Mutational landscape of high-grade B-cell lymphoma with MYC-, BCL2 and/or BCL6 rearrangements characterized by whole-exome sequencing

by Axel Kuenstner, Hanno M. Witte, Joerg Riedl, Veronica Bernard, Stephanie Stoelting, Hartmut Merz, Vito Olschewski, Wolfgang Peter, Julius Ketzer, Yannik Busch, Peter Trojok, Nikolas von Bubnoff, Hauke Busch, Alfred C. Feller, and Niklas Gebauer

Received: July 15, 2021.
Accepted: November 9, 2021.


Publisher's Disclaimer.
E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication. E-publishing of this PDF file has been approved by the authors. After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors' final approval; the final version of the manuscript will then appear in a regular issue of the journal. All legal disclaimers that apply to the journal also pertain to this production process.
Title: Mutational landscape of high-grade B-cell lymphoma with MYC-, BCL2 and/or BCL6 rearrangements characterized by whole-exome sequencing


Affiliations:

¹ Medical Systems Biology Group, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
² Institute for Cardiogenetics, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
³ University Cancer Center Schleswig-Holstein, University Hospital of Schleswig-Holstein, Campus Lübeck, 23538 Lübeck, Germany
⁴ Department of Hematology and Oncology, University Hospital of Schleswig-Holstein, Campus Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
⁵ Department of Hematology and Oncology, Federal Armed Forces Hospital Ulm, Oberer Eselsberg 40, 89081 Ulm, Germany
⁶ Hämatopathologie Lübeck, Reference Centre for Lymph Node Pathology and Hematopathology, Lübeck, Germany
⁷ HLA Typing Laboratory of the Stefan-Morsch-Foundation, 557565 Birkenfeld, Germany
⁸ University of Cologne, Faculty of Medicine and University Hospital Cologne, Institute for Transfusion Medicine, 50937 Cologne, Germany
⁹ Department of Paediatrics, University Hospital of Schleswig-Holstein, Campus Luebeck, 23538 Luebeck, Germany
Scientific category: Lymphoid Neoplasia

Running Title: HGBL-DH/TH mutational landscape

Key Words: High-grade B-cell lymphoma, mutational landscape, whole-exome sequencing

Word Count (Text): 4,432

Word Count (Abstract): 250

References: 52

Tables: 1

Figures: 5

Supplementary Tables: 9

Supplementary Figures: 8

Competing Interests: The authors declare no conflicts of interest.

*These authors contributed equally to this manuscript

**Shared senior authorship

Corresponding Author:
PD Dr. med. Niklas Gebauer

Department of Hematology and Oncology

UKSH Campus Luebeck

Ratzenburger Allee 160

23538 Luebeck

Email: Niklas.Gebauer@uksh.de
Abstract

High-grade B-cell lymphoma accompanied with MYC and BCL2 and/or BCL6 rearrangements (HGBL-DH/TH) poses a cytogenetically-defined provisional entity among aggressive B-cell lymphomas that is traditionally associated with unfavorable prognosis.

To better understand the mutational and molecular landscape of HGBL-DH/TH we here performed whole-exome-sequencing and deep panel next-generation-sequencing (NGS) of 47 clinically annotated cases. Oncogenic drivers, mutational signatures and perturbed pathways were compared with data from follicular lymphoma (FL), diffuse large-B-cell lymphoma (DLBCL) and Burkitt lymphoma (BL).

We find an accumulation of oncogenic mutations in NOTCH, IL6/JAK/STAT and NFkB signaling pathways and delineate the mutational relationship within the continuum between FL/DLBCL, HGBL-DH/TH and BL. Further, we provide evidence of a molecular divergence between BCL2 and BCL6 rearranged HGBL-DH. Beyond a significant congruency with the C3/EZB DLBCL cluster in BCL2 rearranged cases on an exome-wide level, we observe an enrichment of the SBS6 mutation signature in BCL6 rearranged cases. Differential gene set enrichment and subsequent network propagation analysis according to cytogenetically defined subgroups revealed an impairment of TP53 and MYC pathway signaling in BCL2 rearranged cases, whereas BCL6 rearranged cases lacked this enrichment, but instead showed impairment of E2F targets. Intriguingly, HGBL-TH displayed intermediate mutational features in all three aspects.

This study elucidates a recurrent pattern of mutational events driving FL into MYC-driven BCL2 rearranged HGBL, unveiling the mutational pathogenesis of this provisional entity. Through this refinement of the molecular taxonomy for aggressive, germinal-center derived B-cell lymphomas, this calls into question the current WHO-classification system, especially regarding the status of MYC/BCL6 rearranged HGBL.
Introduction

High-grade B-cell lymphoma with MYC-, BCL2 and/or BCL6 (HGBL) rearrangements poses a novel, yet provisional, cytogenetically-defined entity within the current WHO classification of lymphoid tumors. It is presumably allocated in the pathobiological continuum between diffuse large B-cell (DLBCL) and Burkitt lymphoma (BL) (1). The t(8;14)(q24;q32) IgH/MYC rearrangement constitutes the molecular hallmark of BL. This or further derivative chromosomal rearrangements that juxtapose MYC to a genomic enhancer, occur in approximately 10% of DLBCL and have been shown to correlate with inferior clinical outcome (2). The rearrangement is a driver of oncogenesis that is in approx. 50% of cases accompanied by additional rearrangements involving BCL2 and/or BCL6 and referred to as double-hit (DH) or triple-hit (TH) lymphomas (3-11).

While the clinical outcome in HGBL-DH/TH patients is generally poor, recent studies have hinted at a significant impact of MYC translocation partners and defined MYC/Ig-rearrangements to be the most reliable predictors of adverse outcome (2, 7).

In a prior study, we discovered an elevated frequency of TP53 impairment in MYC-driven DH/TH, whose presence was subsequently demonstrated for a subset of patients with a single-hit MYC translocation as well, indicating inferior outcome (12, 13).

By conventional cytogenetics HGBL-DH/TH were shown to recurrently harbor a complex karyotype (4). Data on the genetic basis of this entity, however, remains elusive. Several preliminary studies, predominantly focusing on HGBL-DH/TH with DLBCL morphology, have employed a panel-based next generation sequencing (NGS) approach (14-16). The insights from these studies were all restricted by gene-panel design and the associated clinicopathological data, yet their central assertions included a significant enrichment in mutations affecting CREBBP, BCL2 and KMT2D alongside an overall reflection of the phenotypical gray-zone between DLBCL and BL. Most recently, Cucco et al. elucidated significant aspects of the molecular signature of HGBL in a panel-based sequencing and gene
expression study, employing a 70-gene HaloPlex panel and an array-based gene expression approach. The authors restricted their study to samples with DLBCL morphology that stemmed from a clinical trial and the UK’s population-based Haematological Malignancy Research Network (17).

A comprehensive, exome-wide assessment of oncogenic driver mutations in HGBL-DH/TH, including cases with BL-like morphology is, however, still warranted and of vital importance to the refinement of the pathogenetic understanding of this clinically challenging entity ultimately enabling targeted therapeutic approaches.

We therefore conducted a whole-exome sequencing (WES) study on a large cohort of HGBL-DH/TH, validated by panel-based NGS and supplemented these data with a comprehensive clinicopathological assessment of the study group. Here, we report on oncogenic drivers, somatic copy number alterations (SCNAs) and putative pathway perturbations, thus refining the molecular taxonomy of MYC-driven germinal-center-derived aggressive lymphomas.

**Methods**

**Case selection and clinicopathological characteristics**

In a retrospective approach, we reviewed our institutional database to identify HGBL patients whose primary diagnostic biopsy specimen had been referred to the Reference center for Hematopathology University Hospital Schleswig Holstein Campus Lübeck and Hämatopathologie Lübeck for centralized histopathological panel evaluation between January 2007 and December 2019. For additional Information on clinicopathological work-up, please see Supplementary materials and methods and Supplementary Table 1.

This retrospective study was approved by the ethics committee of the University of Lübeck (reference-no 18-356) and conducted in accordance with the declaration of Helsinki. Patients had given written informed consent regarding routine diagnostic and academic assessment of
their biopsy specimen including molecular studies at the Reference center for Hematopathology and transfer of their clinical data.

**Whole exome and targeted amplicon-based sequencing**

Whole exome sequencing of \( n = 47 \) HGBL-DH/TH samples was performed by a hybrid capture approach with the Agilent SureSelect Human All Exon V6 library preparation kit (Agilent Technologies) followed by Illumina short read sequencing on a NovaSeq platform (Illumina) to an average depth of 304x (standard deviation ±195x; median 234x; sequencing depth was estimated using mosdepth v0.3.2 (18)) by Novogene (UK) Co., Ltd (Supplementary Table 2). Seeking to validate the initial delineation of the exome sequencing-derived mutational landscape in HGBL we employed our in-house custom AmpliSeq panel (Thermo Fisher Scientific, Waltham, Massachusetts, USA) for targeted amplicon sequencing (tNGS), encompassing all coding exons of 43 genes (see Supplementary Table 3) in 21 cases. DNA preparation for validation experiments was extracted from the same sample but in an independent approach from deeper tissue sections. Raw paired-end data (fastq format) was trimmed and quality filtered using FASTP (v0.20.0; minimum length 50bp, max. unqualified bases 30%, trim tail set to 1) (19) and trimmed reads were mapped to GRCh37/hg19 using BWA MEM (v0.7.15) (20). Resulting alignment files in SAM format were cleaned, sorted, and converted into BAM format using PICARD TOOLS (v2.18.4). Single nucleotide variants (SNVs), as well as short insertions and deletions (InDels) were identified following the best practices for somatic mutations calling provided by GATK (21). Somatic copy number aberrations (SCNAs) were identified by CONTROL-FREEC (v11.4) (22). For further details on nucleic acid extraction, panel sequencing, single nucleotide and copy number variant calling please see Supplementary methods.
Mutational deleteriousness and significance, network propagation, gene set enrichment and mutational cluster analysis

The MutSigCV algorithm was employed on WES data to delineate significantly mutated genes within the study cohort while deleteriousness was assessed via the CADD v1.3. The acquired genomic data were then processed through the LymphGen algorithm and underwent manual screening for an enrichment in overlapping aberrations with the molecular clusters proposed by Chapuy et al. followed by validation through a logistic regression framework (23, 24). Cytogenetically defined subgroups (HGBL with MYC and BCL2 aberrations, HGBL with MYC and BCL6 aberrations, and HGBL-TH) underwent differential downstream analysis by a network propagation approach simulating a protein-protein interaction network. Subsequently a gene set variation analysis was performed against HALLMARK gene sets. For details including statistical approaches correlating molecular and clinicopathological findings please see Supplementary methods.

Data availability

Sequencing data in bam format from WES and panel sequencing have been deposited in the European genome-phenome archive (EGA) under the accession number EGAS00001005420.

Results

Clinicopathological characteristics of the study group

We collected 47 cases of HGBL-DH/TH at diagnosis with sufficient FFPE tissue samples for molecular studies (median age 71; range 35 – 89 years) all of which were included in the final analysis, following successful library preparation for WES. There was insufficient clinical follow-up in 9/47 (19%) cases. An underlying HIV infection was clinically excluded in all cases. The majority of patients in our study were male (25/47; 53%) and presented with advanced stage disease (24/38 stage III/IV; 63%) and an adverse prognostic constellation.
Most patients received an intensive CHOP-like therapeutic frontline approach (25/38; 66%). The overall response rate after first line (immuno-) chemotherapy was 76% resembling a general therapeutic response in 29/38 cases.

Table 1 summarizes the baseline characteristics of all HGBL-DH/TH cases included in the current study. Histologically, the predominant morphology was that of DLBCL (NOS) (32/47), however, Burkitt-like morphology and immunophenotype was present in 15/47 patients. Cytogenetically 21/47 cases presented with MYC/BCL2, 17/47 with MYC/BCL6 double-hit constellation and 9 triple-hit lymphomas were included. MYC translocation partner revealed MYC-Ig rearrangement in 8/17 cases. The treatment outcome in our cohort was unfavorable yet in keeping with previous data reported by Rosenwald and colleagues (2). For confirmatory purposes, we included four cases, which were assessed for TP53 mutation status in a previous study and were able to validate both cytogenetic as well as molecular observations (12).

The mutational landscape of HGBL-DH/TH identified by WES

To characterize the mutational landscape in an extensive cohort of HGBL-DH/TH cases, we successfully performed WES in 47 patient-derived tumor biopsies and matched constitutional DNA in seven cases. We further applied the analytical framework outlined above to analyze WES data in the absence of paired germline DNA in the majority of cases. Following the primary identification of SNVs and indels in individual samples and subsequent filtering to correct for FFPE-derived artefacts and spurious mutations, we applied the MutSig2CV algorithm and thereby identified 22 significant candidate driver genes (p < 0.001; 13 genes with q < 0.1; Supplementary Table 4) (25).

All HGBL-DH/TH cases carried mutations in genes of oncogenic relevance according to our bioinformatic annotations. In total, we described 10,092 presumably harmful somatic mutations (cut-off see materials and methods) involving 5,521 genes after variant filtering. Of
these, SNVs and InDels represented 74.1% of the mutations (7,479 SNVs). Among them, missense mutations were the most frequent alterations (85.2%), followed by nonsense (5.7%) and InDels (5.6%), while splice-site mutations posed 3.3% of somatic mutations (Figure 1A).

Displaying an overall intermediate tumor mutational burden (median 3.974; range 1.065 – 18.234 mutations/Mbase; Figure 1B), HGBL-DH/TH revealed no evidence of MSI-related hypermutations, which is in keeping with observations in DLBCL (0.3%), yet differs from other aggressive lymphomas (e.g., primary mediastinal B-cell lymphoma) (26).

Upon comparative analysis of WES and targeted resequencing data, we were able to demonstrate a concordance rate of 92.0% (46/50 in 18 matched samples) of mutational calls, prompting high confidence in mutational calls derived from WES, even in non-germline-paired cases. A comprehensive description of all variants described by WES as well as panel based NGS is provided in Supplementary Tables 5 and 6. Nevertheless, we observed a significant enrichment of non-germline matched samples in non-synonymous SNVs, which prompted us to include significantly mutated genes according to the MUTSigCV analysis, only. As an exception to this rule, we also included MYC mutations below the statistical significance level due to their previously established clinical and functional relevance.

**Recurrent copy number alterations in HGBL-DH/TH**

We investigated our HGBL-DH/TH cohort for SCNAs employing the CONTROL-FREEC (22) algorithm in tumor-normal and tumor-only mode, respectively, followed by GISTIC2.0 (27) analysis. The analysis excluded chromosomes X and Y as well as common benign copy number variants defined by the Broad Institute’s panel of normal. Upon cross-referencing our findings with genomic loci of known oncogenes, tumor-suppressors and elements of significant signaling pathways we identified recurrent copy number gains in oncogenes such as *MEF2B* and *CSF1R*, which have previously been implicated in the pathogenesis of
malignant lymphomas (28, 29). Further, copy number losses in tumor suppressors like \textit{NPM1} were recurrently identified (Figure 2A, B). No significant differences were detected for genes affected by copy number alterations between the three cytogenetically defined subgroups (Fisher exact test \( p > 0.05 \) after Bonferroni correction for multiple testing). Common CNAs, as defined by the above referenced panel of normal were encountered at the expected frequencies (30, 31).

**Significantly mutated candidate driver genes and mutational signatures**

Putative candidate driver genes comprised several genes previously implicated in HGBL-DH/TH pathogenesis, such as \textit{KMT2D}, \textit{CREBBP} and \textit{TP53} alongside several further mutated genes such as \textit{CDKN2A}, \textit{LNP1} or \textit{SI} (14). Established mutational drivers known from other B-cell lymphoproliferative disorders (e.g., FL, DLBCL, BL) were recurrently encountered (e.g., \textit{CCND3}, \textit{ARID1A}) (Figures 2C; 3A) (32, 33). Following SNV and InDel evaluation with \textit{MutSigCV}, a network propagation approach (Figures 3B, C) was employed on significantly mutated genes to delineate the functional implications of significant genetic events on neighboring genes. Additionally, we investigated \textit{MutSigCV} genes using our HGBL-DH/TH (D/THL) cohort, the cytogenetical subgroups (BCL2, BCL6, THL), as well as cohorts of ABC-type DLBCL (\( n = 67(33) \)), GCB-type DLBCL (\( n = 45 (22) \)), BL (\( n = 108(38) \)) and FL (\( n = 199(38) \)) (all retrieved via cBioPortal) and overlapping genes between the five lymphoma subtypes and the three cytogenetical subtypes of HGBL-DH/TH (Figure 3D; E). In our limited cohort, distinctions regarding subtype-specific mutational signatures were found to be marginal among \textit{BCL2/BCL6} status or Burkitt-like vs non-Burkitt like morphology. However, we found \textit{CCND3} mutations, previously reported as driver mutations in BL pathogenesis, to be significantly enriched in HGBL-DH/TH patients with Burkitt-like morphology (9/15 vs 3/28). This observation hints at partially similar molecular paths of pathogenesis between BL and HGBL with Burkitt-like morphology. Further, an enrichment of
mutations affecting CREBBP in HGBL-DH/TH patients with BCL2 rearrangement was observed, which is well in keeping with its proposed fundamental role in FL pathogenesis. Distribution of mutations within selected, significantly mutated genes is depicted in Supplementary Figure 1. Additional profiling of mutational signatures driving HGBL-DH/TH revealed a predominance of the SBS5 (implicated in aging, potential FFPE artifacts and tobacco exposure) signature across all subtypes alongside the emphasized occurrence of the SBS6 signature (implicated in defective DNA mismatch repair (MMR)) in patients with BCL6 rearrangements (Supplementary Figure 2, Supplementary Table 7).

**Comparative analysis of mutational landscape in HGBL-DH/TH, related entities and molecular clusters in DLBCL**

Next, we sought to refine the genomic taxonomy of aggressive germinal-center-derived B-cell lymphomas and to investigate the mutational commonalities and differences between HGBL-DT/TH and other related pathological entities. We subsequently selected cBioPortal cohorts from several entities for their similar or divergent features of B-cell differentiation (FL, GCB-type DLBCL and BL versus ABC-type DLBCL,) and a comparative analysis of candidate mutational drivers in HGBL-DH/TH (as described earlier) and cohorts of ABC-type DLCBL (n = 67), GCB-type DLBCL (n = 45 (23)), BL (n = 108(34)) and FL (n = 199(34)) was conducted, screening for shared as well as mutually exclusive putative driver mutations. Interestingly, we identified one overlapping candidate driver common to all entities (KMT2D). Additionally, CREBBP was found in all entities except ABC-type DLCBL. Mutations affecting EZH2, IRF8 and TNFRSF14 were, however, specifically occured in HGBL-DT/TH and FL/GCB-type DLBCL, while CCND3 mutations appeared to be a pathogenetic feature shared between BL and HGBL-DT/TH. Additionally, TP53 mutations posed a predominant feature of aggressive lymphomas present in all HGBL-DH/TH, GCB-type DLBCL and BL types and therefore most likely acquired during high-grade
transformation (Figure 3D, E). In basic accordance with previous studies, our data suggest a common origin especially for BCL2 rearranged HGBL-DH/TH and FL/GCB-type DLBCL (14, 17).

Upon comparative investigation of our current data and mutational clusters, previously described in DLBCL, we observed a striking predominance of C3/EZB cluster cases in the BCL2 rearranged subgroup according to the integrative molecular classification proposed by Chapuy et al. and Wright et al., respectively. This is in keeping with a significant enrichment of these cases with DLBCL morphology in terms of MYC rearrangement status (Figure 4, Supplementary Figure 3) (23, 24). Complementary to our analysis, employing the LymphGen algorithm (cf. Supplementary Table 8), a logistic regression indicated a significantly different number of mutated genes in C3 between the HGBL subtypes. HGBL harboring only BCL2 were shown to exhibit the highest number of mutated C3 genes, while HGBL with BCL6 alterations had the lowest number of mutated C3 genes (BCL2/6 cases: (p = 6.059*10^{-5}, adj R2 = 0.3108). In contrast to triple hit cases, HGBL with MYC and an isolated additional BCL6 rearrangement showed a significant decrease in the number of mutated C3 genes (p = 2.00*10^{-5}, estimate: -1.7589) (Supplementary Figure 4, Supplementary Table 9). Within the subgroup of BCL6 rearranged cases, the BN2 cluster was more prominent than the EZB cluster. In keeping with their strong affinity towards the C3/EZB cluster, BCL2 rearranged cases exhibited an enrichment for mutations in CREBBP and KMT2D, while BCL6 rearranged cases were, in contrast, enriched for mutations in ARID1A. The vast majority of triple-hit cases was also classified within the EZB cluster.

**Mutational impairment of NOTCH, RTK-RAS and TP53 signaling in HGBL-DH/TH**

Cumulatively, we detected genetic lesions, putatively impairing NOTCH signaling in 74% of HGBL-DH/TH patients. Expanding on previously reported CREBBP, EP300 and DTX1 mutations in HGBL we further identified recurrent mutations affecting NCOR1 and others
NOTCH signaling was thereby the predominant target of somatic mutation in HGBL-DH/TH, albeit with a quite heterogeneous mutational pattern affecting 35/47 patients with lesions in 28/71 genes (Supplemental Figure 6A). Most of these genomic aberrations had been previously reported to be gain-of-function mutations putatively resulting in constitutive NOTCH pathway activation in various types of predominantly germinal-center derived B-cell lymphoma. Several of these mutational hits including NCOR1 and DTX1 have been shown to herald adverse clinical outcome (35, 36).

This remained the case when undertaking a differential downstream analysis within the cytogenetically defined subgroups (HGBL with MYC and BCL2 aberrations, HGBL with MYC and BCL6 aberrations and HGBL-TH), which was prompted by their significantly divergent distribution onto molecular clusters. Through this analysis a mutational signature became apparent that is additionally dominated by impairment of TP53 and MYC signaling in BCL2 rearranged cases. BCL6 rearranged cases lacked this enrichment, while HGBL-TH cases revealed intermediate mutational features. As another recurrent feature across all subgroups we observed alterations, putatively affecting IL6/JAK/STAT signaling in 74% of patients (Supplementary Figure 5). Mutations in PIM1 and SOCS1 were most frequently encountered in our case series and have been previously implicated in HGBL-DH/TH pathogenesis. These genes failed, however, to reach the predefined level of statistical significance within the scope of our MutSigCV analysis and required further investigation in a more comprehensive dataset (17). These candidate driver genes are supplemented by mutations in LTB and STAT3 (both previously identified in HIV-associated plasmablastic lymphoma) among others (37). Beyond this, we observed a relatively dispersed mutational pattern with putative driver events affecting 28 genes within the NOTCH-pathway (Supplemental Figure 6A).

In accordance with previous studies, we found mutations directly impacting NF-kB signaling in 62% of cases (Supplementary Figure 5) (38). While this was among the predominant
pathways identified through our network propagation approach, mutations affecting the pathway were narrowly detectable with only BCL2 harboring mutations in more than three patients (34%) followed by recurrent SNVs and indels in PARP1 (6%) and BIRC3 (4%) and CARD11 (4%).

The network propagation approach further underscored the mutational impairment of the aforementioned pathways alongside WNT and PI3K signaling. These observations are in accordance with preliminary impressions derived from targeted sequencing studies, employing panel-based approaches (14, 17). In addition to the divergent results from our mutational pathway analysis, we identified an enrichment of E2F targets impacted by significantly mutated genes in both double- and triple-hit cases affected by BCL6 rearrangements. Of further interest, we report on highly recurrent mutations in known Activation-induced cytidine deaminase (AID-) and somatic hypermutation (SHM) targets such as PIM1, SOCS1 and others.

Survival analysis
Upon integrated analysis of molecular and clinical data we investigated genomic alterations present in > 15% of patients for their impact on overall survival (OS) and progression-free survival (PFS). Hereby we identified ARID1A mutations to be predictive of worse clinical outcome in our cohort (OS: p = 0.0049; PFS: 0.025). Subsequent Bonferroni correction for multiple testing was performed. Thus, we identified a significant impact of mutations affecting ARID1A which was maintained regarding OS when correction for multiple testing was applied while its primarily significant effect on PFS was reduced to a trend of borderline statistical significance (Figure 5). A Cox proportional hazard model revealed this effect to be independent of the established clinical IPI prognosticators (age, LDH, extra nodal manifestations, stage, and performance status; OS: p < 0.001; HR: 13.989; 95%CI: 3.362 – 58.205; PFS: p = 0.001; HR: 6.648; 95%CI: 2.098 – 21.061). Within the cytogenetically
defined subgroups, we identified no alterations with independent impact on clinical outcome. However, a trend of borderline statistical significance towards inferior outcome in MYC/BCL2 rearranged cases harboring FOXO1 mutations was observed (Supplementary Figure 7).

Discussion
Here we report on whole-exome sequencing data from an extensive cohort of HGBL-DH/TH tumors, which is to the best of our knowledge the hitherto largest cohort and most extensive molecular data set for this entity. Previous reports on HGBL-DH/TH were limited by low sample numbers and/or targeted sequencing approaches. Contrary to this, WES here allowed to systematically define recurrent mutations, predominant mutational signatures and SCNVs in their respective clinicopathological context from which we report three central observations.

Firstly, being the first exome-wide mutational investigation for this rare subtype of lymphoma, we identify a significant overlap of mutational drivers between HGBL-DH/TH and FL as well as GCB-type DLBCL (e.g., TNFRSF14, EZH2 and IRF4) as its high-grade counterpart. Aggressive transformation was associated with the acquisition of mutations in TP53. Moreover, shared features, including CCND3 and CDKN2A mutations underscore a close molecular relation between HGBL-DH/TH and BL (34, 39). This is additionally reflected in the enrichment of HGBL-DH/TH patients with Burkitt-like morphology for CCND3 mutations. Further, we identify a number of significant mutational drivers not captured by previous, panel-based sequencing studies. Most frequently among these, we find SI mutations that have been previously implicated in CLL progression as well as mutations in POU2AF1, which has been recently found to be an augmented target of mutations during aggressive transformation of FL to DLBCL (40, 41). Although MYC did not meet the predefined MutSigCV significance level in our study, we still observed mutations in 19% of
cohort samples (Supplementary Figure 8), which is in agreement with previous panel-based studies (17).

Secondly, upon screening the mutational landscape in HGBL-DH/TH in comparison to the molecular clusters of DLBCL, proposed by Chapuy et al. and the LymphGen algorithm proposed by Wright et al., we unveil a striking overlap of BCL2 rearranged cases with the C3/EZB cluster, which was previously shown to be enriched for MYC rearrangements and oncogenic drivers implicated in FL pathogenesis (23, 24). We argue that a predominant subset of HGBL-DH/TH most likely corresponds to these transformed FL. This offers a potential explanation for the inferior clinical outcome of C3 DLBCL patients, despite their GCB-phenotype, through an enrichment for MYC rearranged HGBL-DH/TH cases. Of note, we find the predominant impairment of TP53 and to a lesser extent MYC signaling in BCL2 rearranged cases to be in keeping with a previous study on an independent set of HGBL-DH/TH, in which we found TP53 mutations to be a recurrent feature of HGBL-DH with BCL2, but not BCL6 rearrangements (12). Intriguingly, we further observed two molecular subtypes in MYC/BCL6 only rearranged cases. While selected cases were categorized within the EZB cluster, several cases revealed an association with the BN2 cluster, potentially hinting at a MYC-driven high-grade transformation of a precursor lesion with an origin within the marginal zone, as previously described (24). In addition to these observations, we found triple-hit cases to reflect EZB lymphomas in the vast majority of cases, potentially hinting at BCL6 rearrangements as late and non-defining events in HGBL-TH lymphomagenesis. Supporting this assumption, Pedrosa et al. have shown DLBCL with BCL2 and BCL6, but without MYC rearrangements to be exclusively associated with the EZB cluster (42). Considering the significantly mutated genes, our observations underscore previous assumptions regarding a molecular divergence between BCL2 and BCL6 rearranged HGBL-DH (14, 17, 43). The predominant mutational distinction between these groups was the presumably FL-derived enrichment for CREBBP mutations in the BCL2 rearranged subgroup.
On an exome-wide level we observed an enrichment of the SBS6 signature (implicated in defective DNA mismatch repair) and a significantly diminished congruency with the C3/EZB DLBCL cluster in the BCL6 rearranged subgroup. A pronounced SBS6 signature in BCL6 rearranged cases is in keeping with previous reports by Gu et al. who described genomic instability as a result of defective MMR and thereby a shorter latency to the development of BCL6-driven diffuse large B-cell lymphoma (DLBCL) in a murine model (44). Of note, these findings fundamentally dispute the combined characterization in the current WHO classification, despite several shared clinical aspects common to all subtypes of HGBL-DH/TH (1, 2). Beyond de novo DLBCL with BCL6 rearrangement, potential alternative explanations for this phenomenon include both clonal evolution and subsequent aggressive transformation from rare cases of BCL6 rearranged marginal zone lymphomas alongside BCL2 non-rearranged/BCL6 rearranged FL, which were previously shown to be characterized by a heterogenous mutational landscape (45, 46). From our data, we further deduce an intermediate role for HGBL-TH, which may indicate two divergent paths of clonal evolution originating from a BCL2 or a BCL6 driven disease with subsequent acquisition of the alternative rearrangement. The predominance of the SBS5 signature across all cytogenetic subtypes is most likely attributable to none-filtered FFPE-artifacts and advanced patient age, as was described recently (47, 48).

Lastly, we describe a pronounced mutational impairment of NOTCH, IL6/JAK/STAT and NFκB signaling pathways and recurrent oncogenetically relevant genes affected by SCNVs (including MEF2B, which was previously shown to be enriched in mutations/aberrations within the C3 DLBCL cluster) thereby systematically characterize the oncogenetic footprint of this subgroup of lymphoma. This is further combined with the identification of novel putative mutational drivers (e.g., NCOR1, DTX1, LTB and STAT3) alongside several previously established mutational hotspots in HGBL-DH/TH. Of note, and in keeping with previous observations by Zhang et al., who described an increased AID activity in double-hit
lymphomas, we observe a significant accumulation of mutations in known AID and SHM-targets such as *PIM1*, *SOCS1* and others (49-51). Moreover, among these significantly mutated genes we describe *ARID1A* which emerges as a potential prognosticator of treatment response and outcome from our correlative assessment of clinical and molecular features of our present cohort, which was found to be independent from previously established clinical prognostic factors.

We acknowledge the shortcomings inherent to the retrospective design of the study alongside the limited availability of germline DNA for matched pair analysis. The latter aspect is reflected in a significantly elevated number of mutations in non-matched samples and an uneven distribution of controlled cases unto the cytogenetic subtypes. This prompted us to limit our subsequent analysis to significantly mutated genes (except for *MYC* and *BCL2* mutations, which were additionally included based on their proven relevance in prior studies (14, 17, 23, 52) and thereby equalizing the abovementioned effect. A minor divergence in mutational calls between WES and amplicon sequencing may be attributable to a diverse clonal architecture with mutationally different subclones as DNA samples for WES and Panel-NGS were isolated from the same biopsies but different tissue sections. Additionally, on average between 78.35% and 99.17% (1st quantile 91.67%, 3rd quantile 96.59%) of the exome targets were covered with at least 40x coverage per sample, while only variants with a minimum coverage of 40x were considered present, which might have led to the exclusion of variants on a low percentage of occasions due to too low WES sequencing coverage. However, this trade-off reduces the number of false positive variant calls and thereby enhances confidence in our calls.

Pairing of our WES-results with RNA-seq data, preferably in an extended, clinically annotated cohort, which was beyond the scope of the present study, would further deepen our molecular understanding of HGBL-DH/TH, especially regarding cases with prominent Burkitt or Burkitt-like morphology.
In summary, our identification of distinct mutational landscapes among HGBL-DH/TH, derived from an exome-wide sequencing approach shows both overlapping and distinctive features compared with germinal center derived lymphomas such as GCB-type DLBCL and low-grade FL as well as BL. Our work further underscores the developing notion of a recurrent pattern of mutational events driving a potentially unidentified preexisting FL into MYC-driven HGBL-DH/TH, offering insight into the molecular pathogenesis of this provisional entity. By refining the molecular taxonomy for aggressive, germinal-center derived B-cell lymphomas, these results call into question the current WHO classification system, especially regarding the status of MYC/BCL6 rearranged HGBL.

Acknowledgments
A.K. and H.B. acknowledge computational support from the OMICS compute cluster at the University of Lübeck. The research was supported by a grant to N.G. by the Stefan-Morsch-Foundation alongside infrastructural support. H.B. acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy— EXC 22167-390884018).

References


Table 1. Clinical characteristics of the study group.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>HGBL (n = 47)</th>
<th>DHL-BCL2 (n = 21)</th>
<th>DHL-BCL6 (n = 17)</th>
<th>THL (n = 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insufficient FU</td>
<td>9 (19.1%)</td>
<td>4 (19.0%)</td>
<td>2 (11.8%)</td>
<td>3 (33.3%)</td>
</tr>
<tr>
<td>Age (yrs.; median + range)</td>
<td>71 (35 – 89)</td>
<td>73 (35 – 88)</td>
<td>72 (35 – 89)</td>
<td>66 (42 – 76)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>22 (46.8%)</td>
<td>8 (38.1%)</td>
<td>11 (64.7%)</td>
<td>3 (33.3%)</td>
</tr>
<tr>
<td>Male</td>
<td>25 (53.2%)</td>
<td>13 (61.9%)</td>
<td>6 (35.3%)</td>
<td>6 (66.7%)</td>
</tr>
<tr>
<td>R-IPI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1 (2.6%)</td>
<td>1 (5.9%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1-2</td>
<td>13 (34.2%)</td>
<td>5 (29.4%)</td>
<td>6 (40.0%)</td>
<td>2 (33.3%)</td>
</tr>
<tr>
<td>&gt;2</td>
<td>24 (63.2%)</td>
<td>11 (64.7%)</td>
<td>9 (60.0%)</td>
<td>4 (66.7%)</td>
</tr>
<tr>
<td>Stage (Ann Arbor)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>5 (13.2%)</td>
<td>2 (11.8%)</td>
<td>3 (20.0%)</td>
<td>-</td>
</tr>
<tr>
<td>II</td>
<td>9 (23.7%)</td>
<td>4 (23.5%)</td>
<td>3 (20.0%)</td>
<td>2 (33.3%)</td>
</tr>
<tr>
<td>III</td>
<td>5 (13.2%)</td>
<td>1 (5.9%)</td>
<td>2 (13.3%)</td>
<td>2 (33.3%)</td>
</tr>
<tr>
<td>IV</td>
<td>19 (50.0%)</td>
<td>10 (58.8%)</td>
<td>7 (46.7%)</td>
<td>2 (33.3%)</td>
</tr>
<tr>
<td>B-Symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>20 (52.6%)</td>
<td>10 (58.8%)</td>
<td>7 (46.7%)</td>
<td>3 (50.0%)</td>
</tr>
<tr>
<td>No</td>
<td>18 (47.4%)</td>
<td>7 (41.2%)</td>
<td>8 (53.3%)</td>
<td>3 (50.0%)</td>
</tr>
<tr>
<td>Extranodal sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>9 (23.7%)</td>
<td>3 (17.6%)</td>
<td>4 (26.7%)</td>
<td>2 (33.3%)</td>
</tr>
<tr>
<td>1-2</td>
<td>28 (73.7%)</td>
<td>14 (82.4%)</td>
<td>10 (66.7%)</td>
<td>4 (66.7%)</td>
</tr>
<tr>
<td>&gt;2</td>
<td>1 (2.6%)</td>
<td>-</td>
<td>1 (6.7%)</td>
<td>-</td>
</tr>
<tr>
<td>ECOG PS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>20 (52.6%)</td>
<td>8 (47.1%)</td>
<td>8 (53.3%)</td>
<td>4 (66.7%)</td>
</tr>
<tr>
<td>≥2</td>
<td>18 (47.4%)</td>
<td>9 (52.9%)</td>
<td>7 (46.7%)</td>
<td>2 (33.3%)</td>
</tr>
<tr>
<td>LDH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>7 (18.4%)</td>
<td>3 (17.6%)</td>
<td>3 (20.0%)</td>
<td>1 (16.7%)</td>
</tr>
<tr>
<td>Elevated</td>
<td>31 (81.6%)</td>
<td>14 (82.4%)</td>
<td>12 (80.0%)</td>
<td>5 (83.3%)</td>
</tr>
<tr>
<td>CNS involvement at diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2 (5.3%)</td>
<td>-</td>
<td>2 (13.3%)</td>
<td>-</td>
</tr>
<tr>
<td>No</td>
<td>36 (94.7%)</td>
<td>17 (100.0%)</td>
<td>13 (86.7%)</td>
<td>6 (100.0%)</td>
</tr>
<tr>
<td>Morphology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLBCL-like</td>
<td>32 (68.1%)</td>
<td>17 (80.9%)</td>
<td>12 (70.6%)</td>
<td>3 (33.3%)</td>
</tr>
<tr>
<td>Burkitt-like</td>
<td>15 (31.9%)</td>
<td>4 (19.0%)</td>
<td>5 (29.4%)</td>
<td>6 (66.7%)</td>
</tr>
<tr>
<td>Frontline therapy regimen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHOP-like</td>
<td>25 (65.8%)</td>
<td>11 (64.7%)</td>
<td>10 (66.7%)</td>
<td>4 (66.7%)</td>
</tr>
<tr>
<td>R-based</td>
<td>31 (81.6%)</td>
<td>14 (82.4%)</td>
<td>12 (80.0%)</td>
<td>5 (83.3%)</td>
</tr>
<tr>
<td>Intensified*</td>
<td>13 (34.2%)</td>
<td>7 (41.1%)</td>
<td>3 (20.0%)</td>
<td>3 (50.0%)</td>
</tr>
<tr>
<td>Less intensive**</td>
<td>9 (23.7%)</td>
<td>4 (23.5%)</td>
<td>4 (26.7%)</td>
<td>1 (16.7%)</td>
</tr>
<tr>
<td>Refusal</td>
<td>1 (2.6%)</td>
<td>-</td>
<td>1 (6.7%)</td>
<td>-</td>
</tr>
</tbody>
</table>

CHOP, cyclophosphamide/hydroxydaunorubicin/vincristine/prednisolone; CNS, central nervous system; DHL, double-hit lymphoma; DLBCL, diffuse large B-Cell Lymphoma; ECOG, Eastern cooperative oncology group; FU, follow-up; HGBL, high grade B-cell lymphoma; LDH, Lactate dehydrogenase; PS, performance status; R, rituximab; R-IPI, revised International Prognostic Index; THL, triple-hit lymphoma; Yrs., years.

*Intensified regimens: B-ALL, GMALL, CHOEP (additional etoposide), EPOCH

**Less intensive regimens: Bendamustin, mini-CHOP (50% dose reduction), rituximab mono
Figure Legends:

Figure 1. Variant classification and mutations per sample. Panel (A) shows the number of variants stratified by variant classification while panel (B) delineates the number of mutations per sample with a median of 153 mutations per sample.

Figure 2. Genomic and mutational landscape in HGBL-DH/TH. The location of SCNAs along the genome is shown in (A) (red bars denote gains; blue bars denote losses; gene names refer to affected oncogenes according to OncoKB). Panel (B) displays oncogenes from panel (A) and SCNA status (red refers to gain, blue refers to loss). Additionally, BCL2 and BCL6 status are shown for each case. Co-oncoplot for genes identified as significant driver genes by MUTSigCV ($p < 0.001$; $n = 22$) in our cohort stratified by cytogenetical subtypes is shown in panel (C); different types of mutations are colour coded and additional covariates are shown below the plot for each sample.

Figure 3. Mutational analysis for significantly mutated genes, network propagation and mutational overlap with related entities. (A) Significance levels for significantly mutated genes in our HGBL-DH/TH cohort, regardless of subgroup (MUTSigCV $p < 0.001$; gene names in orange indicate $q < 0.1$). Panels (B) and (C) show pathway enrichment analysis results for network propagation analysis (see supplementary methods for details) of significant MUTSigCV genes (MUTSigCV $p < 0.001$) for MYC/BCL2 subgroup (MYC/BCL2 genes included: POU2AF1, HVCN1, B2M, TP53, AP3S1, LNP1, CREBBP, GULP1, TNFRSF14) and MYC/BCL6 subgroup (MYC/BCL6 genes included: CDKN2A, CD78B, LNP1, KRTAP13-1, UBE2A, CCND3) against HALLMARK gene sets and NFκB pathway. UpSet plot (D) showing the overlap of MUTSigCV genes using our HGBL-DH/TH (D/THL) cohort, the cytogenetical subgroups (BCL2, BCL6, THL), as well as cohorts of ABC-type DLBCL ($n = 67(34)$), GCB-type DLBCL ($n = 45 (23)$), BL ($n = 108(34)$) and FL ($n = 118(34)$).
199(34)) (all retrieved via cBioPortal); Set size refers to the number of genes per cohort and intersection size shows the number of overlapping genes per comparison. Comparisons are denoted by black points and black connecting lines. Panel (E) shows the overlapping genes between the five lymphoma subtypes and the three cytogenetical subtypes of HGBL-DH/TH; grey denotes mutated genes.

**Figure 4. Allocation of HGBL-DH/TH samples unto the molecular subgroups/clusters of DLBCL, according to LymphGen based on their mutational signature.** Additionally, covariates are shown above the plot for each sample. Row names refer to chromosomal alterations, genes and fusions and are background coloured by their specific subtype (light blue = BN2, light orange = EZB).

**Figure 5. Survival curves according to ARID1A mutational status.** Overall (A) and progression-free survival (B) according to ARID1A mutational status. Numbers at risk alongside hazard ratios and p-values according to log-rank testing are provided. Subsequent Bonferroni correction for multiple testing (all genes with a mutational frequency > 15% were investigated) identified a significant impact of ARID1A mutations regarding OS while its primarily significant effect on PFS was reduced to a trend of borderline statistical significance.
## MYC/BCL2/BCL6

### LymphGen Subtype
- Burkitt-like
- DLBCL

### Prior Fol. Lymphoma
- na
- no
- yes

### Morphology
- Burkitt-like
- DLBCL

### Gender
- female
- male

### Matched normal tissue
- no
- yes

### Number Mutations
- 800
- 600
- 400
- 200
- 0

### Number Clones
- 6
- 4
- 2
- 0

### Alterations
- Mutation
- Truncation
- Amplification
- Translocation

### Genes
- BCL2 Fusion
- BCL6 Fusion
- BCL2
- EZH2
- TNFRSF14
- KMT2D
- CREBBP
- REL
- FAS
- IRF8
- EP300
- Chrom. 12p
- MEF2B
- CIITA
- ARID1A
- GNA13
- STAT6
- PTEN
- Chrom. 21
- EBF1
- GNAI2
- C10orf12
- BCL7A
- HLA–DMB
- S1PR2
- MAP2K1
- FBXO11
- MIR17HG
- BCL6
- NOTCH2
- TNFAIP3
- DTX1
- CD70
- BCL10
- UBE2A
- TMEM30A
- KLF2
- SPEN
- CCND3
- NOL9
- TP63
- ETS1
- HIST1H1D
- PRKCB
- HIST1H2BK
- TRIP12
- KLHL21
- TRRAP
- PABPC1
Supplement

Supplementary Materials and Methods
Clinicopathological characteristics of the study group

Both conventional and immunohistochemical slides were reviewed by a panel of experienced hematopathologists in a multi-step process, where diagnosis was confirmed by two experienced hematopathologists (ACF, HM) in accordance with the current edition of the WHO classification of tumors of the hematopoietic and lymphoid tissues (1). From 80 cases meeting diagnostic and clinical criteria for HGBL-DH/TH with available biopsy specimen, 47 were selected for subsequent genomic analysis, based on tumor DNA quality and library preparation success.

Antibodies and positivity cutoffs employed in the current study are summarized in Supplementary Table 1. Fluorescence in situ hybridization (FISH) for MYC (including MYC-Ig fusion), BCL2, BCL6 and chromogenic in situ hybridization for EBER were performed, as described (2, 3). MYC translocation partner was shown to be an Ig-gene in 45% of cases with sufficient tissue available following DNA extraction (9/20) which is in keeping with previous reports (4).

Clinical information was collected from the original files, and patients’ performance status (Eastern Cooperative Oncology Group [ECOG]), stage, treatment modalities, therapeutic response, pattern of relapse, baseline serum levels of lactate dehydrogenase (LDH), revised international prognostic index (R-IPI)(5) and information on survival were anonymously coded alongside hematopathological assessments. Extent of disease was routinely evaluated according to the Cotswold modifications of the Ann Arbor classification (6).

Extraction of nucleic acids

Tumor as well as germline DNA (when available from biopsies taken from non-involved sites; n = 7) was extracted from three FFPE tissue sections of 5µm thickness using Maxwell® RSC DNA FFPE kit (Promega), according to the manufacturers’ instructions. Quality assessment and quantification was performed on an Agilent 2100 Bioanalyzer system (Agilent Technologies).

Targeted next generation sequencing

Library preparation was carried out according to manufacturers’ instructions and sequencing was performed on the Illumina MiSeq platform (Illumina, San Diego, California, USA) to a median depth of 3569x (s.d. ± 1635).
Panel resequencing data analysis

Resequencing data was processed as described above for whole-exome data, but the remove duplicates step was omitted. Variant calling was done using FREEBAYES (v1.3.2-46-g2c1e395), variants were annotated using ANNOVAR and coverage for each variant was extracted using VCF-QUERY (7). Afterwards, variants were filtered and only variants with a minimum coverage of 100, minimum variant allele frequency of 10% and population allele frequency < 0.001 in Gnomad or PopFreqMAX database were kept for further analysis.

Variant Calling in exome sequencing data

Raw paired-end data (fastq format) was trimmed and quality filtered using FASTP (8) (v0.20.0; minimum length 50bp, max. unqualified bases 30%, trim tail set to 1) and trimmed reads were mapped to GRCh37/hg19 using BWA MEM (v0.7.15)(9). Resulting alignment files in SAM format were cleaned and sorted and converted into BAM format using Picard Tools (v2.18.4). Next, mate-pair information was fixed, duplicates were removed and base quality recalibration was performed using Picard Tools (10) and dbSNP v138. Single nucleotide variants (SNVs) and short insertions and deletions (indels) were identified following the best practices for somatic mutations calling provided by GATK (11) for version 4.1.7 or higher for matched normal and unmatched tumor samples. Briefly, GATKs Mutect2 (12) (v4.1.7.0) algorithm was applied to all BAM files with Gnomad variants as germline resource and the b37 exome panel data as panel of normal. In cases, where matched normal tissue was available (Cases 1-7), Mutect2 was run in matched tumor-normal mode. Afterwards, FFPE read orientation artefacts were identified and removed according to GATKs guidelines and left-aligned filtered variants were annotated using ANNOVAR (13) (v2019Oct24). Coverage for reference and alternative alleles for each variant were extracted using VCF-QUERY (VCFTOOLS v0.1.13(14)). The top 20 frequently mutated genes (FLAGS (15)) were removed from further analysis. Somatic variants were filtered as follows: Minimum coverage of 40, minimum variant allele frequency of 10%, variant allele frequency < 0.001 in Gnomad or PopFreqMAX database. To identify genes that are more often mutated than expected, MUTSigCV (v1.41)(16) was applied and potential driver genes were identified using p < 0.001 and q < 0.1.
Copy number aberrations

The genomic landscape of HGBL-DH/TH was assessed for somatic copy number aberrations (SCNAs) by CONTROL-FREEC (v11.4)(17) using the tumor-only mode for samples without matched and normal tissue and in matched normal-tumor mode for samples with normal tissue available (cases 1-7). The output (ratio and reads per called window) was converted to run GISTIC2.0 (v2.0.23)(18) excluding CNA calls from chromosomes X and Y and excluding known common CNAs using Broad Institute’s panel of normal (ftp://ftp.broadinstitute.org/pub/GISTIC2.0/hg19_support). GISTIC2.0 analysis was performed using default parameters.

Mutational significance and deleteriousness, network propagation, gene set enrichment and mutational cluster analysis

The effect of strong deleterious effects (CADD1.3 phred score > 20) was assessed per sample using a network propagation approach(19) applying a regularized Laplacian kernel based on STRINGdb v11(20) protein-protein interaction network (DIFFUSTATS v1.8.0(21)). Genes affected by strongly deleterious mutations were set to 1, whereas non-mutated genes were set to 0 to model the behavior of the mutation and network diffusion was performed using a parametric method with statistical normalization (z-scores).

The effect of potential driver genes identified by MUTSIGCV on neighboring genes was assessed using a network propagation approach (as described above). Resulting z-scores were used as pre-ranked input for a rank-MANOVA based statistical approach to detect enriched gene sets (MITCH R packages)(22, 23). Gene set enrichment was performed against HALLMARK gene sets and the NF-kB signaling pathway (genes were retrieved from KEGG; entry ID hsa04064). In addition, the acquired genomic data were processed through the LymphGen algorithm and the mutational patterns sequentially underwent cluster analysis and were subsequently screened manually for an enrichment in overlapping aberrations with the molecular clusters proposed by Chapuy et al. (24, 25). Further, a logistic regression framework was employed in order to test for significantly different numbers of mutated genes in a given cluster between HGBL harboring different cytogenetic constellations.
Statistical Analyses

If not stated differently, all statistical analyses were performed using R (v4.1.0) and TIDYVERSE (v1.3.1)(26) for data handling. Filtering of genomic regions was performed using GENOMICRANGES (v1.44.0)(27) and MAFTOOLS (v2.8.0)(28) were used to visualize the data. The number of somatic signatures and the contribution of each signature to each case was estimated based on the YAPSA (v1.18.0) package with calculation of correction factors for WES to avoid biases in the k-mer distribution introduced with the exome capture kit. Progression-free survival and overall survival (PFS, OS) were calculated from the date of diagnosis and censored at last clinical contact. Survival (PFS and OS) according to potential prognostic factors was estimated by means of the Kaplan–Meier method and univariate log-rank test. Survival analysis was carried out employing the R packages survival (v3.2-11) and survminer (v0.4.9).
Künstner et al.: Mutational landscape of high-grade B-cell lymphoma with MYC-, BCL2 and/or BCL6 rearrangements characterized by whole-exome sequencing – Supplement

**Supplementary Table 1.** Antibodies and positivity cutoffs employed in the current study

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Supplier</th>
<th>Clone</th>
<th>Positivity cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bcl2</td>
<td>Lab Vision</td>
<td>100/D5</td>
<td>30%</td>
</tr>
<tr>
<td>Bcl6</td>
<td>Dako</td>
<td>BG-B6p</td>
<td>30%</td>
</tr>
<tr>
<td>CD10</td>
<td>Menarini</td>
<td>56C6</td>
<td>30%</td>
</tr>
<tr>
<td>CD20</td>
<td>Dako</td>
<td>L26</td>
<td>-</td>
</tr>
<tr>
<td>CD30</td>
<td>Dako</td>
<td>BerH2</td>
<td>10%</td>
</tr>
<tr>
<td>CD38</td>
<td>Leica Biosystems</td>
<td>SPC32</td>
<td>-</td>
</tr>
<tr>
<td>CD138</td>
<td>Leica Biosystems</td>
<td>MI15</td>
<td>-</td>
</tr>
<tr>
<td>CD56</td>
<td>Leica Biosystems</td>
<td>CD564</td>
<td>10%</td>
</tr>
<tr>
<td>Kappa</td>
<td>Leica Biosystems</td>
<td>CH15</td>
<td>-</td>
</tr>
<tr>
<td>Lambda</td>
<td>Leica Biosystems</td>
<td>SHL53</td>
<td>-</td>
</tr>
<tr>
<td>MUM-1 (Irf4)</td>
<td>Dako</td>
<td>Mum 1P</td>
<td>30%</td>
</tr>
<tr>
<td>Ki-67</td>
<td>Dako</td>
<td>Mib-1</td>
<td>-</td>
</tr>
</tbody>
</table>

**Supplementary Table 2.** Mean sequencing coverage and proportion of targets covered by 40x or 100x coverage of each sample applied in WES.

**Supplementary Table 3.** Gene list for the custom AmpliSeq panel (Thermo Fisher Scientific, Waltham, Massachusetts, USA) for targeted amplicon sequencing – see separate .xlsx file

**Supplementary Table 4.** Genes found to be significantly mutated in HGBL-DH/TH according to the MutSig2CV algorithm – see separate .xlsx file.

**Supplementary Table 5.** All variants described by WES – see separate .xlsx file.

**Supplementary Table 6.** All variants described by panel based NGS – see separate .xlsx file.
Künstner et al.: Mutational landscape of high-grade B-cell lymphoma with MYC-, BCL2 and/or BCL6 rearrangements characterized by whole-exome sequencing – Supplement

**Supplementary Table 7.** Association between HGBL-DH/TH cases and mutational signatures derived from WES data – see separate .xlsx file.

**Supplementary Table 8.** LymphGen predictions for each case – see separate .xlsx file.

**Supplementary Table 9.** A logistic regression framework assessing the number of C3 subgroup mutations present in MYC rearranged with additional rearrangements of BCL2, BCL6 or both.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
<th>CI_Lower</th>
<th>CI_Upper</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>2.5714</td>
<td>0.2893</td>
<td>8.89</td>
<td>2.21*10^{-11}</td>
<td>1.988</td>
<td>3.1544</td>
<td>44</td>
</tr>
<tr>
<td>BCL2/6</td>
<td>-0.6714</td>
<td>0.4004</td>
<td>-1.677</td>
<td>1.01*10^{-01}</td>
<td>-1.478</td>
<td>0.1356</td>
<td>44</td>
</tr>
<tr>
<td>BCL6</td>
<td>-1.7589</td>
<td>0.3681</td>
<td>-4.778</td>
<td>2.00*10^{-05}</td>
<td>-2.501</td>
<td>-1.0171</td>
<td>44</td>
</tr>
</tbody>
</table>

Multiple R-squared: 0.3407, Adjusted R-squared: 0.3108
F-statistic: 12.21 on 2 and 44 DF, p-value: 6.059e-05
CI = 95% confidence interval, Std. Error = standard error. DF = degrees of freedom.
Supplementary Figure legends

Supplementary Figure 1. Distribution of mutations within selected, significantly mutated genes in the format of lollipop plots.
Supplementary Figure 2. Mutational signature analysis. Panel (A) depicts the profiling of mutational signatures driving HGBL-DH/TH revealed a homogenous predominance of the SBS5 signature alongside the significantly emphasized occurrence of the SBS6 signature (implicated in defective DNA mismatch repair) in patients with BCL6 rearrangements. Additionally, covariates are shown above the plot for each sample. Panel (B) displays the proportion of the SBS6 signature according to BCL6 FISH status, showing an elevated frequency in BCL6 rearranged cases.
Supplementary Figure 3. Allocation of HGBL-DH/TH samples unto the molecular subgroups/clusters of DLBCL, according to Chapuy et al. based on their mutational signature(24). Additionally, covariates are shown above the plot for each sample.
Supplementary Figure 4. Concordance of BCL2 and BCL6 rearranged cases with the C3 DLBCL mutational Cluster according to Chapuy et al. by means of a logistic regression model; *** denotes p < 0.001 (see Supplementary Table 8 for details).
Supplementary Figure 5. Oncoplot depicting distribution of mutations unto the genes of NOTCH, IL6/JAK/STAT and NF-kB signaling pathways. Median of variant allele frequency per gene across mutated samples is shown in the left bar plot; the bar plot on the right-hand side shows the number of mutated samples per gene and pathway. Subtypes were inferred applying the LymphGen algorithm.
Supplementary Figure 6. Mutational distribution unto oncogenic signaling pathways. Fraction of mutated genes and fraction of affected samples unto different oncogenic signaling pathways for the HGBL-DH/TH data set (A), the BCL2 subgroup (B), BCL6 subgroup (C), and the triple hit subgroup (D). While RTK-RAS, NOTCH as well as WNT signaling appear to be ubiquitously predominant targets of mutations accross all cytogenetic subtypes, the TP53 network is especially disrupted in the BCL2 subgroup.

### Table 1: Fraction of Mutated Genes and Affected Samples Unto Different Oncogenic Signaling Pathways

<table>
<thead>
<tr>
<th>Pathway</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTK-RAS</td>
<td>39/85</td>
<td>28/47</td>
<td>18/85</td>
<td>10/71</td>
</tr>
<tr>
<td>NOTCH</td>
<td>28/71</td>
<td>35/47</td>
<td>16/71</td>
<td>9/85</td>
</tr>
<tr>
<td>WNT</td>
<td>26/68</td>
<td>23/47</td>
<td>14/68</td>
<td>4/85</td>
</tr>
<tr>
<td>Hippo</td>
<td>19/38</td>
<td>27/47</td>
<td>10/38</td>
<td>2/85</td>
</tr>
<tr>
<td>PI3K</td>
<td>10/29</td>
<td>10/47</td>
<td>3/3</td>
<td>2/10</td>
</tr>
<tr>
<td>MYC</td>
<td>6/13</td>
<td>15/47</td>
<td>2/6</td>
<td>2/16</td>
</tr>
<tr>
<td>Cell_Cycle</td>
<td>4/15</td>
<td>16/47</td>
<td>4/15</td>
<td>2/16</td>
</tr>
<tr>
<td>NRF2</td>
<td>2/3</td>
<td>3/47</td>
<td>2/3</td>
<td>2/16</td>
</tr>
<tr>
<td>TGF-Beta</td>
<td>2/7</td>
<td>3/47</td>
<td>2/7</td>
<td>2/16</td>
</tr>
<tr>
<td>TPS3</td>
<td>1/6</td>
<td>9/47</td>
<td>1/6</td>
<td>1/16</td>
</tr>
<tr>
<td>TGF-Beta</td>
<td>2/7</td>
<td>3/47</td>
<td>2/7</td>
<td>2/16</td>
</tr>
<tr>
<td>TP53</td>
<td>2/7</td>
<td>2/7</td>
<td>2/7</td>
<td>2/16</td>
</tr>
<tr>
<td>MYC</td>
<td>2/7</td>
<td>2/7</td>
<td>2/7</td>
<td>2/16</td>
</tr>
<tr>
<td>PI3K</td>
<td>1/29</td>
<td>1/16</td>
<td>1/29</td>
<td>1/16</td>
</tr>
</tbody>
</table>

### Table 2: Fraction of Pathway Affected

<table>
<thead>
<tr>
<th>Pathway</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTK-RAS</td>
<td>39/85</td>
<td>28/47</td>
<td>18/85</td>
</tr>
<tr>
<td>NOTCH</td>
<td>28/71</td>
<td>35/47</td>
<td>16/71</td>
</tr>
<tr>
<td>WNT</td>
<td>26/68</td>
<td>23/47</td>
<td>14/68</td>
</tr>
<tr>
<td>Hippo</td>
<td>19/38</td>
<td>27/47</td>
<td>10/38</td>
</tr>
<tr>
<td>PI3K</td>
<td>10/29</td>
<td>10/47</td>
<td>3/3</td>
</tr>
<tr>
<td>MYC</td>
<td>6/13</td>
<td>15/47</td>
<td>2/6</td>
</tr>
<tr>
<td>Cell_Cycle</td>
<td>4/15</td>
<td>16/47</td>
<td>4/15</td>
</tr>
<tr>
<td>NRF2</td>
<td>2/3</td>
<td>3/47</td>
<td>2/3</td>
</tr>
<tr>
<td>TGF-Beta</td>
<td>2/7</td>
<td>3/47</td>
<td>2/7</td>
</tr>
<tr>
<td>TPS3</td>
<td>1/6</td>
<td>9/47</td>
<td>1/6</td>
</tr>
<tr>
<td>TGF-Beta</td>
<td>2/7</td>
<td>3/47</td>
<td>2/7</td>
</tr>
<tr>
<td>TP53</td>
<td>1/6</td>
<td>9/47</td>
<td>1/6</td>
</tr>
<tr>
<td>MYC</td>
<td>2/7</td>
<td>2/7</td>
<td>2/7</td>
</tr>
<tr>
<td>PI3K</td>
<td>1/29</td>
<td>1/16</td>
<td>1/29</td>
</tr>
</tbody>
</table>

### Table 3: Fraction of Samples Affected

<table>
<thead>
<tr>
<th>Pathway</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTK-RAS</td>
<td>39/85</td>
<td>28/47</td>
<td>18/85</td>
</tr>
<tr>
<td>NOTCH</td>
<td>28/71</td>
<td>35/47</td>
<td>16/71</td>
</tr>
<tr>
<td>WNT</td>
<td>26/68</td>
<td>23/47</td>
<td>14/68</td>
</tr>
<tr>
<td>Hippo</td>
<td>19/38</td>
<td>27/47</td>
<td>10/38</td>
</tr>
<tr>
<td>PI3K</td>
<td>10/29</td>
<td>10/47</td>
<td>3/3</td>
</tr>
<tr>
<td>MYC</td>
<td>6/13</td>
<td>15/47</td>
<td>2/6</td>
</tr>
<tr>
<td>Cell_Cycle</td>
<td>4/15</td>
<td>16/47</td>
<td>4/15</td>
</tr>
<tr>
<td>NRF2</td>
<td>2/3</td>
<td>3/47</td>
<td>2/3</td>
</tr>
<tr>
<td>TGF-Beta</td>
<td>2/7</td>
<td>3/47</td>
<td>2/7</td>
</tr>
<tr>
<td>TPS3</td>
<td>1/6</td>
<td>9/47</td>
<td>1/6</td>
</tr>
<tr>
<td>TGF-Beta</td>
<td>2/7</td>
<td>3/47</td>
<td>2/7</td>
</tr>
<tr>
<td>TP53</td>
<td>1/6</td>
<td>9/47</td>
<td>1/6</td>
</tr>
<tr>
<td>MYC</td>
<td>2/7</td>
<td>2/7</td>
<td>2/7</td>
</tr>
<tr>
<td>PI3K</td>
<td>1/29</td>
<td>1/16</td>
<td>1/29</td>
</tr>
</tbody>
</table>
Supplementary Figure 7. Survival curves according to FOXO1 mutational status. Overall (A) and progression-free survival (B) according to FOXO1 mutational status.
Künstner et al.: Mutational landscape of high-grade B-cell lymphoma with MYC-, BCL2 and/or BCL6 rearrangements characterized by whole-exome sequencing – Supplement

Supplementary Figure 8. Oncoplot depicting the mutational spectrum encountered in HGBL-DH/TH prior to filtering through the MutSigCV algorithm.
Künstner et al.: Mutational landscape of high-grade B-cell lymphoma with MYC-, BCL2 and/or BCL6 rearrangements characterized by whole-exome sequencing – Supplement

References


