs as well as the correlated variants using the online resources HaploRegv4, RegulomeDB and Fantom5 for evidence of functional effects (**Supplementary Data 2**). These data revealed regions of active chromatin at 6p25.3, 6q15 and 8q24 risk loci in B-cells. To explore whether there was an association between SNP genotype and transcript levels we performed an eQTL analysis using the GTEx project, MuTHR and blood eQTL data from Westra *et al*. We used SMR analysis to test for a concordance between signals from GWAS and *cis* eQTL for genes within 1 Mb of the sentinel and correlated SNPs ($r^2 > 0.8$) at each locus (**Supplementary Data 1**) and derived b_{XY} statistics, which estimate the effect of gene expression on PCNSL risk. After accounting for multiple testing we were unable to demonstrate any consistently significant eQTL for any of the risk loci examined. Chromatin looping interactions formed between enhancer elements and the promoters of genes they regulate map within distinct chromosomal topological associating domains. To identify patterns of local chromatin patterns, we analyzed promoter capture Hi-C data on the LCL cell line GM12878 as a source of B-cell information. Looping chromatin interactions were shown between non-coding regions at 6p25.3 (rs11646171) with the *IRF4* promoter (**Fig. 2**) and at 8q24.21 (rs13254990) with the *MYC* promoter; both genes with strong relevance to B-cell tumorigenesis.

Using ChIP-seq data on 82 TFs in GM12878 we examined for an over-representation of the binding of TFs at risk loci. Although not statistically significant the strongest TF bindings were shown for *TBLIXR1* that is mutated in 20% of PCNSL (**Supplementary Fig. 6**).

DISCUSSION

To our knowledge this is the first study providing insight into the genetic predisposition to PCNSL. While PCNSL is a specific entity it corresponds pathologically to diffuse large B-cell lymphoma. Hence, it is therefore perhaps not surprising that we identified associations, in common with DLBCL at 6q25.3 and 8q24.21 for PCNSL. However, the absence of associations at the 8q24.21 (rs4733601) and 2p23.3 (rs79480871) risk loci suggests the existence of a distinct developmental pathway for PCNSL, possibly reflective of its etiology. Moreover, analysis of publicly accessible GWAS data on DLBCL (NCI GWAS Stage 1) provides no support for an association between 3p22.1 (rs41289586) with DLBCL (**Supplementary Table 6**).

The 6p25.3 risk SNP rs116446171 (**Fig. 2**), which maps intergenic to *EXOC2* (exocyst complex component 2) and *IRF4* (interferon regulatory factor 4), has been previously shown to influence the risk of DLBCL. *EXOC2* is part of the multi-protein exocyst complex essential for polarized vesicle trafficking and the maintenance and intercellular transfer of viral proteins and virions. Furthermore, the RalB/*EXOC2* effector complex is a component of the TBK-1-dependent innate signaling pathway. While thus far there is no evidence to implicate *EXOC2* in lymphoma, the RalB/*EXOC2* complex may contribute to tumor cell survival. In contrast, *IRF4* has a well-established role in the development of many B-cell malignancies.

The 3p22.1 risk SNP rs41289586 (**Fig. 2**) localizes to exon 6 of the anoctamin 10 gene (*ANO10*) and is responsible for the rare missense change (*ANO10*:c.788G>A, p.Arg263His). Although rs1052501 at 3p22.1 has been reported to be associated with risk of multiple myeloma, another B-cell malignancy, this SNP maps >1Mb away from rs41289586 (pairwise $r^2 = 0.0002$). Inherited defects in *ANO10*, which encodes a calcium-activated chloride channel transmembrane protein are a cause of autosomal recessive spinocerebellar ataxia. While other anoctamins have been implicated in

cancer, to date there is no evidence for the role of *ANO10* in a B-cell malignancy. However, intriguingly, rs41289586 has been associated with regulation of macrophage response and associated with *Borrelia* seropositivity, implicating *ANO10* in the innate immune defense.

In addition to the 6p25.3 and 3p22.1 risk loci we identified promising associations ($P < 2 \times 10^{-7}$), at 6q15 (rs10806425, $P = 1.36 \times 10^{-7}$) and 8q24.21 (rs13254990; $P = 1.33 \times 10^{-7}$) annotating genes with strong relevance to B-cell tumorigenesis (**Table 1, Supplementary Fig. 4**). rs10806425 localizes to intron 1 of the gene encoding *BACH2* (basic leucine zipper transcription factor 2). Loss of heterozygosity of *BACH2* has been reported at a frequency of 20% in B-cell lymphoma. In DLBCL patients with higher *BACH2* expression tend to have a better prognosis. *BACH2* is a key regulator of the pre-BCR checkpoint as well as a tumor suppressor in pre-B acute lymphoblastic leukemia. One mechanism of *BACH2* downregulation in leukemia is the loss of the transcription factor *PAX5*, which is intriguingly, commonly mutated in both PCNSL and B-cell ALL.

The 8q24 SNP rs13254990 localizes to intron 4 of *PVT1*, a non-coding RNA affecting the activation of MYC. Two independent risk loci at 8q24 defined by SNPs rs13255592 and rs4733601 have previously been shown to influence DLBCL. rs13255592 also localizes within intron 4 of *PVT1* and is highly correlated with rs13254990 ($r^2 = 0.98$, $P = 3.81 \times 10^{-7}$). No association between rs4733601, which maps approximately 1.9Mb telomeric to *PVT1*, and PCNSL risk was however shown ($r^2 = 4.21 \times 10^{-5}$, $P = 0.99$; **Supplementary Table 4**). The 8q24.21 128-130Mb genomic interval harbors multiple independent risk loci with different tumor specificities (**Supplementary Table 7**). The strongest additional association for PCNSL being shown by the Hodgkin lymphoma risk SNP rs2019960 ($P = 4.1 \times 10^{-5}$) raising the possibility of an additional risk locus for the disease at 8q24.21.

Although in part speculative, the 6q25.3 association implicates *IRF4* in the development of PCNSL. Through interaction with transcription factors including PU.1, IRF4 controls the termination of pre-B-cell receptor signaling and promotes the differentiation of pro-B cells to small B cells. Furthermore, via BLIMP1 and BCL6, IRF4 controls the transition of memory B cells. The observation that *PVT1* rearrangement occurs frequently in highly aggressive B-cell lymphomas harboring an 8q24 abnormality suggest that germline variation in this region may influence PCNSL risk. The 6q15 association implicates *BACH2* in the development of PCNSL. *BACH2* is an attractive candidate *a priori* for having a role in PCNSL development being a regulator of the antibody response mediating effects through *BLIMP1*, *XBPI1*, *LRF4*, and *PAX5*. Moreover, *BACH2* regulates the activity of the tumor suppressor c-Rel in lymphoma development. Collectively these data are consistent with aberrant B-cell developmental pathways being central for predisposition to PCNSL. The finding of a relationship between ANO10:c.788G>A, p.Arg263His with PCNSL, but not classical DLBCL highlights differences in biological etiology with this lymphoma. Intriguingly, this variant has previously been shown to influence macrophage response thereby implicating *ANO10* in innate immune defense with the development of PCNSL. While not statistically significant, the *HLA-DRA* and *HLA-DRB1* associations are also of relevance as these alleles have previously been shown to influence the human reaction to viral load and EBV infection, respectively. The linkage of these genes to the development of PCNSL is therefore entirely consistent with an infective basis to this B-cell malignancy even though none of the patients we studied were immunocompromised.

Inevitably constrained by the sheer rarity of PCNSL we acknowledge that a limitation of our study has been an inability to replicate study findings in additional series. Another limitation is the absence of case-case study with other DLBCL. This could be performed in future analyses. Our findings are however based on a meta-analysis of two series cohorts of PCNSL. Moreover, despite the rarity of PCBSL by ascertaining patients through the LOC Network has provided us with greater power to detect associations than the smaller studies of other rare lymphomas.

In summary, our findings represent an important step in defining the contribution of common genetic variation to the risk of developing PCNSL. Our observations are notable since the associations highlighted define regions of the genome harboring plausible candidate genes for further investigation. Given the relatively modest size of our analysis, it is highly probable that further studies will discover additional common susceptibility loci. These coupled with functional analyses should provide for an explanation of the biological underpinnings of PCNSL.

DATA AVAILABILITY

Genotype data that support the findings of this study have been deposited in the database of the European Genome-phenome Archive (EGA) with accessions code PRJEB21814. NCI Non-Hodgkin Lymphoma GWAS data was obtained through dbGAP (phs000801.v2.p1). The remaining data are contained within the paper and Supplementary files are available from the authors upon request.

FUNDINGS

The primary source of funding was provided by la Ligue Nationale Contre le Cancer-RE 2015 (K.H.X), French National Institute of Cancer (InCa) LOC Network (K.H.X and C.S). A.A has also been granted with a “poste d’accueil AP-HP-CEA”. K.L is supported by l’Association pour la Recherche sur les Tumeurs Cérébrales (ARTC) and Institute CARNOT – Institut du Cerveau et de la Moelle Epinière (ICM). Finally, also acknowledge support from Cancer Research UK and Bloodwise. A.S. is supported by a clinical fellowship from Cancer Research UK. We are grateful to all investigators and all the patients and individuals for their participation. Samples from AP-HM were retrieved from AP-HM tumor bank, authorization number 2013-1786. We also thank the clinicians, other hospital staff and study staff that contributed to the blood sample and data collection for this study and OncoNeuroTek that provided and prepared DNA samples. Genotypes from NCI Non-Hodgkin Lymphoma GWAS were accessed through dbGaP accession phs000801.v2.p1.

AUTHORSHIP

A.A. and K.H.X., developed the project and provided overall project management; K.L., M.D., R.S.H., K.H.X. and A.A. drafted the manuscript. K.L., A.S., P.J.L. and M.W. performed bioinformatic and statistical analyses; Patient samples and phenotype data were provided by C.D., D.G., K.H.X., C.S. and other members of the LOC Network. M.D., I.D., L.R.P., A.R., D.G.

performed project management and supervised genotyping; M.D., I.D., A.R., M.B. and A.H. performed sequencing and genotyping. A.A., K.H.X., M.D., A.R., L.R.P. and P.B. supervised laboratory management and oversaw genotyping of cases; D.G., M.D., I.D., L.R.P. performed sample management of cases. All authors reviewed and approved the manuscript prior to submission.

CONFLICT OF INTEREST

The remaining authors declare no competing financial interests.

REFERENCES

2. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016; 127(20):2375-2390.
3. Hoang-Xuan K, Bessell E, Bromberg J, et al. Diagnosis and treatment of primary CNS lymphoma in immunocompetent patients: guidelines from the European Association for Neuro-Oncology. *Lancet Oncol*. 2015; 16(7):e322-332.
4. Bessell EM, Dickinson P, Dickinson S, Salmon J. Increasing age at diagnosis and worsening renal function in patients with primary central nervous system lymphoma. *J Neurooncol*. 2011; 104(1):191-193.
5. Villano JL, Koshy M, Shaikh H, Dolecek TA, McCarthy BJ. Age, gender, and racial differences in incidence and survival in primary CNS lymphoma. *Br J Cancer*. 2011; 105(9):1414-1418.
6. O'Neill BP, Decker PA, Tieu C, Cerhan JR. The changing incidence of primary central nervous system lymphoma is driven primarily by the changing incidence in young and middle-aged men and differs from time trends in systemic diffuse large B-cell non-Hodgkin's lymphoma. *Am J Hematol*. 2013; 88(12):997-1000.
7. Bashir R, Luka J, Cheloha K, Chamberlain M, Hochberg F. Expression of Epstein-Barr virus proteins in primary CNS lymphoma in AIDS patients. *Neurology*. 1993; 43(11):2358-2362.
8. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc*. 2010; 5(9):1564-1573.
9. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009; 5(6):e1000529.

10. Genomes Project C, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467(7319):1061-1073.
11. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun*. 2015; 6:8111.
12. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007; 39(7):906-913.
13. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One*. 2013; 8(6):e64683.
14. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet*. 2010; 42(5):436-440.
15. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002; 21(11):1539-1558.
16. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016; 48(5):481-487.
17. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013; 45(6):580-585.
18. Grundberg E, Small KS, Hedman AK, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet*. 2012; 44(10):1084-1089.
19. de Souza N. The ENCODE project. *Nature methods*. 2012; 9(11):1046.
20. Rendeiro AF, Schmidl C, Strefford JC, et al. Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. *Nat Commun*. 2016; 7:11938.
21. Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013; 155(4):934-947.

22. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol.* 2010; 6(12):e1001025.
23. Cowper-Salari R, Zhang X, Wright JB, et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet.* 2012; 44(11):1191-1198.
24. Cerhan JR, Berndt SI, Vijai J, et al. Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. *Nat Genet.* 2014; 46(11):1233-1238.
25. Conde L, Halperin E, Akers NK, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nat Genet.* 2010; 42(8):661-664.
26. Wang SS, Abdou AM, Morton LM, et al. Human leukocyte antigen class I and II alleles in non-Hodgkin lymphoma etiology. *Blood.* 2010; 115(23):4820-4823.
27. Law PJ, Berndt SI, Speedy HE, et al. Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. *Nat Commun.* 2017; 8:14175.
28. Sud A, Thomsen H, Law PJ, et al. Genome-wide association study of classical Hodgkin lymphoma identifies key regulators of disease susceptibility. *Nat Commun.* 2017; In press.
29. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* 2016; 44(D1):D877-881.
30. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012; 22(9):1790-1797.
31. Kawaji H, Severin J, Lizio M, et al. The FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation. *Genome Biol.* 2009; 10(4):R40.
32. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013; 45(10):1238-1243.

33. Mifsud B, Tavares-Cadete F, Young AN, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet.* 2015; 47(6):598-606.
34. Bruno A, Boisselier B, Labreche K, et al. Mutational analysis of primary central nervous system lymphoma. *Oncotarget.* 2014; 5(13):5065-5075.
35. Mukerji J, Olivieri KC, Misra V, Agopian KA, Gabuzda D. Proteomic analysis of HIV-1 Nef cellular binding partners reveals a role for exocyst complex proteins in mediating enhancement of intercellular nanotube formation. *Retrovirology.* 2012; 9:33.
36. Chien Y, Kim S, Bumeister R, et al. RalB GTPase-mediated activation of the IkappaB family kinase TBK1 couples innate immune signaling to tumor cell survival. *Cell.* 2006; 127(1):157-170.
37. Acquaviva J, Chen X, Ren R. IRF-4 functions as a tumor suppressor in early B-cell development. *Blood.* 2008; 112(9):3798-3806.
38. Pathak S, Ma S, Trinh L, et al. IRF4 is a suppressor of c-Myc induced B cell leukemia. *PLoS One.* 2011; 6(7):e22628.
39. Kreher S, Johrens K, Strehlow F, et al. Prognostic impact of B-cell lymphoma 6 in primary CNS lymphoma. *Neuro Oncol.* 2015; 17(7):1016-1021.
40. Broderick P, Chubb D, Johnson DC, et al. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat Genet.* 2011; 44(1):58-61.
41. Renaud M, Anheim M, Kamsteeg EJ, et al. Autosomal recessive cerebellar ataxia type 3 due to ANO10 mutations: delineation and genotype-phenotype correlation study. *JAMA Neurol.* 2014; 71(10):1305-1310.
42. Wanitchakool P, Wolf L, Koehl GE, et al. Role of anoctamins in cancer and apoptosis. *Philos Trans R Soc Lond B Biol Sci.* 2014; 369(1638):20130096.
43. Hammer C, Wanitchakool P, Sirianant L, et al. A Coding Variant of ANO10, Affecting Volume Regulation of Macrophages, Is Associated with Borrelia Seropositivity. *Mol Med.* 2015; 21:26-37.

44. Sasaki S, Ito E, Toki T, et al. Cloning and expression of human B cell-specific transcription factor BACH2 mapped to chromosome 6q15. *Oncogene*. 2000; 19(33):3739-3749.
45. Sakane-Ishikawa E, Nakatsuka S, Tomita Y, et al. Prognostic significance of BACH2 expression in diffuse large B-cell lymphoma: a study of the Osaka Lymphoma Study Group. *J Clin Oncol*. 2005; 23(31):8012-8017.
46. Swaminathan S, Huang C, Geng H, et al. BACH2 mediates negative selection and p53-dependent tumor suppression at the pre-B cell receptor checkpoint. *Nat Med*. 2013; 19(8):1014-1022.
47. 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(5):1869-1875.
48. Rickert RC. New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. *Nat Rev Immunol*. 2013; 13(8):578-591.
49. Schmidlin H, Diehl SA, Blom B. New insights into the regulation of human B-cell differentiation. *Trends Immunol*. 2009; 30(6):277-285.
50. McDonnell TJ, Korsmeyer SJ. Progression from lymphoid hyperplasia to high-grade malignant lymphoma in mice transgenic for the t(14; 18). *Nature*. 1991; 349(6306):254-256.
51. Taub R, Kirsch I, Morton C, et al. Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proc Natl Acad Sci U S A*. 1982; 79(24):7837-7841.
52. Ladanyi M, Offit K, Jhanwar SC, Filippa DA, Chaganti RS. MYC rearrangement and translocations involving band 8q24 in diffuse large cell lymphomas. *Blood*. 1991; 77(5):1057-1063.
53. Muto A, Tashiro S, Nakajima O, et al. The transcriptional programme of antibody class switching involves the repressor Bach2. *Nature*. 2004; 429(6991):566-571.

54. Hunter JE, Butterworth JA, Zhao B, et al. The NF-kappaB subunit c-Rel regulates Bach2 tumour suppressor expression in B-cell lymphoma. *Oncogene*. 2016; 35(26):3476-3484.
55. Hammer C, Begemann M, McLaren PJ, et al. Amino Acid Variation in HLA Class II Proteins Is a Major Determinant of Humoral Response to Common Viruses. *Am J Hum Genet*. 2015; 97(5):738-743.
56. Li Z, Xia Y, Feng LN, et al. Genetic risk of extranodal natural killer T-cell lymphoma: a genome-wide association study. *Lancet Oncol*. 2016; 17(9):1240-1247.
57. Scales M, Jager R, Migliorini G, Houlston RS, Henrion MY. visPIG--a web tool for producing multi-region, multi-track, multi-scale plots of genetic data. *PLoS One*. 2014; 9(9):e107497.

FIGURE AND TABLE LEGENDS

Figure 1: Manhattan plot of association P -values. Shown are the genome-wide $-\log_{10}P$ -values (two-sided) of >10 million successfully imputed autosomal SNPs in 475 cases and 1,134 controls. The red horizontal line represents the genome-wide significance threshold of $P = 5.0 \times 10^{-8}$.

Figure 2: Regional plots of association results and recombination rates for new risk loci for primary cerebral nervous system lymphoma. Results shown for (a) 6p25 and (b) 3q21. Plots (drawn using visPig) show association results of both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates. $-\log_{10}P$ values (y -axis) of the SNPs are shown according to their chromosomal positions (x -axis). The sentinel SNP in each combined analysis is shown as a large circle or triangle and is labelled by its rsID. The color intensity of each symbol reflects the extent of LD with the top genotyped SNP, white ($r^2 = 0$) through to dark red ($r^2 = 1.0$). Genetic recombination rates, estimated using 1000 Genomes Project samples, are shown with a light blue line. Physical positions are based on NCBI build 37 of the human genome. Also shown are the chromatin-state segmentation track (ChromHMM) for lymphoblastoid cells using data from the HapMap ENCODE Project, and the positions of genes and transcripts mapping to the region of association. The top track represents capture Hi-C promoter contacts in GM12878 cells. The colour intensity of each contact reflects the interaction score.

Table 1: Summary results for SNPs associated with primary central nervous system lymphoma risk

† **LOC Network:** Marie-Pierre Moles-Moreau¹, Rémy Gressin², Vincent Delwail³, Franck Morschhauser⁴, Philippe Agapé⁵, Arnaud Jaccard⁶, Hervé Ghesquieres⁷, Adrian Tempescul⁸, Emmanuel Gyan⁹, Jean-Pierre Marolleau¹⁰, Roch Houot¹¹, Luc Fornecker¹², Anna-Luisa Di Stefano¹³, Inès Detrait¹⁴, Amithys Rahimian¹⁵, Mark Lathrop¹⁶, Diane Genet¹⁷, Frédéric Davi¹⁸, Nathalie Cassoux¹⁹, Valérie Touitou²⁰, Sylvain Choquet²¹, Anne Vital²², Marc Polivka²³, Dominique Figarella-Branger^{24,25}, Alexandra Benouaich-Amiel²⁶, Chantal Campello²⁷, Frédéric Charlotte²⁸, Nadine Martin-Duverneuil²⁹, Loïc Feuvret³⁰, Aurélie Kas³¹, Soledad Navarro³², Chiara Villa³³, Franck Bielle³⁴, Fabrice Chretien³⁵, Marie Christine Tortel³⁶, Guillaume Gauchotte³⁷, Emmanuelle Uro-Coste³⁸, Catherine Godfrain³⁹, Valérie Rigau⁴⁰, Myrto Costopoulos¹⁸, Magalie Le Garff-Tavernier¹⁸, David Meyronnet⁴¹, Audrey Rousseau⁴², Clovis Adam⁴³, Thierry Lamy⁴⁴, Cécile Chabrot⁴⁵, Eileen M. Boyle⁴⁶, Marie Blonski⁴⁷, Anna Schmitt⁴⁸.

1. Department of Hematology, Angers University Hospital, Angers, 49033, France.
2. Department of Hematology CHU Grenoble Michallon 38043 Grenoble Cedex 02 – France
3. Service d'Oncologie Hématologique et de Thérapie Cellulaire, CHU de Poitiers, INSERM, CIC 1402, Poitiers, Centre d'Investigation Clinique, Université de Poitiers, Poitiers, France
4. Department of Hematology, CHRU Lille, Lille, 59037, France
5. Institut de Cancérologie – 44800 Saint Herblain – France
6. Department of Hematology CHU Dupuytren 87042 Limoges- France
7. Department of Hematology, University Hospital of Lyon, 69002, Lyon, France
8. Department of Hematology CHU Morvan 29609 Brest Cedex – France
9. Department of Hematology CHU Bretonneau 34044 Tours, France
10. Department of Hematology, University Hospital of Amiens, 80054, Amiens, France.
11. CHU Rennes, Service Hématologie Clinique, F-35033 Rennes, France

12. Department of Hematology CHU Strasbourg 67000 Strasbourg – France
13. Department of Neurology Hôpital Foch 92151 Suresnes – France
14. Inserm, U 1127, ICM, F-75013 Paris, France; CNRS, UMR 7225, ICM, F-75013 Paris, France; Institut du Cerveau et de la Moelle épinière ICM, Paris 75013, France; Sorbonne Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France; AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Service de neurologie 2-Mazarin, 75013 Paris, France.
15. OncoNeuroTek, Institut du Cerveau et de la Moelle épinière, ICM, Paris, F-75013, France.
16. McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada, H3A 0G1.
17. AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Service de neurologie 2-Mazarin, 75013 Paris, France. Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France
18. Department of Biological Hematology, AP-HP, Groupe Hospitalier Pitié-Salpêtrière, 75013, Paris, France.
19. Department of Oncological Ophthalmology, Institut Curie, Paris, France.
20. AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Department of Ophthalmology, Paris, 75013, France.
21. AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Department of Hematology, Paris, 75013, France
22. CNRS, Institut des Maladies Neurodégénératives, UMR 5293, F-33000 Bordeaux, France; Department of Pathology, Bordeaux University Hospital, Bordeaux, France
23. Department of Pathology, CHU Paris-GH St-Louis Lariboisière F.Widal - Hôpital Lariboisière, 75010, Paris, France.

24. Department of pathology and Neuropathology, Hôpital de la Timone, Aix-Marseille Univ, AP-HM, Marseille, 13005, France.
25. AMU, CRO2, 13005, Marseille, 13005, France.
26. Department of Neurology, CHU Toulouse, 31059 Toulouse, France.
27. Service de Neuro-Oncologie, CHU Timone, 13005 Marseille – France.
28. Department of Pathology, CHU Pitié-Salpêtrière, 75013 Paris, France.
29. Department of Neuroradiology, CHU Pitié-Salpêtrière, 75013 Paris, France.
30. Department of Radiotherapy, CHU Pitié-Salpêtrière, 75013 Paris, France.
31. Department of Nuclear Medicine, CHU Pitié-Salpêtrière, 75013 Paris, France.
32. Department of neurosurgery, CHU Pitié-Salpêtrière, 75013 Paris, France.
33. Department of Pathology, Hôpital Foch, 92151 Suresnes – France.
34. Inserm, U 1127, ICM, F-75013 Paris, France ; CNRS, UMR 7225, ICM, F-75013 Paris, France ; Institut du Cerveau et de la Moelle épinière ICM, Paris 75013, France ; Sorbonne Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France ; AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Service de Neuropathologie, 75013 Paris, France.
35. Department of Neuropathology, Centre Hospitalier Sainte Anne, Paris, France.
36. Department of Pathology, Hôpitaux Civils de Colmar, 68024, Colmar Cedex, France.
37. Department of Pathology, Nancy University Hospital, Vandoeuvre-lès-Nancy, France
38. Department of Pathology, CLCC Institut Claudius Regaud, 31059 Toulouse, France.
39. Department of Pathology, CHU Clermont-Ferrand, 63000 Clermont-Ferrand, France.
40. Department of Pathology, CHU Montpellier, 34000 Montpellier, France.
41. Department of Pathology, University Hospital of Lyon, 69002, Lyon, France.
42. Department of Pathology, CHU Angers, 49933 Angers – France.
43. Department of Pathology, CHU Kremlin Bicêtre, 94270 Le Kremlin Bicêtre.
44. CHU Rennes, Rennes, 35033, France.

45. CHU Clermont Ferrand, Clermont Ferrand, 63000, France.
46. Department of Haematology, Lille University Hospital, Lille, 59037, France.
47. CHU Nancy. Nancy, 54500, France.
48. CHU Bordeaux. Bordeaux ,33000, France.

A genome-wide association study identifies susceptibility loci for primary central nervous system lymphoma at 6p25.3 and 3p22.1: a LOC network study.

Karim Labreche^{1,2}, Mailys Daniau^{2,3}, Amit Sud¹, Philip J. Law¹, Louis Royer-Perron^{2,4}, Amy Holroyd¹, Peter Broderick¹, Molly Went¹, Marion Benazra^{2,3}, Guido Ahle⁵, Pierre Soubeyran^{6,7}, Luc Taillandier⁸, Olivier L. Chinot^{9,10}, Olivier Casasnovas¹¹, Jacques-Olivier Bay¹², Fabrice Jardin¹³, Lucie Oberic¹⁴, Michel Fabbro¹⁵, Gandhi Damaj¹⁶, Annie Brion¹⁷, Karima Mokhtari^{2,18,19}, Cathy Philippe²⁰, Marc Sanson^{2,4,19}, Caroline Houillier², Carole Soussain²¹, Khê Hoang-Xuan^{2,4,*}, Richard S. Houlston^{1,*}, Agusti Alentorn^{2,4,*}, LOC Network[†].

¹ Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, Surrey SM2 5NG; UK

² Inserm, U 1127, ICM, F-75013 Paris, France; CNRS, UMR 7225, ICM, F-75013 Paris, France; Institut du Cerveau et de la Moelle épinière ICM, Paris 75013, France; Sorbonne Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France.

³ Institut du Cerveau et de la Moelle épinière, Plateforme iGenSeq, 47 Boulevard de l'Hôpital, 75013 Paris, France.

⁴ AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Service de neurologie 2-Mazarin, 75013 Paris, France. Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France

⁵ Department of Neurology, Hôpitaux Civils de Colmar, 68024, Colmar Cedex, France.

⁶ Department of Medical Oncology, Institut Bergonié, Bordeaux, F-33000, France

⁷ U1218 INSERM Research Unit, Bordeaux, F-33000, France.

⁸ Neuro-oncology Department, Nancy University Hospital and CRAN UMR 7039 CNRS, SBS BEAM Department, Nancy University, Vandoeuvre-lès-Nancy, France.

⁹Department of pathology and Neuropathology, Hôpital de la Timone, Aix-Marseille Univ, AP-HM, Marseille, 13005, France.

¹⁰AMU, CRO2, 13005, Marseille, 13005, France.

¹¹Department of Hematology, Dijon University Hospital, Dijon, 2100, Dijon, France.

¹²Department of Hematology, Clermont-Ferrand University Hospital, Clermont-Ferrand, 63003, France.

¹³Department of Hematology, Cancer Center Henri Henri Becquerel Center, 76000 Rouen, France and INSERM U1245, Cancer Center Henri Becquerel, Institute of Research and Innovation in Biomedicine, University of Normandy, Rouen, 76000, France.

¹⁴Department of Hematology, IUCT – Oncopole, 31100 Toulouse – France.

¹⁵Institut du Cancer Val d’Aurelle 34298 Montpellier Cedex 5 – France

¹⁶Department of Hematology, University Hospital of Caen, Caen, 14033, France.

¹⁷Department of Hematology, CHRU Besançon, Besançon, 25030, France

¹⁸AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Department of Neuropathology Raymond Escourrolle, Paris, F-75013, France

¹⁹OncoNeuroTek, Institut du Cerveau et de la Moelle épinière, ICM, Paris, F-75013, France.

²⁰Neurospin Centre CEA, Saclay 91191, Gif sur Yvette, France.

²¹Department of Hematology, Hôpital René Huguenin, Institut Curie, 92210, Saint-Cloud, France

* These authors jointly supervised this work

¥ Corresponding author: Agusti Alentorn; Tel: 00 33 1 42 16 41 60; email: agusti.alentorn@aphp.fr

Inserm, U 1127, ICM, F-75013 Paris, France; CNRS, UMR 7225, ICM, F-75013 Paris, France; Institut du Cerveau et de la Moelle épinière ICM, Paris 75013, France; Sorbonne Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France. 47, bd de l’hôpital 75013 PARIS - France

Word count : 5,419 words

ABSTRACT

Background: Primary central nervous system lymphoma (PCNSL) is a rare form of extra-nodal non-Hodgkin lymphoma. PCNSL is a distinct subtype of non-Hodgkin lymphoma, with over 95% of tumors belonging to the diffuse large B-cell lymphoma (DLBCL) group. We have conducted a genome-wide association study (GWAS) on immunocompetent patients to address the possibility that common genetic variants influence the risk of developing PCNSL

Methods: We performed a meta-analysis of two new genome-wide association studies of PCNSL totaling 475 cases and 1,134 controls of European ancestry. To increase genomic resolution, we imputed >10 million single-nucleotide polymorphisms (SNPs) using the 1000 Genomes Project combined with UK10K as reference. In addition we performed a transcription factor binding disruption analysis and investigated the patterns of local chromatin patterns by capture Hi-C data.

Results: We identified independent risk loci at 3p22.1 (rs41289586, *ANO10*, $P = 2.17 \times 10^{-8}$) and 6p25.3 near *EXOC2* (rs116446171, $P = 1.95 \times 10^{-13}$). In contrast the lack of an association between rs41289586 and DLBCL, suggests distinct germline predisposition to PCNSL and DLBCL. We found looping chromatin interactions between non-coding regions at 6p25.3 (rs11646171) with the *IRF4* promoter and at 8q24.21 (rs13254990) with the *MYC* promoter, both genes with strong relevance to B-cell tumorigenesis.

Conclusion: To our knowledge this is the first study providing insight into the genetic predisposition to PCNSL. Our findings represent an important step in defining the contribution of common genetic variation to the risk of developing PCNSL.

Importance of the study

Primary CNS lymphomas (PCNSL) are a rare type of diffuse large B-cell lymphoma (DLBCL). Molecular studies in PCNSL patients have revealed similar patterns of molecular characteristics as those in nodal DLBCL. However, it is unknown whether there is a genetic predisposition to PCNSL. We performed a meta-analysis of two new genome-wide association studies of PCNSL analyzing the genotype of 475 patients and over 1000 healthy subjects patients to identify common genetic variants influencing the risk of PCNSL. We have identified independent risk loci associated with PCNSL. This finding advances our understanding of the genetic basis of PCNSL development.

INTRODUCTION

Primary diffuse large B-cell lymphoma of the central nervous system (PCNSL) is a rare tumor that accounts for $\leq 1\%$ of all lymphomas, and approximately 2% of all primary CNS tumors¹. The WHO classification of tumors of hematopoietic and lymphoid tissues recognizes PCNSL as a distinct subtype of non-Hodgkin lymphoma (NHL)², with over 95% of tumors having comparative histology to the diffuse large B-cell lymphoma (DLBCL)³.

Immunocompromised individuals are considered most at risk of PCNSL, however, the incidence of the PCNSL is increasing in the immunocompetent populations and now represent the vast majority of patients⁴⁻⁶. The disease typically follows an aggressive course and despite advances in the treatment of PCNSL is still associated with very high mortality³.

Although PCNSL is strongly linked to Epstein-Barr virus (EBV) infection in immunocompromised patients, its detection is virtually absent in PCNSL from immunocompetent patients and little else is known about its etiology or risk factors in the population⁷. To address the possibility that common genetic variants influence the risk of developing PCNSL, we have conducted a genome-wide association study (GWAS) on immunocompetent patients. Specifically, we performed a meta-analysis of two new GWAS of PCNSL and identify independent SNPs at 3p22.1 and 6p25.3 associated with risk.

MATERIALS AND METHODS

Subjects and ethics

This study was based on two primary GWAS datasets: (1) GWAS-1 comprised 346 immunocompetent HIV negative patients (184 male; median age 68 years) with PCNSL ascertained through the Service de Neurologie Mazarin, Groupe Hospitalier Pitié-Salpêtrière Paris and the Lymphome oculo-cerebral network (LOC) between 2008-2017, which serves all of France. For controls we made use of Illumina HumanHap 660 data on 788 individuals from the SU.VI.MAX (SUpplementation en VItamines et MinerauxAntioXydants) study healthy subjects (women aged 35–60 years; men aged 45–60 years). (2) GWAS-2 comprised 129 immunocompetent HIV negative patients (76 male; median age 69 years) with primary DLBCL CNS tumors ascertained through the Service de Neurologie Mazarin, Groupe Hospitalier Pitié-Salpêtrière Paris and LOC 2001-2007. For controls, we made use of second series of Illumina HumanHap 660 data generated on 346 individuals from the SU.VI.MAX. Collection of patient samples and associated clinico-pathological information was undertaken with written informed consent and ethical review board approval in accordance with the tenets of the declaration of Helsinki. The diagnosis of PCNSL (ICD-10 C83.3; WHO 9690/3) was established in accordance with WHO guidelines and all patient samples were obtained at first diagnosis.

Genotyping and quality control

Constitutional DNA was extracted from venous blood samples using QIAamp DNA Blood Mini Kit (Qiagen) (OncoNeuroTek, Paris) and quantified using Caliper LabchipGX and Nanodrop. Cases were genotyped using the Infinium OmniExpress-24 v1.2 BeadChip array according to the manufacturer's recommendations (Illumina Inc, San Diego, CA, USA). Standard quality control measures were applied to the GWAS⁸. Specifically, individuals with low call rate (<99%) as well as all individuals with non-European ancestry (using the HapMap version 2 CEU, JPT/CHB and YRI

populations as a reference) were excluded. SNPs with a call rate <90% were excluded as were those with a minor allele frequency (MAF) < 0.01 or displaying significant deviation from Hardy-Weinberg equilibrium (*i.e.* $P < 10^{-6}$). GWAS data were imputed to >10 million SNPs with IMPUTE2 v2.3⁹ software using a merged reference panel consisting of data from 1000 Genomes Project (phase 1 integrated release 3, March 2012)¹⁰ and UK10K¹¹. Genotypes were aligned to the positive strand in both imputation and genotyping. Imputation was conducted separately for each GWAS, and in each, the data were pruned to a common set of SNPs between cases and controls before imputation. Poorly imputed SNPs defined by an information measure <0.80 were excluded. Tests of association between imputed SNPs and *P*-values were calculated using logistic regression under an additive genetic model in SNPTESTv2.5¹². The adequacy of the case-control matching and possibility of differential genotyping of cases and controls were evaluated using Quantile-Quantile (Q-Q) plots of test statistics (**Supplementary Fig. 1**). The fidelity of rs116446171, rs41289586, and rs13254990 and rs10806525 imputation was confirmed by direct genotyping either by sequencing or by KasPar allele-specific PCR (**Supplementary Table 1**).

HLA imputation and analysis

To examine if specific coding variants within HLA genes contributed to association signals, we imputed the classical HLA alleles (A, B, C, DQA1, DQB1, DRB1) and coding variants across the HLA region (chr6:29–34 Mb) using SNP2HLA¹³ - <http://www.broadinstitute.org/mpg/snp2hla/>. Imputation was based on a reference panel from the Type 1 Diabetes Genetics Consortium (T1DGC), which comprises genotype data from 5,225 individuals of European descent typed for HLA-A, B, C, DRB1, DQA1, DQB1, DPB1, DPA1 4-digit alleles. A total of 8,961 classical HLA alleles (two- and four-digit resolution) and 1,873 AA markers including 580 AA positions that were ‘multi-allelic’, were successfully imputed (info score >0.8 for variant). Multi-allelic markers were analyzed as binary markers and a meta-analysis was conducted where we tested SNPs, HLA alleles and AAs across the HLA region for association with PCNSL using SNPTESTv2.5¹².

Meta-analysis

Meta-analyses were performed using the fixed-effects inverse-variance method based on the β estimates and standard errors from each study using META v1.6¹⁴. Cochran's Q -statistic to test for heterogeneity, and the I^2 statistic to quantify the proportion of the total variation due to heterogeneity were calculated¹⁵.

eQTL analysis

To examine the relationship between SNP genotype and gene expression we carried out Summary-data-based Mendelian Randomization (SMR) analysis as per Zhu *et al.*, 2016 (<http://cnsgenomics.com/software/smr/index.html>)¹⁶. We used publicly available lymphoblastoid cell line data from the Genotype-Tissue Expression (GTEx)¹⁷ (<http://www.gtexportal.org>) v6p release and MuTHR¹⁸. Briefly, GWAS summary statistics files were generated from the meta-analysis. Reference files were generated from merging 1000 genomes phase 3 and UK10K (ALSPAC and TwinsUK) vcfs. Results from the SMR test for each of the five risk loci are reported in **Supplementary Data 1**. As previously advocated only probes with at least one expression quantity trait loci (eQTL) P -value of $< 5.0 \times 10^{-8}$ were considered for SMR analysis. We set a threshold for the SMR test of $PSMR < 7.57 \times 10^{-4}$ and $PSMR < 2.5 \times 10^{-3}$ corresponding to a Bonferroni correction for 66 tests (66 probes with a top eQTL $P < 5.0 \times 10^{-8}$ across the 5 loci and two LCL eQTL dataset) and 20 tests (20 probes with a top eQTL $P < 5.0 \times 10^{-8}$ across the 5 loci and Muther eQTL dataset) respectively.

Functional annotation

Novel risk SNPs and their proxies (*i.e.* $r^2 > 0.2$ in the 1000 Genomes EUR reference panel) were annotated for putative functional effect based upon histone mark ChIP-Seq data for H3K27ac, H3K4Me1 and H3K27Me3 from GM12878 (LCL)¹⁹ and primary B-cells²⁰. We searched for

overlap with “super-enhancer” regions as defined by Hnisz *et al*²¹, restricting the analysis to the GM12878 cell line and CD19⁺ B-cells. The novel risk SNPs and their proxies ($r^2 > 0.2$ as above) were intersected with regions of accessible chromatin in CLL cells, as defined by Rendeiro *et al*²⁰, which were used as a surrogate for likely sites of transcription factor (TF) binding. SNPs falling within accessible sites (n=47) were taken forward to TF binding motif analysis and were also annotated for genomic evolutionary rate profiling (GERP) score²² as well as bound TFs based on ENCODE project¹⁹ ChIP-Seq data.

Transcription factor binding disruption analysis

To examine enrichment in specific TF binding across risk loci, we adapted the variant set enrichment method of Cowper-Salari *et al*²³. Briefly, for each risk locus, a region of strong LD (defined as $r^2 > 0.8$ and $D' > 0.8$) was determined, and SNPs within were termed the associated variant set (AVS). TF ChIP-Seq uniform peak data were obtained from ENCODE for the GM12878 cell line, which included data for 82 TFs. For each of these marks, the overlap of the SNPs in the AVS and the binding sites was determined to produce a mapping tally. A null distribution was produced by randomly selecting SNPs with the same characteristics as the risk-associated SNPs, and the null mapping tally calculated. This process was repeated 10,000 times, and approximate *P*-values were calculated as the proportion of permutations where the null mapping tally was greater or equal to the AVS mapping tally. An enrichment score was calculated by normalizing the tallies to the median of the null distribution. Thus, the enrichment score is the number of s.d.'s of the AVS mapping tally from the mean of the null distribution tallies.

RESULTS

Association analysis

After quality control, the two GWAS provided SNP genotypes on a total of 475 cases and 1,134 controls (**Supplementary Fig. 1 and 2 - Supplementary Tables 2 and 3**). To increase genomic resolution, we imputed >10 million SNPs using the 1000 Genomes Project¹⁰ combined with UK10K¹¹ as reference. Q-Q plots for SNPs with MAF >0.5% post imputation showed only minimal evidence of over-dispersion (λ values for both GWAS = 1.00; **Supplementary Fig. 3**). Meta-analyzing test results from the two GWAS, we derived joint odds ratios (OR) per-allele and 95% confidence intervals (CI) under a fixed-effects model for each SNP and associated *P*-values.

Genome-wide significant associations ($P < 5 \times 10^{-8}$) were shown for loci at 3p22.1 (rs41289586, $P = 2.17 \times 10^{-8}$) and 6p25.3 (rs116446171, $P = 1.95 \times 10^{-13}$) (**Fig. 1, Table 1**). Conditional analysis of GWAS data showed no evidence for additional independent signals at either of the two risk loci.

Following on from this we examined whether other reported risk loci for DLBCL influenced PCNSL risk. Respective association *P*-values for the 6p21.22-HLA (human leukocyte antigen) (rs2523607) and 2p23.3 (rs79480871) risk SNPs were 0.023 and 0.14 (**Supplementary Table 4**).

HLA alleles

Variation at HLA has been linked to risk of DLBCL and a number of other B-cell tumors²⁴⁻²⁸. The strongest SNP association at 6p21 (HLA) for PCNSL was provided by rs2395192 ($P = 1.81 \times 10^{-7}$), which maps between *HLA-DRA* and *HLA-DRB5* (**Supplementary Fig.5, Table 1**). To obtain additional insight into plausible functional variants within the HLA region, we imputed the classical HLA alleles and amino acid residues using SNP2HLA¹³. No imputed HLA alleles or amino acid positions reached genome-wide significance (**Supplementary Fig. 5**). The strongest coding

changes within the HLA region were observed for the HLA class II alleles DRB1 Ser11Pro (AA_DRB1_11_32660115_SP, $P = 3.35 \times 10^{-6}$) and presence of the haplotype SRG (DRB1_13_32660109_SRG, $P = 3.35 \times 10^{-6}$) (**Supplementary Table 5**).

Functional annotation of risk loci

To gain insight into the biological basis underlying associations at 6p25.3 and other promising risk loci, we first evaluated each of the risk SNPs as well as the correlated variants using the online resources HaploRegv4²⁹, RegulomeDB³⁰ and Fantom5³¹ for evidence of functional effects (**Supplementary Data 2**). These data revealed regions of active chromatin at 6p25.3, 6q15 and 8q24 risk loci in B-cells. To explore whether there was an association between SNP genotype and transcript levels we performed an eQTL analysis using the GTEx project¹⁷, MuTHR¹⁸ and blood eQTL data from Westra *et al*³². We used SMR¹⁶ analysis to test for a concordance between signals from GWAS and *cis* eQTL for genes within 1 Mb of the sentinel and correlated SNPs ($r^2 > 0.8$) at each locus (**Supplementary Data 1**) and derived b_{XY} statistics, which estimate the effect of gene expression on PCNSL risk. After accounting for multiple testing we were unable to demonstrate any consistently significant eQTL for any of the risk loci examined. Chromatin looping interactions formed between enhancer elements and the promoters of genes they regulate map within distinct chromosomal topological associating domains. To identify patterns of local chromatin patterns, we analyzed promoter capture Hi-C data on the LCL cell line GM12878 as a source of B-cell information³³. Looping chromatin interactions were shown between non-coding regions at 6p25.3 (rs11646171) with the *IRF4* promoter (**Fig. 2**) and at 8q24.21 (rs13254990) with the *MYC* promoter; both genes with strong relevance to B-cell tumorigenesis.

Using ChIP-seq data on 82 TFs in GM12878 we examined for an over-representation of the binding of TFs at risk loci. Although not statistically significant the strongest TF bindings were shown for *TBLIXR1* that is mutated in 20% of PCNSL³⁴ (**Supplementary Fig. 6**).

DISCUSSION

To our knowledge this is the first study providing insight into the genetic predisposition to PCNSL. While PCNSL is a specific entity it corresponds pathologically to diffuse large B-cell lymphoma. Hence, it is therefore perhaps not surprising that we identified associations, in common with DLBCL²⁴ at 6q25.3 and 8q24.21 for PCNSL. However, the absence of associations at the 8q24.21 (rs4733601) and 2p23.3 (rs79480871) risk loci suggests the existence of a distinct developmental pathway for PCNSL, possibly reflective of its etiology. Moreover, analysis of publicly accessible GWAS data on DLBCL (NCI GWAS Stage 1) provides no support for an association between 3p22.1 (rs41289586) with DLBCL²⁴ (**Supplementary Table 6**).

The 6p25.3 risk SNP rs116446171 (**Fig. 2**), which maps intergenic to *EXOC2* (exocyst complex component 2) and *IRF4* (interferon regulatory factor 4), has been previously shown to influence the risk of DLBCL²⁴. *EXOC2* is part of the multi-protein exocyst complex essential for polarized vesicle trafficking and the maintenance and intercellular transfer of viral proteins and virions³⁵. Furthermore, the RalB/*EXOC2* effector complex is a component of the TBK-1-dependent innate signaling pathway³⁶. While thus far there is no evidence to implicate *EXOC2* in lymphoma, the RalB/*EXOC2* complex may contribute to tumor cell survival³⁶. In contrast, *IRF4* has a well-established role in the development of many B-cell malignancies³⁷⁻³⁹.

The 3p22.1 risk SNP rs41289586 (**Fig. 2**) localizes to exon 6 of the anoctamin 10 gene (*ANO10*) and is responsible for the rare missense change (*ANO10*:c.788G>A, p.Arg263His). Although rs1052501 at 3p22.1 has been reported to be associated with risk of multiple myeloma⁴⁰, another B-cell malignancy, this SNP maps >1Mb away from rs41289586 (pairwise $r^2 = 0.0002$). Inherited defects in *ANO10*, which encodes a calcium-activated chloride channel transmembrane protein are a cause autosomal recessive spinocerebellar ataxia⁴¹. While other anoctamins have been implicated in

cancer⁴², to date there is no evidence for the role of *ANO10* in a B-cell malignancy. However, intriguingly, rs41289586 has been associated with regulation of macrophage response and associated with *Borrelia* seropositivity⁴³, implicating *ANO10* in the innate immune defense.

In addition to the 6p25.3 and 3p22.1 risk loci we identified promising associations ($P < 2 \times 10^{-7}$), at 6q15 (rs10806425, $P = 1.36 \times 10^{-7}$) and 8q24.21 (rs13254990; $P = 1.33 \times 10^{-7}$) annotating genes with strong relevance to B-cell tumorigenesis (**Table 1, Supplementary Fig. 4**). rs10806425 localizes to intron 1 of the gene encoding *BACH2* (basic leucine zipper transcription factor 2). Loss of heterozygosity of *BACH2* has been reported at a frequency of 20% in B-cell lymphoma⁴⁴. In DLBCL patients with higher *BACH2* expression tend to have a better prognosis⁴⁵. *BACH2* is a key regulator of the pre-BCR checkpoint as well as a tumor suppressor in pre-B acute lymphoblastic leukemia⁴⁶. One mechanism of *BACH2* downregulation in leukemia is the loss of the transcription factor *PAX5*, which is intriguingly, commonly mutated in both PCNSL⁴⁷ and B-cell ALL⁴⁶.

The 8q24 SNP rs13254990 localizes to intron 4 of *PVT1*, a non-coding RNA affecting the activation of MYC. Two independent risk loci at 8q24 defined by SNPs rs13255592 and rs4733601 have previously been shown to influence DLBCL²⁴. rs13255592 also localizes within intron 4 of *PVT1* and is highly correlated with rs13254990 ($r^2 = 0.98$, $P = 3.81 \times 10^{-7}$). No association between rs4733601, which maps approximately 1.9Mb telomeric to *PVT1*, and PCNSL risk was however shown ($r^2 = 4.21 \times 10^{-5}$, $P = 0.99$; **Supplementary Table 4**). The 8q24.21 128-130Mb genomic interval harbors multiple independent risk loci with different tumor specificities (**Supplementary Table 7**). The strongest additional association for PCNSL being shown by the Hodgkin lymphoma risk SNP rs2019960 ($P = 4.1 \times 10^{-5}$) raising the possibility of an additional risk locus for the disease at 8q24.21²⁸.

Although in part speculative, the 6q25.3 association implicates *IRF4* in the development of PCNSL. Through interaction with transcription factors including PU.1, IRF4 controls the termination of pre-B-cell receptor signaling and promotes the differentiation of pro-B cells to small B cells⁴⁸. Furthermore, via BLIMP1 and BCL6, IRF4 controls the transition of memory B cells⁴⁹. The observation that *PVT1* rearrangement occurs frequently in highly aggressive B-cell lymphomas harboring an 8q24 abnormality suggest that germline variation in this region may influence PCNSL risk⁵⁰⁻⁵². The 6q15 association implicates *BACH2* in the development of PCNSL. *BACH2* is an attractive candidate *a priori* for having a role in PCNSL development being a regulator of the antibody response mediating effects through *BLIMP1*, *XBPI*, *LRF4*, and *PAX5*⁵³. Moreover, *BACH2* regulates the activity of the tumor suppressor c-Rel in lymphoma development⁵⁴. Collectively these data are consistent with aberrant B-cell developmental pathways being central for predisposition to PCNSL. The finding of a relationship between ANO10:c.788G>A, p.Arg263His with PCNSL, but not classical DLBCL highlights differences in biological etiology with this lymphoma. Intriguingly, this variant has previously been shown to influence macrophage response⁴³ thereby implicating *ANO10* in innate immune defense with the development of PCNSL. While not statistically significant, the *HLA-DRA* and *HLA-DRB1* associations are also of relevance as these alleles have previously been shown to influence the human reaction to viral load and EBV infection, respectively⁵⁵. The linkage of these genes to the development of PCNSL is therefore entirely consistent with an infective basis to this B-cell malignancy even though none of the patients we studied were immunocompromised.

Inevitably constrained by the sheer rarity of PCNSL we acknowledge that a limitation of our study has been an inability to replicate study findings in additional series. Another limitation is the absence of a case-case study with other DLBCL. This could be performed in future analyses. Our findings are however based on a meta-analysis of two series cohorts of PCNSL. Moreover, despite

the rarity of PCBSL by ascertaining patients through the LOC Network has provided us with greater power to detect associations than the smaller studies of other rare lymphomas⁵⁶.

In summary, our findings represent an important step in defining the contribution of common genetic variation to the risk of developing PCNSL. Our observations are notable since the associations highlighted define regions of the genome harboring plausible candidate genes for further investigation. Given the relatively modest size of our analysis, it is highly probable that further studies will discover additional common susceptibility loci. These coupled with functional analyses should provide for an explanation of the biological underpinnings of PCNSL.

DATA AVAILABILITY

Genotype data that support the findings of this study have been deposited in the database of the European Genome-phenome Archive (EGA) with accessions code PRJEB21814. NCI Non-Hodgkin Lymphoma GWAS data was obtained through dbGAP (phs000801.v2.p1). The remaining data are contained within the paper and Supplementary files are available from the authors upon request.

FUNDINGS

The primary source of funding was provided by la Ligue Nationale Contre le Cancer-RE 2015 (K.H.X), French National Institute of Cancer (InCa) LOC Network (K.H.X and C.S). A.A has also been granted with a “poste d’accueil AP-HP-CEA”. K.L is supported by l’Association pour la Recherche sur les Tumeurs Cérébrales (ARTC) and Institute CARNOT – Institut du Cerveau et de la Moelle Epinière (ICM). Finally, also acknowledge support from Cancer Research UK and Bloodwise. A.S. is supported by a clinical fellowship from Cancer Research UK. We are grateful to all investigators and all the patients and individuals for their participation. Samples from AP-HM were retrieved from AP-HM tumor bank, authorization number 2013-1786. We also thank the clinicians, other hospital staff and study staff that contributed to the blood sample and data collection for this study and OncoNeuroTek that provided and prepared DNA samples. Genotypes from NCI Non-Hodgkin Lymphoma GWAS were accessed through dbGaP accession phs000801.v2.p1.

AUTHORSHIP

A.A. and K.H.X., developed the project and provided overall project management; K.L., M.D., R.S.H., K.H.X. and A.A. drafted the manuscript. K.L., A.S., P.J.L. and M.W. performed bioinformatic and statistical analyses; Patient samples and phenotype data were provided by C.D., D.G., K.H.X., C.S. and other members of the LOC Network. M.D., I.D., L.R.P., A.R., D.G.

performed project management and supervised genotyping; M.D., I.D., A.R., M.B. and A.H. performed sequencing and genotyping. A.A., K.H.X., M.D., A.R., L.R.P. and P.B. supervised laboratory management and oversaw genotyping of cases; D.G., M.D., I.D., L.R.P. performed sample management of cases. All authors reviewed and approved the manuscript prior to submission.

CONFLICT OF INTEREST

The remaining authors declare no competing financial interests.

REFERENCES

1. PM Kluin MD, JA Ferry. Primary diffuse large B-cell lymphoma of the CNS. *IARC Press, Lyon*. 2008:240-241.
2. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016; 127(20):2375-2390.
3. Hoang-Xuan K, Bessell E, Bromberg J, et al. Diagnosis and treatment of primary CNS lymphoma in immunocompetent patients: guidelines from the European Association for Neuro-Oncology. *Lancet Oncol*. 2015; 16(7):e322-332.
4. Bessell EM, Dickinson P, Dickinson S, Salmon J. Increasing age at diagnosis and worsening renal function in patients with primary central nervous system lymphoma. *J Neurooncol*. 2011; 104(1):191-193.
5. Villano JL, Koshy M, Shaikh H, Dolecek TA, McCarthy BJ. Age, gender, and racial differences in incidence and survival in primary CNS lymphoma. *Br J Cancer*. 2011; 105(9):1414-1418.
6. O'Neill BP, Decker PA, Tieu C, Cerhan JR. The changing incidence of primary central nervous system lymphoma is driven primarily by the changing incidence in young and middle-aged men and differs from time trends in systemic diffuse large B-cell non-Hodgkin's lymphoma. *Am J Hematol*. 2013; 88(12):997-1000.
7. Bashir R, Luka J, Cheloha K, Chamberlain M, Hochberg F. Expression of Epstein-Barr virus proteins in primary CNS lymphoma in AIDS patients. *Neurology*. 1993; 43(11):2358-2362.
8. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc*. 2010; 5(9):1564-1573.

9. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009; 5(6):e1000529.
10. Genomes Project C, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010; 467(7319):1061-1073.
11. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun.* 2015; 6:8111.
12. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007; 39(7):906-913.
13. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One.* 2013; 8(6):e64683.
14. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet.* 2010; 42(5):436-440.
15. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002; 21(11):1539-1558.
16. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016; 48(5):481-487.
17. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013; 45(6):580-585.
18. Grundberg E, Small KS, Hedman AK, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet.* 2012; 44(10):1084-1089.
19. de Souza N. The ENCODE project. *Nature methods.* 2012; 9(11):1046.
20. Rendeiro AF, Schmidl C, Strefford JC, et al. Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. *Nat Commun.* 2016; 7:11938.

21. Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013; 155(4):934-947.
22. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol*. 2010; 6(12):e1001025.
23. Cowper-Salari R, Zhang X, Wright JB, et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet*. 2012; 44(11):1191-1198.
24. Cerhan JR, Berndt SI, Vijai J, et al. Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. *Nat Genet*. 2014; 46(11):1233-1238.
25. Conde L, Halperin E, Akers NK, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nat Genet*. 2010; 42(8):661-664.
26. Wang SS, Abdou AM, Morton LM, et al. Human leukocyte antigen class I and II alleles in non-Hodgkin lymphoma etiology. *Blood*. 2010; 115(23):4820-4823.
27. Law PJ, Berndt SI, Speedy HE, et al. Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. *Nat Commun*. 2017; 8:14175.
28. Sud A, Thomsen H, Law PJ, et al. Genome-wide association study of classical Hodgkin lymphoma identifies key regulators of disease susceptibility. *Nat Commun*. 2017; In press.
29. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res*. 2016; 44(D1):D877-881.
30. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012; 22(9):1790-1797.
31. Kawaji H, Severin J, Lizio M, et al. The FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation. *Genome Biol*. 2009; 10(4):R40.

32. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013; 45(10):1238-1243.
33. Mifsud B, Tavares-Cadete F, Young AN, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet.* 2015; 47(6):598-606.
34. Bruno A, Boisselier B, Labreche K, et al. Mutational analysis of primary central nervous system lymphoma. *Oncotarget.* 2014; 5(13):5065-5075.
35. Mukerji J, Olivieri KC, Misra V, Agopian KA, Gabuzda D. Proteomic analysis of HIV-1 Nef cellular binding partners reveals a role for exocyst complex proteins in mediating enhancement of intercellular nanotube formation. *Retrovirology.* 2012; 9:33.
36. Chien Y, Kim S, Bumeister R, et al. RalB GTPase-mediated activation of the IkappaB family kinase TBK1 couples innate immune signaling to tumor cell survival. *Cell.* 2006; 127(1):157-170.
37. Acquaviva J, Chen X, Ren R. IRF-4 functions as a tumor suppressor in early B-cell development. *Blood.* 2008; 112(9):3798-3806.
38. Pathak S, Ma S, Trinh L, et al. IRF4 is a suppressor of c-Myc induced B cell leukemia. *PLoS One.* 2011; 6(7):e22628.
39. Kreher S, Johrens K, Strehlow F, et al. Prognostic impact of B-cell lymphoma 6 in primary CNS lymphoma. *Neuro Oncol.* 2015; 17(7):1016-1021.
40. Broderick P, Chubb D, Johnson DC, et al. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat Genet.* 2011; 44(1):58-61.
41. Renaud M, Anheim M, Kamsteeg EJ, et al. Autosomal recessive cerebellar ataxia type 3 due to ANO10 mutations: delineation and genotype-phenotype correlation study. *JAMA Neurol.* 2014; 71(10):1305-1310.
42. Wanitchakool P, Wolf L, Koehl GE, et al. Role of anoctamins in cancer and apoptosis. *Philos Trans R Soc Lond B Biol Sci.* 2014; 369(1638):20130096.

43. Hammer C, Wanitchakool P, Sirianant L, et al. A Coding Variant of ANO10, Affecting Volume Regulation of Macrophages, Is Associated with Borrelia Seropositivity. *Mol Med.* 2015; 21:26-37.
44. Sasaki S, Ito E, Toki T, et al. Cloning and expression of human B cell-specific transcription factor BACH2 mapped to chromosome 6q15. *Oncogene.* 2000; 19(33):3739-3749.
45. Sakane-Ishikawa E, Nakatsuka S, Tomita Y, et al. Prognostic significance of BACH2 expression in diffuse large B-cell lymphoma: a study of the Osaka Lymphoma Study Group. *J Clin Oncol.* 2005; 23(31):8012-8017.
46. Swaminathan S, Huang C, Geng H, et al. BACH2 mediates negative selection and p53-dependent tumor suppression at the pre-B cell receptor checkpoint. *Nat Med.* 2013; 19(8):1014-1022.
47. 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(5):1869-1875.
48. Rickert RC. New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. *Nat Rev Immunol.* 2013; 13(8):578-591.
49. Schmidlin H, Diehl SA, Blom B. New insights into the regulation of human B-cell differentiation. *Trends Immunol.* 2009; 30(6):277-285.
50. McDonnell TJ, Korsmeyer SJ. Progression from lymphoid hyperplasia to high-grade malignant lymphoma in mice transgenic for the t(14; 18). *Nature.* 1991; 349(6306):254-256.
51. Taub R, Kirsch I, Morton C, et al. Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proc Natl Acad Sci U S A.* 1982; 79(24):7837-7841.
52. Ladanyi M, Offit K, Jhanwar SC, Filippa DA, Chaganti RS. MYC rearrangement and translocations involving band 8q24 in diffuse large cell lymphomas. *Blood.* 1991; 77(5):1057-1063.

53. Muto A, Tashiro S, Nakajima O, et al. The transcriptional programme of antibody class switching involves the repressor Bach2. *Nature*. 2004; 429(6991):566-571.
54. Hunter JE, Butterworth JA, Zhao B, et al. The NF-kappaB subunit c-Rel regulates Bach2 tumour suppressor expression in B-cell lymphoma. *Oncogene*. 2016; 35(26):3476-3484.
55. Hammer C, Begemann M, McLaren PJ, et al. Amino Acid Variation in HLA Class II Proteins Is a Major Determinant of Humoral Response to Common Viruses. *Am J Hum Genet*. 2015; 97(5):738-743.
56. Li Z, Xia Y, Feng LN, et al. Genetic risk of extranodal natural killer T-cell lymphoma: a genome-wide association study. *Lancet Oncol*. 2016; 17(9):1240-1247.
57. Scales M, Jager R, Migliorini G, Houlston RS, Henrion MY. visPIG--a web tool for producing multi-region, multi-track, multi-scale plots of genetic data. *PLoS One*. 2014; 9(9):e107497.

FIGURE AND TABLE LEGENDS

Figure 1: Manhattan plot of association P -values. Shown are the genome-wide $-\log_{10}P$ -values (two-sided) of >10 million successfully imputed autosomal SNPs in 475 cases and 1,134 controls. The red horizontal line represents the genome-wide significance threshold of $P = 5.0 \times 10^{-8}$.

Figure 2: Regional plots of association results and recombination rates for new risk loci for primary cerebral nervous system lymphoma. Results shown for (a) 6p25 and (b) 3q21. Plots (drawn using visPig⁵⁷) show association results of both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates. $-\log_{10}P$ values (y -axis) of the SNPs are shown according to their chromosomal positions (x -axis). The sentinel SNP in each combined analysis is shown as a large circle or triangle and is labelled by its rsID. The color intensity of each symbol reflects the extent of LD with the top genotyped SNP, white ($r^2 = 0$) through to dark red ($r^2 = 1.0$). Genetic recombination rates, estimated using 1000 Genomes Project samples, are shown with a light blue line. Physical positions are based on NCBI build 37 of the human genome. Also shown are the chromatin-state segmentation track (ChromHMM) for lymphoblastoid cells using data from the HapMap ENCODE Project, and the positions of genes and transcripts mapping to the region of association. The top track represents capture Hi-C promoter contacts in GM12878 cells. The colour intensity of each contact reflects the interaction score.

Table 1: Summary results for SNPs associated with primary central nervous system lymphoma risk

† **LOC Network:** Marie-Pierre Moles-Moreau¹, Rémy Gressin², Vincent Delwail³, Franck Morschhauser⁴, Philippe Agapé⁵, Arnaud Jaccard⁶, Hervé Ghesquieres⁷, Adrian Tempescul⁸, Emmanuel Gyan⁹, Jean-Pierre Marolleau¹⁰, Roch Houot¹¹, Luc Fornecker¹², Anna-Luisa Di Stefano¹³, Inès Detrait¹⁴, Amithys Rahimian¹⁵, Mark Lathrop¹⁶, Diane Genet¹⁷, Frédéric Davi¹⁸, Nathalie Cassoux¹⁹, Valérie Touitou²⁰, Sylvain Choquet²¹, Anne Vital²², Marc Polivka²³, Dominique Figarella-Branger^{24,25}, Alexandra Benouaich-Amiel²⁶, Chantal Campello²⁷, Frédéric Charlotte²⁸, Nadine Martin-Duverneuil²⁹, Loïc Feuvret³⁰, Aurélie Kas³¹, Soledad Navarro³², Chiara Villa³³, Franck Bielle³⁴, Fabrice Chretien³⁵, Marie Christine Tortel³⁶, Guillaume Gauchotte³⁷, Emmanuelle Uro-Coste³⁸, Catherine Godfrain³⁹, Valérie Rigau⁴⁰, Myrto Costopoulos¹⁸, Magalie Le Garff-Tavernier¹⁸, David Meyronnet⁴¹, Audrey Rousseau⁴², Clovis Adam⁴³, Thierry Lamy⁴⁴, Cécile Chabrot⁴⁵, Eileen M. Boyle⁴⁶, Marie Blonski⁴⁷, Anna Schmitt⁴⁸.

1. Department of Hematology, Angers University Hospital, Angers, 49033, France.
2. Department of Hematology CHU Grenoble Michallon 38043 Grenoble Cedex 02 – France
3. Service d'Oncologie Hématologique et de Thérapie Cellulaire, CHU de Poitiers, INSERM, CIC 1402, Poitiers, Centre d'Investigation Clinique, Université de Poitiers, Poitiers, France
4. Department of Hematology, CHRU Lille, Lille, 59037, France
5. Institut de Cancérologie – 44800 Saint Herblain – France
6. Department of Hematology CHU Dupuytren 87042 Limoges- France
7. Department of Hematology, University Hospital of Lyon, 69002, Lyon, France
8. Department of Hematology CHU Morvan 29609 Brest Cedex – France
9. Department of Hematology CHU Bretonneau 34044 Tours, France
10. Department of Hematology, University Hospital of Amiens, 80054, Amiens, France.
11. CHU Rennes, Service Hématologie Clinique, F-35033 Rennes, France

12. Department of Hematology CHU Strasbourg 67000 Strasbourg – France
13. Department of Neurology Hôpital Foch 92151 Suresnes – France
14. Inserm, U 1127, ICM, F-75013 Paris, France; CNRS, UMR 7225, ICM, F-75013 Paris, France; Institut du Cerveau et de la Moelle épinière ICM, Paris 75013, France; Sorbonne Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France; AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Service de neurologie 2-Mazarin, 75013 Paris, France.
15. OncoNeuroTek, Institut du Cerveau et de la Moelle épinière, ICM, Paris, F-75013, France.
16. McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada, H3A 0G1.
17. AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Service de neurologie 2-Mazarin, 75013 Paris, France. Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France
18. Department of Biological Hematology, AP-HP, Groupe Hospitalier Pitié-Salpêtrière, 75013, Paris, France.
19. Department of Oncological Ophthalmology, Institut Curie, Paris, France.
20. AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Department of Ophthalmology, Paris, 75013, France.
21. AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Department of Hematology, Paris, 75013, France
22. CNRS, Institut des Maladies Neurodégénératives, UMR 5293, F-33000 Bordeaux, France; Department of Pathology, Bordeaux University Hospital, Bordeaux, France
23. Department of Pathology, CHU Paris-GH St-Louis Lariboisière F.Widal - Hôpital Lariboisière, 75010, Paris, France.

24. Department of pathology and Neuropathology, Hôpital de la Timone, Aix-Marseille Univ, AP-HM, Marseille, 13005, France.
25. AMU, CRO2, 13005, Marseille, 13005, France.
26. Department of Neurology, CHU Toulouse, 31059 Toulouse, France.
27. Service de Neuro-Oncologie, CHU Timone, 13005 Marseille – France.
28. Department of Pathology, CHU Pitié-Salpêtrière, 75013 Paris, France.
29. Department of Neuroradiology, CHU Pitié-Salpêtrière, 75013 Paris, France.
30. Department of Radiotherapy, CHU Pitié-Salpêtrière, 75013 Paris, France.
31. Department of Nuclear Medicine, CHU Pitié-Salpêtrière, 75013 Paris, France.
32. Department of neurosurgery, CHU Pitié-Salpêtrière, 75013 Paris, France.
33. Department of Pathology, Hôpital Foch, 92151 Suresnes – France.
34. Inserm, U 1127, ICM, F-75013 Paris, France ; CNRS, UMR 7225, ICM, F-75013 Paris, France ; Institut du Cerveau et de la Moelle épinière ICM, Paris 75013, France ; Sorbonne Universités, UPMC Université Paris 06, UMR S 1127, F-75013 Paris, France ; AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Service de Neuropathologie, 75013 Paris, France.
35. Department of Neuropathology, Centre Hospitalier Sainte Anne, Paris, France.
36. Department of Pathology, Hôpitaux Civils de Colmar, 68024, Colmar Cedex, France.
37. Department of Pathology, Nancy University Hospital, Vandoeuvre-lès-Nancy, France
38. Department of Pathology, CLCC Institut Claudius Regaud, 31059 Toulouse, France.
39. Department of Pathology, CHU Clermont-Ferrand, 63000 Clermont-Ferrand, France.
40. Department of Pathology, CHU Montpellier, 34000 Montpellier, France.
41. Department of Pathology, University Hospital of Lyon, 69002, Lyon, France.
42. Department of Pathology, CHU Angers, 49933 Angers – France.
43. Department of Pathology, CHU Kremlin Bicêtre, 94270 Le Kremlin Bicêtre.
44. CHU Rennes, Rennes, 35033, France.

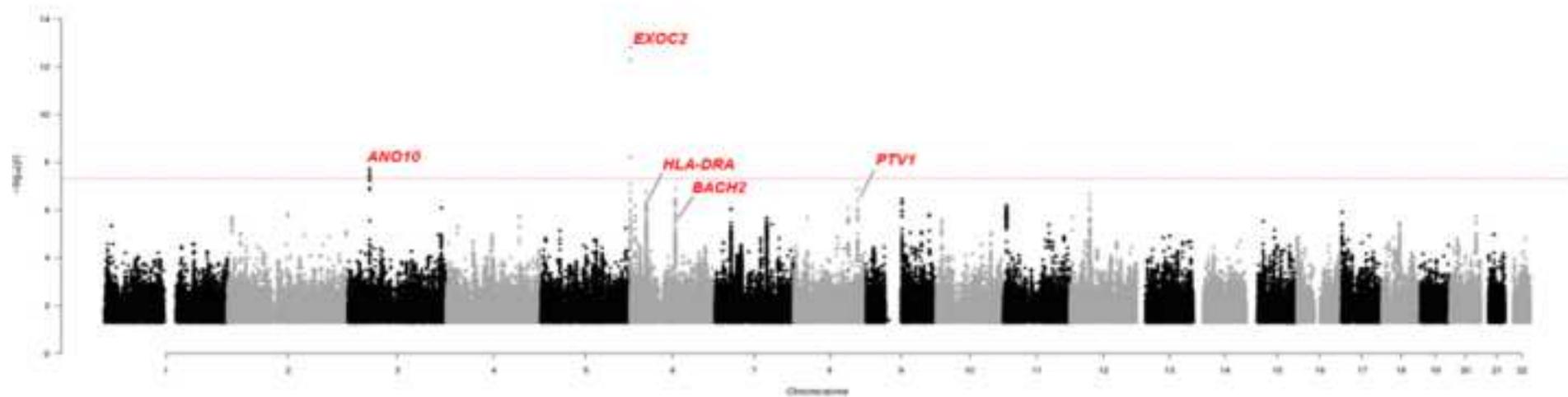
45. CHU Clermont Ferrand, Clermont Ferrand, 63000, France.
46. Department of Haematology, Lille University Hospital, Lille, 59037, France.
47. CHU Nancy. Nancy, 54500, France.
48. CHU Bordeaux. Bordeaux ,33000, France.

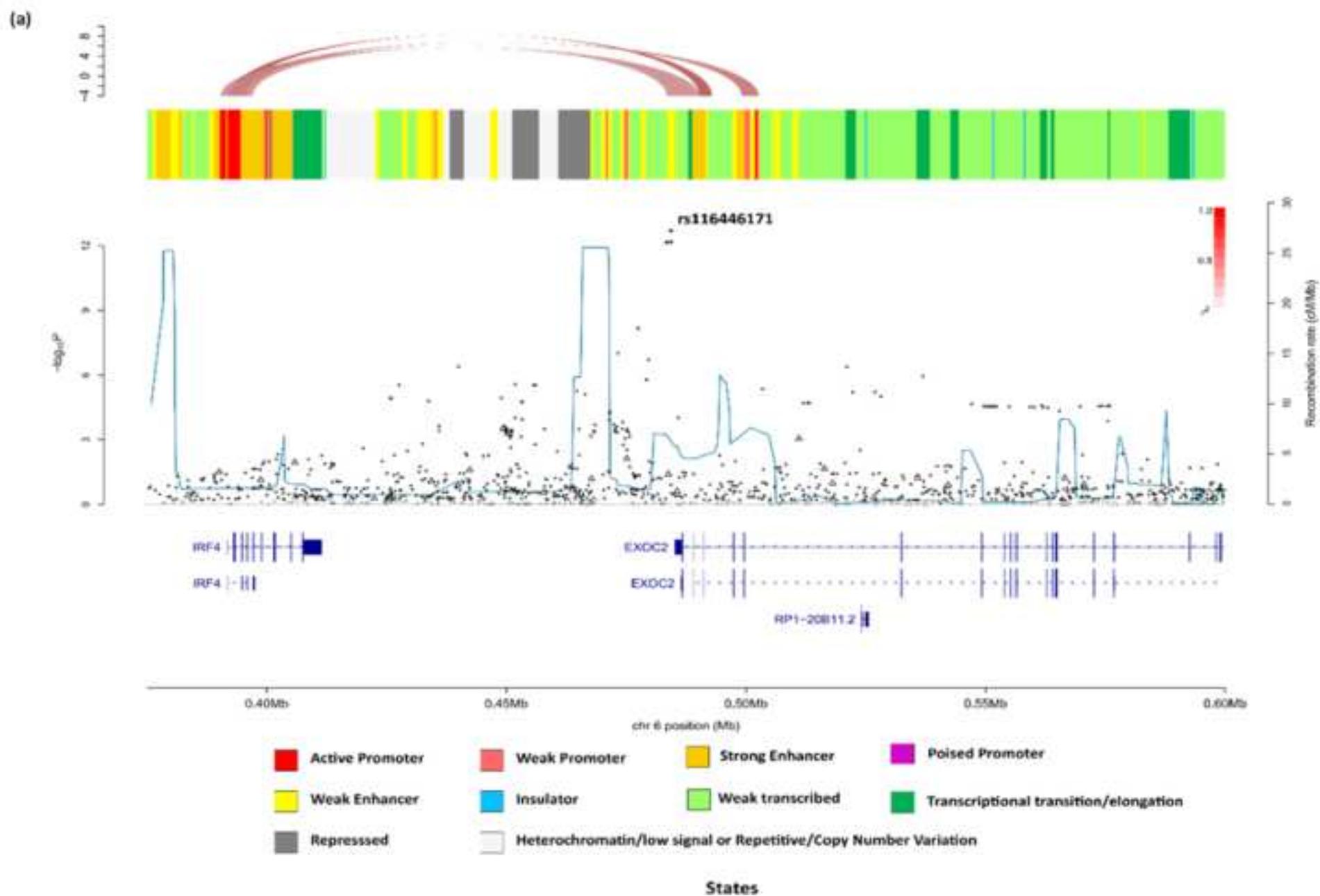
Table 1: Summary results for SNPs associated with primary central nervous system lymphoma risk

Locus	Nearest gene(s)	SNP	Position (bp, hg19)	Risk allele	Dataset	RAF (case;control)	Imputation Info Score	OR	95% CI	P-value
6p25.3	<i>EXOC2</i>	rs116446171	484,453	G	GWAS-1	(0.066; 0.022)	0.85	4.11	(2.47 - 6.85)	5.13x10 ⁻⁸
					GWAS-2	(0.088;0.019)	0.84	7.87	(3.59 - 17.21)	2.36x10 ⁻⁷
					Combined			4.99	(3.26 - 7.65)	1.53x10 ⁻¹³
$I^2=46%$ $P_{het}=0.17$										
3p22.1	<i>ANO10</i>	rs41289586	43,618,558	T	GWAS-1	(0.048;0.017)	0.96	3.42	(1.94 - 6.02)	1.90x10 ⁻⁵
					GWAS-2	(0.065;0.019)	0.98	4.84	(2.10 - 11.13)	2.05x10 ⁻⁴
					Combined			3.82	(2.39 - 6.09)	1.87x10 ⁻⁸
$I^2=0%$ $P_{het}=0.50$										
8q24.21	<i>PTVI</i>	rs13254990	129,076,451	T	GWAS-1	(0.43;0.33)	0.98	1.58	(1.31 - 1.91)	2.21x10 ⁻⁶
					GWAS-2	(0.40;0.32)	0.98	1.44	(1.05 - 1.96)	0.021
					Combined			1.54	(1.31 - 1.81)	1.33x10 ⁻⁷
$I^2=0%$ $P_{het}=0.60$										
6q15	<i>BACH2</i>	rs10806425	90,926,612	C	GWAS-1	(0.68;0.58)	1	1.50	(1.25 - 1.80)	8.93x10 ⁻⁶
					GWAS-2	(0.69;0.59)	1	1.53	(1.14 - 2.05)	0.0045
					Combined			1.51	(1.30 - 1.77)	1.36x10 ⁻⁷
$I^2=0%$ $P_{het}=0.93$										
6p21.32	<i>HLA-DRA</i>	rs2395192	32,447,644	C	GWAS-1	(0.48;0.59)	0.96	1.56	(1.30 - 1.88)	1.65x10 ⁻⁶
					GWAS-2	(0.52;0.60)	0.97	1.38	(1.03 - 1.84)	0.029
					Combined			1.51	(1.29 - 1.76)	1.81x10 ⁻⁷
$I^2=0%$ $P_{het}=0.47$										

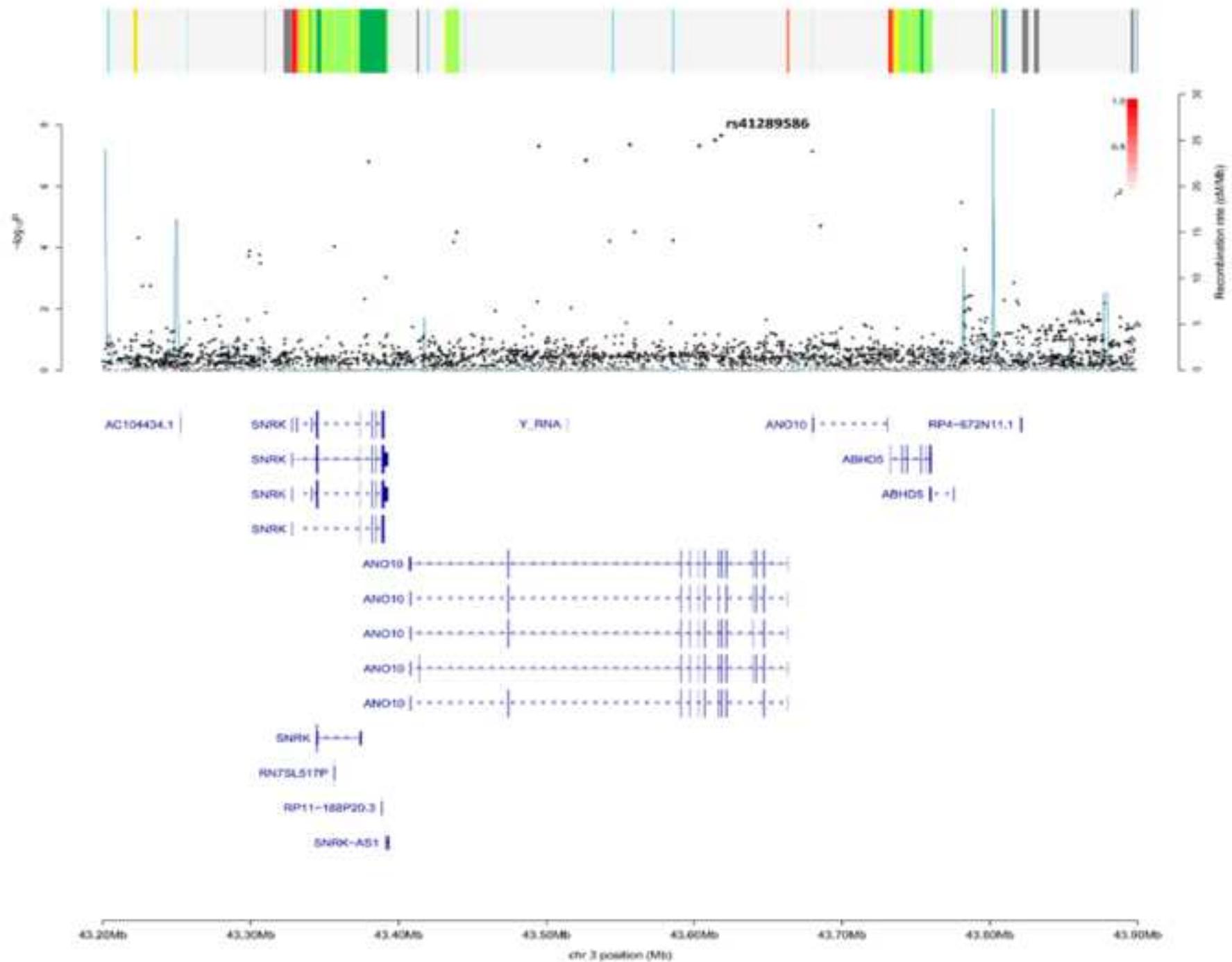
bp, base pair; OR, odds ratio; 95% CI, 95% confidence interval; P_{het} , P -value for heterogeneity; I^2 , proportion of the total variation due to heterogeneity.

RAF is risk allele frequency across all of the GWAS-1 and GWAS-2 datasets, respectively.





(b)

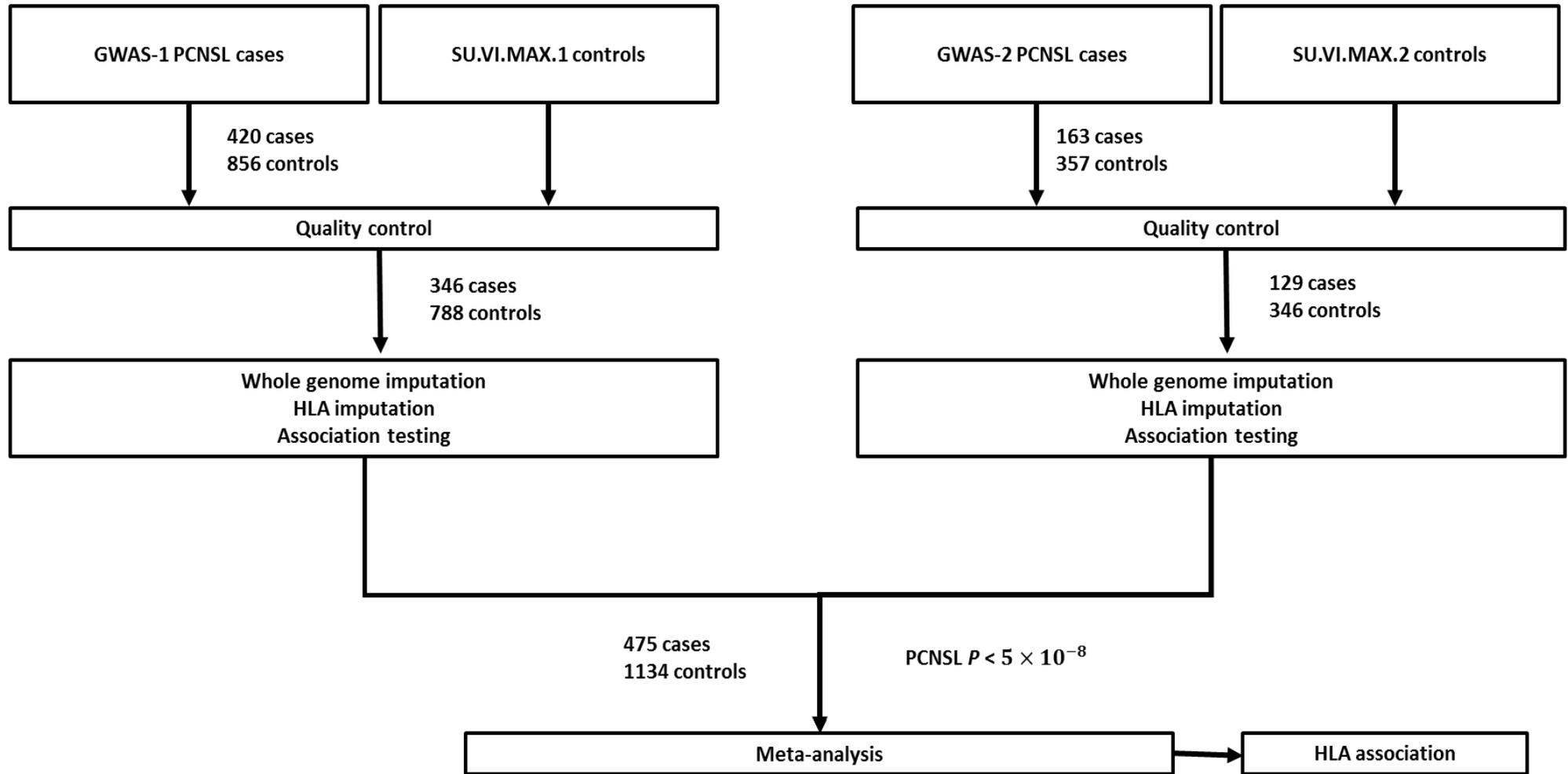


A genome-wide association study identifies susceptibility loci for primary central nervous system lymphoma at 6p25.3 and 3p22.1: a LOC network study group

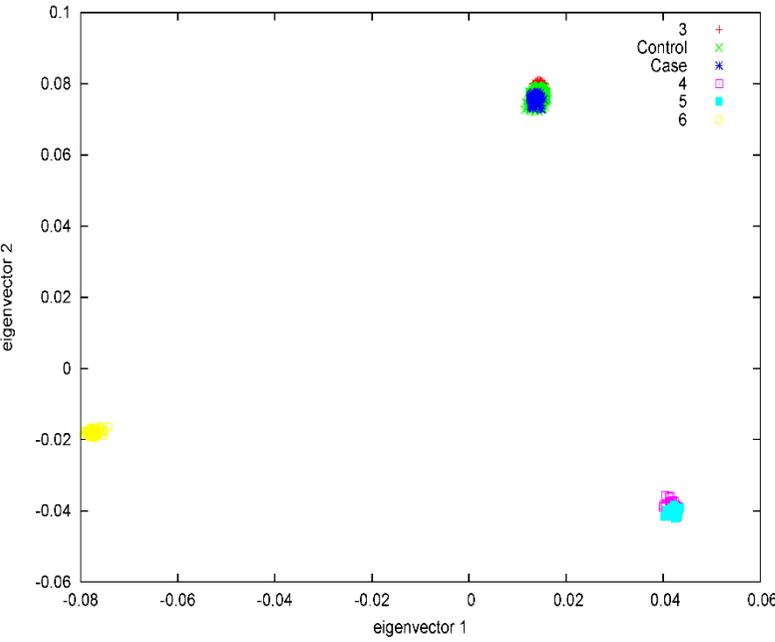
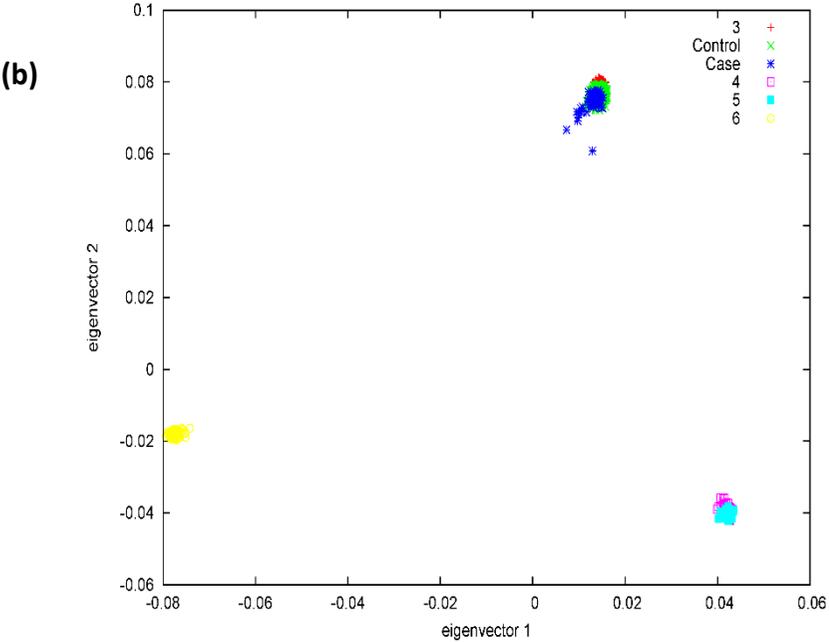
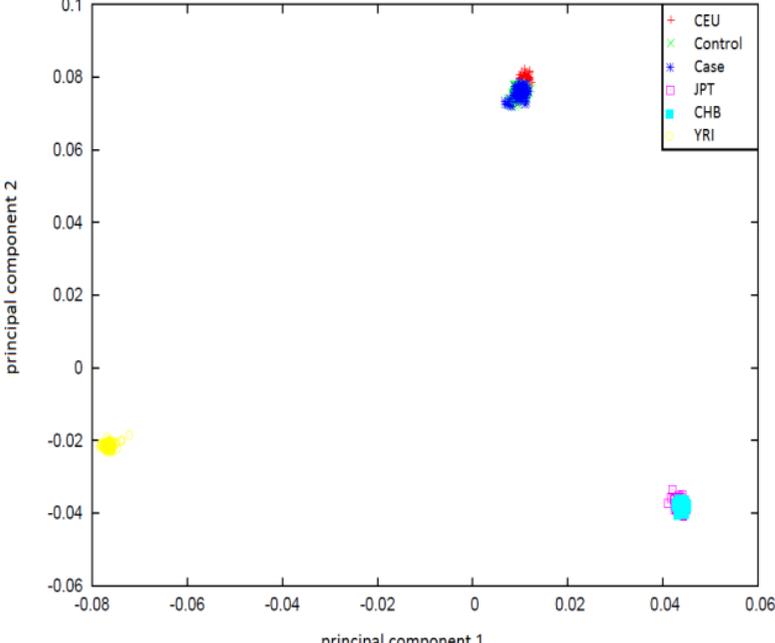
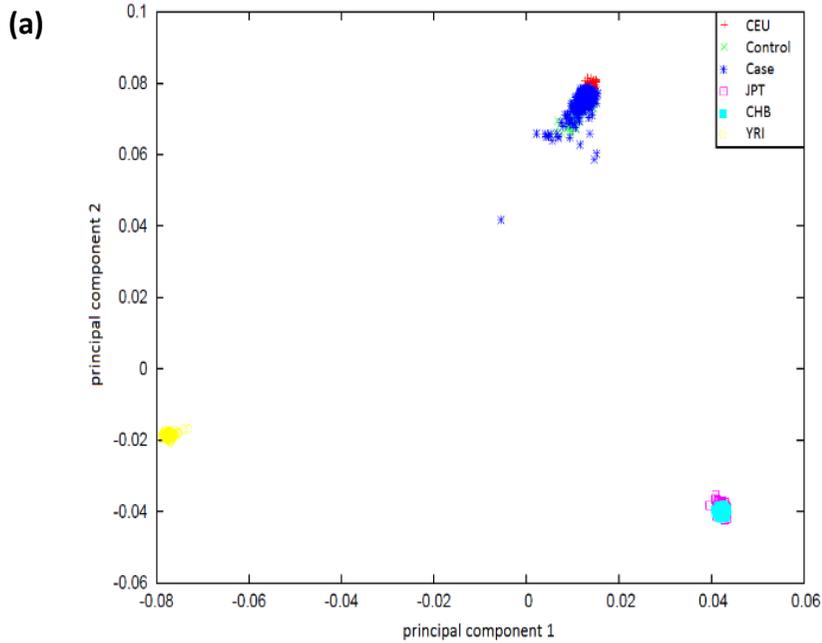
Karim Labreche^{1,2}, Maily Daniau^{2,3}, Amit Sud¹, Philip J. Law¹, Louis Royer-Perron^{2,4}, Amy Holroyd¹, Peter Broderick¹, Molly Went¹, Marion Benazra^{2,3}, Guido Ahle⁵, Pierre Soubeyran^{6,7}, Luc Taillandier⁸, Olivier L. Chinot^{9,10}, Olivier Casasnovas¹¹, Jacques-Olivier Bay¹², Fabrice Jardin¹³, Lucie Oberic¹⁴, Michel Fabbro¹⁵, Gandhi Damaj¹⁶, Annie Brion¹⁷, Karima Mokhtari^{2,18,19}, Cathy Philippe²⁰, Marc Sanson^{2,4,19}, Caroline Houillier², Carole Soussain²¹, Khê Hoang-Xuan^{2,4,*}, Richard S. Houlston^{1,*}, Agusti Alentorn^{2,4,*}, LOC Network[†].

Supplementary Data:

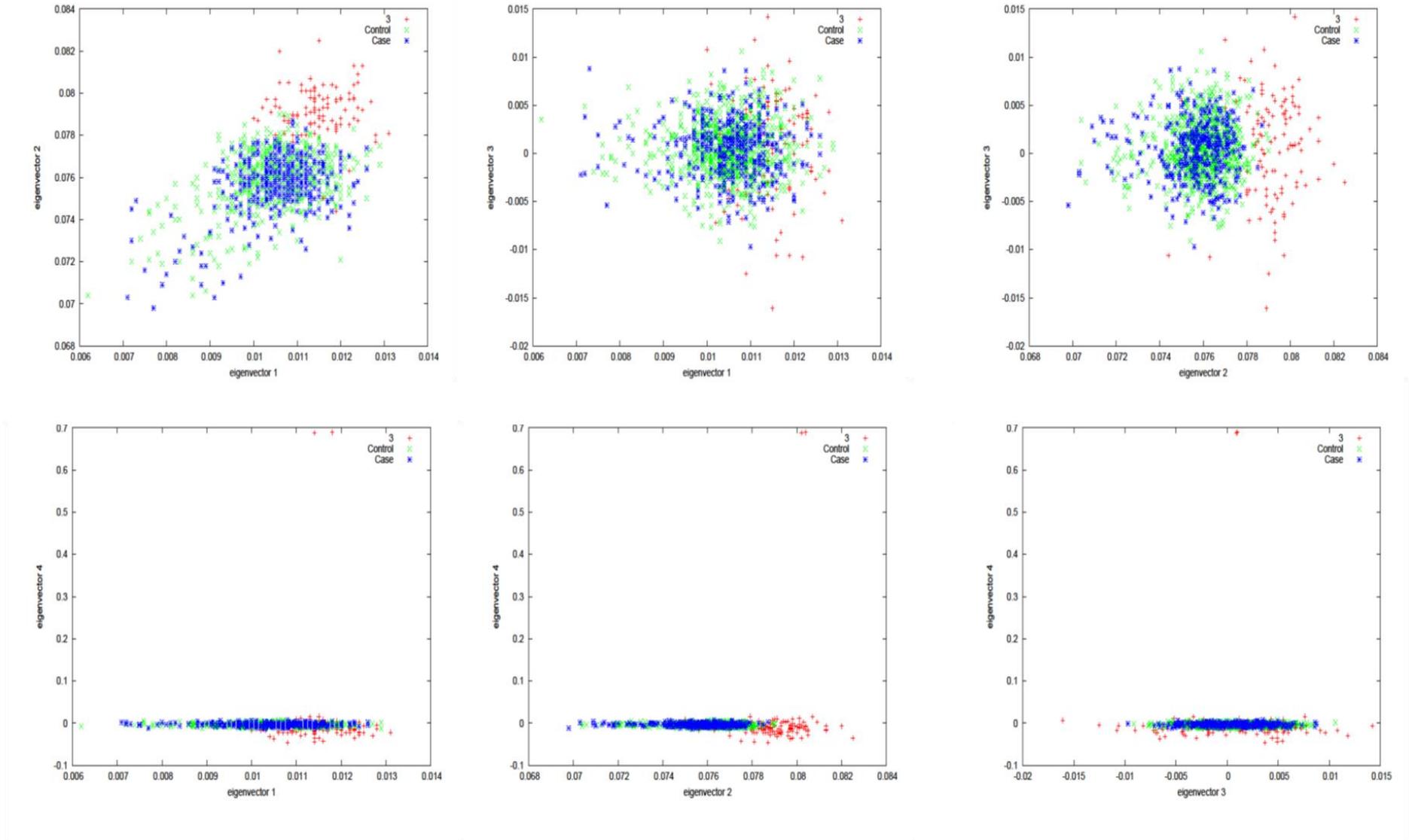
Supplementary Figure 1: Analysis strategy for primary cerebral nervous system lymphoma genome-wide association study. Primary cerebral nervous system lymphoma (PCNSL); Human leucocyte antigen (HLA). Reasons for sample exclusion detailed in **Supplementary Table 1**.



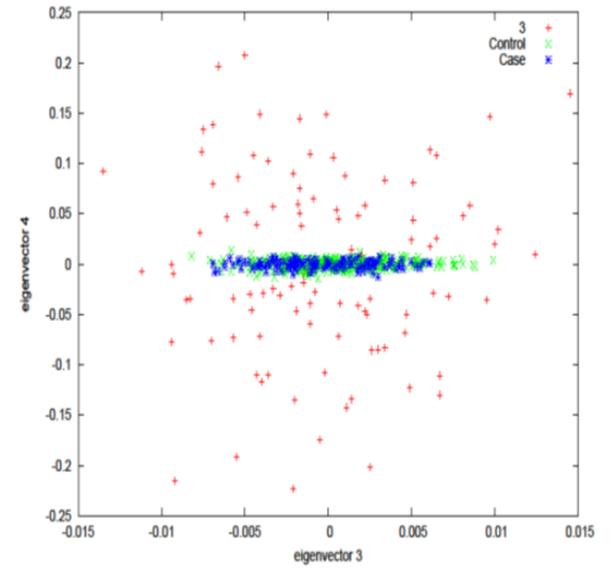
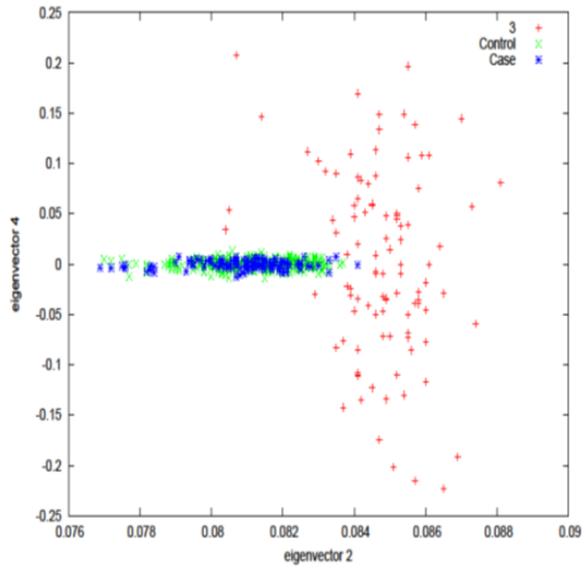
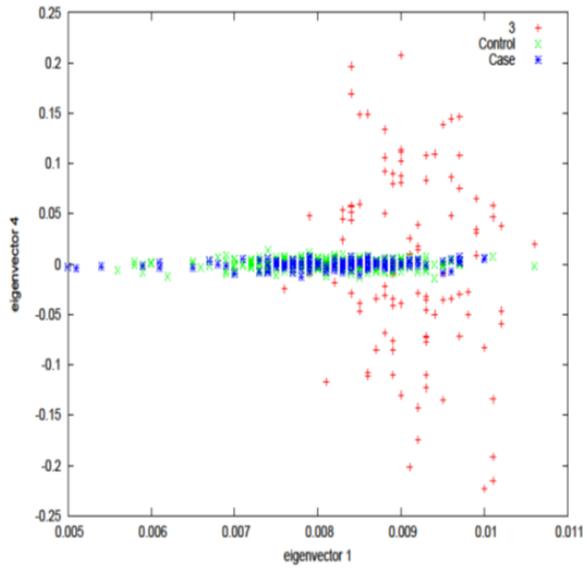
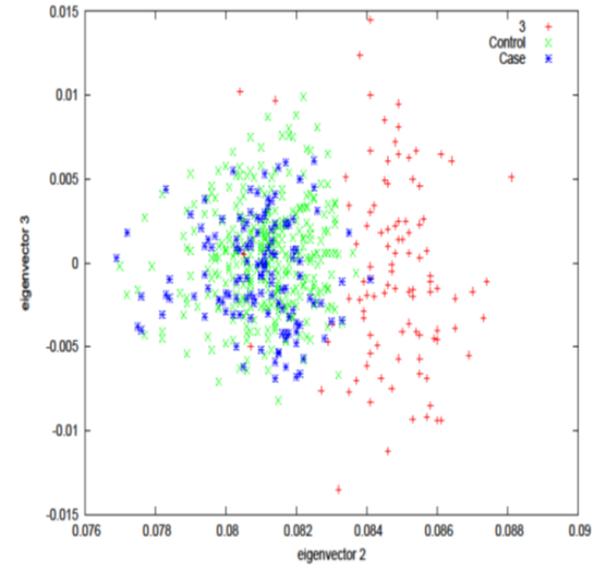
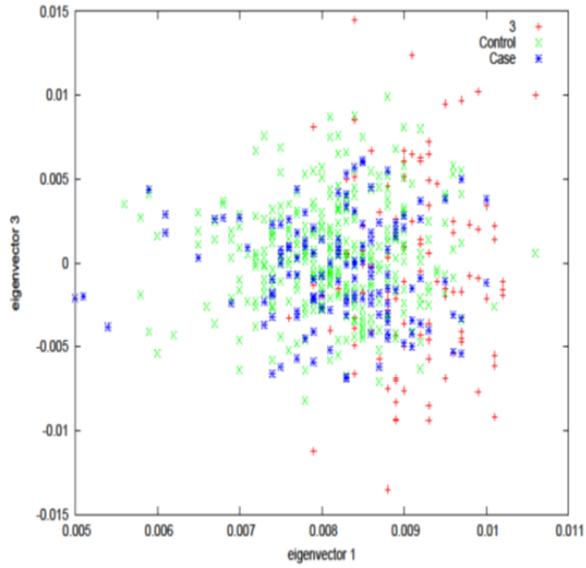
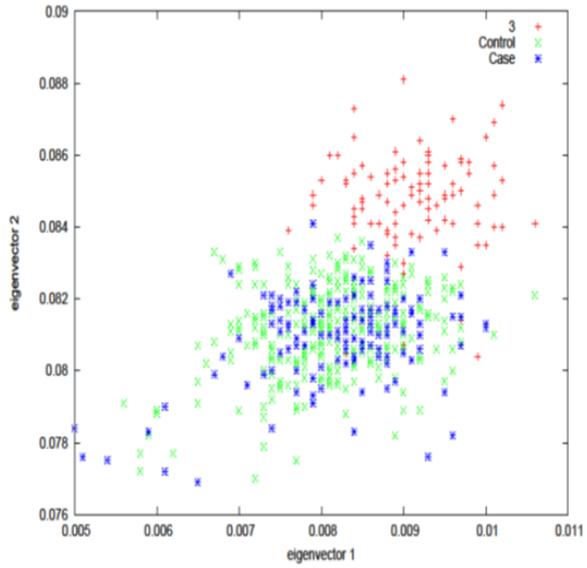
Supplementary Figure 2: Identification of individuals of non-European ancestry in cases and controls. (a) GWAS-1, left before excluding non-European ancestry in cases and controls, right after (b) GWAS-2. The first two principal components of the analysis are plotted. HapMap CEU individuals are plotted in red, JPT individuals are plotted in pink, CHB are plotted in cyan, YRI are plotted in yellow. Cases are plotted in blue, controls plotted in green. PCA plots for (c) GWAS-1 and GWAS-2 (d) without the subjects of non-European ancestry with four principal components after excluding non-European ancestry in cases and controls. HapMap CEU individuals are plotted in red.



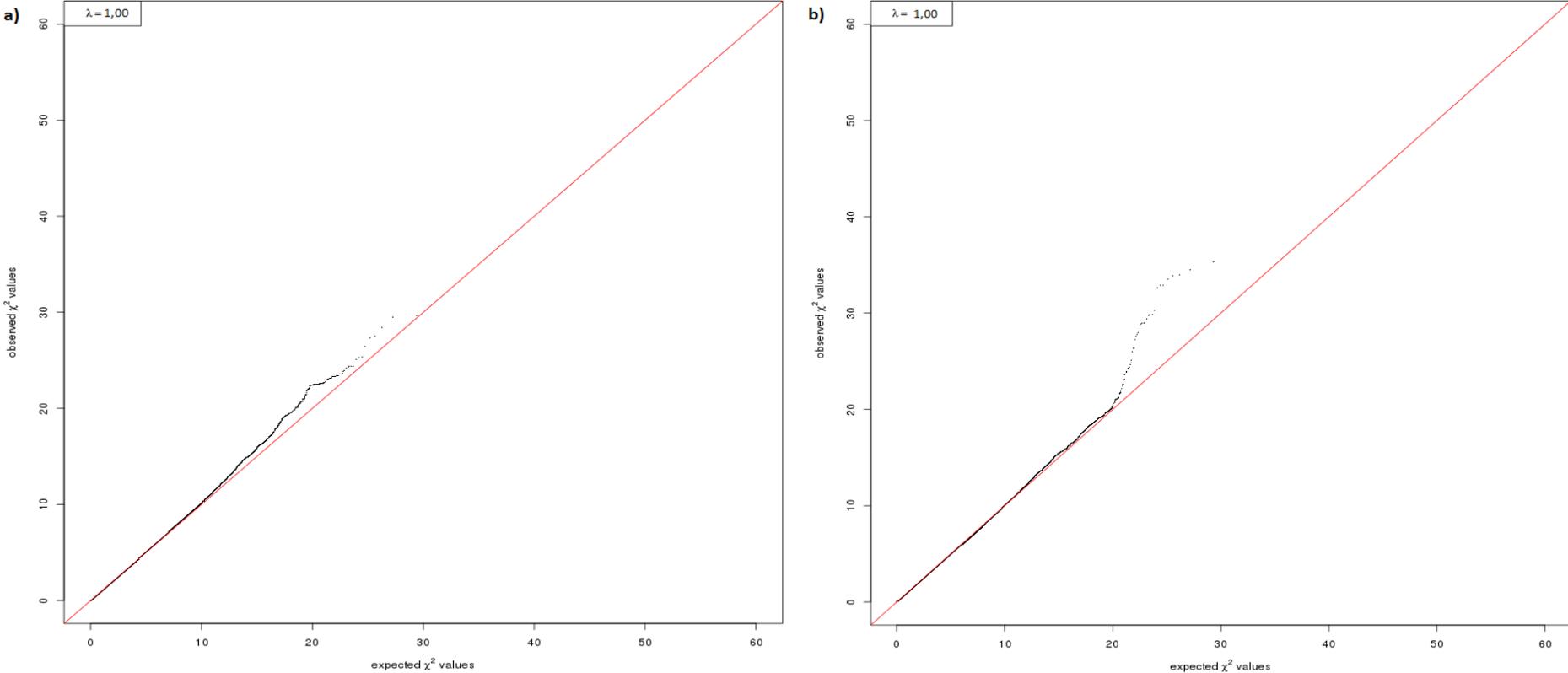
(c)



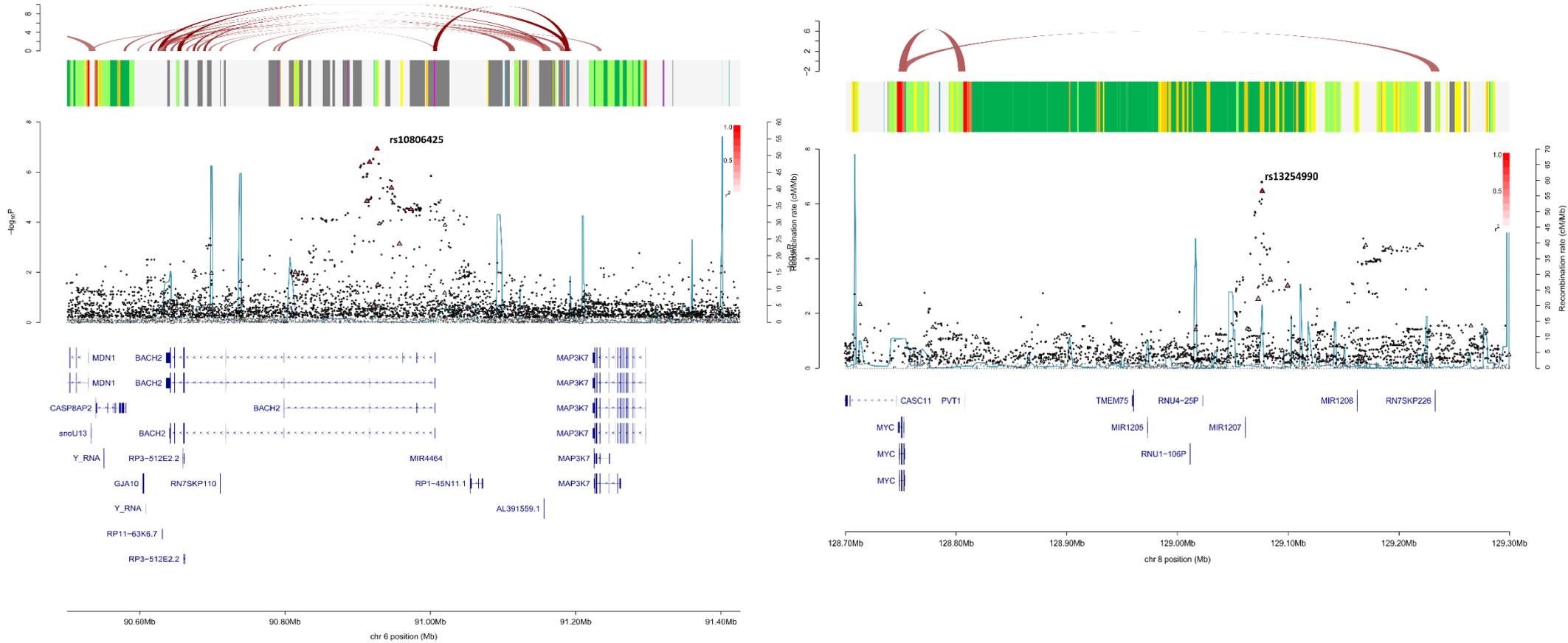
(d)



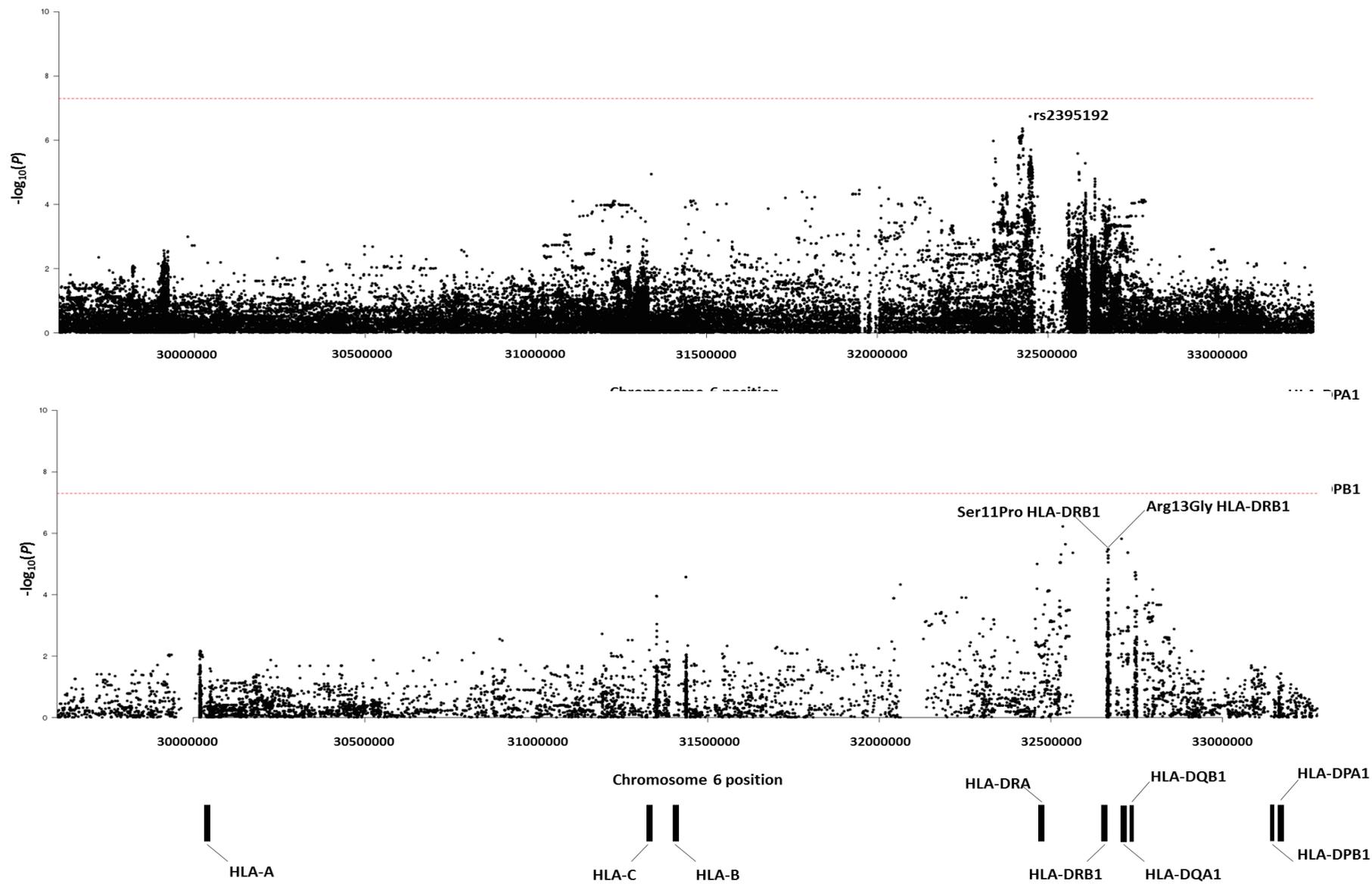
Supplementary Figure 3: Quantile-Quantile (Q-Q) plots of observed and expected χ^2 values of association between SNP genotype and risk of PCNSL after imputation. (a) GWAS-1, (b) GWAS-2. The red line represents the null hypothesis of no true association.



Supplementary Figure 4: Regional plots of association results and recombination rates for loci for primary cerebral nervous system lymphoma. 6q15 (a) and 8q24.21 (b). Plots (drawn using visPig) show association results of both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates. $-\log_{10} P$ -values (y-axis) of the SNPs are shown according to their chromosomal positions (x-axis). The sentinel SNP in each combined analysis is shown as a large circle or triangle and is labeled by its rsID. The color intensity of each symbol reflects the extent of LD with the top genotyped SNP, white ($r^2 = 0$) through to dark red ($r^2 = 1.0$). Genetic recombination rates, estimated using 1000 Genomes Project samples, are shown with a light blue line. Physical positions are based on NCBI build 37 of the human genome. Also shown are the chromatin-state segmentation track (ChromHMM) for lymphoblastoid cells using data from the HapMap ENCODE Project, and the positions of genes and transcripts mapping to the region of association. The top track represents capture Hi-C promoter contacts in GM12878 cells. The colour intensity of each contact reflects the interaction score.



Supplementary Figure 5: Regional association plot of PCNSL risk within the HLA region. Positions are based on NCBI build 36 of the human genome. The $-\log_{10}$ of the combined logistic regression test P -values are plotted against their physical chromosomal position. The broken red line represents the genome-wide level of significance ($P < 5 \times 10^{-8}$).



Supplementary Table 1: (a) Concordance between imputation and either sequencing or direct genotyping for the reported SNPs; r^2 : Pearson product-moment correlation coefficient between imputed and sequenced genotype. (b) Primer sequences; FP: Forward Primer; RP: Reverse Primer; A1: Allele 1; A2: Allele 2; C1: Common 1.

(a)

SNP	Nearest gene(s)	Alleles	Concordance (imputed/sequenced)	r^2	Dataset
rs116446171	EXOC2	GG	3/4	0.91	GWAS-1 (n=296) GWAS-2 (n=109)
		CG	46/46		
		CC	346/355		
rs41289586	ANO10	TT	0/0	0.99	GWAS-1 (n=334)
		CT	33/33		
		CC	300/301		
rs13254990	PVT1	TT	79/81	0.97	GWAS-1 (n=301) GWAS-2 (n=114)
		CT	183/190		
		CC	140/144		
rs10806525	BACH2	CC	182/182	0.99	GWAS-1 (n=306) GWAS-2 (n=114)
		CA	197/199		
		AA	39/39		

(b)

SNP	Primer sequence	Genotyping method
rs41289586	FP : AAGACTCGGCAGCATCTCCAGTGCCTCGTTCAACCTCATCT	Sequencing
	RP : GCGATCGTCACTGTTCTCCCAGGTAATGCGCAACTGT	
rs116446171	A1 : GAAGGTGACCAAGTTCATGCTCCTCGTTAACTTGCTCCAGGTC	KasPar allele-specific PCR
	A2 : GAAGGTGCGGAGTCAACGGATTCTCGTTAACTTGCTCCAGGTG	
	C1 : GTTACAGAACCCTGGTTTAGTGGGTA	
rs13254990	A1 : GAAGGTGACCAAGTTCATGCTCAATGATATGCTAGAAATATTAACCTCTC	KasPar allele-specific PCR
	A2 : GAAGGTGCGGAGTCAACGGATTCCAATGATATGCTAGAAATATTAACCTCTT	
	C1 : AATACTCAAACCTCAGTCTGAGACTGCAT	
rs10806525	A1 : GAAGGTGACCAAGTTCATGCTATTTGGACCAGTAATACTTTATTATGGGT	KasPar allele-specific PCR
	A2 : GAAGGTGCGGAGTCAACGGATTGGACCAGTAATACTTTATTATGGGG	
	C1 : GATGCTACTGAAGATCCTACCACGTA	

Supplementary Table 2: Details of the quality control filters applied to each genome-wide association study. Samples were excluded due to call rate (<90% or failed genotyping), ethnicity (principle components analysis or other samples reported to be not of white, European descent), relatedness (any individuals found to be duplicated or related within or between data sets through identity by state) or sex discrepancy. PCNSL, Primary cerebral nervous system lymphoma.

	GWAS-1 PCNSL		GWAS-2 PCNSL	
	Cases	Controls	Cases	Controls
Pre-quality control	420	856	163	357
Sex discrepancy	9	0	3	NA
Call rate fail	17	21	0	2
Heterozygosity rate	6	24	4	0
Related Individuals	3	0	1	0
Non-European Ancestry	39	23	29	9
Post-quality control [†]	346	788	129	346

[†] filters for quality control were performed simultaneously so numbers for each criteria may not sum to total removed. NA, not applicable

Supplementary Table 3: Details of the quality control filters applied to each genome-wide association (GWAS) study. Genotyped single nucleotide polymorphisms (SNPs) with a call rate <95% were excluded as were those with a minor allele frequency (MAF) <0.01 or displaying significant deviation from Hardy-Weinberg equilibrium (HWE) (i.e. $P < 10^{-5}$). PCNSL, Primary cerebral nervous system lymphoma.

	GWAS-1 PCNSL		GWAS-2 PCNSL	
	Cases	Controls	Cases	Controls
Genotyping Platform	Infinium OmniExpress-24 v1.2 BeadChip	HumanHap 660	Infinium OmniExpress- 24 v1.2 BeadChip	HumanHap 660
Pre-quality control	713014	425190	712331	425190
Call rate fail	3886	NA	7266	NA
HWE fail	NA	10	NA	11
MAF < 0.01	71881	NA	66467	NA
Different call rates between cases and controls	370777	177043	382334	176293
Others*	18107	0	18802	0
Post-quality control [†]	248137	248137	248886	248886

[†] filters for quality control were performed simultaneously so numbers for each criteria may not sum to total removed.

*exclusion of sex chromosome SNPs and triallelic SNPs.

Supplementary Table 4: Previously reported associated SNPs in DLBCL.

Cancer	Locus	Nearest gene(s)	SNP	Previously Risk allele	Previous RAF	Previous OR (95%)	Reported P value	Current Risk allele	Current RAF	GWAS P value	Current OR (95%)
Diffuse Large B Cell Lymphoma ²³	6p25.3	<i>EXOC2</i>	rs116446171	G	0.019	2.20 (1.87-2.59)	2.33×10^{-21}	G	0.036	1.53×10^{-13}	4.99 (3.26-7.65)
	8q24.21	<i>PTV1</i>	rs13255292	T	0.321	1.22 (1.15-1.29)	9.98×10^{-13}	T	0.37	3.81×10^{-7}	1.50 (1.29-1.76)
			rs4733601	A	0.47	1.18 (1.11-1.25)	3.36×10^{-11}	A	0.47	0.99	1.00 (0.86-1.16)
	6p21.33	<i>HLA-B</i>	rs2523607	A	0.120	1.32 (1.21-1.44)	2.40×10^{-10}	A	0.095	0.023	1.34 (1.04-1.73)
	2p23.3	<i>NCOA1</i>	rs79480871	T	0.076	1.34 (1.21-1.49)	4.23×10^{-8}	T	0.067	0.14	1.18 (0.97-1.42)

Supplementary Table 5: HLA associations for primary cerebral nervous system lymphoma. The risk allele is the allele corresponding to the estimated odds ratio. OR, odds ratio; CI, confidence interval; P_{het} , P -value for heterogeneity; I^2 , proportion of the total variation due to heterogeneity.

chromosome	SNP	Position (bp, hg19)	Allele A	Allele B	OR	95% CI	P -value	RAF	P_{het}	I^2
6	rs7754768	32528157	C	T	1.48	(1.27 - 1.73)	5.98×10^{-07}	0.36	0.88	0%
6	rs9271588	32698931	C	T	1.45	(1.25 - 1.68)	1.51×10^{-06}	0.55	0.20	38%
6	rs9268832	32535767	T	C	1.46	(1.25 - 1.7)	2.27×10^{-06}	0.35	0.97	0%
6	AA_DRB1_11_32660115_SP	32660115	A	P	1.44	(1.24 - 1.69)	3.35×10^{-06}	0.60	0.22	33%
6	AA_DRB1_13_32660109_SRG	32660109	A	P	1.44	(1.24 - 1.69)	3.35×10^{-06}	0.60	0.22	33%
6	SNP_DRB1_32660109_GC	32660109	A	P	1.44	(1.24 - 1.69)	3.35×10^{-06}	0.60	0.22	33%
6	SNP_DRB1_32660115_G	32660115	A	P	1.44	(1.24 - 1.69)	3.35×10^{-06}	0.60	0.22	33%
6	SNP_DRB1_32657334	32657334	C	T	1.44	(1.23 - 1.68)	3.94×10^{-06}	0.60	0.23	31%
6	AA_DQA1_34_32717152	32717152	E	Q	1.43	(1.23 - 1.67)	4.26×10^{-06}	0.56	0.26	22%
6	SNP_DQA1_32717151	32717151	G	C	1.43	(1.23 - 1.67)	4.26×10^{-06}	0.56	0.26	22%
6	rs1964995	32557389	G	A	1.44	(1.23 - 1.68)	4.32×10^{-06}	0.60	0.22	35%
6	rs3129891	32523058	A	G	1.55	(1.28 - 1.86)	4.90×10^{-06}	0.20	0.13	56%
6	AA_DRB1_11_32660115_SPD	32660115	A	P	1.44	(1.23 - 1.68)	5.28×10^{-06}	0.61	0.16	49%
6	SNP_DRB1_32660115_GT	32660115	A	P	1.44	(1.23 - 1.68)	5.28×10^{-06}	0.61	0.16	49%
6	AA_DRB1_47_32660007	32660007	Y	F	1.42	(1.22 - 1.65)	6.48×10^{-06}	0.50	0.10	62%
6	SNP_DRB1_32660007	32660007	T	A	1.42	(1.22 - 1.65)	6.48×10^{-06}	0.50	0.10	62%
6	AA_DRB1_13_32660109_SR	32660109	A	P	1.42	(1.22 - 1.66)	8.90×10^{-06}	0.55	0.27	17%
6	rs2213585	32521128	C	T	1.42	(1.22 - 1.67)	8.98×10^{-06}	0.34	0.70	0%
6	rs2213586	32521072	T	C	1.42	(1.22 - 1.67)	8.98×10^{-06}	0.34	0.70	0%
6	rs2227139	32521437	C	T	1.42	(1.22 - 1.67)	8.98×10^{-06}	0.34	0.70	0%
6	rs3763327	32521808	G	C	1.42	(1.22 - 1.67)	8.98×10^{-06}	0.34	0.70	0%
6	rs7192	32519624	T	G	1.42	(1.22 - 1.67)	8.98×10^{-06}	0.34	0.70	0%
6	rs7195	32520517	A	G	1.42	(1.22 - 1.67)	8.98×10^{-06}	0.34	0.70	0%
6	rs2395153	32453573	G	C	1.42	(1.22 - 1.66)	9.96×10^{-06}	0.62	0.37	0%

Supplementary Table 6: Previously reported associated SNPs in DLBCL and current reported associated SNPs in PCNSL and their association statistics from analysis of publicly available “Stage 1 NCI GWAS” DLBCL study cohort (1650 cases and 2665 controls).

Studies	Locus	Nearest genes(s)	SNP	Reported <i>P</i> value	Reported OR (95%)	Stage 1 NCIS GWAS	
						<i>P</i> value	OR (95%)
Diffuse Large B Cell Lymphoma ²³	6p25.3	<i>EXOC2</i>	rs116446171	2.33×10 ⁻²¹	2.20 (1.87–2.59)	1.41×10 ⁻⁸	1.94 (1.54-2.44)
	8q24.21	<i>PVT1</i>	rs13255292	9.98×10 ⁻¹³	1.22 (1.15–1.29)	5.89×10 ⁻⁶	1.20 (1.11-1.30)
			rs4733601	3.63×10 ⁻¹¹	1.18 (1.11–1.25)	1.02×10 ⁻⁴	1.16 (1.07-1.25)
	6p21.33	<i>HLA-B</i>	rs2523607	2.40×10 ⁻¹⁰	1.32 (1.21–1.44)	1.53×10 ⁻¹¹	1.45 (1.30-1.62)
	2p23.3	<i>NCOA1</i>	rs79480871	4.23×10 ⁻⁸	1.34 (1.21–1.49)	6.46×10 ⁻⁴	1.32 (1.21-1.54)
Current GWAS	6p25.3	<i>EXOC2</i>	rs116446171	1.53×10 ⁻¹³	4.99 (3.26–7.65)	1.41×10 ⁻⁸	1.94 (1.54-2.44)
	3p22.1	<i>ANO10</i>	rs41289586	1.53×10 ⁻¹³	3.82 (2.39 - 6.09)	0.26	1.15 (0.89-1.47)
	8q24.21	<i>PVT1</i>	rs13254990	1.87×10 ⁻⁸	1.54 (1.31 - 1.81)	3.54×10 ⁻⁶	1.21 (1.11-1.31)
	6q15	<i>BACH2</i>	rs10806425	1.33×10 ⁻⁷	1.51 (1.30 - 1.77)	2×10 ⁻⁴	1.15 (1.07-1.24)
	6p21.32	<i>HLA-DRA</i>	rs2395192	1.36×10 ⁻⁷	1.51 (1.29 - 1.76)	0.11	0.94 (0.87-1.01)

Supplementary Table 7: Previously reported 8q24.21 cancer associated SNPs and their LD with rs13254990.

Cancer	SNP	Previously reported		Current GWAS		LD with rs13254990		Previously reported OR
		Risk allele	P value	Risk allele	P value	r ²	D'	
Hodgkin's Lymphoma ¹	rs2019960	C	7 × 10 ⁻⁸	C	4.10 × 10 ⁻⁵	2.5 × 10 ⁻²	0.045	1.33
Colorectal Cancer ²	rs6983267	G	5 × 10 ⁻¹⁴	T	0.89	2.1 × 10 ⁻³	0.038	1.20
Prostate Cancer ³	rs1447295	A	6 × 10 ⁻¹⁸	C	0.80	4.05 × 10 ⁻⁵	0.071	1.38
Prostate Cancer ³	rs12682344	G	5 × 10 ⁻¹²	G	0.23	1.14 × 10 ⁻⁵	6 × 10 ⁻⁴	1.95
Prostate Cancer (early onset) ⁴	rs10505477	A	9 × 10 ⁻⁹	G	0.82	1.7 × 10 ⁻⁴	0.041	1.39
Colorectal Cancer ⁵	rs10505477	T	8 × 10 ⁻¹³	G	0.82	1.7 × 10 ⁻⁴	0.041	1.39
Breast Cancer (early onset) ⁶	rs2392780		1 × 10 ⁻⁸	G	0.38	1.0 × 10 ⁻⁵	0.041	1.15
Renal cell carcinoma ⁷	rs6470589	G	5 × 10 ⁻¹¹	G	0.39	1.01 × 10 ⁻⁵	9 × 10 ⁻⁴	1.27
Bladder Cancer ⁸	rs9642880	T	4 × 10 ⁻³⁸	T	0.08	1.73 × 10 ⁻⁵	0.056	1.22
Chronic lymphocytic leukemia ⁹	rs2466035	C	2 × 10 ⁻⁸	C	0.69	3.2 × 10 ⁻⁴	0.088	1.21
Breast Cancer ¹⁰	rs13281615	G	1 × 10 ⁻¹⁷	G	0.93	1.24 × 10 ⁻⁵	0.025	1.08
Breast Cancer ¹⁰	rs11780156	T	3 × 10 ⁻¹¹	T	0.09	6.4 × 10 ⁻³	0.21	1.07
Ovarian Cancer ¹¹	rs10088218	A	1 × 10 ⁻¹⁷	A	0.66	1.2 × 10 ⁻³	0.075	1.19
Prostate Cancer ¹²	rs6983561	C	4 × 10 ⁻¹³	C	0.22	1.46 × 10 ⁻⁵	0.0042	1.61
Prostate Cancer ¹³	rs13254738	C	4 × 10 ⁻¹⁰	A	0.60	8.25 × 10 ⁻⁵	0.0069	1.59
Prostate Cancer ¹⁴	rs4242384	C	3 × 10 ⁻¹⁶	A	0.90	8.27 × 10 ⁻⁵	0.065	1.88
Prostate Cancer ¹⁴	rs1016343	T	4 × 10 ⁻¹⁰	T	0.56	1.08 × 10 ⁻⁴	0.0023	1.31
Glioma ¹⁵	rs4295627	G	5 × 10 ⁻²¹	G	0.27	1.19 × 10 ⁻⁴	0.016	1.40
Breast Cancer ¹⁶	rs1562430	A	3 × 10 ⁻¹¹	C	0.36	1.42 × 10 ⁻⁶	0.035	1.16
Prostate Cancer ¹⁷	rs1456315	A	2 × 10 ⁻²⁹	C	0.65	6.43 × 10 ⁻⁵	0.01	-
Prostate Cancer ¹⁷	rs7837688	T	1 × 10 ⁻²⁵	G	0.71	8.33 × 10 ⁻⁵	0.056	-
Prostate Cancer ¹⁸	rs16902094	G	6 × 10 ⁻¹⁵	G	0.87	3.63 × 10 ⁻⁷	0.014	1.21
Prostate Cancer ¹⁸	rs16901979	A	3 × 10 ⁻¹⁴	A	0.22	4.05 × 10 ⁻⁶	0.01	1.80
Prostate Cancer ¹⁸	rs445114	T	5 × 10 ⁻¹⁰	C	0.95	9.96 × 10 ⁻⁶	0.013	1.14
Glioma ¹⁹	rs891835	G	8 × 10 ⁻¹¹	G	0.80	7.35 × 10 ⁻⁵	0.023	1.24
Colorectal Cancer ²⁰	rs7014346	A	9 × 10 ⁻²⁶	G	0.93	1.1 × 10 ⁻⁴	0.039	1.19
Prostate Cancer ²¹	rs6983267	G	7 × 10 ⁻¹²	T	0.89	2.17 × 10 ⁻⁴	0.038	1.28
Glioma ²²	rs55705857	A	2.24 × 10 ⁻³⁸	A	None	2.72 × 10 ⁻⁴	0.047	2.40

								Labreche et al
Diffuse Large B Cell	rs13255292	T	9.98×10^{-13}	T	3.81×10^{-7}	0.93	0.987	1.22
Lymphoma ²³	rs4733601	A	3.63×10^{-11}	G	0.99	4.21×10^{-5}	0.145	1.18
B Acute Lymphoblastic	rs4617118	G	2×10^{-12}	G	0.46	2.5×10^{-3}	0.085	1.28
Leukemia (B-ALL) ²⁴								

References:

1. Cozen, W. *et al.* A meta-analysis of Hodgkin lymphoma reveals 19p13.3 TCF3 as a novel susceptibility locus. *Nat Commun* **5**, 3856 (2014).
2. Zhang, B. *et al.* Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat Genet* **46**, 533-42 (2014).
3. Knipe, D.W. *et al.* Genetic variation in prostate-specific antigen-detected prostate cancer and the effect of control selection on genetic association studies. *Cancer Epidemiol Biomarkers Prev* **23**, 1356-65 (2014).
4. Lange, E.M. *et al.* Genome-wide association scan for variants associated with early-onset prostate cancer. *PLoS One* **9**, e93436 (2014).
5. Whiffin, N. *et al.* Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum Mol Genet* **23**, 4729-37 (2014).
6. Ahsan, H. *et al.* A genome-wide association study of early-onset breast cancer identifies PFKM as a novel breast cancer gene and supports a common genetic spectrum for breast cancer at any age. *Cancer Epidemiol Biomarkers Prev* **23**, 658-69 (2014).
7. Gudmundsson, J. *et al.* A common variant at 8q24.21 is associated with renal cell cancer. *Nat Commun* **4**, 2776 (2013).
8. Figueroa, J.D. *et al.* Genome-wide association study identifies multiple loci associated with bladder cancer risk. *Hum Mol Genet* **23**, 1387-98 (2014).
9. Berndt, S.I. *et al.* Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nat Genet* **45**, 868-76 (2013).
10. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**, 353-61, 361e1-2 (2013).
11. Pharoah, P.D. *et al.* GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet* **45**, 362-70, 370e1-2 (2013).
12. Xu, J. *et al.* Genome-wide association study in Chinese men identifies two new prostate cancer risk loci at 9q31.2 and 19q13.4. *Nat Genet* **44**, 1231-5 (2012).
13. Cheng, I. *et al.* Evaluating genetic risk for prostate cancer among Japanese and Latinos. *Cancer Epidemiol Biomarkers Prev* **21**, 2048-58 (2012).
14. Schumacher, F.R. *et al.* Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum Mol Genet* **20**, 3867-75 (2011).
15. Sanson, M. *et al.* Chromosome 7p11.2 (EGFR) variation influences glioma risk. *Hum Mol Genet* **20**, 2897-904 (2011).
16. Fletcher, O. *et al.* Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst* **103**, 425-35 (2011).
17. Takata, R. *et al.* Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet* **42**, 751-4 (2010).
18. Gudmundsson, J. *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet* **41**, 1122-6 (2009).
19. Shete, S. *et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet* **41**, 899-904 (2009).
20. Tenesa, A. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* **40**, 631-7 (2008).
21. Eeles, R.A. *et al.* Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* **40**, 316-21 (2008).
22. Enciso-Mora, V. *et al.* Deciphering the 8q24.21 association for glioma. *Hum Mol Genet* **22**, 2293-302 (2013).
23. Cerhan, J.R. *et al.* Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. *Nat Genet* **46**, 1233-8 (2014).
24. Wiemels J.L., et al. GWAS in childhood acute lymphoblastic leukemia reveals novel genetic associations at chromosome 17q12 and 8q24.21. *Nat Commun.* **9**, 286 (2018).



Click here to access/download

Hyperlinked Supplement (Excel tables, videos, etc.)
Supplementary Data 1.xlsx





Click here to access/download

Hyperlinked Supplement (Excel tables, videos, etc.)
Supplementary Data 2.xlsx

