Infect Chemother. 2021 Sep;53(3):489-502
https://doi.org/10.3947/ic.2021.0031
pISSN 2093-2340·eISSN 2092-6448

Infection & Chemotherapy

Original Article

Transcriptome Analysis Identifies Altered Biological Processes and Novel Markers in Human Immunodeficiency Virus-1 Long-Term Non-Progressors

Dayeon Lee [1,2], Cheol-Hee Yoon [3], Sin Young Choi [1,2], Jung-Eun Kim [4], Young-Keol Cho [4], Byeong-Sun Choi [3], and Jihwan Park [1,2]

[1]School of Life Sciences, Gwangju Institute of Science and Technology (GIST), Gwangju, Korea
[2]Anti-Virus Research Center, Gwangju Institute of Science and Technology (GIST), Gwangju, Korea
[3]Division of Chronic Viral Disease, Center for Emerging Virus Research, National Institute of Infectious Disease, National Institute of Health, Cheongju, Korea
[4]Department of Microbiology, University of Ulsan College of Medicine, Seoul, Korea

Received: Mar 19, 2021
Accepted: Sep 3, 2021

Corresponding Authors:
Young-Keol Cho, PhD
Department of Microbiology, University of Ulsan College of Medicine, 88, Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea.
Tel: +82-2-3010-4283
Fax: +82-2-3010-4259
E-mail: ykcho2@amc.seoul.kr

Byeong-Sun Choi, PhD
Division of Chronic Viral Disease, Center for Emerging Virus Research, National Institute of Infectious Disease, National Institute of Health, 187, Osongsaengmyeong 2-ro, Osong-eup, Heungdeok-gu, Cheongju 28159, Chungcheongbuk-do, Korea.
Tel: +82-43-719-8410
Fax: +82-43-719-8459
E-mail: byeongsun@korea.kr

Jihwan Park, PhD
School of Life Sciences, Gwangju Institute of Science and Technology (GIST), 123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, Korea.
Tel: +82-62-715-2503
Fax: +82-62-715-2484
E-mail: jihwan.park@gist.ac.kr

ABSTRACT

Background: The latent reservoir of Human Immunodificiency Virus-1 (HIV-1) has been a major barrier to the complete eradication of HIV-1 and the development of HIV therapy. Long-term non-progressors (LTNPs) are a rare group of patients with HIV-1 who can spontaneously control HIV-1 replication without antiretroviral therapy. Transcriptome analysis is necessary to predict the pathways involved in the natural control of HIV-1, elucidate the mechanisms involved in LTNPs, and find biomarkers for HIV-1 reservoir therapy.
Materials and Methods: In this study, we obtained peripheral blood mononuclear cells from two LTNP subjects at multiple time points and performed RNA-sequencing analyses.
Results: We found that LTNPs and normal subjects had different transcriptome profiles. Functional annotation analysis identified that differentially expressed genes in LTNPs were enriched in several biological pathways such as cell cycle-related pathways and the transforming growth factor-beta signaling pathway. However, genes that were downregulated in LTNPs were associated with immune responses such as the interferon response and IL2-STAT5 signaling. Protein-protein interaction network analysis showed that CD8A, KLRD1, ASGR1, and MLKL, whose gene expression was upregulated in LTNPs, directly interacted with HIV-1 proteins. The network analysis also found that viral proteins potentially regulated host genes that were associated with immune system processes, metabolic processes, and gene expression regulation.
Conclusion: Our longitudinal transcriptome analysis of the LTNPs identified multiple previously undescribed pathways and genes that may be useful in the discovery of novel therapeutic targets and biomarkers.

Keywords: Human immunodeficiency virus; HIV-1 non-progressors; RNA-seq; Gene expression; Transcriptome analysis

ic Infection & Chemotherapy

ORCID iDs
Dayeon Lee
https://orcid.org/0000-0002-8006-8704
Cheol-Hee Yoon
https://orcid.org/0000-0002-5696-0567
Sin Young Choi
https://orcid.org/0000-0002-8188-9731
Jung-Eun Kim
https://orcid.org/0000-0001-9246-0159
Young-Keol Cho
https://orcid.org/0000-0003-0424-8911
Byeong-Sun Choi
https://orcid.org/0000-0002-6128-7995
Jihwan Park
https://orcid.org/0000-0002-5728-912X

Funding
This work was supported by "GIST Research Institute (GRI)," "GIST Research Institute (GRI) Integrated Institute of Biomedical Research (IIBR)," and "Anti-Virus Research Center" grants funded by GIST in 2021 and the Korea National Institute of Health (Grant Number: 2019-NI-067 and 2020-ER5103).

Conflict of Interest
No conflicts of interest.

Author Contributions
Conceptualization: JP, CHY, YKC, BSC. Data collection: SYC, JEK. Data curation: DL, JEK, YKC. Formal analysis: DL. Funding acquisition: JP, BSC. Investigation: CHY, JEK. Supervision: JP, YKC, BSC. Writing-original draft: DL. Writing-review & editing: JP, DL, BSC.

INTRODUCTION

Approximately, 2 million people are infected with Human Immunodeficiency Virus-1 (HIV-1) every year, and about 1.1 million people die from HIV-related diseases. HIV-1 usually infects CD4+ T cell and activates viral replication, leading to cell death and loss of immunocompetence [1]. Recently, various HIV treatments, such as antiretroviral therapy (ART), have been developed and utilized for the treatment of HIV and acquired immunodeficiency syndrome (AIDS) [2]. ART regulates viral replication by targeting certain steps in the HIV-1 life cycle, allowing ART-treated patients to maintain low levels of HIV RNA in the plasma. However, ART is still limited because of its inability to completely eradicate HIV-1. This is because resting memory CD4+ T cells that contain HIV-1 are still present in resting states [3]. The HIV-1 provirus does not replicate in the resting state, but the host still maintains low levels of viral load after ART. The cells in resting states that harbor HIV-1 virus are called the HIV-1 latent reservoir. The latent reservoir remains in the human body for a long time, and it can transition into the active state when the host cell conditions are appropriate for replication [4].

AIDS, which is defined as having CD4+ T cell counts less than 200 cells/μL after HIV infection, takes about 10 years to progress. Rarely, there are patients who can naturally control the progression of HIV infection. This group of patients is called long-term non-progressors (LTNPs). LTNPs have been reported to maintain normal counts of CD4+ T cells and have no clinical symptoms for at least 7 years even without the use of ART [5]. Similar to these LTNPs, a group of about 1% of patients with HIV known as elite control (EC) can maintain extremely low viral loads and CD4+ T cells >500 cells/μL for several years without any treatment [6]. Both these groups can spontaneously control disease progression and are considered to be important functional cure models for the development of HIV therapy.

Transcriptome analysis is necessary to predict the pathways involved in the natural control of HIV-1 and to elucidate the mechanisms utilized by LTNPs in the development of a functional treatment. RNA-sequencing (RNA-seq) is broadly used to identify transcriptional differences between conditions to discover biomarkers and ultimately elucidate the underlying mechanisms of diseases. Through RNA-seq and data analysis, we can find differentially expressed genes (DEGs) that give clues in the identification of LTNP-specific biological mechanisms.

Recent studies have contributed to the development of HIV-1 reservoir therapy via the identification of candidate genes and biomarkers involved in the functional cure of HIV-1 through analysis of the transcriptomes of LTNPs. A previous study performed RNA-seq on the peripheral blood mononuclear cells (PBMCs) of seven infected individuals with general disease progression, before and after ART, as well as 16 LTNPs [7]. The researchers identified 20 interferon-related antiviral signaling genes that could be used as markers of disease progression. It was also confirmed that levels of NFAT1/Elk-1 transcriptional factor-mediated genes and cell membrane proteins related to calcium signaling were elevated in LTNPs. Ding et al. identified mechanisms in LTNPs by integrating and analyzing microarray-based transcriptome data sets consisting of data on CD4+ T cells, CD8+ T cells, whole blood, etc. of 74 people, including 32 elite controllers [8]. The results showed that *CMPK1*, *CBX7*, *EIF3L*, *EIF4A*, and *ZNF395* had high correlations with control of HIV-1 progression. In addition, the reduction of ribosome components and inhibition of translational genes were found to be associated with AIDS progression.

ic Infection & Chemotherapy

Additionally, several transcriptome analyses have been performed to elucidate the mechanisms found in LTNPs and find markers related to the functional cure of HIV [6, 7, 9]. However, because there were few similarities between the previously mentioned studies and the samples were heterogeneous, the mechanisms and markers involved in LTNPs have not been clearly identified. It appears that this limitation is due to the use of variable patient datasets in which longitudinal analyses have not been performed. Furthermore, microarray analysis has low specificity and sensitivity in the identification of DEGs in LTNPs.

In this study, we performed RNA-seq analysis on samples of two LTNP subjects obtained at different time points as well as two healthy PBMC samples. By comparing the transcriptome profiles of the LTNPs with those of the healthy samples, we identified DEGs and elucidated the biological processes involved in LTNPs. Moreover, we found genes that interacted with HIV-1 proteins and their biological functions. Our study provides knowledge on common pathways among LTNPs and insight on the spontaneous control of HIV-1 replication by host cells.

MATERIALS AND METHODS

1. Patients and samples

Two patients who were infected with HIV-1 and showed characteristics of LTNPs for over 10 years were included in this study. The sampling was performed as described in a previous report [10]. Eleven and six PBMC samples were obtained from patient ID 92-23 and patient ID 91-20, respectively (**Supplementary Fig. 1**). Control PBMCs (no HIV) were purchased from STEMCELL (Vancouver, Canada). This study was approved by the Institutional Review Board of the Asan Medical Center (No. 2019-0888) and Korea Disease Control and Prevention Agency (No. 2019-07-06-P-A).

2. RNA isolation

Total RNA was obtained using TRIzol reagent (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's instructions. The quality of the RNA was assessed using Agilent 2100 bioanalyzer with the RNA 6000 Pico Chip (Agilent Technologies, Amstelveen, The Netherlands). The quantity of RNA was calculated using an ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

3. Library preparation and sequencing

The library was prepared using a QuantSeq 3' mRNA-Seq Library Prep Kit (Lexogen, Inc., Vienna, Austria) for the control and test RNAs. Illumina-compatible sequences containing oligo-dT primers were hybridized with the prepared RNAs. Then, reverse transcription was performed. Random primers were containing Illumina-compatible linker sequences were initiated second strand synthesis after degradation of the RNA templates. All reaction components were eliminated using magnetic beads, and the double-stranded library was purified. Complete adapter sequences were added to the amplified library for cluster generation.

The finished library was isolated from the PCR components. NextSeq500 was performed high-throughput sequencing via single-end 75 sequencing (Illumina, San Diego, CA, USA).

4. RNA-sequencing data analysis

Trim Galore is a wrapper tool of Cutadapt and FastQC. It was used to trim the adapters and check the quality of the raw sequence data [11]. The first 13 bp of the Illumina standard

https://doi.org/10.3947/ic.2021.0031

adapter was trimmed (using default options). A custom reference genome was generated by combining the Genome Reference Consortium Human build 38 (GRCh38/hg38) and the HIV-1 genome (MN043583) before aligning the reads to the genome. The trimmed reads were then aligned to the custom reference genome using the alignment software STAR [12]. To increase the mapping rate, limits on the mapped lengths were removed by changing the settings of STAR (--outFilterScoreMinOverLread 0, --outFilterMatchNminOverLread 0).

5. Principal component analysis

Before performing principal component analysis (PCA), variance stabilization was conducted via regularized-logarithm transformation provided by the DESeq2 package in the R software [13]. The similarities were calculated by considering the stabilized count tables of each sample. Each dot represented a sample, and distances between the dots indicated similarities. Shorter distances between two dots indicated that the two samples had similar gene expression profiles. PCA plots were generated in three types according to the following sample information: group information, CD4+ T cell counts, and HIV-1 detection scores. HIV-1 detection scores were calculated according to the presence or absence of HIV-1 proteins in each sample. The average values were calculated for each sample by assigning scores of 1 if the HIV-1 protein was present, 0.5 if there was partial presence, and 0 if the protein was absent. The heatmap was generated by selecting the top genes found by sorting the genes in descending order according to counts in the variance stabilized count tables. The top 500 genes were selected for clustering in the LTNP and healthy groups, and the top 350 genes were selected for clustering the 91-20-LTNP and 91-20-ART.

6. Identification of differentially expressed genes

To quantify the read fragments mapped to the protein-coding genes of the custom genome, featureCounts was used and the count matrix was generated [14]. To increase the percentage of assigned fragments, we changed the settings of STAR (-fraction TRUE). Using the count matrix, the DEGs of each group were identified using the DESeq2 R/Bioconductor packages [13]. Genes with *P*-values <0.05 were selected as DEGs. EnhancedVolcano package in R software was used to visualize the upregulated and downregulated genes of specific groups [15].

7. Gene set enrichment analysis and functional annotation analysis

Gene set enrichment analysis (GSEA) was performed to identify the statistically significant gene sets in the LTNP group [16]. The normalized count table made with DESeq2 was used for the analysis [13]. Hallmark gene sets from Molecular Signatures Database (MSigDB) were used [17]. Only gene sets containing 10 to 500 genes with false discovery rates (FDRs) <0.25 were selected. The functions of the DEGs were annotated using the database for annotation, visualization, and integrated discovery (DAVID) and Gene Ontology (GO) [18]. Biological processes (BP) were included in the GO. The enrichment of the BPs was ranked according to the sets of DEGs. Categories with more than 10 gene counts were selected, and bar plots were drawn. The minimum *P*-value for the significant BP category was 0.05. To draw the bar plot, the –log10 (*P*-value) of each BP was calculated.

8. Construction of the protein-protein interaction network and Gene Ontology analysis

Protein-protein interaction (PPI) network analysis and visualization of the interactions were performed using Cytoscape [19]. The DEGs of the LTNP group and HIV-1 proteins were used to analyze the functional network. An interaction score of 0.4 between the human genes and HIV-1 proteins was selected. Interaction groups were divided according to the distances

https://doi.org/10.3947/ic.2021.0031

ic Infection & Chemotherapy

from the HIV-1 nodes. Search tool for recurring instances of neighbouring genes (STRING) enrichment was used to determine the enrichments of the GO terms of each group [20]. Terms with more than 10 genes and FDRs <0.25 were selected.

RESULTS

1. Global transcriptome analysis of the LTNP, ART-treated, and healthy control samples

The PBMC samples from the two LTNP subjects collected at different time points and from the healthy subjects were analyzed in this study. The PBMC samples from the LTNP patients were collected at about 1-year intervals over 10 years. All patients' characteristics are shown in **Supplementary Table 1**. Subject 92-23 was classified as LTNP, with low levels of viral load and numbers of CD4+ T cells within the normal range (>500 cells/μL) for 10 years without any treatment. We used 10 PBMC samples from 92-23-LTNP. The other subject, 91-20, was diagnosed in 1991. Subject 91-20 divided into two groups, 91-20-LTNP and 91-20-ART. 91-20-LTNP maintained LTNP characteristics for 15 years but experienced a decrease in the number of CD4+ T cells due to virus reactivation typical progression. The numbers of CD4+ T cells of 91-20 decreased to ≤500 cells/μL since 2006, and ART treatment was initiated starting August 2007 (91-20-ART). 10 PBMC samples from 91-20-LTNP and 4 PBMC samples from 91-20-ART were used in this analysis. There was a year where HIV viral load information was not collected from both patients. Hence, we calculated HIV-1 detection scores using HIV-1 *gag, pol, vif, env,* and *nef* detection information from each patient's PBMC samples. The detailed calculation method can be found in the Materials and Methods section. A total of 26 samples (2 controls, 10 samples from 92-23-LTNP, 10 samples from 91-20-LTNP and 4 samples from 91-20-ART) were subjected to RNA sequencing.

PCA was performed to visualize the similarities between the LTNP and healthy control groups (**Fig. 1**). The PCA results showed that the 92-23-LTNP samples involved in the LTNP group were clustered with the control group because of the higher number of CD4+ T cells (**Fig. 1A**). However, the HIV-1 detection scores did not significantly affect the clustering patterns between the samples. On the other hand, the 91-20-LTNP and 91-20-ART samples were separated on the PCA plot, which was attributed to the CD4+ T cell counts rather than the HIV-1 detection scores (**Fig. 1B**). According to the hierarchical clustering, the LTNP samples were clustered together, and samples tended to be clustered according to CD4+ T cell counts (**Fig. 1C**). While the hierarchical clustering results showed that the 91-20-LTNP samples tended to be clustered together, the two 91-20-ART samples that were collected 2–3 years after ART treatment (ID_1091-2010 and ID_1191-204) were found to be clustered with the LTNP samples (**Fig. 1D**).

2. Biological processes associated with LTNPs

We performed GSEA to identify LTNP-specific BPs by comparing the gene expression of the LTNP (92-23-LTNP and 91-20-LTNP) groups with that of the control groups [16]. A normalized count table was used to compare features between the LTNP and control groups. We used 50 specific, well-defined gene sets provided by MSigDB for the GSEA [17]. The results of the analysis from comparing the LTNP and control groups identified that mitotic spindle (198 genes), G2M checkpoint (199 genes), E2F target (199 genes), and transforming growth factor (TGF)-beta signaling (199 genes) were significantly enriched in the LTNP samples (**Fig. 2A**). In contrast, the gene sets highly enriched in the control groups were interferon alpha response (97 genes), IL-2 STAT5 signaling (86 genes), IL-6 JAK STAT3

https://doi.org/10.3947/ic.2021.0031

ic Infection & Chemotherapy

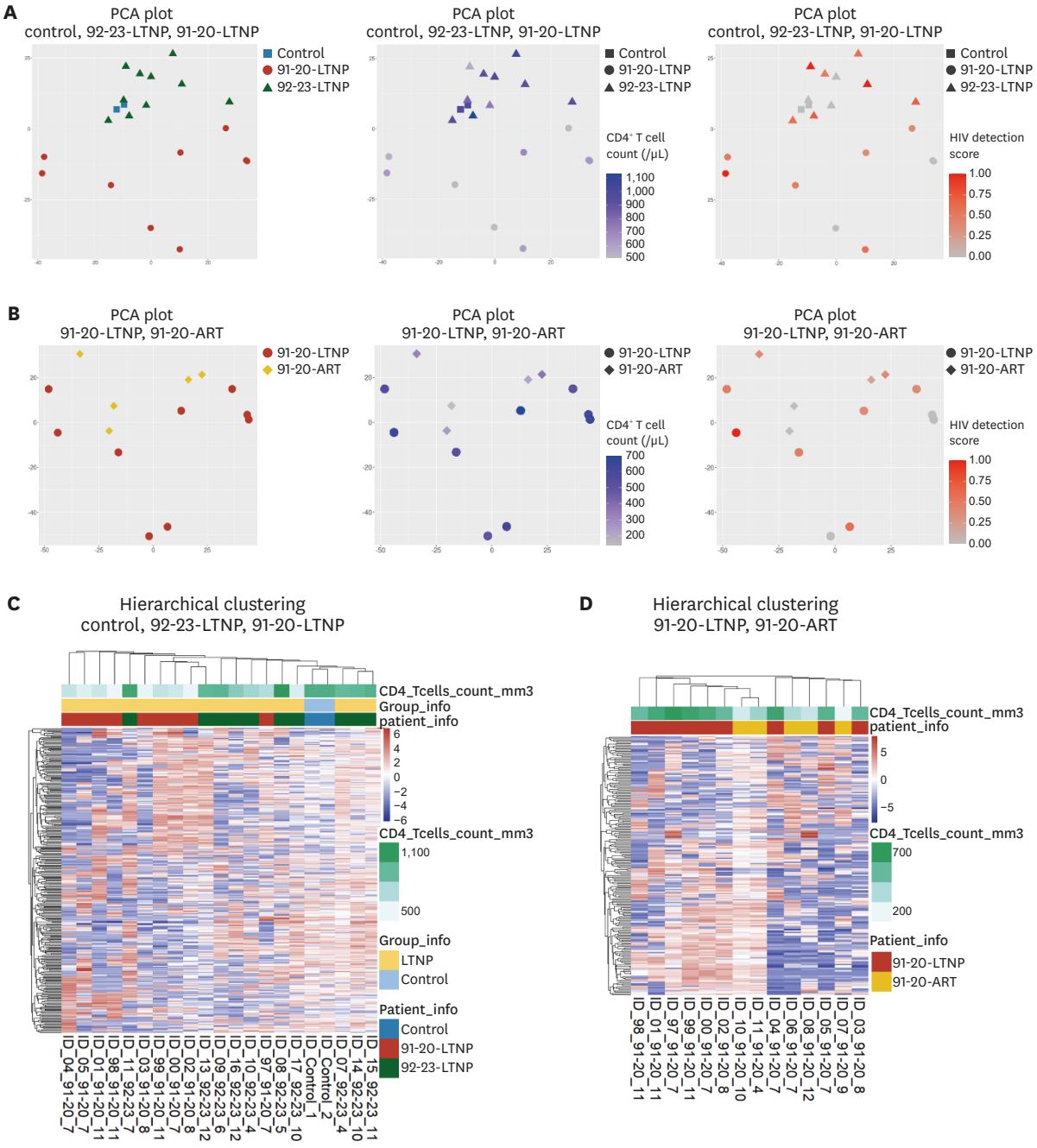

Figure 1. Global transcriptome analysis of LTNP, ART-treated, and healthy control samples.
(A) Principal component analysis plots of the LTNP and healthy (control) RNA-sequencing data (n = 20). Each dot indicates an individual sample. The distance between two dots represents the similarities between the samples. In the left, the characteristics of the samples are indicated by colors. Both of middle and right, the characteristics of the samples are indicated by shapes. In the middle, blue indicates the counts of CD4+ T cells. On the right, red indicates HIV-1 detection scores. (B) Principal component analysis plots of the 91-20-LTNP samples and 91-20-ART samples (n = 14). (C) The top 500 genes were used for the hierarchical clustering of the 92-23-LTNP and 91-20-LTNP and control samples. The dendrogram at the top of the graph shows the similarities between the samples. The color of the heatmap represents the intensities of gene expression. Red indicates strong gene expression, and blue indicates weak gene expression. (D) The top 350 genes were used for the hierarchical clustering of the 91-20-LTNP and 91-20-ART samples.
LTNP, long-term non-progressor; ART, Antiretroviral therapy; HIV, human immunodeficiency virus; PCA, principal component analysis.

signaling (87 genes), and interferon gamma response (200 genes) (**Fig. 2B**). We selected core enrichment genes from each gene set and visualized their expression across the samples with a heatmap (**Fig. 2C**, **2D** and **Supplementary Fig. 2**). The heatmaps clearly showed the

https://doi.org/10.3947/ic.2021.0031

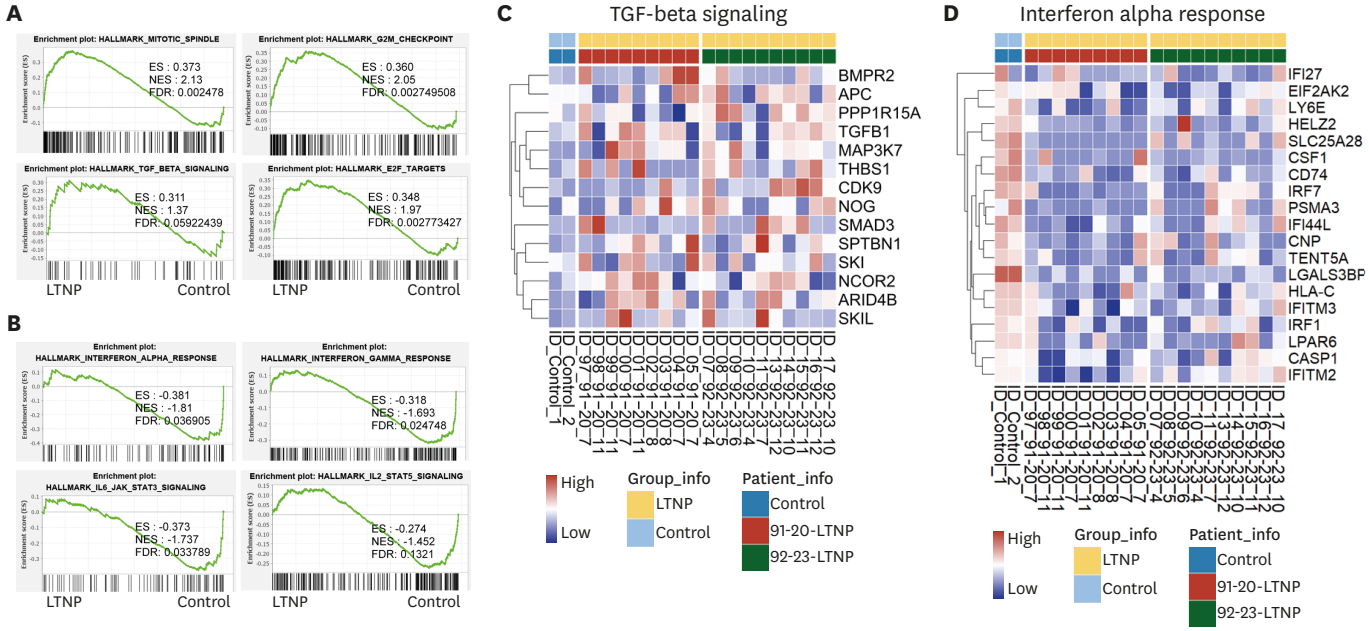

Figure 2. Biological processes associated with the LTNPs.
(A and B) Gene sets enriched in the LTNP (A) and control groups (B). GSEA was performed on the LTNP and control groups. The GSEA plot shows the significant signatures of the LTNP and control groups by using ranked gene lists. The x-axis represents the DEGs which are ranked according to fold changes (LTNPs vs control). The black bars represent enriched genes. The y-axis represents the enrichment scores and ranked list metrics. ES, NES, and FDR are indicated for each gene set. (C and D) Heatmap of the core enrichment genes from the gene sets related to TGF-beta signaling (C) and interferon alpha response (D). The Z-score of each gene was calculated.
LTNP, long-term non-progressor; GSEA, gene set enrichment analysis; DEGs, differentially expressed genes; ES, enrichment scores; NES, normalized enrichment scores; FDR, false discovery rates; TGF, transforming growth factor.

differential expression patterns of the genes that were associated with those biological functions between the LTNP and control samples. These data suggest that cell cycle-related genes and TGF-beta signaling genes were upregulated, whereas immune response genes were downregulated in the LTNP samples.

3. Functional annotation analysis of LTNP-specific DEGs

DEGs were determined by comparing three groups using DESeq2 to identify genes significantly expressed in the LTNP group [13]. As described in the Materials and Methods, the genes were filtered using the cutoff values, *P*-value <0.05 and |logFC| >1. Upregulated and downregulated genes are indicated by red dots on the volcano plots (**Fig. 3A**). We focused on the upregulated genes of the LTNP group to identify LTNP-specific genes. Comparison analysis identified 946, 1116, and 906 upregulated genes between 92-23-LTNP *vs.* control, 91-20-LTNP *vs.* control, and 91-20-LTNP vs 91-20-LTNP, respectively. We identified 263 genes by overlapping the DEGs of 91-20-LTNP with the control and 91-20-LTNP (**Supplementary Table 2**). Functional annotation analysis was performed with DAVID using the DEGs [18]. The analysis showed that the DEGs of LTNPs were enriched in variable BPs (**Fig. 3C**). Among the several BPs, the top seven categories are shown in the plots. The DEGs of 92-23-LTNP were enriched in "DNA repair" (GO:0006281, 20 DEGs), "regulation of cell cycle" (GO:005172, 13 DEGs), and "transcription, DNA-templated" (GO:0006351, 89 DEGs). Additionally, the DEGs of 91-20-LTNP were enriched in "DNA repair" (GO:0006281, 23 DEGs), "cell division" (GO:0051301, 27 DEGs), "transcription, DNA-templated" (GO:0006351, 105 DEGs), and "TGF-beta receptor signaling pathway" (GO:0007179, 10 DEGs) upon comparison with that of the control group. These functional annotation results indicated that the LTNP DEGs were related to regulation of the cell cycle, transcription, and TGF-beta signaling. Moreover, the

https://doi.org/10.3947/ic.2021.0031

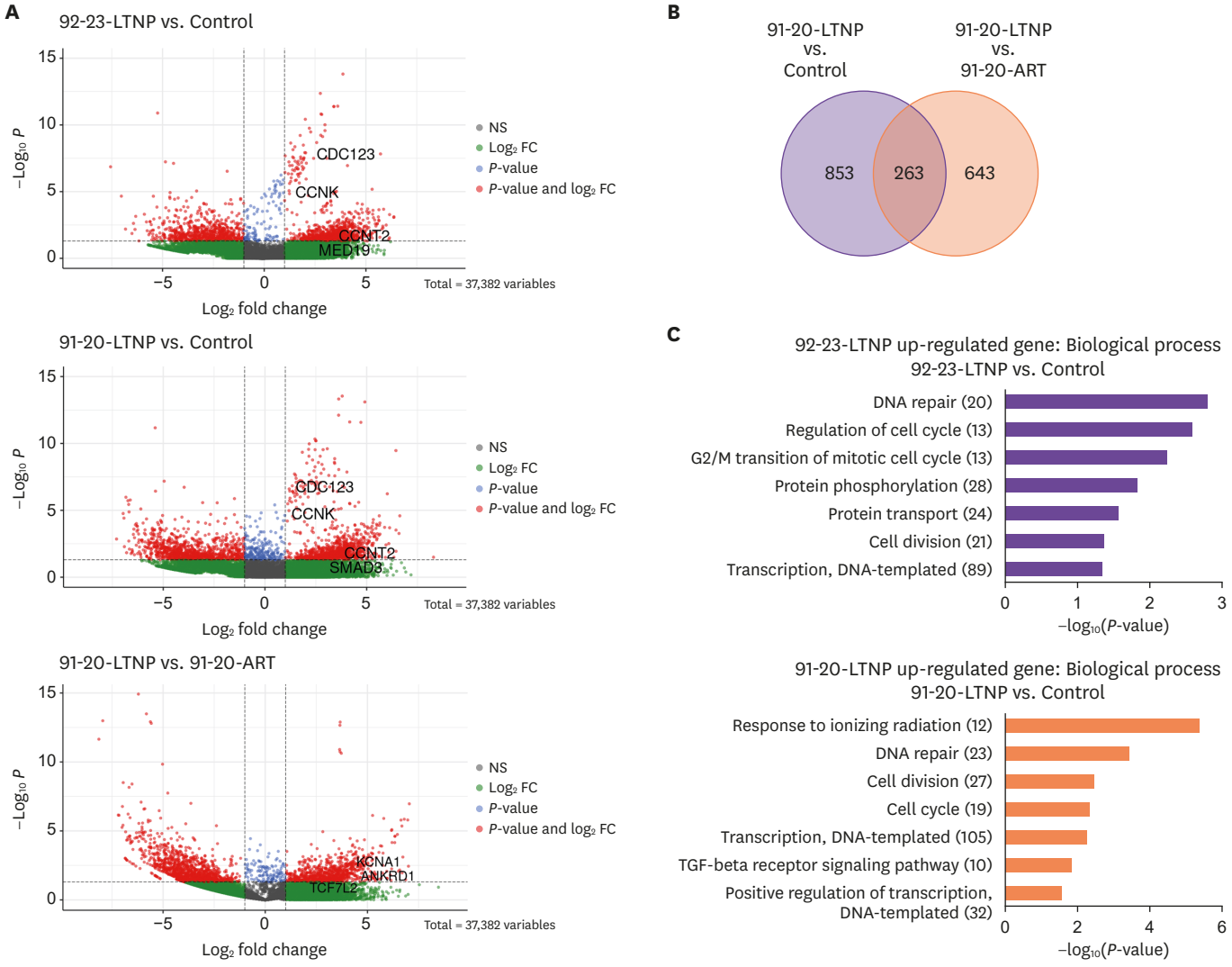

Figure 3. Functional annotation analysis of the LTNP-specific genes.
(A) Volcano plots of the DEGs of 92-23-LTNP vs control, 91-20-LTNP vs control, and 91-20-LTNP vs 91-20-ART. The red dots represent the DEGs of the LTNPs which have *P*-values <0.05 and log2 fold changes >1 or <−1. The upregulated genes are on the right side, and the downregulated genes are on the left side. The names of the genes included in significant biological process are shown in the plot. (B) Venn diagram shows the DEGs (log2 fold change >1 or <−1) of the 91-20-LTNP. (C) Functional annotation of the 92-23-LTNP and 91-20-LTNP using the database for annotation, visualization, and integrated discovery with *P*-values <0.05. The 91-20-LTNP compared with the control and 91-20-ART. The number next to the biological process name represents the number of DEGs involved with it. Each bar indicates the value of -log10 (*P*-value). The biological process is considered significant if the value is large.
LTNP, long-term non-progressor; DEGs, differentially expressed genes; ART, antiretroviral therapy; NS, non-significant; FC, fold change; DNA, deoxyribonucleic acid.

DEGs of 91-20-LTNP enriched in "cellular response to TGF-beta stimulus" (GO:0071560, 6 DEGs), "protein phosphorylation" (GO:0006468, 21 DEGs), and "Wnt signaling pathway" (GO:0016055, 11 DEGs) upon comparison with that of the 91-20-ART. This result showed that TGF-beta signaling-related genes were significantly enriched in 91-20-LTNP. Moreover, we proposed 6 genes, including EFCAB7, IRS1, KCNA1, KCNA2, KCNJ2 and SGCB as a candidate to diagnose LTNP in HIV-1 from experimental results. (**Supplementary Table 2**). A total of 263 LTNP-specific DEGs were used to select the candidate genes. After uploading the LTNP-specific DEGs to Ensembl BioMart, we filtered only the gene corresponding to gene ontology accession "plasma membrane protein complex" (GO:0098797) [21]. Since these candidate genes encode membrane protein and logFC is greater than 2, they can be utilized as potential diagnosis markers for LTNP.

https://doi.org/10.3947/ic.2021.0031

ic Infection & Chemotherapy

4. Protein-protein interaction network analysis reveals the involvement of host cellular pathways in regulating viral processes in LTNPs

PPI analysis was performed using Cytoscape to analyze the interaction network between the human and HIV-1 proteins [19]. The LTNP DEGs were matched with human proteins based on the STRING database and subsequently used for the analysis [20]. A total of 946 and 1,116 DEGs corresponding to 92-23-LTNP and 91-20-LTNP, respectively, were used to create the PPI network, in which 476 genes were found to be common in both patients. The STRING database matched 73 LTNP DEGs to protein information. According to the PPI network, five human proteins (CD8A, MLKL, KLRD1, ASGR1, and CSMD1) directly interacted with the HIV-1 proteins, and four of them (CD8A, MLKL, KLRD1, and ASGR1) corresponded to common DEGs between the two patients (**Fig. 4A**). In addition, most of the other proteins closely interacted with the HIV-1 proteins (TBX21, FCGR2B, CX3CR1, SERPINB8, and MICA), which corresponded to the common LTNP DEGs. A total of 86 nodes were divided into two groups according to their interaction strengths with the HIV-1 proteins. Excluding the HIV-1 protein nodes, group 1 contained 18 nodes and group 2 contained 55 nodes (**Fig 4A**). We identified the enriched BPs in each group using STRING enrichment [20]. Group 1 was mostly enriched in the regulation of immune system processes and response (9 nodes). In addition, the proteins closely interacting with HIV-1 proteins were also enriched in lymphocyte-mediated immunity (4 nodes). The group 2 nodes were mostly enriched in metabolic processes (44 nodes). Especially, 25 nodes with high interaction strengths in group 2 were enriched in the regulation of gene expression. We selected genes from the group 1 and 2 nodes and visualized their expression across the LTNP and control samples using a heatmap (**Fig. 4B** and **4C**). The heatmaps show clear changes in expression of the genes that have direct or indirect interactions with HIV-1 proteins in the LTNP samples, compared with that of the control samples. These results showed that the LTNP DEGs interacting with the HIV-1 proteins were involved in metabolic pathways, transcription regulation, and immune responses.

DISCUSSION

LTNPs exhibit rare phenotypes of HIV-1 infectious populations with long disease progressions. LTNPs can naturally control viremia, a function that researchers have been studying through the gene expression profiles of LTNPs to find specific gene expression patterns that may aid in the development of functional treatments. As there are few studies focusing on common transcription programs or markers, comprehensive research needed to be performed to elucidate the regulatory mechanisms found in LTNPs. In this study, we analyzed samples from several time points to obtain several rare LTNP patient samples and performed transcriptomic analysis related to longitudinal LTNPs (**Supplementary Fig. 1**). We identified functional gene groups and signaling pathways by obtaining and analyzing the RNA-sequencing data of the LTNPs to discover candidates for the development of treatments and identification of biomarkers.

A previous study showed that genes related to the interphase of the mitotic cell cycle and the Ran in mitotic spindle were upregulated in latent cells, in which HIV-1 does not actively replicate [22]. This result suggests that infected cells can control disease progression by regulating the cell cycle and transcription. Consistent with previous studies, our functional annotation analysis revealed that the LTNP DEGs were enriched in the gene sets of the mitotic spindle, G2/M checkpoint, E2F target, and TGF-beta signaling pathway.

https://doi.org/10.3947/ic.2021.0031

ic Infection & Chemotherapy

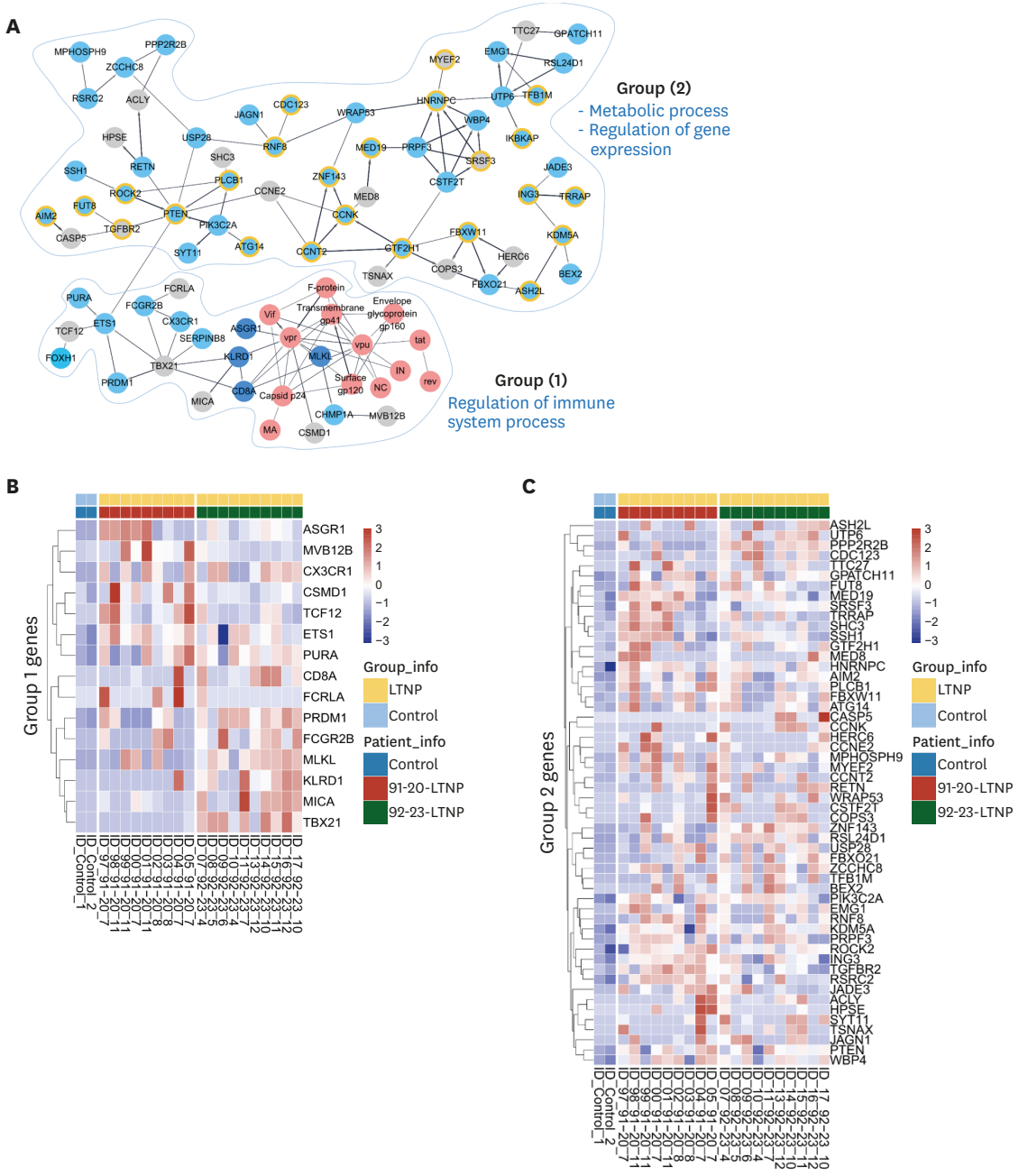

Figure 4. Constructing the protein-protein interaction network of the HIV-1 proteins and DEGs of LTNPs.
(A) Using the STRING database, the interactions between HIV-1 proteins and the LTNP-specific genes were analyzed. The red nodes indicate HIV-1 proteins, the blue nodes indicate the overlapping DEGs of 92-23-LTNP and 91-20-LTNP, and the grey nodes indicate the DEGs of 92-23-LTNP or 91-20-LTNP. The nodes with yellow borders are involved in the regulation of gene expression. The line thicknesses indicate the strengths of interactions. The nodes are divided into two groups according to distance from the HIV-1 node. Gene ontology analysis was performed on each group. (B) Heatmap of the genes corresponding to the blue and grey nodes in group 1. Z-scores of all genes are shown. (C) Heatmap of the genes corresponding to the blue and grey nodes in group 2. Z-scores of all genes are shown. HIV, human immunodeficiency virus; DEGs, differentially expressed genes; LTNP, long-term non-progressor.

Both GSEA and DAVID results showed that the TGF-beta signaling genes were highly expressed in the LTNP group. TGF-beta is known as an anti-inflammatory and anti-immunogenic cytokine that is involved in various pathways related to T cell status regulation. According to our results, the alpha-(1,6)-fucosyltransferase 8 (*FUT8*) gene was upregulated in the LTNP group. This gene encodes FUT8, a protein that stimulates TGF-beta binding

https://doi.org/10.3947/ic.2021.0031

and facilitates downstream TGF-beta signaling [23]. Our PPI analysis results identified that FUT8 interacts with TGF-beta receptor 2 (TGFBR2). High expression of TGF-beta have been observed in HIV-infected cells, in which TGF-beta inhibits HIV-1 expression and replication by inhibiting the HIV reverse transcriptase [24]. In addition, TGF-beta 1 promotes HIV-1 latency by promoting the expression of BLIMP-1, a transcription inhibitor [25]. Furthermore, TGF-beta regulates not only transcription but also cell cycle progression. In normal cells, the STAT3 and STAT5 transcription factors are activated by IL-2 to induce the cell cycle progression of T cells [26]. It modulates the IL-2-mediated pathway and inhibits T cell proliferation by blocking the G1/S phase transition. Through GSEA, we found that IL-2-STAT5 signaling-related genes were downregulated in the LTNP group compared with that of the control group. Moreover, upregulation of TGF-beta induced the anti-immunogenic response and decreased the cytotoxicity of CD8+ T cells [27]. Indeed, dysfunctional CD8+ T cell cannot kill infected CD4+ T cells, allowing LTNPs to maintain normal ranges of CD4+ T cell counts unless there are detectable viral loads. Thus, TGF-beta signaling contributes to the maintenance of LTNP characteristics by regulating transcription, cell cycle progression, and immunogenic function in HIV-1-infected T cells.

PPI analysis showed the interactions between HIV-1 proteins and upregulated genes identified in the LTNP samples. Through analysis of the expression profiles, we found that the cyclin K gene (*CCNK*) was upregulated in the LTNP samples. CCNK is a transcriptional regulator that interacts with Nef (viral protein) and inhibits the replication and transcription of HIV-1, thereby interfering with HIV-1 pathogenesis [28]. The elongation factor b (P-TEFb) is a novel cellular complex involved in viral gene replication and transcription that includes the CDK9-Cyclin T1 (CCNT1) complex [29]. To allow the expression of viral genes, CDK9 and cyclin T1 form a complex. However, when Nef is present, CCNK interacts with Nef and forms a complex with CDK9 instead of CCNT1, inhibiting HIV-1 gene expression and replication. Additionally, the HIV-1 Tat protein enhances HIV-1 transcription by recruiting protein kinases like P-TEFb and activating RNA polymerase II [30]. From our data, we found that the Cyclin T2 (*CCNT2*) gene was upregulated in the LTNP samples. The CCNT2-CDK9 complex is a subunit of P-TEFb and a negative regulator of Tat protein-mediated elongation [31]. Moreover, CCNT2 interacts with the general transcription factor IIH Subunit 1 (GTF2H1), a cell growth suppressor, allowing for disruption of Tat-driven elongation [32, 33]. Consequently, several genes encoding negative regulators of HIV-1 replication and transcription were also upregulated in the LTNP samples. High expression of HIV-1 protein-mediated negative regulators causes HIV-1 to be dysfunctional in the host cells and allows for maintenance of low levels of viral load.

Our data showed that the T-box transcription factor (*TBX21*) gene was upregulated in the LTNP samples. Network analysis showed that the TBX21 protein interacted with FCGR2B, CX3CR1, and SERPIN8.

The gene *TBX21* encodes the protein TBX21, also known as T-bet, which is a protein related to the TH1-type immune response [34]. HIV-1-infected CD4+ T cells use the T-bet dependent pathway. The extracellular Tat protein interacts with other cellular factors to induce the T-bet dependent pathway. T-bet regulates the expression of TH1 cytokines like IFN-γ. After HIV-1 infection, IFN-γ inhibits Tat protein interaction and promotes HIV-1 latency by impeding HIV-1 replication in TH1 cells. Our results suggest that infected CD4+ T cells sustain latency in LTNPs by actively controlling HIV-1 replication through immune responses under the environment containing Tat protein.

https://doi.org/10.3947/ic.2021.0031

ic Infection & Chemotherapy

We performed transcriptome analysis using RNA-sequencing. DEGs of LTNP samples were classified into several BPs. Our results suggest that LTNPs spontaneously regulate the replication and transcription of HIV-1 through the TGF-beta signaling pathway. Interestingly, HIV-1 proteins like Nef or Tat are also involved in the regulation of HIV-1 replication in infected cells by interacting with human proteins. However, this study has some limitations such as the small number of samples used for transcriptome analysis. In addition, although variable cell types are involved in HIV-1 disease progression and latency, we only used bulk RNA-sequencing data, which can only provide the average gene expression of heterogeneous cell types in PBMCs. Therefore, for further study, single cell RNA-sequencing can be utilized to identify the common cell types found in LTNPs and cell type-specific DEGs in LTNPs.

ACKNOWLEDGEMENTS

We are grateful to e-Biogen (Seoul, Korea) for initial analysis of bulk RNA-seq data.

SUPPLEMENTARY MATERIALS

Supplementary Table 1

Clinical characteristisc of LTNP patients involved in the study

Click here to view

Supplementary Table 2

Differentially expressed genes of 91-20-LTNP vs 91-20-ART

Click here to view

Supplementary Figure 1

Click here to view

Supplementary Figure 2

Click here to view

REFERENCES

1. Stevenson M. HIV-1 pathogenesis. Nat Med 2003;9:853-60.
PUBMED | CROSSREF
2. Volberding PA, Deeks SG. Antiretroviral therapy and management of HIV infection. Lancet 2010;376:49-62.
PUBMED | CROSSREF
3. Sengupta S, Siliciano RF. Targeting the latent reservoir for HIV-1. Immunity 2018;48:872-95.
PUBMED | CROSSREF
4. Hill AL, Rosenbloom DI, Fu F, Nowak MA, Siliciano RF. Predicting the outcomes of treatment to eradicate the latent reservoir for HIV-1. Proc Natl Acad Sci U S A 2014;111:13475-80.
PUBMED | CROSSREF
5. Kumar P. Long term non-progressor (LTNP) HIV infection. Indian J Med Res 2013;138:291-3.
PUBMED

6. Blankson JN, Bailey JR, Thayil S, Yang HC, Lassen K, Lai J, Gandhi SK, Siliciano JD, Williams TM, Siliciano RF. Isolation and characterization of replication-competent human immunodeficiency virus type 1 from a subset of elite suppressors. J Virol 2007;81:2508-18.
PUBMED | CROSSREF

7. Díez-Fuertes F, De La Torre-Tarazona HE, Calonge E, Pernas M, Alonso-Socas MDM, Capa L, García-Pérez J, Sakuntabhai A, Alcamí J. Transcriptome sequencing of peripheral blood mononuclear cells from elite controller-long term non progressors. Sci Rep 2019;9:14265.
PUBMED | CROSSREF

8. Ding JW, Ma L, Zhao JY, Xie YL, Zhou JM, Li XY, Cen S. An integrative genomic analysis of transcriptional profiles identifies characteristic genes and patterns in HIV-infected long-term non-progressors and elite controllers. J Transl Med 2019;17:35.
PUBMED | CROSSREF

9. Zhang LL, Zhang ZN, Wu X, Jiang YJ, Fu YJ, Shang H. Transcriptomic meta-analysis identifies gene expression characteristics in various samples of HIV-infected patients with nonprogressive disease. J Transl Med 2017;15:191.
PUBMED | CROSSREF

10. Cho YK, Kim JE, Lee SH, Foley BT, Choi BS. Impact of HIV-1 subtype and Korean Red Ginseng on AIDS progression: comparison of subtype B and subtype D. J Ginseng Res 2019;43:312-8.
PUBMED | CROSSREF

11. Babraham Bioinfomatics. Trim galore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries. Availlable at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. Accessed 27 May 2020.

12. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21.
PUBMED | CROSSREF

13. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
PUBMED | CROSSREF

14. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923-30.
PUBMED | CROSSREF

15. Blighe K, Rana S, Lewis M. EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. Available at: https://bioconductor.org/packages/release/bioc/vignettes/EnhancedVolcano/inst/doc/EnhancedVolcano.html. Accessed 1 September 2020.

16. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.
PUBMED | CROSSREF

17. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015;1:417-25.
PUBMED | CROSSREF

18. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003;4:3.
PUBMED | CROSSREF

19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-504.
PUBMED | CROSSREF

20. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. J Proteome Res 2019;18:623-32.
PUBMED | CROSSREF

21. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Charkhchi M, Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, El Houdaigui B, Fatima R, Gall A, Garcia Giron C, Grego T, Guijarro-Clarke C, Haggerty L, Hemrom A, Hourlier T, Izuogu OG, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Gonzalez Martinez J, Marugán JC, Maurel T, McMahon AC, Mohanan S, Moore B, Muffato M, Oheh DN, Paraschas D, Parker A, Parton A, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Steed E, Szpak M, Szuba M, Taylor K, Thormann A, Threadgold G, Walts B, Winterbottom A, Chakiachvili M, Chaubal A, De Silva N, Flint B, Frankish A, Hunt SE, IIsley GR, Langridge N, Loveland JE, Martin

FJ, Mudge JM, Morales J, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Cunningham F, Yates AD, Zerbino DR, Flicek P. Ensembl 2021. Nucleic Acids Res 2021;49 D1:D884-91.
PUBMED | CROSSREF

22. Yang J, Yang Z, Lv H, Lou Y, Wang J, Wu N. Bridging HIV-1 cellular latency and clinical long-term non-progressor: an interactomic view. PLoS One 2013;8:e55791.
PUBMED | CROSSREF

23. Zhang J, Ten Dijke P, Wuhrer M, Zhang T. Role of glycosylation in TGF-β signaling and epithelial-to-mesenchymal transition in cancer. Protein Cell 2021;12:89-106.
PUBMED | CROSSREF

24. Poli G, Kinter AL, Justement JS, Bressler P, Kehrl JH, Fauci AS. Transforming growth factor beta suppresses human immunodeficiency virus expression and replication in infected cells of the monocyte/macrophage lineage. J Exp Med 1991;173:589-97.
PUBMED | CROSSREF

25. Chinnapaiyan S, Dutta RK, Nair M, Chand HS, Rahman I, Unwalla HJ. TGF-β1 increases viral burden and promotes HIV-1 latency in primary differentiated human bronchial epithelial cells. Sci Rep 2019;9:12552.
PUBMED | CROSSREF

26. Bright JJ, Kerr LD, Sriram S. TGF-beta inhibits IL-2-induced tyrosine phosphorylation and activation of Jak-1 and Stat 5 in T lymphocytes. J Immunol 1997;159:175-83.
PUBMED

27. Theron AJ, Anderson R, Rossouw TM, Steel HC. The role of transforming growth factor beta-1 in the progression of hiv/aids and development of non-aids-defining fibrotic disorders. Front Immunol 2017;8:1461.
PUBMED | CROSSREF

28. Khan SZ, Mitra D. Cyclin K inhibits HIV-1 gene expression and replication by interfering with cyclin-dependent kinase 9 (CDK9)-cyclin T1 interaction in Nef-dependent manner. J Biol Chem 2011;286:22943-54.
PUBMED | CROSSREF

29. Kim JB, Sharp PA. Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. J Biol Chem 2001;276:12317-23.
PUBMED | CROSSREF

30. Parada CA, Roeder RG. Enhanced processivity of RNA polymerase II triggered by Tat-induced phosphorylation of its carboxy-terminal domain. Nature 1996;384:375-8.
PUBMED | CROSSREF

31. Ramakrishnan R, Liu H, Donahue H, Malovannaya A, Qin J, Rice AP. Identification of novel CDK9 and Cyclin T1-associated protein complexes (CCAPs) whose siRNA depletion enhances HIV-1 Tat function. Retrovirology 2012;9:90.
PUBMED | CROSSREF

32. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, Stec E, Ferrer M, Strulovici B, Hazuda DJ, Espeseth AS. Genome-scale RNAi screen for host factors required for HIV replication. Cell Host Microbe 2008;4:495-504.
PUBMED | CROSSREF

33. Jain S, Arrais J, Venkatachari NJ, Ayyavoo V, Bar-Joseph Z. Reconstructing the temporal progression of HIV-1 immune response pathways. Bioinformatics 2016;32:i253-61.
PUBMED | CROSSREF

34. Kulkarni A, Ravi DS, Singh K, Rampalli S, Parekh V, Mitra D, Chattopadhyay S. HIV-1 Tat modulates T-bet expression and induces Th1 type of immune response. Biochem Biophys Res Commun 2005;329:706-12.
PUBMED | CROSSREF