Title

Integrated proteomics and transcriptomics analyses identify novel cell surface markers of HIV latency

Authors

Beliakova-Bethell, Nadejda
Manousopoulou, Antigoni
Deshmukh, Savitha
et al.

Publication Date

2022-08-01

DOI

10.1016/j.virol.2022.06.003

Peer reviewed

HHS Public Access

Author manuscript

Virology. Author manuscript; available in PMC 2023 August 16.

Published in final edited form as:

Virology. 2022 August ; 573: 50–58. doi:10.1016/j.virol.2022.06.003.

Integrated proteomics and transcriptomics analyses identify novel cell surface markers of HIV latency

Nadejda Beliakova-Bethell[a,b,3,*], **Antigoni Manousopoulou**[c,d,3,1], **Savitha Deshmukh**[b], **Amey Mukim**[b,2], **Douglas D. Richman**[a,b], **Spiros D. Garbis**[c,d,4,1], **Celsa A. Spina**[a,b,4]

[a]VA San Diego Healthcare System and Veterans Medical Research Foundation, San Diego, CA, USA

[b]University of California at San Diego, CA, USA

[c]Cancer Sciences Unit, Faculty of Medicine, University of Southampton, Southampton, UK

[d]Proteas Bioanalytics Inc., BioLabs at the Lundquist Institute, Torrance, CA, USA

Abstract

Elimination of the latent HIV cell reservoir may be possible, if the molecular identity of latently infected cells were fully elucidated. We conducted comprehensive molecular profiling, at the protein and RNA levels, of primary T cells latently infected with HIV *in vitro*. Isobaric labelling quantitative proteomics and RNA sequencing identified 1453 proteins and 618 genes, altered in latently infected cells compared to mock-infected controls ($p < 0.05$). Biomarker selection was based on results from integrated data analysis. Relative enrichment for latently infected cells was monitored using flow cytometric sorting and the HIV integrant assay. Antibodies against selected proteins, encoded by *CEACAM1* and *PLXNB2*, enabled enrichment of latently infected cells from cell mixtures by 3–10 fold (5.8 average, $p < 0.001$), comparable to levels obtained with biomarkers reported previously. Individual biomarkers are likely linked to subsets of latently infected cells, and an extended antibody panel will be required to inclusively target the latent HIV reservoir.

[*]Corresponding author. University of California San Diego Department of Medicine Stein, Clinical Research Building Room 330 9500 Gilman Drive, #0679, La Jolla, CA, 92093-0679, USA. nbeliakovabethell@health.ucsd.edu (N. Beliakova-Bethell).
[1]Current affiliation: Proteas Bioanalytics Inc., BioLabs at the Lundquist Institute, Torrance, CA, USA.
[2]Current affiliation: Beckman Coulter Life Sciences, Inc., Mississauga, ON, Canada.
[3]These authors contributed equally to the study.
[4]These authors contributed equally to project conceptualization and supervision.

CRediT authorship contribution statement

Nadejda Beliakova-Bethell: Conceptualization, Supervision, Data curation, Formal analysis, Visualization, Writing – original draft. **Antigoni Manousopoulou:** Investigation, Data curation, Formal analysis, Visualization. **Savitha Deshmukh:** Investigation, Validation, assisted preparing methods section of the manuscript. **Amey Mukim:** Investigation. **Douglas D. Richman:** Resources. **Spiros D. Garbis:** Conceptualization, Methodology, Project administration, writing – proteomics sections of the manuscript. **Celsa A. Spina:** Conceptualization, Methodology, Project administration, Funding acquisition, writing - assisted preparing the original manuscript, All authors participated in editing and approved the final manuscript.

Declaration of interests

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: SDG is the Founder, President and CEO/CTO of Proteas Bioanalytics Inc., BioLabs LA at The Lundquist Institute, Torrance, CA. DDR has consulted for Antiva Biosciences, Inc., Gilead Sciences, Inc., and Merck, Inc. The remaining authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2022.06.003.

Keywords

HIV latency; Biomarkers; Primary CD4$^+$ T cells; Proteome; Quantitative proteomics; Transcriptome; RNA-Seq; *CEACAM1* ; *PLXNB2*

1. Introduction

Human immunodeficiency virus-1 (HIV) establishes a stable reservoir in CD4$^+$ T cells (Chun et al., 1997b; Finzi et al., 1997, 1999; Siliciano et al., 2003; Wong et al., 1997), enabling it to persist in the face of combined modern antiretroviral therapy (ART) and the body's immune response. This latent reservoir, commonly defined as cells bearing quiescent proviruses that retain the capacity to produce infectious particles (Chun et al., 1997a; Eisele and Siliciano, 2012), remains the major obstacle to achieving a functional cure. Multiple mechanistic blocks exist for production of infectious virus, including blocks in RNA splicing, nuclear export and translation; therefore, complete viral quiescence is not requisite in the current definition of HIV latency - only the lack of infectious virion production (Pasternak and Berkhout, 2018). Moreover, the ability of many replication defective viruses to elicit host immune cell activation (Imamichi et al., 2016) has also caused definition of the latent reservoir (also termed "latent/persistent") to be expanded to include cells bearing transcription- and translation-competent provirus (Baxter et al., 2018).

If the molecular identity of this cellular reservoir of latent HIV infection could be fully elucidated, it may be possible to design rational strategies to selectively target such cells for elimination. As an example, proteins expressed on the surface of an infected cell represent suitable targets for antibody-bound immunotoxins (Rawlings et al., 2015). Alternatively, convertible CAR T cells (Herzig et al., 2019) can be designed to selectively attack cells, based on expression of several antigenic markers, without a need for proviral reactivation. In addition, reliable protein marker targets also can be used for accurate quantification *ex vivo* of latent cell reservoirs, without requiring HIV activation. Antibody reagents, raised specifically against such protein biomarkers, have the potential to create major impacts on clinical research in HIV-infected people on suppressive ART by enabling: 1) selective isolation and recovery of viable cells with latent HIV infection to study molecular mechanisms, which control viral latency and reactivation, and 2) evaluation of the efficacy of novel therapeutic interventions *in vivo* to reduce the size of latently infected cell reservoirs.

Previous research efforts have identified candidate biomarkers of latently infected cells, which facilitated various levels of target cell enrichment. Studies that compared transcriptomes of *in vitro* models of HIV latency and mock-infected cell controls proposed CD2 (Iglesias-Ussel et al., 2013) and Fc fragment of IgG receptor IIa (*FCGR2A*, also known as CD32a) (Darcis et al., 2020; Descours et al., 2017) as biomarkers of viral latency. Use of antibodies against lymphocyte activating 3 (*LAG3*), T cell immunoreceptor with Ig and ITIM domains (*TIGIT*), and programmed cell death 1 (PD-1) enriched latently infected cells from persons with HIV (Fromentin et al., 2016). A recent study, using a panel of 48 antibodies with cytometry by time of flight (CyTOF) analysis produced a biomarker panel

that was comprised of eight molecules (Neidleman et al., 2020) - a discovery consistent with the idea that the entire heterogenic HIV cell reservoir cannot be accurately defined by expression of a single antigenic protein. All of these proposed biomarkers facilitated various degrees of enrichment for latently infected cells, but were not able to capture all parts of the entire latent reservoir. Further identification of promising biomarkers should benefit greatly from the inclusion of combined comprehensive molecular characterization of latently infected cells.

Therefore, in the present study, we profiled the entire proteome of primary T cell samples taken from our *in vitro* model of HIV latency, using quantitative proteomics based isobaric stable-isotope labelling and high-definition liquid chromatography hyphenated with mass spectrometry. Compared to comprehensive profiling at the RNA level, which has been conducted previously by microarrays or RNA sequencing (RNA-Seq) (Descours et al., 2017; Iglesias-Ussel et al., 2013; Trypsteen et al., 2019; White et al., 2016), protein-based methods offer the benefit of a more direct transfer of results to antibody-based approaches for recovery of latently infected target cells. Moreover, the integrated analysis of proteome and transcriptome provides an additional level of biomarker validation. In the present study, we combined results front our quantitative proteomic profiling with previous RNA-Seq results from our published transcriptomic profiling study (Trypsteen et al., 2019) that used the same HIV latency model (Beliakova-Bethell et al., 2019; Spina et al., 2013; Trypsteen et al., 2019). Based on these analyses, we selected candidate antigenic marker proteins to screen for the capacity to identify, and enrich for latently infected cells.

2. Materials and methods

2.1. Study participants

Primary $CD4^+$ T cells were used in this study to establish HIV infection *in vitro*. Cells were isolated using negative selection with peripheral blood samples from HIV sero-negative donor volunteers. The protocol was approved by the Institutional Review Boards of the University of California San Diego, and the Veterans Affairs San Diego Healthcare System. All donors provided written informed consent.

2.2. In vitro primary T cell model of HIV latency

The primary T cell model of HIV latency selected for this study has some unique characteristics. Latent viral infection is established in cells of all the known major maturation phenotype subsets via the biologically relevant mechanism of cell-to-cell viral transmission, which may more accurately represent the potential mechanism(s) by which HIV latent infection develops *in vivo*. This *in vitro* primary T cell model of HIV latency has been described previously (Beliakova-Bethell et al., 2019; Spina et al., 2013). The approach generates paired samples of infected and mock-infected cells from the same blood donors that are used for each analytic comparison. Please refer to Supplementary Methods for details.

2.3. Quantitative proteomics sample processing

Paired sample aliquots (latently infected/mock infected) from 10 individual cell donors were recovered at the end of *in vitro* culture (HIV latency model), washed 4 times with phosphate buffered saline (PBS), snap frozen and stored at −80°C. For protein isolation, the frozen cell samples were dissolved in 0.5 M triethylammonium bicarbonate, 0.05% sodium dodecyl sulphate and subjected to pulsed probe sonication (Misonix, Farmingdale, NY, USA). Lysates were centrifuged (16,000 g, 10 min, 4°C), and supernatants were measured for protein content using infrared spectroscopy (Merck Millipore, Darmstadt, Germany). One hundred micrograms of protein per sample were reduced, alkylated using the iTRAQ reagent kit per manufacturer's instructions (Sciex, Inc., Framingham, MA, USA) and subjected to overnight digestion in the dark at room temperature using trypsin at a 1:25 ratio (Promega, Inc., Madison, WI, USA) (White et al., 2015). Peptides were analyzed using multi-dimensional liquid chromatography and tandem mass spectrometry as reported previously (Galanos et al., 2016; Giannogonas et al., 2016; Hanley et al., 2016; Manousopoulou et al., 2017; White et al., 2015).

2.4. Database searches and identification of differentially expressed proteins (DEPs)

Unprocessed raw files were submitted to Proteome Discoverer 1.4 for target decoy searching against the UniProtKB *Homo sapiens* database comprised of 20,159 entries (release date January 2015), allowing for up to two missed cleavages, a precursor mass tolerance of 10 ppm, a minimum peptide length of six and a maximum of two variable (one equal) modifications of: iTRAQ 8-plex (Y), oxidation (M), deamidation (N, Q), or phosphorylation (S, T, Y). Methylthio (C) and iTRAQ (K, Y and *N*-terminus) were set as fixed modifications. False discovery rate (FDR) at the peptide level was set at <0.05. Percent co-isolation to exclude peptides from quantitation was set at 50. Reporter ion ratios from unique peptides only were taken into consideration for the quantitation of the respective protein. The iTRAQ ratios of proteins were median-normalized and $\log_2$ transformed. A one-sample Student's T-Test was performed to identify DEPs between the *in vitro* model of HIV latency and mock-infected cells. In a discovery-driven approach, significance was set at nominal p-value < 0.05, and no fold change cut off was used for initial identification of DEPs. In adherence to the Paris Publication Guidelines for the analysis and documentation of peptide and protein identifications (http://www.mcponline.org/site/misc/ParisReport_Final.xhtml), only proteins adhering to the above mentioned criteria and additionally identified with at least two unique peptides were considered. The quantitative proteomic profiling data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD024014.

2.5. Transcriptomics dataset of differentially expressed genes (DEGs)

Identified DEPs were compared with DEGs identified using the same model of HIV latency (Trypsteen et al., 2019). This RNA-Seq dataset included results from four replicate experiments with paired samples of latently infected and mock-infected cells that were generated in the same manner as in the current quantitative proteomics study, using cells from different blood donors.

2.6. Gene ontology (GO) terms and pathway analysis

The Database for Annotation, Visualisation and Integrated Discovery (DAVID) v6.8 (Dennis et al., 2003) was used to determine cellular localization of identified DEPs. The list of all detected proteins was used as background, and the list of DEPs as gene list; GO terms in the category “cellular component” were used to map subcellular localization. MetaCore analysis (Clarivate Analytics, Philadelphia, PA, USA) was applied to the DEPs and DEGs to identify pathways that were over-represented in the HIV-infected samples *vs* the mock-infected controls. FDR p-values <0.05 were considered significant for pathway discovery, and FDR p-values <0.01 to identify common pathways between DEPs and DEGs.

2.7. Antibody selection for identification and enrichment of latently infected T cells

Antibodies were selected for screening, based on the identification of candidate cellular biomarkers of latent HIV infection through quantitative proteomics (DEPs) and RNA-Seq (DEGs) analyses. Searches of relevant immunologic databases were performed to determine commercial sources of available antibodies that targeted the top 20 candidate biomarkers (https://www.biocompare.com/Antibodies). A list of antibodies for testing was selected, based on reagent availability and suitability to flow cytometry analysis (Table 1). In the fluorescence-activated cell sorting (FACS) enrichment experiments, antibodies were used at concentrations recommended by the manufacturer. For details on screening, staining protocols, and flow cytometry please refer to Supplementary Methods.

2.8. Integrant HIV DNA assay

Cells from the mixed cell populations of our HIV latency model, generated using samples from independent blood donors, or cells from persons with HIV, were sorted live for DNA extractions. Cell samples, taken prior to and after each FACS experiment were collected for comparison of the percentages of latently infected cells that were present before and after enrichment. Cell counts supplied by the sorter were recorded for each recovered cell population, and the copies of integrated HIV DNA were quantified in both the original bulk and enriched cell aliquots, by droplet digital PCR (ddPCR) as described previously (Beliakova-Bethell et al., 2019; Koelsch et al., 2008). The fold enrichment for latently infected cells was determined by dividing the percent viral integrants found in the enriched cell subset by the percent integrants before enrichment. See Supplementary Methods for details.

2.9. Statistical analysis

Flow cytometry FCS raw data were processed using FlowJo version 10 software (FlowJo, LLC, Ashland, OR, USA). Percentages of cells expressing each marker and the proportions of cells with integrated HIV DNA were modeled using beta regression (*betareg* library (Cribari-Neto and Zeileis, 2010) in the Bioconductor tool repository). Graphs were constructed using GraphPad Prism software (GraphPad Software, La Jolla, CA, USA).

3. Results

3.1. Identification of differentially expressed proteins and gene transcripts

Use of our quantitative proteomics approach profiled a total of 10,886 proteins (peptide FDR $p < 0.05$) in primary T cell samples from our *in vitro* model of HIV latency. Of these, 673 were up- and 780 downregulated ($p < 0.05$). See Supplementary Table 1 for a complete list of identified DEPs. Sub-cellular localization of DEPs was determined using GO analysis in DAVID, which showed mapping of: 37.9% to cytoplasm, 36.8% to nucleus, 21.3% to extracellular exosome, and 19.0% to membrane (some proteins localized to more than one cellular compartment). The list of DEPs associated with membrane localization is presented in Supplementary Table 2.

Global transcriptomic analysis by RNA-Seq, using cells from our model of HIV latency, generated from a different set of blood donors, profiled 16,058 genes, which identified 618 DEGs between cells from the HIV latency model and the paired mock-infected cells (Trypsteen et al., 2019). Of these DEGs, 434 were protein-encoding genes. Twenty-one genes in this group exhibited the same modulation trends at both the RNA and protein levels (8 up- and 13 downregulated; color-coded, Fig. 1), and 6 proteins were classified by DAVID as members of the cellular compartment term “membrane” (Fig. 1).

3.2. Pathway enrichment analysis of differentially expressed genes and proteins

Pathway Map Analysis using MetaCore identified 410 pathways enriched for DEPs and 196 pathways enriched for DEGs (FDR $p < 0.05$). Complete lists of all pathways, significantly enriched for DEPs and DEGs, are presented in Supplementary Tables 3 and 4, respectively. Pathways, previously implicated in establishment and maintenance of HIV latency, were found in both datasets. For example, pathways related to p53 signaling that were enriched for DEPs included *DNA damage_p53 activation by DNA damage*, and *DNA damage_role of SUMO in p53 regulation*; and those enriched for DEGs included *suppression of p53 signaling in multiple myeloma, dual role of p53 in transcription deregulation in Huntington’s disease*, and *p53 signaling in prostate cancer* (Table 2). Another highly relevant example is mTOR signaling, with DEP enrichment seen in the pathways *signal transduction _mTORC2 upstream signaling* and *signal transduction _mTORC1 upstream signaling*, and DEGs enriched in *signal transduction _mTORC2 downstream signaling, signal transduction _mTORC2 upstream signaling* and *signal transduction _mTORC1 downstream signaling* pathways (Table 2). Both p53 and mTOR signaling pathways were reported previously to affect the establishment and maintenance of HIV latency (Besnard et al., 2016; White et al., 2016). Pathways describing other processes, such as Wnt and NOTCH signaling, overlapped perfectly between our DEP and DEG datasets (Table 2). Again, components of both Wnt and NOTCH signaling pathways have been shown to contribute to inhibition of HIV replication and establishment of latency (reviewed in (Kulpa et al., 2013)). The combined results from our studies described herein, verify that our research approach can detect at both the RNA and protein levels the same cellular processes, which are known to be modulated in HIV latency.

3.3. Biomarker selection and validation by flow cytometry

The *in vitro* latency model that we employed in our experiments contains 5–20% cells with latent HIV infection (Beliakova-Bethell et al., 2019), dependent on the individual cell donor. Identification of proteins and genes associated with latent infection was population-based and dependent on protein/gene expression by a very small proportion of the total cells analyzed. Therefore, flow cytometry analysis was used to validate the selected biomarkers at the single cell level.

Using results from both quantitative proteomics and RNA-Seq analyses, twelve candidate biomarkers were chosen to test antibody specificity to enrich for latently infected cells (Fig. 2). The selection of protein biomarkers was based on: significance, consistency of detection among replicate experiments, and fold difference in expression between the mock-infected and virus-infected $CD4^+$ T cell samples, which were generated from our model of HIV latency. Fold difference of expression was set at top 5% absolute fold change for either protein or gene expression. Overlap between the quantitative proteomics and RNA-Seq results was also taken into account. Proteins that were predicted to be localized to the plasma membrane and upregulated in the latently infected cell samples, compared to the mock-infected paired cell controls, were selected preferentially. Because of the limitations of quantitative proteomics analysis in proteins exclusively expressed in a single sample type (e.g. protein ratios not derived due to comparisons with zero), RNA-Seq data were used to identify such candidate genes (Fig. 2). Throughout the manuscript, we refer to the selected markers by the official symbols for genes that encode these proteins.

Out of the twelve antibodies that were selected for testing by flow cytometry, eight were able to detect their target proteins in the matched positive control cell lines, while four were not (Fig. 2). The eight antibodies that successfully detected antigen expression in the positive controls were then advanced to examine their reactivity with primary $CD4^+$ T cells taken front our model of HIV latency. Greater percentages of cells, positive for *CEACAM1, STX16*, and *PLXNB* expression, were consistently detected in samples from the latently infected cell cultures, compared with those from their paired mock-infected cell cultures (Supplementary Fig. 1). Average fold changes were: *CEACAM1*, 1.2, $p < 0.001$; *STX16*, 1.8, $p < 0.001$; *PLXNB2*, 1.2, $p < 0.001$. Thus, the upregulation of these three proteins, observed by quantitative proteomics and RNA-Seq analyses, was confirmed at a single-cell level by flow cytometric analysis. In contrast, differential expression of *LGALS1* (average fold change −1.3, $p = 0.197$) and *EPHA2* (average fold change 1.05, $p = 0.907$) proteins between latently infected and mock-infected cell samples was found to be inconsistent across replicate flow cytometry experiments (Supplementary Fig. 2). Interestingly, with the use of flow cytometric detection, both *MX1* (average fold change −1.25, $p < 0.001$) and *CD80* (average fold change −1.3, $p < 0.001$) were consistently expressed on a smaller percentage of cells in latently infected samples (Supplementary Fig. 2), despite being found to be upregulated in quantitative proteomics experiments or RNA-Seq. *LIMK1* expression was detected on average in a greater percentage of cells from latently infected samples (fold change 5, $p = 0.042$); however, the expression was not consistent across experiments (Supplementary Fig. 2) and contrasted with its observed downregulation in the quantitative proteomics study.

Because increases in the percentage of latently infected cells that expressed each individual marker were very modest, we next assessed the capacity of two-by-two antibody combinations to detect differential expression of candidate biomarkers (Fig. 2). We aimed to determine whether such combinations could enhance discrimination of a positive subset selectively in the latently infected cultures and thereby increase success of cell sorting experiments to enrich isolation and recovery of latently infected T cells. Of the combinations examined, dual staining for *CEACAM1* and *PLXNB2* cell surface biomarkers was the most promising (Supplementary Fig. 3). This approach identified a unique subpopulation (*CEACAM1*$^{+bright}$*PLXNB2*$^{+}$) that was consistently increased in the infected cell samples, compared to the paired mock-infected cell samples (average fold change 2.1, $p < 0.001$, Fig. 3A and B).

3.4. Enrichment for cells with latent HIV infection

To test the premise that expression of *CEACAM1* and *PLXNB2* proteins could be used to selectively enrich for latently infected primary CD4^{+} T cells from a mixed cell population, a separate series of experiments were performed (N = 5), using our *in vitro* model of HIV latency to generate the latently infected cells. At the end of culture, a cell aliquot was removed and set aside for subsequent analyses. The remainder of cells were stained with antibodies against *CEACAM1* and *PLXNB2* in preparation for FACS. A live-cell gating region was designed (Fig. 3A) to collect cells that were positive for *PLXNB2*, or had high expression of *CEACAM1*, or expressed both *CEACAM1*$^{+bright}$*PLXNB2*$^{+}$. The isolated cells, recovered by FACS, and the bulk cell aliquot, reserved prior to sorting, were subsequently assayed by integrant DNA analysis. With five replicate experiments that used cells from different donors, HIV DNA copies were on average 5.8 fold higher in the sorted enriched subpopulation (±2.7 standard deviation, range 3–10 fold, $p < 0.001$, Fig. 3C).

In addition, preliminary experiments to use CD4^{+} T cells from persons with long-term ART-suppressed HIV infection exhibited similar levels of latent cell enrichment, using the same selected antibody reagents. Unfortunately, due to the limited numbers of recovered CD4^{+} T cells from three available biobanked CD4^{+} T cell samples (500,000 or less total cells recovered), we were able to quantify integrant DNA in only one sample that contained a relatively large HIV reservoir. Before enrichment using antibodies against *CEACAM1* and *PLXNB2*, the ddPCR-based integrant DNA assay detected 4,167 latently infected cells normalized to one million CD4^{+} T cells. Following *CEACAM1* and *PLXNB2* antibody staining and FACS, 42,424 latently infected cells per one million CD4^{+} T cells were measured, demonstrating 10-fold enrichment. These results were consistent with the observations made, using our *in vitro* model of HIV latency. Although integrated HIV DNA was undetectable in the additional two patient samples (<500,000 cells recovered), we proceeded with single cell RNA-Seq examination. When normalized to total cells sequenced, this approach detected a greater number of cells that expressed low levels of HIV RNA (2.7-fold and 6.0-fold), following enrichment using the *CEACAM1* and *PLXNB2* antibodies.

Taken together, these results demonstrate that combined use of quantitative proteomics and transcriptomics analyses is capable of defining cellular biomarkers, which can facilitate the

identification and selective enrichment of latently infected primary T cells to enable further potential mechanistic studies.

4. Discussion

The results from the present study constitute the most in-depth quantitative proteomic profiling to date of primary $CD4^+$ T cells with latent HIV infection, with 10,886 proteins detected (Supplementary Table 1). The expression levels for 21 proteins in HIV infected *vs* uninfected control $CD4^+$ T cell samples were confirmed at the RNA level, using global transcriptomic analysis (Fig. 1). Discrepant results between proteomics and transcriptomics analyses may be explained, in part, by discrepancies between expression of RNA and proteins at detectable levels (Dumaual et al., 2013). For example, some proteins may be regulated post-translationally via activity of other proteins modulated in latency. Changes in protein expression may also reflect transient changes in RNA expression at an earlier time point, which was not captured in our present design. On the other hand, RNA-Seq is a more sensitive method that was able to detect a greater number of genes (N = 16,058), compared to 10,886 proteins detected by quantitative proteomics. It is therefore plausible that some differences observed by RNA-Seq were not detectable at the protein level. For example, previously reported biomarkers of latency, such as PD-1, *LAG3* and *TIGIT* (Fromentin et al., 2016), were not detected at the protein level in our study. The levels of their expression on latently infected cells generated in our model without activation are likely too low to be detectable by proteomics profiling. Perhaps most relevant, proviral integration frequency in our primary T cell model has been shown to vary in concert with donor-to-donor biologic variation that has ranged from 5 to 20% in our prior studies (Beliakova-Bethell et al., 2019) and 2.4%–22.3% in the cell enrichment experiments of the current study (Fig. 3C). With such ranges in establishment of latent infection, the detection of differential expression in gene transcription or protein production within a mixed cell population would require relatively high expression on individual cells, carrying persistent HIV infection. Therefore, it follows that it would be more difficult to detect any differences in biomarker expression when the size of the infected latent/persistent subpopulation is small (e.g. 2% compared to 20%). Despite this type of limitation, conducting biomarker discovery at both the protein and RNA expression levels provides a higher degree of confidence that overlapping detection of potential targets will lead to more reliable candidates for study; even though, some true biomarkers that are identified by only one method may be missed.

Compared to prior proteomics studies that have used latently infected cells (Berro et al., 2007; Zhang et al., 2019), our study used a primary T cell model system as opposed to cell line models of HIV latency, and quantified the entire proteome. Thus, a greater number of proteins were profiled in a more biologically relevant setting. In these prior studies, membrane proteins were enriched in latently infected and parental cell lines, and DEPs were identified using two-dimensional electrophoresis followed by matrix assisted laser desorption ionization - time of flight (MALDI-TOF) (Berro et al., 2007) or liquid chromatography mass spectrometry analysis (Zhang et al., 2019) of individual protein spots. The two studies, which used ACH-1 and J-Lat cells, identified a limited number of proteins (17 and 13, respectively), and demonstrated no overlap between their findings. Among the combined 30 total proteins found in these cell lines studies, four were identified as DEPs

within our quantitative proteomics experiments: eukaryotic translation initiation factor 2D (*EIF2D*, also known as ligatin) (Berro et al., 2007); phosphomevalonate kinase (*PMVK*); capping actin protein, gelsolin like (*CAPG*); and GrpE like 1, mitochondrial (*GRPEL1*) (Zhang et al., 2019). Such variations between studies are likely attributable to differences in the technologies applied and cell types used. However, one interesting exception was noted. *CAPG* expression was downregulated in both the J-Lat cell line model (Zhang et al., 2019) and our primary $CD4^+$ T cell model of latency (reported here), as well as in resting $CD4^+$ T cells taken from persons with HIV infection and suppressed viremia, compared to uninfected controls (Zhang et al., 2019). Furthermore, we found some overlap in DEPs, identified by our current study, with those from an additional published report (Azzam et al., 2016) that was conducted using memory phenotype $CD4^+$ T cells from persons infected with HIV on long-term suppressive ART: myosin heavy chain 9 (*MYH9*), myosin light chain 6B (*MYL6B*), S100 calcium binding protein A9 (*S100A9*), and heterogeneous nuclear ribonucleoprotein U (*HNRNPU*). The combined results from this group of studies indicate that proteomics studies, conducted in different *in vitro* model systems do have the capability to produce similar results, as well as reflect accurately *in vivo* observations.

On the other hand, proteomics profiling alone is not sufficient to differentiate between the global effects of viral infection on all $CD4^+$ T cells (including uninfected ones) and the protein dysregulation specific to latently infected cells. Assuming a given protein is expressed on latently infected but not uninfected cells, expression of an upregulated DEP is the average between unchanged expression on uninfected cells and relatively high expression on latently infected ones. Alternatively, observed differences in expression may be the result of a bystander effect on cells that did not acquire latent HIV infection, but were exposed to virus in culture. Furthermore, proteomics profiling quantifies total cellular proteins, which may produce discrepant results with flow cytometric evaluations of protein expression on cell surface (e.g. CD80 in this study). In studies conducted *ex vivo* with cells from persons infected with HIV, it is difficult to uncouple the effects of ART from manifestations of latent HIV infection. Because of such limitations, it was important in our study to select candidate biomarkers (Fig. 2) and test their ability with flow cytometry to enrich for latently infected cells from mixed cell populations. Using our model of HIV latency, we found several markers that were expressed on a larger proportion of cells from the latently infected cell cultures, compared with the paired mock-infected control cell cultures (Fig. 3A and B and Supplementary Figs. 1-3). The most promising combination of markers, *CEACAM1* plus *PLXNB2*, facilitated 3–10 fold enrichment of latently infected cells by FACS in 5 independent experiments, using different cell donors to generate the *in vitro* model of latent infection (Fig. 3C) and in proof-of-concept experiments (N = 3) using cells from persons with HIV.

The observed range of enrichment achieved, both in cells from our *in vitro* model of HIV latency and *ex vivo* cells from persons with HIV infection, was comparable to the performance reported in previous publications for other biomarkers of HIV latency. For example, memory $CD4^+$ T cells that simultaneously expressed PD-1, *TIGIT* and *LAG3* in samples from persons with suppressed HIV infection were enriched for viral DNA up to 10 fold (median increase 8.15 fold; range, 4.92–9.59), when compared to total $CD4^+$ T cells (Fromentin et al., 2016). In another report, $CD4^+$ T cells that expressed high

levels of surface CD2 antigen in persons with HIV infection, compared with total $CD4^+$ T cells, contained enriched amounts of HIV DNA copies, with a median increase of 5.7 fold (range, 3- to 10.8 fold) (Iglesias-Ussel et al., 2013). Of note, the model of latency that identified CD2 as a biomarker differed from our model in the mechanism of latency establishment: activated infected cells that returned to quiescence (Iglesias-Ussel et al., 2013), *vs* cell-to-cell viral transmission from productively infected cells to resting cells (our latency model). While CD2 protein expression was detected at the protein level in our study, its expression in latently infected cell samples was not consistently higher than that of the paired mock-infected cells (average fold change 1.09, $p = 0.35$). Variation in experimental results is likely dependent on the different cell culture characteristics of each model used for biomarker discovery - in particular, the mechanisms of latency establishment. Furthermore, variations in the surface expression level of common T cell antigens (i.e. CD2, CD3, CD80) are often linked to parallel changes in cellular activation, differentiation, or maturation regardless of the infection status of the cell. As a relevant example, CD2 expression on the cell surface has been demonstrated to be higher on memory, compared to naïve $CD4^+$ T cells (He et al., 2021). Our own data are consistent with the idea that CD2 does not serve as a unique marker to identify HIV reservoir cells in a memory T cell population.

Most importantly, all biomarkers that are identified through *in vitro* experimentation require subsequent validation with *ex vivo* studies of latently infected cells from persons with persistent HIV infection. In this respect, we are encouraged by the limited preliminary test results obtained with the protein markers *CEACAM1* and *PLXNB2*, identified in our present study. To move forward, though, expanded additional validation testing will be needed when sufficient cell material can be assured in the future not to be a significant limiting factor. Despite these limitations, the results from our current work and the prior studies of others suggest strongly that each of the identified proposed biomarkers defines different subsets of latently infected T cells.

Based on these findings, it is very likely that an extended combined biomarker panel will be needed to capture a complete reservoir profile, represented by different functional cell types. Further studies to test different combinations of markers will be needed to identify a complex signature of HIV latency that can be used for accurate quantification of the majority of the cell reservoir, or to target it for elimination. It is probable that the combination of markers will include those expressed on latently infected cells and not on uninfected ones, as well as those expressed exclusively on uninfected cells (Neidleman et al., 2020). Our work provides a comprehensive molecular platform to facilitate further biomarker selection and testing using samples from persons with HIV, including peripheral blood and tissue compartments.

5. Conclusions

Proteins that are dysregulated during latency vary when identified by different methods and in different cell systems; however, some overlaps do exist between *in vitro* and *ex vivo* studies. Combined proteomics and transcriptomics approaches facilitate identification of candidate biomarkers of latency with higher confidence. Antibodies against selected biomarker proteins, identified in the present study, enriched for latently infected T cells,

to comparable levels achieved with other biomarkers proposed previously. The results from this study are consistent with the idea that each identified biomarker candidate defines only a subset of latently infected cells, and that an extended complex biomarker panel will be required to capture or target a majority of the latent HIV reservoir, containing several different cell types.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Dr. Maile Karris and Ms. Deedee Pacheco for arranging and providing peripheral blood samples from the healthy volunteer donors. We thank the San Diego CFAR [P30 AI036214] Flow Cytometry Core for support in conducting cell sorting and phenotyping analyses and Genomics and Sequencing Core for support with HIV DNA assays. We gratefully acknowledge the PRIDE team for the proteomics data processing and repository assistance. This work was supported by a grant from the National Institutes of Health R21 AI122389 to CAS, and the Janies B. Pendleton Charitable Trust. NBB was supported, in part, by a Career Developmental Award-II (IK2 BX002731) and Merit Review Award (1I01 BX005285) from the Office of Research and Development, Veterans Health Administration, and the National Institutes of Health R56 (AI157755). The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government. The sponsors of this research were not involved in the study design, collection or interpretation of the data, manuscript preparation, or the decision to submit the article for publication.

Data statement

Quantitative proteomic profiling data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD024014. Raw RNA-Seq data used for comparison with quantitative proteomics data are available at Gene Expression Omnibus (GEO) database (URL: https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE114883, and processed data are available in the Supplementary Materials for the original study (Trypsteen et al., 2019). All other relevant data are within the paper and its supplementary files.

References

Azzam S, Schlatzer D, Maxwell S, Li X, Bazdar D, Chen Y, Asaad R, Barnholtz-Sloan J, Chance MR, Sieg SF, 2016. Proteome and protein network analyses of memory T cells find altered translation and cell stress signaling in treated human immunodeficiency virus patients exhibiting poor CD4 recovery. Open Forum Infect. Dis 3.

Baxter AE, O'Doherty U, Kaufmann DE, 2018. Beyond the replication-competent HIV reservoir: transcription and translation-competent reservoirs. Retrovirology 15, 18–18. [PubMed: 29394935]

Beliakova-Bethell N, Mukim A, White CH, Deshmukh S, Abewe H, Richman DD, Spina CA, 2019. Histone deacetylase inhibitors induce complex host responses that contribute to differential potencies of these compounds in HIV reactivation. J. Biol. Chem 294, 5576–5589. [PubMed: 30745362]

Berro R, de la Fuente C, Klase Z, Kehn K, Parvin L, Pumfery A, Agbottah E, Vertes A, Nekhai S, Kashanchi F, 2007. Identifying the membrane proteome of HIV-1 latently infected cells. J. Biol. Chem 282, 8207–8218. [PubMed: 17237230]

Besnard E, Hakre S, Kampmann M, Lim HW, Hosmane NN, Martin A, Bassik MC, Verschueren E, Battivelli E, Chan J, Svensson JP, Gramatica A, Conrad RJ, Ott M, Greene WC, Krogan NJ, Siliciano RF, Weissman JS, Verdin E, 2016. The mTOR complex controls HIV latency. Cell Host Microbe 20, 785–797. [PubMed: 27978436]

Chun T-W, Carruth L, Finzi D, Shen X, DiGiuseppe JA, Taylor H, Hermankova M, Chadwick K, Margolick J, Quinn TC, Kuo Y-H, Brookmeyer R, Zeiger MA, Barditch-Crovo P, Siliciano RF, 1997a. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature 387, 183–188. [PubMed: 9144289]

Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JA, Baseler M, Lloyd AL, Nowak MA, Fauci AS, 1997b. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc. Natl. Acad. Sci. U. S. A 94, 13193–13197. [PubMed: 9371822]

Cribari-Neto F, Zeileis A, 2010. Beta regression in R. J. Stat. Software 34, 1–24.

Darcis G, Kootstra NA, Hooibrink B, van Montfort T, Maurer I, Groen K, Jurriaans S, Bakker M, van Lint C, Berkhout B, Pasternak AO, 2020. CD32(+) CD4(+) T cells are highly enriched for HIV DNA and can support transcriptional latency. Cell Rep. 30, 2284–2296 e2283. [PubMed: 32075737]

Dennis G Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA, 2003. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol. 4, P3. [PubMed: 12734009]

Descours B, Petitjean G, López-Zaragoza J-L, Bruel T, Raffel R, Psomas C, Reynes J, Lacabaratz C, Levy Y, Schwartz O, Lelievre JD, Benkirane M, 2017. CD32a is a marker of a CD4 T-cell HIV reservoir harbouring replication-competent proviruses. Nature 543, 564–567. [PubMed: 28297712]

Dumaual CM, Steere BA, Walls CD, Wang M, Zhang Z-Y, Randall SK, 2013. Integrated analysis of global mRNA and protein expression data in HEK293 cells overexpressing PRL-1. PLoS One 8 e72977–e72977. [PubMed: 24019887]

Eisele E, Siliciano RF, 2012. Redefining the viral reservoirs that prevent HIV-1 eradication. Immunity 37, 377–388. [PubMed: 22999944]

Finzi D, Blankson J, Siliciano JD, Margolick JB, Chadwick K, Pierson T, Smith K, Lisziewicz J, Lori F, Flexner C, Quinn TC, Chaisson RE, Rosenberg E, Walker B, Gange S, Gallant J, Siliciano RF, 1999. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nat. Med 5, 512–517. [PubMed: 10229227]

Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, Chadwick K, Margolick J, Brookmeyer R, Gallant J, Markowitz M, Ho DD, Richman DD, Siliciano RF, 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science 278, 1295–1300. [PubMed: 9360927]

Fromentin R, Bakeman W, Lawani MB, Khoury G, Hartogensis W, DaFonseca S, Killian M, Epling L, Hoh R, Sinclair E, Hecht FM, Bacchetti P, Deeks SG, Lewin SR, Sékaly R-P, Chomont N, 2016. CD4(+) T cells expressing PD-1, TIGIT and LAG-3 contribute to HIV persistence during ART. PLoS Pathog. 12, e1005761. [PubMed: 27415008]

Galanos P, Vougas K, Walter D, Polyzos A, Maya-Mendoza A, Haagensen EJ, Kokkalis A, Roumelioti F-M, Gagos S, Tzetis M, Canovas B, Igea A, Ahuja AK, Zellweger R, Havaki S, Kanavakis E, Kletsas D, Roninson IB, Garbis SD, Lopes M, Nebreda A, Thanos D, Blow JJ, Townsend P, Sørensen CS, Bartek J, Gorgoulis VG, 2016. Chronic p53-independent p21 expression causes genomic instability by deregulating replication licensing. Nat. Cell Biol 18, 777–789. [PubMed: 27323328]

Giannogonas P, Apostolou A, Manousopoulou A, Theocharis S, Macari SA, Psarras S, Garbis SD, Pothoulakis C, Karalis KP, 2016. Identification of a novel interaction between corticotropin releasing hormone (Crh) and macroautophagy. Sci. Rep 6, 23342–23342. [PubMed: 26987580]

Hanley CJ, Noble F, Ward M, Bullock M, Drifka C, Mellone M, Manousopoulou A, Johnston HE, Hayden A, Thirdborough S, Liu Y, Smith DM, Mellows T, Kao WJ, Garbis SD, Mirnezami A, Underwood TJ, Eliceiri KW, Thomas GJ, 2016. A subset of myofibroblastic cancer-associated fibroblasts regulate collagen fiber elongation, which is prognostic in multiple cancers. Oncotarget 7, 6159–6174. [PubMed: 26716418]

He S, Guo J, Fu Y, Spear M, Qin C, Fu S, Cui Z, Jin W, Xu X, Chen W, Shang H, Wu Y, 2021. Prestimulation of CD2 confers resistance to HIV-1 latent infection in blood resting CD4 T cells. iScience 24, 103305–103305. [PubMed: 34765923]

Herzig E, Kim KC, Packard TA, Vardi N, Schwarzer R, Gramatica A, Deeks SG, Williams SR, Landgraf K, Killeen N, Martin DW, Weinberger LS, Greene WC, 2019. Attacking latent HIV with convertible CAR-T Cells, a highly adaptable killing platform. Cell 179, 880–894 e810. [PubMed: 31668804]

Iglesias-Ussel M, Vandergeeten C, Marchionni L, Chomont N, Romerio F, 2013. High levels of CD2 expression identify HIV-1 latently infected resting memory CD4(+) T cells in virally suppressed subjects. J. Virol 87, 9148–9158. [PubMed: 23760244]

Imamichi H, Dewar RL, Adelsberger JW, Rehm CA, O'Doherty U, Paxinos EE, Fauci AS, Lane HC, 2016. Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. Proc. Natl. Acad. Sci. U. S. A 113, 8783–8788. [PubMed: 27432972]

Koelsch KK, Liu L, Haubrich R, May S, Havlir D, Günthard HF, Ignacio CC, Campos-Soto P, Little SJ, Shafer R, Robbins GK, D'Aquila RT, Kawano Y, Young K, Dao P, Spina CA, Richman DD, Wong JK, 2008. Dynamics of total, linear nonintegrated, and integrated HIV-1 DNA in vivo and in vitro. J. Infect. Dis 197, 411–419. [PubMed: 18248304]

Kulpa DA, Brehm JH, Fromentin R, Cooper A, Cooper C, Ahlers J, Chomont N, Sékaly R-P, 2013. The immunological synapse: the gateway to the HIV reservoir. Immunol. Rev 254, 305–325. [PubMed: 23772628]

Manousopoulou A, Gatherer M, Smith C, Nicoll JAR, Woelk CH, Johnson M, Kalaria R, Attems J, Garbis SD, Carare RO, 2017. Systems proteomic analysis reveals that clusterin and tissue inhibitor of metalloproteinases 3 increase in leptomeningeal arteries affected by cerebral amyloid angiopathy. Neuropathol. Appl. Neurobiol 43, 492–504. [PubMed: 27543695]

Neidleman J, Luo X, Frouard J, Xie G, Hsiao F, Ma T, Morcilla V, Lee A, Telwatte S, Thomas R, Tamaki W, Wheeler B, Hoh R, Somsouk M, Vohra P, Milush J, James KS, Archin NM, Hunt PW, Deeks SG, Yukl SA, Palmer S, Greene WC, Roan NR, 2020. Phenotypic analysis of the unstimulated in vivo HIV CD4 T cell reservoir. Elife 9, e60933. [PubMed: 32990219]

Pasternak AO, Berkhout B, 2018. What do we measure when we measure cell-associated HIV RNA. Retrovirology 15, 13–13. [PubMed: 29378657]

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Pérez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF, Ternent T, Brazma A, Vizcaíno JA, 2019. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 47, D442–D450. [PubMed: 30395289]

Rawlings SA, Alonzo F, Kozhaya L, Torres VJ, Unutmaz D, 2015. Elimination of HIV-1-infected primary T cell reservoirs in an in vitro model of latency. PLoS One 10, e0126917. [PubMed: 25993666]

Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, Kovacs C, Gange SJ, Siliciano RF, 2003. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. Nat. Med 9, 727–728. [PubMed: 12754504]

Spina CA, Anderson J, Archin NM, Bosque A, Chan J, Famiglietti M, Greene WC, Kashuba A, Lewin SR, Margolis DM, Mau M, Ruelas D, Saleh S, Shirakawa K, Siliciano RF, Singhania A, Soto PC, Terry VH, Verdin E, Woelk C, Wooden S, Xing S, Planelles V, 2013. An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients. PLoS Pathog. 9.

Trypsteen W, White CH, Mukim A, Spina CA, De Spiegelaere W, Lefever S, Planelles V, Bosque A, Woelk CH, Vandekerckhove L, Beliakova-Bethell N, 2019. Long non-coding RNAs and latent HIV – A search for novel targets for latency reversal. PLoS One 14, e0224879. [PubMed: 31710657]

White CH, Johnston HE, Moesker B, Manousopoulou A, Margolis DM, Richman DD, Spina CA, Garbis SD, Woelk CH, Beliakova-Bethell N, 2015. Mixed effects of suberoylanilide hydroxamic acid (SAHA) on the host transcriptome and proteome and their implications for HIV reactivation from latency. Antivir. Res 123, 78–85. [PubMed: 26343910]

White CH, Moesker B, Beliakova-Bethell N, Martins LJ, Spina CA, Margolis DM, Richman DD, Planelles V, Bosque A, Woelk CH, 2016. Transcriptomic analysis implicates the p53 signaling pathway in the establishment of HIV-1 latency in central memory CD4 T cells in an in vitro model. PLoS Pathog. 12, el006026.

Wong JK, Hezareh M, Gunthard HF, Havlir DV, Ignacio CC, Spina CA, Richman DD, 1997. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. Science 278, 1291–1295. [PubMed: 9360926]

Zhang Y, Shen Y, Yin L, Qi T, Jia X, Lu H, Zhang L, 2019. Plasma membrane proteomic profile discovers macrophage-capping protein related to latent HIV-1. Curr. HIV Res 17, 42–52. [PubMed: 31057110]

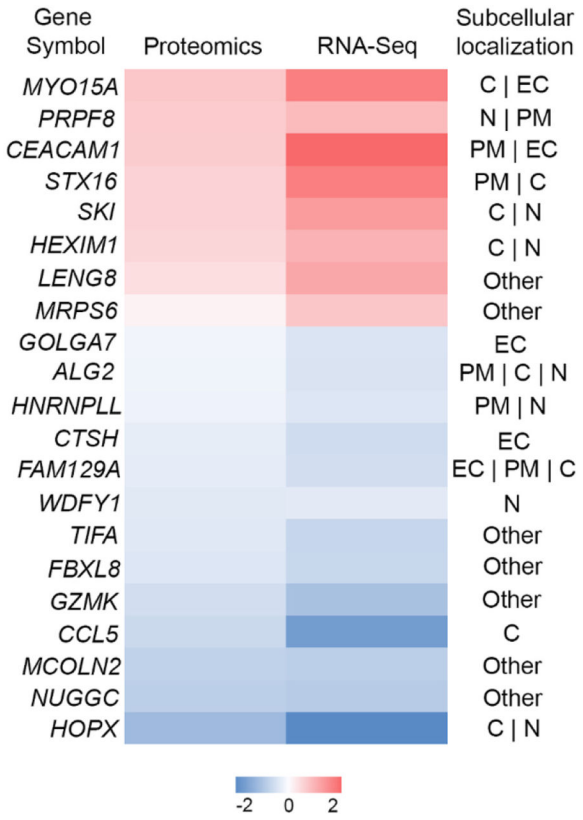

Fig. 1.

Differential gene expression in HIV latency at both protein (quantitative proteomics) and transcript levels (RNA-Seq), modulated in the same direction. Cellular compartment annotation in DAVID: C, cytoplasm; PM, plasma membrane; N, nucleus; EC, extracellular exosome; Other, not a member of listed compartments. Scale bar indicates $\log_2$ transformed fold changes in protein and transcript expression, latently infected over mock infected condition.

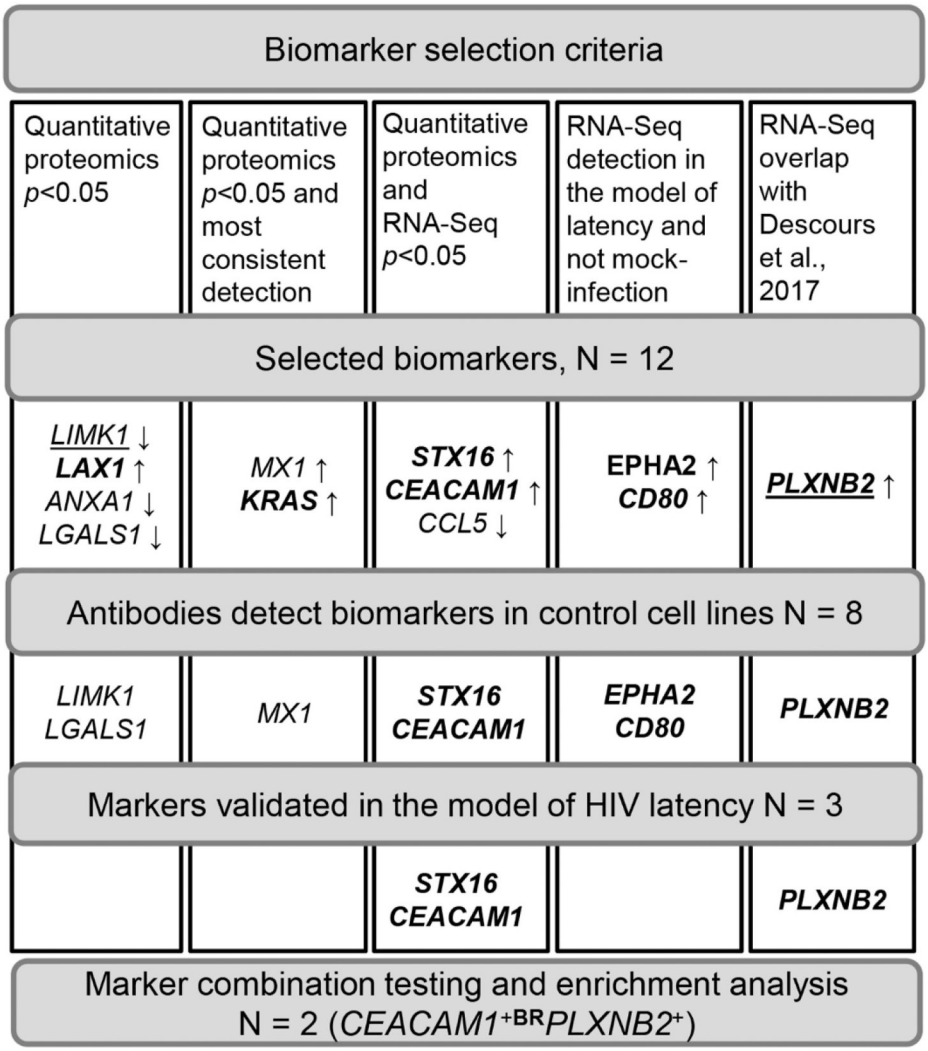

Fig. 2.

Flowchart of biomarker selection and validation. Biomarkers were selected based on significance ($p < 0.05$) in the quantitative proteomics and RNA-Seq experiments, consistency of protein detection among replicate experiments, large difference in expression level, and overlap observed between methods. *Bold*, proteins with predicted localization to the plasma membrane. *Underlined*, proteins/genes with large differences in expression between the latently infected cell samples and the paired mock-infected cell samples (top 5% of absolute fold change). Arrows indicate up- and downregulation in latency. *BR*, bright.

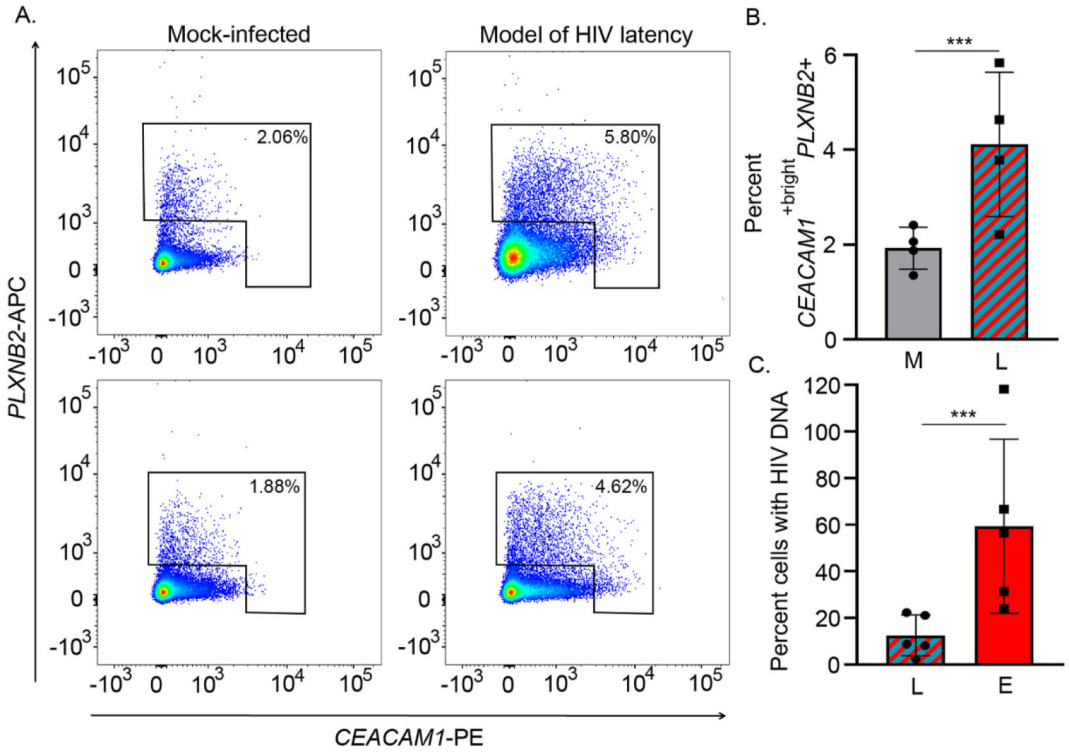

Fig. 3.

Use of combined dual antibody detection of *CEACAM1* and *PLXNB2* expression patterns identifies a unique cell subpopulation that is increased in latent infection and can be used to enrich for latently infected cells. A. Two representative experiments showing dot-plot images of marker expression with over-lay of gated analysis regions. Paired cell samples from mock-infected control and latently infected cell cultures in the primary $CD4^+$ T cell model of HIV latency are shown. *PE*, phycoerythrin; *APC*, allophycocyanin. B. Statistical analysis of the percentages of $CEACAM1^{+bright}PLXNB2^+$ cells within the gated regions in mock-infected (M) vs latently infected (L) samples, averaged over four independent experiments. C. In a separate set of experiments, integrated HIV DNA was quantified by droplet digital PCR in the infected cell samples from model of viral latency (L) and the parallel enriched cell subsets (E), following FACS isolation using antibodies against *CEACAM1* and *PLXNB2*. Data are the mean percentage of cells with HIV DNA from five independent experiments with different cell donors. In B and C, error bars indicate standard deviation. Beta regression was used to determine significance; ***, *p*-value < 0.001.

Table 1

Antibodies used to test enrichment of latently infected cells from mixed cell populations.

Antibody	Isotype	Secondary[a]	Conjugate	Manufacturer (cat. #)
CEACAM1	Mouse IgG2b	N/A	PE	R&D Systems (FAB2244P)
PLXNB1	Mouse IgG2a	N/A	APC	R&D Systems (FAB53291A)
STX16	Rabbit IgG	Goat anti-Rabbi	APC	Thermofisher Scientific (PA5-48340)
LIMK1	Mouse IgG2b	Goat anti-Mouse	PE	Novus Biologicals (NBP2-00748)
LAX1	Rabbit IgG	Goat anti-Rabbit	APC	Abcam (ab133759)
ANXA2	Rabbit IgG	N/A	PE	Abcam (ab210729)
LGALS1	Mouse IgG1	Goat anti-Mouse	PE	Thermofisher Scientific (43–7400)
EPHA2	Mouse IgG2a	N/A	APC	R&D Systems (FAB3035)
CD80	Mouse IgG1	N/A	FITC	R&D Systems (FAB140F-025)
CCL5	Mouse IgG1	N/A	APC	R&D Systems (IC278A)
MX1	Rabbit IgG	Goat anti-Rabbit	APC	Abcam (ab207414)
KRAS	Mouse IgG2b	Goat anti-Mouse	PE	Novus Biologicals (NBP2-59413)

CEACAM1, CEA cell adhesion molecule 1; *PLXNB2*, plexin B2; *STX16*, syntaxin 16; *LIMK1*, LIM domain kinase 1; *LAX1*, lymphocyte transmembrane adaptor 1; *ANXA2*, annexin A2; *LGALS1*, galectin 1; *EPHA2*, EPH receptor A2; *CD80*, CD80 molecule; *CCL5*, C—C motif chemokine ligand 5; *MX1*, MX dynamin like GTPase 1; *KRAS*, KRAS proto-oncogene, GTPase; PE, phycoerythrin; APC, allophycocyanin; FITC, fluorescein isothiocyanate; cat. #, catalog number; N/A, not applicable.

[a]Secondary antibody is not applicable for primary antibodies conjugated to fluorophores.

Table 2

Pathways significantly over-represented (FDR <0.01) in common for DEPs and DEGs.

Pathway	DEPs		DEGs	
	p-value	FDR	*p*-value	FDR
Signal transduction_Calcium-mediated signaling	1.512E-09	5.250E-07	1.143E-05	8.623E-04
Apoptosis and survival_IL-17-induced CIKS-independent signaling pathways	1.861E-06	1.315E-04	2.709E-04	6.525E-03
Pro-inflammatory action of Gastrin in gastric cancer	4.829E-06	2.484E-04	4.303E-04	7.857E-03
Chemotaxis_Lysophosphatidic acid signaling via GPCRs	2.255E-05	6.577E-04	2.376E-04	5.894E-03
Role of neuropeptides in pathogenesis of SCLC	2.348E-05	6.657E-04	3.427E-04	6.821E-03
Oxidative stress_ROS-induced cellular signaling	2.518E-05	6.725E-04	2.836E-04	6.554E-03
Development_Gastrin in cell growth and proliferation	4.946E-05	9.703E-04	2.795E-05	1.416E-03
Proliferative action of Gastrin in gastric cancer	5.176E-05	9.703E-04	5.918E-04	9.709E-03
Immune response_IL-3 signaling via JAK/STAT, p38, JNK and NF-kB	5.592E-05	1.009E-03	9.011E-05	3.125E-03
IGF family signaling in colorectal cancer	1.691E-04	2.195E-03	2.439E-06	3.451E-04
Reproduction_Gonadotropin-releasing hormone (GnRH) signalling	2.542E-04	3.044E-03	1.576E-06	2.549E-04
HBV signaling via protein kinases leading to HCC	3.343E-04	3.628E-03	3.895E-07	1.102E-04
Neurophysiological process_Constitutive and regulated NMDA receptor trafficking	3.517E-04	3.729E-03	2.837E-04	6.554E-03
Nociception_Nociceptin receptor signaling	3.739E-04	3.847E-03	1.786E-05	1.189E-03
Apoptosis and survival_Endoplasmic reticulum stress response pathway	4.149E-04	4.030E-03	1.099E-04	3.501E-03
Development_Keratinocyte differentiation	4.149E-04	4.030E-03	1.300E-05	9.199E-04
Canonical Notch signaling pathway in colorectal cancer	4.689E-04	4.461E-03	3.055E-04	6.651E-03
Development_Positive regulation of WNT/Beta-catenin signaling in the nucleus	5.976E-04	5.355E-03	8.010E-06	7.045E-04
Transcription_CREB signaling pathway	6.465E-04	5.648E-03	3.850E-04	7.263E-03
Inhibition of TGF-beta signaling in lung cancer	7.110E-04	6.029E-03	2.728E-05	1.416E-03
Development_WNT/Beta-catenin signaling in the nucleus	9.559E-04	7.336E-03	2.107E-04	5.555E-03
Development_NOTCH in inhibition of WNT/Beta-catenin-induced osteogenesis	9.998E-04	7.466E-03	1.621E-04	4.586E-03
Development_Positive regulation of WNT/Beta-catenin signaling in the cytoplasm	1.366E-03	9.134E-03	2.206E-10	2.497E-07
IGF signaling in lung cancer	1.374E-03	9.134E-03	2.110E-04	5.555E-03

DEPs, differentially expressed proteins; *DEGs*, differentially expressed genes; *FDR*, false discovery rate adjusted *p*-value.