# Pathognomonic and epistatic genetic alterations in B-cell non-Hodgkin lymphoma

Man Chun John Ma[1¥], Saber Tadros[1¥], Alyssa Bouska[2], Tayla B. Heavican[2], Haopeng Yang[1], Qing Deng[1], Dalia Moore[3], Ariz Akhter[4], Keenan Hartert[3], Neeraj Jain[1], Jordan Showell[1], Sreejoyee Ghosh[1], Lesley Street[5], Marta Davidson[5], Christopher Carey[6], Joshua Tobin[7], Deepak Perumal[8], Julie M. Vose[9], Matthew A. Lunning[9], Aliyah R. Sohani[10], Benjamin J. Chen[11], Shannon Buckley[12], Loretta J. Nastoupil[1], R. Eric Davis[1], Jason R. Westin[1], Nathan H. Fowler[1], Samir Parekh[8], Maher K. Gandhi[7], Sattva S. Neelapu[1], Douglas Stewart[5], Javeed Iqbal[2], Timothy Greiner[2], Scott J. Rodig[13], Adnan Mansoor[5], Michael R. Green[1,14,15]*

[1]Department of Lymphoma and Myeloma, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; [2]Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE, USA; [3]Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE, USA; [4]Department of Pathology and Laboratory Medicine, University of Calgary, Calgary, AB, Canada; [5]Section of Hematology, Department of Medicine, University of Calgary, Calgary, AB, Canada; [6]Northern Institute for Research, Newcastle University, Newcastle upon Tyne, England; [7]Mater Research, University of Queensland, QLD, Australia; [8]Division of Hematology and Medical Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [9]Department of Internal Medicine, Division of Hematology-Oncology, University of Nebraska Medical Center, Omaha, NE, USA; [10]Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; [11]Department of Pathology, University of Massachusetts Medical School, UMass Memorial Medical Center, Worcester, MA, USA; [12]Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, NE, USA; [13]Department of Pathology, Brigham and Womens Hospital, Boston, MA, USA; [14]Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [15]Center for Cancer Epigenetics, University of Texas MD Anderson Cancer Center, Houston, TX, USA.

[¥]Equally contributed

*Corresponding Author

Michael R. Green, Ph.D.

Departments of Lymphoma/Myeloma and Genomic Medicine,

University of Texas MD Anderson Cancer Center,

1515 Holcombe Blvd, Unit 903,

Houston, TX 77030, USA

Ph.: +1-713-745-4244

Email: mgreen5@mdanderson.org

## KEY POINTS

1. Genetic perturbation of the ubiquitin proteasome system is an emerging hallmark of B-cell non-Hodgkin lymphoma (B-NHL).
2. Co-occurring sets of genetic alterations define B-NHL subtypes and likely represent epistatic interactions.

## ABSTRACT

B-cell non-Hodgkin lymphoma (B-NHL) encompasses multiple clinically and phenotypically distinct subtypes of malignancy with unique molecular etiologies. Common subtypes of B-NHL such as diffuse large B-cell lymphoma (DLBCL) have been comprehensively interrogated at the genomic level, but other less common subtypes such as mantle cell lymphoma (MCL) remain sparsely characterized. Furthermore, multiple B-NHL subtypes have thus far not been comprehensively compared to identify conserved or subtype-specific patterns of genomic alterations. Here, we employed a large targeted hybrid-capture sequencing approach encompassing 380 genes to interrogate the genomic landscapes of 755 B-NHL tumors at high depth; primarily including DLBCL, MCL, follicular lymphoma (FL), and Burkitt lymphoma (BL). We identified conserved hallmarks of B-NHL that were deregulated across major subtypes, such as the frequent genetic deregulation of the ubiquitin proteasome system (UPS). In addition, we identified subtype-specific patterns of genetic alterations, including clusters of co-occurring mutations that are pathognomonic. The cumulative burden of mutations within a single cluster were more significantly discriminatory of B-NHL subtypes than individual mutations, implicating likely patterns of genetic epistasis that contribute to disease etiology. We therefore provide a framework of co-occurring mutations that deregulate genetic hallmarks and likely cooperate in lymphomagenesis of B-NHL subtypes.

## INTRODUCTION

Non-Hodgkin lymphomas (NHL) are a heterogeneous group of lymphoid malignancies that predominantly arise from mature B-cells (B-NHL). Although mature B-cell neoplasms encompass 38 unique diagnostic subtypes, over 85% of cases fall within only 7 subtypes[1,2]. Recent next generation sequencing (NGS) studies have shed light onto the key driver mutations in many of these NHL subtypes; for example, large studies of diffuse large B-cell lymphoma (DLBCL) have led to proposed genomic subtypes that have unique etiologies[3-5]. However, many less-common NHL subtypes such as mantle cell lymphoma (MCL) have not been as extensively characterized[6,7]. Furthermore, until recently[3,4] genetic alterations have been considered in a binary fashion as either driver events, which directly promote disease genesis or progression, or passenger events, which have little or no impact on disease biology. In contrast to this principle, most B-NHLs do not result from a single dominant driver but instead result from the serial acquisition of genetic alterations that cooperate in lymphomagenesis. The genetic context of each mutation therefore likely determines its oncogenic potential, and groups of mutations should therefore be considered collectively rather than as singular events. For example, the 'C5' and 'MCD' clusters identified in DLBCL by Chapuy *et al.* and Schmitz *et al.*, respectively, are characterized by the co-occurrence of *CD79B* and *MYD88* mutations[3,4]. In animal models, the *Myd88* L265P mutation was found to promote down-regulation of surface IgM and a phenotype resembling B-cell anergy[8]. However, this effect could be rescued by *Cd79b* mutation, showing that these co-occurring mutations are epistatic[8]. The characterization of other significantly co-occurring genetic alterations are therefore likely to reveal additional important epistatic relationships.

We approached this challenge by performing genomic profiling of 755 B-NHLs across different subtypes. Through this cross-sectional analysis we characterized genomic hallmarks of B-NHL

and identified pathognomonic genetic alterations, including disease-specific mechanisms for deregulating hallmark processes and protein complexes, and sets of significantly co-associated events that represent subtype-specific epistatic genetic alterations. This study therefore provides new insight into how cooperating genetic alterations may contribute to molecularly and phenotypically distinct subtypes of B-NHL.

## METHODS

An overview of our approach is shown in Figure S1. For detailed methods, please refer to supplementary information.

### Tumor DNA samples

We collected DNA for 755 B-NHL tumors, including 199 follicular lymphoma (FL), 196 mantle cell lymphoma (MCL), 148 diffuse large B-cell lymphoma (DLBCL), 108 Burkitt's lymphoma (BL), 45 chronic lymphocytic leukemia / small lymphocytic lymphoma (CLL/SLL), 24 marginal zone lymphoma (MZL), 21 high-grade B-cell lymphoma not otherwise specified (HGBL-NOS), and 14 high-grade B-cell lymphoma with MYC, BCL2 and/or BCL6 rearrangement (DHL) (Table S1). A total of 502 samples were obtained from the University of Nebraska Medical Center, and were prioritized for inclusion in this study if they had fresh/frozen tissue available (n=577) and been previously interrogated by Affymetrix U133 Plus 2.0 gene expression microarrays (n=290)[9-11]. An additional series of 178 formalin-fixed paraffin-embedded (FFPE) tumors were collected from other centers. Samples were de-identified but accompanied by their diagnosis from medical records, plus overall survival time and status (alive, dead) when available. Medical record diagnosis was used in all cases except for those with fluorescence *in situ* hybridization showing translocations in MYC and BCL2 and/or BCL6, which were amended to DHL.

Sequencing results for a subset of 52 BL tumors was described previously[12]. All MCL samples were either positive for *CCND1* translocation by FISH or positive for CCND1 protein expression by IHC, depending on the diagnostic practices of the contributing institution.

Next generation sequencing

A total of 100-1000ng of gDNA was sonicated using a Covaris S2 Ultrasonicator, and libraries prepared using KAPA Hyper Prep Kits (Roche) with TruSeq Adapters (Bioo Scientific) and a maximum of 8 cycles of PCR (average of 6 cycles). Libraries were qualified by TapeStation 4200, quantified by Qubit and 10- to 12-plexed for hybrid capture. Each multiplexed library was enriched using our custom LymphoSeq panel encompassing the full coding sequences of 380 genes that have previously been reported to be somatically mutated in B-cell lymphoma[6,7,13-32] (Table S2, Supplementary Methods), as well as tiling recurrent translocation breakpoints. Enrichments were amplified with 4-8 cycles of PCR and sequenced on a single lane of an Illumina HiSeq 4000 with 100PE reads in high-output mode at the Hudson Alpha Institute for Biotechnology or the MD Anderson Sequencing and Microarray Facility. Variants were called using our previously validated ensemble approach[12,21], copy number alterations identified using off-target reads with the CopyWriteR tool[33], and translocation called using FACTERA[34]. Germline polymorphisms were filtered using dbSNP annotation and the EXAC dataset containing 60,706 healthy individuals[35]. Significantly mutated genes were defined by MutSig2CV[36], significant DNA copy number alterations by GISTIC2[37], and the clonal representation of mutations by ABSOLUTE[38]. Mutation and CNA data are publicly viewable through cBioPortal: https://www.cbioportal.org/study/summary?id=mbn_mdacc_2013. For further details, refer to supplementary methods.

## RESULTS

### Recurrently mutated genes highlight conserved functional hallmarks of B-NHL

We used a 380 gene custom targeted sequencing approach to interrogate the genomes of 755 mature B-NHLs, sequencing to an average depth of 599X (Min, 101X; Max, 1785X; Table S1) for a total yield of 2.06 Tbp. Somatic nucleotide variants (SNVs) and small insertions/deletions (InDels) were identified using an ensemble approach that we have previously validated[21] (Table S3) and significantly mutated genes were identified using MutSig2CV (Table S4). Genes that were significantly mutated in the full cohort or in any one of the 4 subtypes with >100 tumors (BL, DLBCL, FL, MCL), as well as frequently mutated genes that are targets of AID (Table S5, **Figure 1**), were included in downstream analyses. Mutation distributions were classified using a novel metric that classifies them into hotspot, clustered or diffuse (Figure S2). The residue of mutational hotspots and conserved domains targeted by clustered mutations are shown on figure 1. As a proof of principle, genes such as *EZH2* and *MYD88* with known mutational hotspots were classified as 'hotspot', and genes such as *CREBBP* and *TCF3* that accumulate mutations within a single domain were classified as 'clustered'. The mutational burden calculated from our targeted region significantly correlated with that from the whole exome (Figure S3A) and was significantly higher in DLBCL and other high-grade tumors compared to FL and MCL (**Figure 1**; Figure S3B). The mutational signatures were also different between malignancies, but were predominated by mutations attributable to failure of double-strand break repair by homologous recombination[39] (Signature 3; Table S6).

To identify key hallmarks that are deregulated by somatic mutations, we grouped genes using DAVID functional annotation clustering[40]. BCL6 function is not annotated in gene sets within public databases, but was an obvious feature of the recurrently mutated genes, so this hallmark

was assigned according to literature support. The most frequently perturbed processes were chromatin modification and cell signaling, with most of the genes in these categories having been well described in prior reports. Chromatin modifying genes were mutated in 68% of BL, 66% of DLBCL, 91% of FL and 45% of MCL, and included those that encode proteins that catalyze post-translational modifications of histones (*KMT2D*, *CREBBP*, *EZH2*, *EP300*, *WHSC1*, *ASHL1L* and *KMT2A*), components of the SWI/SNF chromatin remodeling complex (*ARID1A*, *SMARCA4*, *BCL7A*), linker histones (*HIST1H1E*, *HIST1H1C*, *HIST1H1B*), and the *TET2* gene. Genes with a role in signaling included those involved in B-cell receptor (*CD79B*, *ID3*, *TCF3, RFTN1*), toll-like receptor (*MYD88*), NFκB (*TNFAIP3*, *CARD11*, *NFKBIE*), Notch (*NOTCH1*, *NOTCH2*), JAK/STAT (*SOCS1*, *STAT6*), PI3K/mTOR (*FOXO1*, *ATP6V1B2*, *APT6AP1*) and G-protein (GNA13, GNAI2) signaling. The *CD79A* and *BCL10* genes were also mutated at a lower frequency that was not significant by MutSig (Figure S4A-B). Among these, the *RFTN1* gene (Figure S4C) is a novel recurrently mutated gene that was mutated in 7.4% of DLBCL and encodes a lipid raft protein that is critical for B-cell receptor signaling[41].

Other processes that are not as well described as genetic hallmarks of B-NHL included BCL6 function and the ubiquitin proteasome system (UPS). Genes associated with BCL6 function encompassed BCL6 itself and its interacting proteins (*BCL6*, *TBL1XR1*, *BCOR*, *SPEN*), regulators of BCL6 activity or expression (*MEF2B*, *IRF8*, *IRF4*), and a critical BCL6 target gene (*PRDM1*). Recurrent mutations of the *NCOR1* and *NCOR2* genes that encode BCL6 co-repressor proteins were also observed (Figure S4D-E), but were not significant by MutSig2CV. Multiple recurrently mutated chromatin modifying genes have also been implicated in BCL6 function and therefore also contribute to this hallmark, including *CREBBP* and *EZH2*[42,43]. The recurrently mutated UPS genes included the *CDC27* gene, which encodes an E3 ligase for CCND1[44] that is mutated in 14% of MCL and has not been previously described in the literature.

The *TMEM51* gene was also identified as a novel recurrently mutated gene in 9% of BL, though the function of this gene is poorly defined. The cross-sectional analysis of multiple lymphoma subtypes therefore identified conserved hallmarks that are recurrently targeted by somatic mutations in B-NHL.

Enrichment of functional hallmarks by structural alterations

The hybrid capture probes utilized in our design also targeted recurrent breakpoint regions in the immunoglobulin heavy- and light-chain loci, as well as the regions of recurrent breakpoints in or near the *BCL2*, *MYC* and *BCL6* genes. Translocations were called using a method that detects discordantly mapped reads[34] and our prior validation of this approach in cases with matched fluorescence in situ hybridization (FISH) data for *MYC* showed that it is 100% specific, but only ~40% sensitive for translocation detection[12]. In addition, we failed to detect *CCND1* translocations using this approach. However, we did observe a significantly higher fraction of *BCL6* translocations (57% [27/47]) partnered to non-immunoglobulin loci (eg. *CIITA*, *RHOH*, *EIF4A2*, *ST6GAL1*; Table S7) compared to *BCL2* (1% [1/114]) and *MYC* (5% [2/38]) translocations (**Figure 2A**; Fisher P-value <0.001). These were more frequent in FL (88% [15/17] of *BCL6* translocations) as compared to DLBCL (39% [9/23] of *BCL6* translocations), presumably because the two immunoglobulin loci in FL are either translocated with the *BCL2* gene or functioning in immunoglobulin expression[45]. We also employed off-target reads to detect DNA copy number alterations (CNAs) in a manner akin to low-pass whole genome sequencing, identified significant peaks of copy gain and losses using GISTIC2[37] (**Figure 2A**; Figure S5; Table S8-9), and defined the likely targets of these CNAs by integrative analysis of matched gene expression profiling (GEP) data from 290 tumors (**Figure 2B-C**, Figure S5, Table S10-11). This identified known CNA targets (**Figure 2D**), including but not limited to deletion of *TNFAIP3* (6q24.2)[46], *ATM* (11q22.3)[47], *B2M* (15q15.5)[48] and *PTEN* (10q23.21)[49], and copy gain

of *REL* and *BCL11A* (2p15), and *TCF4* (18q23)[50]. In addition, we identified novel targets such as deletion of *IBTK* (6q14.1), *UBE3A* (11q22.1) and *FBXO25* (8p23.3), and copy gain of *ATF7* (12q13.13), *UCHL5* (1q31.3), and *KMT2A* (11q23.3). Several CNA peaks, defined as the smallest and most statistically significant region, included genes that were significantly mutated (**Figure 2E**) as well as other genes for which we detected mutations at lower frequencies that were not significant by MutSig (*POU2AF1*, *TP53BP1*, *FAS*, *PTEN*). Furthermore, for several of these genes (*ATM*, *B2M, BIRC3* and *TNFRSF14*), deletions were significantly co-associated with mutations, suggesting that deletion and mutation are complementary mechanisms contributing to biallelic inactivation.

Analyzing targets of translocations and CNAs added to enrichment of functional categories that we observed within recurrently mutated genes, including chromatin modification (*BCL11A*) and signaling (*TCF4*, *PTEN*). This was most prominently observed for genes with a role in apoptosis and the cell cycle (*CDKN2A*, *CDKN2B*, *FAS*, *RPL5*, *DFFB*, *MDM2*, *CDK6*) and the UPS (*CUL4A*, *FBXO25*, *IBTK*, *RNF38*, *UBAP1, UBE3A*, *UBQLN1*, *UCHL5*). Recurrently altered UPS genes collectively promote or suppress the abundance or activity of proteins with roles in other hallmark characteristics (**Figure 3A**). For example; *SOCS1*[51], *BIRC3*[52], *DTX1*[53,54], *IBTK*[55], *TRIM13*[56], *TNFAIP3*[53] and *UBE3A*[53] all inhibit the function of proteins with a role in important B-cell signaling pathways (**Figure 3B**), thereby establishing a functional link between genetic alteration of the UPS and deregulation of signaling and adding an additional layer to the mechanisms by which B-cell receptor signaling is perturbed in B-NHL. The combined analysis of mutations and structural alterations therefore identified large sets of genes that are targeted by genetic alterations and collectively contribute to hallmark features that are commonly deregulated across multiple B-NHL subtypes.

235

236 Disease-specific patterns of genetic alterations

237 Some interesting patterns of disease-specificity were obvious for both mutations and structural
238 alterations. We therefore formally tested the over- or under-representation of these events in
239 each of the 4 subtypes with >100 tumors (BL, DLBCL, FL, MCL), compared to all other tumors
240 in the study (**Figure 4**; Table S12). Using matched GEP data for a subset of cases (Table S1),
241 we also tested the association between genetic alterations and molecularly-defined Burkitt's
242 lymphoma subtypes (n=154; Figure S6-7, Table S13), as well as DLBCL cell of origin subtypes
243 (n=98; Figure S7, Table S14). These analyses showed that the genetic alterations associated
244 with clinically-defined BL were also associated with molecularly-defined BL. Moreover, genetic
245 alterations that we and others have characterized as being over-represented in ABC-like DLBCL
246 such as *CD79B* mutation[57], *MYD88* mutation[58] and *TCF4* copy gain[50] were significantly over-
247 represented in this analysis also.

248

249 We observed some interesting patterns within hallmark characteristics that differ between
250 subtypes. For example, linker histone mutations were present in germinal center B (GCB)-cell
251 derived malignancies (BL, DLBCL, FL) at variable frequencies, but were largely absent from
252 MCL (Figure S8). In contrast, mutations of genes encoding the H3K36 methyltransferases,
253 *WHSC1* (aka *MMSET*/*NSD2*) and *ASH1L*, were frequent in MCL and largely absent from GCB-
254 derived malignancies. We also noted that the SWI/SNF complex was perturbed in different ways
255 in different diseases (**Figure 5**). Specifically, mutations of the *SMARCA4* (aka. BRG1)
256 component of the ATPase module were significantly enriched in BL (24%) compared to other
257 subtypes (4%, Q-value<0.001), while mutations of the *BCL7A* component of the ATPase
258 module were significantly enriched in FL (11%) compared to other subtypes (4%, Q-

value=0.007). In contrast, mutations of *ARID1A* were frequent in both BL (19%) and FL (15%), and DNA copy number gains of *BCL11A* were frequent in both DLBCL (28%) and FL (22%). The SWI/SNF complex is therefore a target of recurrent genetic alterations, as previously suggested[59], but the manner in which this complex is perturbed varies between B-NHL subtypes (**Figure 5**). Similar disease-specific patterns were also observed for signaling genes; for example, *TCF3* and *ID3* have important functions in normal germinal center B-cells[60], but mutations of these genes are specifically enriched within BL and are rarely found in the other GCB-derived malignancies, DLBCL and FL. Similarly, the *ATP6AP1* and *ATP6V1B2* genes that function in mTOR signaling[61,62] are specifically mutated in FL, and the *DUSP2* gene which inactivates ERK1/2[63] and STAT3[64] is specifically mutated in DLBCL. The disease-specific patterns of genetic alterations therefore reveal subtle but important differences in how each subtype of B-NHL perturbs hallmark features.

Pathognomonic sets of co-associated genomic alterations in B-NHL subtypes

We next defined how each genetic alteration co-associated with or mutually-excluded other genetic alterations by pairwise assessments using false-discovery rate (FDR)-corrected Fisher's tests (Table S15). A matrix of the transformed FDR Q-values (-logQ) was used for unsupervised hierarchical clustering to identify clusters of co-associated genetic alterations. Together with patterns of disease-specificity, unsupervised clustering revealed clear groupings of co-associated events for BL, DLBCL, FL and MCL (**Figure 4**). We identified a single cluster that was specifically enriched in DLBCL, included co-associated genetic alterations that were over-represented in the ABC-like, and overlapped with the previously described C5/MCD clusters[3,4]. We also identified a cluster consisting of TP53 mutations and multiple CNAs similar to the C2 subtype reported in DLBCL, but which was enriched in both FL and DLBCL. A cluster with features similar to the C3/EZB DLBCL clusters was also observed, but was specifically enriched

in FL. The remaining two previously reported DLBCL clusters (C1/N2 and C4) were not observed in this analysis, which may be a result of the different clustering strategies used and/or the lower number of DLBCL tumors in this study compared to other prior DLBCL-focused efforts[3,4]. We therefore focused our subsequent description on BL, FL and MCL.

The core set of co-associated genetic alterations in BL were *MYC* translocation and mutation, and mutations of *SMARCA4*, *CCND3* and *ID3*. Mutations of *GNAI2*, *GNA13*, *FOXO1*, *TCF3*, *ARID1A* and *TMEM51* were also clustered with this core set, but were present at lower frequencies. Mutations of *TP53* and copy gain of 1q were frequent in BL, but did not cluster together with other BL-associated genetic alterations. The core set of genetic alterations in FL were *BCL2* translocations and mutations, and mutations of *KMT2D*, *CREBBP* and *EZH2*, in line with prior observations[16,21]. Copy loss of 10q23.31 encompassing the *FAS* and *PTEN* genes, and mutations *EEF1A1*, *TNFRSF14*, *IRF8*, *BCL7A*, *ATP6V1B2* and *ATP6AP1* were present at lower frequencies and clustered with this core set. The MCL cluster included mutations and deletions of *ATM* and *BIRC3*, as well as frequent deletions of 9p21.32 (*CDKN2A* and *CDKN2B*), 9q21.3 (*UBQLN1*), 13q14.2 (miR-15a/16 and TRIM13), 13q34 (*CUL4A* and *ING1*), and 1p21 (*RPL5*). Mutation of *BCOR*, *UBR5*, *SP140*, *CCND1* and *WHSC1* were also significantly over-represented in MCL compared to other subtypes, and were associated with this cluster. These data show that different subtypes of B-NHL are defined by characteristic sets of co-associated genetic alterations that affect multiple hallmarks.

## Combinations of clonal genetic alterations define B-NHL subtypes

Our data have revealed statistical enrichment of individual genetic alterations in subtypes of B-NHL, and pairwise relationships between different genetic alterations that define clusters of

subtype-specific events. However, individual genetic alterations from disease-specific clusters are also observed at variable frequencies in other B-NHL subtypes (**Figure 6A**). We therefore investigated whether the combination of multiple genetic alterations from each cluster, rather than the presence or absence of a single alteration, were pathognomonic. By enumerating the number of genetic alterations from each disease-specific cluster in each tumor, we found that this was indeed the case. Specifically over half of all BL, FL and MCL tumors possessed ≥3 genetic alterations defined by their respective subtype-specific clusters (**Figure 6**). These rates were significantly higher than those observed in other diseases (Fisher P <0.001 for all comparisons). For example, DLBCL and FL share many of the same recurrently mutated genes, particularly when considering the C3/EZB subtype of DLBCL[3,4]. Thus, one or more FL cluster alterations are found in 93% of FL tumors, but also in 54% of DLBCL tumors. However, FL tumors have a significantly higher rate of accumulation of these mutations, with 68% of FL tumors bearing ≥3 of the FL cluster alterations but only 18% of DLBCL tumors bearing ≥3 of these genetic alterations (Fisher P < 0.001). This pattern is also conserved for the BL and MCL clusters, for which the rate of acquiring ≥3 of the cluster alterations was significantly higher in BL or MCL compared to other subtypes of B-NHL, respectively. Thus, B-NHL subtypes are defined by the acquisition of multiple genetic alterations from subtype-associated clusters.

The co-association of multiple mutations may be suggestive of a clonal structure in which early driver mutations are clonal and later driver mutations are present in sub-clones. We therefore utilized ABSOLUTE (Table S16-18) to determine whether co-associated genetic alterations are present at different clonal fractions, and could therefore be phased to create clonal hierarchies. This analysis failed for a small subset of cases and could not be performed for X-linked genes or translocations. However, it clearly revealed the majority of co-associated genetic alterations were present at cancer cell fractions (CCF) >0.9, which is indicative that each mutation is

clonally represented in every cancer cell at the time of sampling. These data suggest that combinations of mutations may therefore be required for effective expansion of lymphoma cells, such that common precursor cells (CPCs) are not detectable within the clonal structure of the clinically-detected tumor. These data therefore support the premise of epistatic interactions between co-occurring genetic alterations in B-NHL.

## DISCUSSION

By performing cross-sectional genomic profiling of a large cohort of tumors, we have defined genes and functional hallmarks that are recurrently targeted by genetic alterations and showed that combinations of genetic alterations are disease-defining features of B-NHL subtypes. Some of the functional hallmarks that we identified have been previously appreciated in other diseases, with a few exceptions. For example, the mutation of genes with roles in chromatin modification are known to be a hallmark of FL[65] and we observed that over 90% of FLs possessed mutations in one or more of the genes in this category. However, these mutations were also observed in two thirds of BL and DLBCL tumors and nearly half of MCLs. There are subtype-specific patterns of chromatin modifying gene alterations, such as those that we described for H3K36 methylation in MCL and the unique patterns of SWI/SNF mutations across GCB-derived B-NHLs. But we suggest that the genetic deregulation of chromatin modification should be considered a general hallmark of B-NHL. In addition, we suggest that the deregulation of BCL6 function and perturbation of the ubiquitin proteasome system are hallmarks of B-NHL that require further investigation. We have highlighted some of the known substrates of recurrently altered UPS genes from the literature, which shows the potential for these genetic alterations to contribute to aberrant B-cell receptor signaling and proliferation. However, many of the genes within these categories have not been functionally studied in the

context of B-cell lymphoma, which represents a significant gap in our understanding of disease etiology.

The role of epistatic interactions between co-occurring genetic alterations is also an emerging field that requires further investigation. These interactions are not uncommon in cancer[66], but our data show that they are pervasive and disease-defining features of the B-NHL genetic landscape. Epistasis between co-associated genetic alterations identified in this study requires formal validation in cell line and/or animal models. However, there are many instances in which co-occurring genetic alterations that we observed have already been shown to cooperate in lymphomagenesis. In addition to the aforementioned example of *MYD88* and *CD79B* mutations, transgenic mouse models of *Ezh2* activating mutations or conditional deletion of *Crebbp* or *Kmt2d* have shown that these events are not alone sufficient for lymphomagenesis[42,67-71]. We and others have observed a co-association between the mutation of these genes and *BCL2* translocations[16,21], and the addition of a *Bcl2* transgene to these murine models indeed promoted lymphoma at a significantly higher rate than that observed with the *Bcl2* transgene alone[42,67-71]. These genetic alterations are therefore significantly more lymphomagenic in combination than they are alone, which provides proof of principal that an epistatic relationship exists between these co-occurring genetic alterations. Future studies focusing on other co-occurring mutations, such as *MYC* translocation and *SMARCA4* mutation in BL, *CREBBP* and *KMT2D* mutation in FL, *TCF4* copy gain and *MYD88* mutation in DLBCL, and *ATM* mutation and *RPL5* deletion in MCL, should therefore be performed to further explore these concepts and define their underlying functional relationship. We suggest that combinations of genetic alterations are likely to more accurately recapitulate the biology of B-NHL than single gene models, and may reveal contextually different functional roles of genetic alterations depending on the co-occurring events.

In conclusion, we have provided a framework of functional hallmarks and co-occurring genetic alterations that are enriched within B-NHL subtypes. These genetic alterations likely represent epistatic interactions that underpin the biology of these tumors, and represent an opportunity for better understanding lymphoma etiology so that we can identify novel rational approaches for therapeutic targeting of the underlying biology.

## ACKNOWLEDGEMENTS

This research was supported by NCI R01CA201380 (MRG), the Nebraska Department of Health and Human Services (LB506 2016-17; MRG), the Jaime Erin Follicular Lymphoma Research Consortium (MRG and SSN), the Futcher Foundation (LN and MRG), the Schweitzer Family Fund (JRW, RED, MRG), a Leukemia and Lymphoma Society Fellow Award (HY), and NCI cancer center support grants to the University of Texas MD Anderson Cancer Center (P30 CA016672) and the Fred & Pamela Buffet Cancer Center (P30 CA036727).

## AUTHOR CONTRIBUTIONS

MCJM and ST performed experiments, analyzed data and wrote the manuscript. AB analyzed data. AB, TBH, HY, QD, DM, KH, NJ, JS, and SG performed experiments. AA, LS, MD, CC, JT, DP, KMV, MAL, ARS, BJC, RB, SSN, LN, RED, JW, SP, MKG, DS, JI, TG, SR, and AM provided samples and/or clinical data. MRG conceived and supervised the study, performed experiments, analyzed the data and wrote the manuscript. All authors reviewed and approved the manuscript.

## DISCLOSURES OF CONFLICTS OF INTEREST

The authors have no conflicts of interest related to this work.

## REFERENCES

1 Swerdlow, S. H. *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127**, 2375-2390, doi:10.1182/blood-2016-01-643569 (2016).

2 Armitage, J. O., Gascoyne, R. D., Lunning, M. A. & Cavalli, F. Non-Hodgkin lymphoma. *Lancet* **390**, 298-310, doi:10.1016/S0140-6736(16)32407-2 (2017).

3 Chapuy, B. *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nature medicine* **24**, 679-690, doi:10.1038/s41591-018-0016-8 (2018).

4 Schmitz, R. *et al.* Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *The New England journal of medicine* **378**, 1396-1407, doi:10.1056/NEJMoa1801445 (2018).

5 Reddy, A. *et al.* Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* **171**, 481-494 e415, doi:10.1016/j.cell.2017.09.027 (2017).

6 Bea, S. *et al.* 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, doi:10.1073/pnas.1314608110 (2013).

7 Zhang, J. *et al.* The genomic landscape of mantle cell lymphoma is related to the epigenetically determined chromatin state of normal B cells. *Blood* **123**, 2988-2996, doi:10.1182/blood-2013-07-517177 (2014).

8 Wang, J. Q. *et al.* Synergistic cooperation and crosstalk between MYD88L265P and mutations that dysregulate CD79B and surface IgM. *The Journal of experimental medicine* **214**, 2759-2776, doi:10.1084/jem.20161454 (2017).

9 Lenz, G. *et al.* Stromal gene signatures in large-B-cell lymphomas. *The New England journal of medicine* **359**, 2313-2323, doi:10.1056/NEJMoa0802885 (2008).

10 Dave, S. S. *et al.* Molecular diagnosis of Burkitt's lymphoma. *The New England journal of medicine* **354**, 2431-2442, doi:10.1056/NEJMoa055759 (2006).

11 Iqbal, J. *et al.* Genome-wide miRNA profiling of mantle cell lymphoma reveals a distinct subgroup with poor prognosis. *Blood* **119**, 4939-4948, doi:10.1182/blood-2011-07-370122 (2012).

12 Bouska, A. *et al.* Adult High Grade B-cell Lymphoma with Burkitt Lymphoma Signature: Genomic features and Potential Therapeutic Targets. *Blood*, doi:10.1182/blood-2017-02-767335 (2017).

13 Khodadoust, M. S. *et al.* Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. *Nature* **543**, 723-727, doi:10.1038/nature21433 (2017).

14 Schmitz, R. *et al.* Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* **490**, 116-120, doi:10.1038/nature11378 (2012).

15 Love, C. *et al.* The genetic landscape of mutations in Burkitt lymphoma. *Nature genetics* **44**, 1321-1325, doi:10.1038/ng.2468 (2012).

16 Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298-303, doi:10.1038/nature10351 (2011).
17 Morin, R. D. *et al.* Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood* **122**, 1256-1265, doi:10.1182/blood-2013-02-483727 (2013).
18 Pasqualucci, L. *et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nature genetics* **43**, 830-837, doi:10.1038/ng.892 (2011).
19 Lohr, J. G. *et al.* Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 3879-3884, doi:10.1073/pnas.1121343109 (2012).
20 Green, M. R. *et al.* Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. *Blood* **121**, 1604-1611, doi:10.1182/blood-2012-09-457283 (2013).
21 Green, M. R. *et al.* Mutations in early follicular lymphoma progenitors are associated with suppressed antigen presentation. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E1116-1125, doi:10.1073/pnas.1501199112 (2015).
22 Okosun, J. *et al.* Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nature genetics* **46**, 176-181, doi:10.1038/ng.2856 (2014).
23 Pasqualucci, L. *et al.* Genetics of follicular lymphoma transformation. *Cell reports* **6**, 130-140, doi:10.1016/j.celrep.2013.12.027 (2014).
24 Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714-726, doi:10.1016/j.cell.2013.01.019 (2013).
25 Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101-105, doi:10.1038/nature10113 (2011).
26 Wang, L. *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *The New England journal of medicine* **365**, 2497-2506, doi:10.1056/NEJMoa1109016 (2011).
27 Kiel, M. J. *et al.* Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. *The Journal of experimental medicine* **209**, 1553-1565, doi:10.1084/jem.20120910 (2012).
28 Rossi, D. *et al.* The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. *The Journal of experimental medicine* **209**, 1537-1551, doi:10.1084/jem.20120904 (2012).
29 Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467-472, doi:10.1038/nature09837 (2011).
30 Treon, S. P. *et al.* MYD88 L265P somatic mutation in Waldenstrom's macroglobulinemia. *The New England journal of medicine* **367**, 826-833, doi:10.1056/NEJMoa1200710 (2012).
31 Gunawardana, J. *et al.* Recurrent somatic mutations of PTPN1 in primary mediastinal B cell lymphoma and Hodgkin lymphoma. *Nature genetics* **46**, 329-335, doi:10.1038/ng.2900 (2014).
32 Vater, I. *et al.* The mutational pattern of primary lymphoma of the central nervous system determined by whole-exome sequencing. *Leukemia* **29**, 677-685, doi:10.1038/leu.2014.264 (2015).
33 Kuilman, T. *et al.* CopywriteR: DNA copy number detection from off-target sequence data. *Genome biology* **16**, 49, doi:10.1186/s13059-015-0617-1 (2015).
34 Newman, A. M. *et al.* FACTERA: a practical method for the discovery of genomic rearrangements at breakpoint resolution. *Bioinformatics* **30**, 3390-3393, doi:10.1093/bioinformatics/btu549 (2014).

35 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).

36 Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-218, doi:10.1038/nature12213 (2013).

37 Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology* **12**, R41, doi:10.1186/gb-2011-12-4-r41 (2011).

38 Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nature biotechnology* **30**, 413-421, doi:10.1038/nbt.2203 (2012).

39 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421, doi:10.1038/nature12477 (2013).

40 Huang, D. W. *et al.* DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic acids research* **35**, W169-175, doi:10.1093/nar/gkm415 (2007).

41 Saeki, K., Miura, Y., Aki, D., Kurosaki, T. & Yoshimura, A. The B cell-specific major raft protein, Raftlin, is necessary for the integrity of lipid raft and BCR signal transduction. *EMBO J* **22**, 3015-3026, doi:10.1093/emboj/cdg293 (2003).

42 Jiang, Y. *et al.* CREBBP Inactivation Promotes the Development of HDAC3-Dependent Lymphomas. *Cancer discovery* **7**, 38-53, doi:10.1158/2159-8290.CD-16-0975 (2017).

43 Beguelin, W. *et al.* EZH2 and BCL6 Cooperate to Assemble CBX8-BCOR Complex to Repress Bivalent Promoters, Mediate Germinal Center Formation and Lymphomagenesis. *Cancer cell* **30**, 197-213, doi:10.1016/j.ccell.2016.07.006 (2016).

44 Pawar, S. A. *et al.* C/EBP{delta} targets cyclin D1 for proteasome-mediated degradation via induction of CDC27/APC3 expression. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9210-9215, doi:10.1073/pnas.0913813107 (2010).

45 Akasaka, T., Lossos, I. S. & Levy, R. BCL6 gene translocation in follicular lymphoma: a harbinger of eventual transformation to diffuse aggressive lymphoma. *Blood* **102**, 1443-1448, doi:10.1182/blood-2002-08-2482 (2003).

46 Kato, M. *et al.* Frequent inactivation of A20 in B-cell lymphomas. *Nature* **459**, 712-716, doi:10.1038/nature07969 (2009).

47 Greiner, T. C. *et al.* Mutation and genomic deletion status of ataxia telangiectasia mutated (ATM) and p53 confer specific gene expression profiles in mantle cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 2352-2357, doi:10.1073/pnas.0510441103 (2006).

48 Challa-Malladi, M. *et al.* 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, doi:10.1016/j.ccr.2011.11.006 (2011).

49 Pfeifer, M. *et al.* PTEN loss defines a PI3K/AKT pathway-dependent germinal center subtype of diffuse large B-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 12420-12425, doi:10.1073/pnas.1305656110 (2013).

50 Jain N, H. K., Tadros S, Fiskus W, Havranek O, Ma M, Bouska A, Heavican T, Kumar D, Deng Q, Moore D, Pak C, Liu C, Gentles A, Hartmann E, Kridel R, Ekstrom-Smedby K, Juliusson G, Rosenquist R, Gascoyne R, Rosenwald A, Giancotti F, Neelapu S, Westin J, Vose J, Lunning M, Greiner T, Rodig S, Iqbal J, Alizadeh A, Davis RE, Bhalla K, Green M. Targetable genetic alterations of TCF4 (E2-2) drive immunoglobulin expression in the activated B-cell subtype of diffuse large B-cell lymphoma. *Sci. Transl. Med.* **In Press** (2019).

51 Ingley, E. *et al.* Csk-binding protein mediates sequential enzymatic down-regulation and degradation of Lyn in erythropoietin-stimulated cells. *The Journal of biological chemistry* **281**, 31920-31929, doi:10.1074/jbc.M602637200 (2006).

52 Hu, S., Alcivar, A., Qu, L., Tang, J. & Yang, X. CIAP2 inhibits anigen receptor signaling by targeting Bcl10 for degredation. *Cell cycle* **5**, 1438-1442, doi:10.4161/cc.5.13.2866 (2006).

53 Compagno, M. *et al.* Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **459**, 717-721, doi:10.1038/nature07968 (2009).

54 Liu, W. H. & Lai, M. Z. Deltex regulates T-cell activation by targeted degradation of active MEKK1. *Molecular and cellular biology* **25**, 1367-1378, doi:10.1128/MCB.25.4.1367-1378.2005 (2005).

55 Liu, W. *et al.* Direct inhibition of Bruton's tyrosine kinase by IBtk, a Btk-binding protein. *Nature immunology* **2**, 939-946, doi:10.1038/ni1001-939 (2001).

56 Joo, H. M. *et al.* Ret finger protein 2 enhances ionizing radiation-induced apoptosis via degradation of AKT and MDM2. *Eur J Cell Biol* **90**, 420-431, doi:10.1016/j.ejcb.2010.12.001 (2011).

57 Davis, R. E. *et al.* Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463**, 88-92, doi:10.1038/nature08638 (2010).

58 Ngo, V. N. *et al.* Oncogenically active MYD88 mutations in human lymphoma. *Nature* **470**, 115-119, doi:10.1038/nature09671 (2011).

59 Krysiak, K. *et al.* Recurrent somatic mutations affecting B-cell receptor signaling pathway genes in follicular lymphoma. *Blood* **129**, 473-483, doi:10.1182/blood-2016-07-729954 (2017).

60 Gloury, R. *et al.* Dynamic changes in Id3 and E-protein activity orchestrate germinal center and plasma cell development. *The Journal of experimental medicine* **213**, 1095-1111, doi:10.1084/jem.20152003 (2016).

61 Okosun, J. *et al.* Recurrent mTORC1-activating RRAGC mutations in follicular lymphoma. *Nature genetics* **48**, 183-188, doi:10.1038/ng.3473 (2016).

62 Wang, F. *et al.* Follicular lymphoma-associated mutations in vacuolar ATPase ATP6V1B2 activate autophagic flux and mTOR. *The Journal of clinical investigation* **130**, 1626-1640, doi:10.1172/JCI98288 (2019).

63 Hu, J. *et al.* MiR-361-3p regulates ERK1/2-induced EMT via DUSP2 mRNA degradation in pancreatic ductal adenocarcinoma. *Cell Death Dis* **9**, 807, doi:10.1038/s41419-018-0839-8 (2018).

64 Lu, D. *et al.* The phosphatase DUSP2 controls the activity of the transcription activator STAT3 and regulates TH17 differentiation. *Nature immunology* **16**, 1263-1273, doi:10.1038/ni.3278 (2015).

65 Green, M. R. Chromatin modifying gene mutations in follicular lymphoma. *Blood* **131**, 595-604, doi:10.1182/blood-2017-08-737361 (2018).

66 Ashworth, A., Lord, C. J. & Reis-Filho, J. S. Genetic interactions in cancer progression and treatment. *Cell* **145**, 30-38, doi:10.1016/j.cell.2011.03.020 (2011).

67 Beguelin, W. *et al.* EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer cell* **23**, 677-692, doi:10.1016/j.ccr.2013.04.011 (2013).

68 Garcia-Ramirez, I. *et al.* Crebbp loss cooperates with Bcl2 over-expression to promote lymphoma in mice. *Blood* **129**, 2645-2656, doi:10.1182/blood-2016-08-733469 (2017).

69 Zhang, J. *et al.* The Crebbp Acetyltransferase is a Haploinsufficient Tumor Suppressor in B Cell Lymphoma. *Cancer discovery*, doi:10.1158/2159-8290.CD-16-1417 (2017).

70 Zhang, J. *et al.* Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nature medicine* **21**, 1190-1198, doi:10.1038/nm.3940 (2015).

71 Ortega-Molina, A. *et al.* The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nature medicine* **21**, 1199-1208, doi:10.1038/nm.3943 (2015).

## FIGURE LEGENDS

**Figure 1: Recurrently mutated genes in B-NHL subtypes.** An oncoplot shows significantly mutated genes across our cohort of 755 B-NHL tumors, arranged according to functional category and frequency. The mutational burden and distribution of mutation types for each case are shown at the top. Amino acid residues for mutational hotspots and/or domains targeted by clustered mutations are shown on the right.

**Figure 2: Structural alterations in B-NHL subtypes. A)** A circos plot shows translocations of the MYC (purple), BCL2 (orange) and BCL6 (green) genes, and GISTIC tracks of DNA copy number gains (red) and losses (blue). **B-C)** Volcano plots of integrative analysis results showing the changes in gene expression of genes within peaks of DNA copy number gain (B) or loss (C). The shaded region in B marks genes with significantly increased expression in tumors with increased copy number compared to those without; the converse is true for the shaded region in C. **D)** An oncoplot with DNA copy number losses and gains ranked according to their frequency shows the distribution of structural alterations across tumors. E) Oncoplots show the overlap of structural alterations and mutations that target the same genes. P-values are derived from a Fisher's exact test (ns, not significant).

**Figure 3: Genetic alterations of the ubiquitin proteasome system (UPS) and potential interplay with BCR signaling. A)** A graphical illustration of UPS gene genetic alterations (mutations, teal; deletions, blue; copy gain, red) and literature supported targets of the UPS genes. The function of UPS gene targets is categorized according to the hallmarks shown in figure 1. **B)** A schematic of the different layers of genetic alterations (mutations, teal; deletions, blue; copy gain, red) affecting signaling pathways in B-NHL. UPS genes are highlighted with a yellow halo.

**Figure 4: Subtype-specific clusters of co-occurring genetic alterations.** The frequency (bar graph) and over/under-representation (blue to yellow scale) of mutations and structural alterations is shown on the left for BL, DLBCL, FL and MCL. Over/under-representation in molecular Burkitt and cell of origin molecular subtypes is shown for each alteration at the bottom. The correlation matrix of co-associated (green) and mutually-exclusive (purple) relationships was clustered and identified groups of co-occurring genetic alterations that were predominantly over-represented in a single B-NHL subtype.

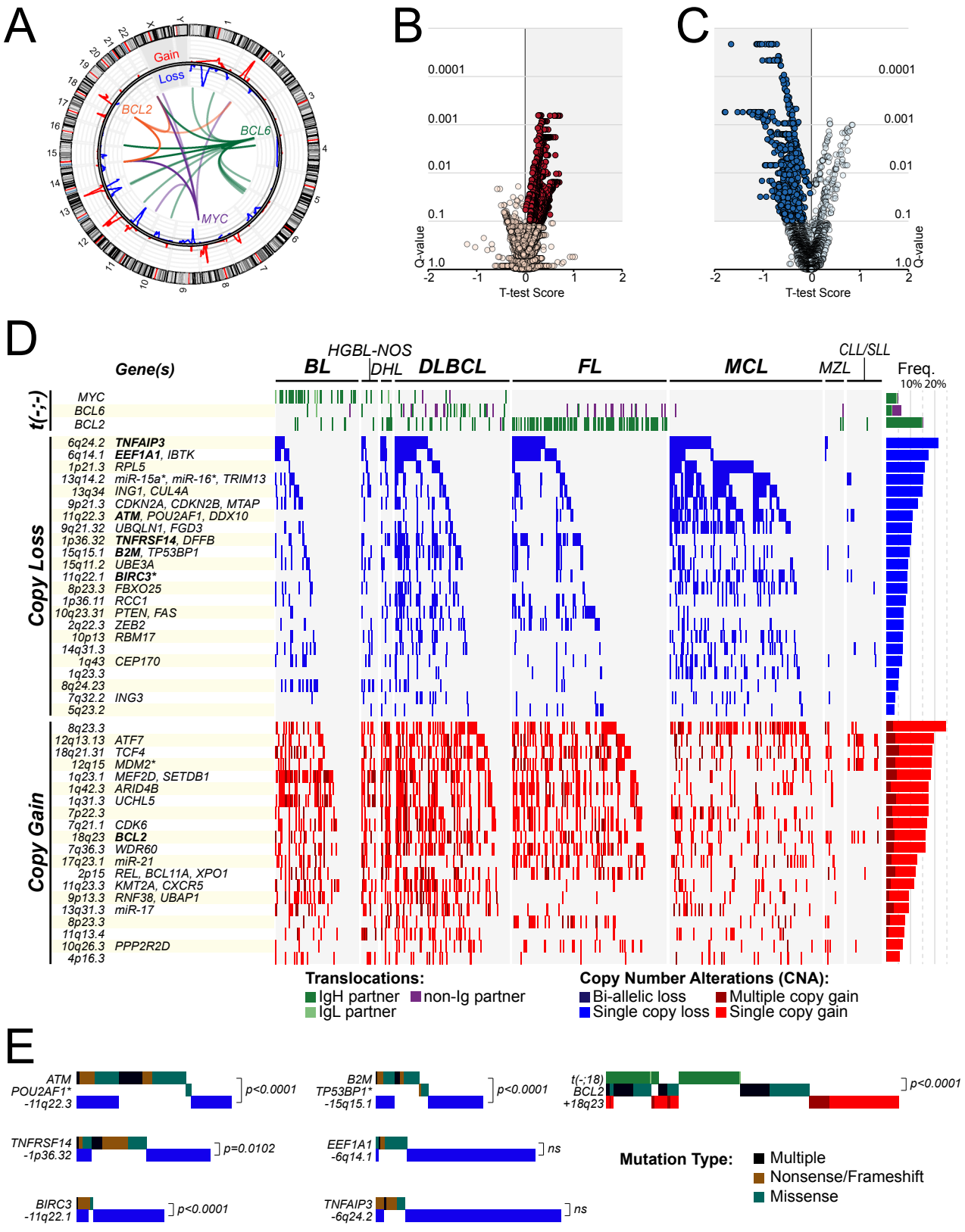
**Figure 5: Subtype-specific patterns of SWI/SNF complex mutations. A)** An oncoplot shows the frequency of genetic alterations in genes that encode components of the SWI/SNF complex. **B)** A schematic of the SWI/SNF complex shows recurrently mutated genes, *ARID1A*, *SMARCA4* and *BCL7A,* and the *BCL11A* gene that is targeted by 2p15 DNA copy number gains. **C-E)** Lollipop plots show the distribution of mutations in the SWI/SNF components ARID1A (C), SMARCA4 (D), and BCL7A (E). **F)** A heatplot shows the location of chromosome 2p DNA copy number gains (red) in all tumors (rows) with 2p15 copy gain (n=92, copy number > 2.2) ordered from highest DNA copy number (top) to lowest (bottom). The *BCL11A* gene is in the peak focal copy gain.

**Figure 6: Cumulative burden of clonal co-occurring mutations. A)** A bar plot shows the frequency of tumors with genetic alterations from each subtype-defining cluster, grouped by the number of co-occurring alterations observed per tumor. The frequency of tumors with 3 or more genetic alterations per tumor is significantly higher in the subtype for which each cluster is named (***Fisher P-value < 0.001). **B-D)** An oncoplot, colored by cancer cell fraction (CCF) from ABSOLUTE, shows the representation and clonality of BL cluster (B), FL cluster (C) and MCL cluster (D) mutations across the dataset. The rate of co-occurrence of multiple genetic alterations in a single tumor (columns) is observably higher in the subtypes for which the

667 clusters are named, and the co-occurring genetics are most often clonal (CCF > 0.9). BL, Burkitt
668 Lymphoma. HG, other high-grade lymphoma (DHL/THL, HGBL-NOS). DLBCL, diffuse large B-
669 cell lymphoma. FL, follicular lymphoma. MCL, mantle cell lymphoma. LG, other low-grade
670 lymphoma (CLL/SLL, MZL).

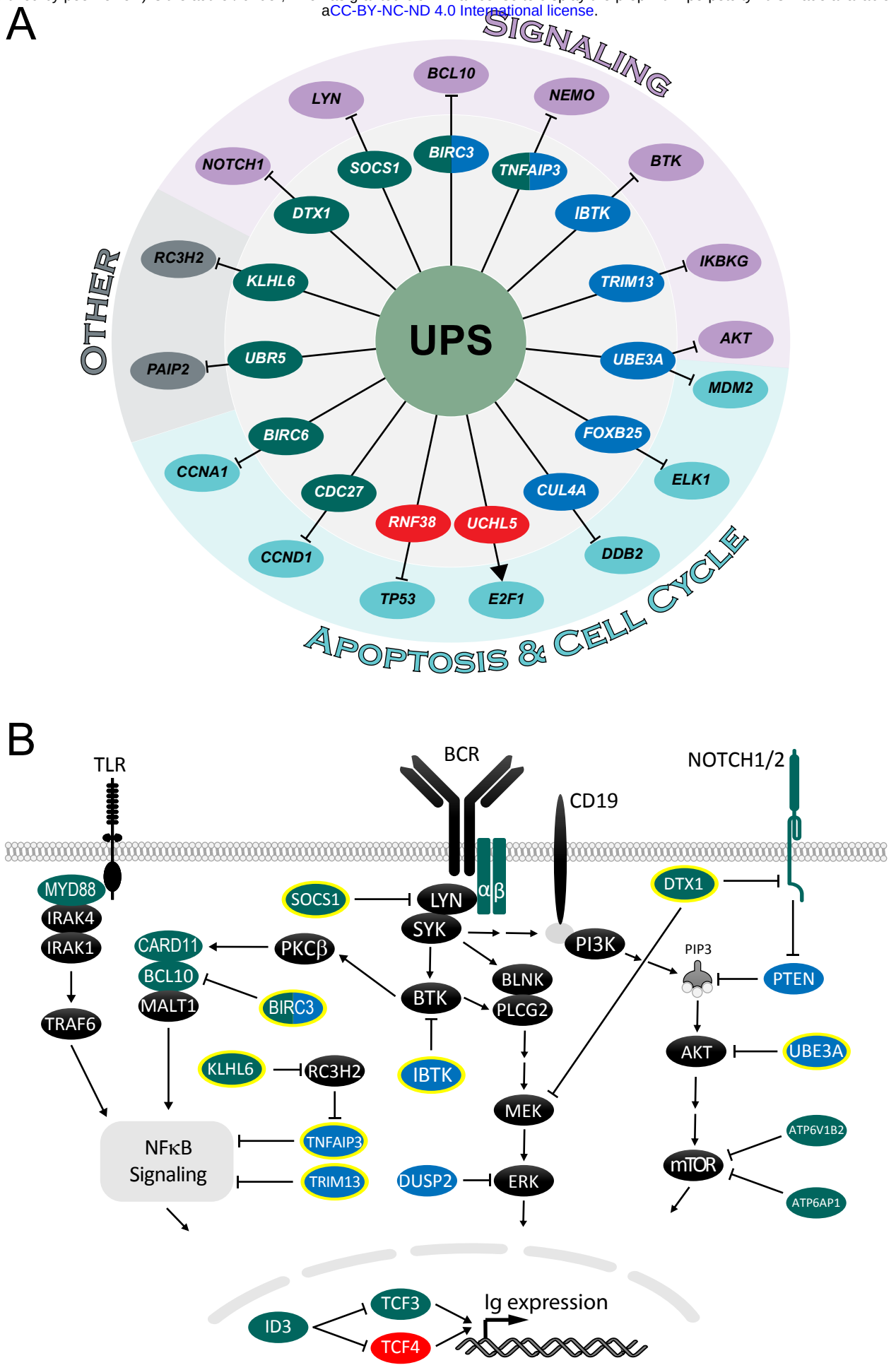
# Figure 1

# Figure 2

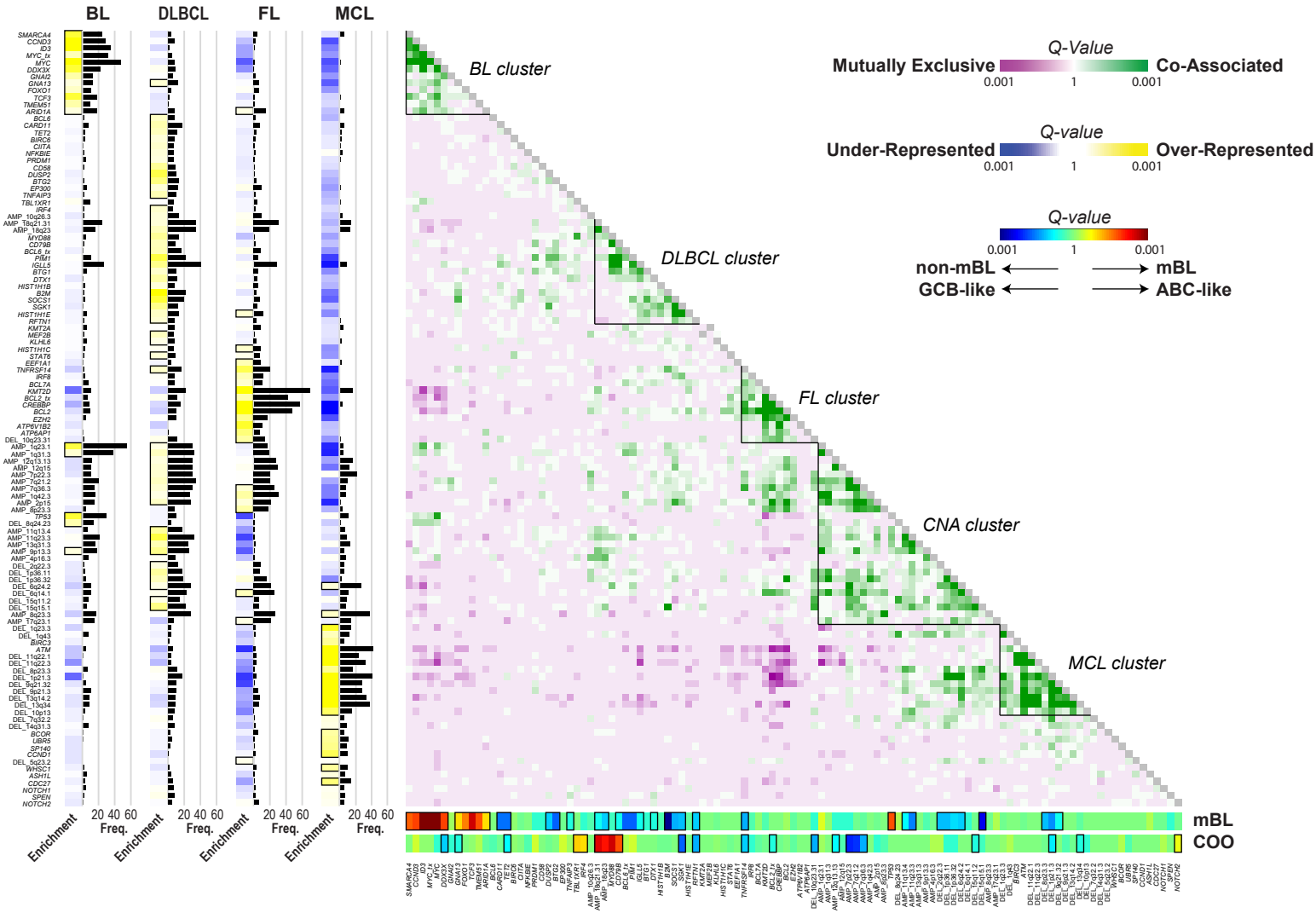


# Figure 3

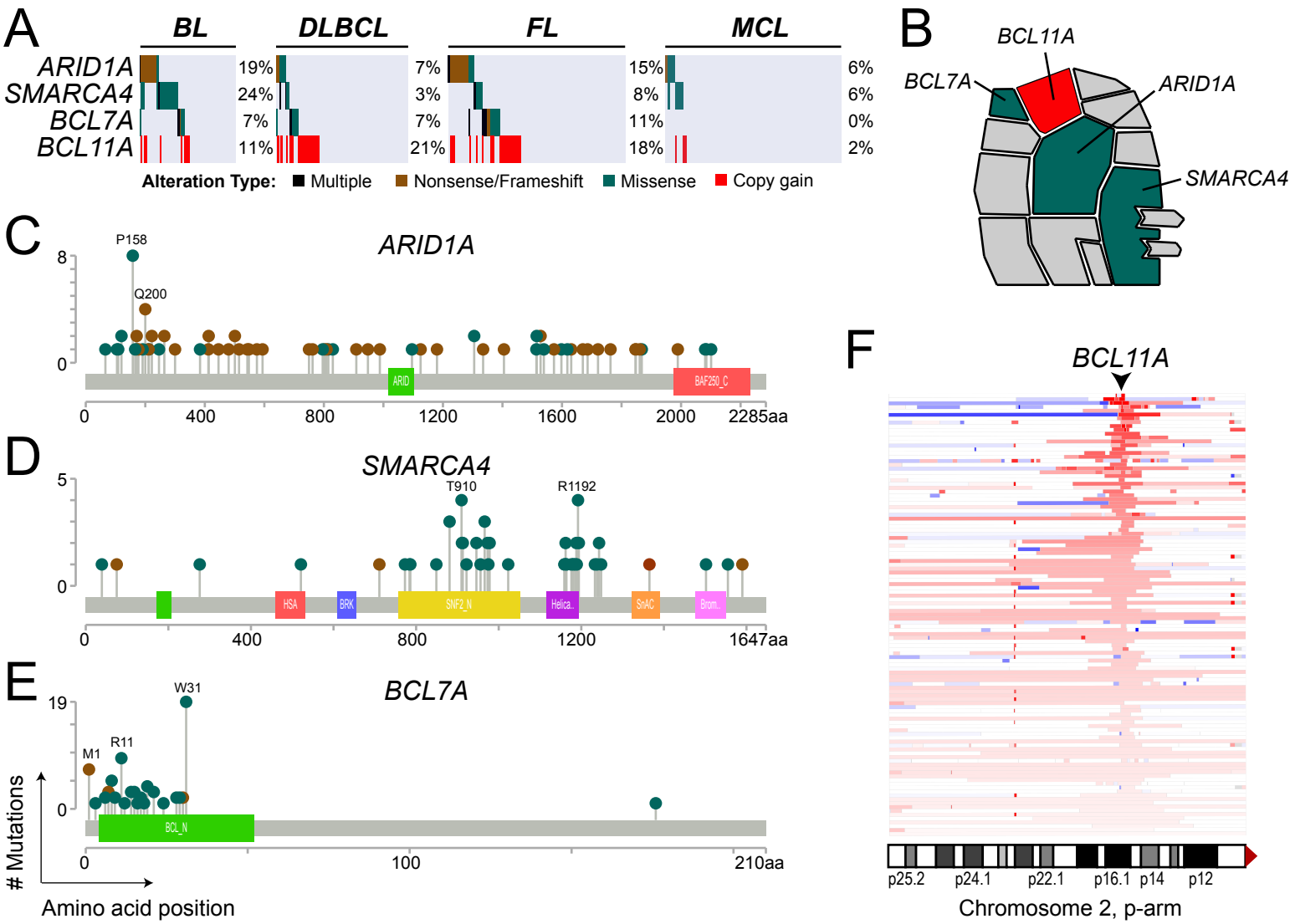




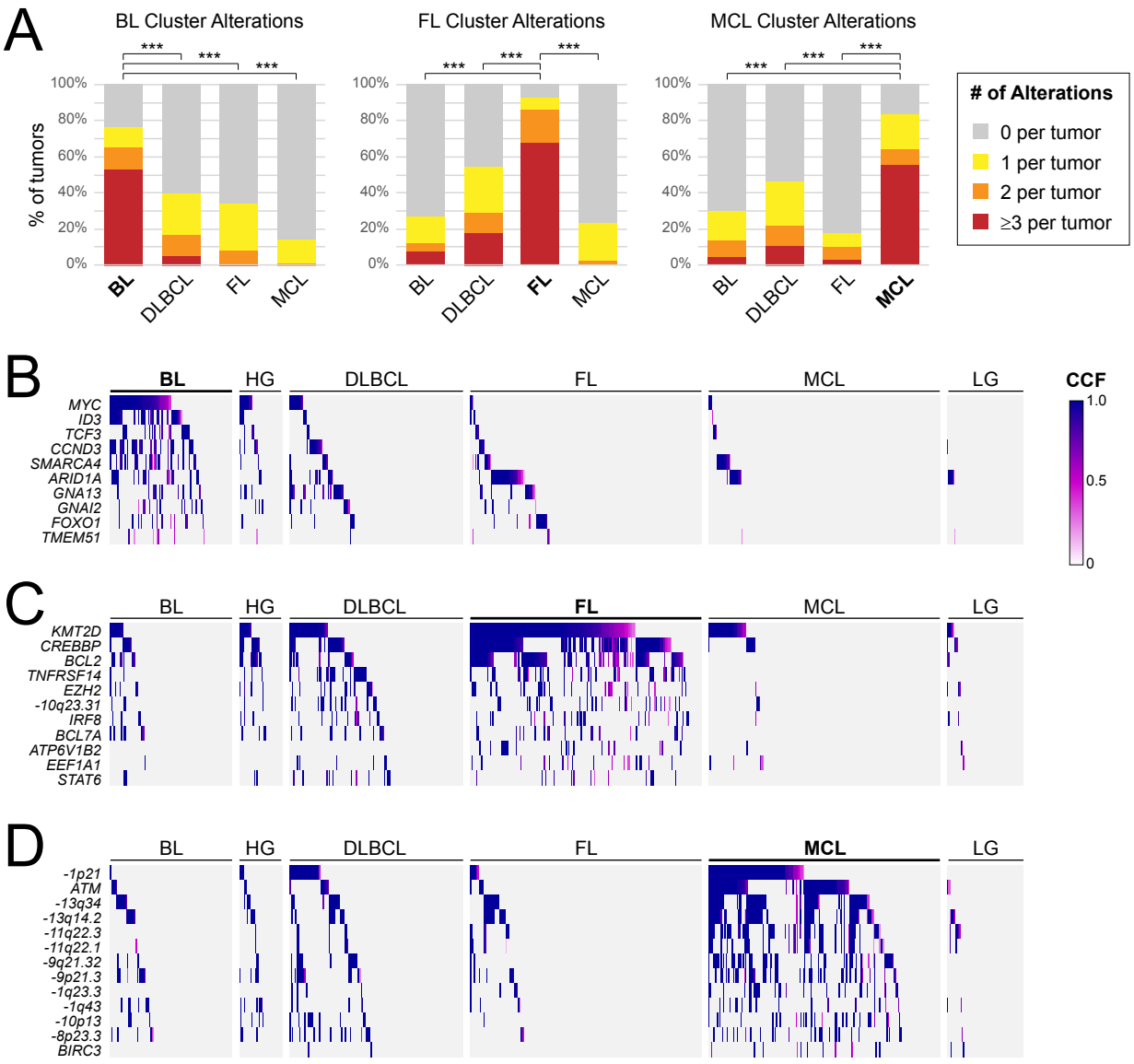
# Figure 4


BL
DLBCL
FL
MCL
Enrichment
Freq.
20 40 60
BL cluster
DLBCL cluster
FL cluster
CNA cluster
MCL cluster
Q-Value
Mutually Exclusive
Co-Associated
0.001
1
0.001
Q-value
Under-Represented
Over-Represented
0.001
1
0.001
Q-value
0.001
1
0.001
non-mBL
mBL
GCB-like
ABC-like
mBL
COO

# Figure 5


A
BL
DLBCL
FL
MCL
ARID1A
SMARCA4
BCL7A
BCL11A
19%
24%
7%
11%
7%
3%
7%
21%
15%
8%
11%
18%
6%
6%
0%
2%
Alteration Type: Multiple Nonsense/Frameshift Missense Copy gain
B
BCL11A
BCL7A
ARID1A
SMARCA4
C
ARID1A
P158
Q200
8
0
ARID
BAF250_C
0
400
800
1200
1600
2000
2285aa
D
SMARCA4
T910
R1192
5
0
HSA
BRK
SNF2_N
Helica..
SnAC
Brom..
0
400
800
1200
1647aa
E
BCL7A
W31
19
0
M1
R11
BCL_N
0
100
210aa
# Mutations
Amino acid position
F
BCL11A
p25.2
p24.1
p22.1
p16.1
p14
p12
Chromosome 2, p-arm

# Figure 6


A
BL Cluster Alterations
FL Cluster Alterations
MCL Cluster Alterations
% of tumors
100%
80%
60%
40%
20%
0%
***
BL
DLBCL
FL
MCL
# of Alterations
0 per tumor
1 per tumor
2 per tumor
≥3 per tumor
B
C
D
HG
LG
CCF
1.0
0.5
0
MYC
ID3
TCF3
CCND3
SMARCA4
ARID1A
GNA13
GNAI2
FOXO1
TMEM51
KMT2D
CREBBP
BCL2
TNFRSF14
EZH2
-10q23.31
IRF8
BCL7A
ATP6V1B2
EEF1A1
STAT6
-1p21
ATM
-13q34
-13q14.2
-11q22.3
-11q22.1
-9q21.32
-9p21.3
-1q23.3
-1q43
-10p13
-8p23.3
BIRC3