American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

SOX11 expression is restricted to EBV-negative Burkitt lymphoma and associates with molecular genetic features

Tracking no: BLD-2023-023242R1

Marta Sureda-Gómez (Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Spain) Ingram Iaccarino (Kiel University, Germany) Anna De Bolòs ((Institut d'Investigacions BiomÃ¨diques August Pi i Sunyer (IDIBAPS); Centro de InvestigaciÃ³n BiomÃ©dica en Red de CÃ¡ncer (CIBERONC), Madrid, Spain.), Spain) Mieke Meyer (University of Kiel, Germany) Patricia Balsas (IDIBAPS, Spain) Julia Richter (Christian-Albrechts University Kiel, Germany) Marta Leonor Rodríguez (IDIBAPS, Spain) Cristina López (Institute of Human Genetics, Ulm University and Ulm University Medical Center, 89091, Ulm, Germany., Germany) Maria Carreras-Caballé (Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Spain) Selina Glaser (Institute of Human Genetics, Ulm University and Ulm University Medical Center, Germany) Ferran Nadeu (Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain., Spain) Pedro Jares (IDIBAPS, Spain) Maria Chiara Siciliano (University of Siena, Italy) Cristiana Bellan (University of Siena, Italy) Salvatore Tornambè (University of Siena, Italy) Roberto Boccacci (University of Siena, Italy) Guillem Clot (University of Barcelona, Barcelona, Spain., Spain) Lorenzo Leoncini (University of siena, Italy) Elías Campo (Department of Hematology Hospital Clinic of Barcelona, University of Barcelona, Barcelona, Spain., Spain) Reiner Siebert (University of Ulm, Germany) Virginia Amador (Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain., Spain) Wolfram Klapper (Department of Pathology, Kiel, Germany)

Abstract:
SRY-related HMG-box gene 11 (SOX11) is a transcription factor overexpressed in mantle cell lymphoma (MCL), a subset of Burkitt lymphomas (BL) and precursor lymphoid cell neoplasms but is absent in normal B-cells and other B-cell lymphomas. SOX11 has an oncogenic role in MCL but its contribution to BL pathogenesis remains uncertain. Here, we observed that the presence of Epstein-Barr virus (EBV) and SOX11 expression were mutually exclusive in BL. SOX11 expression in EBV- BL was associated with an IG::MYC translocation generated by aberrant class switch recombination, while in EBV-/SOX11- tumors the IG::MYC translocation was mediated by mistaken somatic hypermutations. Interestingly, EBV- SOX11 expressing BL showed higher frequency of SMARCA4 and ID3 mutations compared to EBV-/SOX11- cases. By RNA-sequencing, we identified a SOX11-associated gene expression profile, with functional annotations showing partial overlap with the SOX11 transcriptional program of MCL. Contrary to MCL, no differences on cell migration or BCR signaling were found between SOX11- and SOX11+ BL cells. However, SOX11+ BL showed higher adhesion to VCAM-1 than SOX11- BL cell lines. Here we demonstrate that EBV- BL comprises two subsets of cases based on SOX11 expression. The mutual exclusion of SOX11 and EBV, and the association of SOX11 with a specific genetic landscape suggest a role of SOX11 in the early pathogenesis of BL.

Conflict of interest: No COI declared

COI notes: The institution of Dr. WK has received research funding from Roche, Janssen, Amgen, Takeda, Regeneron and InCyte, not related to the submitted work.

Preprint server: No;

Author contributions and disclosures: MSG designed, performed and interpreted all in vitro experiments, bioinformatic and statistical analyses, and wrote the manuscript; II designed and interpreted experiments, performed bioinformatic and statistical analyses; JR, MM, CL, LL, WK and RS provided and analyzed molecular and clinical data of BL patients; PB, MLR, ADB, MC and MM performed in vitro experiments; MCS, ST, MB and RB performed ddPCR, IHC and RNAscope; FN, GC, PJ and SG performed bioinformatic and statistical analyses; EC helped to interpret data; WK and VA designed and supervised experiments, analyzed data and wrote the manuscript; and all authors discussed the results and commented on the manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: The RNA-seq raw data used at our manuscript BLD-2023-023242R1 has been successfully registered with the BioProject database. Our project information will be accessible with the following link: http://www.ncbi.nlm.nih.gov/bioproject/1097853, using our BioProject ID: PRJNA1097853.

Clinical trial registration information (if any):

SOX11 expression is restricted to EBV-negative Burkitt lymphoma and associates with molecular genetic features

Marta Sureda-Gómez[1]*, Ingram Iaccarino[2]*, Anna De Bolòs[1,3], Mieke Meyer[2], Patricia Balsas[1], Julia Richter[2], Marta-Leonor Rodríguez[1], Cristina López[1,3,4], Maria Carreras-Caballé[1], Selina Glaser[4], Ferran Nadeu[1,3], Pedro Jares[1,3,5,6], Guillem Clot[1,3,6], Maria Chiara Siciliano[7], Cristiana Bellan[7], Salvatore Tornambè[7], Roberto Boccacci[7], Lorenzo Leoncini[7], Elias Campo[1,3,5,6], Reiner Siebert[4], Virginia Amador[1,3†#], Wolfram Klapper[2†]

1. Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain.

2. Department of Pathology, Hematopathology Section and Lymph Node Registry, University of Kiel, Kiel, Germany.

3. Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain.

4. Institute of Human Genetics, Ulm University and Ulm University Medical Center, 89091, Ulm, Germany.

5. Department of Hematology Hospital Clinic of Barcelona, Barcelona, Spain.

6. University of Barcelona, Barcelona, Spain.

7. Department of Medical Biotechnologies, Section of Pathological Anatomy, University of Siena, Italy.

* Contributed equally to this work.

† Share senior authorship.

Corresponding author.

RUNNING TITLE: SOX11 and EBV in BL pathogenesis

CORRESPONDING AUTHOR: Dr. Virginia Amador, PhD. Centre Esther Koplowitz (CEK), C/Rosselló 149-153, Barcelona-08036, Spain. E-mail: vamador@recerca.clinic.cat. Phone: +34 932275400 (extension 4582).

The RNA-seq raw data used at our manuscript BLD-2023-023242R1 has been successfully registered with the BioProject database. Our project information will be accessible with the following link: http://www.ncbi.nlm.nih.gov/bioproject/1097853, using our **BioProject ID: PRJNA1097853.**

KEYWORDS: Burkitt Lymphoma (BL), SRY-related HMG-box gene 11 (SOX11), Epstein–Barr Virus (EBV), translocations of *MYC* proto-oncogene

KEY POINTS:

- SOX11 expression and EBV infection occur in alternative subsets of BL with different profile of somatic mutations.
- Among EBV-negative BL *IG*::*MYC* translocation is generated by CSR in SOX11-positive BL and SHM in SOX11-negative tumors.

Text word count: 3989
Abstract word count: 213
Number of tables and figures: 7
Number of references: 49
Scientific category: Lymphoid neoplasia

ABSTRACT

SRY-related HMG-box gene 11 (SOX11) is a transcription factor overexpressed in mantle cell lymphoma (MCL), a subset of Burkitt lymphomas (BL) and precursor lymphoid cell neoplasms but is absent in normal B-cells and other B-cell lymphomas. SOX11 has an oncogenic role in MCL but its contribution to BL pathogenesis remains uncertain. Here, we observed that the presence of Epstein-Barr virus (EBV) and SOX11 expression were mutually exclusive in BL. SOX11 expression in EBV- BL was associated with an *IG*::*MYC* translocation generated by aberrant class switch recombination, while in EBV-/SOX11- tumors the *IG*::*MYC* translocation was mediated by mistaken somatic hypermutations. Interestingly, EBV- SOX11 expressing BL showed higher frequency of *SMARCA4* and *ID3* mutations compared to EBV-/SOX11- cases. By RNA-sequencing, we identified a SOX11-associated gene expression profile, with functional annotations showing partial overlap with the SOX11 transcriptional program of MCL. Contrary to MCL, no differences on cell migration or BCR signaling were found between SOX11- and SOX11+ BL cells. However, SOX11+ BL showed higher adhesion to VCAM-1 than SOX11- BL cell lines. Here we demonstrate that EBV- BL comprises two subsets of cases based on SOX11 expression. The mutual exclusion of SOX11 and EBV, and the association of SOX11 with a specific genetic landscape suggest a role of SOX11 in the early pathogenesis of BL.

INTRODUCTION

Burkitt lymphoma (BL) is a highly proliferative B-cell neoplasm that originates from germinal center B-cells.[1] It is the most common B-cell lymphoma in children and adolescents but also occurs in adults.[2,3] Three clinical variants are distinguished: endemic BL (eBL), sporadic BL (sBL) and immunodeficiency-related BL. eBL is usually positive for Epstein-Barr virus (EBV), occurs mainly in countries of central Africa in which malaria is endemic, and presents with jaw or facial bone involvement in pediatric patients. Clinically sBL differs from eBL as it involves mostly the abdomen (Peyer's patches), head and neck lymph nodes, and in some cases, bone marrow.[1,4,5] Moreover, sBL is less commonly positive for EBV. However, when detected in sBL, EBV is more frequent in adult cases.[6,7]

The genetic hallmark of BL is the *MYC* rearrangement to one of the immunoglobulins (IG) loci, leading to the constitutive overexpression of *MYC*.[8–10] *MYC* dysregulation in B-cells is not sufficient for BL development and additional genomic changes are required.[11,12] Interestingly, several studies have revealed important genetic and molecular differences depending on the EBV status of BL patients.[13–21]

BL is one of the few lymphomas that shows expression of the SRY-related HMG-box gene 11 (SOX11).[22–24] SOX11 expression in BL occurs in 25-55% of tumors predominantly in pediatric patients.[7,24,25] Moreover, SOX11 expression is included in the transcriptional molecular signature used to classify BL.[26] SOX11 expression in MCL is characteristic of the conventional molecular subtype (cMCL) with worse outcome than the SOX11-negative (SOX11-) leukemic non-nodal MCL subtype (nnMCL).[27] In contrast, no association between SOX11 expression and survival has been found in BL.[24] Several *in vitro* and *in vivo* studies have shown the oncogenic role of SOX11 in

the pathogenesis of MCL.[28–34] However, the contribution of SOX11 expression to BL pathogenesis remains unknown.

To understand the relevance of SOX11 in BL, we have investigated the relationship of SOX11 expression with different molecular variables in primary tumors and evaluated the modulation of gene expression profiles (GEP) and functional changes upon SOX11 overexpression and knockout in BL cell lines.

METHODS

Cell lines

SOX11- BL cell lines Ramos and DG75 (ATCC CRL-1596 and ATCC CRL-2625, respectively); and the SOX11-positive (SOX11+) BL cell line BL2 (DSMZ ACC 625) were used to generate stable transduced DG75 ER-SOX11+ and Ramos SOX11+, ectopically overexpressing SOX11, and BL2-SOX11 knockout (KO) BL cell lines. Moreover, we used the stable transduced Z138-SOX11KO[34] and JVM2 (JVM2-SOX11+)[32,34] and its control (Z138CT and JVM2CT) MCL cell lines, previously generated by our laboratory. See more details on cell culture, plasmid, and generation of stable transduced BL cell lines at supplemental methods.

BL patient cohorts

Four previously published BL series were used to correlate SOX11 expression or positivity with different BL molecular characteristics. Duplicated cases between series have been considered and ruled out from one of the duplicated series. The BL Genome Sequencing Project (BLGSP)[21] includes 117 pediatric BL (96 endemic and 21 sporadic) with available RNA-sequencing (seq) and molecular data, including EBV status, breakpoint of *IG::MYC* and recurrently mutated genes, obtained by whole genome sequencing (WGS). Richter et al. cohort[7] includes 138 sBL patients (80 children and 58 adults) with available SOX11 immunohistochemistry (IHC), EBV-encoded small nuclear RNA (EBER) in situ hybridization (ISH) and targeted DNA-seq data (79 cases). Burkhardt et al. cohort[35] includes sBL cases with deep targeted DNA-seq data. ICGC MMML-Seq cases[16] includes 24 pediatric sBL cases with available SOX11 IHC and molecular data, including EBV status, breakpoint of *IG::MYC* and recurrently mutated

genes, obtained by WGS. All pediatric patients were aged <20 years, and the adults ≥20 years.

The study was approved by the Ethics Committee of the Medical Faculty of the University of Kiel (D 429/13) and conducted in accordance with the Decla-ration of Helsinki.

SOX11 status

IHC of SOX11 was performed for 51 pediatric BL cases with available FFPE tissue[35] on an automated strainer (Leica) by using a mouse monoclonal antibody against SOX11 (Cell Marque, MRQ-58) and a pH6 antigen retrieval solution, as previously described.[36,37] SOX11 IHC was previously obtained for ICGC MMML-Seq[16] and Richter et al. cohort.[7] SOX11 was scored positive when at least 10% of lymphoma cells showed unambiguous nuclear staining although in most cases the majority of neoplastic cells rather than small subsets were positive for SOX11. In BLGSP cases for which no IHC was possible,[21] SOX11+ expression was defined as >10.5 $\log_2$ transformed value, obtained by RNA-seq data.

EBV studies

We performed EBER ISH in 51 pediatric BL cases with available FFPE biopsy specimens[35] using Leica Bond-MAX staining systems and reagents (Leica). EBV-positivity was defined as most tumor cells being positive. Presence of EBV traces were investigated by droplet digital PCR (ddPCR) for EBNA1 and BamHI in a cohort of 37 BL specimens negative for EBER (n=12 SOX11+ and n=25 SOX11-) as previously described (supplemental methods).[38,39] Combined IHC for CD10 and the RNAscope for EBNA1 was performed using Leica Bond III automated system (Leica, Germany) in

FFPE BL samples EBER-ISH negative but positive by ddPCR (supplemental methods).[38,39]

Molecular profiling

The IG region involved and the mechanism leading to the *IG::MYC* translocation were evaluated in 89 BL cases from BLGSP[21] and 24 BL cases from ICGC MMML-Seq.[16] *IG::MYC* translocations with breakpoints on the IG loci localized inside or near (<500 base-pairs) of class switch regions of IGH were classified as translocations mediated by class switch recombination (CSR), while those inside V(D)J regions but far from recombination signal sites (RSS) (>15 base-pairs) were classified as mediated by somatic hypermutation (SHM). Translocations with breakpoints localized close to RSS (<15 base-pairs) were classified as mediated by RAG during V(D)J recombination.

For mutational analysis in BL patients, lists of mutated driver genes were obtained from the different publications for a total of 267 patients after excluding overlapped cases.[7,16,21,35] Some of the data was obtained by WGS and other by targeted mutational analysis. For targeted mutational analysis, only mutations with a variant allele frequency (VAF) ≥10% were considered. Only genes that were mutated in ≥10% of the BL cases were used for oncoprint analysis and comparisons between groups.

RNA-sequencing

RNA was obtained using the RNeasy® Mini RNA extraction kit (Qiagen) following manufacturer instructions. RNA quality was checked using a Bioanalyzer (Agilent) and mRNA libraries were prepared using the TruSeq stranded mRNA kit for DG75 cell lines or the TruSeq RNA Library Prep Kit v2 for Ramos and BL2 cell lines

(Illumina). Samples were sequenced on NovaSeq or NextSeq2000 sequencers. For each condition, three technical replicates were analyzed.

RNA-seq data from DG75 cell lines were analyzed using the open source web-based Galaxy.[40] Paired-end fastq files were aligned to the human genome (GRCh38) using HISAT2. Counts files were generated with featureCounts using GRCh38.102.gtf as gene annotation file. For Ramos and BL2 cell lines, single-end sequencing reads were processed and aligned as previously described. Gene count matrix was obtained for BLGSP[21] BL primary cases. RNA-seq data analyses are detailed in supplemental methods.

Statistics

Methods are described in the supplemental methods.

The study was approved by the Ethics Committee of the Medical Faculty of the University of Kiel (D 429/13) and conducted in accordance with the Decla-ration of Helsinki.

RESULTS

SOX11 expression and EBV infection are mutually exclusive in BL

We first investigated the association between *SOX11* expression and different molecular characteristics of the tumors and clinical features of the patients. Using previously published RNA-seq and clinical data from 117 pediatric BLs,[21] we observed significantly higher SOX11 mRNA levels in EBV- than EBV+ tumors (Figure 1A-B, respectively). To confirm this observation, we used 189 cases of two independent series of pediatric and adult sBL patients,[7,35] performing SOX11 IHC and EBER ISH in the FFPE tissue sections of tumors in which these data were not available (Supplemental Table S1 and Supplemental Figure S1A-F). None of the 17 EBV+ BL (0% with CI95% = 0%, 19.5%), whereas 81 of the 172 EBV- BL were positive for SOX11 (47.1% with CI95% = 39.5%, 54.8%), showing mutual exclusivity between SOX11+ and EBV+ (p-value=0.003) (Figure 1C). SOX11 expression was significantly associated with sBL, but not exclusively since 3 BL from endemic areas showed high SOX11 mRNA expression, all of them EBV-, comparable to those observed in SOX11+ sBL cases *(*Figure 1B, red circle). No other molecular or clinical features differed between these three cases compared to the rest of eBL cases (all three were HIV negative (R. Morin, K. Dreval personal communication). Unfortunately, SOX11 IHC was not available for these cases.[21]

Together these results suggest that SOX11 expression delineates a different molecular subtype of EBV- BL. To analyze this hypothesis, we established three different groups of patients according to their EBV and SOX11 status: EBV+ (all SOX11-), EBV-/SOX11- "double negative" and EBV-/SOX11+ and analyzed its association with different molecular BL features.

IG::MYC translocation is predominantly generated by CSR in EBV-/SOX11+ BL and by SHM in EBV-/SOX11- BL

To study the IG partners and the mechanisms involved in the generation of *IG*::*MYC* translocation according to EBV and SOX11 status in BL cases, we used the breakpoints of the *IG*::*MYC* translocation in 24 and 89 pediatric BL patients from the ICGC MMML-Seq[16] and BLGSP[21], respectively, for which these data was available. The SOX11 status was determined by IHC in ICGC MMML-Seq BL cases, and by RNA-seq data in the BLGSP BL cases (Supplemental Figure S1G). We observed that 56/69 (81%) EBV+, 25/26 (96%) EBV-/SOX11+ and 12/18 (67%) EBV-/SOX11- "double-negative" BL carried an *IGH*::*MYC* translocation, and in lower proportion, *IGL*::*MYC* or *IGK*::*MYC* translocations with IG partner frequencies being significantly different in EBV-/SOX11- "double-negative" and EBV-/SOX11+ cases (p-value=0.0134) (Figure 2A, and Supplemental Table S2).

The potential mechanism mediating the translocation, considering the three IG loci, were determined in 105 of the total 113 BL cases for which these data was available, finding significant differences between the three groups (p-value=0.0002). We found the breakpoint in a class switch region of the IGH, in 87% of the EBV-/SOX11+ (21/24) but only in 47% of the EBV-/SOX11- "double-negative" (7/15) BL and 38% of the EBV+ BL (25/66). At the same time, 12.5% of the EBV-/SOX11+ (3/24), 53% of the EBV-/SOX11- "double-negative" (8/15), and 61% of the EBV+ (40/66) BL cases had the breakpoint located in the V(D)J region as a result of SHM process. Only one case showed evidence of acquisition of the translocation by aberrant V(D)J recombination mediated by RAG in the group of EBV+ BLs (Figure 2B and Supplemental Table S2). Pairwise comparisons showed significant differences in the mechanism mediating the translocation between EBV-/SOX11+ and EBV-/SOX11-

"double-negative" (p-value=0.005), and between EBV-/SOX11+ and EBV+ BLs (p-value=4.6e-05).

As previously described,[21] significantly lower AICDA mRNA expression levels were observed in EBV-/SOX11+ (pval=2.4e-09) and EBV-/SOX11- (pval=0.052) compared to EBV+ BLs. However, no significant differences were observed between EBV-/SOX11- and EBV-/SOX11+ BLs (pval=0.21) (Supplemental Figure S2).

Together, these data suggest that among EBV-negative BL SOX11 status is associated with the early pathogenetic event of the *MYC*-translocation.

SOX11+ cases have a distinct mutational landscape among EBV- BL

Several genes are recurrently mutated in BL, promoting oncogenic mechanisms responsible for the development of tumor cells.[14] We combined previously published data on recurrently mutated genes obtained by WGS from 117 pediatric eBL and sBL patients from the BLGSP,[21] and from 24 pediatric sBL cases from the ICGC MMML-Seq,[16] and targeted mutational data on driver BL genes of two different series with 79 pediatric and adults,[7] and 47 pediatric[35] sBL cases. We identified 17 coding genes mutated in ≥10% of BL cases (Figure 3A and Supplemental Table S3).

Then, we analyzed the frequency of mutations of these 17 genes in the three groups of BL patients previously established: EBV+ (all SOX11-), EBV-/SOX11- and EBV-/SOX11+. We observed significant differences in the frequency of *CCND3*, *DDX3X*, *FBXO11*, *FOXO1*, *ID3*, *MYC*, *SIN3A*, *SMARCA4* and *TP53* mutations between these groups of patients (q-value<0.1) (Figure 3B and Supplemental Table S4). Pairwise comparisons showed that EBV-/SOX11- "double negative" and EBV-/SOX11+ BLs share a higher frequency of mutations in *CCND3*, *ID3* and *TP53* genes, and lower in

FOXO1 gene relative to EBV+ BLs (q-value<0.1). However, EBV+ cases showed a significant higher frequency of mutations in *DDX3X* and *SIN3A*, and fewer in *MYC* compared to EBV-/SOX11+, but not to EBV-/SOX11- "double negative" BL (q-value<0.05), suggesting that these differences cannot be attributed to the EBV status alone. In addition, among EBV- cases, EBV-/SOX11+ BL had a significantly higher frequency of mutations in *SMARCA4* and *ID3* (43% and 80%, respectively) compared to EBV-/SOX11- "double negative" (18% and 63%, respectively; q-value=0.14 in both comparisons). As expected *SMARCA4* and *ID3* were also less frequently mutated in EBV+ compared to EBV-/SOX11+ cases (9% and 35%, respectively; q-values<0.001) (Figure 3B and Supplemental Table S5). Thus, both EBV infection[7,20,21] and SOX11 expression in BL are associated with a distinct mutational pattern.

Highly sensitive detection of EBV in SOX11/EBER "double negative" BL

Since SOX11 expression and EBV detected by the gold standard method EBER ISH leaves a third group of BL, being negative for both features ("double negative"), more common in adult patients (Figure 3A), we asked if EBER ISH may miss the detection of EBV in these BL. Thus, a cohort of 37 BL samples that were negative for EBER (n=14 SOX11+ and n=23 SOX11-) were blindly tested for traces of EBV by ddPCR for both EBNA1 and BamHI-W conserved regions of the EBV genome as previously described.[38] Twelve cases (32.4%) were positive for BamHI-W (0.19-18 copies/µl) of which nine were also positive for EBNA1 (0.19-2.7 copies/µl). In six cases the presence of EBV sequences in tumor cells were confirmed by dual staining with RNAscope for EBNA1 and IHC for CD10, showing colocalization of EBV specific signals in CD10+ lymphoma cells. The vast majority of BLs with traces of EBV were SOX11- (10/12; 83.3%), while only two (2/12; 16.7%) were SOX11+ (Table 1).

The biological relevance of EBV-traces may be debatable. However, traces of EBV "corrected" the EBV-status of BL almost exclusively among SOX11- but not among SOX11+ BL, ruling out that we miss a relevant number of EBV+/SOX11+ double positive cases by EBER, and reinforcing that SOX11 and EBV positivity are mutually exclusive in BL. Limited molecular data were available for cases tested for traces of EBV.

Oncogenic pathways regulated by SOX11 in BL cell lines

Mutual exclusivity of SOX11 and EBV status in BL was also demonstrated in the analysis of SOX11 expression in established BL cell lines (Supplemental Figure S3A-B). To identify oncogenic pathways regulated by SOX11 in BL cells, we first used a SOX11- BL cell line, DG75, to ectopically express the SOX11 protein fused to the hormone binding domain of the estrogen receptor (ER-SOX11). Fusion with the ER makes SOX11 activatable upon treatment with 4-Hydroxytamoxifen (4-OHT). As shown in Figure 4A and B, although expressed, the ER-SOX11 protein translocate to the nucleus only when cells are treated with 4-OHT. As we found weak background nuclear expression in absence of 4-OHT, we decided to compare RNA-seq GEPs of DG75 ER-SOX11 and control DG75 ER cells, both treated with 4-OHT for 8 or 24 hours.

Principal component analysis (PCA) showed that the variability between samples was significantly higher due to SOX11 overexpression (85% of the variance, PC1) than by time of induction (8 or 24 hours) (4% of the variance, PC2) (Supplemental Figure S4A), showing that more than 65% of differential expressed genes (DEG) overlapped upon SOX11 expression between 8h and 24h of 4-OHT treatment (Supplemental Figure S4B).

SOX11-specific GEP in DG75 BL cell line, grouping two-time point samples, showed 866 upregulated and 828 downregulated genes in 4-OHT treated DG75 ER-SOX11 compared to control cells (Figure 4C-D and Supplemental Table S6). Pathway enrichment analysis showed that upregulated genes in DG75 ER-SOX11 were enriched in angiogenesis, integrins and G-protein signaling pathways, whereas downregulated genes were enriched in genes related to cadherin and Wnt signaling, among other regulatory pathways (Figure 4E).

Two more cell model systems were generated to investigate the effect of SOX11 expression on BL transcriptome: the SOX11- Ramos BL cell line ectopically overexpressing a FLAG-SOX11 protein (Ramos-SOX11+), and the SOX11+ cell line BL2 where we knocked out the SOX11 gene using the CRISPR-Cas9 gene editing system (BL2-SOX11KO) (Figure 5A). Changes in global gene expression were investigated in both cell systems by RNA-seq (Supplemental Figure S5A-B and S5C-D and Supplemental Tables S7 and S8, respectively). Pathway enrichment analysis performed in Ramos cells showed enrichment of pathways in Ramos-SOX11+ similar than the observed in DG75 ER-SOX11 cell lines (Supplemental Figure S5E)

SOX11-associated BL signature

Using RNA-seq, we overlapped the DEGs obtained between Ramos-SOX11+ and Ramos-CT (Supplemental Table S7); between BL2 CT and BL2-SOX11KO cell lines (Supplemental Table S8), and between DG75 ER-SOX11 and DG75 ER (Supplemental Table S6). 79 genes commonly regulated by SOX11 in at least two different BL cell lines were considered to define a SOX11-associated BL signature (Figure 5B and Supplemental Table S9).

We performed k-means clustering analysis in the previously published 117 pediatric BL primary cases from the BLGSP[21] with RNA-seq data available, using the SOX11-associated BL signature. We observed that BL cases clustered separately according to SOX11 high and low expression levels (Figure 5C). Moreover, most of the genes included in the SOX11-associated BL signature significantly correlated between them and with SOX11 expression in this BL series (Figure 5D). Finally, by GSEA we observed that the SOX11-associated BL signature (Supplemental Table S9) was enriched in $SOX11^{high}$ compared to $SOX11^{low}$ BL cases[21] (SOX11 mRNA cut-off=10.5 log_2 transformed values) (Supplemental Figure S6). These results suggest a similar SOX11 transcriptional activity in cell lines and primary BLs.

SOX11 functional role in MCL and BL

SOX11 directly regulates the transcription of genes involved in MCL oncogenic pathways.[27] To determine whether SOX11 regulated genes in MCL were also modulated in BL, we investigated by GSEA the expression of SOX11-target genes identified in our previously studies in MCL cell lines[28] and primary samples[34] (Supplemental Table S10), in SOX11+ and SOX11- BL cell lines (DG75 ER-SOX11 vs. DG75 ER; Ramos SOX11+ vs Ramos CT; BL2CT vs BL2-SOX11KO) (Figure 6A). On the other hand, the SOX11-associated BL signature was significantly enriched in SOX11+ compared to SOX11- MCL cell lines and primary samples (Supplemental Figures S7A-B, respectively). Together, these results demonstrate that SOX11 regulates common genes in the two lymphoma entities.

Moreover, we observed that 11% (185/1660) of the DEG in DG75 ER-SOX11 BL cell line overlapped with those differentially expressed in Z138 SOX11+ vs Z138-SOX11KO MCL cell line. An 8% (134/1660) of the DEG in DG75 ER-SOX11 BL cell line overlapped with SOX11-direct target genes in Z138 MCL cell line. Furthermore, 22 genes overlapped between the three comparisons. Together, these 297 genes (Figure 6B, red circle) were involved in oxidative stress, heterotrimeric G-proteins, chemokines, and cytokines, integrins, angiogenesis and PDGF signaling pathways (Figure 6C). We validated the upregulation of some of the overlapped genes, specifically the mRNA of PLXNB1, MEX3A and CD24, and the protein levels of MEX3A and CD24 upon SOX11 overexpression and knockout, in both MCL and BL transduced cell lines (Figure 6D-F). The upregulated mRNA levels of PLXNB1, CD24 and MEX3A significantly decreased, reaching similar levels as in DG75 ER control, in DG75 ER-SOX11 cells upon 4-OHT washout (Supplemental Figure S8).

In BL cells, we observed that CXCR5, CCR7 and ITGB7 were significantly upregulated in 4-OHT treated DG75 ER-SOX11 compared to DG75 ER control cells (Figure 6G). However, contrary to MCL, we did not observe a significant higher tumor cell migration towards CXCL13 or adhesion to SNKT stromal cells,[32] nor an increase in the activation of BCR signaling pathway[30] comparing SOX11+ and SOX11- BL cell line models (data not shown). Interestingly, we observed a significantly higher adhesion of SOX11+ to VCAM-1, glycoprotein that interacts with integrin α4β7 (ITGA4 and ITGB7), compared to SOX11- BL cells (Figure 6H).

DISCUSSION

EBV infection is considered a crucial and early event in BL development, particularly in eBL. EBV- and EBV+ BLs exhibit distinct molecular characteristics whereas the clinical features of both groups are variable.[13–18,21] SOX11 is expressed in a broad range of BLs,[22,23] with a higher frequency in pediatric patients.[7] Several studies have described oncogenic functions of SOX11 in MCL.[28–34] However, the functional role of SOX11 in BL and its relation to EBV remains unknown.

Our results have revealed a clear negative association between SOX11 expression and EBV infection. The absence of SOX11 in EBV+ BL seems associated with the presence of the virus rather than the historical epidemiologic subtype as we also observed high SOX11 expression in the few EBV- BL from Africa.

The *IG::MYC* translocation is considered the genetic hallmark of BL. EBV+ BLs mainly acquire the translocation during SHM, whereas the EBV- BLs acquire it through CSR.[16,18,20] Here we add another layer of information. Among EBV- BL, we found differences in IG partners according to SOX11 expression. We observed significant differences between IG partner in *MYC* translocations among the three different groups analyzed, detecting lower proportion of IGH and higher of IGL in EBV-/SOX11+ cases. EBV-/SOX11+ BLs acquired the *IG*::*MYC* translocation during the CSR process. These data suggest that among EBV- BL SOX11-expression is associated with one of the earliest events in the pathogenesis of the tumor, the occurrence of the *MYC* translocation.

Recently, Roco et al suggested that CSR occurs outside the germinal center earlier in the B-cell differentiation process than SHM.[41] One might speculate, that in EBV-/SOX11+ BL predominant CSR-mediated *IG*::*MYC* translocations reflect this situation arising before the cell enters the germinal center. SOX11 represses *BCL6* and

AICDA expression in conventional MCL possibly preventing the entrance of the tumor cell in the germinal center.[29] However, no significant differences in AICDA mRNA levels were observed between EBV-/SOX11- and EBV-/SOX11+ BLs.

BL is the lymphoma entity second to MCL with most SOX11+ cases. Nevertheless, the expression levels in BL are approximately two times lower than in MCL cases,[23] suggesting that its levels in SOX11+ BL are not sufficient to fully block the entrance into the germinal center. However, to confirm this idea needs further functional studies.

EBV+ BL cells have fewer driver gene mutations than EBV-, indicating that EBV infection may relieve the pressure towards selection of mutagenic mechanism.[7,15,16,19–21] Differences in the genetic profile between EBV+ and EBV- BL cases have been previously described,[14-17,21] however we have observed that EBV-/SOX11+ BLs display a distinctive mutational landscape, with significant higher frequency of mutations in *SMARCA4* and *ID3*, compared to EBV-/SOX11- and EBV+ cases. Interestingly, concomitant *SMARCA4* mutations and SOX11 expression has been also observed in MCL cases.[42] In addition, EBV+ BLs exhibited significant higher frequency of mutations in *DDX3X* and *SIN3A*, and fewer in *MYC* compared to SOX11+ but not to SOX11- BLs. These findings suggest that EBV+, EBV-/SOX11- and EBV-/SOX11+ cases might have different oncogenic mechanisms driving their pathogenesis.

By the overlap of SOX11-mediated DEG in three different BL cell lines, we obtained a SOX11-associated BL signature that consistently grouped separately SOX11+ and SOX11- BL cases. Furthermore, SOX11 overexpression in BL recapitulated in part the SOX11-associated transcriptional program found in MCL cells, overlapping with some of the validated pathways directly regulated by SOX11 in MCL[27] as the overexpression of PLXNB1 and CD24 involved in tumor cell

migration,[43,44] and MEX3A involved in the chemoresistance of colorectal cancer quiescent cells.[45] However, *in vitro* experiments suggest that the activation of tumor cell migration[32] or BCR pathways[30] observed in MCL may not be so relevant in BL.

ITGB7 integrin controls the cell homing to Peyer's patches through the binding to VCAM-1.[46,47] We observed a significant upregulation of *ITGB7* gene expression and significantly higher adhesion to VCAM-1 in SOX11+ compared to SOX11- BL cell lines. sBL, the clinical subtype that contains more SOX11+ cases, shows predominantly abdominal tumor presentation and class switch to *IgA* isotype,[16,48] specifically seen in mucosal tissues,[49] such as Peyer's patches, all suggesting that BL SOX11+ cells might have a higher migration to Peyer's patches, through the upregulation of ITGB7, than SOX11- BL cells. However, further studies are needed to validate this last hypothesis.

In conclusion, SOX11 expression and EBV infection occur in alternative subsets of BL with different profile of somatic mutations and different mechanism generating *MYC* translocations. The predominance of IGH class switch mediated *MYC* translocation in SOX11+ BL suggests an earlier development than in SOX11- tumors. Further studies are required to define the functional role of SOX11 in BL.

ACKNOWLEDGEMENTS

This work was supported by the Spanish Ministerio de Ciencia e Innovación (MICINN) (Grant No. PID2021-124048OB-100 to VA), co-financed by the Sub-Directorate General for Evaluation and the European Regional Development Fund. Fondo Europeo de Desarrollo Regional (FEDER-"Una manera de hacer Europa"); Generalitat de Catalunya Suport Grups de Recerca (AGAUR-Senior Research Group; 2021-SGR-0129 to VA); Fundació La Marató de TV3 (201901-30 to VA). MSG was supported by the Agencia de Gestió d'ajuts universitaris i de recerca (AGAUR; investigador novell-FI2018; RH041713). PB was supported by the "Asociación Española contra el cancer" (AECC; Investigadores 2018 fellowship; INVES18042AMAD). ADB was supported by the Agencia de Gestió d'ajuts universitaris i de recerca (AGAUR-PERIS Salut; SLT017/20/000161). MC was supported by the Agencia de Gestió d'ajuts universitaris i de recerca (AGAUR-Investigo2022; 100028TC4), financed by the European Union, Next Generation EU. EC is an Academia Researcher of the "Institució Catalana de Recerca i Estudis Avançats" (ICREA) of the Generalitat de Catalunya. CERCA Programme/Generalitat de Catalunya. CL is supported by a postdoctoral Beatriu de Pinós grant from Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya and by Marie Sklodowska-Curie COFUND program from H2020 (2018-BP-00055). The analyses of RS have been supported by the DFG (SFB1074 project B9). The ICGC MMML-Seq consortium has been supported by the German Ministry of Science and Education in the framework of the ICGC MMML-Seq consortium (01KU1002) and ICGC DE-Mining (01KU1505). The work was additionally supported by the Kinderkrebsinitiative Buchholz, Holm-Seppensen (KKI) continuous reaserch

support to II, JR and WK. II was suported by the Christoph-Schuber-Programm of the KKI.

AUTHORSHIP CONTRIBUTIONS:

MSG designed, performed and interpreted all in vitro experiments, bioinformatic and statistical analyses, and wrote the manuscript; II designed and interpreted experiments, performed bioinformatic and statistical analyses; JR, MM, CL, LL, WK and RS provided and analyzed molecular and clinical data of BL patients; PB, MLR, ADB, MC and MM performed in vitro experiments; MCS, ST, MB and RB performed ddPCR, IHC and RNAscope; FN, GC, PJ and SG performed bioinformatic and statistical analyses; EC helped to interpret data; WK and VA designed and supervised experiments, analyzed data and wrote the manuscript; and all authors discussed the results and commented on the manuscript.

CONFLICT OF INTEREST DISCLOSURES: The authors declare no conflict of interest. The institution of Dr. WK has received research funding from Roche, Janssen, Amgen, Takeda, Regeneron and InCyte, not related to the submitted work.

DATA AVAILABILITY

The RNA-seq data reported in this paper will be available in the Gene Expression Omnibus (GEO) at the time of manuscript publication.

REFERENCES

1. Swerdlow SH, Campo E, Harris NL, et al. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, 4th edition. Lyon, France: IARC Press; 2017.

2. Burkhardt B, Oschlies I, Klapper W, et al. Non-Hodgkin's lymphoma in adolescents: experiences in 378 adolescent NHL patients treated according to pediatric NHL-BFM protocols. *Leukemia*. 2011;25(1):153–160.

3. Miles RR, Arnold S, Cairo MS. Risk factors and treatment of childhood and adolescent Burkitt lymphoma/leukaemia. *Br. J. Haematol.* 2012;156(6):730–743.

4. Burkitt DP. Classics in oncology. A sarcoma involving the jaws in African children. *CA. Cancer J. Clin.* 1972;22(6):345–355.

5. Gabarre J, Raphael M, Lepage E, et al. Human immunodeficiency virus-related lymphoma: relation between clinical features and histologic subtypes. *Am. J. Med.* 2001;111(9):704–711.

6. Satou A, Asano N, Nakazawa A, et al. Epstein-barr virus (EBV)-positive sporadic burkitt lymphoma: An age-related lymphoproliferative disorder? *Am. J. Surg. Pathol.* 2015;39(2):227–235.

7. Richter J, John K, Staiger AM, et al. Epstein–Barr virus status of sporadic Burkitt lymphoma is associated with patient age and mutational features. *Br. J. Haematol.* 2022;196(3):681–689.

8. Zech L, Haglund U, Nilsson K, Klein G. Characteristic chromosomal abnormalities in biopsies and lymphoid-cell lines from patients with Burkitt and non-Burkitt lymphomas. *Int. J. Cancer*. 1976;17(1):47–56.

9. Boerma EG, Siebert R, Kluin PM, Baudis M. Translocations involving 8q24 in Burkitt lymphoma and other malignant lymphomas: a historical review of cytogenetics in the light of todays knowledge. *Leukemia*. 2009;23(2):225–234.

10. Dalla-Favera R, Lombardi L, Pelicci PG, et al. Mechanism of activation and biological role of the c-myc oncogene in B-cell lymphomagenesis. *Ann. N. Y. Acad. Sci.* 1987;511(1):207–218.

11. Schmitz R, Ceribelli M, Pittaluga S, Wright G, Staudt LM. Oncogenic Mechanisms in Burkitt Lymphoma. *Cold Spring Harb. Perspect. Med.* 2014;4(2):.

12. Sander S, Calado DP, Srinivasan L, et al. Synergy between PI3K signaling and MYC in Burkitt lymphomagenesis. *Cancer Cell*. 2012;22(2):167–179.

13. Schmitz R, Young RM, Ceribelli M, et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature*. 2012;490(7418):116–120.

14. Love C, Sun Z, Jima D, et al. The genetic landscape of mutations in Burkitt lymphoma. *Nat. Genet.* 2012;44(12):1321–1325.

15. Kaymaz Y, Oduor CI, Yu H, et al. Comprehensive Transcriptome and Mutational Profiling of Endemic Burkitt Lymphoma Reveals EBV Type-Specific Differences. *Mol. Cancer Res.* 2017;15(5):563–576.

16. López C, Kleinheinz K, Aukema SM, et al. Genomic and transcriptomic changes complement each other in the pathogenesis of sporadic Burkitt lymphoma. *Nat. Commun.* 2019;10(1):1459.

17. Panea RI, Love CL, Shingleton JR, et al. The whole-genome landscape of Burkitt lymphoma subtypes. *Blood*. 2019;134(19):1598–1607.

18. Bellan C, Lazzi S, Hummel M, et al. Immunoglobulin gene analysis reveals 2 distinct cells of origin for EBV-positive and EBV-negative Burkitt lymphomas. *Blood*. 2005;106(3):1031–1036.

19. Abate F, Ambrosio MR, Mundo L, et al. Distinct Viral and Mutational Spectrum of Endemic Burkitt Lymphoma. *PLOS Pathog.* 2015;11(10):e1005158.

20. Thomas N, Dreval K, Gerhard DS, et al. Genetic subgroups inform on pathobiology in adult and pediatric Burkitt lymphoma. *Blood*. 2023;141(8):904–916.

21. Grande BM, Gerhard DS, Jiang A, et al. Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. *Blood*. 2019;133(12):1313–1324.

22. Dictor M, Ek S, Sundberg M, et al. Strong lymphoid nuclear expression of SOX11 transcription factor defines lymphoblastic neoplasms, mantle cell lymphoma and Burkitt's lymphoma. *Haematologica*. 2009;94(11):1563–1568.

23. Mozos A, Royo C, Hartmann E, et al. SOX11 expression is highly specific for mantle cell lymphoma and identifies the cyclin D1-negative subtype. *Haematologica*. 2009;94(11):1555–1562.

24. Wästerlid T, Nordström L, Freiburghaus C, et al. Frequency and clinical implications of SOX11 expression in Burkitt lymphoma. *Leuk. Lymphoma*. 2017;58(7):1760–1763.

25. Deffenbacher KE, Iqbal J, Sanger W, et al. Molecular distinctions between pediatric and adult mature B-cell non-Hodgkin lymphomas identified through genomic profiling. *Blood*. 2012;119(16):3757–3766.

26. Hummel M, Bentink S, Berger H, et al. A Biologic Definition of Burkitt's

Lymphoma from Transcriptional and Genomic Profiling. *N. Engl. J. Med.* 2006;354(23):2419–2430.

27. Beekman R, Amador V, Campo E. SOX11, a key oncogenic factor in mantle cell lymphoma. *Curr. Opin. Hematol.* 2018;25(4):299–306.

28. Vegliante MC, Palomero J, Pérez-Galán P, et al. SOX11 regulates PAX5 expression and blocks terminal B-cell differentiation in aggressive mantle cell lymphoma. *Blood*. 2013;121(12):2175–2185.

29. Palomero J, Vegliante MC, Eguileor A, et al. SOX11 defines two different subtypes of mantle cell lymphoma through transcriptional regulation of BCL6. *Leukemia*. 2016;30(7):.

30. Kuo PY, Jatiani SS, Rahman AH, et al. SOX11 augments BCR signaling to drive MCL-like tumor development. *Blood*. 2018;131(20):2247–2255.

31. Palomero J, Vegliante MC, Rodríguez ML, et al. SOX11 promotes tumor angiogenesis through transcriptional regulation of PDGFA in mantle cell lymphoma. *Blood*. 2014;124(14):2235–2247.

32. Balsas P, Palomero J, Eguileor Á, et al. SOX11 promotes tumor protective microenvironment interactions through CXCR4 and FAK regulation in mantle cell lymphoma. *Blood*. 2017;130(4):501–513.

33. Balsas P, Veloza L, Clot G, et al. SOX11, CD70, and Treg cells configure the tumor immune microenvironment of aggressive mantle cell lymphoma. *Blood*. 2021;138(22):2202–2215.

34. Sureda-Gómez M, Balsas P, Rodríguez M-L, et al. Tumorigenic role of Musashi-2 in aggressive mantle cell lymphoma. *Leukemia*. 2023;37(2):.

35. Burkhardt B, Michgehl U, Rohde J, et al. Clinical relevance of molecular characteristics in Burkitt lymphoma differs according to age. *Nat. Commun.* 2022;13(1):3881.

36. Aukema SM, Hoster E, Rosenwald A, et al. Expression of TP53 is associated with the outcome of MCL independent of MIPI and Ki-67 in trials of the European MCL Network. *Blood*. 2018;131(4):417–420.

37. Croci GA, Hoster E, Beà S, et al. Reproducibility of histologic prognostic parameters for mantle cell lymphoma: cytology, Ki67, p53 and SOX11. *Virchows Arch.* 2020;477(2):259–267.

38. Mundo L, Del Porro L, Granai M, et al. Frequent traces of EBV infection in Hodgkin and non-Hodgkin lymphomas classified as EBV-negative by routine methods: expanding the landscape of EBV-related lymphomas. *Mod. Pathol.* 2020;33(12):2407–2421.

39. Siciliano MC, Tornambè S, Cevenini G, et al. EBV persistence in gastric cancer cases conventionally classified as EBER-ISH negative. *Infect. Agent. Cancer*. 2022;17(1):57.

40. Afgan E, Baker D, Batut B, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 2018;46(W1):W537–W544.

41. Roco JA, Mesin L, Binder SC, et al. Class-Switch Recombination Occurs Infrequently in Germinal Centers. *Immunity*. 2019;51:337–350.

42. Nadeu F, Martin-Garcia D, Clot G, et al. Genomic and epigenomic insights into the origin, pathogenesis, and clinical behavior of mantle cell lymphoma subtypes. *Blood*. 2020;136(12):1419–1432.

43. Ye S, Hao X, Zhou T, et al. Plexin-B1 silencing inhibits ovarian cancer cell migration and invasion. *BMC Cancer*. 2010;10:611.

44. Altevogt P, Sammar M, Hüser L, Kristiansen G. Novel insights into the function of CD24: A driving force in cancer. *Int. J. Cancer*. 2021;148(3):546–559.

45. Álvarez-Varela A, Novellasdemunt L, Barriga FM, et al. Mex3a marks drug-tolerant persister colorectal cancer cells that mediate relapse after chemotherapy. *Nat. cancer*. 2022;3(9):1052–1070.

46. Ruegg C, Postigo AA, Sikorski EE, et al. Role of integrin alpha 4 beta 7/alpha 4 beta P in lymphocyte adherence to fibronectin and VCAM-1 and in homotypic cell clustering. *J. Cell Biol.* 1992;117(1):179–189.

47. Berlin C, Berg EL, Briskin MJ, et al. α4β7 integrin mediates lymphocyte binding to the mucosal vascular addressin MAdCAM-1. *Cell*. 1993;74(1):185–195.

48. López C, Burkhardt B, Chan JKC, et al. Burkitt lymphoma. *Nat. Rev. Dis. Prim.* 2022;8(1):78.

49. Cerutti A, Rescigno M. The Biology of Intestinal Immunoglobulin A Responses. *Immunity*. 2008;28(6):740–750.

TABLES

Table 1. Detection of EBV genome traces obtained by ddPCR of BamHI-W and EBNA1, and RNAscope of EBNA1, in a cohort of 37 BL specimens.

Study ID	EBER ISH	SOX11 IHC	BamHI-W (copies/ul)	EBNA1 (copies/ul)	EBNA1 mRNA (score)
BL-42	EBV-	SOX11-	1.01	0	NA
BL-53	EBV-	SOX11-	6.3	1.7	9
BL-48	EBV-	SOX11-	0	0	NA
BL-54	EBV-	SOX11-	0.19	0	NA
BL-3	EBV-	SOX11-	15	2.6	9
BL-34	EBV-	SOX11-	0	0	NA
BL-17	EBV-	SOX11-	15	1.42	7
BL-1	EBV-	SOX11-	0	0	NA
BL-29	EBV-	SOX11-	18	1.25	6
BL-10	EBV-	SOX11-	0.2	0.2	NA
BL-7	EBV-	SOX11-	0	0	NA
BL-31	EBV-	SOX11-	0	0	NA
BL-56	EBV-	SOX11-	0	0	NA
BL-51	EBV-	SOX11-	0.31	0	NA
BL-46	EBV-	SOX11-	0	0	NA
BL-43	EBV-	SOX11-	0.62	0.25	6
BL-15	EBV-	SOX11-	0	0	NA
BL-89	EBV-	SOX11-	0	0	NA
BL-2	EBV-	SOX11-	0	0	NA
BL-18	EBV-	SOX11-	0	0	NA
BL-113	EBV-	SOX11-	1.31	0.5	NA
BL-119	EBV-	SOX11-	0	0	NA
BL-38	EBV-	SOX11-	0	0	NA
BL-143	EBV-	SOX11+	0	0	NA
BL-144	EBV-	SOX11+	0	0	NA
BL-41	EBV-	SOX11+	0	0	NA
BL-105	EBV-	SOX11+	0.66	0.19	NA
BL-25	EBV-	SOX11+	0	0	NA
BL-12	EBV-	SOX11+	15.9	2.7	11
BL-32	EBV-	SOX11+	0	0	NA
BL-50	EBV-	SOX11+	0	0	NA
BL-116	EBV-	SOX11+	0	0	NA
BL-72	EBV-	SOX11+	0	0	NA

BL-145	EBV-	SOX11+	0	0	NA
BL-146	EBV-	SOX11+	0	0	NA
BL-147	EBV-	SOX11+	0	0	NA
BL-148	EBV-	SOX11+	0	0	NA

FIGURE LEGEND

Figure 1. SOX11 is exclusively expressed in EBV- BL cases. (A-B) SOX11 mRNA expression (log_2 transformed values) according to EBV status (EBV- and EBV+) **(A)** and epidemiological variant (eBL and sBL) **(B)** of 117 pediatric BL cases. Red circle highlights high SOX11 expression in 3 BL from endemic areas. Wilcoxon text was performed to test differences between groups. **(C)** Frequency of SOX11+ and SOX11- patients (by IHC) in an independent series of pediatric and adult sBLs (n=189), according to EBV status. The Cochran-Mantel-Haenszel test was performed to test differences between the frequency in EBV- and in EBV+, while adjusting for cohort and group.

Figure 2. *IG::MYC* translocation in BL primary cases according to SOX11 expression levels and EBV status. (A-B) Frequency of *IGH::MYC, IGK::MYC* and *IGL::MYC* translocations **(A)**, and frequency of translocations generated by CSR, SHM or V(D)J processes **(B)**, in the total group of EBV+, EBV-/SOX11- and EBV-/SOX11+ BLs. Fisher test was performed to evaluate differences between group frequencies.

Figure 3. Mutational profile of BL primary cases according to EBV status and SOX11 expression. (A) Mutational analysis in recurrently mutated driver genes

(mutated in ≥10% of cases) of 267 BL. Mutations, EBV and SOX11 status, age group (pediatric or adult) and epidemiological variant (endemic or sporadic) are shown. The cases in the heatmap are ordered by SOX11 and EBV status. **(B)** Frequencies of mutated cases in EBV+ (n=110), EBV-/SOX11- (n=76) and EBV-/SOX11+ (n=81) for each gene are shown. Fisher's exact test with FDR correction was performed to evaluate differences in the frequencies between groups. **** q-value <0.0001, *** q-value <0.001, ** q-value <0.01, * q-value <0.15.

Figure 4. Gene expression analysis upon SOX11 overexpression in DG75-ER-SOX11 BL cell line. (A) Western blot experiment showing the levels of ER-SOX11 protein in DG75 ER-SOX11 BL cell line. DG75-ER was used as negative SOX11 expressing cell line and tubulin as loading control. **(B)** Immunofluorescence experiments showing the nuclear localization of the SOX11 protein in DG75 ER-SOX11 cells, induced (+) or not induced (-) with 4-OHT for 24h. DG75 ER cell line was used as SOX11 negative control. DAPI mark the cellular nucleus, and merge of the two immunofluorescences images (DAPI and SOX11) was done. **(C)** Heatmap illustrating the scaled expression (Z-score) of 1694 DEG (866 upregulated and 828 downregulated genes; Supplementary Table S7) in DG75 ER-SOX11 compared to DG75 ER cell lines induced with 4-OHT for 8 and 24h, obtained by RNA-seq. Genes with an adjusted P-value <0.1 and absolute log2-transformed fold change >0.65 were considered. **(D)** Volcano plot showing genes differentially expressed, obtained by RNA-seq, upon SOX11 overexpression in DG75 ER-SOX11 compared to DG75 ER BL cell lines treated with 4-OHT. The graph shows on the y-axis –log10(P-value) and on the x-axis the log2-transformed fold change. Genes upregulated and downregulated in DG75 ER-SOX11 vs DG75 ER with an adjusted P-value <0.1 and log_2-transformed fold

change >0.65 or <-0.65 are colored in red and blue, respectively, and genes with an adjusted P-value < 0.00005 and absolute log2-transformed fold change >3 are labeled with their Gene Symbol. **(E-F)** Panther pathway enrichment analysis using DEG upregulated **(E)** and downregulated **(F)** between DG75 ER-SOX11 and DG75 ER after 4-OHT treatment. Number of genes, fold enrichment and –log10(P-value) for each pathway are shown. Only pathways with a p-value <0.05 were considered.

Figure 5. SOX11-associated BL signature found in transduced cell lines is also detected in BL primary cases. (A) Western blot experiment showing the SOX11 protein levels in Ramos SOX11 (SOX11 is FLAG-tagged) cell line, Ramos CT, BL2 CT and BL2-SOX11KO BL cell lines. Tubulin protein was used as **l**oading control. **(B)** Overlap between DEG (adjusted P-value <0.1 and absolute log_2 transformed fold change >0.65) in DG75 ER-SOX11 vs DG75 ER (in purple, 1694 genes, Supplemental Table S6), BL2 CT vs BL2-SOX11KO (in red, 107 genes, Supplemental Table S8) and Ramos SOX11+ vs Ramos CT (in green, 151 genes, Supplemental Table S7). **(C)** Row scaled expression (Z-score) of genes from the SOX11-associated BL signature (79 DEG overlapping between at least two comparisons in Figure 5B) in RNA-seq data of 117 pediatric endemic and sporadic BLs. K-means clustering was performed to separate samples in k=3 groups. *SOX11* expression, EBV status and BL epidemiological variant are shown at the top panel. **(D)** Correlation plot between genes from the SOX11-associated BL signature in RNA-seq data of 117 pediatric endemic and sporadic BLs. Blue and red showed positive and negative Pearson correlation coefficients, respectively. P-values from Pearson correlation are shown: * <0.05, ** <0.01, *** <0.001.

Figure 6. **SOX11 shares similar transcriptional roles in MCL and BL. (A)** Dot plot showing at x-axis the normalized enrichment score (NES), and at y-axis SOX11-target genes identify in our previously studies in MCL cell lines by ChIP-chip (SOX11-direct targets in MCL), upregulated (SOX11 MCL signature-UP) or downregulated (SOX11 MCL signature-Down) in SOX11+ compared to SOX11- MCL cell line (Z138CT vs Z138SOX11KO) and primary samples (cMCL vs nnMCL)), enriched in SOX11+ compared to SOX11- BL cell lines (DG75 RE-SOX11 vs. DG75 ER; Ramos SOX11+ vs Ramos CT; BL2CT vs BL2-SOX11KO). Outlined in colour codes the false discovery rate (FDR). Size of bubbles represent number of enriched genes. **(B)** Overlap between DEG in 4-OHT treated DG75 ER-SOX11 BL cell line (in yellow, 1660 genes with Gene Symbol), upon SOX11 KO in Z138 MCL cell line (in green, 686 genes with Gene Symbol), and SOX11-direct target genes in MCL found by ChIP on chip in Z138 cell line (in blue, 1909 genes with Gene Symbol). **(C)** Pathway enrichment analysis on common genes between DEG in DG75 ER-SOX11 BL cells, and upon SOX11 KO in Z138 MCL cells and SOX11 targets genes obtained by ChIP on chip in Z138 MCL cells (red circle). Number of genes, fold enrichment and –log10(p-value) for each pathway are shown. **(D)** PLXNB1, CD24 and MEX3A relative mRNA expression (normalized to GUSB endogenous control) in BL and MCL SOX11-overexpressing cell lines (left, Ramos-SOX11+, DG75 ER-SOX11 and JVM2-SOX11+), and in SOX11-KO BL and MCL cell lines (right, BL2-SOX11KO and Z138-SOX11KO). Data is shown as mean ± standard deviation of the fold change between SOX11-overexpressing or SOX11-KO and its respective control cell values, obtained in 3 independent experiments. Statistical significance was obtained by unpaired two-tailed Student t-test. **(E)** Western blot experiments showing MEX3A and SOX11 protein levels (ER-SOX11, SOX11-FLAG or endogenous SOX11) in BL2-SOX11KO, DG75 ER-SOX11, and Ramos-SOX11+

and their control cell lines BL2 CT, DG75 ER and Ramos CT. Tubulin was used as a loading control. **(F)** Histograms showing CD24 protein levels analyzed by flow cytometry in Z138- and BL2-SOX11KO MCL and BL cell line models and their respective controls. CD24 staining is shown in filled dark blue histograms for SOX11+ cells and in filled light blue histograms for SOX11- cells, whereas isotype controls are shown in non-filled and long dashed histograms. **(G)** CXCR5, CCR7 and ITGB7 mRNA expression levels (log_2 transformed values) in DG75 ER and DG75 ER-SOX11, obtained by RNA-seq. **(H)** Relative adhesion to VCAM-1 measured as the ratio of fluorescence emission of calcein-labeled cells between those that have been attached to untreated microplate wells precoated with VCAM-1 and those attached in an unspecific way (VCAM-1 adhesion/unspecific cell adhesion in BSA 0.5%). Statistical significance was obtained by unpaired two-tailed Student t-test. * P-value <0.05, ** P-value < 0.01, *** P-value <0.001, **** P-value < 0.0001.

Figure 1

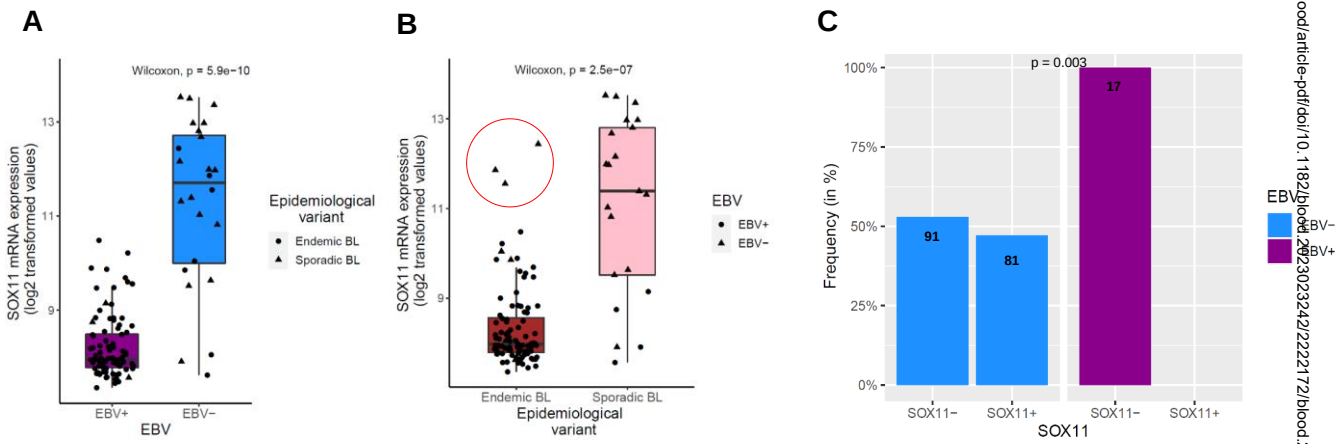


Figure 2

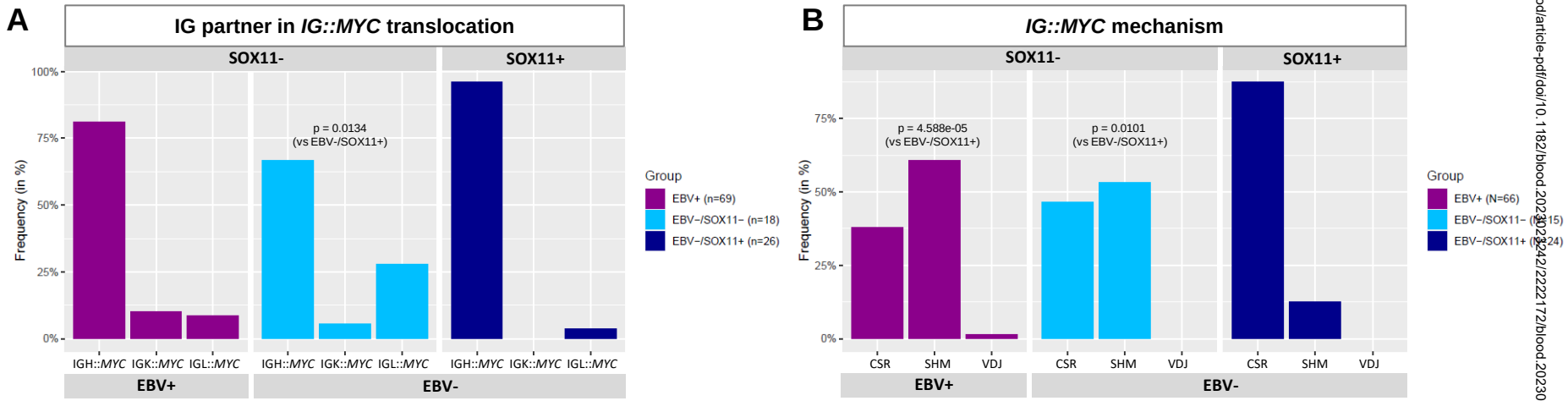


Figure 3

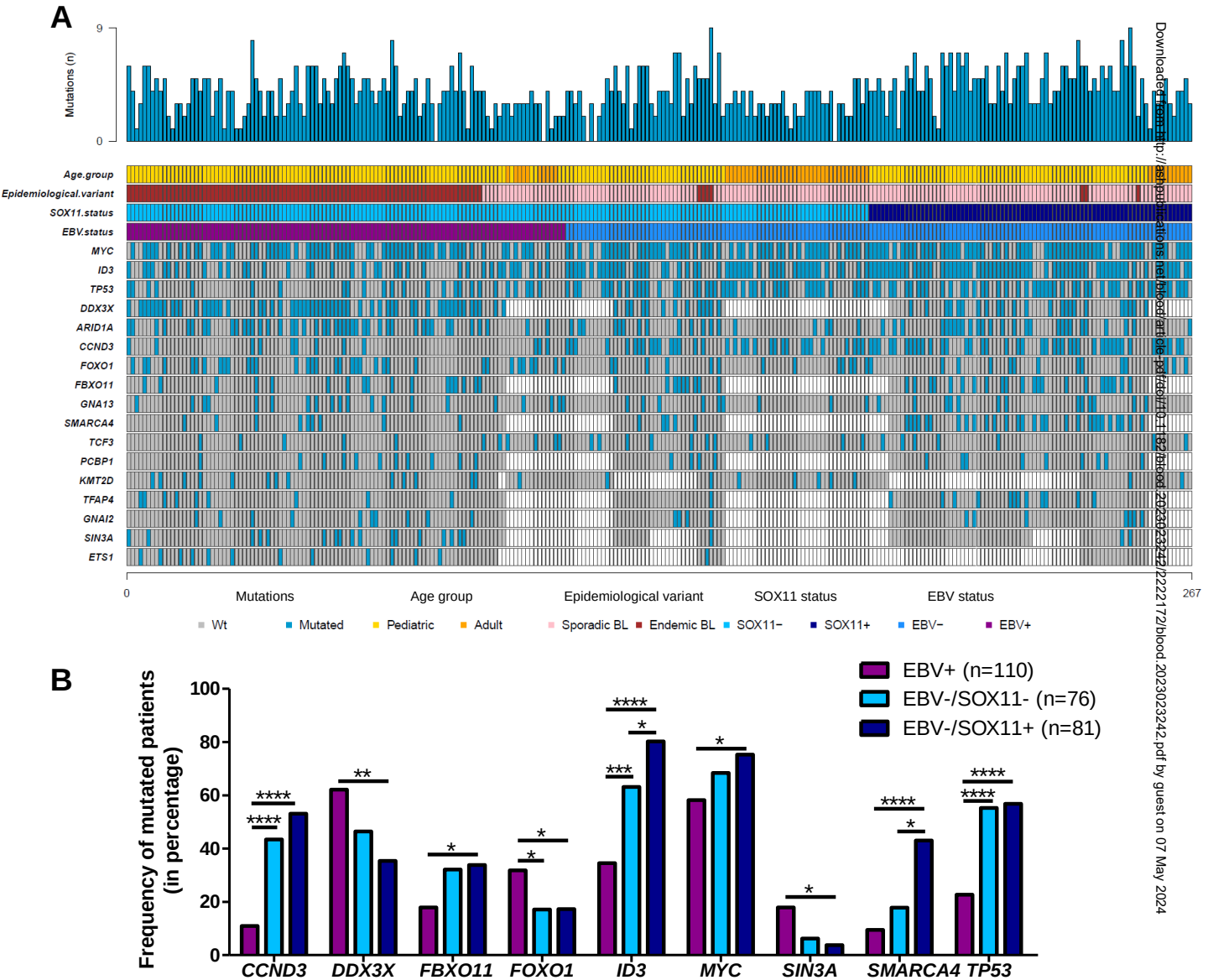


Figure 4

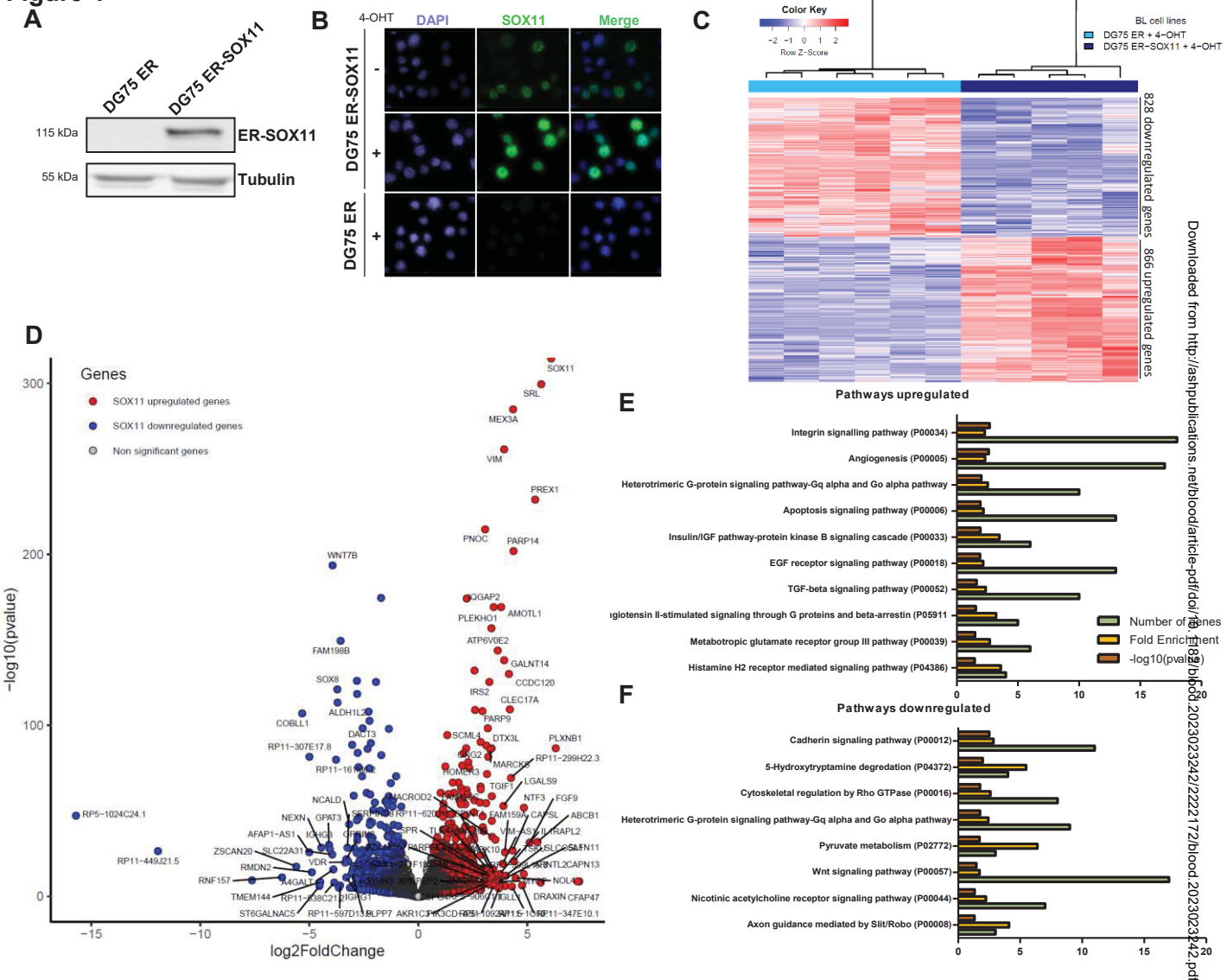


Figure 5

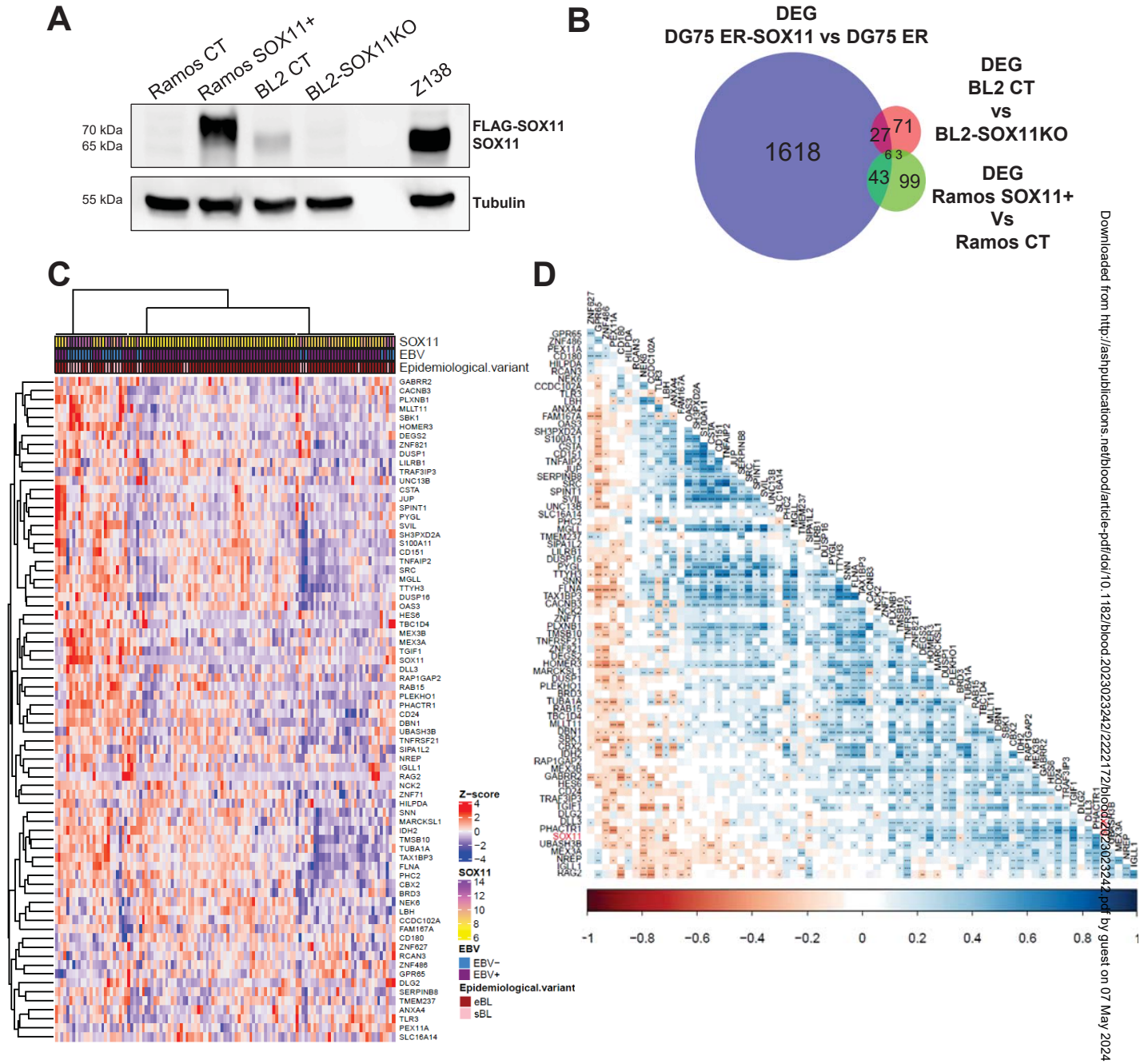
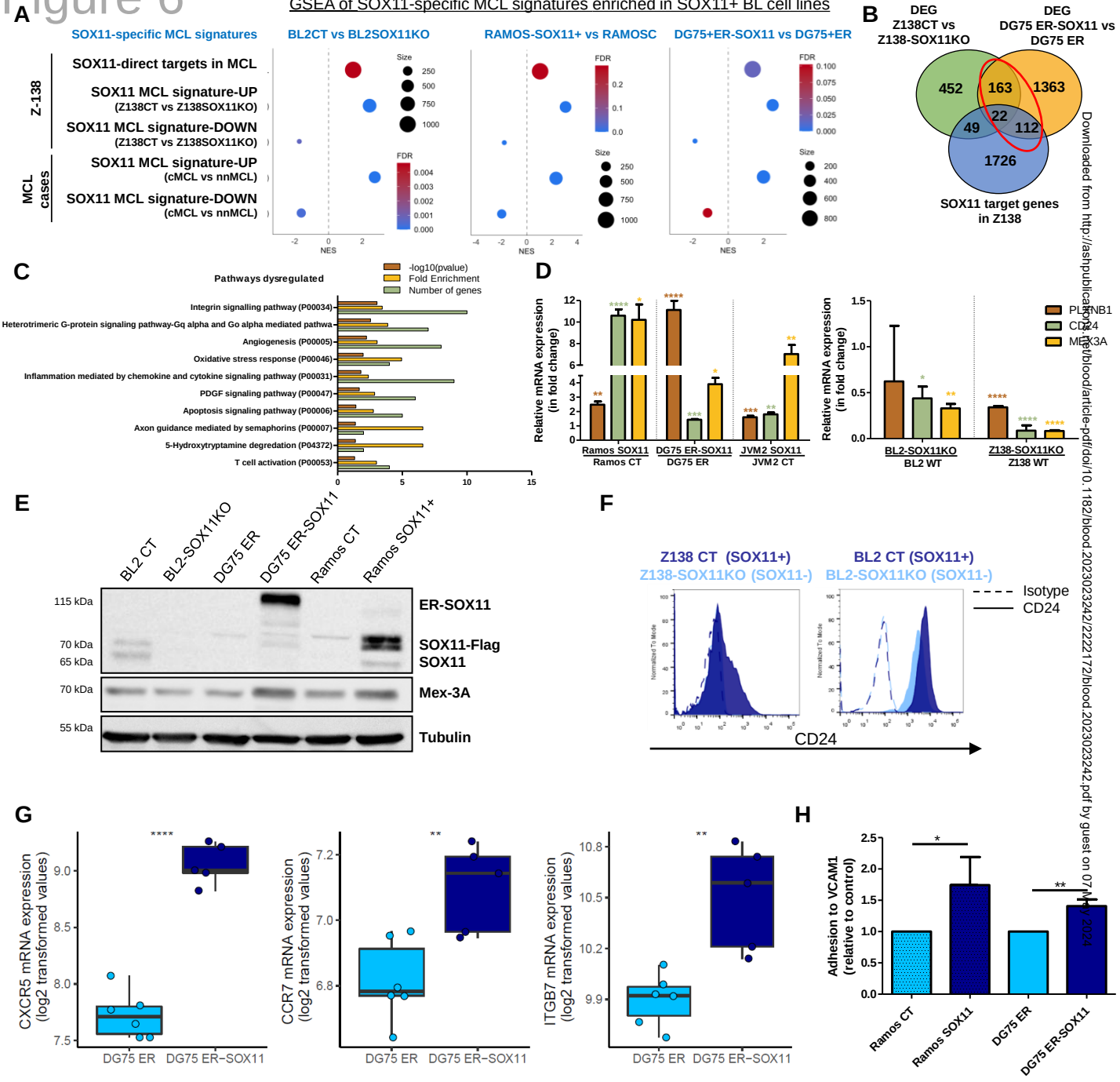
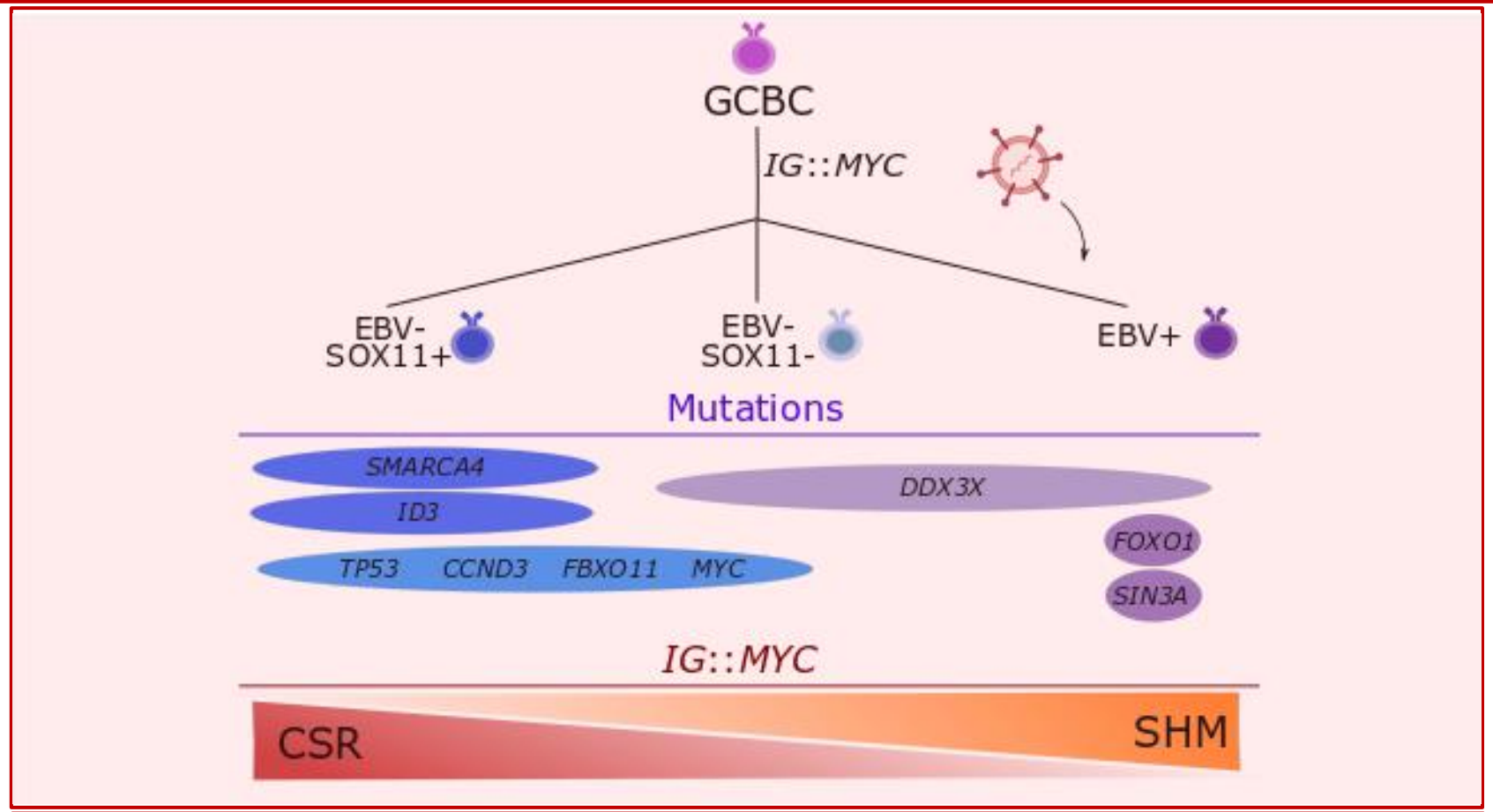


Figure 6



The molecular dichotomy between EBV and SOX11 in BL



Conclusions: SOX11 expression and EBV infection occur in alternative subsets of BL. Expression of SOX11 in EBV- BLs associates with *SMARCA4* and *ID3* mutations and CSR mechanism generating *IG::MYC* translocation.

Sureda-Gómez et al., DOI: