Published in final edited form as:
*Nat Commun*. ; 5: 3904. doi:10.1038/ncomms4904.

# Transient expression of Bcl6 is sufficient for oncogenic function and induction of mature B-cell lymphoma

**Michael R Green**[1,*], **Carolina Vicente-Dueñas**[2,3,*], **Isabel Romero-Camarero**[2,3], **Chih Long Liu**[1], **Bo Dai**[1], **Inés González-Herrero**[2,3], **Idoia García-Ramírez**[2,3], **Esther Alonso-Escudero**[2,3], **Javeed Iqbal**[4], **Wing C Chan**[4], **Elena Campos-Sanchez**[5], **Alberto Orfao**[6], **Belén Pintado**[7], **Teresa Flores**[3,8], **Oscar Blanco**[8], **Rafael Jiménez**[3,9], **Jose Angel Martínez-Climent**[10], **Francisco Javier García Criado**[11], **María Begoña García Cenador**[11], **Shuchun Zhao**[12], **Yasodha Natkunam**[12], **Izidore S Lossos**[13], **Ravindra Majeti**[1], **Ari Melnick**[14], **César Cobaleda**[5], **Ash A. Alizadeh**[1,#], and **Isidro Sánchez-García**[2,3,#]

[1]Divisions of Oncology and of Hematology, School of Medicine, Stanford University, Stanford, CA, USA

[2]Experimental Therapeutics and Translational Oncology Program, Instituto de Biología Molecular y Celular del Cáncer, CSIC/ Universidad de Salamanca, Campus M. de Unamuno s/n, 37007, Salamanca, Spain

[3]Institute of Biomedical Research of Salamanca (IBSAL), Salamanca, Spain

[4]Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE 68198, USA

[5]Centro de Biología Molecular Severo Ochoa, CSIC/Universidad Autónoma de Madrid, c/Nicolás Cabrera, nº 1, Campus de Cantoblanco, 28049, Madrid, Spain

[6]Servicio de Citometría and Departamento de Medicina, Universidad de Salamanca, 37008, Salamanca, Spain

[7]Genetically Engineered Mouse Facility, CNB-CSIC, 28006, Madrid, Spain

[8]Departamento de Anatomía Patológica, Universidad de Salamanca, Salamanca, Spain

[#]To whom correspondence should be addressed. arasha@stanford.edu, isg@usal.es.
[*,#]Should be considered as equal first author or senior author, respectively.

Authorship Contributions

MRG and CVD designed and conducted the majority of experiments, analyzed the data and wrote the manuscript. IRC, BD CLL, IGH, EAE, ECS, RJ, JAMC, FJGC, MBGC, SZ, YN, ISL, AM, RM and CC performed experiments and analyzed the data. BP generated Bcl6-floxed mice, and IGR characterized the Bcl6-floxed and p53 mutant mice. AO analyzed flow cytometry data. TF and OB prepared tissue sections and helped analyze mouse histopathology. SZ and YN performed immunohistochemical characterization of mouse tissues and YN provided expert classification of DLBCL tumors as a trained hematopathologist. JI and WCC provided primary human DLBCL tumor BCL6 genotypes and corresponding gene expression profiles. AAA and ISG conceived the project, designed research, analyzed data, prepared the manuscript, and contributed equally as senior authors in supervising the project. All authors commented upon and edited the final manuscript.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Data Accession Numbers

Gene expression omnibus accession includes; GSE11318, GSE12906, GSE15127, GSE22082, GSE34171, GSE11318, GSE34171, GSE12453, GSE36503, GSE56310.

[9]Departamento de Fisiología y Farmacología, Universidad de Salamanca, Campus M. Unamuno s/n, Salamanca, Spain

[10]Division of Oncology, Center for Applied Medical Research (CIMA), University of Navarra, 31008 Pamplona, Spain

[11]Departamento de Cirugía, Universidad de Salamanca

[12]Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

[13]Division of Hematology-Oncology, University of Miami, Sylvester Comprehensive Cancer Center, Miami, FL 33136, USA

[14]Departments of Medicine and Pharmacology, Weill Cornell Medical College, New York, NY 10021, USA

## Abstract

Diffuse large B-cell lymphoma (DLBCL) is the most common lymphoma and can be separated into two subtypes based upon molecular features with similarities to germinal center B-cells (GCB-like) or activated B-cells (ABC-like). Here we identify gain of 3q27.2 as being significantly associated with adverse outcome in DLBCL and linked with the ABC-like subtype. This lesion includes the BCL6 oncogene, but does not alter BCL6 transcript levels or target-gene repression. Separately, we identify expression of BCL6 in a subset of human hematopoietic stem/progenitor cells (HSPCs). We therefore hypothesize that BCL6 may act by hit-and-run oncogenesis. We model this by transiently expressing Bcl6 within murine HSPCs, and find it causes mature B-cell lymphomas that lack Bcl6 expression and target-gene repression, are transcriptionally similar to post-GCB cells, and show epigenetic changes that are conserved from HSPCs to mature B-cells. Together these results suggest that Bcl6 may function in a hit-and-run role in lymphomagenesis.

## Introduction

As the most common aggressive lymphoma afflicting nearly 30,000 Americans each year, diffuse large Bcell lymphoma (DLBCL) is highly heterogeneous. Current combination therapeutic regimens typically fail in nearly half of all patients with DLBCL, many of whom succumb to their disease. Given the inability to cure many patients with DLBCL, and the significant toxicity of current therapies, better treatment strategies are needed. We previously described a major molecular determinant of this biological and clinical heterogeneity, likely reflecting the cellular origin of tumors. Patients with tumors that have transcriptional profiles related to germinal center B-cells (GCB-like) have a better overall survival than those with tumors having a transcriptional profile related to post-GCB activated B-cells (ABC-like)[1]. This finding has been validated by several groups independently, and the molecular basis for this diversity in DLBCL has been partially deciphered in studies of distinctive genomic aberrations and somatic mutations in DLBCL subtypes.

Genomic studies have defined a subset of alterations that stratify between the two DLBCL subtypes[2,3], with point mutations of histone modifying genes and B-cell receptor signaling components as the prevailing dominant drivers or accelerators of the disease[4]. However,

these alterations are found in only a fraction of patients, and the relationship between more common genetic alterations and DLBCL subtypes remains largely obscure. For example, the most frequent somatic alteration observed in DLBCL, involving genetic translocation of *BCL6*, is arguably the most prominent and paradoxical[5,6]. *BCL6* is a central regulator of germinal center development[7,8], it is more highly expressed in the GCB-like subtype of DLBCL compared to the ABC-like subtype, and is associated with a favorable prognosis[1,9]. Yet genetic translocations of this gene are more prominent in the post-GCB subtype of the disease and associated with adverse outcome[1,10]. Recent findings have implicated Bcl6 in leukemia stem cell survival[11,12] and show its activity may be altered by CREBBP or EP300 mutation[3] at an early stage lymphoma development[13,14]. Separately, genetic and epigenetic aberrations in premalignant hematopoietic progenitors have recently been described in several hematological malignancies, including AML and CLL[15–18]. Together, these findings led us to postulate that *BCL6* may promote tumorigenesis in a manner contrasting that of other traditional oncogenes which act in fully evolved tumor cells and require persistent activity due to oncogene addiction[19].

Somatic DNA copy number alterations (SCNAs) perturb more of the cancer genome than any other somatic alteration, and can alter the gene dosage and subsequent expression of multiple genes in a single alteration[20]. The significance of SCNAs can be assessed from the patterns of broad and focal gains/losses across the genomes of a tumor cohort, allowing potential target genes within conserved regions of DNA copy number gain/loss to be identified. The integration of expression profiling data has additionally allowed putative driver genes within each lesion to be localized by their changes in transcript abundance resulting from altered gene dosage[21]. However, a subset of oncogenes with negative feedback loops may act in a ‘hit-and-run’ fashion; therein, transient expression of the oncogene may induce broad changes to the cancer genome, epigenome, or transcriptome, and be sufficient for oncogenesis in the absence of persistent expression. These ‘hit-and-run’ oncogenes may therefore not be detected by integrative analysis of DNA copy number and gene expression changes, and are difficult to identify in the absence of other genetic alterations targeting the same locus, such as genetic translocations or somatic mutations.

Here we use high resolution analysis of DNA copy number across a large cohort of DLBCL tumors to elucidate recurrent alterations in this disease. We identify gain of the *BCL6* oncogene as being a potential ‘hit-and-run’ oncogene associated with poor outcome and the ABC-like DLBCL subtype. Using transgenic mouse models, we confirm that transient expression of Bcl6 is sufficient to induce aggressive mature B-cell lymphoma that appears transcriptionally similar to activated post-germinal center B-cells.

## Results

### Gain of 3q27.2 is associated with inferior outcome in DLBCL

Using high-resolution DNA copy number profiles of 609 DLBCL tumors analyzed using the GISTIC method, we mapped the landscape of SCNAs in this disease. We identified 22 peaks of significant DNA copy number loss (GISTIC Q-value <0.25) and 17 peaks of significant DNA copy number gain (Figure 1a, Supplementary Table 1). We analyzed the association of each lesion with overall survival in cohorts of patients treated with

combination chemotherapy (CHOP, n=232) or in combination with Rituximab (RCHOP, n=196). Gain of 3q27.2 was the most prognostic lesion and was associated with significantly decreased overall survival in both cohorts (Figure 1b–c, Supplementary Table 1). For 249 cases, matched gene expression profiling data was available and allowed for classification of samples into GCB-like and ABC-like subtypes using the previously defined method (Figure 1d). Gain of 3q27.2 was significantly over-represented in the ABC-like DLBCL subtype (Fisher P-value $=8.1 \times 10^{-8}$), suggesting that it may contribute to the genetic etiology of this subtype and its association with adverse outcome.

The peak of this alteration (chr3:180478352-199501827) contained 134 genes, including the lymphoma oncogene *BCL6*[22] but not a previously hypothesized target of chromosome 3 alterations, *FOXP1*[23] (Figure 2a). The significance of the 3q27.2 peak resulted from a combination of broad events (trisomy 3 and 3q arm-level gain) and recurrent focal gains of 3q27.2 over the *BCL6* locus (Figure 2b). Furthermore, 3q27.2 DNA copy number gains were mutually exclusive of translocations targeting *BCL6* in the 48 tumors for which such data were available[24], suggesting that *BCL6* is a likely target of these lesions (Figure 2b). However, there was no significant increase in either *BCL6* expression or BCL6-target gene[25] repression in cases with 3q27.2 gain compared to those cases with diploid copy number (Figure 2c–d). This same trend was also observed within the context of BCL6 translocations (Figure 2e–f), as previously described[10]. The absence of increased *BCL6* transcript levels in tumors harboring these genomic alterations could not be attributed to the uniformly high expression of BCL6 within DLBCL, since BCL6 transcript levels were found to be significantly lower in DLBCL than in non-malignant B-cells (Supplementary Fig. 1). Together, these data led us to hypothesize that *BCL6* may act in a ‘hit-and-run’ fashion and promote oncogenesis by transient over-expression during an early stage of hematopoietic differentiation or tumor evolution, and that its expression is no longer maintained or required in fully evolved tumor cells.

## BCL6 expression in human hematopoietic precursors

To investigate the potential for *BCL6* to be acting at an early stage of hematopoietic differentiation, we performed gene expression profiling of 183 single sorted human hematopoietic stem/precursor (HSPC) cells from bone marrow aspirates of 2 healthy adults (Supplementary Fig. 2). Single cell profiling was employed in order to allow evaluation of small populations of cells possessing unique gene expression phenotypes that may not be discernible from analysis of bulk populations. Using the beta-2-microglobulin (*B2M*) gene as a reference for successful RNA amplification and gene expression microarray analysis, we observed that BCL6 transcript was detectable in 6.5% (6/92, donor 1) to 7.7% (7/91, donor 2) of single HSPCs (Figure 3a). Notably, HSPCs expressing *BCL6* lacked significant expression of other markers of B-cell differentiation including Pax5, CD19, or CD20, among others. Nonetheless, gene set enrichment analysis found significant repression of BCL6-bound target genes[25] within the population of cells with expressing *BCL6* transcript levels compared to those cells not expressing *BCL6* (Figure 3b). Thus, in a minor subset of adult human HSPCs, BCL6 expression is associated with an early transcriptional program, consistent with similar observations in human cord blood implicating this transcription factor in the commitment of specific progenitors to the lymphoid fate[26].

We therefore hypothesized that *BCL6* may contribute to lymphomagenesis at a stage of development prior to B-cells reaching full maturity. Because *BCL6* shows highly conserved patterns of expression between humans and mice[27], we tested the hypothesis by generating a murine strain that transiently expressed *Bcl6* specifically within HSPCs by placing a floxed *Bcl6* cDNA with an IRES-GFP marker under control of the promoter for the *Ly6A* locus encoding Stem cell antigen-1 (*Sca1*) (*Sca1-Bcl6$^{floxed}$*, Figure 3c)[28]. As expected, using flow cytometric analysis of the co-expressed GFP marker, we detected expression in hematopoietic stem cells (HSC) and multipotential progenitors (MPP), with expression declining in common lymphoid progenitors (CLP) towards barely-detectable expression in pre-proB cells and no detectable expression in proB cells or later stages of B-cell development (Figure 3d; Supplementary Fig. 3). Therefore, this transient expression of *Bcl6* during early hematopoietic development provided a means to evaluate the potential for *Bcl6* to contribute to lymphoma development via a hit-and-run mechanism.

## Transient Bcl6 expression induces mature B-cell lymphoma

*Sca1-Bcl6$^{floxed}$* mice are viable and develop normally, with a typical hematopoietic system early in post-gestational development, and normal germinal center development in response to a T-cell dependent immunogen (Supplementary Fig. 4). However, by 4 weeks of age, increased numbers of hematopoietic stem cells, Lin-Sca1+cKit+ (LSK) and lineage restricted progenitors (Lin-CD48+CD150+), and decreased numbers of common lymphoid progenitors (Lin-Sca1+CD127+cKit+) can be detected (Supplementary Fig. 5). In 8-week old mice, we also noted changes in multiple lymphocyte compartments within the bone marrow, thymus, spleen and peripheral blood (Supplementary Fig. 6). In an ageing mouse colony, *Sca1-Bcl6$^{floxed}$* mice exhibit significantly shortened lifespan compared to wild-type (WT) littermates due to the development of B-cell lymphomas in 50% (21/42) of mice (Figure 4a). These lymphomas are manifested as expanded and confluent white pulp nodules composed of pleomorphic large B-cells resulting in splenomegaly (Figure 4b–c). The lymphoma cells lack Bcl6 expression but express hallmarks of B-cell identity, such as the Pax5 transcription factor (Figure 4d). These results therefore provide the first indication that transient expression of the *Bcl6* oncogene in HSPCs can induce aggressive malignancies of mature B-cells.

## Sca1-Bcl6$^{floxed}$ DLBCL tumors resemble post-GCB cells

We next sought to evaluate the relationship of this model with the trends observed in human DLBCL by assessing whether these malignancies resembled cells at a GCB or post-GCB (ABC/plasmablast) stage of differentiation. Using transcriptome analysis, we identified 750 significantly repressed and 720 significantly induced genes (T-test FDR <0.25) in tumor-bearing spleens from *Sca1-Bcl6$^{floxed}$* mice compared to spleens from WT mice, including increased expression of multiple genes with roles in oncogenesis (Figure 5a, Supplementary Data 1). However, BCL6 target genes were not significantly repressed within these tumors (GSEA FDR = 1.000, Figure 5b). This is consistent with our findings from human DLBCL tumors, and further supports a hit-and-run role for Bcl6 in generating these aggressive B-cell malignancies. Human DLBCL tumors can be reliably classified into GCB or ABC subtypes by gene expression profiling[1], and this can be achieved in murine models due to recent mapping of the transcriptional signatures of normal murine B-cell development[29,30]. Using

the broad transcriptional signature of tumors from *Sca1-Bcl6*$^{floxed}$ mice compared to signatures of normal stages of murine B-cell development, we found these tumors to most significantly align with the post-germinal center ABC/plasmablast stage of differentiation (Hypergeometric enrichment P-value = 0.028, FDR = 0.017, Figure 5c). These changes included increased transcript abundance of the post-germinal center transcription factors *Irf4*, *Prdm1* and *Xbp1* (Figure 5d). In addition, immunohistochemical staining of tumors from *Sca1-Bcl6*$^{floxed}$ mice showed no staining for the GCB marker PNA, but strong staining for Irf4 and the ABC-like DLBCL marker FoxP1 (Figure 5e). These data therefore demonstrate that transient expression of the Bcl6 oncogene within HSPCs of *Sca1-Bcl6*$^{floxed}$ mice is capable of inducing aggressive Bcell tumors that align with a differentiation stage comparable to human ABC-like DLBCL.

## Evidence for HSPCs as lymphoma-initiating cells

To exclude the potential contribution to lymphomagenesis of persistent ectopic Bcl6 expression in the mature B cell compartment we crossed the *Sca1-Bcl6*$^{floxed}$ mice with an mb1-Cre mouse strain. The resulting strain, *Sca1-Bcl6*$^{\Delta}$, maintains expression of *Bcl6* under the *Sca1* promoter in HSPCs, but deletes the exogenous floxed *Bcl6* cDNA upon B-lineage commitment via Cre-recombinase driven by the promoter from mb1 locus encoding the immunoglobulin-associated alpha chain Cd79a (Figure 6a). To validate the efficient deletion of the exogenous floxed Bcl6 cDNA, we sorted B220$^{+}$ cells from bone marrow of young *Sca1-Bcl6*$^{\Delta}$ mice and cultured under conditions to allow the isolation and expansion of a pure population of B220$^{+}$c-Kit$^{+}$ proB cells. Southern blot analysis of DNA from these cells confirmed uniform and efficient deletion of the exogenous floxed *Bcl6* cDNA at the pro-B stage, and therefore all subsequent stages, of B-cell differentiation (Figure 6b, Supplementary Fig. 7).

Importantly, *Sca1-Bcl6*$^{\Delta}$ mice recapitulate the phenotype observed in *Sca1-Bcl6*$^{floxed}$ mice. These mice have a shortened survival compared to wild-type littermates (Figure 6c) due to aggressive B-cell malignancy in 56.50% (13/24) of mice, manifesting as splenomegaly resulting from complete effacement of normal architecture by diffuse B-cell infiltration (Figure 6d–e). Malignant B-cells are primarily IgM$^{+}$ but show evidence of heavy-chain class-switch in a subset of tumors (Supplementary Fig. 8). These mice also showed infiltration of malignant cells into the liver and bone marrow, resulting in disruption of normal architecture (Figure 6e). Tumors showed increased clonality of immunoglobulin rearrangements (Supplementary Fig. 9), and significant similarity to *Sca1-Bcl6*$^{floxed}$ tumors at the transcriptional level (Supplementary Fig. 10). In line with *Sca1-Bcl6*$^{floxed}$ tumors, *Sca1-Bcl6*$^{\Delta}$ tumors also expressed markers of B-cell identity and a post-germinal center stage of differentiation (Supplementary Fig. 10). To identify the tumor repopulating cells for *Sca1-Bcl6*$^{\Delta}$ lymphomas, we purified LSK and B220$^{+}$ cells and transplanted them into sub-lethally irradiated syngeneic recipient mice. Each of the mice transplanted with LSK cells developed a DLBCL that was phenotypically identical to the primary disease. In contrast, the B220+ cells were incapable of inducing lymphoma in secondary recipients, even when injected in a 10 or 100-fold higher number than the LSK cells (Supplementary Table 2), despite these cells being able to transplant disease in other models[31]. This indicates that Bcl6-induced DLBCL in this model is propagated by transformed HSPC cells but not

mature tumor cells, and confirm that activity of the *Bcl6* oncogene restricted to HSPCs can induce malignancies in mice that are of a post-germinal center stage of differentiation.

**p53 loss in Sca1-Bcl6$^{floxed}$ mice does not promote lymphoma—**Prior studies have shown that Bcl6 acts in myeloid leukemia stem cells and normal B-cell development by inactivation of p53 and subsequent sensing of DNA damage[11,12]. We therefore evaluated whether the inactivation of p53 and the accumulation of secondary genetic alterations play a role in this model by crossing *Sca1-Bcl6$^{floxed}$* mice with heterozygous (p53$^{+/-}$) or homozygous (p53$^{-/-}$) p53 knock-out mice. This showed that decrease or loss of p53 did not facilitate B-cell lymphoma development, but instead resulted in myeloid neoplasia (Supplementary Fig. 11). B-cell lymphoma development in *Sca1-Bcl6$^{floxed}$* mice therefore may not proceed via suppression of p53 and the accumulation of secondary genetic lesions as observed in myeloid malignancies. Exome sequencing of *Sca1-Bcl6$^{floxed}$* tumors revealed the accumulation many somatic variants, but none that were recurrent across tumors or that had been implicated in lymphomagenesis. We also noted an absence of DNA copy number abnormalities that are associated with p53 deregulation in lymphoma[32] (Supplementary Fig. 12, Supplementary Data 2). Together these data show that p53 dysfunction in combination with Bcl6 promotes myeloid malignancy, as previously described[11].

## Epigenetic changes associated with transient Bcl6 expression

Bcl6 mediates suppression of target genes via recruitment of factors that epigenetically modify chromatin. We hypothesized that this may elicit changes in DNA methylation, a mark that is capable acting in gene silencing memory[33], and therefore investigated whether this may be a potential mechanism for *Bcl6*-mediated hit-and-run oncogenesis. Using genome-wide DNA methylation profiling by reduced representation bisulfite sequencing (RRBS) of populations of HSPCs and mature B-cells from wild type and *Sca1-Bcl6$^{\Delta}$* mice, we identified broad epigenetic changes associated with the expression of *Bcl6* in HSPCs (Figure 7a, Supplementary Data 3). Importantly, a significant subset of these changes were found to be maintained from HSPCs to mature B-cells in *Sca1-Bcl6$^{\Delta}$* mice, resulting in these populations being epigenetically more similar to each other than to their comparative population from wild-type mice (Figure 7a–b). Genomic regions found to have significant changes in DNA methylation contained a significantly higher representation of Bcl6 DNA binding motifs compared to those regions that showed no change in DNA methylation (Figure 7c), but were not significantly enriched for genes found to be bound by Bcl6 using ChIP-seq analysis of mature human B-cells[25] (Hypergeometric enrichment FDR = 0.990). This trend was particularly notable in genes found to be hypermethylated in *Sca1-Bcl6$^{\Delta}$* HSPCs and mature B-cells compared to their wild type counterparts. Intriguingly, concordantly hypermethylated genes in HSPCs and mature B-cells from *Sca1-Bcl6$^{\Delta}$* mice were significantly enriched for gene sets associated with murine and human stem cells and poor outcome in human DLBCL (Figure 7d; Supplementary Table 3). Although this analysis was not performed on post-GCB tumor B-cells, we suggest that these results implicate epigenetic alterations that are imparted during transient Bcl6 expression and maintained thereafter as a mechanism for Bcl6 hit-and-run oncogenesis.

## Discussion

It is well established that cancer arises via the stepwise acquisition of somatic alterations that transition clones via one or more pre-malignant states to a malignant state[34]. But deconvolution of the stepwise events taking place during tumor cell evolution is difficult because of the many genetic alterations that become clonally dominant by the time of their interrogation within the clinically manifested tumor, and the multitude of avenues by which any given tumor can evolve. However, research within the cancer stem cell field has led to a growing appreciation for the potential of oncogenic events to be acquired by tumor cell precursors that exist at an earlier differentiation state than the evolved tumor clone[16,35]. While several prior studies have implicated aberrations within HSPCs as important for driving neoplasms of mature B-cells including CLL[17] and FL[13,36], there has not yet been any suggestion that a similar mechanism may be relevant for DLBCL.

By high-resolution genomic analysis of a large number of human DLBCL tumors, we identified DNA copy number gain of 3q27.2 as being important for DLBCL disease biology because of its significant association with adverse outcome and the aggressive ABC-like disease subtype. Like many SCNAs, the minimal region of significant copy number gain on 3q27.2 contained multiple genes. However, the propensity for the *BCL6* oncogene to be targeted by other genetic alterations in DLBCL, such as point mutations and translocation that deregulate its expression[24,6,37], and the mutual exclusivity of BCL6 translocation with its DNA copy number gain, strongly suggested that *BCL6* was a target of this alteration. Integrative analysis of DNA copy number and transcriptome data further suggested that *BCL6* may act in a hit-and-run fashion, wherein its transient over-expression is promoted by genetic alteration but is not maintained in the evolved tumor cell population.

The *BCL6* gene encodes a POZ/Zinc finger transcriptional repressor that regulates gene expression in a manner that is independent of the location of its binding site, via interaction with co-repressors that recruit histone deacetylases and induce epigenetic remodeling and heterochromatin formation[38]. The best characterized role of BCL6 is within normal B-cell activation and germinal center formation, but its expression is not limited to the B-cell compartment and contributes to the normal function of other cellular populations including multiple T-cell subsets and macrophages[22,39–42]. Recently, a role for BCL6 has also been described within progenitor populations during normal hematopoiesis and in the function of cancer stem cells in myeloid leukemia[11,12,26]. Laurenti *et al.* identified *BCL6* transcript expression in human cord blood HSPCs and found that extinguishing this expression resulted in decreased numbers of mature B-cells, suggesting that Bcl6 expression in HSPCs has a critical role in early B-cell development[26]. Our observation of BCL6 expression in a subset of adult human bone marrow HSPCs, and the associated coordinate repression of BCL6 target genes within these cells, adds to growing evidence for a role of BCL6 in normal early human hematopoiesis. It is therefore plausible that genetic deregulation of BCL6 within these cells by DNA copy number gain or translocation could potentially contribute to the pathogenesis of diseases like leukemia and lymphoma. The oncogenic function of BCL6 may also potentially act in the germinal center, with repression occurring thereafter as a result of processes that occur during normal B-cell differentiation. However, transgenic mice expressing *Bcl6* in mature B-cells under the control of the immunoglobulin

promoter required repeated immunization in order to induce lymphoma[43]. In contrast, immunization was not required to induce lymphoma in *Sca1-Bcl6$^{floxed}$* or *Sca1-Bcl6$^{\Delta}$* mice that expressed *Bcl6* transiently within HSPCs, suggesting that the lymphomagenic potential of *Bcl6* may be stronger within this compartment.

As patterns of BCL6 expression are tightly conserved between humans and mice[27], we modeled this hypothesis by expression of BCL6 transiently within murine HSPCs. The lymphomas that developed in these mice were histologically similar to human DLBCL and transcriptionally similar to the differentiation stage of human ABC-like tumors in which *BCL6* SCNAs were identified. Also in line with our observations in human ABC-like DLBCL, the murine lymphomas did not show over-expression of Bcl6 or coordinate target gene repression, suggesting that Bcl6 was acting in a hit-and-run fashion. This was further confirmed by Cre-mediated deletion of the exogenous Bcl6 allele upon B-cell maturity, which did not alter the development or phenotype of these tumors. Because BCL6 can alter the activity of the p53 tumor suppressor gene[44], and this is critical for leukemia stem cell survival in chronic myeloid leukemia[12], we investigated whether the lymphomagenic potential of transient Bcl6 expression was via repression of p53 and the accumulation of secondary genetic alterations. Crossing of *Sca1-Bcl6$^{floxed}$* mice with p53$^{+/-}$ and p53$^{-/-}$ mice did not accelerate the development of DLBCL, but instead promoted myeloid malignancies. These myeloid malignancies may have masked the development of lymphoid malignancies that have longer latency periods, but shows that combined deregulation of Bcl6 and p53 has a relatively more profound role in myeloid compared to lymphoid tumorigenesis. Furthermore, lymphomas from *Sca1-Bcl6$^{floxed}$* mice did not show patterns of DNA copy number change or somatic mutation that are associated with p53 malfunction in lymphoma[32], indicating that BCL6 may be acting hit-and-run oncogenesis within this model in a manner that is not promoted by p53 dysfunction. While the exact mechanism by which transient Bcl6 expression promotes oncogenesis remains to be defined, we found some evidence that suggests Bcl6 may function by inducing epigenetic changes that were conserved from HSPCs to mature B-cells in *Sca1-Bcl6$^{\Delta}$* mice. In addition, the reduced polyclonality of B-cell lymphomas from these mice suggested that there may be additional genetic, epigenetic or microenvironmental factors following VDJ recombination that confer a growth advantage to some clones.

Our observations in the mouse model described here are potentially relevant to human disease, providing some evidence for hit-and-run oncogenesis in ABC-like DLBCLs that could be linked to BCL6. Several lines of evidence support such a linkage. First, genetic aberrations targeting *BCL6* in human DLBCL do not result in its significant over-expression or coordinate repression of its target genes. Second, we observed that genes with conserved Bcl6-induced hypermethylation between HSPC and mature B cells of Sca1-Bcl6$^{\Delta}$ mice are significantly enriched for markers distinguishing human DLBCL patients with fatal/refractory disease from those that are cured. Third, correlation between DLBCL subtype and immunoglobulin heavy-chain isotype highlights an important paradox between the differentiation stage and the isotype of the B-cell receptor. Specifically, despite similar levels of AID expression[10], ABC-like DLBCL do not exhibit evidence of productive heavy chain class-switching[45,46] or ongoing somatic mutation[47]. This suggests that the mature

activated B-cell phenotype of ABC-like DLBCL is disconnected from their less mature immunoglobulin genotypes. Finally, in patients presenting with relapsed/refractory disease, DLBCL subtypes exhibit differential sensitivity to salvage chemotherapy regimens as part of autologous stem and progenitor cell transplantation studies[48], raising the hypothesis that differences in therapy might modulate BCL6 expression through epigenetic mechanisms. While our model of hit-and-run oncogenesis is not compatible with those genes that induce 'oncogene-addiction' (eg. MYC[49]), it may be compatible with other genes that are able to induce long lasting epigenetic changes. Whether a subset of human ABC-like DLBCL tumors have their roots in genetic aberrations arising in early HSPCs or in a more mature stem-like lymphoid subpopulation requires further study. Definitive staging of somatic genetic lesions is not obvious because tumors can exhibit phenotypes of one stage of development but contain translocations from prior stages. For example, while BCL6, MYC, BCL2, and BCL1 translocations are all found in mature B-cell lymphomas, these lesions are distinguished by hallmarks such as immunoglobulin recombination signal sequences and junctional additions at translocation breakpoints, suggesting their distinct derivation from early or late events during B-cell development[44,50]. Separately, evidence for lineage plasticity of mature lymphomas[17,36] complicates inferences of cell-of-origin that are based on either gene expression profiles or genotypes alone. It is therefore unclear as to the specific stage of hematopoietic or B-cell differentiation prior to evolved DLBCL that BCL6 acts, and the hierarchy of genetic events contributing to this disease will only be definitively defined by isolating human hematopoietic and tumor cell precursors from patients and identifying the minimal set of genetic lesions they harbor.

Based upon our observations we propose a model wherein an oncogene may act in a hit-and-run fashion within early tumor cell precursors and is no longer required in the evolved tumor cell progeny. Evolved tumor cells may in turn become reliant on alternate survival pathways that are not present within their precursors. We propose that in human ABC-like DLBCL, genetic alterations of *BCL6* may act in a hit-and-run fashion in early precursors, while evolved tumor cells develop reliance on alternative oncogenic mechanisms such as nuclear-factor κB and B-cell receptor signaling pathways[51–52]. This may provide some explanation for the lower sensitivity of mature ABC-like DLBCL cell-lines to BCL6 inhibition[53], despite the high prevalence of BCL6 gene alterations, and the failure of some modern targeted therapies to clear tumor stem cells, despite being effective agents against evolved tumor cells. As a consequence, targeted treatment strategies may need to be altered to accommodate combinations of agents that target oncogenic pathways that are active at both the early and late stages of tumor development. Our findings therefore have important implications for understanding and therapeutically targeting tumor cells.

## Methods

### Integrative analysis of human DLBCL and normal B-cells

All human data was obtained from public databases and associated with informed consent obtained as part of the original source studies. DNA copy number data was obtained from 609 primary DLBCL tumor specimens[23,32,54–56], and analyzed by GISTIC2[57]. This represented all publicly-available high-resolution (>244,000 markers) at the time this study

was undertaken. Survival annotations were available for 232 CHOP-treated patients[23,32] and 196 R-CHOP treated patients[32,54], as described previously. These groups of patients were analyzed separately due to the improved outcome of patients treated with Rituximab. *BCL6* translocation status determined by fluorescence in situ hybridization in 58 tumors with matched gene expression microarray data[24]. Raw.cel files from Affymetrix U133 plus 2.0 gene expression microarray data of 163 tumors with matched DNA copy number data and/or BCL6 translocation status[23,24,32], and for normal B-cell subsets and DLBCL tumors[58], were RMA normalized using the *ExpressionFileCreator* module of GenePattern[59], and batch corrected using ComBat[60]. Cell of origin (COO) subtype was classified using the Wright algorithm of 140 genes[61]. These subtypes aligned with prior classification[23] for 125/163 cases, with the remaining discordances being transition between GCB and Unclassified or ABC and unclassified, but not GCB and ABC. As a sanity-check, COO was found to significantly stratify survival (Log-rank P-value <0.001, Supplementary Fig. 13). For BCL6 expression, the probe intensities of 5 probes specific for BCL6 were averaged (203140_at, 215990_s_at, 228758_at, 236439_at, 239249_at). Comparison of BCL6 expression between was performed using a 2-sided T-test. Gene set enrichment analysis (GSEA) was performed for Bcl6 target genes using a previously described set of ChIP-validated target genes[25] and GSEA-P software[62].

### Single cell gene expression profiling of human HSPCs

Bone marrow from two healthy donors (1 male, 1 female) were obtained purchased from AllCells, LLC. (Emeryville, CA) and sorted on a FACS ARIA II instrument as PI−, Lin−, CD34+, CD38−. Single cells were sorted into Terasaki plates containing alternating rows of wells containing media or Milteny SuperAmp lysis buffer for a total yield of 96 single cells in lysate per donor; wells containing media were visualized under a microscope to provide visual confirmation that single cells were sorted into each well (Supplementary Fig. 2). For all 4 donors, bulk population samples were further sorted into Terasaki plates in additional groups (no further sort, HSC (CD90+, CD45RA−), MPP (CD90−, CD45RA−) and L-MPP (CD90−, CD45RA+)), containing 200–10,000 cells. All antibodies were obtained from BD Biosciences. The lysates were shipped to Miltenyi Biotec (Auburn, CA) and processed using their SuperAmp microarray service, where they amplified the lysates using their SuperAmp technology and hybridized probes made from amplified material onto SurePrint G3 Human Gene Expression 8×60K Microarrays (Aglient). Raw data were quantile normalized in R (Bioconductor preprocessCore library), and replicate microarray probes (representing the same gene) were log2 transformed and combined as the geometric mean. Single cells were classified as expressing BCL6 if their intensity value was 2 standard-deviations above the mean for that donor.

### Generation of Bcl6 transgenic mouse strains

All animal work has been conducted according to relevant national and international guidelines and it has been approved by the Bioethics Committee of University of Salamanca and by the Bioethics Subcommittee of Consejo Superior de Investigaciones Cientificas (CSIC). The Bcl6 floxed vector was generated by inserting the mouse Bcl6–IRES-eGFP cassette flanked by loxP sites into the ClaI site of the pLy6 vector. The transgene fragment was excised from its vector by restriction digestion with NotI, purified and injected (2

ng/mL) into CBAxC57BL/6J fertilized eggs. Transgenic mice were identified by Southern blot analysis of tail snip DNA after EcoRI digestion, using Bcl6 cDNA to detect the transgene. Two independent transgenic lines were generated and analyzed. *Sca1-Bcl6*$^{floxed}$ mice were bred to mb1-Cre (*Cd79*$^{atm1(cre)Reth}$) mice to generate *Sca1-Bcl6*$^{\Delta}$ mice. Upon signs of disease, mice were sacrificed and subjected to standard necropsy procedures. All major organs were examined under the dissecting microscope. Tissue samples were taken from homogenous portions of the resected organ and fixed immediately after excision. Differences in Kaplan-Meier survival plots of transgenic and WT mice were analyzed using the log-rank (Mantel-Cox) test. To test the rearrangement in *Sca1-Bcl6*$^{\Delta}$ mice different samples were analyzed by Southern blot analysis after EcoRI digestion and using *Bcl6* cDNA as probe. PCR was also used for confirmation, with the following primers: ClaI-F2, 5'-TATAAATCTGGCTTGATCAGG-3' and ClaI-R2, 5'-CTGAGGAATTCATGTCTGCC-3'.

### Generation of Bcl6 floxed *p53*$^{-/-}$ and Bcl6 floxed *p53*$^{+/-}$ mice

The heterozygous *p53*$^{+/-}$ mice[63] have been described previously. Heterozygous *p53*$^{+/-}$ mice were bred to *Bcl6 floxed* mice to generate compound heterozygotes. F1 animals were crossed to obtain null *p53*$^{-/-}$ mice hemyizygous for *Sca1-Bcl6*$^{floxed}$ mice. Tumor phenotype was assessed in the first generation of p53 heterozygous and homozygous mice.

### Flow cytometry

Nucleated cells were obtained from total mouse bone marrow (flushing from the long bones), peripheral blood, thymus, or spleen. In order to prepare cells for flow cytometry, contaminating red blood cells were lysed with RCLB lysis buffer and the remaining cells were then washed in PBS with 1% FCS. After staining, all cells were washed once in PBS with 1% FCS containing 2 mg/mL propidium iodide (PI) to allow dead cells to be excluded from both analyses and sorting procedures. The samples and the data were acquired in an AccuriC6 Flow Cytometer and analyzed using Flowjo software. Specific fluorescence of FITC, PE, PI and APC excited at 488 nm (0.4 W) and 633 nm (30 mW), respectively, as well as known forward and orthogonal light scattering properties of mouse cells were used to establish gates. Nonspecific antibody binding was suppressed by preincubation of cells with CD16/CD32 Fc-block solution (BD Biosciences). For each analysis, a total of at least 50,000 viable (PI−) cells were assessed.

The following antibodies were used for flow cytometry: anti-B220 (RA3-6B2), CD3 (145-2C11), CD4 (RM4-5, 1:500), CD8a (53-6.7, 1:500), CD11b/Mac1 (M1/70, 1:200), CD19 (1D3), CD49b (DX5), CD117/c- Kit (2B8, 1:200), CD127/IL-7Rα (A7R34, 1:50), CD135/Flt3 (A2F10.1), Flt3 (A2B10), Ly-6G/Gr1 (RB6-8C5), IgD (11-26c.2a), IgM (R6-60.2), Sca1/Ly-6A/E (E13-161.7, 1:50), CD21 (7G6), CD22 (Lyb-8.2)(Cy34.1), CD23 (B3B4), CD25 (PC61), CD48 (HM48-1), CD150 (TC15-12F12.2) and Ter119 (TER119) antibodies. Unspecific antibody binding was suppressed by preincubation with CD16/CD32 (2.4G2) Fc-block solution (PharMingen). The different hematopoietic progenitors and B cell stages were defined by flow cytometry as shown in Supplementary Figure 6. All antibodies were purchased from BD Biosciences. All antibodies were used at a 1:100 dilution unless otherwise indicated.

### V(D)J recombination

Immunoglobulin rearrangements were amplified by PCR using the primers listed in Supplementary Table 4[49].. Cycling conditions consisted of an initial heat-activation at 95°C followed by 31–37 cycles of denaturation for 1m at 95°C, annealing for 1m at 65°C for heavy chains or 62°C for light chains, and elongation for 1m45s at 72°C. This was followed by a final elongation for 10m at 72°C. To determine the DNA sequences of individual V–(D)J rearrangements, the PCR fragments were isolated from the agarose gel and cloned into the pGEM-Teasy vector (Promega); the DNA inserts of at least ten clones corresponding to the same PCR fragment were then sequenced.

### Immunohistochemistry (IHC)

Tissue samples were taken from homogenous portions of the resected organ by the pathologist and fixed immediately after excision. Samples of each organ were processed into paraffin, sectioned and examined histologically including routinely standard haematoxylin and eosin, and immunohistochemical techniques. Transgenic mice samples were sectioned, dewaxed, and heated in 10 mmol/L sodium citrate buffer for 30 min. Slides were incubated with primary antibodies. The antibodies used included: Bcl6 (N-3, Santa Cruz, 1:100); Pax5 (C-20, Santa Cruz, 1:125), CD21 (A-3, Santa Cruz, 1:125); IgG (A85-I, BD Biosciences, 1:20). Samples were centrally reviewed by a panel of pathologists and diagnosed using uniform criteria based on clinical, histological, immunophenotypical, and molecular characteristics. For comparative studies, age-matched mice were used.

### Analysis of germinal-center formation

Sheep red blood cells ($1–2 \times 10^8$ cells) were injected into the peritoneum of control, *Sca1-Bcl6$^{floxed}$* mice and *Sca1-Bcl6$^{\Delta}$* mice. Ten days later, the spleens were analyzed by immunohistochemistry. The spleens of control, *Sca1-Bcl6$^{floxed}$* mice and *Sca1-Bcl6$^{\Delta}$* mice were isolated, embedded in OCT compound (Sakura) and snap-frozen on dry ice. Cryosections of the spleen were stained with a FITC–anti-IgD antibody (1:100 dilution, BD Biosciences) and biotinylated PNA (1:100 dilution, clone B-1075, Vector Laboratories). FITC–anti-IgD was detected with an alkaline-phosphatase-coupled anti-FITC antibody (Roche), which was visualized by incubation with Fast Red (Sigma). Biotinylated PNA was detected with horseradish peroxidase-conjugated streptavidin (Zymed) followed by incubation with diaminobenzidine (DAB; Sigma).

### Pro-B cell culture

Iscove's modified Dulbecco's medium supplemented with 50 μM β-mercaptoethanol, 1 mM L-glutamine, 2% heat-inactivated fetal calf serum and 0.03% (w/v) primatone RL (Sigma) was used for Pro-B cell-culture experiments. Pro-B cells isolated by MACS-sorting for $B220^+$ (Milteny Biotec) from bone marrow of 2-week-old mice were cultured on -irradiated ST2 cells in IMDM medium containing IL-7 (R&D Systems). The cells were maintain in culture for 1 week and then collected for Southern blot experiments.

### Gene expression microarray analysis of murine tumors

Tumor-bearing spleens were harvested from *Sca1-Bcl6*$^{floxed}$ and *Sca1-Bcl6*$^{\Delta}$ mice and healthy spleens were harvests from control mice; cells were not sorted prior to RNA extraction for this analysis. Total RNA was isolated in two steps using TRIzol (Life Technologies) followed by Rneasy Mini-Kit (Qiagen) purification following the manufacturer's RNA Clean-up protocol with the optional On-column DNase treatment. The integrity and the quality of the RNA were verified by electrophoresis and its concentration measured. Samples were analyzed using Affymetrix Mouse Genome 430 2.0 arrays. Data was normalized as described above, and data sets containing the WT and either the *Sca1-Bcl6*$^{floxed}$ or *Sca1-Bcl6*$^{\Delta}$ samples were filtered separately by median absolute deviation to derive the most variably expressed 10,000 genes for each comparison for use in GSEA and differential gene expression analysis. Differential gene expression analysis was performed using the *ComparativeMarkerSelection* module of GenePattern[59] with T-test statistic, and correcting for multiple hypothesis testing using 1000 permutations. Genes were deemed to be significantly differentially expressed with a False Discovery Rate (FDR) Q-value <0.25 and a fold change ≥1.25. Differentially expressed genes were tested for enrichment of genes associated with normal murine B-cell differentiation states. Gene expression signatures that are specifically up-regulated in pre/pro-B cells, transitional B-cells, follicular or marginal zone B-cells, germinal center B-cells, plasmablasts, and plasma cells were assessed for their overlap with that were up-regulated within tumor specimens using hypergeometric enrichment analysis[29,30]. The relative over-abundance of a specific differentiation state within a tumor sample compared to a wild-type spleen is therefore detected by significant enrichment of the signature that is definitive of that differentiation state.

### DNA methylation profiling

**EpiQuest library construction—**EpiQuest libraries were prepared from 200–500 ng mouse genomic DNA obtained from primary cells ($Sca1^{+}Lin^{-}$ cells from bone marrow, and $B220^{+}$ cells and $Gr1^{+}Mac1^{+}$ from peripheral blood) purified from *Sca1-Bcl6*$^{\Delta}$ mice and/or wild-type mice. DNA were pooled from 6–9 mice to provide one pooled replicate per condition, as performed previously[65,66].The DNA was digested with 60 units of TaqI and 30 units of MspI (New England Biolabs) sequentially. Size-selected TaqI-MspI fragments (40–120 bp and 120–350 bp) were filled in and 3'-terminal-A extended, extracted with a DNA Clean & Concentrator kit (Zymo Research). Ligation to pre-annealed adapters containing 5'-methylcytosine instead of cytosine was performed using the Illumina DNA preparation kit and protocol. Purified, adaptor ligated fragments were bisulphite-treated using the EZ DNA Methylation-Direct Kit (Zymo Research). Preparative-scale PCR (18 cycles) was performed and purified PCR products were subjected to a final size selection on a 4% NuSieve 3:1 agarose gel. SYBR-green-stained gel slices containing adaptor-ligated fragments of 130–210 bp or 210–460 bp in size were excised. Library material was recovered from the gel using a Zymoclean Gel DNA Recovery Kit (Zymo Research) and sequenced on an GAIIx genome analyzer (Illumina), yielding between 43,930,708 and 60,139,994 total reads for each condition.

**Sequence alignments and data analysis—**Sequence reads from bisulphite-treated EpiQuest libraries were identified using standard Illumina base-calling software and then

analyzed using a Zymo Research proprietary computational pipeline. Residual cytosines (Cs) in each read were first converted to thymines (Ts), with each such conversion noted for subsequent analysis. A reference sequence database was constructed from the 36-bp ends of each computationally predicted MspI-TaqI fragment in the 40–220-bp size range. All Cs in each fragment end were then converted to Ts (only the C-poor strands are sequenced in the RRBS process; The converted reads were aligned to the converted reference by finding all 12-bp perfect matches and then extending to both ends of the treated read, not allowing gaps (reverse complement alignments were not considered). The number of mismatches in the induced alignment was then counted between the unconverted read and reference, ignoring cases in which a T in the unconverted read is matched to a C in the unconverted reference. For a given read, the best alignment was kept if the second-best alignment had 2 more mismatches; otherwise the read was discarded as non-unique. The mean CpG coverages ranged between 5–11× and total number of unique mapped reads ranged between 3,378,764–10,295,957 for each condition. The methylation level of each sampled cytosine was estimated as the number of reads reporting a C, divided by the total number of reads reporting a C or T. A bioinformatics pipeline was used to score epigenetic alterations according to strength and significance, and links them to potentially affected genes. To that end, we collected a comprehensive set of regions of interest, which includes promoters, CpG islands and repetitive elements. For each of these regions, the number of methylated and unmethylated CpG observations is determined, and a p-value is assigned using Fisher's exact test. Once all p-values are calculated, multiple-testing correction is performed separately for each region type using the q-value method, which controls the false discovery rate to be below a user-specified threshold (typically 10%). The software pipeline is implemented in Python (alignment processing module) and R (statistical analysis module). Unsupervised clustering was performed on methylation ratio data from each condition using Kendall tau rank correlation.

### Exome sequencing

Next generation sequencing libraries were prepared using NEBnext DNA sample prep kit (New England Biolabs) and exome capture performed using SureSelect XT Mouse All Exon (Agilent). Sequencing was performed using and Genome Analyzer II instrument (Illumina) with 75bp paired-end reads, and one sample per lane. Raw sequences were trimmed by 2bp at their 5’ ends to remove GC bias, and aligned to the murine genome (mm9) using Burrows-Wheeler Aligner (BWA) with default parameters[67]. Somatic mutations and DNA copy number alterations were called with reference to age- and gender-matched controls using VarScan 2.0[68]. Somatic nucleotide variants were discarded if they; (i) did not have a significant p-value ($P \leq 0.05$) determined by VarScan 2.0, (ii) were not determined to be somatic alterations by VarScan 2.0, (iii) were present in ≥10% of reads within the control sample, (iv) were not observed in both strands during sequencing, and (v) were not in the coding region of a gene.

### Bone marrow transplantation experiments

In order to determine the nature of the lymphomagenic cell, bone marrow transplantation experiments were performed. BM LSK or splenic $B220^{+}$ cells were isolated and highly purified from either male *Sca1-Bcl6*$^{\Delta}$ (C57BL/6 x CBA) or male wild-type mice (C57BL/6

x CBA,). Mice were 12 months old. The sorting purity of these cells was analyzed by FACS and determined to be over 98%. In each cohort these cells were injected intravenously into sublethally irradiated (4 Gy) secondary recipient 12-week old male syngenic mice (C57BL/6 x CBA). Diseased recipient mice were sacrificed and assessed for B-cell lymphoma development.

### Bcl6 DNA binding sequence motif analysis

DNA sequences were obtained for genomic regions that were (i) significantly hypermethylated in both HPCs and mature B-cells from *Sca1-Bcl6*$^{\Delta}$ mice compared to identical populations from wild-type mice, (ii) significantly hypomethylated in both HPCs and mature B-cells from *Sca1-Bcl6*$^{\Delta}$ mice compared to identical populations from wild-type mice, and (iii) were not differentially methylated between *Sca1-Bcl6*$^{\Delta}$ mice populations and those from wild-type mice (Supplementary Data 3). Genomic regions were searched for BCL6 DNA binding sequence motifs (TTTNNNGNNATNCTTT)[69] using the CisFinder[70] identify motifs function, with an FDR of 0.2 and a matching threshold of 0.75..

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We are grateful to Dr. Meinrad Busslinger for continuous and generous help and ideas over the years with this project, Dr. Takeshi Tokuhisa for the mouse Bcl6 cDNA, Prof. Michael Reth for the mb1-cre mice, and Dr. E. Dzierzak for the Sca1 promoter. Research in ISG group is supported partially by FEDER and by MICINN (SAF2009-08803 and SAF2012-32810), Junta de Castilla y León (CSI13A08 and proyecto Biomedicina 2009-2010), MEC OncoBIO Consolider-Ingenio 2010 (Ref. CSD2007-0017), NIH grant (R01 CA109335-04A1) and by Group of Excellence Grant (GR15) from Junta de Castilla y Leon. MRG and AA are Special Fellows of the Leukemia and Lymphoma Society. Funding for single cell human HSPC studies was provided by NIH grant (U01HL099999) to RM and AA. AM is supported by NCI R01 CA104348, the Chemotherapy Foundation, the Sam Waxman Cancer Research Foundation, the G&P Foundation and is a Leukemia and Lymphoma Society Scholar. Research at C.C.´s lab is partially supported by FEDER, Fondo de Investigaciones Sanitarias (PI080164), Proyectos Intramurales Especiales (CSIC) and Junta de Castilla y León (SA060A09 and proyecto Biomedicina 2009–2010). AO research is supported by a grant from the Instituto de Salud Carlos III, Ministerio de Sanidad y Consumo, Madrid, Spain (IISCIII-RTICC RD06/0020/0035-FEDER).

## References

1. Alizadeh A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000; 403:503–511. [PubMed: 10676951]
2. Morin R, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011; 476:298–303. [PubMed: 21796119]
3. Pasqualucci L, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature. 2011; 471:189–195. [PubMed: 21390126]
4. Shaffer A, Young R, Staudt L. Pathogenesis of human B cell lymphomas. Ann. Rev. of Immunol. 2012; 30:565–610. [PubMed: 22224767]
5. Chaganti S, et al. Involvement of BCL6 in chromosomal aberrations affecting band 3q27 in B-cell non-Hodgkin's lymphoma. Genes. Chrom. Cancer. 1998; 23:323–327. [PubMed: 9824205]
6. Ye B, et al. Alterations of a zing finger-encoding gene, BCL-6, in diffuse large B-cell lymphoma. Science. 1993; 262:747–750. [PubMed: 8235596]
7. Bihui H, et al. The BCL-6 proto-oncogene controls germinal-centre formation and Th2-type inflammation. Nat. Genet. 1997; 16:161–170. [PubMed: 9171827]

8. Dent A, Shaffer A, Yu X, Allman D, Staudt L. Control of inflammation, cytokine expression, and germinal center formation by BCL-6. Science. 1997; 276:589–592. [PubMed: 9110977]
9. Lossos I, et al. Expression of a single gene, BCL-6, stronly predicts survival in patients with diffuse large B-cell lymphoma. Blood. 2001; 98:945–951. [PubMed: 11493437]
10. Lossos I, Akasaka T, Martinez-Climent J, Siebert R, Levy R. The BCL6 gene in B-cell lymphomas with 3q27 translocations is expressed mainly from the rearranged allele irrespective of the partner gene. Leukemia. 2003; 17:1390–1397. [PubMed: 12835729]
11. Duy C, et al. BCL6 enables Ph+ acute lymphoblastic leukaemia cells to survive BCR-ABL1 kinase inhibition. Nature. 2011; 473:384–388. [PubMed: 21593872]
12. Hurtz C, et al. BCL6-mediated repression of p53 is critical for leukemia stem cell survival in chronic myeloid leukemia. J. Exp. Med. 2011; 208:2163–2174. [PubMed: 21911423]
13. Weigert O, Weinstock DM. The evolving contribution of hematopoietic progenitor cells to lymphomagenesis. Blood. 2012; 120:2553–2561. [PubMed: 22869790]
14. Green M, et al. Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. Blood. 2013; 121:1604–1611. [PubMed: 23297126]
15. Alizadeh A, Majeti R. Surprise! HSC are aberrant in chronic lymphocytic leukemia. Cancer Cell. 2011; 20:135–136. [PubMed: 21840478]
16. Jan M, et al. Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. Sci. Transl. Med. 2012; 4
17. Kikushige Y, et al. Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. Cancer Cell. 2011; 20:246–259. [PubMed: 21840488]
18. Welch J, et al. The origin and evolution of mutations in acute myeloid leukemia. Cell. 2012; 150:264–278. [PubMed: 22817890]
19. Weinstein B. Addiction to oncogenes – the achilles heal of cancer. Science. 2002; 297:63–64. [PubMed: 12098689]
20. Zack T, et al. Pan-cancer patterns of somatic copy number alteration. Nat. Genet. 2013; 45:1134–1140. [PubMed: 24071852]
21. Green M, et al. Integrative analysis reveals selective 9p24.1 amplification, increased PD-1 ligand expression, and further induction via JAK2 in nodular sclerosing Hodgkin lymphoma and primary mediastinal large B-cell lymphoma. Blood. 2010; 116:3268–3277. [PubMed: 20628145]
22. Basso K, Dalla-Favera R. Roles of BCL6 in normal and transformed germinal center B-cells. Immunol. Rev. 2012; 247:172–183. [PubMed: 22500840]
23. Lenz G, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. Proc. Natl. Acad. Sci. 2008; 105:13520–13525. [PubMed: 18765795]
24. Iqbal J, et al. Distinctive patterns of BCL6 molecular alterations and their functional consequences in different subgroups of diffuse large B-cell lymphoma. Leukemia. 2007; 21:2332–2343. [PubMed: 17625604]
25. Basso K, et al. Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal center B cells. Blood. 2010; 115:975–984. [PubMed: 19965633]
26. Laurenti E, et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. Nat. Immunol. 2013; 14:756–763. [PubMed: 23708252]
27. Shay T, et al. Convervation and divergence in the transcriptional programs of the human and mouse immune systems. Proc. Natl. Acad. Sci. USA. 2013; 110:2946–2951. [PubMed: 23382184]
28. Miles C, Sanchez MJ, Sinclair A, Dzierzak E. Expression of the Ly-6E.1 (Sca-1) transgene in adult hematopoietic stem cells and the developing mouse embryo. Development. 1997; 124:537–547. [PubMed: 9053329]
29. Green M, et al. Signatures of murine B-cell development implicate Yy1 as a regulator of the germinal center-specific program. Proc. Natl. Acad. Sci. USA. 2011; 108:2873–2878. [PubMed: 21282644]
30. Romero-Camarero I, et al. Germinal center protein HGAL promotes lymphoid hyperplasia and amyloidosis via BCR-mediated Syk activation. Nat. Comm. 2013; 4

31. Adams J, Strasser A. Is tumor growth sustained by rare cancer stem cells or dominant clones. Cancer Res. 2008; 68:4018–4021. [PubMed: 18519656]
32. Monti S, et al. Integrative analysis reveals an outcome-associated and targetable pattern of p53 and cell cycle deregulation in diffuse large B cell lymphoma. Cancer Cell. 2012; 22:359–372. [PubMed: 22975378]
33. Raynal N, et al. DNA methylation does not stably lock gene expression but instead serves as a molecular mark for gene silencing memory. Cancer Res. 2012; 72:1170–1181. [PubMed: 22219169]
34. Nowell P. The clonal evolution of tumor cell populations. Science. 1996; 194:23–28. [PubMed: 959840]
35. Huntlet B, Gilliand D. Leukaemia stem cells and the evolution of cáncer stem cell research. Nat. Rev. Cancer. 2005; 5:311–321. [PubMed: 15803157]
36. Hart J, et al. Transmission of a follicular lymphoma by allogeneic bone marrow transplantation – evidence to support the existence of lymphoma progenitor cells. Br. J. Haematol. 2006; 136:163–172.
37. Pasqualucci L, Migliazza A, Basso K, Houldsworth J, Chaganti R, Dalla-Favera R. Mutations of the BCL6 proto-oncogene disrupts its negative auto-regulation in diffuse large B-cell lymphoma. Blood. 2003; 101:2914–2923. [PubMed: 12515714]
38. Dent A, Vanaswala F, Toney L. Regulation of gene expression by the poto-oncogene BCL-6. Crit. Rev. Oncol. Hematol. 2002; 41:1–9. [PubMed: 11796228]
39. Chung Y, et al. Follicular regulatory T cells expressing Foxp3 and Bcl-6 suppress germinal center reactions. Nat. Med. 2011; 17:983–988. [PubMed: 21785430]
40. Nurieva R, et al. Bcl6 mediates the development of T follicular helper cells. Science. 2009; 325:1001–1008. [PubMed: 19628815]
41. Ichii H, et al. Role for Bcl-6 in the generation and maintenance of memory CD8+ T cells. Nat. Immunol. 2002; 3:558–563. [PubMed: 12021781]
42. Toney L, et al. BCL-6 regulates chemokine gene transcription in macrophages. Nat. Immunol. 2000; 1:214–220. [PubMed: 10973278]
43. Phan R, Dalla-Favera R. The BCL6 proto-oncogene suppresses p53 expression in germinal-center B cells. Nature. 2004; 432:635–639. [PubMed: 15577913]
44. Cattoretti, et al. Deregulated BCL6 expression recapitulates the pathogenesis of human diffuse large B cell lymphoma in mice. Cancer Cell. 2005; 7:445–455. [PubMed: 15894265]
45. Ruminy P, et al. The isotype of the BCR as a surrogate for the GCB and ABC molecular subtypes in diffuse large B-cell lymphoma. Leukemia. 2011; 25:681–688. [PubMed: 21233831]
46. Lenz G, et al. Aberrant immunoglobulin class switch recombination and switch translocations in activated B cell-like diffuse large B cell lymphoma. J. Exp. Med. 2007; 204:633–643. [PubMed: 17353367]
47. Lossos I, et al. Ongoing immunoglobulin somatic mutation in germinal center B cell-like but not in activated B cell-like diffuse large cell lymphomas. Proc. Natl. Acad. Sci. USA. 2000; 97:10209–10213. [PubMed: 10954754]
48. Thieblemont C, et al. The germinal center/activated B-cell subclassification has a prognostic impact for response to salvage therapy in relapsed/refractory diffuse large B-cell lymphoma: A Bio-CORAL study. J. Clin. Oncol. 2011; 29:4079–4087. [PubMed: 21947824]
49. Jain M, et al. Sustained loss of a neoplastic phenotype by brief inactivation of MYC. Science. 2002; 297:102–104. [PubMed: 12098700]
50. Tsai A, Lu H, Raghavan S, Muschen M, Hsieh C, Lieber M. Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. Cell. 2008; 135:1130–1142. [PubMed: 19070581]
51. Compagno M, et al. Mutations of multiple genes cause deregulation of NF-kB in diffuse large B-cell lymphoma. Nature. 2009; 459:717–720. [PubMed: 19412164]
52. Davis R, et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. Nature. 2010; 463:88–92. [PubMed: 20054396]

53. Cerchietti L, et al. A peptomimetic inhibitor of BCL6 with potent antilymphoma effects in vitro and in vivo. Blood. 2009; 113:3397–3405. [PubMed: 18927431]
54. Scandurra M, et al. Genomic lesions associated with a different clinical outcome in diffuse large B-Cell lymphoma treated with R-CHOP-21. Br. J. Haem. 2010; 151:221–231.
55. Green M, et al. Integrative genomic profiling reveals conserved genetic mechanisms for tumorigenesis in common entities of non-Hodgkin's lymphoma. Genes Chrom. Cancer. 2011; 50:313–326. [PubMed: 21305641]
56. Kato M, et al. Frequent inactivation of A20 in B-cell lymphomas. Nature. 2009; 459:712–716. [PubMed: 19412163]
57. Mermel C, Schumacher S, Hill B, Meyerson M, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011; 12
58. Brune V, et al. Origin and pathogenesis of nodular lymphocyte – predominant Hodgkin lymphoma as revealed by global gene expression analysis. J. Exp. Med. 2008; 205:2251–2268. [PubMed: 18794340]
59. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov J. GenePattern 2.0. Nat. Genet. 2006; 38:500–501. [PubMed: 16642009]
60. Johnson W, Rabinovic A, Li C. Adjusting batch effects in microarray expression data using Empirical Bayes methods. Biostatistics. 2007; 8:118–127. [PubMed: 16632515]
61. Wright G, Tan B, Rosenwald A, Hurt E, Wiestner A, Staudt L. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc. Natl. Acad. Sci. USA. 2003; 100:9991–9996. [PubMed: 12900505]
62. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov J. GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics. 2007; 23:3251–3253. [PubMed: 17644558]
63. Jacks T, et al. Tumor spectrum analysis in p53-mutant mice. Curr. Biol. 1994; 4:1–7. [PubMed: 7922305]
64. Kwon K, et al. Instructive role of the transcription factor E2A in early B lymphopoiesis and germinal center B cell development. Immunity. 2008; 28:751–762. [PubMed: 18538592]
65. Taylor K, et al. Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. Cancer Res. 2007; 67:8511–8518. [PubMed: 17875690]
66. Seisenberger S, et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol. Cell. 2012; 48:849–862. [PubMed: 23219530]
67. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26:589–595. [PubMed: 20080505]
68. Koboldt D, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012; 22:568–576. [PubMed: 22300766]
69. Baron B, et al. BCL6 encodes a sequence-specific DNA binding protein. Genes Chrom. Cancer. 2003; 13:332–224.
70. Sharov A, Ko M. Exhaustive search for over-represented DNA sequence motifs with CisFinder. DNA Res. 2009; 16:261–273. [PubMed: 19740934]

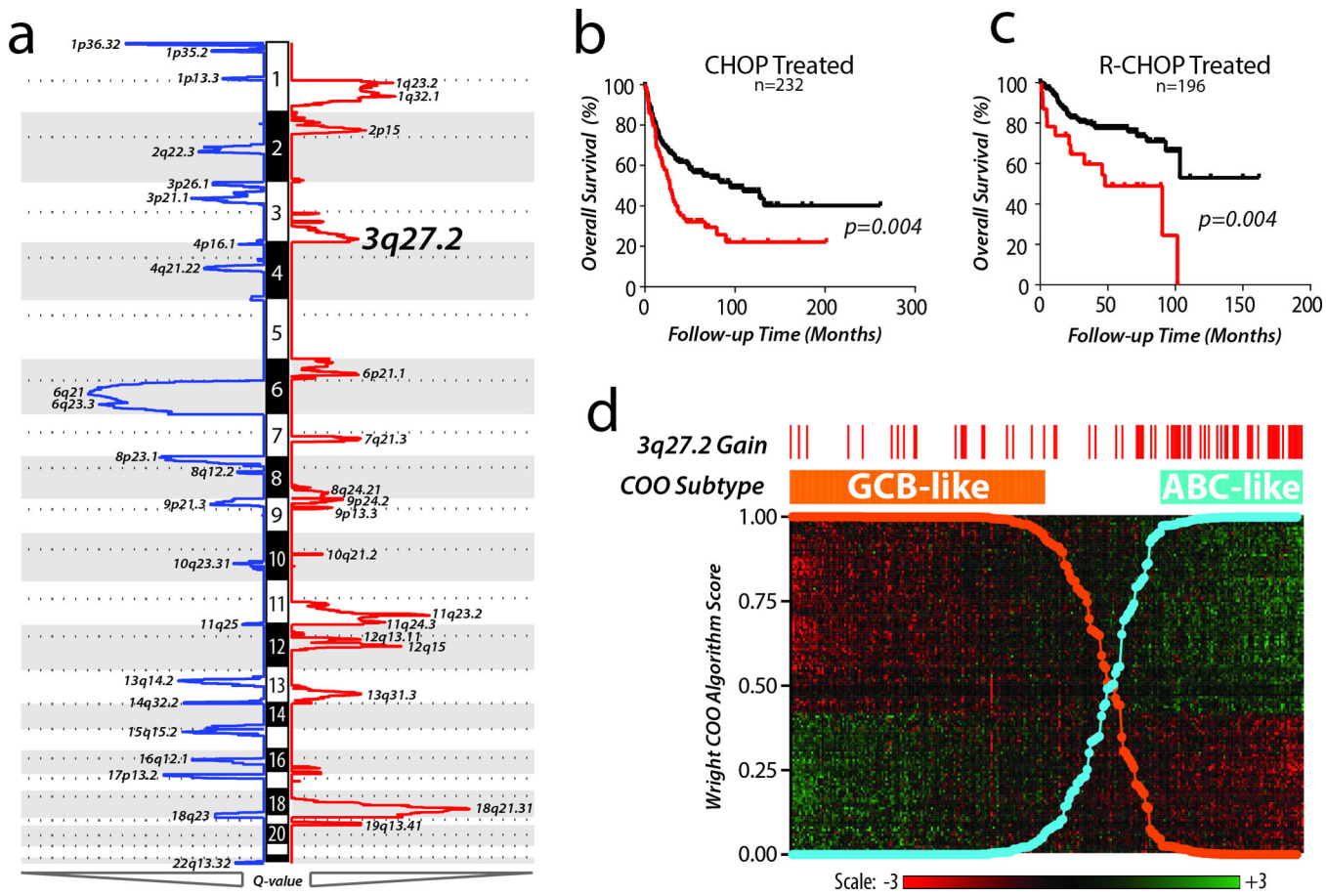


**Figure 1. High resolution DNA copy number profiling of human DLBCL**

**a)** DNA copy number profiles from 609 primary DLBCL tumors were analyzed for significant alterations using the GISTIC algorithm. This algorithm uses the magnitude and frequency of alterations at each position in the genome to assign a GISTIC Q-value, with decreasing values indicating increasing significance. Significant peaks (GISTIC Q-value <0.10) of DNA copy number loss (blue) and gain (red) are annotated with their genomic location. **b)** In 232 patients treated with combination chemotherapy (CHOP) in the absence of Rituximab, presence of 3q27.2 gain (red) was associated with significantly worse overall survival than those with diploid copy number at this region (black). Log-rank P-values were 0.004 and 0.001 for CHOP-treated and R-CHOP-treated cohorts, respectively **c)** Gain of 3q27.2 (red) remained to be associated with significantly worse overall survival compared to diploid 3q27.2 (black) in 196 patients treated with combination chemotherapy plus Rituximab (R-CHOP). **d)** For 249 cases with matched gene expression profiling data, tumors were classified into GCB-like (Orange) and ABC-like subtypes (Blue) based upon the Wright 140 gene algorithm (heat map shown). Gain of 3q27.2, shown by red tick marks for each case in which it was detected, was significantly over-represented in the ABC-like subtype compared to the GCB-like subtype (Fisher P-value = <0.001).

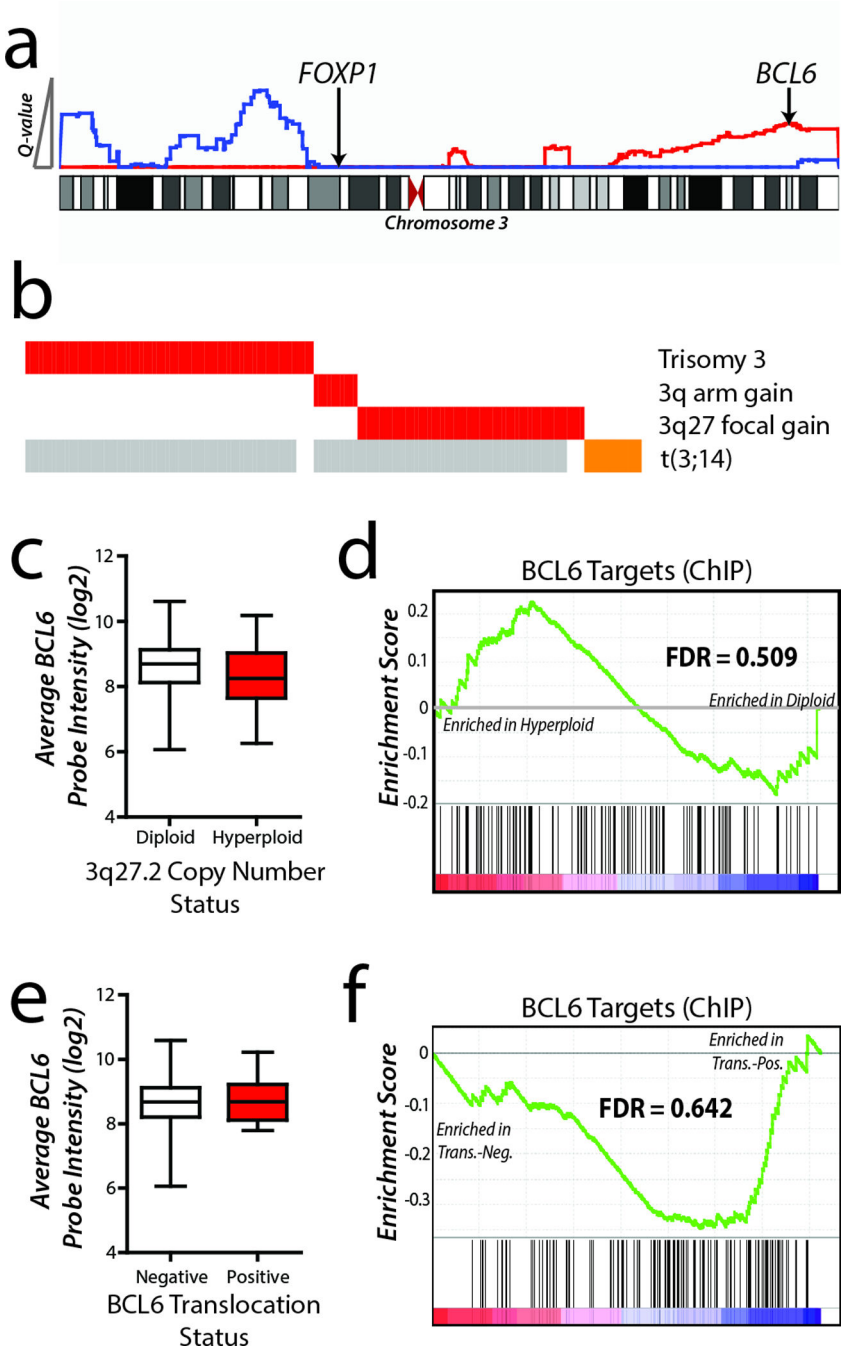


**Figure 2. 3q27.2 gain targets BCL6 but does not alter its expression or target-gene repression**

**a)** GISTIC Q-values are shown for DNA copy number loss (blue) and gain (red) on chromosome 3. The peak of 3q27.2 gain included the *BCL6* oncogene, but not the previously described target *FOXP1*. **b)** Significance of 3q27.2 gain was contributed to by broad gains of chromosome 3 or the 3q arm in 76 tumors, as well as focal gains of 3q27 in 52 tumors. These gains were mutually exclusive to BCL6 translocation in the 21 tumors for which matching DNA copy number and fluorescence in situ hybridization data were available. Grey bars represent data not available. **c)** Increased DNA copy number of *BCL6*

was not associated with increased transcript abundance compared to cases with no gain. Box plots represent the mean +/− the interquartile range with whiskers extending to the minimum and maximum value. **d)** Gene set enrichment analysis (GSEA) of BCL6 target genes showed no significant repression within cases possessing BCL6 DNA copy number gain (GSEA FDR = 0.509). **e)** For the 58 tumors with matching BCL6 translocation status and gene expression profiling data, BCL6 transcript abundance was not increased in tumors with BCL6 translocation compared to tumors without BCL6 translocation. Box plots represent the mean +/− the interquartile range with whiskers extending to the minimum and maximum value. **f)** Tumors with *BCL6* translocation (Trans.-Pos.) showed no significant repression of *BCL6* target genes by GSEA (GSEA FDR = 0.642) compared to tumors without *BCL6* translocation (Trans.-Neg.).

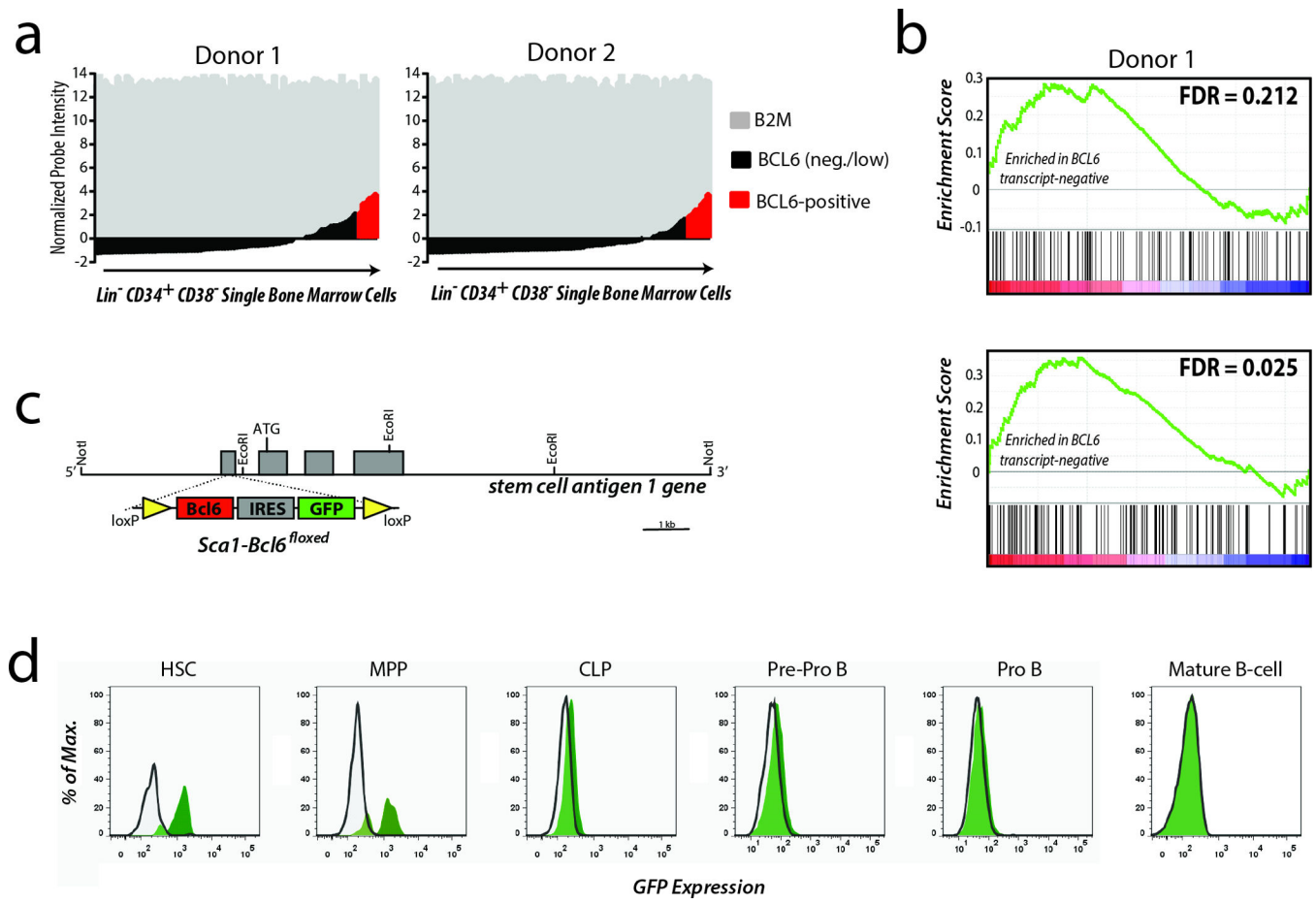


**Figure 3. Expression and activity of BCL6 in human and murine HSPCs**

**a)** Gene expression profiles of single human HSPCs showing expression of the control gene B2M were investigated for expression of BCL6. Positive expression values defined as being 2 standard deviations above the mean were observed in 11/183 single cells, with approximately equal proportions in each donor. **b)** Gene set enrichment analysis of BCL6 target genes in BCL6 transcript-positive cells compared to BCL6 transcript-negative cells in each donor showed an enrichment of target gene expression in BCL6 transcript-negative cells (GSEA FDRs 0.212 and 0.025 for Donor 1 and 2, respectively). This corresponds to repression of target genes in BCL6 transcript-positive cells. **c)** Transient Bcl6 expression within HSPCs was achieved by placing a floxed (yellow loxP sites) Bcl6 cDNA with IRES-GFP reporter under control of the of the promoter for the HSPCspecific gene, stem-cell-antigen 1 (Sca1). **d)** Tracking of the GFP marker for Bcl6 transgene expression during hematopoietic development shows expression is restricted to HSPCs and not in Pro B or mature B-cells. Images are representative of 4 independent experiments. For gating schema, see Supplementary Fig. 3.

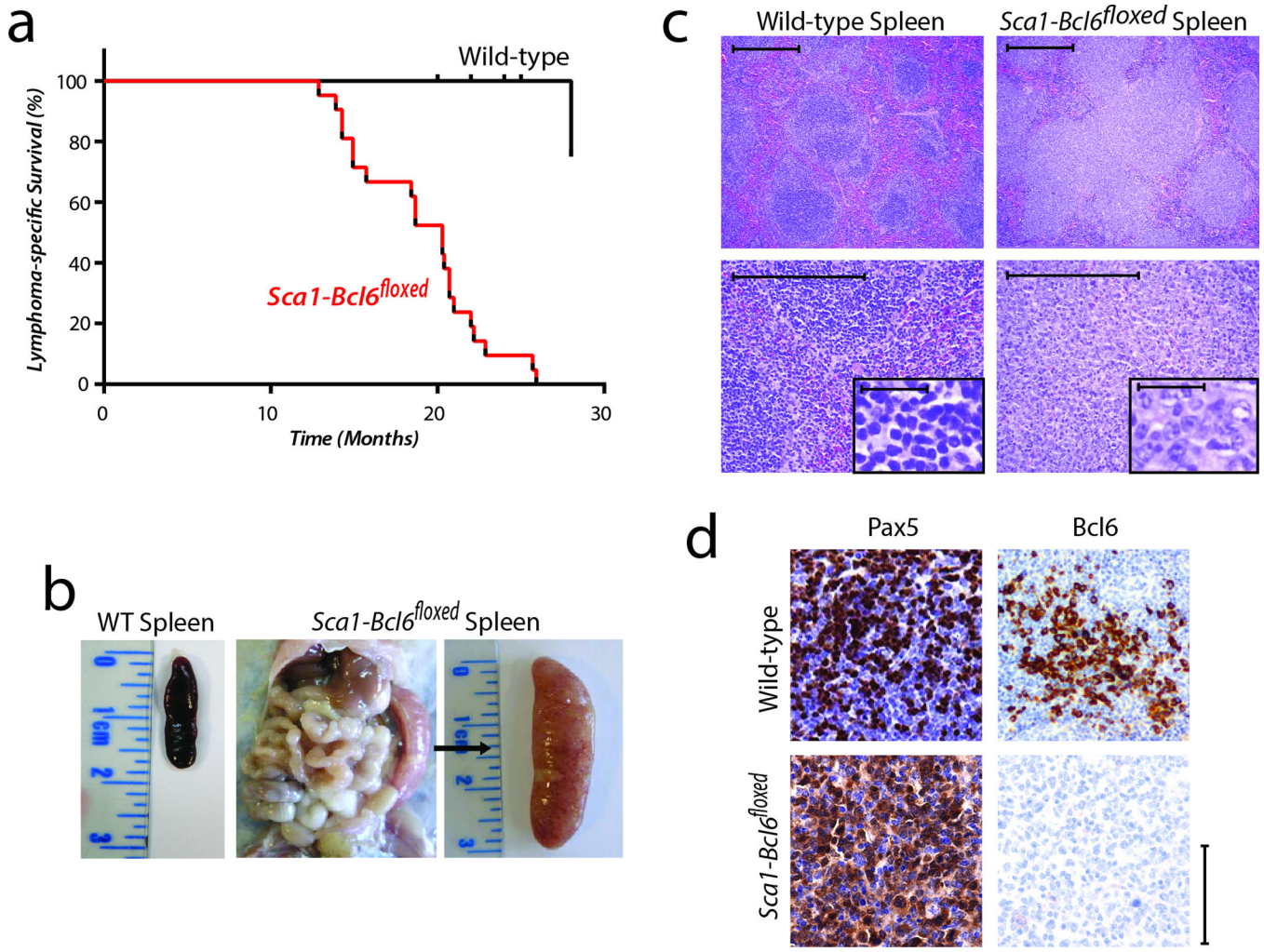


**Figure 4. Expression of the Bcl6 oncogene in hematopoietic progenitor cells (HSPCs) causes aggressive malignancy of mature B-cells that lack Bcl6 protein expression**

**a)** Lymphoma-specific survival of *Sca1-Bcl6*$^{floxed}$ mice (red, n=21), showing a significantly (Log-rank P-value <0.001) shortened lifespan compared to wild-type mice (black, n=20) as a result of mature B-cell malignancies. **b)** Example of splenomegaly observed in 50% (21/42) of *Sca1-Bcl6*$^{floxed}$ mice compared. A spleen from a wild-type mouse is shown for reference. **c)** Hematoxylin and eosin staining of wild-type spleens and tumor-bearing spleens from *Sca1-Bcl6*$^{floxed}$ mice shows loss of normal architecture resulting from effacement with cells morphologically resembling lymphocytes (above = 10×, below = 40×, inset = 400×). Images are representative of ≥3 replicates. Scale bar represents 200μm for all panels. **d)** Immunohistochemistry shows that lymphocytes from *Sca1-Bcl6*$^{floxed}$ tumors bear markers of B-cell identity (Pax5), but lack protein expression of Bcl6. Wild-type spleens obtained from immunized mice. Scale bar represents 100μm for large panels, 50μm for inset.

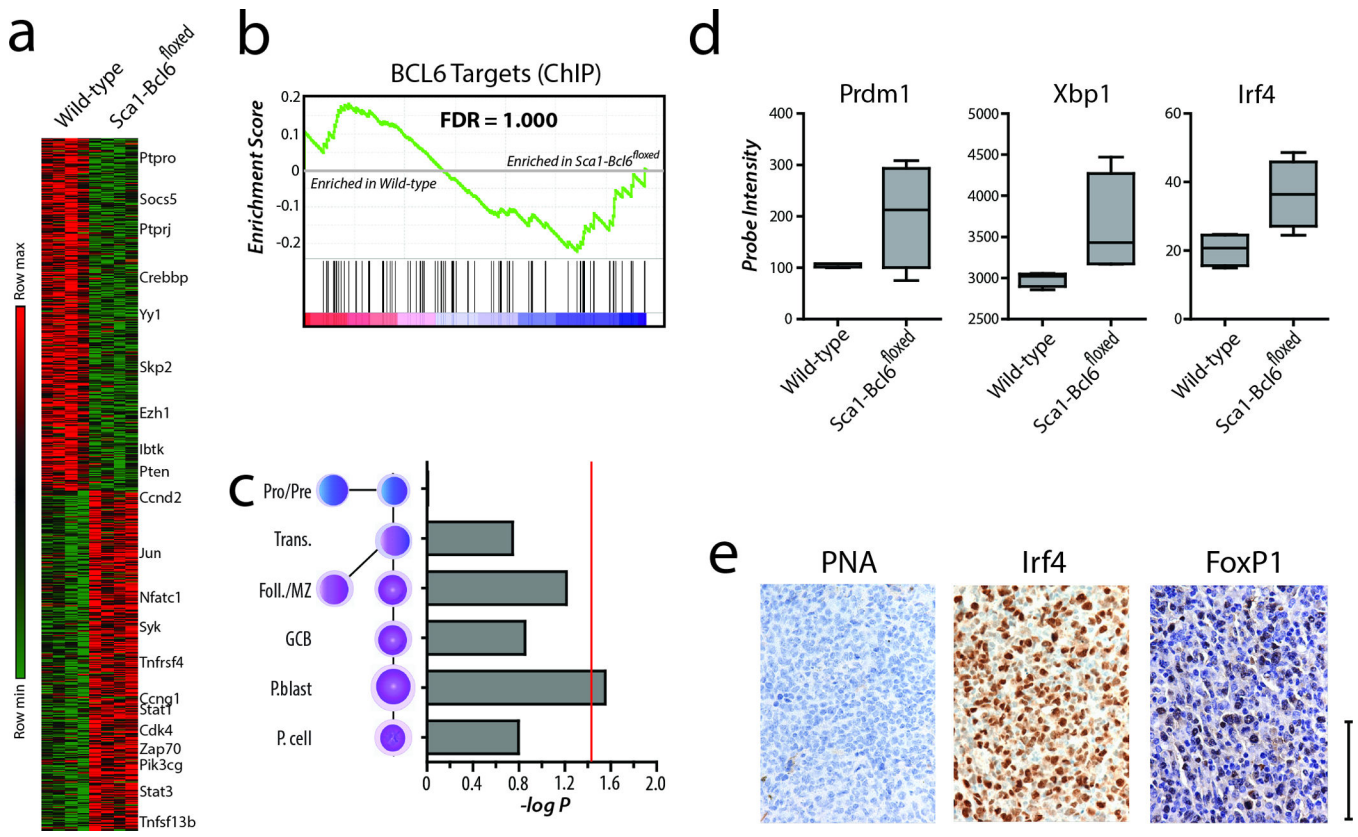


**Figure 5. Tumors from *Sca1-Bcl6*$^{floxed}$ mice resemble post-germinal center B-cells**

**a)** Differential gene expression analysis of tumor-bearing spleens of four *Sca1-Bcl6*$^{floxed}$ mice compared to spleens from four wild-type mice show significant differences. These differences include genes that are involved in B-cell signaling, and cell cycle regulation. **b)** Gene set enrichment analysis identified no significant repression or derepression of Bcl6 target genes in *Sca1-Bcl6*$^{floxed}$ tumor-bearing spleens compared to spleens from wild-type mice (GSEA FDR = 1.000). This is in line with what is observed in human DLBCL that possess amplification of the *BCL6* coding region. **c)** Analysis of differentially expressed genes in *Sca1-Bcl6*$^{floxed}$ tumor-bearing spleens with relation to gene expression signatures of normal murine B-cell differentiation including Pro/Pre-B, transitional (Trans.) B-cells, follicular and marginal zone (Foll./MZ) Bcells, germinal center B-cells (GCB), plasmablasts (P.blast) and plasma cells (P. cell). This shows significant enrichment of the normal plasmablast signature (Hypergeometric enrichment P-value = 0.028, FDR = 0.17). Red line represents a hypergeometric enrichment P-value of 0.05. **d)** Tumors from *Sca1-Bcl6*$^{floxed}$ mice show increased expression of transcription factors that are associated with post-germinal center stages of B-cell differentiation. Box plots represent the mean +/− the interquartile range with whiskers extending to the minimum and maximum value. **e)** Immunohistochemical staining shows negative expression of the germinal center marker peanut agglutinin (PNA), but positive staining for the post-germinal centre transcription factor Irf4, and the ABC-like DLBCL marker FoxP1. Images are representative of ≥3 replicates. All panels are 60×, scale bar represents 100μm.

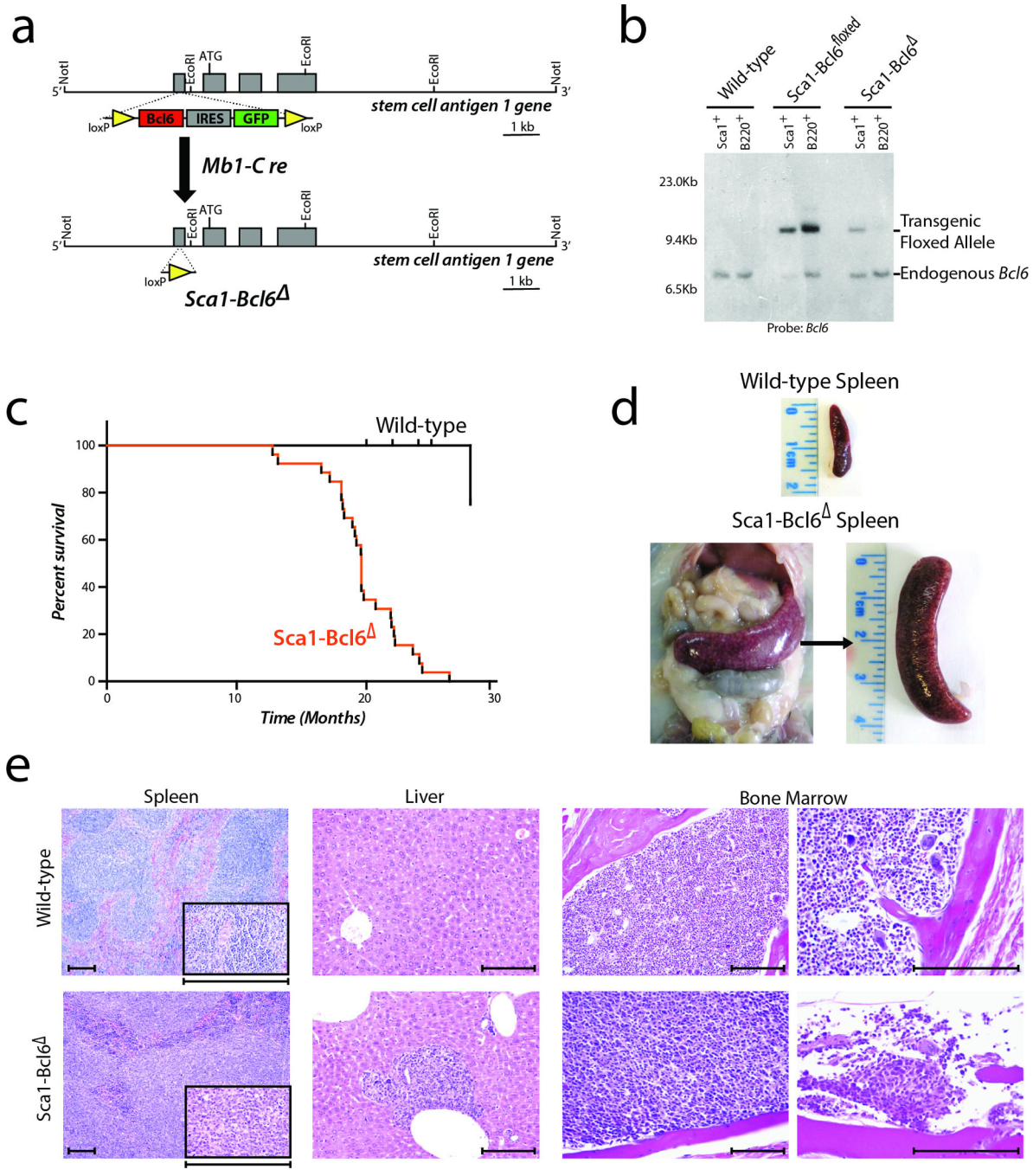


**Figure 6. Cre-mediated deletion of the exogenous Bcl6 allele in precursor B-cells does not alter formation of mature B-cell malignancies**

**a)** Diagramatic representation of the Sca1-Bcl6Δ mice, showing expression of the transgenic Bcl6 allele within HSPCs under control of the Sca1 promoter, followed by Cre-mediated deletion at an early Pro-B cell stage upon Mb1 expression. **b)** Southern blot analysis confirms absence of the transgenic Bcl6 allele within mature (B220+) B-cells of Sca1-Bcl6$^{\Delta}$ mice. Image is representative of 3 replicate experiments. **c)** Lymphoma-specific survival of *Sca1-Bcl6*$^{\Delta}$ mice (n=13) demonstrates a significantly (Log-rank P-value <0.001) shorter

lifespan compared to wild-type mice (n=20). **d)** Example of splenomegaly observed in *Sca1-Bcl6*$^{\Delta}$ mice, with spleen from a wild-type mouse is shown for reference. Images are representative of 13 mice. **e)** Effacement of spleen (100×, inset 400×) by malignant B-cells and infiltrates in the liver (200×) and bone marrow (left 200×, right 400×) result in loss of normal architecture. Images representative are of 13 mice. Scale bar represents 200μm.

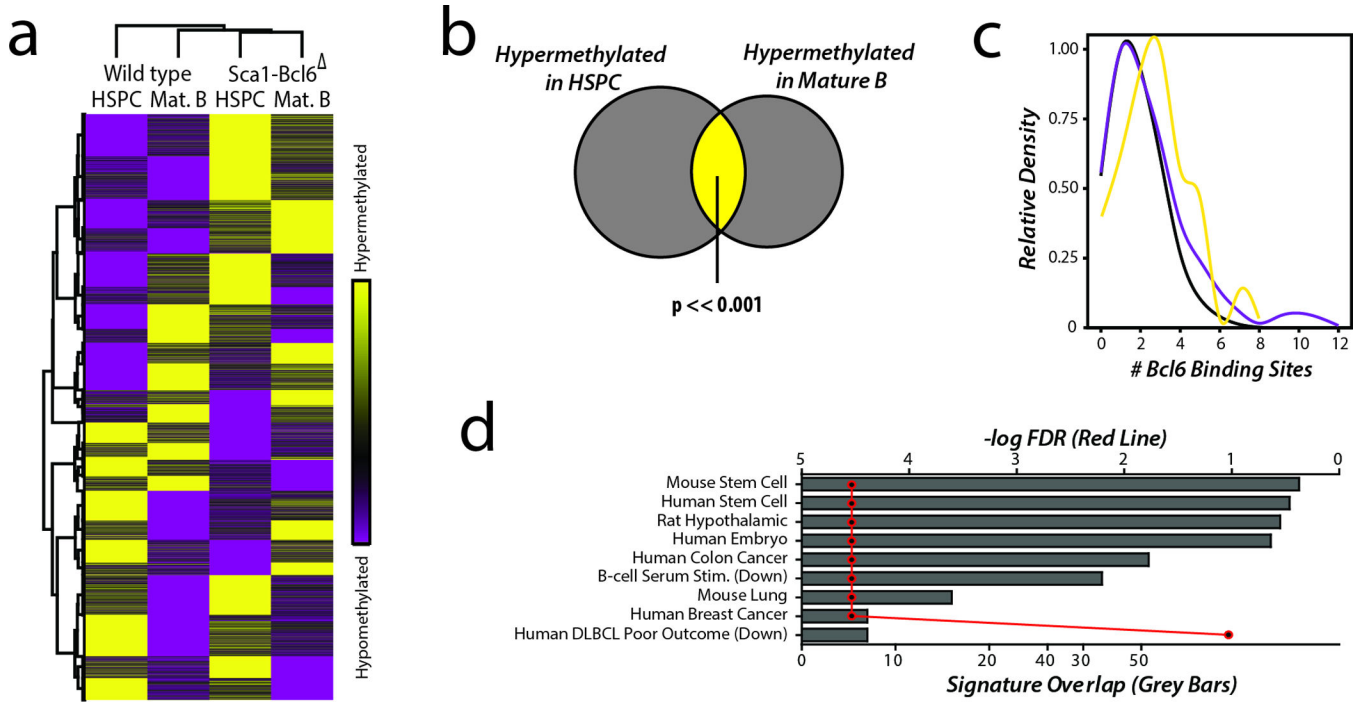


**Figure 7. Bcl6 expression in HPCs induced broad epigenetic changes**

**a)** Unsupervised hierarchical clustering of methylation ratio data from 11258 promoter regions revealed that HSPCs and mature Bcells (Mat. B) from *Sca1-Bcl6*$^{\Delta}$ mice are epigenetically more similar to each other than to their respective counterparts from wild-type mice, again suggesting that Bcl6 acts via an epigenetic mechanism that persists from HSPCs to mature B-cells. Each condition represents a pool of biological replicates from 6–9 mice. **b)** HSPCs and mature B-cells from *Sca1-Bcl6*$^{\Delta}$ mice show significant hypermethylation of a large number of genes compared to the same subsets from wild-type mice (Supplementary Data 3). These included a significant overlap of 470 genes, suggesting that Bcl6 creates an epigenetic signature in the HSPCs that is maintained through differentiation and can be found in the mature B-cell compartment. **c)** Kernel density plot of the number of BCL6 DNA binding sequence motifs identified in genomic regions with significant hypermehtyaltion (yellow) or hypomethylation (purple) in both HSPCs and mature Bcells from *Sca1-Bcl6*$^{\Delta}$ compared to wild-type mice (P-value < 0.05) and in regions with no change in methylation (black) between *Sca1-Bcl6*$^{\Delta}$ and wild-type mice (p>0.95). Hypermethylated regions showed a significant (P-value <0.001) increase in the abundance of BCL6 binding motifs compared to regions with no change in methylation. Hypomethlated regions also showed a significant, but less dramatic, increase in abundance of BCL6 binding motifs (P-value <0.001). P-values calculated by Mann Whitney U test. **d)** Hypergeometric gene set enrichment analysis of the list of genes that were hypermethylated in both HPCs and mature B-cells of *Sca1-Bcl6*$^{\Delta}$ mice showed significantly enrichment (Hypergeometric enrichment FDR < 0.25) of multiple gene sets including those associated with murine and human stem cells, and with poor outcome in DLBCL. The number of overlapping genes between those with conserved hypermethylation and those present in listed genes sets are

shown by grey bars (bottom X-axis), with the corresponding FDR for the hypergeometric enrichment shown by the red line (top X-axis).