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 Hartert¹, Saber Tadros¹,²#, Alyssa Bouska³#, Dalia Moore¹, Tayla Heavican³, Christine Pak⁴, Chih Long Liu⁵, Andrew Gentles⁶, Elena Hartmann⁷,⁸, Robert Kridel⁹, Karin Ekstrom Smedby¹⁰, Gunnar Juliussön¹¹, David Klinkebiel¹², Richard Rosenquist¹³, Randy Gascoyne¹⁴, Andreas Rosenwald⁷,⁸, Martin Bast¹⁵, Julie Vose¹⁵, Matthew Lunning¹⁵, Timothy Greiner³, Scott Rodig⁴, Javeed Iqbal⁶, Ash Alizadeh⁵, Michael R. Green¹,²,¹⁶*

¹Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE, USA; ²Current Affiliation: Department of Lymphoma/Myeloma, University of Texas MD Anderson Cancer Center, Houston, TX, USA; ³Department of Pathology and Immunology, University of Nebraska Medical Center, Omaha, NE, USA; ⁴Department of Pathology, Brigham and Womens Hospital, Harvard Medical School, Boston, MA, USA; ⁵Division of Oncology, Department of Medicine, Stanford University, Stanford, CA, USA; ⁶Department of Radiology, Stanford University, Stanford, CA, USA; ⁷Institute of Pathology, University of Würzburg, Würzburg, Germany; ⁸Comprehensive Cancer Center Mainfranken. Wurzburg, Germany; ⁹Princess Margaret Cancer Center, University of Toronto, Toronto, ON, Canada; ¹⁰Department of Medicine Solna, Clinical Epidemiology Unit, Karolinska Institutet, and Hematology Center, Karolinska University Hospital, Stockholm, Sweden; ¹¹Department of Laboratory Medicine, Stem Cell Center, Lund University, Lund, Sweden; ¹²Department of Biochemistry and Molecular Biology, University of Nebraska Medical Center, Omaha, NE, USA; ¹³Department of Molecular Medicine and Surgery, Karolinska Universitetssjukhuset, Stockholm, Sweden; ¹⁴Center for Lymphoid Cancers, British Columbia Cancer Agency, Vancouver, BC, Canada; ¹⁵Division of Hematology and Oncology, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, NE, USA.; ¹⁶Current 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 \textit{IKZF1} (Mullighan et al. 2009),
\textit{REL} (Joos et al. 1996), \textit{PRDM1} (Pasqualucci et al. 2006), \textit{BCL6} (Green et al. 2014) and \textit{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 \textit{POU2AF1} (OBF1), \textit{NKX3-1}, \textit{NKX3-2}
and \textit{IRF8}, and gains of \textit{PBX1}, \textit{IRF4}, and \textit{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 increased 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 focuses 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

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


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
A quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer cell* **3**: 185-197.


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: 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
changes. C) A diagram of the TCF4 peak downstream from the IGH transcription start site, highlighting strong binding at a likely enhancer element.
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.