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-2023-023242

Marta Sureda-Gómez (IDIBAPS, Spain) Ingram Iaccarino (University Hospital Schleswig-Holstein, Campus Kiel, Germany) Mieke Meyer (University of Kiel, Germany) Patricia Balsas (IDIBAPS, Spain) Julia Richter (Christian-Albrechts University Kiel, Germany) Marta Leonor Rodríguez (IDIBAPS, Spain) Anna De Bolòs (IDIBAPS, Spain) Cristina López (Institute of Human Genetics, Ulm University and Ulm University Medical Center, 89091, Ulm, Germany., Germany) Maria Carreras-Caballé (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) Maria Chiara Siciliano (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 showed 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:

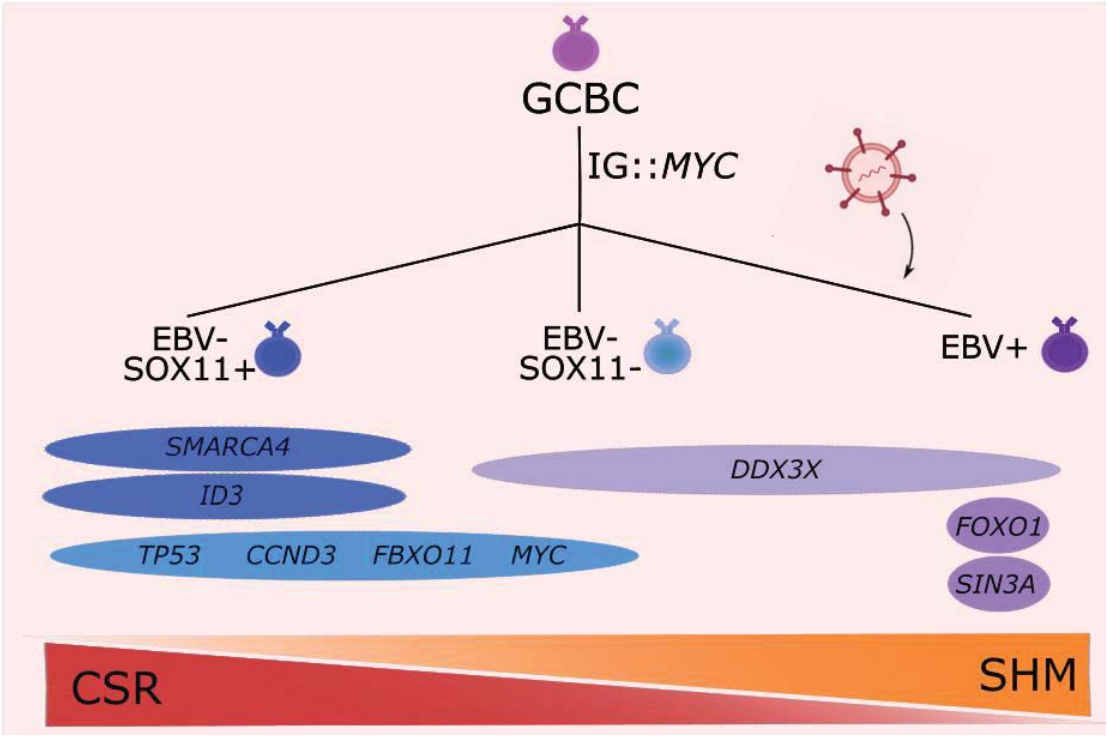
Preprint server: No;

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 and RB performed ddPCR, IHC and RNAscope; FN, GC 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: DEPOSITED RNA-SEQ DATA The RNA-seq data reported in this paper will be available in the Gene Expression Omnibus (GEO) at the time of manuscript publication.

Clinical trial registration information (if any):


GCBC
IG::MYC
EBV-
SOX11+
EBV-
SOX11-
EBV+
SMARCA4
ID3
TP53
CCND3
FBXO11
MYC
DDX3X
FOXO1
SIN3A
CSR
SHM

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

Marta Sureda-Gómez[1]*, Ingram Iaccarino[2]*, Mieke Meyer[2], Patricia Balsas[1], Julia Richter[2], Marta-Leonor Rodríguez[1], Anna de Bolòs[1], Cristina López[1,3,4], Maria Carreras[1], Selina Glaser[4], Ferran Nadeu[1,3], Maria Chiara Siciliano[5], Salvatore Tornambè[5], Roberto Boccacci[5], Guillem Clot[1,3,6], Lorenzo Leoncini[5], Elias Campo[1,3,7], 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 Medical Biotechnologies, Section of Pathological Anatomy, University of Siena, Italy.

6. University of Barcelona, Barcelona, Spain.

7. Department of Hematology Hospital Clinic of Barcelona, University of Barcelona, Barcelona, Spain.

* Contributed equally to this work.
† Contributed equally to this work and share senior authorship.

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).

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: 3924
Abstract word count: 213
Number of tables and figures: 7
Number of references: 52
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 showed 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] Traditionally, three clinical variants are distinguished: endemic BL (eBL), sporadic BL (sBL) and immunodeficiency-related BL. eBL is usually positive for Epstein Bar 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 chromosomal translocation that juxtaposes *MYC* gene 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] Therefore, the recent WHO-HAEM5 classification suggests to subgroup BL tumors according to EBV status rather than in epidemiologic variants, which is long known to suffer from some impreciseness.[22,23]

BL is one of the few lymphoma entities that shows expression of the SRY-related HMG-box gene 11 (SOX11).[24–26] SOX11 expression in BL occurs in 25-55% and associates with the age of BL patients at diagnosis, showing expression predominantly in pediatric cases.[7,26,27] Moreover, SOX11 expression is included in the transcriptional molecular signature used to classify BL.[28] SOX11 expression in MCL is characteristic of the conventional molecular subtype (cMCL) with worse outcome than

the SOX11-negative leukemic non-nodal MCL subtype (nnMCL).[29] In contrast, no association between SOX11 expression and survival has been found in BL.[26] Several *in vitro* and *in vivo* studies have shown the oncogenic role of SOX11 in the pathogenesis of MCL.[30–35] However, the contribution of SOX11 expression to BL pathogenesis remains unknown.

To understand the relevance of SOX11 in BL, we have investigated the association of SOX11 expression with different molecular variables in BL patients using previously published large cohorts. To shed light on the role of SOX11 in the development of BL, we have analyzed changes in the gene expression profile (GEP) and performed functional analysis upon SOX11 overexpression and knockout in BL cell lines.

METHODS

Cell lines

For in vitro studies we used the SOX11-negative BL cell lines Ramos and DG75 (ATCC CRL-1596 and ATCC CRL-2625, respectively); and the SOX11-positive BL cell line BL2 (DSMZ ACC 625). Moreover, we used the stable transduced Z138-SOX11knockout (KO) MCL cell lines, previously generated by our laboratory using the CRISPR-CAS9 genome editing technology, and its control (Z138CT);[34] and the stable lentiviral transduced JVM2 cell line ectopically overexpressing SOX11 (JVM2-SOX11+) and its control JVM2CT.[36] HEK-293T cell line (ATCC CRL-3216) was used for lentivirus production. Cell line authentications were performed by qCell Identity (qGenomics) and *Mycoplasma* contamination was regularly tested.

Cells were cultured at 37 ºC and 5% CO2 in RPMI 1640 with L-glutamine (Corning) or DMEM with L-glutamine (Lonza) (only HEK-293T), supplemented with 10% of fetal bovine serum (FBS) (Gibco), 100 μg/ml streptomycin and 100 U/ml penicillin (Gibco).

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[37] 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[37] 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.[38,39] 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. In BLGSP cases for which no IHC was possible,[21] SOX11-positive expression was defined as >10.5 $\log_2$ transformed value, obtained by RNA-seq data.

EBER ISH

We performed EBER ISH in 51 pediatric BL cases with available FFPE biopsy specimens[37] using Leica Bond-MAX staining systems and reagents (Leica). EBV-positivity was defined as most tumor cells being positive.

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.[7,16,21,37] 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.

ddPCR assay for EBV traces

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).[40,41]

RNAscope for *EBNA1* and CD10 IHC

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).[40,41]

Generation of stable BL cell lines

Plasmid and stable BL cell lines generation are described in supplemental methods.

RNA-seq and RT-qPCR

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.[42] 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.[36] Gene count matrix was obtained for BLGSP[21] BL primary cases. RNA-seq data analyses are detailed in supplemental methods.

For RT-qPCR, cDNA was generated using qScript cDNA Synthesis Kit (Quantabio) and analyzed using Fast SYBR Green Master Mix (Applied Biosystems) and specific primers (Supplemental Table S1) following manufacturer's recommendations.

Statistics

Methods are described in the supplemental methods.

RESULTS

SOX11 expression and EBV infection are mutually exclusive in BL

To understand the clinical relevance of SOX11 in BL, we investigated the association between *SOX11* gene 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 significant 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,[7,37] performing SOX11 IHC and EBER ISH in the FFPE tissue sections of tumors in which these data were not available (Supplemental Table S2 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). 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 S3)

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 S3). 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). 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.[14,44] 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[37] sBL cases. We identified 17 coding genes mutated in ≥10% of BL cases (Figure 3A and Supplemental Table S4).

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 S5). 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 S6). 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.[40] 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), reinforcing that SOX11 and EBV positivity are mutually exclusive in BL. For cases tested for traces of EBV, limited molecular data were available. However, traces of EBV "corrected" the EBV-status of BL almost exclusively among SOX11- but not among SOX11+ BL.

Oncogenic pathways regulated by SOX11 in BL cell lines

In order to identify oncogenic pathways regulated by SOX11 in BL cells, we used the 4-Hydroxytamoxifen (4-OHT) inducible system, that allowed the ectopically overexpressed SOX11 protein fused to the hormone binding domain of the estrogen receptor (ER-SOX11) (Figure 4A) to enter the nucleus when DG75 ER-SOX11 cells were treated with 4-OHT (Figure 4B). 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 S2A), showing that more than 65% of differential expressed genes (DEG) overlapped upon SOX11 expression between 8h and 24h of 4-OHT treatment (Supplemental Figure S2B).

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 S7). 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).

SOX11-associated BL signature

We also generated the Ramos BL cell line overexpressing FLAG-SOX11 protein (RAMOS-SOX11) and knocked out SOX11 in BL2 SOX11+ cells, using the CRISPR-Cas9 gene editing system, generating the BL2-SOX11KO BL cell line (Figure 5A). Using RNA-seq we overlapped the DEGs obtained between Ramos-SOX11+ and Ramos-CT (Supplemental Figure S3A-B and Supplemental Table S8); between BL2 CT and BL2-SOX11KO cell lines (Supplemental Figure S3C-D and Supplemental Table S9), and between DG75 ER-SOX11 and DG75 ER (Supplemental Table S7). 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 S10).

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).

SOX11 functional role in MCL and BL

SOX11 directly regulates the transcription of genes involved in MCL oncogenic pathways.[29] Interestingly, gene set enrichment analysis (GSEA) using customized gene sets showed that SOX11-direct target genes found in Z138 MCL cell line by ChIP-chip[30] and genes upregulated in cMCL/SOX11+ vs nnMCL/SOX11- primary cases, and in Z138CT (SOX11+) vs SOX11 knockout (Z138-SOX11KO) MCL cell lines[36], were enriched in 4-OHT-treated DG75 ER-SOX11+, and in Ramos-SOX11+ and BL2 CT cells compared to their SOX11- respective control cell lines (DG75 ER, Ramos CT and BL2-SOX11KO, respectively). Besides, the downregulated genes in cMCL/SOX11+ vs

nnMCL/SOX11- primary tumors and in Z138 CT vs Z138-SOX11KO MCL cell lines were enriched in the SOX11- BL cell lines (Figure 6A and Supplemental Table S11).

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 deregulation of some of the overlapped genes, specifically *PLXNB1*, *CD24* and *MEX3A*, upon SOX11 overexpression and knockout, at mRNA and protein levels in both, MCL and BL transduced cell lines (Figure 6D-F, respectively), suggesting that SOX11 regulates genes involved in similar signaling pathways in BL and MCL.

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,[32,34] we did not observe a significant higher tumor cell migration towards CXCL13 or adhesion to SNKT stromal cells, nor an increase in the activation of BCR signaling pathway 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).

DISCUSION

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,[24,25] with a higher frequency in pediatric patients.[7] Several studies have described oncogenic functions of SOX11 in MCL.[30–35] 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 to be a feature 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 the region of Africa, reinforcing the concept of molecular subtypes of BL to be associated with EBV.

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.[45] 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 expression in conventional MCL possibly preventing the entrance of the tumor cell in the germinal center.[31] However, the expression levels of SOX11 in MCL are significantly higher than in BL cases,[25] 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] 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. *ID3* mutations are more frequently altered in EBV- than EBV+ BL, but we have also found them significantly higher in EBV-/SOX11+ than in EBV-/SOX11-. Interestingly, concomitant *SMARCA4* mutations and SOX11 expression has been also observed in MCL cases.[46] 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[29] exemplified by the overexpression of PLXNB1 and CD24 involved in tumor

cell migration,[47,48] and MEX3A involved in the chemoresistance of colorectal cancer quiescent cells.[49] Despite similar transcriptional programs of SOX11, *in vitro* experiments suggest that the activation of tumor cell migration or BCR pathways 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.[50,51] 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,[16,43] specifically seen in mucosal tissues,[52] such as Peyer's patches. Together these results suggest 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-positive BL suggests an earlier development than in SOX11-negative 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 and RB performed ddPCR, IHC and RNAscope; FN, GC 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.

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. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leuk. 2022 367*. 2022;36(7):1720–1748.

23. Leoncini L. Epstein–Barr virus positivity as a defining pathogenetic feature of Burkitt lymphoma subtypes. *Br. J. Haematol.* 2022;196(3):468.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

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

30. 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.

31. 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):1580–1599.

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

33. 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.

34. 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.

35. 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.

36. Sureda-Gómez M, Balsas P, Rodríguez ML, et al. Tumorigenic role of Musashi-2 in aggressive mantle cell lymphoma. *Leukemia*. 2023;37:408–421.

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

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

44. Roschewski M, Staudt LM, Wilson WH. Burkitt's Lymphoma. *N. Engl. J. Med.* 2022;387(12):1111–1122.

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

46. 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.

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

48. 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.

49. 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.

50. 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.

51. 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.

52. 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. **(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 S7), BL2 CT vs BL2-SOX11KO (in red, 107 genes, Supplemental Table S9) and Ramos SOX11+ vs Ramos CT (in green, 151 genes, Supplemental Table S8). **(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)** Bubble plot showing the normalized enrichment score (NES, x-axis), FDR (y-axis) and size (number of genes) of SOX11-specific MCL gene sets in DG75 ER-SOX11 vs DG75 ER (left), Ramos-SOX11+ vs Ramos CT (middle), and BL2 CT vs BL2-SOX11KO (right) GSEA comparisons. **(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.

DIGITAL IMAGE

Model of the molecular dichotomy between EBV and SOX11 in BL.

SOX11 expression is exclusive of EBV- BLs. In addition, the *IG::MYC* translocation is mainly acquired by CSR in EBV-/SOX11+ BLs, and by SHM in the EBV-/SOX11- and EBV+ BL cases. Finally, the three BL subgroups show different mutational landscapes, with higher frequency of *SMARCA4* and *ID3* and lower of *DDX3X* mutations in EBV-/SOX11+ compared to EBV-/SOX11- and EBV+ BL cases, and higher frequency of *TP53*, *CCND3, FBXO11* and *MYC* mutations, and lower of *FOXO1* and *SIN3A* in EBV- compared to EBV+ BLs.

Figure 1

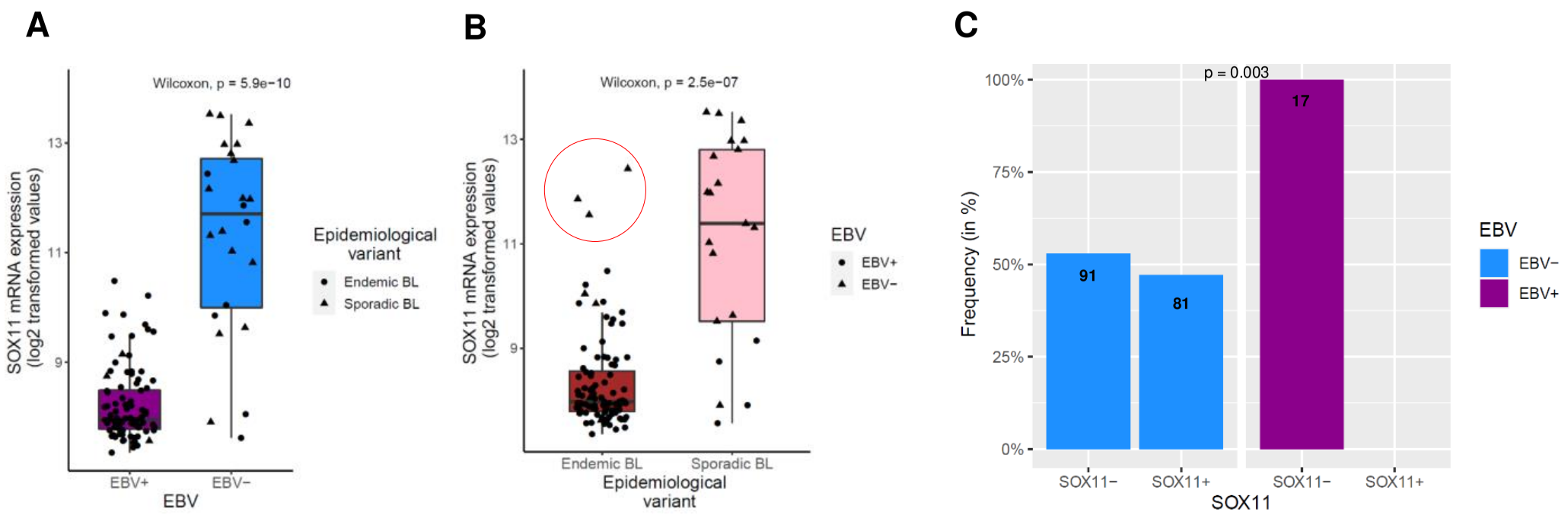

A
Wilcoxon, p = 5.9e−10
SOX11 mRNA expression (log2 transformed values)
13
11
9
EBV+
EBV−
EBV
Epidemiological variant
Endemic BL
Sporadic BL
B
Wilcoxon, p = 2.5e−07
SOX11 mRNA expression (log2 transformed values)
13
11
9
Endemic BL
Sporadic BL
Epidemiological variant
EBV
EBV+
EBV−
C
p = 0.003
100%
75%
50%
25%
0%
Frequency (in %)
91
81
17
SOX11−
SOX11+
SOX11−
SOX11+
SOX11
EBV
EBV−
EBV+

Figure 2

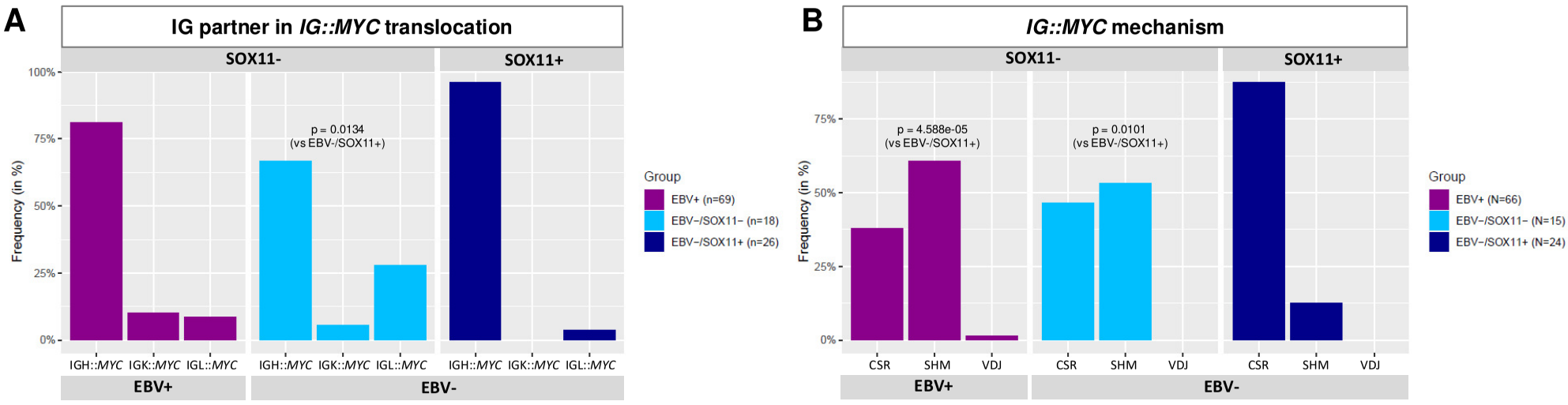
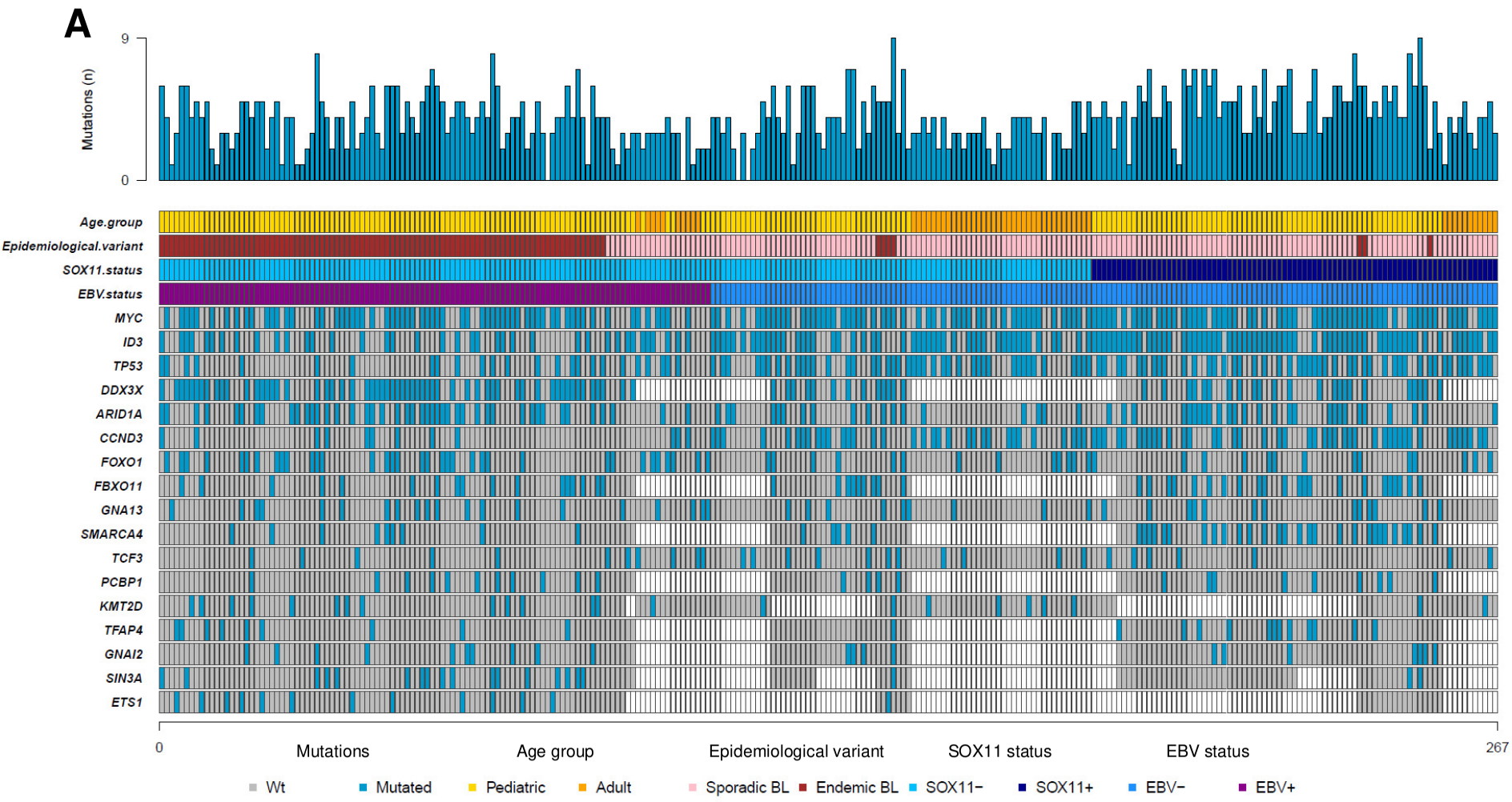

A
IG partner in IG::MYC translocation
SOX11-
SOX11+
100%
75%
50%
25%
0%
Frequency (in %)
p = 0.0134
(vs EBV-/SOX11+)
IGH::MYC
IGK::MYC
IGL::MYC
EBV+
EBV-
Group
EBV+ (n=69)
EBV−/SOX11− (n=18)
EBV−/SOX11+ (n=26)
B
IG::MYC mechanism
SOX11-
SOX11+
75%
50%
25%
0%
Frequency (in %)
p = 4.588e-05
(vs EBV-/SOX11+)
p = 0.0101
(vs EBV-/SOX11+)
CSR
SHM
VDJ
EBV+
EBV-
Group
EBV+ (N=66)
EBV−/SOX11− (N=15)
EBV−/SOX11+ (N=24)

Figure 3


A
Mutations (n)
9
0
Age.group
Epidemiological.variant
SOX11.status
EBV.status
MYC
ID3
TP53
DDX3X
ARID1A
CCND3
FOXO1
FBXO11
GNA13
SMARCA4
TCF3
PCBP1
KMT2D
TFAP4
GNAI2
SIN3A
ETS1
0
Mutations
Age group
Epidemiological variant
SOX11 status
EBV status
267
Wt
Mutated
Pediatric
Adult
Sporadic BL
Endemic BL
SOX11−
SOX11+
EBV−
EBV+

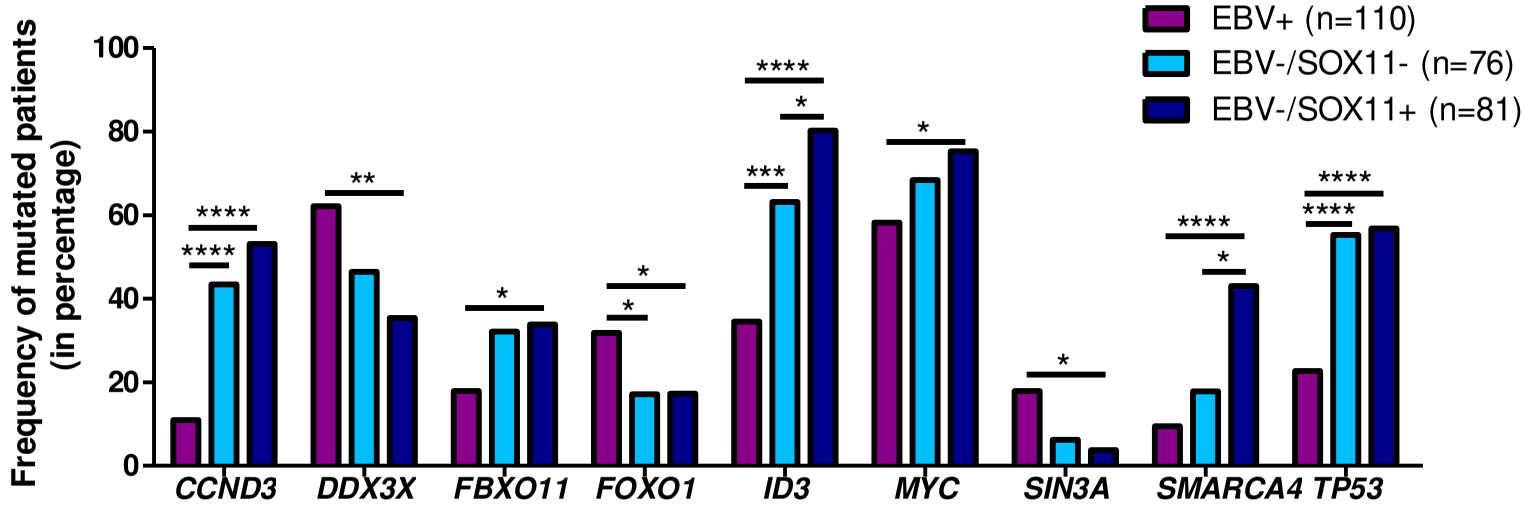

B
EBV+ (n=110)
EBV-/SOX11- (n=76)
EBV-/SOX11+ (n=81)
Frequency of mutated patients (in percentage)
100
80
60
40
20
0
CCND3
DDX3X
FBXO11
FOXO1
ID3
MYC
SIN3A
SMARCA4
TP53

Figure 4

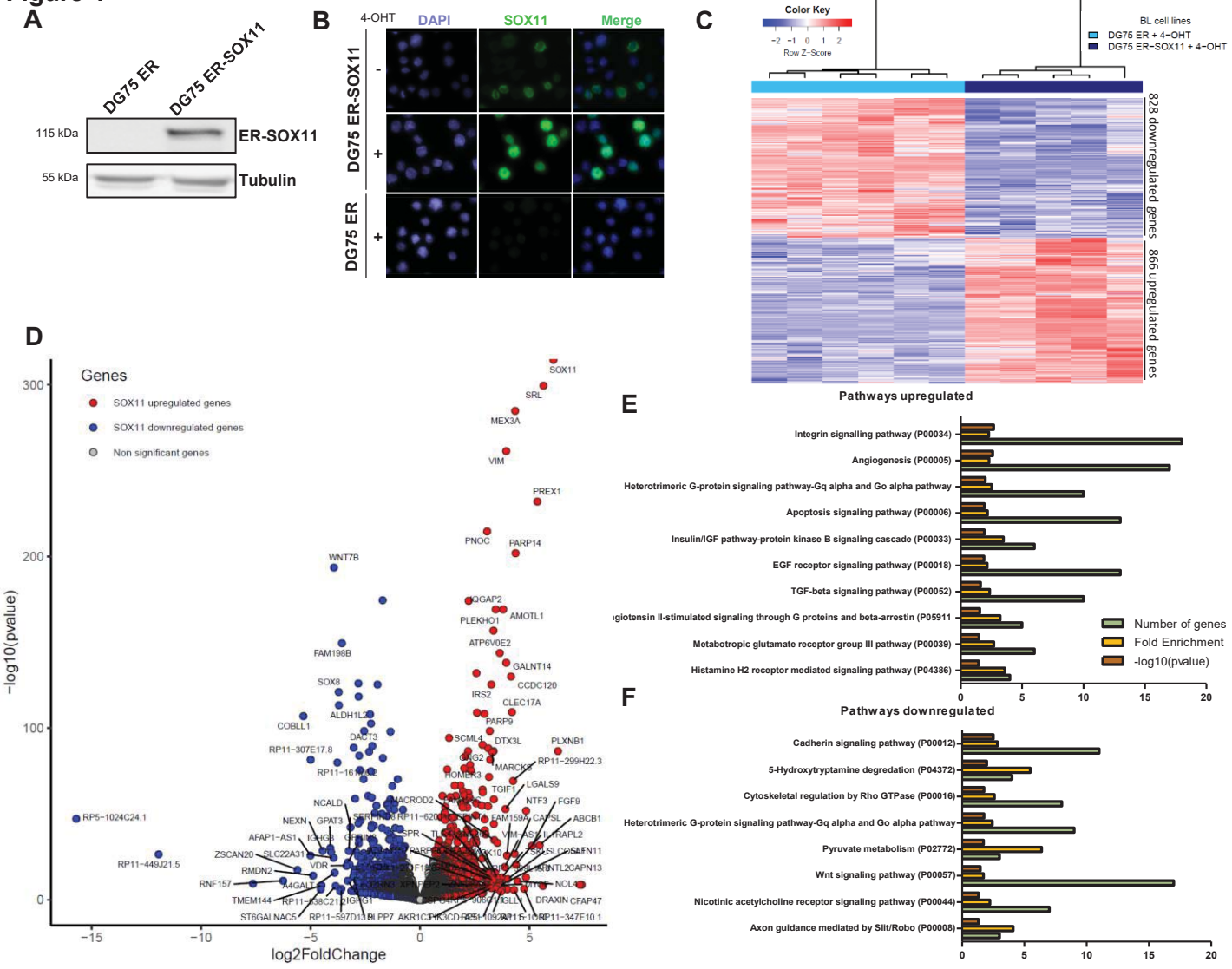

A
DG75 ER
DG75 ER-SOX11
115 kDa
ER-SOX11
55 kDa
Tubulin
B
4-OHT
DAPI
SOX11
Merge
DG75 ER-SOX11
DG75 ER
C
Color Key
Row Z-Score
BL cell lines
DG75 ER + 4−OHT
DG75 ER−SOX11 + 4−OHT
828 downregulated genes
866 upregulated genes
D
Genes
SOX11 upregulated genes
SOX11 downregulated genes
Non significant genes
−log10(pvalue)
log2FoldChange
E
Pathways upregulated
Integrin signalling pathway (P00034)
Angiogenesis (P00005)
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha pathway
Apoptosis signaling pathway (P00006)
Insulin/IGF pathway-protein kinase B signaling cascade (P00033)
EGF receptor signaling pathway (P00018)
TGF-beta signaling pathway (P00052)
Metabotropic glutamate receptor group III pathway (P00039)
Histamine H2 receptor mediated signaling pathway (P04386)
Number of genes
Fold Enrichment
-log10(pvalue)
F
Pathways downregulated
Cadherin signaling pathway (P00012)
5-Hydroxytryptamine degredation (P04372)
Cytoskeletal regulation by Rho GTPase (P00016)
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha pathway
Pyruvate metabolism (P02772)
Wnt signaling pathway (P00057)
Nicotinic acetylcholine receptor signaling pathway (P00044)
Axon guidance mediated by Slit/Robo (P00008)

Figure 5

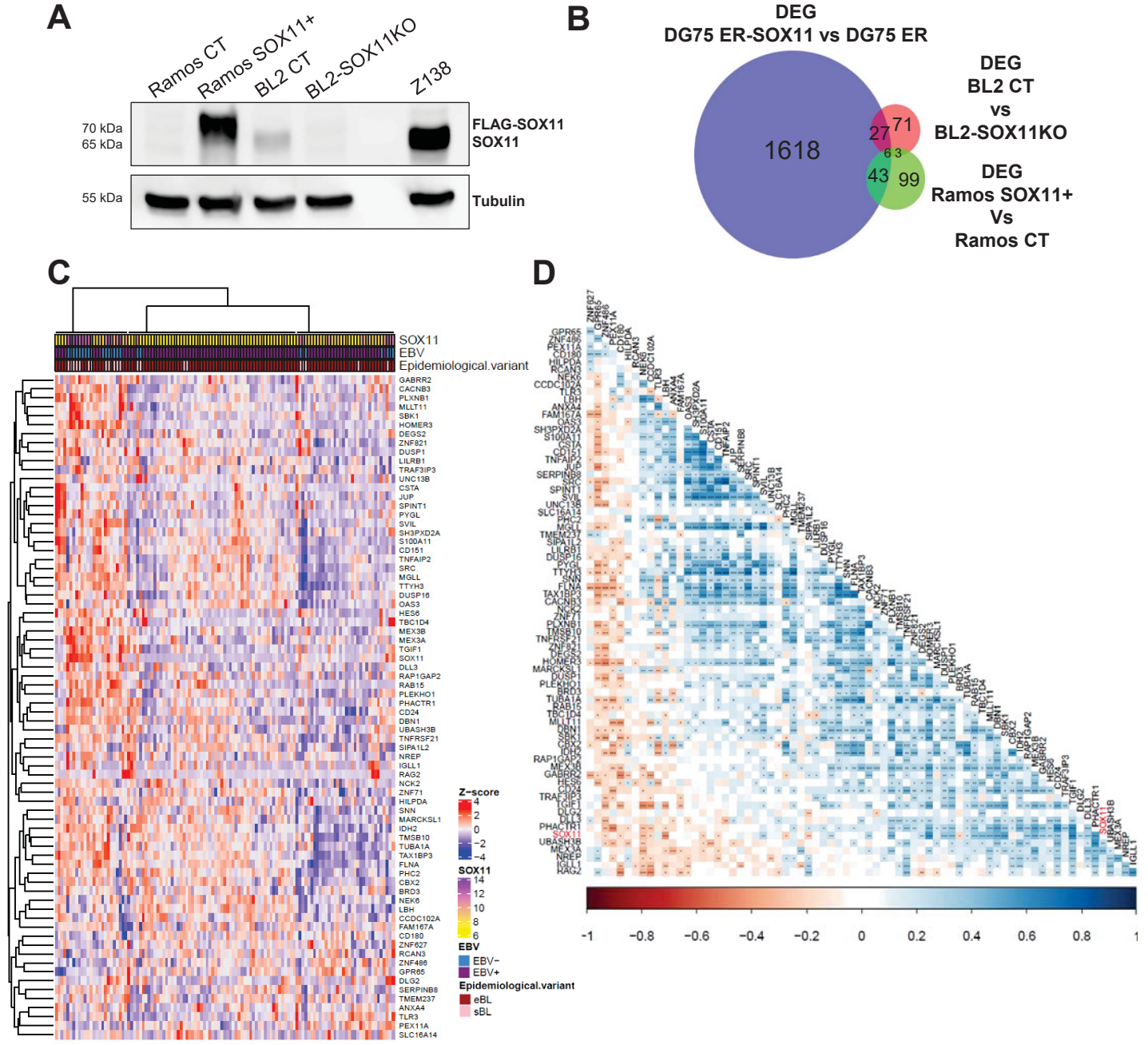
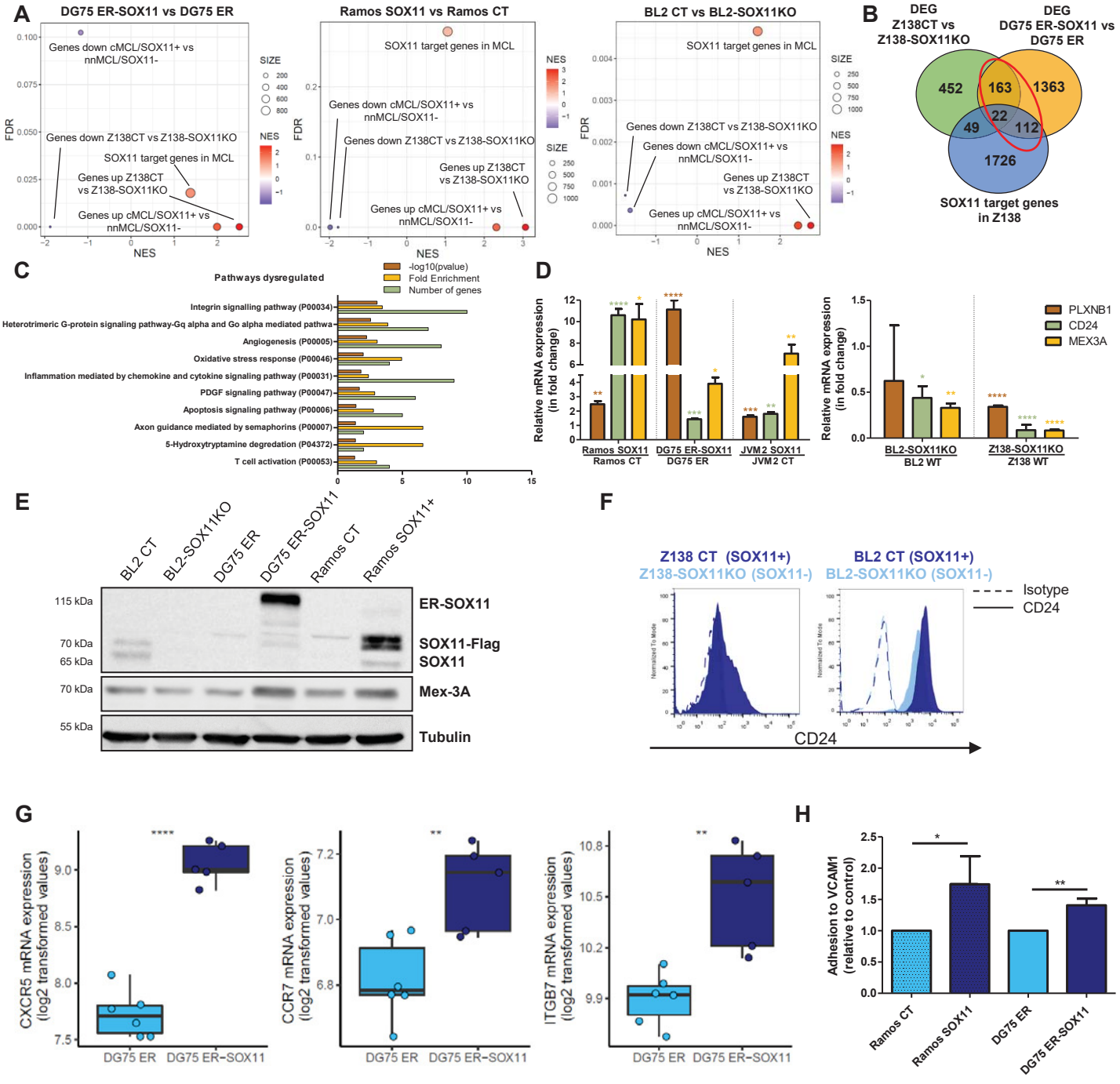

A
Ramos CT
Ramos SOX11+
BL2 CT
BL2-SOX11KO
Z138
70 kDa
65 kDa
FLAG-SOX11
SOX11
55 kDa
Tubulin
B
DEG
DG75 ER-SOX11 vs DG75 ER
1618
27
71
63
43
99
DEG
BL2 CT
vs
BL2-SOX11KO
DEG
Ramos SOX11+
Vs
Ramos CT
C
SOX11
EBV
Epidemiological.variant
Z-score
SOX11
EBV
EBV−
EBV+
Epidemiological.variant
eBL
sBL
D

Figure 6


A
DG75 ER-SOX11 vs DG75 ER
Ramos SOX11 vs Ramos CT
BL2 CT vs BL2-SOX11KO
Genes down cMCL/SOX11+ vs nnMCL/SOX11-
Genes down Z138CT vs Z138-SOX11KO
SOX11 target genes in MCL
Genes up Z138CT vs Z138-SOX11KO
Genes up cMCL/SOX11+ vs nnMCL/SOX11-
FDR
NES
SIZE
B
DEG Z138CT vs Z138-SOX11KO
DEG DG75 ER-SOX11 vs DG75 ER
452
163
1363
22
49
112
1726
SOX11 target genes in Z138
C
Pathways dysregulated
-log10(pvalue)
Fold Enrichment
Number of genes
Integrin signalling pathway (P00034)
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathwa
Angiogenesis (P00005)
Oxidative stress response (P00046)
Inflammation mediated by chemokine and cytokine signaling pathway (P00031)
PDGF signaling pathway (P00047)
Apoptosis signaling pathway (P00006)
Axon guidance mediated by semaphorins (P00007)
5-Hydroxytryptamine degredation (P04372)
T cell activation (P00053)
D
Relative mRNA expression (in fold change)
Ramos SOX11 / Ramos CT
DG75 ER-SOX11 / DG75 ER
JVM2 SOX11 / JVM2 CT
BL2-SOX11KO / BL2 WT
Z138-SOX11KO / Z138 WT
PLXNB1
CD24
MEX3A
E
BL2 CT
BL2-SOX11KO
DG75 ER
DG75 ER-SOX11
Ramos CT
Ramos SOX11+
115 kDa
70 kDa
65 kDa
ER-SOX11
SOX11-Flag
SOX11
Mex-3A
Tubulin
F
Z138 CT (SOX11+)
Z138-SOX11KO (SOX11-)
BL2 CT (SOX11+)
BL2-SOX11KO (SOX11-)
Isotype
CD24
G
CXCR5 mRNA expression (log2 transformed values)
CCR7 mRNA expression (log2 transformed values)
ITGB7 mRNA expression (log2 transformed values)
DG75 ER
DG75 ER-SOX11
H
Adhesion to VCAM1 (relative to control)
Ramos CT
Ramos SOX11
DG75 ER
DG75 ER-SOX11

SUPPLEMENTAL DATA

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†]

SUPPLEMENTAL METHODS

Cell lines

For in vitro studies we used the 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). Moreover, we used the stable transduced Z138-SOX11knockout (KO) MCL cell lines, previously generated by our laboratory using the CRISPR-CAS9 genome editing technology, and its control (Z138CT);[1] and the stable lentiviral transduced JVM2 cell line ectopically overexpressing SOX11 (JVM2-SOX11+) and its control JVM2CT.[2] HEK-293T cell line (ATCC CRL-3216) was used for lentivirus production. Cell line authentications were performed by qCell Identity (qGenomics) and *Mycoplasma* contamination was regularly tested.

Cells were cultured at 37 ºC and 5% CO2 in RPMI 1640 with L-glutamine (Corning) or DMEM with L-glutamine (Lonza) (only HEK-293T), supplemented with 10% of fetal bovine serum (FBS) (Gibco), 100 µg/ml streptomycin and 100 U/ml penicillin (Gibco).

ddPCR assay for EBV traces

Droplet digital PCR (ddPCR) was performed using 200 ng of DNA, 1 × ddPCRSupermix for Probes (BioRad, Hercules, CA, USA), 900 nM of each primer (Life Technologies, Italy), and 250 nM of the probe (Life Technologies, Italy) in a total volume of 22 µL. Among the fragments of the EBV genome, BamHI-W and EBNA1 are the most conserved region of EBV genome. After droplet generation using QX200TM Droplet Generator instrument (BioRad), the generated microdroplets were put into a 96-well plate for amplification. Cycling conditions included preheating at 95 °C for 10 min followed by 40 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 60 s, and

final heating at 98 °C for 10 min. Then, the PCR plate was transferred to a QX200 droplet reader (BioRad), and fluorescence amplitude data were obtained by QuantaSoft software (BioRad). The absolute copy number of each viral assays was calculated by Bio-Rad software and showed as number of copies/µl. As previously reported, we used as biological negative controls a sample of Hair Cell Leukemia (HCL), whom has never been associated with EBV.

RNAscope for EBNA1 and CD10 IHC

After immunohistochemistry (IHC) for CD10 (anti-human murine CD10, PA0270, Leica, Germany) associated with BOND Polymere Refine Red Detection Kit (Leica, Germany), RNAScope for EBNA1 (LS Probe V-EBV-EBNA1, Cat# 485388, Bio-Techne, Italy) was performed on the same slides using the BOND RNAscope Brown Detection reagent kit (Leica, Germany) according to manufacturer instructions. EBNA1 mRNA staining signals were identified as brown punctuate dots, while CD10+ cells were identified as red and counterstaining with hematoxylin. Each sample was quality-controlled for RNA integrity with a probe specific to the housekeeping Cyclophilin B (*PPIB*) mRNA used as positive control. Background staining was evaluated using a negative control probe specific for bacterial dihydrodipicolinate reductase (*dapB*); all Burkitt lymphoma (BL) cases analyzed did not show any dots for the *dapB* in any cells.

Differential expression analysis using RNA-seq data

Gene-level counts were imported to R (version 3.6.3, https://www.r-project.org) for analysis. Differential gene expression was conducted using R/Bioconductor DESeq2 package (version 1.24.0), shrinking the size factor with apeglm method. Genes were considered as differentially expressed when adjusted P-value <0.1 and absolute log2-

transformed fold change >0.65. Lists of differential expressed genes were used for functional annotation analysis using Panther Software. Variance stabilized expression matrix were obtained for gene set enrichment analysis (GSEA), clustering and gene expression analysis.

RT-PCR

For RT-qPCR, cDNA was generated using qScript cDNA Synthesis Kit (Quantabio) and analyzed using Fast SYBR Green Master Mix (Applied Biosystems) and specific primers (Supplemental Table S11) following manufacturer's recommendations.

Plasmid generation

ER-SOX11 expression was achieved using the pEF1α-IRES-AcGFP1 vector (Clontech). Both SOX11 ORF and the sequence of the hormone-binding domain of the mouse Estrogen Receptor (ER) carrying mutations that increase its affinity for the synthetic drug 4-hydroxytamoxifen (4-OHT),[3] was amplified by PCR and cloned in the vector using unique restriction sites. The ER^{TM} sequence was amplified from a plasmid expressing the MycER gene fusion[4] with primers adding the unique restriction sites NheI and HindIII and an improved Kozak sequence. SOX11 sequence was amplified by PCR from the plasmid pcDNA3 SOX11-HA[5] with primers removing the starting ATG and adding the unique restriction sites HindIII and BamHI.

Generation of SOX11-KO

BL2 cells with targeted deletions in the SOX11 gene were produced using a CRISPR/Cas9 strategy. Cas9 ribonucleoprotein (RNP) complexes were formed by combining 50 pmol of EnGen Cas9 NLS protein (New England Biolabs) with 100 pmol

each of two SOX11-targeting gRNAs: gRNA_U3 (GGAGAGCAACCTGCCCCGGG) and gRNA_D (GAAGATCCCGTTCATCCGGG). gRNAs were synthetized in vitro using the GeneArt™ Precision gRNA Synthesis Kit (Invitrogen), according to the manufacturer. Following RNP assembly 10 min at RT, the gRNA-Cas9 complexes were elecroporated by Nucleofection (Lonza) in $5x10^5$ pre-warmed BL2 cells resuspended in Nuceleofector kit V using program C-O9. Cells were left to recover in normal growth medium (RPMI + 10% FBS) for 72 hours and then diluted to 10 cells per ml. Cells were plated in 96 well plates to select single cell clones. Genotyping of expanded single cell clones was performed by PCR using the following primers: SOX11_fw9 (AGCCGCGAAAGCGGGGTGCCGAG) and SOX11_rev8 (CGCGCTCTTCTCTGGGCTCTGG). Presence or absence of small deletions in the SOX11 gene were evaluated by agarose gel electrophoresis and validated by Sanger sequencing of the PCR product.

Western Blot

For western blot, 50-70 μg of total protein was separated by 6–9% SDS–PAGE, transferring them to a 0.45 μm nitrocellulose membrane. Membranes were blocked with 5% of milk powder, in TBS-T for 1 h and incubated overnight at 4 ºC with specific antibodies to detect SOX11 (Cell Marque, MRQ-58, 1/1000 dilution), MEX-3A (Abcam, ab-79046, 1/1000 dilution) or α-Tubulin (Sigma-Aldrich, CP06, 1/5000 dilution). After incubation, washed membranes were incubated 1 h with the corresponding secondary antibody HRP conjugated (DAKO, 1/5000). Washed membrane was incubated with Pierce ECL reagent (Thermo Fisher Scientific) to detect proteins in ImageQuant LAS4000 (Fujifilm).

SOX11 immunofluorescence

DG75 ER and DG75 ER-SOX11 cells induced or not with 150 nM of 4-OHT (H7904, Sigma-Aldrich) for 24 h were incubated for 30 min into a coverslip pre-treated with poly-L-lysine (Sigma-Aldrich). Cells were fixed with 4% of paraformaldehyde for 15 min, washed twice with PBS, permeabilized for 20 min with blocking solution (PBS with 5% FBS and 0.5% Triton), and incubated for 1 h with SOX11 antibody (Abcam, ab170916, 1/100 dilution). Cells were washed twice with PBS, incubated for 1 h with secondary antibody anti-rabbit IgG A488 conjugated (ThermoScientific, A21206, 1/400 dilution), washed twice with PBS and once with MiliQ water, and then dried. A minimal volume of Mounting Medium with DAPI (DUO82040) was added, and samples were observed in Nikon Eclipse 50i microscope.

GSEA

GSEA was performed in variance stabilized expression matrices from RNA-seq data with GSEA v4.0.3. Data was randomized by 1000 permutations gene_set and SOX11-specific MCL customized gene sets previously obtained[5,6] were used for enrichment analysis.

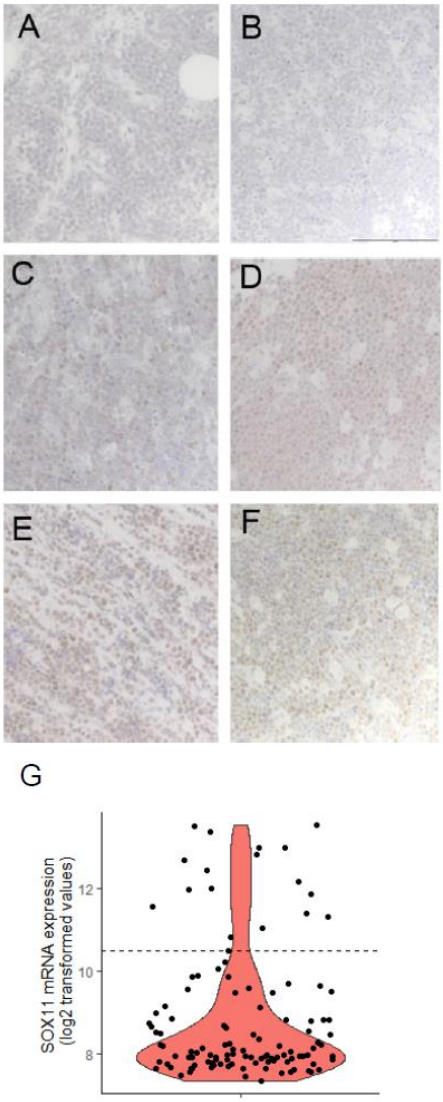
Cell cytometry staining

For extracellular staining of CD24, $2x10^5$ growing BL and MCL cells were washed with PBS and FBS 5% and stained with 2.5 µl of CD24-PE (BD, 555428) or IgG2a kappa isotype-PE (eBioscience, 12-4732-42) for 20-30 min at 4 ºC. Cells were washed with PBS and analyzed in FACS CANTO II (BD). To plot the results, mean fluorescence intensity (MFI) was used.

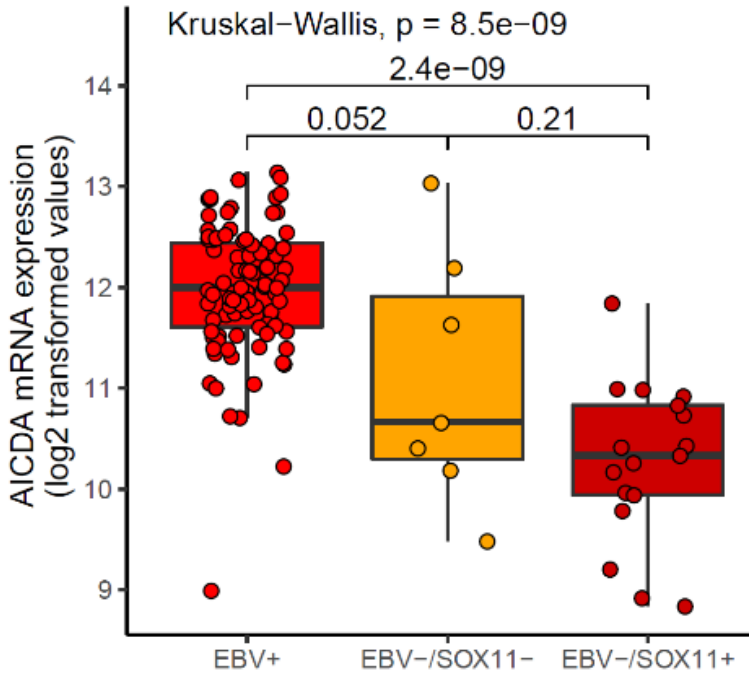
Adhesion assay

Untreated wells of a 96-well microplate were coated with 500 ng/ml of VCAM-1 or with BSA 0.5% in RPMI medium overnight at 4ºC. Next day, BL cells (SOX11-overexpressing and control DG75 (induced with 150 nM of 4-OHT during 24 h) and Ramos cells) were stained with 1 μM of Calcein AM Viability dye (eBioscience) for 30 min. Then, cells were centrifuged and $2x10^5$ BL cells were plated in wells coated with VCAM-1 or BSA 0.5% and incubated during 1 h. After incubation, wells were washed three times with PBS and cells were incubated in 50 μl of RPMI + 50 μl of Triton 2% for 30 minutes. Finally, cells were transferred to a black 96-well microplate and fluorescence emitted by calcein (λ535 nm) was measured in Synergy HT spectrophotometer (Biotek). Relative adhesion was measured as the calcein emitted fluorescence by cells coated with VCAM-1 divided by the fluorescence emitted by calcein-labelled cells adhered in an unspecific way to the well (with BSA 0.5%).

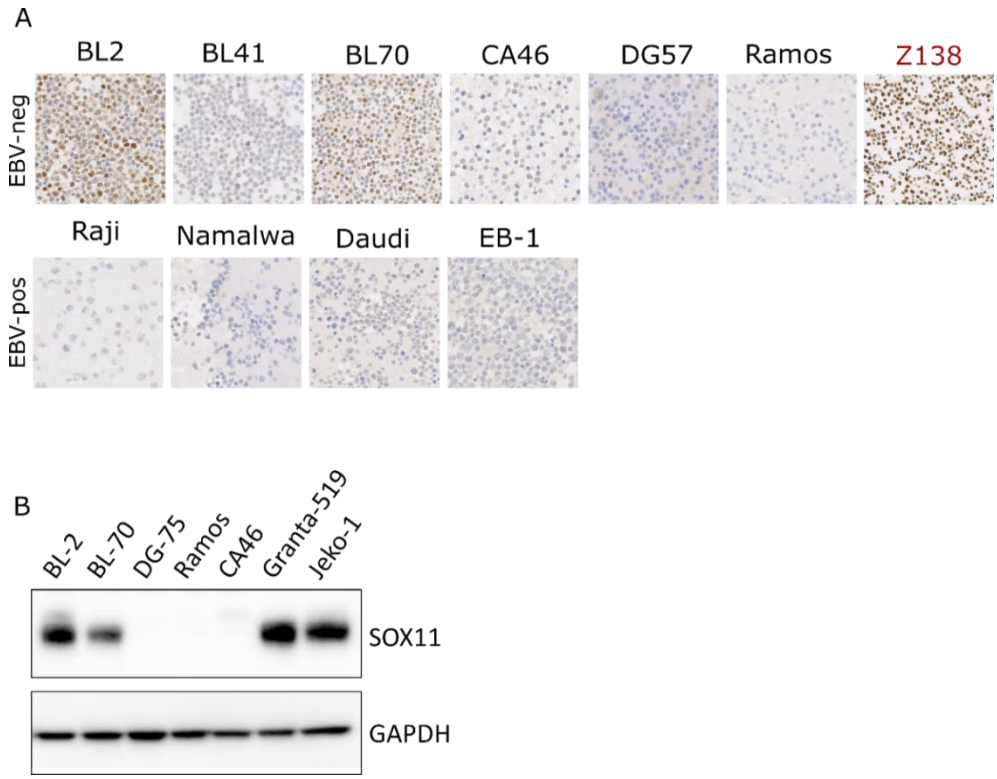
SUPPLEMENTAL FIGURES AND FIGURE LEGENDS



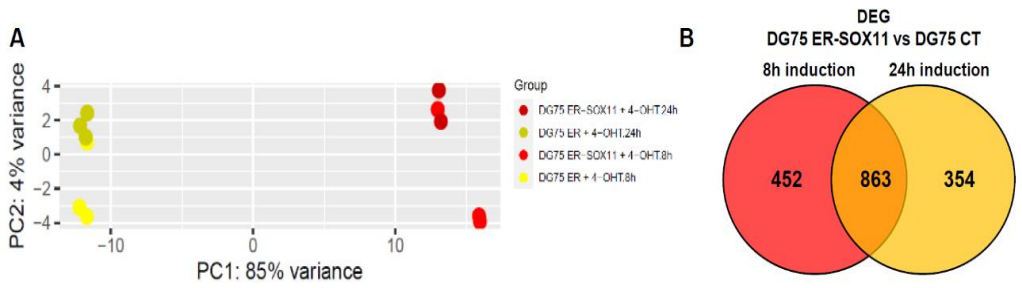
Supplemental Figure S1. SOX11 status in BL cases. SOX11 immunohistochemistry illustrating two cases of negative BL **(A-B)** and the spectrum of SOX11 expression in SOX11+ BL (four independent cases, **C-F**). BL negative for SOX11 showed no nuclear staining at all **(A-B)**. In SOX11 was scored positive when at least 10% of lymphoma cells showed unambiguous nuclear staining (low positive score in **C**). However, positive cases usually showed much higher scores with at least 50% positive nuclei **(D-F)**. Bar in B corresponds to 100µm. Original magnification 400x. **(G)** Illustration of SOX11 mRNA cut-off (10.5 $\log_2$ transformed values, dashed line) obtained from RNA-seq data of 117 BL cases from the BLGSP. SOX11+ status ≥10.5, SOX11- status <10.5.



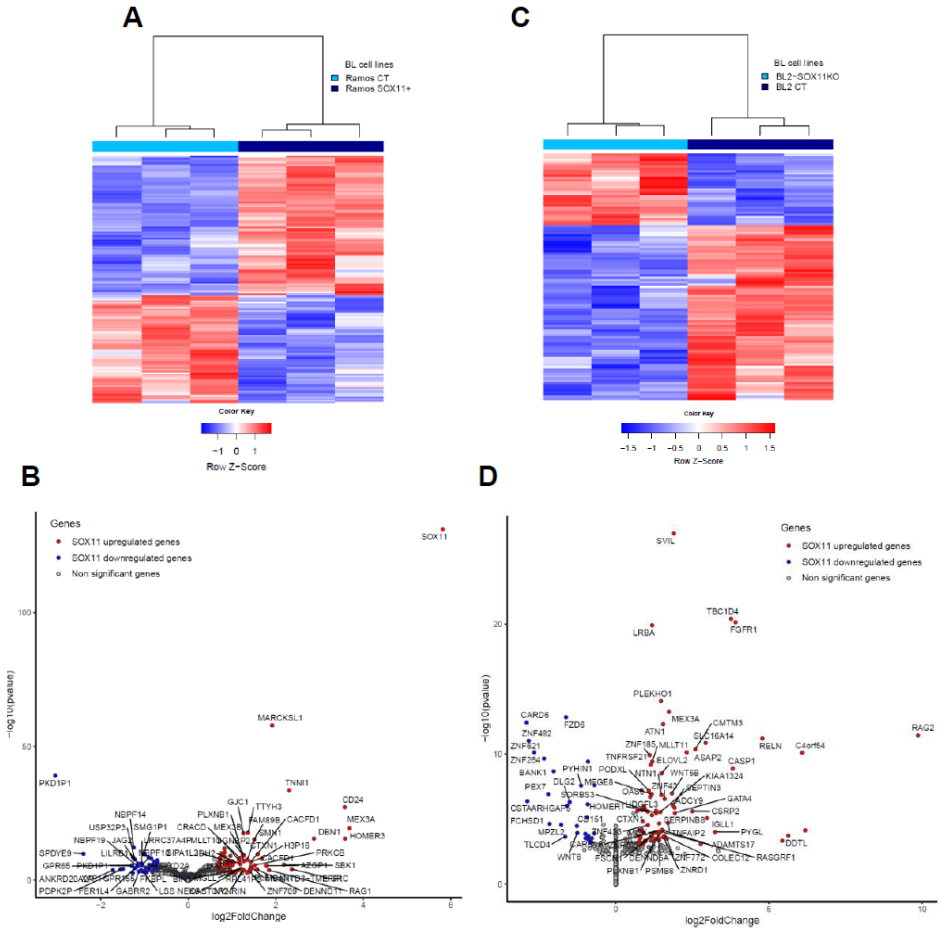
Supplemental Figure S2. AICDA mRNA expression levels (log_2 transformed values) in EBV+ (n=93), EBV-/SOX11- (n=7) and EBV-/SOX11+ (n=17) BL cases. To test differences Kruskall-wallis was performed, and to evaluate pairwise differences, Wilcoxon test was used.



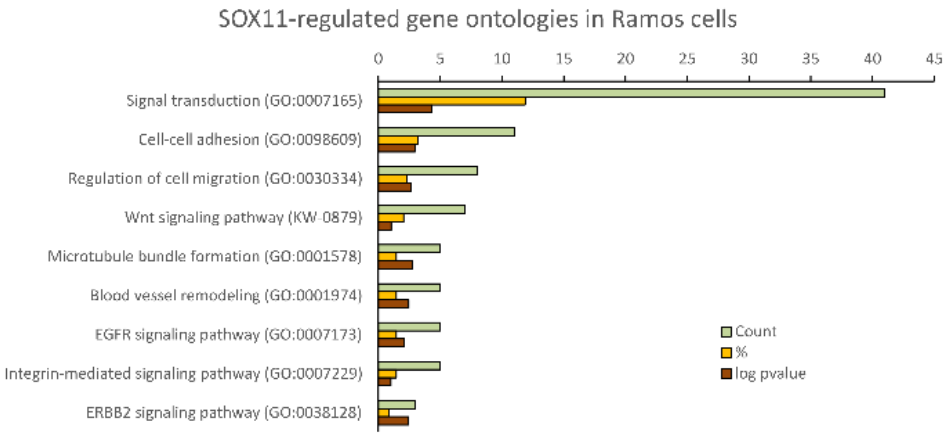
Supplemental Figure S3: Analysis of SOX11 expression in EBV- and EBV+ BL cell lines. (A) SOX11 immunohistochemistry of 11 formalin fixed, and paraffin embedded BL cell lines (cytoblocks) derived into EBV- and EBV+. EBV status according to https://www.dsmz.de. SOX11 immunohistochemistry in a cytoblock of the MCL cell line Z138 is shown as a positive control. **(B)** Western blot analysis of SOX11 expression in a subset of the above cell lines in comparison with the MCL cell lines Granta-519 and Jeko-1. The expression of GAPDH is shown as a loading control.



Supplemental Figure S4. Gene expression profile of DG75-ER-SOX11 BL cell line. **(A)** Principal component analysis on RNA-seq data from DG75-ER-SOX11 and -CT BL cell lines upon 4-OHT induction for 8h and 24h. Principal components 1 and 2 are shown**.** Only the 500 first more variable genes were considered to do the analysis. **(B)** Overlap between differential expressed genes after 8h (in red, 1315 genes) and 24h (in yellow, 1217 genes) of ER-SOX11 induction in DG75 ER-SOX11 BL cell line compared to DG75 CT at 8h and 24h 4-OHT induction.



E



Supplemental Figure S5. Gene expression analysis upon SOX11 overexpression in Ramos SOX11+ and knockout in BL2 BL cell lines. (A and C) Heatmap illustrating the scaled expression (Z-score) of differential expressed genes in **(A)** Ramos SOX11+ compared to Ramos CT cell lines (85 upregulated and 66 downregulated genes)

(Supplemental Table S7), and **(C)** BL2 CT compared to BL2-SOX11KO cell lines (76 upregulated and 31 downregulated genes) (Supplemental Table S8), obtained by RNA-seq. Genes with an adjusted P-value <0.1 and absolute log2-transformed fold change >0.65 were considered. **(B and D)** Volcano plot showing genes differentially expressed, obtained by RNA-seq, upon SOX11 overexpression in **(B)** Ramos SOX11+ compared to Ramos CT cell lines, and **(D)** BL2 CT compared to BL2-SOX11KO cell lines. The graph shows on the y-axis -log (P-value) and on the x-axis the log2-transformed fold change. Genes upregulated and downregulated with an adjusted P-value <0.1 and log2-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)** Panther pathway enrichment analysis using DEG between RAMOS SOX11+ and RAMOS CT cell lines. Number of genes and –log10(P-value) for each pathway are shown.

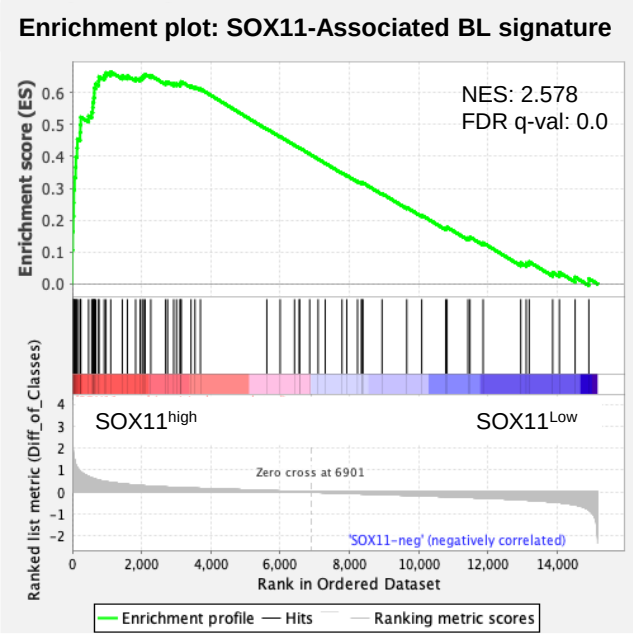
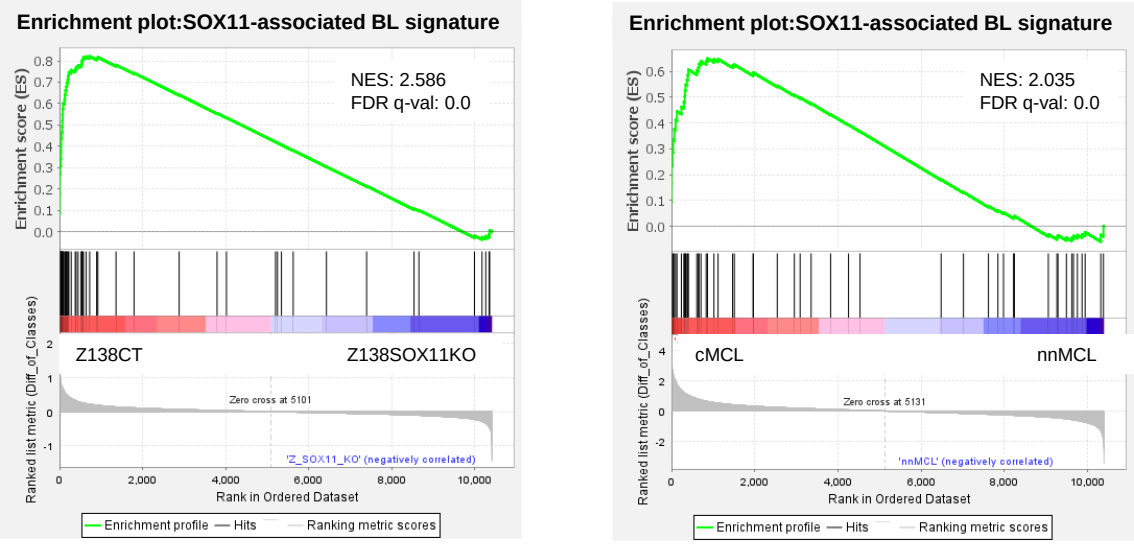
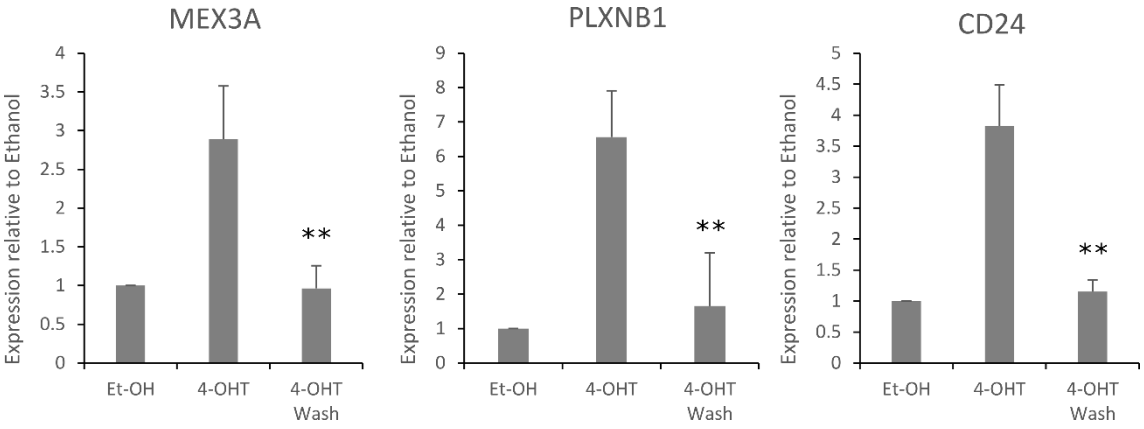


Figure S6. Similar SOX11 transcriptional activity in BL cell lines and primary samples. GSEA showing the enrichment of the SOX11-associated BL signature (Supplemental Table S9) in SOX11^{high} compared to SOX11^{Low} BL primary samples (SOX11 mRNA cut-off=10.5 log_2 transformed values; see Figure S1G). Normalized enrichment score (NES), P value, and false-discovery rate (FDR) are shown. FDR<0.1 indicates statistical significance.



Supplemental Figure S7. Overlap between the expression of a SOX11-specific signature in BL and MCL. GSEA analysis showing enrichment of the SOX11-associated BL signature (Supplemental table S9) in SOX11+ compared to SOX11- **(A)** MCL cell lines (Z138CT vs Z138SOX11KO) and **(B)** cases (cMCL vs nnMCL). Normalized enrichment score (NES), P value, and false-discovery rate (FDR) are shown. FDR<0.1 indicates statistical significance.



Supplemental Figure S8. Dependency on 4-OHT for SOX11-associated gene expression. DG75-ER-SOX11 cells were cultured for 24 hours in presence of either ethanol (Et-OH) or 4-hydroxytamoxifen (4-OHT). After harvesting of cells for gene expression analysis the rest of the cells were washed in complete medium and let them grow for another 24 hours. RNA was extracted from the cells in the above-mentioned conditions and the expression of three SOX11-target genes was analyzed by qPCR. GUSB was used as housekeeping gene. Asterisks indicate a p-value <0.005 in the comparison 4-OHT vs 4-OHT Wash.

SUPPLEMENTAL TABLES

Supplemental Tables are placed in the Supplemental Tables Excel file.

Supplemental Table S1. SOX11 and EBV association (SOX11 status by IHC and EBV status by EBER ISH) in BL pediatric and adult patients.

Supplemental Table S2. IG partner and mechanism leading to *IG::MYC* translocation in BL pediatric cases.

Supplemental Table S3. Mutational status in BL recurrently mutated genes in BL cases.

Supplemental Table S4. Frequency of mutations in BL recurrently mutated genes in EBV+, EBV-/SOX11- and EBV-/SOX11+ BL cases and comparisons between groups.

Supplemental Table S5. Frequency of mutations in BL recurrently mutated genes that shown differences between EBV+, EBV-/SOX11- and EBV-/SOX11+ BL cases. Pairwise comparisons without and with correction (by FDR) are shown.

Supplemental Table S6. Differentially expressed genes upon SOX11 overexpression in DG75 BL cell line with adjusted P-value <0.1 and absolute log2-transformed fold change >0.65.

Supplemental Table S7. Differentially expressed genes upon SOX11 overexpression in Ramos BL cell line with adjusted P-value <0.1 and absolute log2-transformed fold change >0.65.

Supplemental Table S8. Differentially expressed genes upon SOX11 knockout in BL2 BL cell line with adjusted P-value <0.1 and absolute log2-transformed fold change >0.65.

Supplemental Table S9. Genes overlapping between differential expression analysis in DG75 and Ramos SOX11-overexpressing, and BL-SOX11KO cell lines, composing the SOX11-associated signature in BL.

Supplemental Table S10. Customized gene sets SOX11-specific found in MCL used for GSEA in BL cell line models.

Supplemental Table S11. Primers used for qRT-PCR.

REFERENCES

1. 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.

2. Sureda-Gómez M, Balsas P, Rodríguez ML, et al. Tumorigenic role of Musashi-2 in aggressive mantle cell lymphoma. *Leukemia*. 2023;37:408–421.

3. Littlewood TD, Hancock DC, Danielian PS, Parker MG, Evan GI. A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Nucleic Acids Res.* 1995;23(10):1686–1690.

4. Iaccarino I, Hancock D, Evan G, Downward J. c-Myc induces cytochrome c release in Rat1 fibroblasts by increasing outer mitochondrial membrane permeability in a Bid-dependent manner. *Cell Death Differ.* 2003;10(5):599–608.

5. 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.