Focal structural variants revealed by whole genome sequencing disrupt the histone demethylase KDM4C in B-cell lymphomas

Cristina López,[1,2*] Nikolai Schleussner,[3,4,*] Stephan H. Bernhart,[5,6,7] Kortine Kleinheinz,[8] Stephanie Sungalee,[9] Henrike L. Sczakiel,[3,4] Helene Kretzmer,[5,6,7,10] Umut H. Toprak,[11,12,13] Selina Glaser,[1] Rabea Wagener,[1,2] Ole Ammerpohl,[1,2] Susanne Bens,[1,2] Maciej Giefing,[2,14] Juan C. González Sánchez,[15] Gordana Apic,[15] Daniel Hübschmann,[16,17,18] Martin Janz,[3,4] Markus Kreuz,[19] Anja Mottok,[1] Judith M. Müller,[20] Julian Seufert,[11] Steve Hoffmann,[5,6,7,21] Jan O. Korbel,[9] Robert B. Russell,[15] Roland Schüle,[20,22] Lorenz Trümper,[23] Wolfram Klapper,[24] Bernhard Radlwimmer,[25] Peter Lichter,[25] Ralf Küppers,[26] Matthias Schlesner,[11,27] Stephan Mathas[3,4#] and Reiner Siebert[1,2#] and members of the ICGC MMML-Seq Consortium

[1]Institute of Human Genetics, Ulm University and Ulm University Medical Center, Ulm, Germany; [2]Institute of Human Genetics, Christian-Albrechts-University, Kiel, Germany; [3]Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany; [4]Hematology, Oncology and Tumor Immunology, Charité - Universitätsmedizin Berlin, Berlin, Germany, and Experimental and Clinical Research Center, a joint cooperation between the MDC and the Charité, Berlin, Germany; [5]Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig, Germany; [6]Bioinformatics Group, Department of Computer, University of Leipzig, Leipzig, Germany; [7]Transcriptome Bioinformatics, LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany; [8]Department for Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology and Bioquant, University of Heidelberg, Heidelberg, Germany; [9]EMBL Heidelberg, Genome Biology Unit, Heidelberg, , Germany; [10]Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin, Germany; [11]Bioinformatics and Omics Data Analytics (B240), German Cancer Research Center (DKFZ), Heidelberg, Germany; [12]Faculty of Biosciences, Heidelberg University, Heidelberg, Germany; [13]Hopp-Children's Cancer Center at the NCT Heidelberg (KiTZ), Division of Neuroblastoma Genomics (B087), German Cancer Research Center (DKFZ), Heidelberg, Germany; [14]Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; [15]BioQuant and Biochemie Zentrum Heidelberg (BZH), Heidelberg University, Heidelberg, Germany; [16]German Cancer Consortium (DKTK), Heidelberg, Germany; [17]Heidelberg Institute of Stem Cell Technology and Experimental Medicine (HI-STEM), Heidelberg, Germany; [18]Computational Oncology, Molecular Precision Oncology Program, National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Heidelberg, Germany; [19]Institute for Medical Informatics Statistics and Epidemiology, Leipzig, Germany; [20]Klinik für Urologie und Zentrale Klinische Forschung, Klinikum der Albert-Ludwigs-Universität Freiburg, Freiburg, Germany; [21]Leibniz Institute on Ageing-Fritz Lipmann Institute (FLI), Computational Biology, Jena, Germany; [22]BIOSS Centre of Biological Signalling Studies, Albert-Ludwigs-University Freiburg, Freiburg, Germany; [23]Department of Hematology and Oncology, Georg-August-University of Göttingen, Göttingen, Germany; [24]Hematopathology Section, Christian-Albrechts-University, Kiel, Germany; [25]Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany; [26]Institute of Cell Biology (Cancer Research), University of Duisburg-Essen, Essen, Germany, and German Cancer Consortium (DKTK) and [27]Biomedical Informatics, Data Mining and Data Analytics, Augsburg University, Augsburg, Germany

**CL and NS contributed equally as co-first authors.*
[#]SM and RS contributed equally as co-senior authors.

Correspondence:
R. Siebert
reiner.siebert@uni-ulm.de

S. Mathas
stephan.mathas@charite.de

Received: September 13, 2021.
Accepted: March 18, 2022.
Prepublished: April 28, 2022.

https://doi.org/10.3324/haematol.2021.280005

Supplementary data

Focal structural variants revealed by whole genome sequencing disrupt the histone demethylase *KDM4C* in B cell lymphomas

Cristina López[1,2*], Nikolai Schleussner[3,4,*], Stephan H. Bernhart[5,6,7], Kortine Kleinheinz[8], Stephanie Sungalee[9], Henrike L. Sczakiel[3,4], Helene Kretzmer[5,6,7,10], Umut H. Toprak[11,12,13], Selina Glaser[1], Rabea Wagener[1,2], Ole Ammerpohl[1,2], Susanne Bens[1,2], Maciej Giefing[2,14], Juan C. González Sánchez[15], Gordana Apic[15], Daniel Hübschmann[16,17,18], Martin Janz[3,4], Markus Kreuz[19], Anja Mottok[1], Judith M. Müller[20], Julian Seufert[11], Steve Hoffmann[5,6,7,21], Jan O. Korbel[9], Robert B. Russell[15], Roland Schüle[20,22], Lorenz Trümper[23], Wolfram Klapper[24], Bernhard Radlwimmer[25], Peter Lichter[25], ICGC MMML-Seq Consortium, Ralf Küppers[26], Matthias Schlesner[11,27], Stephan Mathas[3,4#], Reiner Siebert[1,2#]

[1]Institute of Human Genetics, Ulm University and Ulm University Medical Center, Ulm, 89081, Germany;
[2]Institute of Human Genetics, Christian-Albrechts-University, Kiel, 24105, Germany;
[3]Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, 13125, Germany,
[4]Hematology, Oncology and Tumor Immunology, Charité - Universitätsmedizin Berlin, Berlin, 12200, Germany, and Experimental and Clinical Research Center, a joint cooperation between the MDC and the Charité, Berlin, 13125, Germany;
[5]Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig, 04107, Germany;
[6]Bioinformatics Group, Department of Computer, University of Leipzig, Leipzig, 04107, Germany;
[7]Transcriptome Bioinformatics, LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, 04107, Germany;
[8]Department for Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology and Bioquant, University of Heidelberg, Heidelberg, 69120, Germany;
[9]EMBL Heidelberg, Genome Biology Unit, Heidelberg, 69117, Germany;
[10]Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin, Germany;
[11]Bioinformatics and Omics Data Analytics (B240), German Cancer Research Center (DKFZ), Heidelberg, 69120, Germany;
[12]Faculty of Biosciences, Heidelberg University, Heidelberg, 69120, Germany;
[13]Hopp-Children's Cancer Center at the NCT Heidelberg (KiTZ), Division of Neuroblastoma Genomics (B087, German Cancer Research Center (DKFZ), Heidelberg, 69120, Germany;
[14]Institute of Human Genetics, Polish Academy of Sciences, Poznan, 60-479, Poland;
[15]BioQuant and Biochemie Zentrum Heidelberg (BZH), Heidelberg University, Heidelberg, 69120, Germany;
[16]German Cancer Consortium (DKTK), Heidelberg, 69120, Germany
[17]Heidelberg Institute of Stem Cell Technology and Experimental Medicine (HI-STEM), Heidelberg, 69120, Germany

[18]Computational Oncology, Molecular Precision Oncology Program, National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Heidelberg, 69120, Germany
[19]Institute for Medical Informatics Statistics and Epidemiology, Leipzig, 04107, Germany;
[20]Klinik für Urologie und Zentrale Klinische Forschung, Klinikum der Albert-Ludwigs-Universität Freiburg, Freiburg, 79104, Germany;
[21]Leibniz Institute on Ageing-Fritz Lipmann Institute (FLI), Computational Biology, Jena, 07745, Germany;
[22]BIOSS Centre of Biological Signalling Studies, Albert-Ludwigs-University Freiburg, Freiburg, 79104, Germany;
[23]Department of Hematology and Oncology, Georg-August-University of Göttingen, Göttingen, 37075, Germany;
[24]Hematopathology Section, Christian-Albrechts-University, Kiel, 24105, Germany;
[25]Division of Molecular Genetics, German Cancer Research Center (DKFZ), 69120, Heidelberg, Germany;
[26]Institute of Cell Biology (Cancer Research), University of Duisburg-Essen, Essen, 45147, Germany, and German Cancer Consortium (DKTK)
[27]Biomedical Informatics, Data Mining and Data Analytics, Augsburg University, Augsburg, 86159, Germany
*contributed equally as co-first author;
# contributed equally as co-senior author.

Correspondence: R.Siebert: reiner.siebert@uni-ulm.de; S.M.: stephan.mathas@charite.de

Supplementary Material and Methods

ICGC MMML-seq cohort

The ICGC MMML-Seq cohort comprises untreated tumor tissue and corresponding germline material (cells from peripheral blood or buffy coats without IGHV clonal rearrangement) obtained with informed consent of the respective patients and/or in minors their legal guardian. The ICGC MMML-Seq study has been approved by the Institutional Review Board of the Medical Faculty of the University of Kiel (A150/10) and Ulm (349/11), and of the recruiting centers. Tumor samples were reviewed by expert hematopathologists and classified according to the WHO 2008 guidelines.[6]

In the framework of the ICGC MMML-Seq network (http://dcc.icgc.org), we mined in the present study data of a total of 186 GC-derived B-cell lymphomas, including 86 follicular Lymphoma (FL), 75 diffuse large B cell lymphoma (DLBCL), 1 primary mediastinal B-cell lymphoma (PMBL), 17 transformed DLBCL (from FL, namely: FL-DLBCL), 1 B-cell lymphoma not otherwise specified (B-NOS), 4 large B-cell lymphoma (LBCL) with *IRF4*-rearrangement, and 2 double hit lymphomas with molecular Burkitt lymphoma (mBL) signature. With the exception of 5 cases (4 LBCL with *IRF4*-rearrangement and 1 PMBL), genomic and transcriptomic data of the cohort have been previously published with regard to other features.[5]

Cell lines

Nineteen B and T-cell non-Hodgkin and 4 classical Hodgkin lymphoma (cHL) derived cell lines were used in the study (Online Supplementary Table S2). Cell lines were obtained from DSMZ (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany) or from Prof. Stephan Mathas. Cell lines were tested negative for mycoplasma contamination, and their authenticity was confirmed by STR analysis using the StemElite ID System (Promega). In addition, copy number data[7] and/or exome data[8] from previously published studies from the 4 cHL herein studied and 2 additional cHL cell lines were included as well as previously reported WGS data from L1236.[9]

METHOD DETAILS

Methods and procedures used in the ICGC MMML-Seq have been described in previous publications of the network and are summarized in the subsequent section.

Sample Processing

The study was performed in accordance with the ICGC guidelines (www.icgc.org). DNA and RNA extraction, the detection and sequencing of immunoglobulin rearrangements, and whole-genome and transcriptome sequencing of the patients have been published previously.[1] The analysis of WGS was performed as previously described[1,3]. Somatic SNVs and indels in matched tumor normal pairs were identified using the DKFZ core variant calling workflows of the ICGC PCAWG project (https://dockstore.org/containers/quay.io/pancancer/pcawg-dkfz-workflow). Initial candidate variants for SNVs in the tumor were generated by samtools and bcftools (version 0.1.19), followed by a lookup of the corresponding positions in the control. The work flow to identify putative somatic variants and indels was performed as published recently.[3] SNVs and indels were annotated using ANNOVAR[10] according to GENCODE gene annotation (version 19) and overlapped with variants from dbSNP (build 141) and the 1000 Genomes Project database. SNVs classified as splicing, non-synonymous changes, stop-gains, and stop-losses were predicted to affect protein function.

Detection of copy number alterations and genomic structural variants

Allele-specific copy-number alterations were analyzed using ACE-seq (allele-specific copy-number estimation from whole genome sequencing; unpublished data). The detailed method was previously published[3]. ACE-seq was used to determine absolute allele-specific copy numbers as well as tumor ploidy and tumor cell content based on coverage ratios of tumor and control as well as the B-allele frequency (BAF) of heterozygous single nucleotide polymorphisms (SNPs). Moreover, ploidies were manually checked and compared with FISH results.

Structural variants (SV) were called using the SOPHIA algorithm (unpublished data) and DELLY v0.5.9 as previously described[3]. The source code of SOPHIA is available at https://bitbucket.org/utoprak/sophia/ and the method detail has been reported previously.[3] We used DELLY algorithm to call simple and complex SVs. A high confident set of somatic SVs of size >1kb, supported by at least four read pairs, and filtered for absence in the paired normal control tissue was derived. Moreover, we removed SVs detected either in ≥1% of a set of 1105 germline samples from healthy samples constructed from the DELLY's consensus germline SVs called in PCAWG normal tissue sample.

To determine the incidence of SVs in Pan-cancer Analysis of Whole Genomes (PCWAG), filtered structural variant calls were generated by SOPHIA for the PCAWG cohorts processed with the same tools and settings as with the lymphoma cohort used in this study. For each patient existence of a *KDM4C* altering event was determined, using a list of SVs with two breakpoints each, as follows: i) If one breakpoint is directly on *KDM4C* (chr9:6682372-7175648, hg19) and the second is off the designated region on any chromosome the event is a *KDM4C* hit; ii) If both breakpoints are on *KDM4C* and they fall on different exons or introns, the event is a KDM4C hit; iii) If both breakpoints are on *KDM4C* and they fall on the same exon, the event is a *KDM4C* hit; iv) If both breakpoints are on the same intron the event is not a *KDM4C* hit.

Integration of different genomic variant types

SNVs, indels, SVs and CNAs were integrated to account for all variant types in the recurrence analysis. Whilst all genes with SNVs or indels in coding regions (nonsynonymous, splicing, frameshift event) and ncRNA were included, SVs and CNAs were handled differently. Any genes between the breakpoints of focal SVs (<1 Mbp) were considered. However, duplications and deletions called by SOPHIA in the range of 10 kbp and 1 Mbp had to be verified by ACEseq, discarding subclonal events with less than 0.7 copy number deviation from the average ploidy. For larger SVs only genes that were directly hit by a breakpoint were considered. Only focal CNA events (<1 Mbp) were taken into account for variant integration, as these are more likely to target specific genes within the affected region than large events such as whole chromosome arm events. To capture the precise target, focal SVs and CNAs were combined and local maxima of overlapping regions with more than one event were identified.

Clonality analysis of *KDM4C* alterations

The cancer cell fraction (CCF; the proportion of cancer cells with a KDM4C alteration among all cancer cells) was calculated as follows:

$$CCF = \frac{VAF * (TCN * TCC + 2 * (1 - TCC))}{TCC * A}$$

TCN: total copy number at locus (i.e. copy number of reference allele plus copy number of alternative allele)

TCC: tumor cell content

A: copy number of alternative allele

VAF: variant allele fraction, calculated as variant reads divided by all reads

- $VAF = \frac{v}{n+v}$

For deletions, variant reads were calculated as the sum of all split reads at both breakpoints, while normal reads were the average of non-variant supporting reads overlapping the breakpoints.

Case 4191799 has an unbalanced translocation affecting the KDM4C locus. It harbours two intact copies of chromosome 9 and one intact copy of chromosome 8. An additional copy of the telomeric part of 9p is translocated to the (8p-truncated) second copy of chromosome 8. As the total copy number is different at the two breakpoints, the CCF calculation has been performed independently for breakpoints *a* and *b*. Variant reads were taken as the sum of all split reads at both breakpoints, while normal reads were the number of non-variant supporting read at each breakpoint. The variant allele fractions were calculated as

- $VAF_a = \frac{v}{n_a+v}, VAF_b = \frac{v}{n_b+v}$

The alternative allele copy number was 1 at both breakpoints, while the reference allele copy number was 1 at breakpoint *a*, and 2 at breakpoint *b*.

For case 4193638 no cancer cell fraction has been determined as the increased copy number at this locus allows for different possible combinations of reference and alternative allele copy numbers.

To evaluate if the observed variant read counts are compatible with the variant being clonal in the cancer cells, 95% confidence intervals for the binomial probability for the observed number of variant reads and total reads at the breakpoint loci were calculated using the binconf function from the R package Hmisc using the score test-based method after Wilson. The number of observations was the sum of the normal reads at breakpoint *a*, the normal reads at breakpoint *b*, and the variant reads. The number of "successes" was the number of variant reads. Next, the expected proportion of variant reads assuming that the variants were clonal was calculated as follows:

- $p = \frac{TCC*A}{TCC*(A+R_{aT}+R_{bT})+(1-TCC)*(R_{aN}+R_{bN})}$

  - R_{aT}, R_{bT}: copy number of reference allele at breakpoint *a* and *b* in tumor
  - R_{aN}, R_{bN}: copy number of reference allele at breakpoint *a* and *b* in normal

Finally, it was evaluated if *p* is within the 95% confidence interval.

Whole-transcriptome analysis

Transcriptome data were mapped with segemehl 0.2.0[11] allowing for spliced alignments and using a minimum accuracy of 90%. Gene expression values were counted using RNAcounter 1.5.2, using the "--nh" option and counting only exonic reads (-t exon). We looked for backsplicing reads from the *KDM4C* locus in the segemehl output.

Gene expression-based GC-B-cell lymphomas classification

The gene expression classifier was performed as previously described[5]. In brief, a set of 23 differentially expressed genes to distinguish between DLBCL of type ABC and GCB has been established previously.[12] We used these genes to classify our cohort based on RNA-Seq data. We applied the thresholds of <0.25 for GCB-DLBCL and >=0.66 for ABC-DLBCL. Samples in between were labeled as unclassified lymphoma. The classifier was applied to all samples with available RNA-Seq data included in the analysis (n=180).

PCR and Sanger sequencing

SNVs and SVs involving *KDM4C* locus were amplified by polymerase chain reaction from the genomic DNA using specific primers. Primer sequences and PCR conditions are available upon request. Amplicons were purified (MinElute 96 UF PCR Purification Kit, Qiagen, Hilden, Germany) and cycle-sequenced using fluorescent dye-termination (Big Dye Terminator V1.1 Cycle Sequencing Kit, Applied Biosystems, Darmstadt, Germany) and an ABI 3100 or ABI 310 automatic capillary genetic analyser. Alternative *KDM4C* fusion transcripts were validated using specific primers to amplify breakpoint fusion sequences. The RNAs from the tumor samples were treated with DNaseI (RNAase-Free, Ambion, Thermofisher Scientific, Darmstadt, Germany). Complementary DNA (cDNA) was synthesized from 1 µg of total RNA using Quanti Tect Reverse Transcription kit (Qiagen, Hilden, Germany) according to manufacturer´s instruction. The cDNAs were amplified by polymerase chain reaction using specific primers and conditions and the amplicons were purified and sequenced as described above.

***KDM4C* fluorescent in situ hybridization (FISH)**

FISH on interphase nuclei was performed on frozen tissue sections and on methanol: acetic acid fixed cells from cell lines. Chromosomal aberrations affecting the *KDM4C* locus were analysed by FISH using two-FISH-probe designs, i.e. using locus-specific and break-apart *KDM4C* probes. The locus-

specific *KDM4C* assay comprises BAC clone RP11-940A10 spanning *KDM4C* labelled in spectrum green which was combined with a control probe for *NOTCH1* gene locus (9q34.3), which including BACs RP11-251M1 and RP11-83N9 both labelled in spectrum orange. The *KDM4C* break-apart assay contains two differentially labelled BAC clones RP11-356C23 (spectrum orange) and RP11-996O17 (spectrum green) flanking the *KDM4C* locus. BAC clones were obtained from Life Technologies, (Darmstadt, Germany). Digital image acquisition, processing, and evaluation were performed using ISIS digital image analysis version 5.0 (MetaSystems, Altussheim, Germany).

DNA constructs

For generation of the pcDNA3-*KDM4C* expression construct, *KDM4C* was amplified by use of primers *KDM4C*_FLAG_BamHI s 5´- GCGGATCCGCCTGCGAGGTGGCCGAGGTGG and *KDM4C*_EcoRIintern as 5´-GCGAATTCTCCAGCCTCCTGGGTTATC for amplification of the 5´-part of the cDNA (referred to as "part A"), and by use of primers *KDM4C*_EcoRIintern s 5´- GCGAATTCATGATCACTTTCCCATATG and *KDM4C*_Stop_XhoI as 5´- GCCTCGAGGCCTACTGTCTCTTCTGGCAC for amplification of the 3´-part of the cDNA. Thereafter, parts A and B were cloned consecutively *via* BamHI and EcoRI as well as EcoRI and XhoI into a modified pcDNA3 construct containing an N-terminal FLAG tag. For the pRTS-1[13] based inducible *KDM4C* expression vector, *KDM4C* was amplified by use of primers *KDM4C*_5´_ATG_XbaI s 5´- GCTCTAGAGCCACCATGGAGGTGGCCGAGGTGGAAAG and *KDM4C*_3´STOP_XbaI as 5´- GCTCTAGACTACTGTCTCTTCTGGCAC using pcDNA3-*KDM4C* as template. The amplified *KDM4C*-product was ligated *via* XbaI into pUC19-Sfi, and mobilized by SfiI digestion for cloning into pRTS-1. For generation of the lentiviral *KDM4C* expression construct, human *KDM4C* was amplified from the pRTS1-*KDM4C* plasmid by use of primers *KDM4C*_AgeI s 5´-CAACCGGTGCCACCATGGAGGTGGCCGAGGTGG and *KDM4C*_XbaI as 5´-CATCTAGACTGTCTCTTCTGGCACTT and cloned into lentiCRISPRv2-EBFP[14] *via* AgeI and XbaI. All DNA constructs were verified by Sanger sequencing.

Cell lines, culture conditions, transfections and transduction, functional analyses

Cell lines were cultured as previously described.[15] Where indicated, 1 μg/ml doxycycline (Dox; Sigma Aldrich, Taufkirchen, Germany) was added. For generation of KDM4C-inducible cells, cells were electroporated in OPTI-MEM I using Gene-Pulser II (Bio-Rad); L1236 with 960 μF and 160 V, KARPAS-422 with 500 μF and 300 V, NAMALWA with 500 μF and 300 V, HEK293 with 960 μF and 180 V. Twenty-four hours after transfection, Hygromycin B (L1236 110 μg/ml; Karpas-422 50 μg/ml; Namalwa 250 μg/ml; Sigma-Aldrich, Taufkirchen, Germany) was added. After 21 – 28 days of culture in the presence of Hygromycin B, cells were suitable for functional assays. Where indicated, GFP+

cells were enriched 72 hours after dox-induction using a FACS Aria. Production of lentiviruses and lentiviral transduction of cells was performed as described[14]. In brief, $5x10^5$ HEK293T/17 cells were seeded in DMEM (10% FCS; 1% Penicillin/Streptomycin; 1% sodium pyruvate; 1% Glutamax) the day before transfection. For transfection, 10 μg of the lentiviral *KDM4C*-plasmid, 5 μg of the packaging plasmid psPAX2 (Addgene plasmid #12260, a kind gift of D. Tron) and 5 μg of the packaging plasmid pCMV-VSV-G (Addgene plasmid #8454, a kind gift of B. Weinberg) were transfected using calcium-phosphate-precipitation. 48 hours after transfection, the viral supernatant was harvested. For viral transduction, $8x10^5$ SU-DHL-1 cells were seeded in 2 ml, and 2 – 3 ml of viral supernatant was added. Centrifugation was carried out with 2,000 g for 90 min at 32 °C. The day after transduction, cells were washed three times with 1 x PBS. The percentage of viable GFP-positive cells was determined by propidium iodide (PI)-staining, and FACS analyses were performed using a FACS Canto II over time at the indicated times. The percentage of PI-negative, GFP-positive cells at each time point was normalized to the percentage of PI-negative, GFP-positive cells at the starting day of the respective assay, as indicated.

Immunoblotting

Protein preparation and immunblotting were performed as described[15]. The following primary antibodies were used: rabbit polyclonal anti-KDM4C (raised against AA1007-1056 of human JMJD2C; generated in-house, R. Schüle laboratory), mouse monoclonal anti-β-actin (A5316; Sigma Aldrich, Taufkirchen, Germany), mouse monoclonal anti-α-tubulin (MCA78A; Serotec, Puchheim, Germany). Filters were incubated with horseradish peroxidase-conjugated secondary antibodies. Bands were visualized with the enhanced chemiluminescence system (Amersham, Freiburg, Germany).

Immunohistochemisty

FFPE slides from L1236 and HEK293 cells, transfected with pcDNA3 as a control or with pcDNA3-KDM4C, were subjected to immunohistochemical staining using homemade anti-KDM4C antibody (R. Schüle laboratory; see immunoblotting section). The FFPE slides were deparaffinized using roti-histol and hydrated in a degree series of ethanol (absolute, 80% and 70%). Thereafter the antigen retrieval was performed using a sodium citrate solution at pressure cooker 10 minutes. FFPE slides were rinsed in TBS and endogenous peroxidase blocked using Triton X-100. Thereafter, the FFPE slides were incubated with anti-KDM4C antibody (1:100) at 4°C overnight and incubated with mouse and rabbit -HRP conjugated detection IHC kit (Abcam) at room temperature following the manufacturer's instructions. To investigate the possible quantification of KDM4C protein expression several variations of IHC were tested without success.

Protein Modelling

Mechismo (http://mechismo.russelllab.org/)[16] was used to predict the potential effect of SVs and SNVs detected by whole genome sequencing.

Supplementary Tables

Supplementary Table S1. Histone methylation genes list analyzed (provided as excel format)

The table includes the histone methylation genes analyzed in the study, showing the function (histone demethylase or methyltransferase), the gene name, the Uniprot ID and in which pathway the protein is involved.

Supplementary Table S2. Genomic aberrations affecting *KDM4C* (provided as excel format).

The table displays the focal structural variants (deletion, duplication, translocation) and single nucleotide variants (SNVs) within the *KDM4C* locus detected in the ICGC cohort and the cell lines analyzed in this study. Moreover, an overview of validated genomic somatic SVs and SNVs in analyzed cohort and alternative KDM4C transcript are displayed in the table. The diagnosis and the cell of origin of the ICGC cases and cell lines explored herein are included in the table. *KDM4C* genomic status reported in COSMIC cell line database (https://cancer.sanger.ac.uk/cell_lines) and the copy number status of *KDM4C* gene in previously published studies in classical Hodgkin (cHL) cell lines using SNP 6.0[7] are highlighted by asterisk (*) and hashtag (#), respectively. *KDM4C* deletions described in L1236 (labelled with dolar ($) were retrieved from whole genome sequencing data previously published[9]. The *KDM4C* validations using Fluorescent in situ hybridization (FISH), and PCR and Sanger sequencing are depicted in the table. *KDM4C* assay comprises BAC clone RP11-940A10 spanning *KDM4C* labelled in spectrum green that was combined with a control probe for *NOTCH1* gene locus (9q34.3) (including BACs RP11-251M1 and RP11-83N9 both labelled in spectrum orange). The *KDM4C* break-apart assay contains the two differentially labelled BAC clones RP11-356C23 (spectrum orange) and RP11-996O17 (spectrum green) flanking the *KDM4C* locus. Note, the exome sequencing data previously published[8] has not reported the homozygous deletion in L1236 nor the truncating mutation in SUP-HD1.

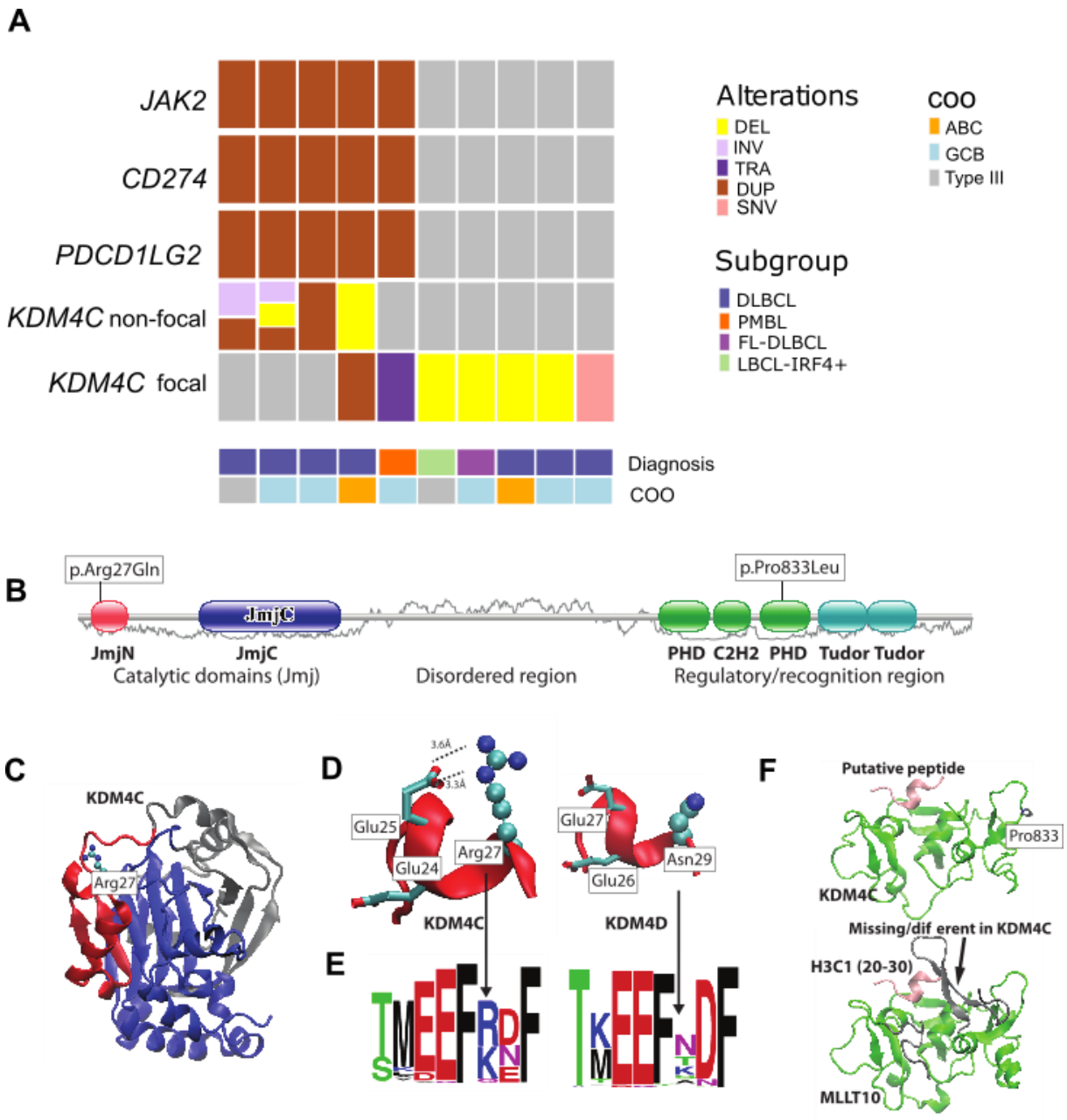
Supplementary Table S3. Targets of KDM4C previously described (provided as excel format).

Supplementary Table S4. Differential expressed genes between *KDM4C* altered cases and *KDM4C* wildtype in FL-DLBCL subgroup (provided as excel format).

Supplementary Figures

Extension of the Legend Figure 2A. Modelling protein of KDM4C aberrations

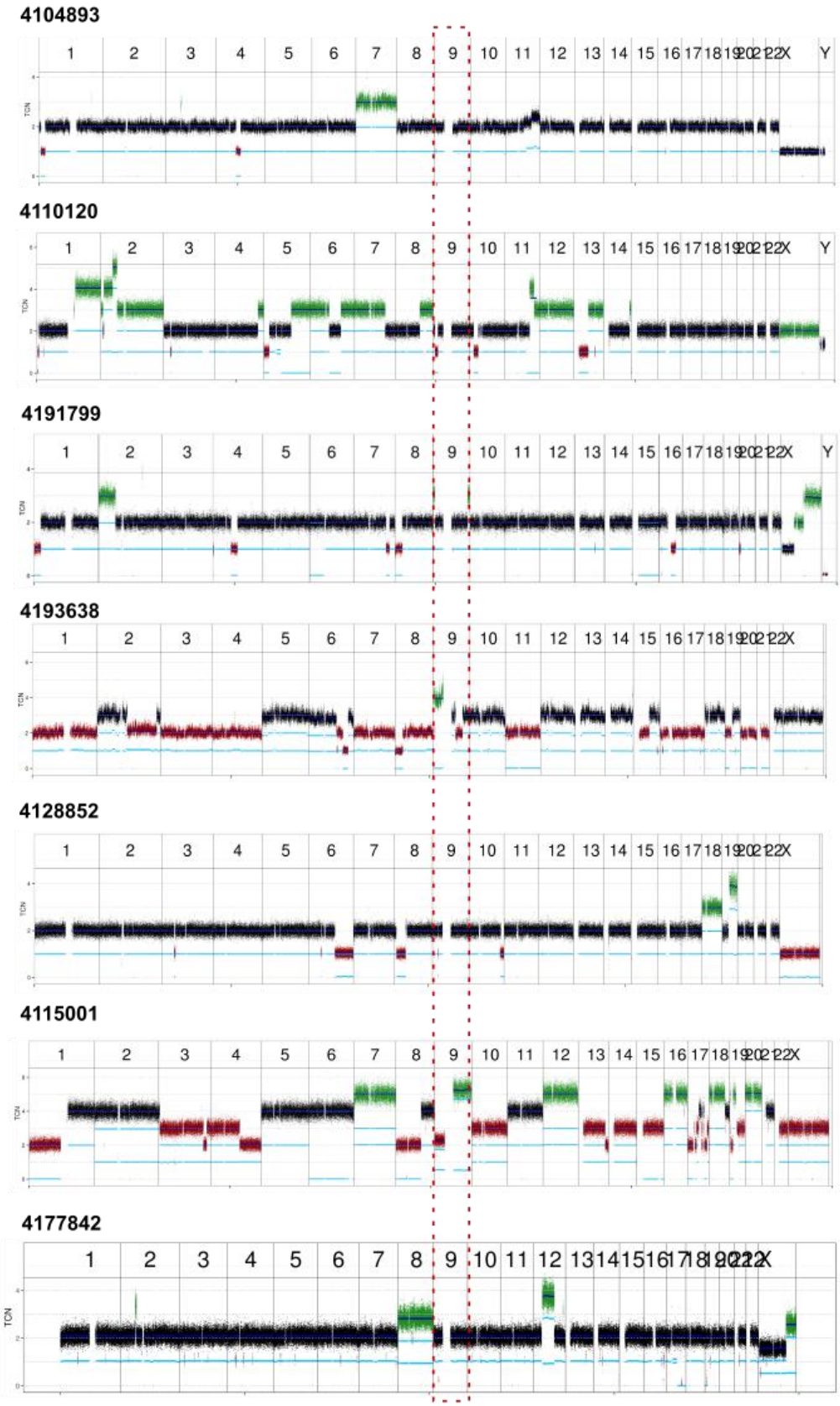
The absence of a catalytic domain in cases 4128852, 41074893, 4115001 can lead to a protein that competes for binding of ligands without catalysis, and argues that this essentially leads to a loss of activity by jamming function of remaining wt-KDM4C proteins. The absence of the recognition domains in case 4191799 will lead to an enzyme that won't necessarily reach its right targets, and could end up acting elsewhere, though it would not hit its intended target as often as the full-length protein. KDM4C in case 4110120 seems unlikely to be functional, due to loss of one catalytic domain and the recognition domains. In addition, case 4177842 harbours a nonsynonymous SNV (c,G80A, p,R27Q) affecting also the JmJH catalytic domain. The predicting impact of this SNV is a pathogenic mutation with consequence at protein level. All predictions were performed using mechismo[16].



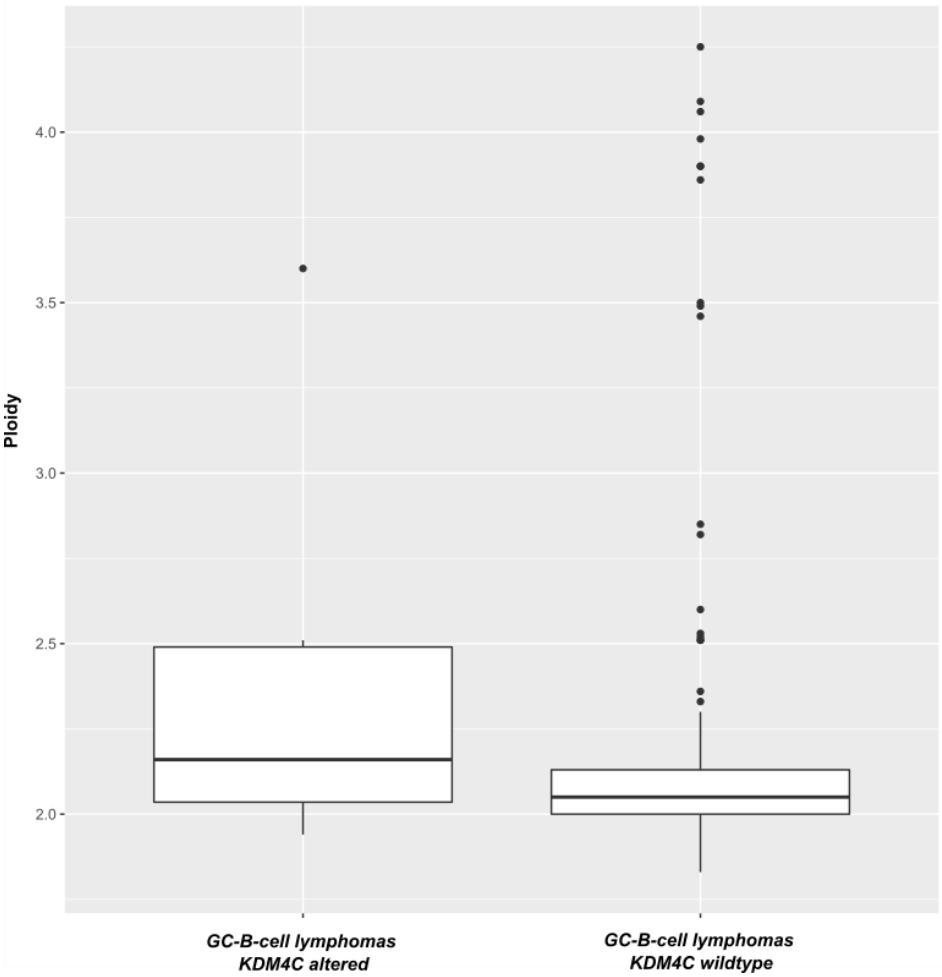
Supplementary Figure S1. Copy number status of genomic cluster in 9p24.1 (*JAK2*, *CD274* and *PDCD1LG2* genes) versus alterations affecting *KDM4C and* KDM4C domain structure and analysis of Arg27 (comparison with the equivalent position in KDM4D) and Pro833.

(A) Oncoprint displaying the cases with copy number gains (focal and non-focal, in contrast to Fig. 1 which only included focal aberrations) involving the *JAK2*, *CD274* and/or *PCD1LG2* genes in GC-derived B-cell lymphomas compared to the genomic status of *KDM4C* in each case. The analyzed genes are ordered from telomere to centromere of chromosome 9. The cases are classified based on diagnosis, and also annotated according to the cell of origin (COO) classification. (B) Domain structure of KDM4C with domains labelled (after Uniprot/Pfam) and showing the position of p.Arg27Gln detected in KDM4C in 4177842 and p.Pro833Leu detected in KARPAS-422. The graph

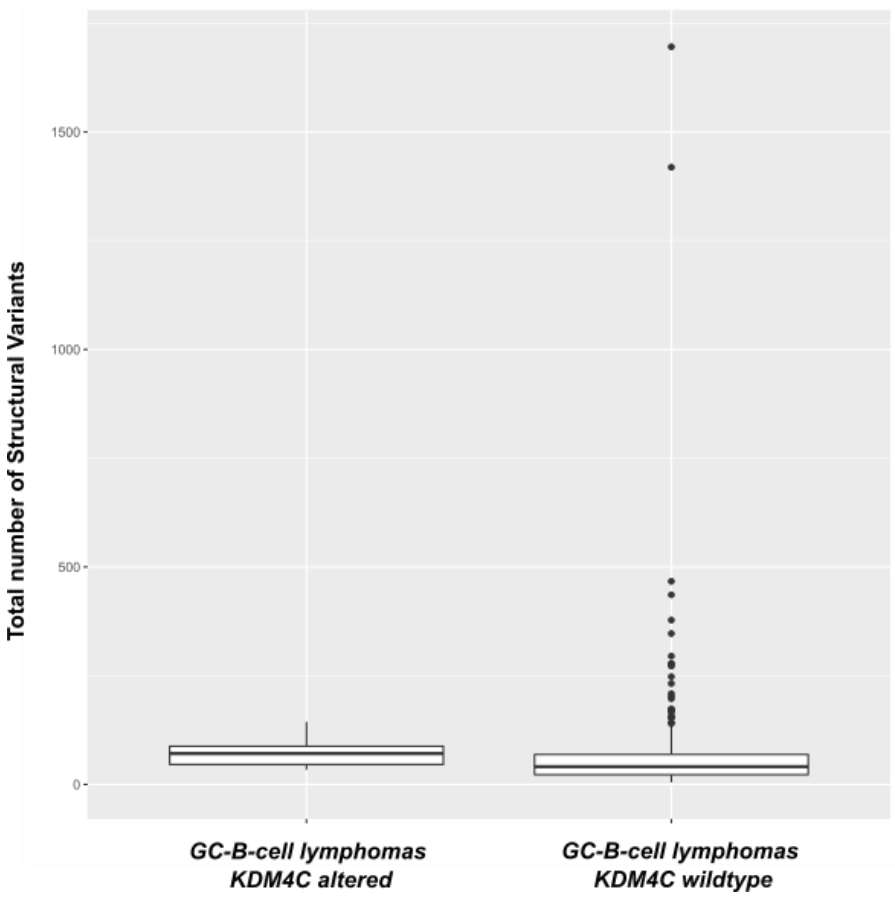
behind the domain is a prediction of disordered sequence from IUPred[17]. (C) KDM4C structure (PDB code 4xdo) showing the JmJN (red) and JmjC (blue) domains together with grey colour denoting the intervening sequence that is also part of the catalytic core. (D) Zoom on the region of Arg27 (spheres) in KDM4C showing Glu24 and Glu25 (sticks), with the putative salt-bridge between Glu25 and Arg27 shown as dashed lines (distances in Angstroms). The equivalent region in KDM4D (PDB code 3dxt) is also shown for comparison. While there are no known interactions here[16] at Arg27, it is adjacent to a pair of Glu residues E24 & E25 which are conserved in the subfamily member KDM4D, where they have been shown to be PARsylated, a modification negatively regulating its activity[18]. Curiously, KDM4D lacks an Arg in the position equivalent to R27 and has an Asn instead, an amino acid similar to Gln. In the KDM4C structure, R27 is hydrogen bonded with one of these Glu residues (PDB:4xdo). It is thus possible that p.R27Q could render KDM4C more like KDM4D in this region, preventing this hydrogen bond and leading to a similar PARsylation and resulting loss-of-function. (E) Sequence motive logos for KDM4C (left) and KDM4D (right) around the positions shown in c. The height of each amino acid symbol denotes its prevalence across orthologs at each position. (F) Model of KDM4C PHD finger domain showing the location of Pro833 (top) in KDM4C detected in the KARPAS-422 cell line and the template structure (MLLT10 bound to H3C1 peptide; PDB:5dah; bottom). The PHD finger is shown in green and the peptides in pink. The grey region in the bottom structure shows a C-terminal, binding-site region that is absent/different in KDM4C and which is contacted by the region around Pro-833. Predictors of variant impact (PolyPhen2, MutationAssessor, SIFT) give conflicting views (from low-impact to damaging) likely because different sequence alignments are used. Position 833 shows a conservation of a small amino acid sequence (Pro, Ser, Ala) across orthologs of KDM4C (EggNOG database), but shows poorer conservation across wider homologs/paralogs (UniProt), where generally a polar amino acid is preferred. This argues for a specific functional role in KDM4C proteins that is different from wider members of this domain family. Inspection of the structure (modelled on PDB:5dah using MODELER; Supplemental Fig. 2) suggests that Pro-833 is involved in orienting the region C-terminal of the PHD finger to form the peptide binding site. Taken together, evolutionary and protein structure predictions also suggest that p.P833L is a loss-of-function variant.



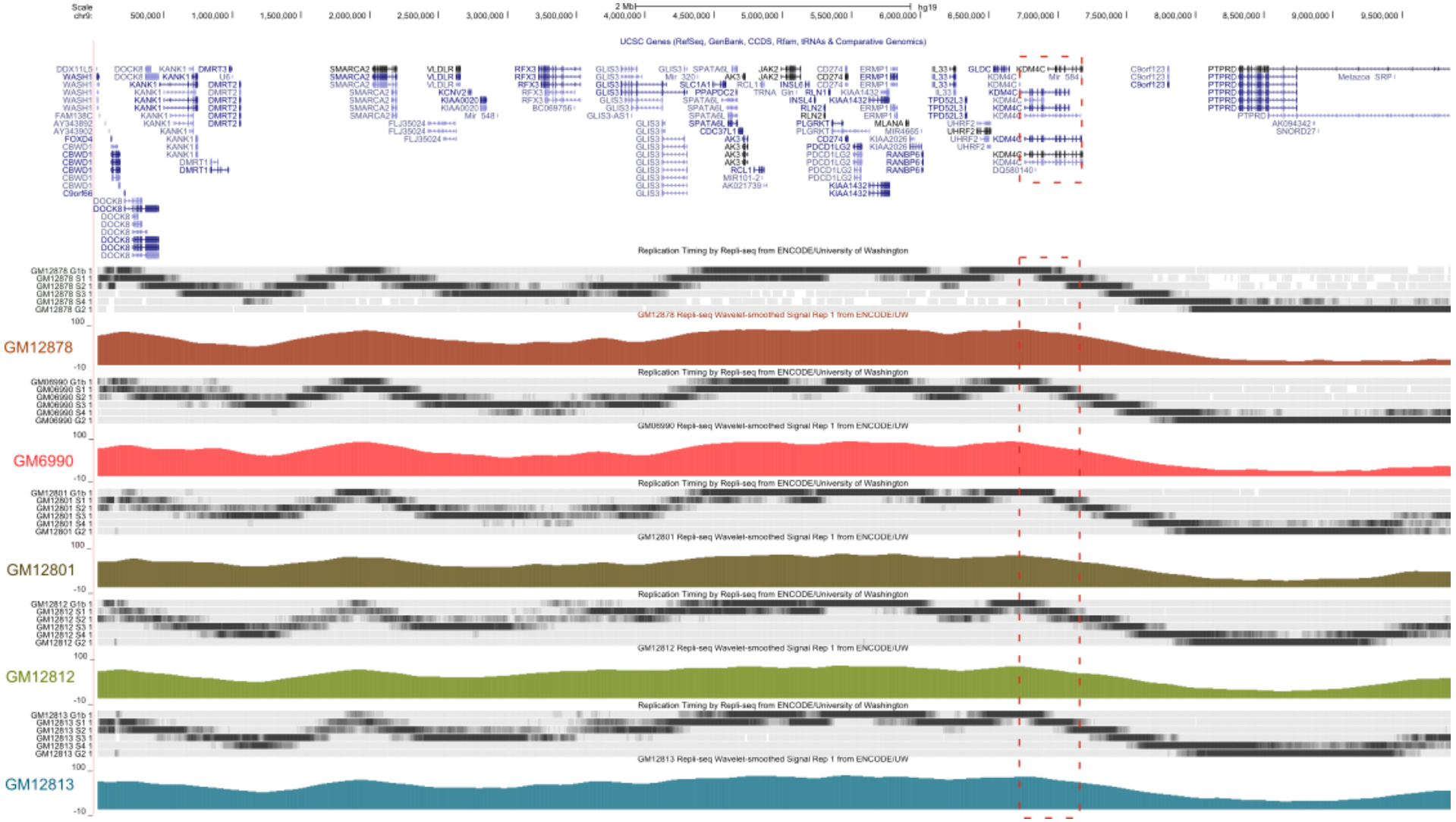
Supplementary Figure S2. Imbalance profiles of germinal center derived B cell lymphomas harboring *KDM4C* aberrations. Losses are indicated in red and gains in green. Chromosome 9, where *KDM4C* locus is located, is highlighted by a red rectangle.



Supplementary Figure S3. Box-plot of overall ploidy in GC-B-cell lymphomas cases with and without *KDM4C* alterations.

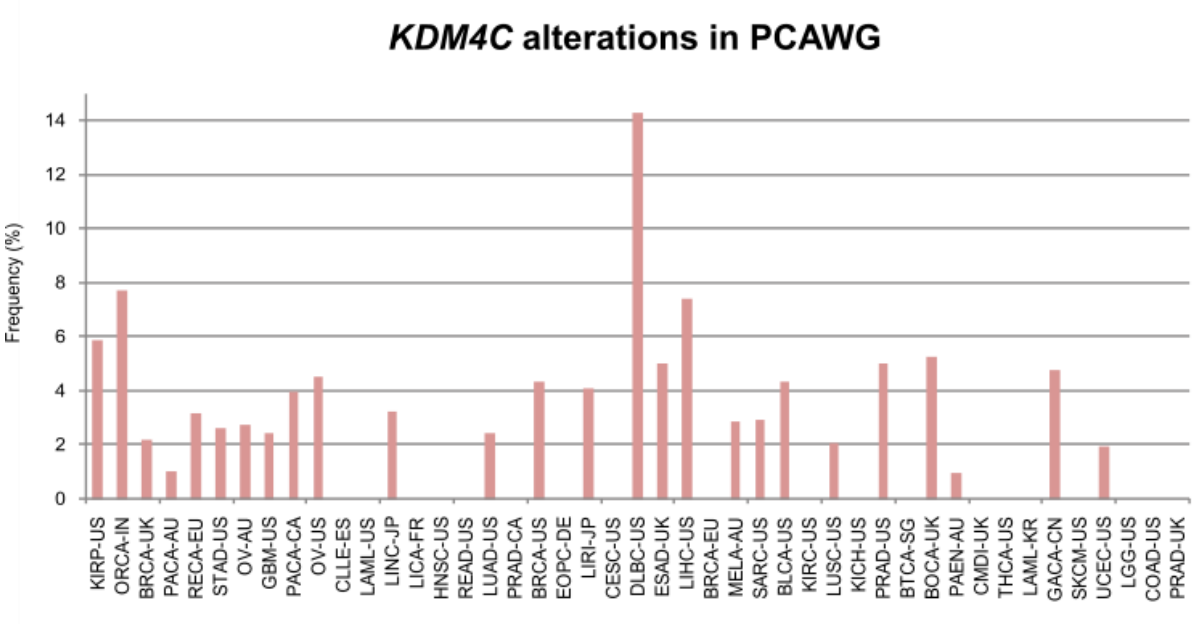


Supplementary Figure S4. Box-plot of the total number of structural variants detected by WGS in lymphomas with and without *KDM4C* alterations.

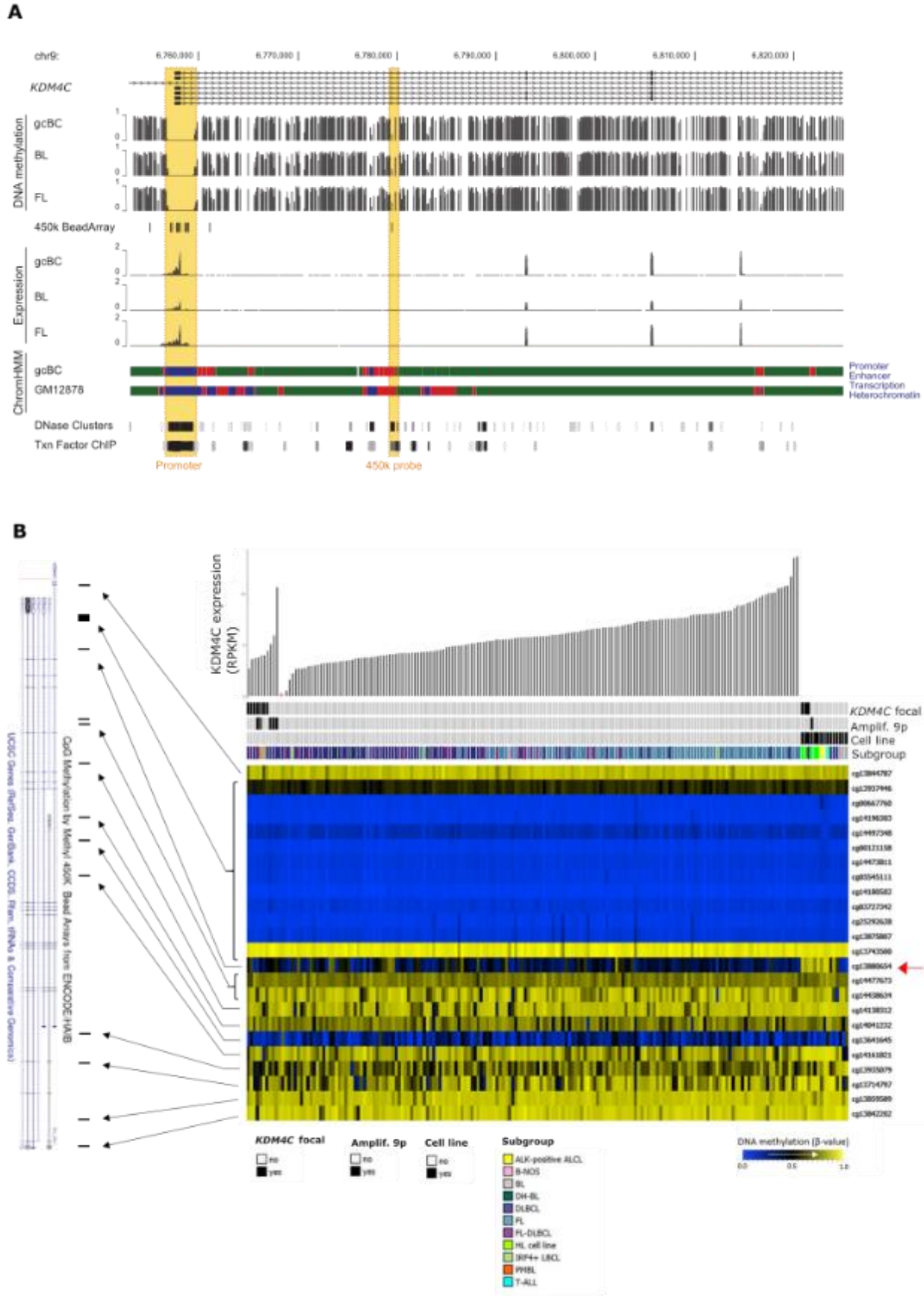


Supplementary Figure S5. Replication timing of the 9p region.

The figure shows the replication timing by Repli-seq from the ENCODE track in the Genome browser (https://genome.ucsc.edu, accessed 10/12/2021) of different lymphoblastoid cell lines[19]. Each line contains the replication timing for the six cell cycle fractions: G1/G1b, S1, S2, S3, S4, and G2. Replication patterns are visualized as a continuous function based on sequencing tag density (*Percentage-normalized Signal*) and as a wavelet-smoothed transform of the six-fraction profile (*Wavelet-smoothed Signal*)[19]. The *KDM4C* locus is highlighted by red rectangle.

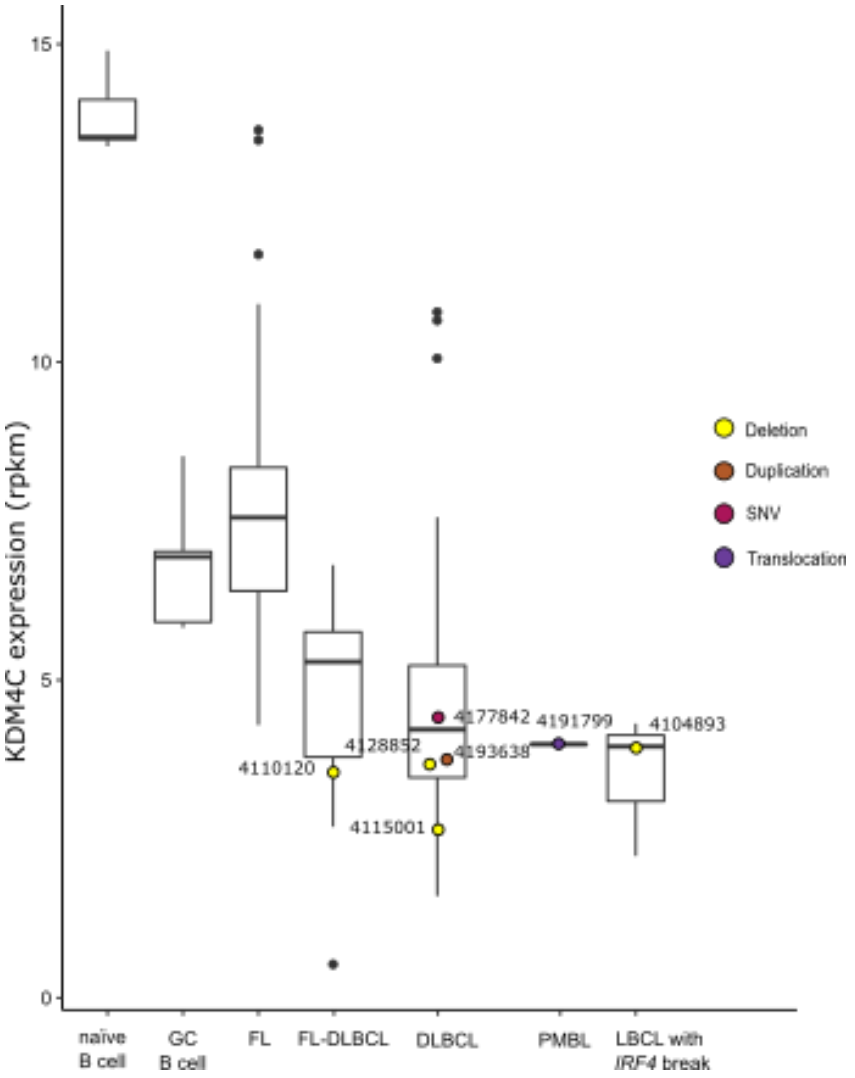


Supplementary Figure S6. Graphic displaying the frequency of structural variants on *KMD4C* in 45 tumor cohorts included PCAWG dataset[20].


A
chr9:
6,760,000
6,770,000
6,780,000
6,790,000
6,800,000
6,810,000
6,820,000
KDM4C
DNA methylation
gcBC
BL
FL
450k BeadArray
Expression
gcBC
BL
FL
ChromHMM
gcBC
GM12878
Promoter
Enhancer
Transcription
Heterochromatin
DNase Clusters
Txn Factor ChIP
Promoter
450k probe
B
KDM4C expression (RPKM)
KDM4C focal
Amplif. 9p
Cell line
Subgroup
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
CpG Methylation by Methyl 450K Bead Arrays from ENCODE/HAIB
KDM4C focal
no
yes
Amplif. 9p
no
yes
Cell line
no
yes
Subgroup
ALK-positive ALCL
B-NOS
BL
DH-BL
DLBCL
FL
FL-DLBCL
HL cell line
IRF4+ LBCL
PMBL
T-ALL
DNA methylation (β-value)
0.0
0.5
1.0

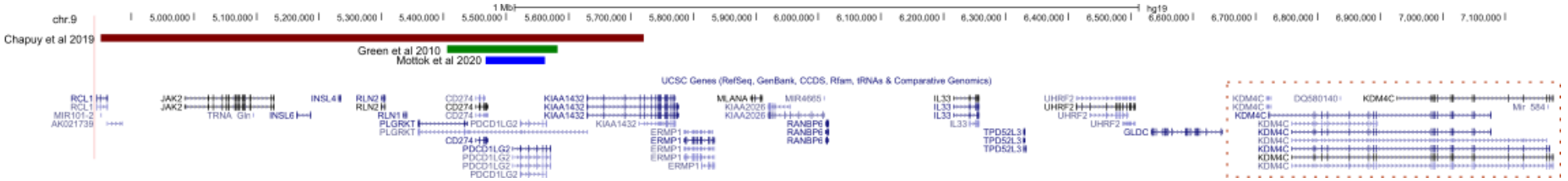
Supplementary Figure S7. Analysis of DNA methylation and chromatin states at the *KDM4C* locus in the ICGC MMML-Seq cohort by mining data from Kretzmer et al., 2015[2].

(A) UCSC genome browser visualisation of the *KDM4C* gene locus showing average DNA methylation and gene expression in germinal center B cells (gcBC), BL and FL as well as chromatin states. The promoter region (left highlight) is enriched in DNase clusters, transcription factor binding sites and is depleted in methylation independent of KDM4C expression levels. The right highlight indicates a differentially methylated 450k Bead Chip methylation array probe cg13880654 hypermethylated in various lymphoma cell lines (see panel b), which overlaps with DNase clusters and transcription factor binding sites and is annotated as an enhancer by ChromHMM. (B) Heatmap displaying the DNA methylation of the CpGs associated with the *KDM4C* gene based on Illumina 450K BeadChip analysis of the ICGC MMML-seq cohort. Twenty-four CpGs within the gene *KDM4C* were identified and displayed for 176 lymphoma samples (10 cases were excluded, not passing the quality control for the analysis) and 15 cell lines. The lymphoma samples and cell lines, respectively, are ordered according to *KDM4C* focal aberration, and/or amplification in 9p24.1 (including *JAK2*, *CD274* and *PDCD1LG2)* status, followed by *KDM4C* gene expression level (is a bar plot at the top of figure S7B), RPKM values, two cases without RNAseq data highlighted with the red asterisk). The 24 CpGs are ordered according to genomic location and mapped to the *KDM4C* locus (left). The CpG 13880654 hypermethylated in lymphoma cell lines except Burkitt lymphoma cell lines is highlighted by an arrow.



Supplementary Figure S8. KDM4C expression in GC-derived B-cell lymphomas.

Mean expression of KDM4C in normal naïve B cell and germinal center derived B cells (GCB cells) from 5 healthy controls and GC-derived B-cell lymphoma, including follicular lymphoma (FL), diffuse large B cell lymphoma (DLBCL), FL-DLBCL, primary mediastinal diffuse large B cell lymphoma (PMBCL), large B-cell lymphoma (LBCL) with *IRF4*-rearrangement. The GC-B-cell lymphomas cases harboring aberrations in *KDM4C* are labeled with the corresponding ICGC MMML-Seq ID and the dot color assigned indicates the type of genomic *KDM4C* aberration (deletion in yellow, duplication in brown, single nucleotide variant (SNV) in dark pink, and translocation in purple).



Supplementary Figure S9. Minimal region of copy number gain on 9p24.1 based on high-resolution studies of PMBL and HL.

Chapuy et al 2019[21] (red color) and Mottok et al., 2019[22] (green color) inferred the copy number aberrations from whole exome sequencing data and used GISTIC algorithm to defined critical region of gain in PMBL cohorts. Green et al 2010[23] analyzed a cohort of HL, DLBCL and PMBL determining the copy number by Affymetrix SNP6.0 microarrays. *KDM4C* locus is highlighted with the red dotted rectangle at the far right. Screenshot based on the USCS browser (https://genome.ucsc.edu, accessed on 09.12.2021).

Remaining members of the ICGC MMML-Seq Consortium

Coordination (C1): Susanne Wagner[2], Gesine Richter[2]
Data Center (C2): Jürgen Eils[20], Jules Kerssemakers[20], Christina Jaeger-Schmidt[20], Ingrid Scholz[20]
Clinical Centers (WP1): Christoph Borst[21], Friederike Braulke[22], Birgit Burkhardt[23], Alexander Claviez[24], Martin Dreyling[25], Sonja Eberth[25], Hermann Einsele[26], Norbert Frickhofen[27], Siegfried Haas[21], Dennis Karsch[28], Nicole Klepl[22], Michael Kneba[28], Jasmin Lisfeld[23], Luisa Mantovani-Löffler[29], German Ott[30], Marius Rohde[31], Christina Stadler[22], Peter Staib[32], Stephan Stilgenbauer[33], Lorenz Trümper[22], Thorsten Zenz[34],
Normal Cells (WPN): Martin-Leo Hansmann[35], Dieter Kube[22], Ralf Küppers[36], Marc A. Weniger[36]
Pathology and Analyte Preparation (WP2-3): Siegfried Haas[37], Michael Hummel[38], Wolfram Klapper[39], Ulrike Kostezka[40], Dido Lenze[38], Peter Möller[41], Andreas Rosenwald[42], Monika Szczepanowski[39],
Sequencing and genomics (WP4-7): Sietse Aukema[2], Vera Binder[43], Arndt Borkhardt[43], Andrea Haake[2], Jessica Hoell[43], Ellen Leich[42], Inga Nagel[2], Jordan Pischimariov[42], Stefan Schreiber[44], Inga Vater[2]
Bioinformatics (WP8-9): Hans Binder[6,7], Benedikt Brors[45], Gero Doose[6,7,8], Jürgen Eils[20], Roland Eils[12,20], Lydia Hopp[6], Helene Kretzmer[6,7,8], David Langenberger[6,7,8], Markus Loeffler[14], Maciej Rosolowski[14], Peter F. Stadler[7].

[20] Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, 69120, Germany;
[21]Department of Internal Medicine/Hematology, Friedrich-Ebert-Hospital, Neumünster;
[22]Department of Hematology and Oncology, Georg-August-University of Göttingen, 37075 Göttingen, Germany;
[23]University Hospital Münster-Pediatric Hematology and Oncology, 48149, Münster, Germany;
[24]Department of Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, 24105, Kiel, Germany;
[25]Department of Medicine III - Campus Grosshadern, University Hospital Munich, Munich, Germany;
[26]University Hospital Würzburg, Department of Medicine and Poliklinik II, University of Würzburg, Würzburg;
[27]Department of Medicine III, Hematology and Oncology, Dr. Horst-Schmidt-Kliniken of Wiesbaden, Wiesbaden;
[28]Department of Internal Medicine II: Hematology and Oncology, University Medical Centre, Campus Kiel, Kiel;
[29]Hospital of Internal Medicine II, Hematology and Oncology, St-Georg Hospital Leipzig, Leipzig, Germany;
[30]Department of Pathology, Robert-Bosch-Hospital, Stuttgart, Germany;
[31]Pediatric Hematology and Oncology, University Hospital Giessen, 35392, Giessen, Germany;
[32]Clinic for Hematology and Oncology, St.-Antonius-Hospital, Eschweiler;
[33]Department for Internal Medicine III, Ulm University, 89081, Ulm, Germany;
[34]National Centre for Tumor Disease, Heidelberg, Germany;
[35]Senckenerg Institute of Pathology, University of Frankfurt Medical School, Frankfurt am Main, 60590, Germany;
[36]Institute of Cell Biology (Cancer Research), Medical School, University of Duisburg-Essen, 45147, Essen, Germany;
[37]Friedrich-Ebert Hospital Neumünster, Clinics for Hematology, Oncology and Nephrology, Neumünster, Germany;
[38]Institute of Pathology, Charité-University Medicine Berlin, 10117, Germany;
[39]Hematopathology Section, Christian-Albrechts-University, 24105, Kiel, Germany;
[40]Comprehensive Cancer Center Ulm (CCCU), University Hospital Ulm, 89081, Ulm, Germany;
[41]Institute of Pathology, University of Ulm and University Hospital of Ulm, 89081, Ulm, Germany;
[42]Institute of Pathology, Comprehensive Cancer Center Mainfranken, University of Würzburg, 97080, Würzburg, Germany;

[43]Medical Faculty, Department of Pediatric Oncology, Hematology and Clinical Immunology, Heinrich-Heine-University, 40225, Düsseldorf, Germany;
[44]Department of General Internal Medicine, University Kiel, Kiel, Germany;
[45]Division of Applied Bioinformatics (G200), German Cancer Research Center (DKZF), 69120, Heidelberg, Germany;

References

1. Richter J, Schlesner M, Hoffmann S, et al. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat Genet*. 2012;44(12):1316-1320.

2. Kretzmer H, Bernhart SH, Wang W, et al. DNA methylome analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. *Nat Genet*. 2015;47(11):1316-1325.

3. López C, Kleinheinz K, Aukema SM, et al. Genomic and transcriptomic changes complement each other in the pathogenesis of sporadic Burkitt lymphoma. *Nat Commun*. 2019;10(1):1459.

4. Doose G, Haake A, Bernhart SH, et al. MINCR is a MYC-induced lncRNA able to modulate MYC's transcriptional network in Burkitt lymphoma cells. *Proc Natl Acad Sci U S A*. 2015;112(38):E5261-70.

5. Hübschmann D, Kleinheinz K, Wagener R, et al. Mutational mechanisms shaping the coding and noncoding genome of germinal center derived B-cell lymphomas. *Leukemia*. 2021;35(7):2002-2016.

6. Swerdlow, S.H., Campo, E., Harris, N.L., Jaffe, E.S., Pileri, S.A., Stein, H., Thiele, J., Vardiman J. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. IARC; 2008.

7. Otto C, Giefing M, Massow A, et al. Genetic lesions of the TRAF3 and MAP3K14 genes in classical Hodgkin lymphoma. *Br J Haematol*. 2012;157(6):702-708.

8. Liu Y, Abdul Razak FR, Terpstra M, et al. The mutational landscape of Hodgkin lymphoma cell lines determined by whole-exome sequencing. *Leukemia*. 2014;28(11):2248-2251.

9. Schneider M, Schneider S, Zühlke-Jenisch R, et al. Alterations of the CD58 gene in classical Hodgkin lymphoma. *Genes Chromosomes Cancer*. 2015;54(10):638-645.

10. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.

11. Hoffmann S, Otto C, Kurtz S, et al. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput Biol*. 2009;5(9):e1000502.

12. Hummel M, Bentink S, Berger H, et al. A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. *N Engl J Med*. 2006;354(23):2419-2430.

13. Bornkamm GW, Berens C, Kuklik-Roos C, et al. Stringent doxycycline-dependent control of gene activities using an episomal one-vector system. *Nucleic Acids Res*. 2005;33(16):1-11.

14. Schleussner N, Merkel O, Costanza M, et al. The AP-1-BATF and -BATF3 module is essential for growth, survival and TH17/ILC3 skewing of anaplastic large cell lymphoma. *Leukemia*. 2018;32(9):1994-2007.

15. Lamprecht B, Walter K, Kreher S, et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat Med*. 2010;16(5):571-579.

16. Betts MJ, Lu Q, Jiang Y, et al. Mechismo: predicting the mechanistic impact of mutations and modifications on molecular interactions. *Nucleic Acids Res*. 2015;43(2):e10.

17. Mészáros B, Erdös G, Dosztányi Z. IUPred2A: Context-dependent prediction of protein

disorder as a function of redox state and protein binding. *Nucleic Acids Res*. 2018;46(W1):W329-W337.

18. Le May N, Iltis I, Amé J-C, et al. Poly (ADP-ribose) glycohydrolase regulates retinoic acid receptor-mediated gene expression. *Mol Cell*. 2012;48(5):785-798.

19. Hansen RS, Thomas S, Sandstrom R, et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci U S A*. 2010;107(1):139-144.

20. Campbell PJ, Getz G, Korbel JO, et al. Pan-cancer analysis of whole genomes. *Nature*. 2020;578(7793):82-93.

21. Chapuy B, Stewart C, Dunford AJ, et al. Genomic analyses of PMBL reveal new drivers and mechanisms of sensitivity to PD-1 blockade. *Blood*. 2019;134(26):2369-2382.

22. Mottok A, Hung SS, Chavez EA, et al. Integrative genomic analysis identifies key pathogenic mechanisms in primary mediastinal large B-cell lymphoma. *Blood*. 2019;134(10):802-813.

23. Green MR, Monti S, Rodig SJ, 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(17):3268-3277.