Correction: 13 February 2008

www.sciencemag.org/cgi/content/full/1152725/DC1

Supporting Online Material for

Identification of Host Proteins Required for HIV Infection Through a Functional Genomic Screen

Abraham L. Brass, Derek M. Dykxhoorn, Yair Benita, Nan Yan, Alan Engelman, Ramnik J. Xavier, Judy Lieberman, Stephen J. Elledge*

*To whom correspondence should be addressed. E-mail: selledge@genetics.med.harvard.edu

Published 10 January 2008 on *Science* Express
DOI: 10.1126/science.1152725

The main PDF includes the following:

Other Supporting Online Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/1152725/DC1)

Tables S2 to S5 as zipped archive 1152725S_Tables.zip

Correction: Reference S28 was inadvertently omitted in the originally posted supporting online material.

Supporting Online Material

Materials and Methods

siRNA screen: To identify host factors required for HIV infection, a high-throughput RNAi-based screen was undertaken on an arrayed library containing 21,121 siRNA pools targeting the vast majority of the human genome (Dharmacon si*ARRAY* siRNA Library (Human Genome, G-005000-05), Thermo Fisher Scientific, Lafayette, CO). Part one of the screen: siRNAs were transiently transfected into the TZM-bl cells at a 50 nM final concentration, using a reverse transfection protocol employing 0.45% Oligofectamine (Invitrogen, Carlsbad, CA) in a 384-well format. The Oligofectamine was diluted in Opti-MEM (Invitrogen) and allowed to incubate ten minutes. The lipid solution was then aliquoted into the wells (9 μl/well) using a liquid handing robot. The plates were spun down at 1,000 RPM and the arrayed siRNAs were added robotically, 1.5 μl of a 1 μM stock per well. After a twenty minute incubation, approximately 440 TZM-bl cells were added per well, in 20 ul of Dulbecco's modified minimal essential media (DMEM, Invitrogen), supplemented with 15% fetal bovine serum (FBS, Invitrogen). The plates were next spun at 1000 RPM and then placed in a tissue culture incubator at 37C and 5% CO2. After 72 h of siRNA-mediated gene knockdown, the medium was removed and the cells were treated with HIV-IIIB at an MOI of 0.5 in 100 μl DMEM with 10% FBS. After an additional 48 h incubation (when silencing is still operative), 20 μl of media was removed and replica plated onto a new 384 well plate containing 1800 TZM-bl cells per well (beginning of part two of screen). The "part one" cells were then fixed with 4% Formalin, permeabilized with 0.2% Triton-X100 and stained for p24, using purified anti-HIV-1 p24 (mab-183-H12-5C, generously provided by the NARRRP, Reagent 3537, kindly contributed by Dr. Bruce Chesebro and Kathy Wehrly) and an Alexa 488 goat anti-mouse secondary (A11001) and rabbit anti-goat tertiary (A11078) antibodies (Invitrogen), and for DNA (Hoechst 33342, Invitrogen). Each step was followed by two washes with buffer containing 10 mM Tris pH 7.5, 150 mM NaCl, 2 mM EDTA pH 8, and 1% FBS. The cells were then imaged on an automated Image Express Micro (IXM) microscope (Molecular Devices) at 4X magnification, using two wavelengths, 488 nm to detect HIV infected cells expressing p24, and 350 nm for nuclear DNA bound by Hoechst 33342. Images were then analyzed using the Metamorph Cell Scoring software program (Molecular Devices Inc.) to determine the total cells per well, and the percentage of p24 positive cells in each well (percent infected). A negative control luciferase siRNA (Luc) and positive control siRNA SMARTpools against CD4 and Rab9p40 (Dharmacon) were present on each plate. In addition wells containing either buffer alone, a non-targeting control siRNA (siCONTROL Non-Targeting siRNA #2, Dharmacon), and an siRNA pool directed against Polo like kinase one (PLK1, Dharmacon) were present on all plates transfected. The screen was performed in duplicate.

Part two: To search for host factors whose depletion leads to defects in producing infectious particles, 20 μl of conditioned media containing HIV from each well in the first round screen was removed prior to fixation and transferred to a new well containing uninfected TZM-bl cells. 20 h later these cells were treated with Gal-Screen chemiluminescence reagent (Applied Biosystems, Foster City, CA), and assessed with an Envision 2 plate reader (Perkin Elmer, Waltham, MA) for Tat-dependent transcription of the stably integrated beta-galactosidase reporter gene. These results were normalized to cell number present in the first round donor well, as recorded by the IXM microscope. Control experiments using HeLa-CD4 cells (which do not contain a Tat-dependent reporter gene) in the recipient wells showed that no significant beta-gal activity was transferred along with the supernatant. siRNA pools were classified hits if they decreased the percentage of p24 positive cells or beta-gal light units by two or greater standard deviations (SD) from the plate mean on both of the duplicate plates, and viable cells were not

decreased by greater than two SDs from the mean of the plate. We next performed a validation screen, in which the four individual oligos comprising each pool were placed into separate wells, and screened again using identical methods as above. siRNA pools were considered validated if one or more of the individual oligos scored - in either part one or two; The percent of infected cells relative to controls for each of the siRNA pools that confirmed is provided in Table S2, column E (values were calculated by averaging the data from the duplicate wells transfected with SMARTpools in the first round of screening; Percent infected, p24 for part one, Beta-gal RLU for part two). Visual spot inspections of control images were done throughout the screen to confirm the accuracy of the automated imaging and cell scoring systems. Confirmation siRNA transfection experiments using TZM-bl cells were done as above and scaled proportionally as needed. A previous screen to detect HDFs using siRNAs directed against a smaller group of human genes, synthesized by a different vendor, has been published, however, no gene list was provided regarding any novel host-viral interactions (*S1*).

Cell lines and culture conditions: TZM-bl, HL2/3 and HeLa-CD4 (HeLa-T4+) HeLa cells were generously provided by the NARRRP, and kindly contributed by Dr. John C. Kappes, Dr. Xiaoyun Wu and Tranzyme Inc. (TZM-bl, (*S2*)), Dr. Richard Axel (HeLa-T4+) and Dr. Barbara K. Felber and Dr. George N. Pavlakis (HL 2/3, (*S3*)). HeLa cells were grown in DMEM supplemented with 10% FBS. TZM-bl cells were chosen due to limitations in experimental methods using more relevant T and macrophage cell lines. They proved useful for screening because they are easily transfected with siRNA, are hardy enough to survive high throughput manipulations and support a full HIV lifecycle to produce infectious virions. Jurkat E6-1 cells (ATCC TIB-152) were grown in RPMI-1640, with 10% FBS and 0.1% beta-mercaptoethanol (Invitrogen).

Viral propagation: HIV-IIIB (HTLV-IIIB/H9 from Dr. Robert Gallo, NARRRP Cat. Reagent 398 (*S4*)) was propagated in the T cell line H9 in DMEM supplemented with 10% heat-inactivated fetal calf serum, 2 mM L-glutamine, 50 U/ml of penicillin and 50μg/ml streptomycin. This HIV-IIIB lab-adapted strain is deficient for Nef, Vpu, and codes for a truncated Vpr (frame shift mutation, which results in a truncated Vpr protein missing its C-terminus, see Table S4 and supporting online text below, Genbank accession number for a molecular clone (HXBc2) derived from this lab strain, K03455). It is important to note, that the IIIB strain is more heterogeneous than just the HXBc2 molecular clone, and represents a population of viral genotypes. However, the majority of these types must be inhibited by the siRNAs in Table S2 (supporting online text). The CCR5-tropic HIV-1-$_{Ba-L}$ strain (from Dr. Suzanne Gartner, Dr. Mikulas Popovic, and Dr. Robert Gallo, NARRRP Cat. Reagent 5104) was propagated on human monocyte-derived macrophage cells. Briefly, peripheral blood mononuclear cells were isolated from whole blood obtained from healthy donors by Ficoll-Hypaque (Pharmacia) density centrifugation. The isolated cells were washed extensively in PBS and plated in RPMI containing 10% heat inactivated human AB serum, 2 mM L-glutamine, 50U/ml of penicillin and 50μg/ml streptomycin and plated a $2x10^6$ cells/ml in 24 well plates. Non-adherent cells were removed after 5 days of culture by washing with warm media. Macrophage cells infected at MOI 0.2 were monitored until >90% of the cells were infected. The virus containing supernatant was harvested by centrifugation (1,500xg for 10 min), aliquoted and stored at -80^oC. Viral titers for both HIV-1 strains were determined by treating Magi (IIIB) or Magi-CCR5 (Bal) cells (NARRRP) with increasing amounts of viral supernatant. 48 h post infection the cells were stained for p24 expression.

Plasmids, shRNA and siRNA Reagents: The coding sequence for Rab6A' was PCR-amplified, fully sequence confirmed as correct, and then recombined into a Gateway-compatible entry

vector using BP-clonase (Invitrogen); This insert was then recombined in frame into a N-terminal GFP fusion expression vector with a Blasticidin selectable marker (gift from Jianping Jin, Harvard Medical School) using LR recombinase (Invitrogen), to produce p203-GFP-Rab6. A GFP only version of the expression vector (p203-GFP) was used as control. The HIV-YFP plasmid was previously described and created by replacing the alkaline phosphatase gene (AP) with YFP (Clontech, Mountain View, CA) in pHIV-APΔenvΔvifΔvpr, which was in turn derived from the NL4-3 molecular clone (Accession number M19921) by deleting *vif* and *vpr* (0.62 kb section removed) and 1.45-kb of *env* (*S5-7*). HIV-YFP contains an intact TAR and is Tat-dependent for transcription (Personal communication, Dr. Richard Sutton, Baylor College of Medicine, Houston). The pHAGE-CMV-ZSG plasmid is a derivative of HRST-CMV, and contains self inactivating LTRs, an internal CMV promoter driving expression of a ZSG reporter gene, a rev response element (RRE), and a woodchuck hepatitis post-transcriptional regulatory element (WPRE, gift of A. Balazs and R. C. Mulligan, Harvard Medical School). The MLV-EGFP plasmid contains an MLV-LTR and the humanized form of Renilla green fluorescence protein (Invitrogen) and was kindly provided by F. Diaz-Griffero and J. Sodroski, Harvard Medical School.

The EcoRI site of pMSCV-puro vector, containing the puromycin resistance gene (Invitrogen) was modified to an MluI site to generate pMSCV-PM (pMSCV-Puro-MluI). shRNAs against Rab6A from the second generation Hannon-Elledge shRNA library (*S8*) were subcloned from the SalI/MluI sites of pSM2c into the XhoI/MluI sites of pMSCV-PM to generate pMSCV-PM-shRNA plasmids, amenable to packaging into retroviruses. The following shRNAs were used (kind gifts of Troy Moore, of Open Biosystems Inc.):

Luciferase control (FF) CGCCTGAAGTCTCTGATTAA

shRab6-1 CTCTTTCACATGTGCTTTA Rab6A 3'UTR 1887-1905
shRab6-2 CCTGCTGAATTTATGTTGT Rab6A 3'UTR 2776-2794
shRab6-3 CCATTGGAATTATCCTTTA Rab6A 3'UTR 1642-1660

The following custom siRNA oligonucleotides (Dharmacon) were used in this study:

Luciferase control (Luc) CGTACGCGGAATACTTCGA

HIV-1 Tat CUGCUUGUACCAAUUGCUAUU

All of the following are Dharmacon siRNAs, catalogue numbers are provided, however in the case of the individual duplex oligos these have been subject to change and we suggest following the sequence information given,

All are human-sequence reagents:

CD4 (SMARTpool M-005234-01), Rab9p40 (SMARTpool M-019457-00), PLK1 (M-003290-01)

In Fig. 2H, I and S3 E, F (Jurkat E6-1 cell experiments)

Rab6-1 D-008975-06 CCAAAGAGCUGAAUGUUAUUU

Rab6-2 D-008975-01 GAGAAGAUAUGAUUGACAUUU

Rab6-3 D-008975-04 GAGCAACCAGUCAGUGAAGUU

In Fig.s S4 C, (TZM-bl cell fusion experiments)

Rab6-1 D-008975-01 GAGAAGAUAUGAUUGACAUUU

Rab6-2 D-008975-04 GAGCAACCAGUCAGUGAAGUU (same as Rab6-3 in Fig. 2 (H, I))

Rab6-3 D-008975-05 AAGCAGAGAAGAUAUGAUUUU

Rab6-4 D-008975-06 CCAAAGAGCUGAAUGUUAUUU

TNPO3-1 D-019949-01 GCAGUGAUAUUUAGGCAUAUU

TNPO3-2 D-019949-02 GGAGAUCCUUACAGUGUUAUU

TNPO3-3 D-019949-03 GAAGGGAUGUGUGCAAACAUU

TNPO3-4 D-019949-04 GAGGGUAUCAGACCUGGUAUU

TNPO3-5 J-019949-09 CGACAUUGCAGCUCGUGUAUU

TNPO3-6 J-019949-10 GAGUGAAGUCGUUGAUCGAUU

TNPO3-7 J-019949-11 UCACCAGGUUGUUUCGAUAUU

TNPO3-8 J-019949-12 GUACAAAACUAACGAUGAAUU

Med28-1 D-014606-01 GCGGAAAGAUGCACUAGUCUU

Med28-2 D-014606-02 GUACUUUGGUGGACGAGUUUU

Med28-3 D-014606-03 UGAGUGGGCUGAUGCGUGAUU

Med28-4 D-014606-04 CAGAAACCAGAGCAAGUUAUU

Vps53-1 D-017048-01 GAAAGGAGAUUUAGAUCAAUU

Vps53-2 D-017048-02 GCAAUUAGAUCACGCCAAAUU

Vps53-3 D-017048-03 AGAAGUACCUCCGAGAAUAUU

Vps53-4 D-017048-04 GCGCCGACCUCUUUGUCUAUU

RanBP2L1 and RanBP2 D-012007-03 GAAGUCCUGCAAUUUAUAAUU

HeLa cells were transfected with siRNAs (50 nM final concentration) using Oligofectamine (Invitrogen) according to the manufacturer's protocol. Transfection of DNA plasmids was performed using Exgene-500 (Fermentas Life Sciences, Burlington, Ontario) per the manufacturer's instructions. Efficiency was determined by cotransfection of MSCV-DSred. Jurkat cells (2e6 per reaction) were transfected with 1.2 uM final concentration of siRNA using a Cell line nucleofactor kit V, with program setting T-14, as per the manufacturer's instructions (Amaxa Biosystems, Cologne, Germany). 72 h after transfection the Jurkat cells were infected with HIV-IIIB at an MOI of 0.2, see Flow cytometry section below for analysis.

Retrovirus production and infection: Retroviruses containing MSCV-PM empty vector (mir30), control (FF) or Rab6 shRNAs (shRab6-1, 2, and 3) were produced by transfecting 293T cells with the specific retroviral plasmid, pCG-Gag-Pol, and pCG-VSV-G using TransIT-293 (Mirus) in OptiMEM per manufacturer's instructions. HIV-YFP was produced by transfecting the HIV-YFP plasmid (kindly given by R. E. Sutton, Baylor School of Medicine) with pCG-VSV-G. p203-GFP-Rab6, p203-GFP, and pHAGE-CMV-ZSG viruses were produced by transfecting the respective pHAGE plasmid, along with pHDM.Hgpm2 (a codon optimized HIV-1 $_{NL4-3}$ Gag-Pol), pHDM-VSV-G, pRC1-CMV-Rev1b, and pMD2btat1b (all kind gifts of J. W. Walsh and R. C. Mulligan, Harvard Medical School). MLV-EGFP virus was prepared by cotransfecting pVPack-GP (Stratagene, La Jolla, CA) and pCG-VSV-G. Retroviruses were harvested 48 h after transfection, filtered with a 0.45 μm filter, titered, and stored at –80^{o}C. For generation of the stable shRab6-KD cell lines, TZM-bl cells were infected at an MOI ~3 using 8 μg/ml polybrene (Sigma). The media was replaced 24 h after infection, and the transduced cell populations were selected with puromycin (Invitrogen) 48 h after infection. To rescue the shRab6-KD cell lines, cells were infected with either p203-GFP-Rab6 or p203-GFP, and 48 h later populations of cells were put under blasticidin selection.

Western Analysis: Whole-cell extracts were prepared by cell lysis, equivalent protein content boiled in SDS sample buffer, resolved by SDS/PAGE, transferred to Immobilon–P membrane (Millipore), and probed with the indicated antibodies. Rabbit anti-Rab6 was from Santa Cruz Biotechnology (C-19, sc-310); mouse monoclonal anti-Med28 7E1 was a very kind gift from Dr. Vijaya Ramesh, (Massachusetts General Hospital).

Quantitative PCR: Total RNA was extracted using an RNeasy Plus RNA isolation kit (Qiagen, Valencia CA). cDNA was generated using a Quantitect Reverse Transcription kit (Qiagen). Specific cDNAs were quantitated by quantitative PCR with the primer combinations listed below, using a QuantiTect SYBR Green PCR Kit (Qiagen) on an ABI 7500 Real Time PCR system following the manufacturer's instruction (Applied Biosystems). Primers were designed using the Roche Applied Science Universal Probe Library web site (Roche, Indianapolis, IN). PCR parameters consisted of 1 cycle of 50°C X 30 s, then 94 °C X 15 s, followed by 40 cycles of PCR at 95 °C X 15 s, 56°C X 30 s, and 72 °C X 30 s. The relative amount of target gene mRNA was normalized to *GAPDH* mRNA. Specificity was verified by melt curve analysis and agarose gel electrophoresis.

Primer sequences

GAPDH 5' GGAGCCAAACGGGTCATCATCTC
GAPDH 3' GAGGGGCCATCCACAGTCTTCT
TNPO3 5' CCTGGAAGGGATGTGTGC
TNPO3 3' AAAAAGGCAAAGAAGTCACATCA
Vps53 5' GAACTGTTGTAAGAGGTCAGACGA

Vps53 3' CACGGGTGATTTCTTTCACTT

HIV Late-RT and integration analyses: HeLa-CD4 cells were transfected with siRNAs as described above. Cells were infected with HIV-IIIB 72 h after transfection, and DNA was extracted using the Hirt method at both 7 h post-infection (hpi) and 24 hpi. Late RT products and integrated HIV DNA were analyzed as described (*S9, 10*). Briefly, Late RT products in extrachromosomal DNA fractions at 7 hpi were analyzed by real-time PCR using MH531/MH532 primers (*S10*). Integrated HIV DNA at 24 hpi was measured by Alu-PCR followed by nested real-time PCR using AE989/AE990 primers (S*9*).

Cell Fusion Assay: TZM-bl shRab6 target cells, were plated in 96-well plates, at 20,000 cells per well, and cultured overnight. The following morning, the media was removed and 15,000 HL2/3 cells were added to each well in fresh media. The co-culture was then incubated at 37C for 6 hours to allow fusion to occur. Fusion was monitored by assaying for Tat-dependent beta-gal reporter gene activation stimulated by HIV-1 Tat from the HL2/3 cells. TZM-bl cells alone were used to determine background luminescence. For cell fusion experiments using siRNA transfected cells, TZM-bl cells were transfected as noted above, and after a 72 h knockdown, the HL2/3 cells were added in fresh media.

Flow Cytometry: To assess coreceptor levels, TZM-bl or Jurkat cells were harvested with cell dissociation enzyme-free PBS-based buffer (Invitrogen), washed and then stained with the indicated antibodies: Mouse monoclonal anti-Human CD4, clone 13B8.2, conjugated with PE or FITC (Beckman Coulter, Fullerton CA), mouse monoclonal anti-Human CXCR4 (CD184), conjugated with PE (BD Biosciences, Franklin Lakes, NJ), mouse monoclonal anti-Human CCR5 (CD195) clone 2D7, (BD Biosciences), or mouse isotype matched PE, FITC or APC-conjugated control antibodies. To determine levels of HIV infection in Jurkat cells, the cells were fixed and permeabilized (Fix and Perm Kit, Invitrogen), then incubated with mouse anti-HIV-1 p24-PE antibody (KC57-PE, Beckman Coulter) or a mouse isotype matched PE control antibody. Fluorescence intensity was analyzed by flow cytometry of 10,000 events (BD LSR II; Beckman Coulter). For cell cycle analysis, collected cells were resuspended in 100 μl (PBS). While vortexing, 2 ml of ice cold 70% (v/v) ethanol were added drop-wise and the suspension was stored at 4C at least overnight. 30 min before FACS, cells were spun down, resuspended in propidium iodine (PI) mix (500 μl PBS, 10 μl RNase [of stock solution of 10 mg/ml], 25 μl PI [of stock solution of 1 mg/ml]).

Bioinformatics Analyses:

Gene Ontology (GO): GO terms and gene annotations were obtained from the gene ontology web site (www.geneontology.org; ontologies revision: 5.508; gene associations revision: Oct/8/2007). Uniprot and VEGA gene identifiers were mapped to NCBI gene identifiers. In cases where multiple IDs matched the same NCBI gene, all GO terms from these IDs were combined and assigned to the NCBI gene.

All GO terms assigned to genes that scored positive in the screen were obtained and tested for over-representation using a hypergeometric distribution as described in the GOHyperGAll module of bioconductor (S*11*). Gene ontology processes high in the tree hierarchy are often too general and contribute little information. For enrichment analysis purposes, we selected biological terms that were assigned to less than 500 human genes. This cutoff was obtained empirically by looking at central nodes in the gene ontology tree.

Biological process: The GO vocabulary is arranged in a tree structure with a single root node. To simplify the representation of terms, terms which were significantly over-represented with a p-value < 0.05 and connected in the tree hierarchy were combined to form an over-represented cluster of connected terms. All the genes annotated within that cluster of terms were represented by the most significant term in the cluster. Gene ontology clusters were ordered based on p-values and if the genes in one cluster were fully contained within another the dominant cluster was accepted. We accepted clusters that contained at least 3 genes.

Molecular function: GO terms for the molecular function category were processed as described above for biological process. However, no clustering of terms based on tree connectivity was performed for this category.

Subcellular localization: The subcellular location of each gene was manually curated based on annotations from Swissprot (*S12*) and Gene Ontology (*S13*). Prediction tools were applied for genes with no annotations. Phobius was used to predict trans-membrane domains (*S14*) ; Maestro to predict mitochondria proteins (*S15*) and TargetP to predict secreted and mitochondria proteins (*S16*).

Microarrays: Gene expression profiles across 79 tissues were obtained from the GNF consortium. Expression profiles from Affymetrix U133A platform and GNF custom probes were used. Expression profiles were normalized using the GCRMA method as implemented in bioconductor (*S17, 18*). Affymetrix MAS5 module of bioconductor was used to identify present or absent transcripts (*S17*) and probes with no single present call across all tissue or highest expression value below log2(100) were removed. Using this approach, we identified 36,549 probes expressed in the GNF dataset in at least one tissue. All calculation and heatmaps were generated based on the set of expressed probes only. Expression profiles were clustered using Cluster 3 (*S19*) and visualized using JavaTreeView (*S20*).

For the purpose of visualization and clustering, a single probe with the largest expression range across all tissues was selected for genes with multiple probes and replicates were collapsed to the average expression value for each probe.

We define immune enrichment as a significantly ($p<0.05$) higher expression level in immune tissues compared to all other tissues. Expression levels were evaluated for each probe separately and enrichment was calculated with the program R (version 2.5). Of the 79 tissues in GNF 22 tissue were classified as immune tissues (Table S3) and for each probe the expression values in immune tissues were tested against expression values in all other tissues. The Wilcoxon rank sum test with two samples (immune vs others) was used to obtain a p-value under the alternative hypothesis that the overall expression in immune tissues is greater than in other tissues. The p-values for all probes were adjusted for multiple testing using the Bonferroni method (*S21*). Similarly, brain enrichment was calculated for each probe by comparing expression in brain tissues to all other tissues. Tissue classification to immune, brain and others is listed in Table S3.

To test the probability of finding a number N of genes enriched in a subset of tissues in GNF two methods were used. First, the tissues were randomly shuffled and the enrichment analysis as described above was performed. This process was iterated 10,000 times and a p-value was obtained by dividing the number times N number of genes or more had a Wilcoxon p-value < 0.05 by 10,000. Secondly, we randomly selected the same number of probes (239 probes for unique genes) as in our subset of genes from the entire GNF microarray and performed an enrichment analysis. Again, this process was iterated 10,000 times and a p-value was obtained by dividing the number of times N genes or more were enriched in immune tissues by 10,000.The higher of the two p-values was accepted as the final p-value.

HIV life cycle map: To place genes in the HIV life cycle map annotation information from

UniProt (*S22*), Gene Ontology (*S13*), NCBI Reference Sequence (*S23*), NCBI OMIM database (*S24* and NCBI GeneRIF (*S25*) was obtained. For each gene the homolog proteins in other species were identified using NCBI HomoloGene (*S25)* and annotation information for the homolog was obtained from NCBI Reference Sequence. This information was integrated into a database framework that summarized the current knowledge for each gene. Literature mining was performed for each gene using an "enhanced" Google and PubMed query via iHop (*S26, 27*). Current knowledge of HDFs and their role in the HIV life cycle was mapped based on reviews (*1-3*). The schematized viral lifecycle was used as an anchor point for manually adding novel and known HDFs. These HDFs were placed on the map based on their cellular function inferred from literature curation and database review. All references (given as Pub Med ID (PMID) number format) and rationale for assignment of 150 genes in the HIV life cycle are available in Table S4.

HIV connectivity network (Figure S7):

The NCBI HIV interaction database contains functional interactions of HIV proteins with human proteins (*S25*). To link HDFs recovered in this study to those in the HIV database, we used NCBI protein-protein interaction data. As a second layer, we also used protein interaction data from other species. NCBI HomoloGene (*S25*) was employed to identify homolog proteins between humans and other species. To uncover protein-protein interactions that are the least likely to occur by chance, we used the hypergeometric distribution without replacement. No correction for multiple hypothesis testing was performed. We obtained a p-value for each gene using R version 2.5 with the following command: phyper(x-1, m, n-m , k, lower.tail = FALSE), where x is the number of HDFs that have a protein interaction with a candidate gene; m, the total number of binding partners for that gene; n total number of human genes that have at least one binding partner; and k, number of HDFs that have at least one binding partners. Proteins with $p < 0.05$ which have an entry in the NCBI HIV interaction database where used to link HDFs to HIV.

Supporting Online Text:

The host ESCRT machinery has been shown to be vital for HIV budding. Of the 28 host proteins published to be involved in this pathway we recovered only one, HRS. Review of our primary screen data revealed that only siRNAs against two more of these factors, Vps4A and 4B, resulted in extensive cell death. LEDGF, a well confirmed HDF important for integration, was not detected in this screen, likely because its intracellular levels greatly exceed those required by the virus (*S9, S28*). However, given the many additional factors, other than insufficient knockdown of the target, involved in producing false negative results (functional redundancy, poor siRNA design, essential gene, off-target toxicities, HIV strain deficient in accessory proteins (please see below), and operator error), we await the results of future improved genetic screens for insights into these and other portions of the viral lifecycle. Furthermore, host factors that might affect the immune response to HIV would likely be missed in this cell-based screen.

As noted above, the HIV-IIIB lab strain used in this study is deficient in Nef, Vpu and a molecular clone derived from this strain (HXBc2) contains a frame shift mutation which codes for a truncated Vpr protein. The predicted HXBc2 Vpr open reading frame would produce a 78 aa protein (wild-type 96 aa full length), with the first 72 residues identical to the NL4-3 wild-type Vpr protein and 6 additional amino acids, from 73-78, encoded by the shifted reading frame (*S29*). This truncated Vpr is missing the six most C-terminal amino acids contained in a previously described deletion mutant, Vpr Δ78-87, which was demonstrated to maintain its interaction with the host factor, VPRBP (S29). A conserved interaction domain Vpr aa 60-78 was defined (underlined below, based on homology to the viral sequence stated in the reference as being amplified from HIV-1/89.6 (*S29, 30*)). A truncated Vpr protein containing aa 1-84 was expressed in 293T cells, but unlike the wild-type Vpr, this mutant was unable to induce a G2 cell cycle arrest (*S31*). Therefore, while the HIV-IIIB Vpr protein may exist at low levels during infection, it is unlikely to mediate its effect by inducing a G2 cell cycle arrest via interactions with VPRBP.

NL4-3 (Genbank acc. no. M19921) Vpr:

MEQAPEDQGPQREPYNEWTLELLEELKSEAVRHFPRIWLHNLGQHIYETY
GDTWAGVEAIIRILQQLLFIHFRIGCRHSRIGVTRQRRARNGASRS

HXBc2 (HIV-IIIB, Genbank acc. no. K03455) Vpr:

MEQAPEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQHIYETY
GDTWAGVEAIIRILQQLLFIHFQNWVST (frame shift creates residues in red)

As discussed in the main text, glycosphingolipids (GSLs) are required for HIV fusion, possibly through direct interaction with HIV gp120. Reducing levels of the GSLs, Gb3 or GM3, inhibits HIV fusion with primary T cells (main text 29-31). In support for the hypothesis that targeting of Rab6 and Vp53 may alter GSL homeostasis and thus HIV infection, we find that viral infection is decreased by depletion of the enzymes that synthesize Gb3 and GM3, A4GALT and SIAT9, respectively (Table S2, Figure 5, and Table S4). Other components of the GSL synthesis pathway found by the screen include a recently identified GSL-transfer protein, FAPP1 (S3*1S32, S3233*) and the small GTPase, ARF1, which targets FAPPs to the Golgi (Fig. 5, Table S4).

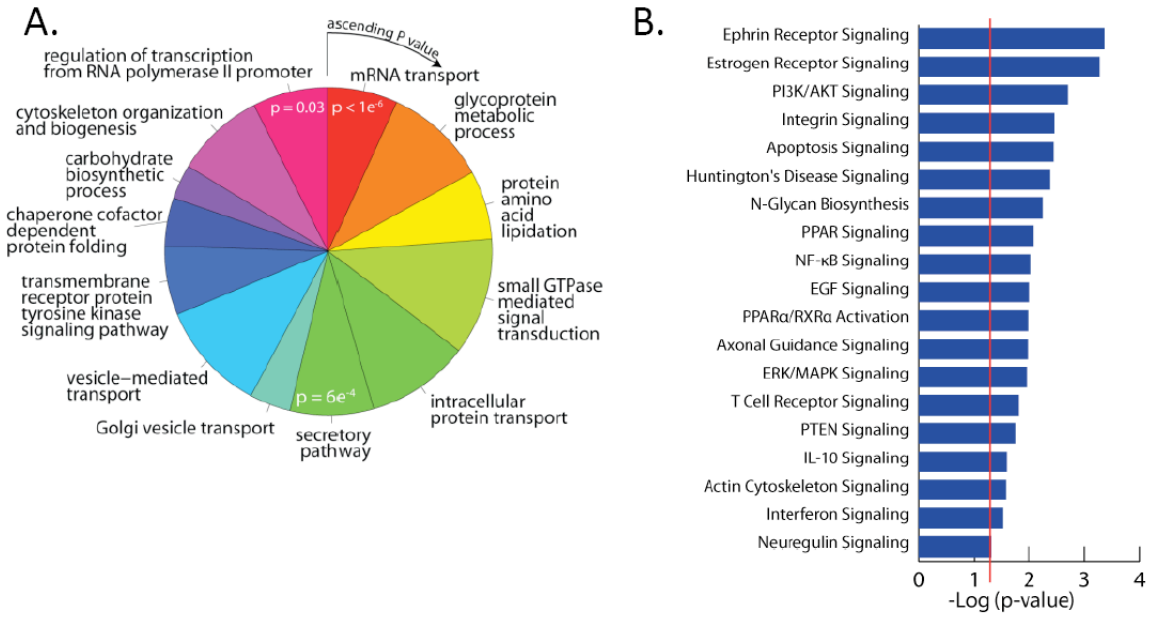

Fig. S1. Bioinformatics analysis. (**A**) Gene ontology biological process. Of the 274 identified genes, 103 were assigned with 136 statistically significant ($p < 0.05$) biological processes. Gene ontology terms were clustered (see methods) and curated manually (Table S3), resulting in 13 processes. The biological processes are ordered clockwise with ascending p-values. (**B)** Pathway enrichment analysis obtained from the Ingenuity program using the right-tailed Fisher's exact test. Threshold is at 1.3 = –log (P = 0.05). Genes assigned by Ingenuity to each pathway are available in Table S3.

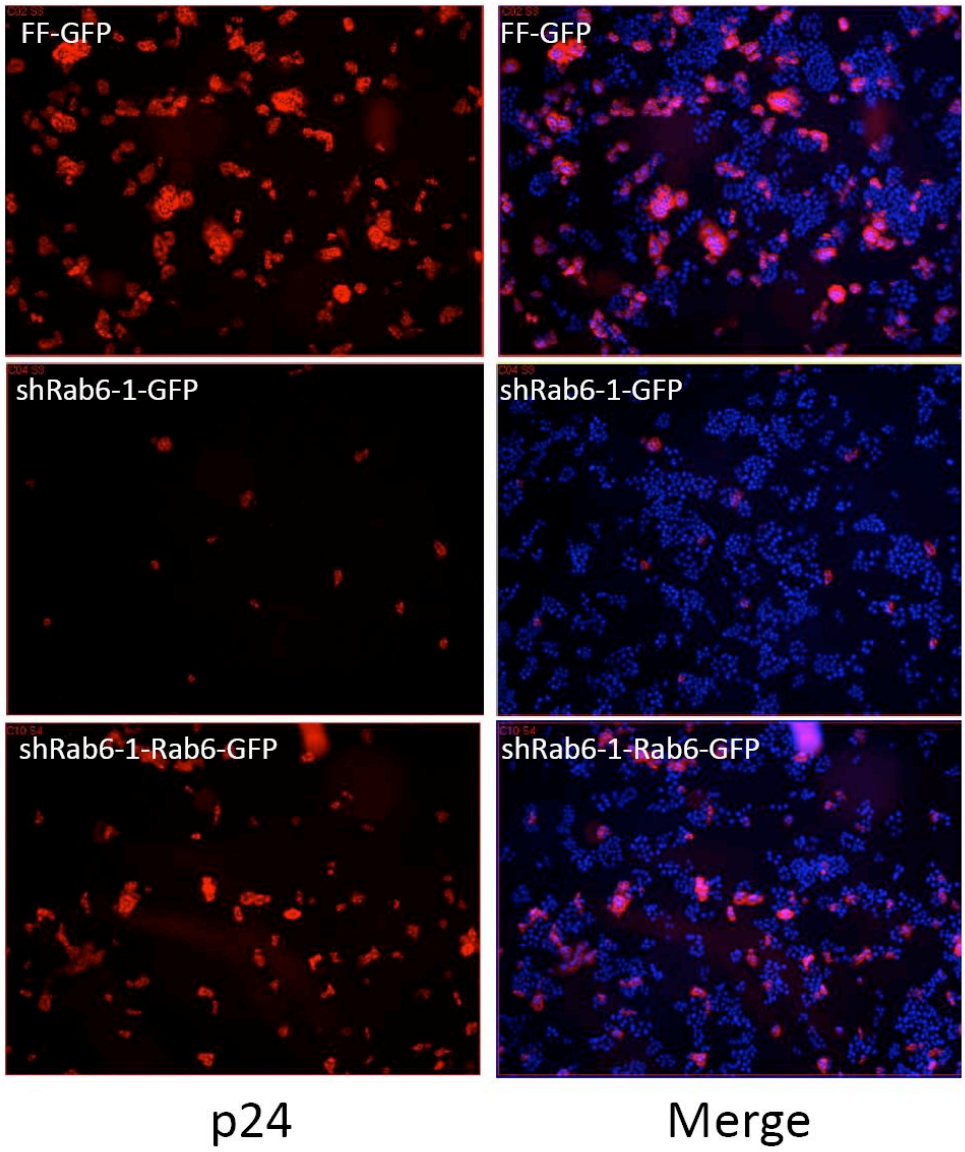

Fig. S2. Rab6 depletion inhibits HIV infection. TZM-bl HeLa cells stably expressing an shRNA against firefly luciferase (FF), or against Rab6 (shRab6-1), and either the control green fluorescence protein (GFP) or the rescue construct containing a Rab6-GFP fusion (Rab6-GFP), were infected with HIV-IIIB. The infection was monitored by staining for p24 expression at 48 h post infection (red). A merged image depicting both p24 and DNA staining (blue) is shown at right (Merge). 4X magnification.

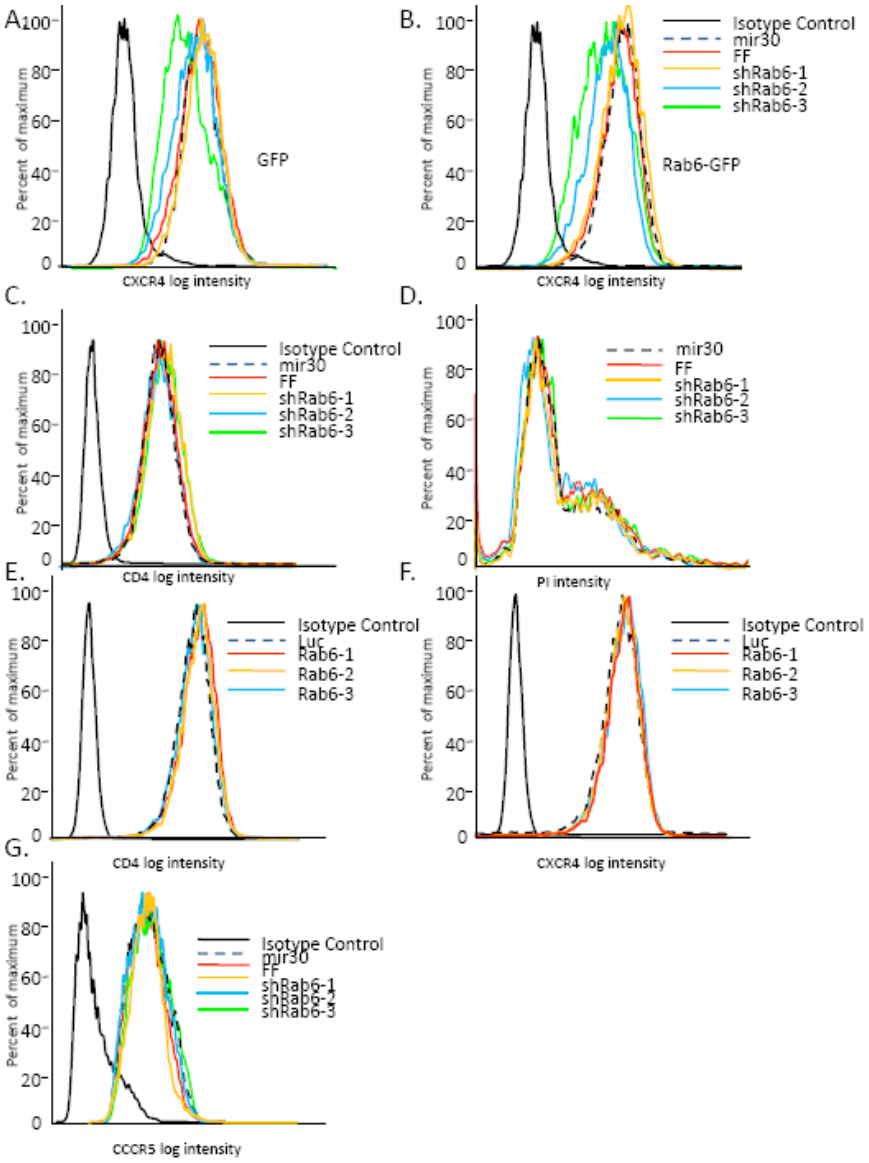

Fig. S3. (A, B) Rab6 depletion does not significantly change CXCR4 levels. TZM-bl HeLa cells stably expressing the indicated shRNAs, and either the control green fluorescence protein (GFP, A) or a Rab6-GFP fusion (Rab6-GFP, B), were stained with anti-CXCR4-phycoerythrin (PE) conjugated antibody, or an Isotype matched-PE control, and analyzed by FACS. (C) Rab6 depletion does not significantly change CD4 levels. Same as in (A, B), but cells were stained with anti-CD4-PE antibody, or an Isotype matched-PE control, and analyzed by FACS. (D) Rab6 depletion does not visibly alter the cell cycle profile. Asynchronous cultures of TZM-bl HeLa cells stably expressing the indicated shRNAs were analyzed by FACS after prodium iodide staining. (E, F) Rab6 depletion does not significantly change receptor levels in Jurkat cells. Jurkat cells transiently transfected with the indicated siRNAs were stained with anti-CD4-PE (E), anti-CXCR4-PE (F) , or an Isotype matched-PE control antibody, 72 h. after transfection, and analyzed by FACS. (G) Rab6 depletion does not alter CCR5 levels. As in (A, B), but cells were stained with anti-CCR5-APC conjugated antibody, or an Isotype matched-APC control, and analyzed by FACS.

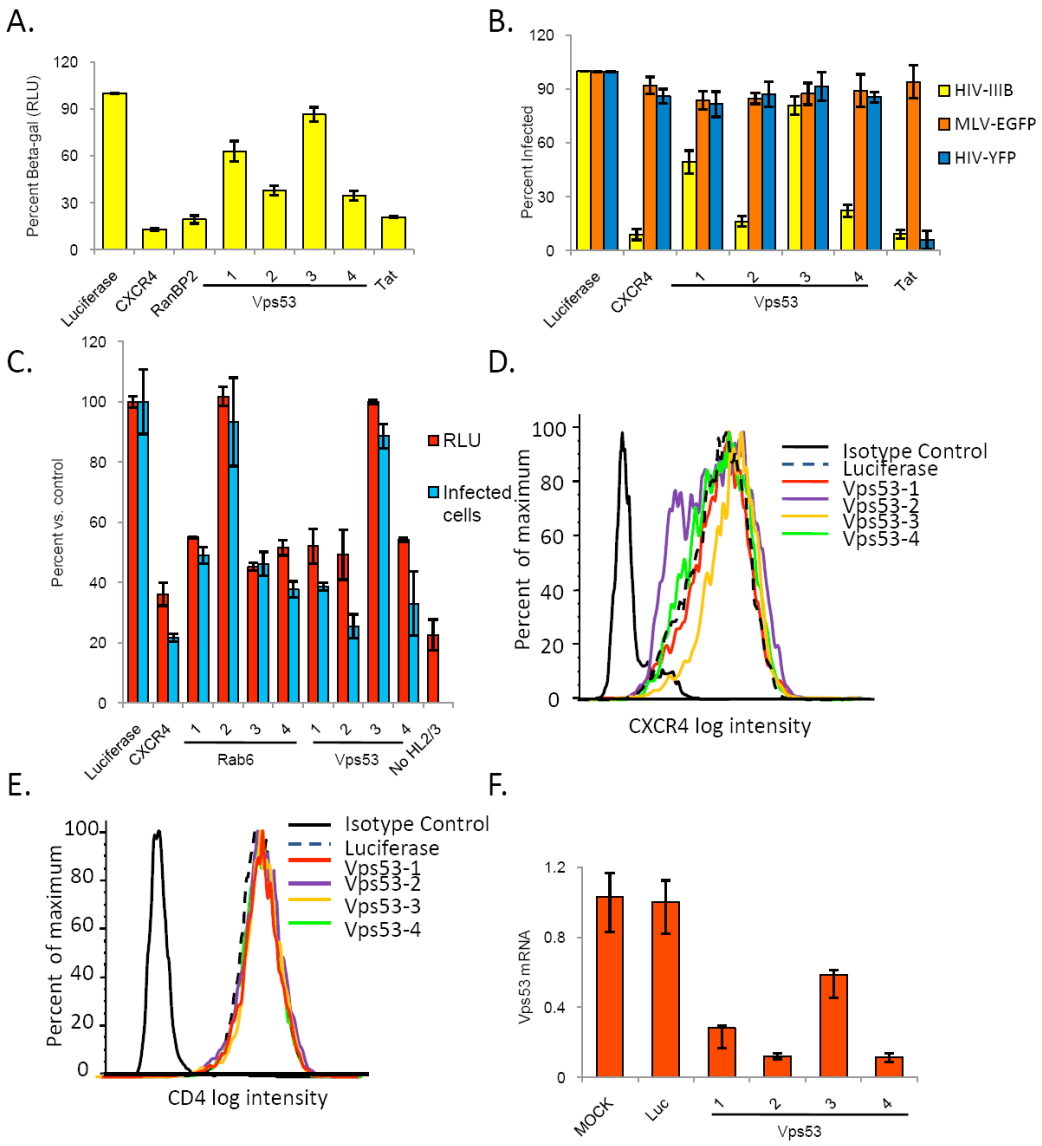

Fig. S4. Targeting of Vps53 inhibits HIV. (A) TZM-bl cells received the noted siRNA treatment. 72 h later these cells were infected with HIV-IIIB; After 20 h of infection, the cells were analyzed for level of Tat activity by determining beta-gal expression in cell lysate. Error bars represent standard deviation of the mean (SD), N=3, throughout. (B) Vps53 depletion inhibits only native-enveloped HIV and not the VSV-G pseudotyped HIV-YFP or MLV-EGFP viruses. TZM-bl HeLa cells were transiently transfected with the indicated siRNAs, and then infected with HIV-IIIB, MLV-EGFP, or HIV-YFP 72 h post transfection. HIV infection was monitored with IF staining for p24 (HIV), or the respective reporter genes at 48 h post infection. (C) Decreased Vps53 or Rab6 levels prevent cell fusion. TZM-bl cells were transfected with the noted siRNAs. 72 h later these transfected cells were layered with HL2/3 cells. The co-culture was then incubated for 6 h. at 37C. This permits fusion between the two cells lines to occur. The relative amount of cell fusion is then quantitated by lysing the cells and determining Tat-dependent beta-gal activity (red bars. Relative light units (RLU)). To illustrate the similarities in

the fusion defect and resistance to HIV infection conferred after siRNA transfection, we have shown the percentage of cells infected vs. controls at 48 h after HIV exposure based on p24 IF levels (blue bars). (D, E) Vps53 depletion does not significantly change host receptor levels. TZM-bl HeLa cells treated with the listed sRNAs against Vps53 (1-4) or Luciferase, after 72 h, the cells were stained with anti-CXCR4-PE (D), anti-CD4-PE conjugated antibody (E), or an isotype matched-PE control antibody, and analyzed by FACS. (F) Vps53 mRNA reduction by siRNAs. TZM-bl cells were transfected with the indicated siRNAs for 72 h, then cDNA was prepared and Vps53 expression levels were measured by qPCR.

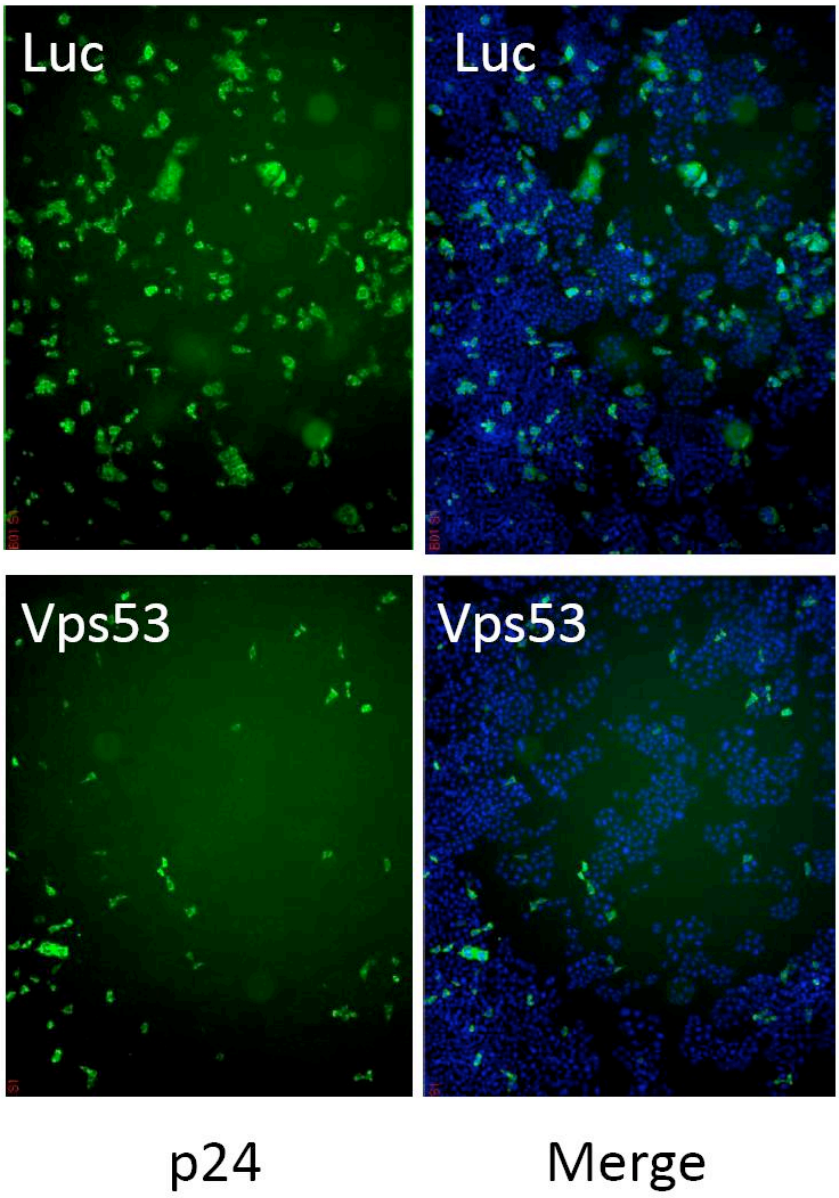

Fig. S5. Vps53 loss curtails HIV infection. TZM-bl cells were treated as described in S4A, with either luciferase (Luc) negative control siRNA or Vps53, siRNA #2 targeting Vps53. Merge denotes combined image for nuclei staining (blue) and anti-p24 HIV Gag protein (green), 48 h after HIV infection. 4X magnification.

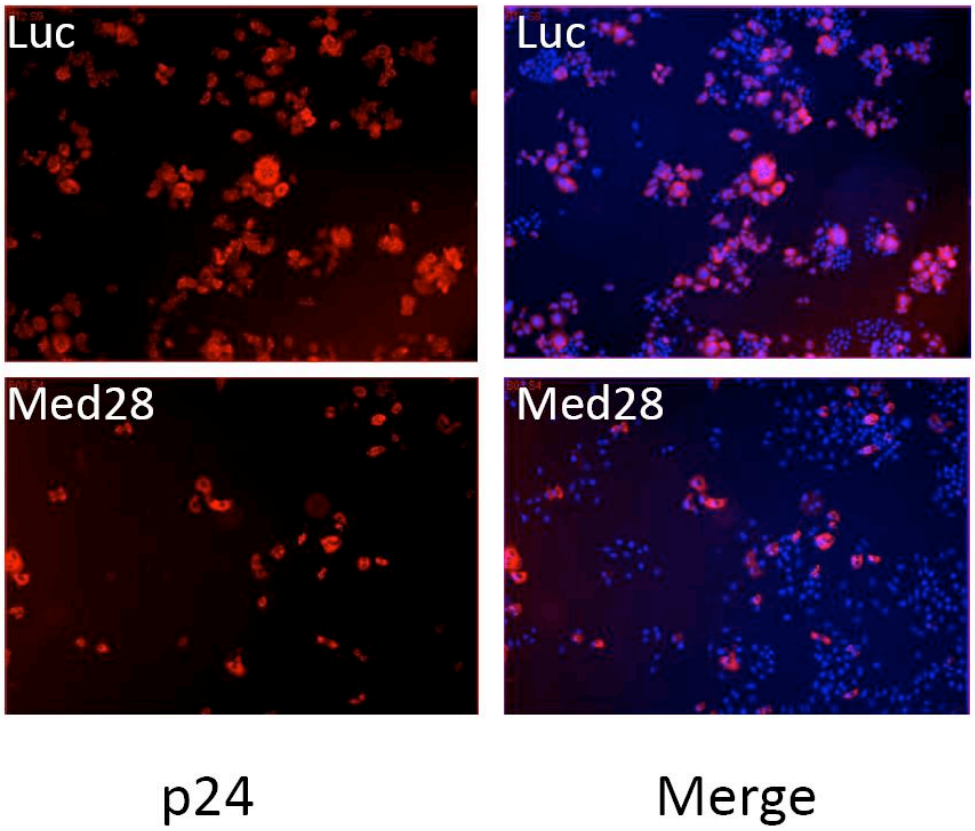

Fig. S6. Med28-depleted cells resist HIV infection. TZM-bl cells were infected with HIV 72 h after siRNA transfection, with either luciferase (Luc) control siRNA, or an siRNA against Med28 (Med28 #2). 48 h after exposure to HIV, the cells were stained for HIV p24 (red), and DNA (blue). Merged p24 and DNA images at right. 4X magnification.

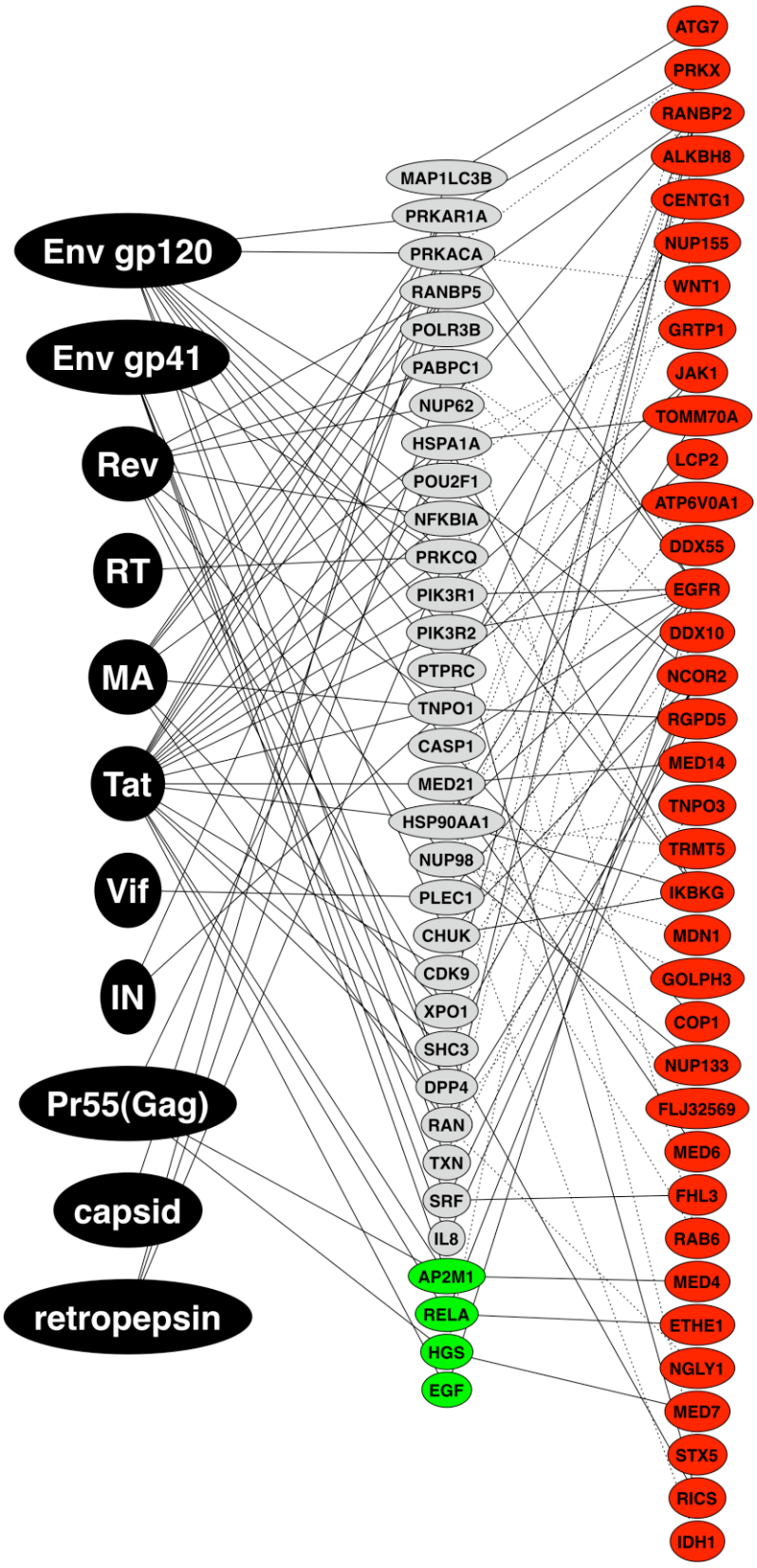

Fig. S7. HIV-HDF connectivity network. To detect interactions between HDFs found in this screen (red) and HIV proteins (black), we identified HDFs that can be linked by protein-protein

interactions to a protein with an entry in the NCBI HIV interaction database (gray (not found in screen) or green (found in screen)). We used human protein-protein interactions (solid line) as well as protein-protein interaction in other species (dotted line) that have human homologs in the NCBI HomoloGene database. To uncover protein-protein interactions that are the less likely to occur by chance, we used the hypergeometric distribution without replacement method, and obtained a p-value for each protein-protein interaction (see methods and Table S5 for analysis results). Only statistically significant protein interactions ($p < 0.05$) are shown here. The nature of the interaction with HIV proteins is listed in Table S5. As noted, HIV-IIIB is deficient in Nef, Vpu and Vpr (truncated protein), and so these HIV proteins were not included in this analysis.

Supplemental Tables

Host Proteins Previously Implicated in HIV Pathogenesis Recovered From siRNA Screen[1,2]		
A4GALT (2/4 siRNAs)	ERCC3 (3)	PSME2 (1)
AKT1 (2)	FBXW11 (4)	PURA (2)
AP2M1 (1)	GCN5L2 (1)	Rab9p40 (3)
Arf1 (2)	H3F3A (1)	RANBP1 (SP)
CD4 (2)	HRS (SP)	RANBP2 (2)
CRTC2 (1)	HTATSF1 (1)	RelA (4)
CRTC3 (3)	IKBG (2)	SIP1 (1)
CTDP1 (1)	La Autoantigen (SSB, SP)	ST3GAL5 (1)
CXCR4 (2)	MAP4 (3)	TFAP4 (SP)
CyclinT1 (SP)	NMT1 (3)	TFE3 (2)
CYCLOB (PPIB, 2)	Nup153(2)	VPRBP (1)
DDX3 (SP)	FAPP1 (1)	ZNRD1 (2)

[1] Numbers in parentheses indicate individual siRNAs out of a total of four possible, that scored on retesting. [2] Evidence provided in Table S4. SP; SMARTpool scored, since the four oligos in the pool were not individually tested.

Table S1. Host proteins previously implicated in HIV pathogenesis recovered from the siRNA screen. 36 genes were classified as previously implicated in HIV biology based on published evidence and/or inclusion in the HIV interaction data base (NCBI). Evidence provided in Table S4.

Table S2. HIV dependency genes. A list of HDFs detected in the screen, and their annotation in UniProt, NCBI Reference Sequence, NCBI HomoloGene and Gene Ontology. The number of individual siRNAs that scored in either part one (column B) or just in part two of the screen (column C) are shown, based on decreasing HIV infection by 2 SD or greater from the mean. Genes that scored with two or more siRNAs in part two only, contain a "Yes" in column D. The percent of infected cells (given as cells positive for p24 for part one, or as Beta-gal RLUs for part two) relative to controls for each of the HDFs is provided in column E. Gene names, synonyms, description and genomic location were obtained from NCBI Reference Sequence (Revision October 2007). UniProt accession numbers were mapped to NCBI Gene IDs by accession numbers provided in UniProt cross-reference file. Gene ontology annotations (Revision October 2007) were obtained from the Gene Ontology Consortium (www.geneontology.org) and mapped to NCBI GeneIDs HIV interactions and their references were obtained from NCBI HIV interaction database. Genes that scored in the initial screen, but whose siRNAs were not rescreened separately, are designated as scoring in part one or part two by a SP (SMARTpool) designation appearing in the appropriate column.

Table S3. Gene specific information used to generate charts in Fig. 1 and Fig. S1.
Fig. 1D: Subcellular location. The subcellular location of each protein was manually curated based on annotation data from Table S2. Fig. S1A: Gene ontology biological processes.

Statistically significant processes were identified and clustered to reduce redundancy. The upper part of the table shows processes that were clustered together based on child-parent relationship in the gene ontology tree. Each cluster of processes is represented in Fig. S1A by a single process. The lower part of the table indicates statistically significant processes that could not be clustered with other adjacent nodes in the Gene ontology tree. For these processes redundancy was reduced by selecting the primary term (lowest p-value) for the same group of genes. Fig. 1E: Gene ontology molecular function. Statistically significant molecular functions were identified and clustered based on the group of genes assigned for each term. A single molecular function was selected to represent each cluster. Fig. S1B: Pathway enrichment analysis obtained from the Ingenuity program. Fig. 1F: Tissue and gene names of GNF heatmap. Tissues labeled from 1 to 79 correspond to the tissues shown from left to right and gene names are arranged in the same order as in the figure. Immune enrichment p-values were obtained by performing a Wilcoxon rank sum test comparing the expression level in immune tissues to expression levels in all other tissues (see methods). Genes with significantly higher expression in immune tissues are indicated in the figure by black bars and are shown here in yellow.

Table S4. Rationale and references for placing genes in HIV life cycle shown in Fig. 5. Based on annotation data from Table S2 and literature review (see methods), each gene was placed based on scientific text mining in the most likely position(s) for HIV dependency.

Table S5. HIV connectivity table for Fig. S7. The HIV interaction database was obtained from NCBI (Dec. 2007 (S*25*)). This database lists functional as well as binding interactions with human proteins. To uncover potential connections between HIV proteins and HDFs with no entry in the HIV database, protein-protein interaction from NCBI were used to link these HDFs to proteins that have an entry in the database (see methods). Human homologs were identified using NCBI HomoloGene and the organisms in which the protein-protein interactions were identified are indicated in the table.

Supplemetary References

S1. D.G. Nguyen, K.C. Wolff, H. Yin, J.S. Caldwell, K.L. Kuhen. *J Virol.* 80, 130 (Jan 2006)
S2. X. Wei *et al.*, *Antimicrob Agents Chemother* 46, 1896 (Jun, 2002).
S3. V. Ciminale, B. K. Felber, M. Campbell, G. N. Pavlakis, *AIDS Res Hum Retroviruses* 6, 1281 (Nov, 1990).
S4. L. Ratner *et al.*, *Nature* 313, 277 (1985)
S5. J. He, N. R. Landau, *J Virol* 69, 4587 (Jul, 1995).
S6. R. Schroers *et al.*, *Mol Ther* 1, 171 (Feb, 2000).
S7. R. E. Sutton, H. T. Wu, R. Rigg, E. Bohnlein, P. O. Brown, *J Virol* 72, 5781 (Jul, 1998).
S8. J. M. Silva *et al.*, *Nat. Genet.* 37, 1281 (2005).
S9. M. C. Shun *et al.*, *Genes Dev* 21, 1767 (Jul 15, 2007).
S10. S. L. Butler, M. S. Hansen, F. D. Bushman, *Nat Med* 7, 631 (May, 2001).
S11. R. C. Gentleman *et al.*, *Genome Biol* 5, R80 (2004).
S12. A. Bairoch, B. Boeckmann, S. Ferro, E. Gasteiger, *Brief Bioinform* 5, 39 (Mar, 2004).
S13. M. Ashburner *et al.*, *Nat Genet* 25, 25 (May, 2000).
S14. L. Kall, A. Krogh, E. L. Sonnhammer, *J Mol Biol* 338, 1027 (May 14, 2004).
S15. S. Calvo *et al.*, *Nat Genet* 38, 576 (May, 2006).
S16. O. Emanuelsson, S. Brunak, G. von Heijne, H. Nielsen, *Nat Protoc* 2, 953 (2007).

S17. R. C. Gentleman et al., Genome Biol 5, R80 (2004).
S18. Z. Wu *et al*., *Journal of the American Statistical Association* 99, 909 (2004).
S19. M. J. de Hoon *et al*., *Bioinformatics* 20, 1453 (2004).
S20. A. J. Saldanha, *Bioinformatics* 20, 3246 (2004).
S21. J. M. Bland et al., *BMJ* 310, 170 (1995).
S22. A. Bairoch *et al*., *Nucleic Acids Res* 33, D154 (2005).
S23. K. D. Pruitt *et al.*, *Nucleic Acids Res* 33, D501 (2005).
S24. V. A. McKusick, *Am J Hum Genet* 80, 588 (2007).
S25. D. L. Wheeler *et al*., *Nucleic Acids Res* (2007).
S26. R. Hoffmann, A. Valencia, *Nat Genet* 36, 664 (2004).
S27. R. Hoffmann, A. Valencia, *Bioinformatics* 21 Suppl 2, ii252 (2005).
S28. M. Llano *et al., Science* 314, 461 (Oct, 2006)
S29. L. Zhao, S. Mukherjee, O. Narayan, *J Biol Chem* 269, 15577 (1994).
S30. R. Collman et al., *J Virol* 66, 7517 (1992).
S31. P. Marzio, S. Choe, M. Ebright, R. Knoblaugh, N. R. Landau*., J Virol* 69, 7909 (1995).
S32. G. D'Angelo *et al.*, *Nature* **449**, 62 (Sep 6, 2007).
S33. A. Godi *et al.*, *Nat Cell Biol* **6**, 393 (May, 2004).