Supporting Information

SI Materials and Methods

Patients

Plasma samples from two distinct cohorts of HIV-1-infected patients were investigated. The HIV lymphoma cohort included 83 plasma samples obtained from HIV-1-infected patients with lymphoma treated at the Centro di Riferimento Oncologico, NCI, Aviano, Italy, from 1997 to 2010 for diffuse large cell B-cell lymphoma (n=33), Burkitt lymphoma (n=19), Immunoblastic lymphoma (n=7), Plasmoblastic lymphoma (n=4), Anaplastic lymphoma (n=1), Primary Effusion lymphoma (n=1), Hodgkin lymphoma (n=17), Non-hodgkin’s Lymphoma not otherwise specified (n=1). The control cohort included 70 residual plasma samples obtained from HIV-1-infected patients undergoing routine monitoring of viremia at the Brescia Civic Hospital (BS), Italy, from 2018 to 2020. All patients were infected with HIV-1 subtype B as determined by the sequence analysis of the p17 region, according to the Los Alamos genotyping algorithm.

The study was conducted in accordance with the Declaration of Helsinki and national standards, and was approved by the Brescia Ethics Committee (NP 3163). Before starting the analyses, the samples were completely anonymized, and the only remaining linkage was with the clinical data included in Tables S1 and S2.

Sanger sequencing

HIV-1 RNA viral load quantification of the control group was done using COBAS® AmpliPrep/COBAS® TaqMan® HIV-1 Test v2.0 automatic system (Roche Molecular System, Inc), according to the manufacturer’s instructions. To perform p17 gene sequences from circulating virions, plasma samples with HIV-RNA values of at least 30 copies/ml were selected. For p17 Sanger
sequence analysis, RNA was extracted using the QIAamp DSP Virus kit (QIAGEN, Heiden, Germany) with a 200 µl plasma starting volume and eluted in 30 µl Buffer AVE. A nested PCR approach was used to amplify a 559-nucleotide (nt)-long amplicon (from nt 326 to nt 885) to evaluate the whole P17 coding sequence. In brief, RNA was reverse transcribed and PCR amplified using the Superscript III One-Step RT-PCR system with PlatinumTaq DNA polymerase (Thermo Fisher Scientific, Carlsbad, CA, USA) in a 50 µl reaction containing 25 µl of reaction mix, 9 µl of MgSO₄, 2 µl of SuperScriptTM III RT/PlatinumTM Taq Mix, 0.2 µM of sense and antisense primers, and 12 µl of extracted RNA. The amplification conditions were as follows: 50°C for 30 min (for reverse transcription) and 94°C for 2 min for Taq DNA polymerase activation, followed by 40 cycles (94°C 15 sec, 55°C 30 sec, 68°C 2 min) and a final cycle at 68°C 7 min. PCR primers used in the reaction were: UGF1, 5′-GTGCCCGTCTGTGTGTG and p24R1, 5′- CATTTCATGGCTGCTTGATG. Then, we subjected 5 µl of PCR product to a nested-PCR reaction, performed using the AmpliTaq Gold™ DNA Polymerase (Thermo Fisher Scientific, Carlsbad, CA, USA) in a total volume of 50 µl containing each internal primer at 0.15 µM (p17F1 5′-AAGGAGAGAGAGGTGCGA and UGR1 5′-AATCTTGTGGGGTGGCTCCTT). Nested-PCR conditions were performed as follows: 95°C for 10 sec, 30 cycles at 94°C for 15 sec, 59°C for 30 sec and 72°C for 2 min, followed by a final extension at 72°C for 7 min. Nested-PCR products were checked on a 1.5% agarose gel, purified using the QIAquick PCR Purification Kit (QIAGEN, Heiden, Germany), quantified with the NanoDrop™ (Thermo Fisher Scientific, Carlsbad, CA, USA) and sequenced by BMR Genomics (Padua, Italy). The derived sequences were analyzed and edited with Geneious software (version 11.1.5) (Biomatters Ltd, New Zealand), using the PNL4.3 reference sequence (Accession number: AF324493).

Bioinformatic analysis

To investigate the variability of vp17 HIV-1 subtype B, a total number of 4015 sequences were retrieved from the Los Alamos HIV database (collection data ranging from 1985 to 2017). All
sequences were initially submitted to a genotyping assignment using the phylogenetic HIV subtyping tool and after excluding low quality genomes four different datasets have been built. In particular, dataset 1 (n=3990) includes all wild type (wt) p17 and vp17 genome sequences sampled from around the world and it has been used to perform statistical analysis; dataset 2, a down-sampled representative dataset (n=1221) includes 441 wt p17s, 470 vp17s carrying aa insertions at position 117-118 (ins117-118 vp17s) and 310 vp17s carrying amino aa at position 125-126 (ins125-126 vp17s) and it has been used to perform the Bayesian time-scaled phylogenetic analysis; dataset 3 (n=2952), includes 2283 wt p17s, 643 ins117-118 vp17s in addition to 26 ins117-118 vp17s generated in this study and it has been used to perform Maximum likelihood (ML) inferences; dataset 4 (n = 2792) includes 2283 wt p17s, 488 ins125-126 vp17s, in addition to 21 ins125-126 vp17s generated in this study and it has been used to perform ML inferences.

Phylogenetic and Bayesian analysis

All the sequences were aligned with HIV-1 subtype B p17 pNL4.3 (AF324492) reference genome. Sequences were aligned with MAFFT (FF-NS-2 algorithm) using default parameters and manually curated, to remove artifacts at the ends and within the alignment, using AliView. Phylogenetic analysis was performed using IQ-TREE (version 1.6.10) under the best fit model according to Bayesian Information Criterion (BIC) indicated by the Model Finder application implemented in IQTREE. The statistical robustness of individual nodes was determined using 1000 bootstrap replicates. In order to investigate the temporal signal in our HIV-1 dataset 2, we regressed root-to-tip genetic distances from this ML tree against sample collection dates using TempEst v 1.5.1. The ML phylogeny was used as a starting tree for Bayesian time-scaled phylogenetic analysis using BEAST 1.10.4. We employed a stringent model selection analysis using both path-sampling and steppingstone models to estimate the most appropriate model combination for Bayesian phylogenetic analysis. The best fitting model was the HKY+G4 substitution model with a Bayesian skyline coalescent model. We computed Monte Carlo Markov chains (MCMC) duplicate runs of 200 million
states each, sampling every 20,000 steps. In order to ensure stationary and adequate effective sample size (ESS) of >200, convergence of MCMC chains was checked using Tracer v.1.7.1. Maximum clade trees were summarized from the MCMC samples using TreeAnnotator after discarding 10% as burn-in.

Cell Cultures and Recombinant Proteins.

Human lymphoma B-cell line Raji was cultured in RPMI 1640 containing 10% (vol/vol) FBS. The coding sequence of HIV-1 isolate BH10 p17 (refp17; aa 1-132) was amplified by PCR with specific primers that allowed us to clone the refp17 sequence into the BamH1 site of the prokaryotic expression vector pGEX-2T (Pharmacia). The vp17 obtained from plasma of patient C7, the ancestor of the sexually transmitted cluster of HIV-1 mutants expressing vp17s, was amplified by PCR and cloned into the same vector. A specific stretch of Glu-Lys was inserted within the refp17 primary sequence between amino acids 114-115 by using the Quick Change Site-Directed Mutagenesis Kit (Stratagene). The recombinant proteins were purified (>98%) by reverse phase fast performance liquid chromatography. The absence of endotoxin contamination (<0.25 endotoxin U/mL) in the protein preparation was assessed by Limulus amoebocyte assay (Associates of Cape Cod Inc.).

Anchorage-Independent Growth Assay

Raji (12,500 cells/well) cells were plated in 12-well plates in 2 ml RPMI containing 5% FBS and 0.35% Sea-Plaque agarose (Lonza), over a 0.7% agarose base. One day after plating, medium containing or not viral proteins (0.01 µg/ml) was added to the top of the layer and replaced every 4 days. After 10 days, 300 µl of 3-[4, 5-Dimethylthiazol-2-y1]-2, 5-diphenyltetrazolium bromide (MTT) (SigmaAldrich) were added to each well and allowed to incubate for 4 h at 37°C. Plates were then placed overnight at 4°C, and colonies>50 µm in diameter were counted.

B cell colony formation assay
Raji cell suspension was sequentially diluted, and cells were seeded by manual pipetting into a 96 well plate at a dilution of 0.5 cells/well. Plates were incubated for 8 days under standard conditions (RPMI medium supplemented with 10% FBS) in the presence or absence of 0.01 µg/ml of viral proteins. Eight days after culture plates were analyzed for single colony formation. The colony area was measured (15 colonies/condition) by using Leica Qwin image analysis software. The same number of colonies (15 colonies/condition) was aseptically harvested from 96-well plates and stained with propidium iodide (PI) to detect PI- viable cells by flow cytometry. Absolute cell counts were obtained by the counting function of the MACSQuant® Analyzer (Miltenyi Biotec, Bergish Gladbach, Germany).

Statistical analysis.

We analyzed dataset 1 that includes the whole genome sequences collected between 1985 and 2017 complete of date and geographical information and evaluated the number of acquired mutations among the genomic position 117-118 and 125-126. We grouped data by three years, starting from 1985, in order to assess whether a linear trend between the number of acquired aa insertions was occurring over time. Data were analyzed for statistical significance using the chi-square test for trend and evaluated with a confidence level (α) of 0.05. Data concerning biological assays were analyzed for statistical significance using one-way ANOVA. Bonferroni's post-test was used to compare data. Differences were considered significant at p<0.05. Statistical tests were performed using GraphPad Prism 8 software (GraphPad).
Fig. S1. Frequency of aa insertions in dominant vp17 sequences of HIV-1-infected patients. (A) Bars represent the percentage of aa insertions in dominant vp17 sequences of HIV-1-infected patients with (Lymphoma patients) or without lymphoma (Control patients) detected at the indicated positions. Aa positions are referred to the subtype B strain BH10 (UniProtKB P04585). No statistically significant difference in the percentage of vp17s was observed among samples obtained from HIV-1-infected patients with or without lymphoma. (B) Bars represent the percentage of aa insertions in dominant
vp17 sequences in each Lymphoma histotype. The numbers appearing at the top of each bar indicate how many samples belong to that group of analysis. No statistically significant difference in the percentage of vp17s was observed among samples obtained from HIV-1-infected patients suffering from different lymphoma histotypes. (C) Bars represent the percentage of aa insertions in dominant vp17 sequences of HIV-1-infected patients with (Lymphoma patients) or without lymphoma (Control patients) related to viremia. The statistical significance was calculated using chi-square test. *P<0.05.; **P<0.01. DLBCL: diffuse large B cell lymphoma; HL: Hodgkin’s lymphoma.
Fig. S2. Pie chart showing percent distribution of aa insertion along the dominant vp17 sequences among HIV-1-infected patients with different Lymphoma histotypes. Charts represent the frequency of aa insertions in (A) patients with diffuse large B cell lymphoma (DLBCL), (B) patients with Immunoblastic/Plasmablastic lymphoma, (C) patients with Burkitt lymphoma (BL), (D) patients with Hodgkin’s lymphoma (HL). Each insertion position is identified by a color code. For each position, the label indicates the type and the frequency of insertion. Aa positions are referred to the subtype B strain BH10 (UniProtKB P04585).
Fig. S3. A mid-point rooted ML phylogenetic tree of ins\textsuperscript{117-118} vp17s. The tree includes ins\textsuperscript{117-118} vp17 sequences retrieved worldwide from Los Alamos database (dataset 3) and 26 ins\textsuperscript{117-118} vp17 sequences obtained in this study from HIV-1-infected lymphoma (n=17) and control (n=9) patients.
Fig. S4. A mid-point rooted ML phylogenetic tree of ins^{125-126} vp17s. The tree includes ins^{125-126} vp17 sequences retrieved worldwide from Los Alamos database (dataset 4) and 21 ins^{125-126} vp17 sequences obtained in this study from HIV-1-infected lymphoma (n=15) and control (n=6) patients.
Fig. S5. Effect of insEK\textsuperscript{114-115} refp17 on B-cell clonogenicity. In the colony formation assay, Raji cells were cultured in the presence or absence of refp17, vp17c7, or insEK\textsuperscript{114-115} refp17 (A) Bright-field images represent the characteristic morphology of two-dimensional colonies (original magnification, x40). (B) The colony area was measured by using Leica Qwin image analysis software. (C) The same number of colonies was aseptically harvested, stained with propidium iodide to detect viable cells by flow cytometry and counted by the counting function of the MACSQuant\textsuperscript{®} Analyzer. (D) In the soft agar assay, Raji cells were incubated in medium containing or not refp17, vp17c7, or insEK\textsuperscript{114-115} refp17. The cell growth was analyzed by using 3-[4, 5-Dimethylthiazol-2-y1]-2, 5-diphenyltetrazolium bromide (MTT). Data are representative of three independent experiments performed in triplicate. The statistical significance was calculated using one-way ANOVA and the Bonferroni’s post-test was used to compare data. NT, not treated cells. **P<0.01.; ***P<0.001.
Supplementary Table 1. Clinical and virological features of HIV-1-infected patients with lymphoma.

| Patient no. | Sex | Age (years) | Year of sample collection | HIV load (copies/ml) | Lymphoma diagnosis | Treatment Naïve/experienced |
|-------------|-----|-------------|---------------------------|----------------------|-------------------|---------------------------|
| L1          | M   | 30          | 1997                      | 50,950               | NHL (DLBCL)       | Experienced               |
| L2          | F   | 32          | 1997                      | 30,010               | NHL (DLBCL)       | Experienced               |
| L3          | M   | 35          | 1997                      | 95,180               | NHL (Burkitt)     | Experienced               |
| L4          | M   | 46          | 1999                      | 295,488              | NHL (Immunoblastic)| Experienced               |
| L5          | M   | 64          | 1999                      | 57,515               | NHL (DLBCL)       | Naïve                     |
| L6          | M   | 44          | 1997                      | 637                  | NHL (DLBCL)       | Experienced               |
| L7          | M   | 47          | 1998                      | >500,000             | NHL (Burkitt)     | Naïve                     |
| L8          | M   | 29          | 1998                      | 424,360              | NHL (Burkitt)     | Experienced               |
| L9          | M   | 34          | 1998                      | 16,027               | NHL (Nos)         | Experienced               |
| L10         | M   | 30          | 1998                      | 12,753               | NHL (Burkitt)     | Experienced               |
| L11         | M   | 37          | 2001                      | 113,275              | NHL (DLBCL)       | Experienced               |
| L12         | M   | 44          | 2001                      | 30,980               | NHL (DLBCL)       | Naïve                     |
| L13         | M   | 47          | 2001                      | 4,822                | NHL (Burkitt)     | Naïve                     |
| L14         | M   | 37          | 1997                      | 10,520               | NHL (Immunoblastic)| Experienced               |
| L15         | M   | 33          | 1997                      | 928                  | NHL (Plasmablastic)| Experienced               |
| L16         | M   | 39          | 1999                      | 199,448              | NHL (DLBCL)       | Naïve                     |
| L17         | M   | 33          | 1999                      | 74,264               | NHL (Burkitt)     | Naïve                     |
| L18         | M   | 52          | 2000                      | 6,806                | NHL (Burkitt)     | Naïve                     |
| L19         | M   | 49          | 2000                      | 8,939                | NHL (Burkitt)     | Naïve                     |
| L20         | F   | 43          | 2001                      | >500,000             | HL                 | Experienced               |
| L21         | F   | 32          | 2002                      | 221,750              | NHL (Immunoblastic)| Naïve                     |
| L22         | M   | 60          | 2002                      | >500,000             | NHL (DLBCL)       | Experienced               |
| L23         | M   | 38          | 2002                      | 211,160              | NHL (DLBCL)       | Experienced               |
| L24         | F   | 32          | 2002                      | 182,526              | NHL (DLBCL)       | Experienced               |
| L25         | M   | 36          | 2002                      | 223,923              | NHL (DLBCL)       | Experienced               |
| L26         | M   | 47          | 2002                      | 535                  | HL                 | Experienced               |
| L27         | M   | 39          | 2002                      | 287                  | NHL (DLBCL)       | Experienced               |
| L28         | F   | 32          | 2000                      | 1,654                | HL                 | Experienced               |
| L29         | F   | 37          | 2000                      | 54,213               | NHL (Primary Effusion Lymphoma) | Naïve |
| L30         | M   | 36          | 2000                      | 2,856                | NHL (Burkitt)     | Experienced               |
| L31         | M   | 52          | 2000                      | 15,507               | NHL (DLBCL)       | Naïve                     |
| L32         | M   | 56          | 2000                      | >500,000             | NHL (Plasmablastic)| Naïve                     |
| L33         | F   | 35          | 2000                      | 349                  | NHL (Immunoblastic, Primary CNS) | Naïve |
| L34         | M   | 39          | 2000                      | 49,597               | HL                 | Naïve                     |
| L35         | M   | 34          | 2001                      | 8,278                | NHL (DLBCL)       | Experienced               |
| L36         | M   | 42          | 2001                      | 2,368                | HL                 | Experienced               |
| L37         | M   | 41          | 2002                      | 40,202               | NHL (DLBCL)       | Experienced               |
| L38         | F   | 32          | 2001                      | 16,383               | HL                 | Experienced               |
| L39         | F   | 43          | 2005                      | >500,000             | NHL (DLBCL)       | Naïve                     |
| L40         | M   | 55          | 2005                      | 208,552              | NHL (Burkitt)     | Naïve                     |
| L41         | F   | 47          | 2005                      | 4,573                | NHL (DLBCL)       | Experienced               |
| L42         | F   | 43          | 2005                      | 74,229               | NHL (DLBCL)       | Naïve                     |
| L43         | M   | 44          | 2006                      | 127                  | NHL (DLBCL)       | Experienced               |
| L44         | M   | 32          | 2006                      | >500,000             | NHL (Immunoblastic)| Naïve                     |
| L45         | F   | 47          | 2006                      | 20,330               | NHL (Plasmablastic)| Experienced               |
| L46         | F   | 37          | 2005                      | 13,044               | NHL (Anaplastic)  | Naïve                     |
| L47         | M   | 39          | 2002                      | 264,764              | NHL (Burkitt)     | Naïve                     |
| L48         | M   | 36          | 2003                      | 973                  | NHL (Plasmablastic)| Experienced               |
| L49         | F   | 42          | 2003                      | 3,501                | NHL (Burkitt)     | Experienced               |
| L50         | M   | 46          | 2004                      | 15,224               | NHL (DLBCL)       | Experienced               |
| L51         | M   | 45          | 2004                      | 114,810              | NHL (DLBCL)       | Experienced               |
| #  | Gender | Age | Year | Value   | Diagnosis                      | Status   |
|----|--------|-----|------|--------|-------------------------------|----------|
| L52| M      | 40  | 2004 | 451    | NHL (DLBCL)                   | Naive    |
| L53| M      | 43  | 2004 | 500,001 | NHL (DLBCL)                   | Experienced |
| L54| M      | 45  | 2004 | 76,797 | NHL (Burkitt)                  | Naive    |
| L55| M      | 38  | 2003 | 4,415  | HL                            | Experienced |
| L56| M      | 29  | 2004 | 521    | HL                            | n.a.     |
| L57| M      | 51  | 2006 | 53,153 | NHL (Burkitt)                  | Experienced |
| L58| M      | 43  | 2006 | 287    | NHL (DLBCL)                   | Experienced |
| L59| M      | 65  | 2006 | 12,036 | NHL (Burkitt)                  | Naive    |
| L60| M      | 41  | 2006 | 2,507  | HL                            | Naive    |
| L61| M      | 42  | 2006 | 13,560 | HL                            | Naive    |
| L62| M      | 29  | 2007 | 358,783 | NHL (Burkitt)                   | Experienced |
| L63| M      | 34  | 2007 | 342    | HL                            | Naive    |
| L64| F      | 34  | 2007 | 359    | NHL (DLBCL)                   | Experienced |
| L65| M      | 63  | 2007 | 72,947 | NHL (DLBCL)                   | Naive    |
| L66| M      | 34  | 2007 | 25,634 | NHL (Burkitt)                  | Naive    |
| L67| M      | 62  | 2008 | 483,891 | NHL (DLBCL)                   | Naive    |
| L68| M      | 46  | 2008 | 92,708 | NHL (Burkitt)                  | Naive    |
| L69| F      | 45  | 2008 | 12,150 | NHL (DLBCL)                   | Experienced |
| L70| M      | 65  | 2009 | 2,237  | NHL (DLBCL)                   | Experienced |
| L71| M      | 46  | 2008 | >500,000 | NHL (DLBCL)                   | n.a.     |
| L72| F      | 42  | 2007 | 16,211 | HL                            | Naive    |
| L73| M      | 59  | 2009 | 388,661 | NHL (DLBCL)                   | Experienced |
| L74| M      | 23  | 2010 | 41,436 | NHL (DLBCL)                   | Naive    |
| L75| M      | 53  | 2009 | 312    | HL                            | Experienced |
| L76| M      | 39  | 2003 | 10,313 | NHL (Immunoblastic)           | Experienced |
| L77| F      | 49  | 2002 | 2,744  | NHL (Immunoblastic, Primary CNS) | Experienced |
| L78| M      | 37  | 1999 | 76,786 | NHL (Burkitt)                  | Experienced |
| L79| F      | 35  | 2001 | >500,000 | NHL (DLBCL)                   | Experienced |
| L80| M      | 41  | 2006 | 166,000 | HL                            | n.a.     |
| L81*| F     | 32  | 2001 | 190,462 | HL                            | Experienced |
| L82| M      | 46  | 2009 | 222,559 | HL                            | n.a.     |
| L83| M      | 52  | 2010 | 202,385 | HL                            | n.a.     |

F: Female; M: Male; NHL: Non-Hodgkin lymphoma; DLBCL: Diffuse Large Cell B-cell lymphoma; CNS: Central Nervous System; HL: Hodgkin lymphoma; n.a.: not available; *Same patient as L38.
Supplementary Table 2. Clinical and virological features of control HIV-1-infected patients.

| Patient no. | Sex | Age (years) | Year of sample collection | HIV load (copies/ml) | Treatment Naïve/experienced |
|-------------|-----|-------------|----------------------------|---------------------|-----------------------------|
| C1          | M   | 52          | 2019                       | 20,826              | Experienced                 |
| C2          | M   | 61          | 2019                       | 2,830,000           | Naïve                       |
| C3          | M   | 24          | 2019                       | 29,562              | Naïve                       |
| C4          | M   | 57          | 2019                       | 465,000             | n.a.                        |
| C5          | M   | 49          | 2019                       | 149,430             | Naïve                       |
| C6          | F   | 56          | 2019                       | 10,265              | Naïve                       |
| C7          | F   | 35          | 2019                       | 7,890               | Experienced                 |
| C8          | M   | 49          | 2019                       | 77,460              | Naïve                       |
| C9          | M   | 41          | 2019                       | 10,619              | Naïve                       |
| C10         | F   | 56          | 2019                       | 96,638              | Naïve                       |
| C11         | M   | 57          | 2019                       | 1,186               | Experienced                 |
| C12         | M   | 39          | 2019                       | 117                 | Naïve                       |
| C13         | F   | 55          | 2019                       | 390,717             | Experienced                 |
| C14         | M   | 54          | 2019                       | 41,300              | n.a.                        |
| C15         | M   | 53          | 2019                       | 174,013             | Naïve                       |
| C16         | M   | 55          | 2019                       | 275                 | Experienced                 |
| C17         | F   | 60          | 2019                       | 122,891             | Naïve                       |
| C18         | M   | 47          | 2019                       | 509                 | Experienced                 |
| C19         | M   | 38          | 2020                       | 93,277              | Naïve                       |
| C20         | M   | 57          | 2018                       | 9,998               | Experienced                 |
| C21         | F   | 53          | 2019                       | 25,473              | Naïve                       |
| C22         | M   | 52          | 2019                       | 38,586              | Naïve                       |
| C23         | F   | 41          | 2019                       | 18,821              | Experienced                 |
| C24         | M   | 38          | 2018                       | 15,750              | Naïve                       |
| C25         | M   | 46          | 2019                       | 23,770              | Experienced                 |
| C26         | M   | 50          | 2019                       | 67                  | Experienced                 |
| C27         | M   | 52          | 2019                       | 37                  | Experienced                 |
| C28         | M   | 69          | 2019                       | 237                 | Experienced                 |
| C29         | M   | 65          | 2019                       | 173                 | Experienced                 |
| C30         | M   | 46          | 2019                       | 2,565               | Experienced                 |
| C31         | F   | 28          | 2020                       | 32,921              | Naïve                       |
| C32         | F   | 49          | 2020                       | 7,808               | Experienced                 |
| C33         | M   | 35          | 2020                       | 20,284              | Naïve                       |
| C34         | M   | 70          | 2020                       | 47                  | Experienced                 |
| C35         | M   | 66          | 2020                       | 173,000             | n.a.                        |
| C36         | M   | 47          | 2020                       | 27,819              | Naïve                       |
| C37         | F   | 28          | 2020                       | 179,565             | n.a.                        |
| C38         | M   | 49          | 2020                       | 380,000             | n.a.                        |
| C39         | M   | 35          | 2020                       | 39                  | n.a.                        |
| C40         | F   | 91          | 2020                       | 57                  | Experienced                 |
| C41         | F   | 58          | 2020                       | 31                  | Experienced                 |
| C42         | F   | 38          | 2020                       | 32,524              | Experienced                 |
| C43         | M   | 55          | 2020                       | 4,610               | Experienced                 |
| C44         | M   | 33          | 2020                       | 110                 | Experienced                 |
| C45         | F   | 56          | 2020                       | 535                 | Experienced                 |
| C46         | M   | 55          | 2020                       | 50,940              | Experienced                 |
| C47         | M   | 70          | 2020                       | 236,247             | Experienced                 |
| C48         | M   | 48          | 2020                       | 3,707,111           | Naïve                       |
| C49         | F   | 42          | 2020                       | 41                  | Experienced                 |
| C50         | M   | 46          | 2020                       | 33,077              | Experienced                 |
| C51         | M   | 45          | 2020                       | 18,390              | Experienced                 |
| C52         | F   | 32          | 2020                       | 316                 | Naïve                       |
| C53         | F   | 62          | 2020                       | 21,748              | Experienced                 |
| C54         | F   | 51          | 2020                       | 446                 | Experienced                 |
|   |   |   |   |   |   |
|---|---|---|---|---|---|
| C55 | F | 48 | 2020 | 156 | Experienced |
| C56 | M | 53 | 2020 | 41 | Experienced |
| C57 | F | 47 | 2020 | 24,561 | Experienced |
| C58 | M | 49 | 2020 | 9,190 | Experienced |
| C59 | F | 63 | 2020 | 10,870 | Experienced |
| C60 | M | 61 | 2020 | 314 | Experienced |
| C61 | M | 52 | 2020 | 51 | Experienced |
| C62 | F | 37 | 2020 | 29,611 | Naïve |
| C63 | M | 60 | 2020 | 407,401 | Naïve |
| C64 | M | 55 | 2020 | 18,381 | n.a. |
| C65 | M | 60 | 2020 | 161,926 | Naïve |
| C66 | M | 56 | 2020 | 299 | Experienced |
| C67 | F | 58 | 2020 | 17,517 | Experienced |
| C68 | M | 42 | 2020 | 348,752 | Experienced |
| C69 | M | 47 | 2020 | 14,591 | Experienced |
| C70 | M | 53 | 2020 | 55 | Experienced |

F: Female; M: Male; n.a.: not available.
Supplementary Table 3. List of Control patients and Lymphoma patients with multiple amino acid insertions in dominant vp17 sequences.

| Patient no. | Type of 1° insertion in vp17 sequences | Position of 1° insertion in vp17 sequences | Type of 2° insertion in vp17 sequences | Position of 2° insertion in vp17 sequences |
|-------------|--------------------------------------|------------------------------------------|--------------------------------------|------------------------------------------|
| C55         | AAAGQ                                | 117-118                                  | SQAS                                 | 125-126                                  |
| C61         | AA                                   | 117-118                                  | SS                                   | 125-126                                  |
| C62         | AQQ                                  | 114-115                                  | G                                    | 127-128                                  |
| L21         | AAA                                  | 117-118                                  | SS                                   | 125-126                                  |
| L23         | AA                                   | 117-118                                  | SSQV                                 | 125-126                                  |
| L36         | AA                                   | 117-118                                  | TGNS                                 | 125-126                                  |
| L45         | ST                                   | 122-123                                  | VSQ                                  | 127-128                                  |
| L47         | A                                    | 117-118                                  | SGN                                  | 125-126                                  |
|                | 1985 - 1987 | 1988 - 1990 | 1991 - 1993 | 1994 - 1996 | 1997 - 1999 | 2000 - 2002 | 2003 - 2005 | 2006 - 2008 | 2009 - 2011 | 2012 - 2014 | 2015 - 2017 |
|----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| **wt p17**     | 103 (89.5%) | 74 (82.2%)  | 103 (88%)   | 114 (93.4%) | 139 (88.5%) | 169 (73.5%) | 401 (71.7%) | 674 (68.4%) | 637 (67.6%) | 551 (89.8%) | 41 (69.5%)  |
| **ins^{117-118}vp17** | 4 (3.6%) | 6 (6.6%) | 6 (5.2%) | 3 (2.5%) | 10 (6.4%) | 35 (15.2%) | 103 (18.4%) | 164 (16.6%) | 180 (19.1%) | 35 (5.7%) | 12 (20.3%) |
| **ins^{125-126}vp17** | 8 (6.9%) | 10 (11.2%) | 8 (6.8%) | 5 (4.1%) | 8 (5.1%) | 26 (11.3%) | 55 (9.8%) | 148 (15%) | 125 (13.3%) | 27 (4.4%) | 6 (10.2%) |