Proteomic Organellar Mapping Reveals Modulation of Peroxisomes by HIV-1

Aaron L. Oom[1,2,3,4], Charlotte A. Stoneham[2,3,4], Mary K. Lewinski[2,4], Alicia Richards[5,8,9,10], Jacob M. Wozniak[1,6,7], David J. Gonzalez[6,7], Nevan J. Krogan[8,9,10], John Guatelli[1,2,3,4]*

[1]Biomedical Sciences Doctoral Program, University of California San Diego, La Jolla, CA 92093, USA

[2]School of Medicine, University of California San Diego, La Jolla, CA 92093, USA

[3]Veterans Medical Research Foundation, La Jolla, CA 92161, USA

[4]VA San Diego Healthcare System, La Jolla, CA 92161, USA

[5]Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037, USA

[6]Department of Pharmacology, University of California San Diego, La Jolla, CA 92093, USA

[7]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA

[8]Department of Cellular and Molecular Pharmacology, University of California San Francisco, CA 94158, USA

[9]Quantitative Biosciences Institute, University of California San Francisco, CA 94158, USA

[10]Institute for Virology and Immunology, J. David Gladstone Institutes, San Francisco, CA 94158, USA

Running title: Effect of HIV-1 on a T Cell Organellar Map

Abbreviations: acetonitrile (ACN), Dynamic Organellar Mapping (D.O.M.), formic acid (FA), fluorescence-assisted cell sorting (FACS), fetal bovine serum (FBS), gene ontology (G.O.), iodoacetamide (IAA), Local Outlier Factor (LOF), mass spectrometry (MS), PBS+0.1% Tween-20 (PBS-T), Principal Component Analysis (PCA), penicillin/streptomycin (pen/strep), serine incorporator 5 (SERINC5), stable isotope labeling by amino acids in cell culture (SILAC), support vector machine (SVM), tris(2-carboxyethyl)phosphine (TCEP), tandem mass tag (TMT), ultra-performance liquid chromatography (UPLC)

*Lead Contact: jguatelli@health.ucsd.edu

Abstract:

The interactions between HIV-1 and the host cell have been extensively studied, but cell-wide organellar changes caused by the virus have not been defined. Here, we aimed to detail perturbations to the host cell spatial proteome following HIV protein expression. Subcellular fractionation followed by mass spectrometric analysis of Jurkat T cells, in which HIV-expression was uniformly induced, identified thousands of cytoplasmic and membrane-bound proteins and enabled their placement within a multi-dimensional analytic space based on the 7 fractions. Spatial proteomic data were analyzed via a support vector machine bagging classifier to assess the movement of proteins between subcellular compartments. The proteins were also examined for absolute movement within 7-dimensional Euclidean space. We found that while some proteins moved between organelles, others remained classified within the same organelle but moved as a group relative to other organelles. The expression of HIV-1 affected the distribution of peroxisomal proteins, causing them to move closer in analytic space to proteins of the endoplasmic reticulum and lysosomes, but only when the viral protein Nef was expressed. These data identify Nef-dependent changes in peroxisomes as a novel perturbation of the host cell by HIV-1.

Introduction:

Spatial proteomics is a small, but growing, field within mass spectrometry (MS) that aims to understand the subcellular localization of the human proteome[1–7]. While initial efforts focused on establishing techniques and reference maps for various cell lines, recent work expanded the field into understanding the whole-cell effects of viral infection using human cytomegalovirus as a prototype[7]. This work led to novel findings on the importance of peroxisomes in herpesvirus infectivity[8], exemplifying the power of these methods for uncovering new and exciting viral biology. To this end, we aimed to map the movements of cellular proteins and organelles caused by the expression of HIV-1 in a T cell model system.

We reasoned that by comparing wild-type (WT) virus to virus lacking the accessory gene *nef* (ΔNef), we could both assess the robustness of our methods and make new observations. Nef is a small (27 kDa), myristoylated membrane-associated accessory protein expressed early during the replication cycle[9,10]. Nef increases viral growth-rate and infectivity[11], and it dysregulates the trafficking of cellular membrane proteins such as CD4, class I MHC, and proteins involved in T cell activation such as CD28[12] and p56-Lck[13]. Some of these activities enable the virus to evade immune detection[14,15]. To enhance infectivity Nef removes Serine Incorporator 5 (SERINC5)—a multi-pass transmembrane protein that inhibits the infectiousness of viral particles[16,17]—from the plasma membrane and sends it to lysosomes for degradation[18]. Due to the lack of an antibody that recognizes endogenous SERINC5, nearly all previous studies have used exogenously expressed, epitope-tagged constructs. We reasoned that spatial proteomic methods, which enable the characterization of endogenous proteins, might reveal the movement of endogenous SERINC5, along with other HIV-1-induced perturbations of the cellular spatial proteome.

To model HIV expression, we used a Jurkat T cell line that harbors a doxycycline-regulated HIV-1 genome. These cells were previously developed by our group to generate nearly homogenous HIV-positive cell populations for MS analysis[19]. Here we use inducible Jurkat T cell lines containing either WT or ΔNef HIV-1_{NL4-3} provirus and compare the spatial proteome of uninduced cells to cells post-induction. We expected SERINC5 and other Nef-targets to move only in cells expressing the WT and not in cells expressing the ΔNef viral genome. Using a 7-step, post-nuclear differential centrifugation procedure and

quantitative MS, we generated 7-dimensional data and analyzed it by several methods including a support vector machine (SVM) organelle-classifier. This analysis of post-nuclear fractions suited our focus on Nef, as Nef modulates the movement of cellular proteins between the plasma membrane and various endosomal membranes including the Golgi and lysosomes. While we were able to detect several known Nef-targets, including endogenous SERINC5, inconsistent mapping often hindered conclusive tracking of their movements following HIV expression. Nonetheless, our data revealed peroxisomes as the major organelle altered by HIV-1 and suggest a link between peroxisomes and Nef function. This finding could ultimately illuminate cryptic mechanisms through which Nef enhances infectivity beyond antagonizing SERINC5[20].

Experimental Procedures:

Experimental design and statistical rationale

All fractionation experiments with mass spectrometric analysis were performed in technical triplicate for each condition (uninduced and induced), with two biological replicates for wild-type and ΔNef NL4-3 Jurkat cells. This yielded a total of 6 uninduced and 6 induced technical replicates for each virus type. Biological replicates were prepared on separate days and analyzed by mass spectrometry on separate days. Western blotting and flow cytometry were performed on each technical replicate. Statistical significance for peroxisome quantification was calculated using Welch's t-test. Statistical significance for centroid-based movement analysis was calculated using Welch's t-test with the Benjamini-Hochberg correction for multiple comparisons with a false detection rate of 2.5%.

Cell culture

The doxycycline-inducible NL4-3 HIV-1 and NL4-3 ΔNef Jurkat cell lines were previously described[19]. The replication-incompetent genome used was based on pNL4-3 but lacked most of the 5' U3 region, encoded a self-inactivation deletion in the 3' LTR, and contained the V3 region from the R5-tropic 51-9 virus[21] to prevent the cell-cell fusion of the Jurkat T cells used herein, which do not express CCR5. Inducible cells were cultured in RPMI 1640 media supplemented with penicillin/streptomycin (pen/strep) and 10% Tet-free fetal bovine serum (FBS), as well as puromycin (1 µg/mL) and G418 (200

µg/mL) to maintain persistence of the tetracycline trans-activator and the inducible genome. Cells were passaged every two days to keep concentrations between $3.5x10^5$ and $1x10^6$ cells/mL. Cells were maintained at 37°C, 5% CO_2, and 95% humidity.

Doxycycline induction and fractionation

On the day prior to fractionation, $2.016x10^9$ cells were plated at $6x10^5$ cells/mL in T75 flasks at a total volume of 40 mL/flask. Half of these cells were induced to express HIV-1/HIV-1ΔNef with doxycycline (1 µg/mL) for 18 hours, while the other half remained uninduced. Following induction, cells of each condition, i.e. uninduced and induced, were split into three technical replicates, and then centrifuged at 500x*g* for 5 min at 4°C. Each technical replicate was pooled into a single 50 mL tube using ice cold 1X PBS, then counted by hemocytometer. From each technical replicate, $3x10^8$ cells were fractionated. Two aliquots of cells were taken from each technical replicate for whole cell western blots and testing induction by flow cytometry.

The fractionation protocol used here is derived from the Dynamic Organellar Maps method[5] with additional centrifugation steps[4] and TMT-based MS analysis as opposed to SILAC[6]. Cells for fractionation were centrifuged at 500x*g* for 5 min at 4°C then resuspended in ice-cold PBS and incubated for 5 min on ice. Cells were again centrifuged at 500x*g* for 5 min at 4°C, then resuspended in ice-cold hypotonic lysis buffer (25 mM Tris-HCl (pH 7.5), 50 mM sucrose, 0.5 mM $MgCl_2$, and 0.2 mM EGTA in water) and incubated for 5 min on ice. Using a 7 mL Dounce homogenizer, cells were homogenized with 20 full strokes of the tight pestle. Cell homogenates were then immediately transferred to a 13 mL (14x89 mm) ultracentrifuge tube with sufficient ice-cold hypertonic sucrose buffer (1.25 M sucrose, 25 mM Tris-HCl (pH 7.5), 0.5 mM $MgCl_2$, and 0.2 mM EGTA in water) to restore 250 mM sucrose concentration. All replicates were then centrifuged at 1,000x*g* for 10 min at 4°C in a Beckman Coulter ultracentrifuge (SW-41 Ti rotor), balancing each tube with balance buffer (250 mM sucrose, 25 mM Tris-HCl (pH 7.5), 0.5 mM $MgCl_2$, and 0.2 mM EGTA in water). Supernatants were transferred to a fresh ultracentrifuge tube, balanced with balance buffer, then fractionated using the following differential centrifugation protocol: 3,000x*g* for 10 min, 5,400x*g* for 15 min, 12,200x*g* for 20 min, 24,000x*g* for 20 min, 78,400x*g* for 30 min, 110,000x*g* for 35 min, and 195,500x*g* for 40 min All centrifugation steps were performed at 4°C with

pellets from each spin being resuspended in SDS buffer (2.5% SDS and 50 mM Tris-HCl (pH 8.0) in water). Fractions were then heated for 10 minutes at 72°C. Protein content of each fraction was quantified in triplicate using a bicinchoninic acid (BCA) protein assay (Thermo-Fisher).

Confirmatory western blots and p24 flow cytometry

Prior to mass spectrometric analysis of fractions, induction and fractionation were evaluated by flow cytometry and western blotting. For p24 flow cytometry, an aliquot of $2x10^6$ cells from each technical replicate were pelleted at 500x*g* for 5 min at 4°C then resuspended in ice-cold FACS buffer (2% FBS and 0.1% sodium azide in 1X PBS). The cells were again pelleted at 500x*g* for 5 min at 4°C then resuspended in Cytofix/Cytoperm reagent (BD Biosciences) and incubated on ice for 30 min Following fixation/permeabilization, cell suspensions were diluted with wash buffer and pelleted at 500x*g* for 5 min at 4°C. Cells were resuspended in p24 primary antibody solution (1:100 dilution of p24-FITC antibody clone KC57 (Beckman Coulter) diluted in perm/wash buffer) and incubated on ice for 30 min in darkness. Ice-cold FACS buffer was added to each sample and cells were pelleted at 500x*g* for 5 min at 4°C. The intracellular p24 was analyzed using an Accuri C6 flow cytometer (BD Biosciences). Uninduced cells had an average p24+ population of 0.27% (S.D. = 0.20) and live cell population of 85.78% (S.D. = 3.37). Induced cells had an average p24+ population of 94.85% (S.D. = 1.23) and live cell population of 79.25% (S.D. = 4.35).

An aliquot of $1x10^7$ cells from each technical replicate was lysed in SDS buffer and probe sonicated on ice until no longer viscous. 3,000x*g* fractions were also probe sonicated. The samples were mixed with 4X loading buffer (200 mM Tris-HCl (pH 6.8), 8% SDS, 40% glycerol, 200 mM tris(2-carboxyethyl)phosphine-HCl (TCEP), and 0.04% bromophenol blue in water) and proteins were then separated on 10% SDS-PAGE gels at a constant 70V. Proteins were transferred to polyvinylidene difluoride (PVDF) membranes for 1 hour using the Trans-Blot turbo (BioRad) system using standard conditions. The membranes were blocked in 5% milk in 1X PBS-T for 30 min at room temperature prior to incubation with primary antibodies diluted in 1% milk and 0.05% sodium azide in 1X PBS-T: sheep anti-Nef (gift from Celsa Spina, diluted 1:3,000), mouse anti-p24 (Millipore, diluted 1:500), Chessie8 (mouse anti-gp41, NIH AIDS Research and Reference Reagent program[22], diluted 1:10,000), rabbit anti-Vpu (NIH

AIDS Research and Reference Reagent program ARP-969, contributed by Dr. Klaus Strebel, diluted 1:1,000), and mouse anti-GAPDH (GeneTex, diluted 1:5,000). The blots were washed and probed with either horseradish peroxidase-conjugated goat anti-mouse, HRP-goat anti-rabbit, or HRP-rabbit anti-sheep secondary (BioRad) diluted 1:3,000, incubating for 1 hour at room temperature on a shaker. Apparent molecular mass was estimated with PageRuler protein standard (Thermo Scientific). Blots were imaged using Western Clarity detection reagent (BioRad) before detection on a BioRad Chemi Doc imaging system with BioRad Image Lab v5.1 software.

Sample digestion for mass spectrometry

Disulfide bonds were reduced with 5 mM TCEP at 30°C for 60 min and cysteines were subsequently alkylated (carbamidomethylated) with 15 mM iodoacetamide (IAA) in the dark at room temperature for 30 min Proteins were then precipitated with 9 volumes of methanol, pelleted and resuspended in 1M urea, 50 mM ammonium bicarbonate. Following precipitation, protein concentration was determined using a BCA protein assay. A total of 0.2 mg of protein was subjected to overnight digestion with 8.0 µg of mass spec grade Trypsin/Lys-C mix (Promega). Following digestion, samples were acidified with formic acid (FA) and subsequently 150 ug peptides were desalted using AssayMap C18 cartridges mounted on an Agilent AssayMap BRAVO liquid handling system, C18 cartridges were first conditioned with 100% acetonitrile (ACN), followed by 0.1% FA. The samples were then loaded onto the conditioned C18 cartridge, washed with 0.1% FA, and eluted with 60% MeCN, 0.1% FA. Finally, the organic solvent was removed in a SpeedVac concentrator prior to LC-MS/MS analysis.

TMT Labeling

Peptide concentration was determined using a Nanodrop, and a total of 15 µg of peptide was then used for TMT labeling, each replicate serving as a multiplex. Briefly, dried peptide sample was resuspended in 200 mM HEPES (pH 8) and incubated for 1 h at room temperature with one of the TMT10-plex reagents (ThermoFisher) solubilized in 100% anhydrous ACN. Reactions were quenched using a 5% hydroxylamine solution at 1-2 µl per 20 µl TMT reagent. The 10 samples (or however many in a multiplex) were then pooled and dried in a SpeedVac. The labeled peptides were resuspended in 0.1%

FA. After sonication for 1min, sample was desalted manually using SepPak, column was first conditioned with 100% ACN, followed by 0.1% FA. Sample was loaded, then washed with 0.1% FA and eluted in a new vial with 60% ACN, 0.1% FA. Finally, the organic solvent was removed using a SpeedVac concentrator prior to fractionation.

High pH Reverse-Phase Fractionation

Dried samples were reconstituted in 20mM ammonium formate (pH ~10) and fractionated using a Waters Acquity CSH C18 1.7 µm 2.1 × 150 mm column mounted on a MClass Ultra Performance Liquid Chromatography (UPLC) system (Waters corp., Milford, MA) at a flow rate of 40 µl/min with buffer A (20 mM ammonium formate pH 10) and buffer B (100% ACN). Absorbance values at 215 nm and 280 nm were measured on Waters UV/Vis spectrophotometer, using a flowcell with a 10 mm path length. Peptides were separated by a linear gradient from 5% B to 25% B in 62.5 min followed by a linear increase to 60% B in 4.5 min and 70% in 3 min and maintained for 7 min before increasing to 5% in 1 min Twenty-four fractions were collected and pooled in a non-contiguous manner into twelve total fractions. Pooled fractions were dried to completeness in a SpeedVac concentrator.

LC-MS^3 Analysis

Dried samples were reconstituted with 0.1% FA and analyzed by LC-MS/MS on an Orbitrap Fusion Lumos mass spectrometer (Thermo) equipped with an Easy nLC 1200 ultra-high pressure liquid chromatography system interfaced via a Nanospray Flex nanoelectrospray source (Thermo). Samples were injected on a C18 reverse phase column (25 cm x 75 um packed with Waters BEH 1.7 um particles) and separated over a 120-min linear gradient of 2-28% solvent B at a flow rate of 300nL/min The mass spectrometer was operated in positive data-dependent acquisition mode.

Parameter settings were set as follows: FT MS1 resolution (120□000) with AGC target of 1e6, ITMS2 isolation window (0.4 m/z), IT MS2 max. inject time (120 ms), IT MS2 AGC (2E4), IT MS2 CID energy (35%), SPS ion count (up to 10), FT MS3 isolation window (0.4 m/z), FT MS3 max. inject time (150 ms), FT MS3 resolution (50□000) with AGC target of 1e5. A TOP10 method was used where each FT MS1 scan was used to select up to 10 precursors for interrogation by CID MS2 with readout in the ion

trap. Each MS2 was used to select precursors (SPS ions) for the MS3 scan which measured reporter ion abundance.

Mass spectrometry spectra identification

The raw MS data were searched with the MaxQuant software[23] (version 1.5.7.5) using the Andromeda search engine[24]. Searches were performed against a UniProt human canonical database appended with HIV-1_{NL4-3} protein sequences. Reporter ion MS3 was selected for searches, with 10plex TMT as a fixed modification and a reporter mass tolerance of 0.003 Da. Trypsin was selected, with a maximum of 2 missed cleavages. A minimum of 7 amino acids was required for peptide identification. The fixed modification specified were carbamidomethylation of cysteine residues and TMT and variable modifications were oxidation of methionine and acetylation of protein N-terminus. Peptides and proteins groups were both filtered to a 1% FDR. Protein groups were filtered for "Only identified by site", "Reverse", and "Contaminant".

SVM and computational analysis

Each technical replicate was row-normalized, then searched for organellar marker proteins. Markers were examined for eight compartments: endoplasmic reticulum, endosomes, Golgi apparatus, large protein complex (a cytosolic catch-all that includes large complexes (e.g. ribosomes, proteasomes) and cytoskeletal elements), lysosomes, mitochondria, peroxisomes, and the plasma membrane. Detected proteins were cross-refenced with markers used by the Borner group[5] as well as with proteins with a singular subcellular localization and an enhanced rating in the Cell Atlas[25]. Final marker proteins were selected for each biological replicate using Local Outlier Factor (LOF) to detect outliers in Principal Component Analysis (PCA) space for each organelle in each technical replicate then keeping inliers common to every technical replicate of a given biological replicate, regardless of induction status. LOF thresholds were individually set for each biological replicate such that the least populated class of markers comprised at minimum 11.

Following marker selection, marker proteins were scaled in a column-wise fashion, e.g. fraction-wise, such that each column ranged from 0 to 1. Remaining proteins were scaled based on this range.

Classification was performed using a 500-member SVM bagging classifier in an iterative fashion. The gamma and cost (C) hyperparameters for the radial basis function kernel were selected using a 10-fold grid search cross-validation in log-space. SVM classification was performed in a recursive manner[26] with the first iteration classifying proteins into all 8 compartments, the second iteration classifying to all but the mitochondria (with its respective markers removed), and the third and final iteration classifying further excluding the large protein complex class and its respective markers. Proteins classified as mitochondrial in the first iteration, proteins classified as large protein complex in the second iteration, and all remaining proteins classified in the third iteration were concatenated together to form the finalized data frame. For each protein, the frequency of its classification into its final class was noted as a confidence measure, i.e. 300 of 500 classifiers in the ensemble outputting mitochondria for a protein is a 60% confidence value of that classification. For each organelle in each technical replicate, proteins in the top 50% of the range of confidence values were kept as classified while proteins with lower confidence values were reverted to an unclassified designation. Classified proteins were visualized using PCA.

Radar plots in Figure 6 were generated using marker proteins common to either all wild-type technical replicates (uninduced and induced) or all ΔNef technical replicates. The average of all pairwise Euclidean distances between every marker of peroxisomes and every marker of a given organelle were calculated and plotted. Centroid-based distance analysis was performed by determining the geometric center of individual proteins in a given set of technical replicates (e.g. WT biological replicate 1, uninduced technical replicates) and calculating p-values with Welch's t-test for movement of the geometric center following induction. Proteins were then ranked by p-value and compared to Benjamini-Hochberg critical values with an FDR of 2.5% for significance. The top 1,000 proteins for a given biological replicate of a genome type were cross-referenced against the top 1,000 proteins of the other biological replicate, and concordant proteins were then analyzed by STRING.

Analyses were carried out in Python 3.7.4 using packages from scikit-learn[27].

Generation of eGFP-SKL stably expressing Jurkat lines

Inducible HIV Jurkat lines stably expressing peroxisomal eGFP were generated by stable transduction of both wild-type and ΔNef cell lines. Lentiviral vectors were produced in HEK293T cells with

of 3rd generation lentivector components; pRSV-Rev (Addgene #12253), pMD2.G (Addgene #12259), and pMDLg/pRRE (Addgene #12251) were generous gifts of Didier Trono. HEK293T cells were plated in 10-cm dishes ($5x10^6$ cells/dish) in Dulbecco's Modified Eagle Medium (DMEM) supplemented with pen/strep and 10% FBS then transfected the following day using Lipofectamine 2000 (ThermoFisher) with equimolar ratios of lentivector components and pLenti-MP2 transfer vector (Addgene #36097, generous gift of Pantelis Tsoulfas) containing peroxisome-localizing eGFP (Addgene #54501, generous gift of Michael Davidson); each of three 10-cm dishes were transfected with a total of 21 µg of plasmids following manufacturer's protocol. Fluorescent protein sequences were subcloned from their original vectors and inserted into the transfer vector by restriction digest. Transfected HEK293T cells were incubated for 48 hours before lentivectors were harvested from cell supernatants. Lentiviral particles were concentrated using Lenti-X concentrator (Takara Bio) following the manufacturer's protocol. Briefly, 3 volumes of clarified supernatants were cooled on ice with 1 volume of concentrator until cooled. Particles were then pelleted by centrifugation at 1,500x*g* for 45 minutes at 4°C and resuspended in 1/10th of the original volume. Concentrated supernatant from one and a half 10-cm dishes of transfected HEK293T cells was used to transduce $2.5x10^7$ inducible Jurkat cells of a given HIV genome by spinoculation at 1,200x*g* for 2 hours at room temperature with polybrene added (8 µg/mL). Following spinoculation, Jurkat cells were resuspended in fresh complete RPMI media and plated at ~$4x10^5$ cells/mL, then incubated for 48 hours. Transduced cells were then sorted by flow cytometry for bulk eGFP-positive cells.

Immunofluorescence and peroxisome quantification

Inducible Jurkat cells were induced as described above for 18 hours. The cells were washed twice with cold PBS, then the plasma membrane stained with wheat germ agglutinin AlexaFluor 647 conjugate (5 µg/mL in PBS) for 15 min on ice in darkness; all subsequent steps were performed in darkness. Following plasma membrane staining, the cells were fixed in cold 4% PFA in PBS (pH 7.4) for 20 min. The cells were washed twice with cold PBS and then with MilliQ water. The cells were resuspended in Mowiol (polyvinyl alcohol) mounting media (prepared in-house) containing DAPI (1 µg/mL) and placed on a microscope slide with a coverslip on top. All slides were allowed to cure overnight at room temperature before being imaged. Images were taken at 100x magnification (1344 x 1024 pixels)

on an Olympus IX81 wide-field microscope fitted with a Hamamatsu CCD camera using the Slidebook software v6 (Imaging Innovations, Inc). Z stacks were generated using 0.5 micron steps then deconvolved using a nearest-neighbor algorithm in the Slidebook software. Peroxisomes were quantified and measured using the 3D Objects Counter[28] and the 3D ROI Manager[29] plugins in FIJI (ImageJ). Those peroxisomes that were more than three times the interquartile range above the third quartile (i.e. Q3 + 3*IQR) were considered to be overexposure artifacts and discarded from counting. A minimum of 50 cells were counted for each condition in each of three biological replicates. Statistical significance was calculated using Welch's t-test.

Results:

Doxycycline-inducible HIV-1$_{NL4-3}$ Jurkat T cells are a scalable and uniform system for subcellular fractionation and proteomic studies.

The WT HIV-1 inducible cells used here were previously generated and used for whole-cell quantitative- and phospho-proteomics[19]. To avoid the formation of syncytia, which could alter the subcellular fractionation and subsequent spatial proteomic data, the inducible HIV-1$_{NL4-3}$ genomes were modified with a CCR5-tropic Env protein to avoid cell-cell fusion between the CCR5-negative Jurkat cells. Due to the high induction rates of HIV-1 expression and the scalability of this culture system, we reasoned that it would be amenable to subcellular fractionation by differential centrifugation with subsequent MS analysis (Fig. 1A). To determine the optimal time-point for analysis following induction of HIV-1 expression, cells were treated with doxycycline for 0, 4, 8, 12, 16, and 18 hours, and the expression of HIV-1 proteins was detected by western blotting and flow cytometry (Fig. 1B and C). WT cells began to express Nef by 4 hours post-induction, and both WT and ΔNef cells expressed p55 Gag precursor (the precursor protein for virion structural proteins) by 8 hours and gp160 (the envelope glycoprotein precursor) by 12 hours. By 18 hours, viral proteins were robustly expressed; about 90-95% of both WT and ΔNef cells were positive by flow cytometry for p24 capsid (a proteolytic product of p55).

Subcellular fractionation was performed 18 hours post-induction; the cells were mechanically ruptured with a Dounce homogenizer in hypotonic solution, then subjected to a differential centrifugation protocol before preparation for quantitative, multiplexed MS analysis. Uninduced and induced cells were

handled in technical triplicate for each biological replicate (n=2). We used a modified version of the Dynamic Organellar Mapping (D.O.M.) protocol[5,6] with additional fractions generated at 110,000x*g* and 195,500x*g* to increase the resolution of the classification analysis. As a quality control before MS, protein yields were quantified for each fraction (Fig. 2A). The post-nuclear fractions accounted for only ~10-15% of total cellular protein, presumably because nuclear proteins and soluble cytoplasmic proteins that failed to pellet at 195,500x*g* were discarded, leaving primarily membranous organelles or organellar fragments and large, cytoplasmic complex proteins in the fractions analyzed. We also observed decreasing protein yields across the fractions, with an increase in the 78,400x*g* fraction, consistent with the original D.O.M. study using HeLa cells[5]. In further support of differential fractionation, differential abundances of viral proteins across the fractions in cells expressing either the WT or ΔNef genomes were observed by western blotting (Fig. 2B). When marker proteins for distinct organelles (defined below) were quantified by MS, variable abundances across the fractions were observed, indicating differential distribution of the organelles among the fractions (Fig. 2C).

To assess the consistency of the protocol, Spearman correlations of the various fractions were compared across biological replicates for WT (Fig. S1 and S2) and ΔNef (Fig. S3 and S4). With offline HPLC fractionation prior to MS, ~75% of all detected proteins are found in a majority—at least 4 of 6 technical replicates—for each condition (uninduced or induced) of both WT and ΔNef expressing cells (Fig. S1 and S3). Moreover, when both biological replicates for each cell line were considered, the protein content of the fractions predominantly clustered by *g*-force rather than by biological replicate or condition (Fig. S2 and S4). Taken together, these findings indicate the depth and reproducibility of this inducible HIV-1 model for subcellular fractionation studies.

HIV-expression minimally affects global protein distribution.

Proteins were classified using a 500-member SVM bagging classifier with recursive filtering to minimize interference from larger and more distinct classes when classifying smaller, membranous organelles (Fig. 3A). For each biological replicate, classifiers were trained on a well-behaved (i.e. tightly clustered in analytical space) set of resident protein markers for each of 8 classes/compartments: endosomes, endoplasmic reticulum, Golgi apparatus, large protein complexes, lysosomes, mitochondria,

peroxisomes, and the plasma membrane (see Experimental Procedures for further details). The large protein complex compartment is a designation used in the Itzhak *et al.* D.O.M. study[5] and here included large, cytoplasmic complexes such as ribosomes and the proteasome, as well as other non-membranous structures, including cytoskeletal elements. To remove low-confidence classifications, we used the proportion of ensemble members that output the final classification ("label") to revert proteins to an "unclassified" group if they were in the lower half of an organelle's range of confidence values for a given technical replicate. The data analysis and markers used for each biological replicate are shown in Supplemental Tables 1-4.

Following SVM classification, labeling concordance was examined by condition and virus type (Fig. 3B and C). Of those proteins detected in all 6 technical replicates for a given condition and virus type, between 80 and 85% were classified the same in a majority of technical replicates (at least 4 out of 6), highlighting the consistency of the SVM classification analysis. Induction of HIV expression did not affect this concordance. Moreover, it had little effect on the overall distribution of proteins among each organellar compartment (Fig. 3D). These distributions suggest that the majority of proteins had a non-membranous organellar localization (large protein complex, shown in red). The remaining proteins were distributed among endosomes (orange), mitochondria (brown), and the ER (blue), while lysosomes (purple), peroxisomes (pink), Golgi apparatus (dark green), and the plasma membrane (light green) each accounted for 5% or less of the total. Although the ensemble SVM approach allows for low-confidence proteins to revert to an unclassified designation, less than 10% of proteins ended up in this category, indicating that the approach confidently classified the majority of proteins.

Nonetheless, certain organelles outperformed others regarding classification concordance. Mitochondrial proteins were the most consistently labeled of any membranous organelle across technical replicates (Fig. S5-S8). Additionally, 5% or less of mitochondria-labeled proteins ultimately reverted to unclassified following SVM classification (Fig. S9). On the other hand, compartments such as the peroxisome showed a greater reversion to unclassified with a majority of peroxisome-labeled proteins classified as peroxisomal in only 1 of 6 technical replicates (Fig. S5-S9). This could be due to more frequent exchange of proteins between peroxisomes and other membrane-bound organelles such as ER, endosomes, and lysosomes, or to a similar behavior of peroxisomes to these organelles in the

fractionation procedure. Further characterization of the classification analysis and concordance by individual organelles are in the Supplemental Figures (Fig. S5-S11).

To further evaluate the performance of the classification pipeline, we cross-referenced our uninduced classified datasets with those of previously published organellar proteomic studies[30–33]. Some disagreement with these studies for both WT (Fig. S12A) and ΔNef (Fig. S12B) was found. This could be due to differences in cell type and cell-fractionation protocols. It could also be due to misclassifications introduced by forcing proteins into a single location (classification) rather than allowing for localization of proteins to more than one organelle.

We inspected the classification of selected known Nef-targets, such as SERINC5 (Fig. 4A), Lck (Fig. 4B), and CD28 (Fig. 4C), as well as Nef itself (Fig. 4D), and Gag (Fig. 4E). Half or more of the induced replicates reverted to unclassified for SERINC5 and CD28, rendering an assessment of HIV-dependent organellar translocations difficult for these proteins. However, SERINC5, Lck, and CD28 are all known to localize to the plasma membrane, and the plasma membrane was a common classification in the uninduced replicates for these proteins, suggesting reasonable assignments in that condition. The classification of Nef to large protein complexes might reflect non-myristoylated Nef which does not associate with membranes but rather stays soluble within the cytoplasm[34]. Why Gag was only classified to endosomes in WT and was predominantly unclassified in ΔNef is unclear, since Nef is not known to affect the subcellular distribution of Gag.

Cell-mapping reveals HIV-1- and Nef-dependent changes in peroxisomes.

Following SVM classification, the proteins detected in cells expressing the WT or ΔNef HIV-1 genomes (and their uninduced counterparts) were visualized using Principal Component Analysis (PCA); the proteins were colored according to their SVM-generated organellar assignment (Fig. 5). As expected, mitochondrial proteins (brown) showed a distinctly separated cluster. The large protein complex class (red) was also well separated from the remaining membranous organelles and exhibited a dispersed distribution, suggesting heterogeneity of its components. At first glance, little difference was apparent between uninduced and induced cell maps. Closer examination of the WT plots, however, revealed a shift in the peroxisome cluster (pink) from adjacent to the plasma membrane cluster (light green) to nestled

between the plasma membrane and ER clusters (blue); see Fig. 5 insets for detail and Fig. S13 for plots with the large protein complex and mitochondrion classes removed for clarity. This peroxisomal shift was less apparent in the ΔNef plots, suggesting a potential Nef phenotype.

PCA-space is not a strict interpretation of high-dimensional spaces. Consequently, we wanted to determine more accurately if our data reflected a shift in the behavior of peroxisomal proteins. To probe this shift in a more systematic and authentic manner, the pairwise distances between organellar marker proteins in the 7-dimensional space were calculated for all conditions and the average values were plotted (Fig. 6A). Following induction of WT HIV-1, these radar plots displayed an increased distance between peroxisomal proteins and mitochondrial, Golgi, and plasma membrane proteins, but a decreased distance between peroxisomal proteins and lysosomal and ER proteins. In contrast, following induction of ΔNef HIV-1, these movements of peroxisomal proteins relative to other organellar proteins were minimal, a result in agreement with the apparent shifts in PCA-space.

To determine if specific peroxisomal proteins were shifting in or out of the organelle, technical replicates of matching conditions were pooled and proteins that were consistently classified (at least 4 out of 6 technical replicates) in one organelle for uninduced cells and a different organelle following HIV-induction were considered hits for label-based movement analysis; a full list of results is found in Supplemental Tables 5 and 6. When these proteins were input to STRING[35] for gene ontology (G.O.) enrichment analysis, no enrichment of peroxisomal proteins was observed in either WT or ΔNef conditions (Fig. 6B). This suggests that the shifts in peroxisomal proteins described above are due to changes in peroxisomes as a whole and their behavior during differential centrifugation, rather than a migration of specific proteins between peroxisomes and other organelles.

To support the notion of peroxisome-wide changes, we used a centroid-based movement analysis. Shifts in protein behavior in the 7-dimensional space defined by fractionation were sought by comparing the spatial center of uninduced technical replicates to that of induced technical replicates for each protein; a full list of results is found in Supplemental Tables 7 and 8. Unlike the compartment-based analysis, this analysis is agnostic of our classification pipeline and can potentially reveal changes missed by the former. Concordant hits of the centroid-based movement analysis between biological replicates were enriched in peroxisomal proteins for cells expressing WT HIV-1, but not for cells expressing ΔNef

(Fig. 6C). From these analyses, we infer that the expression of HIV induces changes at the organellar level, specifically to peroxisomes, that seem independent of movement of specific proteins into or out of the organelle. Similarly, the tyrosine kinase Lck, appeared to move in the centroid-based analysis following induction of WT but not ΔNef, yet its classification as “plasma membrane” did not change (Supplemental Table 6).

To determine whether the above described changes in the spatial proteome represent morphological changes to peroxisomes, we imaged Jurkat cells expressing peroxisome-localized eGFP before and after induction of HIV-1 expression (Fig. 6D). The number of peroxisomes per cell increased modestly upon expression of either WT or ΔNef HIV-1, however the p-value did not reach significance in the case of the WT and in both cases the effect-size was small (Fig. 6E). Moreover, the morphology of the peroxisomes and their distribution within the cells was not appreciably changed by the expression of WT or ΔNef HIV-1. In summary, while these analyses suggest an alteration in the fractionation behavior of peroxisomes due to the expression of HIV-1 encoding Nef, to what exactly that change is due and whether and how it affects peroxisome function remains unclear.

WT centroid analysis suggests Nef-dependent perturbations of enzymes involved in peroxisomal lipid metabolism.

To prioritize potential hits for functional studies, we further examined the 21 proteins designated as peroxisomal in the STRING G.O. analysis of the WT centroid distance analysis hits (Table 1) (Fig. 7). The majority of these proteins show high connectivity to each other (Fig. 7A). Biological process enrichment analysis (Fig. 7B) revealed proteins involved in peroxisomal lipid pathways including ether lipid synthesis (GNPAT and FAR1) and fatty acid β-oxidation (ACOX3, CROT, SCP2, and ABCD3). The synthesis of plasmalogens, a subset of ether lipids, catalyzed by GNPAT, has previously been reported to support the infectivity of herpesvirus[8]; whether it also supports the infectivity of HIV-1 remains to be tested.

Discussion:

We have described here the first organellar map of the cell following HIV expression. From these data, we can begin to picture how HIV alters the cell on an organellar level along with the proteins it manipulates to do so. Our central finding is that the expression of HIV-1 alters the fractionation-behavior of peroxisomes in a Nef-dependent manner.

The model of inducible HIV in Jurkat T cells used here has advantages and disadvantages. One advantage is that the system provides a highly homogenous population of HIV-expressing cells suitable for mass spectrometric analysis[19]. A homogenous population is particularly important in spatial proteomic studies as mixed populations of cells might yield erroneous classifications of proteins due to mixtures of different states[36]. Another advantage is scalability. These experiments required just over $3x10^8$ cells for each technical replicate, or over $1x10^9$ cells for a single biological replicate, to be induced. In our initial attempts with fewer cells, centrifugation at higher RCF (110,000x*g* and 195,500x*g*) yielded insufficient protein mass for quality control and mass spectrometry (data not shown). This highlights an advantage of using this T cell line compared to using primary CD4+ T-cells[37], which would require at least $2x10^9$ cells and ~50% productive infection. A disadvantage of using this T cell system is that the cytoplasmic volume of the cell is relatively small (apparent in the images of Figure 6, which show that most of the cell is occupied by the nucleus). We required an order of magnitude more cells for each technical replicate here than were used in the D.O.M. studies of Itzhak et al., who used HeLa cells with larger cytoplasm. Another disadvantage is that different clonal cell lines were needed to compare different viral genotypes, introducing the possibility that clonal variation influences the results. Mitigating this, we compared the movements of proteins between uninduced and induced states in the case of each cell line: one that expressed the wild-type and one that expressed the Nef-negative viral genome. Nonetheless, clonal variation rather than the expression of Nef could in principle underlie the different response of peroxisomes to HIV-expression in the two cell lines.

Additional technical as well as analytic aspects of our data and results are important to consider. One aspect is the fractionation method. The use of differential centrifugation here and by Itzhak et al. requires the downstream analysis of fewer fractions than gradient fractionation methods[2]. On the other hand, gradient fractionation methods seem to show increased resolution of protein classification[38]. Here, we used additional high-speed centrifugation steps to those described in the D.O.M. method of Itzhak et

al., but whether this yielded improved resolution of proteins is not certain. Additional questions concern the analytical pipeline itself. Spatial proteomics remains a relatively nascent subfield of proteomics and standardized analytical procedures do not yet exist[2,6,7,36,38,39]. We used both an SVM classifier as well as centroid-movement in 7-dimensional space to analyze our data and detect the movement of proteins. Importantly, only the latter can detect changes at the level of whole organelles relative to each other, e.g. the “movement” of peroxisomes relative to other organelles shown here. A difficulty for any organellar classification pipeline is how to classify proteins found in more than one organelle. As noted above, this can lead to erroneous classifications when such proteins are “forced” into a single classification/organellar location[36].

The inconsistency in the classification and movement of known Nef-targets is particularly vexing when considering the high reproducibility of classification across the study as a whole. While some of this inconsistency might be due to the aforementioned “forced-single-classification” problem, we also suspect that similarities of the fractionation properties of the organelles between which Nef-targets move is likely an additional source of error.

Our PCA and centroid movement analyses suggest that the expression of HIV-1 causes cellular proteins to change their behavior not only through organellar translocations but also through other cryptic means that involve an entire organelle - peroxisomes. This change in peroxisomes was also apparent when comparing the relative abundance of peroxisomal markers in the different fractions (see Figure 2C): upon induction of WT, the peak abundance of peroxisomal proteins shifted from the 12,200x*g* fraction to the 78,400x*g* fraction, a shift that was less prominent after the induction of ΔNef. Apparently, HIV-expression has changed peroxisomes such that they require higher *g*-force to pellet. We note that changes in analytic space might or might not correspond to changes in real space within the cell. For an individual protein moving from one organelle to another, a shift in analytic space likely corresponds to an actual shift in intracellular location. On the other hand, shifts of an entire organelle in analytic space might reflect solely that organelle’s fractionation behavior. We imagine that the fractionation behavior of an organelle depends on many factors and likely relies not only on protein content but potentially on lipid content, subtle changes in organelle size and shape, and water or solute content. Nonetheless, future studies that explore peroxisomes in isolation—proteomics and lipidomics of purified peroxisomes—seem

needed to understand the potential changes in peroxisome-composition following expression of HIV-1 proteins.

Beyond these technical considerations, what biological and virological role(s) do peroxisomes play in HIV replication and infection? A previous study using HeLa cells[40] showed a significant decrease in peroxisome number per cell due to Vpu, another HIV accessory protein. The authors suggested that this decrease in peroxisomes might attenuate intracellular innate immune-signaling, a role for peroxisomes described by the Kagan group[41,42]. In contrast, spatial proteomic study of herpesviruses-infection showed an opposite effect on peroxisomes: their numbers increased[8]. This increase in peroxisomes correlated with an increase in infectivity of both human cytomegalovirus (HCMV) and herpes simplex virus type 1, both members of the herpesvirus family, and a knockdown of peroxisomes decreased viral infectivity. In contrast, knockdown of peroxisomes had no effect on the infectivity of adenovirus, which unlike Herpes family viruses is not lipid-enveloped. To explain this, the authors found that HCMV increases the production of a specific lipid class, plasmalogens, that is required for optimal HCMV replication and is made in peroxisomes. Previous work in HIV-lipidomics indicated that plasmalogens are enriched in viral membranes relative to the plasma membrane, at least in the T cell line MT-4[43]. We are tempted to hypothesize based on the data herein and reviewed above that Nef-dependent modulation of peroxisomes contributes to viral infectivity, or that it modulates peroxisome-mediated innate immune-signaling. Notably, previous yeast two-hybrid studies found an interaction between Nef and the peroxisomal protein ACOT8[44–46], although ACOT8 is not among the modulated proteins we identified. Probing the proteins modulated herein should further elucidate Nef's influence on infectivity as well as the broader role of peroxisomes and lipid metabolism during HIV-1 infections.

Data availability:

All code used has been fully documented and annotated in Jupyter notebooks and can be found at the following Github repository along with all raw data: https://gitfront.io/r/oomaaron/eb829861bbedc190fe06b43ff0a5c6239191faba/InducibleHIV-Fractionation/. [Link to public GitHub repository will replace GitFront link following acceptance]

Mass spectrometry data (.RAW files and peptide identification tables) can be found on the ProteomeXchange database using project accession number PXD024716.

Supplemental data:

This article contains supplemental data.

Acknowledgements:

The authors would like to thank the NIH AIDS Reagent Program and The Pendleton Charitable Trust. The authors would like to also thank Hannah Carter and Max Qian for productive discussions on the computational analysis. This work was partially supported by a UC San Diego Center for AIDS Research (UCSD CFAR) Developmental Award, an NIH funded program (P30AI036214). A.LO. is supported by NIAID F31AI141111. C.A.S. is supported by a UCSD CFAR Developmental Award (P30AI036214). M.K.L. was supported by K08AI112394. A.R. and N.J.K. are funded by a grant to the HARC Center from the National Institutes of Health (P50AI150476 to N.J.K.). J.M.W. is supported by NIH/NIGMS Grants T32 GM007752 and NIH/NIAMS T32 AR064194. D.G. is supported by The Collaborative Center of Multiplexed Proteomics at UCSD. J.G. is supported by NIAID R01AI129706. All funding sources had no involvement in the design and execution of this study nor the final manuscript.

Author contributions:

A.L.O., C.A.S., M.K.L., and J.G. conceived the design and scope of this study. M.K.L. created the inducible cell lines used in this study. A.L.O. generated the peroxisomal eGFP cell lines and characterized all cell lines used. A.L.O. and C.A.S. performed the fractionation experiments. A.L.O., J.M.W., and D.G. performed pilot studies for the mass spectrometry analysis. A.R. and N.J.K. performed the mass spectrometry analysis. A.L.O. designed and performed the computational analysis. A.L.O. performed microscopy imaging and analysis for peroxisome confirmation studies. A.L.O. and J.G. wrote the manuscript and all authors reviewed and edited the manuscript.

Declaration of interests: none

1. Christoforou, A. *et al.* A draft map of the mouse pluripotent stem cell spatial proteome. *Nat. Commun.* **7**, 1–12 (2016).

2. Mulvey, C. M. *et al.* Using hyperLOPIT to perform high-resolution mapping of the spatial proteome. *Nat. Protoc.* **12**, 1110–1135 (2017).

3. Nightingale, D. J., Geladaki, A., Breckels, L. M., Oliver, S. G. & Lilley, K. S. The subcellular organisation of Saccharomyces cerevisiae. *Curr. Opin. Chem. Biol.* **48**, 86–95 (2019).

4. Borner, G. H. H. *et al.* Fractionation profiling: a fast and versatile approach for mapping vesicle proteomes and protein-protein interactions. *Mol. Biol. Cell* **25**, 3178–3194 (2014).

5. Itzhak, D. N., Tyanova, S., Cox, J. & Borner, G. H. H. Global, quantitative and dynamic mapping of protein subcellular localization. *Elife* **5**, 1–36 (2016).

6. Itzhak, D. N. *et al.* A Mass Spectrometry-Based Approach for Mapping Protein Subcellular Localization Reveals the Spatial Proteome of Mouse Primary Neurons. *Cell Rep.* **20**, 2706–2718 (2017).

7. Jean Beltran, P. M., Mathias, R. A. & Cristea, I. M. A Portrait of the Human Organelle Proteome In Space and Time during Cytomegalovirus Infection. *Cell Syst.* **3**, 361–373 (2016).

8. Jean Beltran, P. M. *et al.* Infection-Induced Peroxisome Biogenesis Is a Metabolic Strategy for Herpesvirus Replication. *Cell Host Microbe* **24**, 526-541.e7 (2018).

9. Guy, B., Rivière, Y., Dott, K., Regnault, A. & Kieny, M. P. Mutational analysis of the HIV nef protein. *Virology* **176**, 413–425 (1990).

10. Yu, G. & Felsted, R. L. Effect of myristoylation on p27nef subcellular distribution and suppression of HIV-LTR transcription. *Virology* **187**, 46–55 (1992).

11. Chowers, M. Y. *et al.* Optimal infectivity in vitro of human immunodeficiency virus type 1 requires an intact nef gene. *J. Virol.* **68**, 2906–14 (1994).

12. Pawlak, E. N., Dirk, B. S., Jacob, R. A., Johnson, A. L. & Dikeakos, J. D. The HIV-1 accessory proteins Nef and Vpu downregulate total and cell surface CD28 in CD4+ T cells. *Retrovirology* **15**, 6 (2018).

13. Thoulouze, M. I. *et al.* Human Immunodeficiency Virus Type-1 Infection Impairs the Formation of the Immunological Synapse. *Immunity* **24**, 547–561 (2006).

14. Sugden, S. M., Bego, M. G., Pham, T. N. Q. & Cohen, É. A. Remodeling of the host cell plasma membrane by HIV-1 nef and Vpu: A strategy to ensure viral fitness and persistence. *Viruses* **8**, 1–30 (2016).

15. Ramirez *et al.* Plasma Membrane-Associated Restriction Factors and Their Counteraction by HIV-1 Accessory Proteins. *Cells* **8**, 1020 (2019).

16. Rosa, A. *et al.* HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature* **526**, 212–7 (2015).

17. Usami, Y., Wu, Y. & Göttlinger, H. G. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature* **526**, 218–23 (2015).

18. Shi, J. *et al.* HIV-1 Nef antagonizes SERINC5 restriction by downregulation of SERINC5 via the endosome/lysosome system. *J. Virol.* JVI.00196-18 (2018) doi:10.1128/JVI.00196-18.

19. Lapek, J. D., Lewinski, M. K., Wozniak, J. M., Guatelli, J. & Gonzalez, D. J. Quantitative Temporal Viromics of an Inducible HIV-1 Model Yields Insight to Global Host Targets and Phospho-Dynamics Associated with Protein Vpr. *Mol. Cell. Proteomics* **16**, 1447–1461 (2017).

20. Wu, Y. *et al.* Potent enhancement of HIV-1 replication by nef in the absence of SERINC3 and SERINC5. *MBio* **10**, (2019).

21. Chesebro, B., Wehrly, K., Nishio, J. & Perryman, S. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. *J. Virol.* **66**, 6547–54 (1992).

22. Abacioglu, Y. H. *et al.* Epitope Mapping and Topology of Baculovirus-Expressed HIV-1 gp160 Determined with a Panel of Murine Monoclonal Antibodies. *AIDS Res. Hum. Retroviruses* **10**, 371–381 (1994).

23. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

24. Cox, J. *et al.* Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).

25. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science (80-.).* **356**, (2017).

26. Lee, A. J. *et al.* DAFi: A directed recursive data filtering and clustering approach for improving and interpreting data clustering identification of cell populations from polychromatic flow cytometry data. *Cytom. Part A* **93**, 597–610 (2018).

27. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

28. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).

29. Ollion, J., Cochennec, J., Loll, F., Escudé, C. & Boudier, T. TANGO: A generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **29**, 1840–1841 (2013).

30. Rhee, H. *et al.* Proteomic Mapping of Mitochondria. *Science (80-.).* **339**, 1328 (2013).

31. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: An updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251–D1257 (2016).

32. Matheson, N. J. *et al.* Cell Surface Proteomic Map of HIV Infection Reveals Antagonism of Amino Acid Metabolism by Vpu and Nef. *Cell Host Microbe* **18**, 409–423 (2015).

33. Lübke, T., Lobel, P. & Sleat, D. E. Proteomics of the lysosome. *Biochimica et Biophysica Acta - Molecular Cell Research* vol. 1793 625–635 (2009).

34. KAMINCHIK, J. *et al.* Cellular Distribution of HIV Type 1 Nef Protein: Identification of Domains in Nef Required for Association with Membrane and Detergent-Insoluble Cellular Matrix. *AIDS Res. Hum. Retroviruses* **10**, 1003–1010 (1994).

35. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

36. Gatto, L. *et al.* A Foundation for Reliable Spatial Proteomics Data Analysis. *Mol. Cell. Proteomics* **13**, 1937–1952 (2014).

37. Naamati, A. *et al.* Functional proteomic atlas of HIV infection in primary human CD4+ T cells. *Elife* **8**, (2019).

38. Gatto, L., Breckels, L. M. & Lilley, K. S. Assessing sub-cellular resolution in spatial proteomics experiments. *Curr. Opin. Chem. Biol.* **48**, 123–149 (2018).

39. Jadot, M. *et al.* Accounting for Protein Subcellular Localization: A Compartmental Map of the Rat Liver Proteome. *Mol. Cell. Proteomics* **16**, 194–212 (2017).

40. Xu, Z. *et al.* MicroRNAs upregulated during HIV infection target peroxisome biogenesis factors: Implications for virus biology, disease mechanisms and neuropathology. *PLoS Pathog.* **13**, 1–29 (2017).

41. Dixit, E. *et al.* Peroxisomes Are Signaling Platforms for Antiviral Innate Immunity. *Cell* **141**, 668–681 (2010).

42. Odendall, C. *et al.* Diverse intracellular pathogens activate type III interferon expression from peroxisomes. *Nat. Immunol.* **15**, 717–726 (2014).

43. Lorizate, M. *et al.* Comparative lipidomics analysis of HIV-1 particles and their producer cell membrane in different cell lines. *Cell. Microbiol.* **15**, 292–304 (2013).

44. Watanabe, H. *et al.* A novel acyl-CoA thioesterase enhances its enzymatic activity by direct binding with HIV Nef. *Biochem. Biophys. Res. Commun.* **238**, 234–239 (1997).

45. Liu, L. X. *et al.* Binding of HIV-1 nef to a novel thioesterase enzyme correlates with nef- mediated CD4 down-regulation. *J. Biol. Chem.* **272**, 13779–13785 (1997).

46. Cohen, G. B., Rangan, V. S., Chen, B. K., Smith, S. & Baltimore, D. The human thioesterase II protein binds to a site on HIV-1 Nef critical for CD4 down-regulation. *J. Biol. Chem.* **275**, 23097–23105 (2000).

Table 1: STRING-designated peroxisomal proteins from WT centroid distance analysis.

Gene	UniProt ID	WT Marker?	Uninduced Compartment	Induced Compartment
ABCD3	P28288	Y/Y	Peroxisome (6 replicates)	Peroxisome (6 replicates)
ACOX3	O15254	N/N	Peroxisome (5 replicates)	Peroxisome (6 replicates)
ACSL1	P33121	N/N	Inconsistent classification	Inconsistent classification
ACSL4	O60488	N/N	Large Protein Complex (3 replicates)	Large Protein Complex (3 replicates)
ACSL6	Q9UKU0	N/N	Unclassified (4 replicates)	Large Protein Complex (3 replicates)
CAT	P04040	Y/Y	Peroxisome (6 replicates)	Peroxisome (6 replicates)
CROT	Q9UKG9	N/N	Peroxisome (5 replicates)	Peroxisome (4 replicates)
DHRS4	Q9BTZ2	N/N	Inconsistent classification	Large Protein Complex (4 replicates)
FAR1	Q8WVX9	Y/Y	Peroxisome (6 replicates)	Peroxisome (6 replicates)
GNPAT	O15228	Y/Y	Peroxisome (6 replicates)	Peroxisome (6 replicates)
HACL1	Q9UJ83	Y/Y	Peroxisome (6 replicates)	Peroxisome (6 replicates)
HSDL2	Q6YN16	N/N	Large Protein Complex (3 replicates)	Large Protein Complex (5 replicates)
MAVS	Q7Z434	N/Y	Mitochondrion (6 replicates)	Mitochondrion (6 replicates)
NUDT7	P0C024	N/N	Peroxisome (5 replicates)	Peroxisome (5 replicates)
PEX1	O43933	N/Y	Peroxisome (6 replicates)	Peroxisome (3 replicates)
PEX3	P56589	Y/Y	Peroxisome (6 replicates)	Peroxisome (6 replicates)
PEX6	Q13608	N/Y	Peroxisome (6 replicates)	Peroxisome (3 replicates)
PNPLA8	Q9NP80	N/N	Large Protein Complex (3 replicates)	Large Protein Complex (6 replicates)
RAB8B	Q92930	N/N	Plasma membrane (3 replicates)	Plasma membrane (3 replicates)
SCP2	P22307	Y/Y	Peroxisome (6 replicates)	Peroxisome (6 replicates)
ZADH2	Q8N4Q0	N/N	Large Protein Complex (3 replicates)	Large Protein Complex (6 replicates)

The 21 proteins included here are the same as those in the STRING diagram in Fig. 7A. The “WT Marker?” column indicates whether a given protein was used as a marker in either biological replicate, e.g. “Y/N” means a protein was a marker in the first biological replicate but not the second. Those proteins with “inconsistent classification” for their compartment did not achieve a plurality of at least 3 technical replicates.

Figure 1: Inducible HIV-1 Jurkat cell lines yield a near pure population of HIV-expressing cells. A) Equal numbers of doxycycline-inducible wild-type and ΔNef HIV Jurkat cells were induced or left uninduced for 18 hours then fractionated by Dounce homogenization in a hypotonic lysis buffer. Cell homogenates were put through a differential centrifugation protocol, discarding the nuclear pellet (1,000x*g*) and lysing remaining pellets in 2.5% SDS buffer. Fractions were labeled for TMT-10 multiplexing and further offline HPLC fractionation. All multiplexes were run for 3 hours on LC-MS^3. B) Western blot showing induction of HIV p55, gp160, gp41, Nef, and Vpu with a GAPDH loading control. Cells were induced for 0, 4, 8, 12, 16, and 18 hours, lysed, then a portion of these cell lysates was run on 10% SDS-PAGE gels. C) Flow cytometry analysis of remaining sample from 1B. HIV-1 expression peaked at ~95% of cells p24+ by 18 hours.

Figure 2: Inducible Jurkat cells fractionate by differential centrifugation. A) Average percentage of total cellular protein detected in each fraction by BCA protein assay. Bars represent the mean value for a given fraction based on the average from each biological replicate. Error bars are one standard deviation. All BCA assays were performed in technical triplicate on 10-fold dilutions for each biological replicate. B) Western blots for cell fractions of inducible wild-type HIV Jurkat cells (left) and ΔNef HIV Jurkat cells (right), 18 hours post-induction. Blots shown are representative of both biological replicates. C) Parallel coordinate plots for marker proteins from wild-type and ΔNef uninduced and induced fractions. The average scaled abundances for marker proteins held in common across all technical replicates of the indicated condition are plotted by organelle.

Figure 3: Doxycycline induction condition and virus minimally affected protein classification. A) Analytical pipeline for classifying proteins from mass spectrometry data of fractionated cells. First, well-behaved marker proteins were selected for each biological replicate using local outlier factor outlier detection. Next, a Support Vector Machine (SVM) classifier was fitted on training and test subsets, with those parameters inputted in a 500-member SVM bagging classifier. The cross-validation and classifying process was iterated starting with all proteins, then with classified mitochondrial proteins removed, and finally with classified large protein complex proteins removed. Lastly, the number of SVM classifiers in the

plurality of voters for a given protein was used as a pseudo-confidence value to revert low confidence assignments to “unclassified.” B) Classification concordance of proteins detected in all six uninduced or induced wild-type technical replicates; numbers in parentheses indicate the number of proteins considered. C) Classification concordance of proteins detected in all six uninduced or induced ΔNef NL4-3 technical replicates; numbers in parentheses indicate the number of proteins considered. D) Average organellar distribution of proteins by virus type and induction condition. Each bar is the average of six technical replicates.

Figure 4: Classification consistency of select HIV targets and HIV proteins. Proportion of technical replicates classified as the indicated compartment for the denoted viral genome and condition for A) SERINC5, B) Lck, and C) CD28. Proportion of induced technical replicates classified as the indicated compartment for the denoted viral genome for D) Nef and E) Gag.

Figure 5: Principal component analysis plots of wild-type and ΔNef HIV Jurkat cells reveal a shift in peroxisome association. Points are colored by support vector machine ensemble classification results. Plots are representative of six technical replicates for uninduced and induced cells. All plots have been fit to the principal components of the wild-type uninduced data to allow for direct visual comparison. Insets on each plot highlight peroxisomal proteins in the same plot range.

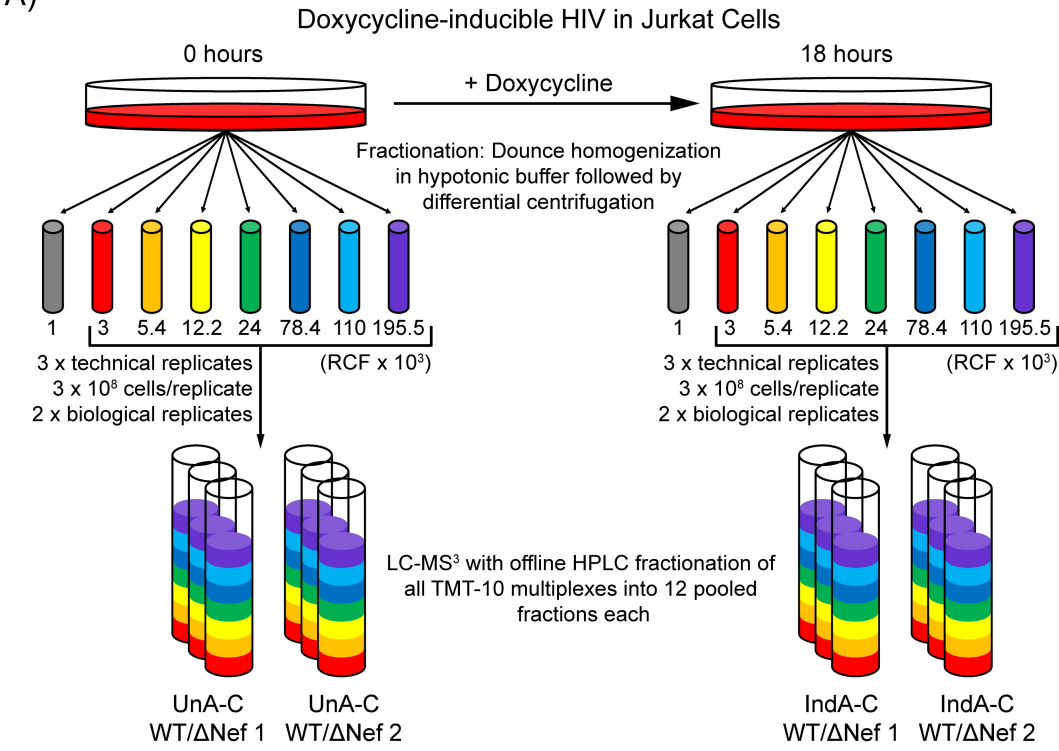
Figure 6: Spatial proteomic data reveals Nef-dependent changes in peroxisomal proteins despite minimal changes in peroxisome morphology or number. A) Radar plots of pairwise Euclidean distances between organellar marker proteins for the indicated conditions. Average pairwise Euclidean distance is shown for uninduced (purple line) and induced (yellow line) marker proteins. B) Bar plots of enriched cellular component gene ontology (G.O.) terms from STRING networks of proteins generated by label-based movement analysis. Proteins included in STRING networks were those that were classified to a given organelle in at least four out of six uninduced technical replicates and were classified to a different organelle in at least four out of six induced technical replicates. G.O. terms are plotted using the negative logarithm of the STRING generated false detection rate (FDR) value. The number of hits for each analysis

is indicated in each plot's title and the number of proteins included in a specific G.O. term is indicated in parentheses for each term. C) Bar plots of enriched cellular component G.O. terms from STRING networks of proteins generated by centroid distance analysis. Proteins included in STRING networks were those that were concordant between individual biological replicates. D) Peroxisome-localized eGFP in uninduced and induced wild-type and ΔNef HIV-1 Jurkat cells. Representative images are shown. E) Characterization of peroxisomes using transduced peroxisomal eGFP in uninduced and induced wild-type and ΔNef HIV-1 Jurkat cells. Peroxisome number per cell was analyzed using Fiji using a minimum of 50 cells for each set of cells in each of three biological replicates (total, pooled number of cells is at least 200 cells/condition). Statistical significance is calculated using a two-sided, independent Student's t-test with Welch's correction for unequal variance.

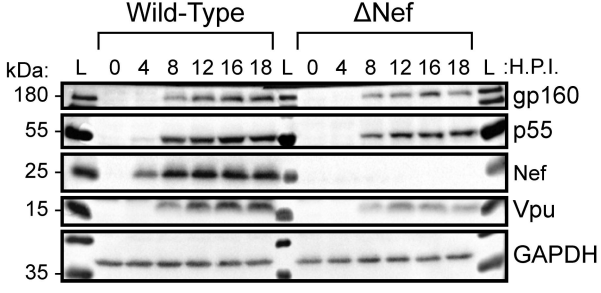
Figure 7: Peroxisomal protein "hits" from the centroid-based analysis of cells expressing wild-type HIV-1 are enriched in proteins involved in beta-oxidation and ether lipid biosynthesis. A) STRING diagram of 21 proteins labeled as "peroxisomal" by STRING in Fig. 6C for wild-type centroid distance analysis. Colored nodes correspond to ether lipid biosynthetic process (red), fatty acid beta-oxidation (blue), long-chain fatty-acyl-CoA metabolic process (green), and peroxisomal membrane transport (yellow). B) Select enriched biological process G.O. terms from STRING diagram in A).

Figure 1

A)


Doxycycline-inducible HIV in Jurkat Cells
0 hours
18 hours
+ Doxycycline
Fractionation: Dounce homogenization in hypotonic buffer followed by differential centrifugation
1 3 5.4 12.2 24 78.4 110 195.5
(RCF x 10³)
3 x technical replicates
3 x 10⁸ cells/replicate
2 x biological replicates
LC-MS³ with offline HPLC fractionation of all TMT-10 multiplexes into 12 pooled fractions each
UnA-C WT/ΔNef 1
UnA-C WT/ΔNef 2
IndA-C WT/ΔNef 1
IndA-C WT/ΔNef 2

B)


Wild-Type
ΔNef
kDa:
L 0 4 8 12 16 18 L 0 4 8 12 16 18 L :H.P.I.
180
55
25
15
35
gp160
p55
Nef
Vpu
GAPDH

C)

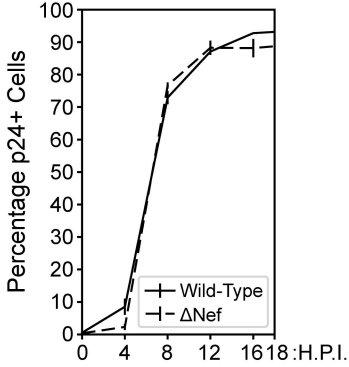

Percentage p24+ Cells
100
90
80
70
60
50
40
30
20
10
0
Wild-Type
ΔNef
0 4 8 12 16 18 :H.P.I.

Figure 2

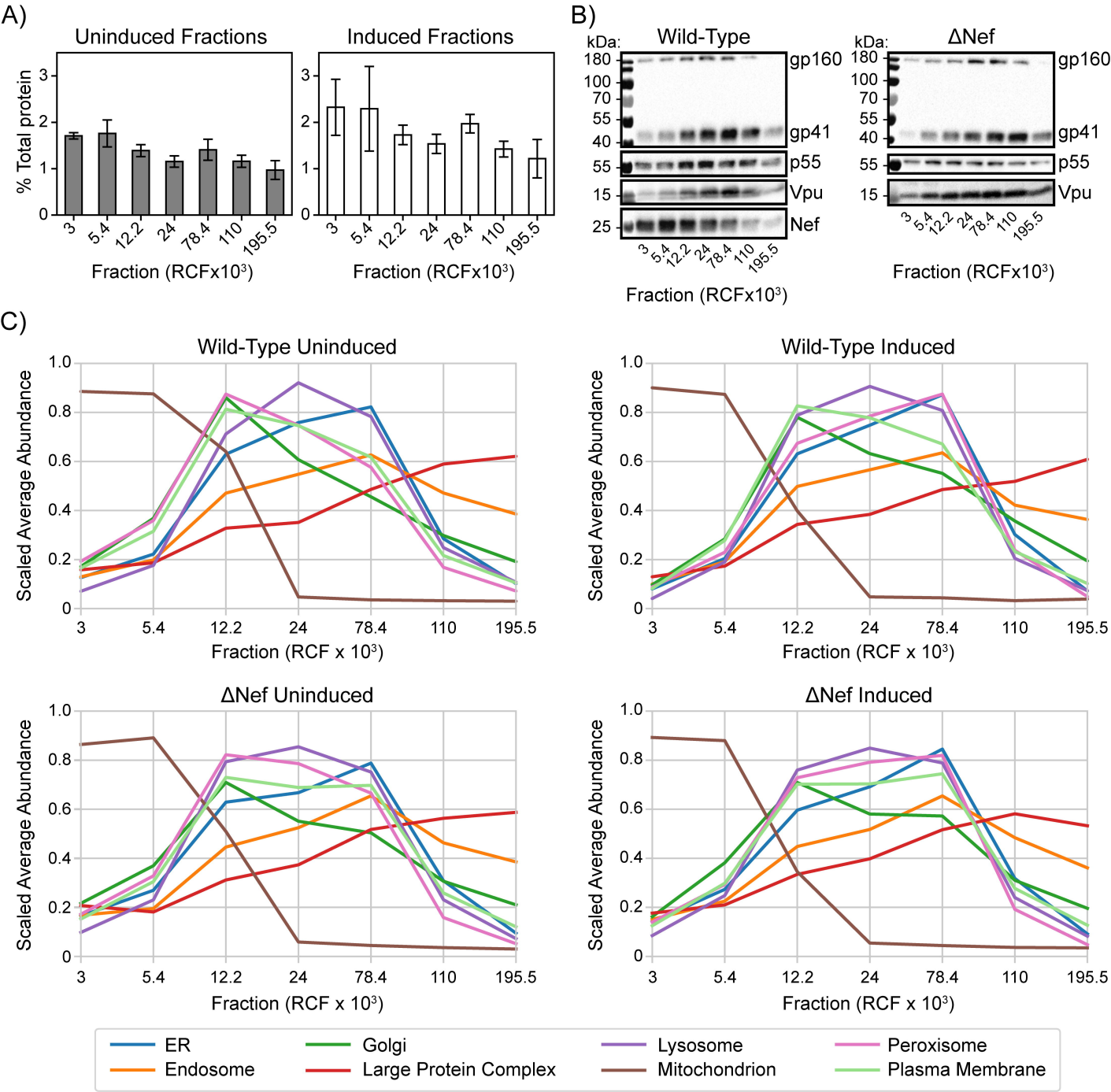
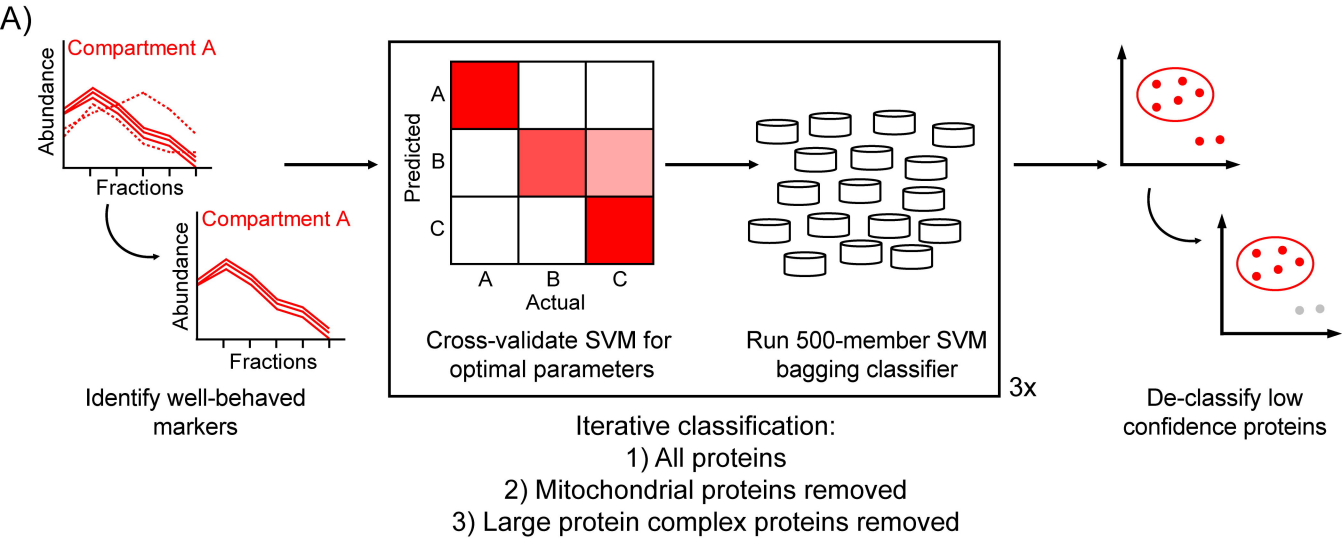
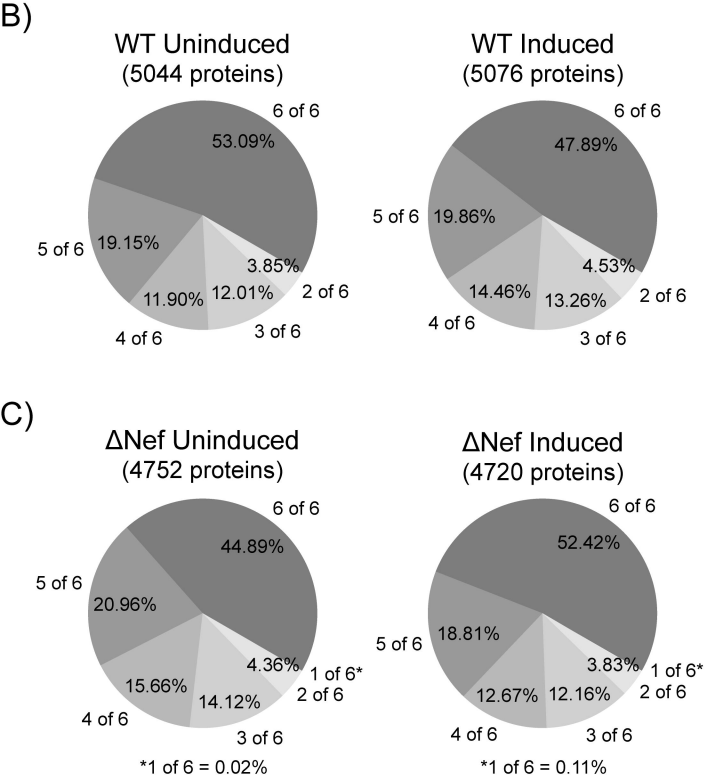

A)
Uninduced Fractions
Induced Fractions
% Total protein
Fraction (RCFx10³)
B)
Wild-Type
ΔNef
kDa:
gp160
gp41
p55
Vpu
Nef
Fraction (RCFx10³)
C)
Wild-Type Uninduced
Wild-Type Induced
ΔNef Uninduced
ΔNef Induced
Scaled Average Abundance
Fraction (RCF x 10³)
ER
Endosome
Golgi
Large Protein Complex
Lysosome
Mitochondrion
Peroxisome
Plasma Membrane

Figure 3


A)
Compartment A
Abundance
Fractions
Compartment A
Abundance
Fractions
Identify well-behaved markers
Predicted
A
B
C
A
B
C
Actual
Cross-validate SVM for optimal parameters
Run 500-member SVM bagging classifier
3x
De-classify low confidence proteins
Iterative classification:
1) All proteins
2) Mitochondrial proteins removed
3) Large protein complex proteins removed


B)
WT Uninduced
(5044 proteins)
6 of 6
53.09%
5 of 6
19.15%
3.85%
11.90%
12.01%
2 of 6
4 of 6
3 of 6
WT Induced
(5076 proteins)
6 of 6
47.89%
5 of 6
19.86%
4.53%
14.46%
13.26%
2 of 6
4 of 6
3 of 6
C)
ΔNef Uninduced
(4752 proteins)
6 of 6
44.89%
5 of 6
20.96%
4.36%
1 of 6*
15.66%
14.12%
2 of 6
4 of 6
3 of 6
*1 of 6 = 0.02%
ΔNef Induced
(4720 proteins)
6 of 6
52.42%
5 of 6
18.81%
3.83%
1 of 6*
12.67%
12.16%
2 of 6
4 of 6
3 of 6
*1 of 6 = 0.11%

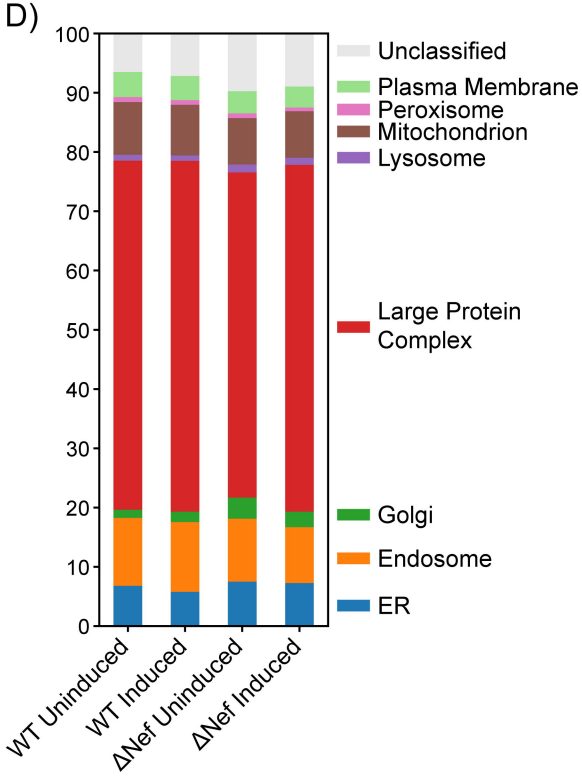

D)
100
90
80
70
60
50
40
30
20
10
0
Unclassified
Plasma Membrane
Peroxisome
Mitochondrion
Lysosome
Large Protein Complex
Golgi
Endosome
ER
WT Uninduced
WT Induced
ΔNef Uninduced
ΔNef Induced

Figure 4

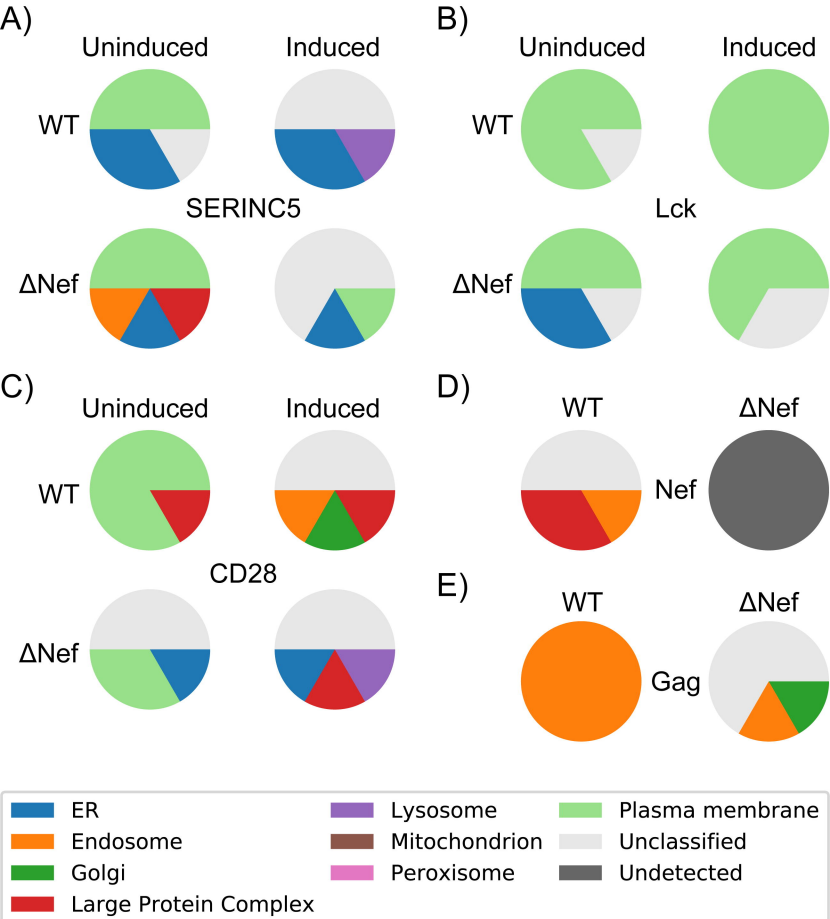

A)
Uninduced
Induced
WT
SERINC5
ΔNef
B)
Uninduced
Induced
WT
Lck
ΔNef
C)
Uninduced
Induced
WT
CD28
ΔNef
D)
WT
ΔNef
Nef
E)
WT
ΔNef
Gag
ER
Endosome
Golgi
Large Protein Complex
Lysosome
Mitochondrion
Peroxisome
Plasma membrane
Unclassified
Undetected

Figure 5

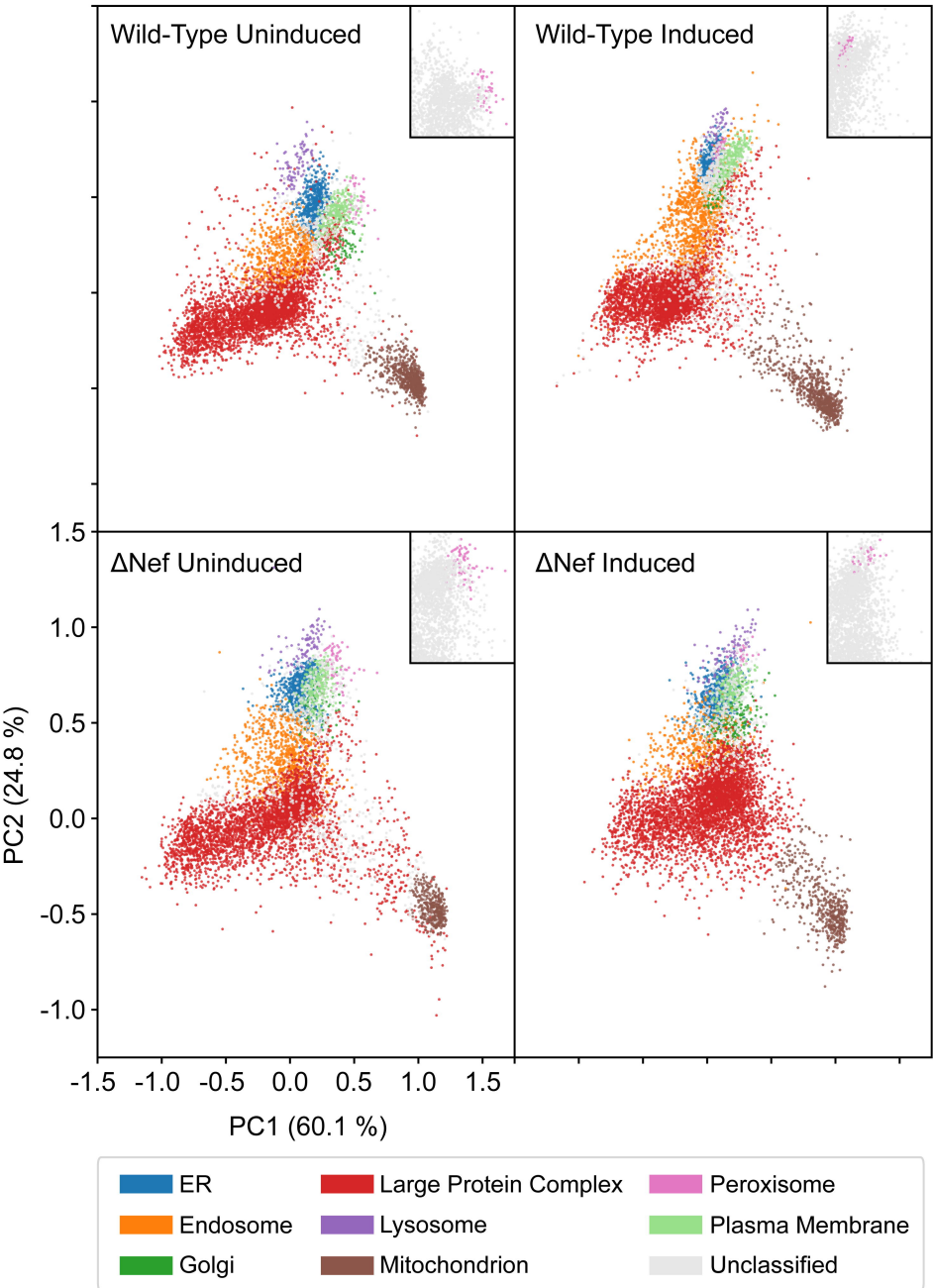

Wild-Type Uninduced
Wild-Type Induced
ΔNef Uninduced
ΔNef Induced
1.5
1.0
0.5
0.0
-0.5
-1.0
PC2 (24.8 %)
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
PC1 (60.1 %)
ER
Large Protein Complex
Peroxisome
Endosome
Lysosome
Plasma Membrane
Golgi
Mitochondrion
Unclassified

Figure 6

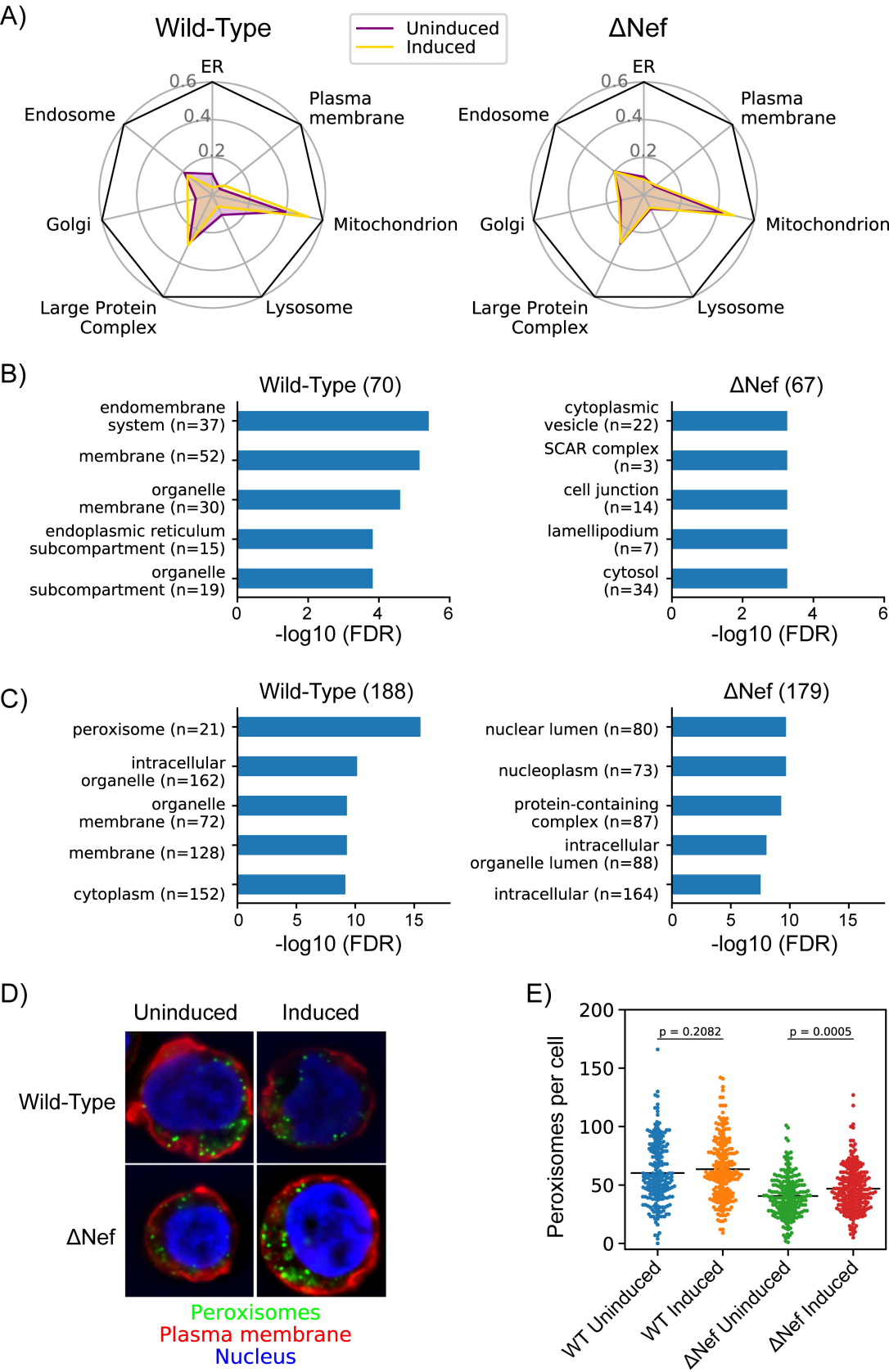
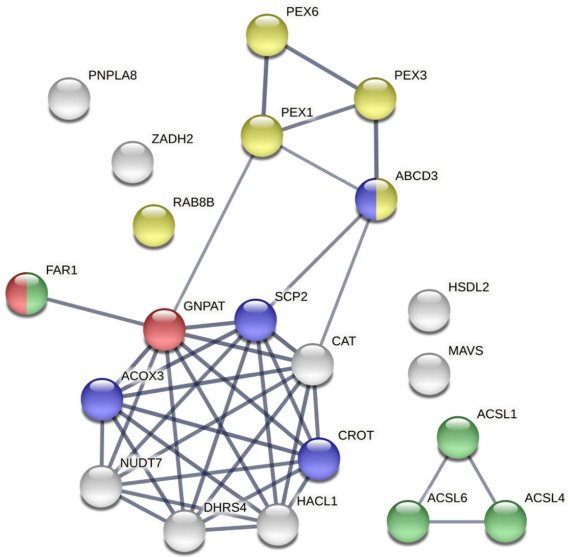
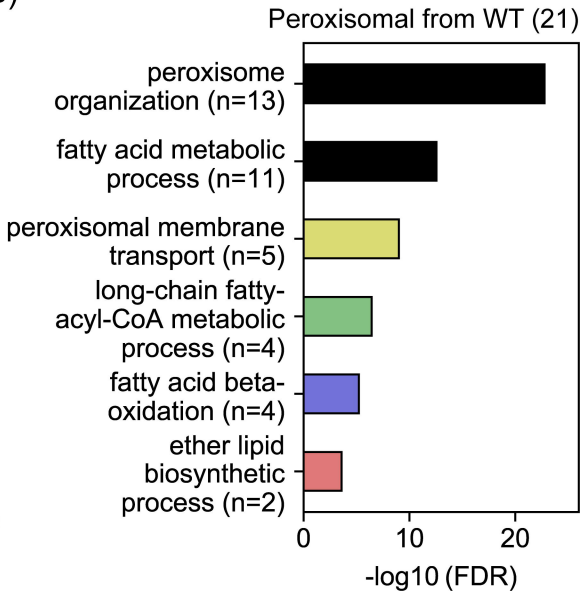

A)
Wild-Type
ΔNef
Uninduced
Induced
ER
Plasma membrane
Mitochondrion
Lysosome
Large Protein Complex
Golgi
Endosome
0.2
0.4
0.6
B)
Wild-Type (70)
endomembrane system (n=37)
membrane (n=52)
organelle membrane (n=30)
endoplasmic reticulum subcompartment (n=15)
organelle subcompartment (n=19)
ΔNef (67)
cytoplasmic vesicle (n=22)
SCAR complex (n=3)
cell junction (n=14)
lamellipodium (n=7)
cytosol (n=34)
-log10 (FDR)
C)
Wild-Type (188)
peroxisome (n=21)
intracellular organelle (n=162)
organelle membrane (n=72)
membrane (n=128)
cytoplasm (n=152)
ΔNef (179)
nuclear lumen (n=80)
nucleoplasm (n=73)
protein-containing complex (n=87)
intracellular organelle lumen (n=88)
intracellular (n=164)
D)
Uninduced
Induced
Wild-Type
ΔNef
Peroxisomes
Plasma membrane
Nucleus
E)
Peroxisomes per cell
p = 0.2082
p = 0.0005
WT Uninduced
WT Induced
ΔNef Uninduced
ΔNef Induced

Figure 7

A)


PEX6
PNPLA8
PEX3
PEX1
ZADH2
ABCD3
RAB8B
FAR1
GNPAT
SCP2
HSDL2
CAT
MAVS
ACOX3
ACSL1
CROT
NUDT7
HACL1
DHRS4
ACSL6
ACSL4

B)


Peroxisomal from WT (21)
peroxisome organization (n=13)
fatty acid metabolic process (n=11)
peroxisomal membrane transport (n=5)
long-chain fatty-acyl-CoA metabolic process (n=4)
fatty acid beta-oxidation (n=4)
ether lipid biosynthetic process (n=2)
0
10
20
-log10 (FDR)