Analysis of the Coding Genome of Diffuse Large B-Cell Lymphoma

Laura Pasqualucci[1,2,3], **Vladimir Trifonov**[4], **Giulia Fabbri**[1], **Jing Ma**[5], **Davide Rossi**[6], **Annalisa Chiarenza**[1], **Victoria A. Wells**[1], **Adina Grunn**[1], **Monica Messina**[1], **Oliver Elliot**[4], **Joseph Chan**[4], **Govind Bhagat**[2,3], **Amy Chadburn**[7], **Gianluca Gaidano**[6], **Charles G. Mulligan**[5], **Raul Rabadan**[4], and **Riccardo Dalla-Favera**[1,2,3,8]

[1]Institute for Cancer Genetics, Columbia University, New York, NY 10032, USA

[2]Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY 10032, USA

[3]Department of Pathology & Cell Biology, Columbia University, New York, NY 10032, USA

[4]Department of Biomedical Informatics and Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032, USA

[5]Department of Pathology, St Jude Children's Research Hospital, Memphis, TN 38105, USA

[6]Division of Hematology, Department of Clinical and Experimental Medicine and IRCAD, Amedeo Avogadro University of Eastern Piedmont, Novara 28100, Italy

[7]Department of Pathology, Northwestern University. Feinberg School of Medicine, Chicago, IL 60611, USA

[8]Department of Genetics & Development, Columbia University, New York, NY 10032, USA

Abstract

Diffuse large B-cell lymphoma (DLBCL) is the most common form of human lymphoma. While a number of structural alterations have been associated with the pathogenesis of this malignancy, the full spectrum of genetic lesions that are present in the DLBCL genome, and therefore the identity of dysregulated cellular pathways, remains unknown. By combining next-generation sequencing and copy number analysis, we show that the DLBCL coding genome contains on average more than 30 clonally represented gene alterations/case. This analysis also revealed mutations in genes not previously implicated in DLBCL pathogenesis, including those regulating chromatin

methylation (*MLL2*, 24% of cases) and immune recognition by T cells. These results provide initial data on the complexity of the DLBCL coding genome and identify novel dysregulated pathways underlying its pathogenesis.

Introduction

Diffuse large B-cell lymphoma (DLBCL), the most common lymphoid malignancy in adulthood, is considered a curable disease in only ~50% of cases[1,2]. This incomplete success indicates the need for further studies aimed at elucidating its pathogenesis through the identification of genes and cellular pathways that are structurally altered in this disease and may represent targets for novel therapeutic approaches.

Current understanding of the biology of DLBCL indicates the existence of several distinct subgroups reflecting the derivation from B-cells at discrete stages of differentiation, namely germinal-center B-cell-like (GCB) DLBCL, activated B-cell-like (ABC) DLBCL, and primary mediastinal B-cell lymphoma (PMBCL)[3–5]. These subtypes are associated in part with distinct genetic lesions, indicating the involvement of separate oncogenic pathways[6,7]. In particular, chromosomal translocations involving the *BCL2* and *MYC* oncogenes and mutations of the *EZH2* methyltransferase are almost exclusively observed in GCB-DLBCL[7–9], while the less curable ABC-DLBCL is preferentially associated with structural alterations disrupting terminal B-cell differentiation (*PRDM1* inactivation)[10–12] and with a variety of lesions that result in the constitutive activation of the NF-κB transcription complex (most commonly, mutations of *TNFAIP3, CARD11, CD79B* and *MYD88*)[13–16]. More recently, chromosomal translocations involving *CIITA* and amplifications/rearrangements of the genes encoding for the CD274/PDL1 and CD273/PDL2 ligands have been reported as recurrent events in PMBCL[17–19]. Additional lesions are common to GCB- and ABC-DLBCL, including *BCL6* translocations[20] and inactivation of the acetyltransferase genes *CREBBP* and *EP300*[21]. Overall, these alterations only affect a fraction of cases, while the full spectrum of genomic lesions that contribute to malignant transformation remains unknown.

Major improvements in sequencing technologies have now provided an unprecedented opportunity to examine the cancer genome for large-scale identification of genomic alterations in a comprehensive and unbiased manner. The present study was aimed at expanding our current knowledge of the number and type of genetic alterations that are present in the DLBCL coding genome by integrating massively parallel whole-exome sequencing (WES) and genome-wide high-density single nucleotide polymorphism (SNP) array analysis. This combined approach allowed the definition of the degree of complexity characterizing the DLBCL coding genome, and revealed the involvement of several previously unrecognized dysregulated genes/pathways.

Results

Mutational load by exome sequencing analysis

In order to determine the load of somatic, non-silent mutations that are clonally represented in the DLBCL genome, we performed massively parallel sequencing of paired tumor and normal DNAs from six untreated, *de novo* DLBCL patients (3 ABC, 1 GCB, 1 non-GCB, 1 unclassified; Supplementary Table 1) by using a hybridization-capture method for the enrichment of non-repetitive protein-coding genes (~85% of the CCDS database) followed by next-generation sequencing using the 454 Genome Sequencer FLX instrument (Supplementary Fig. 1). We also analyzed by direct sequencing the coding exons of selected genes that have been previously implicated in other hematologic cancers, but are not represented in the capture array (e.g., *NOTCH1, MLL2* and *CYLD*)[22–27]. The WES approach produced 11.8 billion base pairs of mapped sequences (~2,3 million reads/case, of average length 305bp), with a mean depth of 9.8× (range: 8.9×–12.8× per case), and on average 77.7% of the target sequence being covered by at least 5 reads (range: 70.8%–81.4%) (Supplementary Table 2). This level of resolution allowed us to provide an initial estimate of the order of magnitude of the mutation load in the disease, and to identify genes that are relevant to its pathogenesis because recurrently mutated across different subtypes (probability of detecting mutations in genes that are affected at ~30% prevalence, 70%) (Methods and Supplementary Note).

Sanger-based re-sequencing of candidate variants in the matched tumor-normal DNAs validated the presence of 96 somatic, non-silent mutations involving 93 distinct genes and present in a major clone (Fig. 1a and Supplementary Tables 3 and 4). The overall mutation load/case varied significantly across samples (mean, 16; range, 5–29) (Figure 1). Mutations were largely represented by single nucleotide substitutions leading to amino acid changes (n=78; 82.2%), but also included in-frame insertions/deletions (n=3), nonsense mutations (n=4), alterations in canonical splice sites (n=7), and frameshift deletions of short nucleotide stretches (n=4), collectively accounting for ~16% of the events (Fig. 1b and Supplementary Table 3). Analogous to the spectrum reported in other cancer types[26,28], we observed a predominance of transitions over transversions (n=58:30; ratio, 1.9) and a preferential targeting of G and C nucleotides (71.6% vs 28.4% affecting A/T nucleotides; *P* value after correction for target exome sequence composition: 0.004) (Fig. 1c); moreover, there was a significant bias toward alterations at 5'-CpG-3' dinucleotides, which account for 10% of all non-synonymous changes (frequency of CpG nucleotides in the target exome, ~3%; p<0.0001)(Fig. 1d) (see Discussion).

The 93 mutated genes identified include most of the ones previously implicated in the pathogenesis of this disease, namely *PRDM1, TNFAIP3, CARD11, CD79B* and *MYD88*, as well as the acetyltransferase gene *CREBBP* (Supplementary Table 3). Of the remaining 87 genes, 26 have never been implicated in cancer to date, while 60 are reported in the COSMIC database and 4 are listed in the Cancer Gene Census database as causally related to cancer (Supplementary Table 3)[21]. Although the functional significance of the mutations found in these 87 genes is largely unknown, 71 of them (81.6%) are expected to alter the

function of the encoded protein, based on two distinct prediction algorithms (Supplementary Table 3).

Copy Number analysis

The same six DLBCL samples (and paired normal DNAs) were then analyzed for the presence of copy number changes by using the Affymetrix SNP6.0 platform, which interrogates ~1.8 million markers, including SNPs and copy number probes. This analysis identified 90 somatically-acquired genomic alterations (66 deletions and 24 gains), with significant variability across individual samples (range: 1–31) (Fig. 2 and Supplementary Table 5). Chromosomes 1, 2, 3 and 6 comprised the highest number of lesions, in agreement with previous studies using chromosomal and array-based CGH[6,29], and ten of the changes (3 losses and 7 gains) involved whole chromosomes or chromosome arms. The SNP array approach correctly identified two deletions of the *PRDM1* gene on chromosome 6q21 (case 2204 and 2210) and a 17p deletion (case 2204), which had been previously detected by FISH analysis. Additional known DLBCL-associated alterations include a focal homozygous deletion of *CDKN2A/CDKN2B* and a 6q23.3 deletion spanning the *TNFAIP3* tumor suppressor. Of the remaining lesions, nine encompassed single genes that likely represent the target of the aberration (Supplementary Table 5). Overall, the detection of 90 copy number alterations, with over 30-fold differences across the 6 patients studied, indicates that the genetic landscape of DLBCL is remarkably heterogeneous, and that largely distinct types of genes are affected in individual cases.

Overall complexity of the DLBCL coding genome

The combination of whole-exome sequencing and copy number data from the 6 index patients, together with FISH analysis for three common chromosomal translocations in lymphoma (*BCL6, BCL2* and *MYC*), provided an integrated snapshot on the complexity of alterations affecting the DLBCL coding genome. In addition to the described mutations and copy number aberrations, chromosomal translocations of *BCL6* were present in two cases, while no alterations were observed at the *MYC* and *BCL2* loci. Few genes were biallelically inactivated by a combination of truncating mutations and/or deletions, as typical of tumor suppressor genes (e.g. *TNFAIP3, PRDM1, CDKN2A/CDKN2B, TMEM30A, CD58* and *IGSF3*). Interestingly, the relative representation of point mutations versus copy number changes varied significantly within individual cases (Fig. 3). Thus, the overall load of tumor-acquired lesions that are present in the major clone was relatively heterogeneous across the 6 cases studied (average load 31.5, range 16–49).

Identification of recurrent targets of point mutations and copy number changes

One criterion to assess the pathogenetic relevance of candidate genetic lesions is to examine their recurrence in the disease, which can involve different modes of alteration, including mutations and copy number changes. To increase our ability to detect relevant tumor-associated genomic alterations across and within different DLBCL subgroups, and to identify additional candidate genes for downstream mutation screening in a larger DLBCL dataset, we first extended the SNP array analysis to 73 DLBCL biopsies representative of the two major subtypes. Consistent with the results obtained from the discovery panel, this

screening confirmed the high degree of complexity of the DLBCL genome, which displayed an average of 24 acquired copy number aberrations/case, with great variability across individual samples (range: 0 to 92), regardless of their DLBCL subtype. Losses were more common than gains (n=1108 and 765, respectively). We identified a total of 325 minimal common regions (MCR) of aberration measuring <1Mb in size and encompassing 1–3 genes, which most likely represent relevant selected targets. Of these, 241 were commonly deleted, with 154 spanning to a single gene, and 84 were gained, and including 20 focal high-level amplifications (in total 474 involved genes).

Gene-annotation clustering analysis using the DAVID algorithm and the list of 560 genes obtained by combining the WES-candidates (n=93) and the MCRs candidates (n=474, 7 of which are in common between the two groups) revealed a significant enrichment in specific functional categories, including regulation of transcription, lymphocyte activation/differentiation, chromatin modification/DNA methylation, and antigen processing and presentation (Supplementary Table 6), suggesting that dysregulation of these biological processes plays a central role in DLBCL pathogenesis.

Out of this large array of candidates, genes were prioritized for further analysis based on one or more of the following criteria: i) mutation frequency in the 6 discovery genomes; ii) participation in both point mutations and focal MCRs of aberration; iii) significance scores, as assessed by two independent statistical approaches for the analysis of copy number data, including GISTIC (which is based on the amplitude and the frequency of occurrence of copy number changes)[30] (Supplementary Fig. 2, Supplementary Tables 7 and 8) and ComFocal, a newly developed algorithm based on the size, amplitude and frequency of the copy number aberration (Supplementary Note and Supplementary Fig. 3); iv) functional annotation. Genes that had been previously implicated in DLBCL or in other hematologic malignancies were also included in the analysis. Based on these criteria, 56 genes were subjected to Sanger-based resequencing of their complete coding exons and consensus splice sites in an independent panel of at least 48 (up to 105) biopsies representative of the main DLBCL subtypes.

Figure 4 shows the prevalence of mutated cases in each of the 56 genes analyzed. Of these, sixteen were never found mutated in the 48 "screening" cases (overall frequency, including the discovery cases: 1/54, 1.8%). The remaining 40 genes harbored mutations in at least one additional patient, and 30 of them (including 14 not previously reported in the disease) were altered in >5% of cases, indicating that our selection criteria had effectively enriched the list for recurrent, thus presumably relevant targets (see Supplementary Tables 9, 10, 12, 14 and 18 for details on the mutations found). These genes point to the involvement of specific biological programs, as described in the following sections.

Frequent alterations in genes controlling chromatin methylation

A prominent feature of the DLBCL genome was the presence of multiple lesions targeting histone/chromatin modification genes. In addition to the recurrent inactivation of the histone acetyltransferases (HATs) *CREBBP* and *EP300*[21], we discovered alterations of several genes involved in the regulation of histone methylation, with mixed-lineage leukemia 2 (*MLL2*) being the most frequently mutated. *MLL2* encodes a histone methyltransferase

(HMT) that controls gene transcription by modifying the lysine-4 position of histone 3 (H3K4) and by promoting PolII-dependent activation of target genes[31]. Targeted re-sequencing of the *MLL2* coding exons in 115 DLBCLs (58 ABC-DLBCLs and 57 GCB-DLBCLs) revealed a total of 33 sequence variants distributed in 28 samples, including 21/92 biopsies and 7/23 cell lines (Fig. 5a). In most cases, the mutations were clearly inactivating events represented by nonsense mutations (n=10), frameshift insertions/deletions (n=11), and a consensus splice site mutation (Fig. 5a,b and Supplementary Table 10). As a consequence, the corresponding *MLL2* alleles are predicted to generate truncated proteins lacking the entire C-terminal cluster of conserved domains (including the SET domain) or significant portions of it (Fig. 5a). Eleven additional missense mutations were distributed along the MLL2 protein, with no apparent clustering (Fig. 5a and Supplementary Table 10). Where available (n=3 patients), analysis of paired normal DNA confirmed their somatic origin, which is strongly suggested for the remaining variants based on their absence in public and our own SNP databases (see Methods). While the functional consequences of these amino acid changes will have to be tested experimentally, five of them were located within or in close proximity to the conserved PHD, FYRN and SET domains, which are central to the MLL2 protein function.

In cell lines, *MLL2* mutations were exclusively associated with a GCB-DLBCL phenotype, where they account for ~50% of samples (n=7/16, versus 0/7 ABC-DLBCL), while their distribution in primary biopsies was not significantly different between the two subgroups (GCB-DLBCL, 27%; ABC-DLBCL, ~20%)(Fig. 5c). With one exception, *MLL2* mutations affected a single allele in all evaluable cases, and were not accompanied by deletion of the second copy (Fig. 5d). This pattern of monoallelic inactivation suggests a role for *MLL2* as a haploinsufficient tumor suppressor, mutated in ~23% of primary DLBCLs (24.3% including cell lines)(see Discussion).

In addition to *MLL2*, several genes involved in histone methylation were targeted by genomic alterations. Our analysis confirmed the occurrence of *EZH2* mutations in 6/107 (5.6%) biopsies and 4/23 lines, almost exclusively of the GCB-type (n=9/63, 14.2%)(Figs. 4 and 6a); these variants all target a specific hotspot residue that has been associated with increased H3K27 trimethylation activity[8,32,33]. In 11 patients, we identified somatic mutations (n=4) or deletions (n=7) of *KDM2B*, a gene encoding for a H3K36 histone demethylase (Fig. 6a, Supplementary Tables 9 and 11). Additionally, individual patients harbored genomic deletions or rearrangements of *MLL3* (n=6), *MLL5* (n=2) and *MLL* (Fig. 6b and Supplementary Table 11).

Since H3K4 methylation has been linked to other chromatin-modifiers, such as HATs and chromatin remodelers[34–38], we looked at the relationship between genetic lesions affecting *MLL2* and *CREBBP/EP300*, which were recently shown to be monoallelically inactivated in up to 39% of DLBCL[21]. Although the number of cases studied does not allow robust statistical analyses, most patients displayed a mutually exclusive involvement of these genes (Fig. 6c), suggesting that alterations at these chromatin modifiers may represent alternative mechanisms converging on a common transcriptional program. Together, these data suggest that disruption of histone modification/chromatin remodeling genes plays a central role in DLBCL pathogenesis (Fig. 6d).

Mutations of genes controlling immune recognition by T cells

A second set of lesions recurrently observed in DLBCL involve immune recognition and antigen presenting functions. Expanding on previous observations from isolated cases, we found frequent inactivating mutations and deletions in the β2-microglobulin (*B2M*) gene. *B2M* encodes a polypeptide found in association with the major histocompatibility complex (MHC) class I on the surface of nearly all nucleated cells[39]. In sixteen samples (13/111 biopsies and 3/23 cell lines) both *B2M* alleles were lost due to homozygous deletions (8 cases), biallelic mutations (4 cases) and hemizygous deletions with inactivating mutation of the second allele (4 cases)(Supplementary Fig. 4 and Supplementary Tables 12 and 13). Nine additional patients harbored monoallelic nonsense or frameshift mutations; however, the relatively low density of probe coverage for this small gene in the array may have prevented the identification of submicroscopic deletions in the second allele. Most mutations are predicted to generate truncated proteins as the result of premature stop codons (n=3), splice site mutations (n=2), out-of-frame indels (n=6) or mutations at the translation-initiating methionine codon (n=6); moreover, 4 missense mutations were found associated with inactivation of the second allele in 4 cases, as documented by sequencing analysis of cloned PCR products, suggesting that they may also be functionally relevant (Supplementary Table 12). Taken together, these lesions predict the loss of B2M expression, which is required for cell surface expression of HLA class I molecules and recognition by cytotoxic T lymphocytes[39].

Focal homozygous deletions (n=4, including one cell line), truncating mutations (n=7, distributed in 5 cases and 1 cell line), in-frame deletions (n=1) and hemizygous deletions (n=10) were also recurrently detected in the *CD58* gene, a member of the immunoglobulin superfamily that functions as a ligand of the CD2 protein on T lymphocytes, participating in their adhesion and activation[40] (Supplementary Fig. 4, Supplementary Tables 14 and 15). Although no copy number data were available to assess the status of the second allele in 3 of the mutated cases, and one patient apparently retained a normal allele, the detection of frequent biallelic inactivation (n=6 samples) suggests that CD58 may play a tumor suppressor function in DLBCL.

A novel finding was the presence of focal deletions in *TNFSF9*, the gene encoding for a transmembrane cytokine that belongs to the tumor necrosis factor (TNF) ligand family (n=10/79 cases, of which 3 homozygous) (Supplementary Table 16). TNFSF9 interacts with a costimulatory receptor molecule in T lymphocytes and follicular dendritic cells (TNFRSF9), and is involved in antigen presentation and in the generation of cytotoxic T cells[41]. Interestingly, *TNFSF9* knock-out mice develop GC-derived lymphoma[42].

Other lesions affecting regulators of immune responses include mutations and genomic breakpoints of the MHC class II transactivator gene *CIITA*, as well as amplifications and breaks in the genes encoding for the receptor immunomodulatory proteins *PDL2* and *PDL1*, often occurring simultaneously (Supplementary Fig. 4a,d). In PMBCL, rearrangements of *CIITA* have been recently shown to cause downregulation of surface HLA class II expression[19], which is associated with reduced tumor cell immunogenicity, while amplifications of the *PDL1* locus (also found in PMBCL and Hodgkin lymphoma) have

been linked to impaired anti-tumor immune responses in several cancers[17,18]. While no clear relationship was observed between alterations at these five loci, their ability to interfere with the interaction between tumor cells and the microenvironment suggests that they may facilitate lymphomagenesis by allowing escape of immuno-surveillance mechanisms.

Alterations in pathways of post-GC differentiation

Our results extend previous reports on the high frequency of genetic lesions affecting a variety of signaling pathways that share their ability to induce the NF-κB transcription complex, including the B-cell receptor (BCR), CD40 and Toll-like receptor pathways[13–16,43], Overall, components of these pathways were found structurally altered in 63% (36/59) ABC- and ~31% (12/41) GCB-DLBCL biopsies (Supplementary Fig. 5a), consistent with the observation that >90% of ABC-DLBCL and ~30% of GCB-DLBCL display nuclear NF-κB, as direct evidence of its constitutive activation[13,44]; these comprise previously identified genes (e.g., the negative regulator *TNFAIP3* and the positive regulators *CARD11, MYD88* and *CD79B*), but also novel targets, such as *ITPKB* and *TRAF3* (Supplementary Fig. 5b,c). NF-κB alterations showed a largely mutually exclusive pattern of distribution (Supplementary Fig. 5d), implicating activation of NF-κB as a major downstream effect shared by the above lesions. Nonetheless, additional consequences of individual lesions may include deregulation of distinct signaling pathways, such as PI3K and MAPK (mutations of *CD79B* and *CARD11*) and/or JAK/STAT (*MYD88*) (Supplementary Fig. 5e).[14,16]

NF-κB alterations were frequently associated with a set of genetic lesions in downstream components of the signaling pathway that regulates GC exit and the generation of plasma cells, ultimately causing a block in terminal B-cell differentiation. Specifically, we confirmed the common biallelic loss of *PRDM1/BLIMP1* in ABC-DLBCL and its mutually exclusive relationship with chromosomal translocations of *BCL6*, which substitute its promoter region making it constitutively active and resistant to IRF4-mediated suppression[45] (Supplementary Fig. 6). Since *PRDM1* is a direct target of BCL6, these translocations will also result in the block of plasma cell differentiation (Supplementary Fig. 5e). However, deregulation of BCL6 has broader consequences, including suppression of DNA damage responses by its direct targets (e.g. p53 and p21)[46,47], which may be provided by other presently unknown lesions in the remaining cases.

Other transcription factors involved in B-cell differentiation emerged as common targets of various numerical or structural aberrations, including *PAX5, IRF4, ETS1* (Supplementary Table 17) and *MEF2B*, a gene highly expressed in the GC[48] which was found mutated in 8/96 biopsies (Fig. 4 and Supplementary Table 18) and rearranged in one case (not shown). Although the functional consequences of the above lesions remain to be defined, these findings indicate that, collectively, over half of all DLBCL harbor defects at one or more genes involved in GC differentiation.

Discussion

The main findings of this study are the initial elucidation of the complexity of the DLBCL coding genome, and the discovery of a novel set of recurrent lesions that may be of relevance for the understanding of the pathogenesis of this malignancy.

The combined set of mutations and CNAs detected in the six DLBCL discovery cases does not allow a final assessment of the precise number of genomic alterations affecting coding genes, but provides information about the order of magnitude of the lesions associated with this malignancy. The estimate of >30 alterations/case emerging from this study reflects only those events with abundant to complete clonal representation, i.e. changes that were likely present during the initial phases of tumor expansion, and thus promoted malignant transformation. Furthermore, this estimate does not include an additional 40% of mutations that may have been missed by the WES approach due to the relatively low depth of coverage (see Methods), as well as chromosomal translocations other than those affecting *BCL6, BCL2* and *MYC* (not detectable by the two methodologies used). Thus, it can be concluded that the coding genome of DLBCL contains <100 lesions on average. Although very approximate, this figure is informative for future studies and may serve as an initial database for the determination of recurrence in additional panels of cases.

When compared to other malignancies, the order of magnitude of lesions detected in DLBCL appears lower than that reported for certain epithelial cancers[49,50]. Among hematologic malignancies, the complexity of the DLBCL genome is not significantly different from multiple myeloma[51], while it appears much more complex than acute myeloid leukemia[52–54] and chronic lymphocytic leukemia[23,27] in terms of both CNAs and mutational load. On the other hand, the observed predominance of transitions over transversions, the preferential targeting of C:G and G:C basepairs and the significant bias toward alterations at CpG dinucleotides is emerging as a common feature shared by most malignancies studied so far, including those of epithelial derivation. This pattern of alterations is generally derived from endogenous biochemical processes, such as the spontaneous deamination of 5-methylcytosine residues. However, it cannot be excluded that the abnormal and/or ectopic activity of activation-induced cytidine deaminase (AICDA) has a role in the generation of DNA lesions affecting these residues; indeed, AICDA has been shown to target the 5' portion of multiple genes –mostly noncoding– in DLBCL, and some of the mutations detected in this study may in fact reflect the activity of aberrant somatic hypermutation (for example, PIM1)[55].

Among the vast array of genetic lesions identified in the DLBCL genome, alterations of MLL family members, and in particular *MLL2*, appear especially frequent. *MLL2* encodes a trimethyltransferase with well-documented influence on the expression of a large number of genes, including homeobox genes. The pattern of monoallelic somatic inactivation observed in DLBCL suggests a role for *MLL2* as a haploinsufficient tumor suppressor, consistent with the observation that monoallelic *MLL2* truncating mutations leading to descreased gene dosage have a pathogenic effect in Kabuki syndrome, a congenital disorder characterized by developmental and intellectual abnormalities[56]. A role for *MLL* genes in malignant transformation is supported by the involvement of *MLL* in the pathogenesis of acute

leukemia[57], although by a distinct and partially unclear mechanism, and by the recent finding of somatic inactivating mutations in both *MLL2* and *MLL3* in various cancers[26,51,58]. Collectively, alterations of chromatin modifying enzymes, including HMT and HAT, emerge as the most frequent alterations associated with DLBCL pathogenesis, being present in over one third of cases independent of disease subtype (GCB-DLBCL, ~50%; ABC-DLBCL, ~30%).

A second novel finding in the DLBCL coding genome is the high frequency of alterations in genes that are involved in immune recognition by T cells. In particular, the observation that bi-allelic inactivation of *B2M* is more common than originally implicated based on only a few cases analyzed[59] suggests a major role for these lesions in causing the loss of HLA class I expression, which constitutes a frequent event in DLBCL[60]. The findings herein provide a mechanistic explanation for a fraction of these cases, and suggest the existence of additional genetic or epigenetic mechanisms preventing the expression of HLA class I molecules. Since it is well established that the lack of these molecules makes cells insensitive to cytotoxic T cell-mediated killing, loss of *B2M* may represent a major mechanism of tumor escape from immune surveillance.

The distribution of recurrent lesions among DLBCL subtypes defined by cell of origin supports the existence of both common and subtype-specific genetic lesions. The former are represented by those inactivating chromatin-modifying functions, including HAT and HMT genes, as well as the newly discovered alterations of immune recognition functions. These alterations may represent common pathways necessary for the development of the DLBCL phenotype, independent of cell of origin. Conversely our results confirm the preferential association of *BCL2* and *MYC* alterations with GCB-DLBCL[7], and of alterations in the NF-κB and BCL6/BLIMP1 axis with ABC-DLBCL[10]. The observed distribution has immediate clinical implications since it suggests the development of therapies that combine drugs targeting commonly altered pathways with those targeting pathways selectively disrupted in DLBCL subtypes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Abramson JS, Shipp MA. Advances in the biology and therapy of diffuse large B-cell lymphoma: moving toward a molecularly targeted approach. Blood. 2005; 106:1164–1174. [PubMed: 15855278]

2. Swerdlow, SH., et al. Lyon: International Agency for Research on Cancer (IARC); 2008. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues.

3. Alizadeh AA, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000; 403:503–511. [PubMed: 10676951]

4. Rosenwald A, et al. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. J Exp Med. 2003; 198:851–862. [PubMed: 12975453]

5. Savage KJ, et al. The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma. Blood. 2003; 102:3871–3879. [PubMed: 12933571]

6. Lenz G, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. Proc Natl Acad Sci U S A. 2008; 105:13520–13525. [PubMed: 18765795]

7. Rosenwald A, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med. 2002; 346:1937–1947. [PubMed: 12075054]

8. Morin RD, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet. 2010; 42:181–185. [PubMed: 20081860]

9. Savage KJ, et al. MYC gene rearrangements are associated with a poor prognosis in diffuse large B-cell lymphoma patients treated with R-CHOP chemotherapy. Blood. 2009; 114:3533–3537. [PubMed: 19704118]

10. Mandelbaum J, et al. BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. Cancer Cell. 2010; 18:568–579. [PubMed: 21156281]

11. Pasqualucci L, et al. Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. J Exp Med. 2006; 203:311–317. [PubMed: 16492805]

12. Tam W, et al. Mutational analysis of PRDM1 indicates a tumor-suppressor role in diffuse large B-cell lymphomas. Blood. 2006; 107:4090–4100. [PubMed: 16424392]

13. Compagno M, et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. Nature. 2009; 459:717–721. [PubMed: 19412164]

14. Davis RE, et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. Nature. 2010; 463:88–92. [PubMed: 20054396]

15. Lenz G, et al. Oncogenic CARD11 mutations in human diffuse large B cell. Science. 2008; 319:1676–1679. [PubMed: 18323416]

16. Ngo VN, et al. Oncogenically active MYD88 mutations in human lymphoma. Nature. 2010

17. Green MR, et al. Integrative analysis reveals selective 9p24.1 amplification, increased PD-1 ligand expression, and further induction via JAK2 in nodular sclerosing Hodgkin lymphoma and primary mediastinal large B-cell lymphoma. Blood. 2010; 116:3268–3277. [PubMed: 20628145]

18. Rui L, et al. Cooperative epigenetic modulation by cancer amplicon genes. Cancer Cell. 2010; 18:590–605. [PubMed: 21156283]

19. Steidl C, et al. MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers. Nature. 2011; 471:377–381. [PubMed: 21368758]

20. Iqbal J, et al. Distinctive patterns of BCL6 molecular alterations and their functional consequences in different subgroups of diffuse large B-cell lymphoma. Leukemia. 2007; 21:2332–2343. [PubMed: 17625604]

21. Pasqualucci L, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature. 2011; 471:189–195. [PubMed: 21390126]

22. Annunziata CM, et al. Frequent engagement of the classical and alternative NF-kappaB pathways by diverse genetic abnormalities in multiple myeloma. Cancer Cell. 2007; 12:115–130. [PubMed: 17692804]

23. Fabbri G, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. J Exp Med. 2011

24. Keats JJ, et al. Promiscuous mutations activate the noncanonical NF-kappaB pathway in multiple myeloma. Cancer Cell. 2007; 12:131–144. [PubMed: 17692805]

25. Mendez-Lago M, Morin R, Andrew J, Mungall A, Chan S, Marra M. Mutations In MLL2 and MEF2B Genes In Follicular Lymphoma and Diffuse Large B-Cell Lymphoma. Blood (ASH Annual Meeting Abstracts). 2010:116–473.

26. Parsons DW, et al. The genetic landscape of the childhood cancer medulloblastoma. Science. 2011; 331:435–439. [PubMed: 21163964]

27. Puente XS, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011

28. Greenman C, et al. Patterns of somatic mutation in human cancer genomes. Nature. 2007; 446:153–158. [PubMed: 17344846]

29. Cigudosa JC, et al. Cytogenetic analysis of 363 consecutively ascertained diffuse large B- cell lymphomas. Genes Chromosomes Cancer. 1999; 25:123–133. [PubMed: 10337996]

30. Beroukhim R, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci U S A. 2007; 104:20007–20012. [PubMed: 18077431]

31. Prasad R, Zhadanov AB, Sedkov Y, Bullrich F, Druck T, Rallapalli R, Yano T, Alder H, Croce CM, Huebner K, Mazo A, Canaani E. Structure and expression pattern of human ALR, a novel gene with strong homology to ALL-1 involved in acute leukemia and to Drosophila trithorax. Oncogene. 1997 Jul 31; 15(5):549–560. [PubMed: 9247308]

32. Sneeringer CJ, et al. Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proc Natl Acad Sci U S A. 2010; 107:20980–20985. [PubMed: 21078963]

33. Yap DB, et al. Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. Blood. 2011; 117:2451–2459. [PubMed: 21190999]

34. Dou Y, et al. Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell. 2005; 121:873–885. [PubMed: 15960975]

35. Ernst P, Wang J, Huang M, Goodman RH, Korsmeyer SJ. MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. Mol Cell Biol. 2001; 21:2249–2258. [PubMed: 11259575]

36. Flanagan JF, et al. Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature. 2005; 438:1181–1185. [PubMed: 16372014]

37. Li H, et al. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature. 2006; 442:91–95. [PubMed: 16728978]

38. Wysocka J, et al. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature. 2006; 442:86–90. [PubMed: 16728976]

39. Cresswell P, Ackerman AL, Giodini A, Peaper DR, Wearsch PA. Mechanisms of MHC class I-restricted antigen processing and cross-presentation. Immunol Rev. 2005; 207:145–157. [PubMed: 16181333]

40. Moingeon P, et al. CD2-mediated adhesion facilitates T lymphocyte antigen recognition function. Nature. 1989; 339:312–314. [PubMed: 2566919]

41. Wang C, Lin GH, McPherson AJ, Watts TH. Immune regulation by 4-1BB and 4-1BBL: complexities and challenges. Immunol Rev. 2009; 229:192–215. [PubMed: 19426223]

42. Middendorp S, et al. Mice deficient for CD137 ligand are predisposed to develop germinal center-derived B-cell lymphoma. Blood. 2009; 114:2280–2289. [PubMed: 19608748]

43. Lenz G, Staudt LM. Mechanisms of Disease: Aggressive Lymphomas. New England Journal of Medicine. 2010; 362:1417–1429. [PubMed: 20393178]

44. Davis RE, Brown KD, Siebenlist U, Staudt LM. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. J Exp Med. 2001; 194:1861–1874. [PubMed: 11748286]

45. Saito M, et al. A signaling pathway mediating downregulation of BCL6 in germinal center B cells is blocked by BCL6 gene alterations in B cell lymphoma. Cancer Cell. 2007; 12:280–292. [PubMed: 17785208]

46. Phan RT, Dalla-Favera R. The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. Nature. 2004; 432:635–639. [PubMed: 15577913]

47. Phan RT, Saito M, Basso K, Niu H, Dalla-Favera R. BCL6 interacts with the transcription factor Miz-1 to suppress the cyclin-dependent kinase inhibitor p21 and cell cycle arrest in germinal center B cells. Nat Immunol. 2005; 6:1054–1060. [PubMed: 16142238]

48. Potthoff MJ, Olson EN. MEF2: a central regulator of diverse developmental programs. Development. 2007; 134:4131–4140. [PubMed: 17959722]

49. Beroukhim R, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010; 463:899–905. [PubMed: 20164920]

50. Pleasance ED, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature. 2010; 463:184–190. [PubMed: 20016488]

51. Chapman MA, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011; 471:467–472. [PubMed: 21430775]

52. Ley TJ, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008; 456:66–72. [PubMed: 18987736]

53. Radtke I, et al. Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. Proc Natl Acad Sci U S A. 2009; 106:12944–12949. [PubMed: 19651601]

54. Walter MJ, et al. Acquired copy number alterations in adult acute myeloid leukemia genomes. Proc Natl Acad Sci U S A. 2009; 106:12950–12955. [PubMed: 19651600]

55. Pasqualucci L, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature. 2001; 412:341–346. [PubMed: 11460166]

56. Ng SB, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010; 42:790–793. [PubMed: 20711175]

57. Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer. 2007; 7:823–833. [PubMed: 17957188]

58. Dalgliesh GL, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010; 463:360–363. [PubMed: 20054297]

59. Jordanova ES, Riemersma SA, Philippo K, Schuuring E, Kluin PM. Beta2-microglobulin aberrations in diffuse large B-cell lymphoma of the testis and the central nervous system. Int J Cancer. 2003 Jan 20; 103(3):393–398. [PubMed: 12471623]

60. Riemersma SA, et al. Extensive genetic alterations of the HLA region, including homozygous deletions of HLA class II genes in B-cell lymphomas arising in immune-privileged sites. Blood. 2000; 96:3569–3577. [PubMed: 11071656]

61. Tchernitchko D, Goossens M, Wajcman H. In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics. Clin Chem. 2004; 50:1974–1978. [PubMed: 15502081]

62. Lin M, et al. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics. 2004; 20:1233–1240. [PubMed: 14871870]

63. Mullighan CG, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature. 2007; 446:758–764. [PubMed: 17344859]

64. Pounds S, et al. Reference alignment of SNP microarray signals for copy number analysis of tumors. Bioinformatics. 2009; 25:315–321. [PubMed: 19052058]

65. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics. 2007; 23:657–663. [PubMed: 17234643]
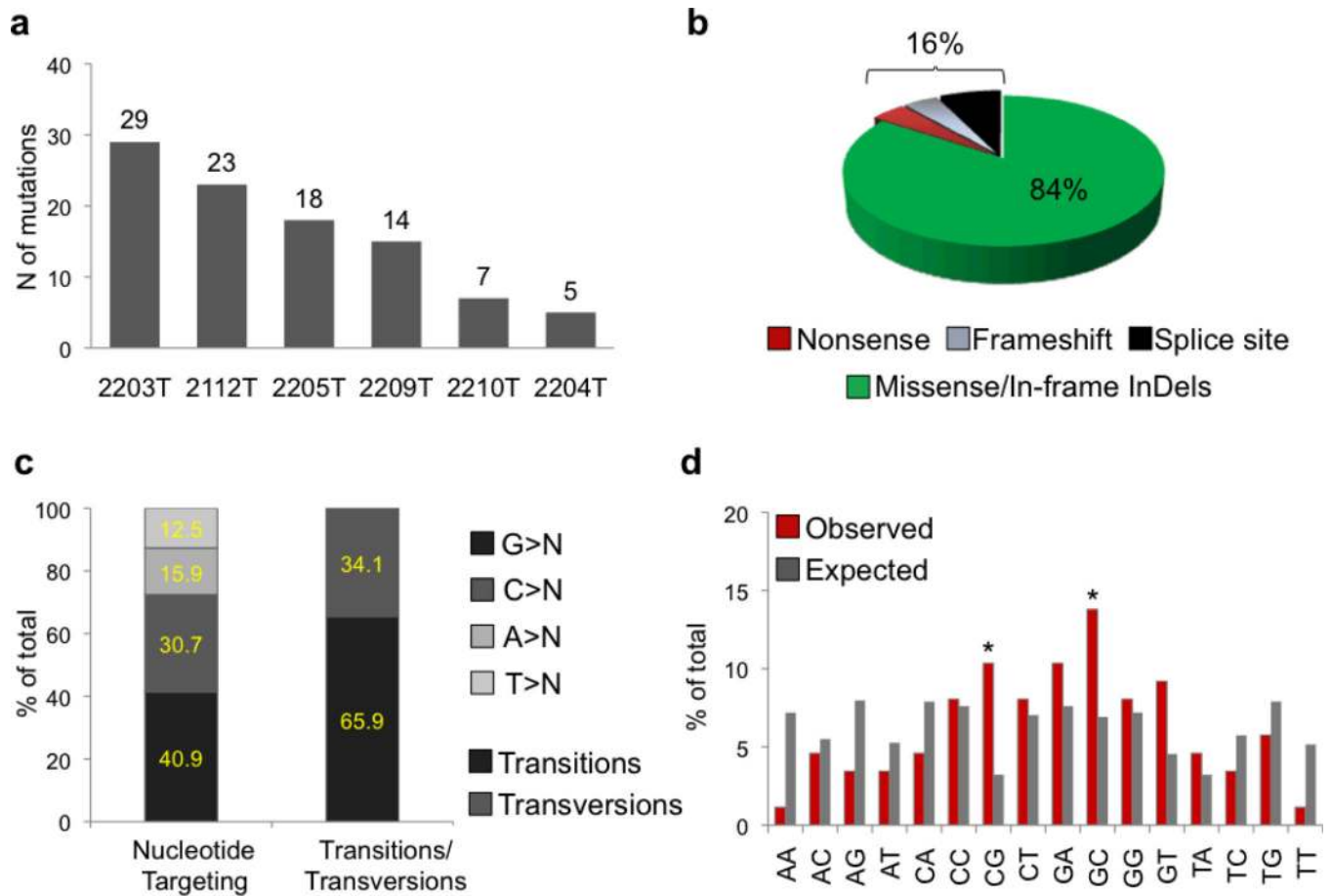
Figure 1. DLBCL non-silent mutation load

Number (**a**) and type (**b**) of Sanger-validated, non-silent mutations identified in the 6 DLBCL discovery patients. **c,** Nucleotide targeting of the DLBCL-associated point mutations (nucleotide composition of the target exome: G, 26.2%; C, 25.7%; T, 22.0%; A, 25.9%). **d,** Observed mutation frequencies at specific dinucleotides (red bars). The expected frequencies (grey bars) correpond to the dinucleotide sequence composition of the target exome. Asterisks denote statistically significant differences in overrepresented changes, as assessed by a Poisson distribution after correction for multiple hypotheses.
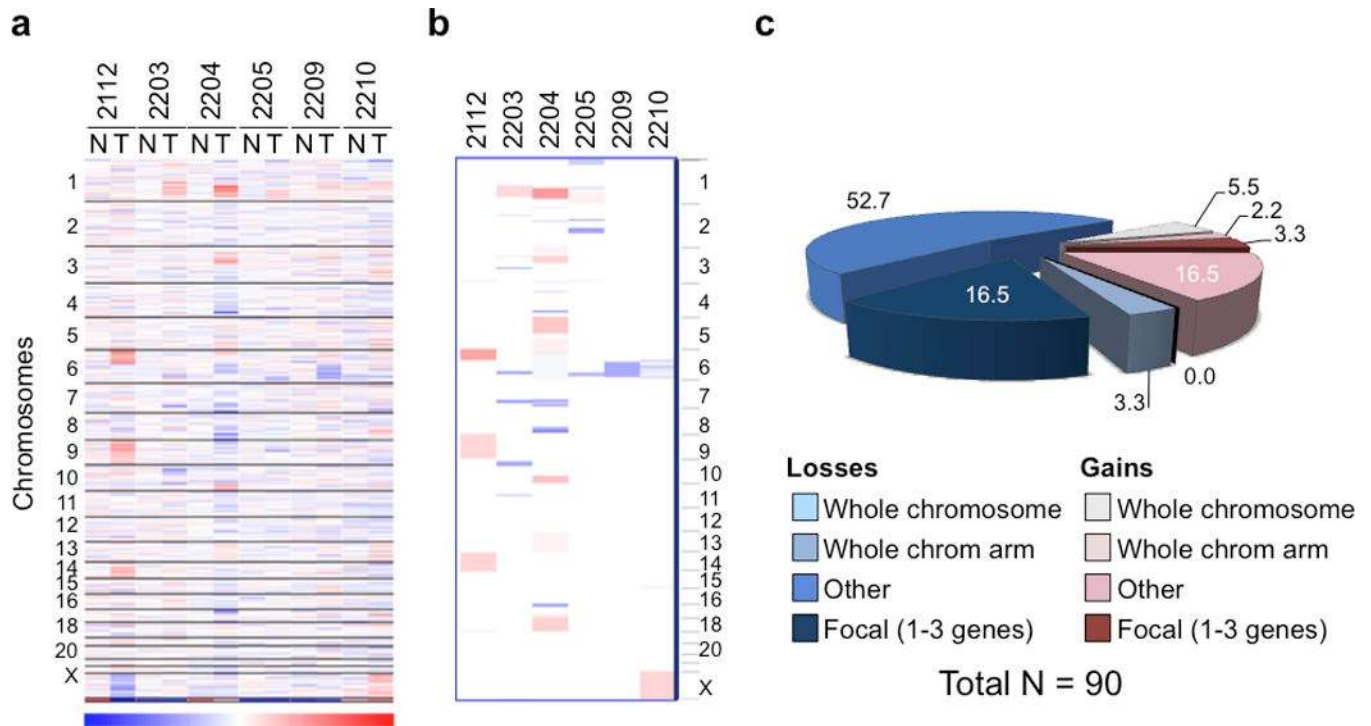
Figure 2. Copy number analysis of the 6 DLBCL discovery cases
a, dChipSNP heatmap showing median-smoothed log2 copy number ratio of all chromosomes in the 6 index DLBCL biopsies and their paired normal DNAs. In the red-blue scale, white corresponds to a normal (diploid) copy number log-ratio, blue is deletion and red is gain. **b,** Curated segmentation data for the 6 DLBCL samples shown in panel **a**, obtained as described in Methods and visualized using the Integrative Genomics Viewer (IGV) software (http://www.broadinstitute.org/igv). **b,** Overall number and frequency of copy number alterations.
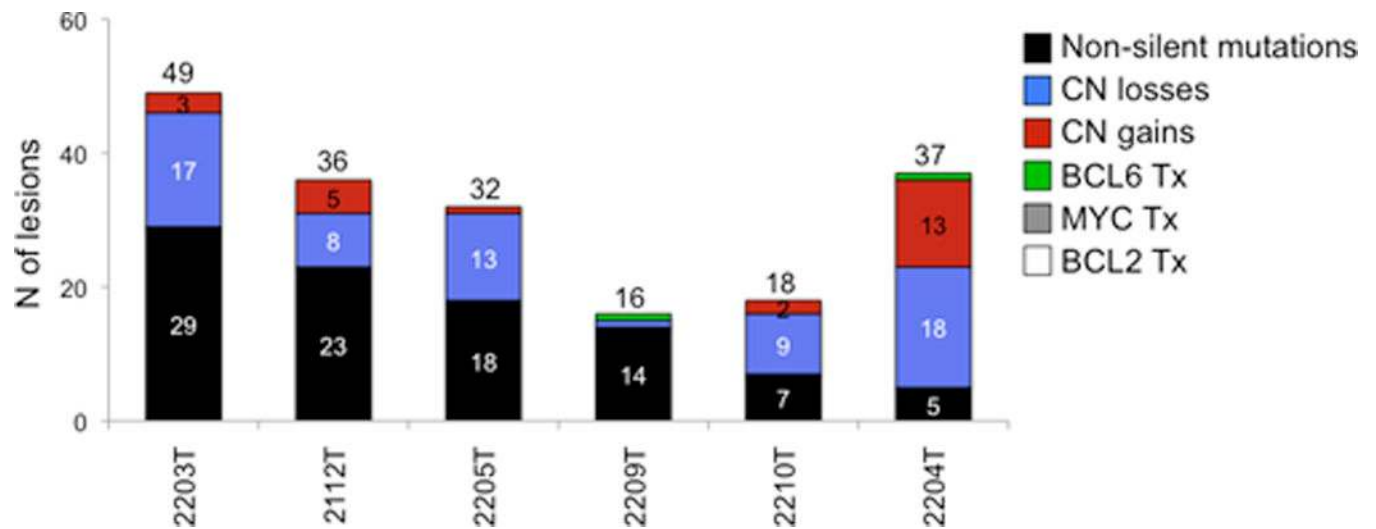
Figure 3. DLBCL harbors a heterogeneous load of numerical and structural genomic aberrations

Combined load of genetic lesions that were present in the major tumor clone of the 6 DLBCL discovery cases (including point mutations, copy number aberrations, and chromosomal translocations at three common target loci).
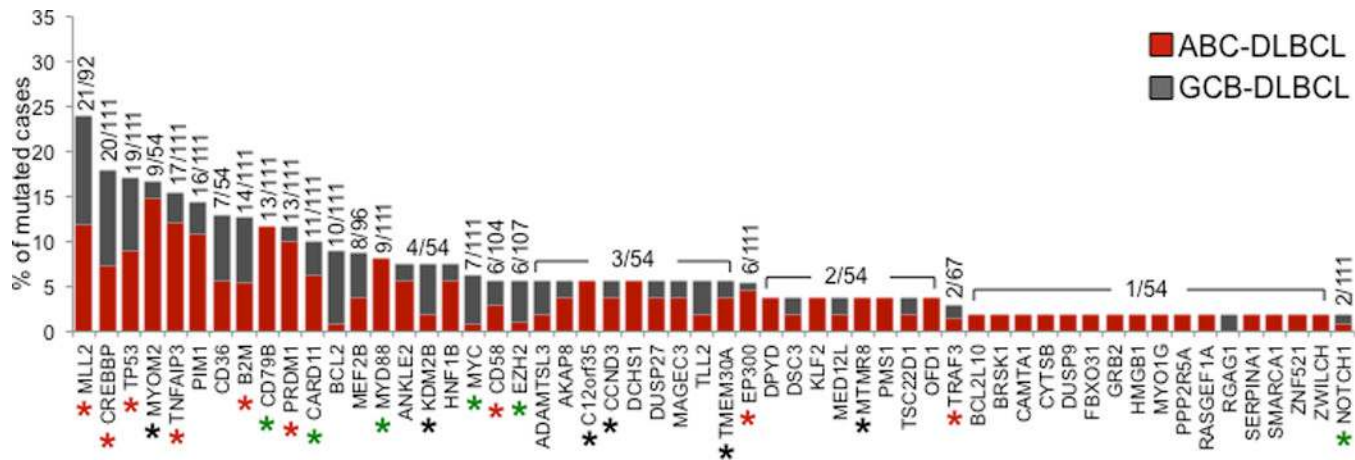
Figure 4. Recurrent mutations in DLBCL

Percentage of DLBCL primary cases harboring mutations in 56 candidate genes after targeted-resequencing of an expanded screening dataset. The final number of mutated cases over total analyzed (including the 6 DLBCL discovery cases) is given for each gene; this number includes validated somatic mutations as well as variants that are not reported in any available SNP databases (see Methods). The somatic origin of the mutations was confirmed by analysis of paired normal DNA in at least one case/gene. Red asterisks denote putative tumor suppressor genes, as suggested by the presence of predominantly inactivating mutations (stop codons, frameshift mutations and experimentally validated missense mutations). Green asterisks denote genes where the oncogenic activity of the mutation was functionally proven. Black asterisks indicate genes harboring mutations of unclear pathogenic role, but predicted to truncate the encoded protein in at least one case.
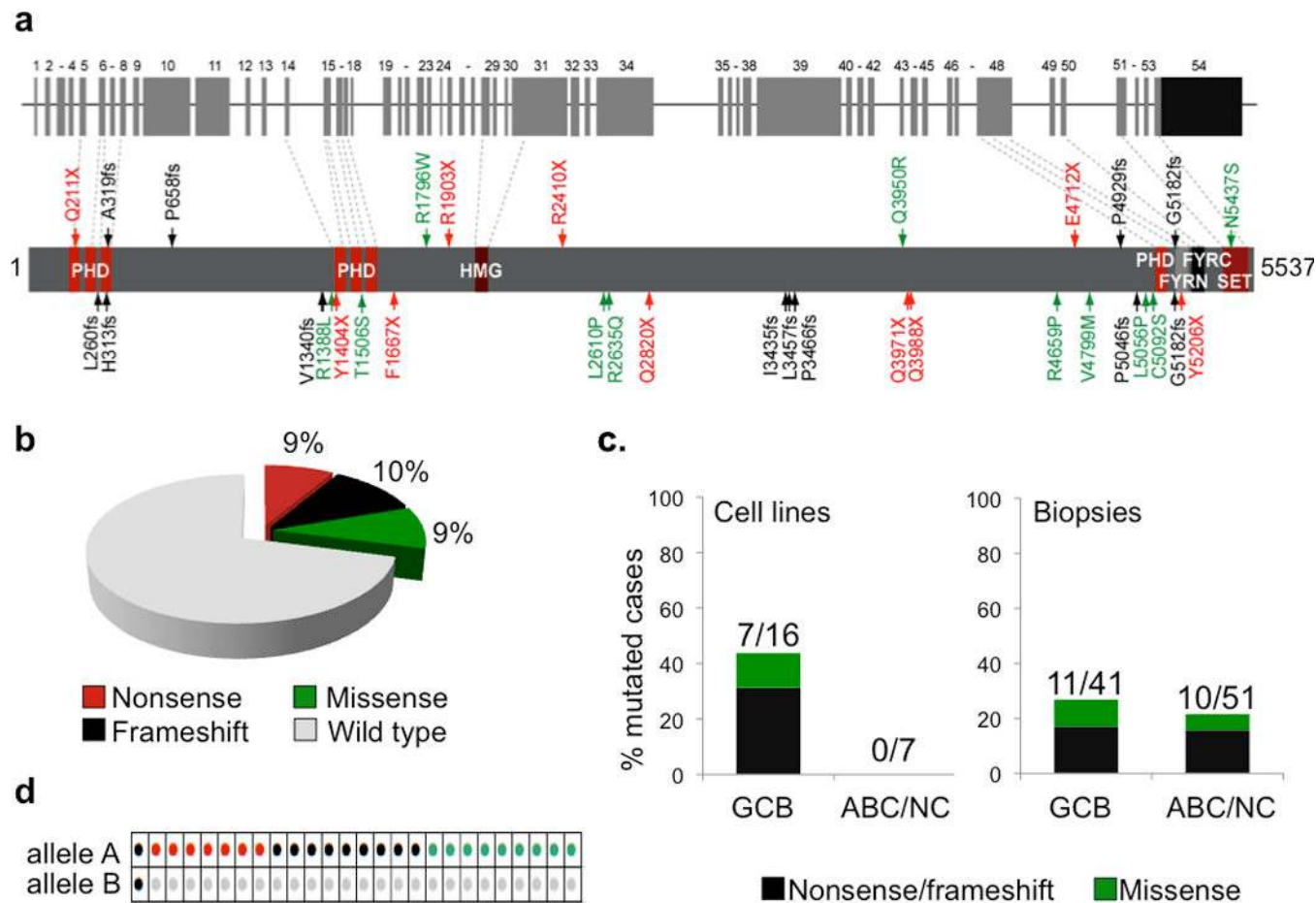
Figure 5. The *MLL2* gene is mutated in a large fraction of DLBCL
a, Schematic diagram of the *MLL2* gene (top) and protein (bottom), with its conserved functional domains (PHD, FYRN, FYRC, SET). In black, untranslated exons. Arrows, color-coded as in (**b**), indicate the position of the mutations found (cell lines on top, primary cases at the bottom). **b,** Overall percentage of DLBCL cases carrying *MLL2* mutations, according to mutation type (cell lines and biopsies). **c,** Prevalence of *MLL2* mutated cases in major DLBCL phenotypic subtypes; numbers on top indicate the actual number of mutated samples over total analyzed. **d,** Allelic distribution of the mutations in 26 of 28 affected samples for which this data could be obtained. Mutations are color-coded as in (**b**).
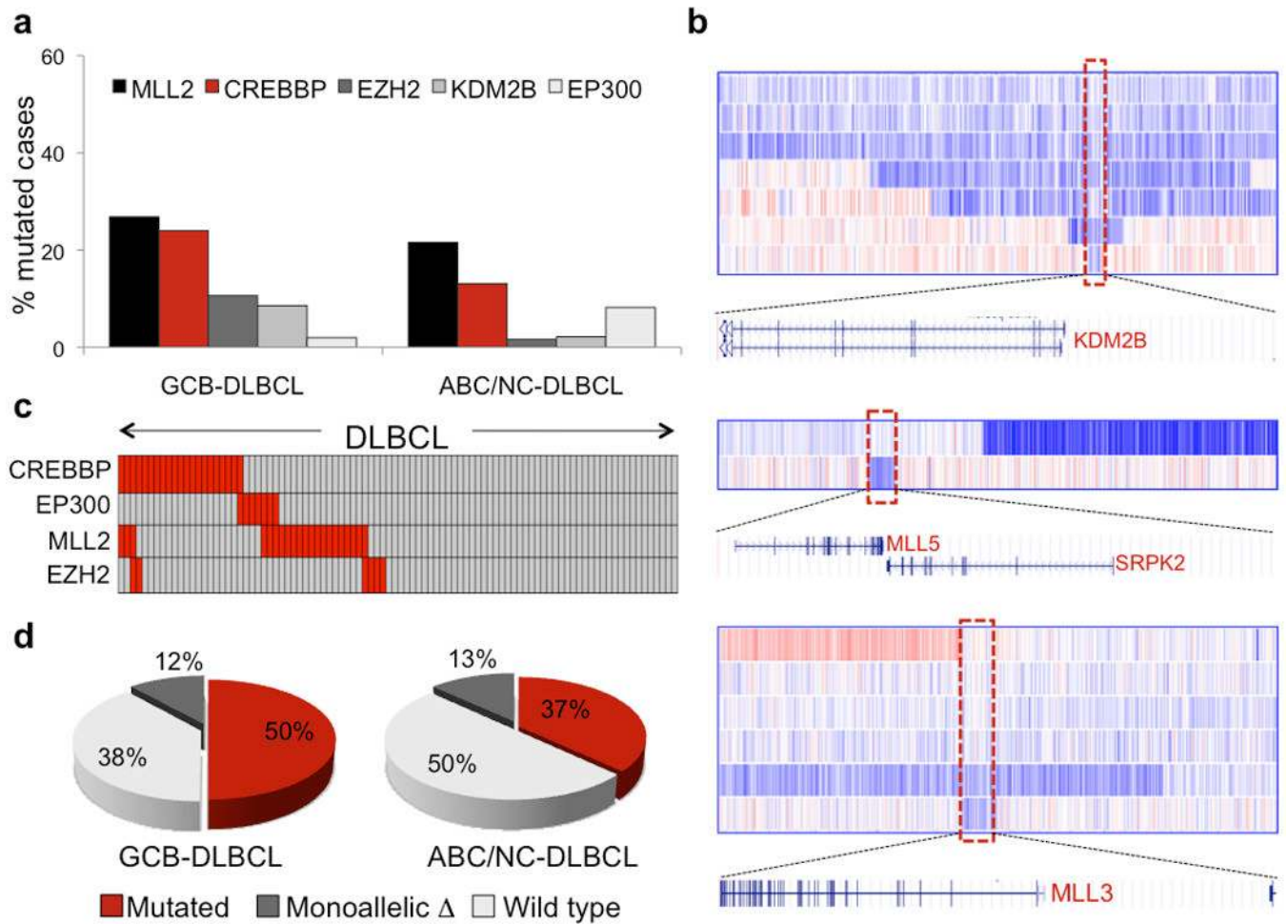
Figure 6. Disruption of histone/chromatin modification genes is a major feature of DLBCL
a, Mutation frequency of five genes encoding for histone-acetyltransferases and histone-methylation modifiers in GCB- and ABC-DLBCL (primary cases only). **b,** dChipSNP heatmap showing median-smoothed log2 copy number ratio for DLBCL biopsies harboring *KDM2B, MLL3* and *MLL5* deletions; in the red-blue scale, white corresponds to a normal (diploid) copy number, blue is deletion and red is gain. Note that, due to the presence of non-tumor cells infiltrating the biopsies, the inferred copy number, and the corresponding color intensity, may vary across samples (see Supplementary Table 11). The boxed area corresponds to the commonly deleted region, which is expanded below the heatmap to show the included genes. **c,** Relationship between *MLL2* mutations and mutations of HATs. In the heatmap, rows correspond to the indicated genes and columns represent individual samples, color-coded according to the gene status (grey, unmutated; red, mutated or deleted). Of the 3 *MLL2*-mutated cases showing simultaneous lesions at *CREBBP*, one harbored an amino acid substitution whose functional significance is presently unclear. Thus, it is possible that this change represents a passenger event or a private germline variant not annotated in currently available databases; alternatively, these mutations may have milder effects, requiring

additional cooperating lesions in the same pathway. **d,** Overall fraction of DLBCL biopsies carrying mutations at one or more of the four histone modification genes shown in (**c**).