Towards single cell transcriptomics in a primary T-cell model of HIV latency

Lesley De Armas, Li Pan,
Siôn Williams, Miami CFAR
Robert Suter, Sylvester Comprehensive Cancer Center
John Harrington, BD Genomics
Mentor: Savita Pahwa, Miami CFAR

UHealth University of Miami Health System | University of Miami Miller School of Medicine

Institute of AIDS and Emerging Infectious Diseases

Abstract

Latent HIV reservoirs represent the largest barrier to cure of HIV/AIDS. CD4+ T cells are the best characterized latent HIV reservoir, yet mechanisms of latency establishment remain poorly understood. In addition, the frequency of latently-infected CD4+ T cells is extremely low in vivo, hindering ex vivo studies. To overcome these hurdles, in vitro models of latency have been developed. Bulk cell studies have revealed differential gene expression between latent and uninfected T cells. Gene expression signatures derived by bulk cell analysis can only provide an average output of potentially heterogeneous cell populations thus masking mechanisms that may be at play within different sub-populations. Single cell gene expression analysis can uncover heterogeneity within phenotypically defined cell subsets. We hypothesize that heterogeneous cell-states or cell-types may respond differently to HIV-1 infection during the establishment of, or release from, latency and that understanding these differences may reveal novel etiology. Our project encompasses two specific aims:

Aim 1. Explore transcriptional heterogeneity in latent HIV-infected CD4+ T cells.

Aim 2. Investigate transcriptional heterogeneity in response to the LRAs IL-15 and Prostratin.

Latency Model

We use a primary cell model of HIV latency in CD4+ T cells. Mononuclear cells derived from tonsil are used as a source of CD4+ T cells. These are infected with a dual reporter virus, DuoFluo, that allows identification and of latent and productively infected cells. mKOrange expression driven by a constitutive EF1a promoter identifies infected cells and eGFP expression driven by the LTR promoter identifies cells with high levels of HIV transcription (**Fig 1.**). Hence productive cells express eGFP and mKO, latent cells mKO only and uninfected cells no fluorescence (**Fig 2.**).

Figure 1. DuoFluo reporter construct

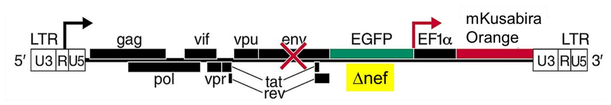
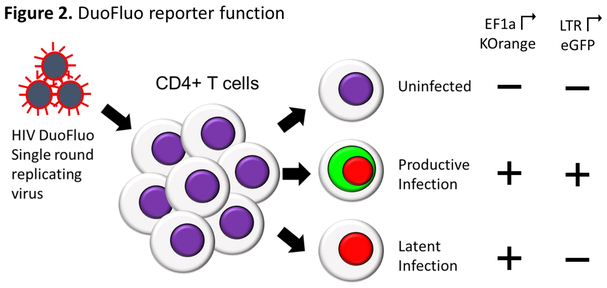


Figure 2. DuoFluo reporter function



The DuoFluo strategy permits collection of abundant uninfected cells and both low abundance latently infected cells and productive cells using flow cytometry from a single round of infection (**Fig 3.**).

Cell Subsets & Infection

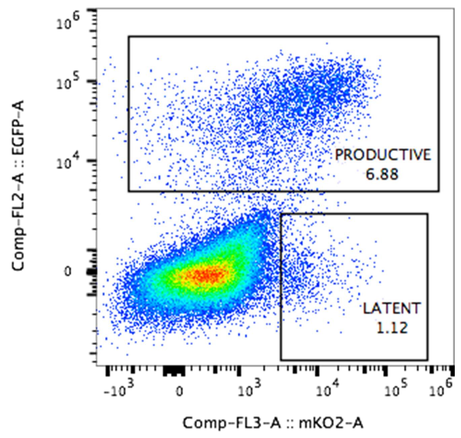
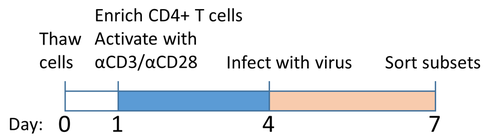


Figure 3. Sorting strategy for productive, latent and uninfected cells

We have developed a 7 day protocol for isolation of the three different cell subsets from a single CD4+ T cell enrichment (**Fig 4.**). T cells are enriched and activated on day 1 and infected with virus at 100ng p24 per 10^6 cells by spinoculation on day 4. Subsets are sorted for analysis on day 7. All experiments in the study will use mononuclear cells derived from the same HIV-1 negative tonsil donor.

Figure 4. Experimental timeline for collection of cell subsets



Single Cell Data Capture

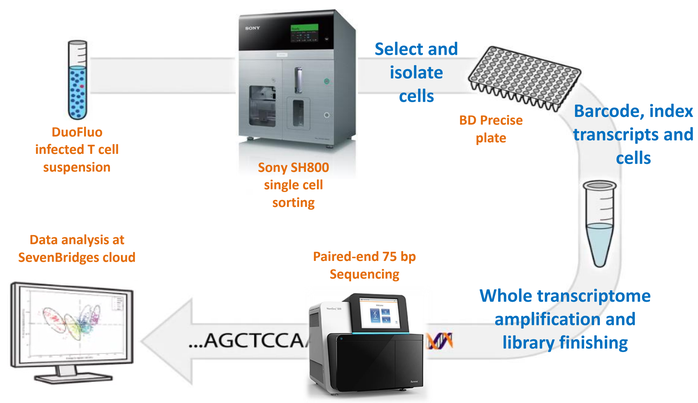


Figure 5. Single cell whole transcriptome data generation using Sony SH800 FACS and BD Precise 3' WTA library prep

To achieve efficient single cell capture we made use of the single cell index sorting capacities of the Sony SH800 FACS sorter at Miami CFAR (**Fig 5.**). Single cell capture into 96-well plates was combined with novel 3' whole transcriptome amplification library prep chemistry from BD Genomics (Precise 3'-WTA). This approach produces low data volume libraries by only sequencing the terminal 3' 75 bp of mRNAs and uses molecular indexing to produce accurate gene expression counts. Initial data analysis was carried out using an automated pipeline at SevenBridges Genomics.

Initial Data

Initial data was collected from a complete experimental run collecting 32 cells of each subset into a single 96-well plate for 3' WTA library prep, sequencing and analysis. Data sanitization, removed cells with extreme expression profiles and genes with high expression or low variance (dispersion scoring and feature selection). This resulted in analysis of 707 genes across 93 cells, 31 from each subset (**Table 1.**).

Table 1.

Parameter	Raw Data	Sanitized Data
Cells	96	93
Genes screened	55,765	707
Genes detected	16,586	707
Median barcodes / cell	7,881	147
Maximum barcodes / cell	16,388	531
Median genes / cell	2,074	59
Maximum genes / cell	3,760	140

Principle component analysis (PCA) revealed an overlay of uninfected and latent profiles and slight separation of productive profiles (**Fig 6A.**). This pattern was anticipated from previous work although it is clear that more data is needed to investigate heterogeneity within the subsets. In addition, alternate visualization such as t-distributed stochastic neighbor embedding (tSNE), may be more appropriate.

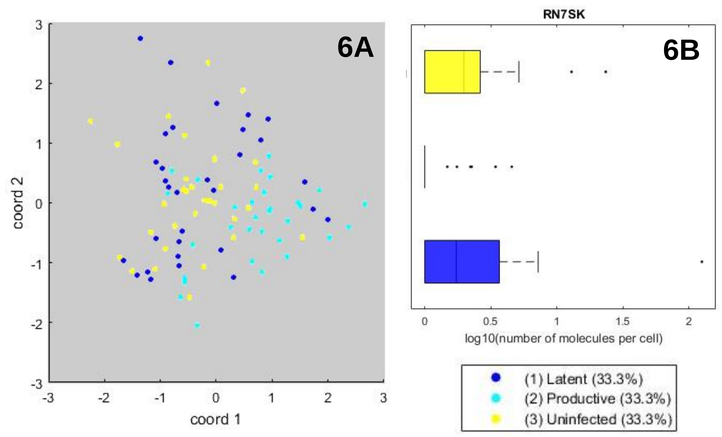


Figure 6A. Principle component analysis (PCA). **6B.** Expression of *RN7SK* across subsets, color key identical for both panels

Preliminary analysis identified some genes involved in HIV-1 biology. For instance, the non-coding RNA RN7SK was expressed at lower levels in productive cells than latent or uninfected cells (**Fig 6B.**). This RNA forms the 7SK small nuclear ribonucleoprotein (snRNP) complex that inhibits early transcript elongation by sequestering the P-TEFb elongation factor. HIV-1 Tat is known to release P-TEFb from the 7SK snRNP complex to promote viral transcription.

Prospects

We now have the experimental procedures to complete Aim 1. Following this Aim 2 will be undertaken using an identical strategy but taking only the sorted latent cell population from day 7 (**Fig 4.**) and exposing them to the LRAs IL-15 or Prostratin, or vehicle, for 24 hrs prior to single cell analysis.

This work is supported by an IAEID State Pilot Award to S Williams (State Contract CODMR).