(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 August 2012 (02.08.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/103081 A1

(51) International Patent Classification:
A61K 39/12 (2006.01)

(21) International Application Number:
PCT/US2012/022349

(22) International Filing Date:
24 January 2012 (24.01.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/435,858 25 January 2011 (25.01.2011) US

(71) Applicants *(for all designated States except US)*: **ALBERT EINSTEIN COLLEGE OF MEDICINE OF YESHIVA UNIVERSITY** [US/US]; 1300 Morris Park Avenue, Bronx, NY 10461 (US). **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US). **WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH** [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHANDRAN, Kartik** [IN/US]; 129 Boerum Place, Apt. 2H, Brooklyn, NY 11201 (US). **WHELAN, Sean** [GB/US]; 236 Broadway, Cambridge, MA 02139 (US). **BRUMMELKAMP, Thijn** [NL/NL]; Stolwijkstraat 60-III, NL-1059 XZ Amsterdam (NL). **CARETTE, Jan** [NL/US]; 751 Clark Way, Palo Alto, CA 94304 (US). **RAABEN, Matthijs** [NL/US]; 76 Myrtle Street, Apt. No. 1, Boston, MA 02114 (US).

(74) Agents: **MILLER, Alan D.** et al.; Amster, Rothstein & Ebenstein LLP, 90 Park Avenue, New York, NY 10016 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD,

[Continued on next page]

(54) Title: METHODS AND ASSAYS FOR TREATING FILOVIRIDAE INFECTIONS

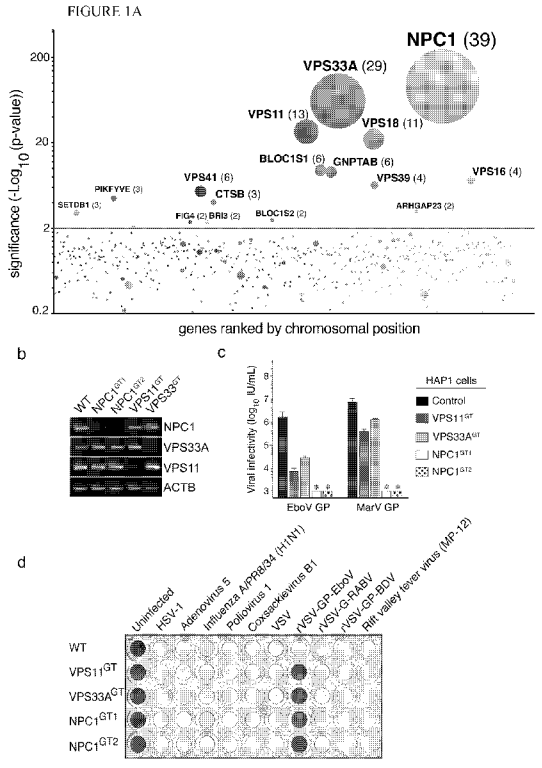

FIGURE 1B-1D

(57) Abstract: Methods and assays for treating a subject with a filovirus infection using an agent that inhibits Niemann-Pick CI (NPCI), VPSII, VPSI6, VPSI8, VPS33A, VPS39, VPS41, BLOCISI, BLOCIS2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4. Methods for screening for an agent that treats and/or prevents infection of a subject with a filovirus, where the methods comprise determining whether the agent inhibits one or more of Niemann-Pick CI (NPCI), VPSII, VPSI6, VPSI8, VPS33A, VPS39, VPS41. BLOCISI, BLOCIS2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4, wherein an agent that inhibits one or more of NPCI, VPSII, VPSI6, VPSI8, VPS33A, VPS39, VPS41, BLOCISI, BLOCIS2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 is a candidate for treating and/or preventing an infection with a filovirus and wherein an agent that does not inhibit NPCI, VPSII, VPSI6, VPSI8, VPS33A. VPS39, VPS41, BLOCISI, BLOCIS2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 is not a candidate for treating and/or preventing an infection with a filovirus.

WO 2012/103081 A1

SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

METHODS AND ASSAYS FOR TREATING FILOVIRIDAE INFECTIONS

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims priority of U.S. Provisional Patent Application No. 61/435,858, filed on January 25, 2011, the content of which is incorporated by reference.

STATEMENT OF GOVERNMENT SUPPORT

[0002] This invention was made with government support under grant numbers AI088027, AI081842, AI057159 and HG004938 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0003] Throughout this application various publications are referred to in superscripts. Full citations for these references may be found at the end of the specification. The disclosures of these publications are hereby incorporated by reference in their entirety into the subject application to more fully describe the art to which the subject invention pertains.

[0004] Infections by the Ebola (EboV) and Marburg (MarV) filoviruses cause a rapidly fatal hemorrhagic fever in humans for which no approved vaccines or antivirals are available[1]. Filovirus entry into cells is mediated by the viral spike glycoprotein (GP), which attaches viral particles to the cell surface, delivers them to endosomes, and catalyzes fusion between viral and endosomal membranes[2]. Additional host factors in the endosomal compartment, including a putative entry receptor, are likely required for viral membrane fusion. However, despite considerable efforts, these critical host factors have defied molecular identification[3-5].

[0005] The present invention addresses the need for methods and assays for treating subjects infected with filoviruses or who are at risk for infection with filoviruses.

SUMMARY OF THE INVENTION

[0006] The present invention provides methods for treating a subject infected with a filovirus or for preventing an infection with a filovirus in a subject at risk for infection with a filovirus, where the methods comprise administering to the subject an agent that inhibits one or more of Niemann-Pick C1 (NPC1), VPS11, VPS16, VPS18, VPS33A, VPS39,

VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 in an amount effective to treat and/or prevent infection with the filovirus.

[0007] The present invention also provides methods for screening for an agent that treats and/or prevents infection of a subject with a filovirus, where the methods comprise determining whether or not the agent inhibits one or more of Niemann-Pick C1 (NPC1), VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4, wherein an agent that inhibits one or more of NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 is a candidate for treating and/or preventing an infection with a filovirus and wherein an agent that does not inhibit NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 is not a candidate for treating and/or preventing an infection with a filovirus.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Figure 1A-1D. Genome-wide haploid genetic screen identifies the HOPS complex and NPC1 as host factors for filovirus entry. a) Genes significantly enriched for gene-trap insertion event in the rVSV-GP-EboV-selected cell population as compared to the non-selected mutagenized cell population. Circles represent genes and their size corresponds to the number of insertions identified in that gene in the rVSV-GP-EboV selected population. Significantly enriched genes (p-value < 0.01) are labeled with gene name. The number of independent insertions is indicated in parentheses. Genes are ranked on the X-axis based on their chromosomal position. b) RT-PCR analysis of the expression levels of NPC1, VPS33A and VPS11 in clones that contain gene trap insertions in the corresponding genes. c) Infectivity of VSV pseudotyped with the indicated filovirus glycoproteins in WT and mutant HAP1 clones. Means ± standard deviation (SD) are shown. EboV, Ebola virus (Zaire), SunV, Sudan virus, MarV, Marburg virus. Asterisks indicate that infectivity was below the limit of detection. d) The indicated HAP1 clones were exposed to a set of unrelated enveloped and nonenveloped viruses including recombinant VSV viruses carrying Rabies or Borna disease virus glycoproteins. Surviving adherent cells were stained with crystal violet.

[0009] Figure 2A-2D. Viral infection mediated by filovirus glycoproteins requires NPC1 but not NPC2. a) Skin fibroblasts from an apparently normal individual (control) and from patients carrying homozygous mutations in NPC1 or NPC2 were stained with filipin to

visualize intracellular cholesterol, or challenged with rVSV-G or rVSV-GP-EboV. Filipin-stained and infected cells were visualized by fluorescence microscopy. Filipin-stained images were inverted for clarity. Hoechst 33342 nuclear counterstain. b) Infectivity of VSV pseudotyped with the indicated viral glycoproteins in control and Niemann-Pick fibroblasts. Asterisks indicate that infectivity was below the limit of detection. c) NPC1 patient fibroblasts stably expressing an empty vector control or human NPC1 were stained with filipin or challenged with rVSV-GP-EboV. d) Infectivity of rVSV-G and rVSV-GP-EboV in Vero cells preincubated for 30 min with the indicated concentrations of U18666A. Scale bars, 200 μm (a, c). Means ± standard deviation (SD) are shown (b, d).

[0010] Figure 3A-3D. Ebola virus entry is arrested at a late step in cells deficient for the HOPS complex and NPC1. a) Viral particles attach and internalize into HOPS- and NPC1-deficient cells. The indicated HAP1 clones were inoculated with rVSV-GP-EboV and examined by transmission electron microscopy. Representative images of early steps in entry are shown. b, In vitro-cleaved rVSV-GP-EboV cannot bypass the block to infection observed in $VPS11^{GT}$, $VPS33A^{GT}$ and $NPC1^{GT}$ cells. Infectivity of mock- or thermolysin-cleaved rVSV-GP-EboV in the indicated mutant HAP1 clones. c) Viral escape into the cytoplasm is blocked in HOPS complex- and NPC1-deficient cells. Wild type HAP1 cells were treated with U18666A (10 μg/ml), and the indicated mutant HAP1 clones were exposed to VSV or rVSV-GP-EboV virus for 3h and processed for VSV M staining. Diffuse M staining indicates successful release of viral nucleocapsids into the cytoplasm. Punctuate staining, indicating viral particles trapped within endosomes and lysosomes, is shown by the arrows. d) Electron micrographs of rVSV-GP-EboV-infected VPS33A- and NPC1-deficient HAP1 cells and NPC1-deficient fibroblasts showing agglomerations of bullet-shaped VSV particles in vesicular compartments. All images were taken at 3 h post-inoculation. Asterisks highlight rVSV-GP-EboV particles in cross-section.

[0011] Figure 4A-4C. NPC1 function is required for infection by authentic Ebola and Marburg viruses. a) Fibroblasts from a healthy individual or an NPC1 patient were exposed to EboV or MarV. Cell supernatants were harvested at the indicated times post-infection and yields of infectious virus were measured. Means ± standard deviation (SD) are shown. Asterisks indicate that infectivity was below the limit of detection. b) Vero cells treated with DMSO vehicle (no drug) or U18666A (20 μM) were exposed to EboV or MarV. Cell supernatants were harvested at the indicated times post-infection and yields of infectious

virus were measured. Means ± standard deviation (SD) are shown. c) A speculative model for the roles of CatB, the HOPS complex, and NPC1 in Ebola virus entry.

[0012] Figure 5A-5C. Generation of HAP1 cells and susceptibility to rVSV-GP-EboV. a) Near-haploid KBM7 cells were coinfected with retroviral vectors expressing OCT4/ SOX2/c-MYC and KLF4 and an adherently growing subclone was identified (HAP1 cells). Karyotypic analysis of HAP1 cells indicates that the majority of cells (27 out of 39 analyzed) is haploid for all chromosomes b) Staining of KBM7 cells and HAP1 cells with pan-hematopoietic markers CD43 and CD45. Stained cells were examined by flow-cytometry. The unstained control is indicated in grey. c) Susceptibility of HAP1 and KBM7 cells to cell-killing by rVSV-GP-EboV.

[0013] Figure 6. Outline of the haploid genetic screen to identify host factors for Ebola virus entry. 100 million early passage HAP1 cells were infected with gene-trap virus and further expanded. A subset of cells was used to characterize the distribution of gene-trap insertion across the human genome. Sequences flanking the gene-traps were amplified, sequenced in parallel and aligned to the human genome. Independent insertion events into annotated genes were counted. 100 million cells were exposed to rVSV-GP-EboV virus and resistant clones were pooled and expanded. Most of these cells were used to amplify sequences flanking the gene-traps, sequence the insertion sites in parallel, and align these sequences to the human genome. A subset of the cells were used to obtain NPC^{GT} and VPS^{GT} cells through subcloning. Gene disruption events in the selected population were compared to the unselected cell population and genes that were significantly enriched for mutations were identified.

[0014] Figure 7A-7C. Identification and characterization of HAP1 cells carrying gene-trap insertions in the NPC1, VPS11 and VPS33A loci. a) Schematic outline of the positions of gene-trap insertions in the corresponding genes. Gene traps were located in the sense orientation in intronic sequences of the 5'-end of the gene and are therefore predicted to disrupt gene function. b) Clonal cell lines carrying the gene-trap insertions in the corresponding loci were identified through subcloning. Genotyping indicates the absence of wild type genomic loci and the presence of gene-trap loci. c) Cells carrying gene-trap insertions in the corresponding loci and wild type HAP1 cells were inoculated with rVSV-GP-EboV, and infected cells were visualized by fluorescence microscopy 12 h later.

[0015] Figure 8A-8B. NPC1 deficiency of HAP1 and CHO cells confers resistance to viral infection mediated by Ebola and Marburg virus glycoproteins. a) Immunoblot blot

analysis of NPC1 in HAP1 cells, HAP1 cells carrying a gene-trap insertion in the NPC1 locus and the same cell line infected with the NPC1-expressing retrovirus. CDK4 was used a loading control. b) Wild type or NPC1-deficient CHO cells were challenged with VSV pseudotyped with the indicated viral glycoproteins, and viral infectivity was measured 24 h later.

[0016] Figure 9. NPC2-mutant fibroblasts derived from a second Niemann-Pick type C patient are susceptible to viral infection mediated by the Ebola virus glycoprotein. Fibroblasts from an apparently normal individual and a Niemann-Pick disease patient carrying homozygous mutations in NPC2 were infected with VSV pseudotypes bearing the indicated viral glycoproteins, and viral infectivity was measured 24 h later. Means ± standard deviation (SD) are shown

[0017] Figure 10A-10B. Clearance of accumulated cholesterol does not render NPC1-deficient cells susceptible to infection by rVSV-GP-EboV. Wild type and NPC1-null CHO cells were cultivated either in normal growth medium (control) or in growth medium containing lipoprotein-depleted fetal bovine serum (depleted) for 6 days. Cells were then stained with filipin to visualize accumulated cholesterol (A) or exposed to rVSV-GP-EboV (B). Filipin-stained or infected cells were visualized by fluorescence microscopy. In each of (A) and (B), top panels are control and bottom panels are depleted.

[0018] Figure 11A-11C. The activities of endosomal cysteine cathepsins B and L are not inhibited in NPC1-defective cells. a) In vitro cleaved rVSV-GP-EboV bypasses the intracellular requirement for cathepsin B (CatB). Infectivity of mock- or thermolysin-cleaved rVSV-GP-EboV in Vero cells treated with the CatB inhibitor CA074. b) Fibroblasts from an apparently normal individual (control) and a Niemann-Pick patient carrying homozygous mutations in NPC1 were lysed at acid pH, and the capacity of these acidic extracts to cleave a fluorogenic peptide substrate for CatB and CatL was measured. Pretreatment of cell extracts with the pan-cysteine protease inhibitor E-64 abolished substrate cleavage, confirming that only cysteine cathepsin activities were being measured. c) Intact control and NPC1-deficient fibroblasts and CHO cells were incubated with a fluorophore-tagged suicide substrate for CatB/CatL. Cells were then lysed and fluorophore-labeled CatB and CatL proteins were detected by SDS-polyacrylamide gel electrophoresis and fluorescence imaging. hCatB and hCatL, human enzymes. cCatB and cCatL, CHO enzymes. sc and hc, single chain and heavy chain forms of CatL.

[0019] Figure 12. Viral membrane fusion is required for VSV M release into the cytoplasm of infected cells. Wild type HAP1 cells were treated with puromycin (5 μg/ml) and inoculated with rVSV-GP-EboV in the presence or absence of bafilomycin A1 (bafA1; 100 nM) or ammonium chloride (NH_4Cl; 20 μg/ml). Cells were fixed 3 h post inoculation and stained with VSV M antibody 23H12. Successful fusion leads to the diffuse M staining throughout the cytoplasm. Failure to fuse leads to discrete punctuate of M staining as shown by bafA1 and NH_4Cl.

[0020] Figure 13. Accumulation of rVSV-GP-EboV viral particles in vesicular compartments in NPC1-deficient cells. Electron micrograph of a second NPC1-deficient HAP1 clone exposed to rVSV-GP-EboV. A large agglomeration of bullet-shaped VSV particles is visible within a vesicular (endosomal) compartment. All images were taken at 3 h post-infection.

[0021] Figure 14A-14B. NPC1 pathway inhibitor U18666A blocks authentic EboV and MarV infection of primary human cells. Human peripheral blood monocyte-derived dendritic cells (DC) (b) and umbilical-vein endothelial cells (HUVEC) (a) were infected in the presence or absence of U18666A (10 μM) at an MOI of 3 and the percentage of infected cells was determined by immunostaining.

[0022] Figure 15A-15B. U18666A acts rapidly to inhibit Ebola virus GP-dependent entry. (a) Time-of-pretreatment experiment: Vero cells were left untreated or treated with U18666A (20 μM) for the indicated times and then exposed to rVSV-GP-EboV. After 1 h, viral entry was terminated by addition of NH_4Cl (20 mM). Viral infectivity was measured 14 h later. Means ± standard deviation (SD) are shown. (b) Time-of-escape experiment: Vero cells were first exposed to rVSV-GP-EboV and then left untreated or treated with U18666A for 1 h at 37°C. After 1 h, viral entry was terminated by addition of NH_4Cl (20 mM). Viral infectivity was determined as above.

[0023] Figure 16. NPC1 pathway inhibitor imipramine blocks authentic EboV and MarV infection. Vero cells were infected in the presence or absence of U18666A (10 μM) at an MOI of 3 and the percentage of infected cells at each timepoint was determined by immunostaining.

[0024] Figure 17A-17B. NPC1 is required for in vivo infection and pathogenesis by EboV (a) and MarV (b). Survival of $NPC1^{+/+}$ and $NPC1^{-/+}$ mice (n=10 for each group) inoculated intraperitoneally (i.p.) with ~1000 pfu of mouse-adapted EboV or MarV.

[0025] Figure 18. NPC1 pathway inhibitor imipramine partially protects mice from EboV. Balb/c mice (n=10 for each group) were inoculated intraperitoneally (i.p.) with ~1000 pfu of mouse-adapted EboV or MarV together with vehicle (PBS) or imipramine at the indicated dose.

[0026] Figure 19. Topological model of NPC1. Domain A contains a sterol-binding domain, but the specific functions of domains C and I are unknown. In the present studies, a flag epitope tag was appended to the C–terminus of NPC1.

[0027] Figure 20A-20C. NPC1 luminal loop domain C is required for filovirus entry, but full-length NPC1 is dispensable. (a) NPC1-null CHO CT43 cells were engineered to express mutant forms of human NPC1-flag lacking domains A, C, or I. Capacity of mutant NPC1 proteins to rescue viral entry and transport lysosomal cholesterol was determined. (*Left*) Infection of NPC1-null CHO CT43 cells expressing mutant NPC1-flag proteins by recombinant VSVs bearing VSV G or filovirus glycoproteins. Infected cells were visualized by fluorescence microscopy. (*Right*) Cholesterol clearance by mutant NPC1-flag proteins in CT43 cells was determined by filipin staining and fluorescence microscopy. Images were inverted for clarity. Scale bars, 20 μm. (b-c) Infectivity of VSV pseudotypes bearing VSV or filovirus glycoproteins (b) and wild type MARV (c) in CT43 cells expressing mutant NPC-flag proteins. SUDV, Sudan virus. Error bars indicate SD. Asterisks indicate values below the limit of detection.

[0028] Figure 21A-21D. NPC1 binds specifically to a cleaved form of the Ebola virus glycoprotein. (a) Co-immunoprecipitation (IP) of NPC1 by EBOV GP. Magnetic beads coated with GP-specific monoclonal antibody KZ52 were incubated with detergent extracts containing no virus (None), uncleaved rVSV-GP, or cleaved rVSV-GP_{CL}. The resulting control or glycoprotein-decorated beads were mixed with cell lysates containing human NPC1-flag at pH 7.5 or pH 5.1 and 4°C. Beads were then retrieved and NPC1-flag in the immune pellets and supernatants was detected by immunoblotting (IB) with an anti-flag antibody. (b) GP_{CL} captures NPC1 in an ELISA. Plates coated with rVSV-GP or rVSV-GP_{CL} were incubated with cell extracts containing NPC1-flag, and bound flag-tagged proteins were detected with an anti-flag antibody. (c-d) GP_{CL} but not GP captures affinity-purified NPC1-flag in an ELISA. (c) NPC1-flag was purified from CT43 CHO cell lysates by flag affinity chromatography and visualized by SDS-PAGE and staining with Krypton infrared protein stain. (d) ELISA plates coated with rVSV-GP or rVSV-GP_{CL} were

incubated with NPC1-flag purified in (c), and bound flag-tagged proteins were detected with an anti-flag antibody.

[0029] Figure 22A-22C. Soluble forms of NPC1 domain C bind directly to GP and selectively neutralize infection by viral particles containing cleaved glycoproteins. (a) The capacity of rVSV-GP and rVSV-GP_{CL} to capture a purified, soluble form of domain C containing flag and hexahistidine tags was determined in an ELISA. (b) The capacity of a purified, soluble form of GP lacking the transmembrane domain (GPΔTM) to associate with purified, soluble domain C was determined by co-immunoprecipitation. (c) rVSV-GP and rVSV-GP_{CL} were preincubated with soluble domain C, and virus-protein mixtures were exposed to Vero cells. Viral infection was enumerated by fluorescence microscopy.

[0030] Figure 23A-23B. A synthetic single-pass membrane protein containing NPC1 domain C can mediate EboV and MarV entry. CT43 cells expressing synthetic membrane proteins containing individual NPC1 luminal domains were exposed to rVSVs bearing uncleaved or cleaved filovirus glycoproteins. Infected cells were visualized (a) and enumerated (b) by fluorescence microscopy. Asterisks in panel b indicate values below the limit of detection.

[0031] Figure 24A-24D. Some possible modes of action of small molecule antivirals targeting NPC1. (a) Binding of EboV and MarV GP to domain C in NPC1 within endosomes or lysosomes is required for viral entry and infection. (b) A compound may direct inhibit GP-NPC1 interaction by binding to either protein. (c) A compound may indirectly inhibit GP-NPC1 interaction by binding to NPC1 (in domain C or elsewhere) or to an associated protein or lipid, thereby inducing a conformational change in NPC1. (d) A compound may induce misfolding of NPC1 or otherwise cause reduction of NPC1 levels within the endo/lysosomal compartment. Note that this figure only illustrate a few possible modes of action.

[0032] Figure 25. A homogeneous electrochemiluminescence assay to screen for inhibitors of the GP-NPC1 interaction. Binding of a GP ectodomain labeled with a SULFO-tag™ (MesoScale Discovery Systems) to NPC1 in immobilized membranes is detected by the emission of light by $*Ru(bpy)_3^{2+}$. This activated species is electrochemically generated at the bottom of the microplate well.

[0033] Figure 26. A homogeneous Alphascreen™ assay (Perkin-Elmer) to screen for inhibitors of the GP-NPC1 interaction. Binding of purified NPC1 domain C tethered to Donor beads and GP ectodomain tethered to Acceptor beads is detected as follows. When

Donor and Acceptor beads are brought into close proximity by the GP-NPC1 interaction, the excitation of the Donor beads provokes the release of singlet oxygen ($^1O_2$), triggering a cascade of energy transfer to the Acceptor beads and resulting in blue-shifted emission.

DETAILED DESCRIPTION OF THE INVENTION

[0034] The present invention provides a method for treating a subject infected with a filovirus or for preventing an infection with a filovirus in a subject at risk for infection with a filovirus comprising administering to the subject an agent that inhibits one or more of Niemann-Pick C1 (NPC1), VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 in an amount effective to treat and/or prevent infection with the filovirus.

[0035] The family *Filoviridae* is a family of viruses including genera *Ebolavirus* and *Marburgvirus*.

[0036] To treat a subject with a filovirus infection means to reduce or stop the spread of filovirus in the subject, or to eliminate the filovirus from the subject, or to reduce or eliminate a sign or symptom of filovirus infection in the subject. Filovirus infection is characterized by hemorrhagic fever, including abnormalities in blood coagulation.

[0037] Subjects who are at risk for infection with filoviruses include subjects who have been exposed to filovirus or are at risk of exposure to filovirus. In addition to the natural occurrence of filoviruses, there is the potential for exposure to these pathogens if they are used as agents of bioterrorism or biological warfare.

[0038] The NPC1 gene encodes NPC1 protein, which is located in the membrane of endosomes and lysosomes and mediates intracellular cholesterol trafficking, in part via binding of cholesterol to its N-terminal domain[38,39]. NPC1 protein has a cytoplasmic C-terminus, 13 transmembrane domains, and 3 large loops in the lumen of the endosome[38] (see Fig. 19). Defects in the NPC1 gene cause Niemann-Pick type C disease[8], a rare autosomal recessive neurodegenerative disorder characterized by over accumulation of cholesterol and glycosphingolipids in endosomal/lysosomal compartments.

[0039] Human NPC1 protein has the amino acid sequence (SEQ ID NO:1) (NCBI Reference Sequence: NM_000271.4):

```
MTARGLALGL LLLLLCPAQV FSQSCVWYGE CGIAYGDKRY NCEYSGPPKP LPKDGYDLVQ      60
ELCPGFFFGN VSLCCDVRQL QTLKDNLQLP LQFLSRCPSC FYNLLNLFCE LTCSPRQSQF     120
LNVTATEDYV DPVTNQTKTN VKELQYYVGQ SFANAMYNAC RDVEAPSSND KALGLLCGKD     180
ADACNATNWI EYMFNKDNGQ APFTITPVFS DFPVHGMEPM NNATKGCDES VDEVTAPCSC     240
```

```
QDCSIVCGPK PQPPPPPAPW TILGLDAMYV IMWITYMAFL LVFFGAFFAV WCYRKRYFVS     300
EYTPIDSNIA FSVNASDKGE ASCCDPVSAA FEGCLRRLFT RWGSFCVRNP GCVIFFSLVF     360
ITACSSGLVF VRVTTNPVDL WSAPSSQARL EKEYFDQHFG PFFRTEQLII RAPLTDKHIY     420
QPYPSGADVP FGPPLDIQIL HQVLDLQIAI ENITASYDNE TVTLQDICLA PLSPYNTNCT     480
ILSVLNYFQN SHSVLDHKKG DDFFVYADYH THFLYCVRAP ASLNDTSLLH DPCLGTFGGP     540
VFPWLVLGGY DDQNYNNATA LVITFPVNNY YNDTEKLQRA QAWEKEFINF VKNYKNPNLT     600
ISFTAERSIE DELNRESDSD VFTVVISYAI MFLYISLALG HMKSCRRLLV DSKVSLGIAG     660
ILIVLSSVAC SLGVFSYIGL PLTLIVIEVI PFLVLAVGVD NIFILVQAYQ RDERLQGETL     720
DQQLGRVLGE VAPSMFLSSF SETVAFFLGA LSVMPAVHTF SLFAGLAVFI DFLLQITCFV     780
SLLGLDIKRQ EKNRLDIFCC VRGAEDGTSV QASESCLFRF FKNSYSPLLL KDWMRPIVIA     840
IFVGVLSFSI AVLNKVDIGL DQSLSMPDDS YMVDYFKSIS QYLHAGPPVY FVLEEGHDYT     900
SSKGQNMVCG GMGCNNDSLV QQIFNAAQLD NYTRIGFAPS SWIDDYFDWV KPQSSCCRVD     960
NITDQFCNAS VVDPACVRCR PLTPEGKQRP QGGDFMRFLP MFLSDNPNPK CGKGGHAAYS    1020
SAVNILLGHG TRVGATYFMT YHTVLQTSAD FIDALKKARL IASNVTETMG INGSAYRVFP    1080
YSVFYVFYEQ YLTIIDDTIF NLGVSLGAIF LVTMVLLGCE LWSAVIMCAT IAMVLVNMFG    1140
VMWLWGISLN AVSLVNLVMS CGISVEFCSH ITRAFTVSMK GSRVERAEEA LAHMGSSVFS    1200
GITLTKFGGI VVLAFAKSQI FQIFYFRMYL AMVLLGATHG LIFLPVLLSY IGPSVNKAKS    1260
CATEERYKGT ERERLLNF                                                  1278
```

[0040] Nucleic acid (mRNA) encoding human NPC1 protein has the nucleotide sequence (SEQ ID NO:2) (NCBI Reference Sequence: NM_000271.4):

```
   1 gaagggcaac acggggacct tgaagcgggg tcgcggcggc gccccagccc gggccaggga
  61 gtcccggcag cggcacctcc cagaaagggc ggagccgacg acgccttctt ccttcctgac
 121 cggcgcgcgc agcctgctgc cgcggtcagc gcctgctcct gctcctccgc tcctcctgcg
 181 cggggtgctg aaacagcccg gggaagtaga gccgcctccg gggagcccaa ccagccgaac
 241 gccgccggcg tcagcagcct tgcgcggcca cagcatgacc gctcgcggcc tggcccttgg
 301 cctcctcctg ctgctactgt gtccagcgca ggtgttttca cagtcctgtg tttggtatgg
 361 agagtgtgga attgcatatg gggacaagag gtacaattgc gaatattctg gcccaccaaa
 421 accattgcca aaggatggat atgacttagt gcaggaactc tgtccaggat tcttctttgg
 481 caatgtcagt ctctgttgtg atgttcggca gcttcagaca ctaaaagaca acctgcagct
 541 gcctctacag tttctgtcca gatgtccatc ctgtttttat aacctactga acctgttttg
 601 tgagctgaca tgtagccctc gacagagtca gtttttgaat gttacagcta ctgaagatta
 661 tgttgatcct gttacaaacc agacgaaaac aaatgtgaaa gagttacaat actacgtcgg
 721 acagagtttt gccaatgcaa tgtacaatgc ctgccgggat gtggaggccc cctcaagtaa
 781 tgacaaggcc ctgggactcc tgtgtgggaa ggacgctgac gcctgtaatg ccaccaactg
 841 gattgaatac atgttcaata aggacaatgg acaggcacct tttaccatca ctcctgtgtt
 901 ttcagatttt ccagtccatg ggatggagcc catgaacaat gccaccaaag gctgtgacga
 961 gtctgtggat gaggtcacag caccatgtag ctgccaagac tgctctattg tctgtggccc
1021 caagccccag cccccacctc ctcctgctcc ctggacgatc cttggcttgg acgccatgta
1081 tgtcatcatg tggatcacct acatggcgtt tttgcttgtg ttttttggag catttttttgc
1141 agtgtggtgc tacagaaaac ggtattttgt ctccgagtac actcccatcg atagcaatat
1201 agctttttct gttaatgcaa gtgacaaagg agaggcgtcc tgctgtgacc ctgtcagcgc
1261 agcatttgag ggctgcttga ggcggctgtt cacacgctgg gggtctttct gcgtccgaaa
1321 ccctggctgt gtcattttct tctcgctggt cttcattact gcgtgttcgt caggcctggt
1381 gtttgtccgg gtcacaacca atccagttga cctctggtca gcccccagca gccaggctcg
1441 cctggaaaaa gagtactttg accagcactt tgggcctttc ttccggacgg agcagctcat
1501 catccgggcc cctctcactg acaaacacat ttaccagcca tacccttcgg gagctgatgt
1561 acccctttgga cctccgcttg acatacagat actgcaccag gttcttgact tacaaatagc
1621 catcgaaaac attactgcct cttatgacaa tgagactgtg acacttcaag acatctgctt
1681 ggcccctctt tcaccgtata acacgaactg caccattttg agtgtgttaa attacttcca
1741 gaacagccat tccgtgctgg accacaagaa aggggacgac ttctttgtgt atgccgatta
1801 ccacacgcac tttctgtact gcgtacgggc tcctgcctct ctgaatgata caagtttgct
1861 ccatgaccct tgtctgggta cgtttggtgg accagtgttc ccgtggcttg tgttgggagg
1921 ctatgatgat caaaactaca ataacgccac tgcccttgtg attaccttcc ctgtcaataa
1981 ttactataat gatacagaga agctccagag ggcccaggcc tgggaaaaag agtttattaa
2041 ttttgtgaaa aactacaaga atcccaatct gaccatttcc ttcactgctg aacgaagtat
2101 tgaagatgaa ctaaatcgtg aaagtgacag tgatgtcttc accgttgtaa ttagctatgc
2161 catcatgttt ctatatattt ccctagcctt ggggcacatg aaaagctgtc gcaggcttct
2221 ggtggattcg aaggtctcac taggcatcgc gggcatcttg atcgtgctga gctcggtggc
```

```
2281 ttgctccttg ggtgtcttca gctacattgg gttgcccttg accctcattg tgattgaagt
2341 catcccgttc ctggtgctgg ctgttggagt ggacaacatc ttcattctgg tgcaggccta
2401 ccagagagat gaacgtcttc aaggggaaac cctggatcag cagctgggca gggtcctagg
2461 agaagtggct cccagtatgt tcctgtcatc cttttctgag actgtagcat ttttcttagg
2521 agcattgtcc gtgatgccag ccgtgcacac cttctctctc tttgcgggat tggcagtctt
2581 cattgacttt cttctgcaga ttacctgttt cgtgagtctc ttggggttag acattaaacg
2641 tcaagagaaa aatcggctag acatcttttg ctgtgtcaga ggtgctgaag atggaacaag
2701 cgtccaggcc tcagagagct gtttgtttcg cttcttcaaa aactcctatt ctccacttct
2761 gctaaaggac tggatgagac caattgtgat agcaatattt gtgggtgttc tgtcattcag
2821 catcgcagtc ctgaacaaag tagatattgg attggatcag tctctttcga tgccagatga
2881 ctcctacatg gtggattatt tcaaatccat cagtcagtac ctgcatgcgg gtccgcctgt
2941 gtactttgtc ctggaggaag ggcacgacta cacttcttcc aaggggcaga acatggtgtg
3001 cggcggcatg ggctgcaaca atgattccct ggtgcagcag atatttaacg cggcgcagct
3061 ggacaactat acccgaatag gcttcgcccc ctcgtcctgg atcgacgatt atttcgactg
3121 ggtgaagcca cagtcgtctt gctgtcgagt ggacaatatc actgaccagt tctgcaatgc
3181 ttcagtggtt gaccctgcct gcgttcgctg caggcctctg actccggaag gcaaacagag
3241 gcctcagggg ggagacttca tgagattcct gcccatgttc ctttcggata accctaaccc
3301 caagtgtggc aaaggggggac atgctgccta tagttctgca gttaacatcc tccttggcca
3361 tggcaccagg gtcggagcca cgtacttcat gacctaccac accgtgctgc agacctctgc
3421 tgactttatt gacgctctga agaaagcccg acttatagcc agtaatgtca ccgaaaccat
3481 gggcattaac ggcagtgcct accgagtatt tccttacagt gtgttttatg tcttctacga
3541 acagtacctg accatcattg acgacactat cttcaacctc ggtgtgtccc tgggcgcgat
3601 atttctggtg accatggtcc tcctgggctg tgagctctgg tctgcagtca tcatgtgtgc
3661 caccatcgcc atggtcttgg tcaacatgtt tggagttatg tggctctggg gcatcagtct
3721 gaacgctgta tccttggtca acctggtgat gagctgtggc atctccgtgg agttctgcag
3781 ccacataacc agagcgttca cggtgagcat gaaaggcagc cgcgtggagc gcgcggaaga
3841 ggcacttgcc cacatgggca gctccgtgtt cagtggaatc acacttacaa aatttggagg
3901 gattgtggtg ttggcttttg ccaaatctca aattttccag atattctact tcaggatgta
3961 tttggccatg gtcttactgg gagccactca cggattaata tttctccctg tcttactcag
4021 ttacataggg ccatcagtaa ataaagccaa aagttgtgcc actgaagagc gatacaaagg
4081 aacagagcgc gaacggcttc taaatttcta gccctctcgc agggcatcct gactgaactg
4141 tgtctaaggg tcggtcggtt taccactgga cgggtgctgc atcggcaagg ccaagttgaa
4201 caccggatgg tgccaaccat cggttgtttg gcagcagctt tgaacgtagc gcctgtgaac
4261 tcaggaatgc acagttgact tgggaagcag tattactaga tctggaggca accacaggac
4321 actaaacttc tcccagcctc ttcaggaaag aaacctcatt ctttggcaag caggaggtga
4381 cactagatgg ctgtgaatgt gatccgctca ctgacactct gtaaaggcca atcaatgcac
4441 tgtctgtctc tccttttagg agtaagccat cccacaagtt ctataccata tttttagtga
4501 cagttgaggt tgtagataca ctttataaca ttttatagtt taaagagctt tattaatgca
4561 ataaattaac tttgtacaca tttttatata aaaaaacagc aagtgatttc agaatgttgt
4621 aggcctcatt agagcttggt ctccaaaaat ctgtttgaaa aaagcaacat gttcttcaca
4681 gtgttcccct agaaaggaag agatttaatt gccagttaga tgtggcatga aatgagggac
4741 aaagaaagca tctcgtaggt gtgtctactg ggttttaact tatttttctt taataaaata
4801 cattgttttc ctaaaaaaaa aaaaaaa
```

[0041] Human vacuolar protein sorting 11 (VPS11) protein has the amino acid sequence (SEQ ID NO:3) (NCBI Reference Sequence: NM_021729.4):

```
MAAYLQWRRF VFFDKELVKE PLSNDGAAPG ATPASGSAAS KFLCLPPGIT VCDSGRGSLV      60
FGDMEGQIWF LPRSLQLTGF QAYKLRVTHL YQLKQHNILA SVGEDEEGIN PLVKIWNLEK     120
RDGGNPLCTR IFPAIPGTEP TVVSCLTVHE NLNFMAIGFT DGSVTLNKGD ITRDRHSKTQ     180
ILHKGNYPVT GLAFRQAGKT THLFVVTTEN VQSYIVSGKD YPRVELDTHG CGLRCSALSD     240
PSQDLQFIVA GDECVYLYQP DERGPCFAFE GHKLIAHWFR GYLIIVSRDR KVSPKSEFTS     300
RDSQSSDKQI LNIYDLCNKF IAYSTVFEDV VDVLAEWGSL YVLTRDGRVH ALQEKDTQTK     360
LEMLFKKNLF EMAINLAKSQ HLDSDGLAQI FMQYGDHLYS KGNHDGAVQQ YIRTIGKLEP     420
SYVIRKFLDA QRIHNLTAYL QTLHRQSLAN ADHTTLLLNC YTKLKDSSKL EEFIKKKSES     480
EVHFDVETAI KVLRQAGYYS HALYLAENHA HHEWYLKIQL EDIKNYQEAL RYIGKLPFEQ     540
AESNMKRYGK ILMHHIPEQT TQLLKGLCTD YRPSLEGRSD REAPGCRANS EEFIPIFANN     600
PRELKAFLEH MSEVQPDSPQ GIYDTLLELR LQNWAHEKDP QVKEKLHAEA ISLLKSGRFC     660
DVFDKALVLC QMHDFQDGVL YLYEQGKLFQ QIMHYHMQHE QYRQVISVCE RHGEQDPSLW     720
EQALSYFARK EEDCKEYVAA VLKHIENKNL MPPLLVVQTL AHNSTATLSV IRDYLVQKLQ     780
```

```
KQSQQIAQDE LRVRRYREET TRIRQEIQEL KASPKIFQKT KCSICNSALE LPSVHFLCGH     840
SFHQHCFESY SESDADCPTC LPENRKVMDM IRAQEQKRDL HDQFQHQLKC SNDSFSVIAD     900
YFGRGVFNKL TLLTDPPTAR LTSSLEAGLQ RDLLMHSRRG T                        941
```

[0042] Nucleic acid (mRNA) encoding human VPS11 protein has the nucleotide sequence (SEQ ID NO:4) (NCBI Reference Sequence: NM_021729.4):

```
   1 ctcacgtgac aaagctcccg gaggtgggag ccctgggcca aaatggcggc ctacctgcag
  61 tggcggcgct tcgttttctt cgacaaggag ctggtgaagg agccgctgag caatgatggg
 121 gccgctcccg gggccacacc tgcttctgga tccgctgctt ccaagttcct ttgcctccct
 181 cctggcatca ctgtctgcga ctcaggccga gggagcctgg tctttggaga tatggaaggc
 241 cagatctggt tcttgccacg ttccctacag cttacaggct tccaagccta caaactacgg
 301 gtgacacacc tgtaccaact gaagcagcac aatattctgg catctgttgg agaagatgaa
 361 gagggcatca acccccttggt taagatctgg aacctggaga agagagatgg tggcaatcca
 421 ctctgcactc gaatcttccc tgctattcca ggaacagagc caactgttgt atcttgtttg
 481 actgtccatg aaaatctcaa ctttatggcc attggtttca cagatggcag tgttacattg
 541 aacaaaggag acatcacccg ggaccggcat agcaagaccc agattttgca caagggcaac
 601 tatcctgtaa ctggattggc ctttcgccaa gcaggaaaga ccactcactt gtttgttgtg
 661 acaacagaga acgtccagtc ctatatagtt tctggaaaag actaccctcg cgtggagttg
 721 gacacccatg gttgtggcct gcgctgctca gccctaagtg acccttctca ggacctgcag
 781 ttcattgtgg ccggggatga gtgtgtctac ttgtaccagc ctgatgaacg tgggccctgc
 841 ttcgcctttg agggccataa gctcattgcc cactggttta gaggctacct tatcattgtc
 901 tcccgtgacc ggaaggtttc tcccaagtca gagtttacca gcagggattc acagagctcc
 961 gacaagcaga ttctaaacat ctatgacctg tgcaacaagt tcatagccta tagcaccgtc
1021 tttgaggatg tagtggatgt gcttgctgag tggggctccc tgtacgtgct gacgcgggat
1081 gggcgggtcc acgcactgca ggagaaggac acacagacca aactggagat gctgtttaag
1141 aagaacctat ttgagatggc gattaacctt gccaagagcc agcatctgga cagtgatggg
1201 ctggcccaga ttttcatgca gtatggagac catctctaca gcaagggcaa ccacgatggg
1261 gctgtccagc aatatatccg aaccattgga aagttggagc catcctacgt gatccgcaag
1321 tttctggatg cccagcgcat tcacaacctg actgcctacc tgcagaccct gcaccgacaa
1381 tccctggcca atgccgacca taccaccctg ctcctcaact gctataccaa gctcaaggac
1441 agctcgaagc tggaggagtt catcaagaaa aagagtgaga gtgaagtcca ctttgatgtg
1501 gagacagcca tcaaggtcct ccggcaggct ggctactact cccatgccct gtatctggcg
1561 gagaaccatg cacatcatga gtggtacctg aagatccagc tagaagacat taagaattat
1621 caggaagccc ttcgatacat cggcaagctg ccttttgagc aggcagagag caacatgaag
1681 cgctacggca agatcctcat gcaccacata ccagagcaga caactcagtt gctgaaggga
1741 ctttgtactg attatcggcc cagcctcgaa ggccgcagcg atagggaggc cccaggctgc
1801 agggccaact ctgaggagtt catccccatc tttgccaata acccgcgaga gctgaaagcc
1861 ttcctagagc acatgagtga agtgcagcca gactcacccc aggggatcta cgacacactc
1921 cttgagctgc gactgcagaa ctgggcccac gagaaggatc cacaggtcaa agagaagctt
1981 cacgcagagg ccatttccct gctgaagagt ggtcgcttct gcgacgtctt tgacaaggcc
2041 ctggtcctgt gccagatgca cgacttccag gatggtgtcc tttaccttta tgagcagggg
2101 aagctgttcc agcagatcat gcactaccac atgcagcacg agcagtaccg gcaggtcatc
2161 agcgtgtgtg agcgccatgg ggagcaggac ccctccttgt gggagcaggc cctcagctac
2221 ttcgctcgca aggaggagga ctgcaaggag tatgtggcag ctgtcctcaa gcatatcgag
2281 aacaagaacc tcatgccacc tcttctagtg gtgcagaccc tggcccacaa ctccacagcc
2341 acactctccg tcatcaggga ctacctggtc caaaaactac agaaacagag ccagcagatt
2401 gcacaggatg agctgcgggt gcggcggtac cgagaggaga ccacccgtat ccgccaggag
2461 atccaagagc tcaaggccag tcctaagatt ttccaaaaga ccaagtgcag catctgtaac
2521 agtgccttgg agttgccctc agtccacttc ctgtgtggcc actccttcca ccaacactgc
2581 tttgagagtt actcggaaag tgatgctgac tgccccacct gcctccctga aaaccggaag
2641 gtcatggata tgatccgggc ccaggaacag aaacgagatc tccatgatca attccagcat
2701 cagctcaagt gctccaatga cagcttttct gtgattgctg actactttgg cagaggtgtt
2761 ttcaacaaat tgactctgct gaccgaccct cccacagcca gactgacctc cagcctggag
2821 gctgggctgc aacgcgacct actcatgcac tccaggaggg gcacttaagc agcctggagg
2881 aagatgtggg caacagtgga ggaccaagag aacagacaca atgggacctg ggcgggcgtt
2941 acacagaagg ctggctgaca tgcccagggc tccactctca tctaatgtca cagccctcag
3001 aactaaagcg gactttcttt ccctgccttc ttatttagtc agcttgccat ccctcctctt
3061 cactagcagt gtagatcatt ccagatcagt gggggagggc acctcagcaa cctctgagtg
3121 tggacaatag ctgctttctt ctctatccaa gagcaccagg ctgtgcttgg gtccttgctc
```

```
3181 tcagagtcta taaataaaag aatataatga tttgggagct taaaaaaaaa aaaaaaaaaa
3241 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaa
```

[0043] Human vacuolar protein sorting 16 (VPS16) protein has the amino acid sequence (SEQ ID NO:5) (NCBI Reference Sequence: NM_022575.2):

```
MDCYTANWNP LGDSAFYRKY ELYSMDWDLK EELRDCLVAA APYGGPIALL RNPWRKEKAA      60
SVRPVLDIYS ASGMPLASLL WKSGPVVSLG WSAEEELLCV QEDGAVLVYG LHGDFRRHFS     120
MGNEVLQNRV LDARIFHTEF GSGVAILTGA HRFTLSANVG DLKLRRMPEV PGLQSAPSCW     180
TVLCQDRVAH ILLAVGPDLY LLDHAACSAV TPPGLAPGVS SFLQMAVSFT YRHLALFTDT     240
GYIWMGTASL KEKLCEFNCN IRAPPKQMVW CSRPRSKERA VVVAWERRLM VVGDAPESIQ     300
FVLDEDSYLV PELDGVRIFS RSTHEFLHEV PAASEEIFKI ASMAPGALLL EAQKEYEKES     360
QKADEYLREI QELGQLTQAV QQCIEAAGHE HQPDMQKSLL RAASFGKCFL DRFPPDSFVH     420
MCQDLRVLNA VRDYHIGIPL TYSQYKQLTI QVLLDRLVLR RLYPLAIQIC EYLRLPEVQG     480
VSRILAHWAC YKVQQKDVSD EDVARAINQK LGDTPGVSYS DIAARAYGCG RTELAIKLLE     540
YEPRSGEQVP LLLKMKRSKL ALSKAIESGD TDLVFTVLLH LKNELNRGDF FMTLRNQPMA     600
LSLYRQFCKH QELETLKDLY NQDDNHQELG SFHIRASYAA EERIEGRVAA LQTAADAFYK     660
AKNEFAAKAT EDQMRLLRLQ RRLEDELGGQ FLDLSLHDTV TTLILGGHNK RAEQLARDFR     720
IPDKRLWWLK LTALADLEDW EELEKFSKSK KSPIGYLPFV EICMKQHNKY EAKKYASRVG     780
PEQKVKALLL VGDVAQAADV AIEHRNEAEL SLVLSHCTGA TDGATADKIQ RARAQAQKK      839
```

[0044] Nucleic acid (mRNA) encoding human VPS16 protein has the nucleotide sequence (SEQ ID NO:6) (NCBI Reference Sequence: NM_022575.2):

```
   1 ctaggtgggt gtcccctcgg tgcttcccag ctgccgtctg caccagccat ggactgctac
  61 acggcgaact ggaacccact cggggactct gccttttacc ggaaatatga gctgtacagc
 121 atggactggg acctgaagga ggaactcagg gattgcctgg tggctgctgc accctatggg
 181 ggccccattg cactgctgag gaacccctgg cggaaggaga aagctgctag tgtgaggcca
 241 gtgctcgata tatactctgc ttccggcatg cctctggcca gcctgctgtg gaagagtgga
 301 cccgtggtgt ccctgggctg gtcagctgag gaggagctgc tctgtgtgca ggaagatggt
 361 gctgtactgg tttatgggct tcatggtgac ttccggagac acttcagcat gggcaatgaa
 421 gtgctccaga accgggttct ggatgcccgg atctttcaca ctgagtttgg ttccggagtg
 481 gccatcctca caggggccca ccgcttcacc ctcagtgcca atgtgggtga cctcaaactc
 541 cgccggatgc cagaggtgcc aggtctgcaa agtgcaccct cctgctggac tgtgctgtgc
 601 caggaccgag tggcacacat tcttctggct gtggggcctg acctttacct cttggaccat
 661 gcagcctgct ccgcagtgac gccccctggc ctggccccag gagtaagcag cttcctacag
 721 atggctgtct ccttcaccta ccgacacctg gcactcttca cagacacagg ctacatctgg
 781 atggggacag catcactcaa ggagaagcta tgtgagttca actgcaacat ccgggcacct
 841 ccaaagcaga tggtctggtg cagccgtcct cgtagcaagg agagggccgt ggtggtggcc
 901 tgggaaaggc ggctgatggt ggtgggcgat gcacccgaga gcatccagtt tgtgctggat
 961 gaggactcct acctggtgcc tgagctcgat ggggtccgca tcttctcccg cagcacccac
1021 gagttcctgc atgaggttcc agcggccagc gaggaaatct tcaaaattgc ctcaatggcc
1081 cccggggcgc tgctcctgga ggctcagaag gagtatgaga aagagagcca gaaggcggac
1141 gagtacctgc gggagatcca ggagctgggc cagctgaccc aggccgtgca gcagtgcatt
1201 gaggctgcag gacatgagca ccagccagac atgcagaaga gtctgctcag ggccgcctcc
1261 ttcggaaagt gtttcctgga cagatttcca cccgacagct tcgtgcacat gtgtcaggac
1321 ctgcgtgtgc tcaatgctgt tcgggactat cacatcggga tcccgctcac ctatagccaa
1381 tataagcagc tcaccatcca ggtgctgctg gacaggctcg tgttgcggag actttacccc
1441 ctggccatcc agatatgcga gtacttgcgc cttcctgaag tacagggcgt cagcaggatc
1501 ctggcccact gggcctgcta caaggtgcaa cagaaggatg tctcagatga ggatgtggct
1561 cgagccatta accagaagct gggggacacg cctggtgtct cttactccga cattgctgca
1621 cgagcctatg gttgtggccg cacggagctg gccatcaagc tgctggagta tgagccacgc
1681 tcaggggagc aggtacccct tctcctaaag atgaagagga gcaaactggc actaagcaag
1741 gccatcgaga gcggggacac tgacctggtg ttcacggtgt tgctgcacct gaagaacgag
1801 ctgaaccgag gagatttttt catgaccctt cggaatcagc ccatggccct cagtttgtac
1861 cgacagttct gtaagcatca ggagctagag acgctgaagg acctttacaa tcaggatgac
1921 aatcaccagg aattgggcag cttccacatc cgagccagct atgctgcaga agagcgtatt
1981 gaggggcgag tagcagctct gcagacagcc gccgatgcct tctacaaggc caagaatgag
2041 tttgcagcca aggctacaga ggatcaaatg cggctcctac ggctgcagcg gcgcctagaa
```

```
2101 gacgagctgg ggggccagtt cctagacctg tctctacatg acacagttac caccctcatt
2161 cttggcggtc acaacaagcg tgcagagcag ctggcacgtg acttccgcat ccctgacaag
2221 aggctctggt ggctgaagct gactgccctg gcagatttgg aagattggga agagctagag
2281 aagttttcca agagcaagaa atcacccatt ggctacctgc cttttgtgga gatctgcatg
2341 aaacaacata acaaatacga agccaagaag tatgcttccc gcgtgggtcc cgagcagaag
2401 gtcaaggctt tgcttcttgt tggcgatgtg gctcaggctg cagatgtggc catcgaacac
2461 cggaatgagg ctgagctgag cctcgtattg tcccactgca cgggagccac agatggggcc
2521 acagctgaca agattcaacg ggccagggca caagcccaga agaagtgagg agtccatcct
2581 gtacatctca agcaaggggt tcctccccta gcacctgggc ttggcagaag ggccatagtt
2641 catccagctc ctcccctaga gcaatgctga ggagcggggg catggtagca gggctgtctg
2701 gttttaaata aagttggaac acttcaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
2761 aaaaaaaaa
```

[0045] Human vacuolar protein sorting 18 (VPS18) protein has the amino acid sequence (SEQ ID NO:7) (NCBI Reference Sequence: NM_020857.2):

```
MASILDEYEN SLSRSAVLQP GCPSVGIPHS GYVNAQLEKE VPIFTKQRID FTPSERITSL      60
VVSSNQLCMS LGKDTLLRID LGKANEPNHV ELGRKDDAKV HKMFLDHTGS HLLIALSSTE     120
VLYVNRNGQK VRPLARWKGQ LVESVGWNKA LGTESSTGPI LVGTAQGHIF EAELSASEGG     180
LFGPAPDLYF RPLYVLNEEG GPAPVCSLEA ERGPDGRSFV IATTRQRLFQ FIGRAAEGAE     240
AQGFSGLFAA YTDHPPPFRE FPSNLGYSEL AFYTPKLRSA PRAFAWMMGD GVLYGALDCG     300
RPDSLLSEER VWEYPEGVGP GASPPLAIVL TQFHFLLLLA DRVEAVCTLT GQVVLRDHFL     360
EKFGPLKHMV KDSSTGQLWA YTERAVFRYH VQREARDVWR TYLDMNRFDL AKEYCRERPD     420
CLDTVLAREA DFCFRQRRYL ESARCYALTQ SYFEEIALKF LEARQEEALA EFLQRKLASL     480
KPAERTQATL LTTWLTELYL SRLGALQGDP EALTLYRETK ECFRTFLSSP RHKEWLFASR     540
ASIHELLASH GDTEHMVYFA VIMQDYERVV AYHCQHEAYE EALAVLARHR DPQLFYKFSP     600
ILIRHIPRQL VDAWIEMGSR LDARQLIPAL VNYSQGGEVQ QVSQAIRYME FCVNVLGETE     660
QAIHNYLLSL YARGRPDSLL AYLEQAGASP HRVHYDLKYA LRLCAEHGHH RACVHVYKVL     720
ELYEEAVDLA LQVDVDLAKQ CADLPEEDEE LRKKLWLKIA RHVVQEEEDV QTAMACLASC     780
PLLKIEDVLP FFPDFVTIDH FKEAICSSLK AYNHHIQELQ REMEEATASA QRIRRDLQEL     840
RGRYGTVEPQ DKCATCDFPL LNRPFYLFLC GHMFHADCLL QAVRPGLPAY KQARLEELQR     900
KLGAAPPPAK GSARAKEAEG GAATAGPSRE QLKADLDELV AAECVYCGEL MIRSIDRPFI     960
DPQRYEEEQL SWL                                                        973
```

[0046] Nucleic acid (mRNA) encoding human VPS18 protein has the nucleotide sequence (SEQ ID NO:8) (NCBI Reference Sequence: NM_020857.2):

```
   1 gcccgcgtca cgggggcggg agtcagctga gctgccgggg cgaggttggg atcacctggc
  61 accggctgaa gggagcctgt gatttttttg tagcgggggc ggggagtaag gtgcaagact
 121 gcgccagatt caaggacgag ggctgcccga ttatctcgct gcataaggca agagcaagag
 181 gatcctcagg attttaaaga ggaggcgacg gctgcaggtt cccaggatct gtcagaggct
 241 ggggagttac agcttccatt ctggggcgac ggggacccgg ggggggtagc ccttttgtaa
 301 tccccaggcc ccggacaaag agcccagagg ccgggcacca tggcgtccat cctggatgag
 361 tacgagaact cgctgtcccg ctcggccgtc ttgcagcccg gctgccctag cgtgggcatc
 421 ccccactcgg ggtatgtgaa tgcccagctg gagaaggaag tgcccatctt cacaaagcag
 481 cgcattgact tcacccccttc cgagcgcatt accagtcttg tcgtctccag caatcagctg
 541 tgcatgagcc tgggcaagga tacactgctc cgcattgact tgggcaaggc aaatgagccc
 601 aaccacgtgg agctgggacg taaggatgac gcaaaagttc acaagatgtt ccttgaccat
 661 actggctctc acctgctgat tgccctgagc agcacggagg tcctctacgt gaaccgaaat
 721 ggacagaagg tacggccact agcacgctgg aaggggcagc tggtggagag tgtgggttgg
 781 aacaaggcac tgggcacgga gagcagcaca ggccccatcc tggtcgggac tgcccaaggc
 841 cacatctttg aagcagagct ctcagccagc gaaggtgggc ttttcggccc tgctccggat
 901 ctctacttcc gcccattgta cgtgctaaat gaagaagggg gtccagcacc tgtgtgctcc
 961 cttgaggccg agcggggccc tgatgggcgt agctttgtta ttgccaccac tcggcagcgc
1021 ctcttccagt tcataggccg agcagcagag ggggctgagg cccagggttt ctcagggctc
1081 tttgcagctt acacggacca cccaccccca ttccgtgagt ttcccagcaa cctgggctac
1141 agtgagttgg ccttctacac ccccaagctg cgctccgcac cccgggcctt cgcctggatg
1201 atgggggatg gtgtgttgta tggggcattg gactgtgggc gccctgactc tctgctgagc
1261 gaggagcgag tctgggagta cccagagggg gtagggcctg gggccagccc acccctagcc
```

```
1321 atcgtcttga cccagttcca cttcctgctg ctactggcag accgggtgga ggcagtgtgc
1381 acactgaccg ggcaggtggt gctgcgggat cacttcctgg agaaatttgg gccgctgaag
1441 cacatggtga aggactcctc cacaggccag ctgtgggcct acactgagcg ggctgtcttc
1501 cgctaccacg tgcaacggga ggcccgagat gtctggcgca cctatctgga catgaaccgc
1561 ttcgatctgg ccaaagagta ttgtcgagag cggcccgact gcctggacac ggtcctggcc
1621 cgggaggccg atttctgctt tcgccagcgt cgctacctgg agagcgcacg ctgctatgcc
1681 ctgacccaga gctactttga ggagattgcc ctcaagttcc tggaggcccg acaggaggag
1741 gctctggctg agttcctgca gcgaaaactg gccagtttga agccagccga acgtacccag
1801 gccacactgc tgaccacctg gctgacagag ctctacctga gccggcttgg ggctctgcag
1861 ggcgacccag aggccctgac tctctaccga gaaaccaagg aatgctttcg aaccttcctc
1921 agcagccccc gccacaaaga gtggctcttt gccagccggg cctctatcca tgagctgctc
1981 gccagtcatg gggacacaga acacatggtg tactttgcag tgatcatgca ggactatgag
2041 cgggtggtgg cttaccactg tcagcacgag gcctacgagg aggccctggc cgtgctcgcc
2101 cgccaccgtg acccccagct cttctacaag ttctcaccca tcctcatccg tcacatcccc
2161 cgccagcttg tagatgcctg gattgagatg ggcagccggc tggatgctcg tcagctcatt
2221 cctgccctgg tgaactacag ccagggtggt gaggtccagc aggtgagcca ggccatccgc
2281 tacatggagt tctgcgtgaa cgtgctgggg gagactgagc aggccatcca caactacctg
2341 ctgtcactgt atgcccgtgg ccggccggac tcactactgg cctatctgga gcaggctggg
2401 gccagccccc accgggtgca ttacgacctc aagtatgcgc tgcggctctg cgccgagcat
2461 ggccaccacc gcgcttgtgt ccatgtctac aaggtcctag agctgtatga ggaggccgtg
2521 gacctggccc tgcaggtgga tgtggacctg gccaagcagt gtgcagacct gcctgaggag
2581 gatgaggaat tgcgcaagaa gctgtggctg aagatcgcac ggcacgtggt gcaggaagag
2641 gaagatgtac agacagccat ggcttgcctg gctagctgcc ccttgctcaa gattgaggat
2701 gtgctgccct tctttcctga tttcgtcacc atcgaccact tcaaggaggc gatctgcagc
2761 tcacttaagg cctacaacca ccacatccag gagctgcagc gggagatgga agaggctaca
2821 gccagtgccc agcgcatccg gcgagacctg caggagctgc ggggccgcta cggcactgtg
2881 gagccccagg acaaatgtgc cacctgcgac ttccccctgc tcaaccgccc tttttacctc
2941 ttcctctgtg gccatatgtt ccatgctgac tgcctgctgc aggctgtgcg acctggcctg
3001 ccagcctaca agcaggcccg gctggaggag ctgcagagga agctgggggc tgctccaccc
3061 ccagccaagg gctctgcccg ggccaaggag gccgagggtg gggctgccac ggcagggccc
3121 agccgggaac agctcaaggc tgacctggat gagttggtgg ccgctgagtg tgtgtactgt
3181 ggggagctga tgatccgctc tatcgaccgg ccgttcatcg acccccagcg ctacgaggag
3241 gagcagctca gttggctgta ggagggtgtc acctttgatg gggggtgggc aatggggagc
3301 agtggcttga acccacttga gaaggctgcc tcctaggctc tgctcagtca tcttgcaatt
3361 gccacactgt gaccacgttg acgggagtag agtagcgctg ttggccagga ggtgtcaggt
3421 gtgagtgtat tctgccagct tttcatgctg ttcttcagag ctgcagttat gccagaccat
3481 cagcctgcct cccagtagag gcccttcacc tggagaagtc agaaatctga cccaattcca
3541 ccccctgcct ctagcacctc ttctgtccct gtcattcccc acacacgtcc tgttcacctc
3601 gagagagaga gagagagagc acctttcttc cgtctgttca ctctgcggcc tctggaatcc
3661 cagctcttct ctctcagaag aagccttctc ttcctcctgc ctgtaggtgt cccagaagtg
3721 agaaggcagc cttcgaagtc ctgggcattg ggtgagaaag tgatgctagt tggggcatgc
3781 ttttgtgcac actctctggg gctccagtgt gaagggtgcc ctggggctga gggccttgtg
3841 gaggatggtc ggtggtggtg atggaggtgg agagcattaa actgtctgca ctgcaaaaaa
3901 aaaaaaaaaa aaaaaaaaaa aa
```

[0047] Human vacuolar protein sorting 33A (VPS33A) protein has the amino acid sequence (SEQ ID NO:9) (NCBI Reference Sequence: NM_022916.4):

```
MAAHLSYGRV NLNVLREAVR RELREFLDKC AGSKAIVWDE YLTGPFGLIA QYSLLKEHEV      60
EKMFTLKGNR LPAADVKNII FFVRPRLELM DIIAENVLSE DRRGPTRDFH ILFVPRRSLL     120
CEQRLKDLGV LGSFIHREEY SLDLIPFDGD LLSMESEGAF KECYLEGDQT SLYHAAKGLM     180
TLQALYGTIP QIFGKGECAR QVANMMIRMK REFTGSQNSI FPVFDNLLLL DRNVDLLTPL     240
ATQLTYEGLI DEIYGIQNSY VKLPPEKFAP KKQGDGGKDL PTEAKKLQLN SAEELYAEIR     300
DKNFNAVGSV LSKKAKIISA AFEERHNAKT VGEIKQFVSQ LPHMQAARGS LANHTSIAEL     360
IKDVTTSEDF FDKLTVEQEF MSGIDTDKVN NYIEDCIAQK HSLIKVLRLV CLQSVCNSGL     420
KQKVLDYYKR EILQTYGYEH ILTLHNLEKA GLLKPQTGGR NNYPTIRKTL RLWMDDVNEQ     480
NPTDISYVYS GYAPLSVRLA QLLSRPGWRS IEEVLRILPG PHFEERQPLP TGLQKKRQPG     540
ENRVTLIFFL GGVTFAEIAA LRFLSQLEDG GTEYVIATTK LMNGTSWIEA LMEKPF         596
```

[0048] Nucleic acid (mRNA) encoding human VPS33A protein has the nucleotide sequence (SEQ ID NO:10) (NCBI Reference Sequence: NM_022916.4):

```
   1 gtgcgctgcc gtaccggtca cgtggacgtt tggtcacgtg actgcgtccg tggtcctccc
  61 gtaggaaccg gcggactcgg ttggcgttgt ggggcagggg gtggtggagc aagatggcgg
 121 ctcatctgtc ctacggccga gtgaacctaa acgtgttgcg cgaggcggtg cgtcgcgagc
 181 tgcgcgagtt cctggacaag tgcgcaggaa gcaaggcaat agtttgggat gaatacctaa
 241 ctggaccctt tggcctgatt gcacagtatt cactattgaa ggaacatgaa gtggaaaaaa
 301 tgttcacact taaaggaaat cgtttgccgg cagctgatgt gaagaatata attttttttg
 361 tcagacccag gctagagttg atggatataa tcgctgaaaa cgtgctcagt gaagatagac
 421 gaggcccaac gagagatttt catattctgt ttgtgccacg ccgtagcctg ttgtgcgaac
 481 agcggttgaa ggatctgggt gtcttgggat cctttattca cagggaggag tacagcttag
 541 atctcattcc attcgatggg gatctcttat ccatggaatc agagggtgca ttcaaagagt
 601 gctacctgga gggtgaccag acgagcctgt accacgcagc caaggggctg atgaccctgc
 661 aagctctgta tggaacgatc ccccagatct ttgggaaagg agaatgcgct cggcaagtgg
 721 ccaatatgat gatcaggatg aagagagagt ttacaggaag ccagaattca atatttcctg
 781 tttttgataa tctcttgttg cttgatcgga atgtggattt attaacacct cttgccactc
 841 agctgacata tgaaggactc attgatgaaa tttatggcat tcagaacagt tatgtgaaat
 901 tacctccaga gaaatttgca cctaagaaac agggcgatgg tggtaaggac ctccccacgg
 961 aagcaaagaa gctgcagctg aattctgcag aggagctcta tgctgagatc cgagataaga
1021 acttcaacgc agttggctct gtgctcagca agaaagcaaa gatcatctct gcagcattcg
1081 aggaaagaca caatgctaag accgtggggg agatcaagca gtttgtttcc cagttgcccc
1141 acatgcaggc agcaaggggc tcgcttgcaa accatacctc aattgcagaa ttgatcaaag
1201 atgtcactac ttctgaagac ttttttgata aattaaccgt ggaacaggag tttatgtctg
1261 gaatagacac tgataaggtc aacaattaca ttgaggattg tatcgcccaa aagcactcgt
1321 tgatcaaggt gttaagacta gtttgcctcc aatccgtgtg taatagtggg ctcaaacaaa
1381 aagttttgga ttattacaaa agagagattc tccagacata cggctatgag cacatattga
1441 ccttacacaa cctggagaag gccggcctgc tgaaaccgca gacgggggggc agaaacaatt
1501 acccaactat acggaaaaca ttacgcctct ggatggatga tgttaatgag caaaacccca
1561 cggacatatc gtatgtgtac agtgggtatg ccccgctcag tgtgcggctg gcccagctgc
1621 tttcccggcc tggctggcgg agcatcgagg aggtcctccg catcctccca gggccccact
1681 ttgaggagcg gcagccactg cccacaggac tgcagaagaa acgtcaaccg ggagaaaacc
1741 gagtgactct gatatttttc cttgggggcg taaccttcgc tgaaattgct gccctgcgat
1801 ttctctccca gttggaagat ggaggtacag aatatgtcat tgccaccact aaactaatga
1861 atggaaccag ttggatagag gctctgatgg aaaaaccttt ctaggatgtt cagaggagac
1921 ttaacaagtg tactgcagaa taaactacct ctttgaagaa attgctgaaa ggaagtaaaa
1981 ccccatagaa gaaacatggg aatacagaat atattctggg gtcagcttct taaattatac
2041 tactgtttac tgctttctcc gtctcttttg tattcccttt ttttttttcc tttgagacgg
2101 agtcttgctc tgtcacccag actagagtgc agtggcacgg tctcacctca ctgcaacctc
2161 cacctcctag gttcaagcaa ttctcctgcc tcagcctcct gagtagctgg gactacaggc
2221 atgcaccacc acacccggct aatttttgta tttttagtag ccatggtgtt tcaccatgtt
2281 ggccaggctg gtctcaaact cctgacctca ggtgatccac ctgcctcggc ctcccacagt
2341 gctgggatta caggcctgag ccaccgtgcc tggcccctaa tctgctgaag aaagaagaat
2401 agaagaaaat caacctgagt aaaagcagca ctggttttga gttttctaag ctcagggtct
2461 tcattagaga cctctggaaa tacattaagg atggtggggg tagataatcc attcagccag
2521 acaaacgggg ccagctctta aaataagaaa gctgagactg aggaggtgaa actgaaaata
2581 aaaagagaaa gttcatcctc taaaaaaaaa aaaaaaaaaa aaaaaaaa
```

[0049] Human vacuolar protein sorting 39 (VPS39) protein has the amino acid sequence (SEQ ID NO:11) (NCBI Reference Sequence: NM_015289.2):

```
MHDAFEPVPI LEKLPLQIDC LAAWEEWLLV GTKQGHLLLY RIRKDVGCNR FEVTLEKSNK      60
NFSKKIQQIH VVSQFKILVS LLENNIYVHD LLTFQQITTV SKAKGASLFT CDLQHTETGE     120
EVLRMCVAVK KKLQLYFWKD REFHELQGDF SVPDVPKSMA WCENSICVGF KRDYYLIRVD     180
GKGSIKELFP TGKQLEPLVA PLADGKVAVG QDDLTVVLNE EGICTQKCAL NWTDIPVAME     240
HQPPYIIAVL PRYVEIRTFE PRLLVQSIEL QRPRFITSGG SNIIYVASNH FVWRLIPVPM     300
ATQIQQLLQD KQFELALQLA EMKDDSDSEK QQQIHHIKNL YAFNLFCQKR FDESMQVFAK     360
LGTDPTHVMG LYPDLLPTDY RKQLQYPNPL PVLSGAELEK AHLALIDYLT QKRSQLVKKL     420
NDSDHQSSTS PLMEGTPTIK SKKKLLQIID TTLLKCYLHT NVALVAPLLR LENNHCHIEE     480
```

```
SEHVLKKAHK YSELIILYEK KGLHEKALQV LVDQSKKANS PLKGHERTVQ YLQHLGTENL     540
HLIFSYSVWV LRDFPEDGLK IFTEDLPEVE SLPRDRVLGF LIENFKGLAI PYLEHIIHVW     600
EETGSRFHNC LIQLYCEKVQ GLMKEYLLSF PAGKTPVPAG EEEGELGEYR QKLLMFLEIS     660
SYYDPGRLIC DFPFDGLLEE RALLLGRMGK HEQALFIYVH ILKDTRMAEE YCHKHYDRNK     720
DGNKDVYLSL LRMYLSPPSI HCLGPIKLEL LEPKANLQAA LQVLELHHSK LDTTKALNLL     780
PANTQINDIR IFLEKVLEEN AQKKRFNQVL KNLLHAEFLR VQEERILHQQ VKCIITEEKV     840
CMVCKKKIGN SAFARYPNGV VVHYFCSKEV NPADT                                875
```

[0050] Nucleic acid (mRNA) encoding human VPS39 protein has the nucleotide sequence (SEQ ID NO:12) (NCBI Reference Sequence: NM_015289.2):

```
   1 ggggttgacg atggctgtgt tgttgaaggg cctgtagccg gggggttcct ggccggatcc
  61 cggtctaccc ttagcccaga ctcgttccgg accccagccc ggcccggaac actctgggcg
 121 agacggcggt ggcaactctc cccttgccgc catgcacgac gctttcgagc cagtgccgat
 181 cctagaaaag ctgcctctgc aaatcgactg tctggctgcc tgggaggaat ggcttcttgt
 241 gggaaccaaa caaggacatc ttcttctcta taggattcgg aaggacgttg gttgcaacag
 301 atttgaagtg acactagaga aatccaataa gaacttctcc aaaaagattc agcagatcca
 361 tgtggtttcc cagtttaaga ttctggtcag cttgttagaa aataacattt atgtccatga
 421 cctattgaca tttcaacaaa tcactacggt ttcaaaggca aagggagcat cactgtttac
 481 ttgtgacctc cagcacacag agaccggtga ggaggtgtta cggatgtgtg tggcagtaaa
 541 aaagaagctg cagctctatt tctggaagga cagggaattt catgaattgc agggggactt
 601 tagtgtgcca gatgtgccca agtccatggc gtggtgtgaa aattctatct gtgtgggttt
 661 caagagagac tactacctaa taagggtgga tggaaagggg tccatcaaag agctctttcc
 721 aacaggaaaa cagctggagc ccttagttgc acctctggca gatggaaaag tggctgtggg
 781 ccaggatgat ctcaccgtgg tactcaatga ggaagggatc tgcacacaga aatgtgccct
 841 gaactggacg gacataccag tggccatgga gcaccagcct ccctacatca ttgcagtgtt
 901 gcctcgatat gttgagatcc gaacatttga accgaggctt ctggtccaaa gcattgaatt
 961 gcaaaggccc cgtttcatta cctcaggagg atcaaacatt atctatgtgg ccagcaatca
1021 ttttgtttgg agactcatcc ctgtccccat ggcaacccaa atccaacaac ttctccagga
1081 caagcagttt gaattggctc tgcagctcgc agaaatgaaa gatgattctg acagtgaaaa
1141 gcagcaacaa attcatcaca tcaagaactt gtatgccttc aacctcttct gccagaagcg
1201 ttttgatgag tccatgcagg tctttgctaa acttggcaca gatcccaccc atgtgatggg
1261 cctgtaccct gacctgctgc ccacagacta cagaaagcag ttgcagtatc ccaacccatt
1321 gcctgtgctc tccggggctg aattggagaa ggctcactta gctctgattg actacctgac
1381 acagaaacga agtcaattgg taaagaagct gaatgactct gatcaccagt caagcacctc
1441 accgctcatg gaaggcactc ccaccatcaa atccaagaag aagctgctac aaatcatcga
1501 caccaccctg ctcaagtgct atctccatac aaatgtggcc ctggtggccc ccttgctacg
1561 cctggagaac aatcactgcc acatcgagga gagcgagcac gtgctaaaga aggctcacaa
1621 gtacagtgag cttatcatcc tgtatgagaa gaaggggctc cacgagaaag ctctgcaggt
1681 gctcgtggac cagtccaaga aagccaactc ccctctgaaa ggccacgaga ggacagtgca
1741 gtatctgcag catctgggca cagaaaacct gcatttgatt ttctcctact cagtgtgggt
1801 gctgagagac ttcccagaag atggcctgaa gatatttact gaagatctcc cggaagtgga
1861 gtctctgcca cgtgatcgag tcctcggctt cttaatagag aattttaagg gtctggctat
1921 tccttatctg gaacacatca tccatgtttg ggaggagaca ggctctcggt tccacaactg
1981 cctgatccag ctatactgtg agaaggtgca aggtctgatg aaggagtatc tcctgtcctt
2041 ccctgcaggc aaaaccccag tcccagctgg agaggaagag ggtgagctgg gagaataccg
2101 gcaaaagctc ctcatgttct tggagatttc cagctactat gatccaggcc ggctcatctg
2161 tgattttccc tttgatggcc tcttagaaga acgagctctc ctgttggggc gcatggggaa
2221 acatgaacaa gctcttttca tttatgtcca catcttgaag gatacaagga tggctgagga
2281 gtactgccac aaacactatg accgaaacaa agatggcaac aaagatgtgt atctgtccct
2341 gcttcggatg tacctgtcgc cccccagcat tcactgcctg gggccaatca agctggaact
2401 actggagcca aaagccaacc tccaggccgc tctgcaggtc ctcgagctac accacagcaa
2461 actggacacc accaaggccc tcaaccttct gccagcaaac actcagatca atgacatacg
2521 catcttcctg gaaaaggtct tggaagaaaa tgcacaaaag aaacggttca atcaagtgct
2581 caagaacctt ctccatgcag aattcctgag ggtccaggaa gagcggattt tacaccagca
2641 ggtgaagtgc atcatcacag aggagaaggt gtgcatggtg tgtaagaaga agattgggaa
2701 cagtgcattt gcaagatacc ccaatggagt ggtcgtccat tacttctgtt ccaaagaggt
2761 aaacccagct gacacttgag cccagcatcc tggggatcca gcggatggac agcttggctc
2821 tcccagagag gtgaaggagc acctggcctt aggaatcctg gctgccacca ccacaaggct
2881 ccccatttgg acattactgg ctatcttgtg ccctggaaca actctgaatt aattagactc
```

```
2941 atggtctggc attgccagct ttttaatggg aaaagagatt agttatacct tataccatta
3001 tgttgtgggc aattccagag aattcagtac ctgcttggtc aggaggatgt gcaccatctt
3061 gcctttgcac accagtcacc tgaacaagga aacttgtcac aagtgtttgt aaccatgggg
3121 ttgttcatca agggcttttc tattaagtac atgacttcac aaggaccgct cagcatggct
3181 cactggagag ttccatgaga gaacagcact caagcttctg gccgcatgga cccgatggct
3241 cgcattctgt gtagtgtttt acgtctccat ggtaactgtg ccctgcaccc ctcggtagcc
3301 gccctgttag ttttcagtct ccttttcttt ctcaccattt atcacttccc tcactgccct
3361 acccaggctt tctctcccac ttccctgact ctgggaataa ctaatattta agcaaggtaa
3421 gatgagaagc aaggggtctc agttctagga atacagtgct agttgattgt caggtatgtt
3481 gtaaatagac cctctttggc catacactcc atgcctagat gcctcggaga gcatcattct
3541 ctgcctaggc aaggccctgc atcccttgcc tcaggccggg ctgagtgtga ctgcagctcc
3601 tgaggatggg cctgccctgt ctggggtatg cgtgatccct agatacatgt tcccacagag
3661 gtgcctgctc cgtcttcgct caccagacac tcaggcaggc tggcttagtc tttgtgcgtg
3721 gcgattttgt gctctgggcc ctttctcttt ttccagccag tttccattca cttgccttac
3781 agcctgccct ggccgtcact ccccagcttt gttcagcaat ggtgtggttg gagagttgtg
3841 ctgggatagc gcaggaaggt gggtcccggc aacacgcagg ggatgagtgg acctggaact
3901 gacaatggcg tgctgccaag tgttcctgag aggtgtttag gcacagcaga ggggacgcgg
3961 ggggcaagaa cagcaggacg ctggtttaaa aataactcac cgccaaacct gtggagcagt
4021 gtggggcatc ctgccagagg tgcacaggct ggagtttcag gcactgcagg ctgatgacac
4081 acagggagag tggccctgcc tcctgtcctc cccggggttt ttgcagactc gaagtctcac
4141 tgcaccagtg tctttgatgg tggtgagggt gggtgatggt gcccagcacc aacagtttta
4201 gtggcctgtc cttgacctgc cgtggtcctt tgtaaactat ggctccatgc tgtgtgacag
4261 atcaacgtgc tgatggtaag tagactaggc ttccccaggc atgccgtccg tgggggcctg
4321 aagagacagt gagtgccatt ggccccattc gcagatgtgg gagactctgc tcaggcctgt
4381 gaggctgggc agcccttcac cagagttcgg aggagcagtg tgtggcgcca cgtcccgact
4441 ggccataccc acacagaagc agtgctgccc ggggcctcat ctgggccagc ttggactctg
4501 cttcctccag gagcagcagg gaagctctgg gccacctccc tggatagcag gaacttgacc
4561 tgccatgtgt gccctgcctt cctggccagc tgtgcttgtt atcttccatt ctcacaaact
4621 gtctttgaag caatagaata aagaatgtgt gttttctttc ctggtataca tacatgatcc
4681 catgctccca agctccattc ttccttccct caactctctg ccctccacag agctatggag
4741 aaggctggag atgaaagctt tgtagtgagg actgataaag atctcatcac tgctccttat
4801 aataaaccta ataaagcaag aaaccaagcc taaaaaaaaa aaaaaaaaaa a
```

[0051] Human vacuolar protein sorting 41 (VPS41) protein has the amino acid sequence (SEQ ID NO:13) (NCBI Reference Sequence: NM_014396.3):

```
MAEAEEQETG SLEESTDESE EEESEEEPKL KYERLSNGVT EILQKDAASC MTVHDKFLAL      60
GTHYGKVYLL DVQGNITQKF DVSPVKINQI SLDESGEHMG VCSEDGKVQV FGLYSGEEFH     120
ETFDCPIKII AVHPHFVRSS CKQFVTGGKK LLLFERSWMN RWKSAVLHEG EGNIRSVKWR     180
GHLIAWANNM GVKIFDIISK QRITNVPRDD ISLRPDMYPC SLCWKDNVTL IIGWGTSVKV     240
CSVKERHASE MRDLPSRYVE IVSQFETEFY ISGLAPLCDQ LVVLSYVKEI SEKTEREYCA     300
RPRLDIIQPL SETCEEISSD ALTVRGFQEN ECRDYHLEYS EGESLFYIVS PRDVVVAKER     360
DQDDHIDWLL EKKKYEEALM AAEISQKNIK RHKILDIGLA YINHLVERGD YDIAARKCQK     420
ILGKNAALWE YEVYKFKEIG QLKAISPYLP RGDPVLKPLI YEMILHEFLE SDYEGFATLI     480
REWPGDLYNN SVIVQAVRDH LKKDSQNKTL LKTLAELYTY DKNYGNALEI YLTLRHKDVF     540
QLIHKHNLFS SIKDKIVLLM DFDSEKAVDM LLDNEDKISI KKVVEELEDR PELQHVYLHK     600
LFKRDHHKGQ RYHEKQISLY AEYDRPNLLP FLRDSTHCPL EKALEICQQR NFVEETVYLL     660
SRMGNSRSAL KMIMEELHDV DKAIEFAKEQ DDGELWEDLI LYSIDKPPFI TGLLNNIGTH     720
VDPILLIHRI KEGMEIPNLR DSLVKILQDY NLQILLREGC KKILVADSLS LLKKMHRTQM     780
KGVLVDEENI CESCLSPILP SDAAKPFSVV VFHCRHMFHK ECLPMPSMNS AAQFCNICSA     840
KNRGPGSAIL EMKK                                                       854
```

[0052] Nucleic acid (mRNA) encoding human VPS41 protein has the nucleotide sequence (SEQ ID NO:14) (NCBI Reference Sequence: NM_014396.3):

```
  1 ctgtcaggtg actctcccgt ggcgccatgg cggaagcaga ggagcaggaa actgggtccc
 61 ttgaagaatc tacagatgag tctgaggaag aagagagcga agaggaaccc aagctgaagt
121 atgaaaggct ttccaatggg gtaactgaaa tacttcagaa ggatgcagct agctgcatga
181 cagtccatga caagtttttg gcattgggca cacattatgg caaggtttat ttacttgatg
```

```
 241 tccaggggaa catcactcag aagtttgatg taagtcctgt gaagataaat cagattagct
 301 tggatgaaag tggagagcac atgggtgtgt gttcagagga tggcaaggtg caggtatttg
 361 gactgtattc tggagaagaa tttcacgaga cttttgactg tcccattaaa attattgctg
 421 tgcacccaca tttcgtgaga tccagttgca agcagtttgt gaccggaggg aagaagctgc
 481 tactgtttga acggtcttgg atgaacagat ggaagtctgc tgttctgcat gaaggggaag
 541 ggaacataag gagtgtgaag tggagaggcc atctgattgc ttgggccaat aatatgggtg
 601 tgaagatttt tgacatcatc tcaaagcaaa gaatcaccaa tgtgccccgg gatgatataa
 661 gtcttcgccc agacatgtat ccctgcagcc tctgctggaa ggacaatgtg acactgatta
 721 ttggctgggg gacttctgtc aaggtgtgct cagtgaagga acggcatgcc agtgaaatga
 781 gggatttgcc aagtcgatat gttgaaatag tgtctcagtt tgaaactgaa ttctacatca
 841 gtggacttgc acctctctgt gatcagcttg ttgtactttc gtatgtaaag gagatttcag
 901 aaaaaacgga aagagaatac tgtgccaggc ctagactgga catcatccag ccactttctg
 961 agacttgtga agagatctct tctgatgctt tgacagtcag aggctttcag gagaatgaat
1021 gtagagatta tcatttagaa tactctgaag gggaatcact tttttacatc gtgagtccga
1081 gagatgttgt agtggccaag gaacgagacc aagatgatca cattgactgg ctccttgaaa
1141 agaagaaata tgaagaagca ttgatggcag ctgaaattag ccaaaaaaat attaaaagac
1201 ataagattct ggatattggc ttggcatata taaatcacct ggtggagaga ggagactatg
1261 acatagcagc acgcaaatgc cagaaaattc ttgggaaaaa tgcagcactc tgggaatatg
1321 aagtttataa atttaaagaa attggacagc ttaaggctat tagtccttat ttgccaagag
1381 gtgatccagt tctgaaacca ctcatctatg aaatgatctt acatgaattt ttggagagtg
1441 attatgaggg ttttgccaca ttgatccgag aatggcctgg agatctgtat aataattcag
1501 tcatagttca agcagttcgg gatcatttga agaaagatag tcagaacaag actttactta
1561 aaaccctggc agaattgtac acctatgaca agaactatgg caatgctctg gaaatatact
1621 taacattaag acataaagac gtttttcagt tgatccacaa gcataatctt ttcagttcta
1681 tcaaggataa aattgtttta ttaatggatt ttgattcaga gaaagctgtt gacatgcttt
1741 tggacaatga agataaaatt tcaattaaaa aggtagtgga agaattggaa gacagaccag
1801 agctacagca tgtgtatttg cataagcttt tcaagagaga ccaccataag gggcagcgtt
1861 accatgaaaa acagatcagt ctttatgctg aatatgatcg accaaactta cttccctttc
1921 tccgagacag tacccattgc ccacttgaaa aggctcttga gatctgtcaa cagagaaact
1981 ttgtagaaga gacagtttat cttctgagcc gaatgggtaa tagccgaagt gccctgaaga
2041 tgattatgga ggaattacat gatgttgata aagcaatcga atttgccaag gagcaagatg
2101 atggagagct gtgggaagat ttgattttat attccattga caaaccacca tttattactg
2161 gcttgttaaa caacattggc acacatgttg acccaattct actgattcac cgtattaagg
2221 aaggaatgga gatccccaat ttgagagatt ccttggttaa aattctgcaa gactacaatt
2281 tgcaaattct gcttcgtgaa ggctgcaaga agattctcgt agctgactct ttgtccttac
2341 tgaagaaaat gcaccgaact caaatgaaag gtgttcttgt tgatgaggag aacatctgtg
2401 agtcgtgcct ttcccctatt cttccatcag atgcagctaa gcccttcagc gtggtggtct
2461 tccattgccg gcacatgttc cacaaggagt gcctgcccat gcccagcatg aactctgctg
2521 cacagttctg caacatctgc agtgctaaga accgtggacc aggaagtgca attttggaga
2581 tgaaaaaata gctcatttct ccttgtcagt ctccttgtca ccactctttt tgagactgtt
2641 tttgcaacaa caaaagcatt tgttgacact cgtgctgtta agagatttgt ttatgtttat
2701 attatactca aaaacaattt cttcatctat tcctgtacta atggtttctc tttgcagttc
2761 acagagaatt tggggctctc ttcatgcctt gaaattttgg ggtccatagt gaatattttg
2821 ttatttattt gtttggctca ttctttatat agtaatggaa acataagtct aggagttaga
2881 aatgaatttt ttagacctta gtaaaaccat ttaaccataa aatggacaac tgagaattct
2941 cccagctgcc tgaaagcgtc gccaactgtg gttatcctgc aagctgctac ctgcaacttg
3001 gacgttgttt ccacgtgctc tgctggctac gattcttgca ttctgggttt ggctttttc
3061 tgtgtcatca actatggtta tcctctaaat aggcatttaa tgaaacattg tacaaattgt
3121 cactcatttg atgacacctg ggaataacat tagcaggctg atgtcctgca ccattatgtt
3181 tactaatcac atgttctgtg tgctgtgacg actgtcaaag agtatctggc catggcggac
3241 actcagcatt tgttgattga ataaatgtta gctcttctca ttgtgaagga ctcactttta
3301 ctgggataaa caaatgcagt taagaattct ggcacccttg taaggaagaa aagagagttc
3361 aacaccttcg agtctgagcg cttgtggcta gagtttgcca ggagggagga aaccagtgac
3421 cctgaaaact gagggtgcct caggagcagt gggaccacct gatgctgaag gacggactaa
3481 tgatgtttcc tcttgccttc tctggtgcct ccattgccct catggaacag agcatatcat
3541 agagggagaa aagtcaaact tgtaattgtg tcttacagtt actggcttca tcttccttgg
3601 gatatatggt catcctctaa tgagtgtaaa agtgcgcaaa acacatcctt attgttcctg
3661 atctcttagt cccataaatg ggaacaaata cagctttctg cttctttctt tttggggaaa
3721 ggacagggtg ctagtgagta ctgacagcat gccagctacc gaagtcaccc agccattccc
3781 atgagcagca gttcatttaa ttgtcacagc gtcgccagga agaagatctg ataaacctag
```

```
3841 gtttacagat aaagaaagca aaatgtagag atgttgttga ggtcacagag gtgactgcct
3901 aacttcagag cagggcttct gatcccttta agaaattaca gggccagccg ggcatggtgg
3961 ctcacgcccg taatcccagg gctttgggag gccttggcag gtggatcacc tgagatcgca
4021 cgttcgagac cagcctgacc aacatggaga aaccccatct ctactaaaaa cacaaattag
4081 ccaggcgtgg tggtacatgc ctgtaatccc agctactcag gaggctgagg caggagaatc
4141 acttgacccc aggagacgta ggttgtggtg agctgagatc gcgccattgc actccagcct
4201 gggcaacaag agcaaaactc cgtctcaaaa aagaaaagaa aagaaaagaa atcatagggc
4261 caagttcaaa ggaaatgcac agaacatatc ttcacattag agttaagaat tctctagcaa
4321 acaacagatt tttttgttgt tgttagtcac aaatacttag aactggaagg ctctttgtta
4381 ttattgaatg tacccctcag ccttctcagc atttccttat cccaagacta gtgtgctttc
4441 tgctacactg ctagttttca gttttgttct tacccaattg ttttttcttt tcaacattac
4501 caatttacag attcagttta ttacatttac attaatcctc acttatgatt tgagcaagct
4561 catttccaga aaagtttact ttaagatcat caataggatt tgctaatttc agtgaagtca
4621 ttttgcttca ggggtaaatt atcctagtta ccaagtccta tttggacata aagaaaatcc
4681 tacttataga aaaggagaaa ataattaaac agtcttcatt tttaagtaac tgatttaaaa
4741 ggaaaataat aaaatatgtt cgtttatcat ttcagaaatt gctgtaacac actggaaaat
4801 tcctgaacaa tatagatttt atcgttaata aaaaacacta gctttcgttc cttagaatgt
4861 cttttctttt gaataaacag tattgggtga ttta
```

[0053] Human BLOC1S1 protein has the amino acid sequence (SEQ ID NO:15) (NCBI Reference Sequence: NM_001487.3):

```
MAPGSRGERS SFRSRRGPGV PSPQPDVTML SRLLKEHQAK QNERKELQEK RRREAITAAT      60
CLTEALVDHL NVGVAQAYMN QRKLDHEVKT LQVQAAQFAK QTGQWIGMVE NFNQALKEIG     120
DVENWARSIE LDMRTIATAL EYVYKGQLQS APS                                   153
```

[0054] Nucleic acid (mRNA) encoding human BLOC1S1 protein has the nucleotide sequence (SEQ ID NO:16) (NCBI Reference Sequence: NM_001487.3):

```
  1 acacagcggt cacgtgacat ggccccgggg agccgaggtg agcgttccag cttccggagc
 61 cggagggggc ccggcgtacc cagcccccag cccgacgtga ccatgctgtc ccgcctccta
121 aaagaacacc aggccaagca gaatgaacgc aaggagctgc aggaaaagag gaggcgagag
181 gctatcactg cagcgacctg cctgacagaa gctttggtgg atcacctcaa tgtgggtgtg
241 gcccaggcct acatgaacca gagaaagctg gaccatgagg tgaagaccct acaggtccag
301 gctgcccaat ttgccaagca gacaggccag tggatcggaa tggtggagaa cttcaaccag
361 gcactcaagg aaattgggga tgtggagaac tgggctcgga gcatcgagct ggacatgcgc
421 accattgcca ctgcactgga atatgtctac aaagggcagc tgcagtctgc cccttcctag
481 cccctgttcc ctcccccaac cctatccctc ctacctcacc cgcaggggga aggagggagg
541 ctgacaagcc ttgaataaaa cacaagcctc cgtttctcaa aaaaaaaaaa
```

[0055] Human BLOC1S2 protein has the amino acid sequence (SEQ ID NO:17) (NCBI Reference Sequence: NM_173809.2):

```
MAAAAEGVLA TRSDEPARDD AAVETAEEAK EPAEADITEL CRDMFSKMAT YLTGELTATS      60
EDYKLLENMN KLTSLKYLEM KDIAINISRN LKDLNQKYAG LQPYLDQINV IEEQVAALEQ     120
AAYKLDAYSK KLEAKYKKLE KR                                               142
```

[0056] Nucleic acid (mRNA) encoding human BLOC1S2 protein has the nucleotide sequence (SEQ ID NO:18) (NCBI Reference Sequence: NM_173809.2):

```
  1 ccggaaacag cgcggggtcc gctatggcgg cggcagccga gggcgtactg gcgacccgga
 61 gtgatgagcc cgcccgagac gatgccgccg tggagacagc tgaggaagca aaggagcctg
121 ctgaagctga catcactgag ctctgccggg acatgttctc caaaatggcc acttacctga
181 ctggggaact gacggccacc agtgaagact ataagctcct ggaaaatatg aataaactca
241 ccagcttgaa gtatcttgaa atgaaagata ttgctataaa cattagtagg aacttaaagg
301 acttaaacca gaaatatgct ggactgcagc cttatctgga tcagatcaat gtcattgaag
361 agcaggtagc agctcttgag caggcagctt acaagttgga tgcatattca aaaaaactgg
421 aagccaagta caagaagctg gagaagcgat gagaaactta tttctatggg acagagtctt
481 tttttttaa tgtggaagaa tgtcttataa aacctgaatc ctgaggctga tgaattgtga
```

```
 541 aaattcctca aaaggaaatt atgctggtca tcacaggaac atctcaacgt tcgagtaaac
 601 tggaggactg tggctattcc tgaaccttct ttgagacaga atccctcaga atctcacact
 661 tataacttcc tacctttac ttgaatgctt tgccatattc aggacagaga ctctcacaaa
 721 gttcagaaaa cagctggact taccagtaaa atcaaatgag aggacctatt ttctctggta
 781 gtggttgatt actacattat tttcttaagt ggctggtttt ttagttacta tgtaaatggt
 841 cgtttttctg ttaatgatgc taatgtgttg taaacaagat tctaaattta aaaaggaaaa
 901 caaaacaaac ttgttctttg cagcttatca ccttgtgaat gtcggtaact tacttttcca
 961 taatattgca aataacataa aatcttaaaa taattccaag ctgagtcttc tagattgagc
1021 agaaatggtg aaaggagtat tgataacttg gcgtatgtga tgggcccctc ttgtttattt
1081 tctatgtgag tcacattgac atgcgatcag tttgggaaat gtgatgaaaa caaagactag
1141 atgggtatgt gtgtttatgt gttgggtagg gaggtgacga ttgccactca taaaataaag
1201 gattttataa aataccttcc tactgtgtat gtaggatttg gggggatctt agggacctaa
1261 tcgacttctt tgcacactaa aaacatcaga caatgggaca tactgactga ccagtctagg
1321 ttgaaagata ggcagcctta cccagaacac aacattagca gctgggaagg tgtctgaggt
1381 ccccaattac atatcccaaa gagtttctta cttctgtttc tgtcatttcc cctctgttcc
1441 agacagtcat catttatcct ctgttcttcc tggactgttg ctgtggtctc ttctcatctt
1501 aactgtccct tccaaccacc ctccacactg ctgccagagc aatcttaaaa atgtaaactg
1561 gccctattac tcctgcttag aaccctgtag tgacttatca tggcctctga aaaatctaga
1621 ctttcattat gcatacaagt cccttgtgtt tttttgtttg tttatagaca gggtctctat
1681 tgtcacccag gctggagtgt ggtggtgtga taatagctca ccttgacctc ctaggctcaa
1741 ctgatcttcc cacctcagcc tcctgaatag ctgggactac cagcacgctc caccatgcct
1801 tgctaattat tttttataca gatggagtct cactatgttg tccaggctgg tcttgaactg
1861 gcctcaagtg attctcttac cacagcctgc cagagtgctg agattgtagg catgaaccac
1921 cgtgcctaac ataaggcccc tatttaaaca tttttctt
```

[0057] Human N-acetylglucosamine-1-phosphate transferase, alpha and beta subunits (GNPTAB) protein has the amino acid sequence (SEQ ID NO:19) (NCBI Reference Sequence: NM_024312.4):

```
MLFKLLQRQT YTCLSHRYGL YVCFLGVVVT IVSAFQFGEV VLEWSRDQYH VLFDSYRDNI      60
AGKSFQNRLC LPMPIDVVYT WVNGTDLELL KELQQVREQM EEEQKAMREI LGKNTTEPTK     120
KSEKQLECLL THCIKVPMLV LDPALPANIT LKDLPSLYPS FHSASDIFNV AKPKNPSTNV     180
SVVVFDSTKD VEDAHSGLLK GNSRQTVWRG YLTTDKEVPG LVLMQDLAFL SGFPPTFKET     240
NQLKTKLPEN LSSKVKLLQL YSEASVALLK LNNPKDFQEL NKQTKKNMTI DGKELTISPA     300
YLLWDLSAIS QSKQDEDISA SRFEDNEELR YSLRSIERHA PWVRNIFIVT NGQIPSWLNL     360
DNPRVTIVTH QDVFRNLSHL PTFSSPAIES HIHRIEGLSQ KFIYLNDDVM FGKDVWPDDF     420
YSHSKGQKVY LTWPVPNCAE GCPGSWIKDG YCDKACNNSA CDWDGGDCSG NSGGSRYIAG     480
GGGTGSIGVG QPWQFGGGIN SVSYCNQGCA NSWLADKFCD QACNVLSCGF DAGDCGQDHF     540
HELYKVILLP NQTHYIIPKG ECLPYFSFAE VAKRGVEGAY SDNPIIRHAS IANKWKTIHL     600
IMHSGMNATT IHFNLTFQNT NDEEFKMQIT VEVDTREGPK LNSTAQKGYE NLVSPITLLP     660
EAEILFEDIP KEKRFPKFKR HDVNSTRRAQ EEVKIPLVNI SLLPKDAQLS LNTLDLQLEH     720
GDITLKGYNL SKSALLRSFL MNSQHAKIKN QAIITDETND SLVAPQEKQV HKSILPNSLG     780
VSERLQRLTF PAVSVKVNGH DQGQNPPLDL ETTARFRVET HTQKTIGGNV TKEKPPSLIV     840
PLESQMTKEK KITGKEKENS RMEENAENHI GVTEVLLGRK LQHYTDSYLG FLPWEKKKYF     900
QDLLDEEESL KTQLAYFTDS KNTGRQLKDT FADSLRYVNK ILNSKFGFTS RKVPAHMPHM     960
IDRIVMQELQ DMFPEEFDKT SFHKVRHSED MQFAFSYFYY LMSAVQPLNI SQVFDEVDTD    1020
QSGVLSDREI RTLATRIHEL PLSLQDLTGL EHMLINCSKM LPADITQLNN IPPTQESYYD    1080
PNLPPVTKSL VTNCKPVTDK IHKAYKDKNK YRFEIMGEEE IAFKMIRTNV SHVVGQLDDI    1140
RKNPRKFVCL NDNIDHNHKD AQTVKAVLRD FYESMFPIPS QFELPREYRN RFLHMHELQE    1200
WRAYRDKLKF WTHCVLATLI MFTIFSFFAE QLIALKRKIF PRRRIHKEAS PNRIRV        1256
```

[0058] Nucleic acid (mRNA) encoding human GNPTAB protein has the nucleotide sequence (SEQ ID NO:20) (NCBI Reference Sequence: NM_024312.4):

```
   1 gctcccggaa gcggcggccg cggcgcggag ccgagcgggc gtccgtcgcc ggagctgcaa
  61 tgagcggcgc ccggaggctg tgacctgcgc gcggcggccc gaccggggcc cctgaatggc
 121 ggctcgctga ggcggcggcg gcggcggcgg cggctcaggc tcctcggggc gtggcgtggc
 181 ggtgaagggg tgatgctgtt caagctcctg cagagacaga cctatacctg cctgtcccac
 241 aggtatgggc tctacgtgtg cttcttgggc gtcgttgtca ccatcgtctc cgccttccag
```

```
 301 ttcggagagg tggttctgga atggagccga gatcaatacc atgttttgtt tgattcctat
 361 agagacaata ttgctggaaa gtcctttcag aatcggcttt gtctgcccat gccgattgac
 421 gttgtttaca cctgggtgaa tggcacagat cttgaactac tgaaggaact acagcaggtc
 481 agagaacaga tggaggagga gcagaaagca atgagagaaa tccttgggaa aaacacaacg
 541 gaacctacta agaagagtga gaagcagtta gagtgtttgc taacacactg cattaaggtg
 601 ccaatgcttg tcctggaccc agccctgcca gccaacatca ccctgaagga cctgccatct
 661 ctttatcctt cttttcattc tgccagtgac attttcaatg ttgcaaaacc aaaaaaccct
 721 tctaccaatg tctcagttgt tgtttttgac agtactaagg atgttgaaga tgcccactct
 781 ggactgctta aaggaaatag cagacagaca gtatggaggg gctacttgac aacagataaa
 841 gaagtccctg gattagtgct aatgcaagat ttggctttcc tgagtggatt tccaccaaca
 901 ttcaaggaaa caaatcaact aaaaacaaaa ttgccagaaa atctttcctc taaagtcaaa
 961 ctgttgcagt tgtattcaga ggccagtgta gcgcttctaa aactgaataa ccccaaggat
1021 tttcaagaat tgaataagca aactaagaag aacatgacca ttgatggaaa agaactgacc
1081 ataagtcctg catatttatt atgggatctg agcgccatca gccagtctaa gcaggatgaa
1141 gacatctctg ccagtcgttt tgaagataac gaagaactga ggtactcatt gcgatctatc
1201 gagaggcatg caccatgggt tcggaatatt ttcattgtca ccaacgggca gattccatcc
1261 tggctgaacc ttgacaatcc tcgagtgaca atagtaacac accaggatgt ttttcgaaat
1321 ttgagccact tgcctacctt tagttcacct gctattgaaa gtcacattca tcgcatcgaa
1381 gggctgtccc agaagtttat ttacctaaat gatgatgtca tgtttgggaa ggatgtctgg
1441 ccagatgatt tttacagtca ctccaaaggc cagaaggttt atttgacatg gcctgtgcca
1501 aactgtgccg agggctgccc aggttcctgg attaaggatg gctattgtga caaggcttgt
1561 aataattcag cctgcgattg ggatggtggg gattgctctg gaaacagtgg agggagtcgc
1621 tatattgcag gaggtggagg tactgggagt attggagttg gacagccctg gcagtttggt
1681 ggaggaataa acagtgtctc ttactgtaat cagggatgtg cgaattcctg gctcgctgat
1741 aagttctgtg accaagcatg caatgtcttg tcctgtgggt ttgatgctgg cgactgtggg
1801 caagatcatt ttcatgaatt gtataaagtg atccttctcc caaaccagac tcactatatt
1861 attccaaaag gtgaatgcct gccttatttc agctttgcag aagtagccaa aagaggagtt
1921 gaaggtgcct atagtgacaa tccaataatt cgacatgctt ctattgccaa caagtggaaa
1981 accatccacc tcataatgca cagtggaatg aatgccacca caatacattt taatctcacg
2041 tttcaaaata caaacgatga agagttcaaa atgcagataa cagtggaggt ggacacaagg
2101 gagggaccaa aactgaattc tacagcccag aagggttacg aaaatttagt tagtcccata
2161 acacttcttc cagaggcgga aatccttttt gaggatattc ccaaagaaaa acgcttcccg
2221 aagtttaaga gacatgatgt taactcaaca aggagagccc aggaagaggt gaaaattccc
2281 ctggtaaata tttcactcct tccaaaagac gcccagttga gtctcaatac cttggatttg
2341 caactggaac atggagacat cactttgaaa ggatacaatt tgtccaagtc agccttgctg
2401 agatcatttc tgatgaactc acagcatgct aaaataaaaa atcaagctat aataacagat
2461 gaaacaaatg acagtttggt ggctccacag gaaaaacagg ttcataaaag catcttgcca
2521 aacagcttag gagtgtctga aagattgcag aggttgactt ttcctgcagt gagtgtaaaa
2581 gtgaatggtc atgaccaggg tcagaatcca cccctggact tggagaccac agcaagattt
2641 agagtggaaa ctcacaccca aaaaaccata ggcggaaatg tgacaaaaga aaagccccca
2701 tctctgattg ttccactgga aagccagatg acaaaagaaa agaaaatcac agggaaagaa
2761 aaagagaaca gtagaatgga ggaaaatgct gaaaatcaca taggcgttac tgaagtgtta
2821 cttggaagaa agctgcagca ttacacagat agttacttgg gctttttgcc atgggagaaa
2881 aaaaagtatt tccaagatct tctcgacgaa gaagagtcat tgaagacaca attggcatac
2941 ttcactgata gcaaaaatac tgggaggcaa ctaaaagata catttgcaga ttccctcaga
3001 tatgtaaata aaattctaaa tagcaagttt ggattcacat cgcggaaagt ccctgctcac
3061 atgcctcaca tgattgaccg gattgttatg caagaactgc aagatatgtt ccctgaagaa
3121 tttgacaaga cgtcatttca caaagtgcgc cattctgagg atatgcagtt tgccttctct
3181 tatttttatt atctcatgag tgcagtgcag ccactgaata tatctcaagt ctttgatgaa
3241 gttgatacag atcaatctgg tgtcttgtct gacagagaaa tccgaacact ggctaccaga
3301 attcacgaac tgccgttaag tttgcaggat ttgacaggtc tggaacacat gctaataaat
3361 tgctcaaaaa tgcttcctgc tgatatcacg cagctaaata atattccacc aactcaggaa
3421 tcctactatg atcccaacct gccaccggtc actaaaagtc tagtaacaaa ctgtaaacca
3481 gtaactgaca aaatccacaa agcatataag gacaaaaaca aatataggtt tgaaatcatg
3541 ggagaagaag aaatcgcttt taaaatgatt cgtaccaacg tttctcatgt ggttggccag
3601 ttggatgaca taagaaaaaa ccctaggaag tttgtttgcc tgaatgacaa cattgaccac
3661 aatcataaag atgctcagac agtgaaggct gttctcaggg acttctatga atccatgttc
3721 cccatacctt cccaatttga actgccaaga gagtatcgaa accgtttcct tcatatgcat
3781 gagctgcagg aatggagggc ttatcgagac aaattgaagt tttggaccca ttgtgtacta
3841 gcaacattga ttatgtttac tatattctca ttttttgctg agcagttaat tgcacttaag
```

```
3901 cggaagatat ttcccagaag gaggatacac aaagaagcta gtcccaatcg aatcagagta
3961 tagaagatct tcatttgaaa accatctacc tcagcattta ctgagcattt taaaactcag
4021 cttcacagag atgtctttgt gatgtgatgc ttagcagttt ggcccgaaga aggaaaatat
4081 ccagtaccat gctgttttgt ggcatgaata tagcccactg accaggaatt atttaaccaa
4141 cccactgaaa acttgtgtgt tgagcagctc tgaactgatt ttacttttaa agaatttgct
4201 catggacctg tcatcctttt tataaaaagg ctcactgaca agagacagct gttaatttcc
4261 cacagcaatc attgcagact aactttatta ggagaagcct atgccagctg ggagtgattg
4321 ctaagaggct ccagtctttg cattccaaag ccttttgcta aagttttgca ctttttttt
4381 ttcatttccc atttttaagt agttactaag ttaactagtt attcttgctt ctgagtataa
4441 cgaattggga tgtctaaacc tatttttata gatgttattt aaataatgca gcaatatcac
4501 ctcttattga caatacctaa attatgagtt ttattaatat ttaagactgt aaatggtctt
4561 aaaccactaa ctactgaaga gctcaatgat tgacatctga aatgctttgt aattattgac
4621 ttcagcccct aagaatgcta tgatttcacg tgcaggtcta atttcaaagg gctagagtta
4681 gtactactta ccagatgtaa ttatgttttg gaaatgtaca tattcaaaca gaagtgcctc
4741 attttagaaa tgagtagtgc tgatggcact ggcacattac agtggtgtct tgtttaatac
4801 tcattggtat attccagtag ctatctctct cagttggttt ttgatagaac agaggccagc
4861 aaactttctt tgtaaaaggc tggttagtaa attattgcag gccacctgtg tctttgtcat
4921 acattcttct tgctgttgtt tagtttgttt tttttcaaac aaccctctaa aaatgtaaaa
4981 accatgttta gcttgcagct gtacaaaaac tgcccaccag ccagatgtga ccctcaggcc
5041 atcatttgcc aatcactgag aattagtttt tgttgttgtt gttgttgttg tttttgagac
5101 agagtctctc tctgttgccc aggctggagt gcagtggcgc aatctcagct cactgcaacc
5161 tccgcctccc gggttcaagc agttctgtct cagccttctg agtagctggg actacaggtg
5221 catgccacca caccctgcta atttttgtat ttttagtaga gacgggggtt ccaccatatt
5281 ggtcaggctt atcttgaact cctgacctca ggtgatccac ctgcctctgc ctcccaaagt
5341 gctgagatta caggcataag ccagtgcacc cagccgagaa ttagtatttt tatgtatggt
5401 taaaccttgg cgtctagcca tattttatgt cataatacaa tggatttgtg aagagcagat
5461 tccatgagta actctgacag gtattttaga tcatgatctc aacaatattc ttccaaaatg
5521 gcatacatct tttgtacaaa gaacttgaaa tgtaaatact gtgtttgtgc tgtaagagtt
5581 gtgtatttca aaaactgaaa tctcataaaa agttaaattt ttgtctgaca aaaaaaaaaa
5641 aaaa
```

[0059] Human phosphoinositide kinase, FYVE finger containing (PIKFYVE) protein has the amino acid sequence (SEQ ID NO:21) (NCBI Reference Sequence: NM_015040.3):

```
MATDDKTSPT LDSANDLPRS PTSPSHLTHF KPLTPDQDEP PFKSAYSSFV NLFRFNKERA      60
EGGQGEQQPL SGSWTSPQLP SRTQSVRSPT PYKKQLNEEL QRRSSALDTR RKAEPTFGGH     120
DPRTAVQLRS LSTVLKRLKE IMEGKSQDSD LKQYWMPDSQ CKECYDCSEK FTTFRRRHHC     180
RLCGQIFCSR CCNQEIPGKF MGYTGDLRAC TYCRKIALSY AHSTDSNSIG EDLNALSDSA     240
CSVSVLDPSE PRTPVGSRKA SRNIFLEDDL AWQSLIHPDS SNTPLSTRLV SVQEDAGKSP     300
ARNRSASITN LSLDRSGSPM VPSYETSVSP QANRTYVRTE TTEDERKILL DSVQLKDLWK     360
KICHHSSGME FQDHRYWLRT HPNCIVGKEL VNWLIRNGHI ATRAQAIAIG QAMVDGRWLD     420
CVSHHDQLFR DEYALYRPLQ STEFSETPSP DSDSVNSVEG HSEPSWFKDI KFDDSDTEQI     480
AEEGDDNLAN SASPSKRTSV SSFQSTVDSD SAASISLNVE LDNVNFHIKK PSKYPHVPPH     540
PADQKEYLIS DTGGQQLSIS DAFIKESLFN RRVEEKSKEL PFTPLGWHHN NLELLREENG     600
EKQAMERLLS ANHNHMMALL QQLLHSDSLS SSWRDIIVSL VCQVVQTVRP DVKNQDDDMD     660
IRQFVHIKKI PGGKKFDSVV VNGFVCTKNI AHKKMSSCIK NPKILLLKCS IEYLYREETK     720
FTCIDPIVLQ EREFLKNYVQ RIVDVRPTLV LVEKTVSRIA QDMLLEHGIT LVINVKSQVL     780
ERISRMTQGD LVMSMDQLLT KPHLGTCHKF YMQIFQLPNE QTKTLMFFEG CPQHLGCTIK     840
LRGGSDYELA RVKEILIFMI CVAYHSQLEI SFLMDEFAMP PTLMQNPSFH SLIEGRGHEG     900
AVQEQYGGGS IPWDPDIPPE SLPCDDSSLL ELRIVFEKGE QENKNLPQAV ASVKHQEHST     960
TACPAGLPCA FFAPVPESLL PLPVDDQQDA LGSEQPETLQ QTVVLQDPKS QIRAFRDPLQ    1020
DDTGLYVTEE VTSSEDKRKT YSLAFKQELK DVILCISPVI TFREPFLLTE KGMRCSTRDY    1080
FAEQVYWSPL LNKEFKEMEN RRKKQLLRDL SGLQGMNGSI QAKSIQVLPS HELVSTRIAE    1140
HLGDSQSLGR MLADYRARGG RIQPKNSDPF AHSKDASSTS SGQSGSKNEG DEERGLILSD    1200
AVWSTKVDCL NPINHQRLCV LFSSSSAQSS NAPSACVSPW IVTMEFYGKN DLTLGIFLER    1260
YCFRPSYQCP SMFCDTPMVH HIRRFVHGQG CVQIILKELD SPVPGYQHTI LTYSWCRICK    1320
QVTPVVALSN ESWSMSFAKY LELRFYGHQY TRRANAEPCG HSIHHDYHQY FSYNQMVASF    1380
SYSPIRLLEV CVPLPKIFIK RQAPLKVSLL QDLKDFFQKV SQVYVAIDER LASLKTDTFS    1440
KTREEKMEDI FAQKEMEEGE FKNWIEKMQA RLMSSSVDTP QQLQSVFESL IAKKQSLCEV    1500
LQAWNNRLQD LFQQEKGRKR PSVPPSPGRL RQGEESKISA MDASPRNISP GLQNGEKEDR    1560
```

```
FLTTLSSQSS TSSTHLQLPT PPEVMSEQSV GGPPELDTAS SSEDVFDGHL LGSTDSQVKE    1620
KSTMKAIFAN LLPGNSYNPI PFPFDPDKHY LMYEHERVPI AVCEKEPSSI IAFALSCKEY    1680
RNALEELSKA TQWNSAEEGL PTNSTSDSRP KSSSPIRLPE MSGGQTNRTT ETEPQPTKKA    1740
SGMLSFFRGT AGKSPDLSSQ KRETLRGADS AYYQVGQTGK EGTENQGVEP QDEVDGGDTQ    1800
KKQLINPHVE LQFSDANAKF YCRLYYAGEF HKMREVILDS SEEDFIRSLS HSSPWQARGG    1860
KSGAAFYATE DDRFILKQMP RLEVQSFLDF APHYFNYITN AVQQKRPTAL AKILGVYRIG    1920
YKNSQNNTEK KLDLLVMENL FYGRKMAQVF DLKGSLRNRN VKTDTGKESC DVVLLDENLL    1980
KMVRDNPLYI RSHSKAVLRT SIHSDSHFLS SHLIIDYSLL VGRDDTSNEL VVGIIDYIRT    2040
FTWDKKLEMV VKSTGILGGQ GKMPTVVSPE LYRTRFCEAM DKYFLMVPDH WTGLGLNC      2098
```

[0060] Nucleic acid (mRNA) encoding human PIKFYVE protein has the nucleotide sequence (SEQ ID NO:22) (NCBI Reference Sequence: NM_015040.3):

```
   1 caaccatgta agcagcttcg cttcctgccg caaccgtccg cggcctgagg agcccaccgc
  61 cgctctcggg ggccgacttc cgggggctga gccgttgaag cggaggctgg ggcgggggcc
 121 agccggcgcg gccggggcag gaggcgcaga ctcatgaaat ggccacagat gataagacgt
 181 ccccaacact ggactctgct aatgatttgc ctcgatctcc tactagtcct tctcatctca
 241 cacactttaa acctttgact cctgatcaag atgagccccc ttttaaatca gcttatagtt
 301 cttttgtaaa tctctttcgt tttaacaaag agagagcaga aggaggccag ggagaacagc
 361 agcctttgag tggaagttgg accagccctc agctcccttc gaggacacag tctgttaggt
 421 cacccacacc ttataaaaag cagcttaatg aggaactcca gcggcgctct tcagcattag
 481 acacaagaag gaaagcagaa cctacctttg gaggtcatga ccctcgtaca gctgttcagc
 541 ttcgaagcct cagcacagta ttaaaacgcc tcaaggaaat catggagggg aaaagccagg
 601 atagtgacct gaaacaatac tggatgccag atagccaatg taaagagtgc tatgactgta
 661 gtgagaaatt tacaaccttt aggcgcagac accattgccg actatgtggg cagattttct
 721 gcagtcgttg ctgtaatcaa gaaatccctg gaaaatttat gggctataca ggagacctcc
 781 gagcttgcac atattgtaga aaaatagcct taagttatgc tcattccaca gacagtaatt
 841 ctattgggga agacttgaat gctctttcag attctgcttg ctctgtgtct gtgcttgatc
 901 caagtgaacc ccgaacacct gttgggagta ggaaagccag ccgtaacata tttttagagg
 961 atgatttggc ctggcaaagt ttgattcatc cagattcctc aaatactcct ctttcaacaa
1021 gacttgtatc tgtgcaagag gatgctggga aatctcctgc tcgaaataga tcagccagca
1081 ttactaacct gtcactggat agatctggtt ctcctatggt accttcatat gagacatctg
1141 tcagtcccca ggctaaccga acatatgtta ggacagagac cactgaggat gaacgcaaaa
1201 ttcttctgga cagtgtgcag ttaaaagacc tgtggaaaaa aatctgccat cacagcagtg
1261 gaatggagtt tcaggatcac cgctactggt tgagaacgca tcccaactgc attgtaggaa
1321 aggaattagt caactggcta atccgaaatg ggcatattgc cacaagggca caagctatag
1381 caattggaca agcaatggtt gatggacgtt ggctggattg tgttagtcat cacgaccagc
1441 ttttcagaga tgagtatgcg ctgtatagac cactgcagag tacagaattt tctgagacgc
1501 cttctcccga cagtgactca gtgaactccg tggaaggaca ctctgagcca tcctggttta
1561 aagacataaa gtttgatgac agtgacacag aacagatagc tgaagaaggt gacgataatt
1621 tggctaattc tgccagtcct agcaagcgca catcagtcag cagtttccag tccacagtgg
1681 acagtgactc agccgcttct atcagcctga acgtggagct ggacaacgtg aacttccata
1741 tcaagaagcc ctccaagtac ccacatgtgc cccctcaccc tgctgaccaa aaagagtatt
1801 tgatttctga cactggagga caacagctct caataagtga cgctttcatc aaagaatcct
1861 tatttaatcg ccgagtagag gaaaaatcca aagagctgcc tttcacacct ttgggctggc
1921 atcataacaa cctggagctc ctgagggagg agaatgggga gaaacaagcc atggagaggt
1981 tgctttcagc taatcataac cacatgatgg cactactcca gcagttgctc catagtgact
2041 cactgtcatc atcttggagg gacatcatcg tgtcattggt ctgccaggtt gttcagacag
2101 tccgacctga tgtcaagaac caggatgatg acatggatat ccgtcagttt gtccacatca
2161 aaaaaatccc aggtggaaag aagtttgatt ctgtggttgt caatggcttt gtttgtacca
2221 agaacattgc acataaaaag atgagttctt gtattaaaaa ccccaaaatt cttctgttga
2281 agtgttccat tgagtatctc tacagagaag aaactaagtt tacttgcatt gatcctattg
2341 tgcttcagga aagggaattc ttgaagaatt atgtccagcg aatagttgat gttcgaccca
2401 ccttggttct tgttgagaaa acagtgtctc ggattgccca ggacatgtta ttggaacatg
2461 gcattacttt ggtcattaat gtaaagtcac aagttttgga acgaatcagt cgaatgaccc
2521 aaggtgattt agtgatgtca atggaccagc tgcttacgaa accacacctg ggcacttgtc
2581 acaaatttta tatgcagata tttcagttgc ctaatgaaca aaccaagaca ctgatgtttt
2641 ttgaaggttg tccacagcac ctaggctgta caatcaagct aagaggaggc tctgattatg
2701 agctggctcg agttaaggag atcctaatat ttatgatctg tgttgcttat cattctcaac
2761 tagaaatatc ctttctcatg gatgaatttg ctatgcctcc cacattaatg caaaaccctt
```

2821 cattccattc cctgattgag ggacgagggc atgaggggc tgtccaagag cagtacggtg
2881 gaggttccat cccctgggat cctgacatcc ctcctgagtc tctgccctgt gatgatagca
2941 gtttgctgga attgaggatt gtgtttgaga agggtgagca ggaaaataaa aatcttccgc
3001 aggctgttgc ctctgtgaag catcaagaac atagcacaac agcttgcccg gcgggtctcc
3061 cttgtgcttt ctttgcacct gtaccggaat cattgttgcc actccctgtg gatgaccaac
3121 aagatgcttt aggcagcgag cagccagaga ctttgcagca aacagttgtg ctgcaggatc
3181 ccaaaagcca gataagagcc tttagagacc ctctacagga tgacactgga ttatatgtta
3241 ctgaggaagt cacctcctct gaagataaac gaaagactta ttctttggcc tttaagcagg
3301 aattaaaaga tgtgatcctc tgtatctccc cagtaatcac attccgagaa ccctttcttt
3361 taactgaaaa ggggatgaga tgctctaccc gagattattt tgcagagcag gtttactggt
3421 ctcctctcct caataaagaa ttcaaagaaa tggagaacag gaggaagaaa cagctgctca
3481 gggatctctc tggacttcag ggcatgaatg gaagtattca ggccaagtct attcaagtct
3541 taccctcaca tgagctagtg agcactagaa ttgctgagca tctgggcgat agccagagct
3601 tgggtagaat gctggccgat tatcgagcca gaggaggaag aattcagccc aaaaattcag
3661 acccttttgc tcattcaaag gatgcatcaa gtacttcaag tggccaatca ggaagcaaaa
3721 atgagggtga tgaagagaga gggcttattc tgagtgatgc tgtgtggtca acaaaggtgg
3781 actgtctgaa tcccattaat caccagagac tttgtgtgct cttcagcagc tcttctgccc
3841 agtccagcaa tgctcctagt gcctgtgtca gtccttggat tgtaacaatg gaattttatg
3901 gaaagaatga tcttacatta ggaatatttt tagagagata ctgtttcagg ccttcttatc
3961 agtgtccaag catgttctgt gataccccca tggtacatca tattcggcgc tttgttcatg
4021 gccaaggctg tgtgcagata atcctgaagg agttggattc tccagtacct ggatatcagc
4081 atacaattct tacatattcc tggtgtagaa tctgcaaaca ggtaacacca gttgttgctc
4141 tttccaatga gtcctggtct atgtcatttg caaaatacct tgaacttagg ttttatgggc
4201 accagtatac tcgcagagcc aacgctgagc cctgtggtca ctccatccat catgattatc
4261 accagtattt ctcctataac cagatggtgg cgtctttcag ttattctccc attcggcttc
4321 ttgaagtatg tgttccactc cccaaaatat tcattaagcg tcaggcccca ttaaaagtgt
4381 cccttcttca ggatctgaag gacttctttc aaaaagtttc acaggtatat gttgccattg
4441 atgaaagact tgcatctttg aaaactgata catttagtaa aacaagagag gaaaaaatgg
4501 aagatatttt tgcacagaaa gagatggaag aaggtgagtt caagaactgg attgagaaga
4561 tgcaagcaag gctcatgtct tcctctgtag ataccccctca gcaactgcag tcggtctttg
4621 agtcactcat tgccaagaaa caaagtctct gtgaagtgct gcaagcttgg aataacaggt
4681 tgcaggacct tttccaacag gaaaagggta gaaagagacc ttcagttcct ccaagtcctg
4741 gaagactgag acaaggggaa gaaagcaaga taagtgcgat ggatgcatct ccacggaata
4801 tttctccagg acttcagaat ggagaaaaag aggatcgctt cttaacaact ttgtccagcc
4861 agagctccac cagttctact catctccaat tgcctacgcc acctgaagtc atgtctgaac
4921 agtcagtggg agggccccct gagctagata cagccagcag ttccgaagat gtgtttgatg
4981 ggcatttgct gggatccaca gacagccaag tgaaggaaaa gtcaaccatg aaagccatct
5041 ttgcaaattt gcttccagga aatagctata atcctattcc atttccttt gatccagata
5101 aacactactt aatgtatgaa catgaacgag tgcccattgc agtctgcgag aaggaaccca
5161 gctccatcat tgcttttgct ctcagttgta aagaataccg aaatgcctta gaggaattgt
5221 ctaaagcgac tcagtggaac agtgccgaag aagggcttcc aacaaatagt acttcagata
5281 gcagaccaaa gagtagcagc cctatcagat tacctgaaat gagtggagga cagacaaatc
5341 gtacaacaga aacagaacca caaccaacca aaaaggcttc tggaatgttg tccttcttca
5401 gagggacagc agggaaaagc cccgatctct cttcccagaa gagagagacc ttacgtggag
5461 cagatagtgc ttactaccag gttgggcaga cgggcaagga ggggaccgag aatcaaggcg
5521 ttgagcctca agatgaagta gatggaggag atacacaaaa gaagcaactc ataaatcctc
5581 atgtggaact tcaattttca gatgctaatg ccaagtttta ctgtcggctc tactatgcgg
5641 gagagtttca taagatgcgt gaagtgattc tggacagcag tgaagaagat ttcattcgtt
5701 ccctctccca ctcatcaccc tggcaggccc ggggaggcaa atcaggagct gccttctatg
5761 caactgagga tgatagattt attttgaagc aaatgcctcg tctggaagtc cagtccttcc
5821 tcgactttgc accacattac ttcaattata ttacaaatgc tgttcaacaa aagaggccca
5881 cggcgttggc caaaattctt ggagtttaca gaattggtta taagaactct cagaacaaca
5941 ctgagaagaa gttagatctc cttgtcatgg aaaatctttt ctacgggaga aagatggcac
6001 aggtttttga tttgaagggc tctcttagga atcggaatgt aaaaactgac actggaaaag
6061 agagttgtga tgtggtcctg ctagatgaaa atctcctaaa gatggttcga gacaaccctc
6121 tatatattcg ttctcattcc aaagctgtgc tgagaacctc gatccatagt gactcccatt
6181 tcctttctag ccacctcatt atagattatt ctttgctggt tgggcgagat gatactagca
6241 atgagctagt agttggaatt atagattata ttcgaacatt tacatgggac aaaaagcttg
6301 agatggttgt gaaatcaaca ggaattttag gtggacaagg taaaatgcca acagtggtgt
6361 ctccggagtt gtacaggact aggttttgtg aggcaatgga caagtatttc ctaatggtac

6421 cagaccactg gacaggcttg ggtctgaatt gctgaaatca agcacatatt ttgaaatgga
6481 ctgtgaagga aaaggggaca ggaacaaagg accaaaaata agctacatgt tttatttctt
6541 catcgtgttc accactgtat gccaaggctt ttcagttctg tggctgttta gactgtccgt
6601 aatggaatgg taaaactcca tgaatttgca ctttggtttt tgatacctgt ggagctgtct
6661 gtaggttggg aagtggcatg aaaatttttct taagctaaaa tacagacatg tttcaaaggg
6721 ctaaagttgg agatgagtag atagggtgaa aaatgggtta aatttgctag cttaattgtt
6781 ttaagaagaa aacagtgtct cataaattga ctatcctggc atcacattta acatgttatc
6841 tacttagaaa gcatttgtag agctgctgaa tttgttttgt gtttttctgt aataatttaa
6901 tgttacttat tatcagaatt tctgaaacct ttacaaaaat tctgatttat tccattaatg
6961 gccagttaaa cacgtgggca tttattgttt tattgaggaa tttgacttaa actgggaatc
7021 ctgtcatgtt gtttatcttt ccagcttgcc tgtttttgag tatgtttgat gtttttaaaa
7081 ttttgtcttc tctgtggatg acaggaggct acagcaatta actttaagcc tccttttaga
7141 gatattttta aagcttgttt aaaatttttg tgcaattcat atattaaatt gcacttactt
7201 gcatacgctc atattctagg gtttttttctc tatttttagg gtatcatagt aaatcattag
7261 taaatgagtc tgtagttact aaaccctaat ggaataatta ttaatgaaag atttttgaaa
7321 tataaaaaat aaattaggcc caatccaaga aattgagtga gaaggaaaca cttgttttat
7381 tcacagaggt aaagtgtctt ttcaatataa ccagcaattt aggtggcatc tataaaataa
7441 aaaatttcta ctgtggacat ccccttttcc aactttctac ataatggcta gttctgacta
7501 ctaagaaatg ttaagaaata ggccaagtgc ggtggctcac gcctataatc ctatcacttt
7561 gggaggccta ggcaggcaga tcacctgagg tcaggagttc aagaccagcc tggccaacat
7621 gaggaaaccc catctctact aaaaagacaa aaaattagct gggcatggtg gcatatacct
7681 gtaaccccaa ctacttgggt ggctgaggca ggagaattgt ttgaacctgg gaggcggagg
7741 ctgcagtgag ctgggattac accattgcat tccagcctgg gcaacagagc aagactgtct
7801 caaaaaacaa aacgaaacaa aaaaagaaag ttattcttag taaggaactt cttgtttaat
7861 agcatttttg tttattttaa aaagtgatca gaagtagtaa actatctttg aggaaatact
7921 gtaaccccag aatatttcct cttgacttct ttttgtaaca aggataattt agggatttat
7981 aaagttgtaa ggatttcact gttttcggac tgcctataat aatagcacat taaccttcac
8041 ataataagaa atctggacaa gttcagttac acagtatgat gaatacttga attaggaaca
8101 ttgtggaaaa tttgctttag agaatcaagg cagtagtttg gtatttggtg cttattaaaa
8161 atgtggtttg ttttgaactg gaagcaagtt gaccaaggac ttatgactaa tgtgatgcta
8221 agttccactt ggcccctttt aaaaacgtgt atgtgccttt tgaagataca caaaacactg
8281 aggattttag ttttgaaatc aaagactatt aaaggagctg tacagaggta aaaaaataaa
8341 tgtggaacat tattaactta ttttgtgtct aggaacaatg gattttgtat ctgatttaaa
8401 atgccaacac tgttttgtct ctgttcattt tttctgtgag gatacttaag gttattattc
8461 ctgtctgttt cctgtactcc cctagtcatg agcacttgaa gtacaaggtg tctcccccta
8521 ggtgcaatta ggttgtttct ttgtttttag tttcaattct atgtgcatag caggaatgct
8581 ccacaggaat ggcttctgac aataatctgt cctgttgatt ttgttttcct tgcccatgac
8641 ttgaacaact gtgttttaaa gtactgtagt ctagtaggta actttgtggc aaaaattttc
8701 aatataatac attctgaaac aatagttgct gccttgcaaa ggtaatctct cattttaaaa
8761 ttggacagta ttaatgaagg ggaaatatac aatttatttc tattgagtgg tagaactata
8821 tgtctggtcc cttgctgctc ttgtttaggc cactatcata gatatatttc aaatattgta
8881 ctactcagtg ttaagtattg aatgactgtt tccctttcct tcaaggccta gagtatattc
8941 tgaaaattta ggaatgagga agaaatctta atacttcctt ccttaacata caacatgagt
9001 cccgagaata attgatagta gcaagagaaa actatgtcag taacatgttg ctttgtataa
9061 aaatcttatt tataaatgtg aagctttttg atgccatcaa aacttattaa aaaataggat
9121 ttacttttttt ctaattctga cctaagaaaa ataatgagaa caagctgttg caagctcttt
9181 tgtagtctat tgaatatttt atagatattc aaaatttcct acaaactata atttttttcca
9241 tgatttagca gtgagtgatt ttctagcttt ggctcttatt aggtattgta aatagtaggg
9301 ttatatcgat atcagctttt gtgatggcat tgtggtcatc agcttcatga cattttaccc
9361 atttgcagtg atcctgtgta aaactgccaa ggaaagtaat tacctgtagg agtttgctga
9421 gcttgaagag tgaaaactgt tgtgaatgag cctgatcata aaacggacca ggccattcat
9481 tattcctcaa gtgttaatat actgacttat gcagtattca aaccatctag tgcaatgttt
9541 ttgtttttgt ttttttttg gtaacacagg tgcagtgtat tatagaaaaa ataaaaacta
9601 caatcattag cagttttaat actgctgtgt cagttttgta aaaaatgtac attatgtctt
9661 ttgacatgtt gaattttaaa ctagggaaat gacattgtaa atcatagtag cctcttttaa
9721 tttaatatga aaaatgccac tatattgaaa gtacttaatg tattgtatat atttctctac
9781 tttggttcta gctattttat atgattgaca tgttatttaa aagataactg ccttgaactt
9841 ttggagactt gtactgtaaa taaagaaatc ttaacaataa actcagaatc tacttactcc
9901 a

[0061] Human FIG4 protein has the amino acid sequence (SEQ ID NO:23) (NCBI Reference Sequence: NM_014845.5):

```
MPTAAAPIIS SVQKLVLYET RARYFLVGSN NAETKYRVLK IDRTEPKDLV IIDDRHVYTQ      60
QEVRELLGRL DLGNRTKMGQ KGSSGLFRAV SAFGVVGFVR FLEGYYIVLI TKRRKMADIG     120
GHAIYKVEDT NMIYIPNDSV RVTHPDEARY LRIFQNVDLS SNFYFSYSYD LSHSLQYNLT     180
VLRMPLEMLK SEMTQNRQES FDIFEDEGLI TQGGSGVFGI CSEPYMKYVW NGELLDIIKS     240
TVHRDWLLYI IHGFCGQSKL LIYGRPVYVT LIARRSSKFA GTRFLKRGAN CEGDVANEVE     300
TEQILCDASV MSFTAGSYSS YVQVRGSVPL YWSQDISTMM PKPPITLDQA DPFAHVAALH     360
FDQMFQRFGS PIIILNLVKE REKRKHERIL SEELVAAVTY LNQFLPPEHT IVYIPWDMAK     420
YTKSKLCNVL DRLNVIAESV VKKTGFFVNR PDSYCSILRP DEKWNELGGC VIPTGRLQTG     480
ILRTNCVDCL DRTNTAQFMV GKCALAYQLY SLGLIDKPNL QFDTDAVRLF EELYEDHGDT     540
LSLQYGGSQL VHRVKTYRKI APWTQHSKDI MQTLSRYYSN AFSDADRQDS INLFLGVFHP     600
TEGKPHLWEL PTDFYLHHKN TMRLLPTRRS YTYWWTPEVI KHLPLPYDEV ICAVNLKKLI     660
VKKFHKYEEE IDIHNEFFRP YELSSFDDTF CLAMTSSARD FMPKTVGIDP SPFTVRKPDE     720
TGKSVLGNKS NREEAVLQRK TAASAPPPPS EEAVSSSSED DSGTDREEEG SVSQRSTPVK     780
MTDAGDSAKV TENVVQPMKE LYGINLSDGL SEEDFSIYSR FVQLGQSQHK QDKNSQQPCS     840
RCSDGVIKLT PISAFSQDNI YEVQPPRVDR KSTEIFQAHI QASQGIMQPL GKEDSSMYRE     900
YIRNRYL                                                               907
```

[0062] Nucleic acid (mRNA) encoding human FIG4 protein has the nucleotide sequence (SEQ ID NO:24) (NCBI Reference Sequence: NM_014845.5):

```
   1 acgtcctcca gccccgctcc cgacgtgagg ggcggggctt gcctggaggc ggggcgcagg
  61 gatccggaaa cacctgatca tctataggtt tagtgcctaa tgggtgttgt tcctggctgg
 121 acttgatgtc cagggcctga ggggttttct cgccgagtct cctggggcgg tccggaggct
 181 cgtgccctgt tgtggggccc ccatttgccg ccgccatgcc cacggccgcc gcccccatca
 241 tcagctcggt ccagaagctg gttctgtatg agactagagc tagatacttt ctagttggga
 301 gcaataatgc agaaacgaaa tatcgtgtct tgaagattga tagaacagaa ccaaaagatt
 361 tggtcataat tgatgacagg catgtctata ctcaacaaga agtaagggaa cttcttggcc
 421 gcttggatct tggaaataga acaaagatgg gacagaaagg atcctcgggc ttatttcgag
 481 cggtttcagc ttttggtgtt gtgggttttg tcaggttctt agaaggctat tatattgtgt
 541 taataactaa aaggaggaag atggcggata ttggaggtca tgcaatctat aaggtcgaag
 601 atacaaatat gatctatata cccaatgatt ctgtacgggt tactcatcct gatgaagcta
 661 ggtatctacg aatatttcaa aatgtggacc tatctagcaa tttttacttt agttacagct
 721 atgatttgtc ccactcactt caatataatc tcactgtctt gcgaatgccc ctggagatgt
 781 taaagtcaga aatgacccag aatcgccaag agagctttga catctttgaa gatgaaggat
 841 taattacaca aggtggaagc ggggtatttg ggatctgtag tgagccttat atgaaatatg
 901 tatggaatgg tgaacttctg gatataatta aaagtactgt gcatcgtgac tggcttttgt
 961 atattattca tgggttctgt gggcagtcaa agctgttgat ctatggacga ccagtgtatg
1021 tcactctaat agctagaaga tccagtaaat ttgctggcac ccgttttctt aaaagaggtg
1081 caaactgtga gggtgatgtt gcaaatgaag tggagactga acaaatactc tgcgatgctt
1141 ctgtgatgtc tttcactgca ggaagttatt cttcatatgt acaagttaga ggatctgtgc
1201 ccttatactg gtctcaggac atttcaacta tgatgcctaa accacctatt acattggatc
1261 aggcagatcc atttgcacat gtggctgccc ttcactttga ccagatgttc cagaggtttg
1321 gctctcccat catcatcttg aatttagtga aggaacgaga gaaaagaaag catgaaagaa
1381 ttctgagtga agaacttgtt gctgctgtga cctatctcaa ccaatttttg cctcctgagc
1441 acactattgt ttatattccc tgggacatgg ccaagtatac caaaagcaag ctgtgtaatg
1501 ttcttgatcg actaaatgtg attgcagaaa gtgtggtgaa gaaaacaggt ttctttgtaa
1561 accgccctga ttcttactgt agcattttgc ggccagatga aaagtggaat gaactaggag
1621 gatgtgtgat tcccactggt cgcctgcaga ctggcatcct tcgaaccaac tgtgtggact
1681 gtttagatcg caccaacaca gcacagttta tggtgggaaa atgtgctctg gcctatcagc
1741 tgtattcact gggactgatt gacaaaccta atctacagtt tgatacagat gcagttaggt
1801 tatttgagga actctatgaa gatcatggtg ataccctatc ccttcagtat ggtggttctc
1861 aacttgttca tcgtgtgaaa acctacagaa agatagcacc atggacccag cactccaaag
1921 acatcatgca aaccctgtct agatattaca gcaatgcttt ttcagatgcc gatagacaag
1981 attccattaa tctcttcctg ggagttttcc atcccactga agggaaacct catctctggg
2041 agctcccaac agatttttat ttgcatcaca aaaataccat gagacttttg ccaacaagaa
2101 gaagttatac ttactggtgg acaccagagg tgataaagca tttaccattg ccctatgatg
```

```
2161 aagttatctg tgctgtgaac ttaaagaagt tgatagtgaa gaaattccac aaatatgaag
2221 aagagattga tatccacaat gagttctttc ggccatatga gttgagcagc tttgatgata
2281 ccttttgctt ggctatgaca agctcagcac gtgactttat gcctaagacc gttggaattg
2341 atccaagtcc atttactgtg cgtaaaccag atgaaactgg aaaatcagta ttgggaaaca
2401 aaagcaatag agaagaagct gtattacagc ggaaaacggc agccagcgcc ccgccgcccc
2461 ccagcgagga ggctgtgtcc agcagctctg aggatgactc tgggactgat cgggaagaag
2521 agggctctgt gtctcagcgc tccactcccg tgaagatgac tgatgcagga gacagtgcca
2581 aagtgaccga gaatgtggtc caacccatga aggagctata tggaattaac ctctcagatg
2641 gcctctcaga agaagatttc tccatttatt caagatttgt tcagctgggg cagagtcaac
2701 ataaacaaga caagaatagc cagcagccct gttctaggtg ctcagatgga gttataaaac
2761 taacacccat ctcggctttc tcgcaagata acatctatga agttcagccc ccaagagtag
2821 acagaaaatc tacagagatc ttccaagccc acatccaggc cagccaaggt atcatgcagc
2881 ccctaggaaa agaggactcc tccatgtacc gagagtacat caggaaccgc tacctgtgaa
2941 aagagcgcag gtccacctgg tggacacgtc tgattagctt agaacctgtc ttgtctcatc
3001 ttcaaaaggt aacttattaa aagtcctttg cgtctgaagc ctttctcctt ttctgtcact
3061 tgcaaattcc aaattatagc taataaagat gactagataa tttgcaaaaa aaaaaaaaaa
3121 aaa
```

[0063] Human Rho GTPase activating protein 23 (ARHGAP23) protein has the amino acid sequence (SEQ ID NO:25) (NCBI Reference Sequence: NM_001199417.1):

```
MNGVAFCLVG IPPRPEPRPP QLPLGPRDGC SPRRPFPWQG PRTLLLYKSP QDGFGFTLRH      60
FIVYPPESAV HCSLKEEENG GRGGGPSPRY RLEPMDTIFV KNVKEDGPAH RAGLRTGDRL     120
VKVNGESVIG KTYSQVIALI QNSDDTLELS IMPKDEDILQ LAYSQDAYLK GNEPYSGEAR     180
SIPEPPPICY PRKTYAPPAR ASTRATMVPE PTSALPSDPR SPAAWSDPGL RVPPAARAHL     240
DNSSLGMSQP RPSPGAFPHL SSEPRTPRAF PEPGSRVPPS RLECQQALSH WLSNQVPRRA     300
GERRCPAMAP RARSASQDRL EEVAAPRPWP CSTSQDALSQ LGQEGWHRAR SDDYLSRATR     360
SAEALGPGAL VSPRFERCGW ASQRSSARTP ACPTRDLPGP QAPPPSGLQG LDDLGYIGYR     420
SYSPSFQRRT GLLHALSFRD SPFGGLPTFN LAQSPASFPP EASEPPRVVR PEPSTRALEP     480
PAEDRGDEVV LRQKPPTGRK VQLTPARQMN LGFGDESPEP EASGRGERLG RKVAPLATTE     540
DSLASIPFID EPTSPSIDLQ AKHVPASAVV SSAMNSAPVL GTSPSSPTFT FTLGRHYSQD     600
CSSIKAGRRS SYLLAITTER SKSCDDGLNT FRDEGRVLRR LPNRIPSLRM LRSFFTDGSL     660
DSWGTSEDAD APSKRHSTSD LSDATFSDIR REGWLYYKQI LTKKGKKAGS GLRQWKRVYA     720
ALRARSLSLS KERREPGPAA AGAAAAGAGE DEAAPVCIGS CLVDISYSET KRRHVFRLTT     780
ADFCEYLFQA EDRDDMLGWI RAIRENSRAE GEDPGCANQA LISKKLNDYR KVSHSSGPKA     840
DSSPKGSRGL GGLKSEFLKQ SAARGLRTQD LPAGSKDDSA AAPKTPWGIN IIKKNKKAAP     900
RAFGVRLEEC QPATENQRVP LIVAACCRIV EARGLESTGI YRVPGNNAVV SSLQEQLNRG     960
PGDINLQDER WQDLNVISSL LKSFFRKLPE PLFTDDKYND FIEANRIEDA RERMRTLRKL    1020
IRDLPGHYYE TLKFLVGHLK TIADHSEKNK MEPRNLALVF GPTLVRTSED NMTDMVTHMP    1080
DRYKIVETLI QHSDWFFSDE EDKGERTPVG DKEPQAVPNI EYLLPNIGRT VPPGDPGSDS    1140
TTCSSAKSKG SWAPKKEPYA REMLAISFIS AVNRKRKKRR EARGLGSSTD DDSEQEAHKP    1200
GAGATAPGTQ ERPQGPLPGA VAPEAPGRLS PPAAPEERPA ADTRSIVSGY STLSTMDRSV    1260
CSGASGRRAG AGDEADDERS ELSHVETDTE GAAGAGPGGR LTRRPSFSSH HLMPCDTLAR    1320
RRLARGRPDG EGAGRGGPRA PEPPGSASSS SQESLRPPAA ALASRPSRME ALRLRLRGTA    1380
DDMLAVRLRR PLSPETRRRR SSWRRHTVVV QSPLTDLNFN EWKELGGGGP PEPAGARAHS    1440
DNKDSGLSSL ESTKARAPSS AASQPPAPGD TGSLQSQPPR RSAASRLHQC L              1491
```

[0064] Nucleic acid (mRNA) encoding human ARHGAP23 protein has the nucleotide sequence (SEQ ID NO:26) (NCBI Reference Sequence: NM_001199417.1):

```
   1 ctgccacccg atgaatggag tcgccttctg cctggtcggg atcccgcccc gcccggagcc
  61 ccggccccca cagctgccac tgggcccaag agatgggtgc tctcctaggc gccccttccc
 121 ctggcagggg ccgaggacgc tgctgctgta caaaagtccc caggacggct ttggcttcac
 181 tctgcgccac ttcatcgtgt acccacccga gtcggccgtg cactgcagcc tgaaggagga
 241 agagaatgga ggccgtggag gaggaccctc cccccggtac cgcctggagc ccatggacac
 301 catctttgtc aagaatgtga aggaagacgg ccctgcccat agggcggggc ttcgcacagg
 361 agaccggctg gtaaaggtga atggggaaag cgtcattggg aagacctact ctcaggtcat
 421 agctctgatc cagaatagtg atgacactct ggagctgtct atcatgccca aggacgagga
 481 catcctccag ctggcctact cccaggatgc ctacctgaaa gggaacgagc cgtattctgg
```

```
 541 agaggcccgc agcatcccag agccaccgcc gatctgctac ccccgcaaga cctacgcccc
 601 tcctgcccgg gcctccacca gggccactat ggtgcctgag cccacctcag cactgcccag
 661 tgaccccccgg agtcctgctg cctggagtga cccggggctc cgtgtgccac ctgctgcccg
 721 tgcccacctg gacaactctt ccttggggat gagccagccc cgccccagcc ctggtgcctt
 781 cccccacctc tcctcggagc cccggacgcc ccgtgccttc ccagagcctg gcagccgggt
 841 gccccccagc agactggagt gccagcaggc cttgtcacac tggctgtcaa accaggtacc
 901 ccgccgggcg ggggagagac ggtgcccagc catggccccc cgggcccgca gcgcctccca
 961 ggaccggttg gaggaggtgg ctgccccccg cccgtggccc tgctccacct cccaggatgc
1021 tttgagccag ctgggccagg agggctggca ccgagctcgc tcagatgact acttgagccg
1081 ggccacccgt tctgccgagg cactggggcc aggggcactg gtgtcacccc gctttgagcg
1141 gtgtggctgg gcttcccagc gttcgtctgc ccgcaccccc gcctgcccaa ctcgggacct
1201 gccagggccc caggccccac ccccgtctgg cctgcagggc ctggatgacc tcgggtacat
1261 cggctaccgg agctacagcc catcattcca gcgccggacc ggcctcctcc atgcgctctc
1321 cttccgggac tcacccttg gggggctgcc taccttcaac ctggcccagt cccctgcgtc
1381 attcccacca gaggcctccg agccacccag ggttgtacgg ccggaaccca gcacccgggc
1441 cctggagcct cctgcggagg atcgcggcga tgaggtggtc ctgaggcaga agcccccgac
1501 gggccgcaag gttcagctga cccccgcaag acagatgaac cttggatttg gtgacgagtc
1561 cccagagcca gaggccagtg ggcgagggga acgcctgggc aggaaggtgg ccccttggc
1621 caccaccgaa gactctctgg cttccatccc ctttattgat gagcccacca gccccagcat
1681 tgacctccaa gccaagcacg tccctgcctc tgctgtggtc tccagtgcca tgaactcagc
1741 ccctgtcctg ggcaccagcc catcttcccc gaccttcact ttcaccctcg gacgccatta
1801 ctcgcaggac tgcagcagca tcaaggctgg ccgccgctcc tcctacctgc tggccatcac
1861 cacggagcgc tccaagtcct gcgatgatgg actcaacacc ttccgcgacg agggccgggt
1921 tctgcggcgc ctgccaaacc gcatacccag cctgcggatg ctccggagct tcttcaccga
1981 cgggtccttg gatagctggg gcacctctga agatgctgac gctccttcta agcgacactc
2041 aacctctgac ctctcagatg cgaccttcag cgatatcagg agagaaggct ggttgtatta
2101 taagcagatt ctcaccaaga aggggaagaa agcgggcagc ggcctgcgcc agtggaagcg
2161 ggtgtacgcc gcgctgcggg cgcgctcgct ctcgctgagc aaggagcggc gggagcccgg
2221 gccggcggcg gcgggggctg cggcggccgg cgcaggtgag gacgaggcgg cgcccgtctg
2281 catcggctcc tgcctcgtgg acatctccta cagcgagacc aagaggaggc acgtgttccg
2341 gctgaccacc gctgacttct gtgaatatct ctttcaggct gaggaccggg atgacatgct
2401 gggctggatc agagcgatcc gggagaacag cagggccgag ggcgaggacc ccggctgtgc
2461 caaccaagct ctgatcagca agaagcttaa cgattatcgc aaagtgagcc atagctctgg
2521 gcccaaagct gattcctccc ccaaaggctc tcgcggcctg gggggcctca agtctgagtt
2581 cctcaagcag agtgcggcac gtggcctcag gactcaggac ctgcccgcag ggagcaagga
2641 tgacagtgct gcagccccca aaacccccctg gggcatcaac atcatcaaga aaaataagaa
2701 ggccgctccg agggcgtttg gggtcaggct ggaggagtgc cagccagcca cggagaacca
2761 gcgcgtcccc ttaatcgtgg ctgcatgctg tcgcattgtg gaggcacgag ggctggagtc
2821 cacaggcatt taccgagtgc ccggcaacaa tgcagtggtg tccagcctac aggagcagct
2881 caaccgcggg cctggtgaca tcaacctgca ggatgagcgc tggcaagacc tcaatgtgat
2941 cagcagcctg ctcaagtcct tcttccgaaa gctgcccgag cctcttttca ctgatgacaa
3001 atacaacgac ttcatcgagg ccaaccgcat tgaggacgcg cgggagcgaa tgaggacgct
3061 gcggaagctg atccgggatc tcccaggaca ctactatgaa acgctcaaat tccttgtggg
3121 ccatctcaag accatcgctg accactctga gaaaaacaag atggaacccc ggaacctggc
3181 cctggtcttt gggccgacac tggtgaggac gtctgaggac aacatgacag acatggtgac
3241 ccacatgcct gaccgctaca agatcgtgga gacactgatc cagcactcag actggttctt
3301 cagtgacgaa gaggacaagg gagagagaac ccctgtgggc gacaaggagc ctcaggcagt
3361 gcccaacatt gagtacctcc tgcccaacat tggcaggaca gtgccccctg gcgacccggg
3421 gtcagattct accacctgta gttcagccaa gtccaagggt tcgtgggccc ccaagaagga
3481 gccgtacgcc cgggagatgc tggcgatctc cttcatctcg gccgtcaacc gcaagcgcaa
3541 gaagcggcgg gaggcgcggg ggctgggcag cagcaccgac gacgactcgg agcaggaggc
3601 gcacaagcct ggggcggggg ccacagcgcc ggggactcag gagcggccgc aggggccgct
3661 gcctggcgcc gtcgcccccg aggcccccgg acgcctcagt cccccggcgg cgccggagga
3721 gcggccggcc gcggacacgc gctccattgt gtcgggctac tccaccctgt ccaccatgga
3781 ccgcagcgtg tgctcgggcg ctagcggtcg gcgggcaggg gcgggggatg aggcggacga
3841 cgagcgtagc gagctgagcc acgtggagac ggacactgag ggcgcggcgg gcgcggggcc
3901 tgggggggcgc ctgacacgcc ggccgtcctt cagctcgcac cacctcatgc cctgcgacac
3961 tctggcgcgc cgccgcctgg cccggggccg cccagacggc gagggcgcgg gccggggcgg
4021 tccccgcgcc ccggagccgc ccggctcggc gtcgtccagc agccaggagt cgctgcggcc
4081 cccggcggcg gcgctggcct cccggccctc gcgcatggag gcgctgcgtc taaggctccg
```

```
4141 cggcacggcg gacgacatgc tcgccgtgcg cctgcggcgg ccgctgtcgc ccgagacccg
4201 gcggcgccgg agcagctggc gccgccacac cgtggtggtg cagagcccgc tgactgacct
4261 caacttcaac gagtggaagg agctgggcgg agggggcccc ccggagcctg cgggcgcgcg
4321 ggcgcacagt gacaacaagg actccggact cagcagcctg gagtccacca aggcgcgggc
4381 cccgtcgtcc gctgcctcgc agccgcccgc gcccggggac acggggtccc tgcagagcca
4441 gccccccgcgc cgctcggccg cctcccgcct gcatcagtgt ctgtgatccc cacctcccgc
4501 gccgctcggg cgccaccccct ccctagagcc cctttggaac caggaggctt caccagcctg
4561 cacctcctct tctgtggccc ctgggtgcat ggtgtgggtg gagggcgcag caggcagtgt
4621 ctctagttgg tgtgctggaa ctggcagggc agaggagaag gctggggccg gactaattga
4681 atggaagggg gttccagagg tgatgagcag aagaggaggg ggcgtgggct gctggggtct
4741 gtgtccctgc acacatgcgc ccgataggtc cttctgagcc tttctgtggc tgcacttggg
4801 gacccttgtg gaccatgggg tgtggctagg gaacccctaa gtttcagact aaaggaaaga
4861 tcctgggtga tgctggcttt ttgcttcttt cttctgccct cccacctcag cttgtaagcg
4921 gggatgtgtg tatgtctggg gagaggaggt gtagggtgcg tatgtccatg gggggagggg
4981 cttgtgtgtg cagtcattgt cccaaggtgt ttccagtagc gacttctgtc cccctatccc
5041 caccctggtc cccactttgc gcccccgggc tccctgcctt tggtgcacac aggatcctgc
5101 ccgcccccct tgccagagcc agagaagggg gttggggcca ttccaaggag gcaggactga
5161 aaccctcacc agggttactc cccaacatcc ttttgcctga gtcaccctct aagcgcttta
5221 accacgggca gctgcctgtt ccccagacag tttttggtgg gggggtcca gggtcccct
5281 tgctggtacc tccctcaccc ctctttttgt ttttccatct gtgcctgttc cttccacagc
5341 ccaggcacac agaagcccac cttcttcccc ttaggaggag ggatagtcaa cacccctgct
5401 gtctctctgt cactcacaca ctgatttatg gggtctgagc tgggctgttc ctgcaggatg
5461 gacaggaccc agcgccctct tctccccaca ggctgtaaat agacttccaa tcaccaggcc
5521 agcccccaca caccctcact cattccaggg aagcccaggt aggtggtgaa cccgctgcca
5581 cgtctatcag tcctcttgtt ttatgcaaag atttactgta aagtagattt ctttccctcc
5641 ctccccccatt cttttattgt aaatattgtc tctaaatgtg taacatatta taaagaattt
5701 ataaggattt ttaaagatgt tttgctcatt tacaaaagtg ttgtaacagt gttggacaaa
5761 gccttccacc ccatgtccgc atggctcctt tcactgtgtc cttgacacac ctctctggca
5821 acaactaaaa tttcctgctt ctgaaaagtc ctgtcttaaa agtacagtct atatcttgga
5881 aataaatagc tttcctcaag gcatgaaaaa aaa
```

[0065] The invention also encompasses splice variants of NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 and/or FIG4.

[0066] Known inhibitors of NPC1 include U18666A[25] (3-β-[2-(diethylamino)ethoxy]androst-5-en-17-one) and the antidepressant imipramine[26] (a tricyclic antidepressant).

[0067] VPS11, VPS16, VPS18, VPS33A, VPS39, and VPS41 are subunits of the homotypic fusion and vacuole protein sorting (HOPS) complex. The mammalian HOPS complex plays a critical role in fusion of endosomes and lysosomes[6]. One or more inhibitors may be used, for example, to inhibit one or more subunits of HOPS.

[0068] PIKFYVE is involved in the biogenesis of endosomes,[14,15] and BLOC1S1 and BLOC1S2 are involved in the biogenesis of lysosomes.[16] GNPTAB is involved in targeting of luminal cargo to the endocytic pathway.[17]

[0069] Inhibition of NPC1,VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 can occur at the level of the protein or at the level of nucleic acid (DNA or RNA) encoding the protein.

[0070] For example, the agent can be an antisense molecule, a ribozyme, or a RNA interference (RNAi) molecule, such as a small interfering RNA (siRNA) molecule, that specifically inhibits expression of NPC1,VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 protein. The agent can be comprised of nucleic acid (e.g., DNA or RNA) or nucleic acid mimetics (e.g., phosphorothionate mimetics) such as those known in the art.

[0071] The agent can also be, for example, an antibody, antibody fragment, aptamer or small molecule that specifically binds to NPC1,VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 and reduces its activity or interferes with its normal function. Antibody fragments include, but are not limited to, $F(ab')_2$ and Fab' fragments and single chain antibodies. $F(ab')_2$ is an antigen binding fragment of an antibody molecule with deleted crystallizable fragment (Fc) region and preserved binding region. Fab' is 1/2 of the $F(ab')_2$ molecule possessing only 1/2 of the binding region. The term antibody is further meant to encompass polyclonal antibodies and monoclonal antibodies. The antibody can be a human antibody or a non-human antibody such as a goat antibody or a mouse antibody. Antibodies can be "humanized" using standard recombinant DNA techniques. Aptamers are single stranded oligonucleotides or oligonucleotide analogs that bind to a particular target molecule, such as a protein. Thus, aptamers are the oligonucleotide analogy to antibodies. However, aptamers are smaller than antibodies. Their binding is highly dependent on the secondary structure formed by the aptamer oligonucleotide. Both RNA and single stranded DNA (or analog) aptamers can be used. Aptamers that bind to virtually any particular target can be selected using an iterative process called SELEX, which stands for Systematic Evolution of Ligands by EXponential enrichment.

[0072] Possible modes of action of antiviral compounds include those illustrated, for example, in Figure 24A-24D.

[0073] Rapidly acting small molecule inhibitors of NPC1,VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 protein may be preferred for treatment of viral infections due to the rapid speed of viral replication.

[0074] It is envisioned that administration of the agent to the subject would normally be limited to periods when the subject either has a filovirus infection or when the subject has been exposed to filovirus or is at risk of exposure to filovirus, in order to minimize any

deleterious effect of administration of the agent. Ebola/marburgvirus infections are typically acute in nature, so drug treatment of infection for only a short period of time is appropriate.

[0075] The agent can be administered to the subject in a pharmaceutical composition comprising a pharmaceutically acceptable carrier. Examples of acceptable pharmaceutical carriers include, but are not limited to, additive solution-3 (AS-3), saline, phosphate buffered saline, Ringer's solution, lactated Ringer's solution, Locke-Ringer's solution, Krebs Ringer's solution, Hartmann's balanced saline solution, and heparinized sodium citrate acid dextrose solution. The pharmaceutically acceptable carrier used can depend on the route of administration. The pharmaceutical composition can be formulated for administration by any method known in the art, including but not limited to, oral administration, parenteral administration, intravenous administration, transdermal administration, intranasal administration, and administration through an osmotic mini-pump. The compounds can be applied to the skin, for example, in compositions formulated as skin creams, or as sustained release formulations or patches.

[0076] The present invention also provides a method for screening for an agent that treats and/or prevents infection of a subject with a filovirus, the method comprising determining whether or not the agent inhibits one or more of Niemann-Pick C1 (NPC1), VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4, wherein an agent that inhibits NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 is a candidate for treating and/or preventing an infection with a filovirus and wherein an agent that does not inhibit NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 is not a candidate for treating and/or preventing an infection with a filovirus.

[0077] The agent used for treatment or in screening can be, for example, an agent that targets domain C of NPC1 or nucleic acid encoding domain C of NPC1. Domain C of NPC1 (Figure 19) is a 248-amino acid domain from residue 373 to residue 620 of SEQ ID NO:1.

[0078] The method can be carried out with respect to NPC1, for example, by measuring cholesterol transport, where a decrease in cholesterol transport in the presence of the agent indicates that the agent inhibits NPC1. The assay can be carried out using a cell line that expresses NPC1.

[0079] NPC1's cholesterol transport function is separable from its viral host factor function. Preferably, the agent selectively targets NPC1's viral host factor function, without blocking NPC1's cholesterol transport function.

[0080] The method can also be carried out, for example, by measuring binding between NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4, and the filovirus or a filovirus glycoprotein (GP), where a decrease in binding in the presence of the agent indicates that the agent inhibits NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4. The method can be carried out, for example, using a enzyme-linked-immunosorbent assay (ELISA). The method can be carried out, for example, using a electrochemiluminescence (ECL) assay.

[0081] The method can also be carried out, for example, by measuring filovirus infection in tissue culture, where a reduction in filovirus infection in the presence of the agent indicates that the agent inhibits NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 and/or FIG4.

[0082] The invention also provides an agent for treating and/or preventing infection of a subject with a filovirus identified by any of the methods disclosed herein for screening for an agent that treats and/or prevents infection of a subject with a filovirus. The invention further provides a pharmaceutical composition for treating and/or preventing infection of a subject with a filovirus comprising a pharmaceutically acceptable carrier and an agent identified by any of the methods disclosed herein for screening for an agent that treats and/or prevents infection of a subject with a filovirus.

[0083] This invention will be better understood from the Experimental Details, which follow. However, one skilled in the art will readily appreciate that the specific methods and results discussed are merely illustrative of the invention as described more fully in the claims that follow thereafter.

EXPERIMENTAL DETAILS

Introduction

[0084] A genome-wide haploid genetic screen in human cells is described for identifying host factors required for EboV entry. The screen uncovered 67 mutations disrupting all six members of the HOPS multisubunit tethering complex, which is involved in fusion of endosomes to lysosomes[6], and 39 independent mutations that disrupt the

endo/lysosomal cholesterol transporter protein Niemann-Pick C1 (NPC1)[7,8]. Cells defective for the HOPS complex or NPC1 function, including primary fibroblasts derived from human Niemann-Pick type C1 disease patients, are resistant to infection by EboV and MarV, but remain fully susceptible to a suite of unrelated viruses. Membrane fusion mediated by filovirus glycoproteins and viral escape from the vesicular compartment were shown to require the NPC1 protein, independent of its known function in cholesterol transport. The findings uncover unique features of the entry pathway used by filoviruses and indicate antiviral strategies to combat these deadly agents.

Methods

[0085] *Summary:* Adherent HAP1 cells were generated by the introduction of OCT4/SOX2/c-Myc and KLF4 transcription factors. 100 million cells were mutagenized using a promotor-less retroviral gene-trap vector carrying a GFP reporter. Cells were not selected for reporter gene expression, and insertion sites were mapped for approximately 1% of the unselected population using parallel sequencing. Cells were exposed to rVSV-GP-EboV and the resistant cell population was expanded and used to sequence insertion sites. Genes that were statistically enriched for mutation events in the selected population were identified, and the roles of selected genes in filovirus entry were characterized.

[0086] *Cells:* KBM7 cells and derivatives were maintained in IMDM supplemented with 10% FCS, L-glutamine, and penicillin–streptomycin. Vero grivet monkey cells and human dermal fibroblasts (Coriell Institute for Medical Research) were maintained in DMEM supplemented with 10% FCS, L-glutamine, and penicillin–streptomycin. Wild type and NPC1-null (CT43) Chinese hamster ovary (CHO) fibroblasts were maintained in DMEM-Ham's F-12 medium (50-50 mix) supplemented with 10% FCS, L-glutamine, and penicillin–streptomycin[21].

[0087] *Viruses:* A recombinant VSV expressing eGFP and EboV GP lacking the mucin domain (Δ309–489) (rVSV-GP-EboV) was recovered and amplified as described previously[11]. Recombinant rVSV-BDV was generously provided by Juan Carlos de la Torre. rVSV-GP-Rabies was generated by replacement of the VSV G ORF in VSV-eGFP (REF PMC 116335) with that of the SAD-B19 strain of Rabies virus, and recombinant virus was recovered and amplified as described[29].

[0088] The following non-recombinant viruses were used: Adenovirus type 5 (ATCC), Coxsackievirus B1 (ATCC), Poliovirus 1 Mahoney (generously provided by Christian

Schlieker), HSV-1 KOS (generously provided by Hidde Ploegh), Influenza A PR8 (Charles Rivers) and Rift valley fever virus MP-12 (generously provided by Jason Wojcechowskyj).

[0089] *Generation of HAP1 cells:* Retroviruses encoding SOX2, C-MYC, OCT4 and KLF4 were produced as described earlier[12]. Concentrated virus was used to infect near haploid KBM7 cells in three consecutive rounds of spin-infection with an interval of 12 hours. Conditions were used that resulted in an infection percentage of >95% of pLIB-EGFP (Clontech) that was taken along in a separate infection as a control. Cells were plated at low density in regular medium (IMDM 10% FCS, L-glutamine, and penicillin–streptomycin). Expression of the four transcription factors markedly changed morphology of the KBM7 cells from round, non-adherent cells typical for CML cells, to more flattened and adherent cells. Colonies were picked and tested for ploidy. One clonally derived cell line (referred to as HAP1) with a haploid DNA content as determined using DNA staining and flow cytometry was further grown and characterized. Karyotyping of this line demonstrated that the majority of the analyzed cells (27/39) were fully haploid, a smaller population (9/39) was haploid for all chromosomes except chromosome 8, like the parental KBM7 cells. Less than 10% (3/39) was diploid for all chromosomes except for chromosome 8 that was tetraploid. All cells carried the Philadelphia chromosome present in the parental KBM7 cells.

[0090] *Haploid genetic screen:* Gene trap virus was produced by transfection of 293T cells in T175 dishes using turbofectin 8 (origene) with a mixture of pGT-GFP, pGT-GFP+1 and pGT-GFP+2 (6.7 μg) combined with 1.7 μg pAdvantage, 2.6 μg CMV-VSVG and 4 μg Gag-pol. The virus-containing supernatant was concentrated using ultracentrifugation for 1.5 h at 25,000 r.p.m. in a Beckman SW28 rotor. To create a mutagenized cell population ~100 million HAP1 cells were infected with the gene-trap virus. After expansion for 7 days, a proportion of the cells was harvested for genomic DNA isolation to create a control dataset containing sequences flanking the gene-trap insertions in unselected cells. For the screen, hundred million mutagenized cells were exposed to rVSV-GP-EboV at an MOI ~100. The resistant colonies that grew out were expanded and ~30 million cells were used for genomic DNA isolation.

[0091] *Sequence analysis of gene trap insertion sites:* Insertion sites were identified en masse by sequencing the genomic DNA flanking gene trap proviral DNA as described before[9]. In short, a control dataset was generated containing insertion sites in mutagenized HAP1 cells before selection with rVSV-GP-EboV. For this purpose genomic DNA was

isolated from ~40 million cells and subjected to a linear PCR followed by linker ligation, PCR and sequencing using the Genome Analyzer platform (Illumina). The insertions sites were mapped on the human genome and insertion sites were identified that were located in genomic regions annotated to contain genes. The insertions in this control dataset comprise of ~400,000 independent insertions that meet this criteria. To generate the experimental dataset, insertions in the mutagenized HAP1 cells after selection with rVSV-GP-EboV were identified using an inverse PCR protocol followed by sequencing using the Genome Analyzer. The number of inactivating mutations (=sense orientation or present in exon) per individual gene was counted as well as the total number of inactivating insertions for all genes. Enrichment of a gene in the screen was calculated by comparing how often that gene was mutated in the screen compared to how often the genes carries an insertion in the control dataset. For each gene a p-value (corrected for false discovery rate) was calculated using the one-sided Fisher exact test.

[0092] *Characterization of the HAP1 mutant lines:* Clonal cell lines with gene trap insertion in NPC1, VPS11 and VPS33A were derived and genomic DNA was isolated using Qiamp DNA mini kit (Qiagen). To confirm that the cells were truly clonal and to confirm the absence of the wild type DNA locus, a PCR was performed with primers flanking the insertion site using the following primers:

NPC-F1, 5'-GAAGTTGGTCTGGCGATGGAG-3' (SEQ ID NO:27);

NPC1-R2, 5'-AAGGTCCTGATCTAAAACTCTAG-3' (SEQ ID NO:28);

VPS33A-F1, 5'-TGTCCTACGGCCGAGTGAACC-3' (SEQ ID NO:29);

VPS33A-R1, 5'-CTGTACACTTTGCTCAGTTTCC-3' (SEQ ID NO:30);

VPS11-F1, 5'-GAAGGAGCCGCTGAGCAATGATG-3' (SEQ ID NO:31);

VPS11-R1, 5'-GGCCAGAATTTAGTAGCAGCAAC-3' (SEQ ID NO:32).

To confirm the correct insertion of the gene trap at the different loci a PCR was performed using the reverse (R1) primers of NPC1, VPS11 and VPS33A in combination with a primer specific for the gene trap vector: PGT-F1; 5'-TCTCCAAATCTCGGTGGAAC-3' (SEQ ID NO:33). To determine RNA expression levels of NPC1, VPS11 and VPS33A in the respective mutants, total RNA was extracted using RNeasy (Qiagen), reverse transcribed using Superscript III (Invitrogen) and PCR amplified using gene specific primers:

VPS11: 5'-CTGCTTCCAAGTTCCTTTGC-3' (SEQ ID NO:34) and

5'-AAGATTCGAGTGCAGAGTGG-3' (SEQ ID NO:35);

NPC1: 5'-CCACAGCATGACCGCTC-3' (SEQ ID NO:36) and

5'-CAGCTCACAAAACAGGTTCAG-3' (SEQ ID NO:37);
VPS33A: 5'-TTAACACCTCTTGCCACTCAG-3' (SEQ ID NO:38) and
5'-TGTGTCTTTCCTCGAATGCTG-3' (SEQ ID NO:39).

[0093] *NPC1 constructs:* Human NPC1 cDNA was ligated in-frame to a triple flag sequence, and the resulting gene encoding C-terminally FLAG-tagged NPC1 was subcloned into the *BamHI* and *SalI* restriction sites of the pBABE-puro retroviral vector[30]. Constructs encoding flag-tagged NPC1 'loop-minus' mutants in pBABE-puro [ΔA, lacking NPC1 amino acid residues 24-252); ΔC, lacking residues 381-611); ΔI, (lacking residues 865-1088)] were generated by replacing the indicated sequence with a *BglII* restriction site. To engineer the individual loop domain constructs, a cassette vector encoding the following sequence elements was first generated and cloned into the *BamHI* and *SalI* sites of pBABE-puro: NPC1 signal peptide (encoding NPC1 amino acid residues 1-24), *MluI* restriction site, the first NPC1 transmembrane domain (residues 267-295), NPC1 C-tail (residues 1252-1278), gly-gly-gly-ser linker, and triple flag tag. Each loop domain (A, residues 25-266; C, residues 373-620; I, residues 854-1098) was cloned into the *MluI* site of this cassette vector. All constructs were verified by automated DNA sequencing.

[0094] *CT43 cell populations stably expressing NPC1 proteins:* For transduction of VH-2 cells, the full-length human NPC1 cDNA (Origene) was cloned into the retroviral vector pMXsIRESblasti-FLAG[10]. For transduction of CHO WT and CT43 cells, the pBABE-puro-based retroviral vectors described above were used. Retroviruses packaging the transgenes were produced by triple transfection in 293T cells, and target cells were directly exposed to sterile-filtered retrovirus-laden supernatants in the presence of polybrene (6 μg/mL). Transduced cell populations were selected with blasticidin (20 μg/mL; for pMX) or puromycin (10 μg/mL; for pBABE-puro).

[0095] *Cell viability assays for virus treatments:* KBM7 and HAP1 cells were seeded at 10,000 cells per well in a 96-well tissue culture plate and treated with the indicated concentrations of rVSV-GP-EboV or left untreated. Three days after treatment the cell viability was measured using an XTT colorimetric assay (Roche) according to manufacturer's protocol. Viability is plotted as percentage viability compared to untreated control. To compare susceptibility of the HAP1 mutants to different viruses, they were seeded at 10,000 cells per well and treated with different cytolytic viruses at a concentration that in pilot experiments was the lowest concentration to produce extensive cytopathic effects. Three days after treatment, viable, adherent cells were fixed with 4% formaldehyde

in phosphate-buffered saline (PBS) followed by staining with 0.5% crystal violet dye in 70% ethanol for 30 min. After three gentle washes with water, air-dried plates were scanned.

[0096] *Viral infectivity measurements:* Infectivities of VSV pseudotypes were measured by manual counting of eGFP-positive cells using fluorescence microscopy at 16-26 h post-infection, as described previously[3]. rVSV-GP-EboV infectivity was measured by fluorescent-focus assay (FFA), as described previously[11].

[0097] *Filipin staining:* Filipin staining to visualize intracellular cholesterol was done essentially as described[31]. Briefly, cells were fixed with paraformaldehyde (3%) for 15 min at room temperature. After three PBS washes, cells were incubated with filipin complex from *Streptomyces filipinensis* (Sigma-Aldrich) (50 μg/mL) in the dark for 1 h at room temp. After three PBS washes, cells were visualized by fluorescence microscopy in the DAPI channel.

[0098] *Measurements of cysteine cathepsin activity:* The enzymatic activities of CatB and CatL in acidified postnuclear extracts of Vero cells, human fibroblasts, and CHO lines were assayed with fluorogenic peptide substrates Z-Arg-Arg-AMC (Bachem Inc., Torrance, CA) and (Z-Phe-Arg)2-R110 (Invitrogen), respectively, as described previously[32]. As a control for assay specificity, enzyme activities were also assessed in extracts pretreated with E-64 (10 μM), a broad-spectrum cysteine protease inhibitor, as previously described[11]. Active CatB and CatL within intact cells were labeled with the fluorescently-labeled activity-based probe GB111 (1 μM) and visualized by gel electrophoresis and fluorimaging, as described previously[33].

[0099] *Purification and dye conjugation of rVSV-GP-EboV:* rVSV-GP-EboV was propagated, purified and labeled with Alexa Fluor 647 (Molecular Probes, Invitrogen Corporation) as described previously[34] with minor modifications. Briefly, Alexa Fluor 647 (Molecular Probes, Invitrogen Corporation) was solubilized in DMSO at 10 mg/mL and incubated at a final concentration of 31.25 μg/ml with purified rVSV-GP-EboV (0.5 mg/ml) in 0.1 M NaHCO3 (pH 8.3) for 90 min. at RT. Virus was separated from free dye by ultracentrifugation. Labeled viruses were resuspended in NTE (10 mM Tris pH 7.4, 100 mM NaCl, 1 mM EDTA) and stored at −80°C.

[00100] *Virus binding/internalization assay:* Cells were inoculated with an MOI of 200-500 of Alexa 647-labeled rVSV-GP-EboV at 4°C for 30 min. to allow binding of virus particles to the cell surface. Cells were subsequent fixed in 2% paraformaldehyde (to

examine virus binding) or following a 2 h incubation at 37°C and an acid wash to remove surface-bound virus. The cellular plasma membrane was labeled by incubation of cells with 1 ug/mL Alexa Fluor 594 wheat germ agglutinin (Molecular Probes, Invitrogen) in PBS for 15 min. at RT. External virus particles were detected using a 1:2000 dilution of antibody 265.1, a mouse monoclonal specific for Ebola GP. The GP antibodies were detected by Alexa 488-conjugated goat anti-mouse secondary antibody (Molecular Probes, Invitrogen). After washing with PBS, cells were mounted onto glass slides using Prolong Antifade Reagent (Invitrogen, Molecular Probes). Fluorescence was monitored with a epifluorescence microscope (Axiovert 200M; Carl Zeiss, Inc.; Thornwood, NY) and representative images were acquired using Slidebook 4.2 software (Intelligent Imaging Innovations; Denver, CO)[34, 35].

[00101] *VSV M protein-release assay:* Cells grown on 12 mm coverslips coated with poly-D-lysine (Sigma-Aldrich) were pre-treated with 5 µg/ml puromycin for 30 min. and inoculated with rVSV at an MOI of 200-500 in the presence of puromycin. After 3 h, cells were washed once with PBS and fixed with 2% paraformaldehyde in PBS for 15 min. at RT. To detect VSV M protein, fixed cells were incubated with a 1:7500 dilution of monoclonal antibody 23H12 (kind gift of Doug Lyles[36]), in PBS containing 1% BSA and 0.1 % Triton X-100 for 30 min. at RT. Cells were washed three times with PBS, and the anti-M antibodies were detected using a 1:750 dilution of Alexa 594-conjugated goat anti-mouse secondary antibodies. In addition, cells were counter-stained with DAPI to visualize nuclei. Cells were washed three times and mounted onto glass slides after which M localization images were acquired using a Nikon TE2000-U inverted epifluorescence microscope (Nikon Instruments, Inc.; Melville, NY). Representative images were acquired with Metamorph software (Molecular Devices).

[00102] *Electron microscopy:* Confluent cell monolayers in 6-well plates were inoculated with rVSV-GP-EboV at a MOI of 200-500 for 3 h. Subsequently, cells were fixed for at least 1 h at RT in a mixture of 2.5% glutaraldehyde, 1.25% paraformaldehyde and 0.03% picric acid in 0.1 M sodium cacodylate buffer (pH 7.4). Samples were washed extensively in 0.1 M sodium cacodylate buffer (pH 7.4) after which they were treated with 1% osmiumtetroxide and 1.5% potassiumferrocyanide in water for 30 min. at RT. Treated samples were washed in water, stained in 1% aqueous uranyl acetate for 30 min., and dehydrated in grades of alcohol (70%, 90%, 2×100%) for 5 min. each. Cells were removed from the dish with propyleneoxide and pelleted at 3,000 rpm for 3 min.. Samples were

infiltrated with Epon mixed with propyleneoxide (1:1) for 2 h at RT. Samples were embedded in fresh Epon and left to polymerize for 24-48 h at 65°C. Ultrathin sections (about 60–80 nm) were cut on a Reichert Ultracut-S microtome and placed onto copper grids. Images were acquired using a Technai G2 Spirit BioTWIN (Fei, Hillsboro, OR) transmission electron microscope.

[00103] *Authentic filoviruses and infections:* Cells were exposed to EBOV-Zaire 1995 or MARV-Ci67 at an MOI of 3 for 1 h. Viral inoculum was then removed and fresh culture media was added. At 48 h post-infection, cells were fixed with formalin, and blocked with 1% bovine serum albumin. EBOV-infected cells and uninfected controls were incubated with EBOV GP-specific monoclonal antibodies 13F6 or KZ52. MARV-infected cells and uninfected controls were incubated with MARV GP-specific monoclonal antibody 9G4. Cells were washed with PBS prior to incubation with either goat anti-mouse IgG or goat anti-human IgG conjugated to Alexa 488. Cells were counterstained with Hoechst stain (Invitrogen), washed with PBS and stored at 4°C. Infected cells were quantitated by fluorescence microscopy and automated image analysis. Images were acquired at 9 fields/well with a 10× objective lens on a Discovery-1 high content imager (Molecular Devices) or at 6 fields/well with a 20× objective lens on an Operetta (Perkin-Elmer) high content device. Discovery-1 images were analyzed with the "live/dead" module in MetaXpress software. Operetta images were analyzed with a customized scheme built from image analysis functions present in Harmony software.

[00104] *Animals and filovirus challenge experiments:* Mouse-adapted MarV Ci67 was provided by Sina Bavari[47]. Female and male BALB/c NPC1$^{+/-}$ mice and BALB/c NPC1 $^{+/+}$ mice (5 to 8 week old) were obtained from Jackson Laboratory (Bar Harbor, ME). Mice were housed under specific-pathogen-free conditions. Research was conducted in compliance with the Animal Welfare Act and other federal statutes and regulations relating to animals and experiments involving animals and adhered to principles stated in the Guide for the Care and Use of Laboratory Animals (National Research Council, 1996). The facility where this research was conducted is fully accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care International. For infection, mice were inoculated intraperitoneally (i.p.) with a target dose of 1000 pfu (30,000 X the 50% lethal dose) of mouse-adapted EboV or mouse-adapted MarV Ci67 virus in a biosafety level 4 laboratory. Mice were observed for 28 days after challenge by study personnel and by an impartial third party. Daily observations included evaluation of mice for clinical symptoms

such as reduced grooming, ruffled fur, hunched posture, subdued response to stimulation, nasal discharge, and bleeding. Serum was collected from surviving mice to confirm virus clearance. Back titration of the challenge dose by plaque assay determined that EboV-infected mice received 900 pfu/mouse and MarV-infected mice received 700 pfu/mouse.

[00105] *GP-NPC1 co-immunoprecipitation (co-IP) assays:* Protein G-coated magnetic beads (20 μL/reaction; Spherotech) were incubated with the GP-specific monoclonal antibody KZ52 (5 μg) for 1 h, washed to remove unbound antibody, and then added to uncleaved or in vitro-cleaved rVSV-GP-EBOV or VSV-GP-EBOV particles (5 μL concentrated virus; 10^7–10^8 infectious units), or to purified EBOV GPΔTM (9 μg) in NTE-CHAPS buffer (10 mM Tris.Cl [pH 7.5], 140 mM NaCl, 1 mM EDTA, 0.5% vol/vol CHAPS (3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate)). Bead-virus mixtures were incubated for 2 h at room temperature, and then added to crude detergent extracts of CHO CT43 cells expressing a flag-tagged NPC1 protein (NPC1-flag) (2×10^6 cell-equivalents in 150 μL), or to purified, soluble NPC1 domain C (5 μg/mL). After overnight incubation with mixing at 4°C, beads were retrieved with a magnet, extensively washed with NTE-CHAPS, and heated in Laemmli sample buffer to elute bound proteins. Solubilized proteins were subjected to SDS-polyacrylamide gel electrophoresis, and NPC1 and GP were detected by immunoblotting with anti-flag (Sigma-Aldrich) and anti-GP1 antibodies, respectively.

[00106] *GP-NPC1 capture ELISA:* 96-well high-binding ELISA plates (Corning) were coated with the GP-specific monoclonal antibody KZ52 (2 μg/mL in PBS), and then blocked with PBS containing 3% bovine serum albumin and 0.5% CHAPS (PBSA-CHAPS). Uncleaved or in vitro-cleaved rVSV-GP or VSV-GP particles solubilized in PBSA-CHAPS buffer were added to the blocked plates, and GP capture was allowed to proceed for 1 h at 37°C. After washing to remove unbound GP, serial dilutions of NPC1-flag partially purified from CT43 cells (0–100 ng/well), crude detergent extracts of 293T cells expressing flag-tagged NPC1 or NPC1L1 proteins (0–2×10^5 cell-equivalents), or purified, soluble domain C (0–40 μg/mL) were added to the wells. After an overnight incubation at 4°C, plates were extensively washed, and bound flag-tagged proteins were detected with an anti-flag antibody-horseradish peroxidase conjugate and Ultra-TMB substrate (Thermo).

[00107] *Affinity purification of NPC1-flag:* CT43 cells expressing NPC1-flag (2×10^8 cells) were harvested and lysed as above, and the extracts were incubated with magnetic

beads coated with anti-flag antibody (0.25 mL) at 4°C with mixing for 12-16 h. Beads were then extensively washed with NTE-CHAPS, and bound proteins were eluted with 10 packed-bead volumes of triple flag peptide (5 mg/mL; Sigma). The eluate was concentrated and buffer-exchanged using a centrifugal concentrator (100 kDa molecular weight cutoff; Pall Biosciences), and NPC1-flag purity was assessed by SDS-PAGE and staining with the Krypton infrared protein-binding dye (Thermo).

[00108] *Generation and purification of soluble domain C and GPΔTM proteins:* A construct engineered to encode NPC1 domain C (residues 372-622) flanked by sequences that form a stable, antiparallel coiled coil, and fused to a preprotrypsin signal sequence and flag and hexahistidine tags at its N–terminus. A plasmid encoding EBOV GPΔTM (residues 1-650) fused to a hexahistidine tag at the C–terminus was kindly provided by G.G. Olinger (USAMRIID). Soluble domain C was expressed in human 293-Freestyle cells (Invitrogen) and purified from conditioned supernatants by nickel affinity chromatography. GPΔTM was expressed in 293-EBNA cells (ATCC) and purified from conditioned supernatants in a similar manner.

[00109] *Neutralization of rVSV-GP-EBOV by soluble domain C:* Uncleaved or cleaved rVSV-GP-EBOV particles were mixed with soluble domain C for 1 h at room temp. Subsequently, the virus mixtures were diluted and exposed to Vero cell monolayers for 1 h at 37°C, at which time NH_4Cl (20 mM) was added to block additional entry events and cell-to-cell spread. Viral infectivity was determined at 12-16 h post-infection by enumerating eGFP-positive cells.

Results

[00110] Haploid genetic screens have previously been used to gain insight into a variety of biological processes relevant to human disease[9,10]. Here this approach was used to explore the cell entry pathway used by filoviruses at an unprecedented level of detail. To interrogate millions of independent gene disruption events in human cells for associated defects in EboV entry, a replication-competent vesicular stomatitis virus bearing the EboV glycoprotein (rVSV-GP-EboV)[11] was used to select for resistant cells. Although this recombinant virus multiplies in and kills most cultured cell lines, it grew poorly in near-haploid KBM7 cell cultures (Figure 5C). To obtain a system suitable for a haploid genetic screen using rVSV-GP-EboV, experiments were undertaken to alter the differentiation state of KBM7 cells[12]. In an unsuccessful attempt to induce pluripotency in KBM7 cells through the expression of OCT4, SOX-2, c-MYC and KLF4[13], a cell line was obtained that was

termed HAP1 (Figure 5A). HAP1 cells grow adherently, can be clonally expanded, and no longer express markers associated with hematopoietic cells (Figure 5B). The majority of cells in early passage cultures of HAP1 cells are haploid for all chromosomes, including chromosome 8 (which is present in two copies in KBM7 cells). Unlike KBM7 cells, HAP1 cells did support robust multiplication of rVSV-GP-EboV and were rapidly killed by it (Figure 5C), thus allowing mutagenesis-based screens for essential filovirus host factors.

[00111] A retroviral promoter-less gene trap vector[10] was used to perform insertional mutagenesis on early-passage HAP1 cells, creating a library of cells with single-gene disruptions. To generate a control dataset, ~800,000 insertion events were mapped in unselected cells using deep sequencing. Next, ~100 million mutagenized cells were exposed to rVSV-GP-EboV. Cells resistant to killing by this virus were expanded as a pool, and insertion sites were mapped using parallel sequencing. Enrichment for mutations in a particular gene was calculated by comparing the gene's mutation frequency in resistant cells to that observed in the unselected control dataset (Figure 6). Similar experiments in KBM7 cells have linked numerous genes to a variety of phenotypes with high confidence[9]. Using this approach, a set of genes enriched for mutations in the rVSV-GP-EboV-resistant cell population were identified (Figures 1A and 4C). Nearly all of the candidate host factors encoded by these genes are involved in the architecture and trafficking of endo/lysosomal compartments, highlighting the central importance of this pathway for EboV GP-dependent infection. The screen identified the endosomal cysteine protease cathepsin B (CatB), the only host factor whose genetic deletion was previously demonstrated to inhibit EboV GP-dependent entry[3]. Further inspection showed that mutations were highly enriched in all 6 subunits of the homotypic fusion and vacuole protein sorting (HOPS) complex (VPS11, VPS16, VPS18, VPS33A, VPS39 and VPS41), for which a total of 67 independent mutations were identified. Like its well-studied yeast counterpart, the mammalian HOPS complex plays a critical role in fusion of endosomes and lysosomes[6]. The identification of all 6 members of the HOPS complex demonstrates the high, and possibly saturating, coverage of the mutagenesis screen. A number of additional genes were also identified whose products are involved in the biogenesis of endosomes (PIKFYVE)[14,15] and lysosomes (BLOC1S1, BLOC1S2)[16], and in targeting of luminal cargo to the endocytic pathway (GNPTAB)[17]. Finally, the single strongest hit obtained, with 39 independent gene-trap insertions, was the Niemann-Pick disease locus NPC1, which encodes an endo/lysosomal cholesterol transporter[8].

[00112] Neither the HOPS complex nor NPC1 has previously been implicated in the entry of any type of virus. To investigate their roles in filovirus entry, the resistant cell population was subcloned to obtain clones deficient for the HOPS subunits VPS11 and VPS33A, and for NPC1 (Figure 7A, 7B). Expression of the corresponding gene products was no longer detected in these clones (Figure 8A). NPC1-, VPS11- and VPS33A-null cells displayed marked resistance to infection by rVSV-GP-EboV and VSV pseudotypes bearing EboV or MarV GP (Figures 1C, and 7C). NPC1-deficient cells were completely refractory to infection by these viruses. Cells that lack a functional HOPS complex or NPC1 were nonetheless fully susceptible to infection by a large panel of other enveloped and nonenveloped viruses, including native VSV and recombinant VSVs bearing the Rabies and Borna disease virus glycoproteins[17] (Figure 1D). These mutant cells also fully supported infection by influenza A virus, which enters cells via late endosomes[19] (Figure 1D). Therefore, deficiency of VPS11, VPS33A, or NPC1 causes resistance specifically to viral entry mediated by filovirus glycoproteins.

[00113] Loss of NPC1 function causes Niemann-Pick disease, a hereditary neurovisceral disorder characterized by the accumulation of cholesterol and sphingolipids within lysosomes[8,20,21]. Tests were conducted of the susceptibility of patient fibroblasts carrying homozygous mutations in NPC1 to filovirus GP-dependent infection. As expected, control cells derived from a healthy individual were readily infected by rVSV-GP-EboV and VSV pseudotyped with GP proteins derived from EboV, Sudan virus, or MarV, whereas NPC1-mutant cells were infected poorly or not at all (Figures 2A, B). By contrast, both types of cells were efficiently infected by native VSV. The susceptibility of NPC1-deficient fibroblasts to rVSV-GP-EboV infection was restored by retroviral expression of wild type NPC1, confirming that loss of the NPC1 protein is responsible for the infection defect (Figure 2C).

[00114] Mutations in a second gene, NPC2, cause identical clinical symptoms and phenocopy the defects in cellular lipid transport[7]. Surprisingly, NPC2-mutant fibroblasts derived from two different patients were susceptible to filovirus GP-dependent infection, despite similar capacities of the NPC2- and NPC1-mutant cells to accumulate cholesterol in lysosomes (Figures 2A, 2B and 9). Furthermore, clearance of accumulated cholesterol from NPC1-null cells by prolonged cultivation in lipoprotein-depleted growth medium did not confer susceptibility to rVSV-GP-EboV infection (Figure 10A, 10B). Thus, resistance of NPC1-deficient cells to rVSV-GP-EboV is not caused by defects in lipid transport *per se*,

consistent with the results of the screen, which did not identify NPC2 as a host factor for EboV entry (Figure 1A).

[00115] Filoviruses display broad mammalian host and tissue tropism and can infect a wide variety of cell types in culture[22,,23]. To determine if NPC1 is generally required for filovirus GP-mediated infection, rVSV-GP-EboV infection was measured in NPC1-null Chinese hamster ovary (CHO) cells[24]. Loss of NPC1 conferred complete resistance to viral infection (Figures 8B and 10B) that could be reversed by expression of human NPC1 (Figure 10B). Therefore, NPC1 plays a critical role in entry mediated by filovirus glycoproteins that is conserved in mammals.

[00116] Filovirus particles can probably bind to a diverse set of cell-surface molecules[4,31], upon which they undergo internalization by a macropinocytosis-like mechanism[32,33], and traffic to late endosomal compartment(s) where GP is cleaved by endosomal cysteine proteases[3]. Cleaved GP then mediates fusion of viral and endosomal membranes, thereby releasing the viral nucleocapsid into the cytoplasm[34]. To determine which step(s) in filovirus entry require the HOPS complex and NPC1, an assessment was conducted of possible defects in attachment and internalization of rVSV-GP-EboV in VPS33A- and NPC1-null HAP1 cells. No significant difference were observed in binding of Alexa 647 fluorophore-labeled rVSV-GP-EboV to wild type and mutant cells at 4°C (not shown). Cells with bound virus were then warmed to 37°C to promote endocytosis and acid-washed to strip non-internalized viral particles from the cell surface. Fluorescent microscopy showed similar levels of internalized rVSV-GP-EboV in wild type and mutant cells (not shown). Consistent with these findings, bullet-shaped VSV particles were readily observed by electron microscopy at the cell periphery and within plasma membrane invaginations resembling nascent macropinosomes (Figure 3A). Therefore, GP-mediated entry is not inhibited at binding or internalization steps in NPC1- or HOPS-defective cells, suggesting a downstream block.

[00117] Cleavage of EboV GP by CatB and/or cathepsin L (CatL) is a prerequisite for viral membrane fusion[3,5]. Mutant HAP1 cells possess normal levels of CatB/CatL enzyme activity (Figure 11B, 11C) and remained refractory to infection by in vitro-cleaved rVSV-GP-EboV particles (Figure 3C) that no longer required CatB/CatL activity within Vero cells (Figure 11A). Therefore, the HOPS complex and NPC1 are likely required at step(s) downstream of GP proteolytic processing.

[00118] The intracellular distribution of the internal VSV M (matrix) protein was used as a marker for successful membrane fusion in VPS33A- and NPC1-null HAP1 cells (Figure 3D). Cells were exposed to native VSV or rVSV-GP-EboV in the presence of puromycin to block protein synthesis, and then fixed and immunostained to visualize the incoming M protein. Productive entry into wild type HAP1 cells caused redistribution of the incoming viral M throughout the cytoplasm (Figure 3C), whereas a membrane fusion block imposed by agents that elevate endosomal pH resulted in punctate M staining (Figure 12). Diffuse M staining was also observed for VSV in U18666A-treated wild type cells, and in HOPS complex- and NPC1-null cells (Figure 3C), consistent with the capacity of VSV to productively infect these cells (Figure 1D). By contrast, only punctate M staining was obtained in drug-treated and mutant cells exposed to rVSV-GP-EboV (Figure 3C). Electron micrographs of mutant cells revealed agglomerations of viral particles within vesicular compartments (Figures 3D and 13), reinforcing the conclusion that fusion and uncoating of the incoming rVSV-GP-EboV is arrested. Therefore, NPC1 and a functional HOPS complex are required for late step(s) in filovirus entry leading to viral membrane fusion.

[00119] The above experiments were done with recombinant or pseudotyped VSV particles bearing filovirus glycoproteins. Because these surrogate systems may not faithfully represent all aspects of filovirus infection, it was tested if infection and multiplication by authentic EboV and MarV are affected in NPC1-mutant patient fibroblasts. Consistent with the findings with VSV particles, yields of infectious viral progeny were profoundly reduced for both viruses in the mutant cells, relative to control fibroblasts (Figure 4A). Therefore, NPC1 is essential for authentic filovirus infection.

[00120] Certain small molecules such as U18666A[25] and the antidepressant imipramine[26] are known to cause a cellular phenotype similar to that observed in Niemann-Pick disease, in part by targeting the NPC1 protein[27,28,29]. Both compounds potently inhibited viral infection mediated by EboV GP but not VSV in Vero grivet monkey cells (Figure 2D for U18666A; imipramine not shown). U18666A inhibited viral infection almost immediately after its addition to Vero cells (<10 min) (Figure 15A) and before significant intracellular accumulation of cholesterol could be observed (>4 h) (data not shown)[30]. Moreover, sensitivity of viral infection to U18666A was lost by ~2 h post-infection, indicating that U18666A inhibits infection at the entry step (Figure 15B).

[00121] The effect of U18666A and imipramine on infection by authentic EboV and MarV was examined. Stark reductions in viral yield were obtained in Vero cells treated with

either drug (Figures 4B and 16). Moreover U18666A greatly reduced infection of human peripheral blood monocyte-derived dendritic cells and umbilical-vein endothelial cells (HUVEC) (Figure 14A, 14B), without affecting cell number or morphology. These findings indicate that filovirus entry and infection is sensitive to perturbation by small-molecule inhibitors of NPC1.

[00122] The effect of NPC1 mutation in lethal mouse models of EboV and MarV infection was assessed. Heterozygous NPC1 ($NPC1^{-/+}$) knockout mice and their wild type littermates were challenged with mouse-adapted EboV or MarV and monitored for 28 days. Whereas $NPC1^{+/+}$ mice rapidly succumbed to infection with either filovirus, $NPC1^{-/+}$ mice were largely protected (Figure 17A, 17B). Therefore, NPC1 is critically required for filovirus *in vivo* pathogenesis.

[00123] Given its efficacy in tissue culture, the protective capacity of imipramine was examined in the lethal mouse model of of EboV infection. Mice administered a single dose of imipramine 2 h before EboV challenge were substantially protected from filovirus challenge. Although the efficacy of imipramine at interrupting NPC1 function *in vivo* was not examined, these findings provide the first evidence that pharmacological inhibition of NPC1 *in vivo* can confer protection against filovirus infection.

[00124] To determine if filovirus entry requires the entire NPC1 protein or can instead be attributed to a discrete region within it, NPC1 deletion mutants individually lacking the large luminal loop domains A, C, and I (Figure 19) were expressed in an NPC1-null cell line (Chinese hamster ovary [CHO] CT43[21]), and their capacity to mediate lysosomal cholesterol transport and viral infection was examined (Figure 20). CT43 cells accumulated lysosomal cholesterol[21], and they were completely resistant to infection by wild type EBOV/MARV and rVSV-GP-EBOV/MARV[10]. As shown previously[10], expression of flag epitope-tagged WT NPC1 (NPC1-flag) in these cells not only corrected their cholesterol transport defect but also rendered them highly susceptible to infection by wild type filoviruses and rVSVs bearing filovirus glycoproteins. All three 'loop-minus' NPC1 mutants were inactive at lysosomal cholesterol transport (Figure 20A), despite their significant localization to LAMP1-positive late endosomal/lysosomal compartments (not shown), confirming that this cellular activity of NPC1 requires all three luminal domains A, C, and I. However, the mutants differed in their capacity to support filovirus GP-mediated entry. Both NPC1-ΔA-flag and NPC-ΔI-flag could mediate entry, albeit at reduced levels relative to WT NPC1-flag. In striking contrast, NPC1-ΔC-flag was unable to rescue viral

entry (Figure 20) even though it resembled the other mutants in expression level and intracellular distribution (not shown). Similar results were obtained in infection assays with wild type MARV (Figure 20C). These findings unequivocally separate NPC1's functions in lysosomal cholesterol transport and filovirus entry. More importantly, they demonstrate that a discrete region within NPC1, the luminal domain C, is essential for EBOV and MARV entry.

[00125] The preceding experiment raised the possibility that filovirus GP uses NPC1 to enter cells by interacting directly with this protein without regard to its normal cellular functions. To examine this hypothesis, it was first tested if EBOV GP could bind to NPC1 in a cell- and membrane-free system. Concentrated rVSV-GP-EBOV particles were solubilized in a nonionic detergent-containing buffer, and the GP protein in these extracts was captured by magnetic beads coated with the GP-specific monoclonal antibody KZ52. These GP-decorated beads did not retrieve NPC1-flag from CT43 detergent extracts in a co-immunoprecipitation (co-IP) assay (Figure 21A). Next rVSV-GP-EBOV was incubated with the bacterial metalloprotease thermolysin to generate a GP intermediate (GP_{CL}) that resembles the product of endo/lysosomal GP cleavage[3,5]. GP_{CL} could capture NPC1-flag at both neutral and acid pH (Figure 21A). Similar results were obtained in a reciprocal co-IP experiment: magnetic beads displaying NPC1-flag captured GP_{CL} but not GP (not shown).

[00126] To confirm these findings, the capacity of rVSV-derived GP and GP_{CL} to capture NPC1-flag from 293T human embryonic kidney cell extracts was examined using an enzyme-linked immunosorbent assay (ELISA). GP and GP_{CL} were captured onto antibody KZ52-coated ELISA plates, and then incubated with CT43 extracts containing NPC1-flag. NPC1-flag bound saturably to wells coated with GP_{CL} but not with GP, consistent with the results from the co-IP assay (Figure 21B). Affinity-purified NPC1-flag (Figure 21C) bound saturably to wells coated with GP_{CL} but not GP in the ELISA, providing evidence that GP_{CL} directly interacts with NPC1 (Figure 21D). Cumulatively, these findings demonstrate that the proteolytic priming of EBOV GP creates, or unmasks, a specific and direct binding site for NPC1.

[00127] It was next tested if NPC1 domain C is not only necessary but also sufficient to mediate EBOV GP_{CL}-NPC1 binding. To examine the GP_{CL}-NPC1 interaction with 'soluble proteins' in the absence of detergent, a soluble, secreted, and biologically-active form of domain C[40] was engineered and its binding to GP_{CL} was tested. Cleaved rVSV-GP_{CL}, but not uncleaved rVSV-GP, captured purified domain C in an ELISA (Figure 22A). Even more

stringently, GP_{CL} derived from a purified, soluble GP protein lacking the transmembrane domain (GPΔTM) co-precipitated purified domain C, whereas uncleaved GPΔTM did not (Figure 22B). Consistent with its capacity to bind directly and stably to GP_{CL}, soluble domain C neutralized infection by rVSV-GP_{CL} but not rVSV-GP in a dose-dependent manner (Figure 22C). Therefore, NPC1 domain C directly and specifically binds to a cleaved form of the EboV glycoprotein.

[00128] Finally, it was asked if a synthetic single-pass membrane protein containing only NPC1 domain C could mediate filovirus entry. Accordingly, NPC1 luminal domains A, C, and I were separately fused to the first transmembrane domain of NPC1, the NPC1 cytoplasmic tail, and a flag tag, and expressed in CT43 cells. All three proteins were expressed to similar levels, and domain A-flag and domain C-flag localized significantly to late endosomes and/or lysosomes (not shown). The capacity of these engineered single-domain transmembrane proteins to mediate viral entry was tested in CT43 cells (Figure 23A, 23B). Remarkably, only domain C-flag afforded measurable, although incomplete, rescue of filovirus GP-dependent entry, in full agreement with the GP_{CL}-binding activity of domain C (Figure 22). Taken together, these results indicate that sequences essential for both the EboV GP binding and entry host factor activities of NPC1 reside within domain C, a 248-amino acid domain of this 1278-amino acid protein that protrudes into the endosomal lumen. These findings, together with other functional data presented herein, also indicate that NPC1 is a critical endosomal receptor for cell entry by the Ebola and Marburg viruses.

[00129] The current work enables the development of small molecule antivirals targeting the NPC1 protein in cells and hosts (Figure 24). A number of possible modes of action for these antivirals are envisioned, only some of which are detailed here. For example, these molecules may (1) directly inhibit the GP-NPC1 virus-receptor interaction during entry (Figure 24A) by blocking the binding site in either protein (Figure 24B); (2) indirectly inhibit the GP-NPC1 virus-receptor interaction during entry by binding to NPC1 and inducing a conformational change in this protein (Figure 24C); (3) indirectly inhibit the GP-NPC1 virus-receptor interaction during entry by binding to an associated cellular component (e.g., protein or lipid) and inducing a conformational change in this protein (not pictured); and/or (4) reduce levels of the NPC1 protein by causing it to misfold, or otherwise targeting it for degradation within cells (Figure 24D).

[00130] The current work enables the development of assays for identification of small molecule inhibitors of the GP-NPC1 interaction by high-throughput screening. For example,

results are provided with an enzyme-linked immunosorbent assay (ELISA) to detect the binding of GP to intact NPC1 or NPC1 domain C-containing fragment (Figures 21B, 21D, 22A), which may be adapted to high-throughput screening. One possible embodiment of such a screening assay is a homogeneous electrochemiluminescence (ECL) assay to measure the binding of purified GP to immobilized endosomal membrane fragments containing the complete NPC1 protein (Figure 25). A second possible embodiment of such a screening assay is a homogeneous assay in which interaction of GP and NPC1 domain C brings two distinct functionalized beads into proximity, resulting in the emission of light at a specific wavelength that can be measured with the appropriate instrumentation. The current work also enables other types of GP-NPC1 interaction assays.

[00131] Global disruption of nonessential human genes as described here has provided a solid genetic framework for understanding the unusual entry pathway used by the Ebola and Marburg viruses. Most of the genes that were identified affect different aspects of lysosome function, suggesting that filoviruses exploit this organelle in a manner distinct from other viruses. By uncovering unanticipated roles for these cellular genes and their products in EboV and MarV entry into host cells, the present work opens new avenues for sorely needed anti-filovirus therapeutics.

REFERENCES

1 Feldmann, H and Geisbert, TW, Ebola haemorrhagic fever. *Lancet* (2010).

2 Lee, JE and Saphire, EO, Ebolavirus glycoprotein structure and mechanism of entry. *Future Virol.* 4 (6), 621 (2009).

3 Chandran, Kartik et al., Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. *Science.* 308 (5728), 1643 (2005).

4 Kuhn, Jens H. et al., Conserved receptor-binding domains of Lake Victoria marburgvirus and Zaire ebolavirus bind a common receptor. *The Journal of biological chemistry* 281 (23), 15951 (2006).

5 Schornberg, Kathryn et al., Role of endosomal cathepsins in entry mediated by the Ebola virus glycoprotein. *Journal of virology* 80 (8), 4174 (2006).

6 Nickerson, Daniel P., Brett, Christopher L., and Merz, Alexey J., Vps-C complexes: gatekeepers of endolysosomal traffic. *Current opinion in cell biology* 21 (4), 543 (2009).

7 Naureckiene, S. et al., Identification of HE1 as the second gene of Niemann-Pick C disease. *Science (New York, N Y)* 290 (5500), 2298 (2000).

[8] Carstea, E. D. et al., Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. *Science (New York, N Y)* 277 (5323), 228 (1997).

[9] Carette et al, JE (submitted).

[10] Carette, Jan E. et al., Haploid genetic screens in human cells identify host factors used by pathogens. *Science (New York, N Y)* 326 (5957), 1231 (2009).

[11] Wong, Anthony C. et al., A forward genetic strategy reveals destabilizing mutations in the Ebolavirus glycoprotein that alter its protease dependence during cell entry. *Journal of virology* 84 (1), 163 (2010).

[12] Carette, Jan E. et al., Generation of iPSCs from cultured human malignant cells. *Blood* 115 (20), 4039 (2010).

[13] Takahashi, Kazutoshi et al., Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131 (5), 861 (2007).

[14] Sbrissa, Diego et al., Core protein machinery for mammalian phosphatidylinositol 3,5-bisphosphate synthesis and turnover that regulates the progression of endosomal transport. Novel Sac phosphatase joins the ArPIKfyve-PIKfyve complex. *The Journal of biological chemistry* 282 (33), 23878 (2007); Ikonomov, Ognian C. et al., PIKfyve controls fluid phase endocytosis but not recycling/degradation of endocytosed receptors or sorting of procathepsin D by regulating multivesicular body morphogenesis. *Molecular biology of the cell* 14 (11), 4581 (2003).

[15] Dell'Angelica, Esteban C., The building BLOC(k)s of lysosomes and related organelles. *Current opinion in cell biology* 16 (4), 458 (2004).

[16] Tiede, Stephan et al., Mucolipidosis II is caused by mutations in GNPTA encoding the alpha/beta GlcNAc-1-phosphotransferase. *Nature medicine* 11 (10), 1109 (2005).

[17] Perez, Mar et al., Generation and characterization of a recombinant vesicular stomatitis virus expressing the glycoprotein of Borna disease virus. *Journal of virology* 81 (11), 5527 (2007).

[18] Sieczkarski, Sara B. and Whittaker, Gary R., Differential requirements of Rab5 and Rab7 for endocytosis of influenza and other enveloped viruses. *Traffic (Copenhagen, Denmark)* 4 (5), 333 (2003).

[19] Pentchev, P. G. et al., A defect in cholesterol esterification in Niemann-Pick disease (type C) patients. *Proceedings of the National Academy of Sciences of the United States of America* 82 (23), 8247 (1985); Blanchette-Mackie, E. J. et al., Type-C Niemann-Pick disease: low density lipoprotein uptake is associated with premature

cholesterol accumulation in the Golgi complex and excessive cholesterol storage in lysosomes. *Proceedings of the National Academy of Sciences of the United States of America* 85 (21), 8022 (1988).

20 Wool-Lewis, R. J. and Bates, P., Characterization of Ebola virus entry by using pseudotyped viruses: identification of receptor-deficient cell lines. *Journal of virology* 72 (4), 3155 (1998); Takada, A. et al., A system for functional analysis of Ebola virus glycoprotein. *Proceedings of the National Academy of Sciences of the United States of America* 94 (26), 14764 (1997).

21 Cruz, J. C., Sugii, S., Yu, C., and Chang, T. Y., Role of Niemann-Pick type C1 protein in intracellular trafficking of low density lipoprotein-derived cholesterol. *The Journal of biological chemistry* 275 (6), 4013 (2000).

22 Liscum, L. and Faust, J. R., The intracellular transport of low density lipoprotein-derived cholesterol is inhibited in Chinese hamster ovary cells cultured with 3-beta-[2-(diethylamino)ethoxy]androst-5-en-17-one. *The Journal of biological chemistry* 264 (20), 11796 (1989).

23 Rodriguez-Lafrasse, C. et al., Abnormal cholesterol metabolism in imipramine-treated fibroblast cultures. Similarities with Niemann-Pick type C disease. *Biochimica et biophysica acta* 1043 (2), 123 (1990).

24 Lange, Y., Ye, J., Rigney, M., and Steck, T., Cholesterol movement in Niemann-Pick type C cells and in cells treated with amphiphiles. *The Journal of biological chemistry* 275 (23), 17468 (2000); Liu, Ronghua, Lu, Peihua, Chu, Joseph W. K., and Sharom, Frances J., Characterization of fluorescent sterol binding to purified human NPC1. *The Journal of biological chemistry* 284 (3), 1840 (2009); Cenedella, Richard J., Cholesterol synthesis inhibitor U18666A and the role of sterol metabolism and trafficking in numerous pathophysiological processes. *Lipids* 44 (6), 477 (2009).

25 Lloyd-Evans, Emyr et al., Niemann-Pick disease type C1 is a sphingosine storage disease that causes deregulation of lysosomal calcium. *Nature medicine* 14 (11), 1247 (2008).

26 Marzi, Andrea et al., DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus. *Journal of virology* 78 (21), 12090 (2004).

[27] Nanbo, Asuka et al., Ebolavirus is internalized into host cells via macropinocytosis in a viral glycoprotein-dependent manner. *PLoS pathogens* 6 (9) (2010); Saeed, Mohammad F., Kolokoltsov, Andrey A., Albrecht, Thomas, and Davey, Robert A., Cellular entry of ebola virus involves uptake by a macropinocytosis-like mechanism and subsequent trafficking through early and late endosomes. *PLoS pathogens* 6 (9) (2010).

[28] Weissenhorn, W. et al., The central structural feature of the membrane fusion protein subunit from the Ebola virus glycoprotein is a long triple-stranded coiled coil. *Proceedings of the National Academy of Sciences of the United States of America* 95 (11), 6032 (1998).

[29] Whelan, S. P., Ball, L. A., Barr, J. N., and Wertz, G. T., Efficient recovery of infectious vesicular stomatitis virus entirely from cDNA clones. *Proceedings of the National Academy of Sciences of the United States of America* 92 (18), 8388 (1995).

[30] Morgenstern, J. P. and Land, H., Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic acids research* 18 (12), 3587 (1990).

[31] Pentchev, P. G. et al., The cholesterol storage disorder of the mutant BALB/c mouse. A primary genetic lesion closely linked to defective esterification of exogenously derived cholesterol and its relationship to human type C Niemann-Pick disease. *The Journal of biological chemistry* 261 (6), 2772 (1986).

[32] Ebert, Daniel H., Deussing, Jan, Peters, Christoph, and Dermody, Terence S., Cathepsin L and cathepsin B mediate reovirus disassembly in murine fibroblast cells. *The Journal of biological chemistry* 277 (27), 24609 (2002).

[33] Blum, Galia et al., Dynamic imaging of protease activity with fluorescently quenched activity-based probes. *Nature chemical biology* 1 (4), 203 (2005).

[34] Cureton, David K. et al., Vesicular stomatitis virus enters cells through vesicles incompletely coated with clathrin that depend upon actin for internalization. *PLoS pathogens* 5 (4), e1000394 (2009).

[35] Ehrlich, Marcelo et al., Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell* 118 (5), 591 (2004).

[36] Lefrancois, L. and Lyles, D. S., The interaction of antibody with the major surface glycoprotein of vesicular stomatitis virus. I. Analysis of neutralizing epitopes with monoclonal antibodies. *Virology* 121 (1), 157 (1982).

[37] Dube D, Brecher MB, Delos SE, Rose SC, Park EW, Schornberg KL, Kuhn JH, White JM. The primed ebolavirus glycoprotein (19-kilodalton GP1,2): sequence and residues critical for host cell binding. J Virol. 2009 Apr;83(7):2883-91. Epub 2009 Jan 14.

[38] Davies JP, Ioannou YA. Topological analysis of Niemann-Pick C1 protein reveals that the membrane orientation of the putative sterol-sensing domain is identical to those of 3-hydroxy-3-methylglutaryl-CoA reductase and sterol regulatory element binding protein cleavage-activating protein. *J Biol Chem.* 2000 Aug 11;275(32):24367-74.

[39] Kwon HJ, Abi-Mosleh L, Wang ML, Deisenhofer J, Goldstein JL, Brown MS, Infante RE. Structure of N-terminal domain of NPC1 reveals distinct subdomains for binding and transfer of cholesterol. *Cell.* 2009 Jun 26;137(7):1213-24.

[40] Deffieu, M. and Pfeffer, S.R. Niemann-Pick type C1 function requires lumenal domain residues that mediate cholesterol-dependent NPC2 binding. *Proc Natl Acad Sci USA.* 2011. 108, 18932-36.

What is claimed is:

1. A method for treating a subject infected with a filovirus or for preventing an infection with a filovirus in a subject at risk for infection with a filovirus comprising administering to the subject an agent that inhibits one or more of Niemann-Pick C1 (NPC1), VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 in an amount effective to treat and/or prevent infection with the filovirus.

2. The method of Claim 1 for treating a subject infected with a filovirus.

3. The method of Claim 1 for preventing an infection with a filovirus in a subject who is at risk for infection with a filovirus.

4. The method of Claim 3, wherein the subject is exposed to a filovirus as the result of bioterrorism or biological warfare.

5. The method of any of Claims 1-4, wherein the agent inhibits NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 protein activity.

6. The method of any of Claims 1-4, wherein the agent inhibits nucleic acid that encodes NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 protein.

7. The method of any of Claims 1-6, wherein the agent targets domain C of NPC1 or nucleic acid encoding domain C of NPC1.

8. The method of any of Claims 1-7, wherein the filovirus is an ebolavirus.

9. The method of any of Claims 1-7, wherein the filovirus is a marburgvirus.

10. A method for screening for an agent that treats and/or prevents infection of a subject with a filovirus, the method comprising determining whether or not the agent inhibits one or more of Niemann-Pick C1 (NPC1), VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4, wherein an agent that inhibits NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 is a candidate for treating and/or preventing an infection with a filovirus and wherein an agent that does not inhibit NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4 is not a candidate for treating and/or preventing an infection with a filovirus.

11. The method of Claim 10, wherein the agent targets domain C of NPC1 or nucleic acid encoding domain C of NPC1.

12. The method of Claim 10 or 11 for screening for an agent that treats an infection of a subject with a filovirus.

13. The method of Claim 10 or 11 for screening for an agent that prevents an infection of a subject with a filovirus.

14. The method of any of Claims 10-13, comprising measuring cholesterol transport, where a decrease in cholesterol transport in the presence of the agent indicates that the agent inhibits NPC1.

15. The method of Claim 14, wherein the method is carried out using a cell line that expresses NPC1.

16. The method of any of Claims 10-13, comprising measuring binding between NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4, and the filovirus or a filovirus glycoprotein (GP), wherein a decrease in binding in the presence of the agent indicates that the agent inhibits NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 or FIG4.

17. The method of Claim 16, wherein the method is carried out using an enzyme-linked-immunosorbent assay (ELISA).

18. The method of any of Claims 10-13, comprising measuring filovirus infection in tissue culture, wherein a reduction in filovirus infection in the presence of the agent indicates that the agent inhibits NPC1, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23 and/or FIG4.

19. The method of any of Claims 10-18, wherein the filovirus is an ebolavirus.

20. The method of any of Claims 10-18, wherein the filovirus is a marburgvirus.

21. An agent for treating and/or preventing infection of a subject with a filovirus identified by the method of any of Claims 10-20.

22. A pharmaceutical composition for treating and/or preventing infection of a subject with a filovirus comprising the agent of Claim 21 and a pharmaceutically acceptable carrier.

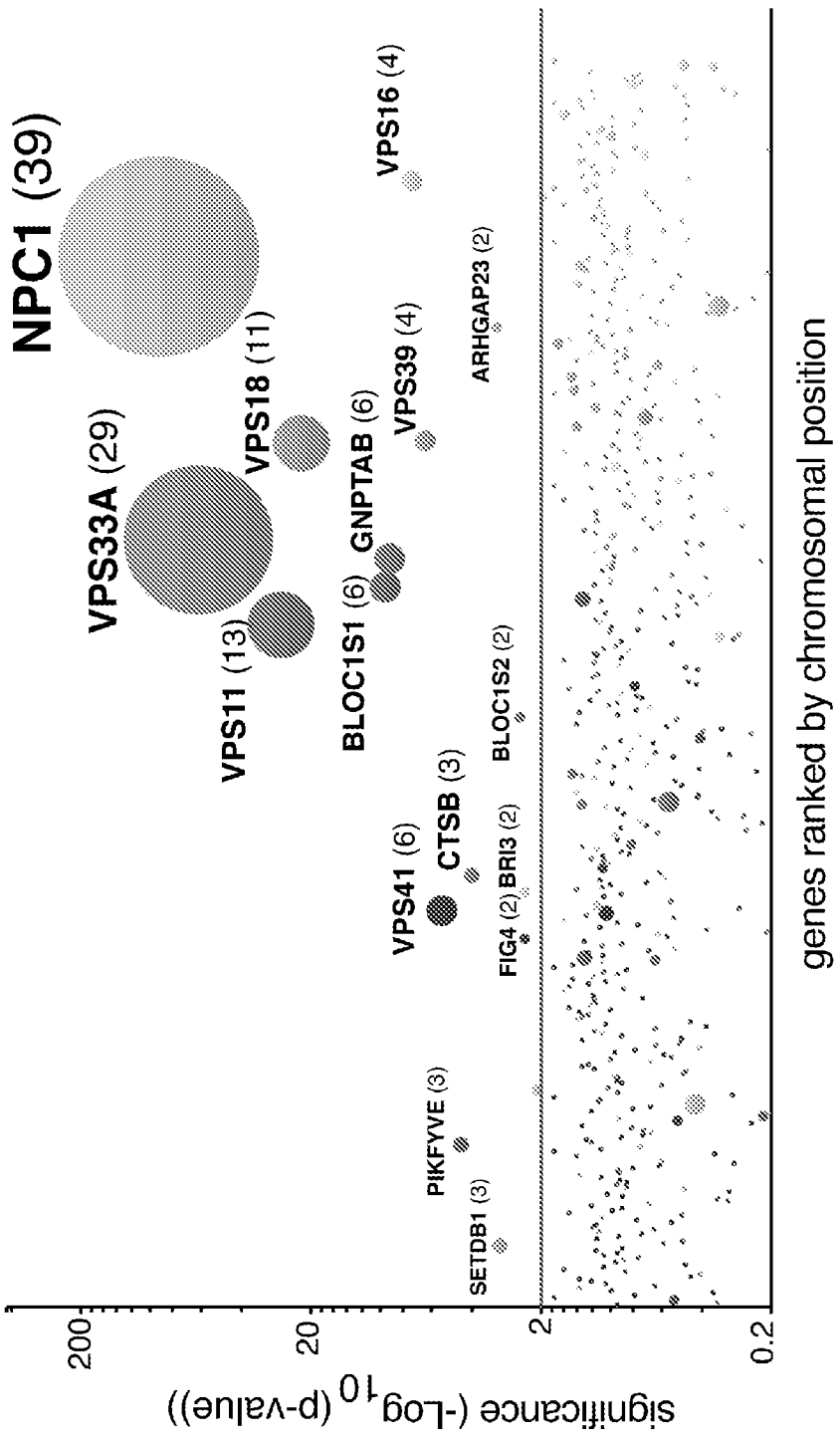
NPC1 (39)
VPS33A (29)
VPS18 (11)
VPS11 (13)
GNPTAB (6)
BLOC1S1 (6)
VPS39 (4)
VPS16 (4)
ARHGAP23 (2)
BLOC1S2 (2)
CTSB (3)
VPS41 (6)
BRI3 (2)
FIG4 (2)
PIKFYVE (3)
SETDB1 (3)
200
20
2
0.2
significance (-Log10(p-value))
genes ranked by chromosomal position

FIGURE 1A

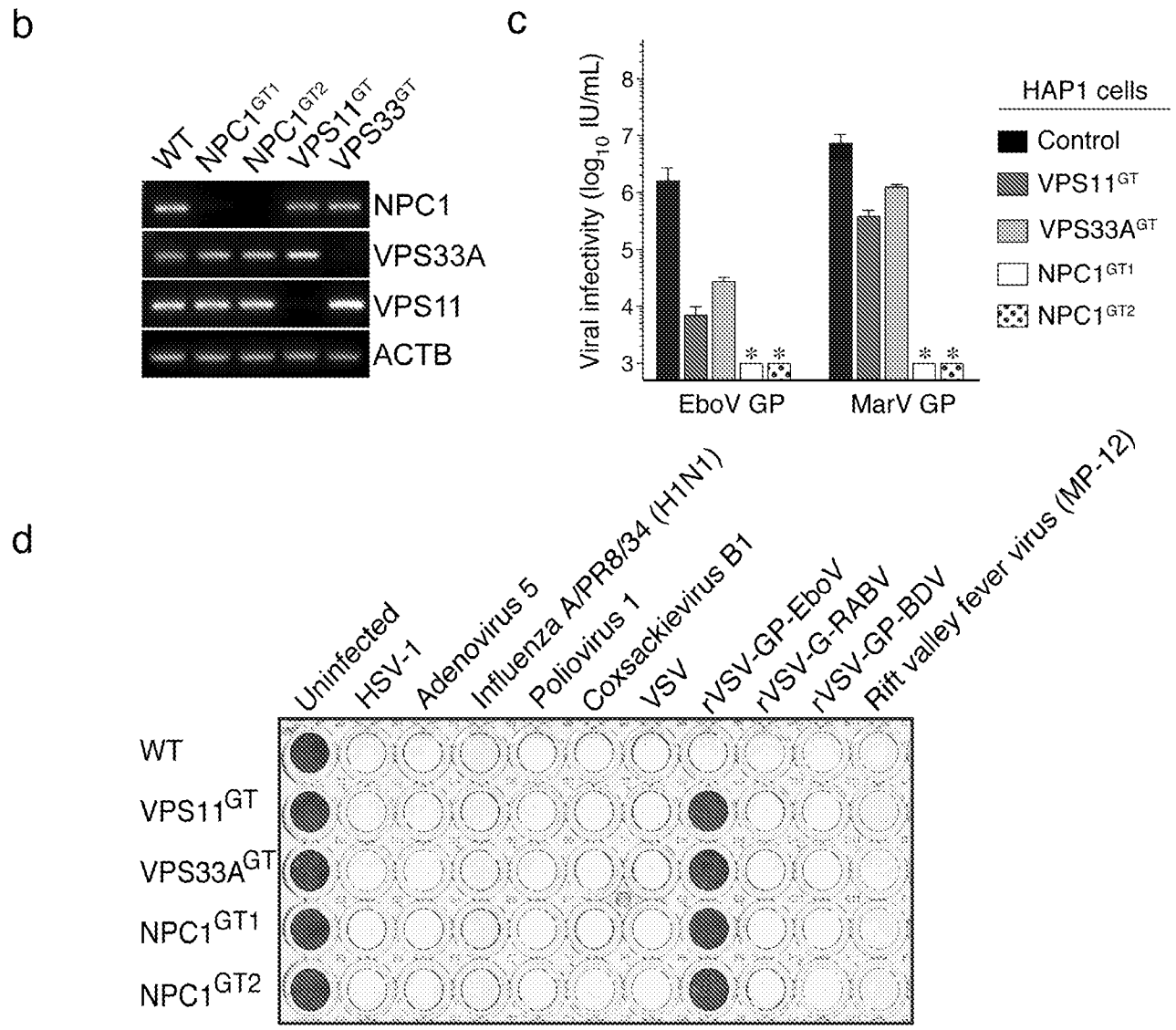
b
WT
NPC1GT1
NPC1GT2
VPS11GT
VPS33GT
NPC1
VPS33A
VPS11
ACTB
c
Viral infectivity (log10 IU/mL)
8
7
6
5
4
3
HAP1 cells
Control
VPS11GT
VPS33AGT
NPC1GT1
NPC1GT2
EboV GP
MarV GP
d
Uninfected
HSV-1
Adenovirus 5
Influenza A/PR8/34 (H1N1)
Poliovirus 1
Coxsackievirus B1
VSV
rVSV-GP-EboV
rVSV-G-RABV
rVSV-GP-BDV
Rift valley fever virus (MP-12)
WT
VPS11GT
VPS33AGT
NPC1GT1
NPC1GT2

FIGURE 1B-1D

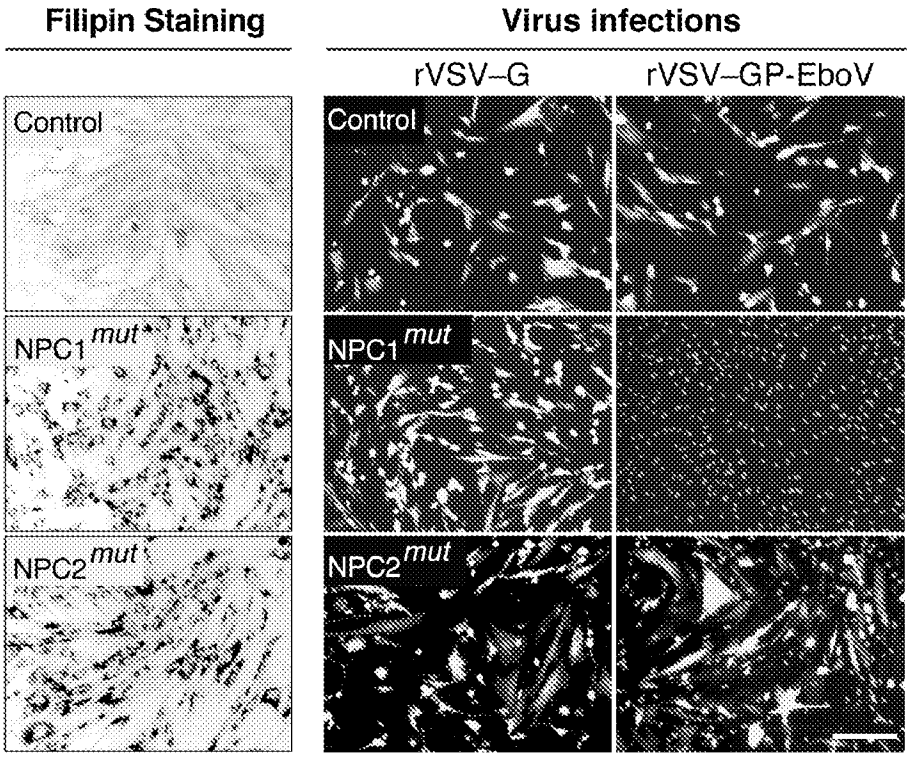
Filipin Staining
Virus infections
rVSV–G
rVSV–GP-EboV
Control
NPC1 mut
NPC2 mut
Control
NPC1 mut
NPC2 mut

FIGURE 2A

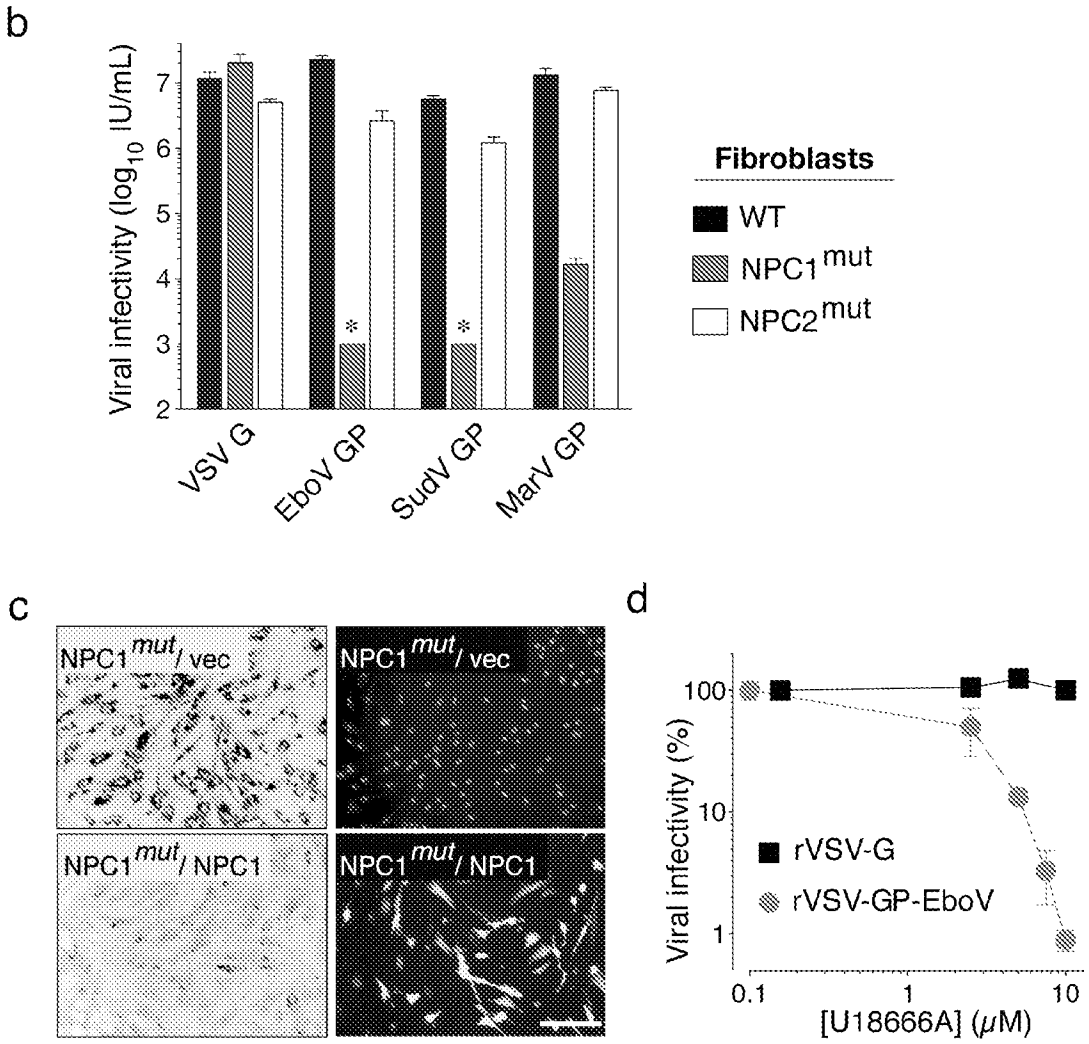
b
Viral infectivity (log_{10} IU/mL)
7
6
5
4
3
2
*
*
VSV G
EboV GP
SudV GP
MarV GP
Fibroblasts
WT
NPC1 mut
NPC2 mut
c
NPC1 mut / vec
NPC1 mut / vec
NPC1 mut / NPC1
NPC1 mut / NPC1
d
Viral infectivity (%)
100
10
1
rVSV-G
rVSV-GP-EboV
0.1
1
10
[U18666A] (μM)

FIGURE 2B-2D

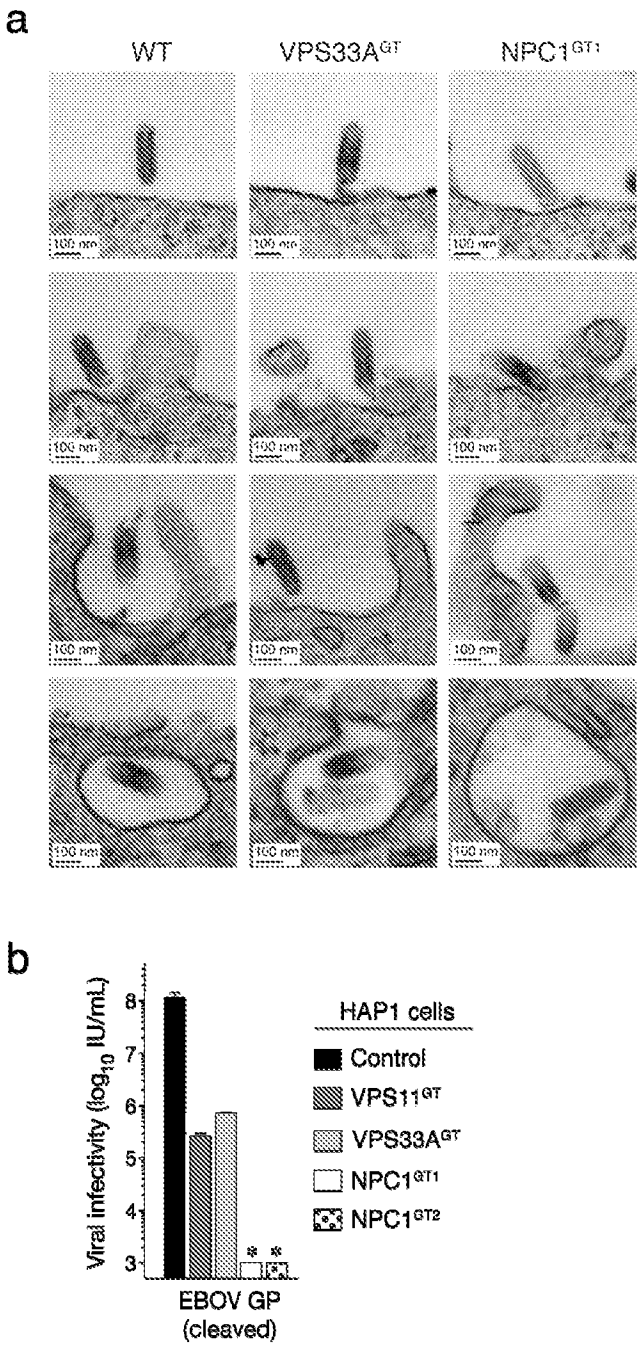
a
WT
VPS33A^GT
NPC1^GT1
100 nm
b
Viral infectivity (log10 IU/mL)
8
7
6
5
4
3
HAP1 cells
Control
VPS11^GT
VPS33A^GT
NPC1^GT1
NPC1^GT2
EBOV GP
(cleaved)

FIGURE 3A-3B

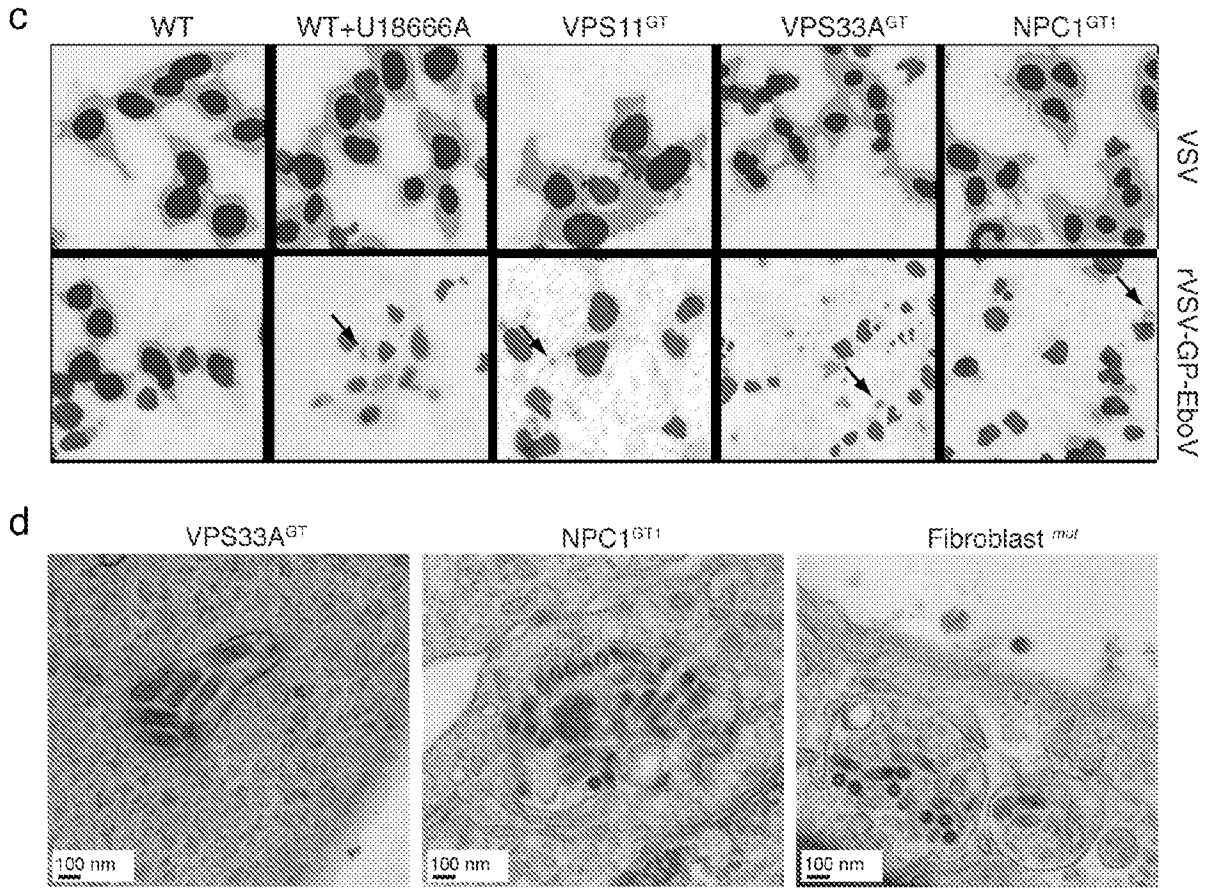
c
WT
WT+U18666A
VPS11GT
VPS33AGT
NPC1GT1
VSV
rVSV-GP-EboV
d
VPS33AGT
NPC1GT1
Fibroblast mut
100 nm
100 nm
100 nm

FIGURE 3C-3D

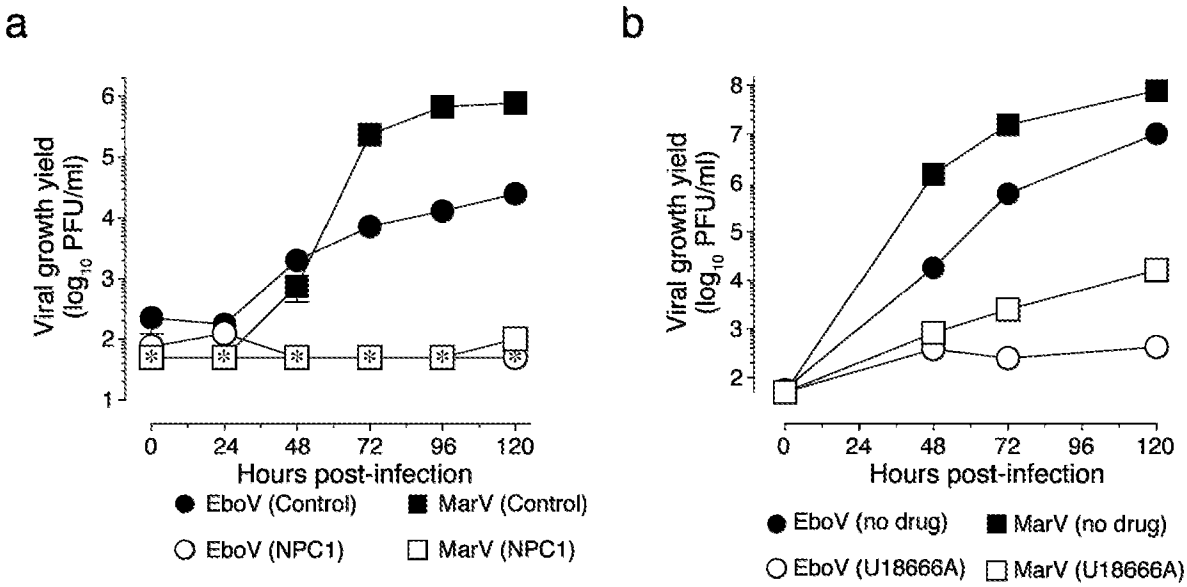
a
Viral growth yield
(log10 PFU/ml)
Hours post-infection
EboV (Control)
MarV (Control)
EboV (NPC1)
MarV (NPC1)
b
Viral growth yield
(log10 PFU/ml)
Hours post-infection
EboV (no drug)
MarV (no drug)
EboV (U18666A)
MarV (U18666A)

FIGURE 4A-4B

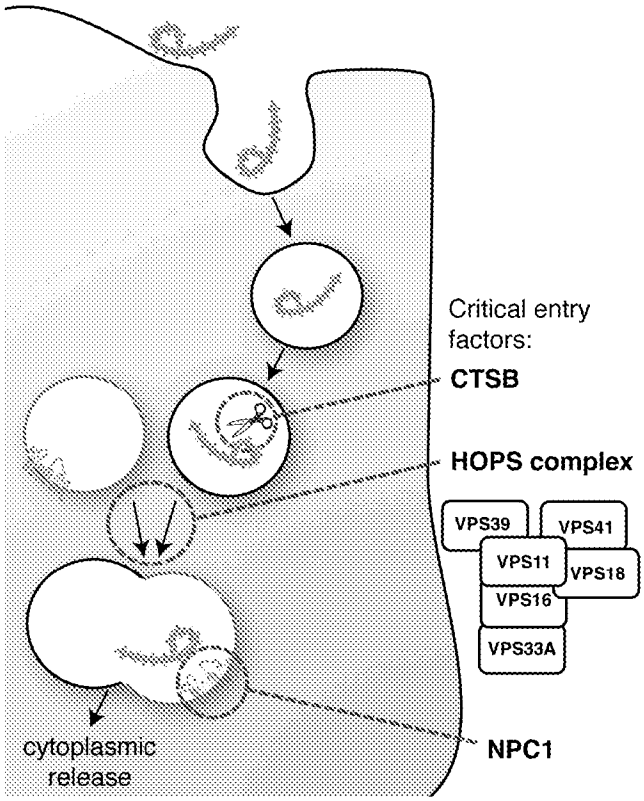
Critical entry factors:
CTSB
HOPS complex
VPS39
VPS41
VPS11
VPS18
VPS16
VPS33A
NPC1
cytoplasmic release

FIGURE 4C

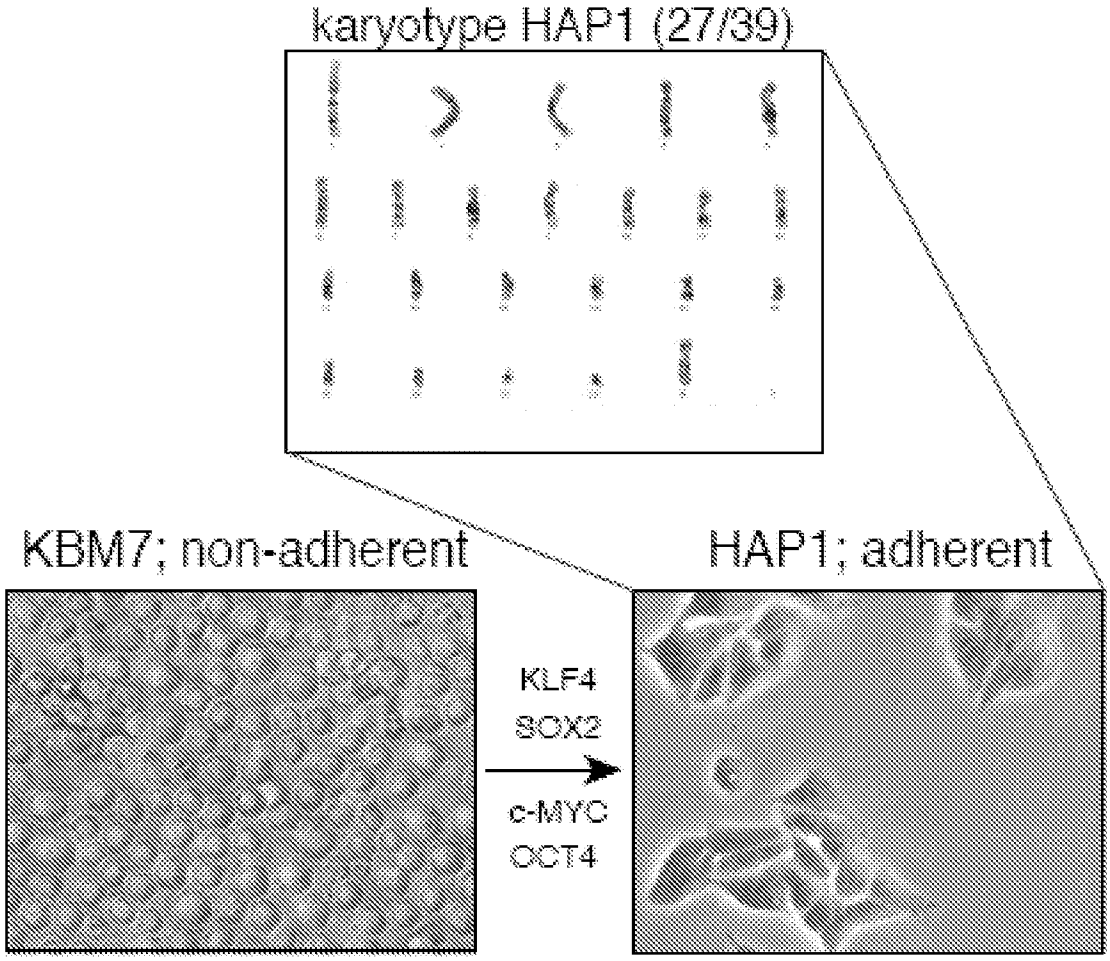
karyotype HAP1 (27/39)
KBM7; non-adherent
HAP1; adherent
KLF4
SOX2
c-MYC
OCT4

FIGURE 5A

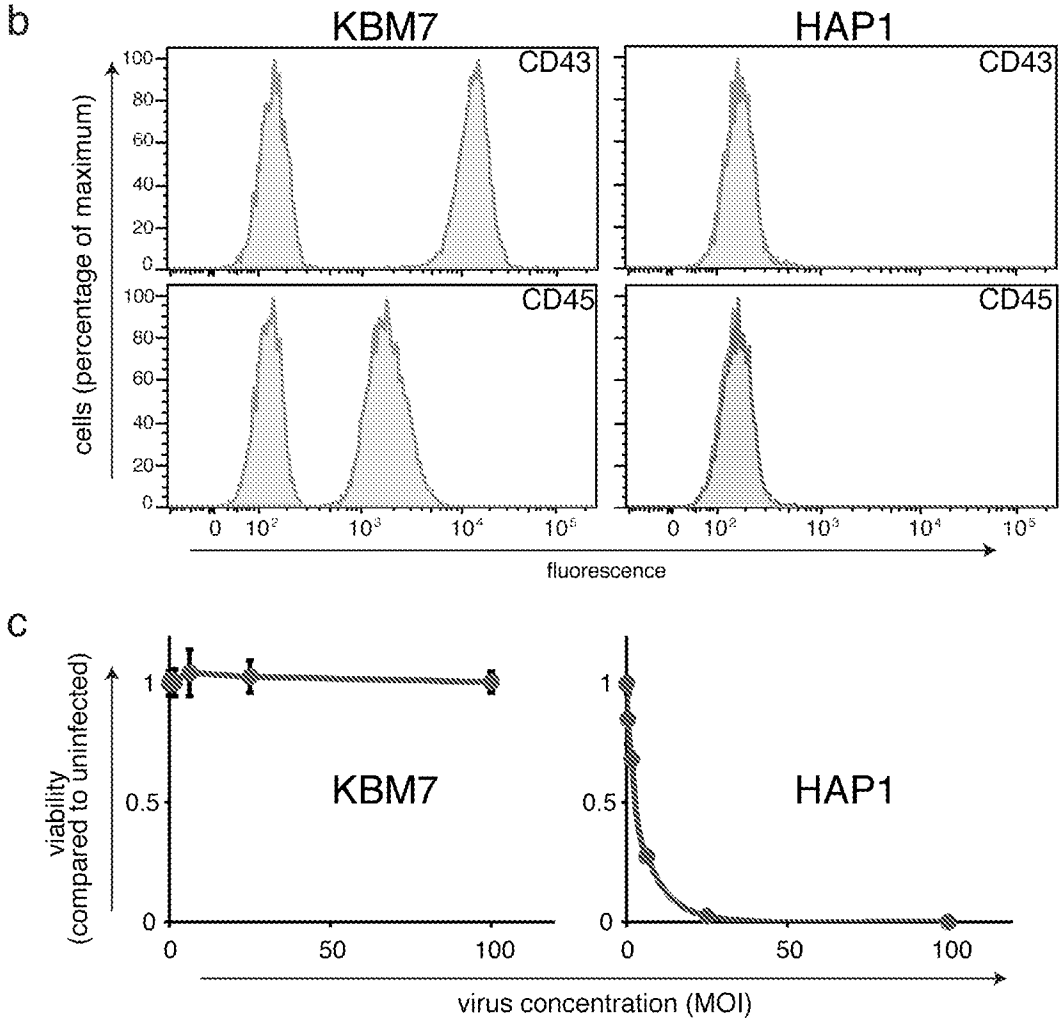
b
KBM7
HAP1
CD43
CD45
cells (percentage of maximum)
fluorescence
c
viability (compared to uninfected)
KBM7
HAP1
virus concentration (MOI)

FIGURE 5B-5C

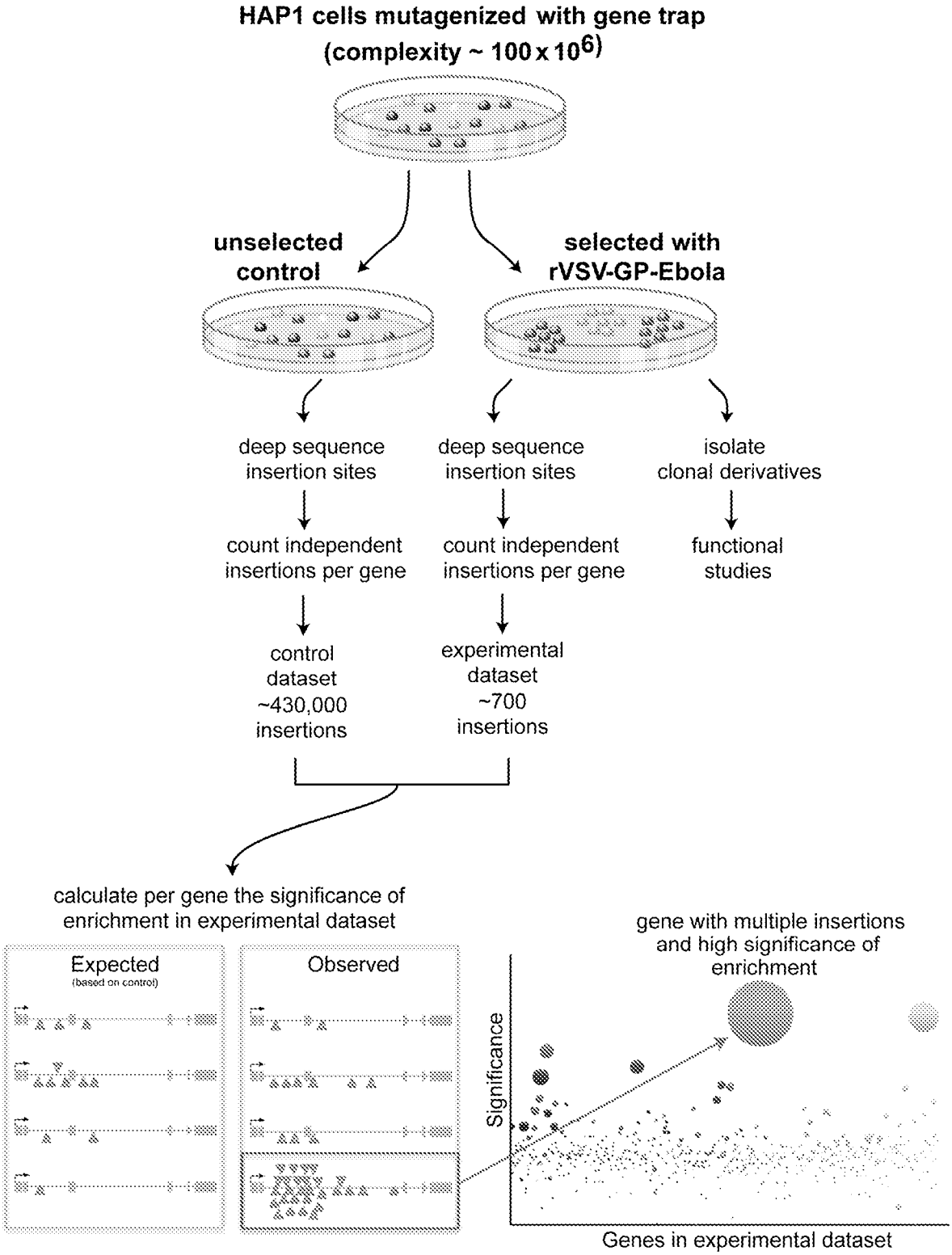
HAP1 cells mutagenized with gene trap
(complexity ~ 100 x 10^6)
unselected control
selected with rVSV-GP-Ebola
deep sequence insertion sites
deep sequence insertion sites
isolate clonal derivatives
count independent insertions per gene
count independent insertions per gene
functional studies
control dataset ~430,000 insertions
experimental dataset ~700 insertions
calculate per gene the significance of enrichment in experimental dataset
Expected
(based on control)
Observed
gene with multiple insertions and high significance of enrichment
Significance
Genes in experimental dataset

FIGURE 6

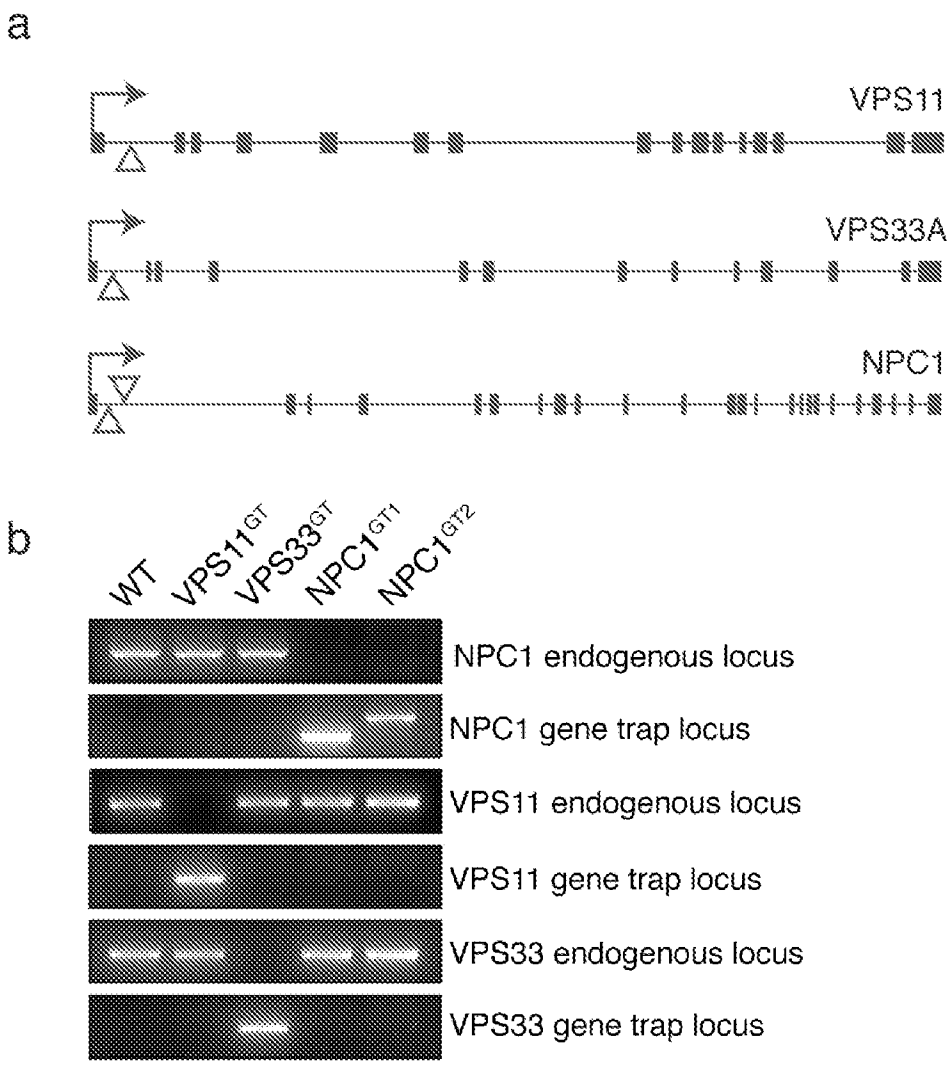
a
VPS11
VPS33A
NPC1
b
WT
VPS11GT
VPS33GT
NPC1GT1
NPC1GT2
NPC1 endogenous locus
NPC1 gene trap locus
VPS11 endogenous locus
VPS11 gene trap locus
VPS33 endogenous locus
VPS33 gene trap locus

FIGURE 7A-7B

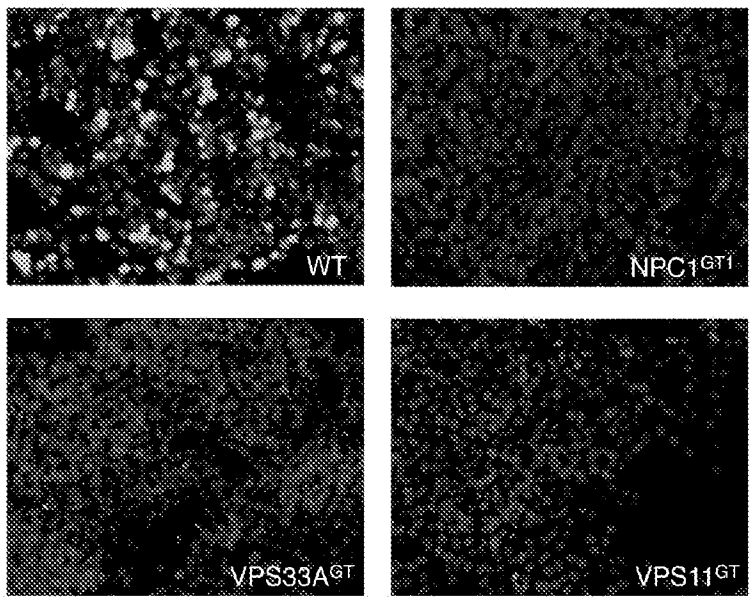

FIGURE 7C

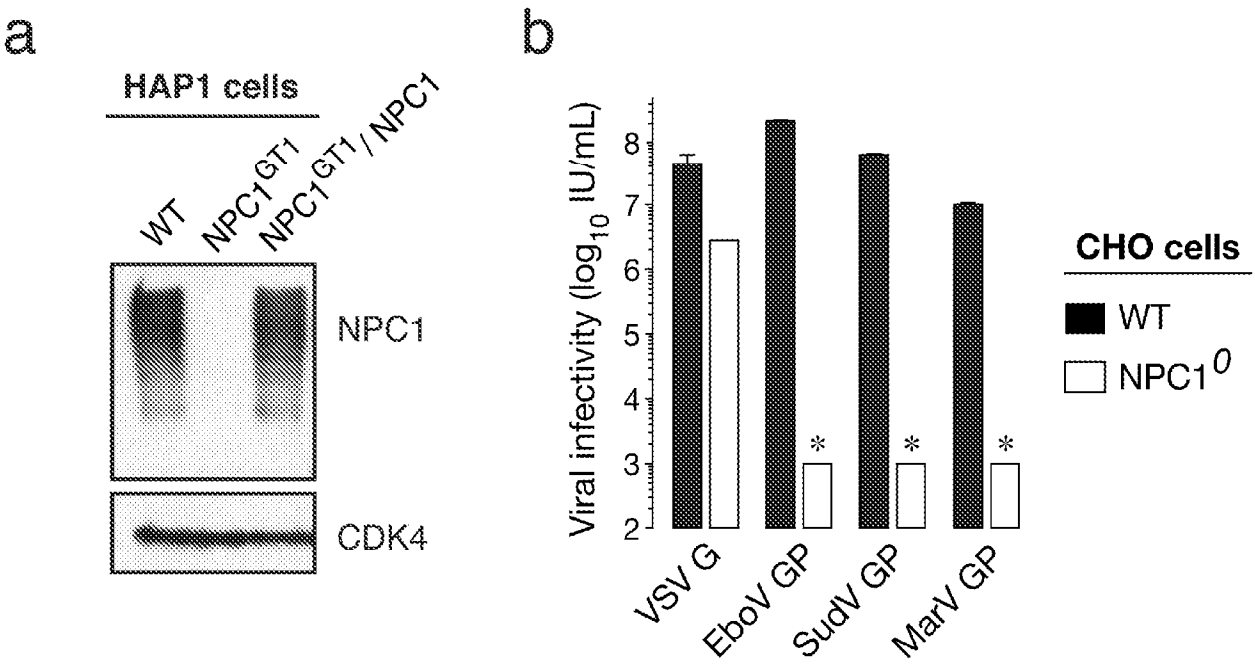
a
HAP1 cells
WT
NPC1GT1
NPC1GT1 / NPC1
NPC1
CDK4
b
Viral infectivity (log10 IU/mL)
2
3
4
5
6
7
8
*
*
*
VSV G
EboV GP
SudV GP
MarV GP
CHO cells
WT
NPC1 0

FIGURE 8A-8B

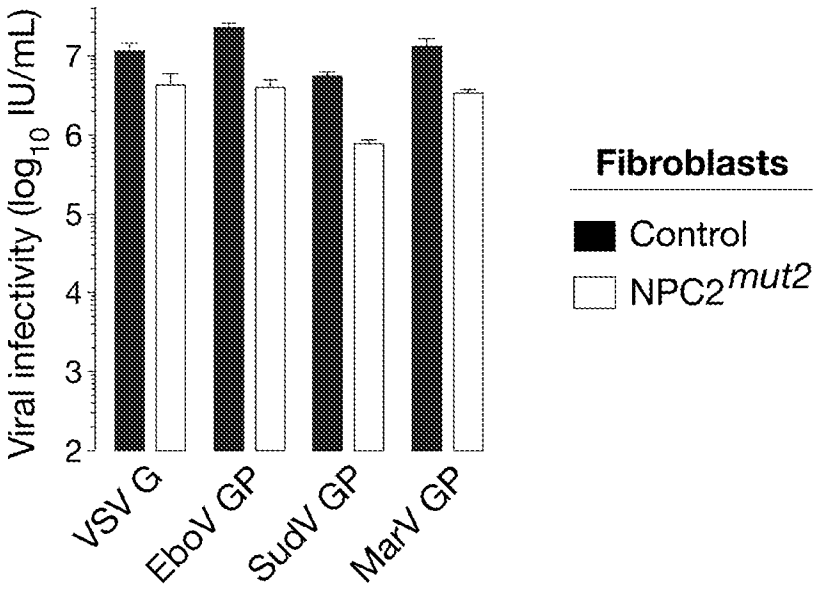
Viral infectivity (log_{10} IU/mL)
7
6
5
4
3
2
VSV G
EboV GP
SudV GP
MarV GP
Fibroblasts
Control
NPC2mut2

FIGURE 9

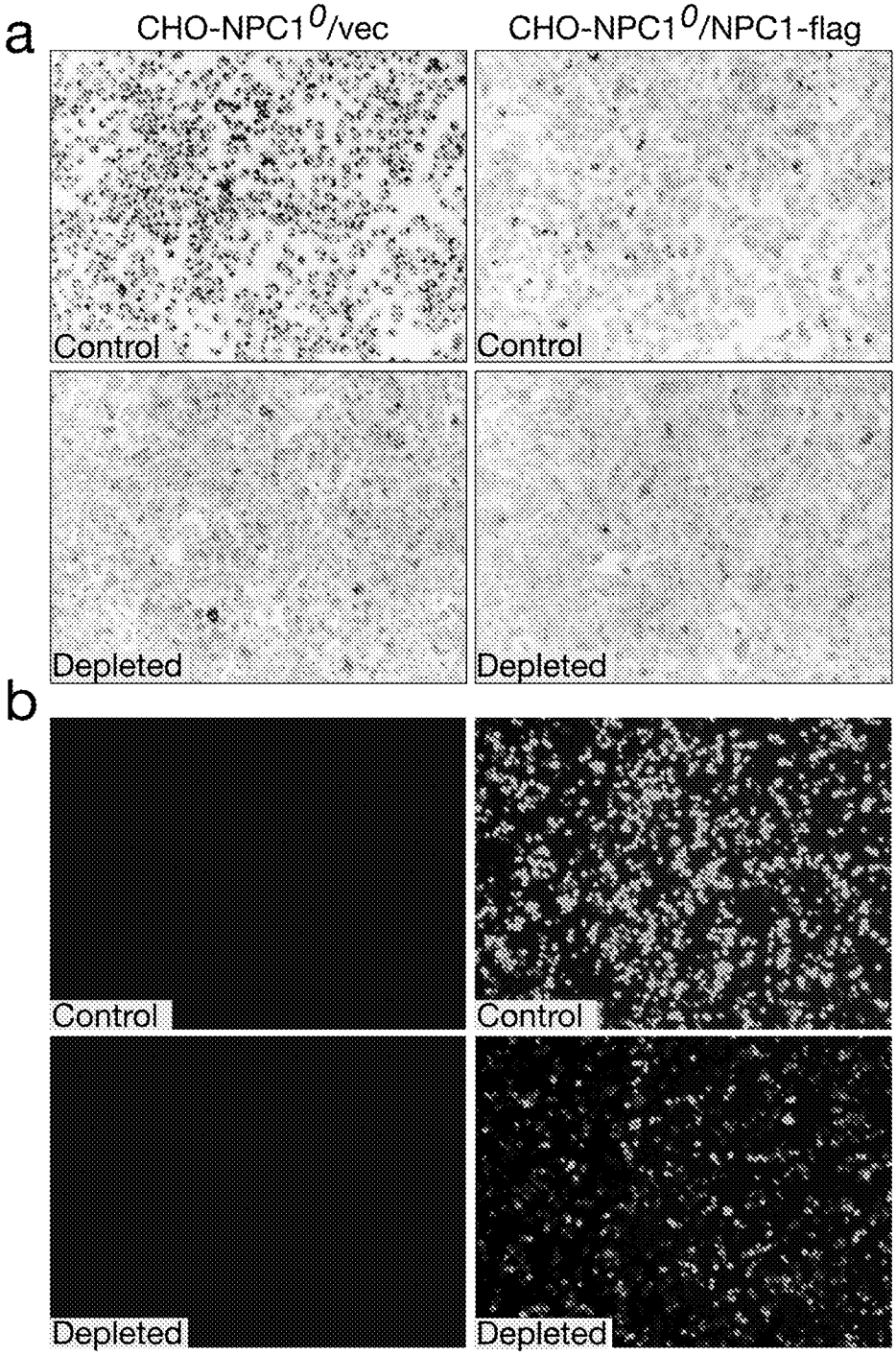

FIGURE 10A-10B

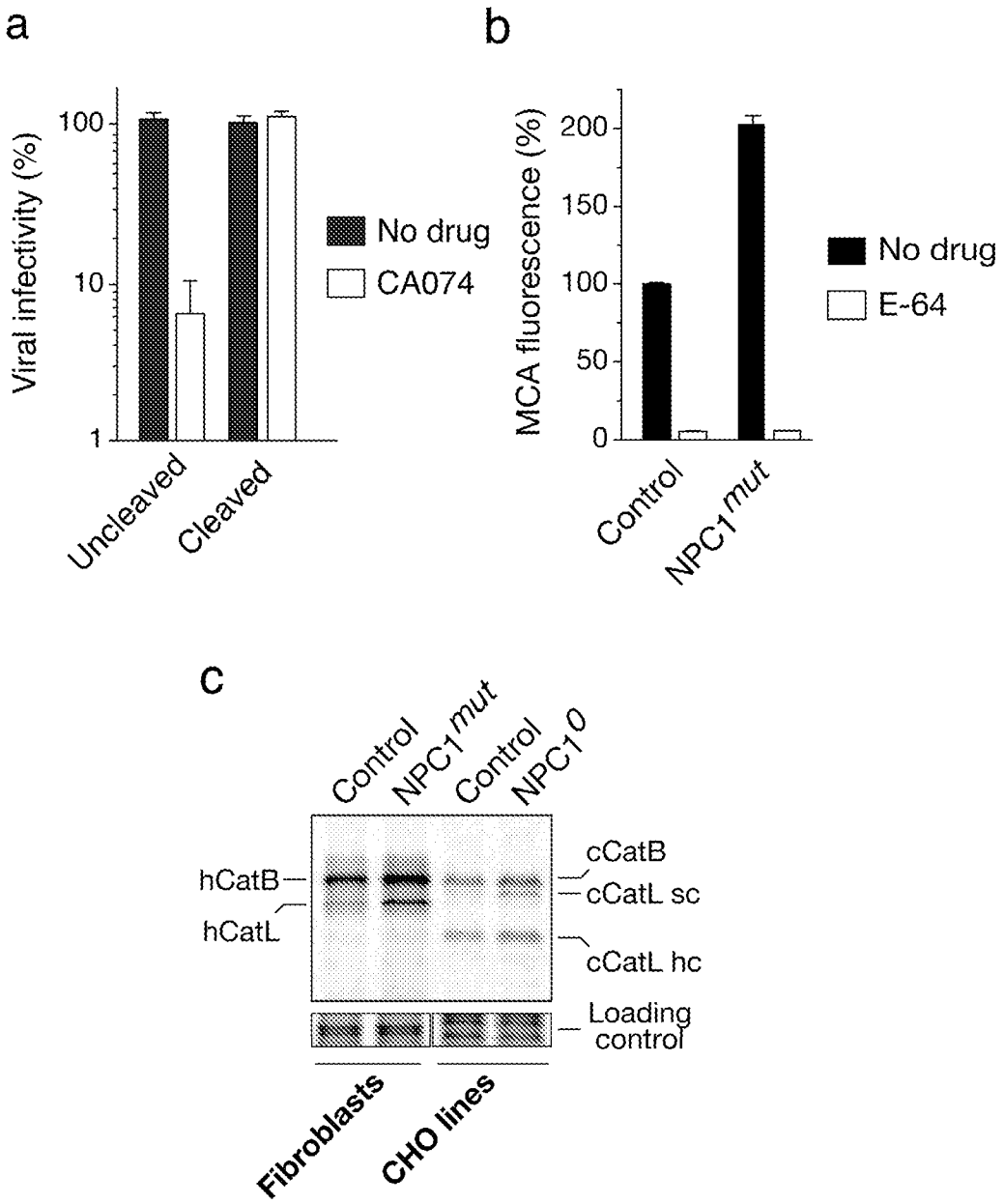

FIGURE 11A-11C

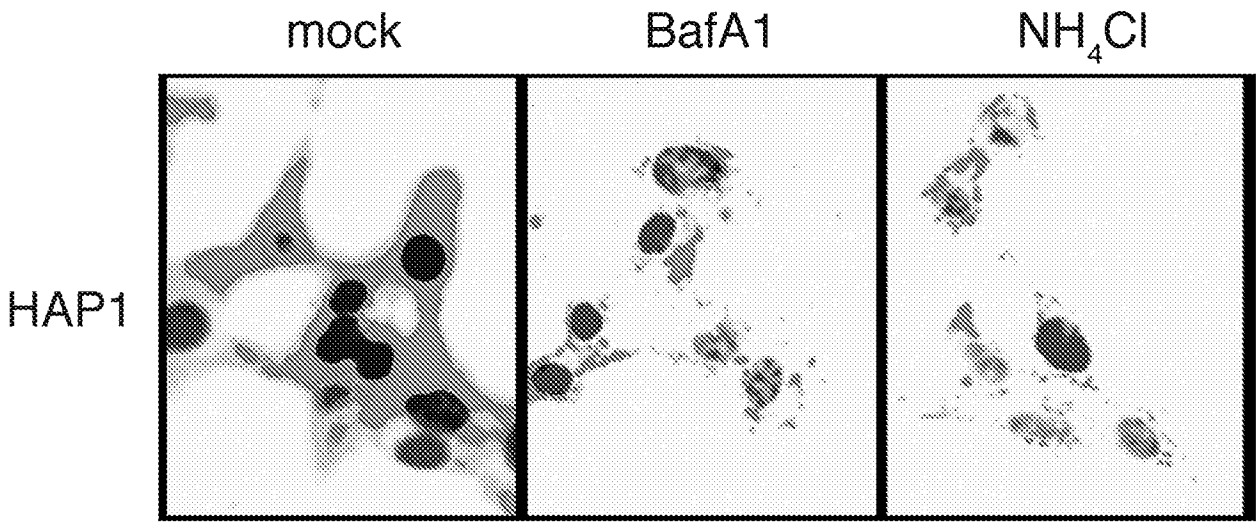

FIGURE 12

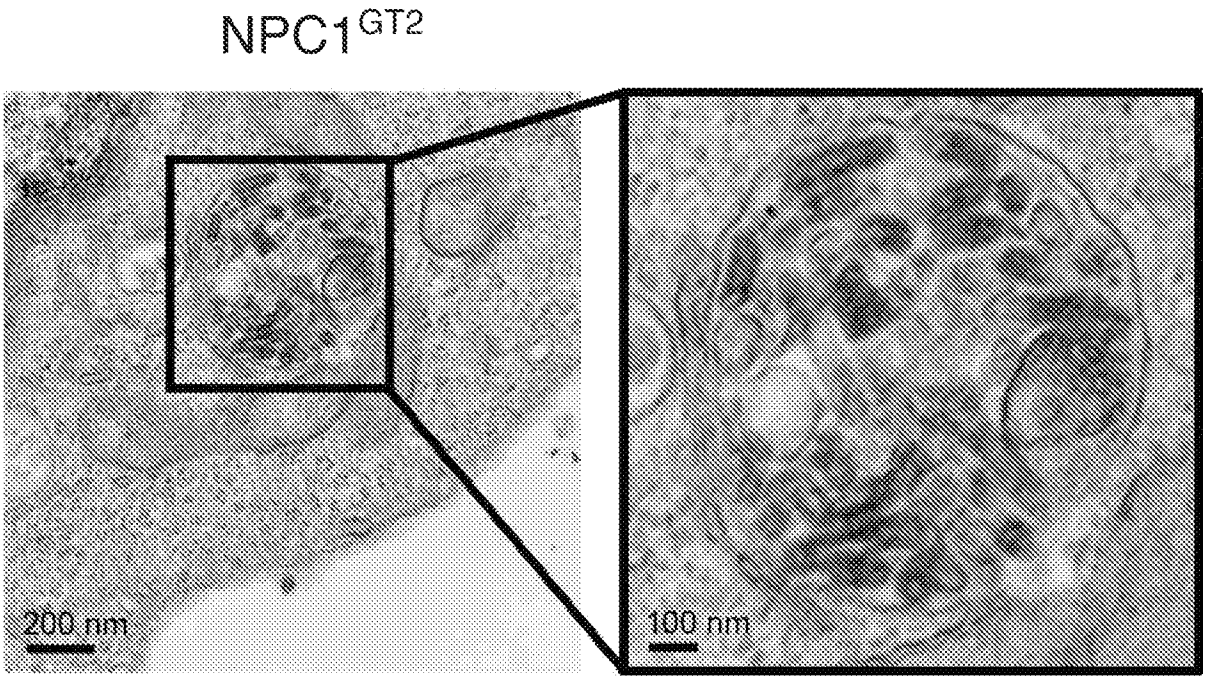

FIGURE 13

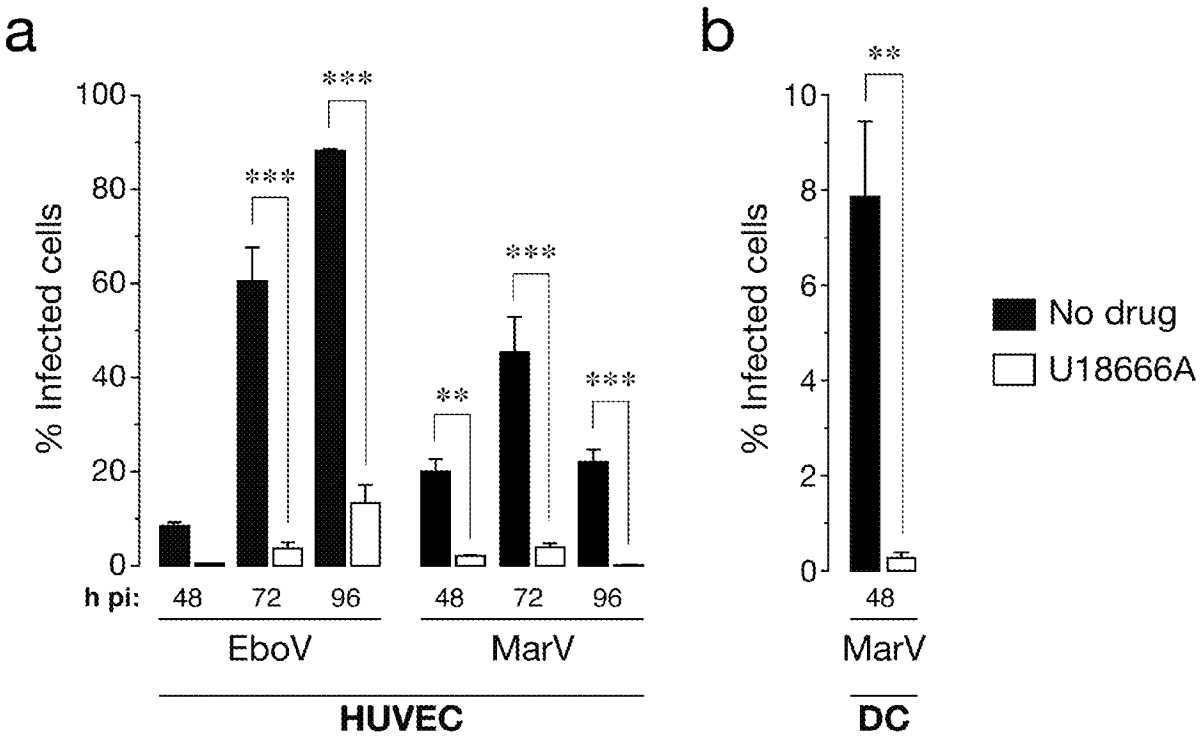

FIGURE 14A-14B

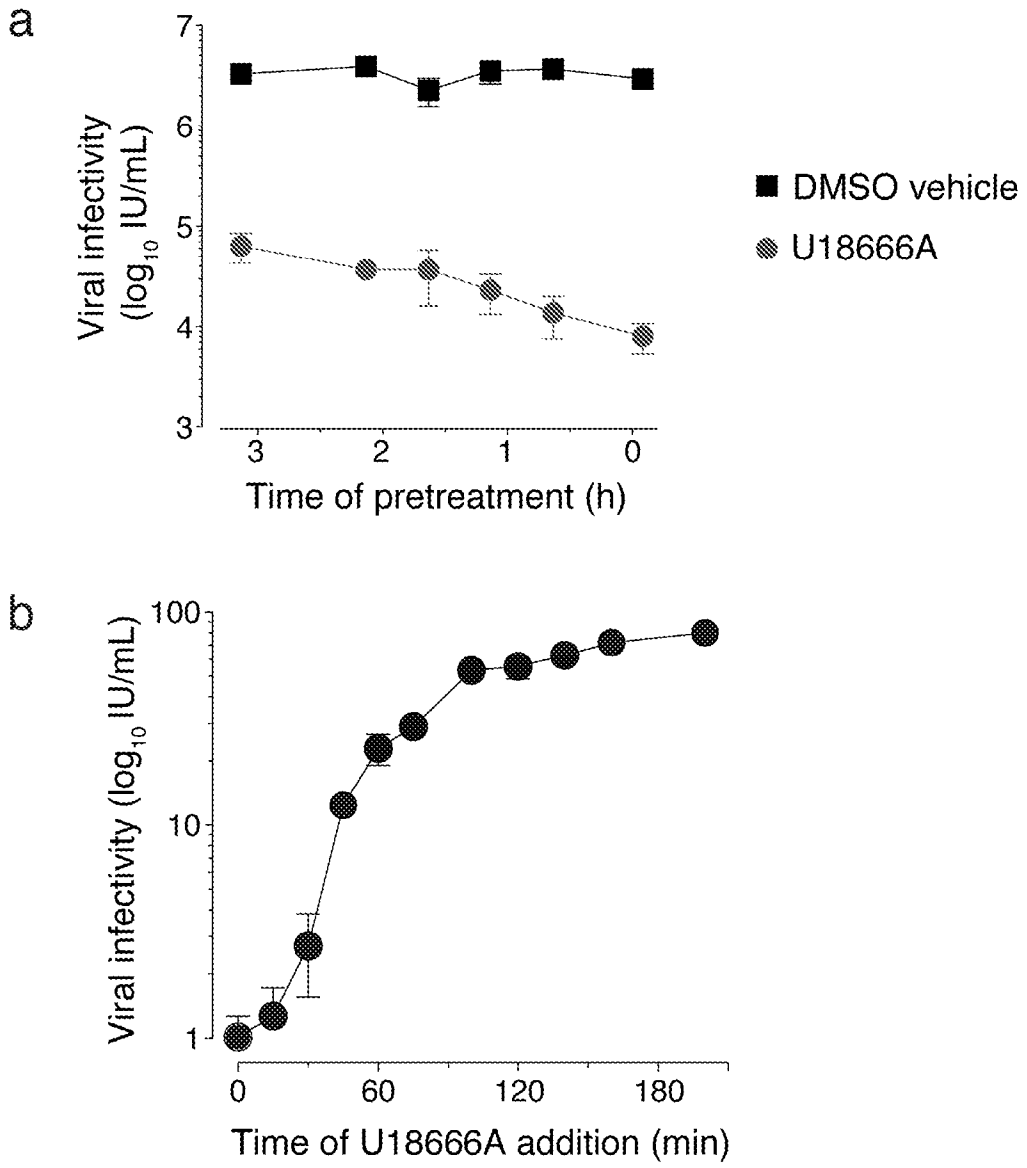

FIGURE 15A-15B

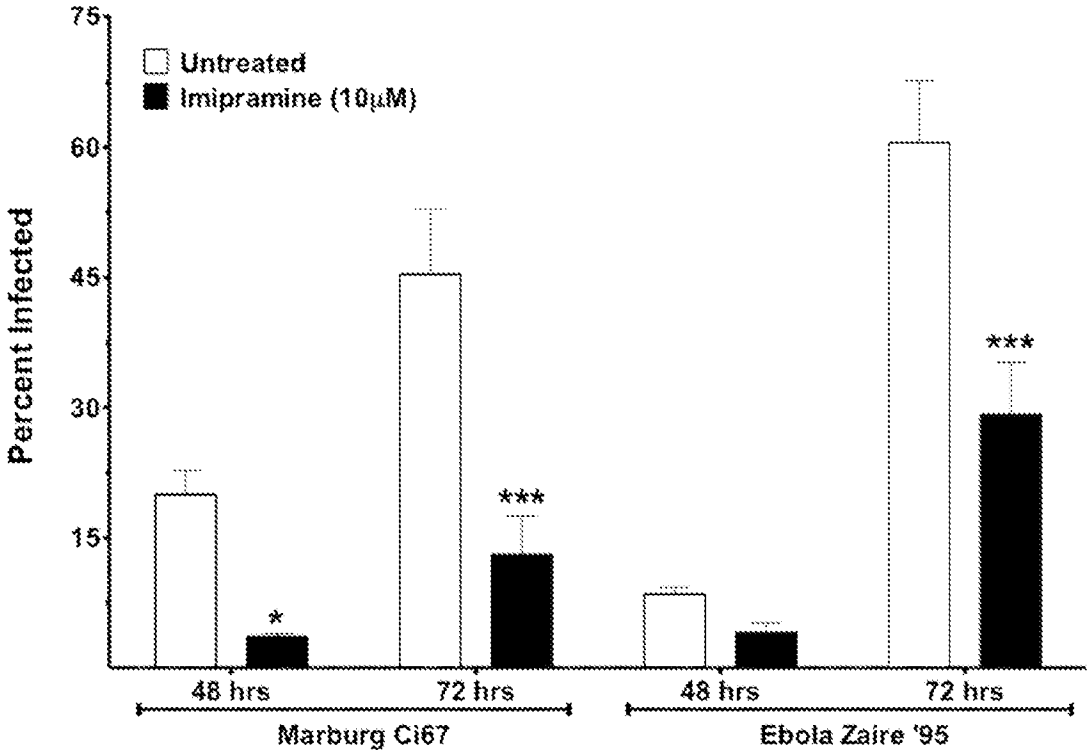
75
60
45
30
15
Percent Infected
Untreated
Imipramine (10μM)
*
***
***
48 hrs
72 hrs
48 hrs
72 hrs
Marburg Ci67
Ebola Zaire '95

FIGURE 16

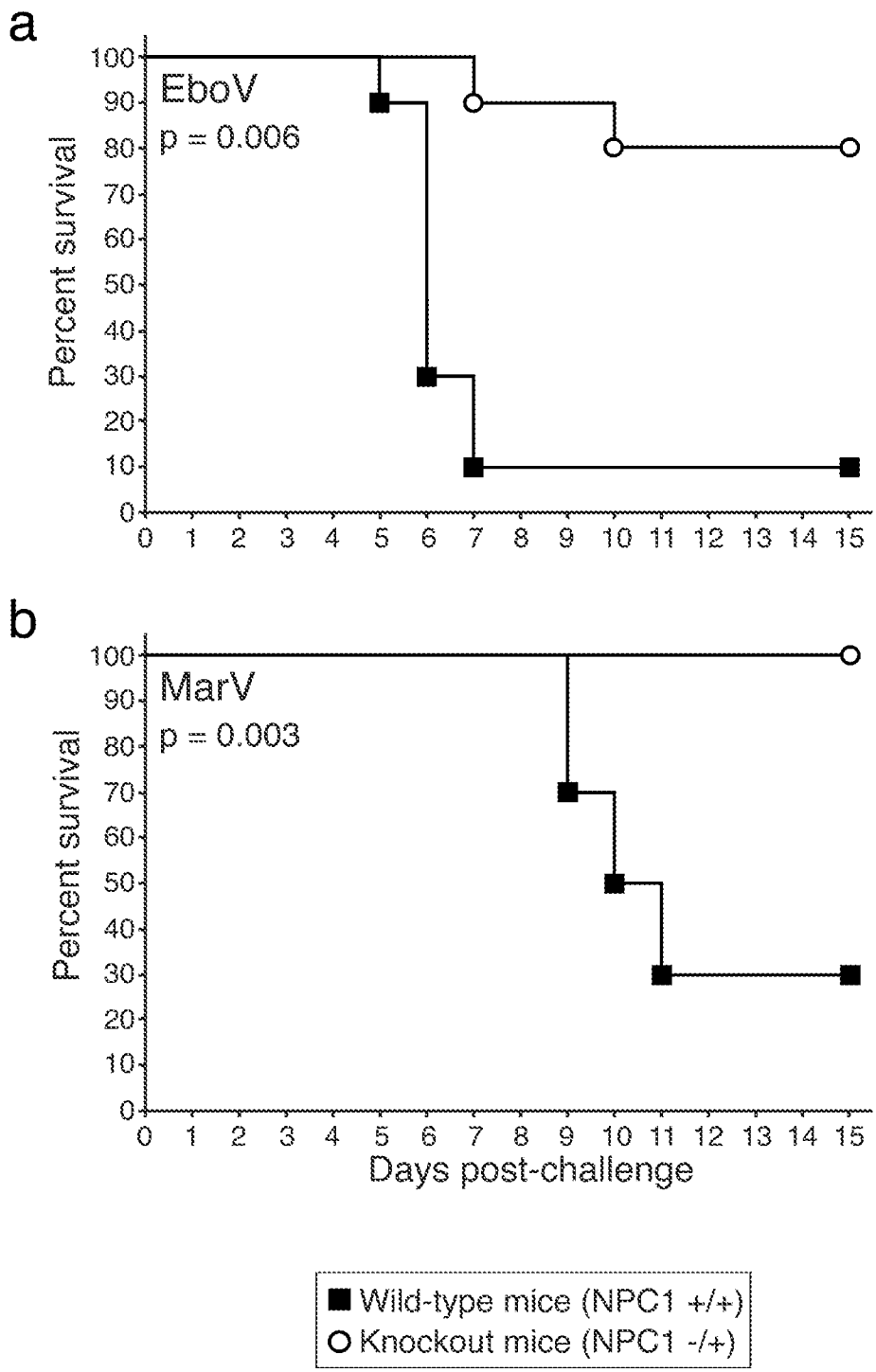

FIGURE 17A-17B

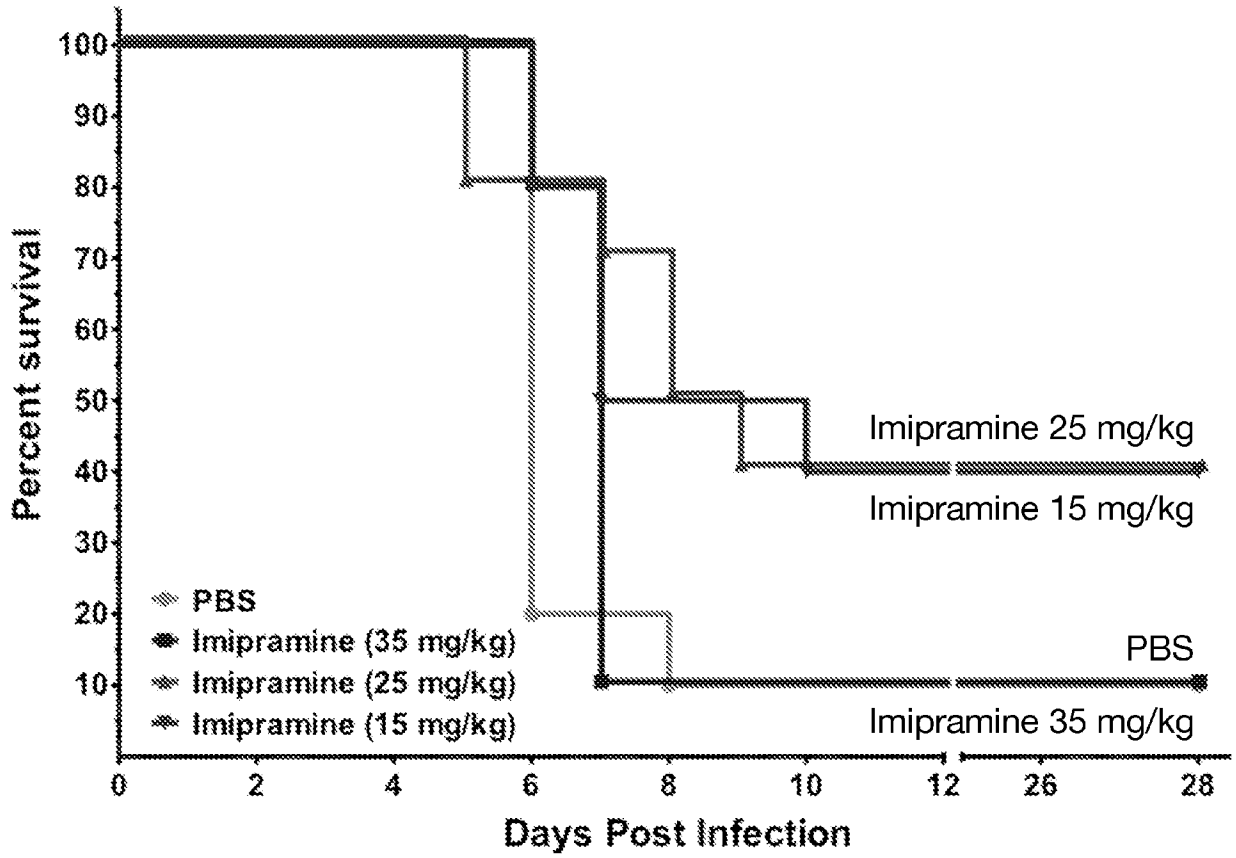
Percent survival
100
90
80
70
60
50
40
30
20
10
0
2
4
6
8
10
12
26
28
Days Post Infection
PBS
Imipramine (35 mg/kg)
Imipramine (25 mg/kg)
Imipramine (15 mg/kg)
Imipramine 25 mg/kg
Imipramine 15 mg/kg
PBS
Imipramine 35 mg/kg

FIGURE 18

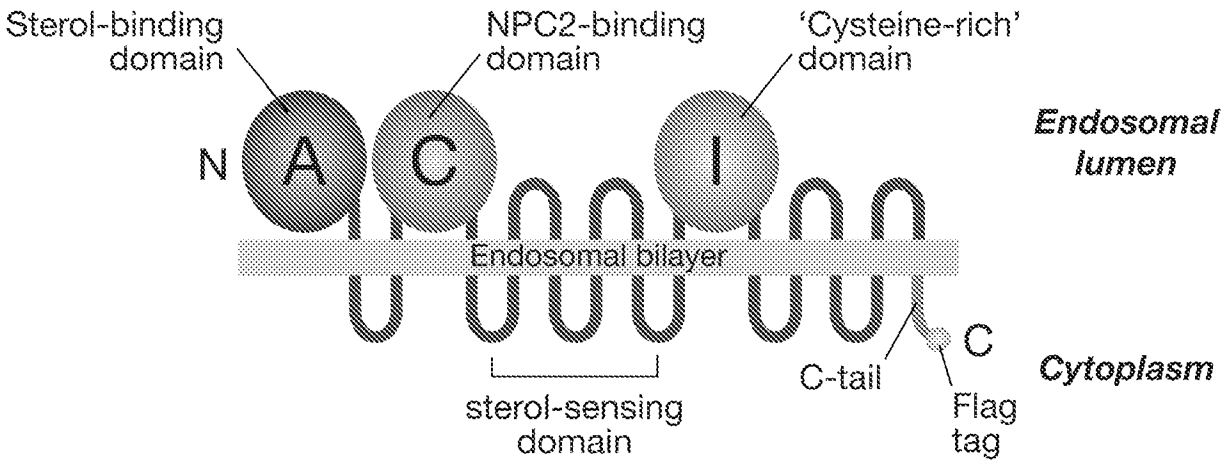
Sterol-binding domain
NPC2-binding domain
'Cysteine-rich' domain
N
A
C
I
Endosomal lumen
Endosomal bilayer
C
Cytoplasm
C-tail
Flag tag
sterol-sensing domain

FIGURE 19

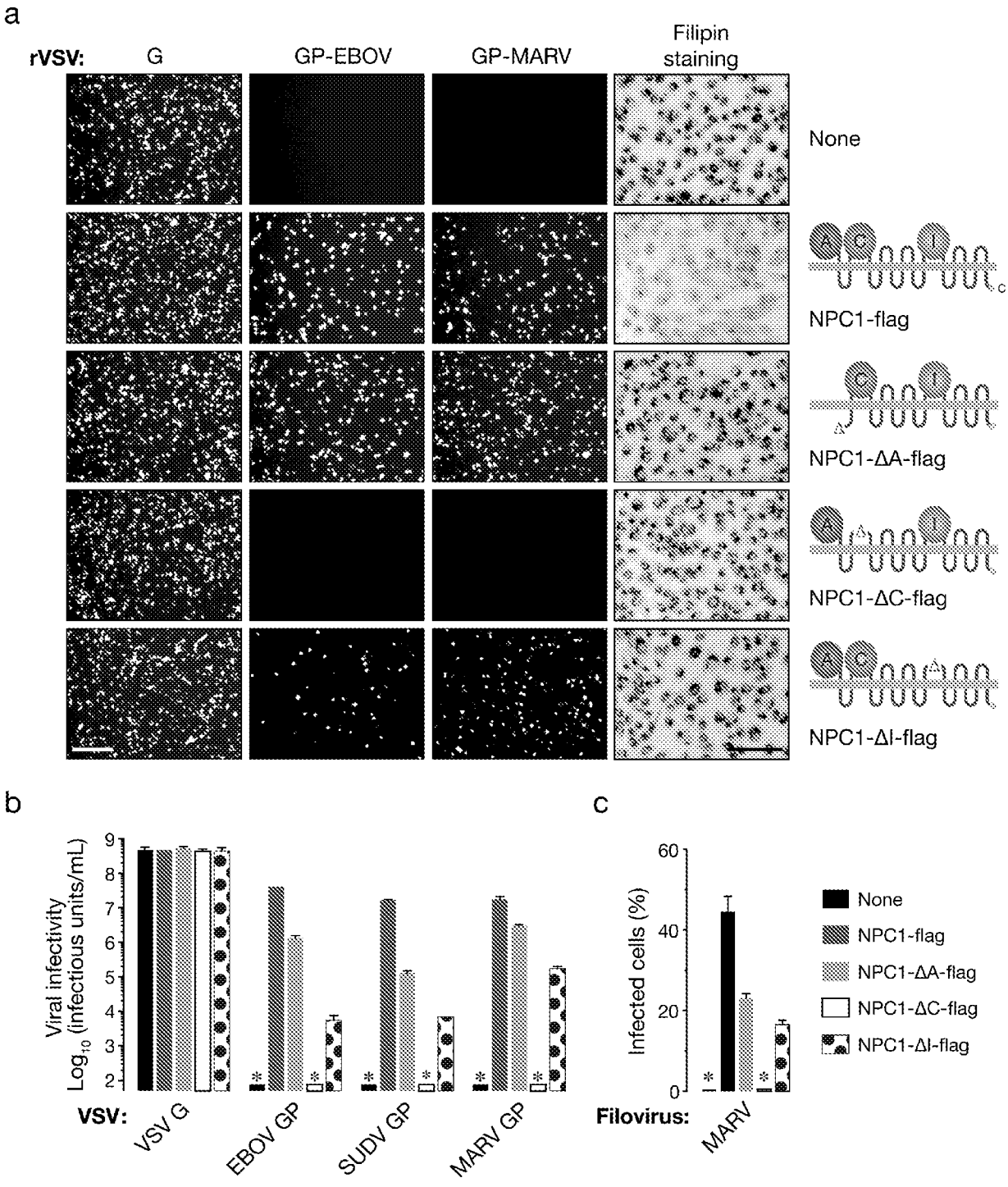

FIGURE 20A-20C

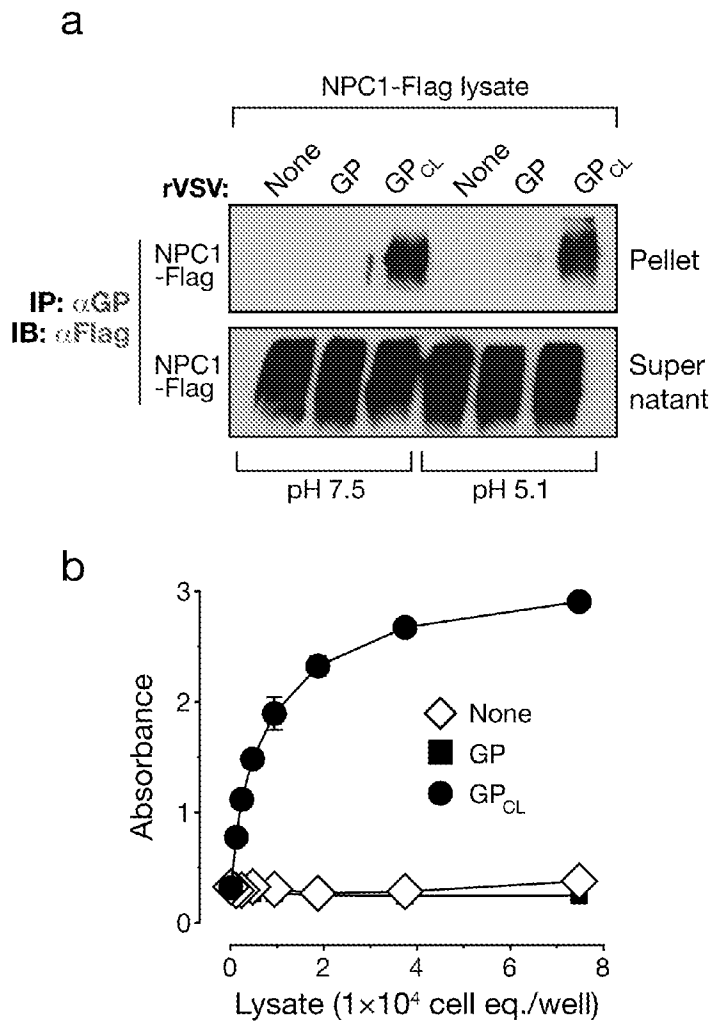

FIGURE 21A-21B

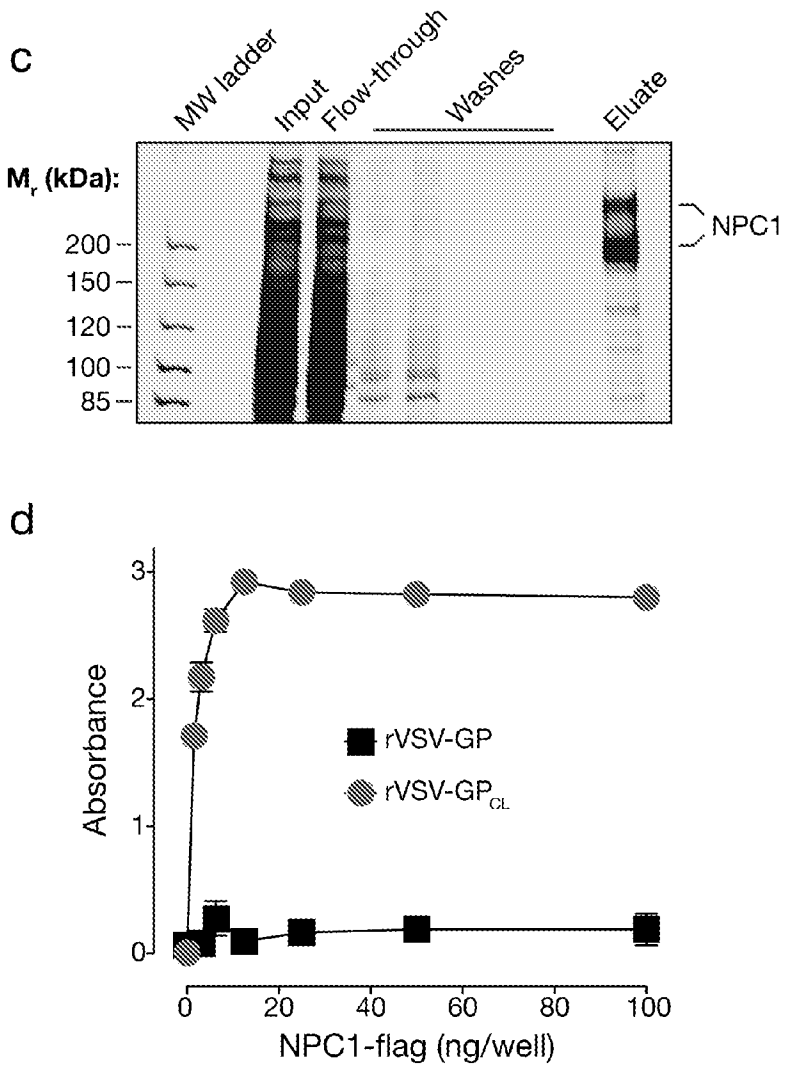

FIGURE 21C-21D

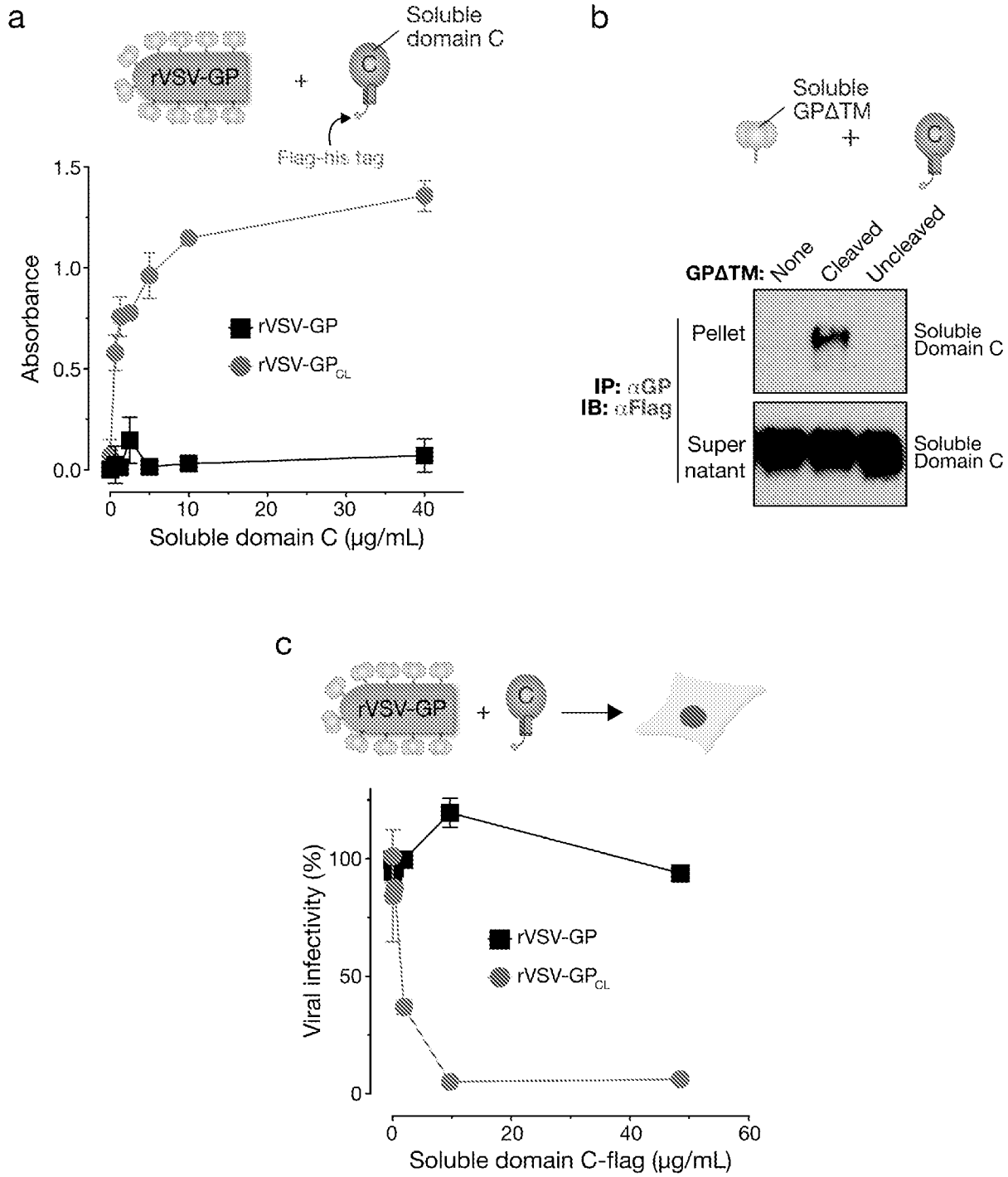

FIGURE 22A-22C

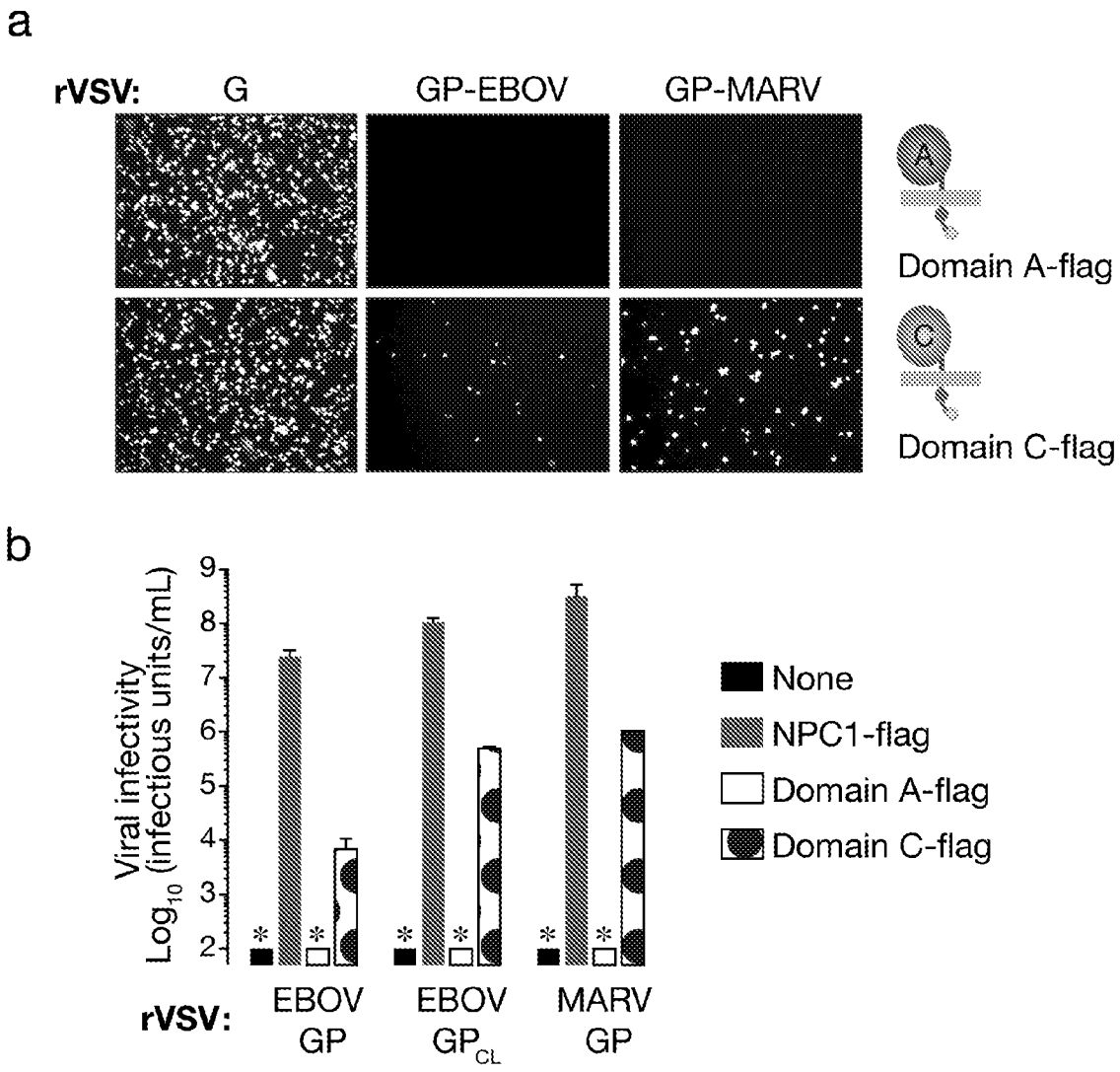

FIGURE 23A-23B

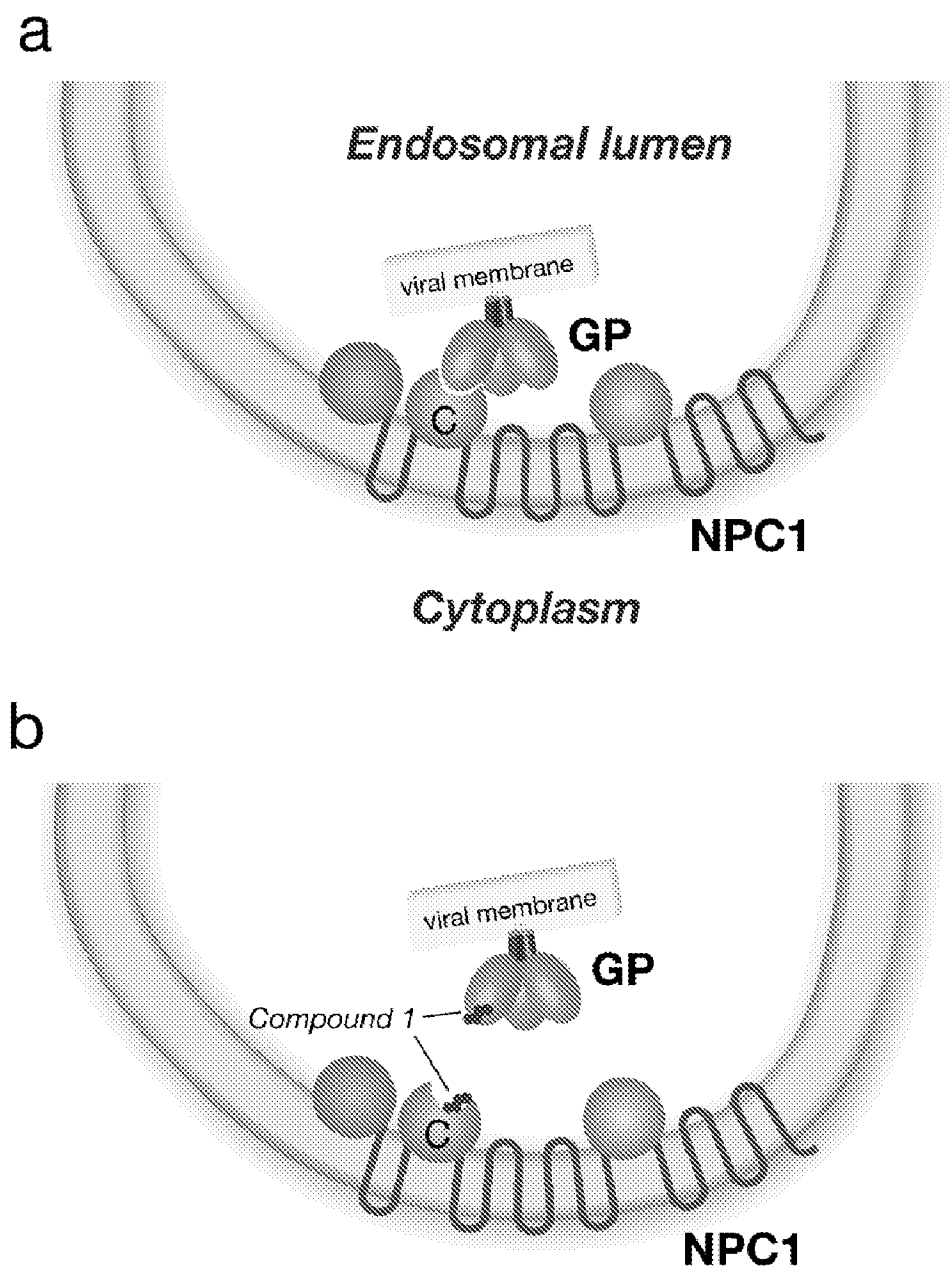

FIGURE 24A-24B

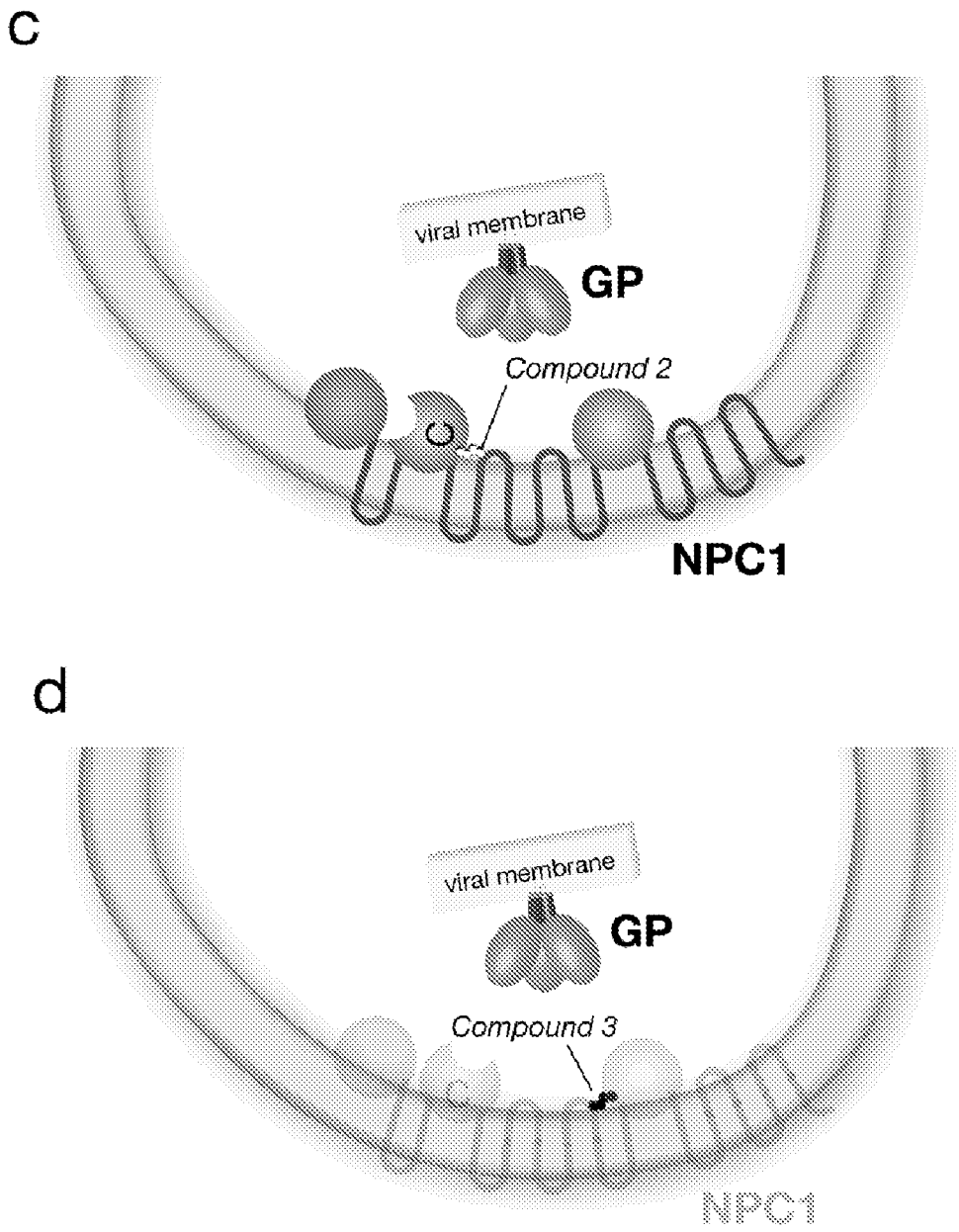

FIGURE 24C-24D

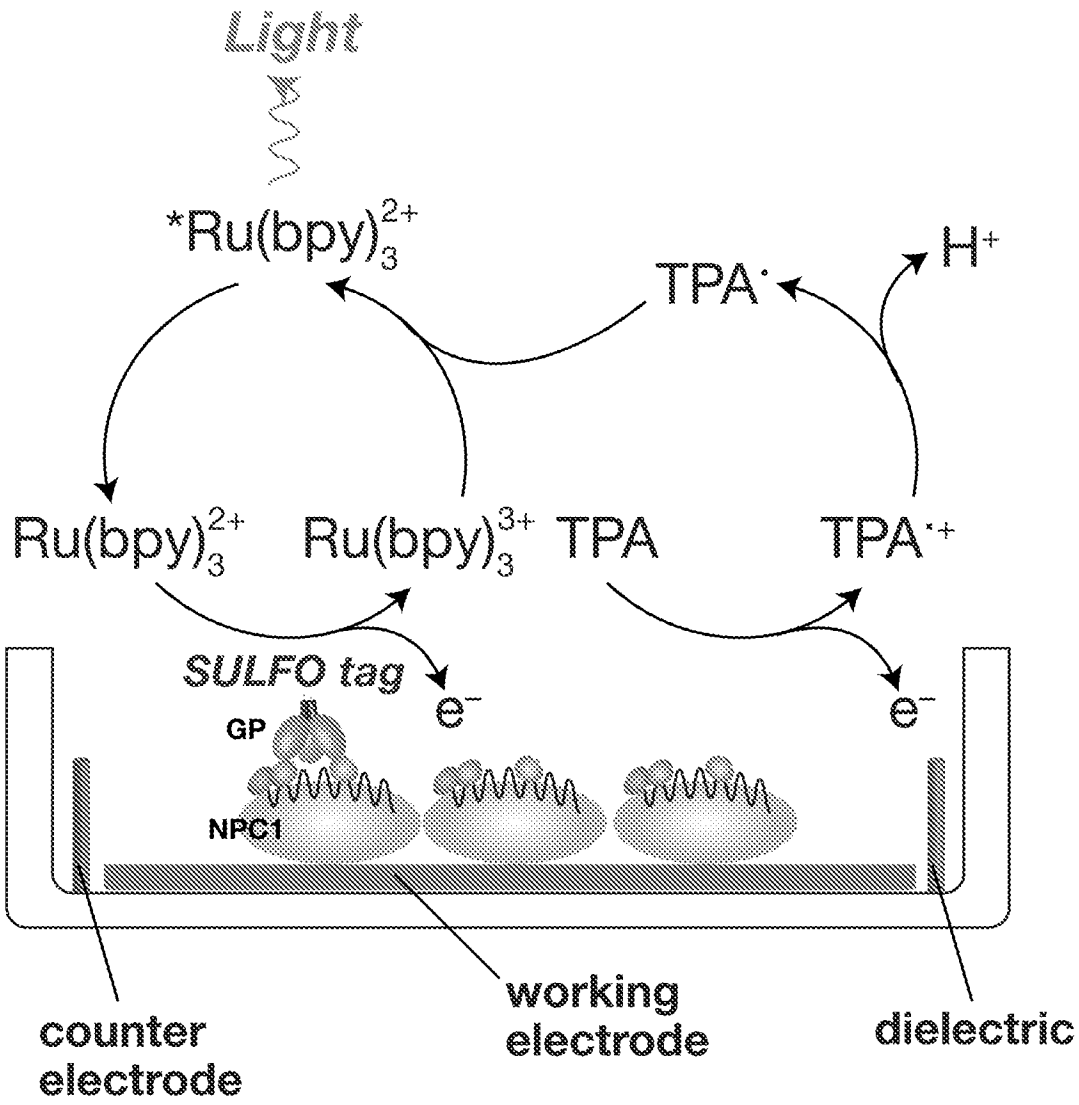
Light
*Ru(bpy)$_3^{2+}$
TPA·
H$^+$
Ru(bpy)$_3^{2+}$
Ru(bpy)$_3^{3+}$
TPA
TPA·+
SULFO tag
e$^-$
e$^-$
GP
NPC1
counter electrode
working electrode
dielectric

FIGURE 25

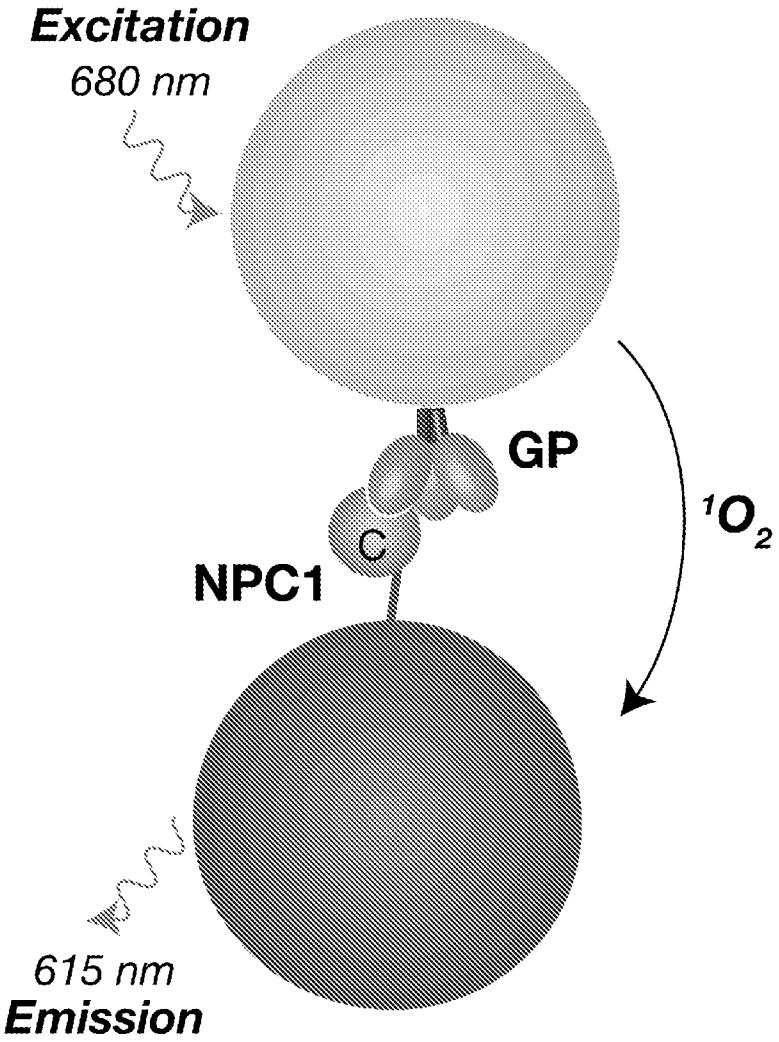

FIGURE 26

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/22349

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - A61K 39/12 (2012.01)
USPC - 424/204.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8)- A61K 39/12 (2012.01);
USPC- 424/204.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC- 424/1.11, 104.1; 435/320.1; 514/44R;
Patents and NPL (classification, keyword; search terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWest (US Pat, PgPub, EPO, JPO), GoogleScholar (PL, NPL), FreePatentsOnline (US Pat, PgPub, EPO, JPO, WIPO, NPL); search terms: filovirus, Marburg, Ebola, Niemann, Pick, NPC1, C1, CI, VPS11, VPS16, VPS18, VPS33A, VPS39, VPS41, BLOC1S1, BLOC1S2, GNPTAB, PIKFYVE, ARHGAP23, FIG4, bioterrorism, chemical, biological, warfare

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	US 2007/0087008 A1 (HODGE et al.) 19 April 2007 (19.04.2007), para [0007]-[0015], [0201], [0205], [0220], [0222], [0228], [0244], [0255], [0277], [0299], [0338], [0374]	1-3, 5/(1-3), 6/(1-3) ---------- 4, 5/(4), 6/(4), 10-13
Y	US 2010/0221357 A1 (OSTROFF) 02 September 2010 (02.09.2010), para [0013], [0070], [0088], [0106]	4, 5/(4), 6/(4), 10-13
X, P ------ Y, P	CARETTE et al. "Ebola virus entry requires the cholesterol transporter Niemann Pick C1." Nature, Letter [online], 24 August 2011 (24.08.2011) [Retrieved on 2011-03-28], Vol. 477, No. 7364, pp 340-343, Retrieved from the Internet: <URL: http://www.hms.harvard.edu/dms/Current/Courses/2011_2012/January/documents/nature10348.pdf>, see entire document, especially Figs. 1-4	1-3, 5/(1-3), 6/(1-3) ---------- 4, 5/(4), 6/(4), 10-13
Y	US 2002/0168771 A1 (GAMERMAN) 14 November 2002 (14.11.2002), para [0089], [0136]	1-6, 10-13

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 March 2012 (28.03.2012)	13 APR 2012
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/22349

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 7-9, 14-22
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.