Supplementary Information

Table of content

Supplementary Table 1. Summary of RNAseq data and mapping statistics

NGS protocol	Time-point (hpi)	Infection	Replicate sample id	rRNA	%	HIV	%	Human	%
mRNAseq	12	Mock	Mock12hr1	304,704	1.3	26	0.0	23,558,521	98.7
			Mock12hr2	305,530	1.2	40	0.0	25,447,126	98.8
			Mock12hr3	242,120	1.3	471	0.0	19,040,603	98.7
		HIV	HIV12hr7	322,300	1.4	3,989,152	17.2	18,916,407	81.4
			HIV12hr8	369,444	1.3	5,313,320	18.8	22,608,978	79.9
			HIV12hr9	215,699	1.4	2,564,062	17.1	12,227,602	81.5
	24	Mock	Mock24hr1	386,502	1.3	24	0.0	28,686,704	98.7
			Mock24hr2	204,187	1.3	428	0.0	15,958,896	98.7
			Mock24hr3	241,182	1.2	41	0.0	19,075,948	98.8
		HIV	HIV24hr7	178,442	1.0	6,276,673	33.6	12,253,141	65.5
			HIV24hr8	224,052	1.1	7,924,233	39.6	11,848,320	59.3
			HIV24hr9	200,501	1.0	7,980,792	39.2	12,182,433	59.8
TotalRNAseq	12	Mock	12_1	5,427,297	37.8	4	0.0	8,925,352	62.2
			12_2	5,342,090	47.1	6	0.0	6,007,577	52.9
			12_3	7,147,336	38.4	1	0.0	11,456,285	61.6
		UV	12_4	7,116,100	43.0	93	0.0	9,429,803	57.0
			12_5	5,248,656	37.7	83	0.0	8,659,522	62.3
			12_6	5,160,715	32.7	76	0.0	10,632,308	67.3
		HIV	12_7	6,798,271	43.4	397,910	2.5	8,468,576	54.1
			12_8	7,455,689	47.5	362,234	2.3	7,865,152	50.2
			12_9	8,074,012	41.6	575,326	3.0	10,747,857	55.4
	24	Mock	24_1	6,926,666	51.8	0	0.0	6,436,364	48.2
			24_3	8,670,783	53.1	17	0.0	7,657,739	46.9
		UV	24_4	6,307,389	49.3	34	0.0	6,484,650	50.7
			24_6	7,056,634	43.7	674	0.0	9,082,503	56.3
		HIV	24_7	6,496,832	35.5	2,279,160	12.5	9,530,047	52.1
			24_8	6,379,398	40.3	1,610,363	10.2	7,835,967	49.5
			24_9	6,677,142	41.2	2,032,685	12.5	7,514,187	46.3

mRNAseq: 75nt single end reads. Total RNAseq: 2x90nt paired-end reads. Mock: cells treated with conditioned medium, as control. HIV: cells infected with intact HIV-1 viruses. UV: cells treated with UV-inactivated HIV-1 virions. rRNA: reads mapped to a collection of human ribosomal RNA sequences. HIV: reads mapped to the reference HIV-1 genome. Human: reads mapped to human reference genome.

Supplementary Table 4. Top 20 drugs from connectivity map search.

Rank	Compound name	Mean of connectivity scores	Number of instances	Enrichment score	Enrichment p-value
1	GW-8510	0.813	4	0.994	0
2	lycorine	-0.699 (-0.629, -0.639, -0.672, -0.694, -0.86)	5	-0.973	0
3	trichostatin A	0.345	182	0.428	0
4	camptothecin	0.82	3	0.994	0.00002
5	alsterpaullone	0.8	3	0.993	0.00002
6	irinotecan	0.792	3	0.992	0.00002
7	mitoxantrone	0.725	3	0.977	0.00004
8	azacitidine	0.752	3	0.974	0.00004
9	daunorubicin	0.698	4	0.93	0.00004
10	resveratrol	0.546	9	0.716	0.00004
11	ciclosporin	0.318	6	0.759	0.00048
12	vorinostat	0.416	12	0.551	0.00048
13	wortmannin	0.239	18	0.452	0.00062
14	methotrexate	0.483	8	0.655	0.00076
15	carbimazole	-0.537	3	-0.909	0.00138
16	SR-95531	-0.416	4	-0.83	0.00155
17	Prestwick-1080	-0.263	4	-0.825	0.00177
18	monensin	0.278	6	0.685	0.00246
19	diazoxide	-0.334	5	-0.731	0.0029
20	theobromine	-0.385	4	-0.794	0.00356

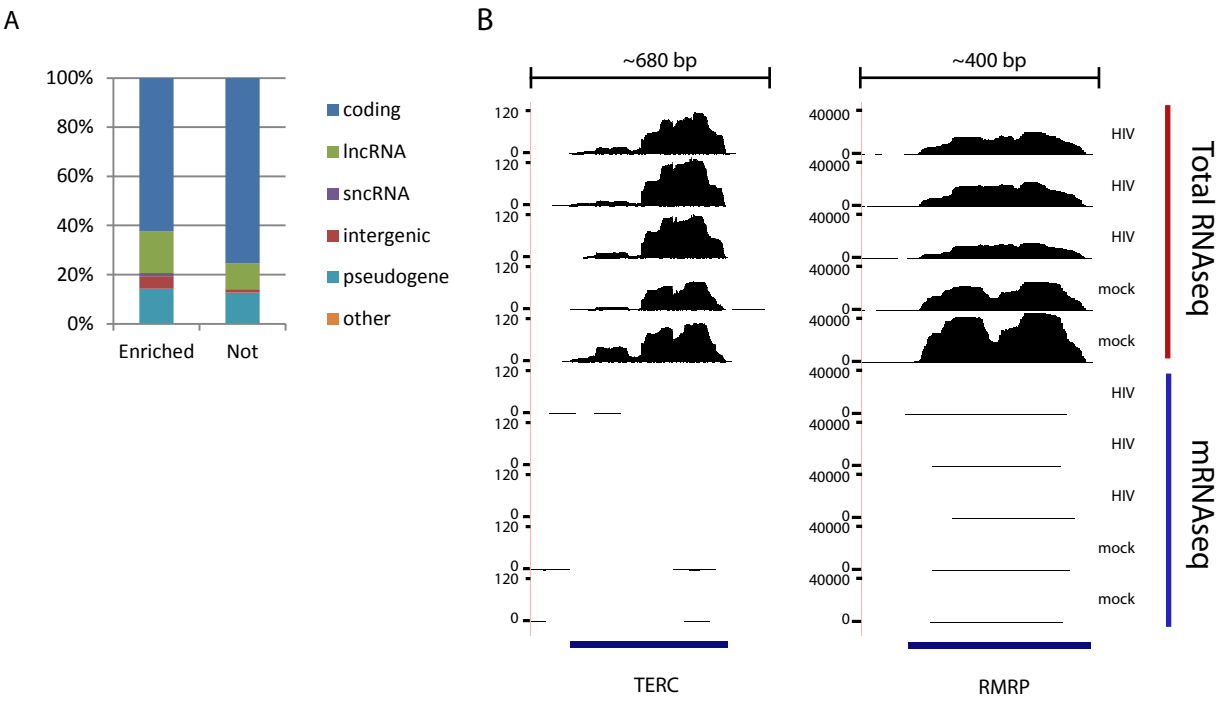
Notes: rows in blue are compounds with negative connectivity scores.Inside parenthesis are the individual connectivity scores for lycorine.

Supplementary Table 5. The enrichment of TF targets in genes with reversed expression changes.

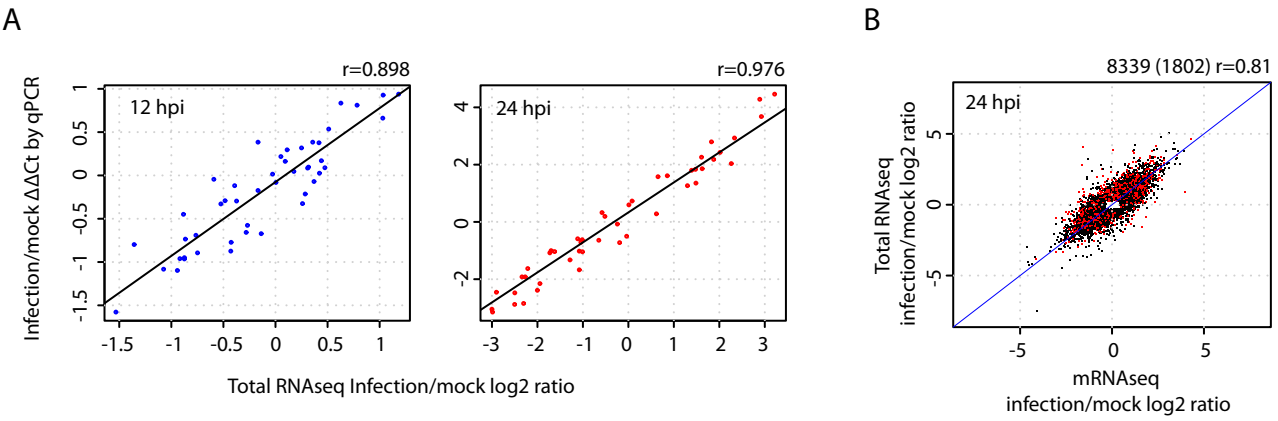
Transcription Factor (TF)	Enrichment p-value	TF DE at 12 hpi	TF DE at 12 hpi and in subgroups a and b
MYC	0.000282801	TRUE	TRUE
ETS1	0.000328471	TRUE	TRUE
IRF1	0.010901002	TRUE	TRUE
ELK4	0.029802343	TRUE	TRUE
GTF2F1	0.003459991	TRUE	FALSE
TAF1	0.003846276	TRUE	FALSE
ELF1	0.013188921	TRUE	FALSE
REST	0.130985877	TRUE	FALSE
BCLAF1	0.154816551	TRUE	FALSE
IRF3	0.325424325	TRUE	FALSE
ZBTB33	0.420432008	TRUE	FALSE
MAX	0.000101078	FALSE	FALSE
ZNF143	0.000737502	FALSE	FALSE
E2F1	0.001299821	FALSE	FALSE
HEY1	0.001971921	FALSE	FALSE
TBP	0.006611031	FALSE	FALSE
SP1	0.016741588	FALSE	FALSE
NRF1	0.045376381	FALSE	FALSE
NR3C1	0.068690462	FALSE	FALSE
E2F4	0.076843823	FALSE	FALSE
JUND	0.079408873	FALSE	FALSE
GTF2B	0.079633892	FALSE	FALSE
SRF	0.085635293	FALSE	FALSE
TAF7	0.089958648	FALSE	FALSE
SETDB1	0.109188848	FALSE	FALSE
SP2	0.132601963	FALSE	FALSE
EP300	0.191966414	FALSE	FALSE
STAT2	0.20196601	FALSE	FALSE
STAT1	0.243178518	FALSE	FALSE
NFYA	0.418352547	FALSE	FALSE
TCF4	0.523971842	FALSE	FALSE
STAT3	0.619749701	FALSE	FALSE

Supplementary Table 6. The list of microarray datasets compiled from GEO.

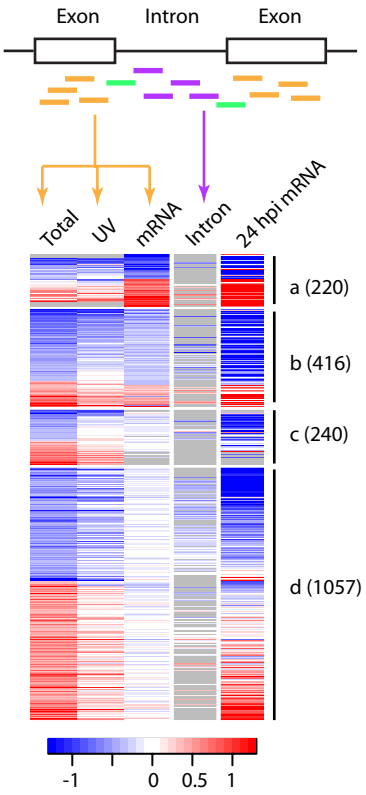
Description	GSE accession	Publication
Treated Jurkat cells	GSE46909	PMID: 23824090
Treated Jurkat cells	GSE10737, GSE10232	PMID: 18773076
Treated MM6 cells	GSE10739, GSE10738, GSE 10233, GSE10234	PMID: 18773076
Regulatory T cell time couse	GSE11292	PMID: 23169000
T effector cell time course	GSE11292	PMID: 23169000
T effector cell expressing GFP	GSE11292	PMID: 23169000
T effector cell over-expressing GARP	GSE11292	PMID: 23169000
CD4+ T cells infected with HIV-1 viruses that lack different viral proteins	GSE12963	PMID: 19050264
CD4+ T cells from HIV positive or negative individuals	GSE9927	PMID: 18077723
Lymph nodes from HIV positive or negative individuals	GSE16363	PMID: 21525393
Pathogen-specific CD4+ T cell sub-populations from the same PBMC	GSE42853	PMID: 23258923
CD4+ T cells from both HIV resistant and HIV low-risk negative individuals	GSE14278	PMID: 20887221
Th1 and Th1Th17 CD4+ T-cells	GSE50175	PMID: 24359430



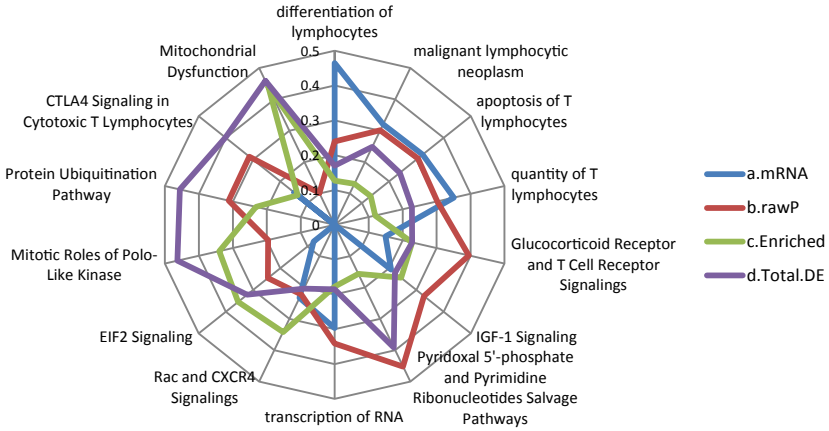
Supplementary Figure 1. Summary of Total RNAseq enriched genes. A. Percentages of gene types in the list of Total RNAseq enriched genes (Enriched) vs. the rest of detected genes (Not). Table 1 shows the classification of gene biotypes. B. Raw read coverage across TERC and RMRP genes, known to transcribe polyA- transcripts, in the samples collected at 24 hpi.



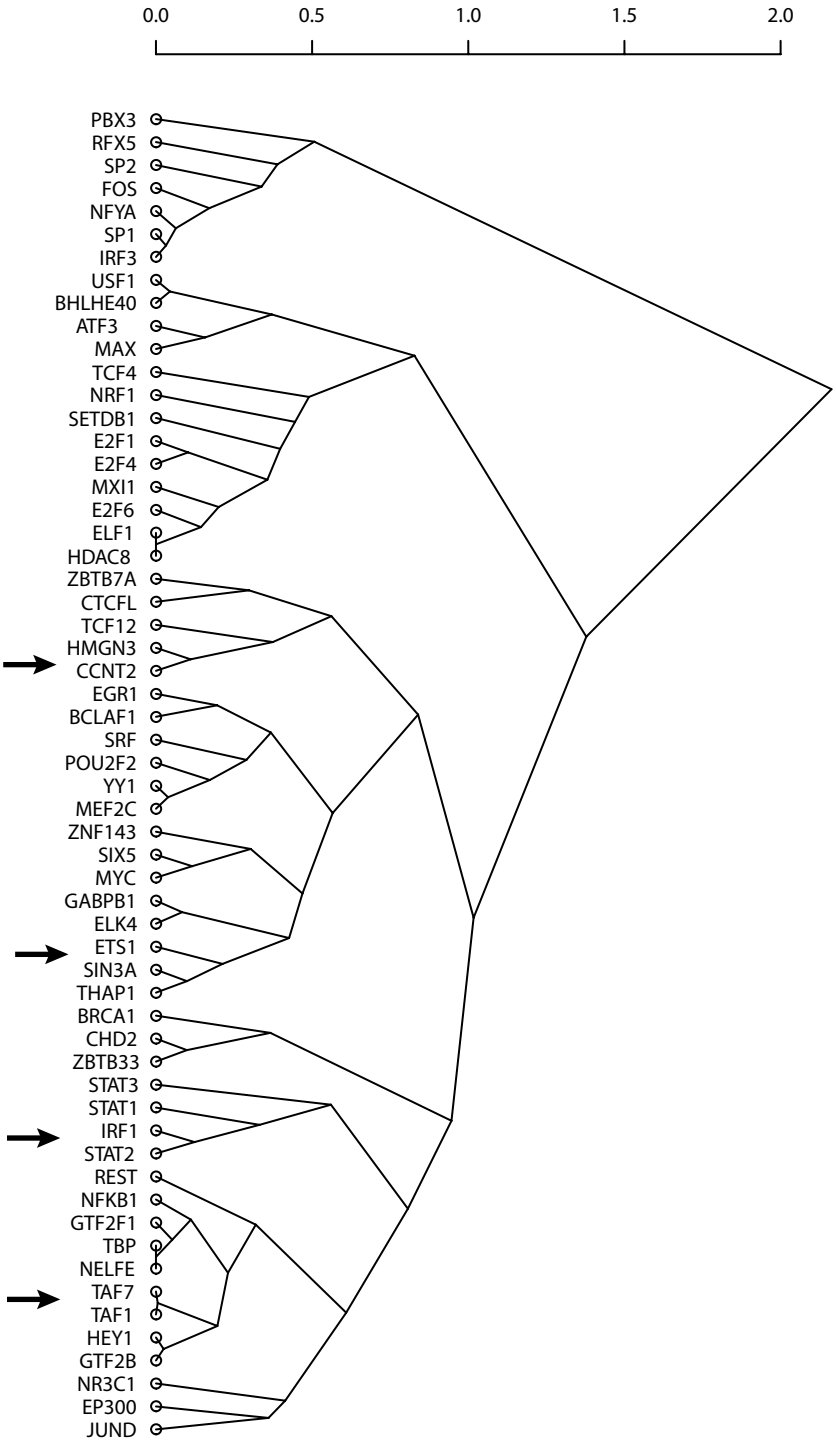
Supplementary Figure 2. Comparison of Total RNAseq measurements with qPCR and mRNAseq measurements. A. Scatter plot showing the agreements between the measurements by qPCR vs. Total RNAseq on the same set of 46 annotated protein-coding genes at both 12 and 24 hpi. The Pearson correlation coefficient between qPCR and Total RNAseq is shown on the top right corner of each scatter plot. B. Scatterplot of host expression changes relative to time matched mocks after HIV-1 infection measured by mRNAseq vs. that by Total RNAseq at 24 hpi. Only genes which were identified as differentially expressed by either RNAseq and detected by both RNAseq were included. Those genes which were enriched by Total RNAseq (described in the main text) are in red. The numbers on the top right corner are in the order of: the total number of genes plotted, the subset of genes which were considered as enriched by Total RNAseq (in red), and the Pearson correlation of log2 infection/ mock ratios of all genes.



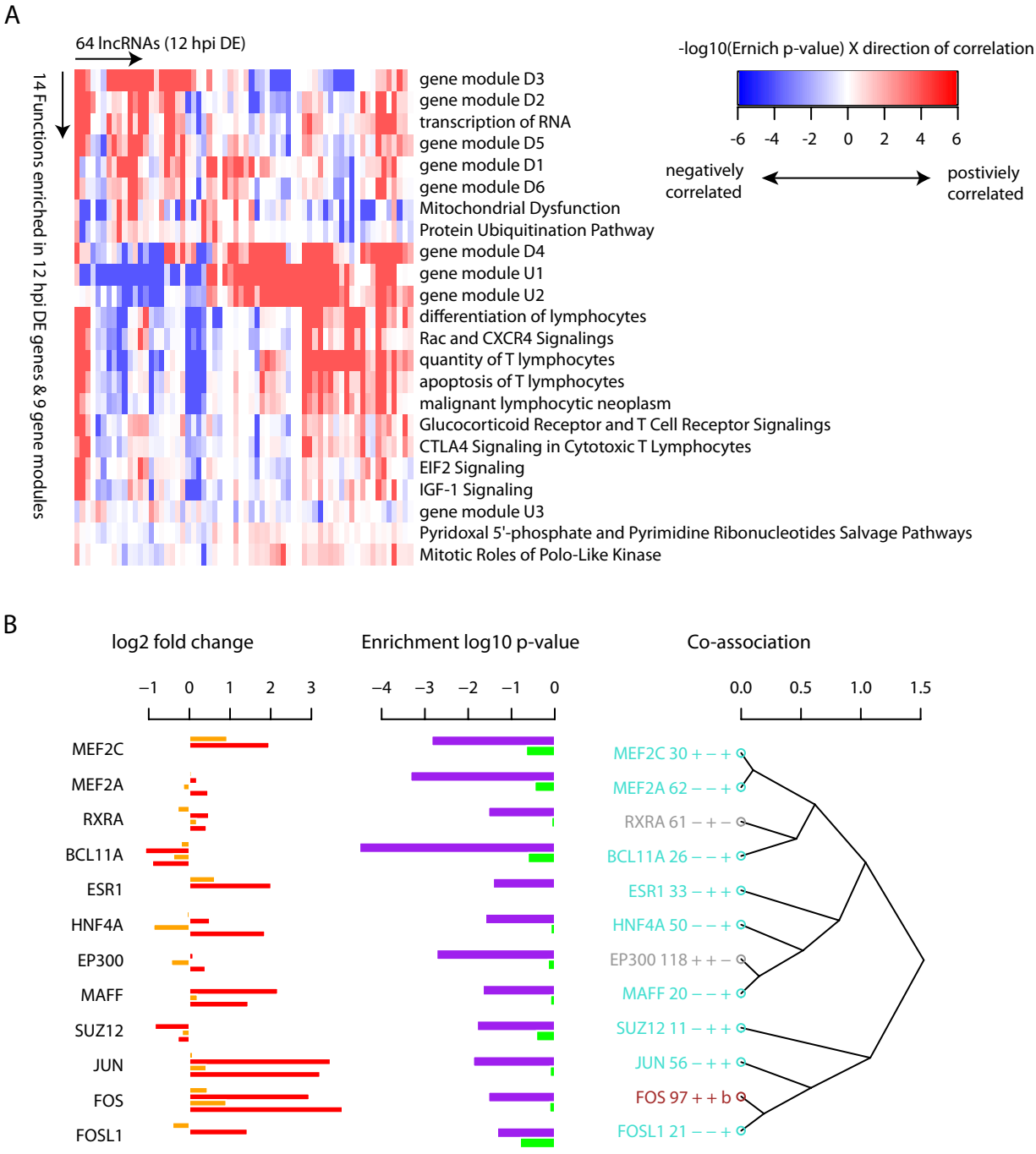
Supplementary Figure 3. Heatmap of the full list of 1,933 DE genes at 12 hpi, similarly shown in Figure 2A.



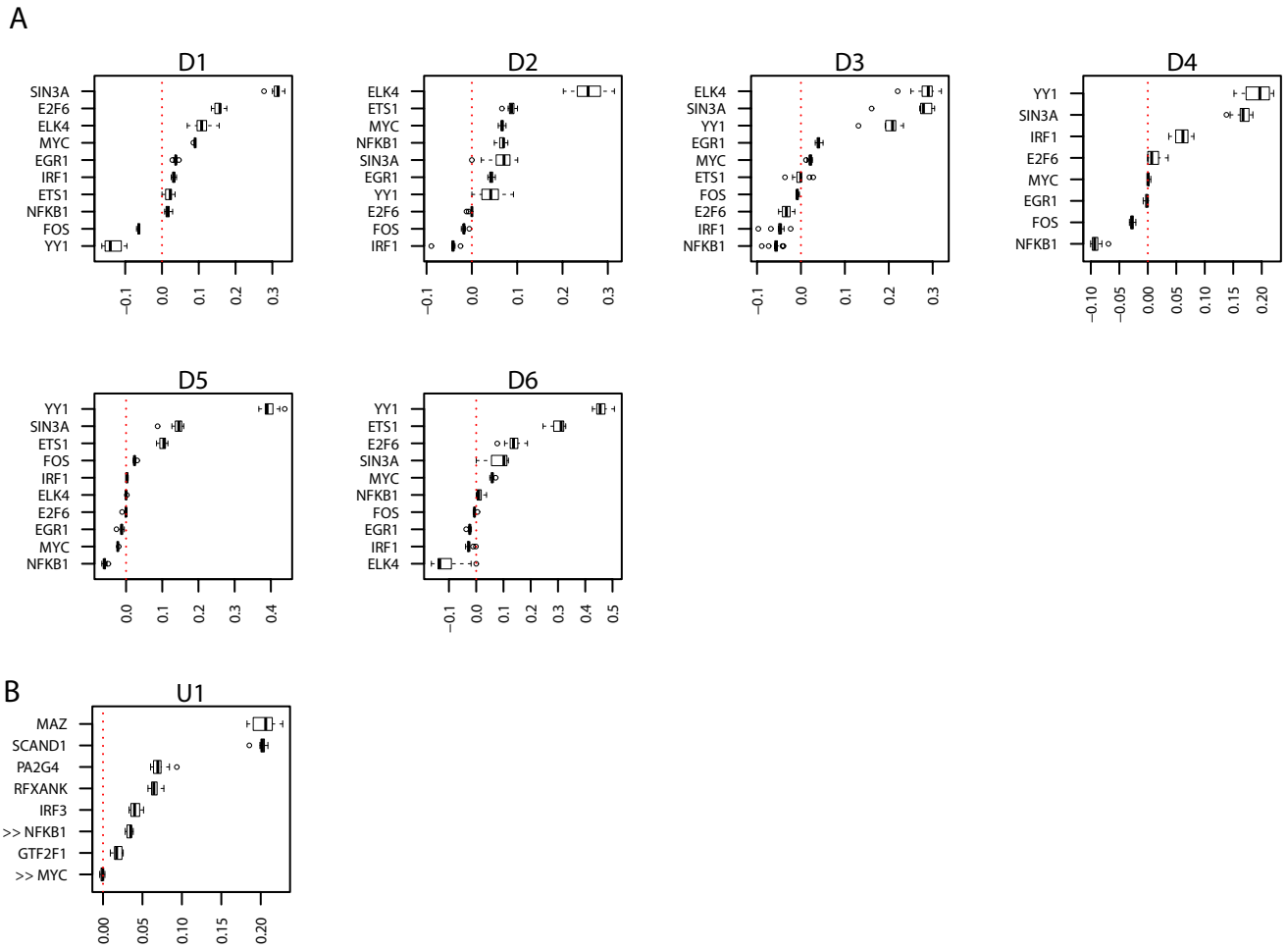
Supplementary Figure 4. Radio plot showing the relative percentages of genes from the four subgroups in Figure 2, for each of 14 enriched functions. Enriched functions were identified in the genes in using IPA. For each of 14 enriched functions, we counted the number of genes from each subgroup. These gene counts were scaled by the total number of genes in the corresponding subgroup, and denoted as normalized gene count. The relative contributions of individual subgroups to an enriched function was calculated as the ratio between its normalized gene count and the sum of normalized gene counts of all four subgroups.



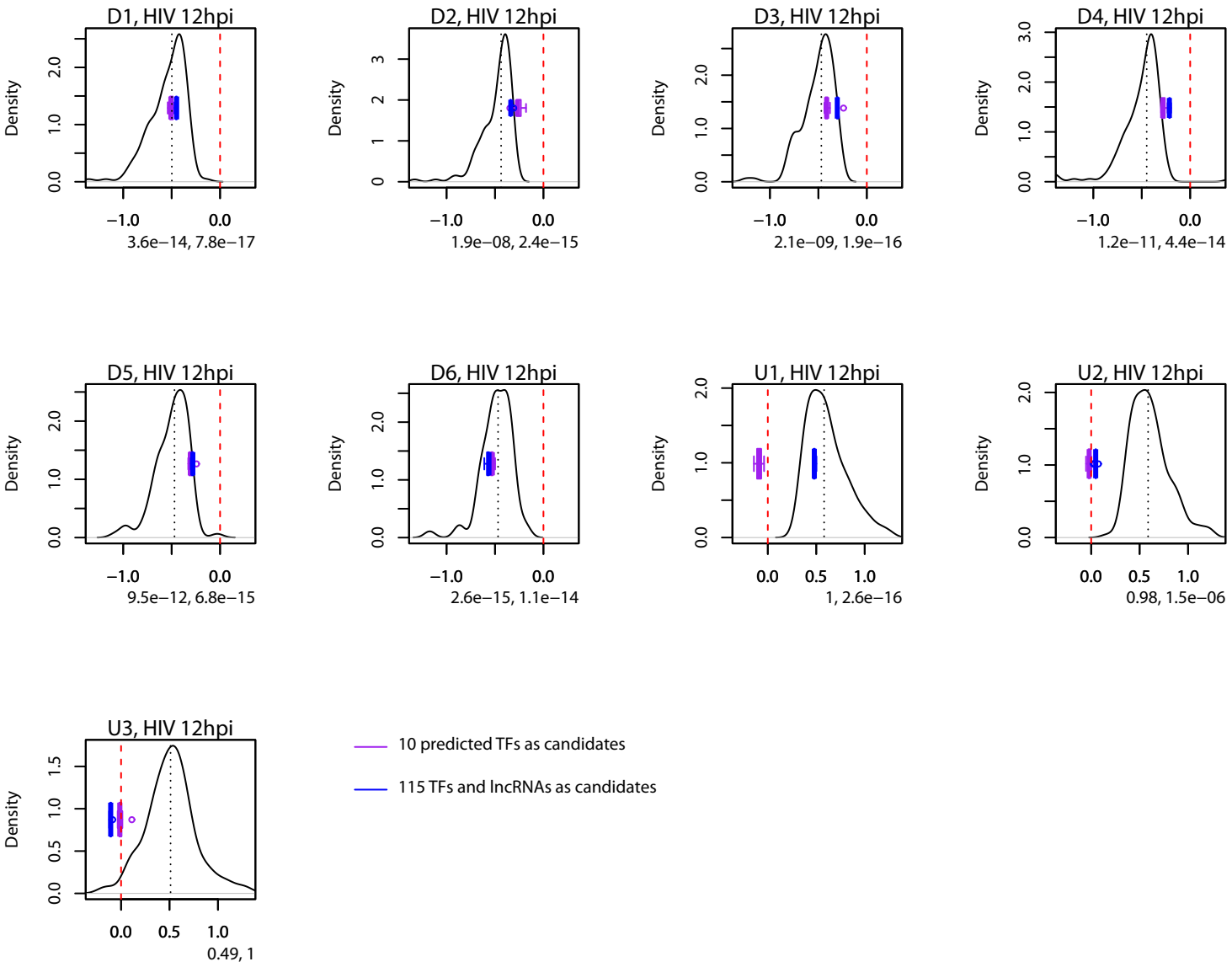
Supplementary Figure 5. Hierarchical clustering of 58 enriched TFs, based on their binding site co-occurrences in the promoters of similar sets of DE genes at 12 hpi during HIV-1 infection. The arrows mark examples of expected pairings like IRF1-STAT1-STAT2-STAT3 and TAF7-TAF1, or previously observed ones like CCNT2-HMGN3, and novel ones like ETS1-SIN3A.



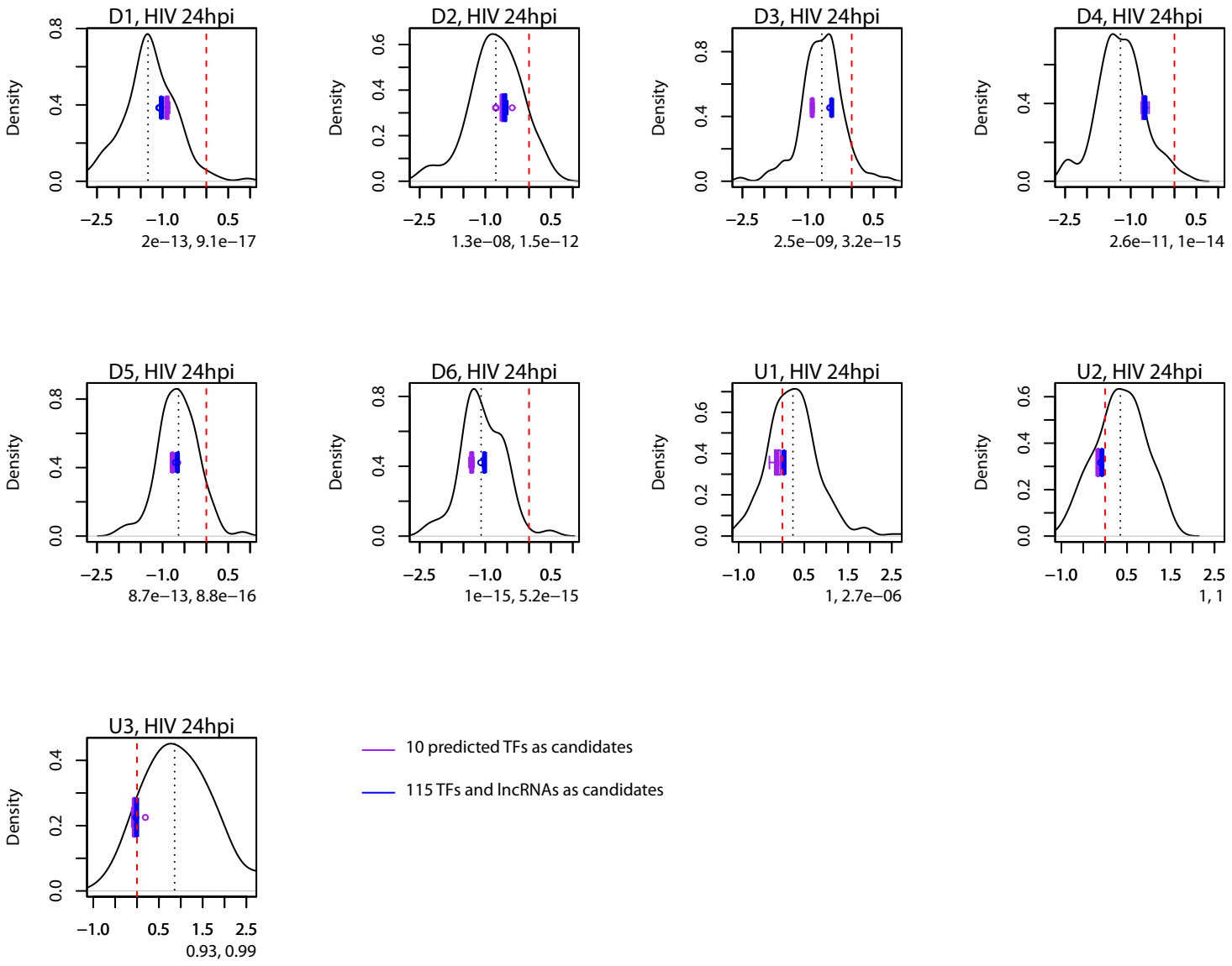
Supplementary Figure 6. LncRNA function and TF inference. A. LncRNA and biological function association matrix. Functions (rows) and lncRNAs (columns) are shown as positively (red), negatively (blue) or not associated (white) with lncRNA tissue expression profiles. The results were limited to functions enriched in genes differentially expressed at 12 hpi and identified gene modules as described later, and lncRNAs differentially expressed at 12 hpi. For the GSEA analysis, we built custom gene sets by extracting out the 12 hpi DE genes from the functions enriched in 12 hpi DE genes (Fig 2E), one gene set for each function. Similarly a gene set was constructed for each gene module. B. Transcription factors (TF) binding sites enriched in promoters of differentially expressed (DE) lncRNAs. Left, the log2 infection/mock expression ratios of TFs enriched in DE lncRNAs but not in non-DE lncRNAs at 12 (orange) and 24 (red) hpi, in the order of mRNAseq and Total RNAseq. Middle, enrichment p-values of TFs enriched in either DE (purple) or non-DE (green) lncRNAs. Right, Hierarchical clustering of TFs based on the degrees of binding site co-occurring in the same promoters. The labels of the leaves are in the order: name, number of target lncRNA, TF enriched in 12 DE genes (+) or not (-), with at least one documented interaction with HIV-1 protein (+) or not, and DE during HIV-1 infection ('b' for 12 hpi DE subgroup b and '+' for DE at 24 hpi) or not (-).



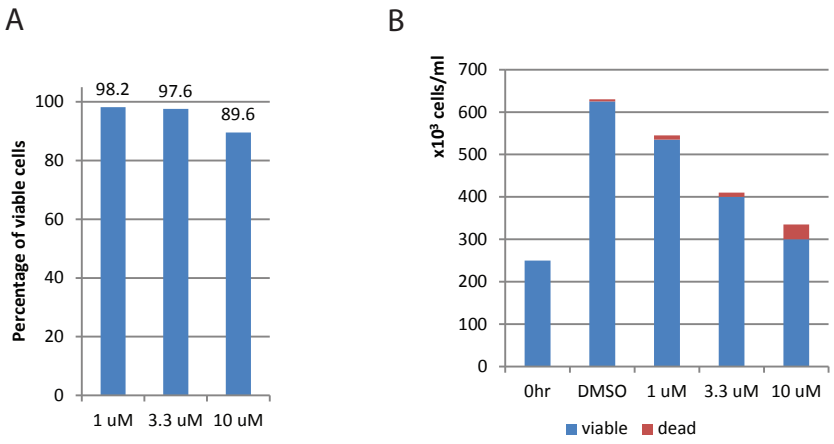
Supplementary Figure 7. Regression coefficients of the predictive models learned from the compendium expression data. A. Similarly as in Figure 6B (top), the regression coefficients of the predictive models learned for gene models ‘D1-D6’ with 10 identified TFs as candidate regulators. B. Similarly as in Figure 6B (top), the regression coefficients of the predictive models learned for gene models ‘U1’ with the expanded set of 115 candidate regulators (102 TFs and 13 lncRNAs) as predictors. The double arrows indicate TFs which were in the set of 10 identified TFs. The regression coefficients for other gene models were omitted due to their inaccurate predictions.



Supplementary Figure 8. Predicted gene module expression changes at 12hpi during HIV infection for all 9 gene modules. Similarly as in Figure 6B (bottom), the boxplot in purple shows the spread of 10 predicted values from the learned models with 10 identified TFs as candidate regulators, and the blue one for the learned models with the expanded set of 115 candidate regulators as predictors. The black line shows the distribution of gene expression changes measured at 12hpi (Total RNAseq).



Supplementary Figure 9. Predicted gene module expression changes at 24hpi during HIV infection for all 9 gene modules. Similarly as in Figure 6B (bottom), the boxplot in purple shows the spread of 10 predicted values from the learned models with 10 identified TFs as candidate regulators, and the blue one for the learned models with the expanded set of 115 candidate regulators as predictors. The black line shows the distribution of gene expression changes measured at 24hpi (mRNAseq).



Supplementary Figure 10. Cell counts under lycorine treatment, without HIV infection. A. Percentage of viable cells in uninfected cultures of SUP-T1 cells treated with increasing concentration of lycorine at 24 hours post treatment. B. Cell counts in the same lycorine treated samples at 24 hours post treatment, in comparison to the starting number of cells (250x103 cells/ml). DMSO is the negative control for lycorine treatment. The dead cells were identified using trypan-blue exclusion assay. A representative experiment is shown here.