

SUPPLEMENTAL DATA

Genetic mechanisms of HLA-I loss and immune escape in Diffuse Large B-Cell Lymphoma

Marco Fangazio^{1†#}, Erik Ladewig^{2#}, Karen Gomez^{2#}, Laura Garcia-Ibanez¹, Rahul Kumar¹, Julie Teruya-Feldstein^{3¶}, Davide Rossi^{4,5,6}, Ioan Filip², Qiang Pan-Hammarström⁷, Giorgio Inghirami⁸, Renzo Boldorini⁹, German Ott¹⁰, Annette M. Staiger^{10,11}, Björn Chapuy¹², Gianluca Gaidano¹³, Govind Bhagat^{14,15}, Katia Basso^{1,14,15}, Raul Rabadan^{2,15,16*}, Laura Pasqualucci^{1,14,15*} and Riccardo Dalla-Favera^{1,14,15,17,18*}

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

² Departments of Systems Biology and Biomedical Informatics, Columbia University, New York, NY, USA

³ Division of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY

⁴ Division of Hematology, Institute of Oncology Research, Bellinzona, Switzerland

⁵ Laboratory of Experimental Hematology, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland

⁶ Faculty of Biomedical Science, Università della Svizzera Italiana, Lugano, Switzerland

⁷ Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden

⁸ Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, NY

⁹ Department of Health Sciences, Division of Pathology, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy

¹⁰ Department of Clinical Pathology, Robert Bosch Krankenhaus, Stuttgart, Germany

¹¹ Institute of Clinical Pharmacology, Stuttgart and University of Tübingen, Tübingen, Germany

¹² Department of Hematology and Oncology, University of Göttingen, Göttingen, Germany

¹³ Department of Translational Medicine, Division of Hematology, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy

¹⁴ Department of Pathology & Cell Biology Columbia University, New York, NY

¹⁵ Herbert Irving Comprehensive Cancer Center, New York, NY

¹⁶ Program for Mathematical Genomics, Columbia University, New York, NY, USA

¹⁷ Department of Microbiology and Immunology, Columbia University, New York, NY

¹⁸ Department of Genetics and Development, Columbia University, New York, NY

[†] Present address: Department of Clinical Biology, Laboratory of Hematology, LHUB-ULB, Université Libre de Bruxelles, Brussels, Belgium

[¶] Present address: Department of Pathology, Molecular & Cell-Based Medicine, Tisch Cancer Institute, Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Equal contribution

* Equal contribution

Correspondence: rd10@cumc.columbia.edu; lp171@cumc.columbia.edu

SUPPLEMENTAL MATERIALS AND METHODS

Cell lines

The DLBCL cell lines OCI-LY3, RC-K8, and WSU-NHL, as well as the lymphoblastoid cell line T2 (ATCC® CRL-1992™)(1) were maintained in Iscove's Modified Dulbecco Medium (IMDM) supplemented with 10% fetal bovine serum (FBS), 100 µg/mL penicillin, 100 µg/mL streptomycin and 2 mM L-glutamine. The OCI-Ly10 cell line was cultured in IMDM with 20% heparinized human plasma, 100 µg/mL penicillin, 100 µg/mL streptomycin, 2 mM L-glutamine, and 55 µM β-mercaptoethanol. We cultured the Phoenix packaging cell line in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS, 100 µg/mL penicillin, 100 µg/mL streptomycin and 2 mM L-glutamine. All cells were maintained in a humidified, 5% CO₂ incubator at 37° C.

Immunohistochemistry analysis of human tissues

FFPE tissue sections from the 657 cases were immunostained for B2M, HLA-I, and HLA-DRB1 protein expression by using standard protocols (2). Briefly, slides were deparaffinized in xylene for 15 min and rehydrated in decreasing concentrations of ethanol before antigen retrieval, which was performed in sodium citrate buffer (pH 6.0) in a microwave oven. Endogenous peroxidase activity was quenched with a solution of phosphate-buffered saline (PBS) containing 3% of hydrogen-peroxide. Sections were then blocked in PBS supplemented with 5% bovine serum albumin and 0.1% tween, followed by overnight incubation at 4° C with antibodies against B2M (1:2000, Dako cat# A0072), HLA-I (HC-10, 1:100, Nordic MUBio, cat# MUB2037P)(3), or HLA-DRB1 (LN-3, 1:100, Abcam, cat# ab233877) in a humidified chamber. Sections were washed three times in PBS/0.1% tween and incubated 1 hr at room temperature with secondary antibodies

conjugated to horseradish peroxidase (HRP) (EnVision, Dako), followed by visualization with 3-amino-9-ethylcarbazole (Sigma).

DNA extraction

High molecular weight genomic DNA was extracted from 74 previously untreated *de novo* DLBCL samples, including 44 fresh frozen biopsies and 30 FFPE biopsies (Dataset S1). Matched normal DNA was obtained for all patients from peripheral blood granulocytes, saliva, or bone marrow using standard phenol chloroform or salting out extraction procedures. The fraction of tumor cells in the tissue section used for molecular studies was estimated to be >55% based on *in-silico* inference by the Sequenza software package (4). DNA was quantitated by the Quant-iT PicoGreen reagent (Invitrogen) and verified for integrity by gel electrophoresis. A prior or concomitant lymphoproliferative disorder was ruled out in all cases by reviewing clinical information and, where available, lymph node and bone marrow biopsies. Thirty-nine of these samples have been used in previously published studies (5-7).

Whole exome and whole genome sequencing

Purified genomic DNA from 64 DLBCL samples and 62 matched normal tissues (~1 µg) was enriched in protein-coding sequences by using the Agilent SureSelect Human 51Mb All Exon V5 kit (Agilent Technologies, Santa Clara, CA), according to the manufacturer's protocol. The resulting target-enriched pool was normalized and combined (4-plex) before high-throughput paired-end (2 × 100 bp) sequencing on the Illumina HiSeq2000 System at Centrillion Biosciences, Inc or at the Beijing Genomics Institute (BGI). For the remaining 10 DLBCL samples (12 normal samples), genomic DNA (~3 µg) was used to build whole genome libraries that were high-

throughput paired-end (2×100 bp) sequenced on the BGISEQ-500 System at the Beijing Genomics Institute (BGI). The WES analysis produced on average 98.6 million passed-filter paired-end reads per tumor sample (range, 52.5 to 142.8) and 76.6 million passed-filter paired-end reads per paired normal sample (range, 44.0 to 131.7). The WGS analysis produced on average 1153.5 million passed-filter paired-end reads per tumor sample (range, 269.3 to 1442.6) and 1137.8 million passed-filter paired-end reads per paired normal sample (range, 602.3 to 1374.9). After filtering for duplicate reads (defined as reads with identical start and orientation), sequences were aligned to the reference human genome hg19 assembly (GRCh37) with the Burrows-Wheeler Aligner tool (version 0.5.9) (6, 7). The mean coverage depth for the WES experiment (i.e., the mean number of reads covering the target exome of a haploid reference) was $80.5\times$ (range, 38.0 to 133.0) for the tumor samples and $62.2\times$ (range, 34.0 to 116.0) for the normal samples, with on average 75.9% and 72.9% of the target exome covered at $>30\times$, respectively (range, 57.5 to 97.5 and 53.8 to 96.3). For the WGS experiment, the mean coverage depth at coding sequences was $40.1\times$ in the DLBCL samples (range, 12.0 to 47.0, with a mean of 83.3% covered at $>30\times$) and $37.3\times$ in the normal samples (range, 16.0 to 49.0, with a mean of 74.9% covered at $>30\times$) ([Dataset S7](#) for details of the exome/genome sequencing performances).

Identification of somatic variants from WES and WGS data

Aligned bam files from all cases underwent PCR and optical duplicate marking, base quality score recalibration, and indel realignment using SAMtools, Picard version 2.0, and Genome Analysis Toolkit (GATK) version 3.0 (<https://gatk.broadinstitute.org/>). Coverage was assessed using BAMQC (8). Sequence variants, including nucleotide substitutions and small insertions or deletions, were obtained independently for each tumor and normal sample with the SAVI

algorithm and were considered of somatic origin if absent in the matched normal sample (9). Only variants present at >15% VAF were considered, to exclude minor subclonal events.

Targeted deep sequencing of the HLA-I regions

Purified tumor/normal genomic DNA from 26 of the 74 DLBCL cases was enriched in HLA-I coding sequences by using a custom capture RNA bait assay developed by Centrillion Biosciences, Inc., and subjected to high-throughput paired-end (2×100 bp) sequencing to an average depth of $1099\times$ (see also [Datasets S8](#) and [S9](#)).

HLA haplotyping and identification of SNVs from sequencing data

HLA genotyping and subsequent mutation calling was performed by applying the Polysolver computational tool to normal and tumor DNA sequencing data obtained by HLA deep sequencing and/or WES, using bam files as inputs, according to published methods (10, 11). Polysolver was also used to apply the standard tools MuTect and Strelka to detect somatic mutations by comparing re-aligned tumor and normal HLA reads (12). HLA alignments were manually inspected, and those loci containing mutations with a score of passing were reported. PCR amplification and Sanger sequencing were utilized to validate 15 representative variants identified by HLA deep sequencing (8 truncating and 7 missense mutations from 9 samples). With one exception, all variants were confirmed as present in the tumor dominant clone, documenting the high specificity of this approach ([Dataset S5](#)). As expected, the sensitivity of the WES approach was lower compared to the deep sequencing approach, with 9/16 Sanger-validated mutations detected by both methods; however, all mutations identified by WES were among those validated by Sanger, demonstrating its high specificity.

LOHHLA

The output of HLA typing (Polysolver) and copy number estimation (Sequenza) were used as input for the LOHHLA tool to identify haplotype-specific copy number (CN) changes of the HLA locus, as described (4, 10, 13). Statistically significant regions ($p < 0.05$) were scrutinized with output plots to help identify a confident set of human leukocyte antigens with loss of heterozygosity.

Copy number analysis

The method of Sequenza was applied to WES or WGS data in order to estimate purity and ploidy as published (4), and to detect the presence of CN aberrations affecting *B2M* and the following genes involved in antigen presentation/immune responses, starting from WES data: *PSMB9* (*LMP2*), *PSMB8* (*LMP7*), *TAP1*, *TAP2*, *NLRC5*, *CD58*, *CD48*, *SLAMF6* (*NTB-A*), *SH2D1A*, *PVR* (*CD155*), *NECTIN2* (*CD112*), *CD274* (*PD-L1*), and *PDCD1LG2* (*PD-L2*). Sequenza also served as an independent method for determining CN changes affecting the HLA loci (14). Mean CN values > 2.3 and < 1.7 were used as cutoffs to call gains and losses, respectively (15).

Sequencing analysis of ASHM target genes

Mutational analysis of 126 previously identified targets of ASHM (16-18) was performed by interrogating the genomic sequences (exons and introns) spanning the first 2kb after the transcription start site on WES/WGS data using SAVI. In order to maintain uniform criteria across WGS- and WES-sequenced samples, only regions covered by the Agilent SureSelect Human 51Mb All Exon V5 kit (Agilent Technologies, Santa Clara, CA) capture platform were considered,

including coding and noncoding exons as well as 500bp of flanking intronic regions that are also captured by current hybridization-based platforms. Mutations were computed in the analysis if somatic in origin, and included intronic variants, changes in untranslated regions, silent mutations, and non-silent mutations, using the following criteria: i) coverage depth ≥ 10 at the mutated position, ii) $\geq 15\%$ variant allele frequency in the tumor, and iii) absent in the paired normal DNA. Individual samples were scored as affected by ASHM when harboring at least 3 variants in one or more of the 126 genes. The following genes were also analyzed by PCR amplification and Sanger sequencing of the first 2kb from the transcription start site: *BCL2*, *BCL6*, *BCL7A*, *CIITA*, *DTX1*, *HIST1H1E*, *ITPKB*, *MYC*, *PAX5*, *PIM1*, *RhoH*, *SIP2*, *SOCS1*.

Assessment of germline homozygosity at HLA-I loci in TCGA and UK Biobank Cohorts

Germline homozygosity at the *HLA-I* loci was assessed in two separate cohorts, including i) 9,623 patients from 30 different types of cancer from the Cancer Genome Atlas (TCGA) project, (<https://www.cancer.gov/tcga>), using previously published HLA-I genotypes (19); and ii) 488,265 individuals from the UK Biobank (<https://www.ukbiobank.ac.uk/>), using the HLA alleles imputed with HLA*IMP:02, as previously described (20). UK Biobank individuals were categorized into different cancer groups by organ location/pathology according to hospital and cancer records using ICD codes from either ICD 10 or ICD 9 (when applicable). Patients in the “normal” category did not have an ICD code of any cancer type. For both TCGA and UK Biobank cohorts, individuals whose HLA-I genotypes showed two identical alleles in at least one HLA-I gene were classified as being homozygous. The number of homozygous genes (varying from one to all three HLA-I genes) was also recorded. The rate of homozygosity in the general population was estimated using normal samples from the GTEx project (<https://gtexportal.org/home/>, dbGaP accession

phs000424.v8.p2). For this purpose, we selected samples from several tissue types (21), as follows: bladder (n=11, representing all samples available for download), brain (arbitrary selection of n=21 samples), lung (n=21), pancreas (n=21) and skin (n=21). For these 95 samples, HLA-I genotypes were inferred from RNA-seq (22).

***In-Silico* neo-antigen prediction**

Somatic coding variants were annotated with Variant Effect Predictor (VEP) version 73 and were used as input in the stand alone neo-antigen prediction algorithms NETMHC (23), NETMHCpan (12), and pickpocket (24). pVAC-Seq within the pVACtools suite was used as a convenient wrapper for aggregating the results of each prediction algorithm and for assessing and prioritizing neo-antigen predictions (25). The set of neo-antigen predictions output by pVAC-Seq were then filtered for several features so that the final candidates only included neo-antigens with i) strong predicted affinity for the MHC-I molecule of the same patient, ii) little to no homology to published mammalian epitopes, and iii) epitopes within genes expressed in DLBCL. Briefly, neo-antigens with predicted affinity <500 nm and stronger than the predicted affinity of the corresponding WT epitope were retained. We then filtered for 9 and 10 kmer sized neoepitopes found within genes expressed in human normal B cell subpopulations (n=3 independent samples each for naïve, memory or GC B cells) (26) and/or in DLBCL primary cases, as assessed based on RNA-seq analysis (<https://www.cancer.gov/tcga>). Finally, we performed neoepitope sequence alignment of the filtered candidates against the set of known epitopes obtained from the Immune Epitope database and analysis resource (IEDB.org). Any candidates with >70% homology to mammalian epitopes were removed from analysis.

Neo-antigen peptides synthesis and HLA-I binding affinity assay

The predicted neo-antigen peptides were custom synthesized at >95% purity (average 98.13%, range 97.17-99.27%) and lyophilized (Thermo Fisher Scientific). Depending on the amino acidic composition, the peptide powders were dissolved in water or in dimethylsulfoxide, and diluted in deionized filtered water in 1mM aliquots and stored at -80° C. We then validated experimentally the binding affinity of the predicted neo-antigens to the HLA-A*02:01 complex (27). In brief, T2 cells were washed twice in 1x PBS and then plated in 24-well cell culture plates at 0.7 x 10⁶ cells/mL in serum free Iscove's Modified Dulbecco's Medium containing 3 µg/mL of urine-isolated beta-2-microglobulin (Millipore-Sigma) and incubated with or without the synthesized peptides for 16 hours in a humidified chamber at 37° C with 5% CO₂. The OVA (257-264) peptide (AnaSpec, Inc.) was used as negative control, and cells without peptide were used for background subtraction. After incubation, cells were washed twice in 1x PBS and stained with PE-conjugated anti-human HLA-I antibody (W6/32, 1:200 Biolegend, cat# 311406) for 20 minutes on ice in the dark, washed twice in PBS followed by analysis on a FACSCanto instrument (Becton Dickinson). Data were then analyzed using the FlowJo software (TreeStar). Background signal was subtracted using the following formula:

$$\text{Normalized MFI} = \frac{(\text{MFI sample} - \text{MFI background})}{\text{MFI background}}$$

Flow Cytometry

Analysis of cell surface HLA-I in DLBCL cell lines was performed using the W6/32 antibody conjugated with R-phycoerythrin (PE), Pacific Blue and Alexa Fluor® 647 fluorochromes (1:200; Biolegend, cat# 311406, cat# 311418, and cat# 311414)(28). Expression of cell surface B2M was determined using the 2M2 antibody conjugated with allophycocyanin (APC), (1:200; Biolegend,

cat# 316312). Cells were washed twice for 5 min in PBS + 0.5% BSA and incubated for 20 min with 0.5 µg of antibody per million cells in 100 µl of PBS + 0.5% BSA on ice. The unbound antibody was removed by washing the cells twice with PBS+0.5% BSA for 5 min followed by acquisition on a FACSCanto instrument (Becton Dickinson). Data were analyzed using the FlowJo software (TreeStar).

HLA-I reconstitution in *B2M* WT cell lines

To reintroduce WT *HLA-I* alleles in the MHC-I negative OCI-LY10 and RC-K8 DLBCL cell lines, cells were first transduced with an amphotropic virus expressing the ecotropic retroviral receptor according to previously published methods (29). WT *HLA-I* alleles (A*02:01, B*18:01, C*04:01) were cloned from the cDNA of primary DLBCL biopsies into the MSCV PIG expression vector (Puromycin IRES GFP) (gift from Scott Lowe; Addgene plasmid # 18751). Virus containing supernatants were produced by transiently transfecting the Phoenix ecotropic cell line according to standard protocols, filtered in 0.45 µm syringe strainers, and used to infect OCI-LY10 and RC-K8 cells expressing ecotropic receptor after addition of 4 µg/mL polybrene. Transduced cells were selected with Puromycin (OCI-LY10, 1.5 µg/mL; RC-K8, 5 µg/mL) for 5 days before analysis. Cells were analyzed for HLA-I surface expression by flow cytometric analysis.

Flow cytometric analysis of mouse B cell subsets

Multi-color flow cytometric analysis of the B cell lymphoid compartment was performed at 3 months of age and at endpoint, as previously reported (30). Briefly, single cell suspensions prepared from lymphoid organs were stained for 20 minutes on ice using different combinations

of fluorescent-labeled antibodies ([Dataset S10](#)). Data were acquired on a FACSCanto™ II or a FACSCalibur™ (BD Biosciences), and analyzed using the FlowJo software (TreeStar).

Immunohistochemistry and immunofluorescence analysis of mouse tissues

FFPE spleen sections were stained with hematoxylin and eosin (Thermo Fisher Scientific) or analyzed by immunohistochemistry using standard protocols (2) and antibodies against B220 (RA3-6B2, 1:400, BD Bioscience cat# BD553086), Cd3 (SP7, 1:800, Thermo Scientific cat# RM-9107), Pax5 (C-20, 1:500, Santa-Cruz cat# sc-1974), Bcl6 (N3, 1:400, Santa-Cruz cat# sc-858) and Irf4 (M17, 1:500, Santa-Cruz cat# sc-6059). Secondary antibodies were as follows: for Cd3 and Bcl6 stained sections, horseradish peroxidase-conjugated goat anti-rabbit secondary antibody (EnVision, Dako), followed by visualization with 3-amino-9-ethylcarbazole (Sigma); for B220 stained slides, alkaline phosphatase (AP)-conjugated streptavidin (Vector Laboratories); for Pax5 and Irf4 stained sections, AP-conjugated rabbit anti-goat secondary antibodies (Southern Biotech). Detection was performed using NBT/BCIP substrate (blue color; Roche) for the AP-conjugates. Immunofluorescence analysis of B2m and Pax5 was performed on FFPE material using a combination of anti-B2m (EP2978Y, 1:250, Abcam cat# ab75853) and anti-Pax5 (C-20, 1:500, Santa-Cruz cat# sc-1974) antibodies. Detection was obtained with Alexa Fluor 488-conjugated donkey anti-rabbit secondary antibodies (1:400, Thermo Scientific cat# A-21206) and, for Pax5, a goat anti-rabbit HRP conjugated secondary antibody followed by Tyramide Signal Amplification Plus Cyanine 3 System (PerkinElmer cat# NEL744001KT). Images were captured using a Nikon Eclipse microscope and the NIS Elements software (Nikon). All images were colored, resized, and merged using Adobe Photoshop (version CS3). Analysis of GC numbers, size, and overall GC area

was performed using Photoshop (version CS3) on images of sections stained with the Bcl6 antibody and scanned with a Leica SCN400 slide scanner.

Statistical analysis

The Fisher's exact test was used to assess whether differences in categorical variables (e.g., the frequency of mutated cases in separate disease groups) were significant. Continuous variables, not assuming normal distribution, were compared by the Mann-Whitney U test (two sided). Statistical significance was defined as *p-value* <0.05. The analysis was performed with the Statistical Package for the Social Sciences (SPSS) software v.19.0 (Chicago, IL).

SUPPLEMENTAL REFERENCES

1. Salter RD, Howell DN, & Cresswell P (1985) Genes regulating HLA class I antigen expression in T-B lymphoblast hybrids. *Immunogenetics* 21(3):235-246.
2. Cattoretti G, *et al.* (2005) Deregulated BCL6 expression recapitulates the pathogenesis of human diffuse large B cell lymphomas in mice. *Cancer cell* 7(5):445-455.
3. Stam NJ, Vroom TM, Peters PJ, Pastoors EB, & Ploegh HL (1990) HLA-A- and HLA-B-specific monoclonal antibodies reactive with free heavy chains in western blots, in formalin-fixed, paraffin-embedded tissue sections and in cryo-immuno-electron microscopy. *Int Immunol* 2(2):113-125.
4. Favero F, *et al.* (2015) Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* 26(1):64-70.
5. Chapuy B, *et al.* (2018) Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nature medicine* 24(5):679-690.
6. de Miranda NF, *et al.* (2014) Exome sequencing reveals novel mutation targets in diffuse large B-cell lymphomas derived from Chinese patients. *Blood* 124(16):2544-2553.
7. Ren W, *et al.* (2018) Genetic landscape of hepatitis B virus-associated diffuse large B-cell lymphoma. *Blood* 131(24):2670-2681.
8. Okonechnikov K, Conesa A, & Garcia-Alcalde F (2016) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32(2):292-294.
9. Trifonov V, Pasqualucci L, Tiacci E, Falini B, & Rabadan R (2013) SAVI: a statistical algorithm for variant frequency identification. *BMC Syst Biol* 7 Suppl 2:S2.
10. Shukla SA, *et al.* (2015) Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol* 33(11):1152-1158.
11. Yewdell JW, Reits E, & Neefjes J (2003) Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nature reviews. Immunology* 3(12):952-961.
12. Jurtz V, *et al.* (2017) NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* 199(9):3360-3368.
13. McGranahan N, *et al.* (2017) Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* 171(6):1259-1271 e1211.
14. Mermel CH, *et al.* (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 12(4):R41.
15. Compagno M, *et al.* (2009) Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* 459(7247):717-721.
16. Pasqualucci L, *et al.* (2001) Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 412(6844):341-346.
17. Vakiani E, *et al.* (2008) Genetic and phenotypic analysis of B-cell post-transplant lymphoproliferative disorders provides insights into disease biology. *Hematol Oncol* 26(4):199-211.
18. Schmitz R, *et al.* (2018) Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *The New England journal of medicine* 378(15):1396-1407.

19. Thorsson V, *et al.* (2018) The Immune Landscape of Cancer. *Immunity* 48(4):812-830 e814.
20. Bycroft C, *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562(7726):203-209.
21. Battle A, Brown CD, Engelhardt BE, & Montgomery SB (2017) Genetic effects on gene expression across human tissues. *Nature* 550(7675):204-213.
22. Orenbuch R, *et al.* (2020) arcasHLA: high-resolution HLA typing from RNAseq. *Bioinformatics* 36(1):33-40.
23. Andreatta M & Nielsen M (2016) Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 32(4):511-517.
24. Zhang H, Lund O, & Nielsen M (2009) The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. *Bioinformatics* 25(10):1293-1299.
25. Hundal J, *et al.* (2019) pVACtools: a computational toolkit to identify and visualize cancer neoantigens. *bioRxiv*:501817.
26. Brescia P, *et al.* (2018) MEF2B Instructs Germinal Center Development and Acts as an Oncogene in B Cell Lymphomagenesis. *Cancer cell* 34(3):453-465 e459.
27. Steinle A & Schendel DJ (1994) HLA class I alleles of LCL 721 and 174 x CEM.T2 (T2). *Tissue Antigens* 44(4):268-270.
28. Parham P, Sehgal PK, & Brodsky FM (1979) Anti-HLA-A,B,C monoclonal antibodies with no alloantigenic specificity in humans define polymorphisms in other primate species. *Nature* 279(5714):639-641.
29. Pear WS, Nolan GP, Scott ML, & Baltimore D (1993) Production of high-titer helper-free retroviruses by transient transfection. *Proceedings of the National Academy of Sciences of the United States of America* 90(18):8392-8396.
30. Meyer SN, *et al.* (2019) Unique and Shared Epigenetic Programs of the CREBBP and EP300 Acetyltransferases in Germinal Center B Cells Reveal Targeted Dependencies in Lymphoma. *Immunity*.

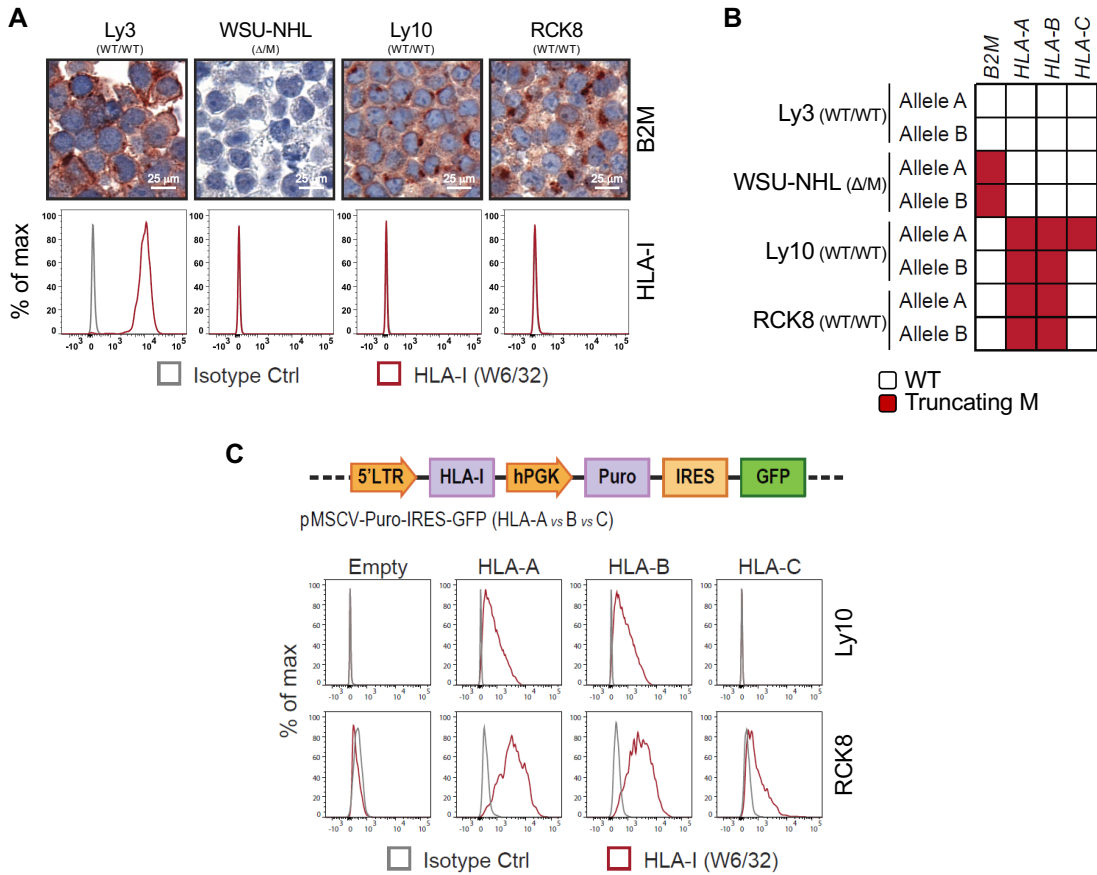


Figure S1. *B2M* WT cell lines with mislocalized B2M protein harbor biallelic alterations in the *hcHLA-I* alleles. (A) Top: immunohistochemistry analysis of B2M in four DLBCL cell lines carrying different genetic configurations of the *B2M* locus (M, mutated; Δ, deleted; WT, wild type) (Scale = 25 μm). Bottom: FACS analysis of the same cell lines with the pan-*HLA-I* antibody (W6/32) (red), or an isotype control (grey). (B) Mutation analysis of the *hcHLA-I* genes in the four cell lines. In the heatmap, white indicates WT alleles and red indicates alleles carrying inactivating mutations. (C) Top: map of the retroviral vector used for the reconstitution experiment in two *B2M*-WT cell lines showing mislocalization of the B2M protein. Bottom: FACS analysis of RCK8 and Ly10 cells, transduced with empty vector or vectors expressing *hcHLA-I* A, -B and -C and stained with the pan-*HLA-I* antibody (W6/32) vs isotype control.

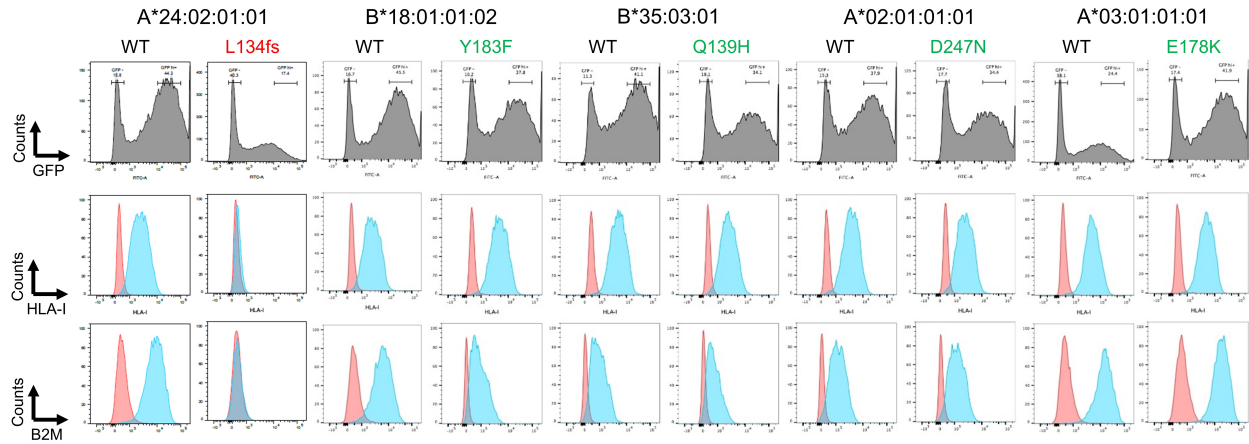


Figure S2. DLBCL-associated *hcHLA-I* missense mutant alleles are able to restore the expression of surface MHC-I complex. FACS analysis of HLA-I (W6/32) and B2M (2M2) expression in the MHC-I negative RC-K8 cell line, transduced with vectors expressing WT and missense mutant versions of different *HLA-I* alleles identified in DLBCL cases. An expression vector for an *hcHLA-I* allele harboring a truncating mutation was used as negative control. Red, untransduced cells; Turquoise, transduced cells.

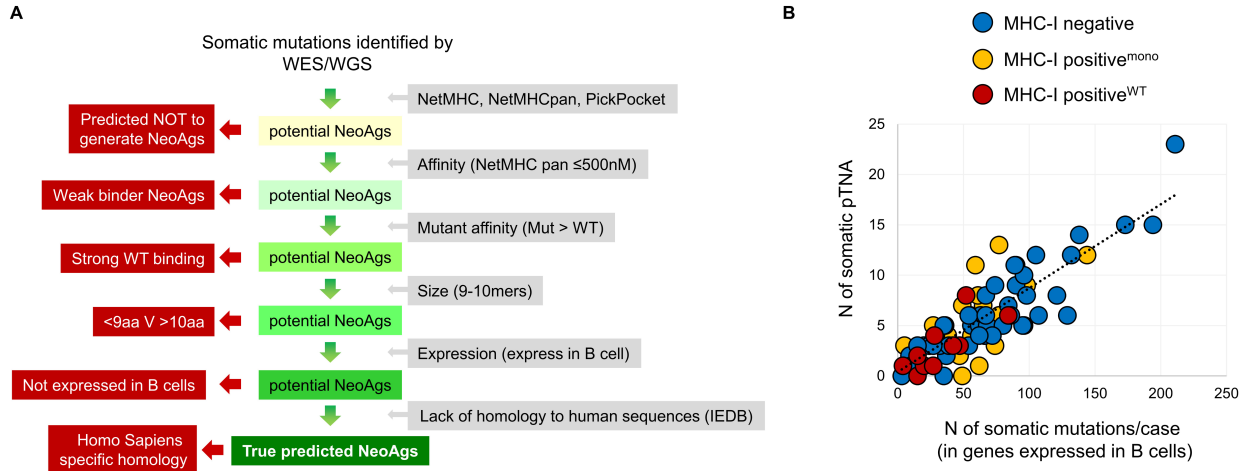
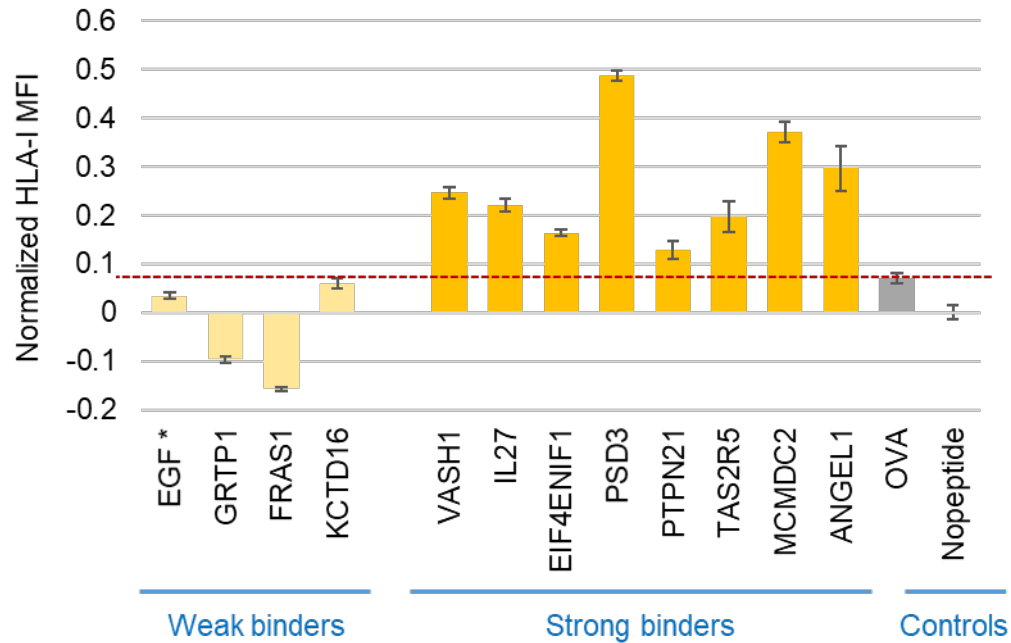


Figure S3. Linear correlation between coding mutational load and neo-antigen load. (A) Schematics of the computational pipeline used for neo-antigen prediction. **(B)** Dot scattered plot illustrating the number of nonsynonymous mutations (y -axis) vs number of predicted tumor neo-antigens (pTNA) (x -axis) in 74 DLBCL cases, color-coded according to MHC-I protein/HLA-I gene status. Dotted line indicates the linear regression line ($R^2=0.714$).



* Predicted to bind HLA-B*18:01

Figure S4. Predicted neo-antigens bind HLA-I and stabilize the MHC-I complex on the cell surface. The histogram represents the normalized HLA-I MFI of T2 cells, analyzed by FACS after co-culture with or without the synthesized peptide neo-antigens and incubation with the anti-human HLA-I antibody W6/32. The OVA (257-264) peptide was used as negative control, and T2 cells co-cultured without peptide were used for background subtraction and normalization.

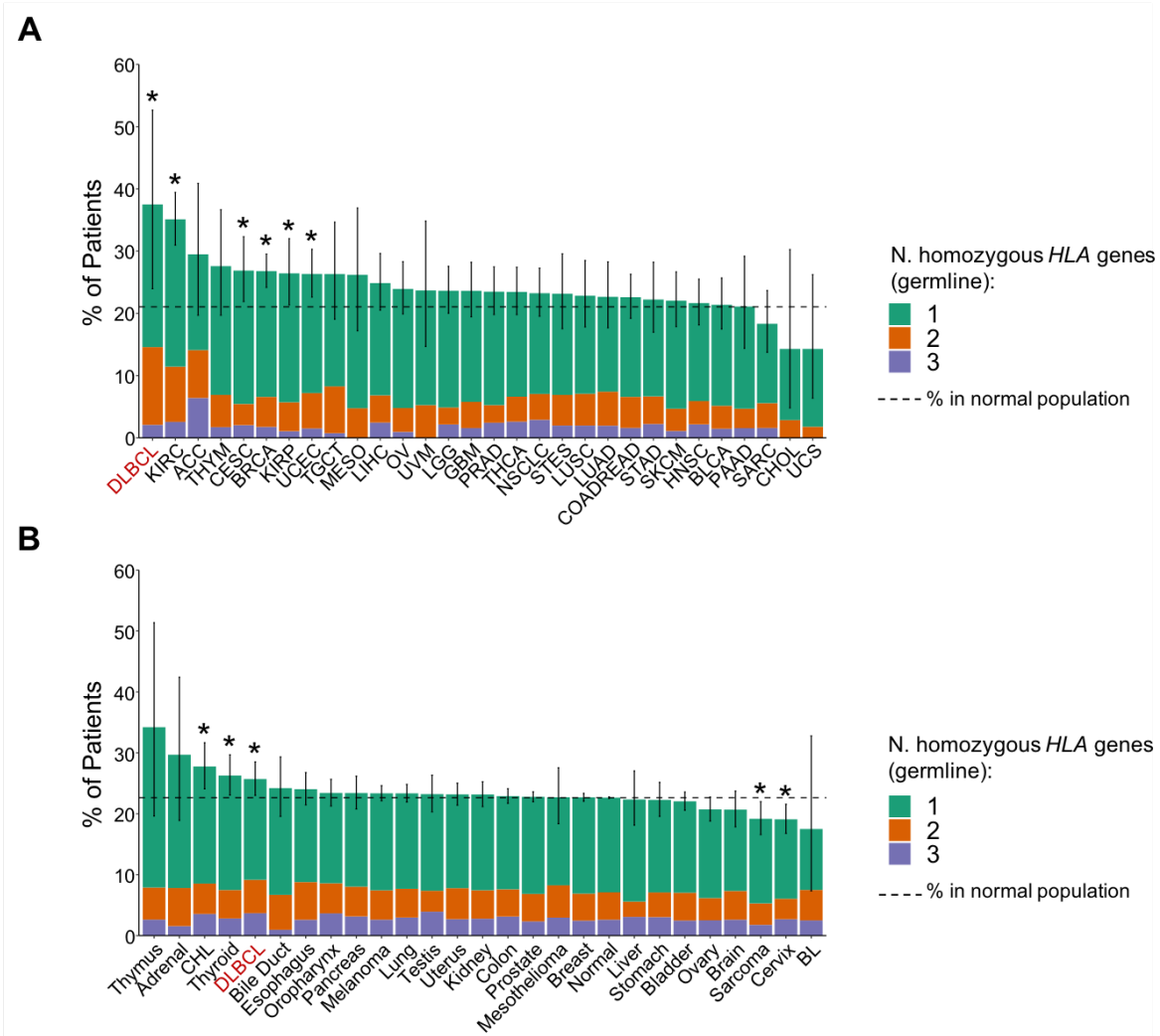


Figure S5. Rates of germline homozygosity in HLA-I genes across different cancers. The relative fraction of patients with one, two, or all three major HLA-I germline genes in homozygosity is shown for each of the indicated tumor types, as assessed in the TCGA (**A**) and UK Biobank (**B**) databases. Cohorts marked with an asterisk (*) had a significantly different homozygosity rate than the background normal rate calculated in healthy individuals from the UK Biobank (23%; used for comparison in the UK Biobank cohort), or from GTEx (21%; used for comparison in TCGA) ($p < 0.05$, binomial test, 95% CI shown). UK BioBank patients: Adrenal (n=64); Bile Duct (n=314); BL, Burkitt lymphoma (n=40); Bladder (n=3026); Brain (n=764); Breast (n=16245); Cervix (n=1074); CHL, Classical Hodgkin lymphoma (n=562); Colon (n=4787); DLBCL, Diffuse large B-cell lymphoma (n=1004); Esophagus (n=1036); Kidney (n=1692); Liver (n=358); Lung (n=3332); Melanoma (n=4504); Mesothelioma (n=339); Oropharynx (n=1478); Ovary (n=1670); Pancreas (n=983); Prostate (n=10243); Sarcoma (n=850); Stomach (n=889); Testis (n=788); Thymus (n=38); Thyroid (n=708); Uterus (n=2144). TCGA patients: ACC, Adrenocortical carcinoma (n=78); BLCA, Bladder urothelial carcinoma (n=407); BRCA, Breast invasive carcinoma (n=1079); CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma (n=294); CHOL, Cholangiocarcinoma (n=35); COADREAD,

Colon adenocarcinoma/Rectum adenocarcinoma (n=562); DLBCL, Diffuse large B-cell lymphoma (n=48); GBM, Glioblastoma multiforme (n=381); HNSC, Head and neck squamous cell carcinoma (n=508); KIRC, Kidney renal clear cell carcinoma (n=507); KIRP, Kidney renal papillary cell carcinoma (n=280); LGG, Brain lower grade glioma (n=512); LIHC, Liver hepatocellular carcinoma (n=366); LUAD, Lung adenocarcinoma (n=256); LUSC, Lung squamous cell carcinoma (n=254); MESO, Mesothelioma (n=84); NSCLC, Non-small cell lung cancer (n=482); OV, Ovarian serous cystadenocarcinoma (n=418); PAAD, Pancreatic adenocarcinoma (n=128); PRAD, Prostate adenocarcinoma (n=494); SARC, Sarcoma (n=251); SKCM, Skin cutaneous melanoma (n=363); STAD, Stomach adenocarcinoma (n=225); STES, Esophagus-stomach cancers (n=203); TGCT, Testicular germ cell tumors (n=133); THCA, Thyroid carcinoma (n=499); THYM, Thymoma (n=116); UCEC, Uterine corpus endometrial carcinoma (n=528); UCS, Uterine carcinosarcoma (n=56); UVM, Uveal melanoma (n=76).

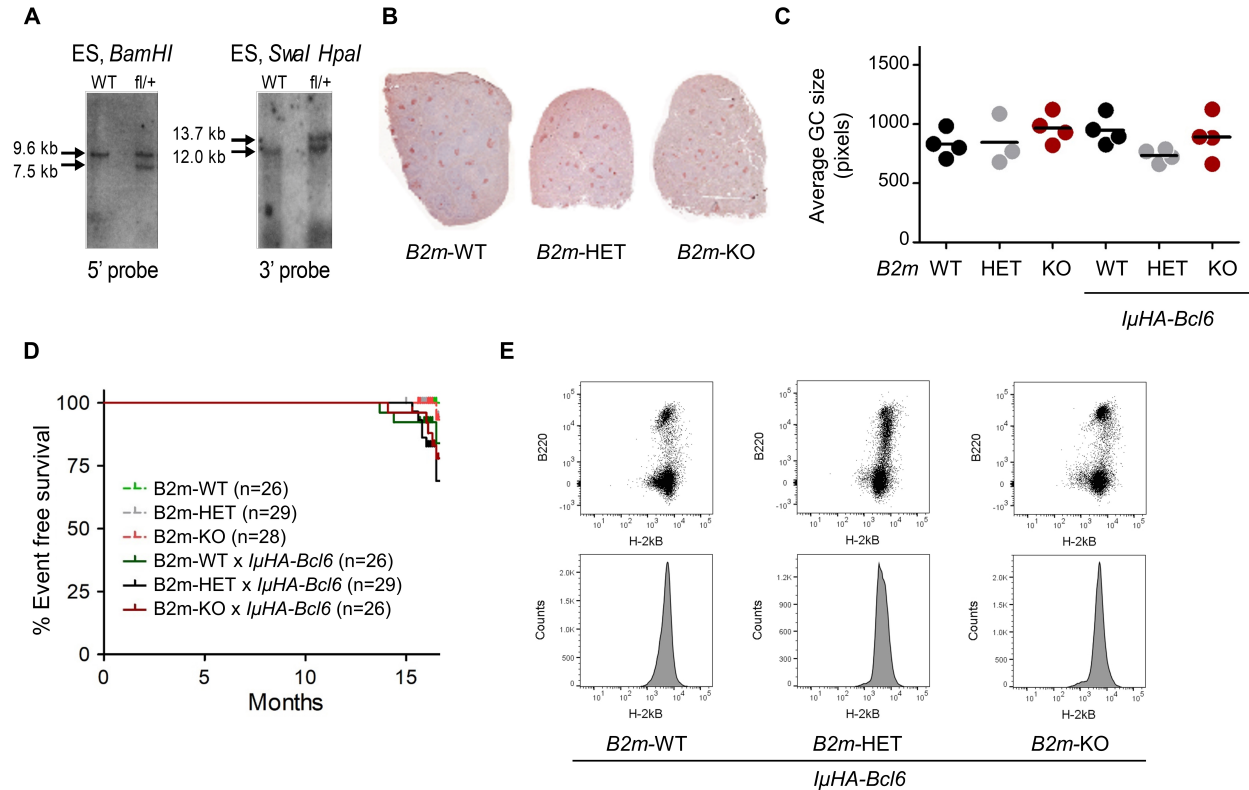


Figure S6. Characterization of *B2m*-KO and *IμHA-Bcl6*/*B2m*-KO compound mice. (A) Left: Southern blot analysis of BamHI digested genomic DNA using a 5' probe confirms the predicted 9.6 kb WT band along with the correctly targeted 7.5 kb band in representative ES cells. Right: Southern blot analysis of Swal/HpaI-double digested DNA using a 3' probe displays the 12 kb WT band and a 13.7 kb band corresponding to the targeted allele in ES cells. **(B)** Immunohistochemistry for Bcl6 on representative spleen sections from mice of the indicated genotypes. Abbreviations: *B2m*-WT (*B2m*^{+/+}/*Cγ1*^{cre/+}), *B2m*-HET (*B2m*^{fl/+}/*Cγ1*^{cre/+}), *B2m*-KO (*B2m*^{fl/fl}/*Cγ1*^{cre/+}). **(C)** Average GC size in mice of the indicated genotypes, measured by pixels quantification of scanned images from Bcl6-stained sections. **(D)** Event-free survival in mice of the indicated genotypes. The total number of animals analyzed is indicated in brackets for each cohort. **(E)** Flow cytometric analysis of B220 and H2kB (MHC-I) in splenic B cells of 15-months old *IμHA-Bcl6*/*B2m*-WT, *IμHA-Bcl6*/*B2m*-HET and *IμHA-Bcl6*/*B2m*-KO mice.

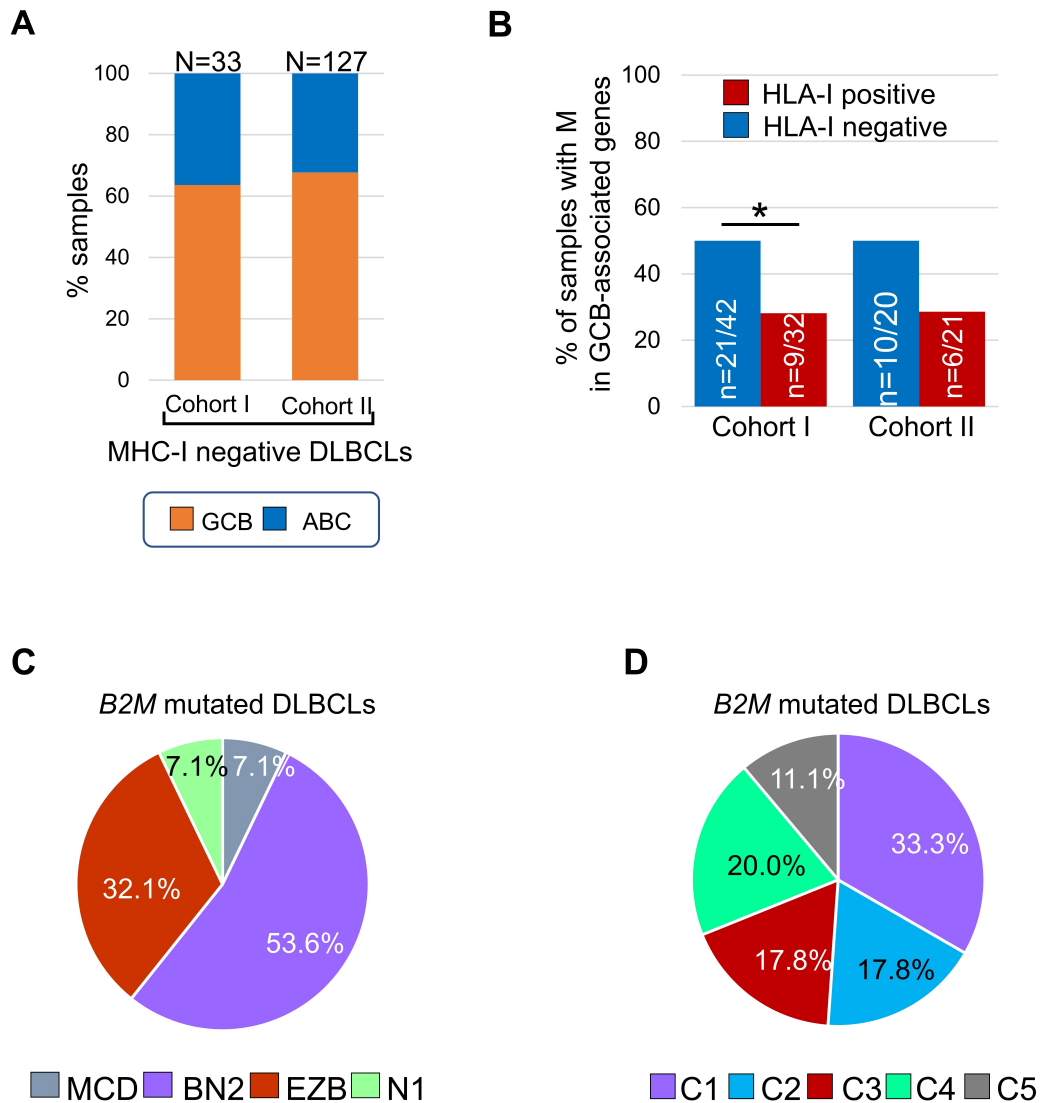


Figure S7. Phenotypic and genetic features of MHC-I defined DLBCL groups. (A) COO classification of MHC-I negative DLBCL in two independent cohorts (cohort I, this study; cohort II, Ennishi et al., Cancer Discovery 2019); (B) Distribution of mutations affecting GCB-associated genes (e.g. *BCL2*, *EZH2*, *GNAI3*, *P2RY8*, and *CREBBP*) in MHC-I^{POS} vs MHC-I^{NEG} DLBCL cases, * p-value <0.05; (C) Genetic classification of *B2M*-mutated DLBCL samples in Schmitz et al., New Engl J Med 2018 (total n=28); (D) Genetic classification of *B2M* mutated DLBCL samples in Chapuy et al., Nature Medicine 2018 (total n=45). In both series, the majority of *B2M* mutated DLBCL were classified as ABC-like BN2/C1, with the GCB-like EZB/C3 class being the second most represented.