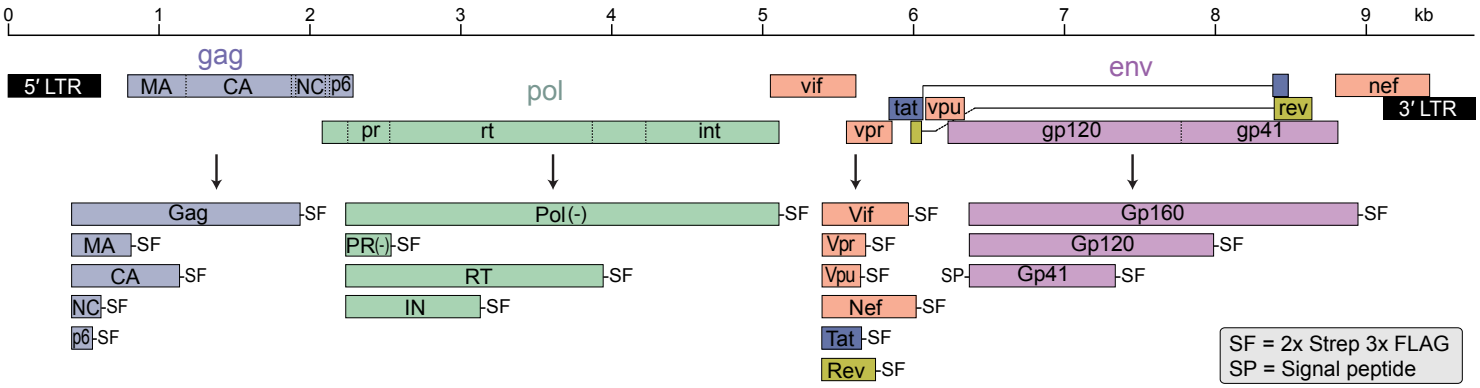
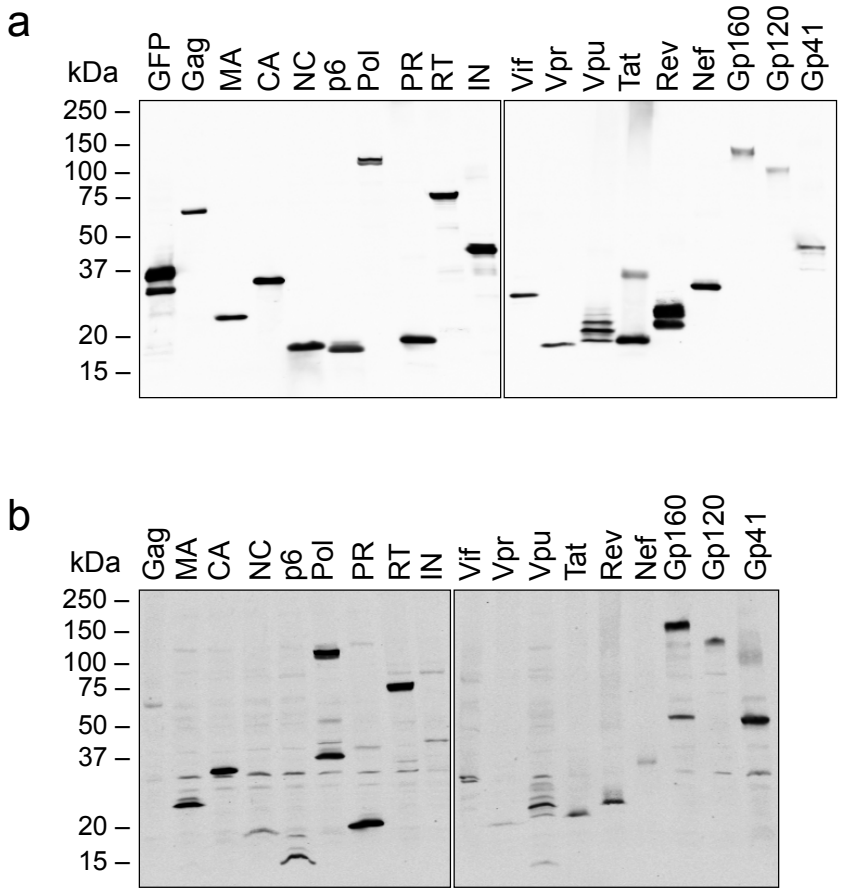
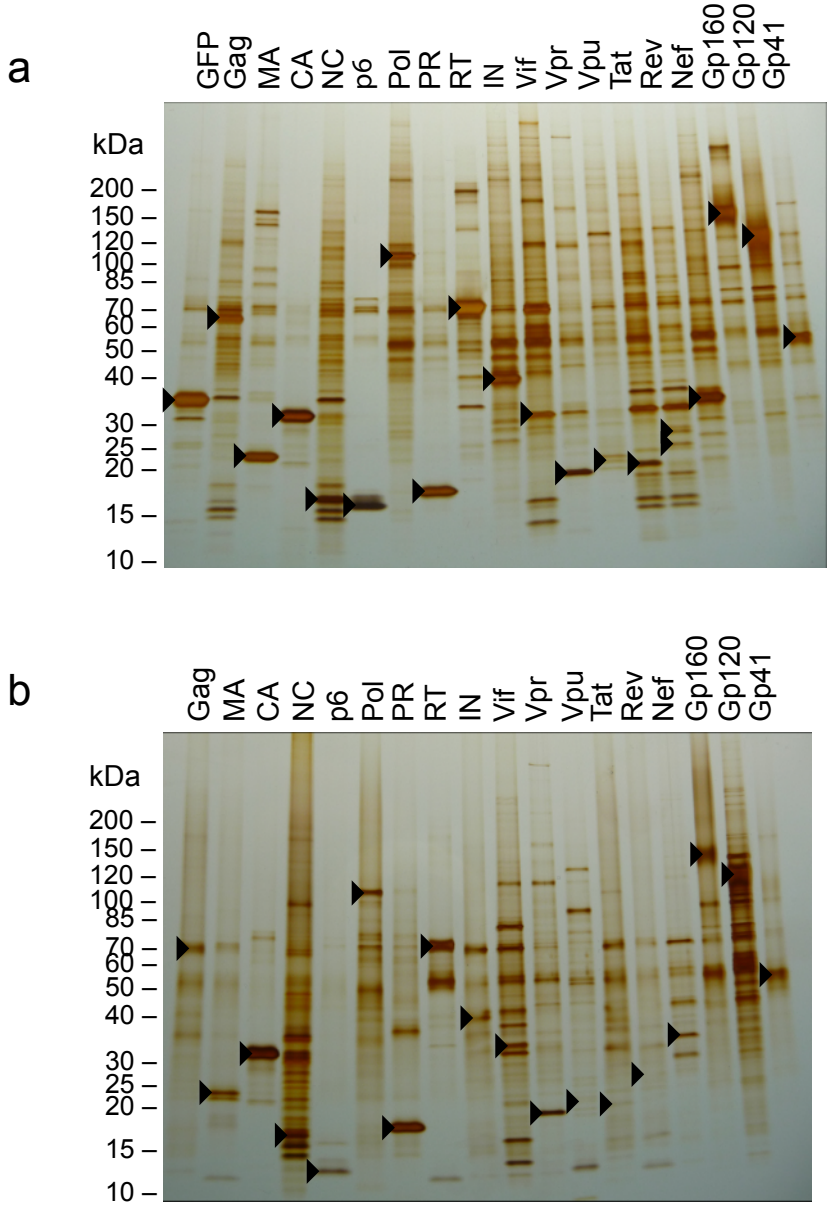
doi:10.1038/nature10719



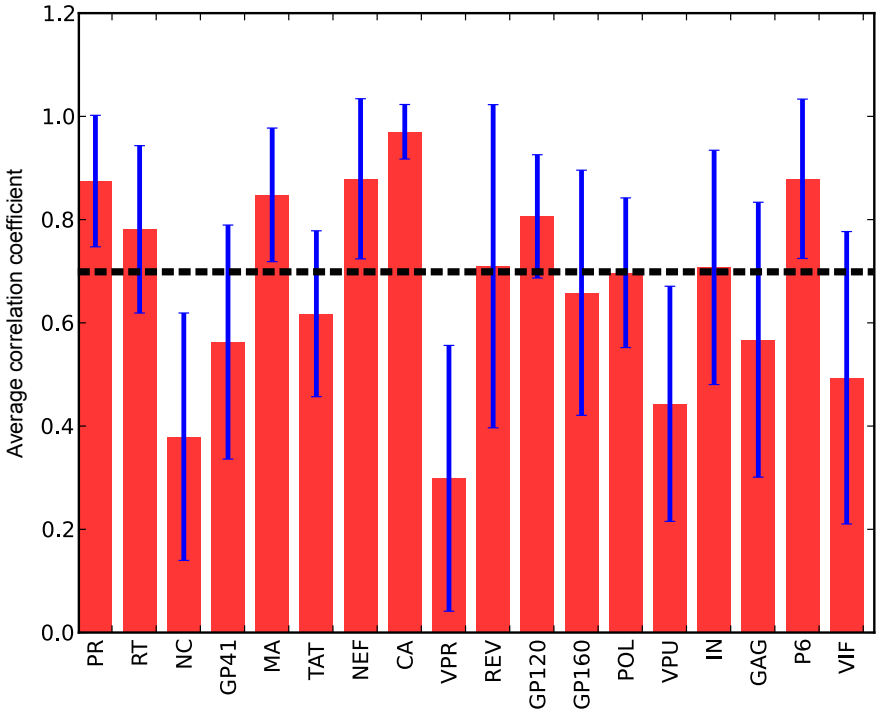
Supplementary Figure 1. Schematic representation of the HIV genome. The 3 polyproteins (Gag, Pol and Env (Gp160)), processed proteins (MA, CA, NC, p6, PR, RT, IN, Gp120 and Gp41), and accessory proteins (Vif, Vpr, Vpu, Nef, Tat and Rev) were cloned and affinity tagged (2xStrepTagII-3xFLAG (SF)) on their C-terminal ends. PR was mutated to be catalytically inactive (PR(-)), and the influenza HA signal peptide sequence was fused to gp41. Sources and amino acid sequences of all factors, many of which are codon-optimized, are in **Supplementary Table 1** and **Supplementary Methods**.



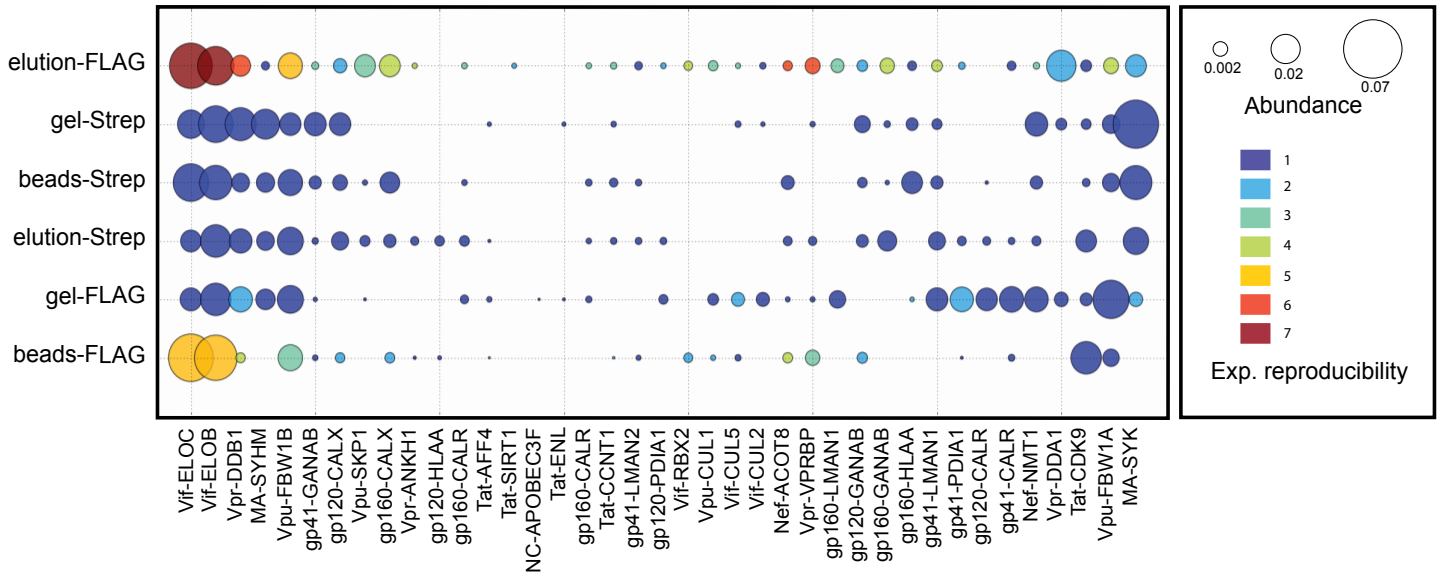
Supplementary Figure 2: Western blot analysis of all 18 HIV proteins. Anti-FLAG western blot analysis of cell lysates of all 18 HIV-SF proteins after being transiently transfected into HEK293 cells (**a**) or induced via tetracycline in Jurkat cells stably expressing each tagged factor (**b**).



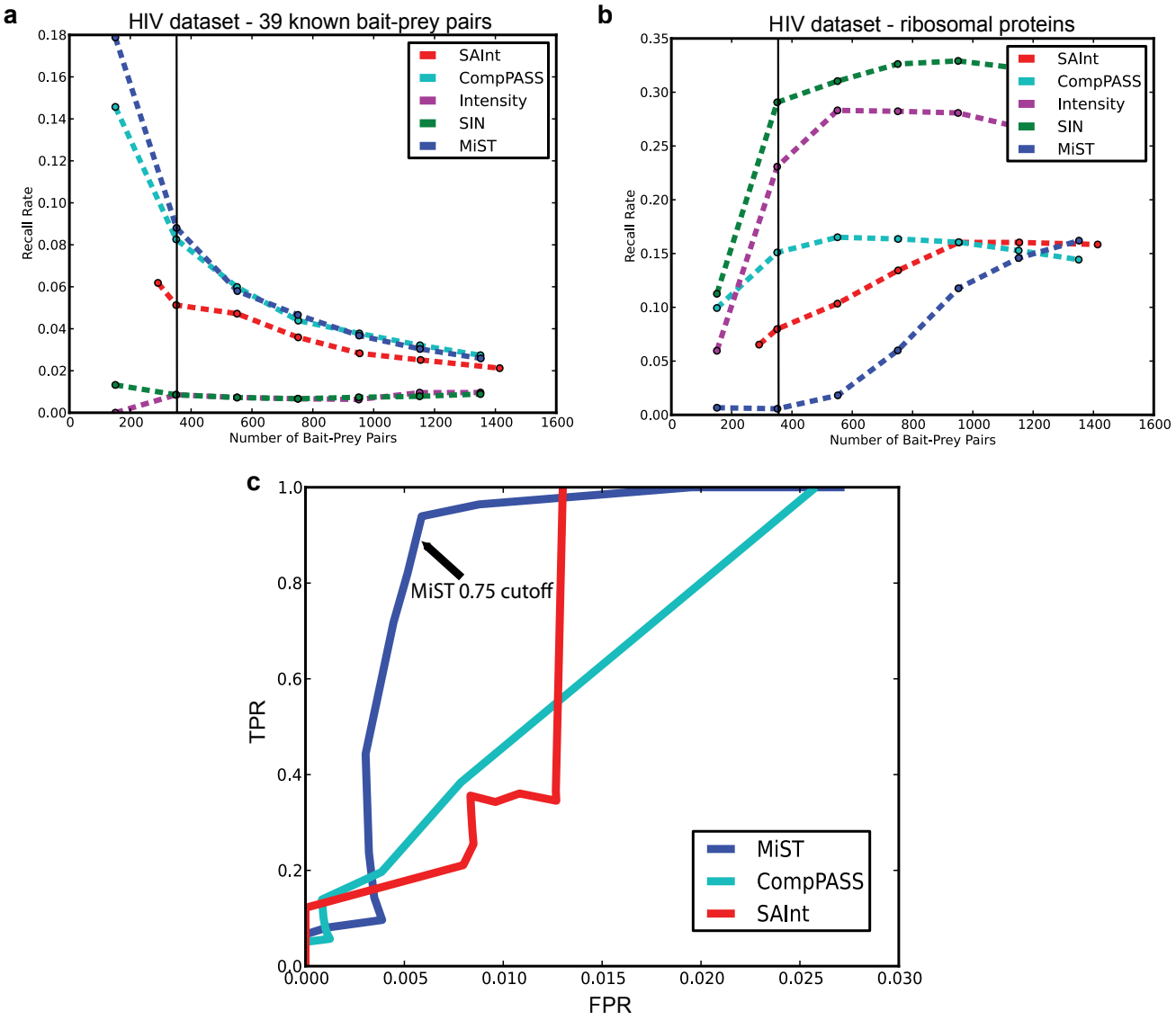
Supplementary Figure 3: Silver stained SDS-PAGE of HIV affinity purifications. Eluates of FLAG affinity purifications from both HEK293 (**a**) and Jurkat (**b**) cells were subjected to SDS-PAGE and stained with silver. The tagged HIV protein in each lane is marked by a black arrow.



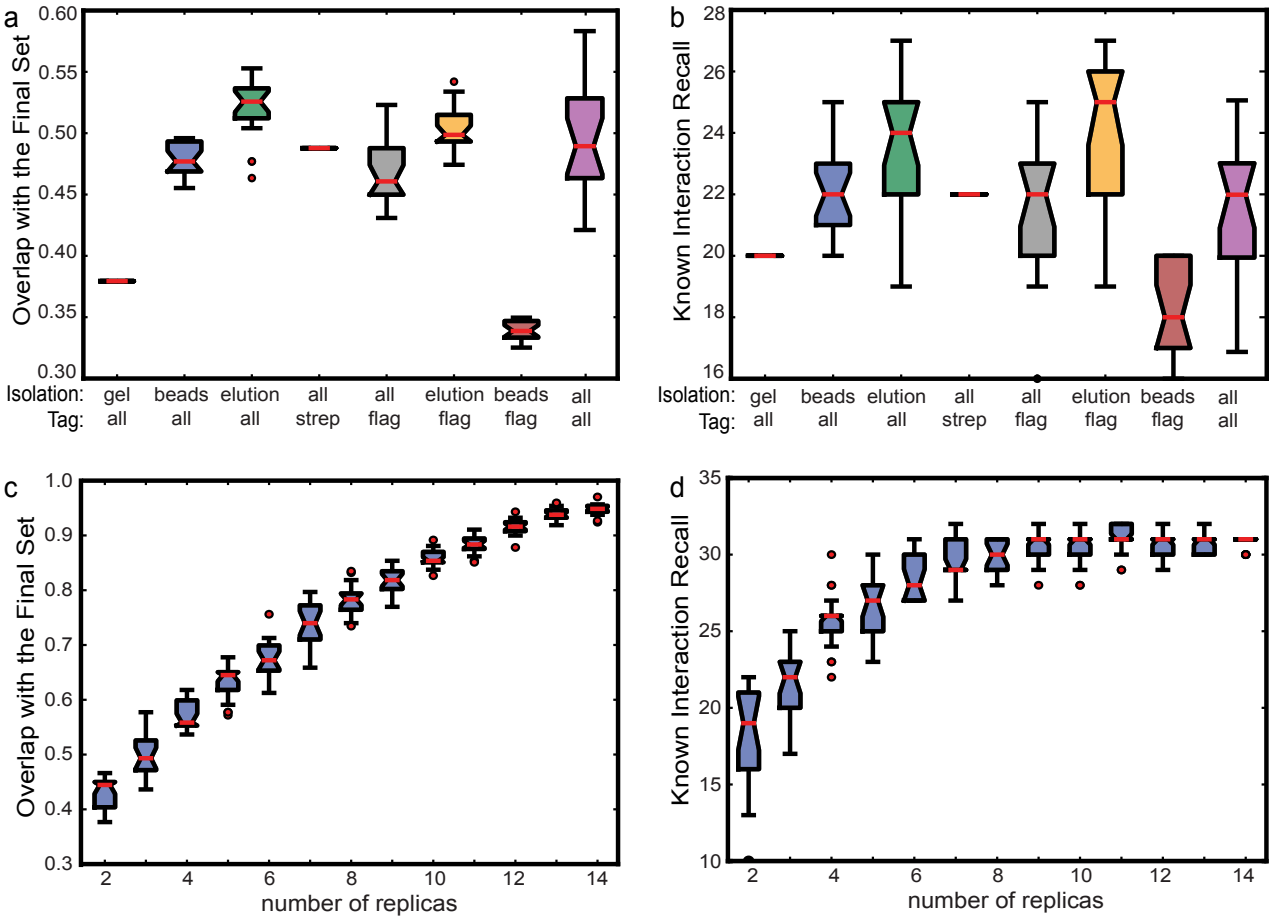
Supplementary Figure 4: Average correlation coefficients for replicate experiments for each HIV protein. For a given bait, we calculated cross-correlations between experiment vectors derived from all combinations of affinity tags and purification protocols, and then determined the average values (bar height) and standard deviations (error bars) of these correlations. The average correlation coefficient across all baits (0.68) is shown with black dashed line.



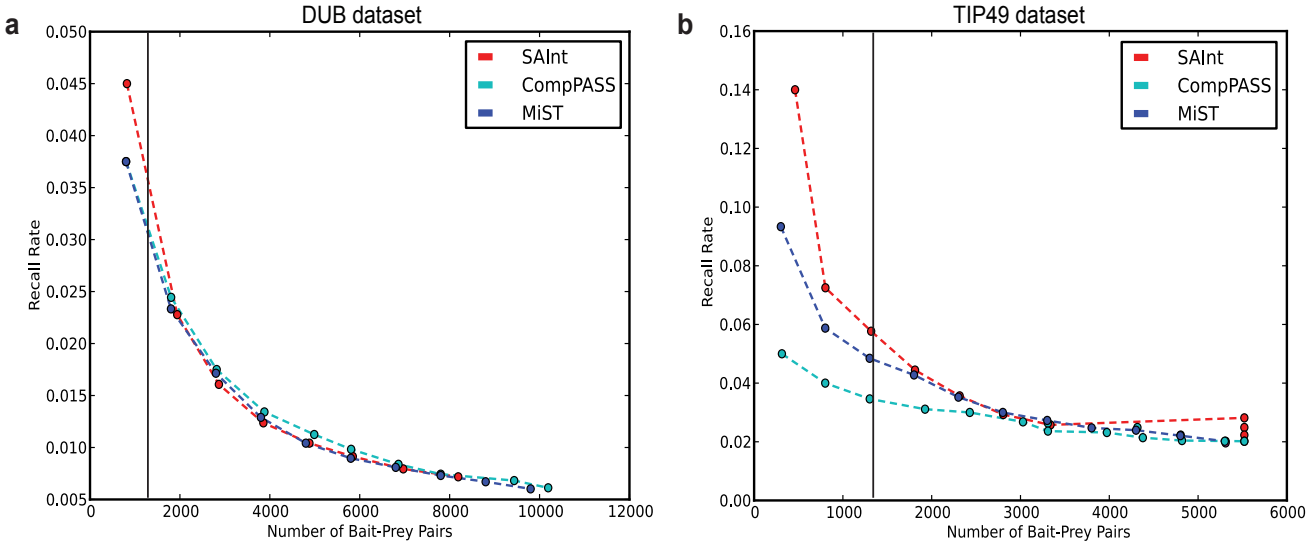
Supplementary Figure 5: Reproducibility and abundance of the well-characterized human-HIV protein pairs. All 39 well-characterized human-HIV protein pairs are shown clustered by their abundances (node size) and reproducibility (node color) for each combination of affinity tag and isolation protocol.



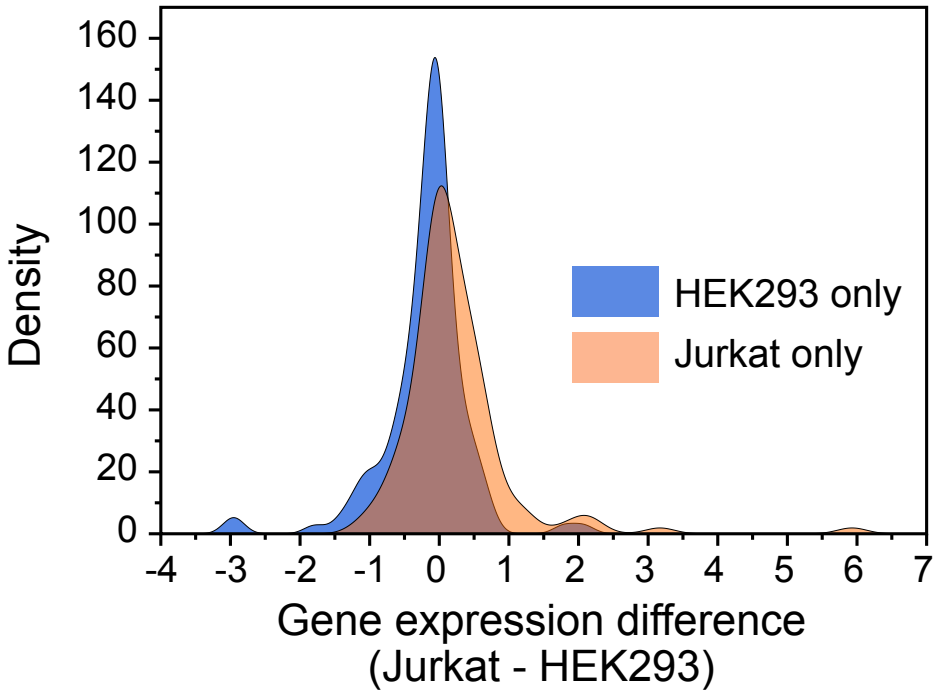
Supplementary Figure 6: Comparison of MiST, CompPASS and SAInt scoring systems to the HIV-human PPI dataset. Using the recall plots, MiST performs better than CompPASS and SAInt when counting (**a**) the number of 39 well-characterized human-HIV pairs (true positives) and (**b**) a set of 1596 interactions involving ribosomal proteins (false positives). The black lines correspond to the MiST score cut-off of 0.75. (**c**), Using the set of true and false positives (see above), an ROC assessment of the false-positive rates (FPR) and true-positive rates (TPR) reveals that MiST behaves superior when compared to COMPass and SAInt. The MiST score cut-off (0.75, marked) falls just before the plateau of the line.



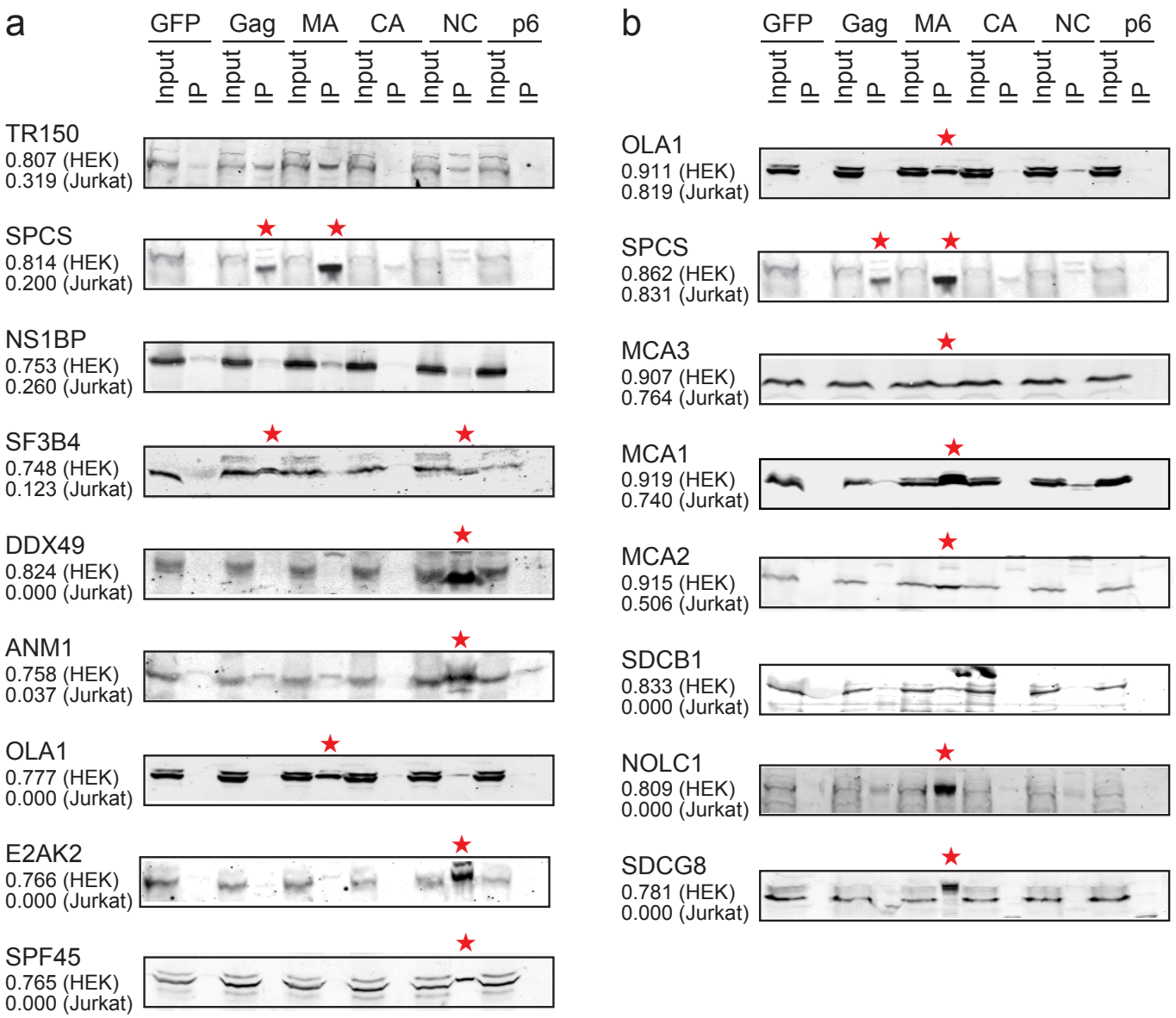
Supplementary Figure 7: Analysis of different experimental protocols and number of replicates. Sets of the bait-prey pairs obtained by resampling of different experimental protocols were compared to the final set of 387 HEK293-derived PPIs (**a**) and to the 39 well-characterized HIV-human PPIs (**b**). Similar analysis was carried out to study the overlap between sets obtained by sampling different number of replicas (**c** and **d**). Box plots identify the middle 50% of the data (box), the median (horizontal red line), median confidence interval (notch), and the extreme points (data points greater than 1•IQR shown with red circles).

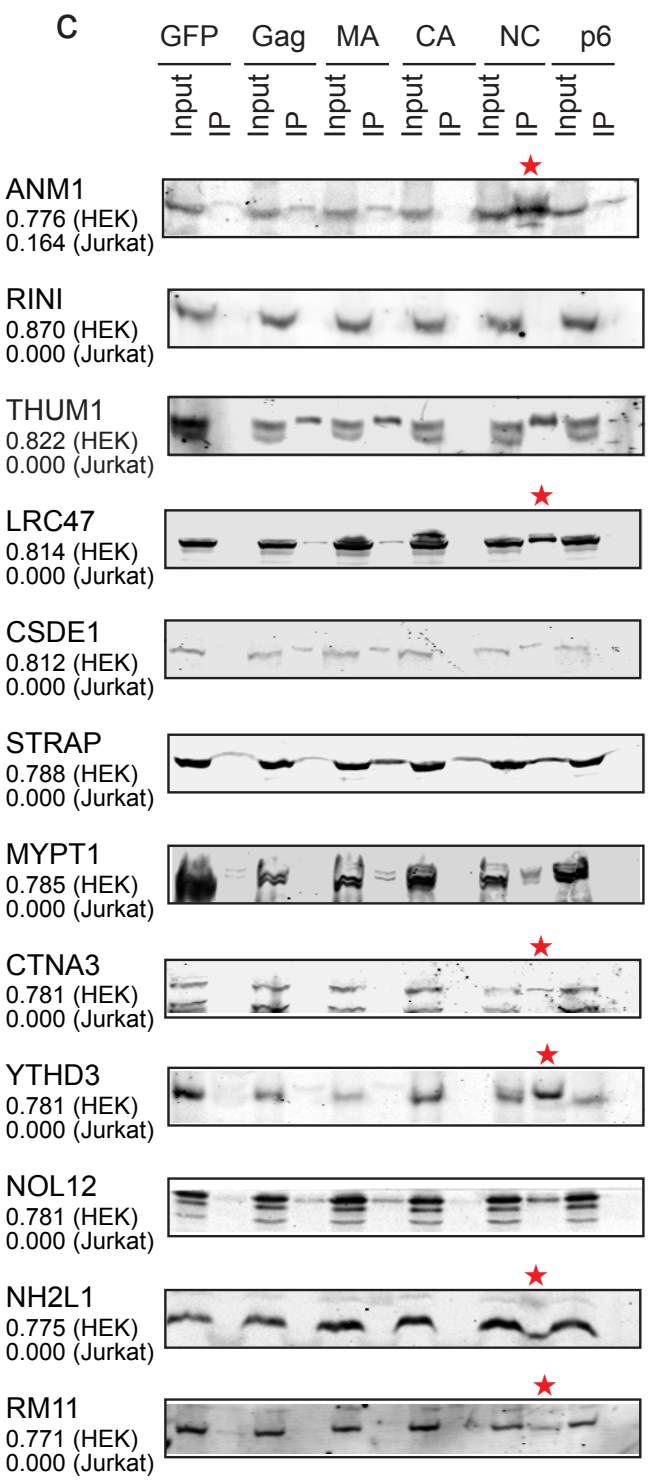


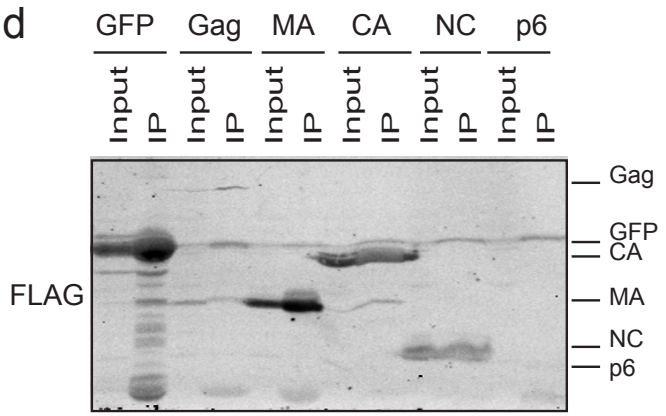
Supplementary Figure 8: Comparison of the MiST, CompPASS and SAInt scoring systems using the DUB (a) and TIP49 (b) datasets. Using two different datasets (DUB and TIP49) and two sets of well-characterized interactions, recall plots were generated to compare the three different scoring systems. The black lines correspond to the cut-offs used in the original studies[36,37].



Supplementary Figure 9: Distribution of the gene expression scores (measured by the difference in z scores of HEK293 and Jurkat cells) for each protein in HEK293 and Jurkat specific interactions. Note that proteins in HEK293 specific interactions (HEK293 only) tend to have shifted expression to the left indicative of higher expression in HEK293 cells, and similarly host proteins in Jurkat specific interactions (Jurkat only) have an expression distribution shifted towards right indicating a higher expression in Jurkat cells. The difference between the distributions was highly significant ($p = 8.16 \bullet 10^{-7}$; Mann-Whitney U test).

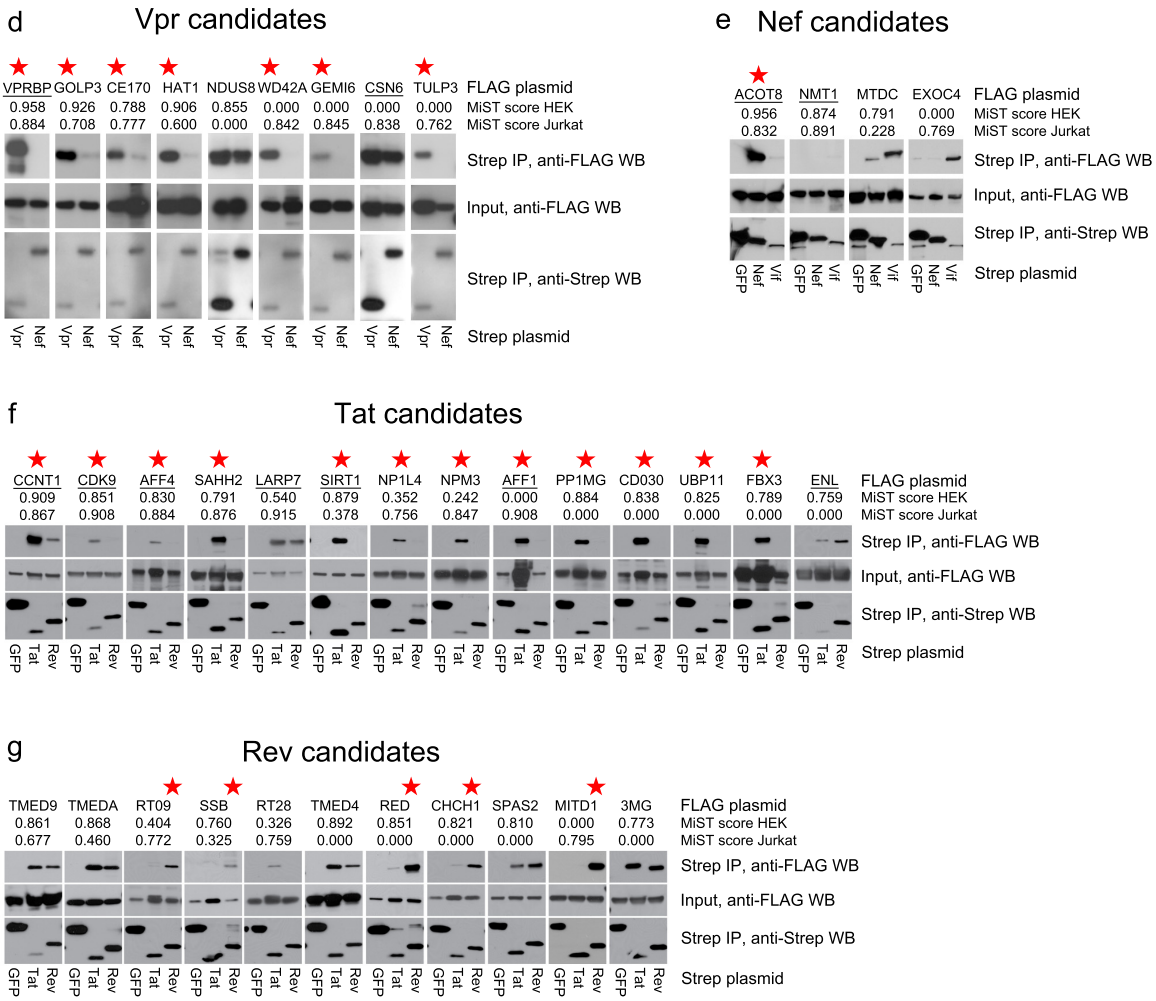

a
GFP
Gag
MA
CA
NC
p6
Input
IP
TR150
0.807 (HEK)
0.319 (Jurkat)
SPCS
0.814 (HEK)
0.200 (Jurkat)
NS1BP
0.753 (HEK)
0.260 (Jurkat)
SF3B4
0.748 (HEK)
0.123 (Jurkat)
DDX49
0.824 (HEK)
0.000 (Jurkat)
ANM1
0.758 (HEK)
0.037 (Jurkat)
OLA1
0.777 (HEK)
0.000 (Jurkat)
E2AK2
0.766 (HEK)
0.000 (Jurkat)
SPF45
0.765 (HEK)
0.000 (Jurkat)
b
GFP
Gag
MA
CA
NC
p6
Input
IP
OLA1
0.911 (HEK)
0.819 (Jurkat)
SPCS
0.862 (HEK)
0.831 (Jurkat)
MCA3
0.907 (HEK)
0.764 (Jurkat)
MCA1
0.919 (HEK)
0.740 (Jurkat)
MCA2
0.915 (HEK)
0.506 (Jurkat)
SDCB1
0.833 (HEK)
0.000 (Jurkat)
NOLC1
0.809 (HEK)
0.000 (Jurkat)
SDCG8
0.781 (HEK)
0.000 (Jurkat)



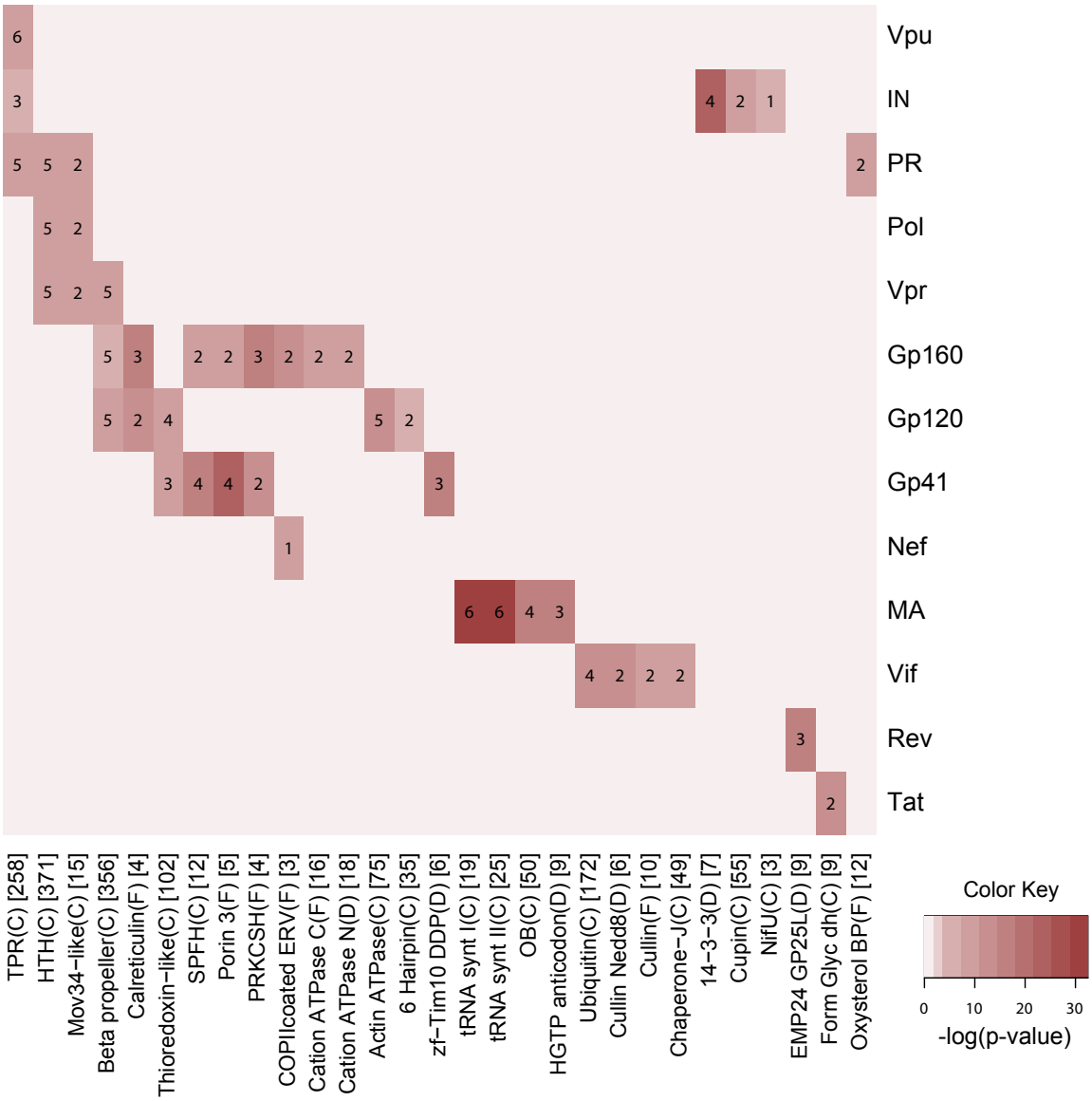


Supplementary Figure 10: Co-immunoprecipitations of interaction candidates using antibodies against the endogenous proteins. SF-tagged HIV proteins or GFP were expressed in 293T cells and pulled down with StrepTactin beads. Cell lysates and eluates were analyzed by western blotting using antibodies against interaction candidates of Gag (**a**), MA (**b**), and NC (**c**) indicated on the left. MiST scores determined by AP-MS analyses from HEK293 and Jurkat cells are given below each candidate for the respective HIV bait. Bands indicating specific interactions are highlighted with a star. Anti-FLAG antibody was used to detect GFP and the HIV proteins (**d**). 17 of the 26 tested interactors were found to be specific.

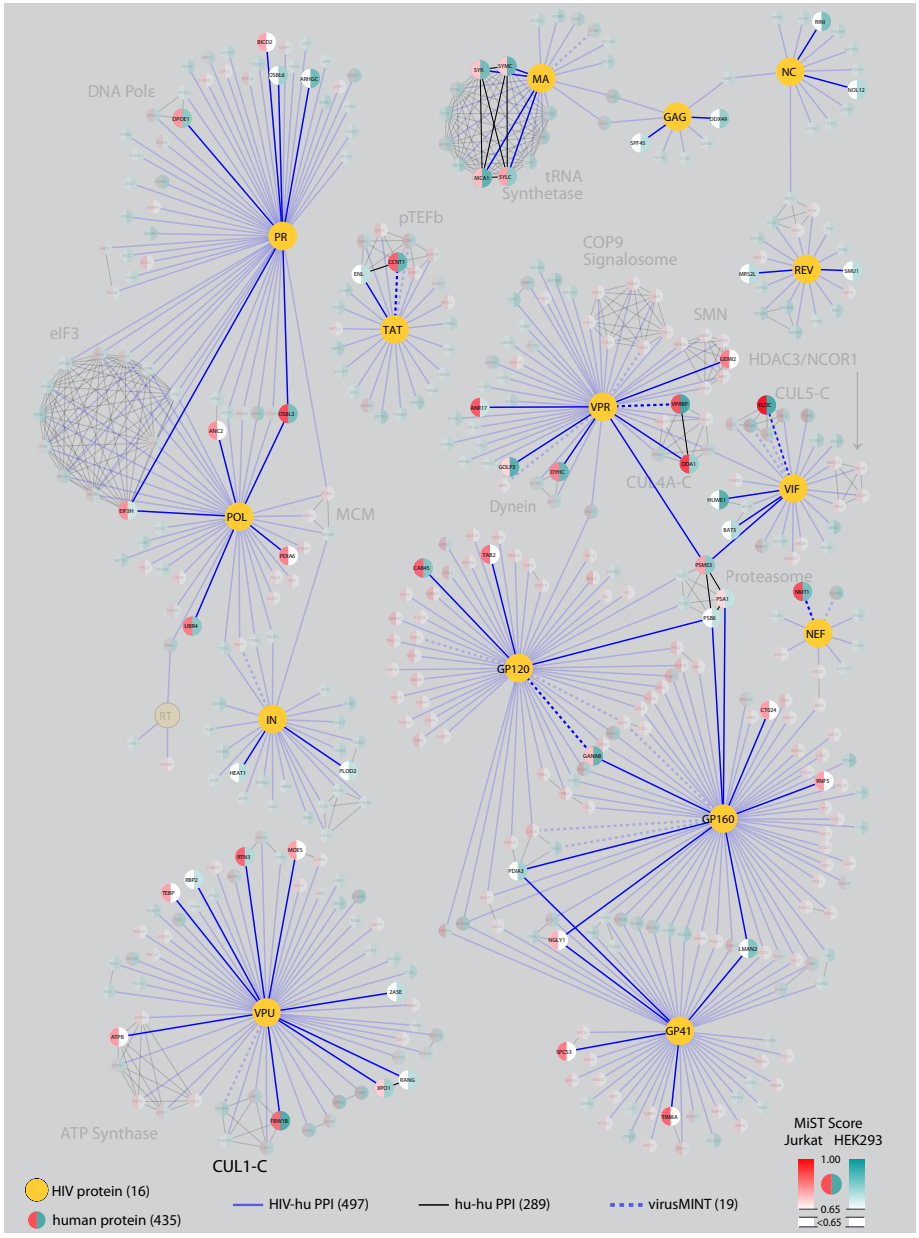
Supplementary Figure 11
page 1of 2



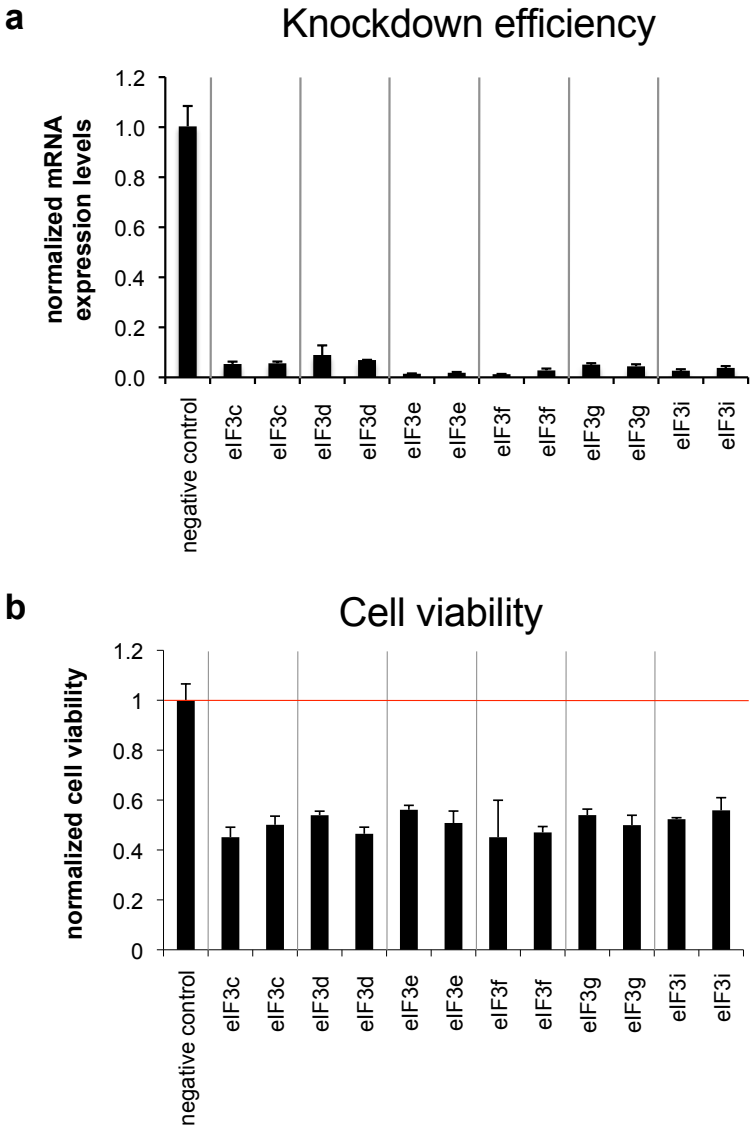
Supplementary Figure 11: Co-immunoprecipitations of interaction candidates using co-expression of tagged human and viral proteins. FLAG-tagged human interaction candidates of PR (**a**), IN (**b**), Vif (**c**), Vpr (**d**), Nef (**e**), Tat (**f**), and Rev (**g**) were coexpressed with StrepTagII-tagged HIV proteins or GFP in 293T cells , and a StrepTactin pulldown was performed. Eluates were analyzed by SDS-PAGE and anti-FLAG (upper panels), as well as anti-Strep-tagII (lower panels) western blotting. Cell lysates were probed against FLAG (middle panels). The MiST scores determined from HEK293 and Jurkat AP-MS analyses are given below each candidate. Interactions that are considered specific are marked with a star. Note that 79% (80/ 101) of the tested host proteins coimmunoprecipitated specifically with its respective HIV bait, while the confirmation rate among previously characterized interactions (underlined) is only 63% (10/16).



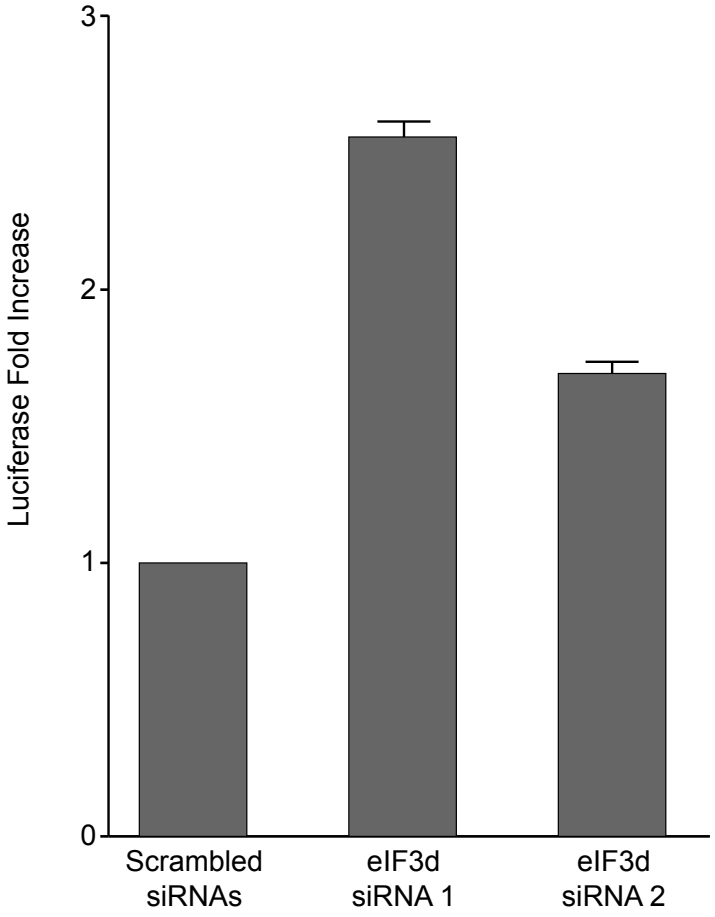
Supplementary Figure 12: A heat-map representing enriched domains (x-axis) that are present in human proteins that were pulled-down with HIV proteins (y-axis). The domain titles include domain name, domain type (C=Clan; D=domain; F=Family), and the number of domain occurrences in the human proteome (in square brackets). Coordinates are colored according to corresponding statistical significance (-log of p-value) of the domain's over-representation. Each square consists also of the count of human proteins that include the relevant domain. Only domains for which the enrichment has a p-value below 0.005 are displayed.



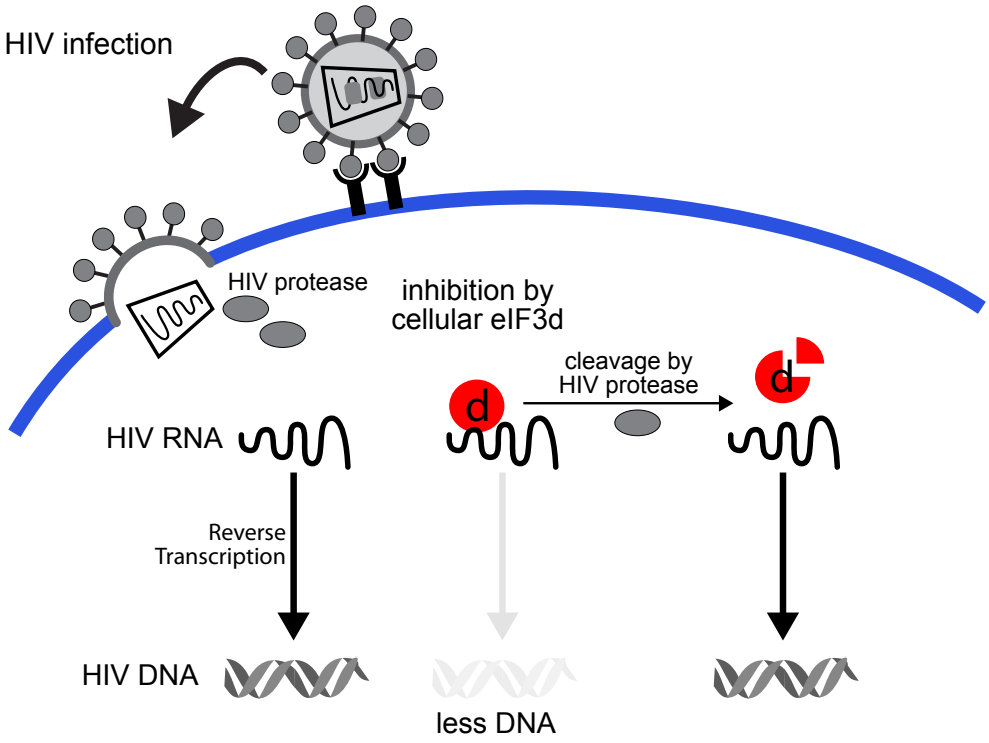
Supplementary Figure 13: Overlap of HIV-1 host factors identified by RNAi screens and AP-MS. Proteins identified both in one of the four published RNAi/HIV-1 infections screens[38-41] (**Supplementary Data 7**) and the HIV-human protein-protein interactions identified in this study are highlighted (for a list of overlapping proteins see **Supplementary Table 6**).



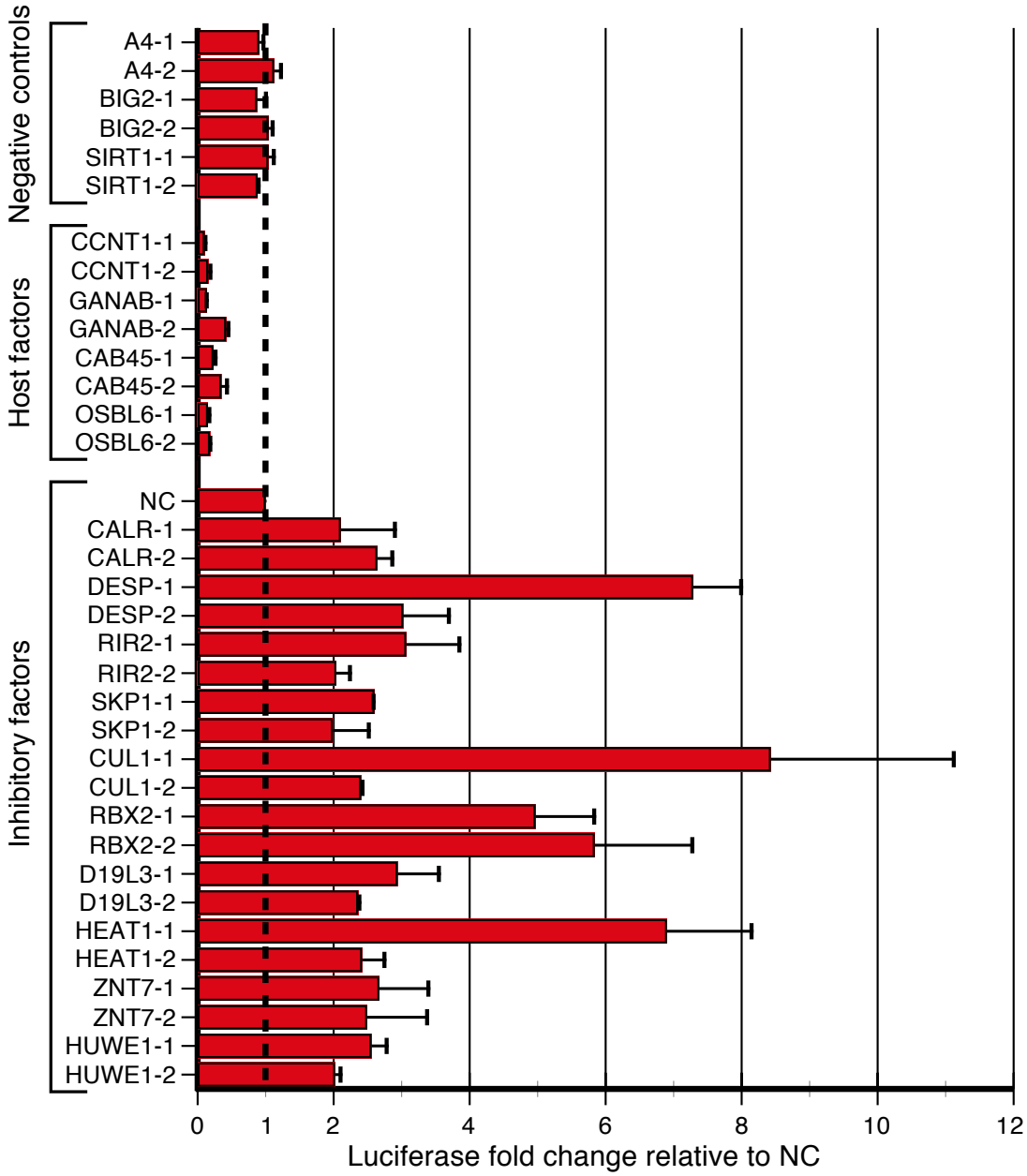
Supplementary Figure 14: Control experiments for infection experiment shown in Fig. 4e, f. a, Knockdown of eIF3 subunits following siRNA treatment of HeLa P4.R5 cells was verified by determining mRNA expression levels using quantitative real-time PCR. Data represents averages and standard deviations of triplicates. **b**, siRNAs targeting eIF3 subunits exhibit comparable cytotoxicity. The viability of siRNA-transfected HeLa P4.R5 cells was determined 72 hours after transfection by measuring intracellular ATP levels using a luminescent substrate. Values were normalized to a panel of control siRNAs specifically designed not to target any gene transcripts and represent averages and standard deviations of triplicates.



Supplementary Figure 15: Control infection experiment for HIV-1 late RT assay. Human 293T cells were transfected with either of two independent siRNAs targeting eIF3d mRNA or with scrambled control siRNAs for 48 hours. The cells were then challenged with an HIV-1 vector encoding luciferase and infection level was quantified based upon luciferase expression. These experiments were performed in triplicate and the values shown represent the mean averages with the standard deviation of the data shown by error bars.

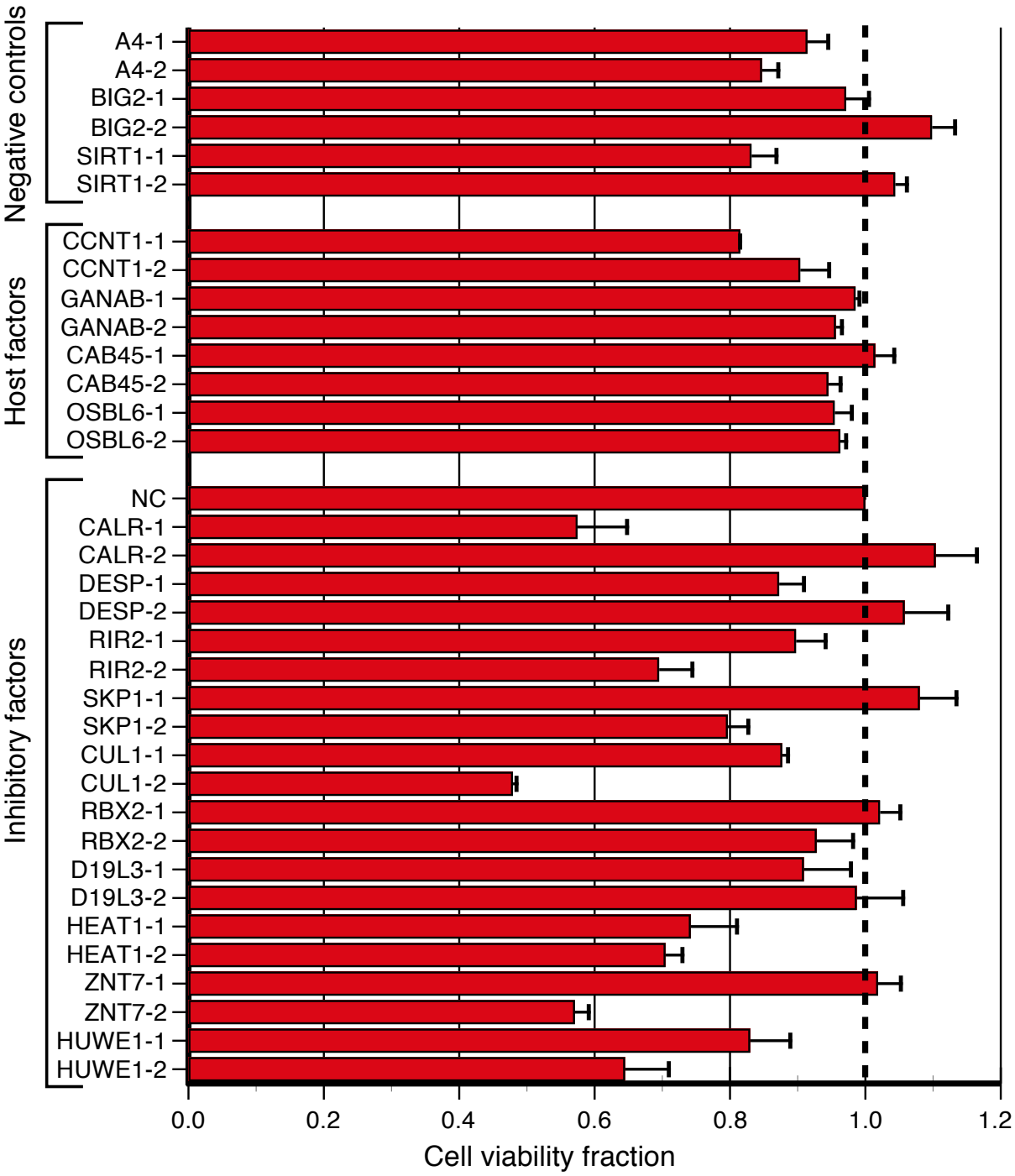


Supplementary Figure 16: Model of the role of eIF3d in the HIV life cycle. One model consistent with the data suggests eIF3d, via its RNA binding domain (RRM), binds to the viral genome, inhibiting the reverse transcription of the HIV RNA. Cleavage of eIF3d by HIV PR could alleviate this inhibition.

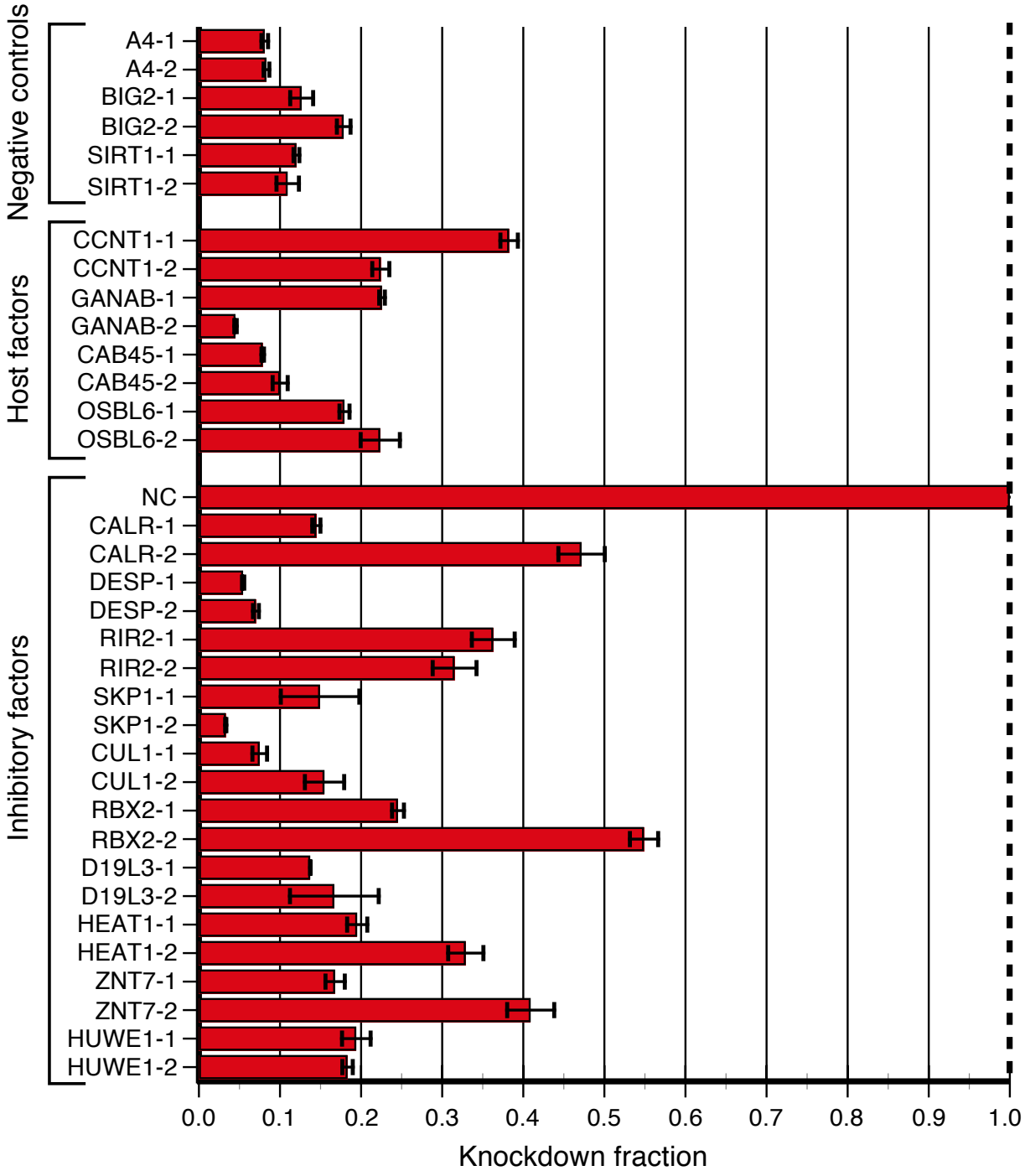


Supplementary Figure 17. Infection experiments for inhibitory factors. Human 293T cells were transfected with either of two independent siRNAs targeting 17 different factors that we found in our AP-MS studies (10 inhibitory factors, 4 previously described host factors[38,39] and 3 factors that have no effect on HIV infection) as well as scrambled control siRNAs for 48 hours. The cells were then challenged with an HIV-1 vector encoding luciferase and infection level was quantified based upon luciferase expression. These experiments were performed 2 to 6 times and the data presented represents the mean with standard deviation shown by error bars. Data from the host factors were normalized as previously described[39]. For the inhibitory factors, both siRNAs for all genes provide a two-fold increase (or greater) in the luciferase assay.

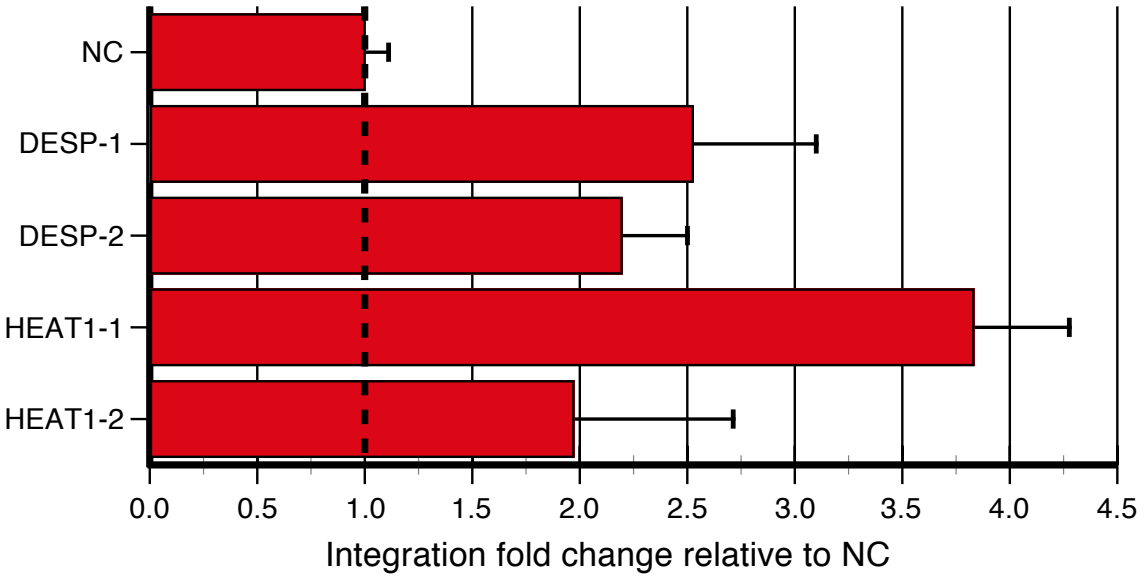
Supplementary Figure 17



Supplementary Figure 18. Toxicity control experiments for infection assays shown in Supplementary Figure 17. The viability of the siRNA-transfected cells was determined 72 hours after infection by measuring intracellular ATP levels using a luminescent substrate. Values are derived from 2 to 6 experiments and were normalized to scrambled siRNAs (NC) specifically designed not to target any genes.



Supplementary Figure 19. Knockdown control experiments for infection assays shown in Supplementary Figure 17. Knockdown of the genes following siRNA treatment was verified by determining mRNA expression levels using real-time PCR. The data represents averages and standard deviations of triplicate experiments and scrambled siRNA (NC) was used as a control and for normalization.



Supplementary Figure 20: Integration assay after knock-down of two inhibitory factors. Two factors (DESP and HEAT1) that have inhibitory effects on HIV infection (**Supplementary Figure 17**) and found to be physically associated with Integrase (**Figure 3**) were also found to have inhibitory effects on integration. Two different siRNAs for each gene were used in a previously published HIV integration assay[39] and scrambled siRNAs were used as a negative control. The experiments were done in triplicate.

Supplementary Tables

Supplementary Table 1: PCR templates used for cloning of HIV-1 ORFs

ORF	Template	Provided by	HIV strain	Reference
gag	pGag-EGFP	M. Resh, G. Pavlakis, NIH AIDS Research and Reference Reagent Program	NL4-3	[43]
ma	pGag-EGFP	see above	NL4-3	[43]
ca	pGag-EGFP	see above	NL4-3	[43]
nc	pGag-EGFP	see above	NL4-3	[43]
p6	pGag-EGFP	see above	NL4-3	[43]
pol	pcDNA3_pc_syn_gp(IIIB)	R. Wagner, University of Regensburg	BH10	[44]
pr	newly synthesized	-	NL4-3	-
rt	newly synthesized	-	NL4-3	-
in	pCEP-IN	Z. Debyser, KU Leuven	NL4-3	[45]
vif	newly synthesized	-	NL4-3	
vpr	pcDNA3-hVpr	M. Lenardo, VRC, NIAID, NIH	NL4-3	[47]
vpu	pcDNA-Vphu	K. Strebel, S. Bour, NIH AIDS Research and Reference Reagent Program	NL4-3	[46]
tat	HXB2 provirus	NIH AIDS Research and Reference Reagent Program	HXB2	-
rev	HXB2 provirus	NIH AIDS Research and Reference Reagent Program	HXB2	-
nef	HXB2 provirus*	NIH AIDS Research and Reference Reagent Program	HXB2	-
gp160	pVRC3300	G. Nabel, VRC, NIAID, NIH	HXB2	[42]
gp120	pVRC3300	see above	HXB2	[42]
gp41	pVRC3300	see above	HXB2	[42]

* HXB2 Nef contains a premature stop at aa 124. This stop was mutated to W to match the subtype B consensus sequence at this position.

Supplementary Table 2: Number of affinity purifications performed and samples analyzed by MS. The columns "HEK APs" and "Jurkat APs" report on how many independent affinity purifications were performed per bait and cell type (biological replicates). Columns "Beads", "Elution", and "Gel" indicate the type of samples that were analyzed by MS for each replicate.

	HEK293								Jurkat								
	HEK APs	Flag			Strep			Total HEK samples analyzed	Jurkat APs	Flag			Strep			Total Jurkat samples analyzed	Total HEK and Jurkat samples analyzed
		Beads	Elution	Gel	Beads	Elution	Gel			Beads	Elution	Gel	Beads	Elution	Gel		
Gag	8	3	8	2	1	1	1	16	7	1	6	0	0	2	0	9	25
MA	5	2	5	1	1	1	1	11	6	0	6	0	0	1	0	7	18
CA	5	2	5	0	1	1	1	10	6	0	6	0	0	1	0	7	17
NC	5	2	5	1	1	1	1	11	3	0	2	0	0	1	0	3	14
p6	5	2	5	1	1	1	1	11	6	0	6	0	0	1	0	7	18
Pol	5	2	5	1	1	1	1	11	4	0	4	0	0	1	0	5	16
PR	6	3	6	2	1	1	1	14	5	1	5	0	0	1	0	7	21
RT	6	3	6	2	1	1	1	14	9	2	8	0	0	2	0	12	26
IN	8	5	8	2	1	1	1	18	5	1	5	0	0	1	0	7	25
Vif	8	5	8	2	1	1	1	18	9	1	8	0	0	2	0	11	29
Vpr	7	4	7	2	1	1	1	16	10	1	9	0	0	2	0	12	28
Vpu	8	5	8	1	1	1	1	17	9	1	8	0	0	2	0	11	28
Tat	6	3	6	2	1	1	1	14	5	1	5	0	0	1	0	7	21
Rev	6	3	6	2	1	1	1	14	6	1	6	0	0	1	0	8	22
Nef	7	4	7	2	1	1	1	16	9	2	8	0	0	2	0	12	28
gp160	5	2	5	1	1	1	1	11	9	0	7	0	0	3	0	10	21
gp120	4	2	3	0	1	1	1	8	2	0	4	0	0	0	0	4	12
gp41	5	2	5	1	1	1	1	11	6	0	4	0	0	2	0	6	17
	109							**241**	**116**							**145**	**386**

Supplementary Table 3: Benchmark of 39 well-characterized HIV-human protein pairs.

HIV protein	Human protein (UniProt ID)	Protein name
MA	P49590	SYHM
MA	Q15046	SYK
NC	Q8IUX4	ABC3F
Vif	Q15370	ELOB
Vif	Q15369	ELOC
Vif	Q93034	CUL5
Vif	Q9UBF6	RBX2
Vif	Q13617	CUL2
Vpr	Q9Y4B6	VPRBP
Vpr	Q16531	DDB1
Vpr	Q8IWZ3	ANKH1
Vpr	Q9BW61	DDA1
Vpu	Q9UKB1	FBW1B
Vpu	Q9Y297	FBW1A
Vpu	P63208	SKP1
Vpu	Q13616	CUL1
Tat	O60563	CCNT1
Tat	P50750	CDK9
Tat	Q9UHB7	AFF4
Tat	Q03111	ENL
Tat	Q96EB6	SIRT1
Nef	O14734	ACOT8
Nef	P30419	NMT1
Gp160	P07237	PDIA1
Gp160	P27797	CALR
Gp160	P27824	CALX
Gp160	Q14697	GANAB
Gp160	P49257	LMAN1
Gp160	P01892	HLA-A
Gp120	P07237	PDIA1
Gp120	P27797	CALR
Gp120	P27824	CALX
Gp120	Q14697	GANAB
Gp120	P01892	HLA-A
Gp41	P07237	PDIA1
Gp41	P27797	CALR
Gp41	P49257	LMAN1
Gp41	Q12907	LMAN2
Gp41	Q14697	GANAB

Supplementary Table 4: Human proteins that include over-represented domains. HIV-1 proteins (Column 1) are followed by human proteins that were purified along with these viral proteins (Column 2). The corresponding Pfam domain accession/name (3rd and 4th columns) are also listed, and, when possible, with their clan membership (5th and 6th columns). Note that a single protein may increase the counts of several domains (in case of multi-domain proteins) (see remarks 2 and 3 below). In addition, in cases where several HIV proteins were pulled down with a single human protein, the domains of the human protein are counted more than once (see remarks 1, 4 and 5 below).

HIV protein	Human protein	Pfam domain accession	Pfam domain name	Pfam Clan accession	Pfam Clan name
IN	1433B	PF00244	14-3-3 domain	(N.A.)	(N.A.)
	1433E				
	1433G				
	1433Z				
	CAND2	PF02985	Heat repeat	CL0020	TPR
	IPO4				
	RNBP6				
Vpu	IPO9	PF03810	Importin-beta N-terminal domain	CL0020	TPR
	XPO1				
	PRKDC	PF02259	FAT domain		
	RBP2	PF00515	Tetratricopeptide repeat		
	LPPRC	PF01535	PPR repeat		
	XPO5	PF08389	Exportin 1-like		
Vif	BAT3	PF00240	Ubiquitin family	CL0027	Ubiquitin superfamily
	ELOB				
	UBL4A				
	SQSTM	PF00564	PB1 domain	(N.A.)	(N.A.)
Rev	TMED4	PF01105	emp24/gp25L/p24 family/GOLD	(N.A.)	(N.A.)
	TMED9				
	TMEDA				
Gp160	CAF1B [1]	PF00400	WD domain	CL0186	Beta propeller clan
	SEH1 [1]				
	LDLR [1]	PF00058	Low-density		
	PON2 [1]	PF01731	Arylesterase		
	ITAL	PF01839	FG-GAP repeat		
MA	AIMP1	PF01588	Putative tRNA binding	CL0021	OB fold
	SYDC [2]	PF01336	OB-fold nucleic acid binding domain		
	SYK [2]				
	SYNM [2]				
	SYEP [2,3]	PF03129	Anticodon binding domain	(N.A.)	(N.A.)
	SYHM [2]				
	SYTC [2]				

	SYIC	PF00133	tRNA synthetases class I (I, L, M, V)	CL0038	Class I aminoacyl-tRNA synthetase, catalytic
	SYLC				
	SYMC	PF09334	tRNA synthetases class		
	SYQ	PF00749	tRNA synthetases class		
	SYRC	PF00750	tRNA synthetases class		
PR	EIF3A [4]	PF01399	PCI domain	CL0123	Helix-turn-helix
	EIF3C [4]				
	EIF3E [4]				
	EIF3M				
	EIF3K [4]	PF10075	COP9 signalosome,		
	DIAP1	PF06367	Diaphanous FH3	CL0020	TPR
	DOHH	PF03130	PBS lyase HEAT-like		
	EIF3L	PF10255	RNA polymerase I-		
	FKBP4	PF00515	Tetratricopeptide repeat		
	PUM2	PF00806	Pumilio-family RNA		
Pol	RECQ1	PF09382	RQC domain	CL0123	Helix-turn-
Vpr	DC1I2	PF00400	WD domain	CL0186	Beta propeller clan
	DCA11				
	DCAF8				
	RBBP7				
	DDB1	PF03178	CPSF A subunit region		
	CSN1	PF01399	COP9 signalosome, subunit CSN8	CL0123	Helix-turn-helix
	CSN2				
	CSN3				
	CSN4				
	CSN8	PF10075	PCI domain		
Gp120	ITA4	PF01839	FG-GAP repeat	CL0186	Beta propeller
	ERP44	PF00085	Thioredoxin	CL0172	Thioredoxin-like
	PDIA5				
	TXND5 [5]				
	MAGT1	PF04756	OST3/OST6 family		
	HS105	PF00012	HSP70	CL0108	Actin-like ATPase Superfamily
	HSP13				
	HSP76				
	HSP77				
	HYOU1				
Gp41	TIM13	PF02953	Tim10/DDP family zinc finger	(N.A.)	(N.A.)
	TIM8A				
	TIM8B				
	FLOT1	PF01145	SPFH domain/ Band 7 family	CL0433	SPFH superfamily
	PHB2				
	PHB				
	STML2				
	TOM40	PF01459	Eukaryotic porin	(N.A.)	(N.A.)

	VDAC1				
	VDAC2				
	VDAC3				
	PDIA3	PF00085	Thioredoxin	CL0172	Thioredoxin-like
	PRDX4	PF00578	AhpC/TSA family		

(1) These proteins were also purified with Gp120.
(2) In addition to the OB-fold nucleic acid binding domain or the Anticodon binding domain, these proteins include domains that belong to the Class II aminoacyl-tRNA and Biotin synthetases Clan (CL0040). Hence, they also contribute to the over-representation of this clan.
(3) In addition to the Anticodon binding domain, this protein includes also a domain that belongs to the Class I aminoacyl-tRNA synthetase, catalytic domain (CL0038). Hence, it also contributes to the over-representation of this clan.
(4) These proteins were also purified with the HIV-1 Pol.
(5) This protein was also purified with Gp41.

Supplementary Table 5: Overlap between our IP-MS data and the HIV-human interactions in VirusMint database. We found 19 overlapping interactions from our study and those in VirusMint[77] and the significance of this overlap is $8 \cdot 10^{-08}$ calculated by hypergeometric distribution. Number of background interactions was defined as the number of unique human proteins detected in our AP-MS experiments (n = 4,219) multiplied by the number of HIV proteins (n = 18).

BAIT	PREY	HEK293 score	JURKAT score	GENE ID	SYMBOL
Gp120	A0A891	0	0.813	3106	HLA-B
Gp120	Q14554	0	0.847	10954	PDIA5
Gp120	Q14697	0.909	0.732	23193	GANAB
Gp160	A0ZUT1	0	0.813	3105	HLA-A
Gp160	P27797	0.915	0.819	811	CALR
Gp160	P27824	0.861	0.667	821	CANX/CALX
IN	O75475	0	0.797	11168	PSIP1
MA	P49590	0.856	0.774	23438	HARS2/SYHM
Nef	O14734	0.956	0.832	10005	ACOT8
Nef	P30419	0.874	0.891	4836	NMT1
Tat	O60563	0.909	0.867	904	CCNT1
Tat	P50750	0.851	0.908	1025	CDK9
Vif	Q15369	0.956	0.972	6921	TCEB1/ELOC
Vif	Q15370	0.96	0.928	6923	TCEB2/ELOB
Vif	Q93034	0.855	0.889	8065	CUL5
Vpr	P13051	0	0.766	7374	UNG
Vpr	Q7L5N1	0	0.838	10980	COPS6/CSN6
Vpr	Q9Y4B6	0.958	0.884	9730	VPRBP
Vpu	Q9Y297	0.9	0.7	8945	BTRC/FBW1A

Supplementary Table 6: Overlap between AP-MS data and published RNAi screens[38-41].

RNAi screen	Prey	GeneID	GENE symbol	HEK293 score	Jurkat score	Bait	Description
KONIG	Gag	84991	RBM17	0.765	0	Q96I25	SPF45_Splicing factor 45
ZHOU	Gag	54555	DDX49	0.824	0	Q9Y6V7	DDX49_Probable ATP-dependent RNA helicase
BRASS	MA	51520	LARS	0.835	0.773	Q9P2J5	SYLC_Leucyl-tRNA synthetase
KONIG	MA	3735	KARS	0.883	0.731	Q15046	SYK_Lysyl-tRNA synthetase
YEUNG	MA	4141	MARS	0.908	0.723	P56192	SYMC_Methionyl-tRNA synthetase
YEUNG	MA	9255	SCYE1	0.919	0.74	Q12904	MCA1_Multisynthetase complex auxiliary component p43
KONIG	NC	6050	RNH	0.87	0	P13489	RINI_Ribonuclease inhibitor
YEUNG	NC	79159	NOL12	0.781	0	Q9UGY1	NOL12_Nucleolar protein 12
BRASS	Pol	8667	EIF3H	0.801	0.868	O15372	EIF3H_Eukaryotic translation initiation factor 3 subunit H
BRASS	Pol	10130	PDIA6	0.408	0.82	Q15084	PDIA6_Protein disulfide-isomerase A6 precursor
KONIG	Pol	29882	ANAPC2	0	0.799	Q9UJX6	APC2_Anaphase-promoting complex subunit 2
KONIG	Pol	26031	OSBPL3	0.826	0.843	Q9H4L5	OSBL3_Oxysterol-binding protein-related protein 3
YEUNG	Pol	23352	UBR4	0.841	0.831	Q5T4S7	UBR4_E3 ubiquitin-protein ligase UBR4
ZHOU	Pol	29882	ANAPC2	0	0.799	Q9UJX6	APC2_Anaphase-promoting complex subunit 2
BRASS	PR	23365	ARHGEF12	0.889	0	Q9NZN5	ARHGC_Rho guanine nucleotide exchange factor 12
BRASS	PR	8667	EIF3H	0.735	0.804	O15372	EIF3H_Eukaryotic translation initiation factor 3 subunit H
KONIG	PR	26031	OSBPL3	0.877	0.893	Q9H4L5	OSBL3_Oxysterol-binding protein-related protein 3
KONIG	PR	114880	OSBPL6	0.812	0	Q9BZF3	OSBL6_Oxysterol-binding protein-related protein 6
ZHOU	PR	23299	BICD2	0	0.771	Q8TD16	BICD2_Protein bicaudal D homolog 2
ZHOU	PR	5426	POLE	0.839	0.794	Q07864	DPOE1_DNA polymerase epsilon catalytic subunit A
BRASS	IN	55127	HEATR1	0.77	0	Q9H583	HEAT1_HEAT repeat-containing protein 1
KONIG	IN	5352	PLOD2	0.754	0	O00469	PLOD2_Procollagen-lysine
BRASS	Vif	10075	HUWE1	0.934	0.324	Q7Z6Z7	HUWE1_E3 ubiquitin-protein ligase HUWE1
KONIG	Vif	6921	TCEB1	0.956	0.972	Q15369	ELOC_Transcription elongation factor B polypeptide 1
YEUNG	Vif	10197	PSME3	0.842	0	P61289	PSME3_Proteasome activator complex subunit 3
ZHOU	Vif	7917	BAT3	0.782	0	P46379	BAT3_Large proline-rich protein BAT3
BRASS	Vpr	64083	GOLPH3	0.926	0.708	Q9H4A6	GOLP3_Golgi phosphoprotein 3
BRASS	Vpr	8487	SIP1	0	0.823	O14893	GEMI2_Survival of motor neuron protein-interacting protein 1
BRASS	Vpr	9730	VPRBP	0.958	0.884	Q9Y4B6	VPRBP_Protein VPRBP
YEUNG	Vpr	26057	ANKRD17	0.742	0.879	O75179	ANR17_Ankyrin repeat domain-containing protein 17
YEUNG	Vpr	79016	DDA1	0.818	0.909	Q9BW61	DDA1_DET1- and DDB1-associated protein 1
YEUNG	Vpr	10197	PSME3	0	0.776	P61289	PSME3_Proteasome activator complex subunit 3
ZHOU	Vpr	1778	DNCH1	0.899	0.818	Q14204	DYHC_Dynein heavy chain
BRASS	Vpu	23291	FBXW11	0.947	0.852	Q9UKB1	FBW1B_F-box/WD repeat-containing protein 11
BRASS	Vpu	5902	RANBP1	0.766	0.574	P43487	RANG_Ran-specific GTPase-activating protein
BRASS	Vpu	5903	RANBP2	0.751	0	P49792	RBP2_E3 SUMO-protein ligase RanBP2
KONIG	Vpu	5529	PPP2R5E	0.764	0	Q16537	2A5E_Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit epsilon isoform
KONIG	Vpu	10728	PTGES3	0	0.777	Q15185	TEBP_Prostaglandin E synthase 3
KONIG	Vpu	5903	RANBP2	0.751	0	P49792	RBP2_E3 SUMO-protein ligase RanBP2

KONIG	Vpu	10313	RTN3	0.785	0.863	O95197	RTN3_Reticulon-3
YEUNG	Vpu	4478	MSN	0	0.779	P26038	MOES_Moesin
ZHOU	Vpu	506	ATP5B	0.375	0.784	P06576	ATPB_ATP synthase subunit beta
ZHOU	Vpu	7514	XPO1	0.813	0.717	O14980	XPO1_Exportin-1
YEUNG	Rev	55234	SMU1	0.763	0	Q2TAY7	SMU1_Smu-1 suppressor of mec-8 and unc-52 protein homolog
ZHOU	Rev	57380	MRS2	0.771	0	Q9HD23	MRS2L_Magnesium transporter MRS2L
BRASS	Tat	904	CCNT1	0.909	0.867	O60563	CCNT1_Cyclin-T1
YEUNG	Tat	4298	MLLT1	0.759	0	Q03111	ENL_Protein ENL
BRASS	Nef	4836	NMT1	0.874	0.891	P30419	NMT1_Glycylpeptide N-tetradecanoyltransferase 1
BRASS	Gp160	55768	NGLY1	0	0.834	Q96IV0	NGLY1_Peptide-N(4)-(N-acetyl-beta-glucosaminyl)asparagine amidase
KONIG	Gp160	23193	GANAB	0.947	0.971	Q14697	GANAB_Neutral alpha-glucosidase AB precursor
KONIG	Gp160	10960	LMAN2	0	0.839	Q12907	LMAN2_Vesicular integral-membrane protein VIP36 precursor
KONIG	Gp160	5682	PSMA1	0.755	0.705	P25786	PSA1_Proteasome subunit alpha type-1
KONIG	Gp160	5694	PSMB6	0.754	0	P28072	PSB6_Proteasome subunit beta type-6 precursor
YEUNG	Gp160	55969	C20orf24	0	0.781	Q9BUV8	CT024_Uncharacterized protein C20orf24
ZHOU	Gp160	2923	PDIA3	0	0.862	P30101	PDIA3_Protein disulfide-isomerase A3 precursor
ZHOU	Gp160	6048	RNF5	0	0.781	Q99942	RNF5_E3 ubiquitin-protein ligase RNF5
KONIG	Gp120	51150	Cab45	0.856	0.861	Q9BRK5	CAB45_45 kDa calcium-binding protein precursor
KONIG	Gp120	23193	GANAB	0.909	0.732	Q14697	GANAB_Neutral alpha-glucosidase AB precursor
KONIG	Gp120	5694	PSMB6	0.754	0.15	P28072	PSB6_Proteasome subunit beta type-6 precursor
YEUNG	Gp120	23118	MAP3K7IP2	0	0.839	Q9NYJ8	TAB2_Mitogen-activated protein kinase kinase kinase 7-interacting protein 2
BRASS	Gp41	55768	NGLY1	0	0.756	Q96IV0	NGLY1_Peptide-N(4)-(N-acetyl-beta-glucosaminyl)asparagine amidase
BRASS	Gp41	60559	SPCS3	0	0.831	P61009	SPCS3_Signal peptidase complex subunit 3
BRASS	Gp41	1678	TIMM8A	0	0.861	O60220	TIM8A_Mitochondrial import inner membrane translocase subunit Tim8 A
KONIG	Gp41	10960	LMAN2	0.86	0	Q12907	LMAN2_Vesicular integral-membrane protein VIP36 precursor
ZHOU	Gp41	2923	PDIA3	0.806	0.441	P30101	PDIA3_Protein disulfide-isomerase A3 precursor

Supplementary Table 7: The N-terminal, but not the C-terminal cleavage product of eIF3d associates with the eIF3 complex. C-terminally 3xFLAG-tagged versions of the N-terminal eIF3d fragment (residues 1-114) and the C-terminal eIF3d fragment (residues 115-548) were expressed in 293T cells and analyzed by AP-MS. The number of peptides detected for each of the eIF3 subunits is shown.

	eIF3dNter-3xFLAG	**eIF3dCter-3xFLAG**
eIF3a	22	0
eIF3b	17	0
eIF3c	12	0
eIF3d	4	39
eIF3e	9	0
eIF3f	9	0
eIF3g	5	0
eIF3h	7	0
eIF3i	6	0
eIF3j	1	0
eIF3k	4	0
eIF3l	8	0
eIF3m	4	0

Supplementary Table 8: siRNA sequences used for knockdown of eIF3 subunits.

siRNA	Target sequence
eIF3c siRNA1	CCCGAGCAGTCTGCGGATGAA
eIF3c siRNA2	ACCGTGATTTCGAGTCCCATA
eIF3d siRNA1	CGCCTCTGTTGCGTACCGTTA
eIF3d siRNA2	TACCAGCGGAATCGAATGAGA
eIF3e siRNA1	ATGGAAGACCTTACACGGTTA
eIF3e siRNA2	CCCAAAGGTCGCGATAATATT
eIF3f siRNA1	CACAATGAGTCAGAAGATGAA
eIF3f siRNA2	AACGGCCGCATGAGCATCAAA
eIF3g siRNA1	CAAGGAGGTCATCAACGGAAA
eIF3g siRNA2	GAGGTCATCAACGGAAACATA
eIF3i siRNA1	AAGGACCCTATCGTCAATGTA
eIF3i siRNA2	CCGGACAGAACGTCCTGTCAA

Supplementary Table 9: Raw values of HIV infection experiment shown in Fig. 4e, f and Supplementary Fig. 14.

siRNA	pNL4-3		HIV-VSVg		Cytotoxicity	
	raw values	normalized values	raw values	normalized values	raw values	normalized values
eIF3c siRNA1	48113 (±8960)	1.238 (±0.23)	4863 (±727)	0.22 (±0.033)	10742 (±954)	0.451 (±0.04)
eIF3c siRNA2	44717 (±10789)	1.15 (±0.278)	8163 (±782)	0.37 (±0.035)	11925 (±827)	0.501 (±0.035)
eIF3d siRNA1	195206 (±2739)	5.022 (±0.07)	136645 (±18435)	6.19 (±0.835)	12838 (±387)	0.539 (±0.016)
eIF3d siRNA2	157134 (±4437)	4.042 (±0.114)	90139 (±12492)	4.084 (±0.566)	11067 (±633)	0.465 (±0.027)
eIF3e siRNA1	161254 (±9085)	4.148 (±0.234)	55521 (±2549)	2.515 (±0.115)	13360 (±423)	0.561 (±0.018)
eIF3e siRNA2	108706 (±6649)	2.796 (±0.171)	28148 (±4455)	1.275 (±0.202)	12099 (±1139)	0.508 (±0.048)
eIF3f siRNA1	110324 (±23209)	2.838 (±0.597)	39358 (±1755)	1.783 (±0.08)	10740 (±3529)	0.451 (±0.148)
eIF3f siRNA2	105832 (±22120)	2.722 (±0.569)	6838 (±1349)	0.31 (±0.061)	11198 (±563)	0.47 (±0.024)
eIF3g siRNA1	24609 (±4387)	0.633 (±0.113)	10675 (±1071)	0.484 (±0.049)	12850 (±573)	0.54 (±0.024)
eIF3g siRNA2	16695 (±2411)	0.429 (±0.062)	3075 (±805)	0.139 (±0.036)	11893 (±943)	0.499 (±0.04)
eIF3i siRNA1	12306 (±2842)	0.317 (±0.073)	6737 (±1732)	0.305 (±0.078)	12460 (±144)	0.523 (±0.006)
eIF3i siRNA2	16346 (±1881)	0.42 (±0.048)	10323 (±882)	0.468 (±0.04)	13305 (±1209)	0.559 (±0.051)
negative control	38874 (±5324)	1 (±0.137)	22074 (±6924)	1 (±0.314)	23811 (±1558)	1 (±0.065)

Supplementary Table 10: Raw values of early RT assay shown in Fig. 4g. The numbers under "siRNA" indicate the two siEIF3Ds; "NC" indicates negative control siRNA.

siRNA	Norm Copies	average	stdev
6	4.577399	3.83	0.59
6	3.052657		
6	3.778994		
6	4.218929		
6	3.53544		
9	2.374325	2.55	0.65
9	2.255004		
9	1.796872		
9	2.829773		
9	3.517763		
nc	1.090787	1.00	0.13
nc	1.071397		
nc	0.837817		

Supplementary Table 11: Raw values of late RT assay shown in Fig. 4g and infection assay (luciferase) shown in Supplementary Fig. 15.

Luciferase	Scrambleds	EIF3D #6		EIF3D #9	
	replicates	replicates	fold increase	replicates	fold increase
	607104	1591411	2.536909146	1088296	1.734880604
	688608	1645503	2.623138593	1066744	1.700524007
	632264	1580856	2.52008315	1033008	1.646744583
	547432		**fold inc avg**		**fold inc avg**
	458440		2.560043629		1.694049731
	551600		**stddev**		**stddev**
	797352		0.055285706		0.044423268
	658840				
	704088				
	avg				
	627303.111				

Q-PCR	Scrambleds				
	replicates PBGD	replicates late HIV RT	late HIV RT/ PBGD	avg	
	36320.52	64497.184	1.7757781	1.510186553	
	38289.848	60297.152	1.574755585		
	50780.51	68094.234	1.340952149		
	41289.53	72056.02	1.745140233		
	46553.52	65180.53	1.400120335		
	48671.31	86453.266	1.776267497		
	49132.74	73865.57	1.503387965		
	44492.844	61235.47	1.376299299		
	61173.71	67228.55	1.098977813		
	EIF3D #6				
	replicates PBGD	replicates late HIV RT	late HIV RT/ PBGD	fold increase	fold inc avg
	34122.664	103299.03	3.027285033	2.00457687	1.787082884
	35668.684	90380.4	2.533886588	1.677863297	**stddev**
	44730.207	113405.12	2.535313999	1.678808485	0.18835591
	EIF3D #9				
	replicates PBGD	replicates late HIV RT	late HIV RT/ PBGD	fold increase	fold inc avg
	40012.99	109560.805	2.738130917	1.813107733	1.655802951
	34630.348	97048.76	2.802419427	1.855677646	**stddev**
	41546.19	81478.88	1.961163707	1.298623474	0.310057952

Supplementary Table 12: siRNA target sequences and values for infection, toxicity and knockdown efficiency for siRNA knockdown experiments in Supplementary Figures 17-19.

		Figure S17 data		Figure S18 data		Figure S19 data	
		HIV-VSVg infection		Cell viability (TOX)		Knockdown (mRNA)	
ID	siRNA target sequence	AVG	stdev	AVG	stdev	AVG	stdev
A4-1	ACCCAATTAAGTCCTACTTTA	0.9056	0.0915	0.9145	0.0336	0.0812	0.0065
A4-2	CAGACTGAACATGCACATGAA	1.1263	0.1289	0.8473	0.0270	0.0833	0.0060
BIG2-1	CAGGGCATGCTGGGAACGTCA	0.8716	0.1624	0.9717	0.0366	0.1267	0.0166
BIG2-2	GTGGCGCTCGATGAAATTAAA	1.0430	0.0893	1.0986	0.0369	0.1786	0.0109
SIRT1-1	CAAGCGATGTTTGATATTGAA	1.0402	0.1089	0.8319	0.0397	0.1203	0.0061
SIRT1-2	CAGGATTATTGTATTTACGTT	0.8829	0.0430	1.0442	0.0201	0.1095	0.0163
CCNT1-1	ACCCAGACAATAGACTATCAA	0.1039	0.0403	0.8153	0.0026	0.3826	0.0132
CCNT1-2	TTGGAACATGTCATCAAGGTA	0.1542	0.0631	0.9039	0.0456	0.2243	0.0130
GANAB-1	CCGGGATGTGCATAACATCTA	0.1384	0.0253	0.9856	0.0086	0.2258	0.0060
GANAB-2	TACCATCTCAGCACAATGATA	0.4184	0.0630	0.9564	0.0119	0.0451	0.0040
CAB45-1	CCAGGAGGTCTTCCTAGGCAA	0.2332	0.0592	1.0146	0.0310	0.0787	0.0040
CAB45-2	CCGGAGGAAGCTGATGGTCAT	0.3466	0.1146	0.9454	0.0209	0.1002	0.0118
OSBL6-1	CACATTCTGAATGAATAAATA	0.1520	0.0499	0.9548	0.0281	0.1795	0.0086
OSBL6-2	CAGGTTGTCAGTGTAAATATT	0.1810	0.0349	0.9631	0.0114	0.2236	0.0267
NC		1		1		1	
CALR-1	CAGTATCTATGCCTATGATAA	2.1059	0.8247	0.5749	0.0760	0.1448	0.0075
CALR-2	TCCCGCTGGATCGAATCCAAA	2.6426	0.2495	1.1040	0.0637	0.4719	0.0311
DESP-1	CAGGGAGATCATGTGGATCAA	7.2869	0.7355	0.8727	0.0395	0.0545	0.0011
DESP-2	AAGGGCCACGGTATTCGCTTA	3.0210	0.7033	1.0581	0.0673	0.0703	0.0061
RIR2-1	GCGGGATTAAACAGTCCTTTA	3.0688	0.8099	0.8977	0.0465	0.3630	0.0289
RIR2-2	CGGGATTAAACAGTCCTTTAA	2.0400	0.2293	0.6954	0.0520	0.3153	0.0295
SKP1-1	AAGATGATGAGAACAAAGAAA	2.6008	0.0166	1.0804	0.0571	0.1494	0.0509
SKP1-2	TCGCAAGACCTTCAATATCAA	1.9888	0.5578	0.7968	0.0332	0.0331	0.0032
CUL1-1	CGCCGTGAATGTGACGAAGGA	8.4298	2.7198	0.8772	0.0115	0.0750	0.0115
CUL1-2	CGGGTTCGAGTACACCTCTAA	2.4034	0.0506	0.4794	0.0082	0.1549	0.0267
RBX2-1	CTCCGGGAGCTCAGGCTCCAA	4.9686	0.8925	1.0218	0.0328	0.2455	0.0098
RBX2-2	CTGTTCAATCATTGAGTGGTA	5.8401	1.4639	0.9281	0.0567	0.5489	0.0199
D19L3-1	CAGAGCGGTTTATCAGATATA	2.9412	0.6359	0.9092	0.0724	0.1373	0.0025
D19L3-2	CAGCGAATGAATATTTACCAA	2.3601	0.0481	0.9873	0.0715	0.1670	0.0572
HEAT1-1	CAACGTGGAGTTGATAGTTCA	6.9027	1.2725	0.7421	0.0712	0.1952	0.0149
HEAT1-2	CCCGATTAGGTGATGATAATA	2.4184	0.3580	0.7048	0.0279	0.3291	0.0242
ZNT7-1	TATGGGTATGTTAGAGCGGAA	2.6721	0.7506	1.0188	0.0366	0.1680	0.0144
ZNT7-2	CCAGCGCATATCATTTAGCAA	2.4927	0.9116	0.5710	0.0233	0.4092	0.0315
HUWE1-1	CCGGCTTTCACCAGTCGCTTA	2.5555	0.2536	0.8299	0.0622	0.1941	0.0201
HUWE1-2	AAGCAGCTTATGGAGATTAAA	2.0209	0.1098	0.6451	0.0675	0.1833	0.0089

Supplementary Table 13: Primers used for knockdown confirmation.

ID	sense (5' -> 3')	anisense (5' -> 3')
A4	CCTTCTCGTTCCTGACAAGTG	CCATCCTCTCCTGGTGTAAGA
BIG2	GGCGCTCGATGAAATTAAAGC	GGACTTGGACTGGCAAGCTA
CALR	GAGCCTGCCGTCTACTTCAAG	CCGTCTCCGTCCAGAAACT
CCNT1	TTCACACAGTTCCCTGGAAATTC	CTCCACTTTAGCTGCTAGAAACA
CUL1	AGCCATTGAAAAGTGTGGAGAA	GCGTCATTGTTGAATGCAGACA
D19L3	ATGATGTCCATCCGGCAAAGA	GCAAAGGGCAATGGTTCCAC
DESP	CCCTGTGATGCTTACCAGAAAA	AAAGGGCTCGCATTTGCTCTT
GANAB	TGGGGATTACCCTTGCTGTG	CTGTCGCTTGCAGAAAGAACT
HEAT1	CCATTGGATGTACTGGCCTGG	GTAAGGCGACAAGTGAATAAGGA
HUWE1	TGCCAGTGCTTGTAAGGAACT	TGGTGACAAATGTTATCTGGTCC
OSBL6	AAATAGGCCAAACCAATGTCCA	TCTGAATATCGAGTGGTGCCTT
RBX2	TGGAAGACGGAGAGGAAACCT	TGAGGGAGAACATCTTGTCGC
RIR2	CACGGAGCCGAAAACTAAAGC	TCTGCCTTCTTATACATCTGCCA
CAB45	GCCAACCACTCGTCCACTC	TGTTCACATCCACCTTGGAAAA
SIRT1	TAGCCTTGTCAGATAAGGAAGGA	TGTTCTGGGTATAGTTGCGAAGT
SKP1	GACCATGTTGGAAGATTTGGGA	TGCACCACTGAATGACCTTTT
ZNT7	GCTTAGGCTTGATTTCCGACT	CCAGAACTTCCGCTCTAACATAC

Supplementary Table 14: Values for integration assay using DESP and HEAT1 siRNAs

ID	Integration AVG provirus	Integration stdev provirus
NC	1.00400	0.11792
DESP-1	2.53089	0.58001
DESP-2	2.19970	0.31215
HEAT1-1	3.83570	0.45216
HEAT1-2	1.97540	0.74922

Supplementary Methods

Plasmids. Codon-optimized versions of the HIV-1 ORFs were kindly provided by Marilyn D. Resh, George Pavlakis, Stephan Bour, Klaus Strebel (obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH), as well as Ralf Wagner, Zeger Debyser, Michael Lenardo and Gary Nabel[42-47]. Codon-optimized version of PR, RT and Vif were newly synthesized (GenScript). For AP-MS all HIV-1 ORFs were subcloned into pcDNA4/TO (Invitrogen) carrying a 3' 2xStrepTagII-TEV-3xFLAG sequence. Four mutations in the original gp41 coding region were fixed by site-directed mutagenesis (T45I, H118Q, I188T and T210S). The catalytic sites in PR and Pol were inactivated by D25N exchange and an influenza HA signal peptide was added 5' to the gp41 coding sequence. For detailed information of plasmid sources see **Supplementary Table 1**. Rev exons were ligated together using a HindIII site present at the end of the first exon (and repeated on primers to amplify the second exon) to make a single message.

Coding regions of human genes were PCR amplified from HEK293 cDNA and cloned into pcDNA4/TO carrying a C-terminal 3xFLAG tag sequence. Identities of ORFs were confirmed by 5' and 3' end sequencing. FlagHA-EIF3F (Addgene plasmid 22546) was kindly provided by Wade Harper[36], FLAG-Bcl2 (Addgene plasmid 18003) by Clark Distelhorst[48], and FLAG-SIRT1 (Addgene plasmid 1791) by Michael Greenberg[49].

Amino acid sequences of HIV-SF proteins used for AP-MS. Underlined are the 2xStrepTagII-TEV-3xFLAG (SF) sequences and the Influenza HA signal peptide (SP). Residues in bold indicate aa exchanges introduced by site-directed mutagenesis.

>Gag-SF
MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQT
GSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAADTGHSNQVSQNYPIVQNIQG
QMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAA
EWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPT
SILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTAC
QGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKIVKCFNCGKEGHTARNCRAPRKKGCWKCGKEG
HQMKDCTERQANFLGKIWPSYKGRPGNFLQSRPEPTAPPEESFRSGVETTTPPQKQEPIDKELYPLTSLR
SLFGNDPSSQGAAAGWSHPQFEKGGGSGGGSGGGSWSHPQFEKGENLYFQGADYKDHDGD*

>MA-SF
MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQT
GSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAADTGHSNQVSQNYGAAAGWSH
PQFEKGGGSGGGSGGGSWSHPQFEKGENLYFQGADYKDHDGDYKDHDIDYKDDDDK*

>CA-SF
MPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQML
KETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLN
KIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAA
TLEEMMTACQGVGGPGHKARVLGAAAGWSHPQFEKGGGSGGGSGGGSWSHPQFEKGENLYFQGADYKDHD
GDYKDHDIDYKDDDDK*

>NC-SF
MQRGNFRNQRKIVKCFNCGKEGHTARNCRAPRKKGCWKCGKEGHQMKDCTERQANGAAAGWSHPQFEKGG
GSGGGSGGGSWSHPQFEKGENLYFQGADYKDHDGDYKDHDIDYKDDDDK*

>p6-SF
MLQSRPEPTAPPEESFRSGVETTTPPQKQEPIDKELYPLTSLRSLFGNDPSSQGAAAGWSHPQFEKGGGS
GGGSGGGSWSHPQFEKGENLYFQGADYKDHDGDYKDHDIDYKDDDDK*

>Pol(D25N)-SF
MPQITLWQRPLVTIKIGGQLKEALL**N**TGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGH
KAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTE
MEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLD
VGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFKKQNPDIVI
YQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKD
SWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELELAENREILKEPVHGVYYD
PSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFK
LPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETFYVDGAANRETKLGKAGYVTN
KGRQKVVPLTNTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDKSESELVNQIIEQLIKKEK
VYLAWVPAHKGIGGNEQVDKLVSAGIRKILFLDGIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASC
DKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPV
KTIHTDNGSNFTSATVKAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAV
FIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRNPLWKGPAKLLWKGEGAVV
IQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDEDLEGGGGWSHPQFEKGGGSGGGSGGGSWSHPQF
EKGENLYFQGADYKDHDGDYKDHDIDYKDDDDK*

PR(D25N)-SF
MPQITLWQRPLVTIKIGGQLKEALL**N**TGADDTVLEEMNLPGRWKPKMIGGIGGFIKVRQYDQILIEICGH
KAIGTVLVGPTPVNIIGRNLLTQIGCTLNFGAAAGWSHPQFEKGGGSGGGSGGGSWSHPQFEKGENLYFQ
GADYKDHDGDYKDHDIDYKDDDDK*

>RT-SF
MPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKKDST
KWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKQKKSVTVLDVGDAYFSVPLDKDFRKYTAFTIPSINNET
PGIRYQYNVLPQGWKGSPAIFQCSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHL
LRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYAGIKVRQL
CKLLRGTKALTEVVPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNL
KTGKYARMKGAHTNDVKQLTEAVQKIATESIVIWGKTPKFKLPIQKETWEAWWTEYWQATWIPEWEFVNT
PPLVKLWYQLEKEPIIGAETFYVDGAANRETKLGKAGYVTDRGRQKVVPLTDTTNQKTELQAIHLALQDS
GLEVNIVTDSQYALGIIQAQPDKSESELVSQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDGLVSAGIRKV
LGAAAGWSHPQFEKGGGSGGGSGGGSWSHPQFEKGENLYFQGADYKDHDGDYKDHDIDYKDDDDK*

>IN-SF
MAFLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHL
EGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTVHTDNGSNFTSTTVKAACWWAGIKQE
FGIPYNPQSQGVIESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDI
QTKELQKQITKIQNFRVYYRDSRDPVWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGD
DCVASRQDEDGAAAGWSHPQFEKGGGSGGGSGGGSWSHPQFEKGENLYFQGADYKDHDGDYKDHDIDYKD
DDDK*

>nVif-SF
MENRWQVMIVWQVDRMRINTWKRLVKHHMYISRKAKDWFYRHHYESTNPKISSEVHIPLGDAKLVITTYW
GLHTGERDWHLGQGVSIEWRKKRYSTQVDPDLADQLIHLHYFDCFSESAIRNTILGRIVSPRCEYQAGHN
KVGSLQYLALAALIKPKQIKPPLPSVRKLTEDRWNKPQKTKGHRGSHTMNGHLEGGGGWSHPQFEKGGGS
GGGSGGGSWSHPQFEKGENLYFQGADYKDHDGDYKDHDIDYKDDDDK*

>hVpr-SF
MEQAPEDQGPQREPYNEWTLELLEELKSEAVRHFPRIWLHNLGQHIYETYGDTWAGVEAIIRILQQLLFI
HFRIGCRHSRIGVTRQRRARNGASRSGAAAGWSHPQFEKGGGSGGGSGGGSWSHPQFEKGENLYFQGADY
KDHDGDYXDHDIDYKDDDDK*

>Tat-SF
MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKALGISYGRKKRRQRRRAHQNSQTHQASLS
KQPTSQPRGDPTGPKELEGGGGWSHPQFEKGGGSGGGSGGGSWSHPQFEKGENLYFQGADYKDHDGDYKD
HDIDYKDDDDK*

>Rev-SF
MAGRSGDSDEELIRTVRLIKLLYQSNPPPNPEGTRQARRNRRRRWRERQRQIHSISERILGTYLGRSAEP
VPLQLPPLERLTLDCNEDCGTSGTQGVGSPQILVESPTVLESGTKEGAAAGWSHPQFEKGGGSGGGSGGG
SWSHPQFEKGENLYFQGADYKDHDGDYKDHDIDYKDDDDK*

>Nef-SF
MGGKWSKSSVIGWPTVRERMRRAEPAADRVGAASRDLEKHGAITSSNTAATNAACAWLEAQEEEEVGFPV
TPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQRRQDILDLWIYHTQGYFPDWQNYTPGPGVRYPLTFG
WCYKLVPVEPDKVEEANKGENTSLLHPVSLHGMDDPEREVLEWRFDSRLAFHHVARELHPEYFKNCGAAA
GWSHPQFEKGGGSGGGSGGGSWSHPQFEKGENLYFQGADYKDHDGDYKDHDIDYKDDDDK*

>gp160-SF
MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPVWKEATTTLLCASDAKAYDTEVHNVWA
THACVPTDPNPQEVVLVNVTENFDMWKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNT
NSSSGRMIMEKGEIKNCSFNISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYSLTSCNTSVITQACPKVSF
EPIPNHYCAPAGFTILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSVNFTDN
AKTIIVQLNTSVEINCTRPNNNTRKRIRIQRGPGRAFVTIGKIGNMRQAHCNISRAKWNNTLKQIDSKLR
EQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTWSTEGSNNTEGSDTITLPCRI
KQIINMWQKVGKAMYAPPISGQIRCSSNITGLLLTRDGGNSNNESEIFRLGGGDMRDNWRSELYKYKVVK
IEPLGVAPTKAKRRVVQREKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIE
AQQHLLQLTVWGIKQLQARTLAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWNHTTWME
WDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKLFIMIVGGLVGLRIVFA
VLSIVNRVRQGYSPLSFQTHLPTPRGPDRPEGIEEEGGERDRDRSIRLVNGSLALIWDDLRSLCLFSYHR
LRDLLLIVTRIVELLGRRGWEALKYWWNLLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAI
RHIPRRIRQGLERILLGLEGGGGWSHPQFEKGGGSGGGSGGGSWSHPQFEKGENLYFQGADYKDHDGDYK
DHDIDYKDDDDK*

>gp120-SF
MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPVWKEATTTLLCASDAKAYDTEVHNVWA
THACVPTDPNPQEVVLVNVTENFDMWKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNT
NSSSGRMIMEKGEIKNCSFNISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYSLTSCNTSVITQACPKVSF
EPIPNHYCAPAGFTILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSVNFTDN
AKTIIVQLNTSVEINCTRPNNNTRKRIRIQRGPGRAFVTIGKIGNMRQAHCNISRAKWNNTLKQIDSKLR
EQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTWSTEGSNNTEGSDTITLPCRI
KQIINMWQKVGKAMYAPPISGQIRCSSNITGLLLTRDGGNSNNESEIFRLGGGDMRDNWRSELYKYKVVK
IEPLGVAPTKAKRRVVQREKRAVGIGALFLGFLGAAGSTMGAASMTGLEGGGGWSHPQFEKGGGSGGGSG
GGSWSHPQFEKGENLYFQGADYKDHDGDYKDHDIDYKDDDDK*

>SP-gp41-SF
MTMKTIIALSYIFCLVFAMLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYL
KDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWNHTTWMEWDREINNYTSLIHSLIEESQNQQEKNE
QELLELDKWASLWNWFNITNWLWYIKLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPTPRGP
DRPEGIEEEGGERDRDRSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWW
NLLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILLGLEGGGGWSHP
QFEKGGGSGGGSGGGSWSHPQFEKGENLYFQGADYKDHDGDYKDHDIDYKDDDDK*

Cell culture and affinity purification. HEK293 cells (ATCC# CRL-1573) and HeLa P4.R5 (acquired from Ned Landau) were maintained in DMEM high glucose/10% FCS plus Pen/Strep. For AP, $2.5x10^6$ cells were seeded in 15 cm plates and the next day transfected with 3-10 µg plasmid using calcium phosphate. 42 h after transfection cells were detached and washed with PBS. Jurkat TRex cells (Invitrogen) were cultured in RPMI/10%FBS plus Pen/Strep and 10 µg/ml Blasticidin. Stable Jurkat cell clones were generated by transfection with the linearized vector, selection with 300 µg/ml Zeocin followed by limiting dilution. For AP, $2.5x10^8$ cells were induced with 1 µg/ml doxycylcin for 16 h. In case of the Vif-SF and Vpr-SF clones, 0.5 µM MG132 (Calbiochem) was added 12 h before harvest. Cells were lysed in 1 ml cold lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5% Nonidet P40, complete protease inhibitor (Roche) and phosphostop (Roche)). Cells were dounced 20x on ice and spun at 2800xg for 20 min. The supernatant was incubated with 60 µl preclearing beads (mouse IgG agarose, SIGMA or Sepharose 4FF, GE Healthcare) for 2 h. The precleared lysate was incubated with 30 µl IP beads over night. FLAG APs were performed with anti-FLAG M2 Affinity Gel (SIGMA) and Strep APs with StrepTactin Sepharose (IBA). The beads were washed 5x with lysis buffer containing 0.05% Nonidet P40 followed by one wash with lysis buffer without detergent. Proteins were eluted with 40 µl 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA containing either 100 µg/ml 3xFLAG peptide (ELIM) and 0.05% RapiGest (Waters), or 2.5 mM

Desthiobiotin (IBA). 4 µl of the eluate was analyzed by 4-20% SDS PAGE (Biorad) and silver staining.

Western blotting. Cell lysates were subjected to SDS-PAGE and transferred to nitrocellulose membranes. Membranes were blocked with 5% milk/TBST and incubated with mouse (SIGMA F1804, 1/1000) or rabbit anti-FLAG (SIGMA F7425, 1/5000), Strep•Tag® II Antibody HRP Conjugate (Novagen 71591, 1/6000), or HIV-1 p24 Monoclonal Antibody (183-H12-5C, 1/50) provided by Dr. Bruce Chesebro and Kathy Wehrly through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH[50]. The following antibodies were used for detection of Gag, MA and NC interaction candidates:

MCA1	Santa-Cruz (sc-66879) 1/1000
MCA2	ProteinTech Group (10424-1-AP) 1/500
MCA3	Santa-Cruz (sc-68325) 1/1000
CSDE1	abcam (ab96124) 1/1000
CTNA3	ProteinTech Group (13974-1-AP) 1/1500
DDX49	Abgent (AP9963a) 1/1000
E2AK2	Santa-Cruz (sc-709) 1/100
NS1BP	ProteinTech Group (14741-1-AP) 1/1500
LRC47	Atlas Antibodies (HPA012018) 1/500
RM11	Cell Signaling Technology 1/1000
NH2L1	GeneTex (GTX109247) 1/1000
NOL12	ProteinTech Group (15426-1-AP) 1/1000
NOLC1	Santa-Cruz (sc-28672) 1/100
OLA1	abcam (ab51077) 1/1000
MYPT1	Millipore (07-672) 1/1000
ANM1	Bethyl Laboratories (A300-723A-1) 1/1500
SPF45	Bethyl Laboratories (A300-723A-1) 1/1500
RINI1	Santa-Cruz (sc-166683) 1/100
SDCB1	Novus Biologicals (NB100-53807) 1/1000
SDCG8	ProteinTech Group (13471-1-AP) 1/1500
SPCS	ProteinTech Group (11551-1-AP) 1/1000
SF3B4	GeneTex (GTX117942) 1/1000
STRAP	Santa-Cruz (sc-130671) 1/100
TR150	Bethyl Laboratories (A300-956A-1) 1/1500
THUM1	Novus Biologicals (H00055623-M01) 1/500
YTHD3	Aviva Systems Biology (ARP55530_P050) 1/1000

Immunoreactive bands were detected either by chemiluminescence (ECL kit; Amersham), or on a digital infrared scanner (LI-COR).

Sample Preparation for Mass spectrometry. For gel-free MS analysis 10 µl of the IP eluate were reduced with 2.5 mM DTT at 60°C for 30 minutes followed by alkylation with 2.5 mM iodoacetamide for 40 minutes at room temperature. 100 ng sequencing grade modified trypsin (Promega) was then added to the sample and incubated overnight at 37°C. The resulting peptides were concentrated on ZipTip C18 pipette tips (Millipore) and eluted in a final 20 ul solution of 0.1% formic acid. For gel-based analysis, 20 µl IP eluate was separated by 4-20% SDS-PAGE and stained with GelCode Blue (Thermo Scientific). Each lane was cut into 15 pieces. Each gel piece was diced into small (1 mm^2) pieces and washed 3x with 25 mM NH_4HCO_3/50% ACN. Gel pieces were dehydrated and incubated with 10 mM DTT in 25 mM NH_4HCO_3 and incubated for 1 hour at 56°C. The supernatant was removed and the gel pieces were incubated with 55 mM iodoacetamide and incubated for 40 minutes. Gel pieces were washed with 25 mM NH_4HCO_3, then 25 mM NH_4HCO_3/50% can and were then dehydrated. 10 ng/µl trypsin in 25 mM NH_4HCO_3 was then added to the gel pieces and incubated overnight at 37°C. Finally, peptides were extracted from the gel pieces with 50% ACN/5% formic acid and the solvent evaporated. The final peptide sample was resuspended in 20 µl 0.1% formic acid.

Mass Spectrometry. All samples were analyzed on a Thermo Scientific LTQ Orbitrap XL mass spectrometer equipped with a nanoACQUITY UPLC (Waters) chromatography system and a nanoelectrospray source. 5 µl of each sample was injected onto a nanoACQUITY Symmetry C18 trap (5 µm particle size, 180 µm x 20 mm) in buffer A (0.1% formic acid in water) at a flow rate of 4 µl/min and then separated over a nanoACQUITY BEH C18 analytical column (1.7 µm particle size, 100 µm x 100 mm) over one hour with a gradient from 2% to 25% buffer B (99.9% ACN/0.1% formic acid) at a flow rate of 0.4 µl/min. The mass spectrometer continuously collected data in a data-dependent manner, collecting a survey scan in the Orbitrap mass analyzer at 40,000 resolution with an automatic gain control (AGC) target of 1×10^6 followed by collision-induced dissociation (CID) MS/MS scans of the 10 most abundant ions in the survey scan in the ion trap with an AGC target of 5,000, a signal threshold of 1,000, a 2.0 Da isolation width, and 30 ms activation time at 35% normalized collision energy. Charge state screening was employed to reject unassigned or 1+ charge states. Dynamic exclusion was enabled to ignore masses for 30 s that had been previously selected for fragmentation. Raw mass spectrometric

data were converted into peaklists using Bioworks 3.3.1 SP1. The spectra were searched using Prospector v.5.3 (http://prospector.ucsf.edu.)[51] against a human-restricted UniProt database (downloaded October 2009) supplemented with HIV protein sequences from 40 strains. Trypsin was specified as the enzyme; one missed cleavage and zero non-specific cleavages at the peptide termini were permitted. Mass accuracy was set to 25 ppm for precursor ions and 0.8 Da for fragment ions. Carbamidomethylation of Cys residues was set as fixed modification, and acetylation of protein N-termini and Met oxidation as variable modifications. Protein Prospector results were filtered by applying a minimum Protein Score of 22.0, a minimum Peptide Score of 15.0, a maximum Protein E-Value of 0.01 and a maximum Peptide E-Value of 0.05.

Development of the MiST scoring system

Introduction

There have been many large-scale affinity tag/purification-mass spectrometry (AP-MS) studies that have targeted different organisms, including bacteria, yeast and human[36,52-56]. Several experimental techniques are available to discriminate between biologically relevant and irrelevant bait-prey pairs derived from this approach (e.g. isotopic labeling , highly repetitive MS experiments , and, of course, in-depth biological characterization). Unfortunately, many of these additional analyses are not readily applicable on large-scale. A variety of computational methods have also been developed to tackle this problem, including the PE[57], SAInt[37] and CompPASS[36] scoring systems. However, the utility of some of these methods to filter irrelevant bait-prey pairs from our HIV-human dataset is rather limited for several reasons. First, many known scoring systems rely on the relative bait and prey abundances and may fail due to the dependence of the abundance on prey in-cell concentration, concentration in isolated samples, affinity to the antibodies, as well as varying propensity of peptides to be detected by MS. Also, scoring algorithms do not always consider the variations of bait-prey pair detected (or variation in amount of detected bait-prey pair) in replicate experiments, making it difficult to filter out contaminants. Furthermore, scores that rely on specificity do not offer a possibility to exclude subset of pull-down experiments with baits that are expected to bind similar preys (e.g. HIV Gag and its processed protein, MA). Finally, development of some scoring depends on a quality of training sets containing true and false biologically relevant complexes, which may not exist.

Therefore, we decided to design a new scoring system that attempts to address many of the aforementioned problems, termed MiST (Mass Spectrometry Interaction Statistics), and apply it to our AP-MS derived HIV-human protein-protein interaction dataset. The MiST scoring system implements three features (abundance, reproducibility, and specificity) derived from the mass spectrometry analysis, and then uses an unsupervised machine learning technique to combine them into a single composite score that ranges from 0 to 1. MiST is available as a web-server at http://salilab.org/mist.

Data quality and integration

In each AP-MS experiment, a single tagged protein (i.e. bait) is used to isolate proteins interacting directly or indirectly with the bait (i.e. prey). Bait-prey pairs do not need to interact physically, nor are all physical interactions necessarily biologically relevant. Each AP-MS experiment defines the composition of potentially multiple complexes with different molecular architectures. Each purification for one given bait is quantified by an "experiment vector" that lists measured amounts of all potential preys *i*. The preys of each bait in our human-HIV dataset were determined in multiple pull-downs and MS runs (varying experimental protocols and personnel), resulting into 8-18 replica experiment vectors per bait (**Supplementary Table 2** numerates the experiments that were carried out during the study).

To assess the reproducibility, and thus ultimately the data quality, we calculated Pearson correlation coefficients of replica experiment vectors for each bait (**Supplementary Fig. 4**). Even though label-free (non-quantitative) mass spectrometry is well known for its irreproducibility, our analysis suggests that experiments were highly reproducible: the average correlation coefficient between the replica experiment vectors is 0.68. We note 4 patterns common to all purifications regardless of the bait used. First, the highest number of detected preys (protein intensity larger than 0) was found by the "gel-FLAG" protocol (i.e. FLAG-tagged bait sample isolation from a gel). Second, the smallest number of detected preys was found by the beads-FLAG protocol. Third, the strongest correlation between different protocols was observed between the elution-FLAG and gel-FLAG protocols. Fourth, the strongest correlation between replicas by the same experimental protocols was observed for the elution-FLAG protocol.

Using a set of 39 well-characterized HIV-human protein-protein interactions (**Supplementary Table 3**), we next carried out a comparison of the data (pre-MiST scored) derived from the different protocols focusing on reproducibility and abundance (**Supplementary Fig. 5**). We saw an overall concordance regardless of protocol used (e.g. Vif-ELOC and Vif-ELOB were reproducible and abundant across different protocols), however some differences were observed. For example, APOBEC3F, a well-characterized partner of Vif, was only observed in the gel-FLAG protocol and SIRT1, a well-studied Tat interactor, was only seen with elution-FLAG. Also, many of the membrane proteins identified were also almost exclusively detected in gel-based experiments (**Supplementary Fig. 5**). When integrating this data, one usually aims to exclude outlying experimental vectors (e.g. irreproducible replicas with Vpr and NC). However, we decided to include and integrate all the information, treating all data equally, rather then apply outlier-filtering for two reasons: i) label-free mass spectrometry experiments are usually not highly reproducible and having more replicas may increase confidence in individual preys that were repetitively detected despite un-correlated experiment vectors, and ii) different types of proteins respond the best to different experimental protocols, and integrating such experiment vectors increases the prey detection coverage.

Definition of input features and MiST score

The amount of prey *i* interacting with bait *b* was quantified using modified SI_N score[58] that is computed from a protein intensity $I_{b,i}$ (not spectral counts as in the original design), total protein intensities of *N* number of preys observed from a single pull-down experiment, $\sum_{i=1}^{N} I_{b,i}$, and the length (number of residues) of the identified prey, L_i, as follows:

$$SI_{N;b,i} = \frac{I_{b,i}}{L_i \cdot \sum_{i=1}^{N} I_{b,i}}$$

The quantity $Q_{b,i,r}$ of bait-prey pair *b,i* in a replica *r* is defined as SI_N score of *b,i* pair normalized by a sum of SI_N scores of all preys from a given pull-down experiment *r* as:

$$Q_{b,i,r} = \frac{SI_{N;b,i,r}}{\sum_{i=1}^{N} SI_{N;b,i,r}}$$

Next, the three features used to define the biological relevance score are calculated as follows. The first feature, **the abundance**, $A_{b,i}$, of a given bait-prey pair *i,b,* is defined as the mean of the bait-prey quantities $Q_{b,i,r}$ over all N_R number of replicas:

$$A_{b,i} = \frac{\sum_{r=1}^{N_R} Q_{b,i,r}}{N_R}$$

The second feature, **the reproducibility**, $R_{b,i}$, of a given bait-prey pair *b,i*, is defined as the normalized entropy of the vector $Q_{b,i}$:

$$R_{b,i} = \frac{\sum_{r=1}^{N_R} Q_{b,i,r} \cdot \log(Q_{b,i,r})}{\log_2(N_R)^{-1}}$$

The third feature, **the specificity**, $S_{b,i}$, of a given bait-prey pair *b,i*, is defined as the proportion of the abundance of prey *i* compared to the abundances of prey *i* for the other N_B number of baits:

$$S_{b,i} = \frac{A_{b,i}}{\sum_{b=1}^{N_B} A_{b,i}}$$

Optionally, MiST can exclude consideration of specificity for baits that are expected to bind similar preys (based on either manual annotation or clustering of pull-downs). The three features were combined into a single composite score (the MiST score) by maximizing the variance in the three features space using the standard principal component analysis (PCA), as implemented in the MDP toolkit[59].

Benchmark and assessment of MiST

To evaluate the accuracy of MiST and compare it with other scoring systems, we used a manually curated list of well-characterized 39 protein pairs (**Supplementary Table 3**). The accuracy of the MiST score was evaluated using this dataset as well as the number of bait-prey pairs that involve ribosomal proteins, an accurate indicator of irrelevant interactions[60,61]. The accuracy of MiST was also compared to those of previously published scoring methods, SAInt and CompPASS , as well as to protein intensities and SI_N scores alone. SAInt was obtained from http://www.nesvilab.org/software.html and used as described in a vignette accompanying the software. For the CompPASS method, we used its D^R scores computed by our implementation of the software based on the description in the original paper.

Using the data derived from HEK293 cells, at a threshold of 0.75, corresponding to 387 protein-protein interactions (PPIs), the recall number of known bait-prey pairs for the SAInt, CompPASS, and MiST scores was 19, 29, and 32, respectively (**Supplementary Fig. 6a,** black line). At the same threshold, the recall number of bait-prey pairs involving ribosomal proteins for the MiST score was only 3, compared to 32 and 75 for SAInt and CompPASS, respectively (**Supplementary Fig. 6b,** black line). In both cases, MiST outperforms CompPASS and SAInt. We also evaluated the performance of MiST, CompPASS, and SAInt by plotting the ROC curve (**Supplementary Fig. 6c**), where true-positive and false-positive rates were assessed using 39 well-characterized human-HIV pairs, and 1596 human-HIV pairs from our dataset involving ribosomal proteins, respectively. Again, MiST performs the best and we have highlighted where on this graph the cut-off of 0.75 falls, which is very close to where the line plateaus. This is in agreement with the threshold defined using the enrichment of the distribution of the MiST scores compared to scores obtained by reshuffling simulation (**Fig. 1d**). The MiST score trained on AP-MS data derived from HEK293 cells was also used to score AP-MS data derived from Jurkat T cells, and therefore, this cut-off was used for both datasets.

Analysis of different experimental conditions and number of replicas

Data derived from the first large-scale, AP-MS mapping of the HIV-human interactome provided us with an opportunity to reassess the importance of the different protocols and number of replicates. First, we determined the variability of different combination of the affinity tag on

the bait and sample isolation procedure using a resampling approach. To this end, for each bait with a given combination of the affinity tag and isolation procedure, three randomly selected replicas obtained by the same experimental protocol were removed from the complete dataset and MiST scores were then re-calculated for each triplicate. This was carried out twenty times, each time randomly selecting a different set of three experiments using the same protocol. The data for all HIV proteins derived from each protocol was combined and analyzed for overlap of MiST scores derived using the entire dataset (**Supplementary Fig. 7a**). With the exception of the beads-FLAG and gel-FLAG/Strep protocol, different protocols recovered ~50% of HIV-human pairs from our final set of HEK293-derived 387 protein-protein interaction pairs (**Supplementary Fig. 7a**), similarly to the results of resampling from all protocols at once (see the right-most box in **Supplementary Fig. 7a**), suggesting a high degree or reproducibility within the experiments. This resampling is slightly biased because not all protocols have three replicas. The gel-FLAG, gel-Strep and bead-FLAG protocols were used less than three times, while the number of the elution-Strep and bead-Strep replicas was exactly three for all baits. Other experimental protocols were repeated more than three times. The elution and elution-FLAG protocols were the only two protocols that, on average, identified more than 193 (50%) of the final set among their 387 top scoring bait-prey pairs. However, the high overlap for the elution-FLAG protocol may be explained by the fact that the majority of experiments were carried out using this protocol. To avoid this bias, we also counted the number of the 39 well-characterized HIV-human PPIs that had MiST scores higher than that of the 387th best-scoring bait-prey pair (**Supplementary Fig. 7b**). We found that the trends we observed in this analysis correlated strongly with the data represented in **Supplementary Fig. 7a**. The elution-FLAG protocol is the combination of an individual bait tag and sample isolation procedure that outperforms others in both metrics, suggesting that it may be the best single protocol to use. However, as already discussed, detection of different types of proteins may be dependent on the use of specific experimental protocols (**Supplementary Fig. 5**).

Next, we examined how the number of replicate experiments affects the number of bait-prey pairs that we find in common with the final set of HEK293-derived PPIs. For each bait, regardless of affinity tag and isolation procedure, a number of replicas obtained by any experimental protocol ranging from 2 to 15 replicates was chosen twenty times randomly, followed by recalculation of the MiST scores. After only three replicas, approximately 50% of

the final set overlaps with an average simulated set (**Supplementary Fig. 7c**). For seven replica experiments, the overlap increases to almost 75%, a point where the number of 39 well-characterized HIV-human PPIs having a MiST score higher than that of the 387th best scoring bait-prey pair levels off (**Supplementary Fig. 7d**). These results suggest that known bait-prey pairs are relatively easy to detect, presumably because the corresponding human proteins are highly abundant and/or bind to the HIV proteins with a higher affinity. In contrast, detection of PPIs involving less abundant and/or weakly interacting human proteins may require up to 10 replicas. In conclusion, if resources are limited, we suggest that one use the elution-FLAG protocol in triplicate for a given set of proteins in conjunction with the MiST scoring system.

Performance of MiST on other large-scale datasets

MiST was initially designed to score host-pathogen datasets, however, to show that it can be readily used for other datasets of varying sizes, interaction densities, and produced by different experimental protocols, we also benchmarked it against the DUB (human deubiquitinating enzymes)[36] and TIP49 (a set of key proteins involved in chromatin remodeling)[62] datasets, both of which were previously analyzed by CompPASS and SAInt . Using a recall plot, a set of high-confidence set of PPIs and a cut-off that was used in a previous comparative analysis[37], MiST was very similar when compared to SAInt and CompPASS in analyzing the DUB dataset (**Supplementary Fig. 8a**, see black line where the threshold was previously defined). Benchmarking all three scoring systems against the TIP49 dataset, we found that the recall rate of MiST was second best after SAInt (**Supplementary Fig. 8b**). These results are noteworthy because the reproducibility component of the MiST scoring systems was designed to work optimally with at least three replicas, which was not the case for DUB (2 replicas per bait) and TIP49 (35 pull-downs with 27 different baits) datasets.

Although we designed MiST to score host-pathogen interactions, and are, in fact using it to analyze other AP-MS derived host-pathogen PPI maps, based on the benchmarks presented here (**Supplementary Fig. 8a, b)**, we feel that it can be successfully applied to a wide variety of different datasets and systems.

Gene expression analysis. To assess whether the cell specificity of HIV-host interactions can be explained by the expression of the host protein in that particular cell type, we analyzed the gene expression profiles of six samples of HEK293 cells, four of which were from Gene Omnibus dataset GDS2426[63] and two samples from NCI60 HEK293 cell lines of BioGPS database[64]. For Jurkat analysis, we used a total of four samples, three of which were replicates from Gene Omnibus dataset GDS2164[65] and one sample from NCI60 BioGPS database. In GDS2426 dataset, we analyzed the expression of HEK293 cells transfected with control plasmid (pcDNA1/Neo; Invitrogen), and in GDS2164 data we analyzed the expression of control Jurkat cells which were untransfected and treated with 10 μM pronasterone.

To assign a z-score to each gene, we performed the following steps:

(1) Determine the average (*avg*), and standard deviation (*std*) of gene expression across all transcripts in a given sample.

(2) Calculate the *z*-score of each transcript *i* by $z_i = (expression_i - avg) / std$.

(3) Calculate the average of *z*-scores over all samples for a given cell type.

(4) Assign the largest z-score across transcripts to its corresponding gene for each cell type.

(5) Plot the difference in *z*-scores of HEK293 and Jurkat cells for proteins appearing as preys in HEK293 and Jurkat specific interactions.

We plotted the distribution of the *z*-score differences ($z_{JURKAT} - z_{HEK293}$) for proteins in HEK293 or Jurkat specific interactions in **Supplementary Fig. 9**.

Functional annotation. We used the Database for Annotation, Visualization and Integrated Discovery (DAVID)[66] to map the Gene Ontology terms for biological functions of proteins that were pulled down with each HIV protein. We used Gene Ontology terms to determine the enriched functional categories. We collapsed the GO terms into 13 biological functions *via* manual curation (**Supplementary Data 4**) to generate a high level overview of the biological role of each HIV protein. The enriched terms were represented in a heatmap by the –log of their

p-values using R. The HIV proteins were hierarchically clustered using R based on their profile of enriched biological terms.

Domain over-representation. For each HIV protein, we listed the frequencies of all domains present in the human proteins that were pulled-down with the viral protein (repeating domains within the same human protein were counted once). To label the domains, we used the Pfam classification for domain clans, when possible[67]. Statistical significance for domain enrichment was obtained by a hypergeometric test, where background domain distribution was based on domain frequency in the human proteome (all human proteins documented in SwissProt[68]). To focus on the most significant results, we used only domains for which the enrichment has a *p*-value below 0.005 (**Supplementary Fig. 12**). We used additional restriction for the data in **Fig. 1g**. Here, we show only cases where the domain appears more than 4 times in the human proteome, and the domain has more than two occurrences in the human proteins that were pulled-down with the HIV proteins.

Comparison to Other Datasets. The statistical significance of the overlap between our final set of HIV-human interactions (MiST>0.75) and those deposited in VirusMint was calculated using hypergeometric test implemented in R. The number of background interactions was defined as the number of unique human proteins detected in our AP-MS experiments (n = 4,219) multiplied by the number of HIV proteins (n = 15). Similarly, we calculated the statistical significance of the overlap between human factors detected in our AP-MS experiments and those detected in genome-wide RNAi screens. In this case, the number of background number of human factors was defined as the total number of human genes (n = 21,121).

Evolutionary analysis. Human proteins identified in this study were mapped to human genome build hg18. Genome-wide alignments to rhesus macaque (rheMac2) were downloaded from the UCSC genome browser (http://genome.ucsc.edu/). Synonymous and non-synonymous substitutions between the two species were identified using custom Perl scripts, with a subset of

substitutions visually inspected in the genome browser. Evolutionary rates for each group of genes considered were measured using the synonymous and non-synonymous rates of evolution (dS and dN, respectively). These patterns were compared to the rest of the genome by 10,000 bootstrap simulations matching the number of genes in a particular study/cell type (constraining each bootstrap replicate to match the distribution of tissue specificity in the observed set of genes from the geneAtlas2 data[69]). In turn, other characteristics of the data were also matched (number of tissues each gene is expressed in, maximal expression values of each gene, mean expression value of each gene across tissues, and GC-content). Each category produced a similar result except for GC-content, which poorly captured the variation in the rates of evolution.

Protease assay. A synthetic HIV-1^{NL4-3} protease gene containing the amino-terminal p6* region was cloned into pMAL-c2x (NEB) and expressed in BL21 (Gold) DE3 cells (Stratagene).Cells were treated for 10 minutes with 100 μM Saquinavir (Roche) prior to induction with 1 mM IPTG for 2 hours. Cells were lysed in 20 mM Tris, 100 mM NaCl, 1 mM DTT, 1 mM EDTA pH 8 buffer and MBP-tagged protein was purified on a MBP-TRAP column (GE Healthcare). Digestion with Factor Xa (NEB) separated the protease from MBP. Saquinavir was removed by addition of 5 M Urea and dialysis into activity buffer (MES pH 6.5)[70]. Active protease was purified on an S75 Superdex (GE Healthcare) and concentrated to 1.2 μM in activity buffer. 2.3 mg/ml eIF3 was digested for 4.5 hours by HIV-1 protease (0.7 μM) in 60 mM MES, 120 mM NaCl pH 5.5. The assay was quenched by addition of LDS-PAGE buffer (Invitrogen) and the complex was size separated on a 4-12% Bis-Tris Gel and silver stained (Pierce). N-terminal sequencing of the eIF3 cleavage product was conducted by the U.C. Davis Molecular Structure facility. The HIV-1 protease substrate specificity consensus sequence was generated from HEK293 peptide proteolysis data published by Schilling and Overall 2008[71] and corrected for natural amino acid abundance.

Infection assays. Individual siRNAs were transfected into HeLa P4.R5 reporter cells using a reverse transfection protocol[72]. The CD4 and CCR5 expressing HeLa P4.R5 cells are a derivative of clone P4[73] and were obtained from Ned Landau. 45 nl RNAiMAX (Invitrogen) diluted in 10

µl OptiMEM (Invitrogen) were added to each well of a 384-well plate containing 0.125 pmol of siRNA in 10 µl of OptiMEM. After 20 min incubation, 1000 HeLa P4.R5 cells in 20 µl DMEM supplemented with 10% fetal bovine serum (FBS, HyClone) were added to each well. After 48 h incubation at 37°C and 5% CO_2, 10 µl virus supernatant containing pNL4-3 or a pNL4-3 derived VSV-G pseudotyped single-cycle reporter virus (pNL43-Luc-E-R+, HIV-1 wild-type Δenv, encoding firefly luciferase GL3) encoding luciferase[74] was added to each well for 24 h. The proviral plasmid pNL4–3 was obtained from Ned Landau. Propagation of pNL43-Luc-E-R+ was described previously described[39]. To quantify infection levels 20 µl detection reagent, BetaGlo (Promega) for pNL4-3 or BriteLite Plus (PerkinElmer) for HIV-VSVg, were added to each well and luminescence of the samples was determined. Cell viability was analyzed by adding 20 µl CellTiter-Glo (Promega) to mock-infected samples 72 h after transfection and subsequently measuring luminescence. In each assay a panel of 40 scrambled siRNAs, specifically designed not to target any gene transcripts, were tested alongside the sample siRNAs. The 10 highest and lowest scoring scrambled siRNAs were ignored to exclude potential off-target effects and the average of the remaining scrambled siRNAs was used as negative control to normalize sample values.

Quantitative RT-PCR. HeLa P4.R5 cells were transfected as described above. 48 hours after transfection the cellular RNA was isolated using the RNeasy 96 Kit (Qiagen). The QuantiTect Reverse Transcription Kit (Qiagen) was used for cDNA synthesis in accordance with the manufacturer's protocol. Real-time PCR was performed using SYBR Green PCR Master Mix (Applied Biosystems) and a standard thermal cycler protocol (50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s and 60°C for 1 min). Results were analyzed using the comparative Ct method[75]. GAPDH served as endogenous control gene and was used to calculate relative expression levels. Individual mRNA expression levels were normalized using the average of two scrambled negative control siRNAs with the following target sequences: 5'-GGTAATTGCGCGTGCAACT-3' and 5'-GGCCGTATCGTAATACTTC-3'.

The following primer pairs were used to determine knock-down levels:

Gene	Sense	Antisense
EIF3C	5'-TGAAGATTCGTGATGTCACCAAG-3'	5'-AGATAGTCCTCTAGGTCAGCCA-3'
EIF3D	5'-ATCCTGCCTAAGAGTGCCAAA-3'	5'-GGGGTTTCTGTGATTTCTGATCC-3'
EIF3E	5'-GAAAACGTCGGCAGGTTCTAA-3'	5'-CAAGCACTGATTCACATTCCCT-3'
EIF3F	5'-AGTGCCGCACAATGAGTCAG-3'	5'-TGTCATGGCCCGTAGCGTA-3'
EIF3G	5'-CGCTGCCCCTACAAGGATAC-3'	5'-CGGCACATACTTCCCTGTCT-3'
EIF3I	5'-GGCCATGAGCGGTCCATTAC-3'	5'-TCTCACCATTCACAGAGTACCAT-3'
GAPDH	5'-GAAGATGGTGATGGGATTTC-3'	5'-GAAGGTGAAGGTCGGAGTC-3'

Early RT assay. Cells (293T; $1.7x10^4$ cells per well, 48 well plate) were treated with siRNA targeting eIF3d or controls (2.5 nM), and after 48 hours infected with HIV NL4-3-luc pseudotyped with VSVG (11 ng). After another 24 hours, total genomic DNA was isolated using the DNAeasy kit (Qiagen). DNA yields for cells knocked down for eIF3d were within 74% of wild type. Products of early reverse transcription were quantified using PCR primers and a Taqman probe that annealed to the HIV R-U5 region[76]. Each sample (10 ng DNA) was tested in 5 biological replicates, each of which was quantified using three separate Q-PCR assays. Sequences of siRNAs and DNA oligonucleotides used are below.

DNA oligonucleotides (written 5' to 3')

HIV LTR (R) Forward:	GCC TCA ATA AAG CTT GCC TTG A
HIV LTR (U5) Reverse:	TCC ACA CTG ACT AAA AGG GTC TGA
LTR Probe:	FAM-GCG AGT GCC CGT CTG TTG TGT GAC TCT GGT AAC TAG CTC GC-Dabcyl

Ribonucleotides used as siRNAs (written 5' to 3')

Hs_EIF3S7_6	CGCCTCTGTTGCGTACCGTTA
Hs_EIF3S7_9	TACCAGCGGAATCGAATGAGA
NC (Qaigen SI03650325)	AATTCTCCGAACGTGTCACGT

Infectivity and late RT assay. In a 96-well plate, siRNAs (Qiagen) were reverse transfected with Lipofectamine RNAiMAX reagent (Invitrogen) in 293T cells at 3000 cells in 100ul per well. At 48 hours post transfection, the cells were infected with 10ul per well of HIV-VSVg reporter virus encoding luciferase. After 24 hours of infection, cells were either treated with Britelite Plus reagent (Perkin Elmer) at a 1:1 ratio to measure luciferase activity by use of a luminometer (Topcount NXT; Packard Instruments), or lysed for analysis by Q-PCR. Cells prepared for Q-PCR were washed with PBS and then resuspended in 100ul per well of lysis buffer (10uM tris-HCl pH8, 1uM EDTA, 0.2uM $CaCl_2$, 0.001% triton X-100, 0.001% SDS, 1mg/ml proteinase K). Cells in lysis buffer were transferred to a MicroAmp Optical 96-well PCR plate (Applied Biosystems) and set in a PCR machine to incubate at 58˚C for 1 hour followed by heat inactivation at 95˚C for 10 minutes. Late HIV RT products were quantified by PCR of the cell lysates and normalized to cell copy number by measuring porphobilinogen deaminase (PBGD). A standard curve was generated for each Q-PCR experiment and samples for all experiments were done in triplicate. The following probes and primers were used:

Probes	**Name**	**5' label**	**3' label**	**Sequence 5' -> 3'**	**Company**
late HIV RT	LRT-P	FAM	TAMRA	CAGTGGCGCCCGAACAGGGA	Applied Biosystems
PBGD	PBGD-VIC	VIC	MGBNFQ	CCGGCAGATTGGAGAGAAAAGCCTGT	Applied Biosystems
Primers					
late HIV RT				TGTGTGCCCGTCTGTTGTGT	Invitrogen
late HIV RT				AACTAGGGAACCCACTGCTTAAG	Invitrogen
PBGD				AAGGGATTCACTCAGGCTCTTTC	Invitrogen
PBGD				GGCATGTTCAAGCTCCTTGG	Invitrogen

Supplementary Discussion:

HIV-human protein-protein interactions identified in this study

In the following text, the major findings of the AP-MS study presented herein are discussed in light of previously reported interactions, with a focus on unbiased studies like yeast-two hybrid screens and AP-MS analyses.

Gag

The HIV-1 Gag protein forms the protein shell of the viral particle. Upon maturation of the virus particle, the Gag polyprotein is cleaved by the viral protease into its subunits MA, CA, NC, p6, as well as two small spacer peptides SP1 and SP2 (not analyzed). The NC domain contains two zinc finger motifs that mediates packaging of the viral RNA into the virus particle. In the absence of HIV RNA, NC binds cellular RNAs[78]. Consistent with this, we found Gag bound to a large number of RNA-associated proteins like RNA helicases, ribosomal proteins, and splicing factors. However, several of these were found associated with other nucleic acid binding proteins, including Tat and Rev, and therefore obtained lower MiST scores. Among these are the RNA helicases DHX9, DDX18, DDX24, and DDX21 which were identified in a previous affinity purification analysis of HIV-1 Gag[79]. One exception is the RNA helicase DDX49, which we found exclusively detected in Gag pulldowns from HEK293 cells, but not Jurkat cells. Two proteins associated both with Gag and its subunit MA: i) OLA1, a poorly characterized ATPase, and ii) the selenocysteine synthase SPCS which belongs to the tRNA synthase family strongly associated with MA.

MA

A surprising finding was that MA strongly associates with the multi-tRNA synthetase complex. HIV utilizes tRNA(Lys) as a primer to initiate the reverse transcription of its genome. Packaging of tRNA(Lys) is known to be mediated via the interaction between lysyl-tRNA synthase (SYK) and Gag. Consistent with our finding, Histidyl-tRNA synthase 2 (SYHM) was identified in a yeast-two hybrid screen to interact with HIV-1 MA[80], however subsequent studies mapped the binding site of lysyl-tRNA synthase to the CA domain of Gag[81,82]. Furthermore, it

has been shown that only SYK is specifically incorporated into the virion whereas the other components of the complex are absent[83], suggesting that SYK may interact with Gag before entering the aminoacyl-tRNA synthetase complex[84]. The relationship between binding of the whole RNA synthetase complex by MA and SYK/tRNA(Lys) packaging therefore warrants further investigation. One possible scenario would be that during assembly, Gag binds SYK specifically through its CA domain, and during virus maturation SYK is sequestered by MA so that CA is liberated to form the viral core and the tRNA(Lys) can be transferred to the primer binding site (PBS) within the HIV RNA.

CA and p6

We found no significant host interactions when we purified affinity tagged versions of CA and p6 in either cell type. The most well-characterized CA binding protein is Cyclophilin A, initially identified in a yeast-two hybrid screen with HIV-1 Gag[85]. We did, in fact, find Cyclophilin A co-purifying with Gag and CA from Jurkat cells, however it was also detected with most of the other HIV proteins, which resulted in a low MiST score. The C-terminal p6 domain of Gag mediates budding of the virus particle by recruiting the ESCRT complex[86], interactions we did not detect. However, these interactions (e.g. with the ESCRT components TSG101[87] and ALIX[88]) may not be stable enough to be detected using our AP-MS approach.

RT

In contrast to most of the other HIV proteins, only a few host proteins have been previously reported to interact with Reverse Transcriptase (RT). Two proteins, AKAP1 and ELAV1, were identified in a yeast-two hybrid screen[89,90], however the validity of the latter was recently questioned[91]. In a phage display screen, beta actin (ACTB) was identified as RT interactor[92]. Both ELAV1 and ACTB were very promiscuous binders in our experiments and therefore considered background, while AKAP1 was not detected at all. To our knowledge, no AP-MS study of HIV-1 RT has been performed so far.

The highest scoring host interactor of RT in our screen was growth factor receptor-bound protein 2 (GRB2), which was detected with high specificity from HEK293 and Jurkat cells. GRB2 is an adaptor protein of various receptors and important for signal transduction leading to activation of the MAP kinase cascade, as well as actin polymerisation through the Arp2/3-N-

WASP pathway[93]. Importantly, phosphorylation of reverse transcription components by MAPK was previously shown to be critical for an early step in HIV-1 infection[94]. Furthermore, several studies have suggested that, upon entry, retroviruses use actin filaments for short-range transport in the cell periphery[95-97]. For example, Bukrinskaya *et al.*[97] showed that viral reverse transcription complexes associate rapidly with the host cell cytoskeleton and suggested that actin microfilaments of the host cell cytoskeleton serve as major sites for reverse transcription in the infected cell. Another group suggested that after membrane fusion, viral gene products might initiate activation of the Arp2/3 complex-dependent actin polymerization behind the viral core to cross the cortical layer[95].

Interestingly, another RT-interactor identified in our study is NHERF, which actually has a similar function to GRB2. It also is a scaffold protein which binds both membrane proteins as well as members of the ezrin/moesin/radixin (ERM) family of membrane-actin cytoskeletal linker proteins, thereby providing a link between the plasma membrane and the cortical actin cytoskeleton[98]. Of note, two members of the ERM family, moesin and ezrin, were shown to block HIV-1 infection before initiation of reverse transcription[99,100]. The authors suggest a model that after fusion, moesin and ezrin may mediate the transition of viral cores from the cortical actin to the microtubule network, allowing reverse transcription and dynein-mediated long-range transport toward the nuclear periphery. Hence a possible scenario is that RT, after fusion, induces actin polymerization and/or phosphorylation of the reverse transcription complex (RTC) to promote activation or transport of the RTC through recruitment of GRB2 and/or NHERF.

IN

HIV integrase (IN) inserts the viral cDNA into the host chromosomal DNA. An important cofactor of this process is the chromatin-associated cellular protein LEDGF/PSIP1 which directly binds IN, directing it to sites of integration[101,102]. Further cellular cofactors of HIV-1 integration reported are BAF, identified in an in vitro reconstitution assay of salt stripped PIC (pre-integration complex) activity[103], and INI1/SNF5, identified in a yeast-two hybrid screen of HIV-1 IN[104]. In our study, LEDGF/PSIP1 was the the only protein specifically associated with IN in Jurkat T cells. Of note, single BAF and INI1/SNF5 peptides were actually detected in IN purifications, but the reproducibility was too low for a significant MiST score. The fact that LEDGF/PSIP1 could not be detected in HEK293 cells using the standard protocol indicates IN

may require a specialized chromatin extraction step for purification. In fact, the LEDGF signal was considerably stronger in IN purifications from nuclear extracts (not shown).

Vif

Vif is an HIV-1 accessory protein that counteracts the antiviral cytidine deaminase APOBEC3G/F which is packaged into the virion by binding to the NC domain of Gag and deaminates the viral DNA during reverse transcription, resulting in viral inactivation. Vif acts as an adaptor between the APOBEC3 and the CUL5-ELOB/C ubiquitin ligase complex, leading to proteasomal degradation of APOBEC3. Consistent with this well-established Vif function, the highest scoring interactors we detected for Vif were the components of the E3 Ub ligase CUL5, ELOB, ELOC, RBX2. Of note, besides CUL5, we also detected CUL2 associated strongly with Vif, especially in Jurkat cells. This is consistent with RBX1, the adaptor protein for Cul2, being found in previous Vif AP-MS studies[105]. Surprisingly, Vif also associated in both cell types with the transcription cofactor PEBB/CBFβ (for a more detailed analysis of this interaction, see Jäger *et al.*, accompanying manuscript).

Another interesting Vif associated complex detected in Jurkat cells is the transcriptional corepressor complex NCOR1/HDAC3/GPS2/TBL1R. It has been reported that histone deacetylase (HDAC) inhibitors are capable of inducing expression of quiescent HIV-1 in latently infected cells, suggesting that silencing of the integrated HIV provirus is regulated by chromatin remodeling[106]. Recently ChIP analysis revealed that HDAC3 binds to the HIV-1 promotor, and RNAi knockdown of HDAC3 combined with suboptimal concentrations of global HDAC enzymatic inhibition leads to an increase in HIV-1 LTR activation[107]. Furthermore, a recent study by Wang *et al*.[108] showed that Vif expression promotes the transition of cells from G_1 to S phase, an effect that is CUL5-independent. The authors suggested that Vif might have a role in reactivation of latent provirus or promote virus expression after reactivation of latently infected cells. While our results do not corroborate Vif's interaction with CDK9 and BRD4, which are thought to be responsible for Vif's effect on cell cycle progression, it seems intriguing to re-address a potential role for Vif in HIV latency through binding of PEBB/CBFβ and/or the HDAC3 complex.

Vif also associated with two proteins that function in autophagy, AMRA1 and SQSTM. While HIV has recently been shown to down regulate autophagy in infected cells[109,110], this process has not been linked to Vif so far. How autophagy may relate to Vif therefore remains to be discovered.

Vpr

A number of different functions during at multiple steps of the viral life cycle have been proposed for the accessory protein Vpr, including modulation of viral reverse transcription, nuclear transport of the PIC, transcriptional activation of the LTR promoter, induction of cell cycle G2 arrest and apoptosis (reviewed in Zhao *et al.*, 2011[111]). Consistent with these diverse functions, Vpr was reported to interact with a number of cellular factors. The best characterized association is with the CUL4A-DDB1 ubiquitin ligase complex, mediated through the DCAF protein VPRBP (DCAF1)[112,113], which also was the strongest association found in our study for Vpr. DDB1 and the DDB1-associated protein DDA1, additional components of this complex, also obtained high MiST scores in both cell types. Besides VPRBP, two other DCAF proteins associated with Vpr, namely WD42A (DCAF8) and WDR23 (DCAF11), likely through their association with DDB. WD42A and WDR23 have been shown to be associated with DDB1 and DDA1 independently of Vpr in previous MS studies[114,115]. We also found UNG (Uracil-DNA glycosylase), which was originally identified as a Vpr interacting protein in a yeast-two hybrid screen[116], and was shown to be a substrate of the Vpr-CUL4A-DDB1 ubiquitin ligase[117,118], as interacting with Vpr in Jurkat cells. Interestingly, components of the COP9 signalosome complex (CSN1-8), which regulates the ubiquitin activity of the CUL4-DDB1 by deneddylation, were found associated with Vpr in Jurkat cells This result contradicts a previous study in which the association of COP9 subunits with Cul4-DDB1-VPRBP E3 complex was inhibited by Vpr in HEK293T cells[119], which led to the suggestion that Vpr enhances the catalytic activity of Cul4–DDB1-VPRBP by interfering with Nedd8 deconjugation by COP9[119]. A potential positive influence of the proteasome inhibitor treatment on this interaction needs to be investigated.

A surprising finding was that Vpr strongly associated with Dynein heavy chain, as well as two intermediate chains in both cell types. Dynein is a multisubunit protein complex, that mediates the retrograde transport of cargo along microtubules from the cell periphery towards

the MTOC. Dynein was reported before to mediate the long-range transport of the HIV reverse transcription complex (RTC) composed of MA, NC, IN, RT, and Vpr toward the nucleus[96], and it was shown that anti-dynein antibodies inhibit a productive HIV infection[120]. However, to date, there has been no direct connection between Vpr and the transfer of the RTC along microtubules.

Another protein found to be strongly associated with Vpr in both cell types is the WD-repeat containing protein RBBP7. Although RBBP7 was shown to co-immunoprecipitate with CUL4-DDB1 when overexpressed[121], it was neither part of the DDB1 nor DDA1 interaction network defined by previous AP-MS studies[114,115], suggesting that RBBP7 might bind Vpr independently of the CUL4A-DDB1 ubiquitin ligase. RBBP7 is known to associate with HAT1, a type B histone acetyltransferase, which also co-purified with Vpr in our study, and of note, was also detected in a previous AP-MS study of HIV-1 Vpr[119]. HAT1 acetylates newly synthesized histones in the cytoplasm and the current model suggests that HAT1/RBBP7 remain associated with the histone during its recruitment from the cytoplasm into the nucleus to the site of chromatin assembly, possibly in the vicinity of DNA double strand breaks[122-124]. Hence, a possible role for RBBP7/HAT1 binding of Vpr could be in the nuclear import of PICs and targeting to open chromatin structures. Another possible function of the Vpr-HAT1 interaction could involve a role in transcriptional activation of the LTR promoter through Vpr.

Several studies have shown that HIV-1 Vpr is able to induce apoptosis in a number of human cell lines and a number of molecules have been reported to be involved in Vpr-induced apoptosis. Interestingly, three Ankyrin repeat domain-containing proteins were detected with Vpr (ANKH1, ANR17, AN13A). ANKH1 was already previously identified in a yeast-two hybrid screen as HIV-1 Vpr interactor, and the authors showed that ANKH1 possesses an antiapoptotic effect [125]. However, to our knowledge no further study on this interaction has been reported. Of note, ANR17 was also detected in another AP-MS study of Vpr[119].

We also found the SMN complex (SMN/GEMI6/GEMI2/DDX20/GEMI4/GEMI7), which is required for the biogenesis of s small nuclear ribonucleoproteins (snRNPs), associated with Vpr in Jurkat cells. GEMI2 was previously identified in a yeast-two hybrid screen to bind HIV-1 IN and reduction of it via RNAi dramatically reduced HIV-1 infection in human primary monocyte-derived macrophages and also reduced viral cDNA synthesis[126], suggesting a role of GEMI2 in reverse transcription. However, this effect was strongly dependent on the presence of

IN and independent of Vpr[127]. Further studies are thus required to test whether besides IN, Vpr might be involved in the effect of GEMI2 on HIV-1 infection.

Vpu

The HIV-1 accessory protein Vpu is the only HIV integral membrane protein besides Env and enhances virus release from cells[128]. Two major roles have been described for Vpu: i) the retention and degradation of newly-synthesized CD4 proteins in the ER[129], and ii) the inactivation of BST2/tetherin, an antiviral factor which causes retention of viral particles on the cell surface. In a yeast-two hybrid screen, the Vpu cytoplasmic tail was found to bind βTrCP/FBW1, a subunit of the SKP1-CUL1-FBW1 (SCF) E3 ubiquitin ligase complex[130]. While binding of the SCF ubiquitin ligase is essential for Vpu mediated ubiquitination and proteasomal degradation of CD4, the mechanism by which Vpu counteracts BST2 seems to be mechanistically distinct[131]. A number of recent publications suggested that binding of FBW1 is not strictly required for Vpu's virus release activity[132-134]. It was suggested that Vpu depletes the plasma membrane pool of BST2, either by interfering with trafficking of newly synthesized BST2 to the cell surface[135], via active internalization of BST2[134], or by interfering with recycling of BST2 to the plasma membrane, all of which ultimately lead to endo-lysosomal degradation[136]. Whatever mechanism is correct, common to all models is that Vpu seems to acts as a regulator of protein transport between the post ER compartments trans-Golgi network (TGN), plasma membrane and endosome. Importantly, the observation that βTrCP/FBW1binding deficient Vpu or Vpu in cells depleted for FBW1 is still partially able to enhance virion release led to the postulation that the cytoplasmic tail of Vpu might recruit not only βTrCP/FBW1 but also an unknown cofactor[132,134].

Besides βTrCP/FBW1 and CD4, three other cellular proteins were reported to interact with HIV-1 Vpu (reviewed in Guatelli, 2009[137]): the potassium channel TASK-1 was identified by structural homology analysis[138], while CD74 (invariant chain of MHCII) and the co-chaperone SGTA were both identified in yeast-two hybrid screens with Vpu[139,140]. So far, none of these proteins were shown to function in Vpu's anti-BST2 of CD4 activity.

In the AP-MS analysis presented herein, Vpu associated strongly with the SCF complex (SKP1, CUL1, FBW1A, FBW1B). The two most strongly associated proteins, LPPRC and SLIRP, form a complex that regulates mitochondrial mRNA synthesis[141]. Accordant to their

function, both proteins are localized predominantly in mitochondria, and have no apparent connection to Vpu activity. It might be noteworthy to point out that LPPRC contains twenty pentatricopeptide motifs, the structure of which has not been determined yet, but is predicted to be similar to the tetratricopeptide repeat (TPR) motif, which is present in SGTA and was shown to be sufficient for binding Vpu[142].

One reason why Vpu and Gp160 have the largest number of associated proteins (Fig. 2d) is likely because they are the only transmembrane proteins in our HIV-1 dataset and therefore the filtering of the membrane-specific background is not as efficient as for the cytosolic or nuclear proteins. To filter out membrane-specific background further, we compared the HIV AP-MS data set with an independent MS data set obtained from purification of five unrelated transmembrane proteins (ADRB2, DRD2, OPRD, V2R, OPRM) under the same conditions as the HIV proteins. Proteins that overlap with this 'membrane background' data set are marked with (**) in Supplementary Data 3. As expected, the overlap was highest with the Vpu and Env candidates. Eleven of the Vpu candidates co-purified also with at least two of the control membrane proteins, mostly subunits of the membrane ATP synthase, as well as the membrane-bound heterocomplex ECHA/ECHB involved in fatty acid β-oxidation.

We also found SCAM3 (Secretory carrier membrane proteins 3) associated with Vpu, which localizes to the trans-Golgi network and endosomal recycling compartment and functions in the cell surface recycling system[143]. Most importantly, SCAM3 was shown to associate with the ESCRT-0 subunit HRS it was suggested that suggest that SCAM3 has a distinct function in parallel with the ESCRTs that regulates receptor degradation[144]. HRS was recently shown to be important for both constitutive and Vpu-enhanced degradation of BST2. Furthermore, Vpu seemed to mediate the association between HRS and BST2[145]. SCAM3 was also identified in an siRNA screen aimed at finding host factors that regulate the Salmonella Typhimurium secretory pathway [146]. Based on the model presented by Janvier *et al*[145], we propose that SCAM3 might be the link that connects BST2 to the ESCRT-mediated lysosomal degradation pathway, triggered by SCF E3 ligase mediated ubiquitination of BST2.

Tat

HIV Tat regulates HIV gene expression by binding to an RNA structure (TAR) in the newly synthesized HIV mRNA. It associates with the P-TEFb complex, whose core is composed

of Cyclin T1 and the cyclin dependent kinase CDK9. CDK9 phosphorylates the C terminal domain of RNA polymerase II, which increases its processivity, enabling full-length HIV transcripts. Cyclin T1 was initially identified as HIV Tat interactor in a GST-Tat purification from HeLa nuclear extracts[147]. In total, more than 370 human proteins have been reported to interact with HIV-1 Tat according to the NIAID HIV-1 Human Protein Interaction Database[148,149].

Recent work, including work from our group, identified additional components of the Tat P-TEFb complex[150-152]. For example, Sobhian et al.[150] characterized two different Tat complexes, an active complex compsed of P-TEFb, AF9, AFF1, AFF4, ELL, ENL, and PAF1, and an inactive complex primarily composed of LARP7 and MEPCE, which also was observed in Tat complexes on the HIV-1 promoter[152]. In the AP-MS screen presented herein, eight out of those ten components were found associated with Tat, with one (ELL) being below the MiST threshold of 0.75 (**Supplemetary Data 3**, full list). AF9 and PAF1 were not detected.

Another previously reported Tat interactor identified in our work is the deacetylase sirtuin 1 (SIRT1), which has been shown to associate with and deacetylate Tat, thereby activating Tat-mediated transactivation[153].

Two other Tat cofactor candidates identified in our study are SAHH2/IRBIT and SAHH3. Originally identified as inositol 1,4,5-trisphosphate binding protein[154], SAHH2 was more recently reported to bind the FIP1 subunit of the cleavage and polyadenylation specificity factor (CPSF) complex[155]. The CPSF complex plays a central role in pre-mRNA 3'-end processing and FIP1 was reported to bind U-rich upstream stimulatory sequence elements (USE) on the pre-mRNA and to recruit the poly(A) polymerase (PAP)[156]. Poly(A) site recognition is critical step during HIV-1 transcription, since the proviral DNA contains poly(A) signals in both the 3' and 5' long terminal repeat (LTR). The two poly(A) sites therefore need to be regulated differentially: 5' LTR poly(A) site formation should be inhibited and efficient processing at the 3' poly(A) site is desired. Two mechanisms for the differential regulation of the two poly(A) sites have been reported. One suggests that suppression of the 5' poly(A) site by the downstream major splice donor site occurs[157], and on the other hypothesizes that USE, which directly binds the CPSF complex[158], stimulates the 3' poly(A) site. A separate connection between HIV-1 Tat and CPSF was reported by de la Vega *et al.* who showed that CPSF binds directly to the HIV-1 LTR promoter and suppresses its basal transcription activity. They further show that Tat

indirectly binds the CPSF-73 subunit via unknown cellular proteins, thereby counteracting its repressive activity[159]. Given the strong association we found between Tat and SAHH2 independently in two cell types, one may speculate, therefore, that SAHH2-FIP1 might provide a link between Tat and CPSF manipulation.

Phosphorylation not only of RNA PolII, but also of other components like Tat, CDK9, NELF, and DSIF play a major role in Tat-induced HIV-1 transcription[160,161]. To date, two phosphatases, protein phosphatase 1 (PP1) and protein phosphatase 2A (PP2A), have been reported to regulate Tat activity[162,163]. PP1 was reported to bind directly to HIV-1 Tat[164] and was proposed to dephosphorylate CDK9[165]. In our study, three PP1 subunits (A, B, 12C) associated weakly with Tat, but also with Gag and NC and therefore all scored below the MiST threshold. A subunit of PP2A associated with HIV Pol, but not Tat, in HEK293 cells. However, PPM1G/PP2Cgamma, a phosphatase not implicated in HIV biology before, strongly associated with Tat in HEK cells (and also weakly in Jurkat). PPM1G is a nuclear Ser/Thr phosphatase reported to associate with histones, the SMN complex and a spliceosome-associated factor YBOX1[166-168]. PPM1G was proposed to modulate alternative splicing of specific pre-mRNAs coregulated by YBOX-1.

Nef

The most enigmatic HIV protein is probably the accessory protein Nef. Nef is a myristoylated 27 kDa protein which is largely dispensable for HIV replication in cell cultures, but accelerates the development of AIDS in patients. Many different functions have been reported for Nef: downregulation of a number of cell surface proteins like CD4 and MHCI[169,170], modulation of T cell receptor signaling[171], apoptosis[172], induction of cytoskeleton changes[173], T cell differentiation[174], B lymphocyte hyperactivation[175], induction of superoxide release[176,177], enhancement of virion infectivity[178], inhibition of fibroblast motility and T cell chemotaxis[179], induction of secretory activity[180], induction of chemokines[181] and impairment of cholesterol efflux from macrophages[182] (reviewed in Foster *et al.*, 2011[183]). In line with these diverse roles, more than eighty host cell proteins have been reported to interact with Nef, the most studied being the (Cdc42/Rac)-activated kinase PAK2, the clathrin adaptor complexes AP1 and AP2, the tyrosine kinases Lck, Fyn, Hck, Lyn, and c-Src and the guanine nucleotide exchange factor VAV. Most of these interactions however are considered low-affinity and not amenable to

standard protein interaction assays. Accordingly, in our AP-MS studies, none of these proteins was co-purified significantly with Nef. VAV was specifically detected with Nef, but not reproducible enough to obtain a high MiST score. The same was true for VATH/NBP1, previously identified in a yeast-two hybrid screen with Nef[184]. DOCK2 and ELMO1, previously identified in an AP-MS study of Nef expressed in Jurkat cells[185], were detected in Jurkat pulldowns of Vpu, but not with Nef.

However, we did find Nef interacting strongly with two proteins in both HEK293 and Jurkat T cells, Acyl-coenzyme A thioesterase 8 (ACOT8) and N-myristoyltransferase 1 (NMT1). NMT1 was previously shown to form abundant and stable complexes with HIV Nef[186] and is likely important for its myristoylation, although a recent report suggested that Nef is preferentially myristoylated by NMT2 in vitro[187]. Interaction of Nef and ACOT8 was also previously identified by three different groups using yeast-two hybrid and Nef affinity purification from T cell lysates[188-190]. Nef was shown to enhance ACOT8 activity *in vitro* and the ACOT8 binding deficient Nef(D123G) mutant was completely defective for CD4 downregulation, MHCI downregulation, and enhancement of infectivity[191]. However, since the ACOT8 binding site lies within the region important for Nef oligomerization, the authors attributed the functional defects rather to an impairment of the oligomerization instead of a loss of ACOT8 association.

ACOT8 belongs to the superfamily of acyl-CoA thioesterases which hydrolyze acyl-CoA thioester compounds into fatty acids and coenzyme A. ACOT8 was shown to have a broad specificity for a large variety of CoA esters. Recently it has become clear that by controlling the ratio of of free and activated fatty acids in the cell, ACOTs are involved in numerous cellular processes, including lipid metabolism, signalling, gene transcription, apoptosis, budding and fusion of intracellular membranes, protein targeting to membranes and enzyme regulation (reviewed in Hunt and Alexson, 2002[192]). Importantly, both the ACOT substrates as well as their products are lipid mediators, acting as ligands for nuclear receptors and therefore are involved in controlling transcription. An interesting class of receptors regulated by free fatty acids and acyl-CoA are the peroxisome proliferator-activated receptors (PPARs). PPARs can act as transcriptional activators or repressors and were shown to play an important role in inflammatory processes in T cells, macrophages and dendritic cells (reviewed in Daynes and Jones, 2002[193]). Intriguingly, there appears to be a high degree of overlap between functions reported for Nef and

cellular processes regulated by PPARs, including differentiation of T cells[194], cytokine production and migration of DCs, regulation of ABCA1 and cholesterol efflux from dendritic cells[195], wound healing, actin remodeling, expression and phosphorylation of PAK2[196], leukocyte adhesion and migration[197], cytokine secretion[198,199], and regulation of inflammatory responses through nuclear transcription factors, such as NF-KappaB, AP-1 and STAT1[200]. Furthermore, evidence for a direct link between HIV replication and PPAR signaling came from observations that PPAR agonists decreased HIV replication in several primary cell systems[201-203], and PPARgamma agonists were shown to mimic the activity of soluble Nef with respect to suppression of hematopoiesis[204].

Together, in light of the strong association between Nef and ACOT8 and the functional congruence between Nef and PPAR signaling it seems worthwhile to further study the physiological role of the Nef-ACOT8 interaction.

Rev

Rev is a 19 kDa HIV accessory protein which mediates the nuclear export of unspliced and incompletely spliced HIV mRNA. To this end, Rev oligomerizes on the Rev responsive element (RRE) within the HIV mRNA and binds the nuclear export factor XPO1/Crm1. The Rev-XPO1 interaction is likely to be transient and was not detected in previous yeast-two hybrid[205,206] or AP-MS[207] studies of Rev. Consistent with this, we did not find XPO1 associated with Rev. Surprisingly, XPO1 associated with Vpu, possibly indirectly through an potential nuclear export sequence (NES) in the F-Box protein βTrCP/FBW1 which was suggested to bind XPO1[208].

A previous AP-MS study of Rev found a prominent interaction between Rev and nucleosome assembly protein 1 (Nap1)[207]. We also found Nap1 (NP1L1, NP1L4) consistently associated with Rev, however it also associated with several other HIV proteins and was therefore considered unspecific.

The only protein we found specifically associated with Rev in both HEK293 and Jurkat cells was Transmembrane emp24 domain-containing protein 9 (TMED9). TMED9 is a member of the p24 protein family of trafficking proteins that function in the early secretory pathway[209]. Two other members of this family, TMED10 and TMED4, also associated with Rev in our study. The p24 proteins are type-I transmembrane proteins localized mainly to the ER-Golgi

compartment, but have also been detected in secretory granules[210] and at the plasma membrane[211].. This class of proteins have a short cytoplasmic tail which binds COPI and COPII proteins thereby mediating vesicle formation, while the luminal part is thought to regulate cargo selection[212,213]. While there is no obvious connection between HIV Rev and ER-Golgi function, several studies have linked HIV RNA export and trafficking to virus assembly[214-217], and Rev has been suggested to not only function in nuclear export of viral RNA, but also in downstream events like Gag assembly, RNA packaging and budding[218]. The mechanisms of HIV-1 RNA transport through the cytoplasm and the site of initial recognition by Gag are still poorly understood. It is possible that these proteins play some role in these latter processes.

Supplementary References

36. Sowa, M.E., Bennett, E.J., Gygi, S.P. & Harper, J.W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **138**, 389-403 (2009).
37. Choi, H. et al. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods* **8**, 70-3 (2010).
38. Brass, A.L. et al. Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319**, 921-6 (2008).
39. König, R. et al. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* **135**, 49-60 (2008).
40. Zhou, H. et al. Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* **4**, 495-504 (2008).
41. Yeung, M.L., Houzet, L., Yedavalli, V.S. & Jeang, K.T. A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. *J Biol Chem* **284**, 19463-73 (2009).
42. Chakrabarti, B.K. et al. Modifications of the human immunodeficiency virus envelope glycoprotein enhance immunogenicity for genetic immunization. *J Virol* **76**, 5357-68 (2002).
43. Hermida-Matsumoto, L. & Resh, M.D. Localization of human immunodeficiency virus type 1 Gag and Env at the plasma membrane by confocal imaging. *J Virol* **74**, 8670-9 (2000).
44. Wagner, R. et al. Rev-independent expression of synthetic gag-pol genes of human immunodeficiency virus type 1 and simian immunodeficiency virus: implications for the safety of lentiviral vectors. *Hum Gene Ther* **11**, 2403-13 (2000).
45. Cherepanov, P. et al. High-level expression of active HIV-1 integrase from a synthetic gene in human cells. *Faseb J* **14**, 1389-99 (2000).
46. Nguyen, K.L. et al. Codon optimization of the HIV-1 vpu and vif genes stabilizes their mRNA and allows for highly efficient Rev-independent expression. *Virology* **319**, 163-75 (2004).
47. Bolton, D.L. & Lenardo, M.J. Vpr cytopathicity independent of G2/M cell cycle arrest in human immunodeficiency virus type 1-infected CD4+ T cells. *J Virol* **81**, 8878-90 (2007).
48. Wang, N.S., Unkila, M.T., Reineks, E.Z. & Distelhorst, C.W. Transient expression of wild-type or mitochondrially targeted Bcl-2 induces apoptosis, whereas transient expression of endoplasmic reticulum-targeted Bcl-2 is protective against Bax-induced cell death. *J Biol Chem* **276**, 44117-28 (2001).
49. Brunet, A. et al. Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. *Science* **303**, 2011-5 (2004).
50. Chesebro, B., Wehrly, K., Nishio, J. & Perryman, S. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. *J Virol* **66**, 6547-54 (1992).
51. Clauser, K.R., Baker, P. & Burlingame, A.L. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem* **71**, 2871-82 (1999).

52. Gavin, A.C. et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631-6 (2006).
53. Ho, Y. et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415**, 180-3 (2002).
54. Krogan, N.J. et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440**, 637-43 (2006).
55. Behrends, C., Sowa, M.E., Gygi, S.P. & Harper, J.W. Network organization of the human autophagy system. *Nature* **466**, 68-76.
56. Butland, G. et al. Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature* **433**, 531-7 (2005).
57. Collins, S.R. et al. Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. *Mol Cell Proteomics* **6**, 439-50 (2007).
58. Griffin, N.M. et al. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol* **28**, 83-9 (2010).
59. Zito, T., Wilbert, N., Wiskott, L. & Berkes, P. Modular Toolkit for Data Processing (MDP): A Python Data Processing Framework. *Front Neuroinformatics* **2**, 8 (2008).
60. Chen, G.I. & Gingras, A.C. Affinity-purification mass spectrometry (AP-MS) of serine/threonine phosphatases. *Methods* **42**, 298-305 (2007).
61. Ewing, R.M. et al. Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* **3**, 89 (2007).
62. Sardiu, M.E. et al. Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc Natl Acad Sci U S A* **105**, 1454-9 (2008).
63. El Hader, C. et al. HCaRG increases renal cell migration by a TGF-alpha autocrine loop mechanism. *Am J Physiol Renal Physiol* **289**, F1273-80 (2005).
64. Wu, C. et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* **10**, R130 (2009).
65. Ndolo, T., George, M., Nguyen, H. & Dandekar, S. Expression of simian immunodeficiency virus Nef protein in CD4+ T cells leads to a molecular profile of viral persistence and immune evasion. *Virology* **353**, 374-87 (2006).
66. Huang da, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).
67. Finn, R.D. et al. Pfam: clans, web tools and services. *Nucleic Acids Res* **34**, D247-51 (2006).
68. Apweiler R, M.M., O'Donovan C, Magrane M, Alam-Faruque Y, et al. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res* **38**, D142-8 (2010).
69. Su, A.I. et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* **101**, 6062-7 (2004).
70. Louis, J.M. et al. Autoprocessing of the HIV-1 protease using purified wild-type and mutated fusion proteins expressed at high levels in Escherichia coli. *Eur J Biochem* **199**, 361-9 (1991).
71. Schilling, O. & Overall, C.M. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* **26**, 685-94 (2008).
72. Charneau, P. et al. HIV-1 reverse transcription. A termination step at the center of the genome. *J Mol Biol* **241**, 651-62 (1994).

73. Clavel, F. & Charneau, P. Fusion from without directed by human immunodeficiency virus particles. *J Virol* **68**, 1179-85 (1994).
74. Connor, R.I., Chen, B.K., Choe, S. & Landau, N.R. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. *Virology* **206**, 935-44 (1995).
75. Schmittgen, T.D. & Livak, K.J. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-8 (2008).
76. Butler, S.L., Hansen, M.S. & Bushman, F.D. A quantitative assay for HIV DNA integration in vivo. *Nat Med* **7**, 631-4 (2001).
77. Chatr-aryamontri, A. et al. VirusMINT: a viral protein interaction database. *Nucleic Acids Res* **37**, D669-73 (2009).
78. Rulli, S.J., Jr. et al. Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J Virol* **81**, 6623-31 (2007).
79. Roy, B.B. et al. Association of RNA helicase a with human immunodeficiency virus type 1 particles. *J Biol Chem* **281**, 12625-35 (2006).
80. Lama, J. & Trono, D. Human immunodeficiency virus type 1 matrix protein interacts with cellular protein HO3. *J Virol* **72**, 1671-6 (1998).
81. Javanbakht, H. et al. The interaction between HIV-1 Gag and human lysyl-tRNA synthetase during viral assembly. *J Biol Chem* **278**, 27644-51 (2003).
82. Kovaleski, B.J. et al. In vitro characterization of the interaction between HIV-1 Gag and human lysyl-tRNA synthetase. *J Biol Chem* **281**, 19449-56 (2006).
83. Halwani, R. et al. Cellular distribution of Lysyl-tRNA synthetase and its interaction with Gag during human immunodeficiency virus type 1 assembly. *J Virol* **78**, 7553-64 (2004).
84. Guo, F., Cen, S., Niu, M., Javanbakht, H. & Kleiman, L. Specific inhibition of the synthesis of human lysyl-tRNA synthetase results in decreases in tRNA(Lys) incorporation, tRNA(3)(Lys) annealing to viral RNA, and viral infectivity in human immunodeficiency virus type 1. *J Virol* **77**, 9817-22 (2003).
85. Luban, J., Bossolt, K.L., Franke, E.K., Kalpana, G.V. & Goff, S.P. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* **73**, 1067-78 (1993).
86. Morita, E. & Sundquist, W.I. Retrovirus budding. *Annu Rev Cell Dev Biol* **20**, 395-425 (2004).
87. VerPlank, L. et al. Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). *Proc Natl Acad Sci U S A* **98**, 7724-9 (2001).
88. Strack, B., Calistri, A., Craig, S., Popova, E. & Gottlinger, H.G. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell* **114**, 689-99 (2003).
89. Lemay, J. et al. AKAP149 binds to HIV-1 reverse transcriptase and is involved in the reverse transcription. *J Mol Biol* **383**, 783-96 (2008).
90. Lemay, J. et al. HuR interacts with human immunodeficiency virus type 1 reverse transcriptase, and modulates reverse transcription in infected cells. *Retrovirology* **5**, 47 (2008).
91. Ahn, J. et al. The RNA binding protein HuR does not interact directly with HIV-1 reverse transcriptase and does not affect reverse transcription in vitro. *Retrovirology* **7**, 40 (2010).
92. Hottiger, M. et al. The large subunit of HIV-1 reverse transcriptase interacts with beta-actin. *Nucleic Acids Res* **23**, 736-41 (1995).

93. Carlier, M.F. et al. GRB2 links signaling to actin assembly by enhancing interaction of neural Wiskott-Aldrich syndrome protein (N-WASp) with actin-related protein (ARP2/3) complex. *J Biol Chem* **275**, 21946-52 (2000).
94. Jacque, J.M. et al. Modulation of HIV-1 infectivity by MAPK, a virion-associated kinase. *Embo J* **17**, 2607-18 (1998).
95. Komano, J., Miyauchi, K., Matsuda, Z. & Yamamoto, N. Inhibiting the Arp2/3 complex limits infection of both intracellular mature vaccinia virus and primate lentiviruses. *Mol Biol Cell* **15**, 5197-207 (2004).
96. McDonald, D. et al. Visualization of the intracellular behavior of HIV in living cells. *J Cell Biol* **159**, 441-52 (2002).
97. Bukrinskaya, A., Brichacek, B., Mann, A. & Stevenson, M. Establishment of a functional human immunodeficiency virus type 1 (HIV-1) reverse transcription complex involves the cytoskeleton. *J Exp Med* **188**, 2113-25 (1998).
98. James, M.F., Beauchamp, R.L., Manchanda, N., Kazlauskas, A. & Ramesh, V. A NHERF binding site links the betaPDGFR to the cytoskeleton and regulates cell spreading and migration. *J Cell Sci* **117**, 2951-61 (2004).
99. Haedicke, J., de Los Santos, K., Goff, S.P. & Naghavi, M.H. The Ezrin-radixin-moesin family member ezrin regulates stable microtubule formation and retroviral infection. *J Virol* **82**, 4665-70 (2008).
100. Naghavi, M.H. et al. Moesin regulates stable microtubule formation and limits retroviral infection in cultured cells. *Embo J* **26**, 41-52 (2007).
101. Ciuffi, A. et al. A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* **11**, 1287-9 (2005).
102. Cherepanov, P. et al. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* **278**, 372-81 (2003).
103. Chen, H. & Engelman, A. The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc Natl Acad Sci U S A* **95**, 15270-4 (1998).
104. Kalpana, G.V., Marmon, S., Wang, W., Crabtree, G.R. & Goff, S.P. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* **266**, 2002-6 (1994).
105. Yu, X. et al. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056-60 (2003).
106. Van Lint, C., Emiliani, S., Ott, M. & Verdin, E. Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. *Embo J* **15**, 1112-20 (1996).
107. Keedy, K.S. et al. A limited group of class I histone deacetylases acts to repress human immunodeficiency virus type 1 expression. *J Virol* **83**, 4749-56 (2009).
108. Wang, J. et al. HIV-1 Vif promotes the G1- to S-phase cell-cycle transition. *Blood* **117**, 1260-9 (2011).
109. Kyei, G.B. et al. Autophagy pathway intersects with HIV-1 biosynthesis and regulates viral yields in macrophages. *J Cell Biol* **186**, 255-68 (2009).
110. Zhou, D. & Spector, S.A. Human immunodeficiency virus type-1 infection inhibits autophagy. *Aids* **22**, 695-9 (2008).
111. Zhao, R.Y., Li, G. & Bukrinsky, M.I. Vpr-Host Interactions During HIV-1 Viral Life Cycle. *J Neuroimmune Pharmacol* (2011).

112. Zhao, L.J., Mukherjee, S. & Narayan, O. Biochemical mechanism of HIV-I Vpr function. Specific interaction with a cellular protein. *J Biol Chem* **269**, 15577-82 (1994).
113. Belzile, J.P. et al. HIV-1 Vpr-mediated G2 arrest involves the DDB1-CUL4AVPRBP E3 ubiquitin ligase. *PLoS Pathog* **3**, e85 (2007).
114. Olma, M.H. et al. An interaction network of the mammalian COP9 signalosome identifies Dda1 as a core subunit of multiple Cul4-based E3 ligases. *J Cell Sci* **122**, 1035-44 (2009).
115. Jin, J., Arias, E.E., Chen, J., Harper, J.W. & Walter, J.C. A family of diverse Cul4-Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1. *Mol Cell* **23**, 709-21 (2006).
116. Bouhamdan, M. et al. Human immunodeficiency virus type 1 Vpr protein binds to the uracil DNA glycosylase DNA repair enzyme. *J Virol* **70**, 697-704 (1996).
117. Schrofelbauer, B., Hakata, Y. & Landau, N.R. HIV-1 Vpr function is mediated by interaction with the damage-specific DNA-binding protein DDB1. *Proc Natl Acad Sci U S A* **104**, 4130-5 (2007).
118. Ahn, J. et al. HIV-1 Vpr loads uracil DNA glycosylase-2 onto DCAF1, a substrate recognition subunit of a cullin 4A-ring E3 ubiquitin ligase for proteasome-dependent degradation. *J Biol Chem* **285**, 37333-41 (2010).
119. Hrecka, K. et al. Lentiviral Vpr usurps Cul4-DDB1[VprBP] E3 ubiquitin ligase to modulate cell cycle. *Proc Natl Acad Sci U S A* **104**, 11778-83 (2007).
120. Fontenot, D.R. et al. Dynein light chain 1 peptide inhibits human immunodeficiency virus infection in eukaryotic cells. *Biochem Biophys Res Commun* **363**, 901-7 (2007).
121. He, Y.J., McCall, C.M., Hu, J., Zeng, Y. & Xiong, Y. DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4-ROC1 ubiquitin ligases. *Genes Dev* **20**, 2949-54 (2006).
122. Parthun, M.R. Hat1: the emerging cellular roles of a type B histone acetyltransferase. *Oncogene* **26**, 5319-28 (2007).
123. Ejlassi-Lassallette, A., Mocquard, E., Arnaud, M.C. & Thiriet, C. H4 replication-dependent diacetylation and Hat1 promote S-phase chromatin assembly in vivo. *Mol Biol Cell* **22**, 245-55 (2011).
124. Qin, S. & Parthun, M.R. Histone H3 and the histone acetyltransferase Hat1p contribute to DNA double-strand break repair. *Mol Cell Biol* **22**, 8353-65 (2002).
125. Miles, M.C. et al. Molecular and functional characterization of a novel splice variant of ANKHD1 that lacks the KH domain and its role in cell survival and apoptosis. *Febs J* **272**, 4091-102 (2005).
126. Hamamoto, S., Nishitsuji, H., Amagasa, T., Kannagi, M. & Masuda, T. Identification of a novel human immunodeficiency virus type 1 integrase interactor, Gemin2, that facilitates efficient viral cDNA synthesis in vivo. *J Virol* **80**, 5670-7 (2006).
127. Nishitsuji, H. et al. Augmentation of reverse transcription by integrase through an interaction with host factor, SIP1/Gemin2 Is critical for HIV-1 infection. *PLoS One* **4**, e7825 (2009).
128. Klimkait, T., Strebel, K., Hoggan, M.D., Martin, M.A. & Orenstein, J.M. The human immunodeficiency virus type 1-specific protein vpu is required for efficient virus maturation and release. *J Virol* **64**, 621-9 (1990).
129. Willey, R.L., Maldarelli, F., Martin, M.A. & Strebel, K. Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. *J Virol* **66**, 7193-200 (1992).

130. Margottin, F. et al. A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. *Mol Cell* **1**, 565-74 (1998).
131. Schubert, U. et al. The two biological activities of human immunodeficiency virus type 1 Vpu protein involve two separable structural domains. *J Virol* **70**, 809-19 (1996).
132. Tervo, H.M. et al. beta-TrCP is dispensable for Vpu's ability to overcome the CD317/Tetherin-imposed restriction to HIV-1 release. *Retrovirology* **8**, 9 (2011).
133. Dube, M. et al. Antagonism of tetherin restriction of HIV-1 release by Vpu involves binding and sequestration of the restriction factor in a perinuclear compartment. *PLoS Pathog* **6**, e1000856 (2010).
134. Iwabu, Y. et al. HIV-1 accessory protein Vpu internalizes cell-surface BST-2/tetherin through transmembrane interactions leading to lysosomes. *J Biol Chem* **284**, 35060-72 (2009).
135. Andrew, A.J., Miyagi, E. & Strebel, K. Differential Effects of Human Immunodeficiency Virus Type 1 Vpu on the Stability of BST-2/Tetherin. *J Virol* **85**, 2611-9 (2011).
136. Hauser, H. et al. HIV-1 Vpu and HIV-2 Env counteract BST-2/tetherin by sequestration in a perinuclear compartment. *Retrovirology* **7**, 51 (2010).
137. Guatelli, J.C. Interactions of viral protein U (Vpu) with cellular factors. *Curr Top Microbiol Immunol* **339**, 27-45 (2009).
138. Hsu, K., Seharaseyon, J., Dong, P., Bour, S. & Marban, E. Mutual functional destruction of HIV-1 Vpu and host TASK-1 channel. *Mol Cell* **14**, 259-67 (2004).
139. Hussain, A., Wesley, C., Khalid, M., Chaudhry, A. & Jameel, S. Human immunodeficiency virus type 1 Vpu protein interacts with CD74 and modulates major histocompatibility complex class II presentation. *J Virol* **82**, 893-902 (2008).
140. Geraghty, R.J., Talbot, K.J., Callahan, M., Harper, W. & Panganiban, A.T. Cell type-dependence for Vpu function. *J Med Primatol* **23**, 146-50 (1994).
141. Sasarman, F., Brunel-Guitton, C., Antonicka, H., Wai, T. & Shoubridge, E.A. LRPPRC and SLIRP interact in a ribonucleoprotein complex that regulates posttranscriptional gene expression in mitochondria. *Mol Biol Cell* **21**, 1315-23 (2010).
142. Dutta, S. & Tan, Y.J. Structural and functional characterization of human SGT and its interaction with Vpu of the human immunodeficiency virus type 1. *Biochemistry* **47**, 10123-31 (2008).
143. Castle, A. & Castle, D. Ubiquitously expressed secretory carrier membrane proteins (SCAMPs) 1-4 mark different pathways and exhibit limited constitutive trafficking to and from the cell surface. *J Cell Sci* **118**, 3769-80 (2005).
144. Aoh, Q.L., Castle, A.M., Hubbard, C.H., Katsumata, O. & Castle, J.D. SCAMP3 negatively regulates epidermal growth factor receptor degradation and promotes receptor recycling. *Mol Biol Cell* **20**, 1816-32 (2009).
145. Janvier, K. et al. The ESCRT-0 Component HRS is Required for HIV-1 Vpu-Mediated BST-2/Tetherin Down-Regulation. *PLoS Pathog* **7**, e1001265 (2011).
146. Mota, L.J., Ramsden, A.E., Liu, M., Castle, J.D. & Holden, D.W. SCAMP3 is a component of the Salmonella-induced tubular network and reveals an interaction between bacterial effectors and post-Golgi trafficking. *Cell Microbiol* **11**, 1236-53 (2009).
147. Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H. & Jones, K.A. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92**, 451-62 (1998).

148. Fu, W. et al. Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res* **37**, D417-22 (2009).
149. Ptak, R.G. et al. Cataloguing the HIV type 1 human protein interaction network. *AIDS Res Hum Retroviruses* **24**, 1497-502 (2008).
150. Sobhian, B. et al. HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. *Mol Cell* **38**, 439-51 (2010).
151. He, N. et al. HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. *Mol Cell* **38**, 428-38 (2010).
152. D'Orso, I. & Frankel, A.D. RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. *Nat Struct Mol Biol* **17**, 815-21 (2010).
153. Pagans, S. et al. SIRT1 regulates HIV transcription via Tat deacetylation. *PLoS Biol* **3**, e41 (2005).
154. Ando, H., Mizutani, A., Matsu-ura, T. & Mikoshiba, K. IRBIT, a novel inositol 1,4,5-trisphosphate (IP3) receptor-binding protein, is released from the IP3 receptor upon IP3 binding to the receptor. *J Biol Chem* **278**, 10602-12 (2003).
155. Kiefer, H. et al. Inositol 1,4,5-triphosphate receptor-binding protein released with inositol 1,4,5-triphosphate (IRBIT) associates with components of the mRNA 3' processing machinery in a phosphorylation-dependent manner and inhibits polyadenylation. *J Biol Chem* **284**, 10694-705 (2009).
156. Kaufmann, I., Martin, G., Friedlein, A., Langen, H. & Keller, W. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *Embo J* **23**, 616-26 (2004).
157. Ashe, M.P., Pearson, L.H. & Proudfoot, N.J. The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *Embo J* **16**, 5752-63 (1997).
158. Gilmartin, G.M., Fleming, E.S., Oetjen, J. & Graveley, B.R. CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes Dev* **9**, 72-83 (1995).
159. de la Vega, L. et al. The 73 kDa subunit of the CPSF complex binds to the HIV-1 LTR promoter and functions as a negative regulatory factor that is inhibited by the HIV-1 Tat protein. *J Mol Biol* **372**, 317-30 (2007).
160. Endo-Munoz, L., Warby, T., Harrich, D. & McMillan, N.A. Phosphorylation of HIV Tat by PKR increases interaction with TAR RNA and enhances transcription. *Virol J* **2**, 17 (2005).
161. Garber, M.E. et al. CDK9 autophosphorylation regulates high-affinity binding of the human immunodeficiency virus type 1 tat-P-TEFb complex to TAR RNA. *Mol Cell Biol* **20**, 6958-69 (2000).
162. Ammosova, T. et al. Expression of a protein phosphatase 1 inhibitor, cdNIPP1, increases CDK9 threonine 186 phosphorylation and inhibits HIV-1 transcription. *J Biol Chem* **286**, 3798-804 (2011).
163. Faulkner, N.E., Lane, B.R., Bock, P.J. & Markovitz, D.M. Protein phosphatase 2A enhances activation of human immunodeficiency virus type 1 by phorbol myristate acetate. *J Virol* **77**, 2276-81 (2003).
164. Ammosova, T. et al. Nuclear targeting of protein phosphatase-1 by HIV-1 Tat protein. *J Biol Chem* **280**, 36364-71 (2005).

165. Ammosova, T., Washington, K., Debebe, Z., Brady, J. & Nekhai, S. Dephosphorylation of CDK9 by protein phosphatase 2A and protein phosphatase-1 in Tat-activated HIV-1 transcription. *Retrovirology* **2**, 47 (2005).
166. Allemand, E., Hastings, M.L., Murray, M.V., Myers, M.P. & Krainer, A.R. Alternative splicing regulation by interaction of phosphatase PP2Cgamma with nucleic acid-binding protein YB-1. *Nat Struct Mol Biol* **14**, 630-8 (2007).
167. Petri, S., Grimmler, M., Over, S., Fischer, U. & Gruss, O.J. Dephosphorylation of survival motor neurons (SMN) by PPM1G/PP2Cgamma governs Cajal body localization and stability of the SMN complex. *J Cell Biol* **179**, 451-65 (2007).
168. Kimura, H. et al. A novel histone exchange factor, protein phosphatase 2Cgamma, mediates the exchange and dephosphorylation of H2A-H2B. *J Cell Biol* **175**, 389-400 (2006).
169. Singh, R.K., Lau, D., Noviello, C.M., Ghosh, P. & Guatelli, J.C. An MHC-I cytoplasmic domain/HIV-1 Nef fusion protein binds directly to the mu subunit of the AP-1 endosomal coat complex. *PLoS One* **4**, e8364 (2009).
170. Schaefer, M.R., Wonderlich, E.R., Roeth, J.F., Leonard, J.A. & Collins, K.L. HIV-1 Nef targets MHC-I and CD4 for degradation via a final common beta-COP-dependent pathway in T cells. *PLoS Pathog* **4**, e1000131 (2008).
171. Neri, F., Giolo, G., Potesta, M., Petrini, S. & Doria, M. The HIV-1 Nef protein has a dual role in T cell receptor signaling in infected CD4+ T lymphocytes. *Virology* **410**, 316-26 (2011).
172. Rasola, A., Gramaglia, D., Boccaccio, C. & Comoglio, P.M. Apoptosis enhancement by the HIV-1 Nef protein. *J Immunol* **166**, 81-8 (2001).
173. Stolp, B., Abraham, L., Rudolph, J.M. & Fackler, O.T. Lentiviral Nef proteins utilize PAK2-mediated deregulation of cofilin as a general strategy to interfere with actin remodeling. *J Virol* **84**, 3935-48 (2010).
174. Dorival, C. et al. HIV-1 Nef protein expression in human CD34+ progenitors impairs the differentiation of an early T/NK cell precursor. *Virology* **377**, 207-15 (2008).
175. Swingler, S. et al. Evidence for a pathogenic determinant in HIV-1 Nef involved in B cell dysfunction in HIV/AIDS. *Cell Host Microbe* **4**, 63-76 (2008).
176. Salmen, S. et al. HIV-1 Nef associates with p22-phox, a component of the NADPH oxidase protein complex. *Cell Immunol* **263**, 166-71 (2010).
177. Olivetta, E. et al. HIV-1 Nef induces p47(phox) phosphorylation leading to a rapid superoxide anion release from the U937 human monoblastic cell line. *J Cell Biochem* **106**, 812-22 (2009).
178. Pizzato, M. et al. Dynamin 2 is required for the enhancement of HIV-1 infectivity by Nef. *Proc Natl Acad Sci U S A* **104**, 6812-7 (2007).
179. Stolp, B. et al. HIV-1 Nef interferes with host cell motility by deregulation of Cofilin. *Cell Host Microbe* **6**, 174-86 (2009).
180. Muratori, C. et al. Massive secretion by T cells is caused by HIV Nef in infected cells and by Nef transfer to bystander cells. *Cell Host Microbe* **6**, 218-30 (2009).
181. Swingler, S. et al. HIV-1 Nef mediates lymphocyte chemotaxis and activation by infected macrophages. *Nat Med* **5**, 997-103 (1999).
182. Mujawar, Z. et al. Human immunodeficiency virus impairs reverse cholesterol transport from macrophages. *PLoS Biol* **4**, e365 (2006).

183. Foster, J.L., Denial, S.J., Temple, B.R. & Garcia, J.V. Mechanisms of HIV-1 Nef Function and Intracellular Signaling. *J Neuroimmune Pharmacol* (2011).
184. Lu, X., Yu, H., Liu, S.H., Brodsky, F.M. & Peterlin, B.M. Interactions between HIV1 Nef and vacuolar ATPase facilitate the internalization of CD4. *Immunity* **8**, 647-56 (1998).
185. Janardhan, A., Swigut, T., Hill, B., Myers, M.P. & Skowronski, J. HIV-1 Nef binds the DOCK2-ELMO1 complex to activate rac and inhibit lymphocyte chemotaxis. *PLoS Biol* **2**, E6 (2004).
186. Hill, B.T. & Skowronski, J. Human N-myristoyltransferases form stable complexes with lentiviral nef and other viral and cellular substrate proteins. *J Virol* **79**, 1133-41 (2005).
187. Seaton, K.E. & Smith, C.D. N-Myristoyltransferase isozymes exhibit differential specificity for human immunodeficiency virus type 1 Gag and Nef. *J Gen Virol* **89**, 288-96 (2008).
188. Cohen, G.B., Rangan, V.S., Chen, B.K., Smith, S. & Baltimore, D. The human thioesterase II protein binds to a site on HIV-1 Nef critical for CD4 down-regulation. *J Biol Chem* **275**, 23097-105 (2000).
189. Liu, L.X. et al. Binding of HIV-1 Nef to a novel thioesterase enzyme correlates with Nef-mediated CD4 down-regulation. *J Biol Chem* **272**, 13779-85 (1997).
190. Watanabe, H. et al. A novel acyl-CoA thioesterase enhances its enzymatic activity by direct binding with HIV Nef. *Biochem Biophys Res Commun* **238**, 234-9 (1997).
191. Liu, L.X. et al. Mutation of a conserved residue (D123) required for oligomerization of human immunodeficiency virus type 1 Nef protein abolishes interaction with human thioesterase and results in impairment of Nef biological functions. *J Virol* **74**, 5310-9 (2000).
192. Hunt, M.C. & Alexson, S.E. The role Acyl-CoA thioesterases play in mediating intracellular lipid metabolism. *Prog Lipid Res* **41**, 99-130 (2002).
193. Daynes, R.A. & Jones, D.C. Emerging roles of PPARs in inflammation and immunity. *Nat Rev Immunol* **2**, 748-59 (2002).
194. Klotz, L. et al. The nuclear receptor PPAR gamma selectively inhibits Th17 differentiation in a T cell-intrinsic fashion and suppresses CNS autoimmunity. *J Exp Med* **206**, 2079-89 (2009).
195. Hanley, T.M., Blay Puryear, W., Gummuluru, S. & Viglianti, G.A. PPARgamma and LXR signaling inhibit dendritic cell-mediated HIV-1 capture and trans-infection. *PLoS Pathog* **6**, e1000981 (2010).
196. Tan, N.S. et al. The nuclear hormone receptor peroxisome proliferator-activated receptor beta/delta potentiates cell chemotactism, polarization, and migration. *Mol Cell Biol* **27**, 7161-75 (2007).
197. Ramirez, S.H. et al. Activation of peroxisome proliferator-activated receptor gamma (PPARgamma) suppresses Rho GTPases in human brain microvascular endothelial cells and inhibits adhesion and transendothelial migration of HIV-1 infected monocytes. *J Immunol* **180**, 1854-65 (2008).
198. Gosset, P. et al. Peroxisome proliferator-activated receptor gamma activators affect the maturation of human monocyte-derived dendritic cells. *Eur J Immunol* **31**, 2857-65 (2001).
199. Clark, R.B. et al. The nuclear receptor PPAR gamma and immunoregulation: PPAR gamma mediates inhibition of helper T cell responses. *J Immunol* **164**, 1364-71 (2000).

200. Kostadinova, R., Wahli, W. & Michalik, L. PPARs in diseases: control mechanisms of inflammation. *Curr Med Chem* **12**, 2995-3009 (2005).
201. Hayes, M.M., Lane, B.R., King, S.R., Markovitz, D.M. & Coffey, M.J. Peroxisome proliferator-activated receptor gamma agonists inhibit HIV-1 replication in macrophages by transcriptional and post-transcriptional effects. *J Biol Chem* **277**, 16913-9 (2002).
202. Skolnik, P.R., Rabbi, M.F., Mathys, J.M. & Greenberg, A.S. Stimulation of peroxisome proliferator-activated receptors alpha and gamma blocks HIV-1 replication and TNFalpha production in acutely infected primary blood cells, chronically infected U1 cells, and alveolar macrophages from HIV-infected subjects. *J Acquir Immune Defic Syndr* **31**, 1-10 (2002).
203. Lemoine, M., Capeau, J. & Serfaty, L. PPAR and Liver Injury in HIV-Infected Patients. *PPAR Res* **2009**, 906167 (2009).
204. Prost, S. et al. Human and simian immunodeficiency viruses deregulate early hematopoiesis through a Nef/PPARgamma/STAT5 signaling pathway in macaques. *J Clin Invest* **118**, 1765-75 (2008).
205. Bogerd, H.P., Fridell, R.A., Madore, S. & Cullen, B.R. Identification of a novel cellular cofactor for the Rev/Rex class of retroviral regulatory proteins. *Cell* **82**, 485-94 (1995).
206. Fritz, C.C., Zapp, M.L. & Green, M.R. A human nucleoporin-like protein that specifically interacts with HIV Rev. *Nature* **376**, 530-3 (1995).
207. Cochrane, A. et al. Stable complex formation between HIV Rev and the nucleosome assembly protein, NAP1, affects Rev function. *Virology* **388**, 103-11 (2009).
208. Nelson, D.E. & Laman, H. A competitive binding mechanism between SKP1 and exportin 1 (CRM1) controls the localization of a subset of F-box proteins. *J Biol Chem* (2011).
209. Strating, J.R. & Martens, G.J. The p24 family and selective transport processes at the ER-Golgi interface. *Biol Cell* **101**, 495-509 (2009).
210. Hosaka, M. et al. A subset of p23 localized on secretory granules in pancreatic beta-cells. *J Histochem Cytochem* **55**, 235-45 (2007).
211. Blum, R. & Lepier, A. The luminal domain of p23 (Tmp21) plays a critical role in p23 cell surface trafficking. *Traffic* **9**, 1530-50 (2008).
212. Dominguez, M. et al. gp25L/emp24/p24 protein family members of the cis-Golgi network bind both COP I and II coatomer. *J Cell Biol* **140**, 751-65 (1998).
213. Sohn, K. et al. A major transmembrane protein of Golgi-derived COPI-coated vesicles involved in coatomer binding. *J Cell Biol* **135**, 1239-48 (1996).
214. Swanson, C.M., Puffer, B.A., Ahmad, K.M., Doms, R.W. & Malim, M.H. Retroviral mRNA nuclear export elements regulate protein function and virion assembly. *Embo J* **23**, 2632-40 (2004).
215. Jin, J. et al. Distinct intracellular trafficking of equine infectious anemia virus and human immunodeficiency virus type 1 Gag during viral assembly and budding revealed by bimolecular fluorescence complementation assays. *J Virol* **81**, 11226-35 (2007).
216. Jin, J., Sturgeon, T., Weisz, O.A., Mothes, W. & Montelaro, R.C. HIV-1 matrix dependent membrane targeting is regulated by Gag mRNA trafficking. *PLoS One* **4**, e6551 (2009).
217. Sherer, N.M., Swanson, C.M., Papaioannou, S. & Malim, M.H. Matrix mediates the functional link between human immunodeficiency virus type 1 RNA nuclear export

elements and the assembly competency of Gag in murine cells. *J Virol* **83**, 8525-35 (2009).

218. Urcuqui-Inchima, S. et al. Production of HIV Particles Is Regulated by Altering Sub-Cellular Localization and Dynamics of Rev Induced by Double-Strand RNA Binding Protein. *PLoS One* **6**, e16686 (2011).