

# RNA sequencing of CD4 T-cells reveals the relationships between lncRNA-mRNA co-expression in elite controller vs. HIV-positive infected patients

Chaoyu Chen<sup>1</sup>, Xiangyun Lu<sup>1</sup>, Nanping Wu<sup>Corresp. 1</sup>

<sup>1</sup> The First Affiliated Hospital, College of Medicine, Zhejiang University, State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou, Zhejiang, China

Corresponding Author: Nanping Wu  
Email address: flwnp@zju.edu.cn

**Background:** Elite controller refers to a patient with human immunodeficiency virus infection with an undetected viral load in the absence of highly active antiretroviral therapy. Studies on gene expression and regulation in these individuals are limited but significant, and have helped researchers and clinicians to understand the interrelationships between HIV and its host.

**Methods:** We collected CD4 T-cell samples from two elite controllers (ECs), two HIV-positive infected patients (HPs), and two healthy controls (HCs) to perform second-generation transcriptome sequencing. Using the Cufflinks software, we calculated the Fragments Per Kilobase of transcript per Million fragments mapped (FPKM) and identified differentially expressed (DE) mRNAs and long non-coding RNAs (lncRNAs), with corrected P value < 0.05 (based on a false discovery rate (FDR) < 0.05). We then constructed a protein-protein interaction network using cytoHubba and a long non-coding RNA-mRNA co-expression network based on the Pearson correlation coefficient.

**Results:** In total, 1109 linear correlations of DE lncRNAs targeting DE mRNAs were found and several interesting interactions were identified as being associated with viral infections and immune responses within the networks based on these correlations. Among these lncRNA-mRNA relationships, hub mRNAs including *HDAC6*, *MAPK8*, *MAPK9*, *ATM* and their corresponding annotated co-expressed lncRNAs presented strong correlations with the MAPK-NF-kappa B pathway, which plays a role in the reactivation and replication of the virus.

**Conclusions:** Using RNA-sequencing, we systematically analyzed the expression profiles of lncRNAs and mRNAs from CD4+ T cells from ECs, HPs, and HCs, and constructed a co-expression network based on the relationships among DE transcripts and database annotations. This was the first study to examine gene transcription in elite controllers and to study their functional relationships. Our results provide a reference for subsequent functional verification at the molecular or cellular level.

# **RNA sequencing of CD4 T-cells reveals the relationships between lncRNA-mRNA co-expression in elite controller vs. HIV-positive infected patients**

Chaoyu Chen<sup>1</sup>, Xiangyun Lu<sup>1</sup>, Nanping Wu<sup>1</sup>

<sup>1</sup>State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University

Corresponding Author:

Nanping Wu

The First Affiliated Hospital, School of Medicine, Zhejiang University, 79 Qingchun Road, Hangzhou 310003, Zhejiang, China

flwnp@zju.edu.cn

# Abstract

Background: Elite controller refers to a patient with human immunodeficiency virus infection with an undetected viral load in the absence of highly active antiretroviral therapy. Studies on gene expression and regulation in these individuals are limited but significant, and have helped researchers and clinicians to understand the interrelationships between HIV and its host.

Methods: We collected CD4 T-cell samples from two elite controllers (ECs), two HIV-positive infected patients (HPs), and two healthy controls (HCs) to perform second-generation transcriptome sequencing. Using the Cufflinks software, we calculated the Fragments Per Kilobase of transcript per Million fragments mapped (FPKM) and identified differentially expressed (DE) mRNAs and long non-coding RNAs (lncRNAs), with corrected P value < 0.05 (based on a false discovery rate (FDR) <0.05). We then constructed a protein-protein interaction network using cytoHubba and a long non-coding RNA-mRNA co-expression network based on the Pearson correlation coefficient.

Results: In total, 1109 linear correlations of DE lncRNAs targeting DE mRNAs were found and several interesting interactions were identified as being associated with viral infections and immune responses within the networks based on these correlations. Among these lncRNA-mRNA relationships, hub mRNAs including *HDAC6*, *MAPK8*, *MAPK9*, *ATM* and their corresponding annotated co-expressed lncRNAs presented strong correlations with the MAPK-NF-kappa B pathway, which plays a role in the reactivation and replication of the virus.

Conclusions: Using RNA-sequencing, we systematically analyzed the expression profiles of lncRNAs and mRNAs from CD4+ T cells from ECs, HPs, and HCs, and constructed a co-

expression network based on the relationships among DE transcripts and database annotations. This was the first study to examine gene transcription in elite controllers and to study their functional relationships. Our results provide a reference for subsequent functional verification at the molecular or cellular level.

**Keywords:** Transcriptome, CD4 T-cell, network, elite controller, HIV

# Introduction

Human immunodeficiency virus (HIV) has been studied for more than 30 years, and research has found that HIV invades host CD4+ cells, integrates its own DNA into the host genome, and establishes a reservoir in the early stage, followed by massive replication, which destroys the normal immune system function of the host, and triggers various concurrent symptoms [1-2]. Ultimately, HIV infection can cause the host to die. Although highly active antiretroviral therapy (HAART) can almost completely reconstitute the immune function of the host, it cannot eradicate the latent HIV pool in the host, which always maintains a low level of virus replication [3]. However, among HIV-infected individuals, a small number of patients are found to be inherently resistant to viral replication, and spontaneously induce suppression of the latent pools without HAART, resulting in an undetectable viral load in plasma for a long period [4-5]. These patients, termed elite controllers (ECs), have attracted significant research interest. Studying the characteristics of the autoimmune factors of ECs is anticipated to identify important

factors that control virus replication. Such valuable information could lead to novel methods to treat and alleviate HIV infection.

Long-noncoding RNAs (lncRNAs) are transcribed RNA molecules greater than 200 nt in length that regulate gene expression by diverse, but as yet not completely understood, mechanisms [6]. Although the function of most lncRNAs is unknown, several have been shown to regulate gene expression at multiple levels, from DNA to phenotype [7]. Through a variety of means, including *cis* (near the site of lncRNA production) or *trans* (co-expressed with their target gene) mechanisms, lncRNAs play a vital role in many biological processes [8]. Studies on lncRNAs have become a hotspot in current non-coding RNA research.

Recent studies focused on the role of lncRNAs in HIV pathogenesis, especially the relationship between lncRNA regulation of gene expression and viral infection, replication, and latency [9-12]. A few lncRNAs have been characterized and proven to be closely associated with HIV, for example, the knockdown of host lncRNA *NEAT1* enhanced virus production by increasing nucleus-to-cytoplasm export of Rev-dependent instability element (INS)-containing HIV-1 mRNAs [13]. The knockdown of lncRNA *NRON* enhanced HIV-1 replication through increased activity of nuclear factor of activated T-cells (NFAT) and the viral long terminal repeat (LTR) [14]. Similarly, lncRNA *MALAT1* releases epigenetic silencing of HIV-1 replication by displacing the polycomb repressive complex 2 from binding to the LTR promoter [15]. Furthermore, an HIV-encoded antisense lncRNA, *ASP-L*, was proven to promote latency HIV [11, 16]. In addition, *HEAL* is a broad enhancer of multiple HIV-1 strains because depletion of *HEAL* inhibited X4, R5, and dual-tropic HIV replication, which was

rescued by HEAL overexpression [17]. Although the mechanism of lncRNA function is sometimes elusory and unpredictable, many current studies, which are limited to the screening and functional prediction stage based on chip or sequencing results, also provide us with valuable information and form the basis for mechanistic research.

In the present study, we conducted a transcriptomics investigation for HIV ECs, to identify differences in the transcriptional expression profiles between ECs and HIV-positive infected patients as compared with healthy controls. The obtained lncRNA-mRNA co-expression network revealed the possible role of the functional relationships between lncRNAs and mRNAs in the ability of ECs to inhibit viral replication.

## Methods

### Ethics statement

Ethical approval was granted by the Ethics Committee of the First Affiliated Hospital, College of medicine, Zhejiang University. The Reference number is 2017-338. All subjects provided written, informed consent. In addition, all participating medical institutions provided local institutional review board approval.

### Cohort and patients

In total, we recruited 196 patients into our cohort, from different provinces of China, who were diagnosed in Qingchun Hospital of Zhejiang Province from January 2000 to January 2018, and received, or volunteered to reject HAART. The viral loads of the patients were detected using

the Cobas Taqman system (Roche, Basle, Switzerland). The healthy controls (HCs) were randomly selected from hospital clinics and were not suffering from any diseases. We chose two ECs, two HIV-positive infected patients (HPs), and two HCs of similar ages, sex, and traditional risk factors to perform second generation transcriptome sequencing. ECs were defined as HIV-infected patients with an undetectable viral load in the absence of HAART for nearly 10 years.

### **Sample collection and preparation**

Peripheral blood mononuclear cells (PBMCs) were isolated from 3–5 mL EDTA-K2+ anticoagulant venous blood using density gradient centrifugation. Primary CD4+ T cells were purified through negative selection using a CD4+ T cell isolation Kit (Miltenyi Biotec, Bergisch, Gladbach, Germany). According to flow cytometry analysis, the purity of the separated CD4+ T cells was more than 90%. Total RNA was extracted using the TRIzol reagent (Life Technologies, Carlsbad, CA, USA). RNA purity was checked using a NanoPhotometer® spectrophotometer (IMPLEN, München, Germany). The RNA concentration was measured using a Qubit® RNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies) and its integrity was assessed using an RNA Nano 6000 Assay Kit for the Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA).

### **RNA library construction and RNA sequencing**

Ribosomal RNA was removed using a Epicentre Ribo-zero™ rRNA Removal Kit (Epicentre, Madison, WI, USA), and the rRNA free residue was cleaned using ethanol precipitation. Subsequently, sequencing libraries were generated using the rRNA depleted RNA and an NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (NEB, Ipswich, MA, USA)

following manufacturer's recommendations. After cDNA synthesis and purification, clustering of the index-coded samples was performed on a cBot Cluster Generation System using a TruSeq PE Cluster Kit v3-cBot-HS (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. The libraries were then sequenced on an Illumina Hiseq 4000 platform and raw data were generated.

### **RNA sequencing analysis**

Clean data were obtained by removing reads containing adapters, reads only containing poly-N, and low quality reads from the raw data. After quality control, paired-end clean reads were aligned to the reference genome downloaded from website using HISAT2 (v2.0.4) [18]. Cufflinks (v2.1.1) was used to calculate the FPKM (expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) values of both lncRNAs and mRNAs in each sample [19]. FPKM is the number of fragments from a gene per thousand bases in each million fragments. It takes into account the sequencing depth and the effect of gene length on the fragment count (Figure S3, S4). Transcripts with FPKM values below 1.0 in any one of the samples were excluded. A summary of gene FPKM values were obtained by combining the FPKM values of transcripts. Cuffdiff (v2.1.1) provided statistical routines to determine differential expression in the digital transcript or gene expression data using a model based on the negative binomial distribution. Transcripts with a corrected P value  $< 0.05$  (false discovery rate (FDR)  $< 0.05$ ) were assigned as differentially expressed. We plotted a heatmap to observe the clustering between the samples and the genes using Multi Experiment Viewer (v4.9.0). All raw data was uploaded in the *.bam* format and stored in the SRA database at the NCBI, with the



147 accession number PRJNA575767.

# 148 **LncRNA-mRNA co-expression network analysis**

149 A *Trans* role is defined as an lncRNA binding to a target DNA as an RNA:DNA heteroduplex, as  
 150 an RNA:DNA:DNA triplex, or RNA recognition of specific chromatin-like complex surfaces  
 151 [20]. We calculated the expression correlation between lncRNAs and mRNAs using the Pearson  
 152 correlation test for target gene prediction and the results were expressed as Pearson correlation  
 153 coefficients. To narrow the scope of investigation and focus on the most important transcripts,  
 154 we first assessed the interaction among all differentially expressed (DE) mRNAs using the  
 155 STRING database (v11.0) and constructed a protein-protein interaction (PPI) network using the  
 156 cytoHubba package in Cytoscape (v3.7.1). After considering 12 synthetic algorithms, the  
 157 intersection of the top 50 genes yielded 24 hub genes, which occupied a central position in the  
 158 network of all DE mRNAs. After identifying and retaining the DE mRNAs and lncRNAs in all  
 159 lncRNA-mRNA relations, we screened 1109 pairs of exact relationships between the two groups  
 160 and constructed a co-expression network based on the Pearson relation coefficient (all  $> 0.95$ ,  $p$   
 161  $< 0.001$ ). Then, we identified the lncRNAs and their associated mRNAs for further functional  
 162 enrichment analysis. The PPI and lncRNA-mRNA network diagrams were drawn using the  
 163 Cytoscape software (v3.7.1).

# 164 **Gene ontology and KEGG pathways analysis**

165 By jointly using the David database (v6.8) and KOBAS database (v3.0), gene ontology (GO)  
 166 enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis of all  
 167 DE mRNAs and DE lncRNA-co-expressed DE mRNAs were implemented to interpret the

biological meaning of the transcripts. GO terms and KEGG pathways with corrected P values less than 0.05 (FDR<0.05) were considered significantly enriched by the differentially expressed RNAs.

# **Quantitative real-time PCR validation**

We used quantitative real-time PCR (qRT-PCR) with SYBR green analysis to validate the accuracy of the sequencing results. To validate random DE mRNAs, total RNA was extracted using TRIzol reagent (Life Technologies). QRT-PCR was performed with the iQ™ SYBR® Green Supermix (Bio-Rad, Hercules, CA, USA). *ACTB* (encoding beta-actin) was used as the endogenous control for the mRNA analysis. The expression levels of mRNA were calculated based on the change in cycling threshold using the of  $2^{-\Delta Ct}$  method [21]. GraphPad Prism 7 (GraphPad Software, Inc., La Jolla, CA, USA) was used to perform the qRT-PCR statistical analysis. The non-parametric Mann-Whitney U test was used to compare between group distributions.

# **Results**

## **Subjects**

A total of 196 individuals infected with HIV were enrolled in our cohort, including patients under treatment (181, 92.4%), HIV-positive infected patients (13,6.6%) and elite controllers (2, 1.0%). In addition, we recruited two healthy individuals from the out-patient department of the hospital. Most patients were infected with the virus through heterosexual contact, followed by intravenous infection, and homosexual transmission. Based on the basic principle of intra-group

identity, we determined three men and three women who were assigned to each of the three study groups as study subjects. Their ages ranged from 39 to 54 and were free of other diseases, such as tuberculosis, diabetes, or hepatitis. Specific details of the clinical characteristics of the participants are shown in Table S1. Descriptive statistics are reported as counts (percentage) for dichotomous and categorical variables, and the median (the 25th percentile and 75th percentile) for continuous variables. The experimental design and analysis process of this research are shown in a flowchart (Figure 1).

# **Identification of DE transcripts between HCs and HPs**

First, we scanned the expression profiles of all transcripts in CD4 T-cells from two HCs, two ECs, and two HPs. In all, more than 82000 mRNA transcripts and more than 20000 corresponding genes were identified. Likewise, nearly 15000 lncRNAs and about 3500 corresponding gene symbols were identified. Differential expression of transcripts between HCs and HPs was first considered. Altogether, 3602 mRNA transcripts and 383 lncRNA transcripts were differently expressed according to the threshold of fold change  $> 2$  and corrected p value  $< 0.05$  (FDR  $< 0.05$ ), among which 2313 mRNAs and 207 lncRNAs were upregulated and 1289 mRNAs and 176 lncRNAs were downregulated. Similarly, we compared gene expression differences between HCs and ECs, but paid more attention to ECs vs. HPs. The comparison with HCs helped to provide a basic reference background. All transcript FPKM values and summary of gene FPKM values were included in Table S2. The lists of FPKM values and regulatory of all DE mRNAs and DE lncRNAs in HCs vs. HPs and HCs vs. ECs are shown in Table S3 and Table S4, respectively.

## Identification of DE mRNAs between ECs and HPs

Among all transcripts, 2936 transcripts were differentially expressed according to the threshold of fold change  $> 2$  and corrected p value  $< 0.05$  ( $FDR < 0.05$ ). There were 1327 mRNAs that were upregulated and 1609 that downregulated in ECs vs. HPs. GO and KEGG pathway analysis was carried out to annotate the function of these DE mRNAs. Certain terms and pathways that might be associated with infection and immunity were significantly enriched (Figure S1). These terms and pathways included negative regulation of interferon-beta production (GO:0032688), cellular component disassembly involved in execution phase of apoptosis (GO:0006921), regulation of defense response to virus by virus (GO:0050690), negative regulation of type I interferon production (GO:0032480), RIG-I-like receptor signaling pathway (hsa04622), and antigen processing and presentation (hsa04612). Note that when considering the overall expression of a gene but encountering conflicting expression levels of multiple transcripts, we considered the summary of post-merger FPKM values synthetically and prioritized the more statistically significant transcript as the study object.

## Identification of DE lncRNAs between ECs and HPs

Among all transcripts, 3543 official gene symbols that matched in the NONCODE database had corresponding lncRNA transcripts. Three main categories of lncRNAs accounted for the majority, including antisense lncRNAs (2537/14853, 17.1%), intronic lncRNAs (8152/14853, 54.9%), and long intergenic noncoding RNAs (lincRNAs; 3344/14853, 22.5%). There were 151 upregulated lncRNAs and 160 downregulated lncRNAs in ECs vs. HPs according to the threshold of fold change  $> 2$  and  $p < 0.05$ . In both the mRNA and lncRNA clusters (Figure 2),

the samples showed marked intra-group correlations and inter-group differences, which further confirmed the specificity of the samples. In terms of the overall profiles of mRNAs encoding proteins with specific biological functions, the HC and EC groups showed closer clustering and were relatively distant from the HP group. The lists of FPKM values and regulatory situations of all DE mRNAs and lncRNAs in ECs vs. HPs are shown in Table S5.

### **LncRNA-mRNA co-expression network analysis**

Almost 0.9 million co-expression relationships were discovered, and the exact form was expressed as a linear correlation whereby one lncRNA targets one mRNA. Based on previous DE analysis, 1109 DE lncRNAs targeting DE mRNAs were identified in all correlations (Table S6). Figure 3 shows the network constructed using these relationships, in which 142 DE lncRNAs were co-expressed with 127 DE mRNAs. To investigate the importance ranking of these DE mRNAs, and find the genes at the center of the functions of gene expression, we constructed a PPI network. Figure 4 shows the top 50 hub genes that playing an important role in interaction of 127 DE mRNAs, in which many genes functioned centrally and could be frequently regulated by other transcripts. Considering the differences in the outputs of different algorithms, we identified those mRNAs that were commonly detected by all 12 algorithms and focused on those mRNAs that were regulated by lncRNAs and their functional enrichment analysis. Table S7 shows the list of 24 intersecting hub mRNAs among the top 50 calculated using the 12 algorithms. In Figure 3, we identified 16 of 24 intersecting hub mRNAs in the diagram, which helped us to quickly locate the target protein coding genes and easily access the enriched functions of these genes. Next, we selected the 16 hub mRNAs and their associated lncRNAs from Figure 3 and constructed Figure

5. We then identified the top 10 GO terms and pathways of all DE mRNAs most closely related to infection and immunity in Figure 3. Many significant functions are co-enriched, such as immune response-regulating signaling pathway (GO:0002764) and cellular response to interleukin-1 (GO:0071347) (Figure 6). It should also be pointed out that when there were various many-to-many regulatory relationships, the most statistically significant regulatory relationship was selected as the main study object.

### **Immune regulation and viral infection associated lncRNA-mRNA networks**

We further investigated the specific functions of the co-expressed hub mRNAs shown in Figure 5 and found several interesting regulatory relationships. Certain genes were enriched in pathways that are closely related to immune regulation and viral infection. These genes and corresponding pathways are shown in Table 1. There were numerous annotations of enriched pathways recurring for different mRNAs, which meant these hub mRNAs had a close functional relationship. The corresponding annotated lncRNAs, including *C3orf35*, *TMEM254-AS1*, *ARRDC3-AS1*, *LINC00202-1*, *TRAF3IP2-AS1*, *RN7SL15P* and *RP4-622L5.7*, were co-expressed with these mRNAs, which suggested several new regulatory relationships that might play an important role in controlling virus replication.

### **Dicussion**

In the present study, we systematically analyzed the expression profiles of lncRNAs and mRNAs from CD4<sup>+</sup> T cells from ECs, HPs, and HCs, and constructed a co-expression network based on the relationships among DE transcripts and database annotations. The functions of most

273 lncRNAs have not been well characterized. Therefore, the gene functions in this research mainly  
 274 referred to those of mRNAs and their expression products. The potential functions of candidate  
 275 lncRNAs and the lncRNAs-associated biological processes in ECs were predicted using the  
 276 lncRNA-mRNA network and functional enrichment analysis. We identified certain genes that  
 277 play a significant role in the ability of ECs to inhibit the replication of latent virus. Most of these  
 278 genes were downregulated in the HCs and ECs and upregulated in the HPs or vice versa, which  
 279 was consistent with their clinical characteristics, in that ECs maintained a similar gene  
 280 expression pattern to the HCs, which distinguished them from the HPs.

281 To the best of our knowledge, the present study was the first to predict lncRNA-mRNA  
 282 interactions for HIV ECs, in which many genes had not been previously reported to function in  
 283 HIV-associated biological processes. The *MAPK* (mitogen activated protein kinase) family was  
 284 found to be associated with the reactivation and replication of the virus in many studies [22-23],  
 285 and the HIV-1 virus inhibition activity of *MAPK-NF- $\kappa$ B/AP-1* pathway inhibitors was the result  
 286 of the negative regulation of HIV-1 LTR promoter activity [24]. In a recent study, *HDAC6*  
 287 (histone deacetylase 6) overexpression significantly enhanced the expression of pro-  
 288 inflammatory cytokines, such as TNF- $\alpha$  (tumor necrosis factor alpha), interleukin (IL-1) $\beta$ , and  
 289 IL-6, with a concomitant reduction in acetylated  $\alpha$ -tubulin. HIV-1 trans-activator of transcription  
 290 (Tat) can upregulate the expression of various chemokine genes including *IL-1*, *IL-6*, *CCL2* (C-C  
 291 motif chemokine ligand 2), *CXCL8* (C-X-C motif chemokine ligand 8), and *CXCL10* (C-X-C  
 292 motif chemokine ligand 10). In addition, *HDAC6* overexpression increased the activation  
 293 of MAPK factors including extracellular signal-regulated kinase (ERK), Jun terminal kinase

(JNK), and mitogen-activated protein kinase 14 (p38). Similarly, *HDAC6* overexpression resulted in activation of the nuclear factor kappa B (NF- $\kappa$ B) and activator protein 1 (AP-1) signaling pathways[25-26]. In our results, low expression levels of *HDAC6* and *MAPK8* and downregulation of IL-1 and IL-6 associated receptors in ECs indicated suppressed HADC6-MAPK-NF- $\kappa$ B/AP-1 signaling and fewer immune response disorders. Inhibition of histone deacetylase contributes to reverse the latency of HIV. Ataxia Telangiectasia Mutated (*ATM*) is also involved in the NF-kappa B signaling pathway (hsa04064) by activating p38MAPK and p53 sequentially to induce a pro-apoptotic pathway. Ariumi et.al [27] indicated that *ATM* could enhance HIV-1 replication by stimulating the action of the Rev viral posttranscriptional regulator, and Lau et al. [28] demonstrated that a novel and specific small molecule inhibitor of ATM kinase activity, KU-55933, is capable of suppressing the replication of both wild-type and drug-resistant HIV-1. Interestingly, some studies considered that HIV-encoded protein Vpu contributes to the attenuation of the anti-viral response by partial inactivation of interferon regulatory factor 3 (IRF3) while host cells undergo apoptosis [29] By contrast, another study reported that Vpu affects interferon expression by inhibiting NF- $\kappa$ B activity without affecting IRF3 levels or activity [30]. Thus, the role of IRF3 in HIV infection and replication remains controversial. The above hub mRNAs are related to annotated lncRNAs including *C3orf35*, *TMEM254-AS1*, *ARRDC3-AS1*, *LINC00202-1*, *TRAF3IP2-AS1* and to unannotated lncRNAs (Figure 5), which might control HIV replication by regulating the expression of these genes. To date, no study has been published showing that these lncRNAs function in the pathways identified in the present study.



Alternatively, some hub genes (Figure 4) were not regulated by lncRNAs but interacted with each other and participated in biological processes such as HIV life circle, inflammation and immune activation and cytokines. Lysine demethylase 6A (KDM6A; also called UTX-1) enzymatic activity is required for the viral protein Tat to remove a repressive mark H3K27me3 in the HIV-1 long terminal repeat (LTR) and promoted HIV-1 gene expression by enhancing both NF- $\kappa$ B p65's nuclear translocation and its binding to the HIV-1 LTR. H3K27 demethylase activity is required for increased HIV-1 transactivation induced by *KDM6A* [31]. In our study, the low expression level of *KDM6A* in the ECs might be a possible causes of their low viral loads. *TRAF3IP2* encodes TRAF3 interacting protein 2, which is involved in regulating responses to cytokines by members of the Rel/NF-kappaB transcription factor family. TRAF3IP2 interacts with TRAF proteins (tumor necrosis factor receptor-associated factors) and either I-kappaB kinase or MAP kinase to activate either NF-kappaB or Jun kinase. We found that lncRNA *TRAF3IP2-AS1* was an anti-sense non-coding RNA of *TRAF3IP2* that was upregulated in the ECs, and was co-expressed with downregulated *MAPK8*. This suggested there may be a regulatory relationship among *TRAF3IP2-AS1*-TRAF3IP2-MAPK8 that plays an important role in the suppression of viral replication.

HAART cannot eradicate the latent virus in the host; therefore, the genes identified in the present study might play a non-negligible regulatory role in the relationship between the latent virus and the host. Latent virus is the biggest obstacle to the elimination of the disease. Thus, understanding the processes that contribute to its persistence, such as inflammation and immune activation, are crucial for the remission and cure of HIV [32-33].

There are also some limitations to this study. First, because of the limited amount of RNA in the samples, the sample size of this study was relatively small. Although we selected eight and six mRNAs for verification using qRT-PCR for 16 genes and 40 samples three times, respectively and found that the expression change trends of 13 of them were consistent with the RNA sequencing results, these results lack of sufficient statistical significance because of the small sample size. Part of the verification analysis is shown in Figure S2. Second, we only assessed gene expression of CD4<sup>+</sup> T cells, which, although they are the main targets of HIV and are vitally important, they are only one of the many cell types that are infected by HIV. Further research should be performed to validate the function and mechanism of the genes in a larger sample and in more cell types.

## Conclusions

This study identified several important differentially expressed genes associated with the elite controller phenomenon, using RNA-sequencing and bioinformatics analysis to explore how ECs spontaneously control virus replication. Some differentially expressed genes were identified and enriched for meaningful GO terms and KEGG pathways related with viral infection and immune responses. After filtering lncRNAs and co-expressed mRNAs using Pearson correlation test, the functions of the lncRNAs were predicted using an lncRNA-mRNA network comprising the functional annotations of mRNAs. This study forms the basis for subsequent cellular and molecular studies, and provides new targets for gene-targeted therapy in the future.

## Additional Information and Declarations

### Supplementary materials

Figure S1. Gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment of all differentially expressed (DE) mRNAs. The height of the histogram correlated positively with the significance of the annotation, and the line graph represents the number of genes enriched under the annotation.

Figure S2. The mRNA-seq data valiated using qRT-PCR.

Figure S3. The box plot of FPKM distribution among different groups. Horizontal coordinates show the group names, the ordinate shows the  $\log_{10}$  (FPKM-1), each area of the box map corresponds to five statistics (top to bottom are the maximum, top quartile, median, down quartile, and minimum value).

Figure S4. The density map of FPKM distribution among different groups. Horizontal coordinates of  $\log_{10}$  (FPKM-1) and the ordinate showing the density of the genes.

Table S1. Clinical characteristics of the participants.

Table S2. All transcripts' FPKM values and a summary of gene FPKM values.

Table S3. Lists of differentially expressed (DE) mRNAs and long non-coding RNAs (lncRNAs) in healthy controls (HCs) vs. HIV-positive infected patients (HPs).

Table S4. Lists of DE mRNAs and lncRNAs in HCs vs. elite controllers (ECs).

Table S5. Lists of DE mRNAs and lncRNAs in the HPs vs. ECs comparison. The transcripts of 24 hub genes and lncRNAs in Figure 5 were annotated with a colored background.

Table S6. The 1109 co-expression relationships as linear correlations whereby one lncRNA targets one mRNA.

Table S7. The intersection list of the top 50 hub-mRNAs calculated using 12 algorithms.

## Competing interests

The authors declare that no competing interests exist.

## Authors' contributions

Chaoyu Chen, Xiangyun Lu and Nanping Wu were involved in the conception and design of the study. Chaoyu Chen and Xiangyun Lu were involved in sample preparation, data acquisition, and organization. Chaoyu Chen conducted the data analysis, and all authors were responsible for data interpretation. Chaoyu Chen was the major contributor to the writing of the manuscript. All authors read and approved the final manuscript.

## Funding

This study was supported by grants from the Mega-Project for National Science and Technology Development under the "Study on comprehensive prevention and treatment of AIDS in demonstration areas, 13th, Five-Year Plan of China" (NO.2017ZX10105001-005) and "Research and application of appropriate treatment and prevention strategies for children with AIDS, 13th, Five-Year Plan of China" (NO.2018ZX10302-102). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

## Human Ethics

Please refer to the ethics statement in the methods section.

## Acknowledgements

We would like to thank all of the participants in the cohort. In addition, we would like to thank the native English speaking scientists of Elixigen Company (Huntington Beach, California) for editing our manuscript.

## References

1. Sabin, C. A. , & Lundgren, J. D. . (2013). The natural history of hiv infection. Current opinion in HIV and AIDS, 8(4).
2. Eisele, E. , & Siliciano, R. . (2012). Redefining the viral reservoirs that prevent hiv-1 eradication. Immunity, 37(3), 377-388.
3. Murray, A. J. , Kwon, K. J. , Farber, D. L. , & Siliciano, R. F. . (2016). The latent reservoir for hiv-1: how immunologic memory and clonal expansion contribute to hiv-1 persistence. The Journal of Immunology, 197(2), 407-417.
4. Migueles, S. A. , & Connors, M. . (2010). Long-term nonprogressive disease among untreated hiv-infected individuals clinical implications of understanding immune control of hiv. JAMA The Journal of the American Medical Association, 304(2), 194-201.
5. Gonzalo-Gil, E. , Ikediobi, U. , & Sutton, R. E. . (2017). Mechanisms of virologic control and clinical characteristics of hiv+ elite/viremic controllers. The Yale journal of biology and

418 medicine, 90(2), 245-259.

419 6. Wilusz, J. E. , Sunwoo, H. , & Spector, D. L. . (2009). Long noncoding rnas: functional  
420 surprises from the rna world. *Genes & Development*, 23(13), 1494-1504.

421 7. Wapinski, O. , & Chang, H. Y. . (2011). Long noncoding rnas and human disease. *Trends in*  
422 *Cell Biology*, 21(6), 354-361.

423 8. Wang, K. C. , & Chang, H. Y. . (2011). Molecular mechanisms of long noncoding  
424 rnas. *Molecular Cell*, 43(6), 904-914.

425 9. Trypsteen, W. , Mohammadi, P. , Van Hecke, C. , Mestdagh, P. , Lefever, S. , Saeys, Y. , De  
426 Bleser, P. , Bandesompele, J. , Ciuffi, A. , Bandekerckhove, L. , & De Spiegelaere, W. .  
427 (2016). Differential expression of lncrnas during the hiv replication cycle: an underestimated  
428 layer in the hiv-host interplay. *Scientific Reports*, 6, 36111.

429 10. Lazar, D. C. , Morris, K. V. , & Saayman, S. . (2015). The emerging role of long non-coding  
430 rnas in hiv infection. *Virus Research*, S0168170215300277.

431 11. Saayman, S. , Ackley, A. , Turner, A. M. W. , Famiglietti, M. , Bosque, A. , Clemson, M. ,  
432 Planelles, V. , & Morris, K. V. . (2014). An hiv-encoded antisense long noncoding rna  
433 epigenetically regulates viral transcription. *Molecular Therapy*, 22(6), 1164-1175.

434 12. Nair, M. , Sagar, V. , & Pilakka-Kanthikeel, S. . (2016). Gene-expression reversal of lncrnas  
435 and associated mrnas expression in active vs latent hiv infection. *Scientific Reports*, 6, 34862.

436 13. Zhang, Q. H. , Chen, C. Y. , Yedavalli, V. S. R. K. , & Jeang, K. T. . (2013). Neat1 long  
437 noncoding rna and paraspeckle bodies modulate hiv-1 posttranscriptional expression. *Mbio*, 4(1),  
438 00596-12.

- 439 14. Imam, H. , Shahr Bano, A. , Patel, P. , Holla, P. , & Jameel, S. . (2015). The lncrna nron  
440 modulates hiv-1 replication in a nfat-dependent manner and is differentially regulated by early  
441 and late viral proteins. *Scientific Reports*, 5, 8639.
- 442 15. Qu Di., Sun Wei-Wei., Li Li., Ma Li., Sun Li., Jin Xia., Li Taisheng., Hou Wei., Wang Jian-  
443 Hua.(2019). Long noncoding RNA MALAT1 releases epigenetic silencing of HIV-1 replication  
444 by displacing the polycomb repressive complex 2 from binding to the LTR promoter. *Nucleic  
445 Acids Res.*, 47(6), 3013-3027. doi:10.1093/nar/gkz117
- 446 16. Kobayashi-Ishihara, M. , Yamagishi, M. , Hara, T. , Matsuda, Y. , Takahashi, R. , Miyake, A.  
447 , Nakano, K. , Yamochi, T. , Ishida, T. , & Watanabe, T. . (2012). Hiv-1-encoded antisense rna  
448 suppresses viral replication for a prolonged period. *Retrovirology*, 9(1), 38.
- 449 17. Chao Ti-Chun., Zhang Qiong., Li Zhonghan., Tiwari Shashi Kant., Qin Yue., Yau Edwin.,  
450 Sanchez Ana., Singh Gatikrushna., Chang Kungyen., Kaul Marcus., Karris Maile Ann Young.,  
451 Rana Tariq M.(2019). HEALThe Long Noncoding RNA Regulates HIV-1 Replication through  
452 Epigenetic Regulation of the HIV-1 Promoter. *mBio*, 10(5), undefined.  
453 doi:10.1128/mBio.02016-19
- 454 18. Pertea, M. , Kim, D. , Pertea, G. M. , Leek, J. T. , & Salzberg, S. L. . (2016). Transcript-level  
455 expression analysis of rna-seq experiments with hisat, stringtie and ballgown. *Nature  
456 Protocols*, 11(9), 1650-1667.
- 457 19. Trapnell, C. , Williams, B. A. , Pertea, G. , Mortazavi, A. , Kwan, G. , Van Baren, M. J. ,  
458 Salzberg, S. L. , Wold, B. J. , & Pachter L. . (2010). Transcript assembly and quantification by  
459 rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature*

- 460 Biotechnology, 28(5), 511-515.
- 461 20. Hung, T. , & Chang, H. Y. . (2010). Long noncoding rna in genome regulation. RNA  
462 Biology, 7(5), 582-585.
- 463 21. Pertea, M. , Kim, D. , Pertea, G. M. , Leek, J. T. , & Salzberg, S. L. . (2016). Transcript-level  
464 expression analysis of rna-seq experiments with hisat, stringtie and ballgown. Nature  
465 Protocols, 11(9), 1650-1667.
- 466 22. Wang Xue., Sun Bing., Mbondji Christelle., Biswas Santanu., Zhao Jiangqin., Hewlett  
467 Indira.(2017). Differences in Activation of HIV-1 Replication by Superinfection With HIV-1 and  
468 HIV-2 in U1 Cells. J. Cell. Physiol., 232(7), 1746-1753. doi:10.1002/jcp.25614
- 469 23. Prasad, S. , & Tyagi, A. K. . (2015). Curcumin and its analogues: a potential natural  
470 compound against hiv infection and aids. Food Funct., 6(11), 10.1039.C5FO00485C.
- 471 24. Gong Jian., Shen Xi-hui., Chen Chao., Qiu Hui., Yang Rong-ge.(2011). Down-regulation of  
472 HIV-1 infection by inhibition of the MAPK signaling pathway. Virol Sin, 26(2), 114-22.  
473 doi:10.1007/s12250-011-3184-y
- 474 25. Youn, G. S. , Lee, K. W. , Choi, S. Y. , & Park, J. . (2016). Overexpression of hdac6 induces  
475 pro-inflammatory responses by regulating ros-mapk-nf- $\kappa$  b/ap-1 signaling pathways in  
476 macrophages. Free Radical Biology and Medicine, 97, 14-23.
- 477 26. Nookala, A. , & Kumar, A. . (2014). Molecular mechanisms involved in hiv-1 tat-mediated  
478 induction of il-6 and il-8 in astrocytes. Journal of Neuroinflammation, 11(1), 214.
- 479 27. Ariumi, Y. , & Trono, D. . (2006). Ataxia-telangiectasia-mutated (atm) protein can enhance  
480 human immunodeficiency virus type 1 replication by stimulating rev function. Journal of



481 Virology, 80(5), 2445-2452.

482 28. Lau, A. , Swinbank, K. M. , Ahmed, P. S. , Taylor, D. L. , Jackson, S. P. , Smith, G. C. M. ,  
483 & O'Connor M. J. . (2005). Suppression of hiv-1 infection by a small molecule inhibitor of the  
484 atm kinase. *Nature Cell Biology*, 7(5), 493-500.

485 29. Park, S. Y. , Waheed, A. A. , Zhang, Z. R. , Freed, E. O. , & Bonifacino, J. S. . (2014). Hiv-1  
486 vpu accessory protein induces caspase-mediated cleavage of irf3 transcription factor. *Journal of*  
487 *Biological Chemistry*, 289(51), 35102-35110.

488 30. Manganaro, L. , De Castro, E. , Maestre, A. M. , Olivieri, K. , García-Sastre, Adolfo,  
489 Fernandez-Sesma, A. , & Simon, V. . (2015). Hiv vpu interferes with nf- $\kappa$  b activity but not with  
490 interferon regulatory factor 3. *Journal of Virology*, JVI.01596-15.

491 31. Zhang, H-S. , Du, G-Y. , Liu, Y. , Zhang, Z-G. , Zhou, Z. , Li, H. , Dai, K-Q. , Yu, XpY. , &  
492 Gou, X-M. . UTX-1 regulates Tat-induced HIV-1 transactivation via changing the methylated  
493 status of histone H3. *The international journal of biochemistry and cell biology*, 2016.

494 32. Dahabieh, M. S. , Battivelli, E. , & Verdin, E. . (2015). Understanding hiv latency: the road  
495 to an hiv cure. *Annual Review of Medicine*, 66(1), 407-421.

496 33. Massanella, M. , Rémi Fromentin, & Chomont, N. . (2015). Residual inflammation and viral  
497 reservoirs: alliance against an hiv cure. *Current opinion in HIV and AIDS*, 11(2).

498

499 **Figure 1. Flowchart of the analysis procedure.** lncRNA, long non-coding RNA; EC, elite  
500 controller; HP, HIV-positive infected patient; DE, differentially expressed; GO, gene ontology;  
501 KEGG, Kyoto Encyclopedia of Genes and Genomes.

**Figure 2. Heatmaps of differentially expressed (DE) mRNAs and long non-coding RNAs (lncRNAs).**

Hierarchical clustering analysis of sequencing detected mRNAs and lncRNAs with their expression abundance. Patient groups are on the horizontal axis, and mRNA and lncRNA genes are grouped along the vertical axis. Pinkish-purple bars indicate increased abundance of the corresponding genes, and blue bars indicates decreased abundance. White bars indicate that the corresponding genes were not detected. A) Expression abundance of mRNA genes between HC vs. EC vs. HP. B) Expression abundance of lncRNA genes between HC vs. EC vs. HP.

**Figure 3. Long non-coding RNA (lncRNA)-mRNA co-expression network in elite controllers (ECs) and HIV-positive infected patients (HPs).** The network was based on the Pearson relation coefficient of an lncRNA targeting an mRNA. Round shapes correspond to mRNAs, and squares to lncRNAs. The size of the shape is positively related to the P value. Red means upregulation and green means downregulation. A solid line indicates a positive correlation and a dotted line indicates a negative correlation. All hub-mRNAs and annotated lncRNAs are tagged with large labels. All lncRNAs and their co-expressed mRNAs are represented.

**Figure 4. Protein-protein interaction (PPI) network of differentially expressed (DE) mRNAs.**

Top 50 hub mRNAs were included in using Closeness algorithm. The intensity of the color indicates the density of the correlation. The closer to red, the more central the gene is. The size of the circle represents the significance of a gene.

**Figure 5. Long non-coding RNA (lncRNA)-Hub mRNA co-expression network in elite controllers (ECs) and HIV-positive infected patients (HPs).** All hub mRNAs are presented in Fig 3 together with their associated lncRNAs. A) *HDAC6*, *SRP54*, *GNG2*, *RPS2*, *IRF3* and their associated lncRNAs. B) *SMARCE1* and its associated lncRNAs. C) *MAPK8* and its associated lncRNAs. D) *GFM1* and its associated lncRNAs. E) *SMC4* and its associated lncRNAs. F) *ARID1B* and its associated lncRNAs. G) *ATM* and *LNC\_001013*. H) *AP3B1* and *LNC\_007860*. I) *RPS3A* and *LNC\_007987*. J) *TRRAP* and *LNC\_009875*. K) *EP400* and *LNC\_001221*. L) *MAPK9* and *LNC\_000086*.

**Figure 6. Gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment classification of the predicted targeting mRNAs of the long non-coding RNA (lncRNA)-mRNA network.** The 10 most significantly enriched KEGG pathways and the 10 most significantly enriched GO terms in the biological process, cellular components, and molecular function categories. A) Top 10 GO terms enriched for predicted targeting mRNAs of the lncRNA-mRNA network. B) Top 10 KEGG terms enriched for predicted targeting mRNAs of the lncRNA-mRNA network.

**Table 1. Immune regulation and viral infection associated lncRNA-mRNA networks.**

**Table 1** (on next page)

Immune regulation and viral infection associated lncRNA-mRNA networks.

**Table1. Immune regulation and viral infection associated lncRNA-mRNA networks**

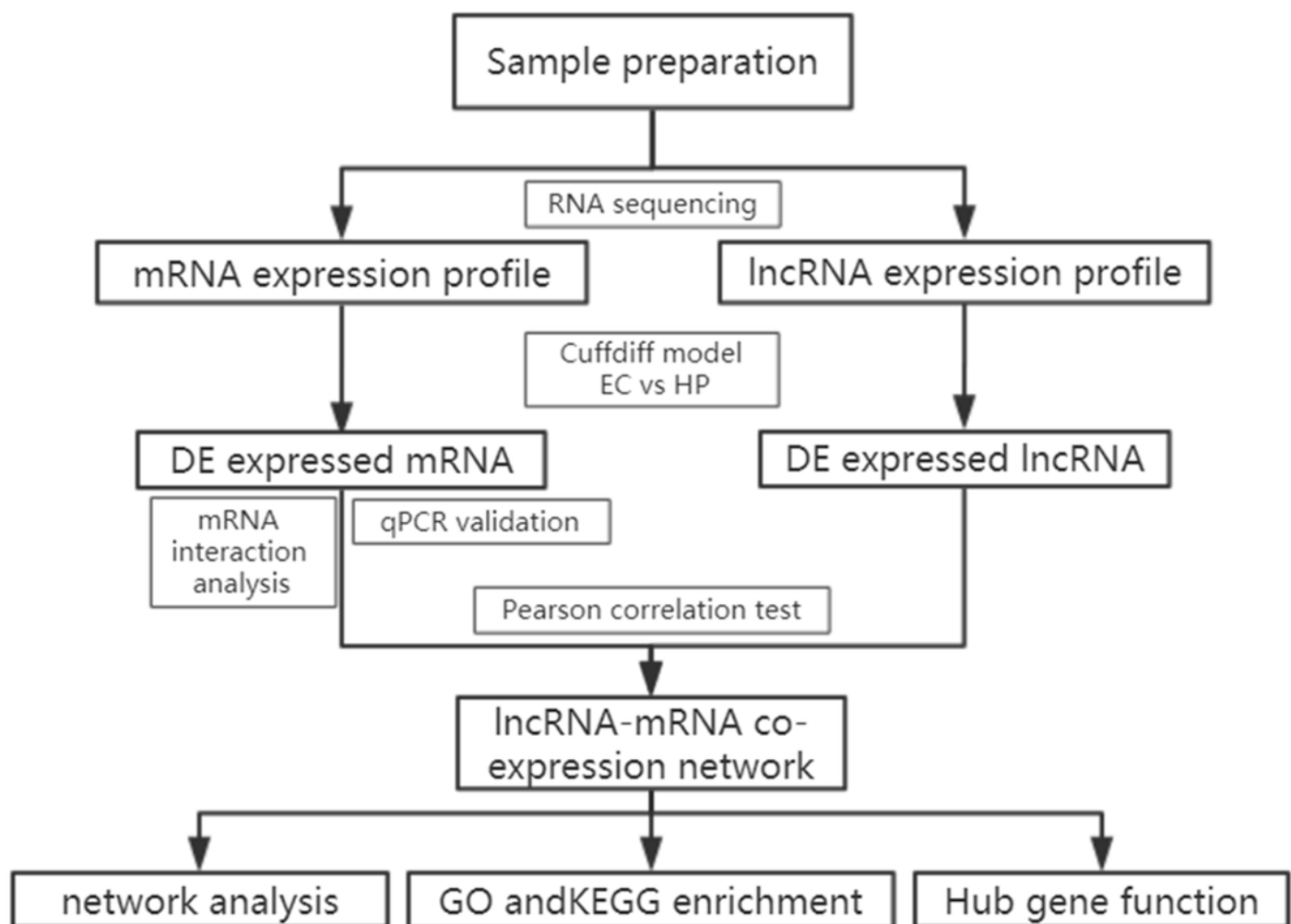
Hub Gene	Function Annotation	KEGG Entry
GNG2	Human cytomegalovirus infection	hsa05163
	Chemokine signaling pathway	hsa04062
	Kaposi sarcoma-associated herpesvirus infection	hsa05167
	Human immunodeficiency virus 1 infection	hsa05170
ATM	NF-kappa B signaling pathway	hsa04064
	FoxO signaling pathway	hsa04068
	Apoptosis	hsa04612
	Human T-cell leukemia virus 1 infection	hsa05166
	Human immunodeficiency virus 1 infection	hsa05170
HDAC6	Viral carcinogenesis	hsa05203
MAPK8	Apoptosis - multiple species	hsa04215
	NOD-like receptor signaling pathway	hsa04621
	Fc epsilon RI signaling pathway	hsa04664
	RIG-I-like receptor signaling pathway	hsa04622
	Inflammatory mediator regulation of TRP channels	hsa04750
	Toll-like receptor signaling pathway	hsa04620
SMC4	Epstein-Barr virus infection	hsa05169
TRRAP	Human T-cell leukemia virus 1 infection	hsa05166
IRF3	RIG-I-like receptor signaling pathway	hsa04622
	Herpes simplex infection	hsa05168
	Human papillomavirus infection	hsa05165

	Influenza A	hsa05164
	NOD-like receptor signaling pathway	hsa04621
	Toll-like receptor signaling pathway	hsa04620
MAPK9	TNF signaling pathway	hsa04668
	Focal adhesion	hsa04510
	IL-17 signaling pathway	hsa04657
	Th1 and Th2 cell differentiation	hsa04658
	Th17 cell differentiation	hsa04659
	T cell receptor signaling pathway	hsa04660

# Figure 1

Flowchart of the analysis procedure.

lncRNA, long non-coding RNA; EC, elite controller; HP, HIV-positive infected patient; DE, differentially expressed; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

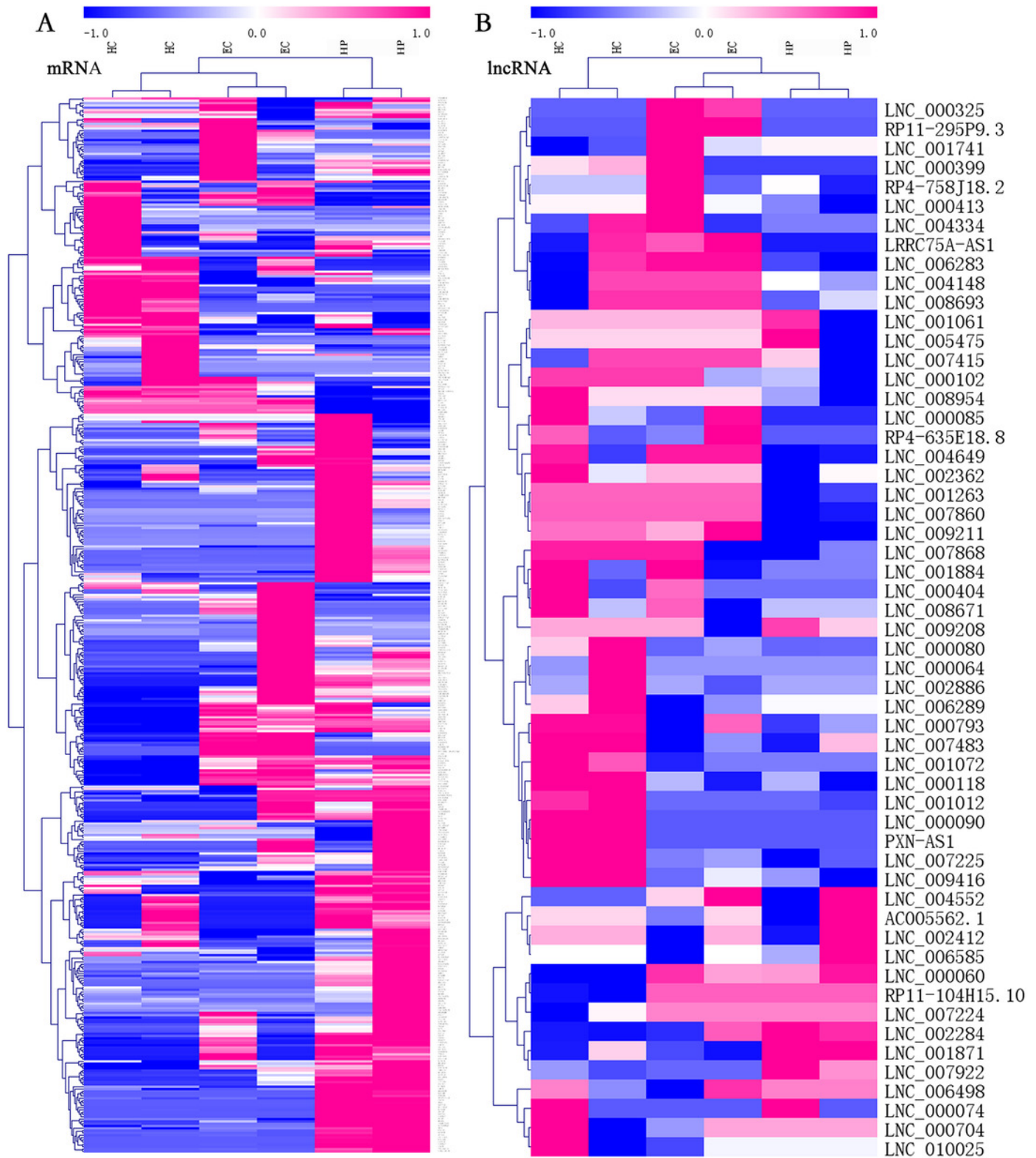


# Figure 2

Heatmaps of differentially expressed (DE) mRNAs and long non-coding RNAs (lncRNAs).

Hierarchical clustering analysis of sequencing detected mRNAs and lncRNAs with their expression abundance. Patient groups are on the horizontal axis, and mRNA and lncRNA genes are grouped along the vertical axis. Pinkish-purple bars indicate increased abundance of the corresponding genes, and blue bars indicates decreased abundance. White bars indicate that the corresponding genes were not detected. A) Expression abundance of mRNA genes between HC vs. EC vs. HP. B) Expression abundance of lncRNA genes between HC vs. EC vs. HP.



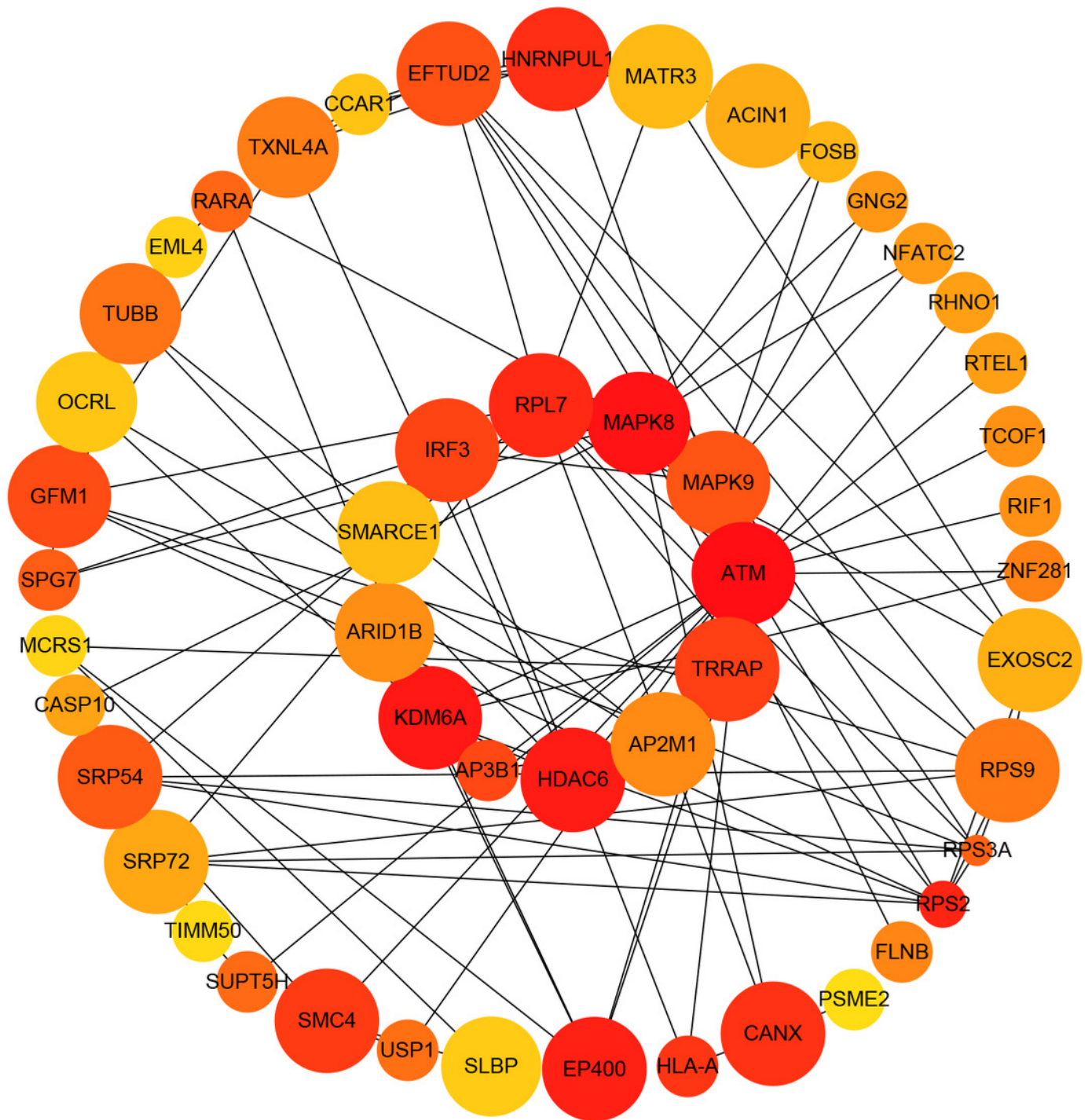




# Figure 4

Protein-protein interaction (PPI) network of differentially expressed (DE) mRNAs.

Top 50 hub mRNAs were included in using Closeness algorithm. The intensity of the color indicates the density of the correlation. The closer to red, the more central the gene is. The size of the circle represents the significance of a gene.

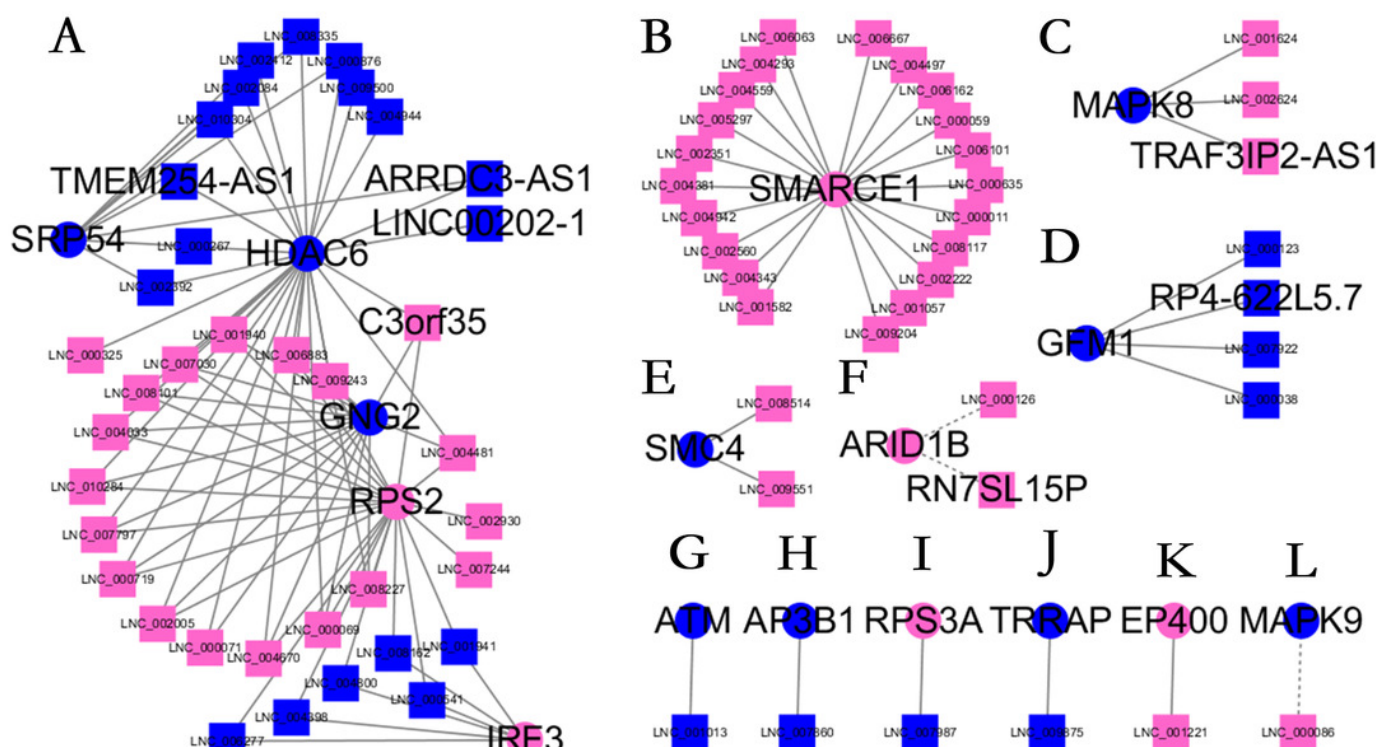




# Figure 5

Long non-coding RNA (lncRNA)-Hub mRNA co-expression network in elite controllers (ECs) and HIV-positive infected patients (HPs).

All hub mRNAs are presented in Fig 3 together with their associated lncRNAs. A) *HDAC6*, *SRP54*, *GNG2*, *RPS2*, *IRF3* and their associated lncRNAs. B) *SMARCE1* and its associated lncRNAs. C) *MAPK8* and its associated lncRNAs. D) *GFM1* and its associated lncRNAs. E) *SMC4* and its associated lncRNAs. F) *ARID1B* and its associated lncRNAs. G) *ATM* and *LNC\_001013*. H) *AP3B1* and *LNC\_007860*. I) *RPS3A* and *LNC\_007987*. J) *TRRAP* and *LNC\_009875*. K) *EP400* and *LNC\_001221*. L) *MAPK9* and *LNC\_000086*.



# Figure 6

Gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment classification of the predicted targeting mRNAs of the long non-coding RNA (lncRNA)-mRNA network.

The 10 most significantly enriched KEGG pathways and the 10 most significantly enriched GO terms in the biological process, cellular components, and molecular function categories. A) Top 10 GO terms enriched for predicted targeting mRNAs of the lncRNA-mRNA network. B) Top 10 KEGG terms enriched for predicted targeting mRNAs of the lncRNA-mRNA network.

