haematologica
Journal of the Ferrata Storti Foundation

Epstein-Barr Virus and immune status imprint the immunogenomics of non-Hodgkin lymphomas occurring in immune-suppressed environments

by Marine Baron, Karim Labreche, Marianne Veyri, Nathalie Desire, Amira Bouzidi, Fatou Seck-Thiam, Frederic Charlotte, Alice Rousseau, Veronique Morin, Cecilia Nakid-Cordero, Baptiste Abbar, Alberto Picca, Marie Le Cann, Noureddine Balegroune, Nicolas Gauthier, Ioannis Theodorou, Mehdi Touat, Veronique Morel, Franck Bielle, Assia Samri, Agusti Alentorn, Marc Sanson, Damien Roos-Weil, Corinne Haioun, Elsa Poullot, Anne Langlois De Septenville, Frederic Davi, Amelie Guihot, Pierre-Yves Boelle, Veronique Leblond, Florence Coulet, Jean-Philippe Spano, Sylvain Choquet, and Brigitte Autran.
Collaborative Groups: Ideation Study Group (Ideation Study Group)

Received: September 20, 2023.
Accepted: May 22, 2024.

Citation: Marine Baron, Karim Labreche, Marianne Veyri, Nathalie Desire, Amira Bouzidi, Fatou Seck-Thiam, Frederic Charlotte, Alice Rousseau, Veronique Morin, Cecilia Nakid-Cordero, Baptiste Abbar, Alberto Picca, Marie Le Cann, Noureddine Balegroune, Nicolas Gauthier, Ioannis Theodorou, Mehdi Touat, Veronique Morel, Franck Bielle, Assia Samri, Agusti Alentorn, Marc Sanson, Damien Roos-Weil, Corinne Haioun, Elsa Poullot, Anne Langlois De Septenville, Frederic Davi, Amelie Guihot, Pierre-Yves Boelle, Veronique Leblond, Florence Coulet, Jean-Philippe Spano, Sylvain Choquet, and Brigitte Autran. Collaborative Groups: Ideation Study Group (Ideation Study Group). Epstein-Barr Virus and immune status imprint the immunogenomics of non-Hodgkin lymphomas occurring in immune-suppressed environments. Haematologica. 2024 June 6. doi: 10.3324/haematol.2023.284332 [Epub ahead of print]

Epstein-Barr Virus and immune status imprint the immunogenomics of non-Hodgkin lymphomas occurring in immune-suppressed environments

Marine Baron[1,2*], Karim Labreche[3*], Marianne Veyri[4#], Nathalie Désiré[3#], Amira Bouzidi[5], Fatou Seck-Thiam[3], Frédéric Charlotte[6], Alice Rousseau[1], Véronique Morin[1], Cécilia Nakid-Cordero[1], Baptiste Abbar[1], Alberto Picca[1], Marie Le Cann[7], Noureddine Balegroune[2], Nicolas Gauthier[2], Ioannis Theodorou[8], Mehdi Touat[9], Véronique Morel[2], Franck Bielle[10], Assia Samri[1], Agusti Alentorn[9], Marc Sanson[9], Damien Roos-Weil[2], Corinne Haioun[11], Elsa Poullot[12], Anne Langlois de Septenville[13], Frédéric Davi[13], Amélie Guihot[1], Pierre-Yves Boelle[3], Véronique Leblond[2], Florence Coulet[14], Jean-Philippe Spano[4+], Sylvain Choquet[2+], Brigitte Autran[1+], on behalf the IDeATIon study group.

Study group: Baptiste Abbar[4], Isabelle Brocheriou[6], Jacques Cadranel, Jérôme Denis[5], Erell Guillerm[13], Ahmed Ibdaih[8], Stéphanie Jouannet[5], Jean-Marc Lacorte[5], Anne-Geneviève Marcelin[14], Alberto Picca[2], Kahina Belkhir[4], Cécilia Nakid-Cordero[2]

* These authors contributed equally.

# These authors contributed equally.

+ These authors contributed equally.

Affiliations

[1]Sorbonne Université, INSERM U1135, Center for Immunology and Infectious Diseases (CIMI), Department of Immunology, AP-HP, Hôpital Pitié-Salpêtrière, F-75013 Paris, France

[2]Sorbonne Université, Department of Clinical Haematology, AP-HP, Hôpital Pitié-Salpêtrière, F-75013 Paris, France

[3]Sorbonne Université, CinBioS, UMS 37 PASS Production de données en Sciences de la vie et de la Santé, INSERM, 75013 Paris, France

[4] Sorbonne Université, INSERM, Pierre et Louis Institute of Epidemiology and Public Health, F-75013 Paris France, Theravir Team, Department of Medical Oncology, AP-HP, Hôpital Pitié-Salpêtrière, F-75013 Paris, France

[5]Sorbonne Université, INSERM, Research Unit on Cardiovascular and Metabolic Disease UMR ICAN, Department of Endocrine Biochemistry and Oncology, AP-HP, Hôpital-Pitié-Salpêtrière, F-75013 Paris, France

[6]Sorbonne Université, Department of Anatomy and Pathologic Cytology, AP-HP, Hôpital Pitié-Salpêtrière, F-75013 Paris, France

[7] Department of Clinical Haematology, AP-HP, Hôpital Kremlin Bicêtre, F-94270 Le Kremlin, France

[8] Department of Immunology, Hôpital Robert Debré, F-75019, Paris, France

[9] Sorbonne Université, INSERM, CNRS, Brain and Spine Institute, ICM, Department of Neurology 2-Mazarin, AP-HP, Hôpital Pitié-Salpêtrière, F-75013 Paris, France

[10] Sorbonne Université, Department of Neuropathology, AP-HP, Hôpital Pitié-Salpêtrière, F-75013, Paris, France

[11]Lymphoid malignancies unit, AP-HP, Mondor Hospital, F-94000 Créteil, France

[12]Department of Anatomy and Pathologic Cytology, AP-HP, Mondor Hospital, F-94000 Créteil, France

[13]Sorbonne Université, INSERM, Centre de Recherche des Cordeliers, Department of Biological Hematology, AP-HP, Hôpital Pitié-Salpêtrière, Paris, France

[14]Sorbonne Université, INSERM, Saint-Antoine Research Center, Microsatellites Instability and Cancer, CRSA, Department of Medical Genetics, AP-HP, Pitié-Salpêtrière Hospital, F-75013 Paris, France

Running heads: Immunogenomics of immune-suppressed NHL

Corresponding author: Marine Baron, 47-83 boulevard de l'Hôpital 75013 Paris, France. E-mail: marine.baron@aphp.fr, phone number: 33 1 42 16 23 46; fax number: 33 1 42 17 27 87

Authorship Contributions: B. Autran, JP.S, M.S., V.L, S.C. and A.G. designed the study, F.C., F.B. and E.P. performed the histopathological diagnosis, A.B. and F.C. performed the sequencing analysis, K.L. and P-Y.B. developed the bioinformatic pipeline, K.L., N.D. and F.S-T. performed the bioinformatic analysis, K.L performed the statistical analysis, N.B. collected the biological samples, C.N-C developed the immunological analysis, M.B., A.R. and V.M. performed the immunological analysis, F.D, A.L de S., I.T. assisted in interpreting the results, M.B and K.L. analyzed all the data, D.R-W., M. L. C., S.C., V.L., V.M., N.G., M.T, A.A., M.S., B. Abbar, A.P. and C.H. took care of the patients, M.V. is the project manager of the IDeATIoN project, M.B. wrote the manuscript, B. Autran. supervised all the work, all the authors commented and reviewed the manuscript.

Acknowledgements: We acknowledge all the patients included in the study. We deeply thank MSDAvenir and Fondation pour la Recherche Médicale for their financial support. We thank the CRB (Centre de Ressources Biologiques) of Paris-Saclay and J. Tisserand, the CRB-C SU and Rihab Jrad, the PRB (Plateforme de Ressources Biologiques) of Hôpital Mondor, and Dr C. Barau, Dr. A. Jebanesan, and Dr. MC. Morcelet for giving and transferring tumor samples. We thank the biobank Onconeurotek of AP-HP Sorbonne Université and Amel Dridi-Aloulou for transferring brain samples. We thank Elodie Courret for its support for biological samples.

Funding: The IDeaTIon project is financially supported by the MSDAvenir endowment fund (grant number: DS-2017-0018). This work was in part supported by the grant INCa-DGOS-Inserm_12560 of the SiRIC CURAMUS (Cancer United Research Associating Medicine, University and Society integrated cancer research program), and by the Fondation pour la Recherche Médicale, grant number FDT202106013113 to MB.

Disclosure of Conflicts of Interest: No disclosure.

Data sharing statement: Data are available upon reasonable request. The procedures carried out with the French data privacy authority (CNIL, Commission nationale de l'informatique et des libertés) and the european regulation (GDPR) do not provide for the transmission of the database, nor do the information and consent documents signed by the patients. Consultation by the editorial board or interested researchers of individual participant data that underlie the results reported in the article after deidentification may nevertheless be considered, subject to prior determination of the terms and conditions of such consultation and in respect for compliance with the applicable regulations. For all inquiries, please contact the following: drc-secretariat-promotion@aphp.fr.

The R code (version 4.2.1) utilized in our research is accessible upon a reasonable request. Please reach out to the corresponding author to obtain access

ABSTRACT

Non-Hodgkin lymphomas (NHL) commonly occur in immune-deficient (ID) patients, both HIV-infected and transplanted, and are often EBV-driven with cerebral localization, raising the question of tumor immunogenicity, a critical issue for treatment responses. We investigated the immunogenomics of 68 lymphoproliferative disorders from 51 ID (34 post-transplant, 17 HIV+) and 17 immunocompetent patients. Overall, 72% were Large B Cells Lymphoma (LBCL) and 25% were primary central-nervous-system lymphoma (PCNSL) while 40% were EBV-positive. Tumor whole-exome and RNA sequencing, along with a bioinformatics pipeline allowed analysis of tumor mutational burden (TMB), tumor landscape and microenvironment (TME) and prediction of tumor neoepitopes. Both TMB (2.2 *vs* 3.4/Mb, p=0.001) and neoepitopes numbers (40 *vs* 200, p=0.00019) were lower in EBV-positive than in EBV-negative NHL, regardless of the immune status. In contrast both EBV and the immune status influenced the tumor mutational profile, with *HNRNPF* and *STAT3* mutations exclusively observed in EBV-positive and ID NHL, respectively. Peripheral blood T-cell responses against tumor neoepitopes were detected in all EBV-negative cases but in only half EBV-positive ones, including responses against *IgH*-derived MHC-class-II restricted neoepitopes. The TME analysis showed higher CD8 T cell infiltrates in EBV-positive vs EBV-negative NHL, together with a more tolerogenic profile composed of Tregs, type-M2 macrophages and an increased expression of negative immune-regulators.

Our results highlight that the immunogenomics of NHL in patients with immunodeficiency primarily relies on the tumor EBV status, while T cell recognition of tumor- and *IgH*-specific neoepitopes is conserved in EBV-negative patients, offering potential opportunities for future T cell-based immune therapies.

INTRODUCTION

Non-Hodgkin lymphomas (NHL) are highly prevalent in severe immune-deficient (ID) patients such as HIV-infected patients and transplant recipients. In both cases, the incidence is higher than in immunocompetent individuals, with a standardized incidence ratio of 11.5 for HIV-related lymphomas and 8 for post-transplant lymphoproliferative disorders (PTLD)[1–3]. The Epstein-Barr virus (EBV) is present in 50–70% of the ID NHL and has been established as a strong oncogene[4,5]. Several targeted sequencing approaches have shown that the mutational landscape of EBV-positive NHL differs from that of EBV-negative NHL, with lower number of mutated genes[6–11]. In addition, ID patients display alterations of the normally potent cellular immunity directed against viral antigens[12,13], potentially leading to immune escape[14–17]. Finally, almost 10% of ID patients, the majority of whom EBV+, have a central-nervous-system (CNS) localization compared to 1% in the immunocompetent population[4,5,18]. As the CNS acts as an immunologically isolated site with very specific immune features, it is hypothesized that the tumor-specific and EBV-specific immune responses are even less active in this tissue, thus favoring immune escape.

The remarkable successes of immunotherapies in solid tumors relay on the level of tumor immunogenicity[19–21]. The tumor neoantigens (NeoAg) that surround tumor mutations and are presented by the patients' MHC molecules allow the immune system to distinguish cancer from noncancer cells and emerge as major factors for anti-tumor immunity, response to immunotherapies and the development of targeted treatments. Indeed, the tumor mutational burden (TMB) and the abundance of predicted immunogenic mutations are associated with higher levels of responses to immune checkpoint inhibitors (ICPi) and increased patient survival[22–24] while personalized NeoAg vaccine approaches have successfully induced NeoAg-specific T cells associated with clinical responses[19,25–28]. However limited data exist for B-cell malignancies which are very specific due to the ability of tumor cells to present

tumor NeoAg with MHC class-II molecules and to produce highly mutated immunoglobulins (Ig) that may be immunogenic. Furthermore, studies in immunocompetent patients suggested that NeoAg derived from the lymphoma Ig heavy- or light-chain variable regions with a strong bias for an MHC-II presentation may be particularly critical immune targets[29,30]. However, the recognition of these peculiar Ig-based neoantigens has never been analyzed in ID patients.

We hypothesized that the immunogenomics and the tumor microenvironment (TME) of NHL occurring in ID patients might be influenced by the lower immune pressure, the frequent EBV oncogenesis and the immune-privileged sites in these patients. To further investigate this hypothesis, we developed the IDeATIon project, a multicentric, prospective study of PTLD and HIV-related lymphoproliferative disorders occurring in a large series of ID patients or in immune-privileged sites such as the CNS. We took advantage of whole tumor DNA and RNA sequencing to compare the TMB, numbers of tumor neoepitopes, neoepitope-specific T cell responses and TME as a function of both tumor EBV status and patient immune status.

METHODS

Patient cohort

From 2019 to 2022, consecutive HIV-infected patients or transplant recipients with treatment-naïve lymphoproliferative disorders were enrolled in the IDeATIon project (ClinicalTrials.gov NCT03706625). Immunocompetent patients with diffuse large B-cell lymphomas not otherwise specified were included as comparators. Diagnostic tissue biopsies and blood were collected at lymphoma diagnosis and before any treatment (**Supplementary materials)**. All patients gave written informed consent (CRB authorization no. 18.06.46). The

protocol was approved by the French national IRB (no. 2018-A01099-46) and the "Commission Nationale de l'Informatique et des Libertés" (no. 918222) and performed in accordance with the Declaration of Helsinki.

Immunoglobulin neoepitope prediction

Using RNAseq fastq files, sequences and frequencies of the Ig productive clonotypes were derived from MixCR-V3.0[31]. Reads were aligned to reference V, D, J and C genes of BCR. Each final IgH chain clonotype was identified by a CDR3 sequence. The FR3-CDR3-FR4 sequence of the dominant clonotype (≥15% of all clonotypes) was then tested for binding affinity score in NetMHC 4.0[32] and NetMHCIIpan 3.2[33] asking for 8–10 mers and 15 mers for MHC-I and II binding respectively. The number of neoepitopes with a score ≤500 nM was determined.

Neoepitopes-specific T cell expansion and functional validation

The most relevant non-Ig derived neoepitopes were selected upon their presentation by expressed HLA molecules or beta-2 microglobulin (RPKM ≥1), and the highest pVACseq priority score for SNV (**Supplementary materials),** and the highest binding affinity score for indels derived from the same somatic variant. When numbers of candidate neoepitopes were >60, the stability filter (NetMHCstab) was used to select the 60 strongest binders.

For Ig-derived neoepitopes, 8–14 mer peptides surrounding the mutation for MHC-I peptides and 25 mer peptides overlapping by 20–24 amino-acids and spanning the complete mutated predicted sequences for MHC-II peptides were selected.

A maximum of 60 peptides corresponding to neoepitopes were synthetized per patient (GeneCust, Boynes, France). Frozen PBMCs were thawed and co-cultured with each patient's personalized pooled peptides (2 µg/mL) in RPMI medium enriched with IL2 (10 UI/mL) (Roche, Bâle, Suisse), IL-7 (25 ng/mL) and IL-15 (25 ng/mL) (BioLegend, San Diego, CA).

On day 10, cell reactivity was tested in a triplicate IFN-γ ELISPOT assay after stimulation with personalized pooled peptides (2 µg/mL per pool), medium alone and phytohemagglutinin as negative and positive controls, respectively. When enough cells were available, additional assays were performed with MHC-I or MHC-II pooled peptides, or each individual peptide. After a 20-hour stimulation, numbers of spot-forming-cells (SFC) were read and mean triplicate SFC numbers were calculated, after background subtraction, with a positivity threshold of 50 SFC/10^6 cells[12,13].

Gene expression and cell type abundance profiling

Tumors were digitally quantified for immune genes with a targeted gene panel including T cell function, and positive and negative immune regulation (**Supplementary materials**). Immune deconvolution analysis was performed with CIBERSORTx[34] version, using expression matrix estimating 22 immune cell types using 547 signature genes (LM22).[35]

Whole exome and RNA sequencing, MHC-I and II restricted neoepitope prediction, differential gene expression, gene ontology enrichment analysis, T cell receptor analysis and statistical analysis are detailed in **Supplementary materials**.

RESULTS

Patient characteristics

Sixty-eight patients with NHL were enrolled, including 51 immunodeficient (ID) patients, either transplant recipients (N=34 PTLD) or HIV-positive (N=17), and 17 immunocompetent (IC) patients with diffuse large B cell lymphoma (DLBCL) (**Table 1**). The median age was 57 years and 72% of patients were male. Diseases were systemic in 75% and cerebral in 25% cases. Two patients with both systemic and CNS localization were classified as systemic disease. The most frequent lymphoproliferative disorder subtypes were Large B Cell

Lymphoma (LBCL) (72%: 34 DLBCL not otherwise specified (NOS) + 12 Primary CNS Lymphoma (PCNSL) + 2 DLBCL NOS HHV8+ and 1 HGBL NOS), followed by polymorphic lymphoproliferation (9%), Burkitt lymphoma (BL) and plasmablastic lymphoma (PL) (6% each). The tumors were EBV-positive in 40% cases, all in the ID group (47% and 65% in PTLD and HIV+ patients, respectively). Both CNS and systemic localizations were equally distributed between EBV-positive and negative NHL. All but one CNS NHL from ID patients were EBV-positive. Among the eight EBV+ tumors tested for EBV antigen expression and latency status, three tumors displayed latency I, 2 patients displayed latency II and 3 other ones displayed latency III status. No significant differences were observed between EBV-positive and -negative NHL regarding time from transplantation to NHL diagnosis (6.9 [IQR:2.85-13.11] vs 10.2 [IQR:8.80-20.41] years, p=0.06), median CD4 counts at NHL diagnosis (266 and 222/mm^3), time from diagnosed HIV infection to NHL diagnosis (4.1 [IQR:2.49-21.53] *vs* 21.6 [IQR:4.92-23.94] years, p=0.7) and overall survival (OS) (NA and 5.3 years, p=0.21) (**Supplementary Figure 1**). The median follow-up from NHL diagnosis was 1.8 years.

The tumor mutational burden is lower in EBV-positive compared to EBV-negative NHL

We investigated the TMB in a WES analysis of paired tumor and normal samples (**Methods in Supplementary data; Supplementary Figure 2**). The median TMB was 3.01/Mb [IQR : 1.81-5.06] similar to the previously described burden for DLBCL patients (**Supplementary Figure 3**)[36]. However, the median TMB was lower in the EBV-positive compared to EBV-negative NHL, both among the whole (ID+IC) NHL population (2.2 *versus* 3.4/Mb, p=0.001) and among the ID only NHL (2.2 *versus* 4.6/Mb, p=0.0075) (**Figure 1.A**). Similarly, when restricting the analysis to the predominant LBCL subgroup (n=49), a lower median TMB was observed in EBV-positive compared to EBV-negative LBCL both among the whole (ID+IC)

population (2.2 *versus* 3.4/Mb, p=0.00071) (**Figure 1.A**) and among the ID patients only (1.9 *versus* 4.9, p=0.00078) (**Supplementary Figure 4.A**). There were no TMB differences according to immune status when comparing IC, PTLD and HIV patients both in the overall population (**Figure 1.B**) and in the EBV-negative one (**Supplementary Figure 4.B)**. By contrast, the TMB differed slightly according to disease localization (3.2 in systemic *versus* 2.3 in CNS disease, p= 0.046) (**Figure 1.C)**. Importantly, a higher TMB was associated with a significantly longer median OS in EBV-negative patients (OS=NA for TMB >3/Mb *vs* 2.2 years for TMB <3/Mb, p=0.01) while there was no difference in EBV-positive ones (**Figure 1.D**) (**Supplementary methods** for the determination of the TMB cut point).

EBV, immune status and disease localization impact the mutational profile of NHL.

Next, we deciphered the mutational profile of these lymphomas and observed some frequently mutated genes such as *PCLO* (25% cases), *CSMD3* (24%), *TP53* (24%), *FAT4* (19%) and *MYD88* (16%) as expected (**Supplementary Figure 5.A)**. However when focusing on the LBCL we observed a dysbalance in the mutational profile depending on both EBV and patients immune status. Indeed in EBV-positive LBCL the most frequently mutated genes were: *TYW1* (27%) and *STAT3, HNRNPF, TNRC18, PARP10* and *KDM3A* (20% each), while *TP53* (35%), *MYD88* (32%), *CD79B* (24%) and *FOXQ1, STAT3, KRT3* and *ZNF574* (15% each) were the most frequently mutated in EBV-negative ones (**Figure 2.A and Supplementary Table 1)**. Furthermore, some mutations were observed only in EBV-positive NHL, such as *HNRNPF* (20%), or only in EBV-negative ones, such as *MYD88* and *CD79B* mutations (32 and 24% respectively). Finally, *TP53* mutations were observed almost exclusively in EBV-negative NHL (35% vs 7%, p=0.001). After false discovery rate (FDR) correction, no significant differences remained yet.

The immune status (IC or ID) co-influenced this tumor profile with *TP53* present in 34% of ID-LBCL (vs 12% in IC-DLBCL), all but one of which were in the EBV-negative group. Furthermore, the *STAT3*, *TYW1* and *KDM3A* mutations were found only in ID-LBCL (25%, 25% and 19%, respectively), regardless of EBV status, while they were absent in IC-DLBCL. **(Figure 2.B and Supplementary Table 2).** As expected, the disease localization had also an impact, with *TP53* mutations found only in systemic diseases (37% *vs* 0%, p=0.009, fdr=0.03) and *MYD88* and *CD79B* mutations being more frequent in CNS than in systemic diseases (50% *vs* 11%, p=0.006, fdr=0.03, and 43% *vs* 6%, p=0.04, fdr=0.03, respectively) (**Supplementary Figure 5.B**). Finally, the 6 polymorphic lymphoproliferations (5 EBV+ and 1 EBV-) showed a distinct mutational profile with *PCLO* as the most recurrently mutated gene (**Supplementary Figure 5.C**). Within the 4 BL, the 2 genes recurrently mutated were: *MYC* and *IGLL5* (**Supplementary Figure 5.D**).

In addition, we identified significant recurrent Somatic Copy Number Alterations (SCNAs) on the WES data (Methods in **Supplementary data**). When focusing on the 47 LBCL, we successfully identified a total of 13 arm-level and 12 focal regions displaying copy gain, along with 5 arm-level and 6 focal regions with copy loss (arm q value $\leq$ 0.1, focal regions q value $\leq$ 0.25 **Supplementary Figure 6.A.**). These SCNAs observed frequencies ranged from 2% to 55%. In IC DLBCL, SCNAs encompassed amplifications of 1q, 18q, 21q, 7q22, 8q24, and 19q13, and deletions involving 6q, 17p, 1p36, and 1p13, consistently with prior reports [37]. Additionally, we observed SCNAs in genes previously reported in DLBCL [38], including *PIK3CA* (25%), *PRDM1* (38%), *MYC* (19%), *CDKN2A* (21%), *PTEN* (14%), *ETV6* (42%), *STAT6* (42%), *beta-2 microglobulin* (*B2M)* (23%), and *TP53* (42%) (**Supplementary Table 3**). Furthermore, we uncovered previously unreported amplifications of chromosome 16 in 34% of LBCL patients in our dataset.

Notably, factors influencing those SCNAs included the EBV status with focal amplification peaks, in EBV-positive cases, of the 9q34.13, a region encompassing the *NOTCH1* and *JAK2* genes (**Supplementary Figure 6.B**). The immune status as well influenced SCNAs with more significantly frequent amplification at 1q22 (33%;q-value=0.0185/29%;q-value> 0.25), 8q24.3 (26%;q-value=0.04/17% q-value > 0.25), 9q34.3 (47%;q-value=0.0008/24%;q-value > 0.25), 11q13.1 (43%;q-value=0.0032/23% ;q-value>0.25), 17q25.3 (36%;q-value=0.0219/29%;q-value>0.25), 22q13.31 (33%;q-value=0.1577/5% q-value > 0.25) genomic regions in the ID-LBCL vs IC-DLBCL groups. Additionally, some deletions were prevalent at 1p36.32 (23%;q-value=0.0361/13%;q-value>0.25) and 15q12 (36%;q-value=0.0018/29% q-value > 0.25) (**Supplementary Figure 6.C**), affecting genes like IL10, MYC, CD45, IRF8, IL17RA, CDKN2C, among others. Systemic localizations also played a role with more frequent deletions affecting the human-leukocyte-antigen (HLA) locus (6p21.31;9%;q-value=0.22//0%;q-value >0.25), BCL6 (3q29; 19%;q-value=0.19/0%;q-value >0.25) and Immunoglobulin genes (15q12; 37%;q-value=0.005/28%;q-value >0.25), together with amplifications of the 8q24 locus (33%;q-value=0.096/%;28%;q-value >0.25) **Supplementary Figure 6.D.** Systemic localizations also played a role with more frequent deletions affecting the human-leukocyte-antigen (HLA) locus (6p21.31), *BCL6* (3q29) and Immunoglobulin genes (15q12), together with amplifications of the 7q22 and 8q24 locus (**Supplementary Figure 6.D**).

The EBV status influences the tumor immunogenicity dominated by MHC class-II restricted neoepitopes

We first predicted the neoepitopes derived from non-Ig tumor variants by including parameters from the RNA sequencing. In the 30 tumors with available RNAseq (Methods in **Supplementary data; Supplementary Figure 2.B**), we found a median of 149 neoepitopes

per tumor, with lower numbers in EBV-positive compared to EBV-negative NHL (40 *vs* 200, p=0.00016), independently of the immune status (**Figure 3.A**). These values were not influenced by the expression levels of the MHC molecules and B2M that did not differ between EBV-positive and negative NHL, ID and IC patients and systemic *versus* CNS localization (**Data not shown)**. Furthermore, very few down-regulations (RPKM <1) of HLA class I and class II were observed (3% and 9% of tumors, respectively) and no B2M down-regulation or mutations were observed despite the presence of heterozygote deletions in four tested patients. Most neoepitopes were predicted to be presented by MHC class-II molecules with a median class-II/ class-I CMH ratio of 4, independently of the EBV status.

As RNAseq was not available in more than half tumors, we conducted the same analysis solely based on WES. Similarly, we observed lower numbers of neoepitopes in EBV-positive compared to EBV-negative tumors (159 *versus* 501, p=0.0024) (**Figure 3.B**), with a highly significant positive correlation between WES and RNA-based results (r=0.93, p<0.001) (**Figure 3.C**). Importantly, the neoepitopes numbers were positively correlated to the TMB (r=0.8, p<0.001) (**Figure 3.C**). Furthermore clonal-derived neoepitopes accounted for 89% of the entire WES predicted neoepitopes with significantly higher neoepitopes numbers originating from clonal compared to subclonal populations (median per patient: 301 and 32, respectively, p=0.02) (**Supplementary Figure 7**).

Then, we analyzed the neoepitopes derived from Ig variable regions genes by using the available RNAseq data from the FR3-CDR3-FR4 region. A dominant tumor Ig clonotype (≥15%) was found in 24 NHL patients, with frequencies ranging from 15% to 100% (median 81%). A median of 15 neoepitopes derived from the tumor heavy chain variable region genes was found, independently of the EBV (15 and 18 for EBV-positive and negative samples, respectively) or immune status (15 and 15 for ID and IC samples, respectively) (**Figure 3.D**).

These Ig-derived neoepitopes were also predicted to be more frequently presented by the MHC class II, with an MHC class-II/class-I ratio of 1.5.

Circulating T-cells specific for tumor neoepitopes, including *Ig*-derived neoepitopes, are detectable in 71% of NHL

To confirm that the predicted NHL neoepitopes could drive effective anti-tumor immunity in ID patients, we analyzed the peripheral blood T cell responses against the corresponding peptides in 14 samples (13 ID and 1 IC patient). We tested a similar number of variant-derived peptides from EBV-positive and -negative tumors with a median of 48 peptides per patient. Neoepitope-specific responses were detectable in 10 out of the 14 (71%) patients samples tested, among whom 93% were ID and 50% had CNS disease. These neoepitope-specific responses were more frequently detectable, though non significantly, in the EBV-negative patients, as observed among all 6 (5 ID + 1 IC, 100%) tested cases, than in EBV-positive ones as observed in only 4 of the 8 (50%) ID cases tested (p=0.08). The level of T cell responses to tumoral neoepitopes did not change with the EBV antigen (EBNA-2, LMP1 and EBER) expression. The median magnitude of positive T-cell responses was 746 SFC/10^6 cells per patient's peptide pool, and above 500 SFC/10^6 cells in the 7 ID patient samples (**Figure 4 and Supplementary Figure 8)**. The magnitudes above SFC >500/10^6 cells were associated with a trend towards higher, though non significantly different, time from transplantation or HIV diagnosis to NHL diagnosis and CD4 count in HIV patients, (12 *vs* 8 years, 27 *vs* 9 years, 135 vs 1087/mm^3, respectively, p>0.4).

The *Ig*-derived neoepitopes were tested in 5 cases with a median number of 22 tested neoepitopes per patient (16 MHC-II and 8 MHC-I restricted) and were recognized in 3 cases with a 1.5- to 3-fold higher magnitude (median 697 SFC/10^6 cells) than against autologous

non-*Ig* neoepitopes. All recognized Ig epitopes were localized in the heavy chain region and predicted to be MHC-class II restricted.

Finally, there were no neoepitopes shared between tumors, although some derived from frequently mutated genes (in ≥10% NHL) such as *FAT4*, *SETD1B*, *ZFP36L1*, *HMCN1*, *MTMR1*, *UBR4*, *KLF2*, *RBMX1*, *ALMS1* and *C2ORF49*. These tumor-derived neoepitopes were not found in the public immune epitope database (IEDB) for T cell epitopes from pathogens and autoimmunity.

EBV drives the tumor microenvironment.

These immunogenicity results prompted us to analyze the intra-tumoral TCR repertoire in RNA seq data of 27 patients' samples with sufficient numbers of TCRβ reads (≥100). The EBV-positive NHL contained higher numbers of unique productive clonotypes than the EBV-negative ones (412 *versus* 134, p=0.003) (**Figure 5).** These numbers were also higher among ID NHL compared to the IC ones, both in the EBV-negative NHL (157 *versus* 90, p=0.04) and in the overall population (**Supplementary Figure 9**). In contrast, the repertoire diversity was similar between the NHL subgroups (**Supplementary Figure 10**) without any shared dominant clonotypes between patients. Finally, the TCR sequences with clone frequencies >10%, suggestive of tumor-neoantigen selection, did not show any evidence of known antigen specificity using the VDJ database.

We then analyzed the TME immune cell composition to determine whether such higher TCR abundance in ID NHL was associated with selected T cell populations (**Supplementary Figure 11**). The proportions of CD8+ T cells, regulatory T cells (Tregs), resting NK cells, type-M1 and M2 macrophages and monocytes were significantly higher in EBV-positive compared to EBV-negative NHL both in the whole cohort (IC+ID) and tended to be higher in

the ID cohort only, but were not influenced by the immune status or disease localization (**Figure 6.A and Supplementary Figure 12, 13 and 14)**. Conversely, the proportion of B-cells was decreased in EBV-positive NHL compared to EBV-negative ones (14 vs 52%, p=0.000008) (data not shown). In EBV-positive NHL, there was a positive correlation between the proportions of CD8 T cells and of activated memory CD4 T cells, and the TMB (r=0.79, p=0.048; r=0.96, p=0.024, respectively) (**Figure 6.B**). A similar trend of association was observed among EBV-negative cases, though not statistically significant.

Finally, we studied the expression of a selected customized panel of relevant genes involved in intra-tumoral immune responses. Whereas there was no difference for genes involved in positive immune regulation or T-cell lymphocyte function, a higher non-significant expression of genes involved in negative immune regulation was observed in EBV-positive NHL samples (**Figure 7**).

DISCUSSION

Our comprehensive large study of the immunogenomics of NHL occurring in ID patients revealed that the presence of EBV influences both the tumor mutational burdens and profiles, along with the immune status, but also influences the tumor variant immunogenicity and the microenvironment. Altogether, a limitation of our study may be the size of our series of NHL, particularly in the context of RNA sequencing analysis (n=30 sample) for MET studies and ELISPOT analysis (n=14 samples) for neoepitopes-specific T cell responses studies, though the largest so far to be studied at the unbiased immunogenomics level. In addition, some heterogeneity in the lymphoproliferative disorders studied might have confounded some analyses given that the major lymphoma subgroup of LBCL and the systemic localization constituted respectively only 72% (n=49) and 75% (n=51) of the entire cohort. The median

onset of EBV-positive PTLD among transplant recipients seemed to be quite later than usually described and may reflect the evolving use of the immunosuppressive regimens and the aging of the patients[5,39–41].

Our unbiased WES-based analysis allowed us to demonstrate a lower TMB in EBV-positive NHL of immunodeficient patients compared to EBV-negative ones, confirming previously suggested data obtained by targeted approaches[6–8,42]. Our data further indicate that the strong viral signaling present in EBV-positive NHL reduces the need for other driver tumor mutations in lymphomagenesis, in accordance with the concept of EBV alone acting as a strong oncogene in infected B cells, especially in the context of low immune pressure[5]. In line with this hypothesis, we showed that the numbers of tumor neoepitopes determined by these mutations were also lower in EBV-positive NHL, suggesting that each additional mutation increases the probability of generating a significant neoantigen, thereby contributing to the overall tumor immunogenicity.

Additionally, the presence of EBV also drives specific mutational profiles. The finding that *HNRNPF* was significantly mutated only in EBV-positive NHL might be in accordance with data showing that other ribonucleoproteins from the same HNRNPK family bind to the EBV nuclear antigen-2 (EBNA-2) and enhance viral LMP2A antigen expression[43]. We also confirmed previous targeted approaches reporting the almost complete lack of *MYD88*, *CD79B* or *TP53* mutations in EBV-positive NHL. More importantly, our WES approach enabled us to show that the immunodeficient status was associated with *STAT3* and *TYW1* mutations regardless of the EBV status, with *STAT3* being mutated only among ID patients, both HIV-positive and transplant recipients. The STAT3 protein is involved in a key signaling pathway modulating multiple physiological processes and inflammatory responses, especially in the B-cell lineage. *STAT3* mutations, which are rare in IC EBV-negative DLBCL had been reported in half EBV-positive diseases, such as PL or polymorphic lymphoproliferations,

particularly in HIV-positive patients[44–47]. However, data for *STAT3* mutations in PTLD patients are limited[48]. Our findings suggest for the first time that chronic immune stimulation and/or inflammation in these two immunodeficient settings associated with permanent antigenic stimuli (allogeneic transplant or HIV) might favor such *STAT3* mutations, thus highlighting the potential of inhibitors of JAK-STAT signaling as a promising treatment option. In addition, STAT3 mutations have been reported very recently to be linked with a "hot" microenvironment in PCNSL[49].

The prediction and validation of the most relevant neoepitopes were allowed by our development of a robust bioinformatic method based on tumor DNA and RNA sequencing, along with *in silico* algorithms and inferred rules for tumor neoepitope immunogenicity. In order to consider all steps of the antigen processing required for successful neoepitope selection[50], we included factors in our pipeline that are involved in the presentation machinery, such as the HLA molecules and B2M expression, in addition to key parameters for effective anti-tumor immune responses[20,51,52]. The strong correlation observed between neoepitope numbers selected from WES and RNA sequencing should facilitate future routine use. In addition to these non-viral neoepitopes, we showed, as previously suggested in IC patients[29,30], that NHL *Ig*-derived neoantigens contain immunodominant neoantigens, mostly presented by MHC-class II, in these ID patients. Finally, as EBNA-2, the immunodominant EBV antigen was expressed in only three of the eight EBV+ tumors evaluated, and as this or the latency state I or II status did not change the neoepitopes-specific immune response in the tumors, these data, though limited, suggest the EBV antigen expression does not modify tumor neoepitopes-specific immune response. Yet it will be essential in the future to characterize both the anti-EBV and anti-neoepitopes immune responses in order to better assess the burden of EBV.

The higher T cell infiltrate observed in the TME analysis of EBV-positive NHL might be in accordance with the strong EBV immunodominance, regardless of the immune status and without evidence for enrichment in specific TCR clonotypes. This T cell microenvironment was composed of both CD8 and activated CD4 s. In addition, a tolerogenic profile, composed of Tregs, type M2 macrophages and a higher expression of negative immune regulation molecules, tended to predominate in EBV-positive NHL, thus confirming that EBV might promote a more tolerogenic TME[16,17].

In conclusion, our exhaustive analysis of the immunogenomics characteristics of NHL occurring in immunodeficient patients shows the major influence of EBV on tumor mutational burden and profile and on tumor neo-antigenicity. Despite the lack of frequent or public NHL driver mutations and neoepitopes preventing the development of shared immune strategies as in other cancers[26,27], the existence of T cell responses in these immunodeficient contexts, directed against non-viral tumor neoepitopes, and particularly against *IgH* ones in ID patients, could pave the way for the development of future Ig-based immune therapies.

REFERENCES

1. Engels EA. Cancer in Solid Organ Transplant Recipients: There Is Still Much to Learn and Do. Am J Transplant. 2017;17(8):1967-1969.

2. Hernández-Ramírez RU, Shiels MS, Dubrow R, Engels EA. Cancer risk in HIV-infected people in the USA from 1996 to 2012: a population-based, registry-linkage study. Lancet HIV. 2017;4(11):e495-e504.

3. Blosser CD, Haber G, Engels EA. Changes in cancer incidence and outcomes among kidney transplant recipients in the United States over a thirty-year period. Kidney Int. 2021;99(6):1430-1438.

4. Carbone A, Vaccher E, Gloghini A. Hematologic cancers in individuals infected by HIV. Blood. 2022;139(7):995-1012.

5. Dharnidharka VR, Webster AC, Martinez OM, Preiksaitis JK, Leblond V, Choquet S. Post-transplant lymphoproliferative disorders. Nat Rev Dis Primers. 2016;2(1):15088.

6. Finalet Ferreiro J, Morscio J, Dierickx D, et al. EBV-Positive and EBV-Negative Posttransplant Diffuse Large B Cell Lymphomas Have Distinct Genomic and Transcriptomic Features: Genetics of PTLD. Am J Transplant. 2016;16(2):414-425.

7. Gandhi MK, Hoang T, Law SC, et al. EBV-associated primary CNS lymphoma occurring after immunosuppression is a distinct immunobiological entity. Blood. 2021;137(11):1468-1477.

8. Menter T, Juskevicius D, Alikian M, et al. Mutational landscape of B-cell post-transplant lymphoproliferative disorders. Br J Haematol. 2017;178(1):48-56.

9. Yoon H, Park S, Ju H, et al. Integrated copy number and gene expression profiling analysis of epstein-barr virus-positive diffuse large b-cell lymphoma. Genes Chromosomes Cancer. 2015;54(6):383-396.

10. Kaulen LD, Denisova E, Hinz F, et al. Integrated genetic analyses of immunodeficiency-associated Epstein-Barr virus- (EBV) positive primary CNS lymphomas. Acta Neuropathol. 2023;146(3):499-514.

11. Guney E, Lucas C-HG, Singh K, et al. Molecular profiling identifies at least 3 distinct types of posttransplant lymphoproliferative disorder involving the CNS. Blood Adv. 2023;7(13):3307-3311.

12. Nakid-Cordero C, Arzouk N, Gauthier N, et al. Skewed T cell responses to Epstein-Barr virus in long-term asymptomatic kidney transplant recipients. PLoS One. 2019;14(10):e0224211.

13. Nakid-Cordero C, Choquet S, Gauthier N, et al. Distinct immunopathological mechanisms of EBV-positive and EBV-negative posttransplant lymphoproliferative disorders. Am J Transplant. 2021;21(8):2846-2863.

14. Green MR, Rodig S, Juszczynski P, et al. Constitutive AP-1 Activity and EBV Infection Induce PD-L1 in Hodgkin Lymphomas and Post-transplant Lymphoproliferative Disorders: Implications for Targeted Therapy. Clin Cancer Res. 2012;18(6):1611-1618.

15. Nicolae A, Pittaluga S, Abdullah S, et al. EBV-positive large B-cell lymphomas in young patients: a nodal lymphoma with evidence for a tolerogenic immune environment. Blood. 2015;126(7):863-872.

16. Keane C, Tobin J, Gunawardana J, et al. The tumour microenvironment is immuno-tolerogenic and a principal determinant of patient outcome in EBV-positive diffuse large B-cell lymphoma. Eur J Haematol. 2019;103(3):200-207.

17. Keane C, Gould C, Jones K, et al. The T-cell Receptor Repertoire Influences the Tumor Microenvironment and Is Associated with Survival in Aggressive B-cell Lymphoma. Clin Cancer Res. 2017;23(7):1820-1828.

18. Dierickx D, Habermann TM. Post-Transplantation Lymphoproliferative Disorders in Adults. N Engl J Med. 2018;378(6):549-562.

19. Li L, Goedegebuure SP, Gillanders WE. Preclinical and clinical development of neoantigen vaccines. Ann Oncol. 2017;28(Suppl_12):xii11-xii17.

20. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. Science. 2015;348(6230):69-74.

21. Yarchoan M, Johnson BA, Lutz ER, Laheru DA, Jaffee EM. Targeting neoantigens to augment antitumour immunity. Nat Rev Cancer. 2017;17(4):209-222.

22. Brown SD, Warren RL, Gibb EA, et al. Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. Genome Res. 2014;24(5):743-750.

23. Rizvi H, Sanchez-Vega F, La K, et al. Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. J Clin Oncol. 2018;36(7):633-641.

24. Rizvi NA, Hellmann MD, Snyder A, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science. 2015;348(6230):124-128.

25. Leidner R, Sanjuan Silva N, Huang H, et al. Neoantigen T-Cell Receptor Gene Therapy in Pancreatic Cancer. N Engl J Med. 2022;386(22):2112-2119.

26. Li F, Deng L, Jackson KR, et al. Neoantigen vaccination induces clinical and immunologic responses in non-small cell lung cancer patients harboring EGFR mutations. J Immunother Cancer. 2021;9(7):e002531.

27. Platten M, Bunse L, Wick A, et al. A vaccine targeting mutant IDH1 in newly diagnosed glioma. Nature. 2021;592(7854):463-468.

28. Sahin U, Derhovanessian E, Miller M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature. 2017;547(7662):222-226.

29. Khodadoust MS, Olsson N, Wagar LE, et al. Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. Nature. 2017;543(7647):723-727.

30. Khodadoust MS, Olsson N, Chen B, et al. B-cell lymphomas present immunoglobulin neoantigens. Blood. 2019;133(8):878-881.

31. Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods. 2015;12(5):380-381.

32. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics. 2016;32(4):511-517.

33. Jensen KK, Andreatta M, Marcatili P, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology. 2018;154(3):394-406.

34. Steen CB, Liu CL, Alizadeh AA, Newman AM. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. Methods Mol Biol. 2020;2117:135-157.

35. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453-457.

36. Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-421.

37. Chapuy B, Stewart C, Dunford AJ, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat Med. 2018;24(5):679-690.

38. Schmitz R, Wright GW, Huang DW, et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. N Engl J Med. 2018;378(15):1396-1407.

39. Luskin MR, Heil DS, Tan KS, et al. The Impact of EBV Status on Characteristics and Outcomes of Posttransplantation Lymphoproliferative Disorder. Am J Transplant. 2015;15(10):2665-2673.

40. Trappe RU, Dierickx D, Zimmermann H, et al. Response to Rituximab Induction Is a Predictive Marker in B-Cell Post-Transplant Lymphoproliferative Disorder and Allows Successful Stratification Into Rituximab or R-CHOP Consolidation in an International, Prospective, Multicenter Phase II Trial. J Clin Oncol. 2017;35(5):536-543.

41. Zimmermann H, Koenecke C, Dreyling MH, et al. Modified risk-stratified sequential treatment (subcutaneous rituximab with or without chemotherapy) in B-cell Post-transplant lymphoproliferative disorder (PTLD) after Solid organ transplantation (SOT): the prospective multicentre phase II PTLD-2 trial. Leukemia. 2022;36(10):2468-2478.

42. Ferla V, Rossi FG, Goldaniga MC, Baldini L. Biological Difference Between Epstein-Barr Virus Positive and Negative Post-transplant Lymphoproliferative Disorders and Their Clinical Impact. Front Oncol. 2020;10:506.

43. Gross H, Hennard C, Masouris I, et al. Binding of the Heterogeneous Ribonucleoprotein K (hnRNP K) to the Epstein-Barr Virus Nuclear Antigen 2 (EBNA2) Enhances Viral LMP2A Expression. PLoS One 2012;7(8):e42106.

44. Chapman JR, Bouska AC, Zhang W, et al. EBV-positive HIV-associated diffuse large B cell lymphomas are characterized by JAK/STAT (STAT3) pathway mutations and unique clinicopathologic features. Br J Haematol. 2021;194(5):870-878.

45. Sarkozy C, Hung SS, Chavez EA, et al. Mutational landscape of gray zone lymphoma. Blood. 2021;137(13):1765-1776.

46. Ramis-Zaldivar JE, Gonzalez-Farre B, Nicolae A, et al. MAPK and JAK-STAT pathways dysregulation in plasmablastic lymphoma. Haematologica. 2021;106(10):2682-2693.

47. Liu Z, Filip I, Gomez K, et al. Genomic characterization of HIV-associated plasmablastic lymphoma identifies pervasive mutations in the JAK-STAT pathway. Blood Cancer Discov. 2020;1(1):112-125.

48. Leeman-Neill RJ, Soderquist CR, Montanari F, et al. Phenogenomic heterogeneity of post-transplant plasmablastic lymphomas. Haematologica. 2022;107(1):201-210.

49. Hernández-Verdin I, Kirasic E, Wienand K, et al. Molecular and clinical diversity in primary central nervous system lymphoma. Ann Oncol.2023;34(2):186-199.

50. Garcia Alvarez HM, Koşaloğlu-Yalçın Z, Peters B, Nielsen M. The role of antigen expression in shaping the repertoire of HLA presented ligands. iScience. 2022;25(9):104975.

51. Wells DK, van Buuren MM, Dang KK, et al. Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. Cell. 2020;183(3):818-834.

52. De Mattos-Arruda L, Vazquez M, Finotello F, et al. Neoantigen prediction and computational perspectives towards clinical benefit: recommendations from the ESMO Precision Medicine Working Group. Ann Oncol. 2020;31(8):978-990.

Table 1: Patients characteristics at diagnosis

Patients characteristics	Overall NHL (n=68)	EBV-negative NHL (n=41)	EBV-positive NHL (n=27)	P
Median age at diagnosis(range)	57 (21-85)	59 (21-85)	57 (30-76)	0.2
Sex : M/F	49/19	28/13	21/6	0.42
Immune status : n (%) - Transplant recipients - HIV patients - IC patients	 34 (50) 17 (25) 17 (25)	 18 (44) 6 (15) 17 (41)	 16 (60) 11 (41) NA	0.234
Allograft type: n (%) - Kidney - Liver - Heart - Heart/lung	 12 (35) 12 (35) 9 (27) 1 (3)	 4 (22) 9 (50) 5 (28) 0	 8 (50) 3 (19) 4 (25) 1 (6)	0.09
Time from transplantation to NHL diagnosis: years, median (range)	9 (0.4-32)	10 (2-33)	7 (0.4-28)	0.06
Time from diagnosed HIV infection to NHL diagnosis: years, median (range)	9 (0-35)	22 (0.1-33)	4 (0-35)	0.7
HIV viral load at NHL diagnosis: cp/mL(range) (available for 14 HIV patients)	78 (<20-1700000)	121 (<20-1700000)	40 (<20-353000)	0.26
CD4 count at NHL diagnosis: /mm3 (range) (available for 14 HIV patients)	245 (24-1846)	223 (160-1846)	267 (24-596)	0.22
WHO classification: n (%) - LBCL* - Polymorphic lymphoma - Burkitt Lymphoma - Plasmablastic lymphoma - Marginal zone lymphoma - Plasmocytoma - Mantle cell lymphoma	 49 (72) 6 (9) 4 (6) 4 (6) 2 (3) 2 (3) 1 (1)	 34 (83) 1 (2) 2 (5) 1 (2) 1 (2) 1 (2) 1 (2)	 15 (55) 5 (19) 2 (7) 3 (11) 1 (4) 1 (4) 0	0.01
Cell of origin (DLBCL NOS) Available for - GC : n (%) - Non GC : n(%)	 43 24 (56) 19 (44)	 31 20 (65) 11 (35)	 12 4 (33) 8 (67)	0.09
EBV status: n(%) - Positive - Negative	 27 (40) 41 (60)	 0 41 (100)	 27 (100) 0	NA
Disease localization: n (%) - Systemic - Central nervous system	 51 (75) 17 (25)	 32 (78) 9 (22)	 19 (70) 8 (30)	0.57
Extranodal localization for systemic disease: n (%) Available for - Yes - No	 48 43 (90) 5 (10)	 37 34 (92) 3 (7)	 19 17 (90) 2 (10)	1

LDH above upper limit: n (%), available for 50 patients	37 (74)	20 (74)	18 (78)	0.15
Ann Arbor Stage for systemic disease: n(%) Available for: - I-II - III-IV	 46 10 (22) 36 (78)	 28 5 (18) 23 (82)	 18 5 (28) 13 (72)	0.4
Age-adjusted IPI for systemic disease: n (%) Available for: - 0-1 - 2-3	 44 14 (32) 29 (68)	 26 8 (31) 18 (69)	 17 6 (35) 11 (65)	1

Abbreviations: EBV, Epstein Barr Virus; NHL, non-Hodgkin lymphoma; IC, immunocompetent; ID, immunodeficient; M, male; F, female; LBCL, large B cell lymphoma; GC: germinal center, EBV, Epstein-Barr Virus; LDH, lactate deshydrogenase; IPI, international prognosis index; NA: not applicable. *: LBCL included 34 diffuse large B-cell lymphoma (DLBCL) not otherwise specified (NOS) + 12 primary central nervous system lymphoma (PCNSL) + 2 DLBCL NOS HHV8-positive + 1 high-grade B-cell lymphoma NOS.

FIGURE LEGENDS

Figure 1. The tumor mutational burden is lower in EBV-positive non-hodgkin lymphoma (NHL) compared to EBV-negative ones. Tumor mutational burden (defined as the number of mutations per Megabase; log_{10}) according to **A.** the EBV status among the 68 NHL patients on the left, the 51 immunodeficient NHL patients in the middle, and the 49 large B cell lymphoma (LBCL) patients on the right. Red and green color denote EBV-negative and positive NHL respectively. **B.** the immune status. Salmon, blue and yellow colors denote immunocompetent, transplant and HIV patients respectively. **C.** the disease localization. Pink and blue color denote systemic and central nervous system (CNS) localization respectively. *Wilcoxon Test.* **D and E**. Overall survival depending on TMB>3/Mb among EBV-negative (D) and positive (E) diseases. *Kaplan Meier*. **Abbreviations**: NHL: non-hodgkin lymphoma, LBCL: large B cell lymphoma, ID: immunodeficient, CNS: central nervous system..

Figure 2. The mutational landscape of non-hodgkin lymphoma (NHL) differs by EBV and immune status. A. Co-oncoplot of the most recurrently mutated genes (≥15%) within the 49 large B-cell lymphoma (LBCL) samples according to the EBV status (EBV-negative on the left, EBV-positive on the right). *TP53*, *MYD88* and *HNRNFP* were differently mutated between the 2 groups but these results lost their significant value after false discovery rate (FDR) correction was applied. **B** .Co-oncoplot of the most recurrently mutated genes (≥15%) within the 49 LBCL samples according to the immune status (immunocompetent on the left, immunodeficient on the right). *STAT3*, *TYW1* and *MYD88* were differently mutated between the 2 groups but these results lost their significant value after FDR correction was applied. *Fisher exact test* were used to compare categorical data. **Abbreviations** : ID: immunodeficient, IC: immunocompetent, PTLD: post-transplant lymphoproliferative disorder.

Figure 3. **The number of neoepitopes is lower in EBV-positive non-hodgkin lymphoma (NHL) compared to EBV-negative ones.** Number of predicted neoepitopes (log_{10}) **A.** from the non *Ig* variants within the 31 RNA samples, according to the EBV status on the left and the immune status (in EBV-negative NHL) on the right. *Wilcoxon test*. **B.** from the non *Ig* variants within the 66 whole exome sequencing (WES) samples (2 samples were excluded because of absence of germline assessment) according to the EBV status. *Wilcoxon test*. **C.** Correlation study between the number of predicted neoepitopes from the RNA and the WES data on the left, and between the number of predicted neoepitopes from the RNA data and the tumor mutational burden on the right. *Spearman correlation*. **D**. Number of predicted neoepitopes (log_{10}) from *Ig* variants within the 24 RNA samples (7 samples were excluded because of dominant *IgH* clone <15%). *Wilcoxon test*. **Abbreviations** : NHL: non-hodgkin lymphoma, IC: immunocompetent; ID: immunodeficient, WES: whole exome sequencing.

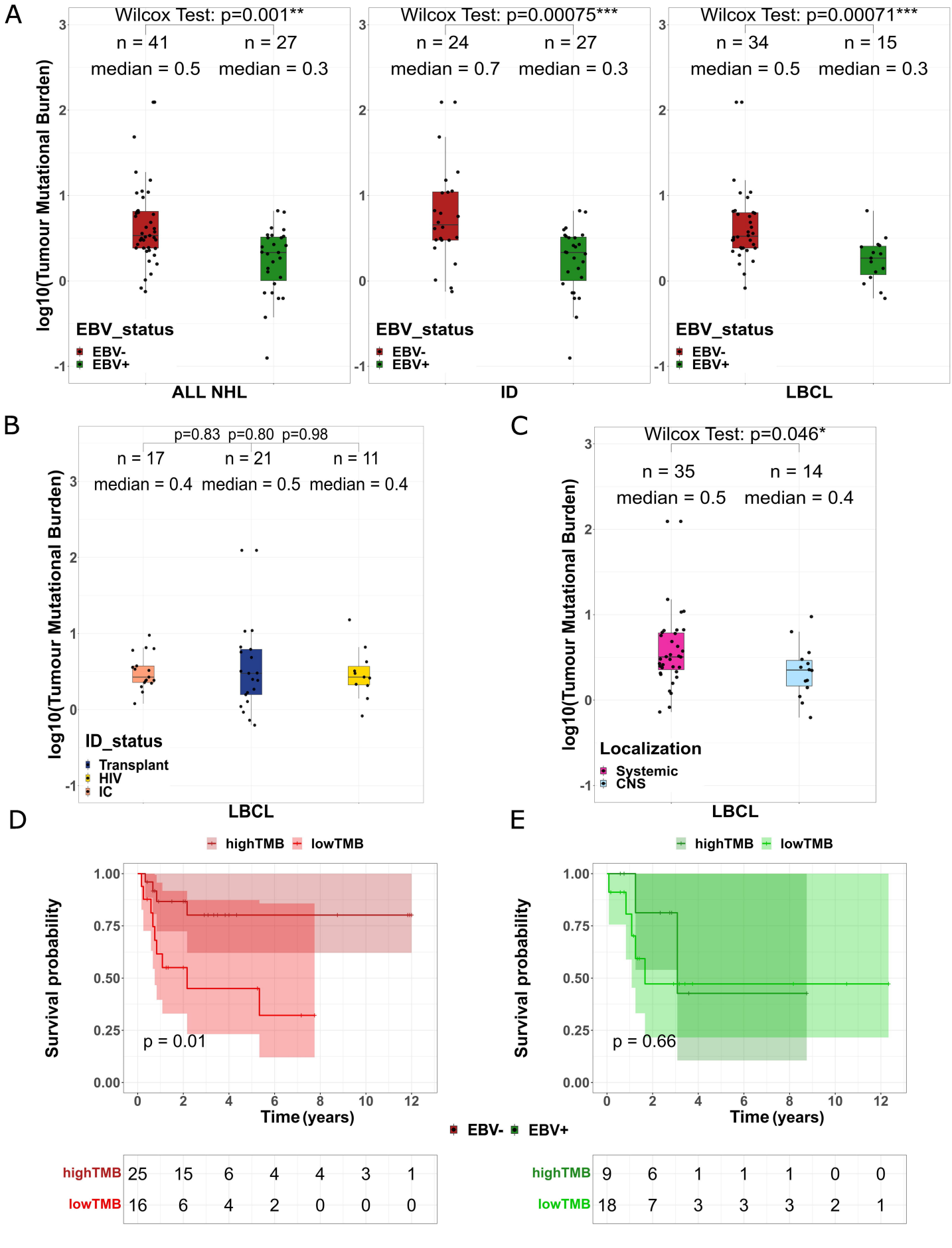
Figure 4. Neoepitopes-specific T cells are detectable among 71% cases, including responses directed against Ig-derived neoepitopes. Number of IFN-γ spots (/10^6 cells) after peptide stimulation for the 14 tested patients (represented on the *x* axis). Green squares denote responses directed against the complete pool, red circles denote responses directed against Ig-derived neoepitopes and black circle denote responses directed against individual non-Ig neoepitopes. Thawed peripheral blood mononuclear cells (PBMC) were co-cultured with personalized pooled peptides during 10 days and next tested for reactivity using IFN-γ enzyme linked immunoSpot (ELISPOT) assays. Patients were all tested for their personalized pooled peptides (named "complete pool") and eventually for each individual peptide if the number of cells were adequate (named as the mutated gene). The mean numbers of spot forming cells (SFC) from triplicates were normalized to number of IFN-γ spots detected per $1x10^6$ PBMC after background subtraction, and ELISPOT-IFN-γ positivity threshold was 50 SFC per million cells. **Abbreviations** : IC : immunocompetent, CNS : central nervous system,

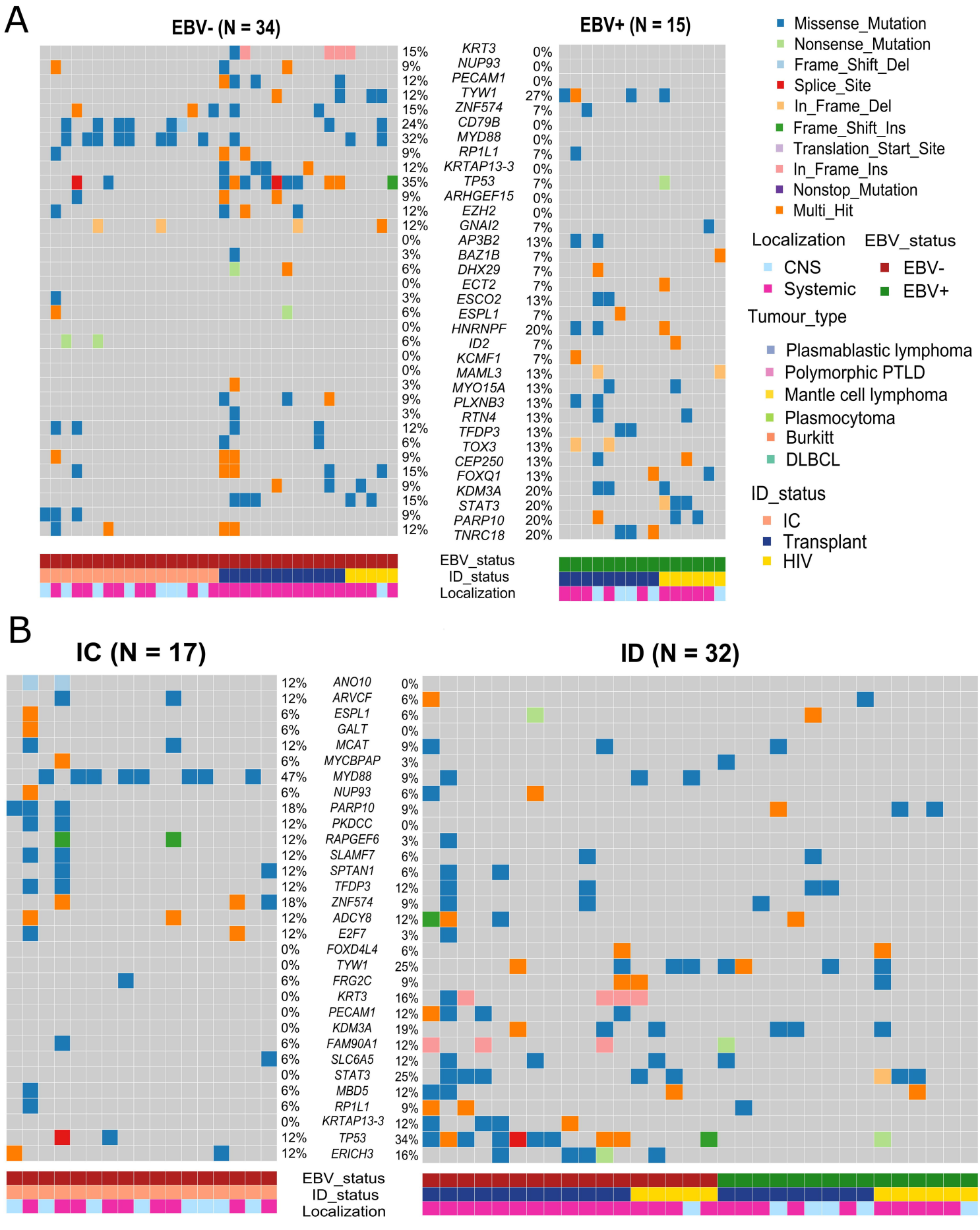
DLBCL : diffuse large B cell lymphoma, NHL : non-hodgkin lymphoma, BL: Burkitt lymphoma, PL: plasmablastic lymphoma, TMB: tumor mutational burden.

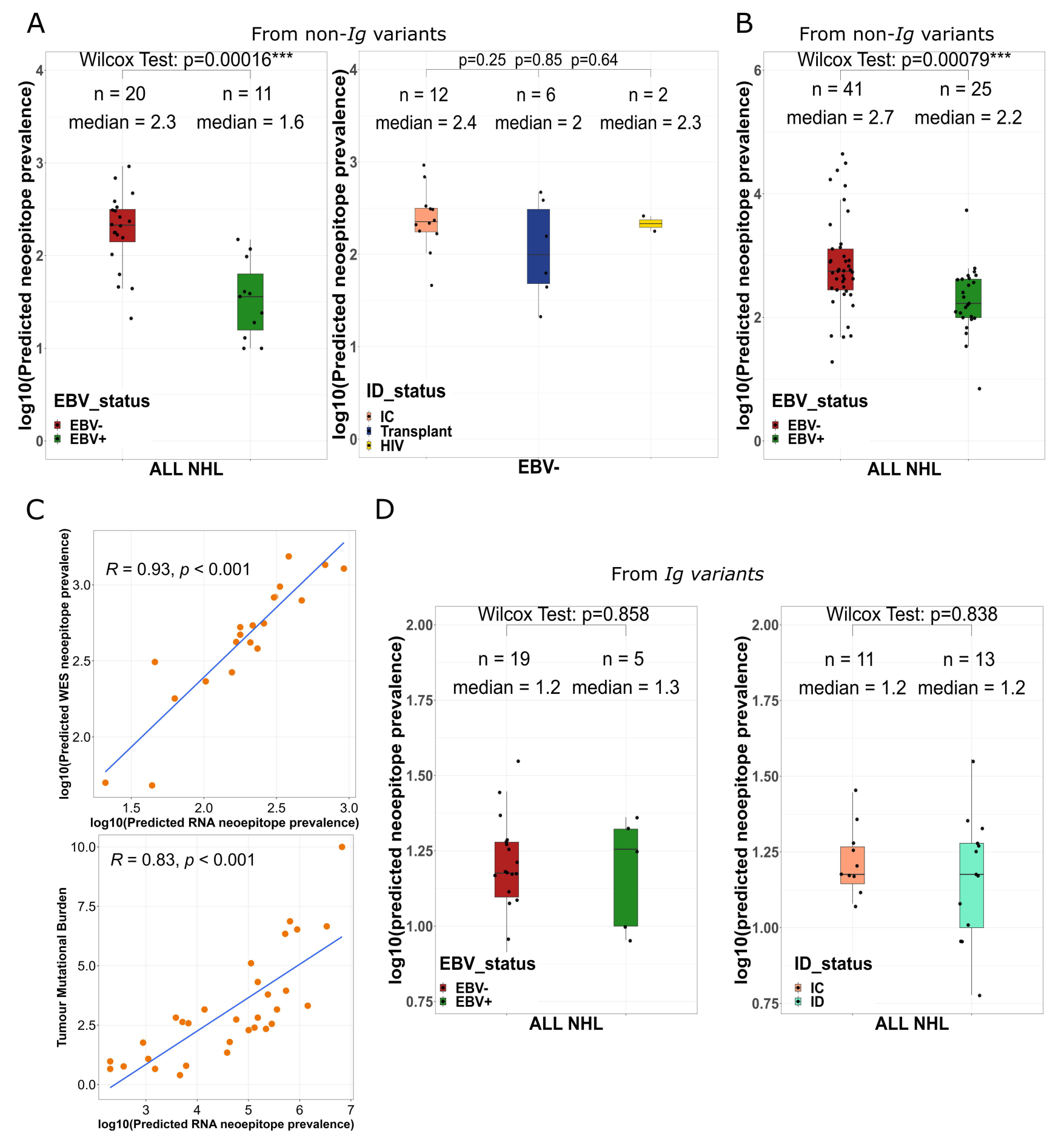
Figure 5. The intra-tumoral T cell receptor (TCR) repertoire diversity does not differ between EBV-positive and negative non-hodgkin lymphoma (NHL). Number of unique productive TCR- β clonotypes according to the EBV status (left), the immune status (middle) and the disease localization (right). *Wilcoxon Test.* **Abbreviations** : IC: immunocompetent; ID: immunodeficient; CNS: central nervous system, NHL: non-hodgkin lymphoma, TCR: T cell receptor.

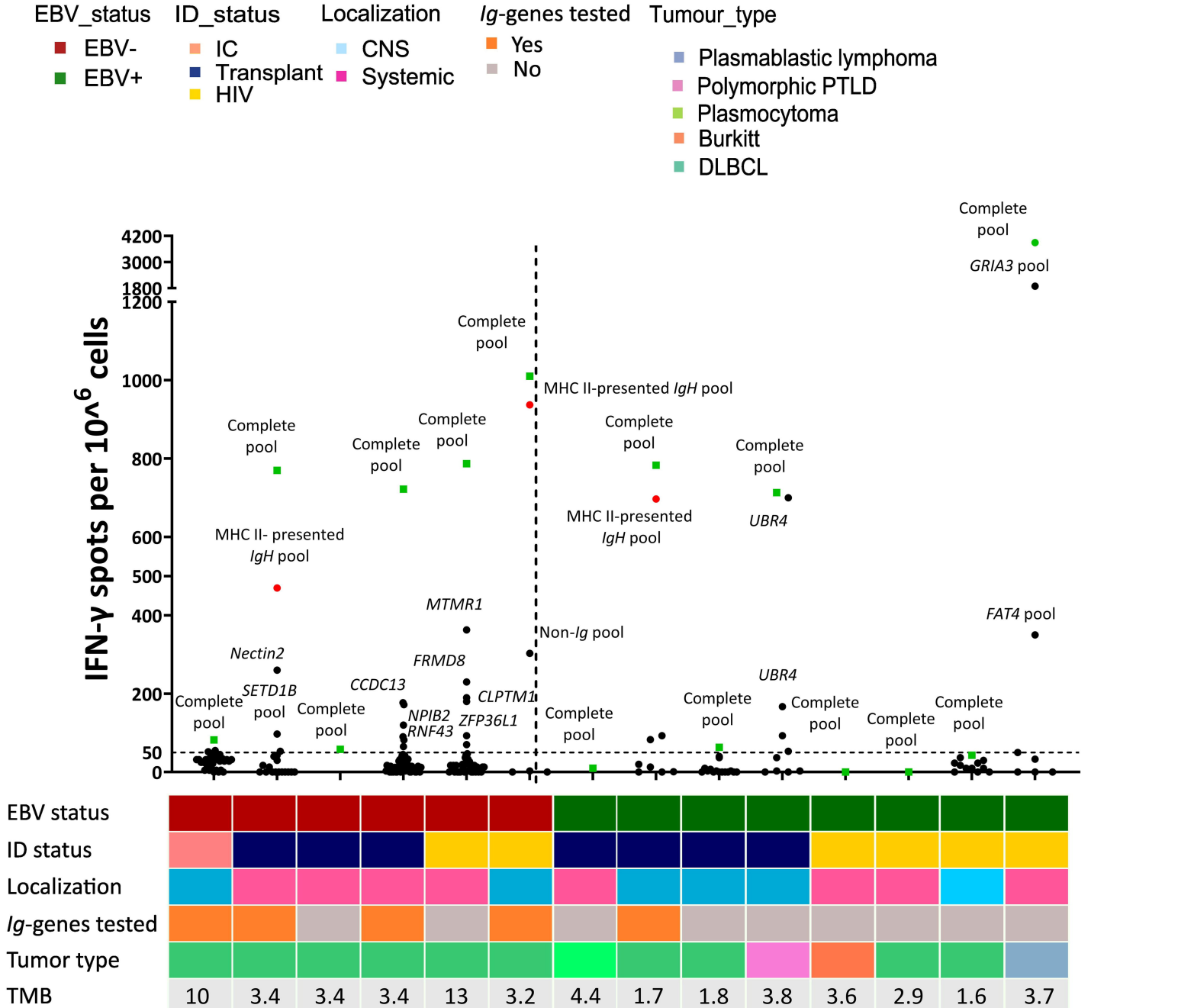
Figure 6. **EBV drives the tumor microenvironment in non-hodgkin lymphoma (NHL) in immune-suppressed and immune- competent patients. A.** Cell type abudance assessed with CIBERSORTx, according to the EBV status. *Wilcoxon Test.* **B**. Correlation study between tumor mutational burden (TMB) and memory resting CD4 T cell and CD8 T cell within EBV-positive NHL (green circles, upper) and EBV-negative NHL (red circles, lower). *Spearman correlation*. **Abbreviations** : NHL : non-hodgkin lymphoma, TMB : tumor mutational burden.

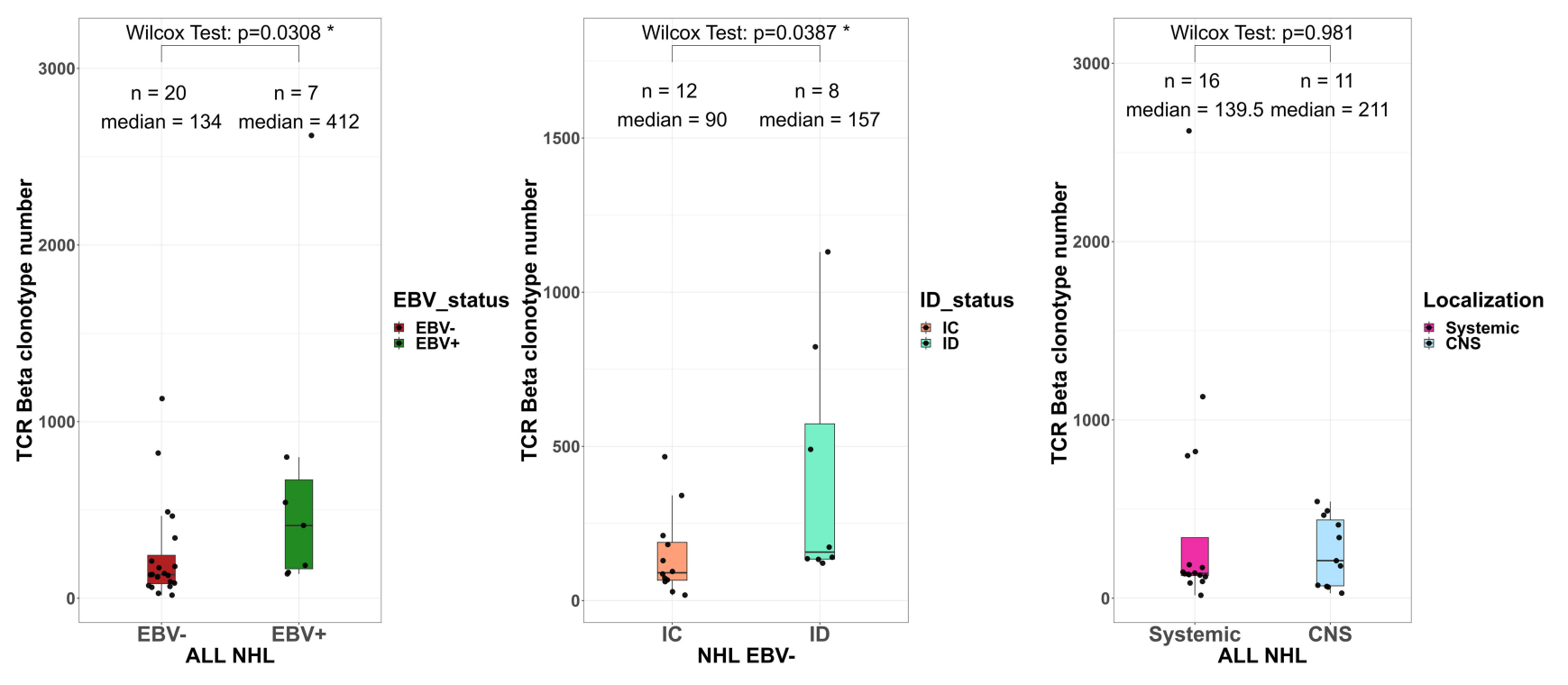
Figure 7. **The gene expression profiling tends to differ from EBV positive and negative non-hodgkin lymphoma (NHL).** Single sample gene set enrichment analysis (ssGSEA) scores of negative immune stimulation (upper panel), positive immune stimulation (middle panel) and T-cell function (lower panel). SsGSEA calculates separate enrichment scores for each pairing of a sample and gene set. Each ssGSEA enrichment score represents the degree to which the genes in a particular gene set are coordinately up- or down-regulated within a sample. *Wilcoxon test.* **Abbreviations** : non-hodgkin lymphoma, single sample gene set enrichment analysis : ssGSEA.

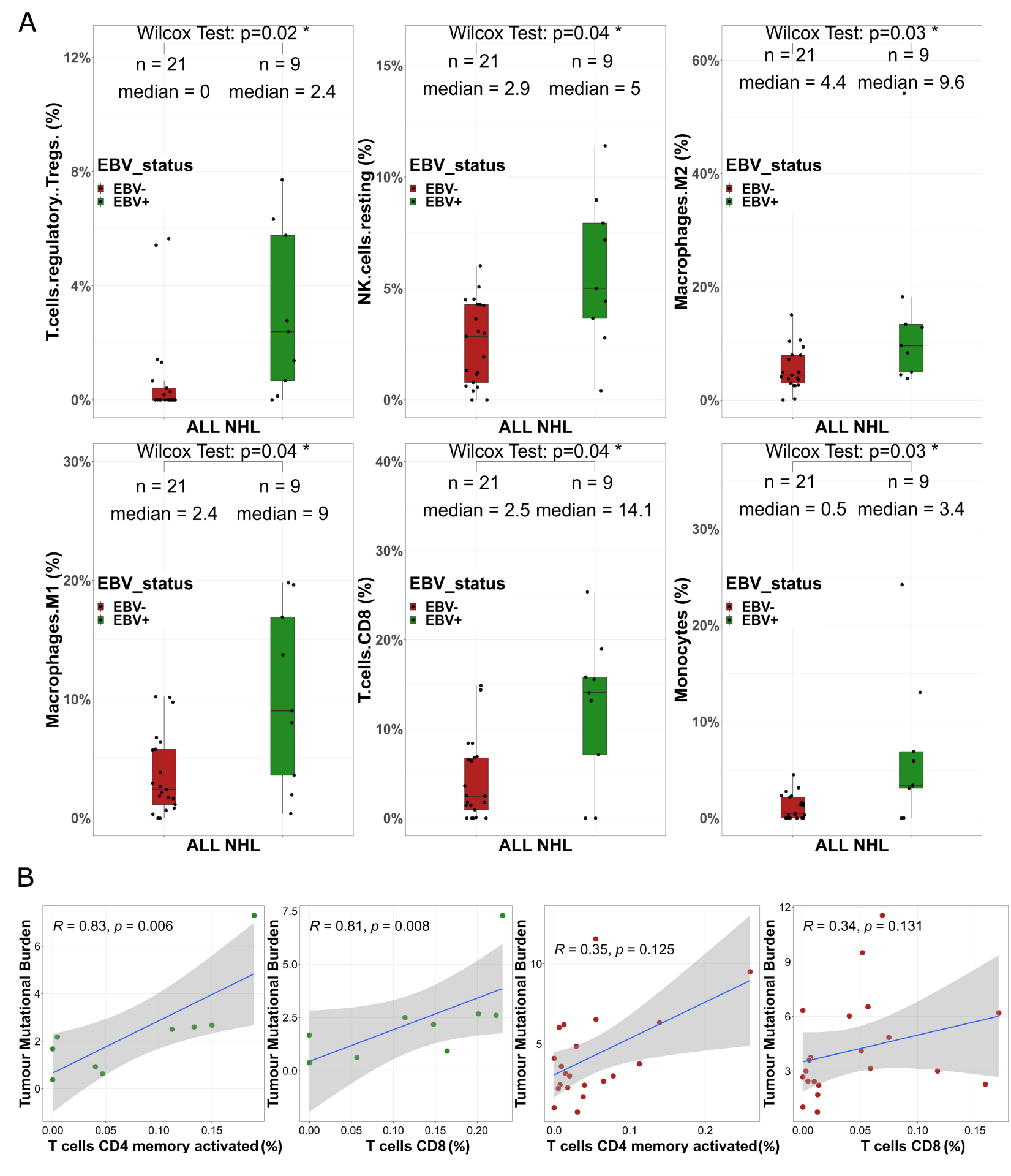

A
Wilcox Test: p=0.001**
n = 41
n = 27
median = 0.5
median = 0.3
Wilcox Test: p=0.00075***
n = 24
n = 27
median = 0.7
median = 0.3
Wilcox Test: p=0.00071***
n = 34
n = 15
median = 0.5
median = 0.3
log10(Tumour Mutational Burden)
EBV_status
EBV-
EBV+
ALL NHL
ID
LBCL
B
p=0.83 p=0.80 p=0.98
n = 17
n = 21
n = 11
median = 0.4 median = 0.5 median = 0.4
ID_status
Transplant
HIV
IC
LBCL
C
Wilcox Test: p=0.046*
n = 35
n = 14
median = 0.5
median = 0.4
Localization
Systemic
CNS
LBCL
D
highTMB
lowTMB
Survival probability
p = 0.01
Time (years)
highTMB 25 15 6 4 4 3 1
lowTMB 16 6 4 2 0 0 0
E
highTMB
lowTMB
Survival probability
p = 0.66
Time (years)
EBV-
EBV+
highTMB 9 6 1 1 1 0 0
lowTMB 18 7 3 3 3 2 1

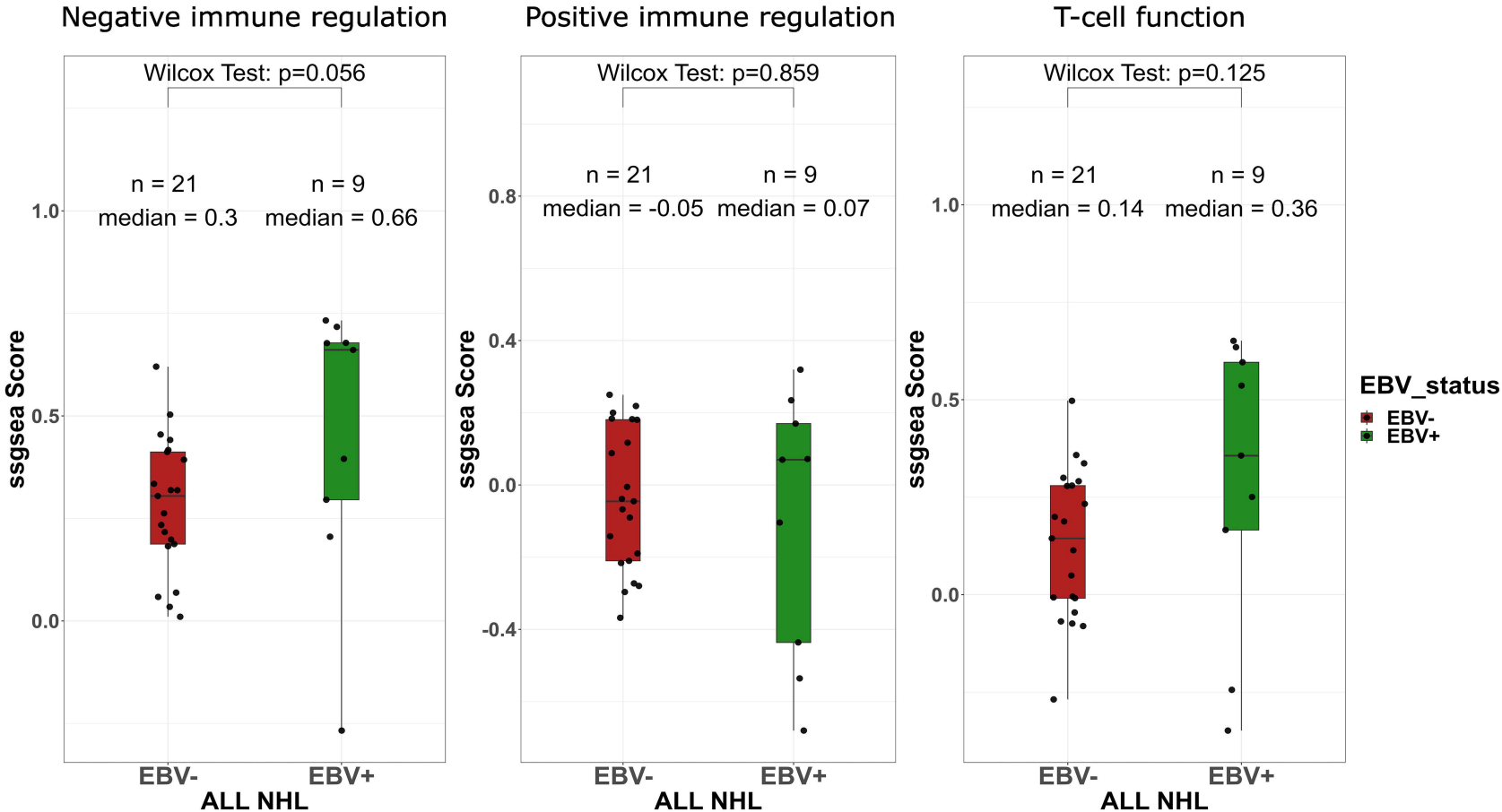

A
EBV- (N = 34)
EBV+ (N = 15)
15% KRT3 0%
9% NUP93 0%
12% PECAM1 0%
12% TYW1 27%
15% ZNF574 7%
24% CD79B 0%
32% MYD88 0%
9% RP1L1 7%
12% KRTAP13-3 0%
35% TP53 7%
9% ARHGEF15 0%
12% EZH2 0%
12% GNAI2 7%
0% AP3B2 13%
3% BAZ1B 7%
6% DHX29 7%
0% ECT2 7%
3% ESCO2 13%
6% ESPL1 7%
0% HNRNPF 20%
6% ID2 7%
0% KCMF1 7%
0% MAML3 13%
3% MYO15A 13%
9% PLXNB3 13%
3% RTN4 13%
12% TFDP3 13%
6% TOX3 13%
9% CEP250 13%
15% FOXQ1 13%
9% KDM3A 20%
15% STAT3 20%
9% PARP10 20%
12% TNRC18 20%
EBV_status
ID_status
Localization
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Del
Splice_Site
In_Frame_Del
Frame_Shift_Ins
Translation_Start_Site
In_Frame_Ins
Nonstop_Mutation
Multi_Hit
Localization
CNS
Systemic
EBV_status
EBV-
EBV+
Tumour_type
Plasmablastic lymphoma
Polymorphic PTLD
Mantle cell lymphoma
Plasmocytoma
Burkitt
DLBCL
ID_status
IC
Transplant
HIV
B
IC (N = 17)
ID (N = 32)
12% ANO10 0%
12% ARVCF 6%
6% ESPL1 6%
6% GALT 0%
12% MCAT 9%
6% MYCBPAP 3%
47% MYD88 9%
6% NUP93 6%
18% PARP10 9%
12% PKDCC 0%
12% RAPGEF6 3%
12% SLAMF7 6%
12% SPTAN1 6%
12% TFDP3 12%
18% ZNF574 9%
12% ADCY8 12%
12% E2F7 3%
0% FOXD4L4 6%
0% TYW1 25%
6% FRG2C 9%
0% KRT3 16%
0% PECAM1 12%
0% KDM3A 19%
6% FAM90A1 12%
6% SLC6A5 12%
0% STAT3 25%
6% MBD5 12%
6% RP1L1 9%
0% KRTAP13-3 12%
12% TP53 34%
12% ERICH3 16%
EBV_status
ID_status
Localization


A
From non-Ig variants
Wilcox Test: p=0.00016***
n = 20
median = 2.3
n = 11
median = 1.6
log10(Predicted neoepitope prevalence)
EBV_status
EBV-
EBV+
ALL NHL
p=0.25 p=0.85 p=0.64
n = 12
median = 2.4
n = 6
median = 2
n = 2
median = 2.3
ID_status
IC
Transplant
HIV
EBV-
B
From non-Ig variants
Wilcox Test: p=0.00079***
n = 41
median = 2.7
n = 25
median = 2.2
EBV_status
EBV-
EBV+
ALL NHL
C
R = 0.93, p < 0.001
log10(Predicted WES neoepitope prevalence)
log10(Predicted RNA neoepitope prevalence)
R = 0.83, p < 0.001
Tumour Mutational Burden
log10(Predicted RNA neoepitope prevalence)
D
From Ig variants
Wilcox Test: p=0.858
n = 19
median = 1.2
n = 5
median = 1.3
log10(predicted neoepitope prevalence)
EBV_status
EBV-
EBV+
ALL NHL
Wilcox Test: p=0.838
n = 11
median = 1.2
n = 13
median = 1.2
ID_status
IC
ID
ALL NHL


EBV_status
EBV-
EBV+
ID_status
IC
Transplant
HIV
Localization
CNS
Systemic
Ig-genes tested
Yes
No
Tumour_type
Plasmablastic lymphoma
Polymorphic PTLD
Plasmocytoma
Burkitt
DLBCL
IFN-γ spots per 10^6 cells
4200
3000
1800
1200
1000
800
600
500
400
200
50
0
Complete pool
GRIA3 pool
MHC II-presented IgH pool
MHC II- presented IgH pool
UBR4
MTMR1
Non-Ig pool
FAT4 pool
Nectin2
SETD1B
FRMD8
CCDC13
CLPTM1
NPIB2
RNF43
ZFP36L1
EBV status
ID status
Localization
Ig-genes tested
Tumor type
TMB
10
3.4
3.4
3.4
13
3.2
4.4
1.7
1.8
3.8
3.6
2.9
1.6
3.7


Wilcox Test: p=0.0308 *
n = 20
median = 134
n = 7
median = 412
TCR Beta clonotype number
3000
2000
1000
0
EBV-
EBV+
ALL NHL
EBV_status
EBV-
EBV+
Wilcox Test: p=0.0387 *
n = 12
median = 90
n = 8
median = 157
1500
1000
500
0
IC
ID
NHL EBV-
ID_status
IC
ID
Wilcox Test: p=0.981
n = 16
median = 139.5
n = 11
median = 211
3000
2000
1000
0
Systemic
CNS
ALL NHL
Localization
Systemic
CNS


A
Wilcox Test: p=0.02 *
n = 21 n = 9
median = 0 median = 2.4
T.cells.regulatory..Tregs. (%)
Wilcox Test: p=0.04 *
n = 21 n = 9
median = 2.9 median = 5
NK.cells.resting (%)
Wilcox Test: p=0.03 *
n = 21 n = 9
median = 4.4 median = 9.6
Macrophages.M2 (%)
Wilcox Test: p=0.04 *
n = 21 n = 9
median = 2.4 median = 9
Macrophages.M1 (%)
Wilcox Test: p=0.04 *
n = 21 n = 9
median = 2.5 median = 14.1
T.cells.CD8 (%)
Wilcox Test: p=0.03 *
n = 21 n = 9
median = 0.5 median = 3.4
Monocytes (%)
EBV_status
EBV-
EBV+
ALL NHL
B
R = 0.83, p = 0.006
R = 0.81, p = 0.008
R = 0.35, p = 0.125
R = 0.34, p = 0.131
Tumour Mutational Burden
T cells CD4 memory activated (%)
T cells CD8 (%)


Negative immune regulation
Wilcox Test: p=0.056
n = 21
median = 0.3
n = 9
median = 0.66
ssgsea Score
1.0
0.5
0.0
EBV-
EBV+
ALL NHL
Positive immune regulation
Wilcox Test: p=0.859
n = 21
median = -0.05
n = 9
median = 0.07
ssgsea Score
0.8
0.4
0.0
-0.4
EBV-
EBV+
ALL NHL
T-cell function
Wilcox Test: p=0.125
n = 21
median = 0.14
n = 9
median = 0.36
ssgsea Score
1.0
0.5
0.0
EBV-
EBV+
ALL NHL
EBV_status
EBV-
EBV+

SUPPLEMENTARY METHODS AND MATERIALS

Methods:

Patient cohort

Patients were enrolled from the Hematology or Neuro-Oncology departments at Pitié-Salpêtrière Hospital, Paris, France, or the lymphoid malignancies unit at Henri Mondor Hospital, Créteil, France. The lymphoma subtype was defined on the diagnostic tumor biopsy according to the 2016 WHO classification[1]. The EBV status was assessed using EBV-encoded small RNA (EBER) *in situ* hybridization analysis. In 8 cases where enough tissue was available, an immune-histochemistry analysis of the EBER, LMP-1 and EBNA-2 EBV antigens expression was performed in order to assess the EBV latency status as follows: latency I: EBER+, LMP-1-, EBNA-2-, latency II: EBER+, LMP-1+, EBNA-2- and latency III: EBER+, LMP1+, EBNA-2+.

Whole exome sequencing

WES was performed on tumor genomic DNA from tumor tissue biopsies and on germline DNA from blood. DNA was isolated from fresh frozen or formalin-fixed paraffin-embedded (FFPE) biopsies and from blood using a QIAamp-DNA Mini-Kit, FFPE Tissue-Kit or blood kit (Qiagen) respectively. Libraries were prepared and hybrid-captured using the SeqCap-EZ MedExome-Enrichment Kit (Roche) with 200 ng DNA input. Sequencing was performed on an Illumina Novaseq system with 150-bp paired-end reads. Raw paired-end fastq files were pre-processed using fastp[2] for adapter trimming and quality filtering. The filtered reads were aligned to the hg38 human reference genome using BWA-MEM[3]. The mean sequencing coverage across targeted bases was 200× and 150× for tumour and germline, respectively (**Supplementary Figure 2.A**). Somatic single nucleotide variants (SNVs) and indels were

called using MuTecT[4] and Strelka2[5], respectively. We excluded potential oxidative damage-induced and low base quality mutations using in-house scripts[6]. For two patients with missing germline DNA, we used Mutect2 (v4.1.4.1) and Panel of Normal (PoN) was generated, made from 38 normal samples DNA (derived from healthy tissue). Variants detected (marked PASS, coverage AD >=10 in both normal and tumor, at least 3 tumor variant reads, VAF >=0.05, gnomAD_AF <= 0.0001) were annotated using VEP[7]. Candidate driver genes and significantly mutated genes were defined using OncodriveCLUST[8].

Detection of copy number alteration (CNV)

We employed FACETS[9] tool to accurately infer CNV and tumour purity estimation for the paired samples (n=66). Moreover, GISTIC2[10] was used for the identification of recurrent copy number alterations. By leveraging data from high-throughput genomic profiling technologies, GISTIC2 aids in the detection of focal and arm-level copy number variations that are statistically significant across a cohort of samples. GISTIC2's allow to discern between true somatic events and random fluctuations in copy number.

CCF inference

PyCloneVI[11] was employed to estimate the cancer cell fractions within each tumor sample. FACETS CNV segments and raw variant allele frequencies were provided as input, along with a specified mutation prevalence threshold of 0.05 to exclude low-frequency variants. The 'pyclone-vi fit' and 'pyclone-vi write-results-file' command was executed with default settings.

We determined if a peptide was clonal for each patient by looking at the mutation's CCF that forms the peptide. A peptide is considered clonal when its CCF is 0.8 or higher.

RNA sequencing

RNA was extracted from fresh frozen biopsies using an RNeasy Micro-Kit (Qiagen). Libraries were prepared from 500 ng RNA. After end-repair, A-tailing, ligation and purification, sequencing was performed on an Illumina Novaseq with 150-bp paired-end reads. Reads were aligned to the human hg38 reference genome after an index was generated using STAR v2.7.2[12] by applying per-sample two-pass mapping. The generated BAM files were pre-processed according to GATK v4.1 RNA-seq best practice (SplitNcigarReads and BQSR), duplicates reads were marked then removed using STAR option '--bamRemoveDuplicatesType UniqueIdentical --bamRemoveDuplicatesMate2basesN 15' and Samtools, respectively. Finally, gene counts were obtained using Htseq[9] (RNAseq library size range: 24.8-154.9 Million reads, see **Supplementary Figure 2.B** and transformed into CPM (Counts Per Million) values.

Gene expression and cell type abundance profiling

T cell function included/ *IFN-γ, TNF-α, CD107a, granzyme* B and *perforin,* positive immune regulation included: *ICOS, ICOSL, CD28, CD40, CD40L, OX0, 0X40L* and *4.1BB,* and negative immune regulation included: *PD-1, PD-L1, PD-L2, LAG-3, TIM-3, Galectin9, TIGIT, CD47, IL-10, TGF-β* and *IDO*. Each sample enrichment scores and gene set pairing was computed separately using Single-sample Gene Set Enrichment Analysis (ssGSEA)[13] (https://www.gsea-msigdb.org/gsea/index.jsp).

Differential gene expression (DGE) and gene ontology (GO) enrichment analysis

CPM values were normalized using the Trimmed Mean of M-values normalization method from edgeR package[14]. To filter out low read counts, a CPM threshold of 0.5 was applied, equivalent to a count of 10 for the library sizes in this dataset. Subsequently, the data were log2 normalized (see **Supplementary Figure 2.C**). DGE analysis was performed with edgeR. DE genes were identified using a p-value cut-off of 5% without fold-change cut-off. A GO analysis

was used to predict putative biological functions based on DGE. The DE genes between CNS vs Systemic, EBV+ vs EBV- and ID vs IC groups were inserted to the goana function in limma-R packages[15] with focus on the biological process ontology.

Since the MHC class-I and class-II locus expression was computed using seq2HLA and measured in RPKM, we opted to calculate the *B2M* expression in RPKM as well.

MHC class-I and class-II restricted neoepitope prediction

Neoepitopes were predicted using the Ideation@SiRIC pipeline combining several software packages. First, we used seq2HLA to determine MHC class-I and class-II types [PMID: 23259685] using default parameters for all 31 RNAseq tumors and 68 normal-WES preprocessed fastq files using fastp. Next, somatic mutation-filtered VCF files were annotated by Variant Effect Predictor (Version 99) with default parameters and additionally with the 'Frameshift' and 'wild-type' plugins (from the pVACtools suite [version 3.0.3 PMID: 31907209]). The annotated non-synonymous mutations were extracted for downstream analysis. For each variant, the transcript expression levels (from RNA sequencing data) were then added using vcf-expression-annotator from VAtools v5.0.1. All parameters were then processed with the pVACseq for neoantigen prediction. For each pVACseq run, epitope prediction was done by the NetMHC [PMID: 18463140] NetMHCpan and NetMHCIIpan [PMID: 32406916] algorithms packed in the pVACseq toolkit; epitope length was set to 8–10 amino-acids long for class-I and 15 for class-II presentation with default parameters for all other settings. Predicted neoepitopes were filtered based on coverage >10X, DNA VAF ≥10%, transcript level expression ≥ 0.5 counts per million (CPM), TSL (Transcript Support Level) =1 and median affinity binding ≤500 nM. When RNA samples were unavailable, selection from the WES data used all of these parameters except the transcript level expression. The number

of neoepitopes per tumor was defined as the overall number of unique filtered mutant peptide sequences per tumor.

Neoepitope priority score: the priority score is one way to prioritize neoepitope candidates proposed by pVACseq: each of the following 4 criteria are assigned a rank-ordered value: B= rank of the mutant IC50 binding affinity/ F= rank of fold change between MT and WT allelles/ M= rank of mutant allele expression/ D= rank of tumor DNA VAF. The score is calculated with the following formula: B + F + (M*2) + (D/2), and then converted to a rank.

T cell receptor (TCR) analysis

From the RNAseq fastq files, the productive TCRβ clonotypes numbers and sequences and their frequencies were generated with MixCR V3.0[16] (https://mixcr.readthedocs.io/). Reads were aligned to reference TCR V, D, J and C genes. Each final clonotype was identified by a unique CDR3 sequence and clonotype count. Further analyses were restricted to samples with numbers of productive TCRs >100. The TCR repertoire abundance was defined as the number of unique productive clonotypes. Shannons' entropy index, computed using vdjtools V1.2.1 software[17], was used to compute the TCR repertoire entropy based on the frequency of particular sequences and normalized after division by clonotype log numbers (lower values indicate greater diversity). The clonotypes with frequencies ≥10% were considered to be suggestive of a tumor-neoantigen selection and investigated for known antigen specificities in public dataset (VDJdb)[18].

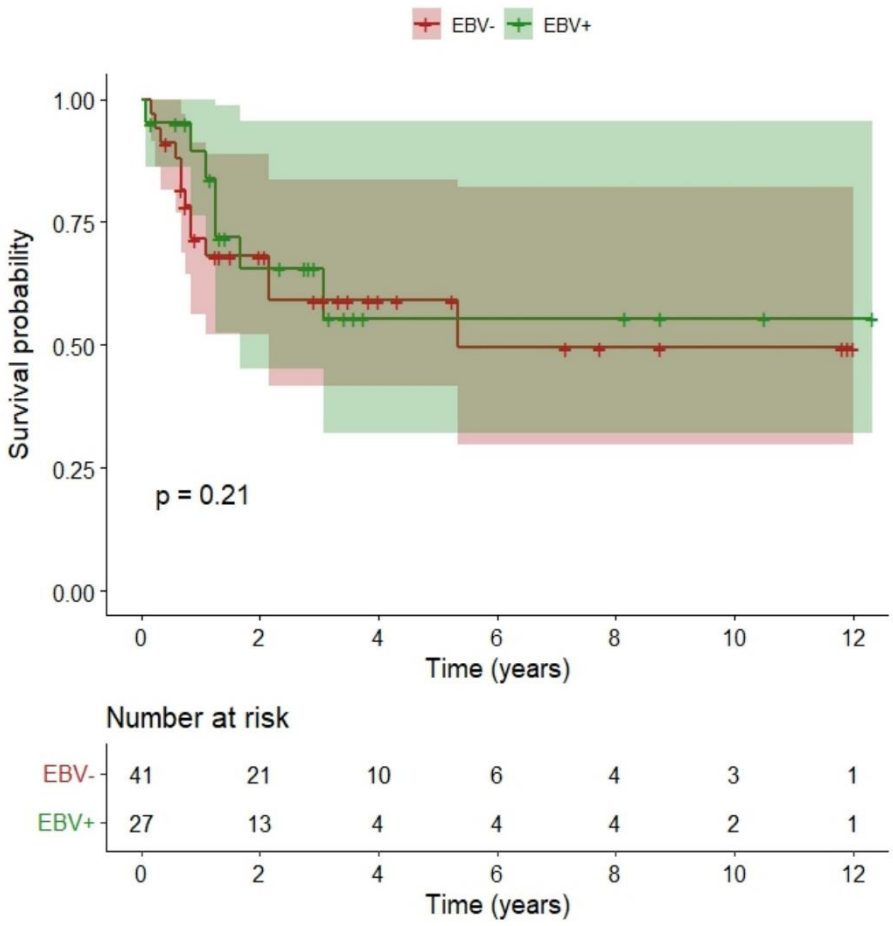
Statistical Analysis

Statistical analysis was performed using R (version 4.2.1). The Fisher exact test was used to compare categorical data and the Wilcoxon test was used to compare quantitative variables between groups. All tests were 2-sided at the threshold of p=0.05. In the case of multiple testing, we controlled FDR at the 5% level. Survival plots were generated using the Kaplan–Meier

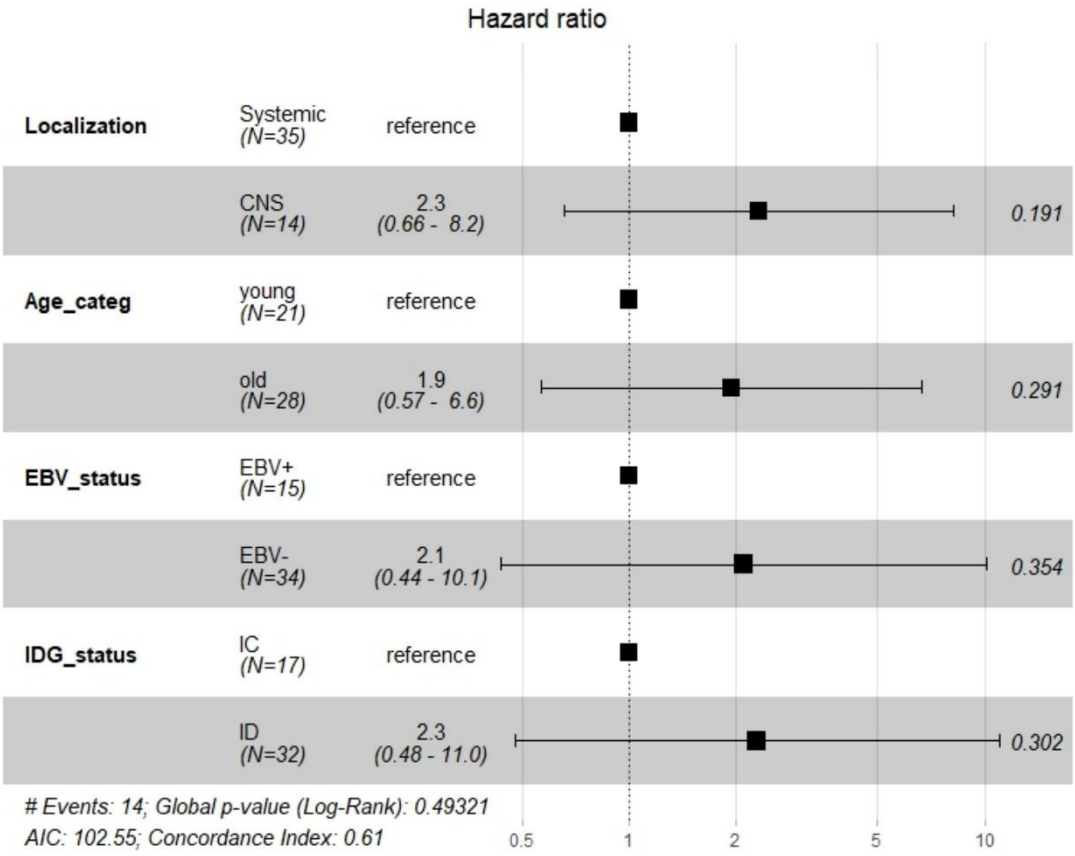
method. The Log-rank test was used to compare survival curves between groups. The Cox proportional-hazards regression model was used to compute hazard ratios summarizing the association between survival and age, EBV status, immune status, and localization. We made use of the “surv_cutpoin” function from survminer R package to determine the optimal TMB cutpoint that corresponds to the most significant relation with survival.

Figures

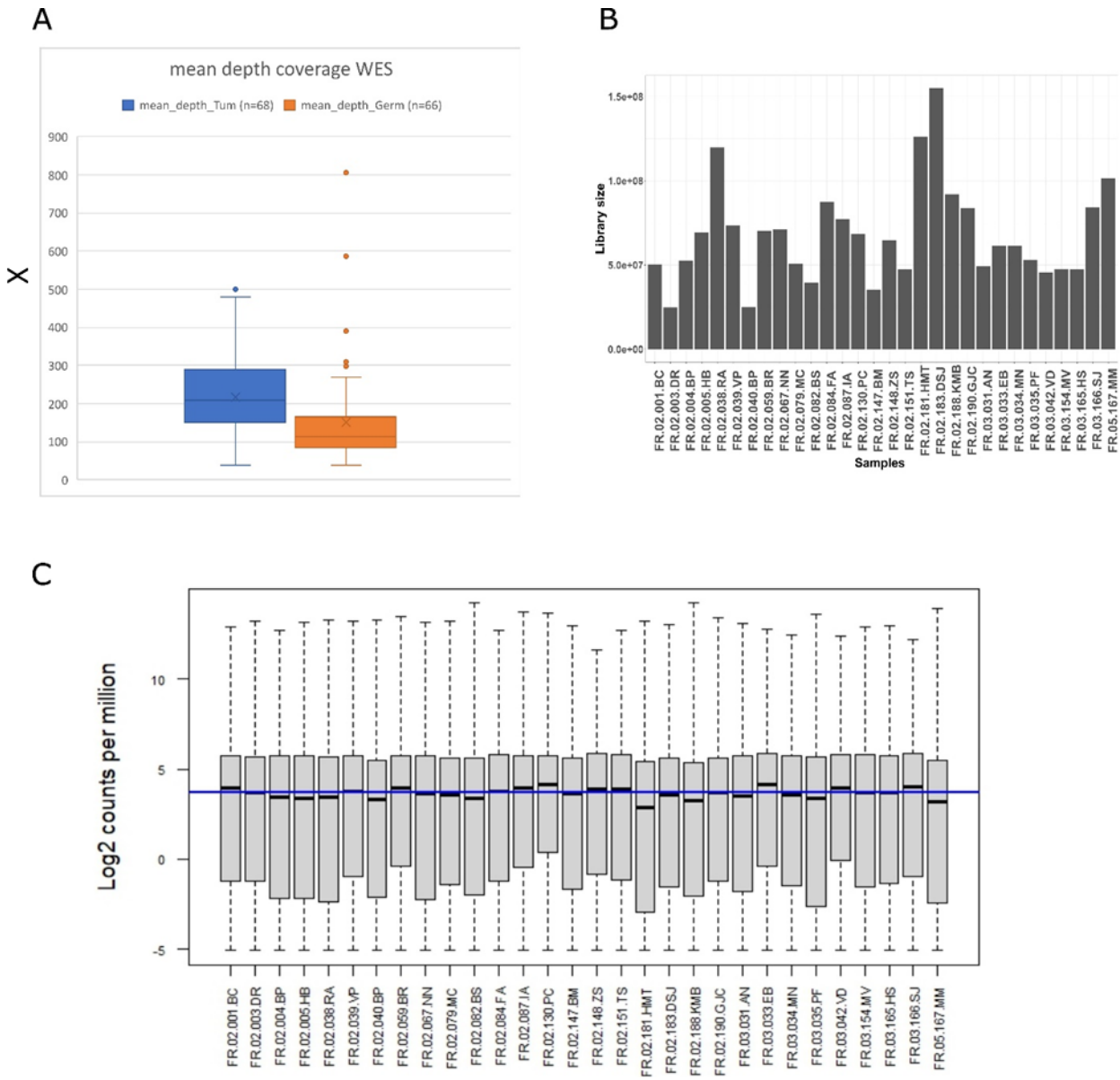
A


EBV-
EBV+
1.00
0.75
0.50
0.25
0.00
Survival probability
p = 0.21
0
2
4
6
8
10
12
Time (years)
Number at risk
EBV- 41 21 10 6 4 3 1
EBV+ 27 13 4 4 4 2 1
0
2
4
6
8
10
12
Time (years)

B

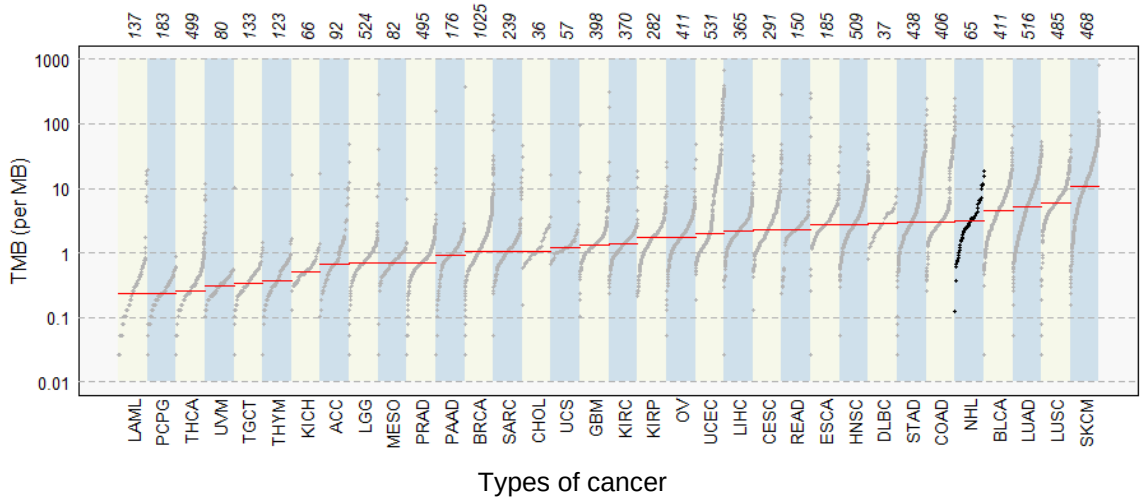

Hazard ratio
Localization
Systemic (N=35)
reference
CNS (N=14)
2.3 (0.66 - 8.2)
0.191
Age_categ
young (N=21)
reference
old (N=28)
1.9 (0.57 - 6.6)
0.291
EBV_status
EBV+ (N=15)
reference
EBV- (N=34)
2.1 (0.44 - 10.1)
0.354
IDG_status
IC (N=17)
reference
ID (N=32)
2.3 (0.48 - 11.0)
0.302
Events: 14; Global p-value (Log-Rank): 0.49321
AIC: 102.55; Concordance Index: 0.61
0.5
1
2
5
10

Supplementary Figure 1. A. Overall survival within NHL patients according to the EBV status (negative: red line; positive: green line) with adjustments made for patient age. **B.** Table representing a Cox regression model using tumor localization, age stratified into an old (>= 58 years old) and a young group (< 58 years old), EBV status, immune status stratified into immunocompetent (IC) group and immune-deficient group (ID: HIV + PTLD).



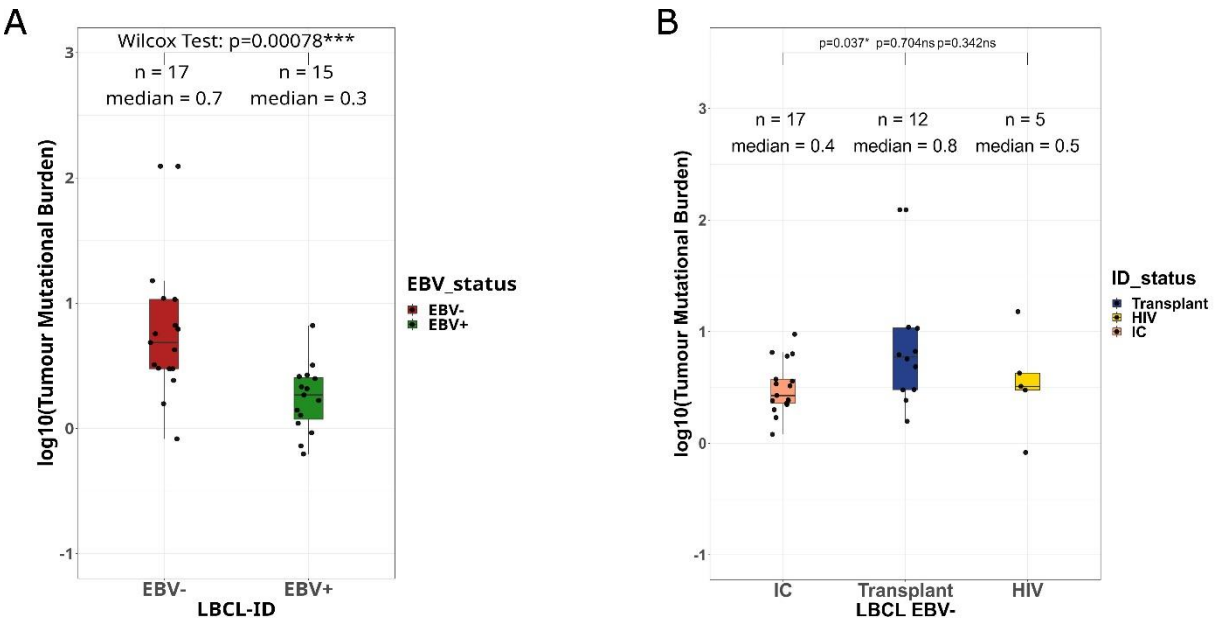
Supplementary Figure 2. A. Coverage of exome sequencing for tumors and their normal counterparts. Boxes are divided by median values. Length of boxes corresponds to interquartile

range and whiskers correspond to 1.5 interquartile ranges. **B**. RNA seq library size. **C.** Distribution of Log2 Count Per Million (CPM) across tumors.

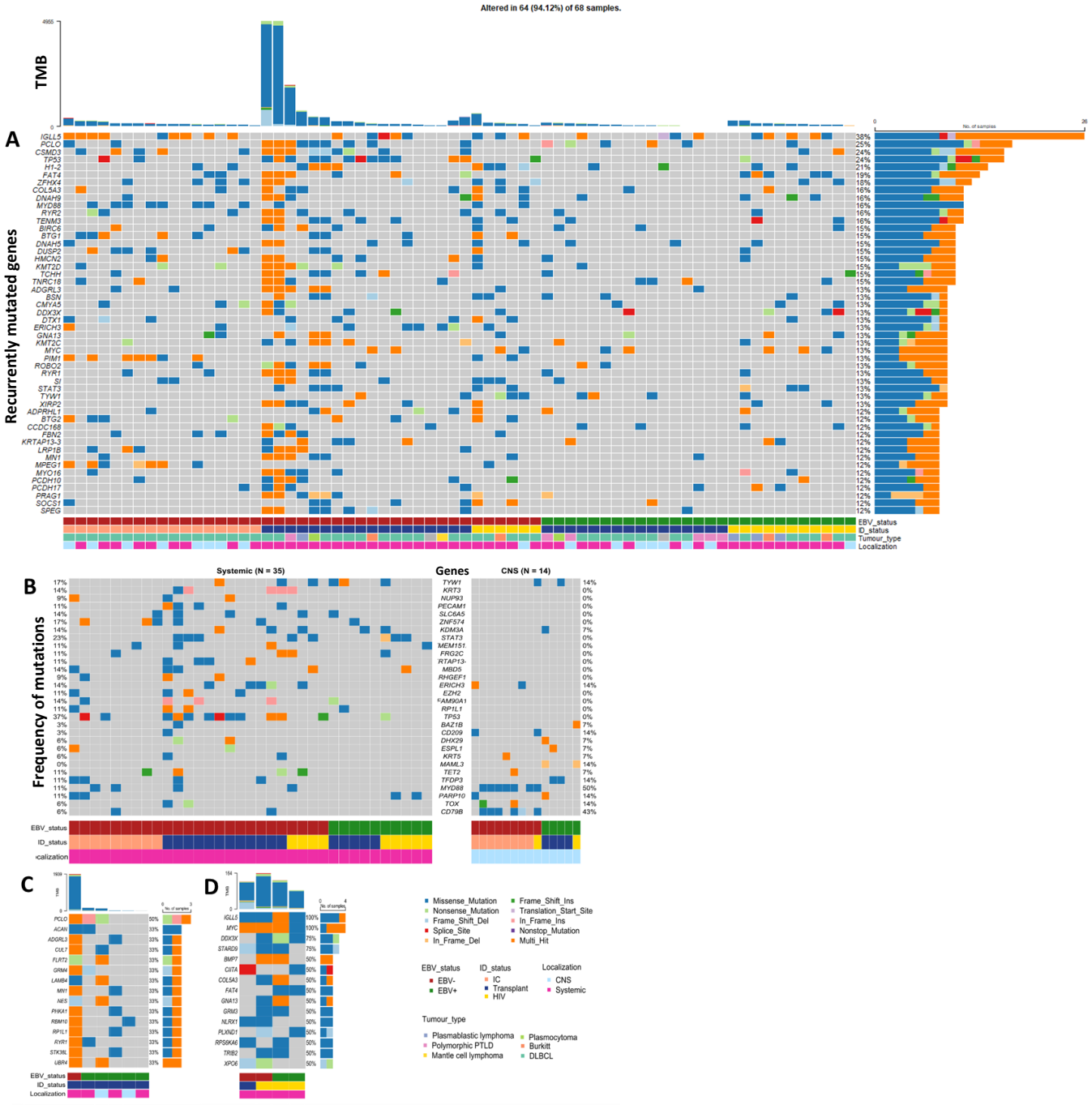


Supplementary Figure 3. **Tumor mutational burden**. The prevalence of somatic mutations in our cohort of NHL patients is illustrated among the ones of human cancer types described previously by [19]. The x axis shows the different cancers ordered based on their median numbers of somatic mutations. Our cohort is named "NHL" and appears close to the "DLBCL" one. The TMB is defined as the number of mutations per megabase. LAML: Acute Myeloid Leukemia, PCPG: Pheochromocytoma and Paraganglioma, THCA: Thyroid carcinoma, UVM: Uveal Melanoma, TGCT: Testicular Germ Cell Tumors, THYM: Thymoma, KICH: Kidney Chromophobe, ACC: Adrenocortical carcinoma, LGG: Brain Lower Grade Glioma, MESO: Mesothelioma, PRAD: Prostate adenocarcinoma, PAAD: Pancreatic adenocarcinoma, BRCA: Breast invasive carcinoma, SARC: Sarcoma, CHOL: Cholangiocarcinoma, UCS: Uterine Carcinosarcoma, GBM: Glioblastoma multiforme, KIRC: Kidney renal clear cell carcinoma, KIRP: Kidney renal papillary cell carcinoma, OV: Ovarian serous cystadenocarcinoma, UCEC: Uterine Corpus Endometrial Carcinoma, LIHC: Liver hepatocellular carcinoma, CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma, READ: Rectum adenocarcinoma, ESCA: Esophageal carcinoma, HNSC: Head and Neck squamous cell

carcinoma, DLBC: Lymphoid Neoplasm Diffuse Large B-cell Lymphoma, STAD: Stomach adenocarcinoma, COAD: Colon adenocarcinoma, NHL: Non-Hodgkin Lymphoma (IDEATION Cohort), BLCA: Bladder Urothelial Carcinoma, LUAD: Lung adenocarcinoma, LUSC: Lung squamous cell carcinoma, SKCM: Skin Cutaneous Melanoma.



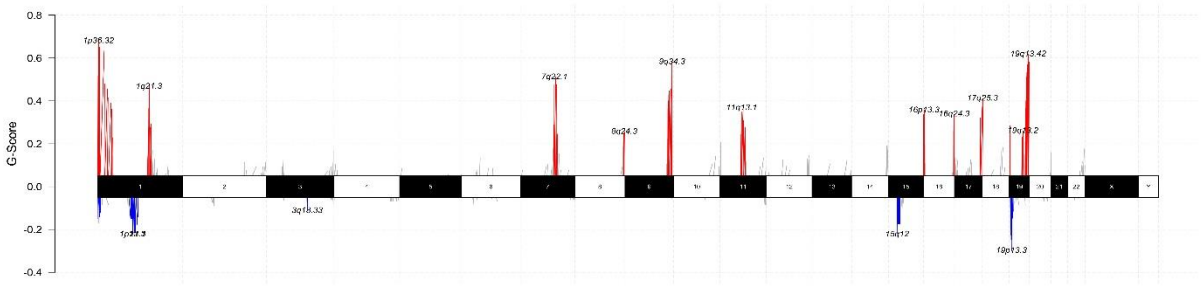
Supplementary Figure 4. Tumor mutational burden (TMB) according to **A**. the EBV status among the immunodeficient (ID) patients with large B cell lymphoma (LBCL) (n=32). **B**. and to the immune status among the EBV-negative LBCL patients (n=34): IC, transplant and HIV patients are illustrated with salmon, blue and HIV bars respectively. The TMB is defined as the number of mutations per megabase. IC: immunocompetent. *Wilcoxon Test*.



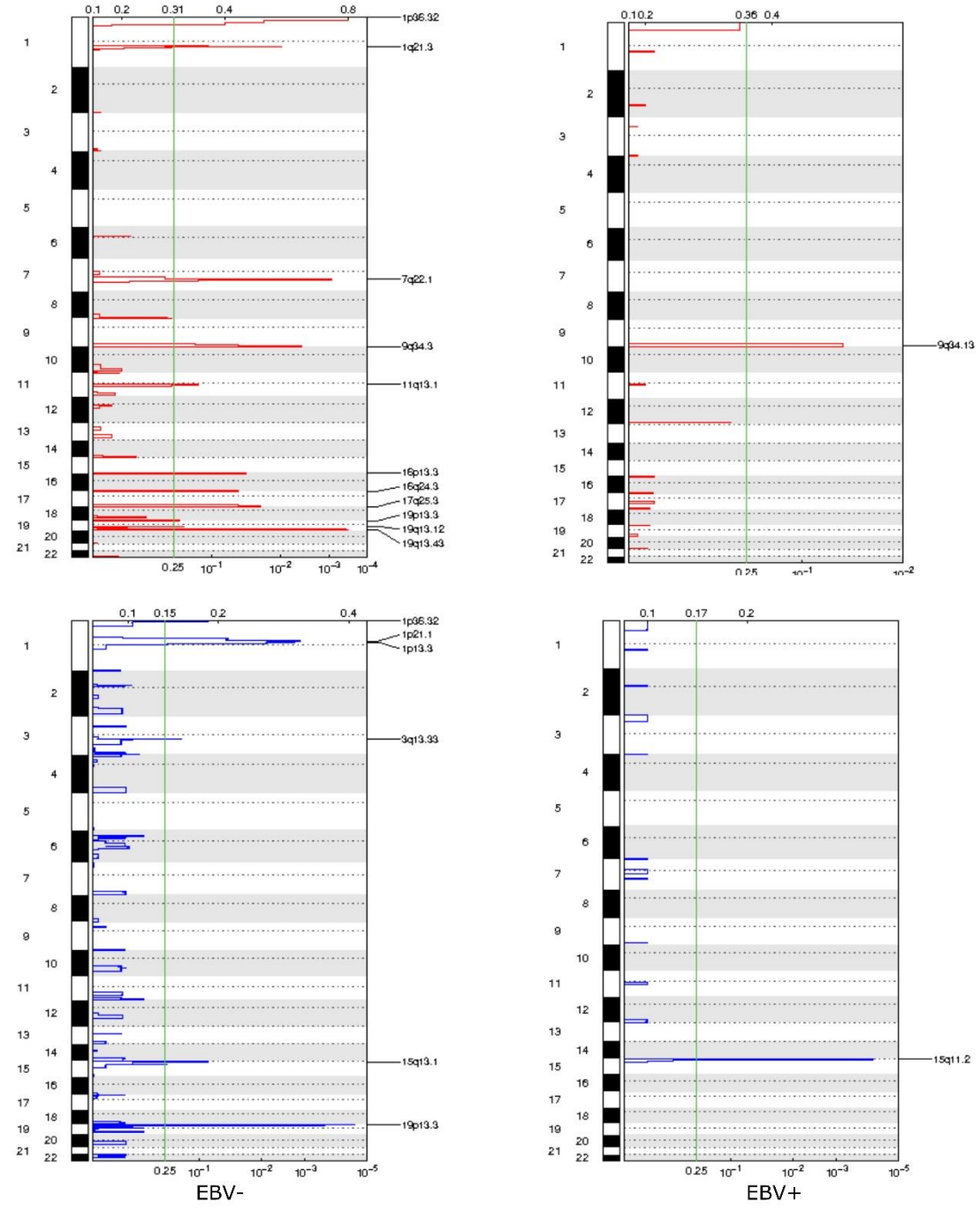
Supplementary Figure 5. **A**. Oncoplot of the top 50 most recurrently mutated genes within the overall population of 68 NHL patients. Tumor Mutational Burden (TMB) is represented as the total number of coding mutations per megabase per tumor. **B**. Co-oncoplot of the most recurrently mutated genes within the 49 LBCL samples with clinical annotations and according to the disease localization. **C**. Oncoplot of the most recurrently mutated genes within the 6 polymorphic PTLD with clinical annotations. **D.** Oncoplot of the most recurrently mutated

genes within the 4 Burkitt Lymphoma (BL) with clinical annotations. CNS: central nervous system. *Fisher exact test.*

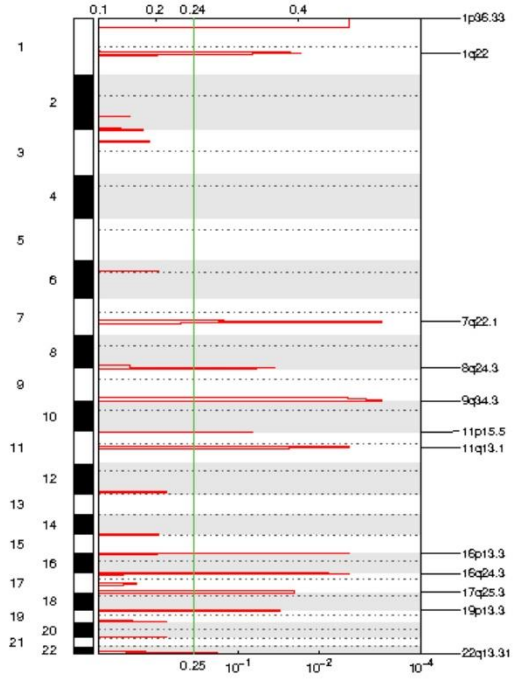
A

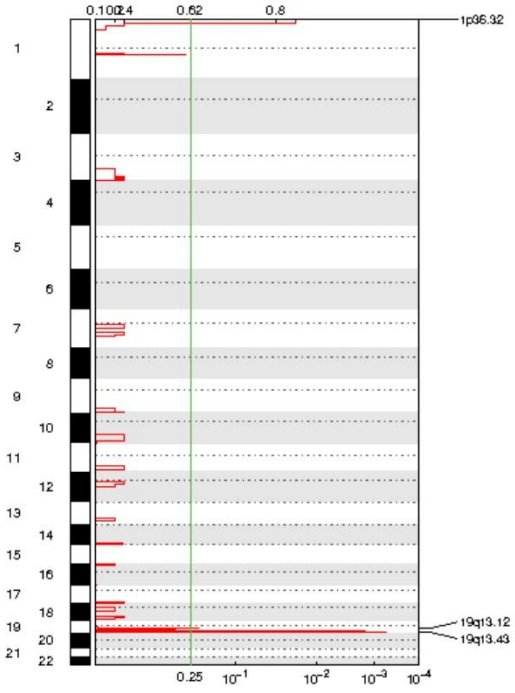

G-Score
1p36.32
1q21.3
7q22.1
9q34.3
9q34.3
11q13.1
16p13.3
16q24.3
17q25.3
19q13.42
19q13.2
3q13.33
15q11.2
19p13.3

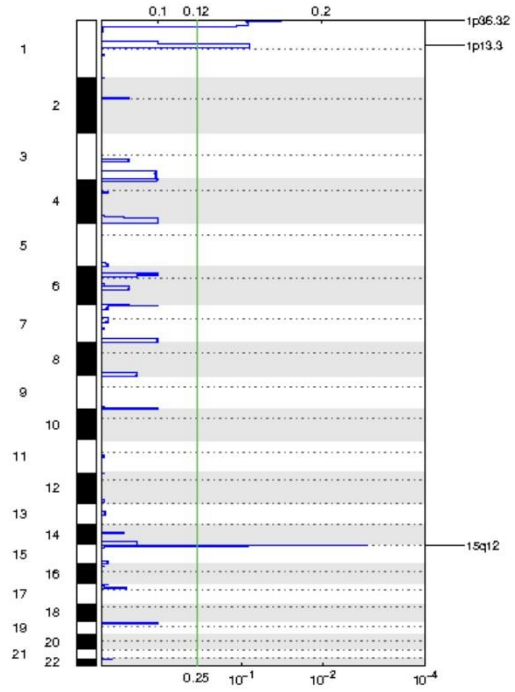
B


1p36.32
1q21.3
7q22.1
9q34.3
11q13.1
16p13.3
16q24.3
17q25.3
19p13.3
19q13.12
19q13.43
9q34.13
1p36.32
1p21.1
1p13.3
3q13.33
15q13.1
19p13.3
15q11.2
EBV-
EBV+

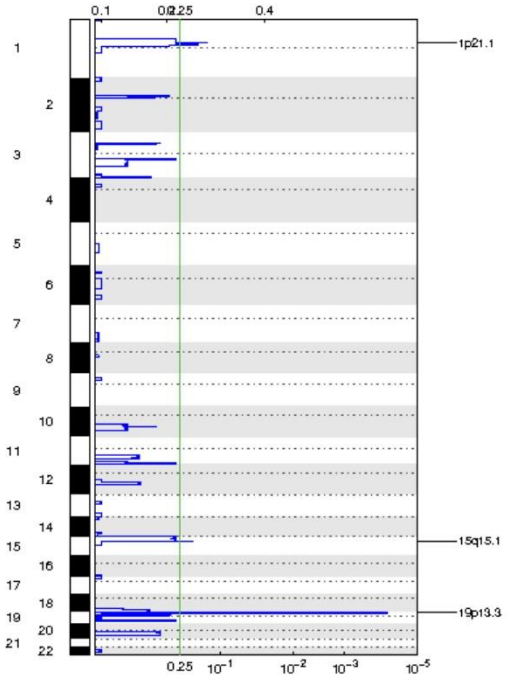
C


0.1
0.2
0.24
0.4
1p36.33
1q22
7q22.1
8q24.3
9q34.3
11p15.5
11q13.1
16p13.3
16q24.3
17q25.3
19p13.3
22q13.31
0.25
10-1
10-2
10-4


0.1002.4
0.62
0.8
1p36.32
19q13.12
19q13.43
0.25
10-1
10-2
10-3
10-4

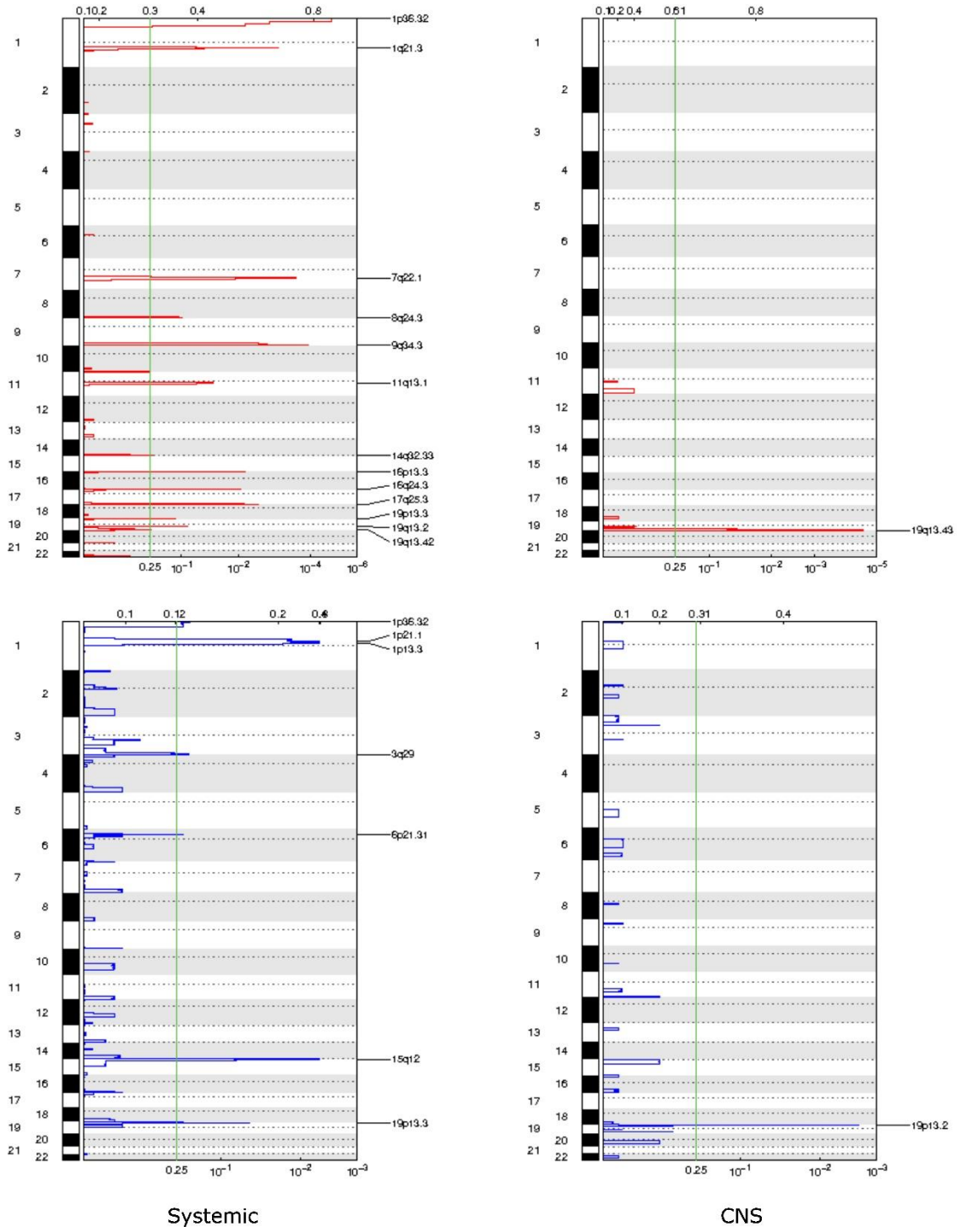

0.1
0.12
0.2
1p36.32
1p13.3
15q12
0.25
10-1
10-2
10-4

ID


0.1
0.25
0.4
1p21.1
15q15.1
19p13.3
0.25
10-1
10-2
10-3
10-5

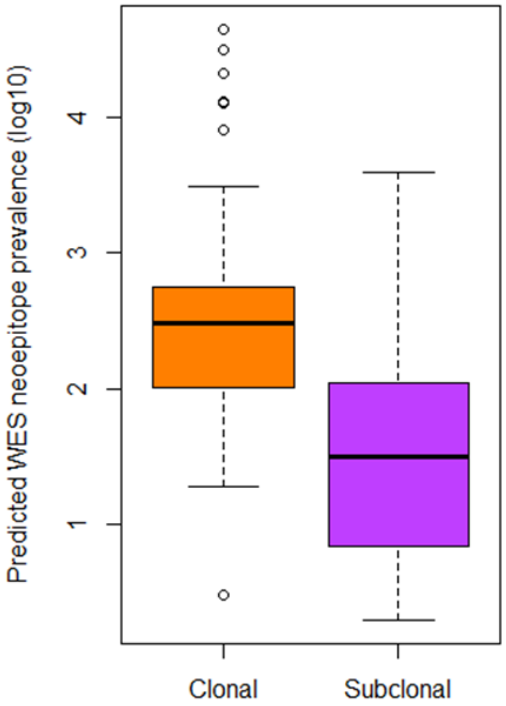
IC

D

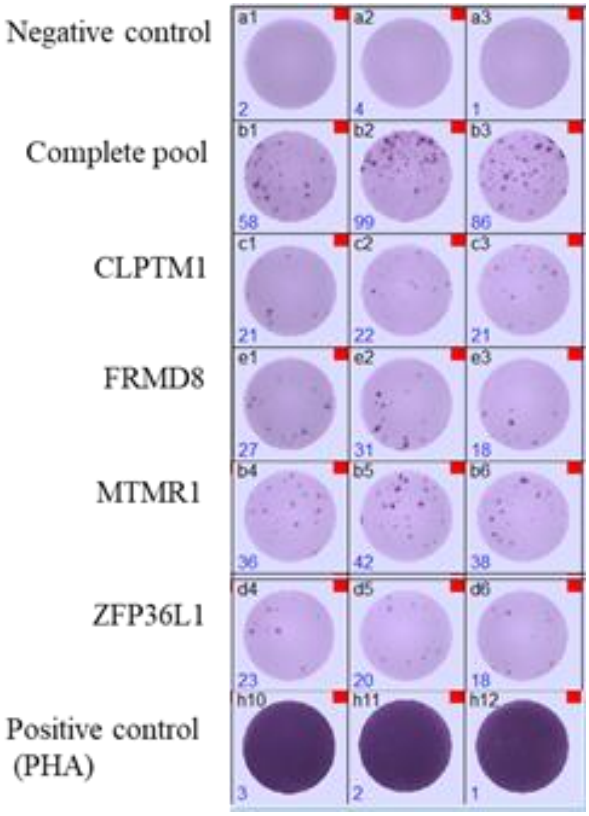


Supplementary Figure 6. **A**. **Frequency of arm-level somatic copy number alterations (SCNAs) among LBCL tumors.** dashed line indicating the threshold of arm-level SCNA frequency at 20% or higher. **B, C and D**. **Comparisons of GISTIC2.0-defined recurrent focal copy number** gains (red) and losses (blue). Chromosomes are shown on the

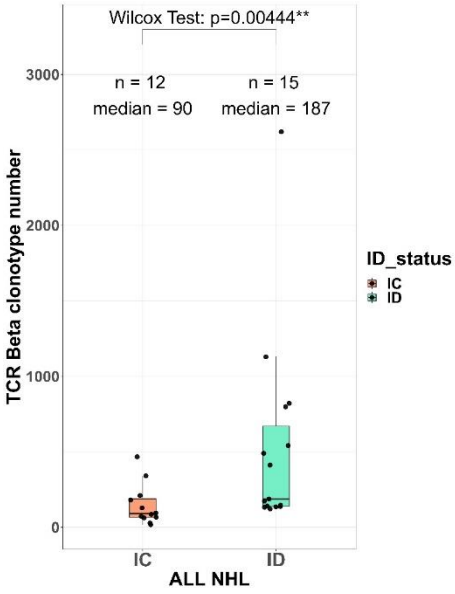
vertical axis. Green line denotes q-value of 0.25. Somatic copy number alterations (SCNAs) are labeled with their associated cytoband. **A.** Based on EBV status **B.** based on immune status. **C.** Based on disease localization. CNS: central nervous system; ID: immunocompetent; ID: immunodeficient.



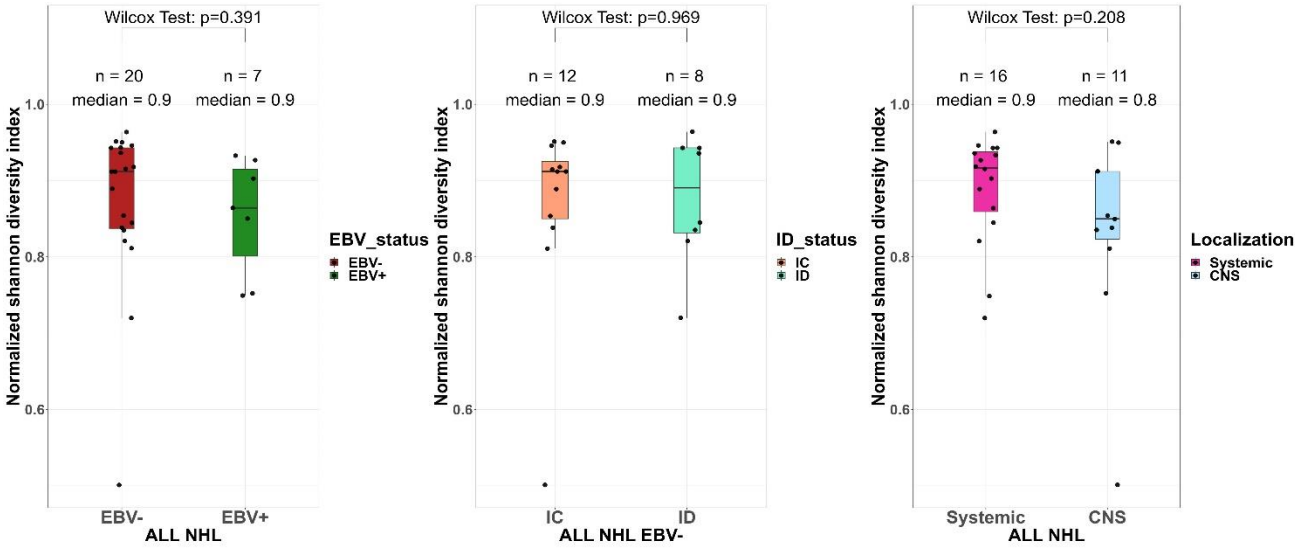
Supplementary Figure 7. Number of Clonal and Subclonal predicted neoepitopes (log_{10}) per patient based on WES sequencing.



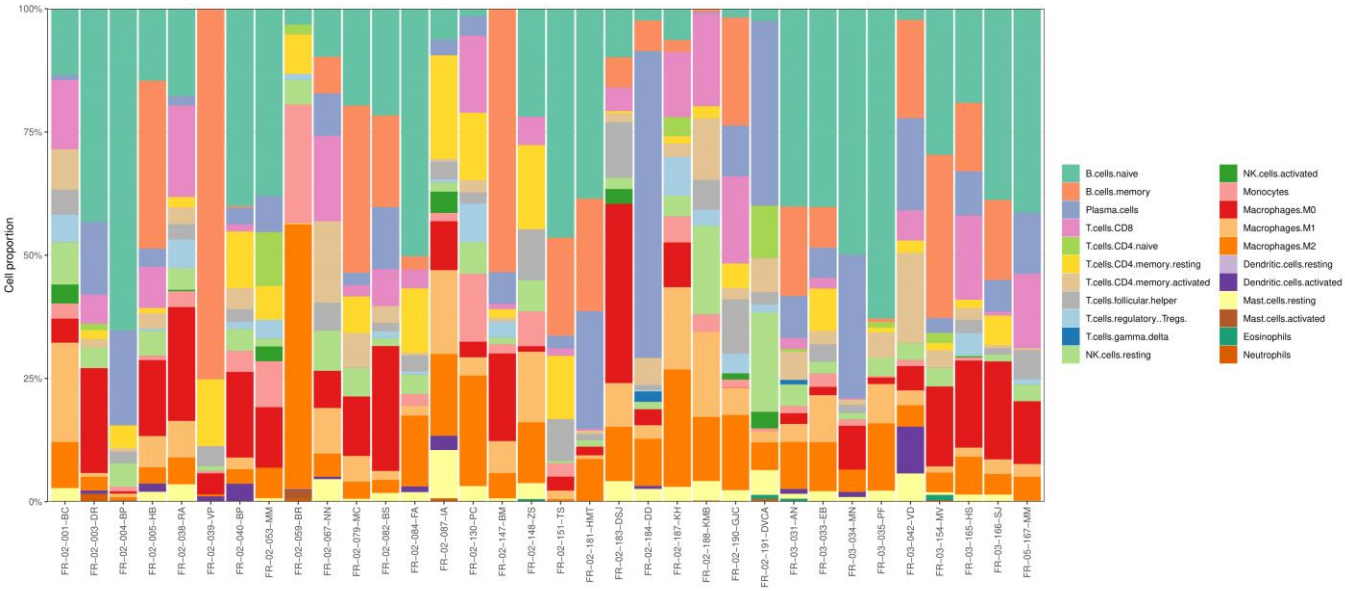
Supplementary Figure 8. Example ELISPOT results among one HIV patient with EBV-negative NHL: the patient was tested for 41 different peptides (only positive results against CLPTM1, FRMD8, MTMR1 and ZFP36L1 peptides are shown), negative control (at the top) and positive control (at the bottom). Thawed PBMC were co-cultured with personalized pooled peptides during 10 days and next tested for reactivity using IFN-γ enzyme linked immunoSpot (ELISPOT) assays. Patients were all tested for their personalized pooled peptides (named "complete pool") and eventually for each individual peptide if the number of cells were adequate (named as the mutated gene). The mean numbers of spot forming cells (SFC) from triplicates were normalized to number of IFN-γ spots detected per $1x10^6$ PBMC after background subtraction, and ELISPOT-IFN-γ positivity threshold was 50 SFC per million cells.



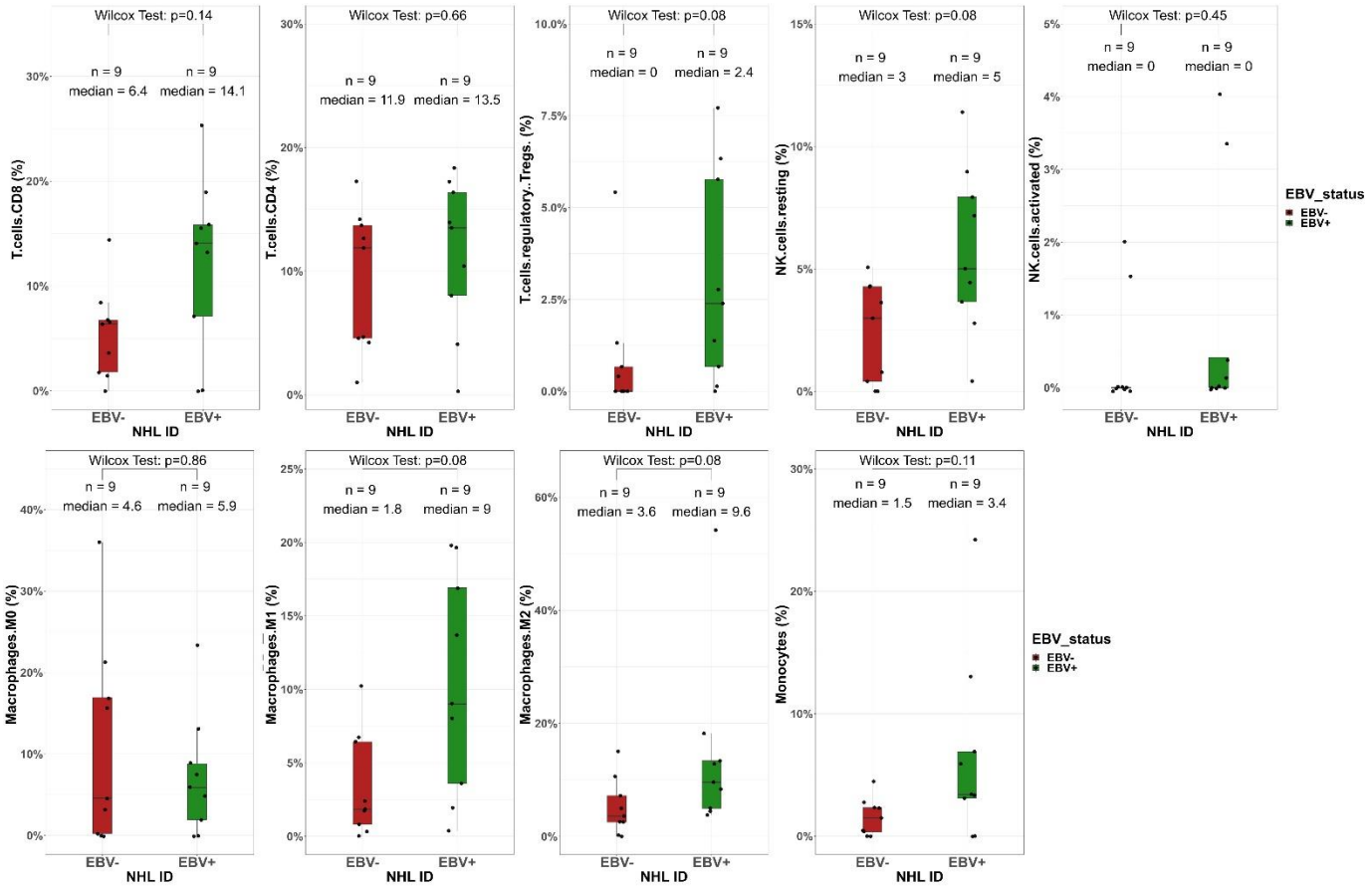
Supplementary Figure 9. Intra-tumoral TCR repertoire abundance according to the immune status within 27 NHL samples. *Wilcoxon Test*.



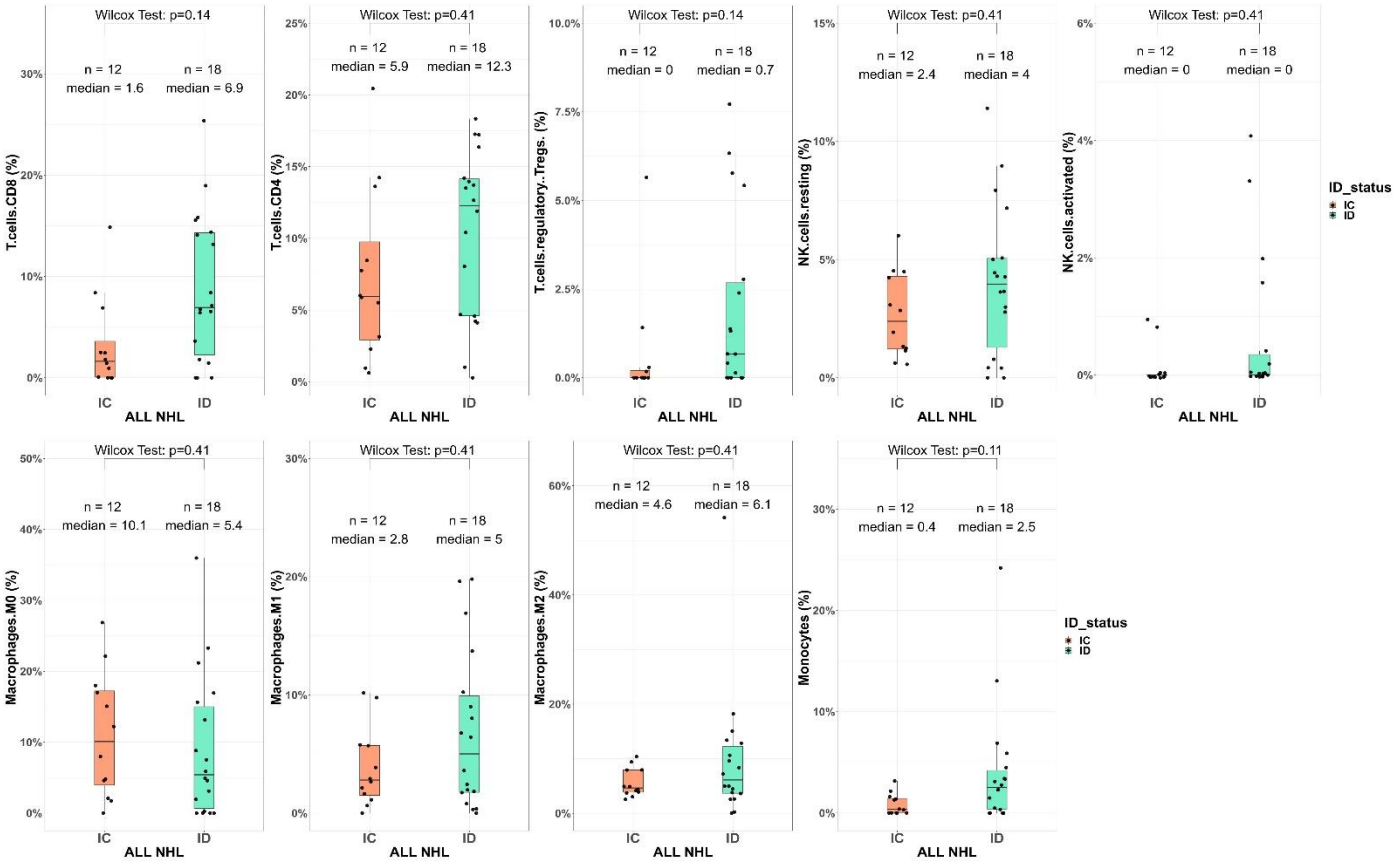
Supplementary Figure 10. Normalized Shannon index diversity index according to the EBV status (left), the immune status (middle) and the disease localization (right). Shannons' entropy was calculated using vdjtools V1.2.1 software[17] and was divided by (log (number of clonotypes)) to be independent of the abundance: lower values indicate more diversity. IC: immunocompetent; ID: immunodeficient; CNS: central nervous system. *Wilcoxon Test*



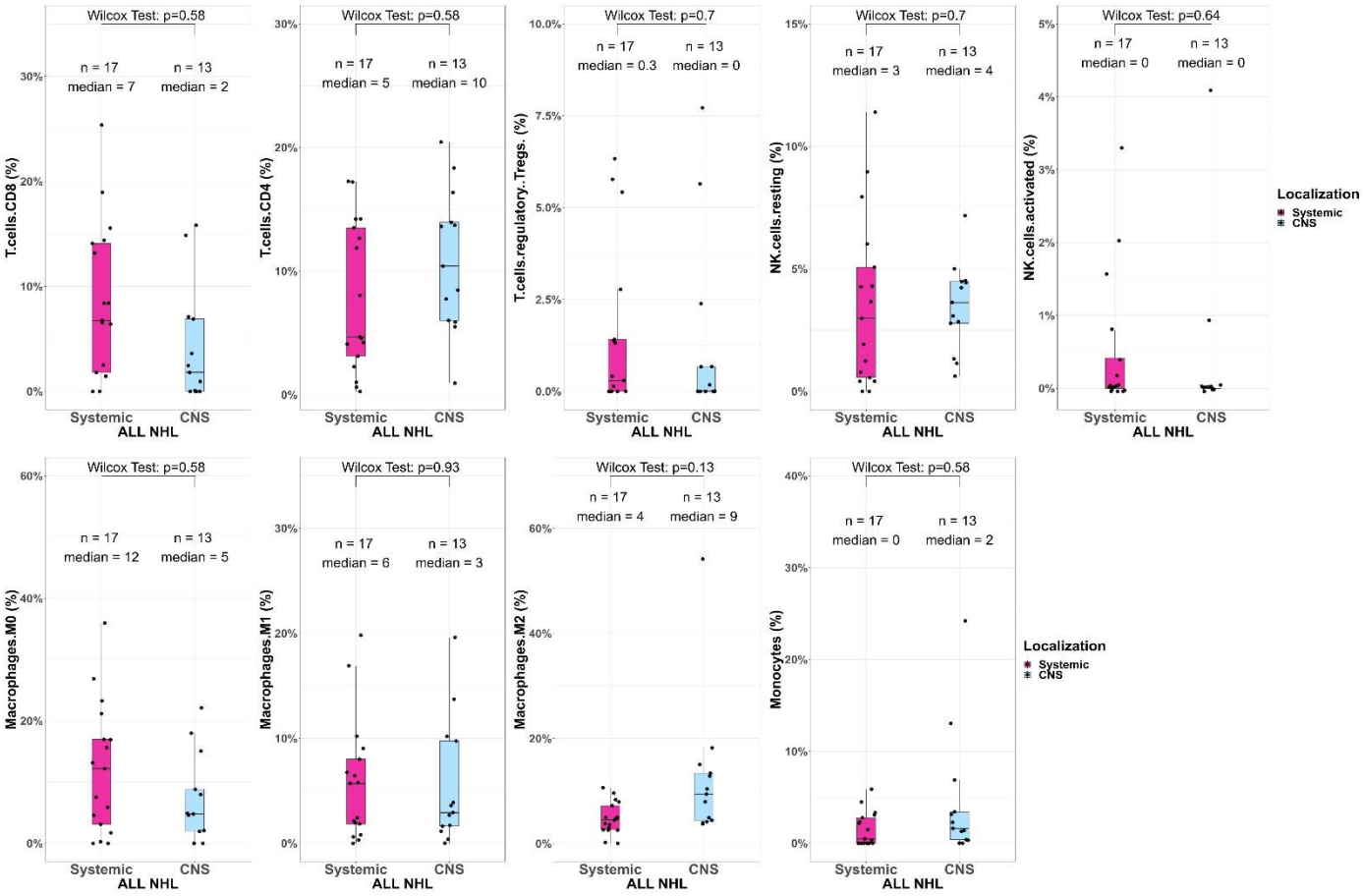
Supplementary Figure 11. Cell type abundance distribution for each patient using CIBERSORTx.



Supplementary Figure 12. Cell type abundance according to the EBV status among the ID patients only. Cell type abundance was assessed using CIBERSORTx.system. *Wilcoxon Test*.



Supplementary Figure 13. Cell type abundance according to the immune status. Cell type abundance was assessed using CIBERSORTx. IC: immunocompetent; ID: immunodeficient. *Wilcoxon Test*.



Supplementary Figure 14. Cell type abundance according to the disease localization. Cell type abundance was assessed using CIBERSORTx. CNS: central nervous system. *Wilcoxon Test*.

Tables

Supplementary Table 1. Comparison of the mutational frequencies stratified by EBV status. Mutational frequencies in the 34 EBV-negative lymphomas were compared to the 15 EBV-positive ones. Genes are ranked by uncorrected p values (with corrected p values shown in the adjacent column).

Hugo_Symbol	EBV-	EBV+	pval	OR	CI. low	CI. up	Adj.Pval
MYD88	11	0	0.01124097	Inf	Inf	1.352539871	0.3484702
HNRNPF	0	3	0.02469605	0.0000000	0.9941339	0.000000000	0.3827888
TP53	12	1	0.04263089	7.3914358	348.7086709	0.900912662	0.4405192
CD79B	8	0	0.08683169	Inf	Inf	0.831634188	0.4613095
AP3B2	0	2	0.08928571	0.0000000	2.2852181	0.000000000	0.4613095
MAML3	0	2	0.08928571	0.0000000	2.2852181	0.000000000	0.4613095
ESCO2	1	2	0.21846505	0.2046464	4.2426528	0.003250142	0.6768950
MYO15A	1	2	0.21846505	0.2046464	4.2426528	0.003250142	0.6768950
RTN4	1	2	0.21846505	0.2046464	4.2426528	0.003250142	0.6768950
TYW1	4	4	0.22720046	0.3752253	2.3835301	0.058448793	0.6768950
EZH2	4	0	0.29833959	Inf	Inf	0.291905276	0.6768950
KRTAP13-3	4	0	0.29833959	Inf	Inf	0.291905276	0.6768950
PECAM1	4	0	0.29833959	Inf	Inf	0.291905276	0.6768950
KRT3	5	0	0.30569453	Inf	Inf	0.412788484	0.6768950
KDM3A	3	3	0.35330099	0.3956178	3.3716730	0.046175510	0.6845207
PARP10	3	3	0.35330099	0.3956178	3.3716730	0.046175510	0.6845207
ARHGEF15	3	0	0.54325879	Inf	Inf	0.180556834	0.8872066
NUP93	3	0	0.54325879	Inf	Inf	0.180556834	0.8872066
TOX3	2	2	0.57635598	0.4147095	6.2831643	0.027324750	0.8872066
CEP250	3	2	0.63519543	0.6354177	8.4440603	0.064516010	0.8872066
PLXNB3	3	2	0.63519543	0.6354177	8.4440603	0.064516010	0.8872066
ZNF574	5	1	0.65177745	2.3768993	122.1232631	0.232289392	0.8872066
TNRC18	4	3	0.65987565	0.5407780	4.2502025	0.077917168	0.8872066
STAT3	5	3	0.68686965	0.6951469	5.1882401	0.113021613	0.8872066
DHX29	2	1	1.00000000	0.8774500	55.2479011	0.042342197	1.0000000
ESPL1	2	1	1.00000000	0.8774500	55.2479011	0.042342197	1.0000000
ID2	2	1	1.00000000	0.8774500	55.2479011	0.042342197	1.0000000
FOXQ1	5	2	1.00000000	1.1181454	13.2050937	0.156202537	1.0000000
GNAI2	4	1	1.00000000	1.8455237	98.3507357	0.161980906	1.0000000
RP1L1	3	1	1.00000000	1.3469268	76.1064554	0.097813307	1.0000000
TFDP3	4	2	1.00000000	0.8692637	10.7397691	0.108142054	1.0000000

Supplementary Table 2. Comparison of the mutational frequencies stratified by immune status. Mutational frequencies in the 32 ID LBCL were compared to the 15 IC DLBCL. Genes are ranked by uncorrected p values (with corrected p values shown in the adjacent column).

Hugo_Symbol	ID	IC	pval	OR	CI. up	CI. low	Adj. Pval
MYD88	3	8	0.004574082	0.1228121	0.644494	0.01724404	0.1326484
STAT3	8	0	0.038368700	Inf	Inf	1.02551071	0.3708974
TYW1	8	0	0.038368700	Inf	Inf	1.02551071	0.3708974
KDM3A	6	0	0.079848305	Inf	Inf	0.66525774	0.4791059
TP53	11	2	0.105125508	3.8314402	40.600568	0.68100497	0.4791059
ANO10	0	2	0.115646259	0.0000000	2.767552	0.00000000	0.4791059
PKDCC	0	2	0.115646259	0.0000000	2.767552	0.00000000	0.4791059
KRT3	5	0	0.148789334	Inf	Inf	0.50449696	0.5393613
E2F7	1	2	0.273122015	0.2496756	5.148406	0.00398254	0.6855032
RAPGEF6	1	2	0.273122015	0.2496756	5.148406	0.00398254	0.6855032
KRTAP13-3	4	0	0.283656478	Inf	Inf	0.35569410	0.6855032
PECAM1	4	0	0.283656478	Inf	Inf	0.35569410	0.6855032
PARP10	3	3	0.405458996	0.4905080	4.138715	0.05794602	0.8398793
ZNF574	3	3	0.405458996	0.4905080	4.138715	0.05794602	0.8398793
FOXD4L4	2	0	0.537414966	Inf	Inf	0.09910238	0.8924406
ARVCF	2	2	0.602031377	0.5076208	7.640914	0.03367628	0.8924406
SLAMF7	2	2	0.602031377	0.5076208	7.640914	0.03367628	0.8924406
SPTAN1	2	2	0.602031377	0.5076208	7.640914	0.03367628	0.8924406
FAM90A1	4	1	0.646250113	2.2520240	119.418697	0.19933525	0.8924406
MBD5	4	1	0.646250113	2.2520240	119.418697	0.19933525	0.8924406
SLC6A5	4	1	0.646250113	2.2520240	119.418697	0.19933525	0.8924406
ADCY8	4	2	1.000000000	1.0699424	13.119541	0.13457850	1.0000000
ERICH3	5	2	1.000000000	1.3799652	16.160751	0.19512568	1.0000000
ESPL1	2	1	1.000000000	1.0652809	66.787013	0.05166123	1.0000000
FRG2C	3	1	1.000000000	1.6394834	92.199004	0.11990355	1.0000000
MCAT	3	2	1.000000000	0.7800302	10.287934	0.07993619	1.0000000
NUP93	2	1	1.000000000	1.0652809	66.787013	0.05166123	1.0000000
RP1L1	3	1	1.000000000	1.6394834	92.199004	0.11990355	1.0000000
TFDP3	4	2	1.000000000	1.0699424	13.119541	0.13457850	1.0000000

IC: immunocompetent; ID: immunodeficient; Adj. P value: adjusted P value after FDR correction.

Supplementary Table 3 is available in a separate excel spreadsheet. All GISTIC2 thresholded SCNA by gene.

References

1. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016;127(20):2375–2390.
2. Chen X, Chang C-W, Spoerke JM, et al. Low-pass Whole-genome Sequencing of Circulating Cell-free DNA Demonstrates Dynamic Changes in Genomic Copy Number in a Squamous Lung Cancer Clinical Cohort. *Clin Cancer Res*. 2019;25(7):2254–2263.
3. Adalsteinsson VA, Ha G, Freeman SS, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun*. 2017;8(1):1324.
4. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213–219.
5. Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;15(8):591–594.
6. Costello M, Pugh TJ, Fennell TJ, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res*. 2013;41(6):e67.
7. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122.
8. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics*. 2013;29(18):2238–2244.
9. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res*. 2016;44(16):e131.
10. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011;12(4):R41.
11. Gillis S, Roth A. PyClone-VI: scalable inference of clonal population structures using whole genome data. *BMC Bioinformatics*. 2020;21(1):571.
12. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
13. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.
14. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010;11(3):R25.
15. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
16. Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. 2015;12(5):380–381.
17. Shugay M, Bagaev DV, Turchaninova MA, et al. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput Biol*. 2015;11(11):e1004503.
18. Goncharov M, Bagaev D, Shcherbinin D, et al. VDJdb in the pandemic era: a compendium of T cell receptors specific for SARS-CoV-2. *Nat Methods*. 2022;19(9):1017–1019.
19. Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415–421.