PNAS
www.pnas.org

Supplementary Information for

Epstein-Barr virus reprograms human B lymphocytes immediately in the prelatent phase of infection

Paulina Mrozek-Gorska, Alexander Buschle, Dagmar Pich, Thomas Schwarzmayr, Ron Fechtner, Antonio Scialdone, and Wolfgang Hammerschmidt

Wolfgang Hammerschmidt
Email: hammerschmidt@helmholtz-muenchen.de

Antonio Scialdone
Email: antonio.scialdone@helmholtz-muenchen.de#

This PDF file includes:

Supplementary Information (SI)
Captions for SI Datasets 1 to 5
Figures S1 to S11A,B with legends
References for SI citations

www.pnas.org/cgi/doi/10.1073/pnas.1901314116

Supplementary Information Text

Materials and Methods

Eukaryotic cell lines

B95-8 (1), Raji (2), Elijah cells (3), and LCLs (derived from EBV transformed B cells) were cultivated in RPMI-1640 medium supplemented with 8 % FCS, 100 µg/ml streptomycin, 100 U/ml penicillin, 1 mM sodium pyruvate, 100 nM sodium selenite, and 0.43 % α-thioglycerols at 37°C and 5 % CO_2. HEK293 2089 cells (4) were cultured in the same medium with 100 µg/ml hygromycin.

Virus supernatants from the HEK293 2089 cell line

HEK293 2089 on 13 cm dishes at around 80 % confluency were used for transient transfection with the two plasmids, p509 and p2670, to induce EBV's lytic cycle. p509 expresses the BZLF1 gene under the control of the CMV promoter (5), which triggers lytic phase reactivation. p2670 contains the BALF4 gene under the CMV promoter to increase the virus titer (6). 6 µg p509 and 6 µg p2670 in RPMI medium were mixed with PEI Max at a 1:6 ratio. After 15 min incubation, the mix was dropped onto the HEK293 2089 cells prepared with fresh medium without selection and the cells were incubated for another three days. The supernatants from the transfected cells were collected and centrifuged for 8 min at 300 g and at 1200 g for 8 min. The virus stocks were stored at 4°C and used to determine the virus concentration with Raji cells as described in detail (7).

Collection and quantification of virus supernatant from the B95-8 cell line

B95-8 cells were cultivated at high density to promote the spontaneous reactivation of EBV's lytic phase and virus release. Virus stocks were prepared as described above and virus titers were estimated using Elijah cells. The cells were incubated with different amounts of virus supernatant at 4°C for 3 hours and later stained with an Alexa 647 coupled antibody directed against the viral glycoprotein gp350. The fractions of gp350-positive cells with B95-8 virus stock were compared with cells incubated with a calibrated reference of 2089 virus by FACS.

Extraction of B cells from adenoid tissue

The adenoid biopsies were rinsed with PBS, transferred to a sterile petri dish, and mechanically disintegrated with two sterile scalpels and PBS. The cell suspension was filtered through a 100 µm mesh cell strainer. This procedure was repeated several times to recover a maximum of single cells. The volume of the collected cells was increased to 50 ml with PBS, the cells were mixed and sedimented at 300 g for 10 min. The cell pellet was resuspended in 30 ml PBS and supplemented with 0.5 ml defibrinated sheep blood for T cells rosetting. 15 ml Ficoll Hypaque was added underneath the cell suspension to obtain two clearly separated phases. The samples were centrifuged at 500 g for 30 min and the cells were carefully collected from the turbid interphase and transferred to a new 50 ml

tube, which was filled with PBS. The cells were washed with PBS three times, using decreasing centrifugation parameters (450, 400 and 300 g for 10 min). Finally, the cell pellet was resuspended in fresh, pre-warmed medium. The cells were counted and directly used for further experiments.

Purification of B cells on MACS columns

All the steps were performed on ice and with pre-cooled solutions. Cells obtained from adenoid biopsies were centrifuged (300 g for 10 min) and washed once with ice-cold FACS staining buffer (PBS, 0.5 % BSA, 1 mM EDTA). Untouched B cells from human PBMCs were isolated according to the manufacturer's protocol using the B Cell Isolation Kit II (Miltenyi Biotec). The collected cell fractions were stained with an appropriate antibody marker (e.g. CD19 for B cells) to control the purity of the B cell preparation.

Primary B cells infection with EBV

Infection of primary B cells was performed using wildtype EBV (2089 or B95-8) with a multiplicity of infection (MOI) 0.1 overnight. This value of MOI provides the optimal conditions with respect to the number of emerging B blasts from EBV-infected primary naïve B lymphocytes (8). The next day the cells were centrifuged (300 g at RT for 10 min) to remove unbound virus and were resuspended in fresh medium.

TMRE and Annexin V detection

For the detection of these parameters, cells from PBMC were purified using the B Cell Isolation Kit II (Miltenyi Biotec). Uninfected or infected B cells ($5x10^5$) were resuspended in 1 ml fresh pre-warmed medium and transferred to polypropylene FACS tubes. TMRE (100 nM final concentration) was added and the samples were incubated protected from light at 37°C for 30 min. The cells were washed two times with FACS buffer and resuspended in 500 µl Annexin V staining buffer together with APC-coupled Annexin V in the dark for 10 min. Daily measurements with a FACS Fortessa instrument (BD) were performed for the first eight days after EBV infection (Fig. 1A and SI Appendix, Fig. S1).

Glucose analogue uptake analysis using 2-NBDG

The experiments were performed with B cells from PBMC purified with the B Cell Isolation Kit II (Miltenyi Biotec). Non-infected cells and cells infected with wt EBV (B95-8) were collected on a daily basis as indicated (Fig. 1B). About $5x10^5$ B cells were seeded into a 24-well plate and resuspended in 500 µl of glucose-free medium supplemented with 1 % FCS. The cells were incubated with the glucose analogue 2-NBDG (2-[N-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)amino]-2-deoxy-D-glucose) (100 µM) for one hour or were left untreated (0 µM). Eventually, the cells were washed two times, resuspended in 300 µl ice-cold PBS and measured on a FACS Fortessa instrument using the FITC channel (Fig. 1B and SI Appendix, Fig. S2).

FACS sorting for naïve B cells

Primary B cells were sorted using a FACS Aria IIIu device (BD) to isolate naïve and resting human B lymphocytes from adenoid biopsies. Briefly, the cells after Ficoll gradient were washed once, resuspended in FACS staining buffer to a final concentration of $1x10^8$ cells/ml and stained using fluorophore-coupled antibodies against CD38 (eBioscience (#25-0389-42) and IgD (BD Pharmingen (#555778) for 60 min in the dark. The samples were washed once with ice-cold FACS staining buffer, resuspended in the same buffer, and filtered through a 35 µm mesh cell strainer to obtain single cell suspensions. Sorting was performed with a 70 µm nozzle, a velocity of about 8,000 events/second and a sorting mask of "4-way purity". The gating criteria included: (i) living cells, (ii) single cells, and (iii) CD38-negative and IgD-positive cells.

Prior to the subsequent analyses, viable, infected B cells were physically sorted as indicated in Fig. 1C using the 100 µm nozzle and a velocity of about 5,000 events/second with the sorting mask of "4-way purity".

Analysis of the cells' diameter

The diameter of the cells was recorded using uninfected cells and B cells infected with wt EBV at different days post infection as indicated in SI Appendix, Figure S3A with the Tali Image-Based Cytometer machine (Invitrogen). The obtained images were captured using the Quick Count program and default settings after recalibration with Tali Calibration Beads and further analysed with the Image J software. Four images with around 200 cells each were analysed at each time point. Defined objects with a circularity >0.75 and an aspect ratio <1.3 were identified as cells. The radius of each identified cell was determined according to the calibration beads with 1.9 mm radius on average, and the cells' diameter was calculated.

Cell cycle analysis

The sorted naïve B cells were infected with wt EBV (2089) at MOI 0.1. The next day cells were washed and around $5x10^5$ cells were seeded per time point. Prior to cell harvest, the cells were always incubated with 3.1 µl BrdU dilution (1:32) per 100 µl of cells and incubated at 37°C for 1 hour in the dark. Afterwards, the cells were harvested and washed with 1 ml of FACS staining buffer. The cell pellet was resuspended in 100 µl of BD Cytofix/Cytoperm buffer followed by incubation for 15-30 min on ice. Next, the cells were washed with 1 ml FACS staining buffer, then frozen in freezing medium (90 % FCS, 10 % DMSO) and stored at -80°C for further analysis. After collecting all samples, the cells were thawed, re-fixated with 100 µl BD Cytofix/Cytoperm buffer for 5 min on ice, washed with 1 ml BD Perm/Wash buffer, and subsequently resuspended in 70 µl PBS and 30 µl of DNase. The mix was incubated at 37°C for 1 h, the cells were washed with 1 ml BD Perm/Wash buffer, resuspended in 50 µl BD Perm/Wash buffer and stained with 1 µl anti-BrdU-APC coupled antibody for 20 min at RT. After staining, the cells were washed with BD Perm/Wash buffer, resuspended in 20 µl 7-AAD solution and 500µl FACS staining buffer. After 10 min the samples were measured with a FACS Fortessa instrument to determine the cell cycle distribution of the cells (Fig. 1D).

Calculating the cell division index (DI)

Sorted naïve B cells were resuspended in warm PBS at a concentration 10^6 cells/ml and mixed with diluted CellTrace violet (ThermoFischer Scientific, 1 µM final concentration). The cells were incubated in a water bath at 37°C and protected from light for 20 min. Afterwards, 50 ml of complete, pre-warmed culture medium was added to the cells to quench the unbound dye followed by 5 min incubation at 37°C. Finally, the cells were spun down at 300 g for 10 min and resuspended in fresh, pre-warmed medium. Prepared cells were infected with two wt EBV strains (2089 and 6008) at an MOI 0.1 and incubated overnight. On the next day, the cells were spun down, resuspended in fresh medium and aliquoted for daily measurements depending on the initial number of cells. Measurements were performed daily on a FACS Fortessa instrument to determine the dilution of CellTrace violet dye as an indicator of cell division. The division index DI documents the average number of cell divisions that cells in the starting population have undergone (SI Appendix, Fig. S3B). The index was calculated using the FlowJo Proliferation Tool software.

Western blot protein quantification with the Wes platform technology

Cell lysates were prepared from a defined number of cells that had been sorted for surface markers of naïve B cell (see Materials and Methods) immediately after sorting (day 0). The remaining naïve B cells were infected with EBV and an MOI of 0.1 and defined cell numbers were collected daily thereafter (day 1, 2, 3, 5, and 8 p.i.) by FACS sorting of only intact, living cells (according to forward-sideward scatter criteria). Similarly, non-infected naïve B cells that had been kept in cell culture medium were sorted for intact cells one, two and three days after their preparation from adenoid tissue. 200,000 cells per time point and sample were lysed in 20 µl of 1x Cell Lysis Buffer (Cell Signaling Technology #9803) freshly supplemented with 1mM PMSF. Briefly, cell pellets were mixed with cell lysis buffer, incubated for 10 min on ice and then sonicated (5 min, Bioruptor on HIGH settings). The lysates were cleared by high speed centrifugation and then used directly to prepare samples for Simple Western with a Wes instrument (Protein Simple, San Jose, USA). The lysates from infected and non-infected cells at the indicated time points were analyzed with antibodies directed against SYK (R&D Systems, #MAB7166; 1:50), CDK4 (GeneTex, #GTX102993, 1:50), MYC (Cell signaling, #5605, 1:50), IMPDH2 (R&D Systems, #MAB8349, 1:50), PGAM1 (Novusbio, #NBP1-49532SS, 1:400), beta-Actin (Novusbio, #NB600-501SS, 1:50), ABCG1 (abcam, #ab52617, 1:10). Antibodies were diluted using the manufacturer's Antibody Diluent 2. Protein separation was performed using the Wes 12-230 kDa separation module with prefilled plates and capillary cartridges (Protein Simple #SM-W002) followed by immunodetection with anti-mouse and anti-rabbit detection modules (Protein Simple, #DM-001 and #DM-002). Lysates were diluted with the 0.1X Sample Buffer (from separation Kit) to enable the detection with high confidence and were prepared for electrophoresis according to the manufacturer's recommendations with the instrument set for 90 min incubation with the primary antibodies. Quantification (area under the curve) and data analysis were performed using Compass for SW software (version 4.0) and the data were further normalized (according

to the dilution factor) using Excel data sheet calculations such that the signal intensities correspond to the number of cells loaded per lane (50,000 cells/lane/antibody). Next, the data were transformed to levels of relative protein expression by setting day 0 to 1 and plotted against the indicated time points.

FACS staining for surface B cell markers

Naïve B cells were sorted from B cell preparation from adenoid tissue (see Materials and Methods) and were infected with EBV (see Materials and Methods) or were kept only in culture for three days in full medium without viral infection. 2-3*10^5 cells (non-infected cells (day0) and cells collected every day after EBV infection or only culture condition) were used for single staining with different surface markers with directly coupled antibodies as indicated: CD19 (APC-conjugated, BioLegend, #302212), CD20 (PE-conjugated, BD, #556633), CD23 (PE-conjugated, BD, #555711), CD27 (APC-conjugated, BD, #337169), CD30 (APC-conjugated, BioLegend, #333909), CD38 (PE-Cy7-conjugated, eBioscience, #25-0389-42), CD62L (PE-Cy7, BioLegend, #304822), CD138 (APC-conjugated, Invitrogen, #17-1389-41), IgM (Pacific Blue-conjugated, BioLegend, #314514), and isotype control (APC- conjugated, BD, #554681) according to manufacturer's recommendations. Briefly, cells were washed once with FACS staining buffer (PBS, 0.5% BSA, 2mM EDTA) and resuspended in 50 µl same buffer. After adding antibodies, samples were mixed well and incubated at 4°C in the dark for 45-60 min. After this time, cells were washed with 1 ml FACS staining buffer and finally resuspended in 200 µl FACS staining buffer followed by measurements on FACS Fortessa instrument with indicated channels. Data were recorded at different days post EBV infection (day one to four, eight, and 12) and for non-infected cells that had been kept in culture for one, two or three days after their initial preparation. The FACS data were analyzed using the FlowJo software (version 10.4) by extracting Mean Fluorescent Intensity (MFI) signals and normalizing to the negative controls. The normalized MFI were plotted for each marker relative to the indicated time points.

Preparation of cell lysates

Preparation of whole cell extracts was performed with exactly $1x10^6$ uninfected cells or with cells collected at different time points after EBV infection. The cells were washed once with PBS (6 min at 300 g and 4°C) and lysed in a volume of 200 µl RIPA buffer for 30 minutes on ice. The lysates were sonicated with a Bioruptor Sonicator (on HIGH setting, 4 x 5 min, on ice). Afterwards, the samples were centrifuged to remove cellular debris (20 min at maximal speed at 4°C) and the supernatants were transferred to new tubes. The protein concentrations of the lysates were determined using the Bradford's protein assay (EMD Millipore) (Fig. 1E). The absorbance was measured at 595 nm on a photometer and protein concentrations were determined using a BSA standard curve.

Sample collection for time-course RNA-seq experiments

B cells were purified from adenoid biopsies and sorted to enrich for naïve B cells. The "Day 0" samples with $1x10^6$ uninfected naïve B cells were washed two times with

PBS and resuspended in 1 ml Trizol reagent. The samples were snap frozen in liquid nitrogen and stored for further analysis at -80°C. The remaining cells were bulk infected (MOI 0.1) with wt EBV (2089). On the next day, the cells were spun down to remove unbound viral particles and resuspended in fresh medium supplemented with Ciprobay (1:200, Bayer) and Cyclosporine A (1 µg/ml, Sigma). Cells were aliquoted for the harvest on different days after infection. At the indicated time points indicated in Figure 1C the cells were collected and resuspended in FACS staining buffer. Living cells were sorted according to the established gating strategy to obtain one million cells. The cells were washed two times with PBS, resuspended in Trizol reagent and snap frozen in liquid nitrogen. Frozen cells were further used for RNA extraction.

RNA isolation

Total RNA was extracted from $1x10^6$ sorted and frozen samples (day zero to five, day eight, and day 14 p.i.). Briefly, the samples in Trizol were thawed on ice and mixed with 200 µl chloroform, vortexed and centrifuged with maximal speed for 20 min at 4°C. The aqueous phase was carefully transferred to a new sterile RNase-free tube and combined with an equal volume of RNA-free absolute ethanol. Samples were then loaded onto RNeasy columns and total RNA extraction was performed using the RNAeasy Kit (Qiagen) including DNase treatment. Total RNA quality and concentration were measured by capillary electrophoresis using the Shimadzu MultiNA Electrophoresis System for DNA/RNA analysis (Fig. 1F).

Library preparation and sequencing

Different amounts of RNA were used (30 ng from Day0, 50 ng from Day1, 100 ng from Day2, and 200 ng from samples collected at Day3-5, 8, and 14) to generate the sequencing libraries. The first-strand cDNA was synthesized using oligo(dT) primer. cDNA samples were fragmented and Illumina TruSeq sequencing adapters were ligated to the 5´ and 3´ends of the cDNA fragments. Barcoded cDNA samples were finally amplified by PCR using a proof-reading polymerase. The primers used for PCR amplification were designed for TruSeq Dual-Index sequencing according to the instructions by Illumina. Aliquots of size-selected cDNAs (200-600 bp) were measured by capillary electrophoresis to confirm the correct size and to determine the concentration of each sample. All prepared libraries were sequenced together (paired-end, 100 bases) using two entire flow cells of the Illumina HiSeq4000 platform **(**Institute of Human Genetics, Helmholtz Zentrum München).

Bioinformatic analysis

Transcript quantification by Salmon

Transcripts were quantified by running Salmon version v0.9.1 (9) in the quasi-mapping-based mode. First, a transcriptome index was created for the set of human reference transcripts (GRCh37.p13) and a set of (manually assembled) 2089 EBV transcripts. Then, the quantification step was carried out with the "quant" function, correcting for the sequence-specific biases ("--seqBias" flag) and the fragment-level GC biases ("--gcBias" flag). Finally, the transcript level abundances were aggregated to gene level counts.

Quality control and data normalization

As a first assessment of sample quality, the total number of gene counts, the number of detected genes, and the ratio between the reads mapped to viral genes and those mapped to endogenous genes were computed. The quality control identified very consistent values of these quantities between samples but for sample 8_1 (SI Appendix, Fig. S4A-C). An additional PCA analysis confirmed the bias in the sample 8_1, which was excluded from the analysis (SI Appendix, Fig. S4D).

All the samples that passed the quality control were normalized for sequencing depth using size factors (10). After normalization, a principal component analysis (PCA) was performed on the log-transformed expression matrix of all genes with 1 added as pseudocount (Fig. 2A). For Figure 2B, we considered as "cell cycle genes" those included in the Gene Ontology term GO:0007049 ("cell cycle") and as "B cell activation genes" the genes differentially expressed during B cell activation (11). All genes that were present in both lists were excluded from the analysis. This resulted in 580 cell cycle genes and 7,617 B cell activation genes.

Identification of differentially expressed genes

The DESeq2 (12) R package (version 1.16.1) was used to identify differentially expressed genes between pairs of samples from different days post infection. Before running DESeq2, genes with an average expression lower than 50 normalized counts were excluded. Next, all genes that were significantly differentially expressed (DE) at a false discovery rate <0.1 and had an estimated fold-change >2 or <0.5 in at least one pair-wise comparison were selected. This resulted in a set of 11,178 DE genes, which were then clustered (see below).

Among all expressed and analysed genes with read counts ≥ 50, only 241 genes were not significantly DE between any of the time points after infection, suggesting that these were not regulated during infection. This number corresponds to 1.85 % of all genes tested (n=13,007).

MA plots visualizing the log fold change of gene expression versus mean normalized counts (SI Appendix, Fig. S7) were obtained during DESeq2 analysis. The intersections of DE genes between any given pair of time points was visualized by UpSet plots (Fig. 2C) (UpSetR R package, version 1.3.3).

Gene clustering

The list of 11,178 DE genes identified above were clustered according to their expression patterns along the infection time course. We calculated a distance matrix between genes as: $\sqrt{(1-\rho)/2}$, where ρ is the Spearman's correlation coefficient between pairs of genes across all samples (13). Hierarchical clustering was performed on this distance matrix ("hclust" function in R, with the "average" aggregation method), followed by the dynamic hybrid cut algorithm to estimate the number of clusters (dynamicTreeCut package v1.63-1, with minimum cluster size of 500 and "deepSplit" parameter equal to 0, as suggested by our robustness analysis, see below). The analysis resulted in six clusters with specific expression patterns of cellular genes along the time-course of EBV infection (Fig. 4).

We assessed the robustness of the clusters obtained from the dynamic hybrid cut algorithm at all possible values of the "deepSplit" parameter. To do so, first we computed the values of the Pearson Gamma and the average Silhouette width of the different clustering; both these parameters suggested that the best clustering was obtained with deepSplit=0, as indicated by the location of their maximum (SI Appendix, Fig. S8B,C).

We also checked how robust the clusters were to gene sub-setting. More specifically, for each value of "deepSplit", we clustered the dataset 100 times after randomly removing 10 % of genes and computed the values of the Pearson Gamma and the average Silhouette width (SI Appendix, Fig. S8D,E). Again, the most robust clusters were found with deepSplit=0. The statistics of Pearson Gamma and Silhouette width were calculated with the "clust.stats" function from the "fpc" R package (version 2.1-11.1).

The six gene clusters were visualized with a t-SNE representation ("RtSNE" R package version 0.13) of the distance matrix with perplexity set to 50 (SI Appendix, Fig. S8A).

GO analysis

Each group of genes assigned to one of the six clusters was analysed with the gene ontology (GO) online tool (GOrilla) (14) to characterize specific biological processes that are enriched in each cluster with respect to the list of all DE genes.

To increase the specificity of the enriched GO term we set the p-value threshold to $\leq 10^{-3}$ and harmonized the redundancy in identified GO terms by additional filtering with the "reduce and visualize Gene Ontology" (REViGO) online tool with settings, which defined medium similarity, database for Homo sapiens, and SimRel for calculating the semantic similarity of the GO terms (15). R scripts created by the REViGO tool were employed to create tree plots from the enriched GO terms (SI Appendix, Fig S9).

Heatmaps of log10 transformed expression data

Heatmaps in SI Appendix, Figures S6 and S10 were generated using "pheatmap" R package version 1.0.10 with default settings on log10 transformed expression data from the top 100 cellular genes contributing to the first two principal components or only viral genes (SI Appendix, Fig. S6 and S10E).

Principal Component Analysis of phenotypic data

The Principal Component Analysis (PCA) matrix was calculated based on the normalized means of the results derived from seven different read-outs: TMRE binding, 2-NBDG uptake, cell diameter, cell division index, S-phase fraction of cells, protein content, and total RNA content. The results of the TMRE uptake were expressed as ratio of the TMRE_high positive cell population versus all TMRE positive cells (TMRE_high and TMRE_low). Similarly, the fraction of cells that actively sequester 2-NBDG and the fraction of cells in the S phase were calculated. The average from three to four biological replicates comprising cell diameter, division index, protein content, and total RNA content were normalized according to the maximal value identified in the single experiments. After normalization, the matrix with all adjusted data from the seven different experiments was used to perform the PCA in R. The first two principal components (PC1 and PC2) were plotted (Fig. 1G) to identify the directions explaining most of the variance in the metabolic and phenotypic data.

PCA with viral genes

Normalized read counts were filtered considering only significant DE viral genes and the principal component analysis (PCA) was performed in R using the log-transformed expression matrix excluding the day 0 time point, where no viral genes were expressed (SI Appendix, Fig. S10B).

Clustering of viral genes

From the significantly DE genes identified by DESeq2, we selected the 27 viral genes and considered them for clustering. A distance matrix was created as explained above (see the previous paragraph "Gene clustering"). In the next step, hierarchical clustering ("hclust") and dynamic hybrid cut algorithm ("dynamicTreeCut") were used on the distance matrix (minimum cluster size equal to 5 and deepSplit=0) to estimate the identity and the number of clusters. We found three clusters that were named with Roman numerals. The expression values of viral genes were plotted as mean normalized expression versus time-points (days p.i.) after EBV infection (SI Appendix, Fig. S10D).

Data availability

Raw RNA-seq data are available at ArrayExpress. The data can be accessed via this link:

https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-7805/

Processed RNA-seq data can be downloaded from the following website, which also provides a user-friendly R Shiny App to visualize and explore the data and the expression of individual cellular and viral genes:

http://ebv-b.helmholtz-muenchen.de/

Code availability

All data were analyzed with standard programs and packages, as detailed above. Code is available on request.

Captions for Datasets 1 to 5

SI Dataset 1. Top 1,000 genes with their contributions to the first principal component and the corresponding cumulative contribution in the PCA of Figure 2A.

SI Dataset 2. Top 1,000 genes with their contributions to the second principal component and the corresponding cumulative contribution in the PCA of Figure 2A.

SI Dataset 3. Contributions of viral genes to the first principal component in the PCA of SI Appendix, Figure 10B.

SI Dataset 4. Contributions of viral genes to the second principal component in the PCA of SI Appendix, Figure 10B.

SI Dataset 5. List of B cell activation genes (from ref. 11) that are differentially expressed along the EBV infection time course, with the corresponding cluster they are in. Genes having "NA" in the "cluster" column are not differentially expressed along the infection time course.

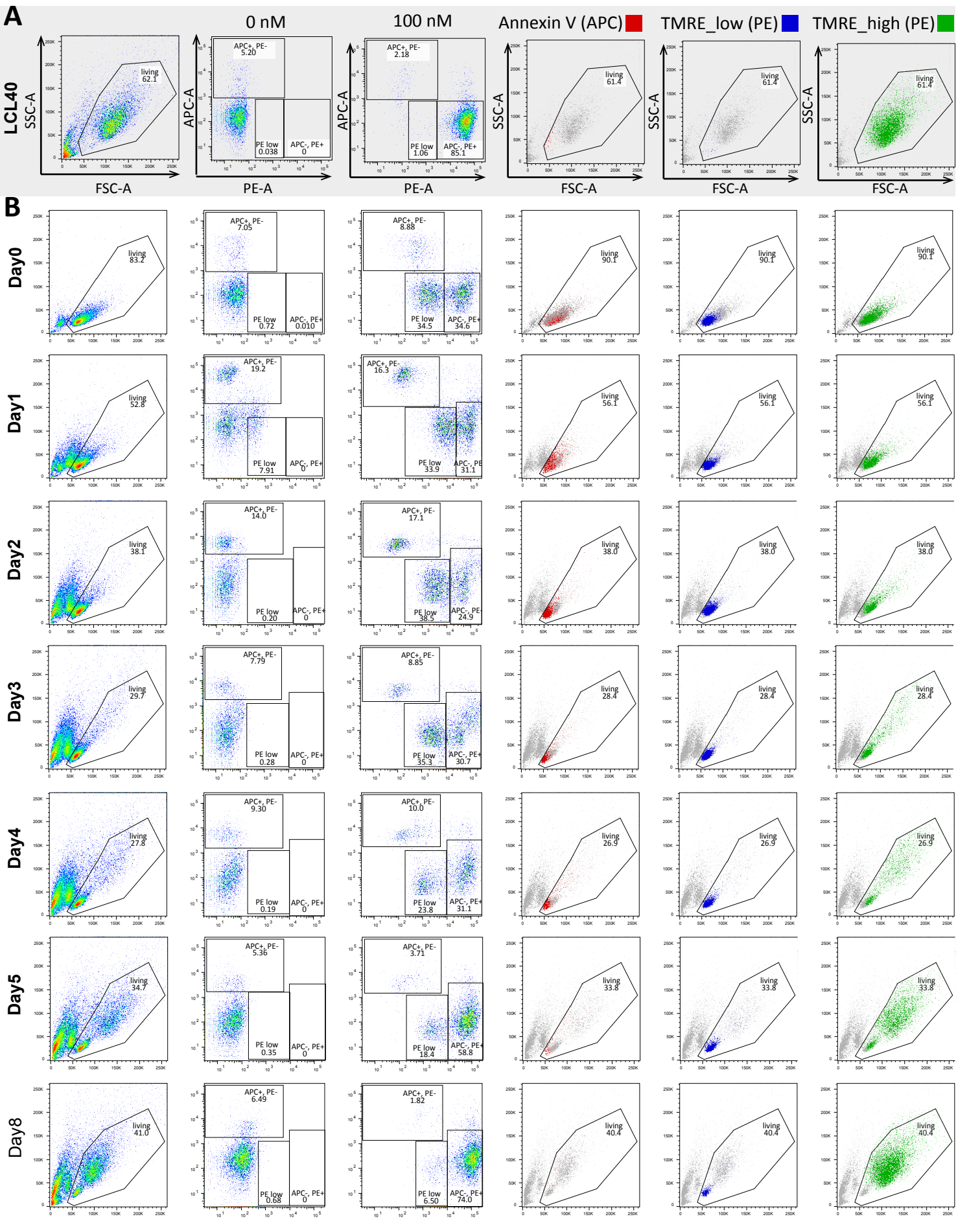

A
0 nM
100 nM
Annexin V (APC)
TMRE_low (PE)
TMRE_high (PE)
LCL40
SSC-A
FSC-A
APC-A
PE-A
living 62.1
APC+, PE- 5.20
PE low 0.038
APC-, PE+ 0
APC+, PE- 2.18
PE low 1.06
APC-, PE+ 85.1
living 61.4
B
Day0
living 83.2
APC+, PE- 7.05
PE low 0.72
APC-, PE+ 0.010
APC+, PE- 8.88
PE low 34.5
APC-, PE+ 34.6
living 90.1
Day1
living 52.8
APC+, PE- 19.2
PE low 7.91
APC-, PE+ 0
APC+, PE- 16.3
PE low 33.9
APC-, PE+ 31.1
living 56.1
Day2
living 38.1
APC+, PE- 14.0
PE low 0.20
APC-, PE+ 0
APC+, PE- 17.1
PE low 36.5
APC-, PE+ 24.9
living 38.0
Day3
living 29.7
APC+, PE- 7.79
PE low 0.28
APC-, PE+ 0
APC+, PE- 8.85
PE low 35.3
APC-, PE+ 30.7
living 28.4
Day4
living 27.8
APC+, PE- 9.30
PE low 0.19
APC-, PE+ 0
APC+, PE- 10.0
PE low 23.8
APC-, PE+ 31.1
living 26.9
Day5
living 34.7
APC+, PE- 5.36
PE low 0.35
APC-, PE+ 0
APC+, PE- 3.71
PE low 18.4
APC-, PE+ 58.8
living 33.8
Day8
living 41.0
APC+, PE- 6.49
PE low 0.68
APC-, PE+ 0
APC+, PE- 1.82
PE low 6.50
APC-, PE+ 74.0
living 40.4

Figure S1

Fig. S1. Annexin V binding and mitochondrial activity during the first eight days after EBV infection.

The FACS plots show the rate of early apoptosis (Annexin V) and mitochondrial activity (TMRE) of the cells at early time points after EBV infection. The percentages of positive cells measured by FACS are indicated.

This first row of panels shows an established LCLs as a control, the remaining rows display uninfected primary B cells (Day0) or cells on different days p.i.. In the first column, the forward and sideward scatter of the different cell samples are compared. The next two columns analyze untreated cells (0 nM) and cells incubated with 100 nM TMRE for 30 min. Columns four to six show the results after back-gating of the cell populations indicating the fraction of cells that were Annexin V positive (red), TMRE_low positive (blue) or TMRE_high positive (green). TMRE incorporation revealed two clearly positive populations in uninfected B cells (Day0). The population of TMRE_high cells became dominant from day four p.i. indicating a high mitochondrial activity of stably infected cells. The figure shows the results of one representative experiment out of four biological replicates.

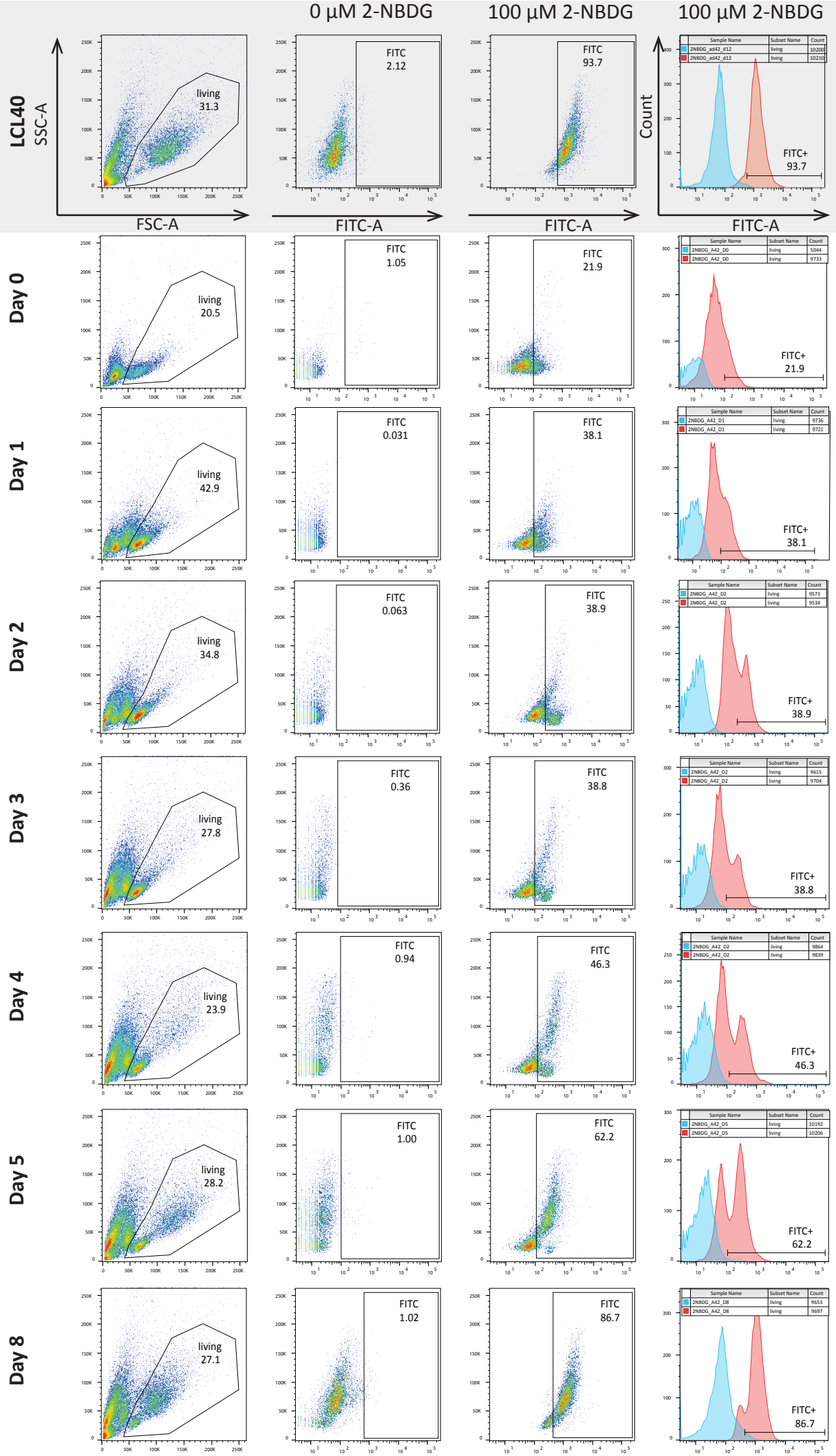

0 µM 2-NBDG
100 µM 2-NBDG
100 µM 2-NBDG
LCL40
SSC-A
FSC-A
FITC-A
Count
Day 0
living 20.5
FITC 1.05
FITC 21.9
FITC+ 21.9
Day 1
living 42.9
FITC 0.031
FITC 38.1
FITC+ 38.1
Day 2
living 34.8
FITC 0.063
FITC 38.9
FITC+ 38.9
Day 3
living 27.8
FITC 0.36
FITC 38.8
FITC+ 38.8
Day 4
living 23.9
FITC 0.94
FITC 46.3
FITC+ 46.3
Day 5
living 28.2
FITC 1.00
FITC 62.2
FITC+ 62.2
Day 8
living 27.1
FITC 1.02
FITC 86.7
FITC+ 86.7

Figure S2

Fig. S2. Uptake of the glucose analogue (2-NBDG) during B cells infection with EBV.

FACS analysis of the glucose uptake into B cells at different days after EBV infection. Panel A shows an established LCL as a reference, the remaining B panels represent uninfected B cells and cells infected for different time periods as indicated. The first column compares forward and sideward scatter of cells incubated with 2-NBDG. The next three columns show the analysis of untreated cells (0 μM 2-NBDG) and cells incubated with 100 μM of 2-NBDG. During the first three days after viral infection the cells did not show significant changes in glucose uptake. Starting from day four p.i., the glucose uptake increased substantially to more than 85 % cells with high glucose uptake. The figure depicts one experiment out of four performed with B cells from different donors.

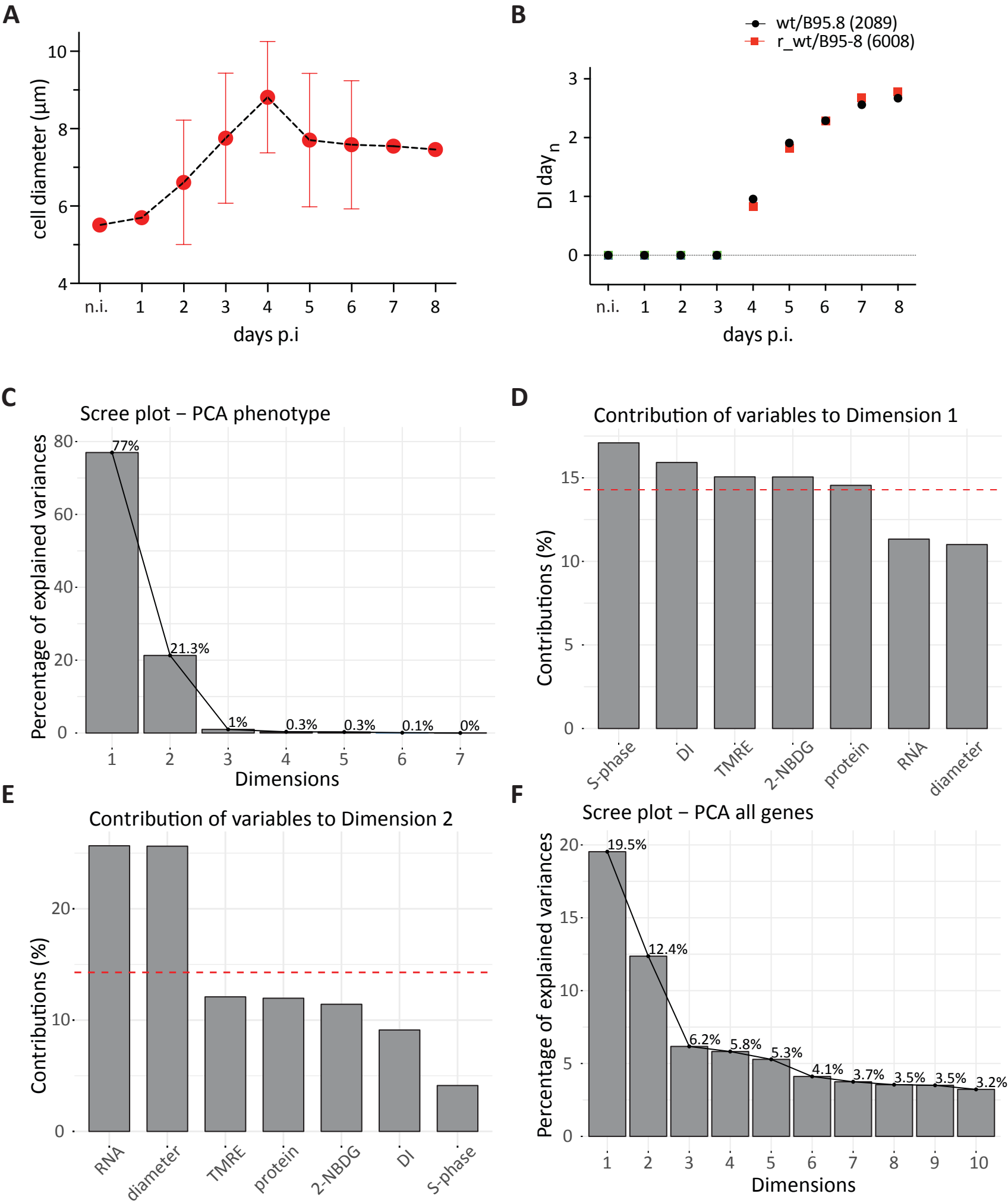

A
cell diameter (µm)
days p.i
B
wt/B95.8 (2089)
r_wt/B95-8 (6008)
DI day n
days p.i.
C
Scree plot – PCA phenotype
Percentage of explained variances
77%
21.3%
1%
0.3%
0.3%
0.1%
0%
Dimensions
D
Contribution of variables to Dimension 1
Contributions (%)
S-phase
DI
TMRE
2-NBDG
protein
RNA
diameter
E
Contribution of variables to Dimension 2
Contributions (%)
RNA
diameter
TMRE
protein
2-NBDG
DI
S-phase
F
Scree plot – PCA all genes
Percentage of explained variances
19.5%
12.4%
6.2%
5.8%
5.3%
4.1%
3.7%
3.5%
3.5%
3.2%
Dimensions

Figure S3

Fig. S3. Cell size and dynamics of cell divisions during EBV infection.

A. The panel shows the changes of the cellular diameter during the first eight days after EBV infection. Mean and standard deviation were calculated from the images of about thousand cells at the specified time points p.i..

B. Sorted naïve B cells were treated with the intracellular dye CellTrace Violet (CTV) and subsequently infected with two wild type strains of EBV, wt/B95-8 (2089) or r_wt/B95-8 (6008), with an MOI 0.1. On a daily basis, the cells were analyzed by FACS and based on the dilution of the CTV the Division Index (DI) was determined and plotted. The graph shows one representative example out of three experiments with different B cell donors.

C. Scree plot for the PCA of the phenotypic data shown in Figure 1G.

D. and E. Contributions of each variable to the first (D) and second (E) principal components in the PCA shown in Figure 1G. The horizontal red dashed line marks the value of contribution if all variables contributed equally.

F. Scree plot for the PCA of all genes shown in Figure 2A.

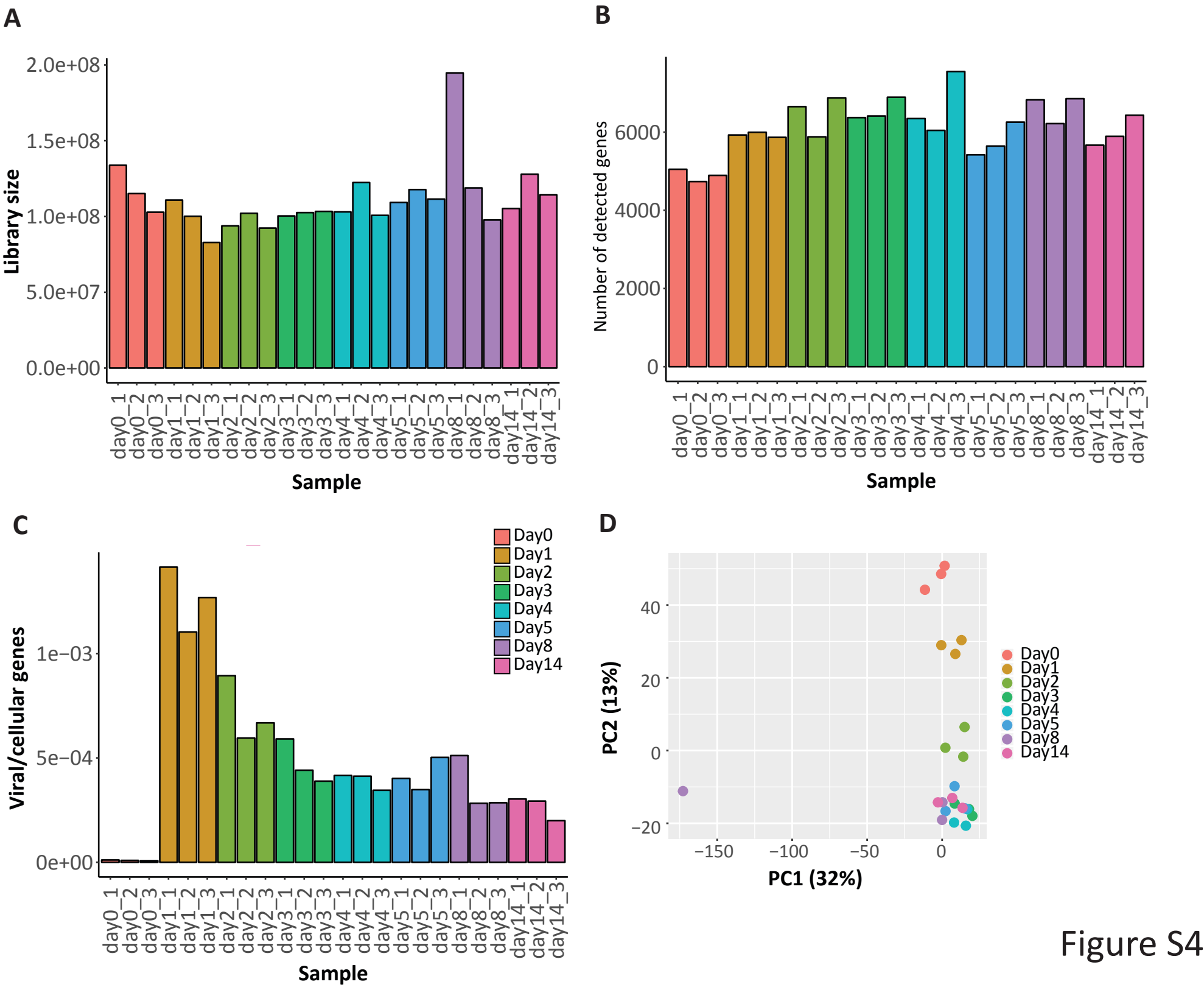

A
Library size
2.0e+08
1.5e+08
1.0e+08
5.0e+07
0.0e+00
Sample
B
Number of detected genes
6000
4000
2000
0
Sample
C
Viral/cellular genes
1e−03
5e−04
0e+00
Sample
Day0
Day1
Day2
Day3
Day4
Day5
Day8
Day14
D
PC2 (13%)
40
20
0
−20
−150
−100
−50
0
PC1 (32%)
Day0
Day1
Day2
Day3
Day4
Day5
Day8
Day14

Figure S4

Fig S4. Quality control of sequencing libraries.

A. The quality control graph provides information about the size of the sequenced libraries prepared from the different samples from three independent donors. The samples show comparable sequencing depth with the exception of Day8_1. This sample was further excluded from analysis to minimize a possible bias from this data. See also panel D of this figure.

B. The bar plot depicts the numbers of detected genes with a threshold of > 10 reads per million (RPM) in each sequenced sample. The numbers of detected genes were comparable.

C. Ratios of the numbers of viral genes versus the numbers of cellular genes at each day for three replicates are shown. Viral genes were absent in uninfected cells (Day0) as expected, in contrast to day one and two, when the ratio of detected viral genes was very high. From day three on, the numbers of detected EBV genes decreased to reach a stable level at day four p.i..

D. Principal Component Analysis (PCA) performed on sequencing results from different time points p.i.. Each dot represents a biological replicate and different colors show different days after EBV infection as indicated. The percentage of total variances explained by the first two principal components is indicated in parentheses. The graph displays the bias introduced by sample 8_1, which determines the whole variances in PC1.

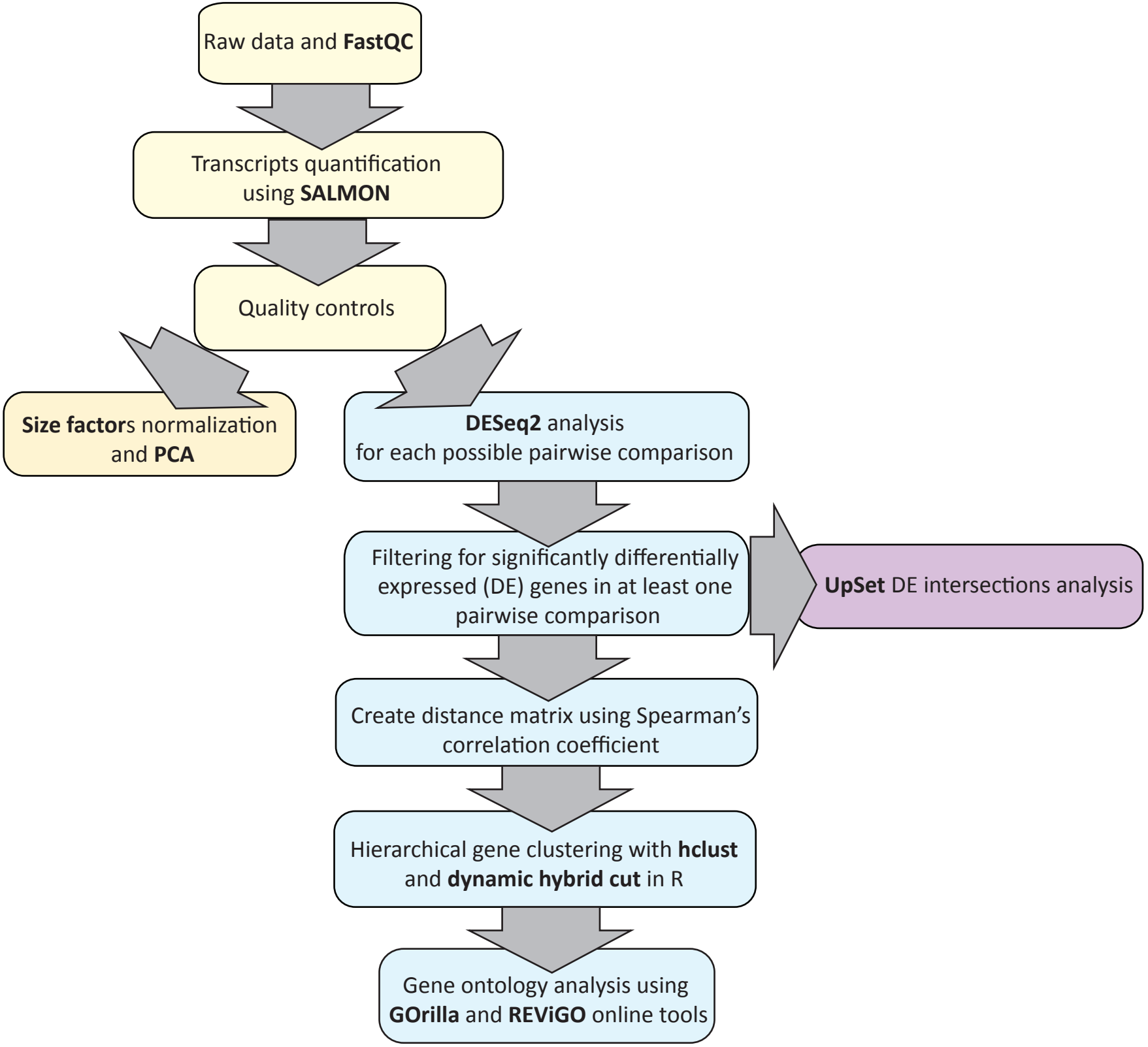
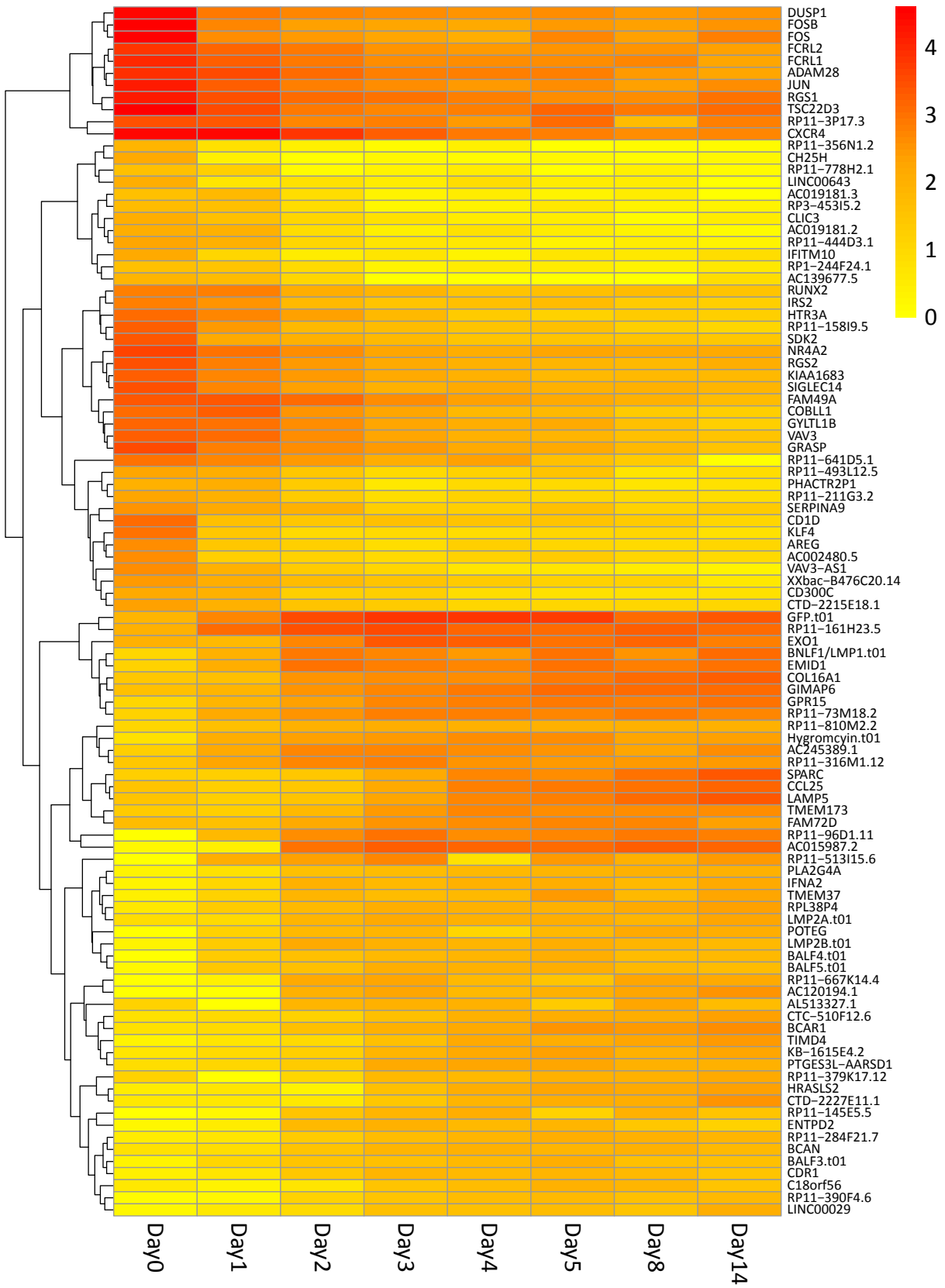

Raw data and **FastQC**
Transcripts quantification using **SALMON**
Quality controls
Size factors normalization and **PCA**
DESeq2 analysis for each possible pairwise comparison
Filtering for significantly differentially expressed (DE) genes in at least one pairwise comparison
UpSet DE intersections analysis
Create distance matrix using Spearman's correlation coefficient
Hierarchical gene clustering with **hclust** and **dynamic hybrid cut** in R
Gene ontology analysis using **GOrilla** and **REViGO** online tools

Figure S5

Fig. S5. Schematic workflow of the bioinformatics analyses after RNA-seq.

The scheme shows the individual steps of the bioinformatic workflow starting with the raw data analysis and transcript quantification, followed by data normalization and PCA analysis. Finally, DE analysis together with an Upset intersection analysis and genes clustering concluded the analyses (more details can be found in Materials and Methods).

A


PC1
log10 expression
4
3
2
1
0
DUSP1
FOSB
FOS
FCRL2
FCRL1
ADAM28
JUN
RGS1
TSC22D3
RP11−3P17.3
CXCR4
RP11−356N1.2
CH25H
RP11−778H2.1
LINC00643
AC019181.3
RP3−453I5.2
CLIC3
AC019181.2
RP11−444D3.1
IFITM10
RP1−244F24.1
AC139677.5
RUNX2
IRS2
HTR3A
RP11−158I9.5
SDK2
NR4A2
RGS2
KIAA1683
SIGLEC14
FAM49A
COBLL1
GYLTL1B
VAV3
GRASP
RP11−641D5.1
RP11−493L12.5
PHACTR2P1
RP11−211G3.2
SERPINA9
CD1D
KLF4
AREG
AC002480.5
VAV3−AS1
XXbac−B476C20.14
CD300C
CTD−2215E18.1
GFP.t01
RP11−161H23.5
EXO1
BNLF1/LMP1.t01
EMID1
COL16A1
GIMAP6
GPR15
RP11−73M18.2
RP11−810M2.2
Hygromcyin.t01
AC245389.1
RP11−316M1.12
SPARC
CCL25
LAMP5
TMEM173
FAM72D
RP11−96D1.11
AC015987.2
RP11−513I15.6
PLA2G4A
IFNA2
TMEM37
RPL38P4
LMP2A.t01
POTEG
LMP2B.t01
BALF4.t01
BALF5.t01
RP11−667K14.4
AC120194.1
AL513327.1
CTC−510F12.6
BCAR1
TIMD4
KB−1615E4.2
PTGES3L−AARSD1
RP11−379K17.12
HRASLS2
CTD−2227E11.1
RP11−145E5.5
ENTPD2
RP11−284F21.7
BCAN
BALF3.t01
CDR1
C18orf56
RP11−390F4.6
LINC00029
Day0
Day1
Day2
Day3
Day4
Day5
Day8
Day14

B

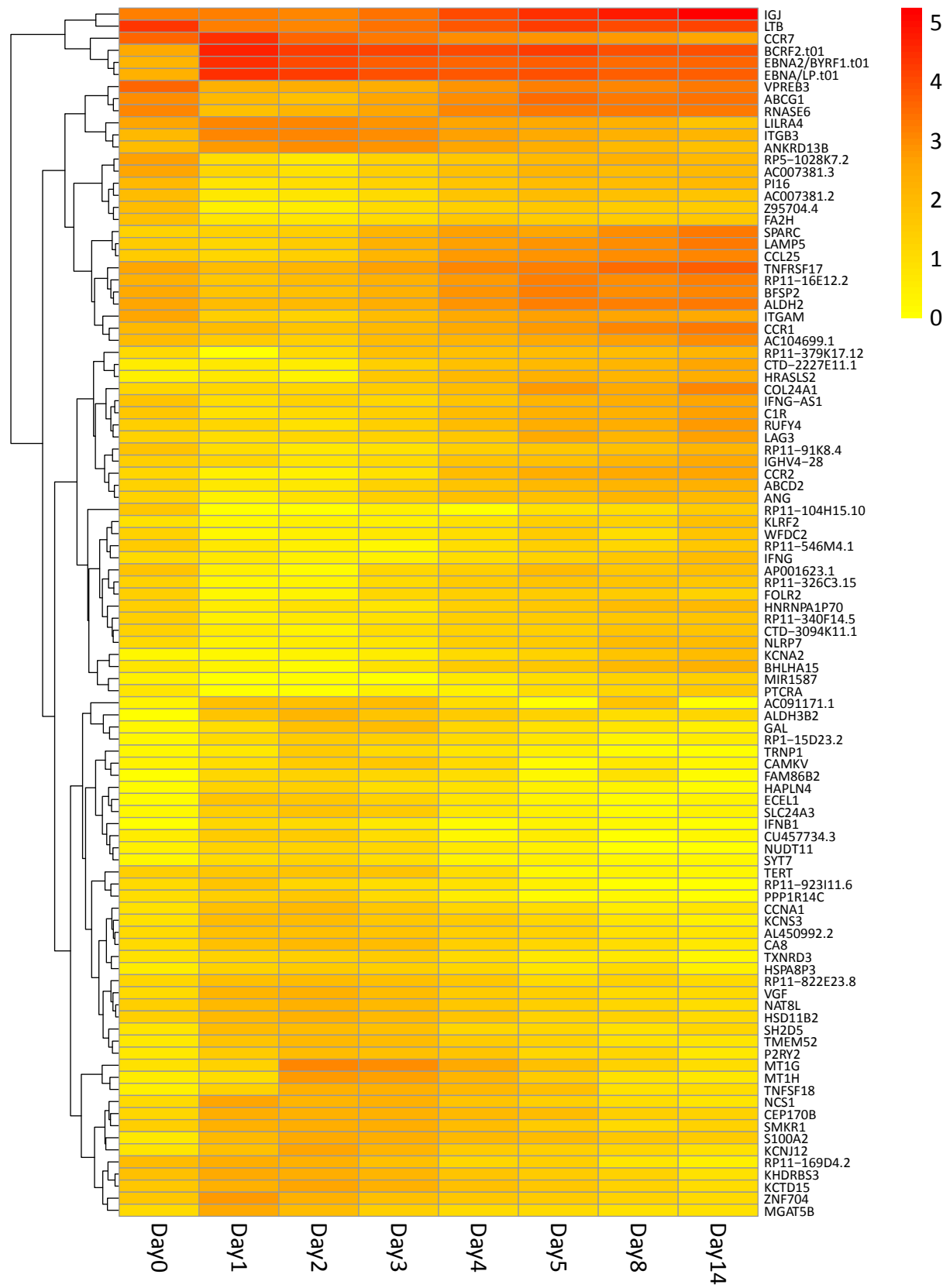

PC2
log10 expression
5
4
3
2
1
0
IGJ
LTB
CCR7
BCRF2.t01
EBNA2/BYRF1.t01
EBNA/LP.t01
VPREB3
ABCG1
RNASE6
LILRA4
ITGB3
ANKRD13B
RP5−1028K7.2
AC007381.3
PI16
AC007381.2
Z95704.4
FA2H
SPARC
LAMP5
CCL25
TNFRSF17
RP11−16E12.2
BFSP2
ALDH2
ITGAM
CCR1
AC104699.1
RP11−379K17.12
CTD−2227E11.1
HRASLS2
COL24A1
IFNG−AS1
C1R
RUFY4
LAG3
RP11−91K8.4
IGHV4−28
CCR2
ABCD2
ANG
RP11−104H15.10
KLRF2
WFDC2
RP11−546M4.1
IFNG
AP001623.1
RP11−326C3.15
FOLR2
HNRNPA1P70
RP11−340F14.5
CTD−3094K11.1
NLRP7
KCNA2
BHLHA15
MIR1587
PTCRA
AC091171.1
ALDH3B2
GAL
RP1−15D23.2
TRNP1
CAMKV
FAM86B2
HAPLN4
ECEL1
SLC24A3
IFNB1
CU457734.3
NUDT11
SYT7
TERT
RP11−923I11.6
PPP1R14C
CCNA1
KCNS3
AL450992.2
CA8
TXNRD3
HSPA8P3
RP11−822E23.8
VGF
NAT8L
HSD11B2
SH2D5
TMEM52
P2RY2
MT1G
MT1H
TNFSF18
NCS1
CEP170B
SMKR1
S100A2
KCNJ12
RP11−169D4.2
KHDRBS3
KCTD15
ZNF704
MGAT5B
Day0
Day1
Day2
Day3
Day4
Day5
Day8
Day14

Figure S6

Fig. S6. Heatmaps of the top 100 genes that contribute to PC1 or PC2.

We considered 50 genes with the most positive and 50 genes with the most negative loadings of the first two principal components (PC1 and PC2). The log10 transformed mean expression values of the combined 100 genes (on the right side) were plotted onto the heatmap at different days after EBV infection (bottom). Panel A represents genes selected from PC1 and panel B displays genes contributing to PC2. Color coding from yellow to red generates a range of transformed expression values from the lowest to the highest, respectively.

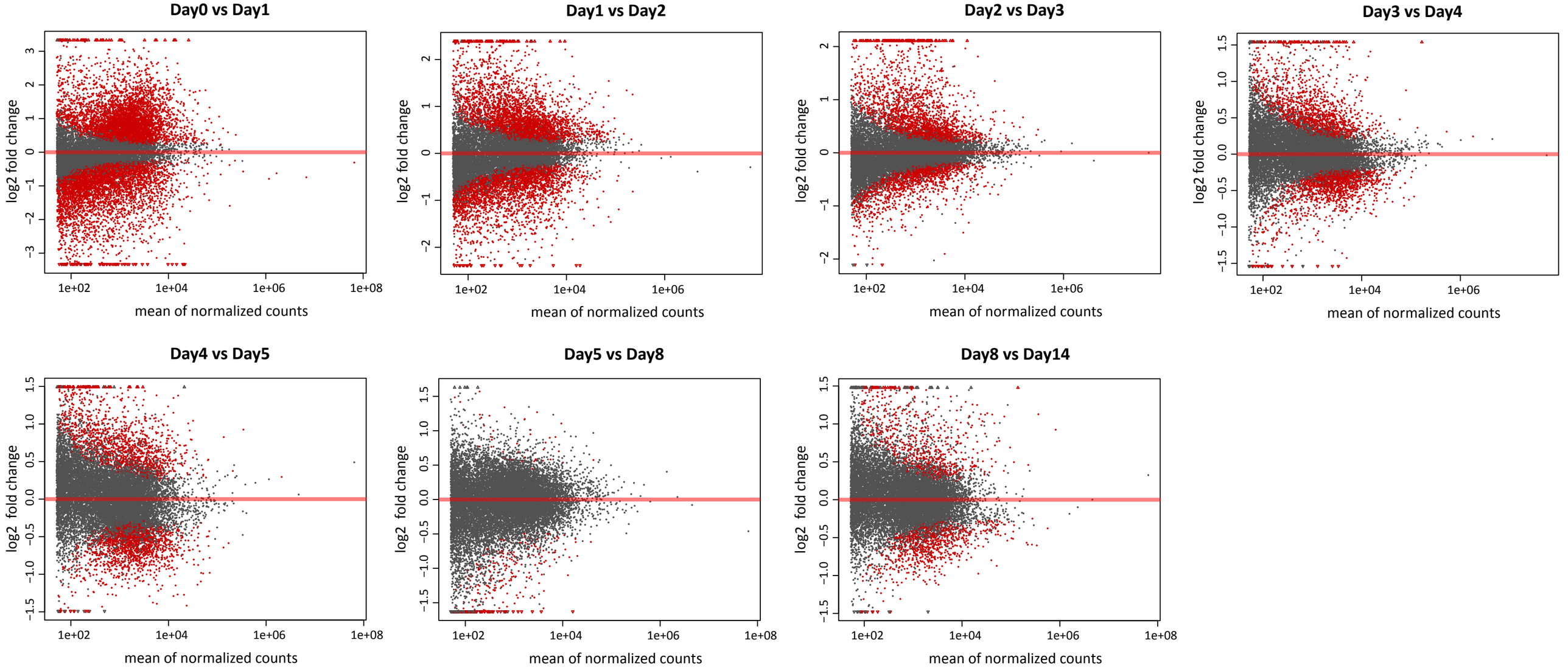

Day0 vs Day1
Day1 vs Day2
Day2 vs Day3
Day3 vs Day4
Day4 vs Day5
Day5 vs Day8
Day8 vs Day14
log2 fold change
mean of normalized counts
1e+02
1e+04
1e+06
1e+08

Figure S7

Fig. S7. MA plots of gene expression on consecutive days after EBV infection.

The MA plots show the comparison of gene expression levels between the consecutive days p.i (Day0 vs Day1, Day1 vs Day2, etc.). The x-axes display the means of normalized counts and the y-axes depict the log fold changes between the time points indicated above the plot. Every dot represents a single gene, genes that passed the significance threshold (FDR) <0.1 are shown in red.

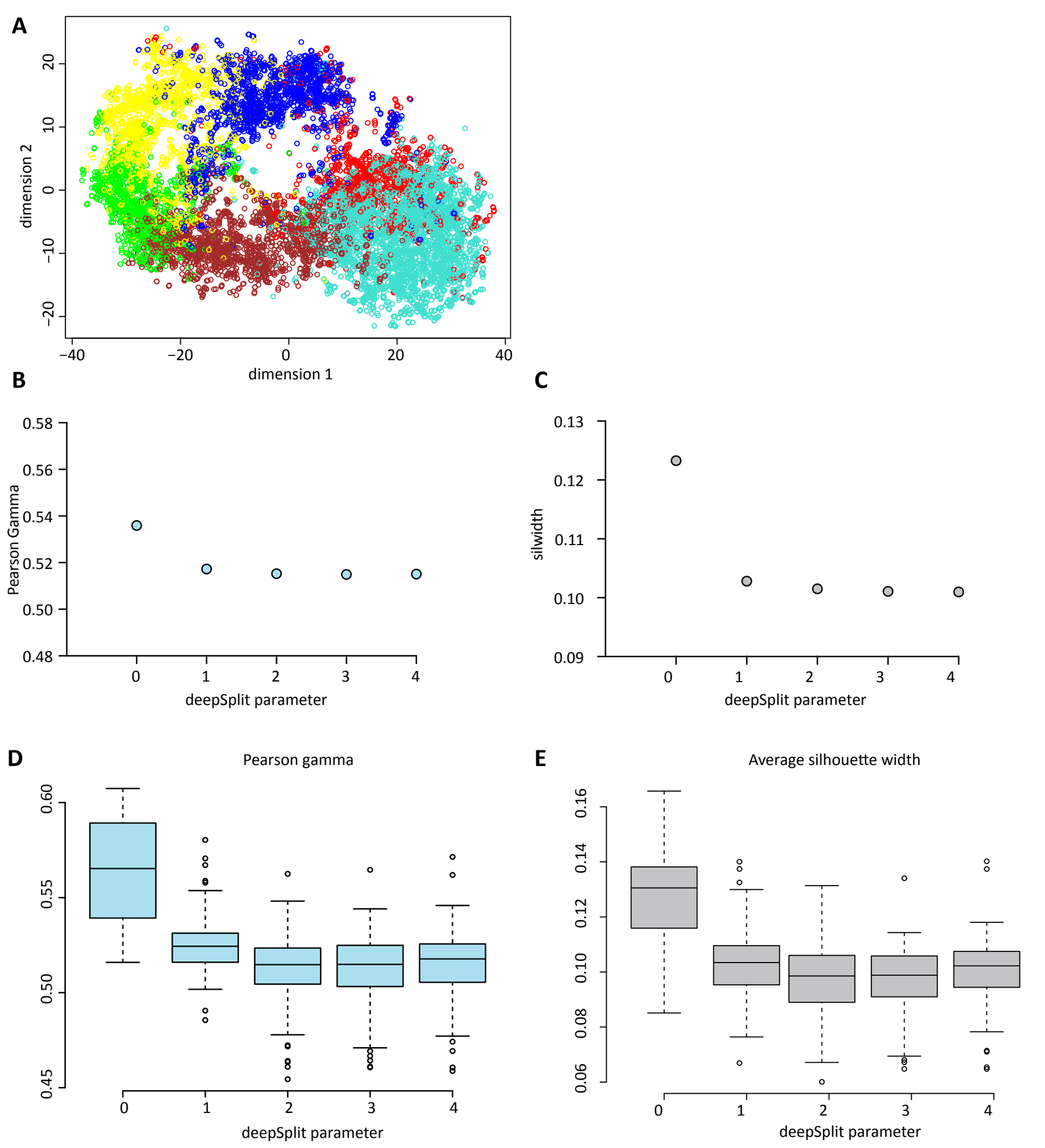

A
dimension 2
dimension 1
B
Pearson Gamma
deepSplit parameter
C
silwidth
deepSplit parameter
D
Pearson gamma
deepSplit parameter
E
Average silhouette width
deepSplit parameter

Figure S8

Fig. S8. Clustering robustness analysis.

A. t-SNE plot of all identified DE genes in the course of EBV infection. The colors indicate the six clusters estimated by the hierarchical clustering and the dynamic hybrid cut algorithm.

B. and C. The dynamic hybrid cut algorithm was used with all possible values of the "deepSplit" parameter with all DE genes. The values of the Pearson Gamma (PG) (B) and silhouette width (silwidth) (C) were calculated using the R function "cluster.stats" from the "fpc" package and plotted against the different values of the "deepSplit" parameters. PG provides an approximation of the dissimilarity in the clustering, so when genes are in different clusters they should strongly correlate with high dissimilarity. Average silwidth emphasises the separation between the clusters and also their neighbouring clusters. Higher values PG and silwidth correspond to better clustering configurations. Here, PG and silwidth parameters for "deepSplit"=0 suggest the best clustering configuration according to the specificity and sensitivity of identified clusters.

D. and E. The same analysis as in panel B and C was performed with 100 randomly selected sub-samples of genes. The dynamic hybrid cut was performed and clustering quality was assessed by Pearson Gamma and silhouette width parameters. The x-axis of the plots depicts all used "deepSplit" values, whereas the y-axis displays values for PG (D) or silwidth (E). Both boxplots represent the median and Inter Quartile Range (IQR) of 50 % of the data. The upper whisker defines the highest value according to the area of +1.5xIQR, whereas lower whisker indicates the lowest values within -1.5xIQR. The outliers are represented as circles. Based also on this randomized analysis, the "deepSplit"=0 gives the highest values for both, PG and silwidth, which confirmed the good clustering quality and the robustness of the clusters to changes in the number of genes.

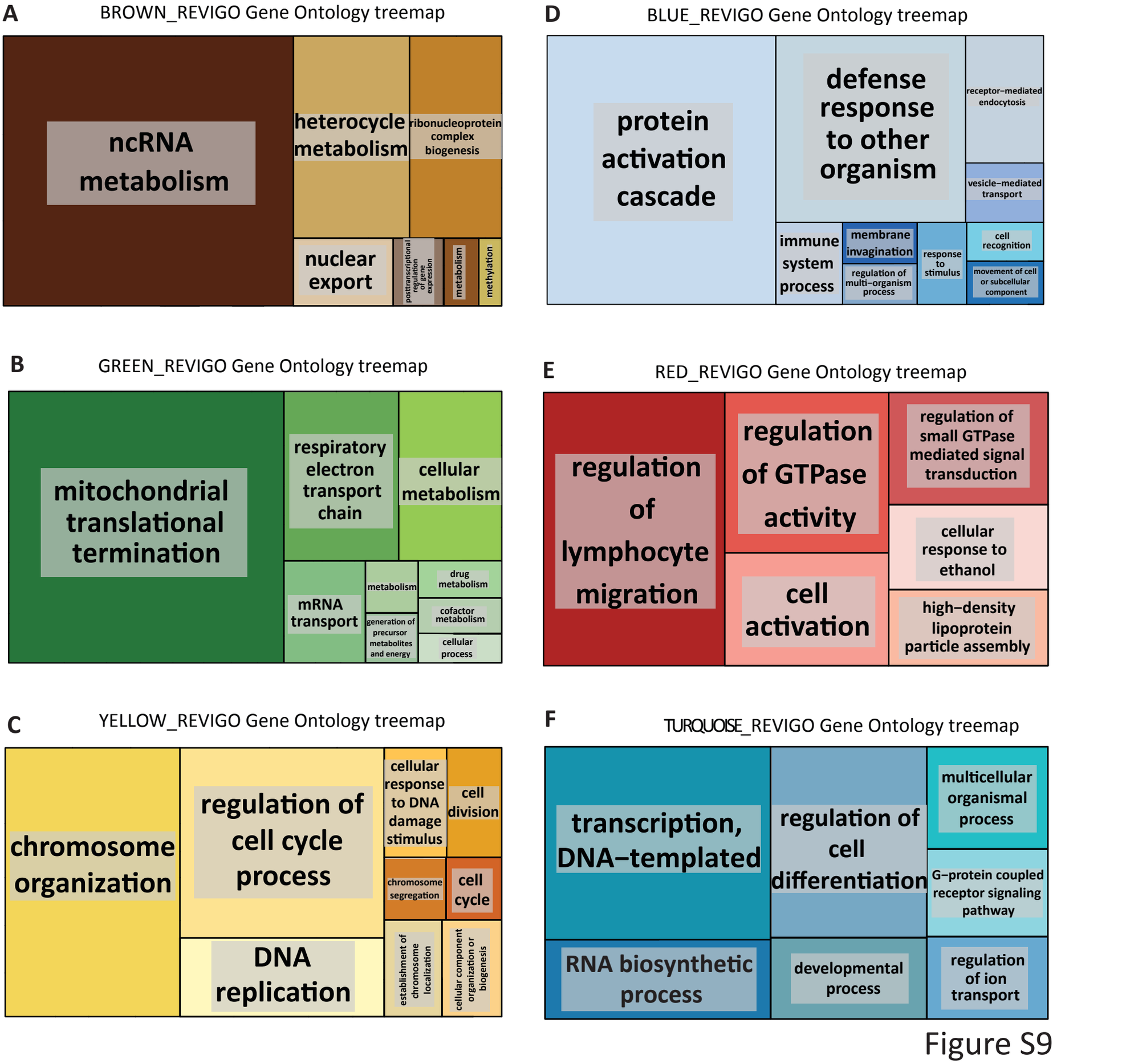
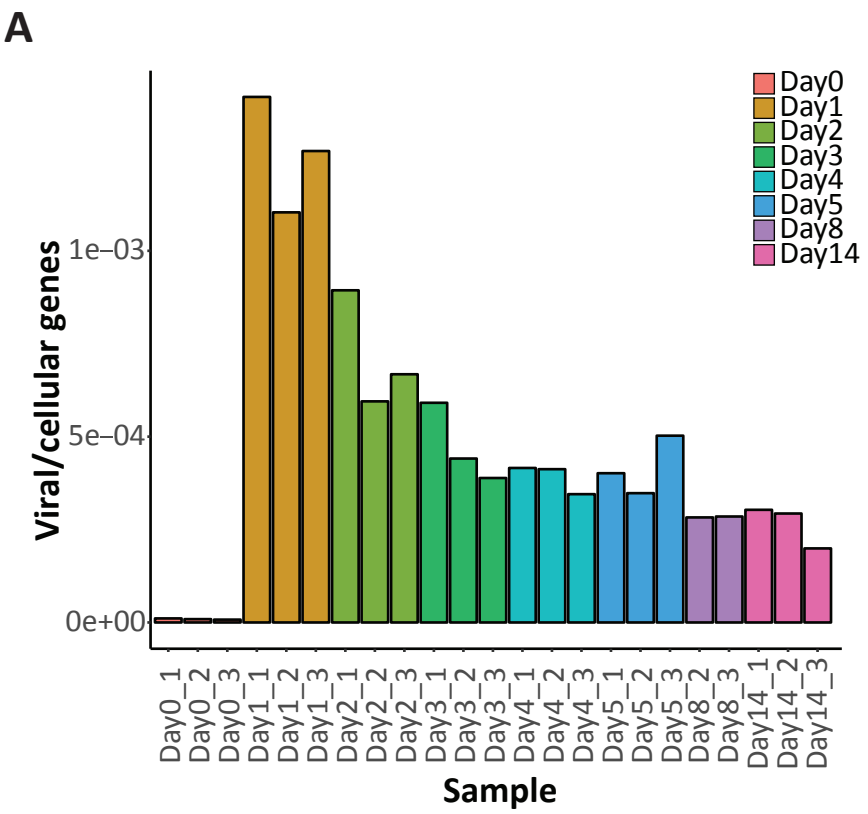
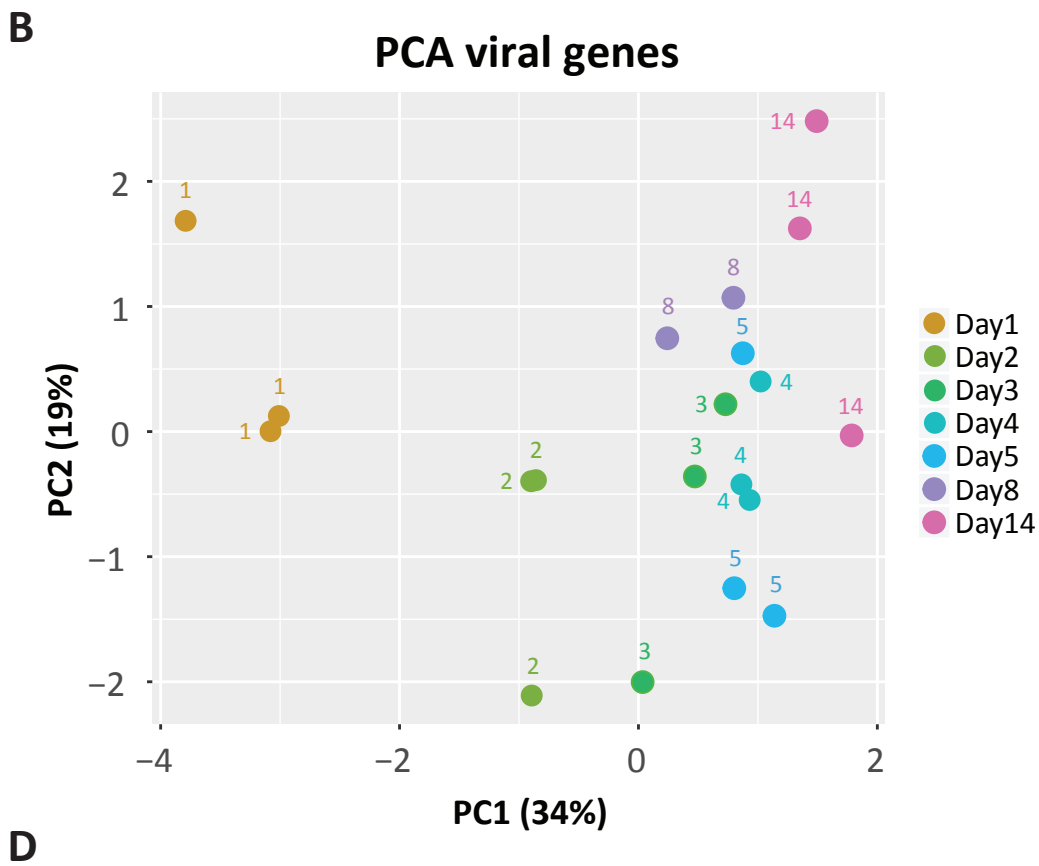

A
BROWN_REVIGO Gene Ontology treemap
ncRNA metabolism
heterocycle metabolism
ribonucleoprotein complex biogenesis
nuclear export
posttranscriptional regulation of gene expression
metabolism
methylation
D
BLUE_REVIGO Gene Ontology treemap
protein activation cascade
defense response to other organism
receptor−mediated endocytosis
vesicle−mediated transport
immune system process
membrane invagination
regulation of multi−organism process
response to stimulus
cell recognition
movement of cell or subcellular component
B
GREEN_REVIGO Gene Ontology treemap
mitochondrial translational termination
respiratory electron transport chain
cellular metabolism
mRNA transport
metabolism
drug metabolism
cofactor metabolism
generation of precursor metabolites and energy
cellular process
E
RED_REVIGO Gene Ontology treemap
regulation of lymphocyte migration
regulation of GTPase activity
regulation of small GTPase mediated signal transduction
cellular response to ethanol
cell activation
high−density lipoprotein particle assembly
C
YELLOW_REVIGO Gene Ontology treemap
chromosome organization
regulation of cell cycle process
cellular response to DNA damage stimulus
cell division
chromosome segregation
cell cycle
DNA replication
establishment of chromosome localization
cellular component organization or biogenesis
F
TURQUOISE_REVIGO Gene Ontology treemap
transcription, DNA−templated
regulation of cell differentiation
multicellular organismal process
G−protein coupled receptor signaling pathway
RNA biosynthetic process
developmental process
regulation of ion transport

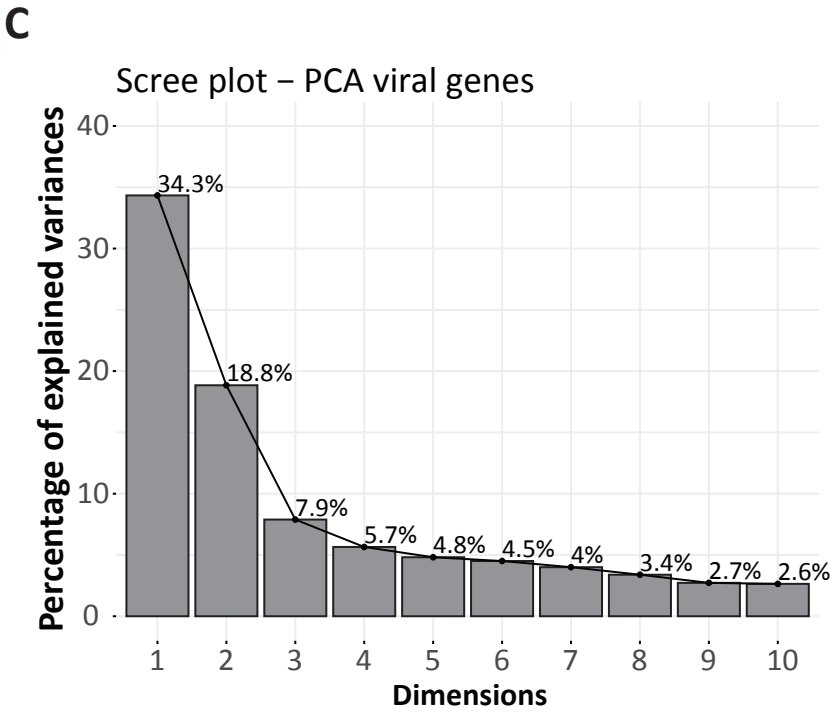
Figure S9

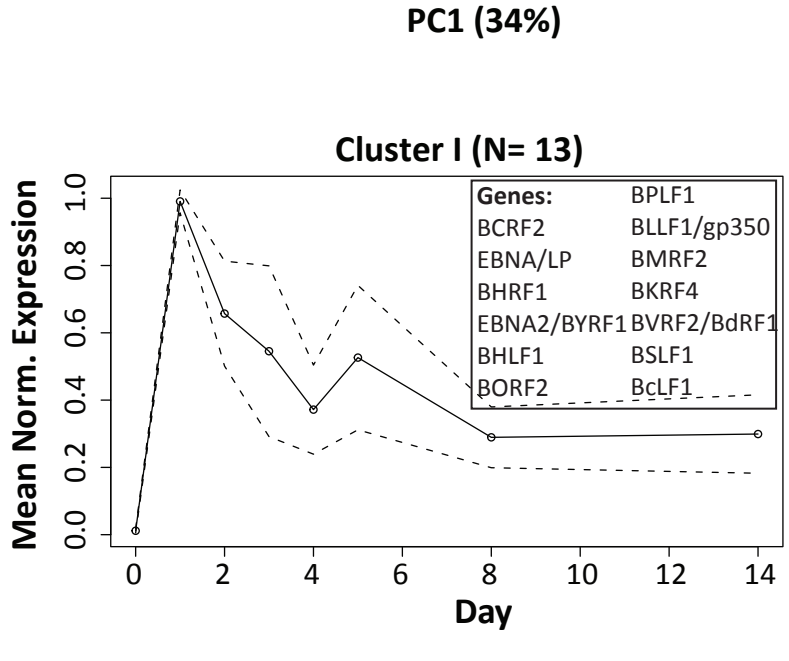
Fig. S9. Visualization of cluster-specific GO enrichments.

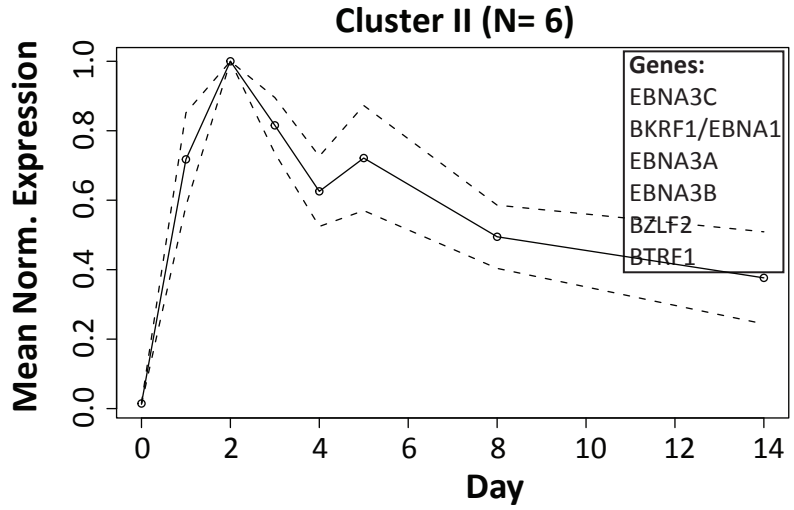
The treemaps represent the GO enrichment analyses after reducing its redundancy using REViGO (http://revigo.irb.hr/). The plots contain rectangles of different sizes indicating the absolute log10 of the calculated p-values of the shown GO terms. The color background of the tree maps relates to the color name of the clusters as indicated: the brown cluster (panel A), the green cluster (panel B), the yellow cluster (panel C), the blue cluster (panel D), the red cluster (panel E), and the turquoise cluster (panel F).

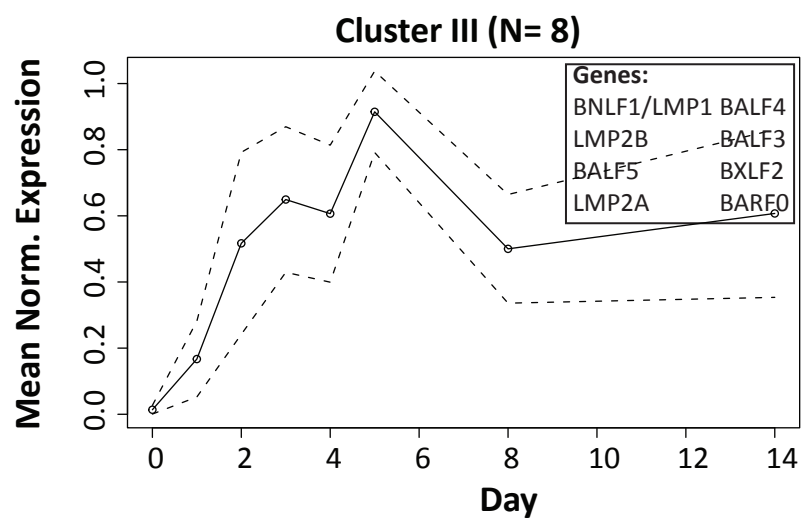

A
Viral/cellular genes
1e−03
5e−04
0e+00
Day0
Day1
Day2
Day3
Day4
Day5
Day8
Day14
Day0_1
Day0_2
Day0_3
Day1_1
Day1_2
Day1_3
Day2_1
Day2_2
Day2_3
Day3_1
Day3_2
Day3_3
Day4_1
Day4_2
Day4_3
Day5_1
Day5_2
Day5_3
Day8_2
Day8_3
Day14_1
Day14_2
Day14_3
Sample


B
PCA viral genes
PC2 (19%)
PC1 (34%)
Day1
Day2
Day3
Day4
Day5
Day8
Day14


C
Scree plot − PCA viral genes
Percentage of explained variances
34.3%
18.8%
7.9%
5.7%
4.8%
4.5%
4%
3.4%
2.7%
2.6%
Dimensions


D
Cluster I (N= 13)
Mean Norm. Expression
Day
Genes:
BCRF2
EBNA/LP
BHRF1
EBNA2/BYRF1
BHLF1
BORF2
BPLF1
BLLF1/gp350
BMRF2
BKRF4
BVRF2/BdRF1
BSLF1
BcLF1


Cluster II (N= 6)
Mean Norm. Expression
Day
Genes:
EBNA3C
BKRF1/EBNA1
EBNA3A
EBNA3B
BZLF2
BTRF1


Cluster III (N= 8)
Mean Norm. Expression
Day
Genes:
BNLF1/LMP1
LMP2B
BALF5
LMP2A
BALF4
BALF3
BXLF2
BARF0

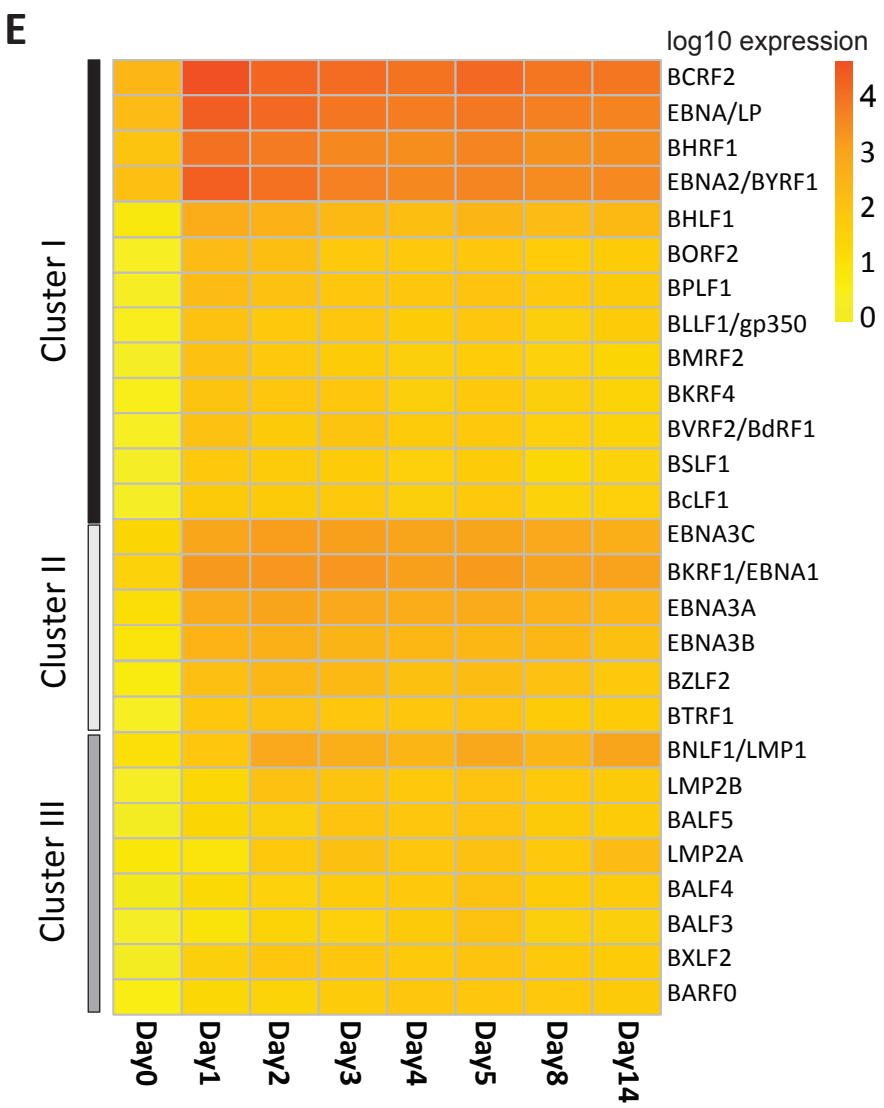

E
log10 expression
Cluster I
Cluster II
Cluster III
BCRF2
EBNA/LP
BHRF1
EBNA2/BYRF1
BHLF1
BORF2
BPLF1
BLLF1/gp350
BMRF2
BKRF4
BVRF2/BdRF1
BSLF1
BcLF1
EBNA3C
BKRF1/EBNA1
EBNA3A
EBNA3B
BZLF2
BTRF1
BNLF1/LMP1
LMP2B
BALF5
LMP2A
BALF4
BALF3
BXLF2
BARF0
Day0
Day1
Day2
Day3
Day4
Day5
Day8
Day14

Figure S10

Fig. S10. Viral genes analysis in the course of EBV infection.

A. The ratios between counts allocated to viral genes versus counts allocated to cellular genes were calculated at different time points p.i.. As expected, no viral genes were present in the uninfected, naïve B cells (Day 0). On day one, the relative abundance of viral genes peaked but decreased subsequently during the process of EBV infection and establishment of latency.

B. Principal Component Analysis (PCA) performed on viral genes. Each dot represents a biological replicate and different colors show different time points after EBV infection as indicated. The percentage of total variances explained by the first two principal components is indicated in parentheses.

C. Scree plot for the PCA of viral genes shown in panel B.

D. The dynamics of expression of viral genes in each of the clusters identified. The hierarchical clustering analysis (see Figure 3) determined three main viral cluster (I, II, and III), which display a dynamic viral gene expression during the early phase of EBV infection. The plots are presented as mean normalized expression levels (y-axis) versus days p.i. (x-axis). The number N of the genes in each cluster is indicated at the top of each plot and the insets contain the lists of the viral genes included in the depicted cluster.

E. Heat map of normalized log10-mean expression values of significantly DE viral genes. Colors from yellow to red indicate the range of calculated values from low expression level to high expression level, respectively. Three clusters I, II, and III are indicated on the left and refer to the three plots in panel D.

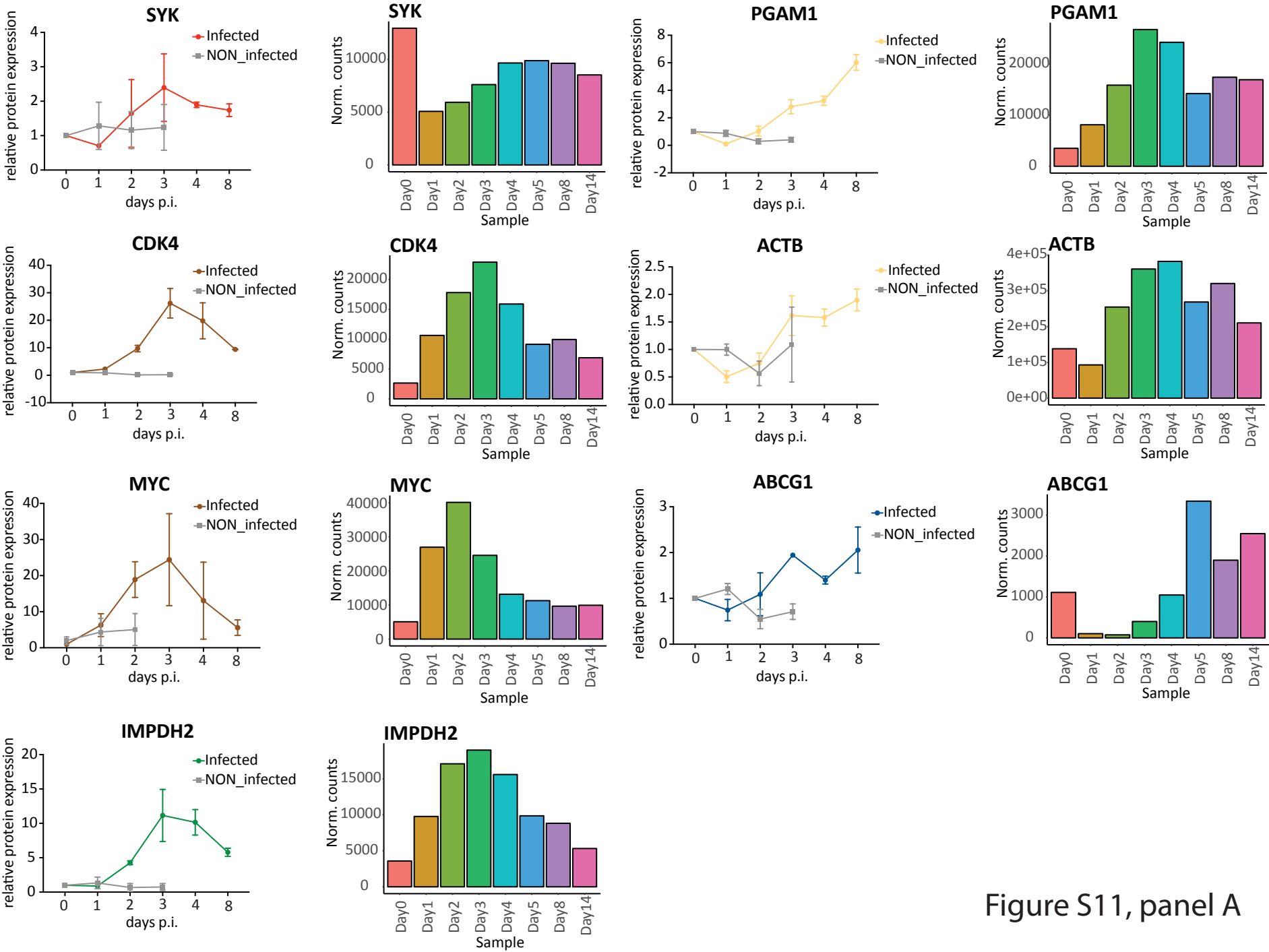


Figure S11, panel A

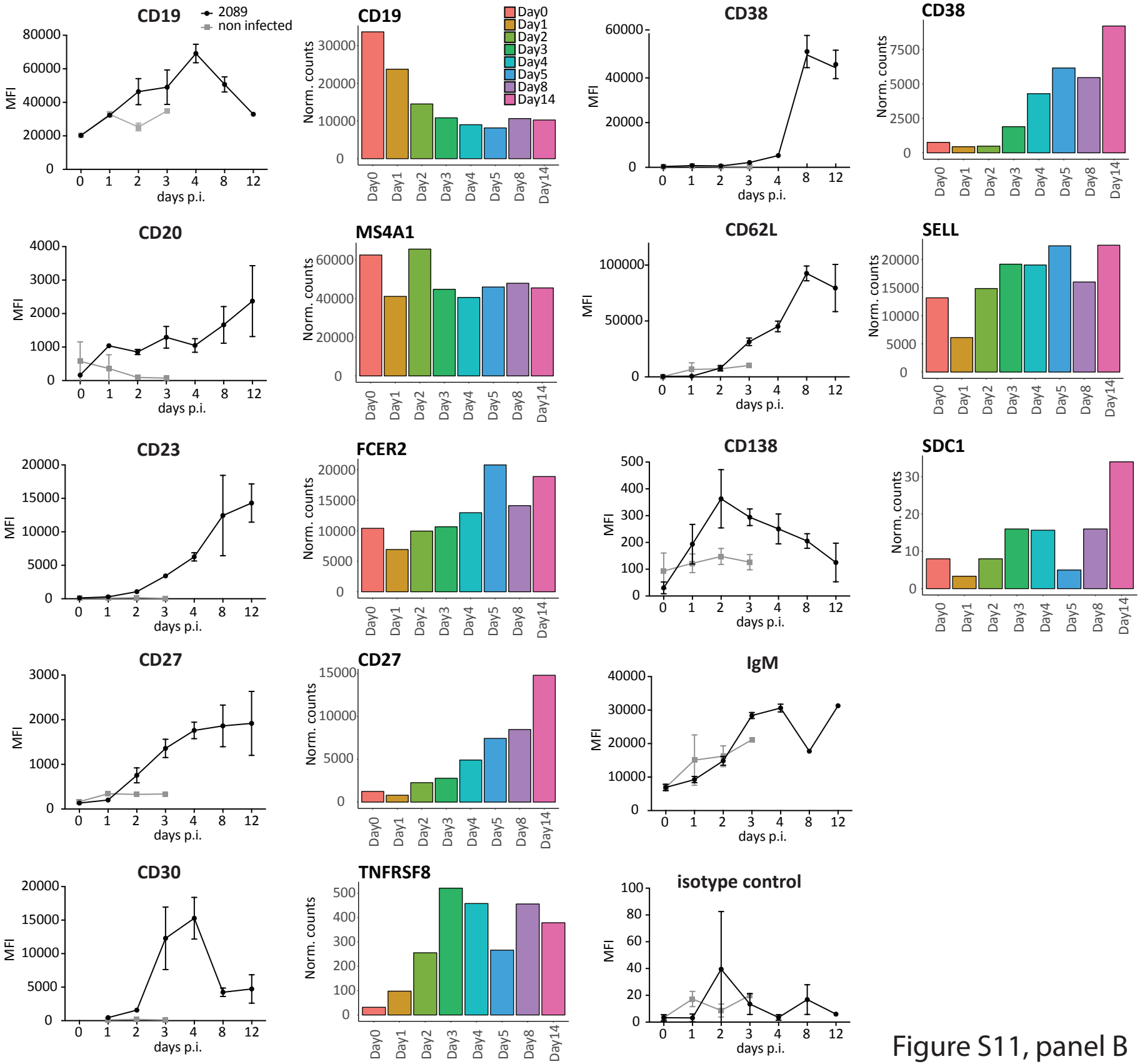


Figure S11, panel B

Fig. S11. Protein expression of selected genes in non-infected and EBV-infected B cells as a function of time.

A. Quantitative Western blot immunodetection. Shown are the relative expression levels of SYK, CDK4, MYC, IMPDH2, PGAM1, ACTB, and ABCG1 proteins. Sorted naïve B cells were analyzed in their non-infected states on plastic in complete cell culture medium for up to three days after preparation and cultivation (grey lines) or after EBV infection for up to eight days (colored lines). The protein levels were determined based on lysates obtained from defined numbers of intact viable cells at each time point and condition (i.e. infected and NON_infected). The differently colored lines of the infected samples indicate their cluster identities (see Figure 4). The calculation of the relative intensity of the bands was performed using the Compass Simple Western software provided by the Wes platform (Protein Simple, San Jose, USA) and the data were expressed relative to the intensities on day 0, which was set to 1. The graphs represent the mean and SD from at least two independent biological replicates. On the right panels, the mean normalized counts of gene expression of the analyzed proteins are displayed extracted from the RNA-seq analysis (see Materials and Methods).

B. Cell surface marker expression. Time course analysis of surface markers with directly labeled antibodies with specificities for CD19, CD20, CD23, CD27, CD30, CD38, CD62L, CD138, or IgM immunoglobulin are shown together with an example of an appropriate isotype control. Infected cells were analyzed at the indicated timepoints (day zero to four, eight, and 12 p.i.) after EBV infection (dark line) or sorted naive B cells were left uninfected and cultured for three days in full cell culture medium on plastic (grey line). The right panels provide the mean normalized counts of gene expression of the protein markers from the RNA-seq analysis (see Materials and Methods). Different B cell surface markers were measured and calculated as Mean Fluorescent Intensity (MFI) normalized to the negative control. Shown are mean values with corresponding SD. The analysis was performed on the cells from two independent donors.

References

1. Miller G, Shope T, Lisco H, Stitt D, Lipman M (1972) Epstein-Barr virus: transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes. *Proc Natl Acad Sci U S A* 69:383-387.
2. Pulvertaft JV (1964) Cytology of Burkitt's Tumour (African Lymphoma). *Lancet* 1:238-240.
3. Rowe M, Rooney CM, Rickinson AB, Lenoir GM, Rupani H, Moss DJ, Stein H, Epstein MA (1985) Distinctions between endemic and sporadic forms of Epstein-Barr virus-positive Burkitt's lymphoma. *Int J Cancer* 35:435-441.
4. Delecluse HJ, Hilsendegen T, Pich D, Zeidler R, Hammerschmidt W (1998) Propagation and recovery of intact, infectious Epstein-Barr virus from prokaryotic to human cells. *Proc Natl Acad Sci U S A* 95:8245-8250.
5. Hammerschmidt W, Sugden B (1988) Identification and characterization of oriLyt, a lytic origin of DNA replication of Epstein-Barr virus. *Cell* 55:427-433.
6. Neuhierl B, Feederle R, Hammerschmidt W, Delecluse HJ (2002) Glycoprotein gp110 of Epstein-Barr virus determines viral tropism and efficiency of infection. *Proc Natl Acad Sci U S A* 99:15036-15041.
7. Steinbrück L, Gustems M, Medele S, Schulz TF, Lutter D, Hammerschmidt W (2015) K1 and K15 of Kaposi's sarcoma-associated herpesvirus are partial functional homologues of latent membrane protein 2A of Epstein-Barr virus. *J Virol* 89:7248-7261.
8. Pich D, Mrozek-Gorska P, Sugimoto A, Akidil E, Bouvet M, Hamperl S, Ling P, Hammerschmidt W (2019) The first days in the life of naïve human B-lymphocytes infected with Epstein-Barr virus. *https://www.biorxiv.org/content/10.1101/666297v1*
9. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14:417-419.
10. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11:R106.
11. Kassambara A, Rème T, Jourdan M, Fest T, Hose D, Tarte K, Klein B (2015) GenomicScape: an easy-to-use web tool for gene expression data analysis. Application to investigate the molecular events in the differentiation of B cells into plasma cells. *PLoS Comput Biol* 11:e1004077.
12. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.
13. Dongen SV, Enright AJ (2012) Metric distances derived from cosine similarity and Pearson and Spearman correlations. *arXiv*
14. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48.
15. Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800.