Published in final edited form as:
Curr Opin HIV AIDS. 2023 September 01; 18(5): 246–256. doi:10.1097/COH.0000000000000809.

Single-cell multi-omics understanding of HIV-1 reservoir at the epigenetics, transcriptional, and protein levels

Michelle Wong[1,*], **Yulong Wei**[1,*], **Ya-Chi Ho**[1]

[1]Department of Microbial Pathogenesis, Yale University School of Medicine, New Haven, CT 06519, United States

Abstract

Purpose of review: The success of HIV-1 eradication strategies relies on in-depth understanding HIV-1-infected cells. However, HIV-1-infected cells are extremely heterogeneous and rare. Single-cell multi-omic approaches are required to resolve the heterogeneity and rarity of HIV-1-infected cells.

Recent findings: Advancement in single-cell multi-omic approaches enabled HIV-1 reservoir profiling across the epigenetic (ATAC-seq), transcriptional (RNA-seq), and protein levels (CITE-seq). Using HIV-1 RNA as a surrogate, ECCITEseq identified enrichment of HIV-1-infected cells in clonally expanded cytotoxic CD4+ T cells. Using HIV-1 DNA PCR-activated microfluidic sorting, FIND-seq captured the bulk transcriptome of HIV-1 DNA+ cells. Using targeted HIV-1 DNA amplification, PheP-seq identified surface protein expression of intact versus defective HIV-1-infected cells. Using ATAC-seq to identify HIV-1 DNA, ASAP-seq captured transcription factor activity and surface protein expression of HIV-1 DNA+ cells. Combining mapping HIV-1 DNA by ATAC-seq and HIV-1 RNA mapping by RNA-seq, DOGMAseq captured the epigenetic, transcriptional, and surface protein expression of latent and transcriptionally active HIV-1-infected cells. To identify reproducible biological insights and authentic HIV-1-infected cells and avoid false-positive discovery of artifacts, we reviewed current practices of single-cell multi-omic experimental design and bioinformatic analysis.

Summary: Single-cell multi-omic approaches may identify innovative mechanisms of HIV-1 persistence, nominate therapeutic strategies, and accelerate discoveries.

Keywords

HIV-1 reservoir; HIV-1 latency; mechanisms of HIV-1 persistence; clonal expansion dynamics; single-cell RNA-seq; single-cell ATAC-seq; T cell receptor sequencing; bioinformatic analysis; HIV-1 genome mapping; HIV-1 cure strategies

Correspondence: ya-chi.ho@yale.edu.

[*]Equal contribution

Conflicts of interest

None.

Introduction

Despite suppressive antiretroviral therapy (ART), HIV-1 integrates into the chromosome of the latent reservoir, primarily CD4+ T cells [1–3] and persists lifelong [4, 5]. Single-genome HIV-1 proviral genome profiling [6], in the past decade [7–12], provided in-depth understanding of the intact and defective HIV-1 genome landscape in the reservoir and immune selection pressure on HIV-1-infected cells [13]. However, the development of successful HIV-1 cure strategies relies on targeting cellular markers that are specifically expressed in HIV-1-infected cells without damaging uninfected cells. Thus, the field has moved forward to understanding the cellular environment of HIV-1 infected cells beyond HIV-1 genome profiling.

Understanding HIV-1-infected cells is extremely challenging.

First, HIV-1-infected cells are extremely rare: only ~1/10^6 CD4+ T cells harbor infectious HIV-1 [2, 3, 14], while ~186–879/10^6 CD4+ T cells are HIV-1-infected [15, 16] but harbor mainly defective HIV-1 [6]. Second, during latency, HIV-1-infected cells are transcriptionally silent and thus cannot be distinguished from uninfected cells. Using HIV-1 RNA [17] or Env protein expression [18] as a surrogate, HIV-1-infected cells can be captured for transcriptome profiling. However, these methods require exogenous stimulation *ex vivo* and the cellular states captured no longer reflect the *in vivo* state. Third, no cellular markers can specifically distinguish the rare HIV-1-infected cells from uninfected cells. While HIV-1-infected cells are enriched in some subpopulations of CD4+ T cells, involving T cell activation (HLA-DR [19]), exhaustion (such as PD-1, TIGIT, LAG-3 [20, 21]), migration (such as integrin α4β1 (VLA-4) or integrin β1 (CD49d)[22, 23]) and differentiation into central memory [24], effector memory [7], Th1 [11], cytotoxic T cells [25, 26], T follicular helper (TFH) cells [27], and survival (such as BIRC5 [28] and Bcl-2 [29]), these markers are not specific enough as therapeutic targets. Fourth, HIV-1-infected cells are highly heterogeneous, reflecting the diverse differentiation, polarization, and exhaustion states of CD4+ T cells [30]. Fifth, the persistence of HIV-1-infected cells is a dynamic process and involves multiple cell survival, proliferation, and immune evasion mechanisms that remain elusive. HIV-1-infected cells not only persist but proliferate over time[31–35]. Upon reactivation, HIV-1-infected cells should presumably die of viral cytopathic effects or immune clearance. However, some HIV-1-infected cells may not die of viral cytopathic effect upon reactivation [36] and may resist immune clearance [29]. Finally, transcriptome-based profiling of HIV-1-infected cells is extremely challenging. CD4+ T cells have low RNA content, compared with other cell types frequently used for single-cell profiling, such as cancer, neuron, and embryo. Thus, technology advancements are urgently needed to understand mechanisms of HIV-1 persistence in both latent and transcriptionally active states, identify the rare HIV-1-infected cells, resolve the heterogeneity of HIV-1-infected cells, and nominate cellular markers that can distinguish HIV-1-infected cells from uninfected cells for therapeutic interventions.

Understanding the diverse epigenetic regulators and immune programs of the heterogeneous CD4+ T cells

Bulk RNA-seq captures the 99.9% of uninfected cells and does not reflect the rare HIV-1-inected cells.—Given the heterogeneity, rarity, and the lack of selection markers for HIV-1-infected cells, investigating the cellular environment of HIV-1-infected cells requires single-cell profiling. Bulk RNA-seq, for example, captures cellular environment of the 99.9% of uninfected cells and thus is irrelevant to our understanding of HIV-1-infected cells. The advancement of single-cell transcriptomic profiling captures the heterogeneous immune cell phenotypes and identifies rare cells of interest [37].

The heterogeneous T cell phenotype is defined by master transcription factors and immune effector gene expression.—The heterogeneous polarization, differentiation, proliferation, and migration states of CD4+ T cells, and the decision between plasticity and fate commitment, is determined by antigen stimulation (including T cell receptor (TCR) signaling strength and costimulatory molecules) and cytokine cues at the local environment where priming occurs [38]. These signals trigger the expression of master transcription factors (such as Tbet for Th1, GATA3 for Th2, RORγt for Th17, FOXP3 for Treg, and Bcl-6 for TFH), which dictate the cellular transcriptional program by binding to the promoter of genes involving polarization, differentiation, effector function, migration, and survival. These transcription factors regulate gene expression and thus dictate the phenotype of the cell, as characterized by transcriptome signatures or protein expression (such as IFNγ expression for Th1, IL-4 expression for Th2, IL-17 for Th17, TGF-β for Treg, and IL-21 for TFH). Understanding the cellular environment at all three aspects of the central dogma of molecular biology – epigenetic regulation by transcription factors at the DNA level, the transcriptional landscape at the RNA level, and protein expression – provides key insights for mechanisms of HIV-1 persistence and a genome-wide search for therapeutic targets.

Tracking the unique T cell receptor (TCR) sequence identifies the temporal-spatial dynamics of CD4+ T cells.—Antigen specificity in T cells is determined by the hypervariable loops (CDR3 region) of the T cell receptor α and β chains which form the center of the antigen-binding site. T cell diversity is determined by the D and J gene segment rearrangement at the CDR3 region, unlike somatic hypermutation in the V region in B cells. Given the diversity of T cell repertoire, different T cells having the same TCR sequence originate from the proliferation of the same mother cell. A T cell clone, i.e. different T cells having the same TCR sequence, respond to the same cognate antigen. T cell clone size, i.e. many cells (large T cell clone) versus few cells (small T cell clone) having the same TCR sequence, reflect *in vivo* proliferation of the T cell clone. By tracking different T cells having the same TCR at different time (temporal dynamics) or at different anatomical locations (spatial dynamics), the temporal-spatial T cell clonal expansion dynamics *in vivo* can be delineated [39].

Technology advancement enables our understanding of the cellular environment at the genome-wide level

At the epigenetic level, assay for transposase-accessible chromatin with sequencing (ATAC-seq)[40] captures the epigenome by identifying genes having increased accessibility and the transcription factors that bind to *cis*-regulatory elements (such as promoters) that regulate respective gene expression. Briefly, Tn5 transposase binds to open chromatin regions. By sequencing DNA sequences accessible to Tn5, ATAC-seq identifies genes having increased accessibility and transcription factor binding footprints (transcription factor activity) [41, 42] (Figure 1).

At the transcriptome level, RNA-seq captures genome-wide non-targeted snapshot of the transcriptional program in each cell [37]. TCR sequence, which defines T cell clonality, can be captured by RNA-seq by targeted amplification [43, 44].

At the protein level, cellular indexing of transcriptomes and epitopes (CITE-seq)[45] and RNA expression and protein sequencing assay (REAP-seq)[46] capture cellular surface protein expression in addition to RNA-seq by staining cells with antibodies tagged with DNA barcodes. Unlike flow cytometry or CyTOF antibodies which can only examine up to ~40 surface proteins because of fluorophore spectrum overlaps or the number of metal isotopes, CITE-seq can profile >100 surface protein expression at the same time, as long as an appropriate antibody is available. While the surface protein expression can be profiled by DNA-tagged antibodies, profiling intracellular proteins remains challenging. While HIV-1-infected cells producing HIV-1 proteins can be identified by flow cytometry-based approaches such as HIV-Flow [22, 47], coupling HIV-1 intracellular protein detection to intracellular protein expression, surface protein expression, transcriptome, and epigenome at the single-cell level remains lacking. Further technology advancement, such as single-cell detection of intracellular proteins by INs-seq [48] or single-cell proteomics [49] that can reach the throughput of hundreds of thousands of cells and specificity for HIV-1 Gag or Env detection (both of which can be nonspecific), may further advance our understanding of the protein expression of HIV-1-infected cells.

Single-cell multi-omic profiling provides unprecedented understanding of cell states across the central dogma of molecular biology.—By combining two modalities such as T-ATAC-seq (ATAC-seq and RNA-seq [50]) and ASAP-seq (ATAC-seq and CITE-seq)[51]) or three modalities – TEA-seq [52] and DOGMA-seq [51] (ATAC-seq, RNA-seq, and CITE-seq), ECCITE-seq [44] (RNA-seq, CITE-seq, and TCR sequencing) in the same single cells, the epigenetic regulation, transcriptional landscape, protein expression, and T cell clonal expansion dynamics can be captured in the same single cells. Given comparable results of single-cell profiling between fresh and cryopreserved samples [53, 54], single-cell multi-omic profiling provides feasibility, scalability, and flexibility for clinical studies.

Identifying transcriptionally active HIV-1-infected cells upon *ex vivo* stimulation

Given the rarity of HIV-1-infected cells, the most effective way to profile HIV-1 reservoir would be to isolate the rare HIV-1-infected cells from the 99.9% of uninfected cells.

However, because there are no cellular markers that can distinguish HIV-1-infected cells from uninfected cells, identifying HIV-1-infected cells requires *ex vivo* stimulation to induce HIV-1 RNA or protein expression as a surrogate. While HIV-1-infected cells can be identified by HIV-1 RNA (by Flow-FISH) or viral protein staining (such as intracellular staining of HIV-1 p24), RNA is degraded upon formaldehyde fixation or high temperature (such as PCR). Thus, these orthogonal methods are not compatible with single-cell RNA-seq. The single-cell transcriptome of HIV-1-infected cells were first captured using flow cytometric sorting of HIV-1 RNA+ cells (HIV-1 SortSeq by our group, using 192 HIV-1-targeting probes for a fluorescent in situ hybridization (FISH), after 16 hours of PMA/ionomycin stimulation [17]) or HIV-1 Env+ cells (LURE by the Nussenzweig group, using antibody against HIV-1 Env for magnetic enrichment, 40 hours of PHA stimulation [18]). However, *ex vivo* activation cannot capture the *in vivo* state of latent and transcriptionally active HIV-1-infected cells. Nevertheless, both studies found that HIV-1-infected cells are highly heterogeneous.

Identifying transcriptionally active HIV-1-infected cells without *ex vivo* stimulation

To profile the *in vivo* state of HIV-1-infected cells, total CD4+ T cells were profiled by single-cell multi-omics, and HIV-1-infected cells were identified by bioinformatically mapping transcriptome to HIV-1 genome. Jack Collora in our group used ECCITE-seq to capture HIV-1 RNA, transcriptome, surface protein expression, and TCR in the same single cell [25]. We found that HIV-1-infected cells are larger in T cell clone size, stable over time, and predominantly (75%) cytotoxic CD4+ T cells. By paired TCR sequencing, we found that cytotoxic CD4+ T cells are the most clonally expanded cells *in vivo*. The high proliferative nature of cytotoxic CD4+ T cells promotes the clonal expansion of HIV-1-infected cytotoxic CD4+ T cells. Of note, all cytotoxic immune effectors (such as cytotoxic CD4+ T cells, cytotoxic CD8+ T cells, and natural killer cells) express granzyme B inhibitor Serpin B9 to prevent self-inflicted injury when lytic granules (such as granzyme B) fall back to the cytotoxic immune effector themselves and induce cell death [55]. By residing in cytotoxic CD4+ T cells, HIV-1 may survive cytotoxic CD8+ T cell killing because HIV-1-infected cytotoxic CD4+ T cells express Serpin B9 that inhibit granzyme B killing. In parallel, Nussenzweig group used TCR Vβ antibodies to isolate CD4+ T clones known to enrich for HIV-1-infected cells from HIV-1+ individuals and profiled these CD4+ T cell clones by CITE-seq [26]. HIV-1 RNA+ cells were also identified by mapping transcriptome to HIV-1 genome. They also found that the clonally expanded CD4+ T cell clones are also enriched for cytotoxic CD4+ T cells. Overall, by single-cell profiling CD4+ T cells and identifying HIV-1 RNA+ cells by mapping transcriptome to HIV-1 genome, both studies found that HIV-1-infected cytotoxic CD4+ T cells are preferentially clonally expanded *in vivo*.

T cell clonality tracking identified enrichment of HIV-1 in cytotoxic CD4+ T cell clones

The clonal expansion of HIV-1-infected cells are typically examined by using unique HIV-1 integration sites in the human genome [34, 35] or by phylogenetic analysis of the highly variable region of the HIV-1 genome, such as *env*, *pol*, or the near full-length HIV-1 genome [31–33], or both [9, 36]. By tracking TCR clonality, i.e. identifying different CD4+ T cells having the same TCR sequence, with paired single-cell transcriptome, we can identify the cell states of CD4+ T cell clones harboring HIV-1 RNA+ cells at different time points, or before versus after antigen stimulation [25]. CD4+ T cell clones having the same TCR sequences respond to the same antigen stimulation. Thus, the transcriptional landscape of HIV-1+ CD4+ T cell clones reflects the cytokine and immune responses that drive the proliferation of the respective CD4+ T cell clone and the HIV-1-infected cells within them. Of note, depending on when HIV-1 infects the respective CD4+ T cell clone, only part (not all) of the CD4+ T cell clones contain HIV-1-infected cells, unless HIV-1 infects the original CD4+ T cells before proliferation started [25, 26]. Overall, tracking TCR clonality within single-cell multi-omics profiling provides a powerful tool for studying the clonal expansion dynamics of HIV-1-infected cells.

Identifying the cellular landscape of latent HIV-1-infected cells

While using HIV-1 RNA as a surrogate identifies transcriptionally active HIV-1-infected cells, the cellular state of latent HIV-1-infected cells (which do not express HIV-1 RNA) remains unknown. HIV-1 latently infected cells would presumably appear indistinguishable from uninfected cells. Distinguishing latent HIV-1-infected cells from uninfected cells is a top priority for designing therapeutic strategies targeting the latent reservoir without damaging the uninfected cells. Four recent paralleled studies attempted to tackle this major challenge in the field.

FIND-seq.—Clark *et al.* in the Abate group and the Boritz group and designed a PCR-activated sorting microfluidic machinery to sort out HIV-1 DNA+ cells for bulk transcriptome profiling (FIND-seq)[56]. Briefly, Clark and Abate built a microfluidic device to encapsulate CD4+ T cells with RT-PCR reagents in hydrogel compartments. HIV-1 DNA+ cells emit green fluorescence upon HIV-1 *gag* PCR amplification and are sorted by a unique microfluidic device built in-house. While this study identified transcriptome signatures for latently infected cells, the PCR reaction (with cycles of heat for denaturing DNA templates) damages RNA quality, and transcriptome can only be obtained from pools of 100 cells, not at the single cell level. Of note, these HIV-1 *gag* DNA+ cells are still likely to be defective, as defective HIV-1 proviruses account for 88–98% of HIV-1 DNA+ cells [6, 12].

PheP-seq.—Sun *et al.* in the Lichterfeld group applied a commercially available platform (Mission Bio) to profile near full-length HIV-1 DNA and surface protein expression at the single cell level (PheP-seq)[57]. Briefly, cells were stained with DNA-tagged surface protein antibodies and partitioned into single cells. HIV-1 DNA was amplified by 18 sets of primers to reconstruct the near full-length HIV-1 proviral sequences. For individuals having known HIV-1 integration sites, additional individual-specific integration site-specific primers were added to the DNA amplification to capture the junction between the known integration site and HIV-1. This study identified cellular surface protein signatures of presumably intact versus defective HIV-1 proviruses.

ASAP-seq.—Wu *et al.* in the Betts group applied ASAP-seq to identify HIV-1 DNA and host chromatin accessibility (by ATAC-seq), and surface protein expression [58]. Vahedi group first proposed the concept of using transposase to identify cells harboring lentiviral

DNA, using lentiviral transduced CAR-T cells as an example [59]. Briefly, ASAP-seq captures both chromatin accessibility (by Tn5 transposase binding to the chromatin) and surface protein expression (using DNA-tagged antibodies) at the same time. When Tn5 binds to HIV-1 DNA, these HIV-1 DNA+ cells can be bioinformatically identified. This study identified host transcription factor activity (derived from chromatin accessibility) in addition to surface protein expression of HIV-1 DNA+ cells [58]. Of note, the majority of these HIV-1 DNA+ cells are also likely to be defective. Further, Tn5 cannot capture HIV-1 proviruses integrated into low accessibility locations such as heterochromatin. Nevertheless, as opposed to targeted protein capture, the addition of ATAC-seq extended our understanding of HIV-1 latently infected cells from targeted surface profiling to the genome-wide level.

DOGMA-seq.—These substantial advancements still cannot capture the single-cell transcriptional landscape of latent HIV-1-infected cells. Yulong Wei in our group used DOGMA-seq (ATAC-seq, RNA-seq, and surface protein expression) and identified the single-cell epigenetic state, transcriptional program, and surface protein expression of latent and transcriptionally active [60]. Briefly, by mapping Tn5 capture of HIV-1 DNA (by ATAC-seq) and by mapping the transcriptome to HIV-1 RNA+ cells, we identified the single-cell programs of HIV-1-infected cells. The paired information of HIV-1 DNA and HIV-1 RNA identified latent (HIV-1 DNA+ RNA–) versus transcriptionally active (HIV-1 RNA+) HIV-1-infected cells. We identified four distinct cellular states of HIV-1-infected cells into cytotoxic CD4+ T cells (by transcription factor Eomes), activated cells (by interferon response factors (IRF) transcription factors), migratory cells (by AP-1 (Jun/Fos) transcription factors), and cell death. Overall, by advancing single-cell multi-omic profiling to ATAC-seq, RNA-seq, and surface protein within the same single cells, we identified the cellular state of latent and transcriptionally active HIV-1-infected cells across the central dogma of molecular biology – from DNA, RNA, to proteins.

Considerations for single-cell multi-omic experimental design

Unlike genome-wide association studies (GWAS) and clinical trials that require large numbers of study participants, single-cell multi-omic studies are typically small in sample size. By using rigorous statistical (such as using adjusted P-values by Benjamini-Hochberg procedures [61] to control for false discovery rate) and bioinformatic approaches (such as batch effect correction), single-cell multi-omic profiling from relatively few number of participants can generate biologically reproducible and meaningful insights. However, biological replicates for at least three samples, both for *in vitro* models or clinical samples, are required to reach biologically significant results, provide statistical rigor, and avoid false discovery of individual differences. To increase sequence quality, dead cells should be removed by magnetic-based depletion. Fc receptor blockade and isotype controls should be used for surface protein staining. To reduce the cost, different samples can be labeled by hashtag antibodies and pooled. The number of cells pooled should be gauged based on doublet rate, cost, and the number of cells estimated to capture the rare HIV-1-infected cells. Profiling uninfected cells (as opposed to using existing datasets) is a critical negative control to ensure HIV-1 mapping does not lead to false positive calling of HIV-1-infected cells. Finally, single-cell multi-omic results should be validated by orthogonal wet-lab experiments to cross-check whether genes identified provide biologically meaningful and reproducible results.

Bioinformatic practices for single-cell multiomics in the context of HIV-1 research

Advanced bioinformatic analysis is essential for identifying reproducible and biologically important gene expression, rather than making false-positive discovery of signatures that are biologically misleading.—The unique strength of single-cell epigenome (by ATAC-seq) and transcriptome (by RNA-seq) profiling is the genome-wide understanding of individual cells which enables discovery of new cellular factors enriched in HIV-1-infected cells, identification of mechanism of HIV-1 persistence, and nomination of therapeutic targets. The caveat is that single-cell ATAC-seq and RNA-seq results are sparse, capturing 25,000 – 50,000 reads per cell. While additional read depth can be achieved, sequencing at 25,000 reads per cell typically reach saturation, indicating that additional reads may not substantially increase additional information. ATAC-seq and RNA-seq results thus require careful bioinformatic analysis beyond following so-called "default" settings. Calling specific subsets of cells having unique overexpression of certain genes or claiming the identification of new cell types requires rigorous steps, including dead cell removal, doublet removal, batch effect correction, careful cell type annotation, and the use of statistically rigorous methods to determine differentially expressed genes. Single-cell RNA-seq datasets are sparse, frequently capturing the highly expressed genes (such as housekeeping genes). These highly expressed genes may not reflect the unique cell type and should be analyzed with caution.

Here we provide a non-exhaustive, entry-point, bioinformatics pre-processing guide for single-cell multiomics analysis, from raw reads to count matrix, removal of low-quality cells, doublets detection, data normalization, batch effect correction, cell clustering, visualization, and annotation (Table 1), following best practice guides [62, 63] and benchmarking studies [64–68] in the single-cell field.

Removal of doublets.—Droplets containing more than one cell (doublets or multiplets) often have higher RNA read. While using the number of RNA reads per cell to remove doublets is one way to remove doublets, it is not rigorous enough in the context of identification of authentic HIV-1-infected cells. Such method is subjected to technical sequencing variation and can vary by cell size [62]. Additional doublet removal methods need to be applied, by identifying doublets containing mutually exclusive canonical biomarkers [73], genotype (single nucleotide polymorphism, SNP) information from different individuals [74, 75], chromosome diploidy (ATAC)[83], or different cell hashing antibodies [74].

Normalization and batch effect correction (integration).—To ensure cellular profiles are comparable between single cells, expression datasets need to be normalized across cells to adjust for variance in sequencing depth [70]. However, normalization (different cells within the same sample) does not correct for batch effects (technical noise between experiments), which can mask true biological insights [79]. Batch effect correction methods, such as Harmony [76], fastMNN [77], or scVI [78], should be applied to examine and remove batch effects.

Data visualization.—Single-cell RNA-seq-based profiling are high-dimensional – each cell can have various (from none, to low, to high) levels of RNA expression of ~20,000 human genes. To gain biological insights, dimension reduction methods plot each cell on a two-dimension state. Principal component analysis (PCA) plots depict cellular clusters based on the top 2 principal components (highly variable genes), PCA1 versus PCA2. Yet, PCA plots are not sufficient to capture the similarity versus differences between cells. To group cell clusters that share biological similarities (eg. plotting CD4+ T cells together away from myeloid cells), cells with similar cellular profiles are grouped (graph-based construction of interconnected cell "communities" using K-nearest neighbor [89] followed by Louvain [90] or Leiden [91] modularity optimization) and then visualized in two-dimensional space using dimension-reduction graphical approaches such as UMAP [89], tSNE [92], or PHATE [93].

Cluster resolution.—The resolution of clusters, such as broad stroke separation between CD4+ T cells from CD8+ T cells versus granular separation of CD4+ T cells into different differentiation and polarization phenotypes, requires careful determination based on biological insight, not by an arbitrary default setting. For example, Clustering Trees [94] can be used to test whether the number of clusters identified reached exhaustive nomination of clusters to resolve biological heterogeneity.

Cell type annotation.—Reference mapping tools such as Azimuth [69] and scType [95] are useful to identify major cell types. However, these reference mapping tools should not be the only method used for cell type annotation. For CD4+ T cells, biological insight into the expression of key transcription factors, cytokines, effector molecules, and surface protein expression needs to be carefully examined (as gene expression violin plots or dot plots) needs to be applied for each cluster.

Integration of multi-modal single-cell data can be achieved by Weighted Nearest Neighbors (WNN) in Seurat v4 [69], a dictionary learning approach in Seurat v5 [86], or a sequential integration approach in StabMap [88].

Biological insight is required for sanity check.—Data from different biological replicates should be pooled and bioinformatically analyzed together and visualized on the same dimension reduction plot. If different biological replicate forms distinct clusters, batch effect is likely present. However, the presence of over-correction of batch effects, as shown by having biologically distinct cell types (such as *ex vivo* activated CD4+ T cells versus unstimulated CD4+ T cells) merged into the same cluster, indicates that different batch effect correction methods should be used to avoid masking biological findings. Researchers should constantly review cellular phenotype and bioinformatic threshold based on in-depth understanding of immunology, rather than false reporting novel cell types.

Identifying authentic HIV-1-infected cells requires rigorous examinations

Mapping to reference genome.—Mapping HIV-1 DNA reads to ATAC-seq and HIV-1 RNA reads to RNA-seq datasets enables unbiased identification of HIV-1-infected cells regardless of the cell type, not only in blood but also in tissues, such as microglia in the brain [96]. Yet, defining the rare HIV-1-infected cells has to be stringent and rigorous to prevent false positive discovery of cell types as a result of sequencing artifacts. By mapping ATAC-seq [58] or RNA-seq reads [25] to the HXB2 reference (for which the clinical and mechanistic implications of HIV-1 mutations are annotated), the identification of HIV-1 RNA transcripts in single cells acts as a surrogate for transcriptionally active HIV-1-infected cells [17, 25]. Increasing the breadth of reference genomes by mapping transcriptomic reads to autologous HIV-1 sequences in addition to HXB2 further increases mapping rate by approximately 20% and increases detection of spliced HIV-1 RNA [25]. HIV-1 transcripts may also be mapped to reference sequences derived from the Los Alamos HIV Database [97], with consideration for the relevant clade and tropism for the sample type and origin.

Sensitivity: the sparsity of HIV-1 reads per cell and the rarity of HIV-1-infected cells in clinical samples.—In previous studies of HIV-1+ individuals under suppressive ART, ~0.13% memory CD4+ T cells harbor HIV-1 DNA as captured by ATAC-seq [58] and ~0.02% CD4+ T cells harbor HIV-1 RNA as captured by RNA-seq [25]. Both HIV-1 DNA and RNA sequences are sparse: the ~9,719 bp of HIV-1 provirus only account for 0.00016% of the 6 × 10^9 bp of diploid human genome and may not be captured by Tn5 transposase in ATAC-seq. Further, Tn5 binds to accessible regions and cannot capture HIV-1 integrated into heterochromatin. Adding a transposase that can target heterochromatin, such as TnH3 in scGET-seq [98] may potentially target HIV-1 integrated into heterochromatin. HIV-1 RNA reads in unstimulated cells are sparse and may reflect transcription of defective HIV-1 [13]. Drop-out of transcripts or stochasticity of expression should be considered [6]. Of note, identification the short reads of HIV-1 DNA or RNA cannot infer genome intactness, unless targeted DNA amplification [57] or long-read RNA sequencing [99] identifies all genomic elements (both cis-regulatory element (such as ψ packaging signal) and protein coding regions) to be intact, without hypermutation, internal deletion, or point mutations [6]. Of note, the HIV-1 5' leader sequence and packaging signal contain secondary RNA elements that are essential for HIV-1 replication competence, such as primer binding site (PBS), dimerization initiation site (DIS), major splice donor (MSD), and packaging stem loops [100, 101]. The definition of intactness of HIV-1 genome requires biological insights into how mutations or deletions affect these cis-acting elements [6, 13].

Specificity: avoid false positive detection of HIV-1-infected cells.—Given the rarity of HIV-1-infected cells and interest in analysis of such a small population, robust validation of any reference sequence and HIV-1 mapping is imperative to minimize the effects of false positives. First, HIV-1 genome identified should be validated by mapping HIV-1 reads through NCBI Blastn [102] or HIV Blast [97] to exclude detection of human-derived transcripts or laboratory contaminants. For example, as enlisted by Liu *et al.* [17], some of HIV-1 sequences deposited in the public database contain human genome, which may lead to false discovery of HIV-1-infected cells. Second, while HIV-1 integration sites can sometimes be captured by identifying the junction of HIV-1 DNA and human genome [58] or the junction of HIV-1-host chimeric RNA through aberrant splicing [17], sequencing artifacts have to be identified and removed. There cannot be any additional sequences between the human genome and the HIV-1 genome, as these additional sequences are artifacts generated during the PCR amplification during library preparation, as described by Sherrill-Mix *et al.* in the Bushman group [103]. HIV-1-host RNA junctions have to be either immediately before the 5’ LTR, immediately after the 3’ LTR (reflecting read-through transcription), or at the canonical splice junctions between human genome and HIV-1 [17]. While HIV-1 may activate cryptic splice sites [17, 104], these HIV-1-host splice junctions should presumably follow GT|AG rules of splicing. Third, during sequencing, index hopping which incorrectly assigns reads to the wrong sample within a sequencing lane [105]. This results in false positive calling of uninfected cells as infected cells. Finally, identification of one HIV-1 read per cell may not be sufficient to define HIV-1-infected cells. The use of a minimum of 2–5 HIV-1 reads per cell is often required to correctly define HIV-1-infected cells, as defined by no detectable HIV-1 reads in uninfected samples. Overall, the use of uninfected samples is an essential negative control to set appropriate threshold for HIV-1 genome mapping.

Conclusion

Advancement in single-cell technologies revolutionized our understanding of mechanisms of HIV-1 persistence as well as immunology, cancer, development, and human diseases. The unique value for the HIV-1 field is that single-cell multi-omics is one of the few approaches that can resolve the heterogeneity and rarity of HIV-1-infected cells which do not have cellular markers for specific enrichment. Genome-wide profiling enables unbiased identification of cell types and identify innovative mechanisms. Multi-omic profiling, from epigenetic regulators (DNA by ATAC-seq), transcriptional programs (RNA by RNA-seq), to cellular markers and therapeutic targets (protein by CITE-seq), broadened our understanding of HIV-1 reservoir across the central dogma of molecular biology. TCR clonal tracking informs the clonal expansion dynamics of HIV-1-infected cells. Cautions need to be taken to make mechanistically impactful, biologically reproducible, and statistically rigorous finding. Bioinformatic analysis for single-cell multi-omics should be designed based on the biological question, rather than following default settings. Defining the rare HIV-1+ cells requires rigorous procedures to avoid sequencing and mapping artifacts. Overall, constant learning of single-cell studies from fields outside of HIV-1 research, such as bioinformatics, biotechnologies, human cell atlas, immunology, and cancer, will accelerate groundbreaking discoveries of mechanisms of HIV-1 persistence and the development of therapeutic strategies.

Financial support and sponsorship

This work is supported by Yale Top Scholar, NIH R01 AI141009, NIH R01 AI174863, NIH R61/R33 DA047037, NIH P01 AI169768, NIH R37 AI147868, NIH R01 DA051906, NIH R01AI145164, NIH UM1 DA051410, NIH U01 DA053628, NIH CHEETAH P50 AI150464, NIH REACH Martin Delaney Collaboratory UM1 AI164565, NIH BEAT-HIV Martin Delaney Collaboratory UM1 AI164570, Natural Sciences and Engineering Research Council of Canada (NSERC) Postdoctoral Fellowship (PDF557234) (Y.W.), AIDS and Cancer Specimen Resource Pilot Grant (Y.W.).

References and recommended reading

[1]. Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JA, Baseler M, Lloyd AL, Nowak MA, Fauci AS, Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy, Proc Natl Acad Sci U S A 94(24) (1997) 13193–7. [PubMed: 9371822]

[2]. 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, Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy, Science 278(5341) (1997) 1295–300. [PubMed: 9360927]

[3]. Wong JK, Hezareh M, Gunthard HF, Havlir DV, Ignacio CC, Spina CA, Richman DD, Recovery of replication-competent HIV despite prolonged suppression of plasma viremia, Science 278(5341) (1997) 1291–5. [PubMed: 9360926] *The above are the three back-to-back studies defining HIV-1 latent resevoir.

[4]. Crooks AM, Bateson R, Cope AB, Dahl NP, Griggs MK, Kuruc JD, Gay CL, Eron JJ, Margolis DM, Bosch RJ, Archin NM, Precise Quantitation of the Latent HIV-1 Reservoir: Implications for Eradication Strategies, J Infect Dis 212(9) (2015) 1361–5. [PubMed: 25877550]

[5]. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, Kovacs C, Gange SJ, Siliciano RF, Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells, Nat Med 9(6) (2003) 727–8. [PubMed: 12754504] *The above two studies both measured the half-life of the HIV-1 latent reservoir to be around 44 months.

[6]. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, Lai J, Blankson JN, Siliciano JD, Siliciano RF, Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure, Cell 155(3) (2013) 540–51. [PubMed: 24243014] **The first HIV-1 near full-length HIV-1 proviral landscape profiling study that demonstrated definitions of intact versus defective proviruses, including large internal deletions, hypermutaions, packaging signal defects, and point mutations.

[7]. Hiener B, Horsburgh BA, Eden JS, Barton K, Schlub TE, Lee E, von Stockenstrom S, Odevall L, Milush JM, Liegler T, Sinclair E, Hoh R, Boritz EA, Douek D, Fromentin R, Chomont N, Deeks SG, Hecht FM, Palmer S, Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants, Cell Rep 21(3) (2017) 813–822. [PubMed: 29045846]

[8]. Einkauf KB, Osborn M, Gao C, Parsons E, Jiang C, Lian X, Sun X, Blackmer JE, Rosenberg ES, Yu X, Lichterfeld M, Evolutionary dynamics of HIV reservoir cells via a novel single-cell multiomics assay, Conferences on Retroviruses and Opportunistic Infections (2021).

[9]. Jiang C, Lian X, Gao C, Sun X, Einkauf KB, Chevalier JM, Chen SMY, Hua S, Rhee B, Chang K, Blackmer JE, Osborn M, Peluso MJ, Hoh R, Somsouk M, Milush J, Bertagnolli LN, Sweet SE, Varriale JA, Burbelo PD, Chun TW, Laird GM, Serrao E, Engelman AN, Carrington M, Siliciano RF, Siliciano JM, Deeks SG, Walker BD, Lichterfeld M, Yu XG, Distinct viral reservoirs in individuals with spontaneous control of HIV-1, Nature 585(7824) (2020) 261–267. [PubMed: 32848246]

[10]. Einkauf KB, Lee GQ, Gao C, Sharaf R, Sun X, Hua S, Chen SM, Jiang C, Lian X, Chowdhury FZ, Rosenberg ES, Chun TW, Li JZ, Yu XG, Lichterfeld M, Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy, J Clin Invest 129(3) (2019) 988–998. [PubMed: 30688658]

[11]. Lee GQ, Orlova-Fink N, Einkauf K, Chowdhury FZ, Sun X, Harrington S, Kuo HH, Hua S, Chen HR, Ouyang Z, Reddy K, Dong K, Ndung'u T, Walker BD, Rosenberg ES, Yu XG, Lichterfeld M, Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells, J Clin Invest 127(7) (2017) 2689–2696. [PubMed: 28628034]

[12]. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, Lai J, Strain MC, Lada SM, Hoh R, Ho YC, Richman DD, Deeks SG, Siliciano JD, Siliciano RF, Defective proviruses rapidly accumulate during acute HIV-1 infection, Nat Med 22(9) (2016) 1043–9. [PubMed: 27500724]

[13]. Pollack RA, Jones RB, Pertea M, Bruner KM, Martin AR, Thomas AS, Capoferri AA, Beg SA, Huang SH, Karandish S, Hao H, Halper-Stromberg E, Yong PC, Kovacs C, Benko E, Siliciano RF, Ho YC, Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape, Cell Host Microbe 21(4) (2017) 494–506.e4. [PubMed: 28407485] **A key study demonstrating that defective HIV-1 proviruses can be transcribed and translated. When major splice donor site of HIV-1 is mutated, HIV-1 can activate cryptic HIV-1 splice donor sites and splice into canonical splice acceptor sites.

[14]. 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, Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection, Nature 387(6629) (1997) 183–188. [PubMed: 9144289]

[15]. Eriksson S, Graf EH, Dahl V, Strain MC, Yukl SA, Lysenko ES, Bosch RJ, Lai J, Chioma S, Emad F, Abdel-Mohsen M, Hoh R, Hecht F, Hunt P, Somsouk M, Wong J, Johnston R, Siliciano RF, Richman DD, O'Doherty U, Palmer S, Deeks SG, Siliciano JD, Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies, PLoS Pathog 9(2) (2013) e1003174.

[16]. Chun T-W, Justement JS, Lempicki RA, Yang J, Dennis G, Hallahan CW, Sanford C, Pandya P, Liu S, McLaughlin M, Ehler LA, Moir S, Fauci AS, Gene expression and viral prodution in latently infected, resting CD4+ T cells in viremic versus aviremic HIV-infected individuals, Proceedings of the National Academy of Sciences 100(4) (2003) 1908–1913.

[17]. Liu R, Yeh YJ, Varabyou A, Collora JA, Sherrill-Mix S, Talbot CC Jr., Mehta S, Albrecht K, Hao H, Zhang H, Pollack RA, Beg SA, Calvi RM, Hu J, Durand CM, Ambinder RF, Hoh R, Deeks SG, Chiarella J, Spudich S, Douek DC, Bushman FD, Pertea M, Ho YC, Single-cell transcriptional landscapes reveal HIV-1-driven aberrant host gene transcription as a potential therapeutic target, Sci Transl Med 12(543) (2020).***HIV-1 SortSeq is the first single-cell study using HIV-1 RNA as a surrogate to identify HIV-1-infected cells. Cells were activated with PMA/ionomycin for 16 hours ex vivo to induce HIV-1 RNA expression. This study identified HIV-1-driven aberrant transcription into the host RNA as a mechanims of HIV-1 persistence. The supplemental information provides a list of HIV-1 reads that contain human genome, which may cause false-positive discovery of HIV-1-infected cells.

[18]. Cohn LB, da Silva IT, Valieris R, Huang AS, Lorenzi JCC, Cohen YZ, Pai JA, Butler AL, Caskey M, Jankovic M, Nussenzweig MC, Clonal CD4(+) T cells in the HIV-1 latent reservoir display a distinct gene profile upon reactivation, Nat Med 24(5) (2018) 604–609. [PubMed: 29686423] ***LURE is the first single-cell study using HIV-1 Env protein as a surrogate to identify HIV-1-infected cells. Cells are activated with PHA for 40 hours ex vivo to induce HIV-1 Env protein expression.

[19]. Lee E, Bacchetti P, Milush J, Shao W, Boritz E, Douek D, Fromentin R, Liegler T, Hoh R, Deeks SG, Hecht FM, Chomont N, Palmer S, Memory CD4 + T-Cells Expressing HLA-DR Contribute to HIV Persistence During Prolonged Antiretroviral Therapy, Front Microbiol 10 (2019) 2214. [PubMed: 31611857]

[20]. 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 RP, Chomont N, CD4+ T Cells Expressing PD-1, TIGIT and LAG-3 Contribute to HIV Persistence during ART, PLoS Pathog 12(7) (2016) e1005761.

[21]. Pardons M, Baxter AE, Massanella M, Pagliuzza A, Fromentin R, Dufour C, Leyre L, Routy JP, Kaufmann DE, Chomont N, Single-cell characterization and quantification of translation-competent viral reservoirs in treated and untreated HIV infection, PLoS Pathog 15(2) (2019) e1007619.

[22]. Dufour C, Richard C, Pardons M, Massanella M, Ackaoui A, Murrell B, Routy B, Thomas R, Routy JP, Fromentin R, Chomont N, Phenotypic characterization of single CD4+ T cells harboring genetically intact and inducible HIV genomes, Nat Commun 14(1) (2023) 1115. [PubMed: 36849523]

[23]. Gantner P, Buranapraditkun S, Pagliuzza A, Dufour C, Pardons M, Mitchell JL, Kroon E, Sacdalan C, Tulmethakaan N, Pinyakorn S, Robb ML, Phanuphak N, Ananworanich J, Hsu D, Vasan S, Trautmann L, Fromentin R, Chomont N, HIV rapidly targets a diverse pool of CD4(+) T cells to establish productive and latent infections, Immunity 56(3) (2023) 653–668 e5. [PubMed: 36804957]

[24]. Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, Yassine-Diab B, Boucher G, Boulassel MR, Ghattas G, Brenchley JM, Schacker TW, Hill BJ, Douek DC, Routy JP, Haddad EK, Sekaly RP, HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation, Nat Med 15(8) (2009) 893–900. [PubMed: 19543283]

[25]. Collora JA, Liu R, Pinto-Santini D, Ravindra N, Ganoza C, Lama JR, Alfaro R, Chiarella J, Spudich S, Mounzer K, Tebas P, Montaner LJ, van Dijk D, Duerr A, Ho YC, Single-cell

multiomics reveals persistence of HIV-1 in expanded cytotoxic T cell clones, Immunity 55(6) (2022) 1013–1031 e7. ***ECCITEseq is the first method used to identify HIV-1-infected cells without ex vivo activation. ECCITEseq also mapped the clonal expansion dynamics of HIV-1+ T cell clones. This study identified enrichment of HIV-1-infected cells in proliferative cytotoxic CD4+ T cells.

[26]. Weymar GHJ, Bar-On Y, Oliveira TY, Gaebler C, Ramos V, Hartweger H, Breton G, Caskey M, Cohn LB, Jankovic M, Nussenzweig MC, Distinct gene expression by expanded clones of quiescent memory CD4(+) T cells harboring intact latent HIV-1 proviruses, Cell Rep 40(10) (2022) 111311. ***This study used TCR Vβ to sort for CD4+ T cell clones known to enrich for HIV-1-infected cells. This study also identified enrichemnt of HIV-1-infected cells in cytotoxic CD4+ T cells.

[27]. Perreau M, Savoye A-L, De Crignis E, Corpataux J-M, Cubas R, Haddad EK, De Leval L, Graziosi C, Pantaleo G, Follicular helper T cells serve as the major CD4 T cell compartment for HIV-1 infection, replication, and production, Journal of Experimental Medicine 210(1) (2012) 143–156. [PubMed: 23254284]

[28]. Kuo HH, Ahmad R, Lee GQ, Gao C, Chen HR, Ouyang Z, Szucs MJ, Kim D, Tsibris A, Chun TW, Battivelli E, Verdin E, Rosenberg ES, Carr SA, Yu XG, Lichterfeld M, Anti-apoptotic Protein BIRC5 Maintains Survival of HIV-1-Infected CD4(+) T Cells, Immunity 48(6) (2018) 1183–1194 e5. [PubMed: 29802019]

[29]. Ren Y, Huang SH, Patel S, Alberto WDC, Magat D, Ahimovic D, Macedo AB, Durga R, Chan D, Zale E, Mota TM, Truong R, Rohwetter T, McCann CD, Kovacs CM, Benko E, Wimpelberg A, Cannon C, Hardy WD, Bosque A, Bollard CM, Jones RB, BCL-2 antagonism sensitizes cytotoxic T cell-resistant HIV reservoirs to elimination ex vivo, J Clin Invest 130(5) (2020) 2542–2559. [PubMed: 32027622]

[30]. Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, Olsen BN, Mumbach MR, Pierce SE, Corces MR, Shah P, Bell JC, Jhutty D, Nemec CM, Wang J, Wang L, Yin Y, Giresi PG, Chang ALS, Zheng GXY, Greenleaf WJ, Chang HY, Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion, Nat Biotechnol 37(8) (2019) 925–936. [PubMed: 31375813]

[31]. Lorenzi JC, Cohen YZ, Cohn LB, Kreider EF, Barton JP, Learn GH, Oliveira T, Lavine CL, Horwitz JA, Settler A, Jankovic M, Seaman MS, Chakraborty AK, Hahn BH, Caskey M, Nussenzweig MC, Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA, Proc Natl Acad Sci U S A 113(49) (2016) E7908–E7916. [PubMed: 27872306]

[32]. Bui JK, Sobolewski MD, Keele BF, Spindler J, Musick A, Wiegand A, Luke BT, Shao W, Hughes SH, Coffin JM, Kearney MF, Mellors JW, Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir, PLoS Pathog 13(3) (2017) e1006283.

[33]. Hosmane NN, Kwon KJ, Bruner KM, Capoferri AA, Beg S, Rosenbloom DI, Keele BF, Ho YC, Siliciano JD, Siliciano RF, Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics, J Exp Med 214(4) (2017) 959–972. [PubMed: 28341641]

[34]. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, Spindler J, Ferris AL, Mellors JW, Kearney MF, Coffin JM, Hughes SH, HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells, Science 345(6193) (2014) 179–83. [PubMed: 24968937]

[35]. Wagner TA, McLaughlin S, Garg K, Cheung CY, Larsen BB, Styrchak S, Huang HC, Edlefsen PT, Mullins JI, Frenkel LM, HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection, Science (New York, N.Y.) 345(6196) (2014) 570–573. [PubMed: 25011556]

[36]. Einkauf KB, Osborn MR, Gao C, Sun W, Sun X, Lian X, Parsons EM, Gladkov GT, Seiger KW, Blackmer JE, Jiang C, Yukl SA, Rosenberg ES, Yu XG, Lichterfeld M, Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses, Cell (2022).

[37]. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D, Trombetta JJ, Gennert D, Gnirke A, Goren A, Hacohen N, Levin

JZ, Park H, Regev A, Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells, Nature 498(7453) (2013) 236–40. [PubMed: 23685454]

[38]. Sallusto F, Cassotta A, Hoces D, Foglierini M, Lanzavecchia A, Do Memory CD4 T Cells Keep Their Cell-Type Programming: Plasticity versus Fate Commitment? T-Cell Heterogeneity, Plasticity, and Selection in Humans, Cold Spring Harb Perspect Biol 10(3) (2018).

[39]. Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, McNamara KL, Granja JM, Sarin KY, Brown RA, Gupta RK, Curtis C, Bucktrout SL, Davis MM, Chang ALS, Chang HY, Clonal replacement of tumor-specific T cells following PD-1 blockade, Nat Med 25(8) (2019) 1251–1259. [PubMed: 31359002] ***This is an important study demonstrating that tracking T cell clones and T cell phenotype between blood and the tumor identified that the effect of PD-1 blockade on reinvigorating tumor-infiltrating T cells.

[40]. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ, Single-cell chromatin accessibility reveals principles of regulatory variation, Nature 523(7561) (2015) 486–90. [PubMed: 26083756]

[41]. Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY, Greenleaf WJ, Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation, Cell 173(6) (2018) 1535–1548.e16. [PubMed: 29706549]

[42]. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position, Nat Methods 10(12) (2013) 1213–8. [PubMed: 24097267]

[43]. Han A, Glanville J, Hansmann L, Davis MM, Linking T-cell receptor sequence to functional phenotype at the single-cell level, Nat Biotechnol 32(7) (2014) 684–92. [PubMed: 24952902]

[44]. Mimitou EP, Cheng A, Montalbano A, Hao S, Stoeckius M, Legut M, Roush T, Herrera A, Papalexi E, Ouyang Z, Satija R, Sanjana NE, Koralov SB, Smibert P, Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells, Nat Methods 16(5) (2019) 409–412. **ECCITEseq captures RNA-seq, CITE-seq, and TCR.

[45]. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, Smibert P, Simultaneous epitope and transcriptome measurement in single cells, Nat Methods 14(9) (2017) 865–868. [PubMed: 28759029] **CITEseq captures RNA-seq and surface protein expression.

[46]. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, Moore R, McClanahan TK, Sadekova S, Klappenbach JA, Multiplexed quantification of proteins and transcripts in single cells, Nature Biotechnology 35(10) (2017) 936–939.

[47]. Gantner P, Buranapraditkun S, Pagliuzza A, Dufour C, Pardons M, Mitchell JL, Kroon E, Sacdalan C, Tulmethakaan N, Pinyakorn S, Robb ML, Phanuphak N, Ananworanich J, Hsu D, Vasan S, Trautmann L, Fromentin R, Chomont N, HIV rapidly targets a diverse pool of CD4(+) T cells to establish productive and latent infections, Immunity 56(3) (2023) 653–668.e5. [PubMed: 36804957]

[48]. Katzenelenbogen Y, Sheban F, Yalin A, Yofe I, Svetlichnyy D, Jaitin DA, Bornstein C, Moshe A, Keren-Shaul H, Cohen M, Wang SY, Li B, David E, Salame TM, Weiner A, Amit I, Coupled scRNA-Seq and Intracellular Protein Activity Reveal an Immunosuppressive Role of TREM2 in Cancer, Cell 182(4) (2020) 872–885.e19.

[49]. Bennett HM, Stephenson W, Rose CM, Darmanis S, Single-cell proteomics enabled by next-generation sequencing or mass spectrometry, Nature Methods 20(3) (2023) 363–374. [PubMed: 36864196]

[50]. Satpathy AT, Saligrama N, Buenrostro JD, Wei Y, Wu B, Rubin AJ, Granja JM, Lareau CA, Li R, Qi Y, Parker KR, Mumbach MR, Serratelli WS, Gennert DG, Schep AN, Corces MR, Khodadoust MS, Kim YH, Khavari PA, Greenleaf WJ, Davis MM, Chang HY, Transcript-indexed ATAC-seq for precision immune profiling, Nat Med 24(5) (2018) 580–590. [PubMed: 29686426] **T-ATAC-seq combines single-cell RNA-seq and ATAC-seq.

[51]. Mimitou EP, Lareau CA, Chen KY, Zorzetto-Fernandes AL, Hao Y, Takeshima Y, Luo W, Huang T-S, Yeung BZ, Papalexi E, Thakore PI, Kibayashi T, Wing JB, Hata M, Satija R, Nazor KL, Sakaguchi S, Ludwig LS, Sankaran VG, Regev A, Smibert P, Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells, Nature Biotechnology

39(10) (2021) 1246–1258. ***This paper described both ASAP-seq (ATAC-seq + CITE-seq) and DOGMA-sea (ATAC-seq + RNA-seq + CITE-seq).

[52]. Swanson E, Lord C, Reading J, Heubeck AT, Genge PC, Thomson Z, Weiss MDA, Li X.-j., Savage AK, Green RR, Torgerson TR, Bumol TF, Graybuck LT, Skene, Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq, eLife 10 (2021) e63632. [PubMed: 33835024] ***TEA-seq (ATAC-seq + RNA-seq + CITE-seq) and DOGMA-seq are conceptually similar.

[53]. Denisenko E, Guo BB, Jones M, Hou R, de Kock L, Lassmann T, Poppe D, Clément O, Simmons RK, Lister R, Forrest ARR, Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows, Genome Biology 21(1) (2020) 130. [PubMed: 32487174]

[54]. Wohnhaas CT, Leparc GG, Fernandez-Albert F, Kind D, Gantner F, Viollet C, Hildebrandt T, Baum P, DMSO cryopreservation is the method of choice to preserve cells for droplet-based single-cell RNA sequencing, Scientific Reports 9(1) (2019) 10699.

[55]. Kaiserman D, Bird PI, Control of granzymes by serpins, Cell Death & Differentiation 17(4) (2010) 586–595. [PubMed: 19893573]

[56]. Clark IC, Mudvari P, Thaploo S, Smith S, Abu-Laban M, Hamouda M, Theberge M, Shah S, Ko SH, Pérez L, Bunis DG, Lee JS, Kilam D, Zakaria S, Choi S, Darko S, Henry AR, Wheeler MA, Hoh R, Butrus S, Deeks SG, Quintana FJ, Douek DC, Abate AR, Boritz EA, HIV silencing and cell survival signatures in infected T cell reservoirs, Nature 614(7947) (2023) 318–325. [PubMed: 36599978] ***FIND-seq used HIV-1 DNA PCR-activated sorting to profile bulk RNA transcriptome (pool of 100 cells).

[57]. Sun W, Gao C, Hartana CA, Osborn MR, Einkauf KB, Lian X, Bone B, Bonheur N, Chun TW, Rosenberg ES, Walker BD, Yu XG, Lichterfeld M, Phenotypic signatures of immune selection in HIV-1 reservoir cells, Nature 614(7947) (2023) 309–317. [PubMed: 36599977] ***PheP-seq used targeted HIV-1 DNA amplification to capture near full-length of HIV-1 genome and determine the intact versus defective HIV-1 proviruses, with paired information of surface protein expression. In participants having known HIV-1-integration sites, primers were designed to capture the specific integration sites.

[58]. Wu VH, Nordin JML, Nguyen S, Joy J, Mampe F, Del Rio Estrada PM, Torres-Ruiz F, González-Navarro M, Luna-Villalobos YA, Ávila-Ríos S, Reyes-Terán G, Tebas P, Montaner LJ, Bar KJ, Vella LA, Betts MR, Profound phenotypic and epigenetic heterogeneity of the HIV-1-infected CD4(+) T cell reservoir, Nat Immunol 24(2) (2023) 359–370. [PubMed: 36536105] ***ASAP-seq captures HIV-1 DNA by ATAC-seq with paired protein profiles.

[59]. Wang W, Fasolino M, Cattau B, Goldman N, Kong W, Frederick MA, McCright SJ, Kiani K, Fraietta JA, Vahedi G, Joint profiling of chromatin accessibility and CAR-T integration site analysis at population and single-cell levels, Proc Natl Acad Sci U S A 117(10) (2020) 5442–5452. [PubMed: 32094195] **Initial concept of using ATAC-seq to map lentiviral DNA.

[60]. Wei YD, Timothy C; Collora Jack A; Ma Haocong K.; Pinto-Santini Delia; Lama Javier; Alfaro Ricardo; Duerr Ann C.; Ho Y-C, Single-cell epigenetic, transcriptional, and protein states of HIV reservoir, Conferences on Retroviruses and Opportunistic Infections, Late Breaker Abstract 142, Seattle, WA, 2023. ***DOGMAseq captures both latent and transcriptionally active HIV-1 DNA with paired ATAC-seq, RNA-seq, and CITE-seq.

[61]. Benjamini Y, Hochberg Y, Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society Series B-Statistical Methodology 57(1) (1995) 289–300. *Single-cell multi-omic analysis needs to be adjusted by multiple hypothesis testing. An adjusted P value by Benjamini-Hochberg procedure is commonly used (as opposed to Bonferonni correction).

[62]. Heumos L, Schaar AC, Lance C, Litinetskaya A, Drost F, Zappia L, Lücken MD, Strobl DC, Henao J, Curion F, Aliee H, Ansari M, Badia-i-Mompel P, Büttner M, Dann E, Dimitrov D, Dony L, Frishberg A, He D, Hediyeh-zadeh S, Hetzel L, Ibarra IL, Jones MG, Lotfollahi M, Martens LD, Müller CL, Nitzan M, Ostner J, Palla G, Patro R, Piran Z, Ramírez-Suástegui C, Saez-Rodriguez J, Sarkar H, Schubert B, Sikkema L, Srivastava A, Tanevski J, Virshup I, Weiler P, Schiller HB, Theis FJ, Single-cell Best Practices C, Best practices for single-cell analysis

across modalities, Nature Reviews Genetics (2023). ***An updated and clear comparison of best practices of signle-cell bioinformatic analysis.

[63]. Luecken MD, Theis FJ, Current best practices in single-cell RNA-seq analysis: a tutorial, Molecular Systems Biology 15(6) (2019) e8746. ***An earlier version of single-cell bioinformatic analysis, which is still very helpful.

[64]. You Y, Tian L, Su S, Dong X, Jabbari JS, Hickey PF, Ritchie ME, Benchmarking UMI-based single-cell RNA-seq preprocessing workflows, Genome Biology 22(1) (2021) 339.

[65]. Xi NM, Li JJ, Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data, Cell Systems 12(2) (2021) 176–194.e6. [PubMed: 33338399] **Comparison of doublet detection methods. Using RNA account alone to detect doublets is not rigorous enough and may lead to false discovery of cell types (that are actually doublets).

[66]. Yu L, Cao Y, Yang JYH, Yang P, Benchmarking clustering algorithms on estimating the number of cell types from single-cell RNA-sequencing data, Genome Biology 23(1) (2022) 49. [PubMed: 35135612]

[67]. Soneson C, Robinson MD, Bias, robustness and scalability in single-cell differential expression analysis, Nature Methods 15(4) (2018) 255–261. [PubMed: 29481549] **Comparison of differential expression analysis methods.

[68]. Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M, Chen J, A benchmark of batch-effect correction methods for single-cell RNA sequencing data, Genome Biology 21(1) (2020) 12. [PubMed: 31948481] **Comparison of differential expression analysis methods.

[69]. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R, Integrated analysis of multimodal single-cell data, Cell 184(13) (2021) 3573–3587 e29. [PubMed: 34062119] **Seurat v4 described the use of Weighted Nearest Neighbor (WNN) to integrate multiple modalities together.

[70]. Ahlmann-Eltze C, Huber W, Comparison of transformations for single-cell RNA-seq data, Nature Methods 20(5) (2023) 665–672. [PubMed: 37037999]

[71]. Lun ATL, Bach K, Marioni JC, Pooling across cells to normalize single-cell RNA sequencing data with many zero counts, Genome Biology 17(1) (2016) 75. [PubMed: 27122128]

[72]. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, Wan E, Wong S, Byrnes L, Lanata CM, Gate RE, Mostafavi S, Marson A, Zaitlen N, Criswell LA, Ye CJ, Multiplexed droplet single-cell RNA-sequencing using natural genetic variation, Nature Biotechnology 36(1) (2018) 89–94.

[73]. Germain PL, Lun A, Garcia Meixide C, Macnair W, Robinson MD, Doublet identification in single-cell sequencing data using scDblFinder, F1000Res 10 (2021) 979. [PubMed: 35814628] *scDblFinder is highly recoomended in the Best Practice recoomendations (Heumos 2023) to remove doublets.

[74]. McGinnis CS, Murrow LM, Gartner ZJ, DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors, Cell systems 8(4) (2019) 329–337.e4. [PubMed: 30954475]

[75]. Wolock SL, Lopez R, Klein AM, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data, Cell Systems 8(4) (2019) 281–291.e9. [PubMed: 30954476]

[76]. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, P.-r. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony, Nature Methods 16(12) (2019) 1289–1296. [PubMed: 31740819]

[77]. Haghverdi L, Lun ATL, Morgan MD, Marioni JC, Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors, Nat Biotechnol 36(5) (2018) 421–427. [PubMed: 29608177]

[78]. Gayoso A, Lopez R, Xing G, Boyeau P, Valiollah Pour Amiri V, Hong J, Wu K, Jayasuriya M, Mehlman E, Langevin M, Liu Y, Samaran J, Misrachi G, Nazaret A, Clivio O, Xu C, Ashuach T, Gabitto M, Lotfollahi M, Svensson V, da Veiga Beltrame E, Kleshchevnikov V, Talavera-López C, Pachter L, Theis FJ, Streets A, Jordan MI, Regier J, Yosef N, A Python library for probabilistic analysis of single-cell omics data, Nature Biotechnology 40(2) (2022) 163–166.

[79]. Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF, Strobl DC, Zappia L, Dugas M, Colomé-Tatché M, Theis FJ, Benchmarking atlas-level data integration in single-cell genomics, Nat Methods 19(1) (2022) 41–50. [PubMed: 34949812]

[80]. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS, Model-based analysis of ChIP-Seq (MACS), Genome Biol 9(9) (2008) R137. [PubMed: 18798982]

[81]. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R, Single-cell chromatin state analysis with Signac, Nat Methods 18(11) (2021) 1333–1341. [PubMed: 34725479]

[82]. Bravo González-Blas C, Minnoye L, Papasokrati D, Aibar S, Hulselmans G, Christiaens V, Davie K, Wouters J, Aerts S, cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data, Nat Methods 16(5) (2019) 397–400. [PubMed: 30962623]

[83]. Thibodeau A, Eroglu A, McGinnis CS, Lawlor N, Nehar-Belaid D, Kursawe R, Marches R, Conrad DN, Kuchel GA, Gartner ZJ, Banchereau J, Stitzel ML, Cicek AE, Ucar D, AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data, Genome Biology 22(1) (2021) 252. [PubMed: 34465366]

[84]. Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ, Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity, Cell 177(7) (2019) 1873–1887.e17. [PubMed: 31178122]

[85]. Johnson WE, Li C, Rabinovic A, Adjusting batch effects in microarray expression data using empirical Bayes methods, Biostatistics 8(1) (2007) 118–27. [PubMed: 16632515]

[86]. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, Srivastava A, Molla G, Madad S, Fernandez-Granda C, Satija R, Dictionary learning for integrative, multimodal and scalable single-cell analysis, Nature Biotechnology (2023). ***Seurat v5 for fast integration of large datasets and multiple modalities.

[87]. Mulè MP, Martins AJ, Tsang JS, Normalizing and denoising protein expression data from droplet-based single cell profiling, Nat Commun 13(1) (2022) 2099. [PubMed: 35440536]

[88]. Ghazanfar S, Guibentif C, Marioni JC, Stabilized mosaic single-cell data integration using unshared features, Nature Biotechnology (2023).***This is a back-to-back study with Seurat v5, also providing integration of multiple modalities.

[89]. McInnes LH, Melville J,J, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, arXivLabs (2018) arXiv:1802.03426.

[90]. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 2008(10) (2008) P10008.

[91]. Traag VA, Waltman L, van Eck NJ, From Louvain to Leiden: guaranteeing well-connected communities, Scientific Reports 9(1) (2019) 5233. [PubMed: 30914743]

[92]. van der Maaten LH, G, Visualizing Data using t-SNE, Journal of Machine Learning Research 9(86) (2008) 2579–2605.

[93]. Moon KR, van Dijk D, Wang Z, Gigante S, Burkhardt DB, Chen WS, Yim K, Elzen AVD, Hirn MJ, Coifman RR, Ivanova NB, Wolf G, Krishnaswamy S, Visualizing structure and transitions in high-dimensional biological data, Nat Biotechnol 37(12) (2019) 1482–1492. [PubMed: 31796933]

[94]. Zappia L, Oshlack A, Clustering trees: a visualization for evaluating clusterings at multiple resolutions, GigaScience 7(7) (2018).***This is a back-to-back study with Seurat v5, also providing integration of multiple modalities.

[95]. Ianevski A, Giri AK, Aittokallio T, Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data, Nature Communications 13(1) (2022) 1246.

[96]. Plaza-Jennings AL, Valada A, O'Shea C, Iskhakova M, Hu B, Javidfar B, Ben Hutta G, Lambert TY, Murray J, Kassim B, Chandrasekaran S, Chen BK, Morgello S, Won H, Akbarian S, HIV integration in the human brain is linked to microglial activation and 3D genome remodeling, Mol Cell 82(24) (2022) 4647–4663.e8. [PubMed: 36525955]

[97]. Apetrei CH, Rambaut B, Wolinsky A, Brister S, Keele JR, Faser B, C, HIV Sequence Compendium, 2021.

[98]. Tedesco M, Giannese F, Lazarević D, Giansanti V, Rosano D, Monzani S, Catalano I, Grassi E, Zanella ER, Botrugno OA, Morelli L, Panina Bordignon P, Caravagna G, Bertotti A, Martino G, Aldrighetti L, Pasqualato S, Trusolino L, Cittaro D, Tonon G, Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin, Nature Biotechnology 40(2) (2022) 235–244.

[99]. Gallardo CM, Wang S, Montiel-Garcia DJ, Little SJ, Smith DM, Routh AL, Torbett BE, MrHAMER yields highly accurate single molecule viral sequences enabling analysis of intra-host evolution, Nucleic Acids Res 49(12) (2021) e70. [PubMed: 33849057] ***This is an elegant study using Nanopore long-range sequencing to capture HIV-1 RNA.

[100]. Keane SC, Heng X, Lu K, Kharytonchyk S, Ramakrishnan V, Carter G, Barton S, Hosic A, Florwick A, Santos J, Bolden NC, McCowin S, Case DA, Johnson BA, Salemi M, Telesnitsky A, Summers MF, RNA structure. Structure of the HIV-1 RNA packaging signal, Science 348(6237) (2015) 917–21. [PubMed: 25999508]

[101]. Keane SC, Van V, Frank HM, Sciandra CA, McCowin S, Santos J, Heng X, Summers MF, NMR detection of intermolecular interaction sites in the dimeric 5'-leader of the HIV-1 genome, Proc Natl Acad Sci U S A 113(46) (2016) 13033–13038. [PubMed: 27791166]

[102]. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, Basic local alignment search tool, J Mol Biol 215(3) (1990) 403–10. [PubMed: 2231712]

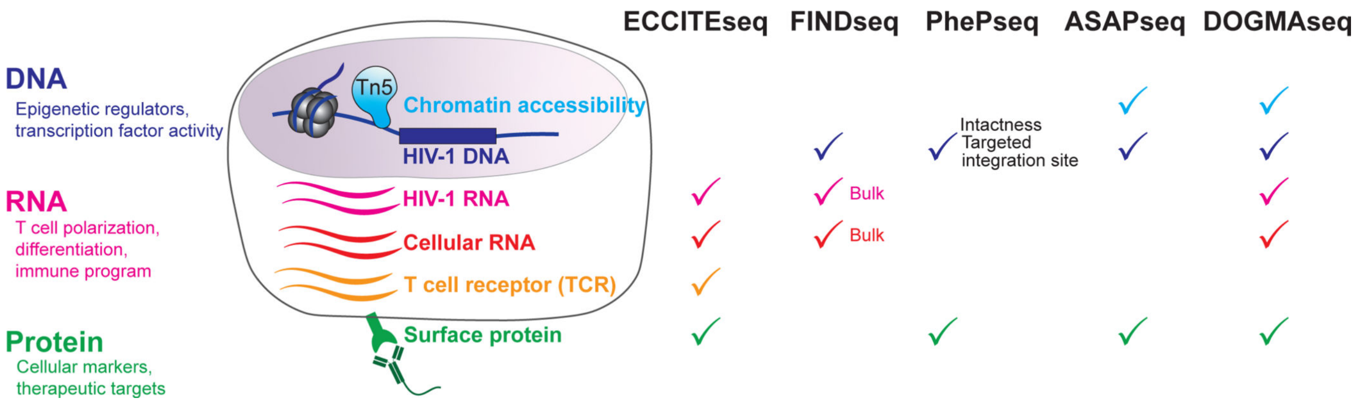
[103]. Sherrill-Mix S, Ocwieja KE, Bushman FD, Gene activity in primary T cells infected with HIV89.6: intron retention and induction of genomic repeats, Retrovirology 12 (2015) 79. [PubMed: 26377088] ***The Additional File 7 in this paper described how to identify unknown sequences between human and HIV-1 reads reflect that are sequencing artifacts.

[104]. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD, Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing, Nucleic Acids Res 40(20) (2012) 10345–55. [PubMed: 22923523]

[105]. Farouni R, Djambazian H, Ferri LE, Ragoussis J, Najafabadi HS, Model-based analysis of sample index hopping reveals its widespread artifacts in multiplexed single-cell RNA-sequencing, Nature Communications 11(1) (2020) 2704. **Index hopping is a critical concern when calling the rare HIV-1-infected cells. Caution needs to be taken to distinguish false discovery of HIV-1-infected cells because of PCR artifacts that generates index hopping.

A box of key points

- Single-cell multi-omic profiling, from epigenetic regulators (DNA by ATAC-seq), transcriptional programs (RNA by RNA-seq), to cellular markers and therapeutic targets (protein by CITE-seq), broadened our understanding of HIV-1 reservoir across the central dogma of molecular biology.
- TCR clonal tracking identifies the clonal expansion dynamics of HIV-1-infected cells.
- Bioinformatic analysis should be designed based on the biological question with stringent quality control and rigorous statistical testing, rather than following default settings.
- Defining the rare HIV-1+ cells requires rigorous procedures to avoid false-positive discoveries because of sequencing and mapping artifacts.



	ECCITEseq	FINDseq	PhePseq	ASAPseq	DOGMAseq
Chromatin accessibility				✓	✓
HIV-1 DNA		✓	✓ Intactness Targeted integration site	✓	✓
HIV-1 RNA	✓	✓ Bulk			✓
Cellular RNA	✓	✓ Bulk			✓
T cell receptor (TCR)	✓				
Surface protein	✓		✓	✓	✓

Figure 1.

Single-cell multi-omics understanding of HIV-1 reservoir at the epigenetics, transcriptional, and protein level

Table 1.

Guide for single-cell multi-omic bioinformatic analysis

Processing scRNA-seq datasets	**Tools recommended**	**Description**
Generate count matrix of cells by genes	Cellranger (10X Genomics)	Single-cell RNA
Remove low-quality cells	Seurat [69]	High mitochondrial gene content, low number of gene features, low unique molecular identifier (UMI) count per cell
Data normalization	Logarithm transformation [70], Scran [71]	See [70] for a benchmark comparison of normalization approaches
Demultiplexing	Freemuxlet (github), Demuxlet [72]	Genotype-based demultiplexing approaches for pooled samples
Doublet removal	scDblFinder [73], DoubletFinder [74], Scrublet [75]	Compare real cells to artificially generated doublets
Batch effect correction	Harmony [76], fastMNN [77], scVI [78]	See [79] for benchmark comparisons of different integration methods.
Processing scATAC-seq datasets	**Tools recommended**	**Description**
Generate count matrix of cells by peaks	Cellranger-ARC (10X Genomis)	Single-cell ATAC Single-cell multiome ATAC + gene expression
Peak recall	MACS2 [80]	Refine Cellranger-ARC defined regions of open chromatin
Remove low-quality cells	Signac [81]	Low transcription start site (TSS) enrichment score, low peak region fragment count per cell, poor nucleosome signal
Data normalization	Latent Semantic Indexing (Signac [81] Latent Dirichlet allocation (cisTopic [82])	Normalize across cells to correct for differences in sequencing depth and peak counts
Doublet removal	scDblFinder [73], AMULET [83]	AMULET identifies cells that violate chromosome diploidy (high number of positions with read count > 2).
Batch effect correction	LIGER [84], ComBat [85]	See [79] for benchmark comparisons of different integration methods.
Processing surface protein expression (CITE-seq) datasets	**Tools recommended**	**Description**
Generate count matrix of cells by genes	Cellranger (10X Genomics)	Single cell gene expression with feature barcoding
Data normalization	Centered Log Ratio transformation (CLR, Seurat)[86], DSB [87]	DSB leverages empty droplets to determine background noise and isotype expression to correct cell to cell expression variations
Demultiplexing and doublet removal	MULTIseqDemux [74], HTODemux [45](Seurat)	When hashtag oligos (HTOs) are used for pooled samples
Batch effect correction	Reciprocal PCA (Seurat)[86]	Conservative integration approach for high biological state variability from cell to cell
Integrating multi-modal data	**Tools recommended**	**Description**
	Weighted Nearest Neighbors (WNN) in Seurat v4 [69], Seurat v5 [86], or StabMap [88]	Instead of having ATAC-seq, RNA-seq, and CITE-seq in separate parts, combining cell features of different modalities into one plot provides integrated understanding of cell states