HIV-1 gene expression during reversal of latency using RNAseq with probe-enrichment

Phillip Tomezsko[1], Silvi Rouskin[2], Daniel Kuritzkes[3,4], Athe Tsibris[3,4]

1. Program in Virology, Harvard Medical School, Boston, MA, USA 2. Whitehead Institute, Cambridge, MA, USA 3. Brigham and Women's Hospital, Boston, MA, USA 4.Harvard Medical School, Boston, MA, USA

00356

Abstract

Background- Transcriptomic analysis of the human and HIV-1 expression profile that is essential for successful reactivation of latently infected cells promises to help inform the next generation of latency reversal agents. However, because of the rarity of latently infected cells, the HIV-1 genome is poorly covered by bulk RNAseq. To address this limitation, we developed an RNAseq method with probe-based enrichment of HIV-1 reads.

Methods- Resting, non-naïve CD4 T cells were isolated from leukapheresis samples from four HIV-1 Eradication and Latency Study (HEAL) participants. 15 million cells were treated for 24 hours with: 1) unstimulated 2) PMA-ionomycin (iono) 3) romidepsin (rmd) 4) bryostatin (bryo) 5) IL-15 6) rmd/bryo. Total RNA was extracted using Trizol. RNA was poly-A selected and libraries were generated following an adapted TruSeq library generation protocol. A custom set of tiling probes was used to enrich HIV-1 and control gene PCR constructs. The unenriched and enriched libraries were sequenced on NextSeq, 40x40 paired-end reads. The reads were aligned to the human transcriptome and HIV-1 genome. A custom script was used to count reads per gene and per region of the HIV-1 genome.

Results- For both host control genes and HIV-1, we observed an average 500-fold enrichment after probe capture. HIV-1 reads aligned across all regions of the genome. PMA-iono and combination rmd/bryo had the greatest increase in HIV-1 transcription. To test the reproducibility of the probe-enrichment, we performed linear regression on normalized RNAseq reads from cells treated with PMA-iono. We observed a Pearson's R2 of 0.95 for total RNAseq between two participants and 0.97 for enriched RNAseq between the same participants. We also found evidence of hypermutated HIV-1 RNAseq reads in the enriched samples.

Conclusions- Our approach to analyzing host and HIV transcriptomes leverages next-generation sequencing to investigate latency reactivation. Probe-based enrichment allowed RNAseq quantification of HIV-1 reads from resting memory CD4+ T cells without the need for sorting of HIV-infected cell populations. We were able to measure HIV-1 transcription after reactivation from latency using a variety of latency reversal agents and compare HIV-1 gene expression across conditions. Analysis of differential host gene expression will yield insight into host factors necessary for HIV-1 reactivation in latently infected resting memory CD4+ T cells in persons with HIV.

Introduction

• The HIV-1 reservoir is transcriptionally silenced; latency reversal is a strategy to expose latently infected cells to the immune system

• Latently infected cells are extremely rare and RNAseq is unable to adequately capture HIV-1 expression

• The goal of this study was to develop a probe-based enrichment of HIV-1 RNAseq reads in order to quantify expression in the latent reservoir under a variety of *ex vivo* stimulation conditions

• RNAseq and enrichment will give information about the host transcriptome as well as HIV-1 genome coverage

Methods

• Leukapharesis samples were obtained from four participants from the HIV-1 Eradication after Latency (HEAL) cohort

• To generate HIV-1 tiling probes, proviral sequences were determined using single genome amplification (SGA)

• Ten human gene sequences were used for tiling probe design in order to quantify after enrichment; genes selected to represent a range of expression levels fm low to high

• Non-naive, resting (nn-r) CD4⁺ T cells were isolated from PBMC and tested with the following conditions for 24 hrs:
  - Unstimulated (U)
  - 50 ng/mL PMA and 1 µM ionomycin (P)
  - 20 nM romidepsin (R)
  - 10 nM bryostatin-1 (B)
  - 1.5 nM IL-15 (I)
  - 20 nM romidepsin and 1 nM bryostatin-1 (RB)

• mRNA was extracted and used for RNAseq library generation with unique molecular identifiers

• HIV-1 and control gene cDNA was enriched using tiling probes

• HIV-1 and human reads were aligned and quantified, then used for differential gene expression analysis

• RNAseq results were compared to RT-qPCR for *gag* caRNA from *ex vivo* stimulated nn-rCD4⁺ T cells from the same participants

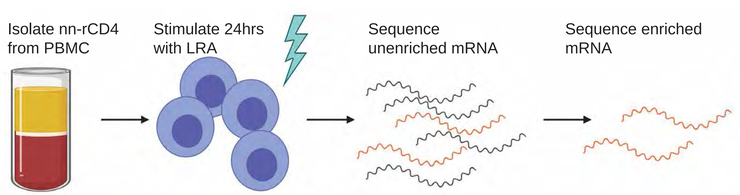


Figure 1 - Schematic of RNAseq with probe enrichment

We developed a **probe-based enrichment RNAseq** technique to study primary latently infected CD4⁺ T cells after *ex vivo* stimulation. We identified *families of differentially expressed transcription factors* after reversal of latency with different treatment conditions. Enrichment revealed coverage across the HIV-1 genome and showed that the *splicing of HIV-1 transcripts did not change with increased expression*.

Latency reversal agents induce distinct genetic states

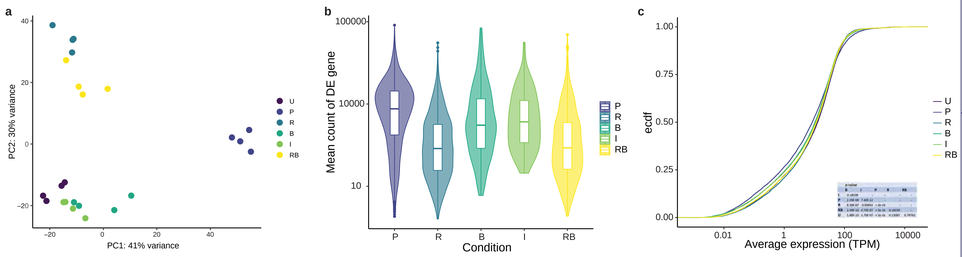


Figure 2 - Differential gene expression analysis of nn-rCD4+ T cells from ART-treated HEAL participants. a) RNAseq libraries from stimulated nn-rCD4⁺ T cells from four HEAL participants were used for differential gene expression analysis using DESeq2. The PCA plot was derived from DESeq2 using normalized gene expression. b) Mean RNAseq counts for differentially expressed genes in specified condition compared to unstimulated cells. c) Empirical cumulative distribution plot of average expression between four HEAL participants for each gene. p value table is for Wilcoxon rank-sum test with Benjamini-Hochberg correction applied.

Transcription factor expression patterns after latency reversal

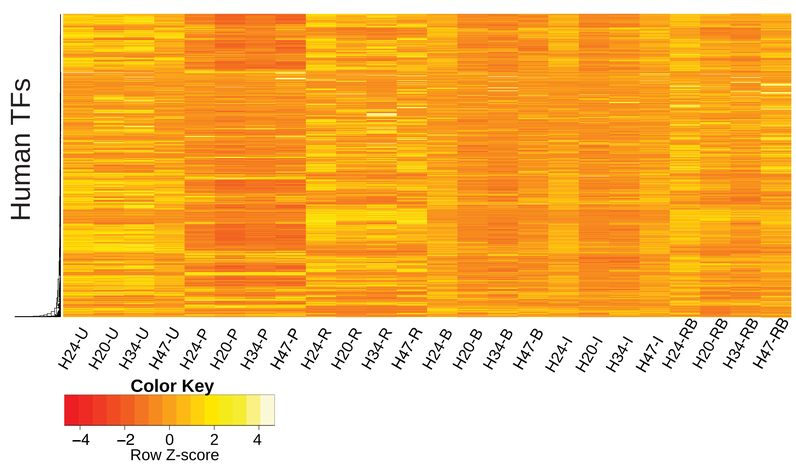


Figure 3 - Heat map of all human transcription factors with coverage in at least one condition. RNAseq coverage was normalized by gene length and sequencing depth per condition.

Probe-based enrichment proportionally enriches HIV-1 RNA

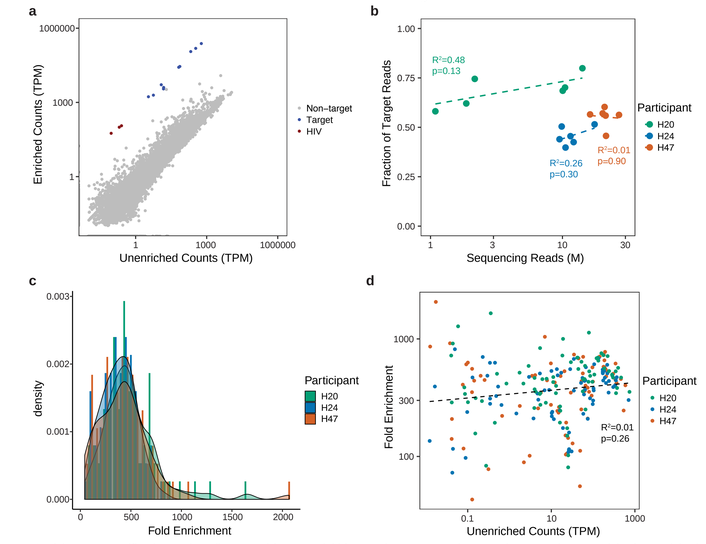


Figure 4 - Enrichment quantification and quality control for potential biases. a) Representative sample comparing unenriched to enriched libraries after normalization. HIV reads are broken up by region: unspliced, singly spliced and multiply spliced. Target genes are 10 human control genes. b) Comparison of fraction target (host control and HIV) to sequencing depth of the enriched libraries for each participant in which enrichment was successful. c) Histogram and density overlay of fold-enrichment of target genes (host control and HIV) for each condition after normalization. d) Fold-enrichment compared to unenriched, normalized counts for target genes (host control and HIV).

Comparison of HIV-1 expression from RNAseq and RT-qPCR

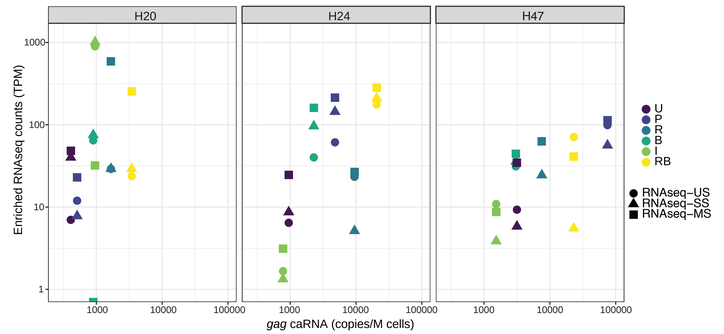


Figure 5 - Each panel shows the relationship between RT-qPCR results and RNAseq, displayed by region of the HIV-1 genome, for the specified participant.

Conclusions

• PMA/iono treatment resulted in more genes with high RNA expression levels, along with a more narrow profile of upregulated TFs, compared to other latency reversal agents.

• The average expression of significantly upregulated genes after reactivation with PMA/iono were ~two-fold higher than the average expression of genes upregulated by romidepsin and romidepsin/bryostatin-1

• Two families of TFs were upregulated in PMA/iono: NF-κB and bZIP (AP-1).

• Probe-based enrichment proportionally enriches HIV-1 RNA and is not biased by sequencing depth of the enriched library or count of the gene in the unenriched library.

• No relationship between splicing pattern of HIV-1 and *gag* caRNA was observed, suggesting that splicing is not rate-limiting step after poly-adenylation.

We gratefully acknowledge the efforts of the HEAL participants and study team. Funding for this project was provided by R21 131757 (to AT)