iScience, Volume *27*

Supplemental information

Distinguishable topology of the task-evoked functional genome networks in HIV-1 reservoirs

Janusz Wiśniewski, Kamil Więcek, Haider Ali, Krzysztof Pyrc, Anna Kula-Păcurar, Marek Wagner, and Heng-Chang Chen

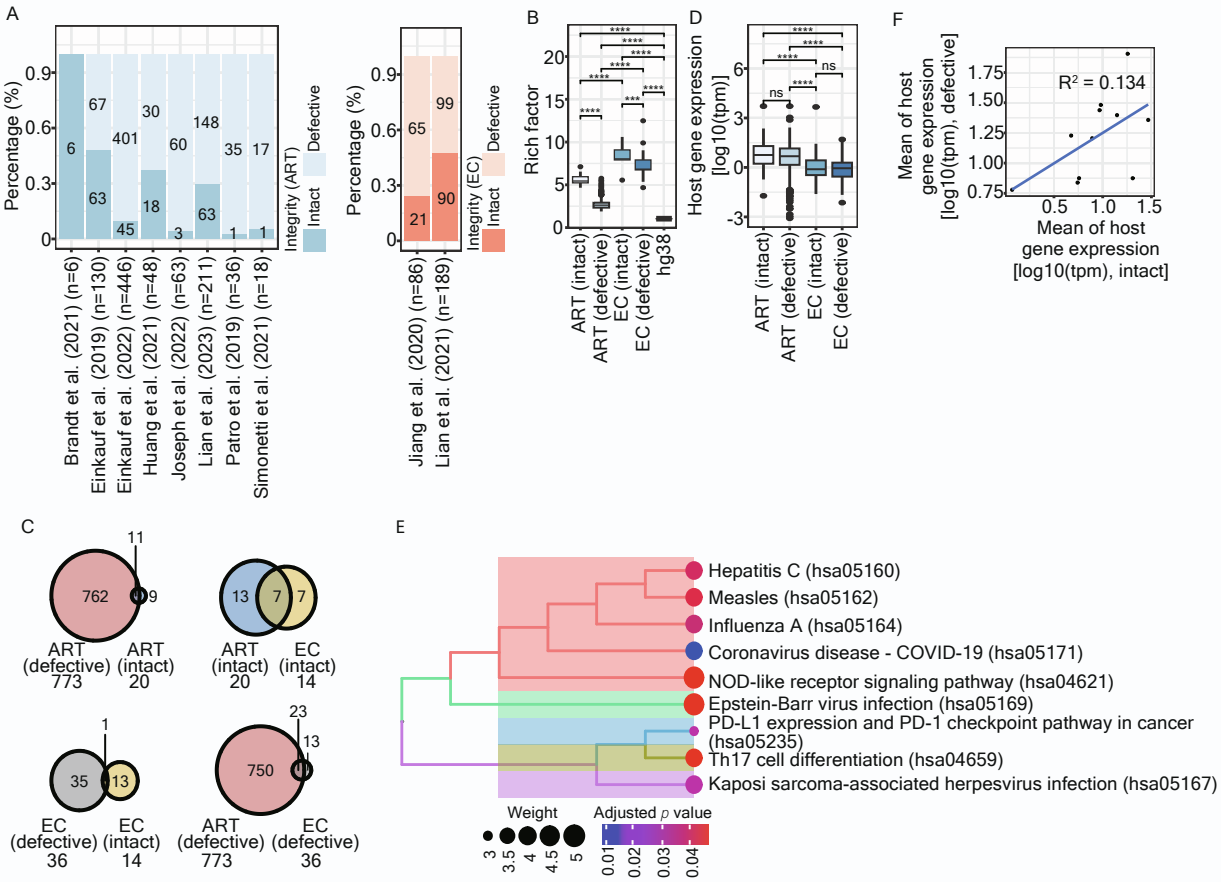

Figure S1. Characteristics of enriched signatures in ART-treated patients and elite controllers, related to Figure 1.

(**A**) A stacked bar plot representing the proportion of genes harboring intact versus defective proviruses in research articles. The left panel lists articles in which provirus-targeted genes were retrieved from ART-treated patients, while the right panel lists articles where provirus-targeted genes were retrieved from elite controllers. (**B**) A box plot displaying the enrichment of signatures represented by rich factors. Details on the commands used for rich factors calculation and the inclusion of a random control "hg38" are described in **STAR Methods**. Statistical significance was determined using the Wilcoxon test in R with default options. The mean and median values for different groups are provided: ART-intact: Mean, 5.655; Median, 5.444. ART-defective: Mean, 2.725; Median, 2.621. EC-intact: Mean, 8.365; Median, 8.030. EC-defective: Mean, 7.304; Median, 6.829. hg38 (control): Mean, 1.039; Median, 1.039. (**C**) Venn diagram representing the overlap of enriched signatures between different comparisons, with the total number of enriched signatures indicated under the name of respective groups. (**D**) A box plot, represented on a logarithmic scale, representing host gene expression (tpm) of the genes retrieved from enriched signatures. The mean and median values for different groups are provided: ART-intact: Mean, 1.87; Median, 0.754. ART-defective: Mean, 1.421; Median, 0.672. EC-intact: Mean, 2.016; Median, -0.108. EC-defective: Mean, 0.432; Median, -0.049 (values on a logarithmic scale). Statistical significance was determined using the Wilcoxon test in R with default options. (**E**) A clustering tree plot representing KEGG pathways enriched by 43 genes retrieved from signatures uniquely enriched in reservoirs harboring intact proviruses in ART-treated patients. (**F**) A scatter plot representing the mean transcriptional state of eleven common signatures enriched in both reservoirs harboring intact and defective proviruses in ART-treated patients. Each dot represents a unique enriched immunologic signature. Calculations from panels (**B**) and (**D**) were represented using box plots, in which the median (the vertical line in the box) of data distribution and the two whiskers were marked. Significance levels are denoted as follows: * *p* 0.05, ** *p* 0.01, *** *p* 0.001, **** *p* 0.0001.

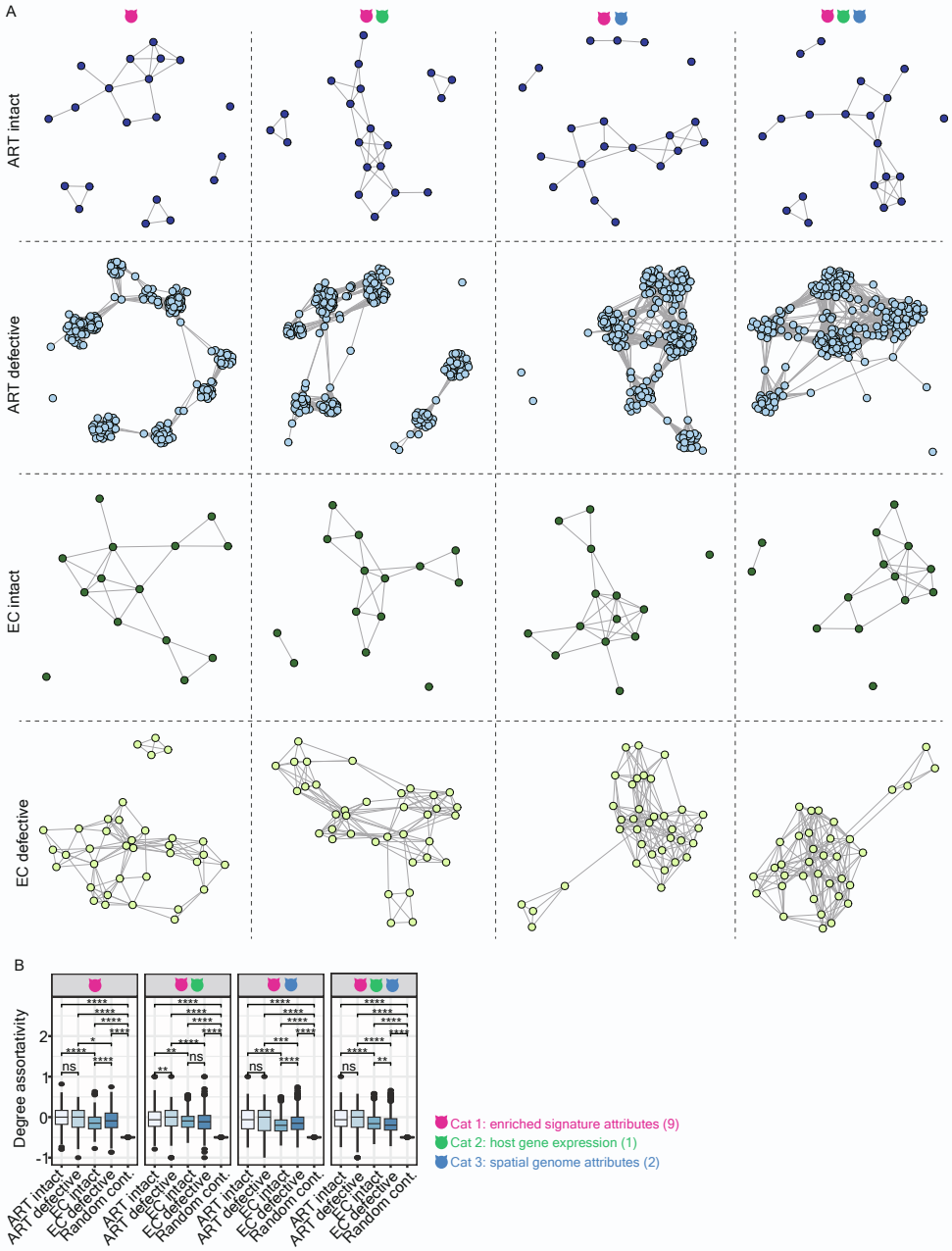

Figure S2. Individual network topologies, related to Figure 2.

(**A**) Graphs illustrating the topological properties of the network structured by enriched signatures (vertices) linked by edges, representing correlation coefficients using Cat 1, Cat 1 and 2, Cat 1 and 3, and all attributes (ordered from left to right). The graphs, from top to bottom, illustrate the networks associated with intact and defective proviruses in ART-treated patients and elite controllers in a sequential order. Enriched signatures (vertices in the graphs) in reservoirs harboring intact and defective proviruses in ART-treated patients and elite controllers are labeled in dark blue, light blue, dark green, and light green, respectively. Predictor variables among different category attributes are described in detail in **STAR Methods**. Vertices represent individual enriched signatures, while edges represent the correlation coefficient between two adjacent enriched signatures. (**B**) A box plot representing the degree assortativity coefficient of individual networks constructed based on different combinations of category attributes using the bootstrapping method detailed in **STAR Methods**. Statistical significance was determined using the Wilcoxon test in R with default options; calculations were represented using box plots, in which the median (the vertical line in the box) of data distribution and the two whiskers were marked. Significance levels are denoted as follows: * *p* 0.05, ** *p* 0.01, *** *p* 0.001, **** *p* 0.0001.

Figure S3. Topological interactions of two distinct networks, related to Figure 2.
(**A**) Bipartite graphs illustrating the interaction between two adjacent enriched signatures present in distinct networks using Cat 1, Cat 1 and 2, Cat 1 and 3, and all attributes (ordered from left to right). The graphs, from top to bottom, illustrate the topological interactions of the networks harboring intact (dark blue) versus defective (light blue) proviruses in ART-treated patients, the networks harboring intact (dark green) versus defective (light green) proviruses in elite controllers, the networks harboring intact proviruses in ART-treated patients (dark blue) versus elite controllers (dark green), and the networks harboring defective proviruses in ART-treated patients (light blue) versus elite controllers (light green). Predictor variables among different category attributes are described in detail in **STAR Methods**. Vertices represent individual enriched signatures, and edges represent the correlation coefficient between two adjacent enriched signatures. (**B**) A box plot representing the correlation coefficients between two adjacent and distinct enriched signatures present in different networks using Cat 1, Cat 1 and 2, Cat 1 and 3, and all attributes (ordered from left to right). Statistical significance was determined using the Wilcoxon test in R with default options. (**C**) A box plot representing the Euclidean distance between two adjacent and distinct enriched signatures present in different networks using Cat 1, Cat 1 and 2, Cat 1 and 3, and all attributes (ordered from left to right). Calculations from panels (**B**) and (**C**) were represented using box plots, in which the median (the vertical line in the box) of data distribution and the two whiskers were marked. Statistical significance was determined using the Wilcoxon test in R with default options. Significance levels are denoted as follows: * *p* 0.05, ** *p* 0.01, *** *p* 0.001, **** *p* 0.0001.

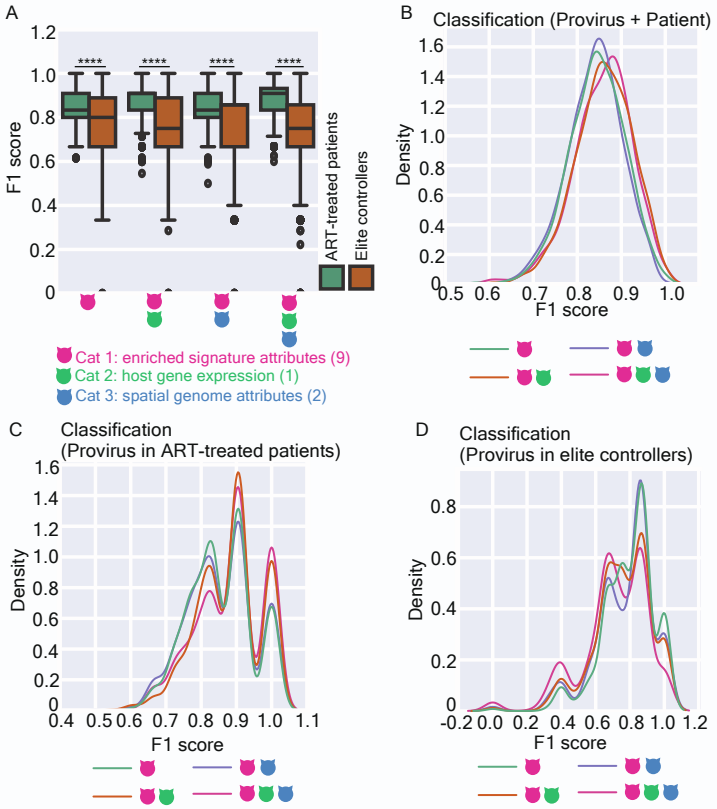

Figure S4. Random forest-based classifiers for predicting networks harboring intact versus defective proviruses, related to Figure 3.

(**A**) A dodged box plot representing the accuracy of classification for networks associated with intact versus defective proviruses using random forest algorithms based on Cat 1, Cat 1 and 2, Cat 1 and 3, and all attributes (ordered from left to right). Statistical significance was determined using the Wilcoxon test in R with default options. Calculations were represented using box plots, in which the median (the vertical line in the box) of data distribution and the two whiskers were marked. Significance levels are denoted as follows: * p 0.05, ** p 0.01, *** p 0.001, **** p 0.0001. (**B**, **C**, **D**) A density plot representing the prediction quality of individual models constructed using different combinations of category attributes to classify networks associated with intact versus defective proviruses in ART-treated patients and elite controllers (**B**), networks associated with intact versus defective proviruses in ART-treated patients (**C**), and networks associated with intact versus defective proviruses in elite controllers (**D**). Interpretation of panels (**C**) and (**D**) is provided in **STAR Methods**.

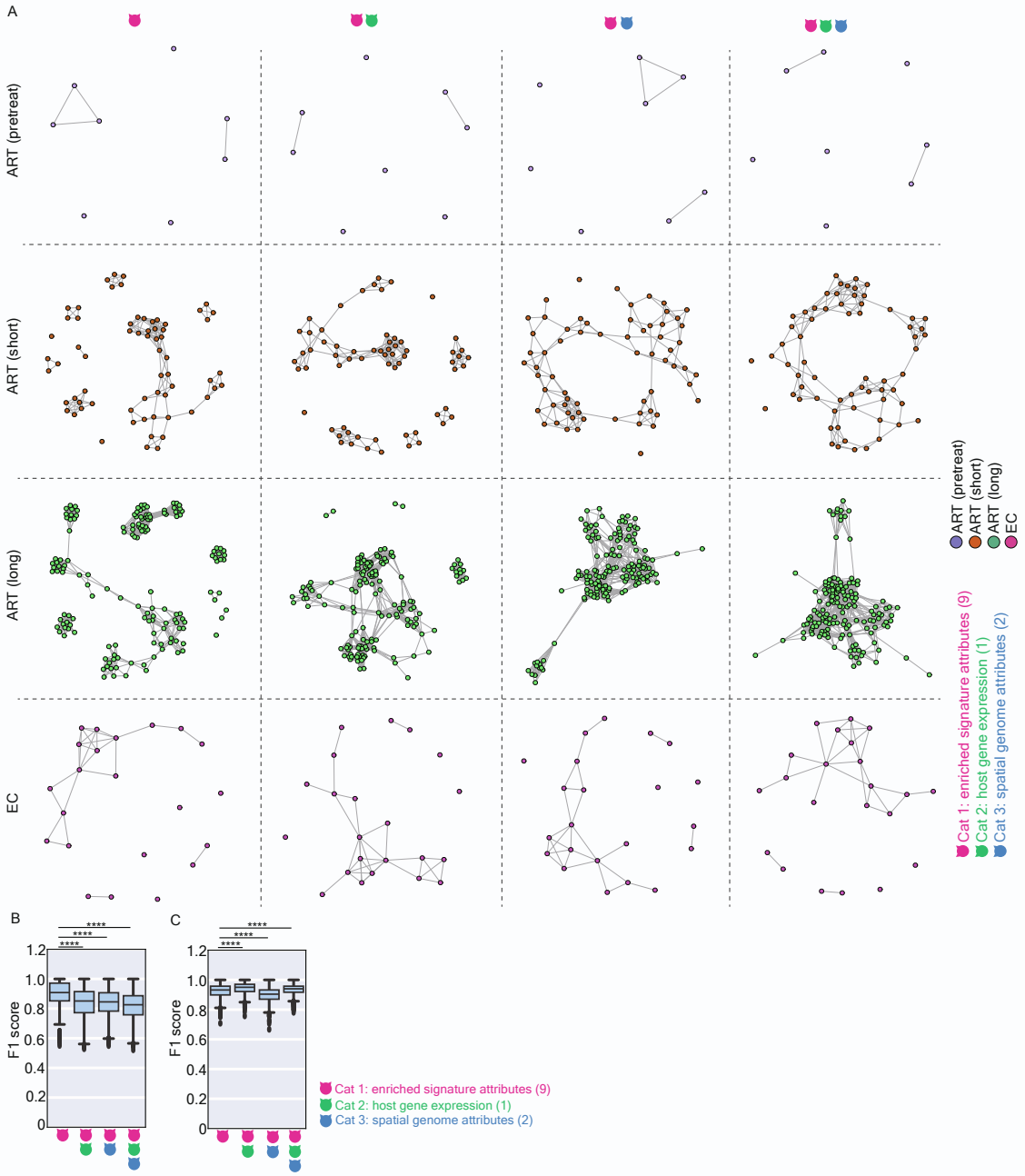

Figure S5. Longitudinal dynamics of the networks alongside HIV-1 infections associated with ART and in elite controllers, related to Figure 4.

(**A**) Graphs illustrate the topological properties of the network structured by enriched signatures (vertices) linked by edges, representing correlation coefficients using Cat 1, Cat 1 and 2, Cat 1 and 3, and all attributes (ordered from left to right). The graphs, from top to bottom, illustrate the networks structured by signatures enriched in pretreatment HIV-1-infected individuals, patients subjected to a short and long period of ART, and elite controllers. Enriched signatures (vertices in graphs) in pretreatment HIV-1-infected individuals, patients subjected to a short and a long period of ART, and elite controllers are labeled in violet, marron, green, and deep pink, respectively. Predictor variables among different category attributes are described in **STAR Methods**. Vertices represent individual enriched signatures, while edges represent the correlation coefficient between two adjacent enriched signatures. (**B**) A box plot representing the prediction power for classifying networks in pretreatment HIV-1-infected individuals, patients subjected to a short and a long period of ART. The F1 score was calculated based on 1,000 times of individual train-test splits in models. (**C**) A box plot representing the prediction power for classifying networks in HIV-1-infected individuals subjected to a short and a long period of ART, and elite controllers. The F1 score was calculated based on 1,000 times of individual train-test splits in models. Statistical significance was determined using the Wilcoxon test in R with default options; calculations from panels (**B**) and (**C**) were represented using box plots, in which the median (the vertical line in the box) of data distribution and the two whiskers were marked. Significance levels are denoted as follows: * p 0.05, ** p 0.01, *** p 0.001, **** p 0.0001.

Table S1. List of enriched immunologic signatures harboring intact proviruses in ART-treated patients, related to Figure 1.

Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
NAKAYA_PBMC_FLUARIX_FLUVIRIN_AGE_18_50YO_CORRELATED_WITH_HAI_28DY_RESPONSE_AT_3DY_NEGATIVE	16/164	364/21355	0,00000002	0,00009860	0,00009226	288/7109/54664/378938/51735/10782/3738/246175/3716/54856/51176/79066/9692/54665/7574/2521	16
NAKAYA_PBMC_FLUARIX_FLUVIRIN_AGE_18_50YO_CORRELATED_WITH_HAI_28DY_RESPONSE_AT_7DY_NEGATIVE	17/164	475/21355	0,00000016	0,00034464	0,00032248	288/6777/7109/378938/6774/10782/1997/84196/23042/11273/246175/9882/54856/51176/196528/23049/2521	17
GSE5099_DAY3_VS_DAY7_MCSF_TREATED_MACROPHAGE_DN	10/164	183/21355	0,00000157	0,00225735	0,00211220	1416/57690/378938/8418/23397/57459/246175/8832/51176/3183	10
GSE40225_WT_VS_RIP_B7X_DIABETIC_MOUSE_PANCREATIC_CD8_TCELL_UP	10/164	199/21355	0,00000333	0,00300956	0,00281604	64853/51466/63892/55770/9873/79230/54476/4306/28968/23049	10
GSE2770_IL4_ACT_VS_ACT_CD4_TCELL_2H_DN	10/164	200/21355	0,00000348	0,00300956	0,00281604	116984/51735/23214/9135/64375/84166/10425/8178/4939/10613	10
GSE21927_BALBC_VS_C57BL6_MONOCYTE_SPLEEN_UP	9/164	189/21355	0,00001603	0,01069483	0,01000715	64853/10521/83852/1147/104/11273/51176/54811/28968	9
GSE11057_CD4_CENT_MEM_VS_PBMC_UP	9/164	197/21355	0,00002228	0,01069483	0,01000715	89894/6777/55870/23048/2035/55690/8832/9882/90987	9

GSE11057_NAIVE_CD4_VS_PBMC_CD4_TCELL_UP	9/164	197/21355	0,00002228	0,01069483	0,01000715	6777/51466/9252/23214/246175/54856/79066/90987/54665	9
GSE3982_MAST_CELL_VS_BCELL_DN	9/164	197/21355	0,00002228	0,01069483	0,01000715	51466/63892/953/8879/3738/10664/3716/9692/4939	9
GSE2935_UV_INACTIVATED_VS_LIVE_SENDAI_VIRUS_INF_MACROPHAGE_DN	8/164	179/21355	0,00007487	0,03235208	0,03027183	953/493/10111/54476/11273/246175/3683/10613	8
GSE32255_UNSTIM_VS_4H_LPS_STIM_DC_UP	8/164	186/21355	0,00009786	0,03489092	0,03264742	953/23048/3738/140609/54476/8832/9882/4306	8
GSE16450_CTRL_VS_IFNA_12H_STIM_IMMATURE_NEURON_CELL_LINE_UP	8/164	197/21355	0,00014558	0,03489092	0,03264742	63892/9873/64754/23315/283450/2530/55809/2521	8
GSE25087_TREG_VS_TCONV_FETUS_UP	8/164	198/21355	0,00015074	0,03489092	0,03264742	953/23048/81671/672/196/29110/64375/3683	8
GSE11057_CD4_EFF_MEM_VS_PBMC_UP	8/164	199/21355	0,00015604	0,03489092	0,03264742	288/89894/6777/51466/9252/3835/2035/54856	8
GSE17974_1H_VS_72H_UNTREATED_IN_VITRO_CD4_TCELL_DN	8/164	199/21355	0,00015604	0,03489092	0,03264742	23048/90861/26263/10664/2108/2175/9978/10694	8
GSE21927_C26GM_VS_4T1_TUMOR_MONOCYTE_BALBC_DN	8/164	199/21355	0,00015604	0,03489092	0,03264742	51466/7328/23315/1997/2035/6872/283450/81856	8
GSE10240_IL22_VS_IL17_STIM_PRIMARY_BRONCHIAL_EPITHELIAL_CELLS_UP	8/164	200/21355	0,00016149	0,03489092	0,03264742	51735/1147/26263/84196/9978/29110/7700/54665	8
GSE1460_DP_VS_CD4_THYMOCYTE_DN	8/164	200/21355	0,00016149	0,03489092	0,03264742	7699/51466/8879/104/9882/196/92822/4939	8

GSE22601_IMMATURE_CD4_SINGLE_POSITIVE_VS_CD8_SINGLE_POSITIVE_THYMOCYTE_DN	8/164	200/21355	0,00016149	0,03489092	0,03264742	116984/8897/6774/51735/3091/54739/196/64375	8
GSE5589_WT_VS_IL6_KO_LPS_AND_IL10_STIM_MACROPHAGE_45MIN_UP	8/164	200/21355	0,00016149	0,03489092	0,03264742	7109/9738/3835/23397/90861/672/57153/10613	8

Table S3. List of enriched immunologic signatures harboring intact proviruses in elite controllers, related to Figure 1.

Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
GSE11057_CD4_EFF_MEM_VS_PBMC_UP	8/81	199/21355	0,00000089	0,00289053	0,00251375	114804/9252/51466/3835/6777/288/89894/54856	8
GSE40225_WT_VS_RIP_B7X_DIABETIC_MOUSE_PANCREATIC_CD8_TCELL_UP	7/81	199/21355	0,00001066	0,01004741	0,00873772	4926/55770/63892/9873/51466/64853/23049	7
GSE31082_DN_VS_CD4_SP_THYMOCYTE_DN	7/81	200/21355	0,00001101	0,01004741	0,00873772	8675/8897/55209/23214/83852/54664/116984	7
NAKAYA_PBMC_FLUARIX_FLUVIRIN_AGE_18_50YO_CORRELATED_WITH_HAI_28DY_RESPONSE_AT_7DY_NEGATIVE	10/81	475/21355	0,00001234	0,01004741	0,00873772	4926/10782/6774/378938/1997/6777/7109/288/23049/54856	10
GSE25146_UNSTIM_VS_HELIOBACTER_PYLORI_LPS_STIM_AGS_CELL_DN	6/81	166/21355	0,00004040	0,02630925	0,02287983	27327/10521/953/55870/8418/54739	6
GSE26488_CTRL_VS_PEPTIDE_INJECTION_OT2_THYMOCYTE_DN	6/81	183/21355	0,00006950	0,02636130	0,02292510	64744/9873/8418/51735/57459/64853	6
GSE5099_DAY3_VS_DAY7_MCSF_TREATED_MACROPHAGE_DN	6/81	183/21355	0,00006950	0,02636130	0,02292510	378938/57690/8418/23397/57459/1416	6
GSE11057_NAIVE_CD4_VS_PBMC_CD4_TCELL_UP	6/81	197/21355	0,00010434	0,02636130	0,02292510	55628/9252/51466/23214/6777/54856	6

GSE16450_CTRL_VS_IFNA_12H_STIM_IMMATURE_NEURON_CELL_LINE_UP	6/81	197/21355	0,00010434	0,02636130	0,02292510	27327/55628/63892/9873/23315/64754	6
GSE20366_TREG_VS_NAIVE_CD4_TCELL_HOMEOSTATIC_CONVERSION_DN	6/81	199/21355	0,00011028	0,02636130	0,02292510	57514/64744/9873/57690/23048/7109	6
GSE22601_IMMATURE_CD4_SINGLE_POSITIVE_VS_CD8_SINGLE_POSITIVE_THYMOCYTE_DN	6/81	200/21355	0,00011335	0,02636130	0,02292510	8897/6774/54739/3091/51735/116984	6
GSE30971_WBP7_HET_VS_KO_MACROPHAGE_UP	6/81	200/21355	0,00011335	0,02636130	0,02292510	114804/5034/57514/8897/6774/3091	6
GSE5142_CTRL_VS_HTERT_TRANSDUCED_CD8_TCELL_EARLY_PASSAGE_CLONE_UP	6/81	200/21355	0,00011335	0,02636130	0,02292510	4926/114804/10521/81671/3091/6777	6
GSE6674_ANTI_IGM_VS_CPG_STIM_BCELL_UP	6/81	200/21355	0,00011335	0,02636130	0,02292510	10782/283237/9252/51466/7328/54664	6