DNA copy number gains of *TCF4* (E2-2) are associated with poor outcome and the activated B-cell-like subtype of diffuse large B-cell lymphoma

Running title: TCF4 Gains in DLBCL

Keenan Hartert[1], Saber Tadros[1,2#], Alyssa Bouska[3#], Dalia Moore[1], Tayla Heavican[3], Christine Pak[4], Chih Long Liu[5], Andrew Gentles[6], Elena Hartmann[7,8], Robert Kridel[9], Karin Ekstrom Smedby[10], Gunnar Juliusson[11], David Klinkebiel[12], Richard Rosenquist[13], Randy Gascoyne[14], Andreas Rosenwald[7,8], Martin Bast[15], Julie Vose[15], Matthew Lunning[15], Timothy Greiner[3], Scott Rodig[4], Javeed Iqbal[3], Ash Alizadeh[5], Michael R. Green[1,2,16]*

[1]Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE, USA; [2]Current Affiliation: Department of Lymphoma/Myeloma, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [3]Department of Pathology and Immunology, University of Nebraska Medical Center, Omaha, NE, USA; [4]Department of Pathology, Brigham and Womens Hospital, Harvard Medical School, Boston, MA, USA; [5]Division of Oncology, Department of Medicine, Stanford University, Stanford, CA, USA; [6]Department of Radiology, Stanford University, Stanford, CA, USA; [7]Institute of Pathology, University of Würzburg, Würzburg, Germany; [8]Comprehensive Cancer Center Mainfranken. Wurzburg, Germany; [9]Princess Margaret Cancer Center, University of Toronto, Toronto, ON, Canada; [10]Department of Medicine Solna, Clinical Epidemiology Unit, Karolinska Institutet, and Hematology Center, Karolinska University Hospital, Stockholm, Sweden; [11]Department of Laboratory Medicine, Stem Cell Center, Lund University, Lund, Sweden; [12]Department of Biochemistry and Molecular Biology, University of Nebraska Medical Center, Omaha, NE, USA; [13]Department of Molecular Medicine and Surgery, Karolinska Universitetssjukhuset, Stockholm, Sweden; [14]Center for Lymphoid Cancers, British Columbia Cancer Agency, Vancouver, BC, Canada; [15]Division of Hematology and Oncology, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, NE, USA.; [16]Current Affiliation: Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, TX, USA.

([#]Equally contributed, *Corresponding author)

* To whom correspondence should be addressed:
Dr. Michael Green
Department of Lymphoma/Myeloma,
University of Texas MD Anderson Cancer Center,
Houston, TX 77030, USA
Phone: +1-713-745-4244
Email: mgreen5@mdanderson.org

Key words: Transcription factor, Lymphoma, Leukemia, B-cell

ABSTRACT

Diffuse large B-cell lymphoma (DLBCL) is the most common lymphoma and can be classified into two prognostically distinct molecular subtypes based upon transcriptional profiling. The activated B-cell (ABC)-like subtype is associated with a poor clinical outcome, and chronic activation of B-cell receptor signaling. Although many genetic alterations have been identified that contribute to this phenotype, these only account for a minority of cases. We employed public high-resolution DNA copy number profiles from 673 tumors to define the landscape of somatic copy number alterations (SCNAs) in DLBCL. Using integrative analysis of gene-expression profiling data, we found DNA-binding transcription factors to be a significantly enriched targets of SCNAs in DLBCL. We extended upon this observation in an additional 2,506 tumors from 6 other histologies, and found SCNA of transcription factors to be pervasive across B-cell malignancies. Furthermore, co-segregating SCNAs targeting transcription factors were associated with adverse patient outcome and the ABC-like subtype of DLBCL. This included a novel target of DNA copy number gain, *TCF4* (E2-2). Gains of *TCF4* were associated with a transcriptional signature that included increased expression of B-cell receptor signaling components. In a validation cohort of 124 DLBCL tumors interrogated by targeted sequencing, we found that *TCF4* DNA copy number gains significantly co-associate with somatic mutation of *CD79B* and *MYD88*. Together, these data suggest that SCNA of transcription factor genes are an important feature of B-cell malignancies, and these alterations may contribute to the ABC-like phenotype of DLBCL in tandem with other previously defined somatic alterations.

INTRODUCTION

B-cell development involves a complex set of processes that have evolved to yield a diverse but self-tolerant repertoire of immunoglobulin producing cells. These processes are tightly regulated by transcription factor networks and feedback loops that control proliferation, apoptosis, and progression of B-cells through the distinct stages of differentiation. Loss of homeostatic control via genetic aberrations during B-cell development can result in B-cell malignancies that bear resemblance to normal stages of differentiation (Shaffer et al. 2002; Kuppers 2005). These corresponding normal stages of development, or cell of origin, are part of the criteria for classifying entities of B-cell leukemia and lymphoma that are distinct with respect to their epidemiology, etiology, morphology, immunophenotype, and clinical behavior (Campo et al. 2011). For example, B-cell acute lymphoblastic leukemia (B-ALL) is characterized by disordered early B-cell development and the accumulation of precursor B-cells, whereas Multiple Myeloma (MM) is a malignancy of plasma cells – one of the terminal stages of B-cell differentiation.

The most common form of lymphoma, diffuse large B-cell lymphoma (DLBCL), can also be sub-classified into two prognostically distinct subsets based upon molecular similarities to B-cell developmental stages (Alizadeh et al. 2000). The germinal center B-cell-like (GCB-like) subtype of DLBCL shows transcriptional similarities to normal germinal center B-cells, whereas the activated B-cell-like (ABC-like) subtype shows transcriptional similarities to activated post-germinal center B-cells or plasmablasts. Importantly, patients with ABC-like DLBCL have significantly worse overall survival compared to patients with GCB-like DLBCL, when treated with either combination chemotherapy (CHOP) or combination chemotherapy with the addition of rituximab (R-CHOP) (Lenz et al. 2008b). A number of genetic alterations have been attributed to the ‘activated’ phenotype of ABC-like DLBCL through modification of signaling downstream of the B-cell receptor (BCR). These include activating mutations of the *CD79A/B*, *CARD11*, or *MYD88* genes (Lenz et al. 2008a; Davis et al. 2010; Ngo et al. 2011). However, these mutations only

account for a fraction of ABC-like DLBCL cases and much of the genetic basis for these tumors therefore remains to be defined.

A common mechanism for tumorigenesis is the gain or loss of DNA encoding oncogenes or tumor suppressor genes, respectively. These somatic copy number alterations (SCNA) can be profiled at high-resolution using single nucleotide polymorphism (SNP) microarrays or next-generation sequencing (NGS)-based techniques. Several prior studies have interrogated SCNAs in B-cell malignancies, but these were typically restricted in number and limited to a single histological subtype. Here, we show through the analysis of high-resolution DNA copy number profiles 3,179 tumors of 7 different histologies that transcription factors are perturbed by SCNAs in B-cell malignancies, and may contribute to the etiology of the ABC-like subtype of DLBCL.

RESULTS

Transcription factors are targeted by DNA copy number gains in DLBCL

We used the GISTIC2 algorithm (Mermel et al. 2011) to analyze high resolution DNA copy number profiles from 673 DLBCL tumors. This algorithm integrates the frequency of alterations across a cohort, and the magnitude of these alterations, to allow the identification of statistically significant regions of DNA copy number gain or loss. The 'peak' of each alteration is defined as the smallest region with the highest significance of DNA copy gain/loss and contains a minimal set of genes that likely includes the target(s) that drive the alteration (Beroukhim et al. 2007; Beroukhim et al. 2010). This method identified a total of 13 significant DNA copy number gains that contained 770 genes in their peaks, and 22 significant DNA copy number losses that contained 1450 genes in their peaks (Figure 1a; Table S1).

The peaks of SCNAs identified by GISTIC contain one or more likely driver genes of each SCNA, but are also likely to contain passenger genes. In order to delineate between driver and passenger

genes, we performed an integrative analysis of 249 tumors with matched DNA copy number and gene expression microarray data (Table S2). This analysis identified 632 genes with significantly reduced gene transcript abundance in tumors with DNA copy number loss at GISTIC peaks compared to tumors without (FDR Q-value < 0.25, Figure 1b), and 435 genes with significantly higher gene transcript abundance in tumors with DNA copy number gain at a GISTIC peak compared to tumors without (FDR Q-value < 0.25, Figure 1c). Among these genes were well described targets of DNA copy number alterations in DLBCL, such as copy number losses with reduced expression of *TP53*, *TNFAIP3*, *PTEN* and *RB1*, and gains with increased expression of *REL*, *BCL11A* and *MDM2*. To identify cellular processes that are significantly altered by SCNAs, we performed hypergeometric enrichment analysis of, (i) genes that were targeted by DNA copy number loss and showed an associated significant reduction in expression, and (ii) genes that were targeted by DNA copy number gain and showed an associated significant increase in expression. This revealed a significant enrichment of cancer hallmark characteristics (Hanahan and Weinberg 2011) such as apoptosis and proliferation (Figure 1c, Table S3) among genes targeted by DNA copy number loss, in line with what is frequently observed in a variety of malignancies. However, analysis of the genes targeted by DNA copy number gain revealed a unique enrichment of genes possessing transcription factor activity (Figure 1c, Table S3). The most highly significant enrichment was ‘Transcription Factor Activity, sequence-specific DNA binding’ and ‘Nucleic acid binding transcription factor’ gene ontologies (FDR<0.001) that each included 58 unique genes that were targeted by DNA copy gain and showed an associated increase in transcript abundance in DLBCL. This highlights SCNA of transcription factor genes as a novel feature of DLBCL genomes.

The DNA copy number landscape in B-cell malignancies includes alterations of multiple developmentally regulated transcription factors.

Due to the potentially important role of transcription factor DNA copy number alterations in DLBCL, and the role of transcription factor networks in regulating B-cell differentiation, we next explored the DNA copy number landscape of other B-cell malignancies that align with different stages of B-cell differentiation. We included all subtypes of B-cell malignancy for which we could collect high resolution DNA copy number data from ≥100 tumors (Table S1). This included B-cell acute lymphoblastic leukemia (B-ALL, n=534, Figure S1), follicular lymphoma (FL, n=368, Figure S2), splenic/nodal marginal zone lymphoma (MZL, n=295, Figure S3), mantle cell lymphoma (MCL, n=206, Figure S4), chronic lymphocytic leukemia (CLL, n=544, Figure S5), and multiple myeloma (MM, n=559, Figure S6). These malignancies show contrasting DNA copy number landscapes at the level of aneusomies and arm-level gains and losses (Figure S7), leading us to analyze them individually using GISTIC rather than collectively. For example, B-ALL and MM tumors each showed patterns of hyperdiploidy, but these targeted alternate chromosomes in each disease. Analysis of data from each individual disease using GISTIC highlighted varying levels of genomic complexity and different peaks of DNA copy number gain and loss (Figure 2). Importantly, we observed DNA copy number alterations of multiple transcription factors that have been previously implicated in regulating B-cell development (Figure 2). These included those that have been previously described to be targeted by SCNAs, such as *IKZF1* (Mullighan et al. 2009), *REL* (Joos et al. 1996), *PRDM1* (Pasqualucci et al. 2006), *BCL6* (Green et al. 2014) and *MYC* (Rao et al. 1998). In addition, we identified novel DNA copy number alterations of transcription that have a defined role in B-cell development, such as loss of *POU2AF1* (OBF1), *NKX3-1*, *NKX3-2* and *IRF8*, and gains of *PBX1*, *IRF4*, and *TCF4* (E2-2). The analysis of an additional set of T-cell malignancies (n=261, Figure S8), myeloid (n=272, Figure S9), and solid tumors (n=1510, Figure S10) (Beroukhim et al. 2010) allowed us to classify these as either disease-specific, B-cell-specific, lymphoid-specific, or multi-cancer alterations. This extends upon prior descriptions

of genetic alterations of transcription factors as key events in lymphomagenesis, and highlights transcriptional circuits that may be perturbed in specific subtypes or multiple subtypes of B-cell malignancies.

Co-segregating transcription factor alterations are associate with cell of origin subtype and poor outcome in DLBCL.

Our observation of different patterns of transcription factor alteration across subtypes of B-cell malignancies that align with diverse stages of B-cell development led us to hypothesize that such alterations may also contribute to the etiology of the GCB-like and ABC-like molecular subtypes of DLBCL. To identify characteristic patterns of SCNAs across DLBCL tumors, we used unsupervised hierarchical clustering of SCNAs that were present in ≥10% of DLBCL tumors (13 gains and 22 losses) and tumors that possessed ≥2 SCNAs (n=481; Figure 3a). As expected, SCNAs located on the same chromosome, but identified as unique GISTIC peaks, often co-segregate. This is driven by tumors that possess broad SCNAs, including arm-level gains/losses and aneuploidies, which may span multiple GISTIC peaks. The peaks are often defined by a subset of cases that possess focal copy number gain/loss over the smaller region, and likely define targets of both focal and broad SCNAs involving that chromosome. In addition to those on the same chromosome, we also observed co-segregation between DNA copy number gains on 3 different chromosomes; 3q27, 18q21, and 19p13.3 (Figure 3a). These alterations have previously been defined as targeting the *BCL6* transcription factor (Green et al. 2014), the anti-apoptotic *BCL2* oncogene (Monni et al. 1997) and the *SPIB* transcription factor (Lenz et al. 2008c), respectively. We performed cell of origin subtype classification on the 249 samples with matched gene expression microarray and DNA copy number data, and identified associations between the GCB-like or ABC-like subtypes and the presence of each SCNA using a Fisher test. This revealed that the three co-segregating DNA copy number gains at 3q27, 18q21 and 19p13.3 were also significantly associated with the ABC-like subtype of DLBCL (Figure 3b, e). As this subtype has

an inferior prognosis compared to the GCB-like subtype, we also assessed the association between SCNAs and overall survival in cohorts treated with combination chemotherapy (CHOP, n=232) or combination chemotherapy with rituximab (R-CHOP, n=197) using a log-rank test (Figure 3c-d). This showed that these lesions were also each associated with inferior overall survival, as confirmed by Kaplan-Meier analysis (Figure 3f).

Gains of 18q21 target the *TCF4* (E2-2) transcription factor.

The involvement of the *BCL6* transcription factor at 3q27 (Green et al. 2014) and the *SPIB* transcription factor at 19p13.3 (Lenz et al. 2008c) has been previously demonstrated. We therefore focused on 18q21 gains. In contrast to prior observations, GISTIC analysis showed that the most significant peak of DNA copy number gain did not include the *BCL2* oncogene (Figure 4a). The 1.58Mbp peak identified by GISTIC included 3 genes, all of which showed significantly increased gene expression in association with these alterations. This included *TXNL1*, *WDR7*, and the *TCF4* transcription factor (Figure 4b). The *TCF4* gene has been previously implicated in regulating immunoglobulin gene expression through binding of the immunoglobulin enhancer (Gloury et al. 2016; Wohner et al. 2016), and we therefore hypothesized that it may have a potential role in promoting BCR signaling in ABC-like DLBCL. The localization of the GISTIC peak to the *TCF4* rather than *BCL2* was driven by the 15% of cases with 18q21 DNA copy number gains that included the *TCF4* gene but not the *BCL2* gene. In contrast, only 11% of cases with 18q21 DNA copy number gains included the *BCL2* gene but not the *TCF4* gene. However, the majority of tumors with 18q21 DNA copy number gains included both the *TCF4* and *BCL2* genes, highlighting both of these genes as likely important targets of this alteration (Figure 4c). This is supported by the significant increase in *BCL2* expression in ABC-like DLBCL tumors associated with 18q21 DNA copy number gains (Figure S11).

TCF4 Gains Are Associated with Increased Immunoglobulin Expression and Co-Segregate with *CD79B* and *MYD88* Mutations

Chronic activation of immunoglobulin signaling is an important feature of ABC-like DLBCL (Davis et al. 2010) and we hypothesized that *TCF4* copy gain may contribute to this phenotype. To investigate this, we performed a genome-wide assessment of the transcriptional consequences of *TCF4* DNA copy number gain. Differential gene expression analysis identified 411 genes with significantly increased and 251 genes with significantly decreased gene transcript abundance associated with *TCF4* gain (Fold-change ≥ 1.2, FDR Q-value < 0.1; Figure 5a; Table S4). In line with the previous implication of *TCF4* in regulating immunoglobulin heavy chain expression (Ernst and Smale 1995; Gloury et al. 2016; Wohner et al. 2016), we observed significantly increase Ig heavy chain μ (IGHM) gene expression in cases with *TCF4* DNA copy number gain. In addition, we observed increased expression of multiple other signaling components downstream of the BCR, including *CD79A*, *LYN*, *BLNK*, *CARD11* and *PIK3CA*. To provide additional support for a direct role of TCF4 in inducing these changes, we utilized publicly available TCF4 ChIP-seq data (Ceribelli et al. 2016) to determine whether TCF4 binds to the regulatory regions of these genes. This revealed a peak of TCF4 binding at the transcription start-sites of genes with increased expression associated with TCF4 DNA copy number gain (Figure 5b), including binding in intronic region of the immunoglobulin heavy-chain locus (Figure 5c), supporting their likely transcriptional regulation by TCF4.

BCR signaling is also perturbed by somatic mutations of *CARD11*, *CD79B* and *MYD88* in ABC-like DLBCL (Lenz et al. 2008a; Davis et al. 2010; Ngo et al. 2011), and the relative representation of these alterations may impact responses to inhibitors of the BCR signaling pathway such as Ibrutinib (Wilson et al. 2015). We therefore assessed the representation of these alterations in an additional cohort of 124 DLBCL tumors that we interrogated by targeted NGS, for which 70 had available cell of origin subtype information (Garcia-Ramirez et al. 2017). Somatic mutations were

identified in *CARD11* (16%, 20/124), *CD79B* (9%, 11/124) and *MYD88* (13%, 16/124) at similar frequencies to prior descriptions (Lenz et al. 2008a; Davis et al. 2010; Ngo et al. 2011). In addition, we utilized the CopyWriteR algorithm (Kuilman et al. 2015) to interrogate SCNAs, and identified *TCF4* gains in 17% (21/124) of cases (Figure 6a). Mutations in *CARD11* did not appear to associate with either subtype, but mutations in *CD79B* and *MYD88* and DNA copy number gains in *TCF4* were more prevalent in the ABC-like subtype as expected. Using a Fisher test, we also observed a significant co-association between *TCF4* gains, *CD79B* mutations and *MYD88* mutations, suggesting that these may not be redundant mechanisms for activating BCR signaling, but may function in tandem. This is in line with recent data showing that *CD79B* and *MYD88* mutations act synergistically to promote BCR signaling (Wang et al. 2017). Together, these results implicate *TCF4* DNA copy number gains in the BCR signaling phenotype of ABC-like DLBCL and suggest that they may function together with other BCR signaling mutations.

DISCUSSION

Deregulation of transcription factor activity has long been linked to the etiology of B-cell malignancies as a result of the recurrent targeting of transcription factors by translocations. This includes fusion of the *TCF3* and *PBX1* genes (Mullighan 2012) and deletion of *IKZF1* in B-ALL (Mullighan et al. 2009), the activation of *MYC*, *BCL6*, and *PAX5* and by translocations with the immunoglobulin locus in mature B-cell lymphoma (Willis and Dyer 2000), and translocation of *IRF4* in multiple myeloma (Iida et al. 1997). We have extended upon these observations by showing that DNA-binding transcription factors are enriched targets for DNA copy number gains in DLBCL, and that unique transcription factors are perturbed across the differentiation spectrum of B-cell malignancies.

Although many of the targets of SCNAs in DLBCL have been previously described by other studies (Monti et al. 2012) the larger numbers of tumors that were analyzed by this study allowed

us to identify novel targets such as *TCF4*. Furthermore, we are the first to show that transcription factors are an enriched target of DNA copy number gains in this disease, and that co-segregating transcription factor alterations are associated with the ABC-like subtype and poor outcome. We have recently shown that transient activity of *BCL6* linked to DNA copy number gains at 3q27 are sufficient to induce lymphoma resembling ABC-like DLBCL in mice (Green et al. 2014), providing a validation of the association between this SCNA and the ABC-like subtype that we observed here. However, it also highlights that integrative analysis may overlook a small number of SCNA driver genes if they have regulatory feedback loops and do not require persistent expression or suppression to have an oncogenic effect. In addition, GISTIC peaks identify the genes that are altered with the highest magnitude across the highest frequency of tumors. This may also result in some genes that drive specific SCNAs in a minority of cases to be overlooked. For example, although the *BCL2* oncogene is not identified with the peak of the 18q21 DNA copy number gain within the DLBCL cohort in this study because of the 15% of tumors that show gains of *TCF4* but not *BCL2*, it likely has a 'driver' role for large SCNAs of 18q in the 74% of cases in which it was co-amplified along with *TCF4*, and is likely the main driver of small SCNAs in this region in the 11% of cases in which *BCL2* is amplified but *TCF4* is not. This study therefore overlooks a small number of likely driver genes, but instead focusses on the most significant set of genes with the greatest amount of evidence for being drivers of SCNAs.

In addition to DNA copy number gains of 3q27, we found that gains of 19p13 and 18q21 co-segregated within a set of cases that were highly enriched for the ABC-like DLBCL subtype. This subtype is characterized by chronically active B-cell receptor signaling and constitutive NFκB activity that is, in part, driven by somatic mutations of genes such as *CD79B* (Davis et al. 2010), *CARD11* (Lenz et al. 2008a) and *MYD88* (Ngo et al. 2011). However, these mutations only account for a subset of cases. DNA copy number gain of the *SPIB* transcription factor gene at 19p13 has been previously identified as being associated with the ABC-like subtype and was

postulated to contribute to the pathogenesis of these tumors by promoting the expression of genes with a role in NFκB signaling. We are the first to describe an association between SCNA of the *TCF4* transcription factor gene at 18q21 and the ABC-like DLBCL subtype. This gene is part of the E2 family of transcription factors and has a well-defined role in regulating the expression of immunoglobulin genes and other developmental programs during B-cell differentiation (Gloury et al. 2016; Wohner et al. 2016). We noted that immunoglobulin heavy-chain genes also showed higher expression in cases with *TCF4* DNA copy number gain, and that this region is bound by *TCF4* in other cell types, suggesting that this alteration may promote higher surface immunoglobulin expression and contribute towards chronic active signaling. However, the association between genetic alterations of *TCF4* and the ABC-like DLBCL subtype would best be validated in a transgenic mouse model to confirm this hypothesis, and validation of TCF4 target genes should additionally be performed in both non-malignant human B-cells and primary DLBCL tumor cells. This may be an important area of investigation, as other genetic alterations that affect B-cell receptor signaling have been linked to the variable clinical efficacy of inhibitors of this pathway such as Ibrutinib (Wilson et al. 2015). We noted that *CD79B* and *MYD88* mutations significantly co-segregated, supporting recent findings that they act cooperatively to promote BCR signaling (Wang et al. 2017). Importantly, we observed that *TCF4* DNA copy number gains also co-segregate with both *CD79B* and *MYD88* mutation, and we therefore suggest that these alterations may also act cooperatively to promote BCR signaling and the ABC-like DLBCL phenotype.

A recent study highlighted a potential alternative avenue for therapeutic targeting of TCF4 using bromodomain inhibitors, as BRD4 was found to regulate *TCF4* expression (Ceribelli et al. 2016). However, *BCL2* over-expression has been highlighted as a potential resistance mechanism to bromodomain inhibitors (Bui et al. 2017), and is also targeted by 18q21 DNA copy number gains in the majority of cases. Importantly, this study also showed synergy between the BCL2 inhibitor,

Venetoclax, and a bromodomain inhibitor, ABBV-075, suggesting that combination of these agents may be an attractive therapeutic strategy for DLBCL – particularly those tumors harboring 18q21 DNA copy number gains.

In conclusion, we have provided evidence for an important role of DNA copy number alterations of transcription factors in B-cell malignancies, particularly the ABC-like subtype of DLBCL. We highlight *TCF4* as a novel target for DNA copy number gains of 18q21, and suggest that these alterations may function in tandem with other mutations to promote the expression and activity of BCR signaling components.

METHODS

DNA copy number data acquisition and processing

Publicly available data for single nucleotide polymorphism microarrays and array comparative genome hybridization platforms with 200,000 markers were downloaded from the gene expression omnibus (www.ncbi.nlm.nih.gov/geo/). These included Affymetrix 250K and SNP 6.0 platforms, and the Agilent 244K platform. The data originated from 36 studies, including 9 unique B-cell malignancies aligning with different stages of B-cell development (Rosenwald et al. 2003; Kuiper et al. 2007; Mullighan et al. 2007; Lenz et al. 2008c; Mullighan et al. 2008a; Mullighan et al. 2008b; Paulsson et al. 2008; Avet-Loiseau et al. 2009; Compagno et al. 2009; Kato et al. 2009; Kawamata et al. 2009; Dickens et al. 2010; Kuiper et al. 2010; Lilljebjorn et al. 2010; Mosca et al. 2010; Walker et al. 2010; Challa-Malladi et al. 2011; Chapman et al. 2011; Green et al. 2011; Gunnarsson et al. 2011; Oricchio et al. 2011; Rinaldi et al. 2011; Fresquet et al. 2012; Monti et al. 2012; Rohde et al. 2012; Bea et al. 2013; Bodker et al. 2013; Chigrinova et al. 2013; Harder et al. 2013; Salaverria et al. 2013; Bouska et al. 2014; Diakos et al. 2014; Li et al. 2014; Messina et al. 2014; Okosun et al. 2014; Green et al. 2015). For comparison, raw data were also acquired from an additional 31 studies including 261 T-cell malignancies, 272 myeloid malignancies, and 1510 solid tumors (Table S5) (Kuiper et al. 2007; Chen et al. 2008; Chiang et al. 2008; Haverty et al. 2008; Scotto et al. 2008; Akagi et al. 2009; Bass et al. 2009; Castro et al. 2009; Forshew et al. 2009; Goransson et al. 2009; Haverty et al. 2009; Hu et al. 2009; Northcott et al. 2009; Radtke et al. 2009; Reid et al. 2009; Schraders et al. 2009; Sos et al. 2009; Tosello et al. 2009; Beroukhim et al. 2010; Bullinger et al. 2010; Hartmann et al. 2010; Nowak et al. 2010; Poage et al. 2010; Huang et al. 2011; Micke et al. 2011; Miller et al. 2011; Dulak et al. 2012; Kondo et al. 2012; Rucker et al. 2012; Wilkerson et al. 2012; Yamagishi et al. 2012). For Affymetrix microarrays, raw cel files were extracted and copy number predicted using the Affymetrix Copy Number Analysis for Genechip (CNAG) tool, with reference to data from 100 caucasian HapMap samples (Table S5). Agilent data was analyzed using BioConductor, as previously described (Lenz et al. 2008c).

Data for all arrays were represented as Log2 copy number change and segmented using the circular binary segmentation (CBS) tool on GenePattern (Reich et al. 2006). Peaks of significant DNA copy number loss and gain were identified using GISTIC2.0 (Mermel et al. 2011). The thresholds utilized for DNA copy number gain and loss were 0.1 copies, over a region encompassing 100 markers.

Integrative Analysis and Hypergeometric Enrichment Analysis

Raw cel files for matched Affymetrix U133 Plus 2.0 gene expression microarray data were obtained for 249 DLBCL (GSE11318 and GSE34171) from the Gene Expression Omnibus (Lenz et al. 2008c; Monti et al. 2012). Data were RMA normalized and the two DLBCL datasets batch-corrected using ComBat (Johnson et al. 2007). Differential gene expression analysis was performed between samples with or without a given lesion using a Students T-test, and corrected for multiple hypothesis testing by the Benjamini-Hochberg method. Genes with a fold-change ≥1.2 in the direction of the copy number alteration and a false-discovery rate <0.25 were considered statistically significant. To define targets of SCNA, the differential gene expression analysis was limited to only the genes within the GISTIC-defined peak of each SCNA. To identify the signature associated with 18q21 DNA copy number gain, differential gene expression analysis was performed on the 10,000 most variably expressed genes across the dataset as determined by median absolute deviation. Significantly altered genes targeted by DNA copy number gains and losses were analyzed separately for enriched gene ontology biological process (GO-BP) terms using DAVID (Huang et al. 2007), with an FDR<0.25 and a fold-change direction corresponding to the direction of the copy number alteration being considered significant.

Survival and Cell of Origin Subtype Association

For visualization of co-association, each lesion was scored as a binary present/absent classification in each case and clustered using Spearman's rank correlation. Only SCNAs that

were present in ≥10% of DLBCL tumors (13 gains and 22 losses) and tumors that possessed ≥2 SCNAs (n=481) were included. DLBCL tumors were classified by cell of origin using the Wright algorithm (Wright et al. 2003). Association between SCNAs and COO subtype were determined using a Fisher exact test. Overall survival data was collated from previous studies for 232 DLBCL patients treated with CHOP combination chemotherapy (Lenz et al. 2008c; Monti et al. 2012), and 197 DLBCL patients treated with CHOP combination chemotherapy plus the anti-CD20 monoclonal antibody rituximab (R-CHOP) (Scandurra et al. 2010; Monti et al. 2012). Associations between the presence of a SCNA and overall survival were assessed using a log-rank test.

ChIP-Seq Data Analysis

Previously described chromatin immunoprecipitation sequencing data for TCF4 (Ceribelli et al. 2016) were downloaded from the gene expression omnibus (GSE75650). Raw FASTQ files were mapped to the human genome (UCSC hg19) using BWA-Mem (Li 2013), and sorted and deduplicated using Picard tools. The physical location of transcription start sites for genes with significantly increase expression associated with 18q21 DNA copy number gains were collated using the UCSC Table Browser (Karolchik et al. 2004). Raw data were visualized using the integrative genomics viewer (Robinson et al. 2011), and heatmaps were created for visualization of TCF4 signal at transcription start sites using EaSeq (Lerdrup et al. 2016).

Identification of Somatic Mutations and DNA Copy Number Alterations from Next Generation Sequencing Data.

Targeted sequencing of 380 genes that are recurrently mutated in B-cell lymphoma was performed previously for 124 primary DLBCL tumors from the University of Nebraska Medical Center (Garcia-Ramirez et al. 2017). Raw FASTQ files were mapped to the human genome (UCSC hg19) using BWA-Mem (Li 2013), realigned around InDels and sorted using the Genome Analysis Toolkit (McKenna et al. 2010), and duplicates removed using Picard tools. Single

nucleotide variants and small insertion/deletions were identified using the GATK Unified Genotyper (McKenna et al. 2010) and VarScan2 (Koboldt et al. 2012), with only variants being called by both algorithms being considered true. VCF files were annotated using SeattleSeq (Ng et al. 2009), and single nucleotide variants annotated in dbSNP or variants falling within regions of tandem repeats were removed. Variants were required to be within a region of ≥100X coverage and have ≥3 supporting reads on each strand. Somatic copy number alterations from targeted sequencing data were identified using CopyWriteR (Kuilman et al. 2015), with 100kb bins.

REFERENCES

Akagi T, Shih LY, Kato M, Kawamata N, Yamamoto G, Sanada M, Okamoto R, Miller CW, Liang DC, Ogawa S et al. 2009. Hidden abnormalities and novel classification of t(15;17) acute promyelocytic leukemia (APL) based on genomic alterations. *Blood* **113**: 1741-1748.

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X et al. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**: 503-511.

Avet-Loiseau H, Li C, Magrangeas F, Gouraud W, Charbonnel C, Harousseau JL, Attal M, Marit G, Mathiot C, Facon T et al. 2009. Prognostic significance of copy-number alterations in multiple myeloma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **27**: 4585-4590.

Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, Kim SY, Wardwell L, Tamayo P, Gat-Viks I et al. 2009. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nature genetics* **41**: 1238-1242.

Bea S, Valdes-Mas R, Navarro A, Salaverria I, Martin-Garcia D, Jares P, Gine E, Pinyol M, Royo C, Nadeu F et al. 2013. Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 18250-18255.

Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, Vivanco I, Lee JC, Huang JH, Alexander S et al. 2007. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 20007-20012.

Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M et al. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature* **463**: 899-905.

Bodker JS, Gyrup C, Johansen P, Schmitz A, Madsen J, Johnsen HE, Bogsted M, Dybkaer K, Nyegaard M. 2013. Performance comparison of Affymetrix SNP6.0 and cytogenetic 2.7M whole-genome microarrays in complex cancer samples. *Cytogenet Genome Res* **139**: 80-87.

Bouska A, McKeithan TW, Deffenbacher KE, Lachel C, Wright GW, Iqbal J, Smith LM, Zhang W, Kucuk C, Rinaldi A et al. 2014. Genome-wide copy-number analyses reveal genomic abnormalities involved in transformation of follicular lymphoma. *Blood* **123**: 1681-1690.

Bui MH, Lin X, Albert DH, Li L, Lam LT, Faivre EJ, Warder S, Huang X, Wilcox D, Donawho CK et al. 2017. Preclinical characterization of BET family bromodomain inhibitor ABBV-075 suggests combination therapeutic strategies. *Cancer research* doi:10.1158/0008-5472.CAN-16-1793.

Bullinger L, Kronke J, Schon C, Radtke I, Urlbauer K, Botzenhardt U, Gaidzik V, Cario A, Senger C, Schlenk RF et al. 2010. Identification of acquired copy number alterations and uniparental disomies in cytogenetically normal acute myeloid leukemia using high-resolution single-nucleotide polymorphism analysis. *Leukemia* **24**: 438-449.

Campo E, Swerdlow SH, Harris NL, Pileri S, Stein H, Jaffe ES. 2011. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood* **117**: 5019-5032.

Castro P, Creighton CJ, Ozen M, Berel D, Mims MP, Ittmann M. 2009. Genomic profiling of prostate cancers from African American men. *Neoplasia* **11**: 305-312.

Ceribelli M, Hou ZE, Kelly PN, Huang DW, Wright G, Ganapathi K, Evbuomwan MO, Pittaluga S, Shaffer AL, Marcucci G et al. 2016. A Druggable TCF4- and BRD4-Dependent Transcriptional Network Sustains Malignancy in Blastic Plasmacytoid Dendritic Cell Neoplasm. *Cancer cell* **30**: 764-778.

Challa-Malladi M, Lieu YK, Califano O, Holmes AB, Bhagat G, Murty VV, Dominguez-Sola D, Pasqualucci L, Dalla-Favera R. 2011. Combined genetic inactivation of beta2-Microglobulin and CD58 reveals frequent escape from immune recognition in diffuse large B cell lymphoma. *Cancer cell* **20**: 728-740.

Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, Harview CL, Brunet JP, Ahmann GJ, Adli M et al. 2011. Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**: 467-472.

Chen Y, Takita J, Choi YL, Kato M, Ohira M, Sanada M, Wang L, Soda M, Kikuchi A, Igarashi T et al. 2008. Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* **455**: 971-974.

Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, LeBlanc AC, Donovan DJ, Thung SN, Sole M et al. 2008. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. *Cancer research* **68**: 6779-6788.

Chigrinova E, Rinaldi A, Kwee I, Rossi D, Rancoita PM, Strefford JC, Oscier D, Stamatopoulos K, Papadaki T, Berger F et al. 2013. Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome. *Blood* **122**: 2673-2682.

Compagno M, Lim WK, Grunn A, Nandula SV, Brahmachary M, Shen Q, Bertoni F, Ponzoni M, Scandurra M, Califano A et al. 2009. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **459**: 717-721.

Davis RE, Ngo VN, Lenz G, Tolar P, Young RM, Romesser PB, Kohlhammer H, Lamy L, Zhao H, Yang Y et al. 2010. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463**: 88-92.

Diakos C, Xiao Y, Zheng S, Kager L, Dworzak M, Wiemels JL. 2014. Direct and indirect targets of the E2A-PBX1 leukemia-specific fusion protein. *PloS one* **9**: e87602.

Dickens NJ, Walker BA, Leone PE, Johnson DC, Brito JL, Zeisig A, Jenner MW, Boyd KD, Gonzalez D, Gregory WM et al. 2010. Homozygous deletion mapping in myeloma samples identifies genes and an expression signature relevant to pathogenesis and outcome. *Clinical cancer research : an official journal of the American Association for Cancer Research* **16**: 1856-1864.

Dulak AM, Schumacher SE, van Lieshout J, Imamura Y, Fox C, Shim B, Ramos AH, Saksena G, Baca SC, Baselga J et al. 2012. Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis. *Cancer research* **72**: 4383-4393.

Ernst P, Smale ST. 1995. Combinatorial regulation of transcription II: The immunoglobulin mu heavy chain gene. *Immunity* **2**: 427-438.

474 Forshew T, Tatevossian RG, Lawson AR, Ma J, Neale G, Ogunkolade BW, Jones TA, Aarum J, Dalton J, Bailey
475 S et al. 2009. Activation of the ERK/MAPK pathway: a signature genetic defect in posterior fossa
476 pilocytic astrocytomas. *J Pathol* **218**: 172-181.
477 Fresquet V, Robles EF, Parker A, Martinez-Useros J, Mena M, Malumbres R, Agirre X, Catarino S, Arteta D,
478 Osaba L et al. 2012. High-throughput sequencing analysis of the chromosome 7q32 deletion
479 reveals IRF5 as a potential tumour suppressor in splenic marginal-zone lymphoma. *British journal*
480 *of haematology* **158**: 712-726.
481 Garcia-Ramirez I, Tadros S, Gonzalez-Herrero I, Martin-Lorenzo A, Rodriguez-Hernandez G, Moore D, Ruiz-
482 Roca L, Blanco O, Alonso-Lopez D, De Las Rivas J et al. 2017. Crebbp loss cooperates with Bcl2
483 over-expression to promote lymphoma in mice. *Blood* doi:10.1182/blood-2016-08-733469.
484 Gloury R, Zotos D, Zuidscherwoude M, Masson F, Liao Y, Hasbold J, Corcoran LM, Hodgkin PD, Belz GT, Shi
485 W et al. 2016. Dynamic changes in Id3 and E-protein activity orchestrate germinal center and
486 plasma cell development. *The Journal of experimental medicine* **213**: 1095-1111.
487 Goransson H, Edlund K, Rydaker M, Rasmussen M, Winquist J, Ekman S, Bergqvist M, Thomas A, Lambe
488 M, Rosenquist R et al. 2009. Quantification of normal cell fraction and copy number neutral LOH
489 in clinical lung cancer samples using SNP array data. *PloS one* **4**: e6057.
490 Green MR, Aya-Bonilla C, Gandhi MK, Lea RA, Wellwood J, Wood P, Marlton P, Griffiths LR. 2011.
491 Integrative genomic profiling reveals conserved genetic mechanisms for tumorigenesis in
492 common entities of non-Hodgkin's lymphoma. *Genes, chromosomes & cancer* **50**: 313-326.
493 Green MR, Kihira S, Liu CL, Nair RV, Salari R, Gentles AJ, Irish J, Stehr H, Vicente-Duenas C, Romero-
494 Camarero I et al. 2015. Mutations in early follicular lymphoma progenitors are associated with
495 suppressed antigen presentation. *Proceedings of the National Academy of Sciences of the United*
496 *States of America* **112**: E1116-1125.
497 Green MR, Vicente-Duenas C, Romero-Camarero I, Long Liu C, Dai B, Gonzalez-Herrero I, Garcia-Ramirez
498 I, Alonso-Escudero E, Iqbal J, Chan WC et al. 2014. Transient expression of Bcl6 is sufficient for
499 oncogenic function and induction of mature B-cell lymphoma. *Nature communications* **5**: 3904.
500 Gunnarsson R, Mansouri L, Isaksson A, Goransson H, Cahill N, Jansson M, Rasmussen M, Lundin J, Norin S,
501 Buhl AM et al. 2011. Array-based genomic screening at diagnosis and during follow-up in chronic
502 lymphocytic leukemia. *Haematologica* **96**: 1161-1169.
503 Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* **144**: 646-674.
504 Harder L, Eschenburg G, Zech A, Kriebitzsch N, Otto B, Streichert T, Behlich AS, Dierck K, Klingler B, Hansen
505 A et al. 2013. Aberrant ZNF423 impedes B cell differentiation and is linked to adverse outcome of
506 ETV6-RUNX1 negative B precursor acute lymphoblastic leukemia. *The Journal of experimental*
507 *medicine* **210**: 2289-2304.
508 Hartmann S, Gesk S, Scholtysik R, Kreuz M, Bug S, Vater I, Doring C, Cogliatti S, Parrens M, Merlio JP et al.
509 2010. High resolution SNP array genomic profiling of peripheral T cell lymphomas, not otherwise
510 specified, identifies a subgroup with chromosomal aberrations affecting the REL locus. *British*
511 *journal of haematology* **148**: 402-412.
512 Haverty PM, Fridlyand J, Li L, Getz G, Beroukhim R, Lohr S, Wu TD, Cavet G, Zhang Z, Chant J. 2008. High-
513 resolution genomic and expression analyses of copy number alterations in breast tumors. *Genes,*
514 *chromosomes & cancer* **47**: 530-542.
515 Haverty PM, Hon LS, Kaminker JS, Chant J, Zhang Z. 2009. High-resolution analysis of copy number
516 alterations and associated expression changes in ovarian tumors. *BMC Med Genomics* **2**: 21.
517 Hu N, Wang C, Ng D, Clifford R, Yang HH, Tang ZZ, Wang QH, Han XY, Giffen C, Goldstein AM et al. 2009.
518 Genomic characterization of esophageal squamous cell carcinoma from a high-risk population in
519 China. *Cancer research* **69**: 5908-5917.

Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC et al. 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic acids research* **35**: W169-175.

Huang YT, Lin X, Chirieac LR, McGovern R, Wain JC, Heist RS, Skaug V, Zienolddiny S, Haugen A, Su L et al. 2011. Impact on disease development, genomic location and biological function of copy number alterations in non-small cell lung cancer. *PloS one* **6**: e22961.

Iida S, Rao PH, Butler M, Corradini P, Boccadoro M, Klein B, Chaganti RS, Dalla-Favera R. 1997. Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma. *Nature genetics* **17**: 226-230.

Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**: 118-127.

Joos S, Otano-Joos MI, Ziegler S, Bruderlein S, du Manoir S, Bentz M, Moller P, Lichter P. 1996. Primary mediastinal (thymic) B-cell lymphoma is characterized by gains of chromosomal material including 9p and amplification of the REL gene. *Blood* **87**: 1571-1578.

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC Table Browser data retrieval tool. *Nucleic acids research* **32**: D493-496.

Kato M, Sanada M, Kato I, Sato Y, Takita J, Takeuchi K, Niwa A, Chen Y, Nakazaki K, Nomoto J et al. 2009. Frequent inactivation of A20 in B-cell lymphomas. *Nature* **459**: 712-716.

Kawamata N, Ogawa S, Gueller S, Ross SH, Huynh T, Chen J, Chang A, Nabavi-Nouis S, Megrabian N, Siebert R et al. 2009. Identified hidden genomic changes in mantle cell lymphoma using high-resolution single nucleotide polymorphism genomic array. *Experimental hematology* **37**: 937-946.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research* **22**: 568-576.

Kondo Y, Nagai K, Nakahata S, Saito Y, Ichikawa T, Suekane A, Taki T, Iwakawa R, Enari M, Taniwaki M et al. 2012. Overexpression of the DNA sensor proteins, absent in melanoma 2 and interferon-inducible 16, contributes to tumorigenesis of oral squamous cell carcinoma with p53 inactivation. *Cancer Sci* **103**: 782-790.

Kuilman T, Velds A, Kemper K, Ranzani M, Bombardelli L, Hoogstraat M, Nevedomskaya E, Xu G, de Ruiter J, Lolkema MP et al. 2015. CopywriteR: DNA copy number detection from off-target sequence data. *Genome biology* **16**: 49.

Kuiper RP, Schoenmakers EF, van Reijmersdal SV, Hehir-Kwa JY, van Kessel AG, van Leeuwen FN, Hoogerbrugge PM. 2007. High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. *Leukemia* **21**: 1258-1266.

Kuiper RP, Waanders E, van der Velden VH, van Reijmersdal SV, Venkatachalam R, Scheijen B, Sonneveld E, van Dongen JJ, Veerman AJ, van Leeuwen FN et al. 2010. IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. *Leukemia* **24**: 1258-1264.

Kuppers R. 2005. Mechanisms of B-cell lymphoma pathogenesis. *Nature reviews Cancer* **5**: 251-262.

Lenz G, Davis RE, Ngo VN, Lam L, George TC, Wright GW, Dave SS, Zhao H, Xu W, Rosenwald A et al. 2008a. Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* **319**: 1676-1679.

Lenz G, Wright G, Dave SS, Xiao W, Powell J, Zhao H, Xu W, Tan B, Goldschmidt N, Iqbal J et al. 2008b. Stromal gene signatures in large-B-cell lymphomas. *The New England journal of medicine* **359**: 2313-2323.

Lenz G, Wright GW, Emre NC, Kohlhammer H, Dave SS, Davis RE, Carty S, Lam LT, Shaffer AL, Xiao W et al. 2008c. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 13520-13525.

568 Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. 2016. An interactive environment for agile analysis
569 and visualization of ChIP-sequencing data. *Nature structural & molecular biology* **23**: 349-357.
570 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWAMEM. *arXiv* **1303**:
571 3997.
572 Li H, Kaminski MS, Li Y, Yildiz M, Ouillette P, Jones S, Fox H, Jacobi K, Saiya-Cork K, Bixby D et al. 2014.
573 Mutations in linker histone genes HIST1H1 B, C, D, and E; OCT2 (POU2F2); IRF8; and ARID1A
574 underlying the pathogenesis of follicular lymphoma. *Blood* **123**: 1487-1498.
575 Lilljebjorn H, Soneson C, Andersson A, Heldrup J, Behrendtz M, Kawamata N, Ogawa S, Koeffler HP,
576 Mitelman F, Johansson B et al. 2010. The correlation pattern of acquired copy number changes in
577 164 ETV6/RUNX1-positive childhood acute lymphoblastic leukemias. *Hum Mol Genet* **19**: 3150-
578 3158.
579 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S,
580 Daly M et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-
581 generation DNA sequencing data. *Genome research* **20**: 1297-1303.
582 Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. 2011. GISTIC2.0 facilitates
583 sensitive and confident localization of the targets of focal somatic copy-number alteration in
584 human cancers. *Genome biology* **12**: R41.
585 Messina M, Del Giudice I, Khiabanian H, Rossi D, Chiaretti S, Rasi S, Spina V, Holmes AB, Marinelli M, Fabbri
586 G et al. 2014. Genetic lesions associated with chronic lymphocytic leukemia chemo-refractoriness.
587 *Blood* **123**: 2378-2388.
588 Micke P, Edlund K, Holmberg L, Kultima HG, Mansouri L, Ekman S, Bergqvist M, Scheibenflug L, Lamberg
589 K, Myrdal G et al. 2011. Gene copy number aberrations are associated with survival in histologic
590 subgroups of non-small cell lung cancer. *J Thorac Oncol* **6**: 1833-1840.
591 Miller S, Rogers HA, Lyon P, Rand V, Adamowicz-Brice M, Clifford SC, Hayden JT, Dyer S, Pfister S,
592 Korshunov A et al. 2011. Genome-wide molecular characterization of central nervous system
593 primitive neuroectodermal tumor and pineoblastoma. *Neuro Oncol* **13**: 866-879.
594 Monni O, Joensuu H, Franssila K, Klefstrom J, Alitalo K, Knuutila S. 1997. BCL2 overexpression associated
595 with chromosomal amplification in diffuse large B-cell lymphoma. *Blood* **90**: 1168-1174.
596 Monti S, Chapuy B, Takeyama K, Rodig SJ, Hao Y, Yeda KT, Inguilizian H, Mermel C, Currie T, Dogan A et al.
597 2012. Integrative analysis reveals an outcome-associated and targetable pattern of p53 and cell
598 cycle deregulation in diffuse large B cell lymphoma. *Cancer cell* **22**: 359-372.
599 Mosca L, Fabris S, Lionetti M, Todoerti K, Agnelli L, Morabito F, Cutrona G, Andronache A, Matis S, Ferrari
600 F et al. 2010. Integrative genomics analyses reveal molecularly distinct subgroups of B-cell chronic
601 lymphocytic leukemia patients with 13q14 deletion. *Clinical cancer research : an official journal of*
602 *the American Association for Cancer Research* **16**: 5641-5653.
603 Mullighan CG. 2012. Molecular genetics of B-precursor acute lymphoblastic leukemia. *The Journal of*
604 *clinical investigation* **122**: 3407-3415.
605 Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J,
606 Pounds SB et al. 2007. Genome-wide analysis of genetic alterations in acute lymphoblastic
607 leukaemia. *Nature* **446**: 758-764.
608 Mullighan CG, Miller CB, Radtke I, Phillips LA, Dalton J, Ma J, White D, Hughes TP, Le Beau MM, Pui CH et
609 al. 2008a. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature*
610 **453**: 110-114.
611 Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, Downing JR. 2008b. Genomic analysis of the
612 clonal origins of relapsed acute lymphoblastic leukemia. *Science* **322**: 1377-1380.
613 Mullighan CG, Su X, Zhang J, Radtke I, Phillips LA, Miller CB, Ma J, Liu W, Cheng C, Schulman BA et al. 2009.
614 Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *The New England journal of*
615 *medicine* **360**: 470-480.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE et al. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**: 272-276.

Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao W, Lim KH, Kohlhammer H, Xu W, Yang Y, Zhao H et al. 2011. Oncogenically active MYD88 mutations in human lymphoma. *Nature* **470**: 115-119.

Northcott PA, Nakahara Y, Wu X, Feuk L, Ellison DW, Croul S, Mack S, Kongkham PN, Peacock J, Dubuc A et al. 2009. Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nature genetics* **41**: 465-472.

Nowak D, Ogawa S, Muschen M, Kato M, Kawamata N, Meixel A, Nowak V, Kim HS, Kang S, Paquette R et al. 2010. SNP array analysis of tyrosine kinase inhibitor-resistant chronic myeloid leukemia identifies heterogeneous secondary genomic alterations. *Blood* **115**: 1049-1053.

Okosun J, Bodor C, Wang J, Araf S, Yang CY, Pan C, Boller S, Cittaro D, Bozek M, Iqbal S et al. 2014. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nature genetics* **46**: 176-181.

Oricchio E, Nanjangud G, Wolfe AL, Schatz JH, Mavrakis KJ, Jiang M, Liu X, Bruno J, Heguy A, Olshen AB et al. 2011. The Eph-receptor A7 is a soluble tumor suppressor for follicular lymphoma. *Cell* **147**: 554-564.

Pasqualucci L, Compagno M, Houldsworth J, Monti S, Grunn A, Nandula SV, Aster JC, Murty VV, Shipp MA, Dalla-Favera R. 2006. Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *The Journal of experimental medicine* **203**: 311-317.

Paulsson K, Cazier JB, Macdougall F, Stevens J, Stasevich I, Vrcelj N, Chaplin T, Lillington DM, Lister TA, Young BD. 2008. Microdeletions are a general feature of adult and adolescent acute lymphoblastic leukemia: Unexpected similarities with pediatric disease. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 6708-6713.

Poage GM, Christensen BC, Houseman EA, McClean MD, Wiencke JK, Posner MR, Clark JR, Nelson HH, Marsit CJ, Kelsey KT. 2010. Genetic and epigenetic somatic alterations in head and neck squamous cell carcinomas are globally coordinated but not locally targeted. *PloS one* **5**: e9651.

Radtke I, Mullighan CG, Ishii M, Su X, Cheng J, Ma J, Ganti R, Cai Z, Goorha S, Pounds SB et al. 2009. Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 12944-12949.

Rao PH, Houldsworth J, Dyomina K, Parsa NZ, Cigudosa JC, Louie DC, Popplewell L, Offit K, Jhanwar SC, Chaganti RS. 1998. Chromosomal and gene amplification in diffuse large B-cell lymphoma. *Blood* **92**: 234-240.

Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. 2006. GenePattern 2.0. *Nature genetics* **38**: 500-501.

Reid JF, Gariboldi M, Sokolova V, Capobianco P, Lampis A, Perrone F, Signoroni S, Costa A, Leo E, Pilotti S et al. 2009. Integrative approach for prioritizing cancer genes in sporadic colon cancer. *Genes, chromosomes & cancer* **48**: 953-962.

Rinaldi A, Mian M, Chigrinova E, Arcaini L, Bhagat G, Novak U, Rancoita PM, De Campos CP, Forconi F, Gascoyne RD et al. 2011. Genome-wide DNA profiling of marginal zone lymphomas identifies subtype-specific lesions with an impact on the clinical outcome. *Blood* **117**: 1595-1604.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature biotechnology* **29**: 24-26.

Rohde KH, Veiga DF, Caldwell S, Balazsi G, Russell DG. 2012. Linking the transcriptional profiles and the physiological states of Mycobacterium tuberculosis during an extended intracellular infection. *PLoS pathogens* **8**: e1002769.

Rosenwald A, Wright G, Wiestner A, Chan WC, Connors JM, Campo E, Gascoyne RD, Grogan TM, Muller-Hermelink HK, Smeland EB et al. 2003. The proliferation gene expression signature is a

664 quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer*
665 *cell* **3**: 185-197.
666 Rucker FG, Schlenk RF, Bullinger L, Kayser S, Teleanu V, Kett H, Habdank M, Kugler CM, Holzmann K,
667 Gaidzik VI et al. 2012. TP53 alterations in acute myeloid leukemia with complex karyotype
668 correlate with specific copy number alterations, monosomal karyotype, and dismal outcome.
669 *Blood* **119**: 2114-2121.
670 Salaverria I, Royo C, Carvajal-Cuenca A, Clot G, Navarro A, Valera A, Song JY, Woroniecka R, Rymkiewicz
671 G, Klapper W et al. 2013. CCND2 rearrangements are the most frequent genetic events in cyclin
672 D1(-) mantle cell lymphoma. *Blood* **121**: 1394-1402.
673 Scandurra M, Mian M, Greiner TC, Rancoita PM, De Campos CP, Chan WC, Vose JM, Chigrinova E, Inghirami
674 G, Chiappella A et al. 2010. Genomic lesions associated with a different clinical outcome in diffuse
675 large B-Cell lymphoma treated with R-CHOP-21. *British journal of haematology* **151**: 221-231.
676 Schraders M, van Reijmersdal SV, Kamping EJ, van Krieken JH, van Kessel AG, Groenen PJ, Hoogerbrugge
677 PM, Kuiper RP. 2009. High-resolution genomic profiling of pediatric lymphoblastic lymphomas
678 reveals subtle differences with pediatric acute lymphoblastic leukemias in the B-lineage. *Cancer*
679 *genetics and cytogenetics* **191**: 27-33.
680 Scotto L, Narayan G, Nandula SV, Arias-Pulido H, Subramaniyam S, Schneider A, Kaufmann AM, Wright JD,
681 Pothuri B, Mansukhani M et al. 2008. Identification of copy number gain and overexpressed genes
682 on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in
683 progression. *Genes, chromosomes & cancer* **47**: 755-765.
684 Shaffer AL, Rosenwald A, Staudt LM. 2002. Lymphoid malignancies: the dark side of B-cell differentiation.
685 *Nature reviews Immunology* **2**: 920-932.
686 Sos ML, Michel K, Zander T, Weiss J, Frommolt P, Peifer M, Li D, Ullrich R, Koker M, Fischer F et al. 2009.
687 Predicting drug susceptibility of non-small cell lung cancers based on genetic lesions. *The Journal*
688 *of clinical investigation* **119**: 1727-1740.
689 Tosello V, Mansour MR, Barnes K, Paganin M, Sulis ML, Jenkinson S, Allen CG, Gale RE, Linch DC, Palomero
690 T et al. 2009. WT1 mutations in T-ALL. *Blood* **114**: 1038-1045.
691 Walker BA, Leone PE, Chiecchio L, Dickens NJ, Jenner MW, Boyd KD, Johnson DC, Gonzalez D, Dagrada GP,
692 Protheroe RK et al. 2010. A compendium of myeloma-associated chromosomal copy number
693 abnormalities and their prognostic value. *Blood* **116**: e56-65.
694 Wang JQ, Jeelall YS, Humburg P, Batchelor EL, Kaya SM, Yoo HM, Goodnow CC, Horikawa K. 2017.
695 Synergistic cooperation and crosstalk between MYD88L265P and mutations that dysregulate
696 CD79B and surface IgM. *The Journal of experimental medicine* doi:10.1084/jem.20161454.
697 Wilkerson MD, Yin X, Walter V, Zhao N, Cabanski CR, Hayward MC, Miller CR, Socinski MA, Parsons AM,
698 Thorne LB et al. 2012. Differential pathogenesis of lung adenocarcinoma subtypes involving
699 sequence mutations, copy number, chromosomal instability, and methylation. *PloS one* **7**:
700 e36530.
701 Willis TG, Dyer MJ. 2000. The role of immunoglobulin translocations in the pathogenesis of B-cell
702 malignancies. *Blood* **96**: 808-822.
703 Wilson WH, Young RM, Schmitz R, Yang Y, Pittaluga S, Wright G, Lih CJ, Williams PM, Shaffer AL, Gerecitano
704 J et al. 2015. Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma.
705 *Nature medicine* **21**: 922-926.
706 Wohner M, Tagoh H, Bilic I, Jaritz M, Poliakova DK, Fischer M, Busslinger M. 2016. Molecular functions of
707 the transcription factors E2A and E2-2 in controlling germinal center B cell and plasma cell
708 development. *The Journal of experimental medicine* **213**: 1201-1221.
709 Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. 2003. A gene expression-based method
710 to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proceedings of the*
711 *National Academy of Sciences of the United States of America* **100**: 9991-9996.

712 Yamagishi M, Nakano K, Miyake A, Yamochi T, Kagami Y, Tsutsumi A, Matsuda Y, Sato-Otsubo A, Muto S,
713 Utsunomiya A et al. 2012. Polycomb-mediated loss of miR-31 activates NIK-dependent NF-kappaB
714 pathway in adult T cell leukemia and other cancers. *Cancer cell* **21**: 121-135.
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738

FIGURES AND LEGENDS

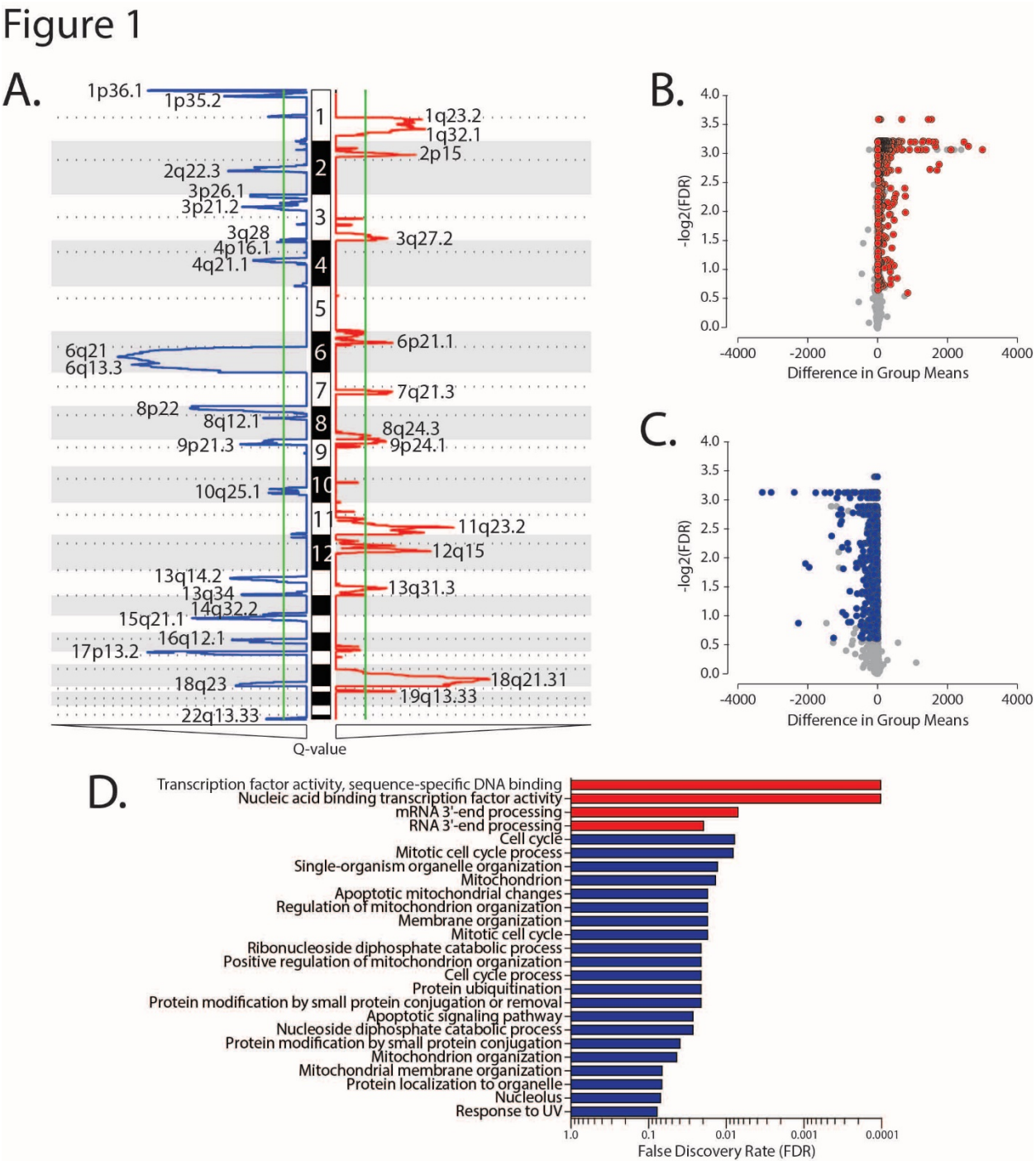

Figure 1: Transcription factors are significantly enriched targets of DNA copy number gains in DLBCL. A) A GISTIC plot shows significant peaks of DNA copy number loss (blue) and gain (red) identified from 673 DLBCL tumors, and labelled by their cytogenetic position. The green

line indicates a significance threshold of FDR = 0.25. **B)** A volcano plot of transcriptional changes of genes within peaks of DNA copy number gain in patients with each lesion compared to patients with diploid copy number. Each point represents a single gene, with red points being significantly increased in association with DNA copy number gain (FDR<0.25, fold-change≥1.2). **C)** A volcano plot of transcriptional changes of genes within peaks of DNA copy number loss in patients with each lesion compared to patients with diploid copy number. Each point represents a single gene, with blue points being genes that are significantly reduced in association with DNA copy number loss (FDR<0.25, fold-change≥1.2). **D)** Hypergeometric enrichment analysis of genes targeted by DNA copy number gain (red) and loss (blue). Genes with transcription factor activity were significantly enriched as targets of DNA copy number gain.

Figure 2

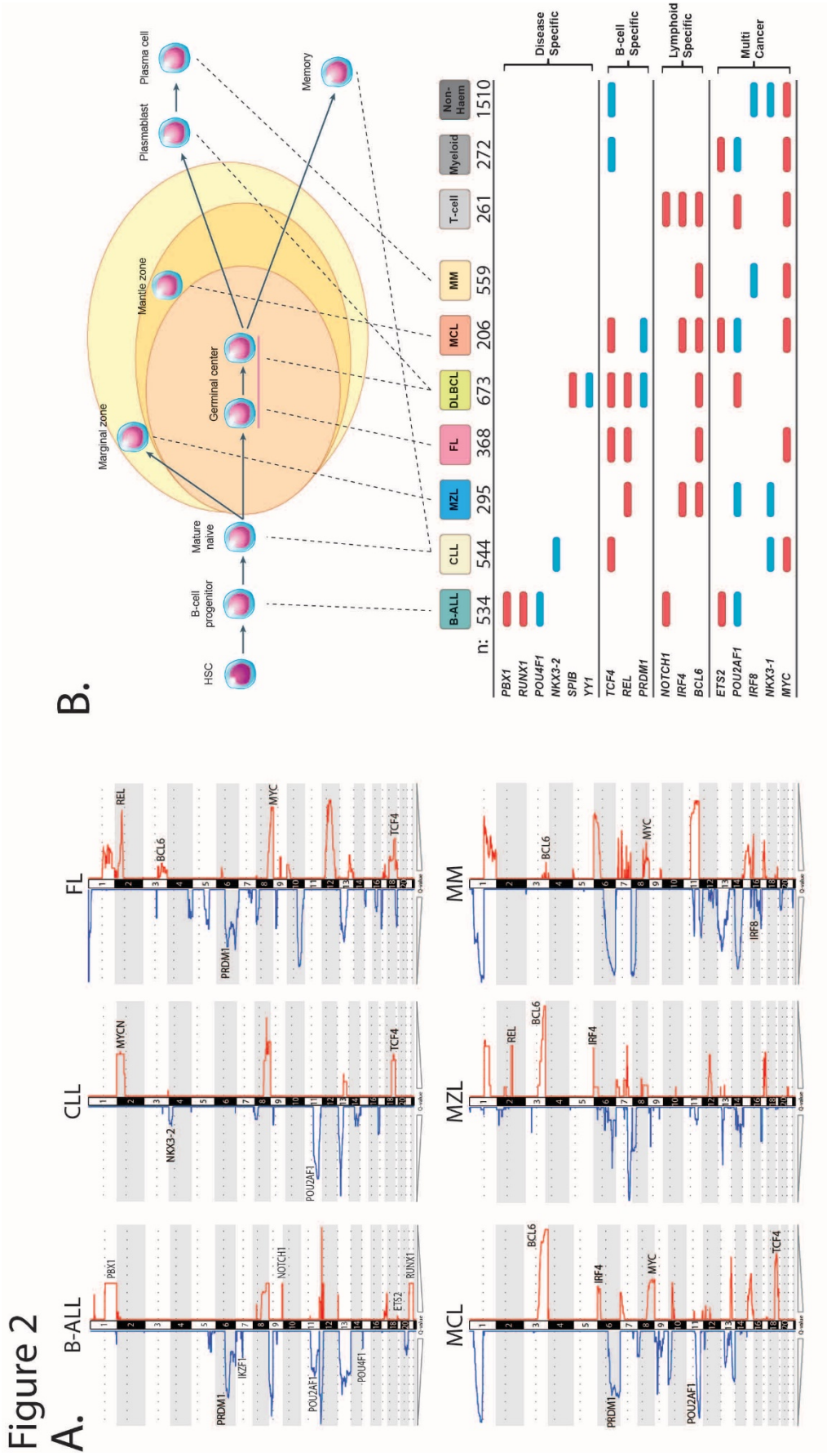

Figure 2: Transcription factor genes are targeted by DNA copy number alterations across the spectrum of B-cell malignancies. A) GISTIC analysis of 534 B-cell acute lymphoblastic leukemia (B-ALL), 544 chronic lymphocytic leukemia (CLL), 368 follicular lymphoma (FL), 206 mantle cell lymphoma (MCL), 295 splenic/nodal marginal zone lymphoma (MZL) and 559 multiple myeloma. Peaks of DNA copy number gain (red) or loss (blue) that target transcription factors with a well-defined role in one or more stages of B-cell development are annotated with the gene symbol of the transcription factor. **B)** The representation of transcription factor alterations across the spectrum of B-cell malignancies is shown with reference to their putative 'normal counterpart'. DNA copy number alterations of some transcription factors are specific to a single disease. Others are found in multiple B-cell malignancies but not in T-cell, myeloid or non-haematological tumors. Alterations of *NOTCH1*, *IRF4* and *BCL6* are found in tumors of both the B- and T-lineage. DNA copy number changes encompassing other transcription factors are found across cancers of multiple lineages.

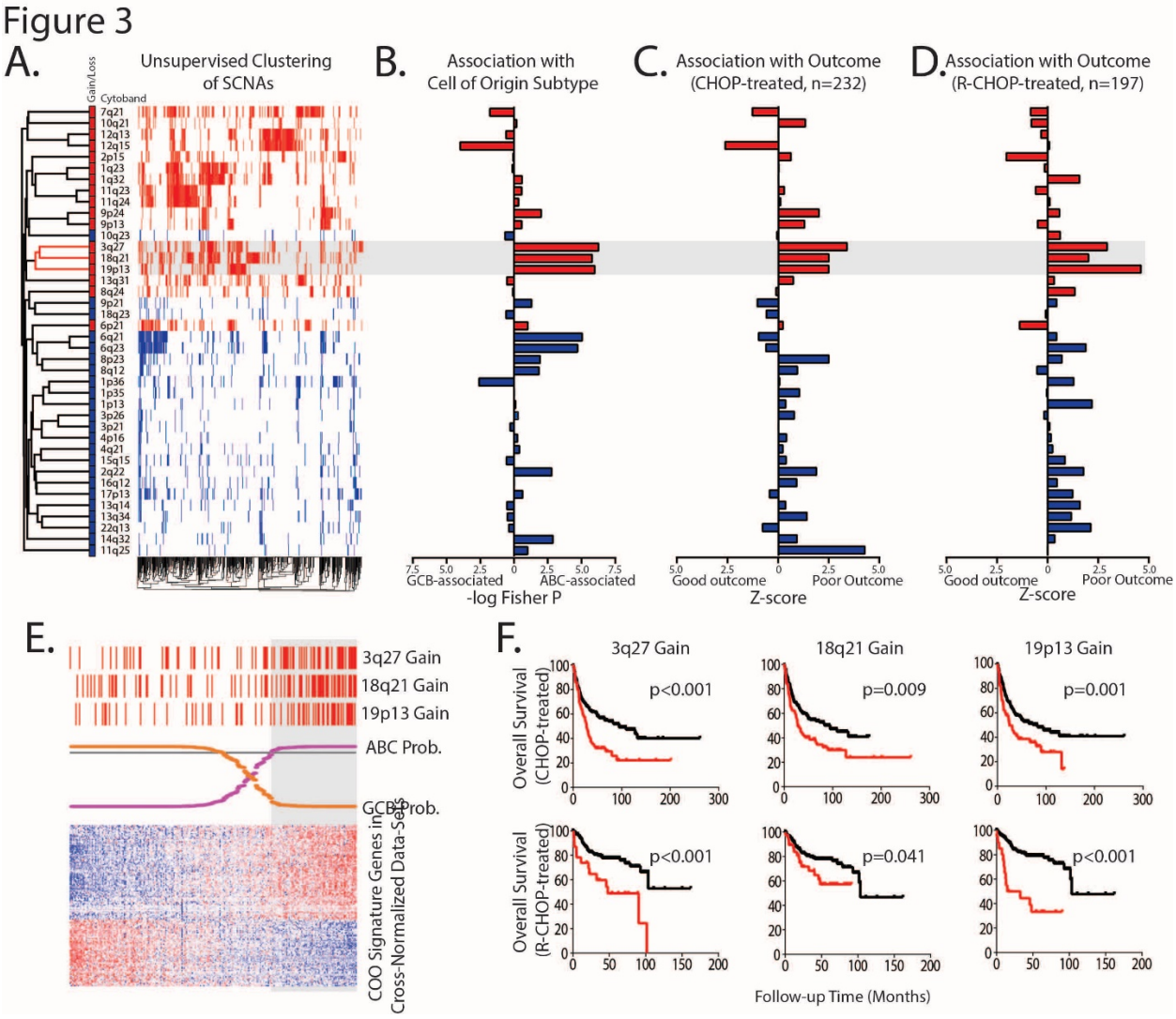

Figure 3: Co-segregating transcription factor alterations associate with the ABC-like subtype and poor outcome in DLBCL. A) Unsupervised hierarchical clustering of DNA copy number gains (red) and losses (blue) in DLBCL tumors with ≥2 SCNAs. DNA copy number gains of 3q27, 18q21 and 19p13 can be seen to cluster together, indicating co-segregation within the same tumors. **B)** A bar graph illustrates the Fisher P-value for association of each alteration with either the GCB-like (left) or ABC-like (right) cell of origin subtype. DNA copy number gains of 3q27, 18q21 and 19p13 are each significantly associated with the ABC-like subtype. **C)** A bar graph illustrates the Z-score for association of each alteration with good (left) or poor (right) outcome in 232 DLBCL patients treated with the CHOP combination chemotherapy regimen.

DNA copy number gains of 3q27, 18q21 and 19p13 are each significantly associated with poor outcome. **D)** A bar graph illustrates the Z-score for association of each alteration with good (left) or poor (right) outcome in 197 DLBCL patients treated with the CHOP regimen plus rituximab (R-CHOP). DNA copy number gains of 3q27, 18q21 and 19p13 are each significantly associated with poor outcome. **E)** A heatmap showing the cell of origin subtype signatures and scores determined by the 140-gene Wright algorithm, and the presence of 3q27, 18q21 and 19p13 DNA copy number gains (top, red) in tumors classified as ABC-like (right, shaded grey). This shows that the majority of ABC-like tumors possess one or more of these lesions, but they are not exclusive to this subtype. **F)** Kaplan-Meier curves show the overall survival of DLBCL patients treated with CHOP (above) or R-CHOP (below), separated by the presence (red) or absence (black) of DNA copy number gains at 3q27 (left), 18q21 (middle), or 19p13 (right). P-values were calculated using a log-rank test.

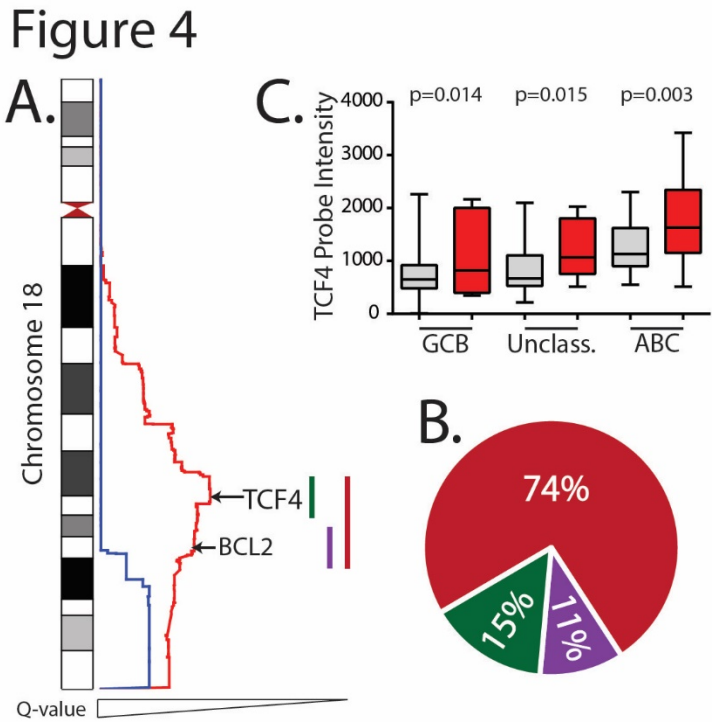

Figure 4: DNA copy number gains of 18q21 target the *TCF4* (E2-2) gene. A) The GISTIC scores for DNA copy number gain (red) and loss (blue) are shown for chromosome 18. It can be seen that the peak with the lowest (most significant) Q-value calculated by GISTIC includes the *TCF4* gene, but not *BCL2*. **B)** A pie graph of cases with 18q DNA copy number gains, separated into those that show gains of both *TCF4* and *BCL2* (red), gains of *TCF4* only (green) and gains of *BCL2* only. Although the peak including *TCF4* is the most significant by GISTIC, the majority of tumors harbor DNA copy number gains that encompass both *TCF4* and *BCL2*. **C)** Box plots show the gene expression microarray probe intensity of TCF4 in cases with diploid 18q21 (grey) or *TCF4* DNA copy number gains (red), separated by cell of origin subtype. *TCF4* shows significantly higher expression in cases with DNA copy number gain, irrespective of molecular subtype, but a higher level in ABC-like DLBCL tumors. P-values were calculated using a 2-tailed student's T-test.

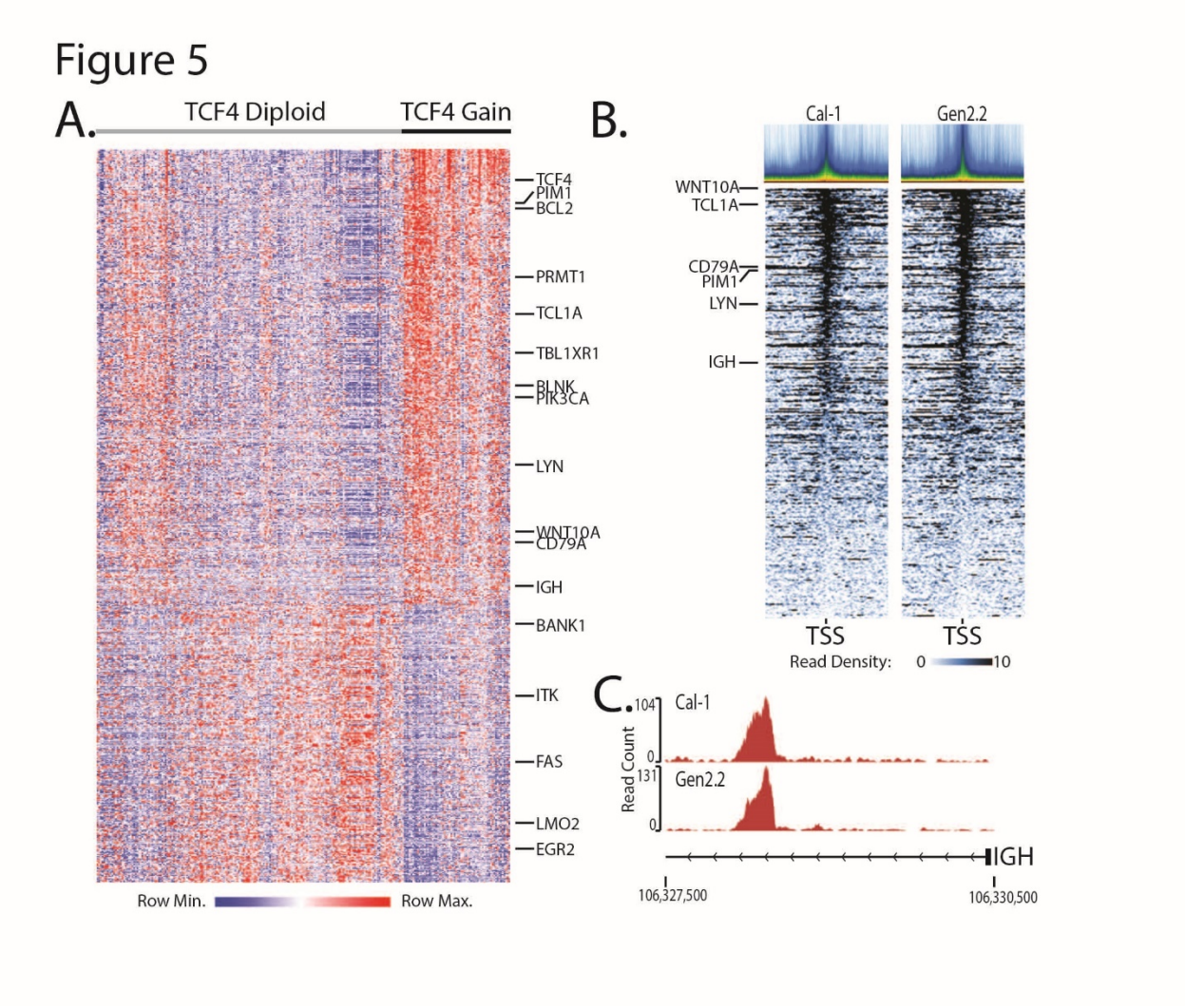

Figure 5: DNA copy number gain of TCF4 is associated with increased expression of B-cell receptor signaling components. A) A heatmap shows significantly differentially expressed genes between cases with diploid *TCF4* and *TCF4* DNA copy number gain. This increased expression of components of the B-cell receptor signaling pathway in tumors with TCF4 gain, including *IGH*, *CD79A*, *LYN*, *BLNK*, and *PIK3CA*. **B)** A heatmap of TCF4 binding to the transcription start site (TSS, +/- 10kb) of genes with increased expression associated with *TCF4* DNA copy number gain. This confirms binding of TCF4 to the regulatory region of the majority of genes with increased expression associated with DNA copy number gain of *TCF4*, including IGH, and therefore indicates a direct link between *TCF4* copy gain and the associated transcriptional

858 changes. **C)** A diagram of the TCF4 peak downstream from the *IGH* transcription start site,
859 highlighting strong binding at a likely enhancer element.
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883

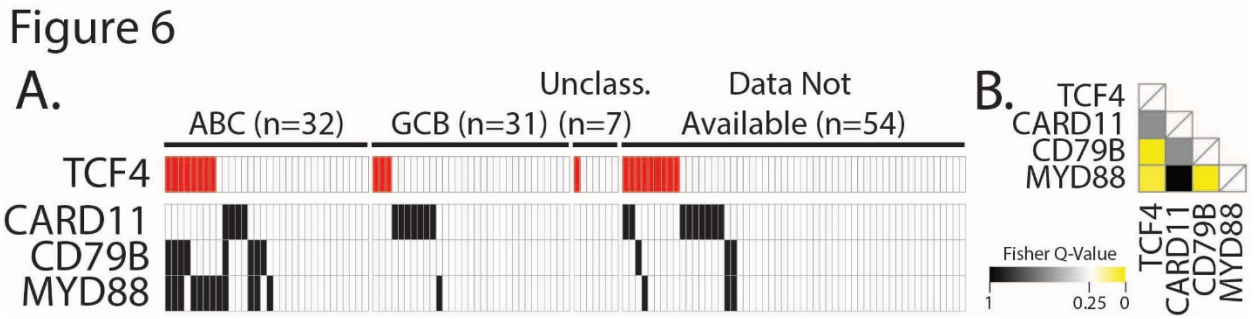

Figure 6: *TCF4* DNA copy number gain co-associated with other BCR signaling mutations in ABC-like DLBCL. A) A heatmap shows the presence (colored) or absence (white) of *TCF4* DNA copy number gain and somatic mutations of *CARD11*, *CD79B* and *MYD88* in 124 DLBCL tumors analyzed by targeted sequencing. The data are ordered by cell of origin subtype for the 70 cases with available matched gene expression profiling data. **B)** Fisher test Q-values are shown for co-association between each of the lesions in A. It can be seen that *TCF4* DNA copy number gain significantly co-associates with both *CD79B* and *MYD88* mutation, and that *CD79B* and *MYD88* mutations also significantly co-associate.