Bioinformatics and HIV Latency

Angela Ciuffi · Pejman Mohammadi ·
Monica Golumbeanu · Julia di Iulio · Amalio Telenti

Abstract Despite effective treatment, HIV is not completely eliminated from the infected organism because of the existence of viral reservoirs. A major reservoir consists of infected resting CD4+ T cells, mostly of memory type, that persist over time due to the stable proviral insertion and a long cellular lifespan. Resting cells do not produce viral particles and are protected from viral-induced cytotoxicity or immune killing. However, these latently infected cells can be reactivated by stochastic events or by external stimuli. The present review focuses on novel genome-wide technologies applied to the study of integration, transcriptome, and proteome characteristics and their recent contribution to the understanding of HIV latency.

This article is part of the Topical Collection on *HIV Pathogenesis and Treatment*

A. Ciuffi (✉) · P. Mohammadi
Institute of Microbiology, University Hospital of Lausanne (CHUV), University of Lausanne, Bugnon 48, 1011 Lausanne, Switzerland
e-mail: angela.ciuffi@chuv.ch

M. Golumbeanu
Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland

M. Golumbeanu
Swiss Institute of Bioinformatics, Basel, Switzerland

J. di Iulio
Department of Genetics, Harvard Medical School, Boston, MA, USA

A. Telenti
Genetic Medicine, J. Craig Venter Institute, La Jolla, CA 92037, USA

Introduction

The study of biological processes, including viral infections, benefits from novel technologies for exploration and discovery. In the past decade, new developments have focused on genome-wide analyses, including high-throughput sequencing technologies for genome and transcriptome analyses, as well as mass spectrometry-based technologies for proteomic studies. These technologies aim at providing the most comprehensive snapshot picture of a specific cellular content, thereby generating large amounts of data. In turn, large-scale datasets require the development of tools and methods to handle and analyze these data adequately. This review focuses on experimental and bioinformatic methodologies applied to the topic of HIV latency.

HIV Latency

HIV infection is considered today a chronic disease, thanks to the efficacy of antiretroviral therapy (ART) [1, 2•, 3–7]. ART aims at blocking multiple viral replication steps and enzymatic activities, i.e., entry, reverse transcriptase, integrase, and protease, thereby limiting virus escape through selection of variants. ART results in undetectable viremia in the plasma of HIV+ individuals but fails to achieve complete virus eradication, hence requiring life-long treatment [1, 2•, 3–7]. HIV persistence is explained by the existence of reservoirs that are established early upon HIV exposure [8••, 9–11]. Two types of reservoirs have been described that are not mutually exclusive and that are likely to co-exist in some individuals. The first reservoir consists in anatomical sanctuaries, such as the central nervous system, gut, or lungs, where ART drug concentration may be suboptimal due to incomplete penetration [12–14]. Infection may persist in these sites due to either partial inhibition of viral replication or to cell-to-cell spread of

the virus [12, 13, 15]. The second reservoir relies on the persistence of HIV in long-lived cells in a latent and thus "hidden" state. This relies on HIV's ability to stably insert its genome in the host cell chromosome, thereby ensuring its life-long association with the infected host cell. This latent cell reservoir consists mostly of resting memory CD4+ T cells, but other cells have been shown to contribute as well, including naïve CD4+ T cells, macrophages, dendritic cells, and hematopoietic progenitor cells [16–22]. Latently infected CD4+ T cells can be induced, i.e., reactivated, and are thus able to produce particles through stochastic transcription or through immune activation, thereby leading to intermittent detectable viremia or viremia rebound after ART cessation [2•, 3, 23••].

The reservoir of latently infected CD4+ T cells is currently considered to be the major obstacle to HIV cure. Global scientific efforts have focused on infected CD4+ T cells in order to understand the nature of latency, i.e., nature of post-integration blocks hampering viral particle production, as well as designing new strategies to stimulate cells to exit latency. For this purpose, multiple models have been developed to characterize various mechanisms that can contribute to HIV latency [24, 25]. Models have also been useful to test reactivation compounds inducing viral particle production from the latently infected CD4+ T cells [26••, 27–29]. The present review focuses on recent genome-wide analyses performed on the models of HIV latency (Table 1).

The Inducible Reservoir

Available methods overestimate or underestimate the size of the latent viral reservoir. This is mainly due to (i) the use of blood samples as the source of the latent reservoir, (ii) the major presence of cells carrying defective viral genomes which obscures the proportion of latently infected cells, and (iii) the incomplete success of current methods used to reactivate latently infected cells in vitro. Indeed, it has been estimated that out of one million blood-purified resting CD4+ T cells from ART-treated individuals, on average, ~1000 cells (ranging approximately between 100 and 2000 cells) were infected by HIV and thus contained proviral DNA [30••]. However, only a small proportion of these HIV+ cells (11.7 %) carry replication-competent viral genome sequences and are thus inducible [31••]. Using phytohemagglutinin, interleukin-2, and irradiated peripheral blood mononuclear cells to stimulate CD4+ T cells in a viral outgrowth assay, only 1 % of HIV+ cells were successfully induced, thereby illustrating that current reactivation protocols and methods stimulate only a fraction of the total inducible reservoir [30••, 31••]. Other methods of stimulation using anti-CD3/CD28 and interleukin-7 for 7 days were able to induce particle production from 1.5 % of HIV+ cells (ranging from 0.6 to 2.4 % in the 13 patients tested) [23••]. Finally, reactivation of

latently infected cells may occur stochastically and spontaneously at a frequency of 0.041 % (0.03–0.15 %) [23••].

These studies indicate that the latency is complex as stimulation does not lead to 100 % reactivation of latently infected cells [26••, 31••]. Cillo et al. also show that, under their stimulation conditions, 7.5 % of infected cells were expressing cell-associated RNA, an almost 40× increase in HIV transcription compared to unstimulated resting CD4+ T cells. The gap between the successful viral particle production and viral transcription suggests that transcriptional latency cannot recapitulate all aspects of the viral reservoir and that other post-integration blocks exist, consistent with multiple recent studies using latency-reactivating agents and multiple models of HIV latency [23••, 26••, 32••].

Integration Site Location and HIV Latency

Reverse transcription of the viral RNA genome and its integration in the host cell chromosome are two hallmarks of retroviruses. In the past decade, thanks to the availability of the human genome sequence, many efforts have been focused on the understanding of the preferential site of HIV integration, as well as its consequences on the host cell, mostly regarding insertional mutagenesis [33–35]. HIV has been shown to integrate preferentially into active transcription units, with no preference for exons or introns, neither for orientation [34, 36, 37]. The site of integration may have consequences for both HIV transcription and host gene transcription because of chromatin arrangement and RNA interference [35, 38]. Therefore, the site of integration can affect the balance between viral transcriptional success and latency, as well as establish a balance between cell death and clonal expansion.

Analysis of integration site distribution commonly uses a three-step experimental approach: (i) DNA fragmentation (sonication, restriction enzymes), (ii) linker ligation, and (iii) PCR amplifications using primer annealing in the long terminal repeat (LTR) and primer annealing in the ligated linker [39–41]. The amplicons are sequenced using next-generation sequencing technologies, generating millions of short reads (ranging usually from ~20 bp to a few hundreds of bp). Virus sequences are removed and trimmed sequences are aligned to the human genome to locate the integration site. This method allows the efficient capture of the preferential pattern of integration site distribution according to the human reference genome annotations. However, it fails to quantify the frequency of integration sites occurring at the exact same location due to the bias caused by the PCR amplification step [42••]. This limitation was recently addressed using random shearing of the DNA or through the usage of random decamer primers tailed with a U5 sequence that allowed discriminating between two identical integration site locations originating from an

Table 1 Genome-wide analyses on HIV latency models

HIV latency models	Analysis type	Experimental approach	Computational tools/approaches	Reference
T cell lines	Integration	DNA digestion (AvrII/XbaI/SpeI or MseI) Linker ligation First PCR (LTR-linker) Nested PCR TOPO cloning and transformation Sanger sequencing	Trimming: custom Perl scripts Genome alignment: BLAT Integration site inclusion criteria: starting at junction with HIV LTR, mapping to the human genome reference sequence with >98 % identity Annotation databases: RefSeq (Acembly, GenScan, Unigene)	Lewinski et al. [51] (Sherrill-Mix et al. [42••])
	Proteome	Two-dimensional gel electrophoresis matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF)	Protein identification: ProFound, Prospector, and Mascot software	Berro et al. [95]
	Proteome	Solid-phase extraction of glycopeptides (SPEG) High-performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS)	Raw data quality evaluation: Trans-Proteomic Pipeline (TPP) Peptide evaluation: INTERACT and Peptide Prophet Label-free quantitation with MaxQuant identification of differentially expressed glycoproteins using a twofold cutoff ratio and a Student's t test with p value ≤0.05	Yang et al. [96]
Bcl-2-modified CD4+ T cell model	Integration	DNA digestion (PstI) Dilution and circularization by ligation Inverse PCR amplification Sequencing	Sequence alignment using the UCSC Bioinformatics Human Genome database Integration site inclusion criteria: HIV LTR sequence	Shan et al. [54] Sherrill-Mix et al. [42••]
CD4+ T cell co-culture with H80 feeder model	Transcriptome	Total RNA extraction Spike-in addition (ERCC Exfold) mRNA-Seq library preparation (TruSeq) mRNA sequencing (Illumina HiSeq 2000)	Trimming: Cutadapt (Illumina adapter removal) Quality filtering: prinseq (low quality and PHRED score reads filtering) Genome and transcriptome alignment: Tophat Differential expression analysis: DESeq Enrichment tests: Fisher's exact test Online visualization tool: LITCHi[1] (GuavaH[2])	Mohammadi et al. [32••]
Direct infection of resting CD4+ T cells	Integration	DNA digestion (Tsp509I or MseI) Linker ligation First PCR (LTR-linker) Nested PCR with barcoded primers containing sequencing adapters 454 sequencing	Integration analysis done by web-based tool InSiPiD[3] Trimming Genome alignment: BLAT, hg18 Integration site inclusion criteria: LTR sequence, mapping to the human genome reference sequence within the first 3 base pairs (bp) and a unique alignment with >98 % identity Annotation database: RefSeq Statistical analysis: one-tailed Student's t test, one-way ANOVA, logistic regression	Pace et al. [52] Sherrill-Mix et al. [42••]
	Integration	DNA digestion (MseI) Linker ligation First PCR (LTR-linker) Nested PCR with barcoded primers containing sequencing adapters Ion Torrent sequencing (Ion PGM 200 sequencing, Ion 316 chip)	Integration analysis done by web-based tool InSiPiD[3] Trimming: hiReadsProcessor R package Genome alignment: BLAT, hg19 Integration site inclusion criteria: LTR and primer sequences, mapping to the human genome reference sequence within the	Sherrill-Mix et al. [42••]

Table 1 (continued)

HIV latency models	Analysis type	Experimental approach	Computational tools/approaches	Reference
			first 3 bp and with 98 % identity, quality score >15 for sequences shorter than 24 bases Annotation databases: UCSC, Encode, RNA-Seq. ChIP-Seq Statistical analysis: *R*	
	Transcriptome	RNA extraction RNA amplification Hybridization to Illumina Human-Ref8 (v3) BeadChip (microarray) Chip scanning with Illumina BeadStation 500GX scanner using Illumina BeadStudio (version 3) software	Raw data processing and filtering using Bioconductor (R), Benjamini and Hochberg *p* value correction with a 0.05 false discovery rate Analysis: Limma package (R) Pathway enrichment analysis: Ingenuity Pathway Analysis (IPA) software	Evans et al. [78]
Patient-derived ex vivo model	Integration	DNA shearing (Covaris Adaptive Focused Acoustics) into 300–500-bp fragments End repair (Epicentre End-it DNA End Repair) dA addition (NEB dA-tailing kit) Linker ligation First PCR (LTR-linker) Nested PCR with N6-barcoded primers containing sequencing adapters Illumina sequencing: 2×150-bp paired-end (MiSeq) or 2×105-bp paired-end (HiSeq)	Trimming: custom Perl scripts Genome alignment: BLAT, hg19 Integration site inclusion criteria for read 1: LTR primer sequence, 5 last bp of the LTR sequence followed with >20-bp DNA sequence with an average quality score >20, mapping to the human genome reference sequence within the first 3 bp and with >95 % identity Integration site inclusion criteria for read 2: alignment to the human genome reference sequence on the opposite strand compared to read 1 and within 1 kb Integration sites are considered different if the breaking point differs from >3 bp Gene ontology analysis: GREAT Statistical analysis: Fisher's exact test	Maldarelli et al. [44•]
	Integration	Linear amplification using forward primer in HIV 3′ end (env or nef region) Second-strand synthesis using a random decamer primer tailed with HIV U5 sequence End repair by trimming (NEB exonuclease I first followed by MyTaq fill-in) Denaturation, annealing, and extension to generate a panhandle structure (LTR sequence double-stranded and loop single-stranded DNA as host DNA) PCR amplification with one specific LTR primer (R region) Nested PCR with inner LTR primer (R or U5 region) Sequencing using downstream U5 primer	Trimming: Sequencher Genome alignment: Bowtie2 (and BLAT), hg19 Annotation databases: RefSeq genes, UCSC known genes Gene ontology analysis: TopGO Statistical analysis: Fisher's exact test, Holm-Bonferroni correction, Cochran-Armitage test	Wagner et al. [45•]

[1] *LITCHi* latent infection of T cells by HIV (http://litchi.labtelenti.org/)

[2] *GuavaH* genomic utility for association and viral analyses in HIV (http://www.guavah.org/) [100]

[3] *InSiPiD* integration site pipeline and database (http://www.bushmanlab.org/tools)

amplicon sequenced twice or from two individual integration events occurring at the same chromosomal position [43, 44•, 45•].

Maldarelli et al. and Wagner et al. observed that the same integration site was identified multiple times in latently infected cells, suggesting that a single infected cell was clonally expanded over time [44•, 45•]. The location of the proviral insertion in these clonally expanded infected cells was enriched in genes involved in cell division cycle or cancer, supporting the notion that the site of viral integration may promote homeostatic proliferation. It is important to note that, despite this observation, cancer due to HIV insertional mutagenesis is not a major issue in HIV-infected individuals [36, 46]. Alternative hypotheses explaining the observed cell clonal expansion include (i) the infection of stem cells [20, 21, 47, 48] and (ii) the possible survival advantage of cells carrying defective viral genomes [31••, 46]. It would thus be interesting to know if the expanded cell clones are part of the inducible or defective viral reservoir. This would guide viral eradication strategies to account for those cells and explore ways to specifically inhibit their proliferation.

The link between the integration site and HIV latency is considered to be at the transcriptional level, resulting in silencing via epigenetic regulation of the chromatin environment or transcriptional interference [25, 49, 50]. Spatial features of the integration site associated with transcriptional latency have been shown to favor (i) heterochromatin (centromeric alphoid repeats) and intergenic regions [51, 52], (ii) flanking cellular genes that are highly expressed [51, 53], and (iii) the same orientation as cellular genes [54]. In contrast, Dahabieh et al. suggested that NF-kB activity and cell activation state correlate with viral gene expression efficiency rather than genomic integration features [55].

The role of integration site location on HIV transcriptional activity has been recently assessed in five primary cell models of HIV latency, identifying features specific for each model [42••]. Viral silencing was stronger if the host gene expression was high in the Jurkat latent model and in active CD4+ T cells, while the contrary was observed in a central memory CD4+ T cell model. The same orientation integration bias in the Bcl-2 latency model was confirmed but was not identified in the other latency models [54]. Integration in intergenic regions was favored in active or resting CD4+ cell models only. In contrast, integration in alphoid repeats was preferentially identified in the Jurkat latency model, confirming a previous study, and in resting and central memory cell models [51]. Finally, only the H4K12ac acetylation epigenetic mark correlated with latency in active, resting, and central memory CD4+ cell models but not in the Bcl-2 and Jurkat cell line models. This study highlighted characteristics specific for each model but failed to identify common features of integration site selection specific of HIV latency across all latency models, leaving open the debate about the contribution of integration site

location to transcriptional silencing [42••]. Single-cell analysis may facilitate understanding of the link existing between integration site location and viral transcription.

Transcriptome and HIV Latency

Transcriptome analyses capture a snapshot of RNA molecules within the cell at a specific time. The mRNA molecules will be translated shaping the proteome content of the cell and therefore its functionality. Although the transcriptome only partially correlates with the proteome, it still provides a valuable approximation of the cell condition at a specific time [56–58]. The amount of specific mRNA species results from a balance between the production rate and decay rate. The former is determined by an interplay of regulatory processes acting at transcription initiation, elongation, and termination steps, and the latter is influenced by proteins binding to the 3′ untranslated region (3′ UTR) and the activity of regulatory RNAs such as microRNAs and noncoding RNAs [59–68]. External stimuli will affect this balance, tipping it towards one direction or another [69, 70]. An additional layer of complexity resides in alternative splicing, for both HIV and host cells [59, 71–73].

Transcriptional silencing of HIV has been considered as a major cause of latency [25, 49, 50]. The majority of studies investigating transcriptional silencing have focused on mechanisms acting on the viral LTR promoter, highlighting the role of the viral Tat protein, cellular proteins, histone post-translational modifications, and chromatin remodeling [25, 49, 50, 74–77]. However, only a few studies have focused on the transcriptome analysis of latently infected cells. Evans et al. developed a new primary model of HIV latency, which consists in co-culturing resting CD4+ T cells with myeloid dendritic cells (mDC) for 24 h, followed by CCR5-tropic HIV-GFP infection [78]. Latently infected cells were defined as nonproliferating CD4+ T cells not expressing viral-encoded GFP and were thus sorted by fluorescence-activated cell sorting (FACS) as SNARF-1[high] and GFP− cells after 5 days post-infection. This model provided evidence that CD4+ T cell direct contact with mDC, but not plasmacytoid DC (pDC), facilitates establishment of HIV latency. The study used microarrays to compare latently infected CD4+ T cells versus mock-infected CD4+ T cells co-cultured with mDC and identified only a limited number of genes, mostly interferon I-regulated genes (IFIT1, IFI27, and OAS1), as differentially expressed. These genes may represent latency biomarkers but may also originate from noninfected exposed cells.

Gene arrays are being replaced by high-throughput sequencing technologies, as these do not require prior knowledge about gene sequences and provide a higher accuracy. Current methods differ in the library preparation and in the targeted RNA species. For example, total RNA sequencing

(including small RNAs) allows capturing a complete picture of RNA within a cell, coding or noncoding [79], but requires a high coverage and does not discriminate between nuclear immature transcripts and cytoplasmic mature transcripts. Messenger RNA-Seq (mRNA-Seq) is based on poly(A) capture, thereby requiring less sequencing depth, and will reflect more faithfully the coding transcriptome, although it will not account for coding transcripts that are not poly-adenylated [80, 81]. Recent improvements in library preparation protocols include determination of RNA strand specificity as well as of splicing variants of the coding transcripts. The analysis of these sequences requires computational tools to align the transcripts to the genome and identify transcript isoforms.

The analysis by Evans et al. highlighted the importance of analyzing purified populations of latently infected resting CD4+ T cells in order to identify specific latency biomarkers [78]. Because of the difficulty to isolate and obtain an uncontaminated population of latently infected cells, only a limited number of studies investigated the cellular transcriptome of HIV latency in primary models. Mohammadi et al. used a primary CD4+ T cell model infected with an attenuated viral construct. Successfully infected cells were isolated and purified by FACS and allowed to revert to a resting, hence latent state, by co-culture with H80 feeder cells [32••]. They analyzed the transcriptome of the infected cells over time by mRNA-Seq, thereby following establishment of latency. Cellular activation or resting state was shown to be a major driver of transcriptional differences between cells. Except for the persistent detection of viral transcripts, only a very limited number of host genes were found to be differentially expressed between noninfected and latently infected cells, suggesting that HIV had a very minor impact on the transcriptome of the resting cell. Latently infected cells were also exposed for 8 and 24 h to diverse stimuli to reactivate latency, including SAHA (vorinostat), disulfiram, and IL-7. Although some of these molecules were impacting viral transcription, they failed to stimulate viral protein expression, underscoring the possibility that post-transcriptional blocks would also contribute to the latent phenotype. This study indicated that HIV latency reflects transcriptional and post-transcriptional blocks, including nuclear export and translation [26••, 32••, 82]. The multiplicity of mechanisms leading to HIV latency was also illustrated recently by the comparison of several latency-reactivating agents and their different efficiencies across latency models [26••].

Proteome and HIV Latency

The genetic flow of information in the cell attains a key principal level with the translation of messenger RNA into functional protein. Each protein is modulated through tissue-specific dynamic processes such as alternative splicing, post-translational modifications, interactions with other molecules, and formation of proteomic complexes, which delineate the multidimensional complex nature of the proteome. A recently published blueprint of the human proteome reported approximately 3×10^6 unique peptides covering 84 % of the annotated coding genome [83]. Analyzing the proteins of a biological system gives the opportunity to understand and reveal the intricacies of many cellular functional pathways, inaccessible through genomic or transcriptomic studies.

During HIV infection, as in the case of any viral infection, the proteome of the infected host cell constitutes a key layer where a plethora of responses and interactions with the viral components take place [84, 85]. HIV orchestrates various cellular processes by interacting and manipulating the functional blocks on the host cell in order to promote viral replication and assure its survival [86]. On the host side, various studies have identified changes in post-translational modifications, localization, and amount of proteins upon infection by HIV [87, 88]. In addition to taking advantage of the host cellular proteome, HIV brings along an arsenal of 15 viral proteins that interfere with the host cell, establish interactions, and carry diverse post-translational modifications in order to assure key steps of the virus life cycle.

Recent technological advances and the advent of mass spectrometry (MS)-based proteomics have made a strong positive impact on virology and have helped uncover and characterize many proteomic responses triggered by viral infection [89, 90]. A standard MS experiment consists of a succession of steps: sample protein fractionation into smaller peptides, separation, ion detection, and relative or absolute protein quantitation. The general pipeline can be adapted to address more specific questions such as studying alterations in the phosphoproteome [91]. Online databases and search algorithms are used to map the obtained peptides to putative proteins. The resulting proteomic MS data has additional complexity since, often, more than one candidate protein is associated to the measurements from a set of peptides. A necessary step in MS data pre-processing is to filter out proteins with few, uncertain peptide measurements. MS experiments allow obtaining relative or absolute protein measurements between two different conditions (e.g., HIV-infected and mock-infected cells) which need to be normalized.

Jager et al. used affinity purification coupled with MS (AP-MS) to construct a map of HIV-host interactions at the proteome level. The authors devised a scoring system coined MiST which incorporates the reproducibility, abundance, and specificity of a protein to filter out nonspecific interactions. They

reported on 497 high-confidence interactions, mainly enriched in transcription and regulation of viral protein post-translational modifications [92]. Another study employed isobaric tags for relative and absolute quantitation (iTRAQ) coupled with tandem MS to determine the host proteome response to HIV infection in a time-dependent manner [93]. Differentially expressed proteins were identified between HIV- and mock-infected samples at three different time points post-infection. Interestingly, the groups of differentially expressed proteins differed between early and late time points, capturing an early response enriched for cell proliferation, protein synthesis, DNA recombination, repair, and maintenance versus a late reaction linked to T cell activation [93]. Stable isotope labeling by amino acids in cell culture (SiLAC) profiling of phosphopeptides during HIV entry in the host cell has revealed responses to HIV infection at the level of the host phosphoproteome, notably suggesting that HIV affects the phosphorylation of serine-rich host proteins to ease its replication and subsequent release [94•].

A series of host transcription factors, such as NF-kB, NFAT, or Sp1, bind to the HIV promoter, thereby modulating viral transcriptional activity in infected CD4+ T cells. Their nuclear availability varies between active and resting, latently infected cells [25, 95, 96]. Furthermore, reversible post-translational modifications of histones, such as deacetylation, have been found to play an important role in the establishment and maintenance of latency by modeling the neighboring chromatin organization at the viral LTR promoter [25, 27, 49, 50]. At the membrane level, 17 proteins involved in cell survival pathways or in the regulation of cellular adherence and transfer were found to be disrupted in latently infected cells [97]. The same study also showed that targeted treatment with inhibitors of several of these proteins, precisely XIAP and BTK, influenced the viability of latently infected cells [97]. Understanding the molecular pathways contributing to HIV latency at the proteome level may help in designing novel treatments to eliminate the quiescent virus.

Little is known about HIV-host interactions and the host proteome response during latency and reactivation. To date, no study has compared the whole proteomes of latent and activated HIV-infected cells. Glycoproteomic tandem MS has been employed to profile the disruption in glycoprotein secretion from latently HIV-infected T cells and to identify aberrantly secreted proteins in plasma from infected patients. Six glycoproteins, L-selectin, neogenin, galectin-3-binding protein, CECR1, ICOS, and the phospholipid transfer protein, were found to be overly secreted by latently infected T cells and elevated in patient plasma [98]. Although the mechanisms through which these proteins contribute to HIV latency remain elusive, this study represents a starting point to explore how latency alters protein secretion of the host cell. Another recent study focused on characterizing histone post-translational modifications in SupT1 cell lines infected with HIV, UV-inactivated HIV, and mock-infected in order to find a link to the transcriptional behavior of the cell. Using nano-LC-MS/MS, a set of histone modifications was detected in HIV- versus mock-infected cells, providing insight on how the virus modulates chromatin accessibility to assure its transcription during infection [99]. This study illustrates the epigenetic mechanisms used to manipulate chromatin to ensure transcriptional silencing during latency. Improvements of measurement investigation techniques and joint analyses of all post-translational modifications should help in assembling a complete view of the whole cellular proteome, thereby informing on the identity and the role of specific modified proteins in the process of latency.

Conclusions

Unraveling the different aspects of latency will be a major step towards efficient HIV treatment and eradication. However, HIV latency and its reactivation have proven to be complex processes with many confounding factors. Genome-wide data provide a snapshot of the latently infected cell, at different molecular levels, i.e., integration site, transcriptome, and proteome. However, several aspects may influence the analysis: (i) The resting status of the cell limits the possibility of identifying unique markers of latency, (ii) the presence of a large number of infected cells carrying defective viral genome sequences obscures the features of the latent viral reservoir that can be reactivated with the goal of cure, and (iii) the analysis of high-throughput genome-wide data is particularly challenging due to the large amount of data as well as systematic technological and biological noise that may lead to high false-positive rates. Adequate bioinformatic tools are required to account for these effects in order to process and analyze the data. In-depth and dynamic genome-wide studies can provide essential contributions to the understanding of establishment of latency during resting condition, maintenance of latency, and the steps towards full reactivation.

Compliance with Ethics Guidelines

Conflict of Interest   Angela Ciuffi, Pejman Mohammadi, Monica Golumbeanu, Julia di Iulio, and Amalio Telenti declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent   This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

• Of importance

•• Of major importance

1. Siliciano JD, Siliciano RF. The latent reservoir for HIV-1 in resting CD4+ T cells: a barrier to cure. Curr Opin HIV AIDS. 2006;1: 121–8.

2.• Durand CM, Blankson JN, Siliciano RF. Developing strategies for HIV-1 eradication. Trends Immunol. 2012;33:554–62. *Excellent review and overview of the obsacle of the HIV reservoir and the strategies to eliminate it.*

3. Barouch DH, Deeks SG. Immunologic strategies for HIV-1 remission and eradication. Science. 2014;345:169–74.

4. Barton KM, Burch BD, Soriano-Sarabia N, Margolis DM. Prospects for treatment of latent HIV. Clin Pharmacol Ther. 2013;93:46–56.

5. Abou-El-Enein M, Bauer G, Reinke P, Renner M, Schneider CK. A roadmap toward clinical translation of genetically-modified stem cells for treatment of HIV. Trends Mol Med. 2014;20(11): 632–42.

6. Kent SJ, Reece JC, Petravic J, Martyushev A, Kramski M, De Rose R, et al. The search for an HIV cure: tackling latent infection. Lancet Infect Dis. 2013;13:614–21.

7. Coiras M, Lopez-Huertas MR, Perez-Olmeda M, Alcami J. Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. Nat Rev Microbiol. 2009;7:798–812.

8.•• Whitney JB, Hill AL, Sanisetty S, Penaloza-MacMaster P, Liu J, Shetty M, et al. Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. Nature. 2014;512:74–7. *This study shows that upon viral infection the viral reservoir is established very early (in less than three days) in an animal model. This study also shows that HIV is detectable in lymph nodes but not in blood of some animals, suggesting that lymph nodes may be an early and important reservoir.*

9. Zaikos TD, Collins KL. Long-lived reservoirs of HIV-1. Trends Microbiol. 2014;22:173–5.

10. Eisele E, Siliciano RF. Redefining the viral reservoirs that prevent HIV-1 eradication. Immunity. 2012;37:377–88.

11. Coleman CM, Wu L. HIV interactions with monocytes and dendritic cells: viral latency and reservoirs. Retrovirology. 2009;6:51.

12. Fletcher CV, Staskus K, Wietgrefe SW, Rothenberger M, Reilly C, Chipman JG, et al. Persistent HIV-1 replication is associated with lower antiretroviral drug concentrations in lymphatic tissues. Proc Natl Acad Sci U S A. 2014;111:2307–12.

13. Costiniuk CT, Jenabian MA. Cell-to-cell transfer of HIV infection: implications for HIV viral persistence. J Gen Virol. 2014;95: 2346–55.

14. Gray LR, Roche M, Flynn JK, Wesselingh SL, Gorry PR, Churchill MJ. Is the central nervous system a reservoir of HIV-1? Curr Opin HIV AIDS. 2014;9:552–8.

15. Denton PW, Long JM, Wietgrefe SW, Sykes C, Spagnuolo RA, Snyder OD, et al. Targeted cytotoxic therapy kills persisting HIV infected cells during ART. PLoS Pathog. 2014;10:e1003872.

16. Soriano-Sarabia N, Bateson RE, Dahl NP, Crooks AM, Kuruc JD, Margolis DM, et al. The quantitation of replication-competent HIV-1 in populations of resting CD4+ T cells. J Virol. 2014; 88: 14070-7.

17. Kumar A, Abbas W, Herbein G. HIV-1 latency in monocytes/macrophages. Viruses. 2014;6:1837–60.

18. Bacchus C, Cheret A, Avettand-Fenoel V, Nembot G, Melard A, Blanc C, et al. A single HIV-1 cluster and a skewed immune homeostasis drive the early spread of HIV among resting CD4+ cell subsets within one month post-infection. PLoS ONE. 2013;8: e64219.

19. Archin NM, Bateson R, Tripathy MK, Crooks AM, Yang KH, Dahl NP, et al. HIV-1 expression within resting CD4+ T cells after multiple doses of vorinostat. J Infect Dis. 2014;210:728–35.

20. Nixon CC, Vatakis DN, Reichelderfer SN, Dixit D, Kim SG, Uittenbogaart CH, et al. HIV-1 infection of hematopoietic progenitor cells in vivo in humanized mice. Blood. 2013;122:2195–204.

21. McNamara LA, Onafuwa-Nuga A, Sebastian NT, Riddell J, Bixby D, Collins KL. CD133+ hematopoietic progenitor cells harbor HIV genomes in a subset of optimally treated people with long-term viral suppression. J Infect Dis. 2013;207:1807–16.

22. Narasipura SD, Kim S, Al-Harthi L. Epigenetic regulation of HIV-1 latency in astrocytes. J Virol. 2014;88:3031–8.

23.•• Cillo AR, Sobolewski MD, Bosch RJ, Fyne E, Piatak Jr M, Coffin JM, et al. Quantification of HIV-1 latency reversal in resting CD4+ T cells from patients on suppressive antiretroviral therapy. Proc Natl Acad Sci U S A. 2014;111:7078–83. *This study measured the fraction of the inducible reservoir upon exposure to different conditions.*

24. Pace MJ, Agosto L, Graf EH, O'Doherty U. HIV reservoirs and latency models. Virology. 2011;411:344–54.

25. Hakre S, Chavez L, Shirakawa K, Verdin E. HIV latency: experimental systems and molecular models. FEMS Microbiol Rev. 2012;36:706–16.

26.•• Spina CA, Anderson J, Archin NM, Bosque A, Chan J, Famiglietti M, et al. An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients. PLoS Pathog. 2013;9:e1003834. *A comparative study of different HIV latency cell models and their response to multiple latency reactivating agents or stimuli.*

27. Shirakawa K, Chavez L, Hakre S, Calvanese V, Verdin E. Reactivation of latent HIV by histone deacetylase inhibitors. Trends Microbiol. 2013;21:277–85.

28. Archin NM, Margolis DM. Emerging strategies to deplete the HIV reservoir. Curr Opin Infect Dis. 2014;27:29–35.

29. Xing S, Siliciano RF. Targeting HIV latency: pharmacologic strategies toward eradication. Drug Discov Today. 2013;18:541–51.

30.•• Eriksson S, Graf EH, Dahl V, Strain MC, Yukl SA, Lysenko ES, et al. Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. PLoS Pathog. 2013;9:e1003174. *A comparative study of different methods to estimate the size of the viral reservoir.*

31.•• Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell. 2013;155: 540–51. *This study quantified the proportion of HIV+ cells carrying defective versus replication-competent proviral sequences.*

32.•• Mohammadi P, di Iulio J, Munoz M, Martinez R, Bartha I, Cavassini M, et al. Dynamics of HIV latency and reactivation in a primary CD4+ T cell model. PLoS Pathog. 2014;10:e1004156. *This study characterized the joint viral and cellular transcriptome in a latency model over time. Post-transcriptional block and cellular activation status were identified as components of latency.*

33. Papayannakos C, Daniel R. Understanding lentiviral vector chromatin targeting: working to reduce insertional mutagenic potential for gene therapy. Gene Ther. 2013;20:581–8.

34. Ciuffi A, Bushman F: Retroviral integration target site selection. In: HIV-1 integrase: mechanism of action and inhibitor design. edn 1. Edited by Neamati N: John Wiley & Sons, Inc.; 2011:51–65. [Binghe W (Series Editor): Wiley Series in Drug Discovery and Development

35. Desfarges S, Ciuffi A: Viral integration and consequences on host gene expression. In: Viruses: essential agents of life. Edited by Witzany G: Springer Science+Business Media; 2012.

36. Ciuffi A. Mechanisms governing lentivirus integration site selection. Curr Gene Ther. 2008;8:419–29.

37. Desfarges S, Ciuffi A. Retroviral integration site selection. Viruses. 2010;2:111–30.

38. Ciuffi A, Telenti A. State of genomics and epigenomics research in the perspective of HIV cure. Curr Opin HIV AIDS. 2013;8:176–81.

39. Ciuffi A, Barr SD. Identification of HIV integration sites in infected host genomic DNA. Methods. 2011;53:39–46.

40. Ciuffi A, Ronen K, Brady T, Malani N, Wang G, Berry CC, et al. Methods for integration site distribution analyses in animal cell genomes. Methods. 2009;47:261–8.

41. Gabriel R, Eckenberg R, Paruzynski A, Bartholomae CC, Nowrouzi A, Arens A, et al. Comprehensive genomic access to vector integration in clinical gene therapy. Nat Med. 2009;15: 1431–6.

42.•• Sherrill-Mix S, Lewinski MK, Famiglietti M, Bosque A, Malani N, Ocwieja KE, et al. HIV latency and integration site placement in five cell-based models. Retrovirology. 2013;10:90. *A comparative study of latency models to identify genomic features of integration site locations specific of HIV latency.*

43. Firouzi S, Lopez Y, Suzuki Y, Nakai K, Sugano S, Yamochi T, et al. Development and validation of a new high-throughput method to investigate the clonality of HTLV-1-infected cells based on provirus integration sites. Genome Med. 2014;6:46.

44.• Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, et al. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science. 2014;345:179–83. *This study identified clonal expansion of some infected cells in HIV+ individuals.*

45.• Wagner TA, McLaughlin S, Garg K, Cheung CY, Larsen BB, Styrchak S, et al. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science. 2014;345:570–3. *This study identified clonal expansion of some infected cells in HIV+ individuals.*

46. Margolis D, Bushman F. HIV/AIDS. Persistence by proliferation? Science. 2014;345:143–4.

47. Buzon MJ, Sun H, Li C, Shaw A, Seiss K, Ouyang Z, et al. HIV-1 persistence in CD4+ T cells with stem cell-like properties. Nat Med. 2014;20:139–42.

48. Powers JM, Trobridge GD. Identification of hematopoietic stem cell engraftment genes in gene therapy studies. J Stem Cell Res Ther. 2013;2013.

49. Van Lint C, Bouchat S, Marcello A. HIV-1 transcription and latency: an update. Retrovirology. 2013;10:67.

50. Mbonye U, Karn J. Transcriptional control of HIV latency: cellular signaling pathways, epigenetics, happenstance and the hope for a cure. Virology. 2014;454–455:328–39.

51. Lewinski MK, Bisgrove D, Shinn P, Chen H, Hoffmann C, Hannenhalli S, et al. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. J Virol. 2005;79:6610–9.

52. Pace MJ, Graf EH, Agosto LM, Mexas AM, Male F, Brady T, et al. Directly infected resting CD4+ T cells can produce HIV Gag without spreading infection in a model of HIV latency. PLoS Pathog. 2012;8:e1002818.

53. Han Y, Lassen K, Monie D, Sedaghat AR, Shimoji S, Liu X, et al. Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. J Virol. 2004;78:6122–33.

54. Shan L, Yang HC, Rabi SA, Bravo HC, Shroff NS, Irizarry RA, et al. Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. J Virol. 2011;85:5384–93.

55. Dahabieh MS, Ooms M, Brumme C, Taylor J, Harrigan PR, Simon V, et al. Direct non-productive HIV-1 infection in a T-cell line is driven by cellular activation state and NFkappaB. Retrovirology. 2014;11:17.

56. Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, et al. Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet. 2011;7:e1001393.

57. Low TY, van Heesch S, van den Toorn H, Giansanti P, Cristobal A, Toonen P, et al. Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. Cell Rep. 2013;5:1469–78.

58. Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. BMC Genomics. 2009;10:161.

59. Bentley DL. Coupling mRNA processing with transcription in time and space. Nat Rev Genet. 2014;15:163–75.

60. Groen JN, Capraro D, Morris KV. The emerging role of pseudogene expressed non-coding RNAs in cellular functions. Int J Biochem Cell Biol. 2014;54:350–5.

61. Groen JN, Morris KV. Chromatin, non-coding RNAs, and the expression of HIV. Viruses. 2013;5:1633–45.

62. Swaminathan S, Kelleher AD. MicroRNA modulation of key targets associated with T cell exhaustion in HIV-1 infection. Curr Opin HIV AIDS. 2014;9:464–71.

63. Chiang K, Rice AP. MicroRNA-mediated restriction of HIV-1 in resting CD4+ T cells and monocytes. Viruses. 2012;4:1390–409.

64. Fowler L, Saksena NK. Micro-RNA: new players in HIV-pathogenesis, diagnosis, prognosis and antiviral therapy. AIDS Rev. 2013;15:3–14.

65. Swaminathan G, Navas-Martin S, Martin-Garcia J. MicroRNAs and HIV-1 infection: antiviral activities and beyond. J Mol Biol. 2014;426:1178–97.

66. Cox JE, Sullivan CS. Balance and stealth: the role of noncoding RNAs in the regulation of virus gene expression. Ann Rev Virol. 2014;1:89–109.

67. Saayman S, Ackley A, Turner AM, Famiglietti M, Bosque A, Clemson M, et al. An HIV-encoded antisense long noncoding RNA epigenetically regulates viral transcription. Mol Ther. 2014;22:1164–75.

68. Chang H, Lim J, Ha M, Kim VN. TAIL-seq: genome-wide determination of poly(A) tail length and 3′ end modifications. Mol Cell. 2014;53:1044–52.

69. Weake VM, Workman JL. Inducible gene expression: diverse regulatory mechanisms. Nat Rev Genet. 2010;11:426–37.

70. Voss TC, Hager GL. Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat Rev Genet. 2014;15: 69–81.

71. Ocwieja KE, Brady TL, Ronen K, Huegel A, Roth SL, Schaller T, et al. HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. PLoS Pathog. 2011;7: e1001313.

72. Imbeault M, Giguere K, Ouellet M, Tremblay MJ. Exon level transcriptomic profiling of HIV-1-infected CD4(+) T cells reveals virus-induced genes and host environment favorable for viral replication. PLoS Pathog. 2012;8:e1002861.

73. Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, et al. Landscape and variation of RNA secondary structure across the human transcriptome. Nature. 2014;505:706–9.

74. Donahue DA, Wainberg MA. Cellular and molecular mechanisms involved in the establishment of HIV-1 latency. Retrovirology. 2013;10:11.

75. Lusic M, Giacca M. Regulation of HIV-1 latency by chromatin structure and nuclear architecture. J Mol Biol. 2014. doi:10.1016/j.jmb.2014.07.022.

76. Park J, Lim CH, Ham S, Kim SS, Choi BS, Roh TY. Genome-wide analysis of histone modifications in latently HIV-1 infected T cells. AIDS. 2014;28:1719–28.

77. Jadlowsky JK, Wong JY, Graham AC, Dobrowolski C, Devor RL, Adams MD, et al. Negative elongation factor is required for the maintenance of proviral latency but does not induce promoter-proximal pausing of RNA polymerase II on the HIV long terminal repeat. Mol Cell Biol. 2014;34:1911–28.

78. Evans VA, Kumar N, Filali A, Procopio FA, Yegorov O, Goulet JP, et al. Myeloid dendritic cells induce HIV-1 latency in non-proliferating CD4+ T cells. PLoS Pathog. 2013;9:e1003799.

79. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science. 2007;316:1484–8.

80. Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. Genomewide characterization of non-polyadenylated RNAs. Genome Biol. 2011;12:R16.

81. Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet. 2008;9:843–54.

82. Lassen KG, Ramyar KX, Bailey JR, Zhou Y, Siliciano RF. Nuclear retention of multiply spliced HIV-1 RNA in resting CD4+ T cells. PLoS Pathog. 2006;2:e68.

83. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. Nature. 2014;509:575–81.

84. Zhang L, Zhang X, Ma Q, Zhou H. Host proteome research in HIV infection. Genom, Proteome Bioinforma. 2010;8:1–9.

85. Zhu J, Davoli T, Perriera JM, Chin CR, Gaiha GD, John SP, et al. Comprehensive identification of host modulators of HIV-1 replication using multiple orthologous RNAi reagents. Cell Rep. 2014;9:752–66.

86. Malim MH, Bieniasz PD. HIV restriction factors and mechanisms of evasion. Cold Spring Harb Perspect Med. 2012;2:a006940.

87. DeBoer J, Jagadish T, Haverland NA, Madson CJ, Ciborowski P, Belshan M. Alterations in the nuclear proteome of HIV-1 infected T-cells. Virology. 2014;468-470:409–20.

88. Haverland NA, Fox HS, Ciborowski P. Quantitative proteomics by SWATH-MS reveals altered expression of nucleic acid binding and regulatory proteins in HIV-1-infected macrophages. J Proteome Res. 2014;13:2109–19.

89. Greco TM, Diner BA, Cristea IM. The impact of mass spectrometry-based proteomics on fundamental discoveries in virology. Ann Rev Virol. 2014;1:581–604.

90. Weekes MP, Tomasec P, Huttlin EL, Fielding CA, Nusinow D, Stanton RJ, et al. Quantitative temporal viromics: an approach to investigate host-pathogen interaction. Cell. 2014;157:1460–72.

91. Altelaar AF, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet. 2013;14:35–48.

92. Jager S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, et al. Global landscape of HIV-human protein complexes. Nature. 2012;481:365–70.

93. Navare AT, Sova P, Purdy DE, Weiss JM, Wolf-Yadlin A, Korth MJ, et al. Quantitative proteomic analysis of HIV-1 infected CD4+ T cells reveals an early host response in important biological pathways: protein synthesis, cell proliferation, and T-cell activation. Virology. 2012;429:37–46.

94.• Wojcechowskyj JA, Didigu CA, Lee JY, Parrish NF, Sinha R, Hahn BH, et al. Quantitative phosphoproteomics reveals extensive cellular reprogramming during HIV-1 entry. Cell Host Microbe. 2013;13:613–23. *This paper is the first phosphoproteomic study in a natural infection design, i.e. X4-tropic HIV on primary CD4+ T cells.*

95. Pan X, Baldauf HM, Keppler OT, Fackler OT. Restrictions to HIV-1 replication in resting CD4+ T lymphocytes. Cell Res. 2013;23:876–85.

96. Zack JA, Kim SG, Vatakis DN. HIV restriction in quiescent CD4(+) T cells. Retrovirology. 2013;10:37.

97. Berro R, de la Fuente C, Klase Z, Kehn K, Parvin L, Pumfery A, et al. Identifying the membrane proteome of HIV-1 latently infected cells. J Biol Chem. 2007;282:8207–18.

98. Yang W, Zhou J-Y, Chen L, Ao M, Sun S, Aiyetan P, et al. Glycoproteomic analysis identifies human glycoproteins secreted from HIV latently infected T cells and reveals their presence in HIV+ plasma. Clin Proteome. 2014;11:9.

99. Britton LM, Sova P, Belisle S, Liu S, Chan EY, Katze MG, et al. A proteomic glimpse into the initial global epigenetic changes during HIV infection. Proteomics. 2014;14:2226–30.

100. Bartha I, McLaren PJ, Ciuffi A, Fellay J, Telenti A. GuavaH: a compendium of host genomic data in HIV biology and disease. Retrovirology. 2014;11:6.