RESEARCH ARTICLE   Open Access

Meta-analysis of gene expression profiles in long-term non-progressors infected with HIV-1

Sun Young Lee, Yong Kwang Park, Cheol-Hee Yoon, Kisoon Kim and Kyung-Chang Kim[*]

Abstract

Background: In the absence of antiretroviral treatments (ARTs), a small group of individuals infected with HIV, including long-term non-progressors (LTNPs) who maintain high levels of CD4+ T cells for more than 7–10 years in the absence of ART and in particular a subgroup of LTNPs, elite controllers (ECs), who have low levels of viremia, remain clinically and/or immunologically stable for years. However, the mechanism of stable disease progression in LTNPs and ECs needs to be elucidated to help those infected with HIV-1 remain healthy. In this study, to identify the characteristics of gene expression profiles and biomarkers in LTNPs, we performed a meta-analysis using multiple gene expression profiles among LTNPs, individuals infected with HIV-1 without ART, individuals infected with HIV-1 with ART, and healthy controls.

Methods: The gene expression profiles obtained from the Gene Expression Omnibus (GEO) microarray data repositories were classified into three groups: LTNPs versus healthy controls (first group, 3 studies), LTNPs versus patients infected with HIV-1 without ART (second group, 3 studies), and LTNPs versus patients infected with HIV-1 with ART (third group, 3 studies). In addition, we considered a fourth group, patients infected with HIV-1 without ART versus healthy controls (3 studies), to exclude genes associated with HIV-1 infection in the three groups. For each group, we performed a meta-analysis using the *RankProd* method to identify and compare the differentially expressed genes (DEGs) in the three groups.

Results: We identified the 14 common DEGs in the three groups when comparing them with each other. Most belonged to immune responses, antigen processing and presentation, the interferon-gamma-mediated signaling pathway, and T cell co-stimulation. Of these DEGs, *PHLDA1* was up-regulated and *ACTB* and *ACTG1* were down-regulated in all three groups. However, the rest of the up- or down-regulated genes were discordant in the three groups. Additionally, *ACTB* and *ACTG1* are known to inhibit viral assembly and production, and *THBS1* is known to inhibit HIV-1 infection.

Conclusions: These results suggest that significant genes identified in a meta-analysis provide clues to the cause of delayed disease progression and give a deeper understanding of HIV pathogenesis in LTNPs.

Keywords: HIV-1, Long-term non-progressors, Gene expression profile, Biomarkers, Meta-analysis

* Correspondence: joytodeath@korea.kr
Division of Viral Disease Research, Center for Infectious Disease Research,
Korea National Institute of Health, 187 Osongsaengmyeong 2-ro, Cheongju,
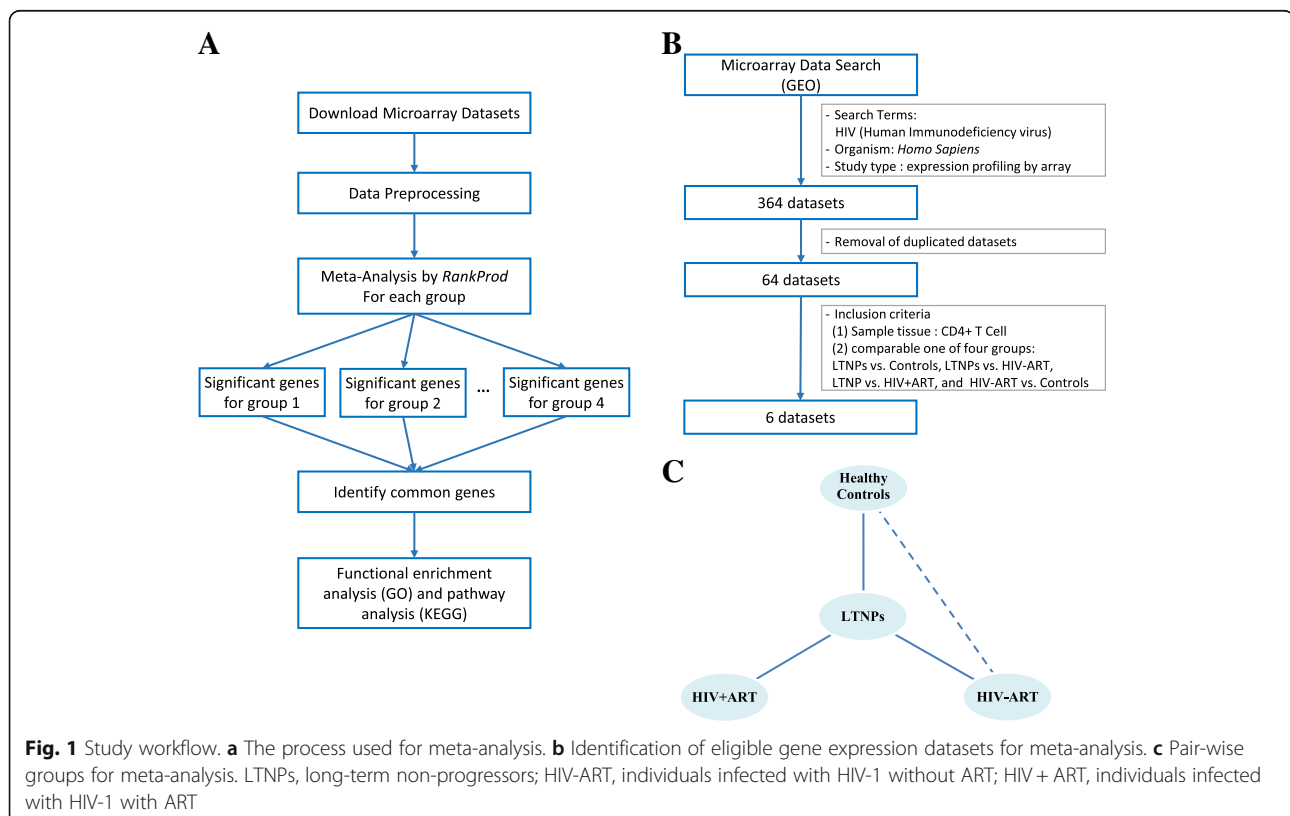Chungbuk 28159, Republic of Korea

Background

In the absence of antiretroviral treatments (ARTs), most individuals infected with HIV exhibit progressive active viral transcription and replication, which result into the loss of CD4+ T cells, appearance of clinical symptoms, and disease progression. In contrast, a small group of individuals infected with HIV, including long-term non-progressors (LTNPs) who maintain high levels of CD4+ T cells for more than 7–10 years in the absence of ART and especially a subgroup of LTNPs, elite controllers (ECs), who have low levels of viremia, remain clinically and/or immunologically stable for years. Understanding the mechanisms of delayed disease progression in LTNPs would be valuable for HIV disease management and treatment, but they remain poorly understood. Previous studies using gene expression profiling [1–5] and meta-analysis [6] in LTNPs have identified that the genes related to interferon responses are a signature for progression. It is usually not advisable to directly combine or compare the gene expression values from different gene expression data sets because of their inherent heterogeneity (i.e., different platforms, protocols, and so on). Instead, the integration of various expression data sets is performed on the summary-level data such as *P* values, effect sizes, or gene ranks. Meta-analysis, a statistical approach that combines results from independent but related studies, has the potential to increase both the statistical power and generalizability of single-study analysis [7] and is most commonly used for the purpose of detecting differentially expressed genes (DEGs) [8]. The application of meta-analysis to clinical datasets has demonstrated that it achieves more reliable identification than an individual analysis and more robust rank products in gene ranking, which lead to much higher reproducibility among independent studies [9]. Thus, to identify the characteristics of gene expression profiles and biomarkers or factors in LTNPs or understand the factors specifically associated with LTNPs, we performed a meta-analysis of gene expression profiles in LTNPs. The gene expression profiles obtained from the Gene Expression Omnibus (GEO) microarray data repositories [10] were classified into four groups. For each group, we performed the meta-analysis using the *RankProd* [11] method to identify DEGs in LTNPs and then compared the DEGs of each group. In this study, we performed a meta-analysis of gene expression profiles in LTNPs to identify the characteristics of delayed disease progression and provide a deeper understanding of HIV pathogenesis.

Methods

Procedure

We performed the analysis in accordance with the workflow illustrated in Fig. 1a for the meta-analysis of gene

Fig. 1 Study workflow. **a** The process used for meta-analysis. **b** Identification of eligible gene expression datasets for meta-analysis. **c** Pair-wise groups for meta-analysis. LTNPs, long-term non-progressors; HIV-ART, individuals infected with HIV-1 without ART; HIV + ART, individuals infected with HIV-1 with ART

expression profiles in LTNPs and to identify candidate genes that characterized LTNPs. First, we obtained the microarray datasets from GEO microarray data repositories. After removing duplicates and irrelevant datasets, six microarray datasets were considered and classified into four groups according to the inclusion criteria. Next, we identified the significant genes in each group using the meta-analysis method *RankProd* [11] and selected potential biomarkers for LTNPs from these significant genes. Finally, Functional enrichment analysis (Gene ontology (GO)) and pathway analysis (Kyoto Encyclopedia of Genes and Genomes, KEGG) were performed.

Microarray datasets

As shown Fig. 1b, we searched the gene expression profiles using the keywords "Human Immunodeficiency virus/HIV" in GEO microarray data repositories. The datasets were selected using the following criteria: tissue section with peripheral blood mononuclear cell (PBMC); gene expression data; and the species *Homo sapiens*. After removing duplicates and irrelevant datasets, a total of 64 microarray datasets were considered. Of these, six datasets were finally selected in accordance with the inclusion criteria. The inclusion criteria were as follows: tissue with CD4+ T cell and mutual comparability in four groups (Fig. 1c), including LTNPs versus healthy controls (LTNPs vs. Controls), LTNPs versus patients

infected with HIV-1 without ART (LTNPs vs. HIV-ART), LTNPs versus patients infected with HIV-1 with ART (LTNPs vs. HIV + ART), and patients infected with HIV-1 without ART versus healthy controls (HIV-ART vs. Controls). Here, patients infected with HIV-1 without ART were chronic patients 1 year after HIV infection. Six datasets were organized into four groups (Table 1): LTNPs versus healthy controls (first group, 3 studies, LC), LTNPs versus patients infected with HIV-1 without ART (second group, 3 studies, LH), LTNPs versus patients infected with HIV-1 with ART (third group, 3 studies, LA), and patients infected with HIV-1 without ART versus healthy controls (fourth group, 3 studies, HC).

Meta-analysis

Meta-analyses are necessary to efficiently integrate and validate related datasets produced by independent groups. Meta-analysis step were performed using R version 3.5.0 [12] and Bioconductor 3.6 (http://www.bioconductor.org) with packages, *RankProd* [11] and *limma* [13]. To identify DEGs in samples between LTNPs and others, the datasets of each group were computed using the *RankProd* [11] method, which is a nonparametric approach based on ranks of fold changes (FC). In this approach, FC ratios were computed for all possible pair-wise comparisons for each dataset. The ranks of the ratios were then used to generate a rank product for

Table 1 Summary of datasets analyzed in the meta-analysis

Series Dataset GEO ID		No. of Samples		Sample Source	Platform
		Case	Control		
		LTNPs	Controls		
1 (LC)	GSE6740 [1]	5	5	CD4+ T cell	GPL96, Affymetrix Human Genome U133A array
	GSE18233 [51]	16	8	CD4+ T cell	GPL6884, Illumina Human WG-6 v3.0 expression bead chip
	GSE23879 [2]	12	12	CD4+ T cell	GPL8432, Illumina HumanRef-8 WG-DASL V3.0
		LTNPs	HIV-ART		
2 (LH)	GSE6740 [1]	5	5	CD4+ T cell	GPL96, Affymetrix Human Genome U133A array
	GSE18233 [51]	16	73	CD4+ T cell	GPL6884, Illumina Human WG-6 v3.0 expression bead chip
	GSE28128 [52]	14	27	CD4+ T cell	GPL8432, Illumina HumanRef-8 WG-DASL V3.0
		LTNPs	HIV + ART		
3 (LA)	GSE18233 [51]	16	29	CD4+ T cell	GPL6884, Illumina Human WG-6 v3.0 expression bead chip
	GSE23879 [2]	12	15	CD4+ T cell	GPL8432, Illumina HumanRef-8 WG-DASL V3.0
	GSE53527 [53]	11	10	CD4+ T cell	GPL10558, Illumina HumanT-12 V4.0 expression bead chip
		HIV-ART	Controls		
4 (HC)	GSE6740 [1]	5	5	CD4+ T cell	GPL96, Affymetrix Human Genome U133A array
	GSE18233 [51]	70	8	CD4+ T cell	GPL6884, Illumina Human WG-6 v3.0 expression bead chip
	GSE9927 [54]	11	9	CD4+ T cell	GPL570, Affymetrix Human Genome U133 plus2.0 array

GEO Gene expression omnibus, *GSE* Gene expression series, *No* Number, *HIV-ART* Patients infected with HIV-1 without ART, *HIV + ART* Patients infected with HIV-1 with ART, *LTNPs* Long-term non-progressors, *LC* LTNPs vs. healthy controls, *LH* LTNPs vs. HIV-ART, *LA* LTNPs vs. HIV + ART, *HC* HIV-ART vs. healthy controls

each gene. Rank products performs well when datasets had low numbers of samples [14]. Permutation tests were then performed to assess the null distributions of the rank products within each dataset. The whole process was repeated multiple times and required powerful computers for the large datasets. We have tested *RankProd* [11] in a workstation (MacPro, 12-core Intel Xeon E5 2.7GHz processors, 64GB memory, macOS Sierra; Apple, Inc., Cupertino California, USA). We used microarray datasets from GEO and these datasets have been pre-processed. The expression values of microarray data were processed using quantile normalization. Data were annotated after converting the gene and probe IDs to the corresponding Entrez IDs and the expression values of same Entrez ID aggregated with mean value. The normalized datasets were then subjected to *limma* [13] to search for DEGs. The intensity values for each probe set were log2 transformed for data integrity.

Functional enrichment analysis of DEGs

Functional enrichment analysis for DEGs was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID) [15]. The DAVID is the most common tool to analyze the functional enrichment of genes and identify the significantly enriched GO categories [16]. We also performed the pathway enrichment analysis utilizing the KEGG pathway.

Results

Identification of DEGs in the individual microarray data of each group

We calculated DEGs for the individual microarray data in the four groups (Additional file 1). Genes with fold changes ($|\log_2 FC| >= 1$) and *adj. p*-values < 0.05 were considered as differentially expressed. Individual analysis revealed little information correlated with LTNPs and the absence of common DEGs except in the fourth group. It is usually not advisable to directly combine or compare the gene expression values from different gene expression datasets because of their inherent heterogeneity.

Identification of significant genes using meta-analysis in each group

We identified DEGs using meta-analysis in the four groups (Fig. 2a and Additional file 2). The list of DEGs is provided in Additional file 3. To consider only features of LTNPs, we excluded DEGs of the fourth group from DEGs of the three other groups. According to the results of our meta-analysis, 330 genes, 886 genes, and 141 genes were identified to be differentially expressed in the LC group, LH group, and LA group, respectively. Moreover, the 10 most significantly up-regulated and down-regulated genes were identified in each group (Fig. 2b-d). Genes that were identified in the HC group

were removed from the list of up- and down-regulated genes.

Functional enrichment analysis and enriched KEGG pathways for each group

GO analysis and KEGG pathway analysis for the functional investigation of DEGs were carried out. The significantly top enriched GO terms for DEGs in the LC group included immune response, signal transduction, and positive regulation of cell proliferation (Fig. 3a). In the LH group, translation initiation, viral transcription, and immune response were included (Fig. 3b). Interferon-gamma-mediated signaling pathway and immune response were included in the LA group (Fig. 3c). According to our expectations, immune response was significantly enriched in all three groups.
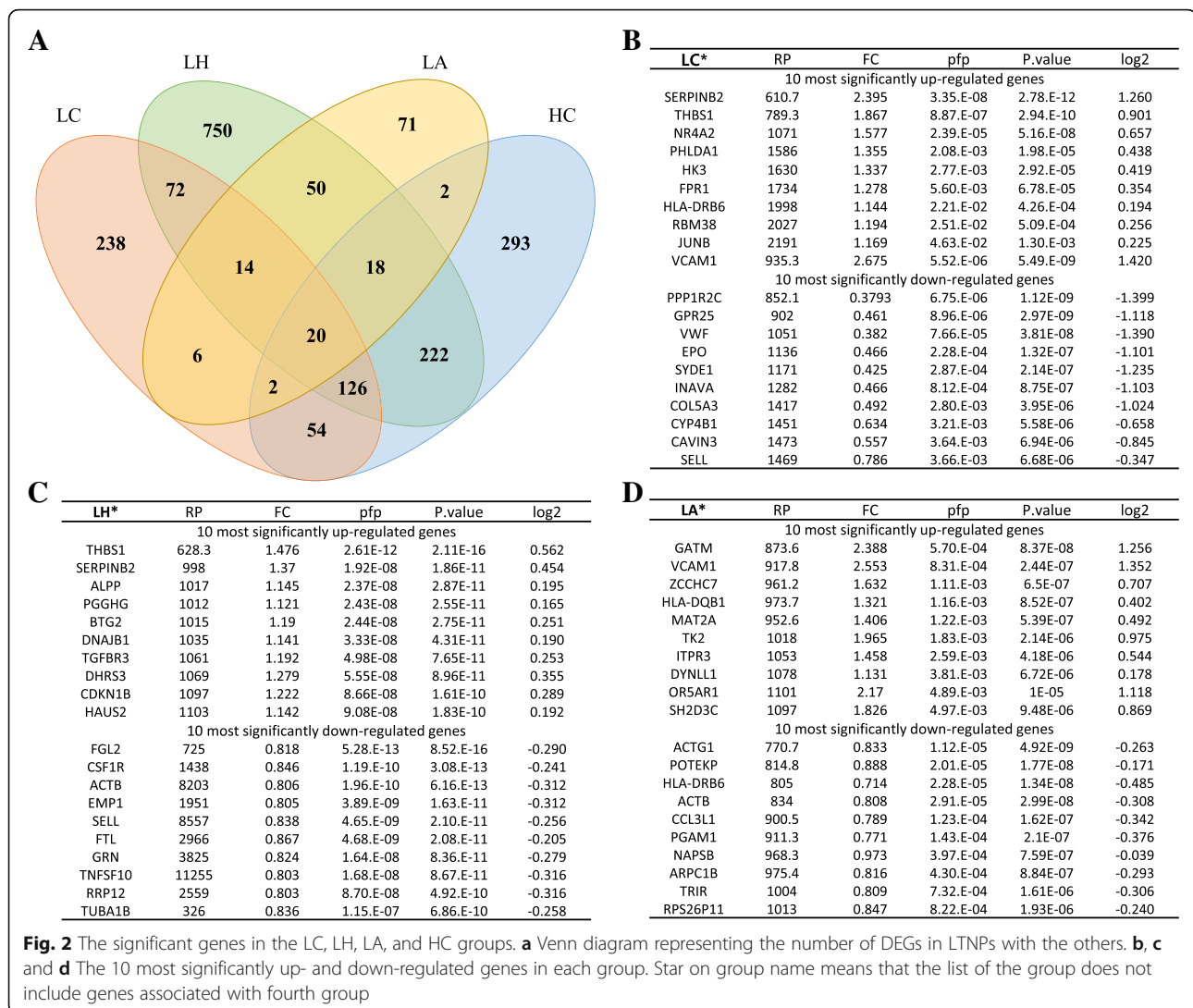
In the pathway analysis, we identified the most significant pathways for DEGs in the three groups. These pathways were as follows: pathways including cytokine-cytokine receptor interaction and Jak-STAT signaling pathway in the LC group, antigen processing and presentation in the LH group, and platelet activation in the LA group (Fig. 4a-c).

Functional enrichment analysis and enriched KEGG pathways derived from the common DEGs

The significant genes found to be common in all three groups were considered significant genes for LTNPs (Fig. 5a). As a result, 14 genes were identified as common DEGs. Of them, *PHLDA1* was up-regulated and *ACTB* and *ACTG1* were down-regulated in all three groups. However, the up- and down-regulation of the rest of the genes were discordant in the three groups (Fig. 5b). As shown in Fig. 6, antigen processing and presentation, immune response, and immune system process were significantly enriched in the 14 common DEGs related to LTNPs. Moreover, phagosome and Rap1 signaling pathways were identified as the significant pathways for the 14 common DEGs.

Discussion

LTNPs are a small group of patients who exhibit natural control of HIV-1 viral replication and lack of clinical progression. That is, LTNPs maintain normal counts of CD4+ T cells and controlled viremia without ART for many years. In individuals infected with HIV-1, viral control is a major characteristic of the clinical outcome. Understanding the molecular mechanisms involved in the HIV pathogenesis of LTNPs will provide clues to identify the cause of delayed disease progression in LTNPs. Genome wide association studies (GWAS) were utilized to discover genetic factors and pathways involved in HIV-1 control [17–20]. Tsiara et al. [21] performed meta-analysis of GWAS to investigate the

A Venn diagram with groups LC, LH, LA, HC showing numbers: 238, 72, 750, 50, 71, 2, 293, 14, 18, 20, 6, 2, 126, 222, 54.

B

LC*	RP	FC	pfp	P.value	log2
10 most significantly up-regulated genes					
SERPINB2	610.7	2.395	3.35.E-08	2.78.E-12	1.260
THBS1	789.3	1.867	8.87.E-07	2.94.E-10	0.901
NR4A2	1071	1.577	2.39.E-05	5.16.E-08	0.657
PHLDA1	1586	1.355	2.08.E-03	1.98.E-05	0.438
HK3	1630	1.337	2.77.E-03	2.92.E-05	0.419
FPR1	1734	1.278	5.60.E-03	6.78.E-05	0.354
HLA-DRB6	1998	1.144	2.21.E-02	4.26.E-04	0.194
RBM38	2027	1.194	2.51.E-02	5.09.E-04	0.256
JUNB	2191	1.169	4.63.E-02	1.30.E-03	0.225
VCAM1	935.3	2.675	5.52.E-06	5.49.E-09	1.420
10 most significantly down-regulated genes					
PPP1R2C	852.1	0.3793	6.75.E-06	1.12E-09	-1.399
GPR25	902	0.461	8.96.E-06	2.97.E-09	-1.118
VWF	1051	0.382	7.66.E-05	3.81E-08	-1.390
EPO	1136	0.466	2.28.E-04	1.32E-07	-1.101
SYDE1	1171	0.425	2.87.E-04	2.14E-07	-1.235
INAVA	1282	0.466	8.12.E-04	8.75E-07	-1.103
COL5A3	1417	0.492	2.80.E-03	3.95E-06	-1.024
CYP4B1	1451	0.634	3.21.E-03	5.58.E-06	-0.658
CAVIN3	1473	0.557	3.64.E-03	6.94.E-06	-0.845
SELL	1469	0.786	3.66.E-03	6.68.E-06	-0.347

C

LH*	RP	FC	pfp	P.value	log2
10 most significantly up-regulated genes					
THBS1	628.3	1.476	2.61E-12	2.11E-16	0.562
SERPINB2	998	1.37	1.92E-08	1.86E-11	0.454
ALPP	1017	1.145	2.37E-08	2.87E-11	0.195
PGGHG	1012	1.121	2.43E-08	2.55E-11	0.165
BTG2	1015	1.19	2.44E-08	2.75E-11	0.251
DNAJB1	1035	1.141	3.33E-08	4.31E-11	0.190
TGFBR3	1061	1.192	4.98E-08	7.65E-11	0.253
DHRS3	1069	1.279	5.55E-08	8.96E-11	0.355
CDKN1B	1097	1.222	8.66E-08	1.61E-10	0.289
HAUS2	1103	1.142	9.08E-08	1.83E-10	0.192
10 most significantly down-regulated genes					
FGL2	725	0.818	5.28.E-13	8.52.E-16	-0.290
CSF1R	1438	0.846	1.19.E-10	3.08.E-13	-0.241
ACTB	8203	0.806	1.96.E-10	6.16.E-13	-0.312
EMP1	1951	0.805	3.89.E-09	1.63.E-11	-0.312
SELL	8557	0.838	4.65.E-09	2.10.E-11	-0.256
FTL	2966	0.867	4.68.E-09	2.08.E-11	-0.205
GRN	3825	0.824	1.64.E-08	8.36.E-11	-0.279
TNFSF10	11255	0.803	1.68.E-08	8.67.E-11	-0.316
RRP12	2559	0.803	8.70.E-08	4.92.E-10	-0.316
TUBA1B	326	0.836	1.15.E-07	6.86.E-10	-0.258

D

LA*	RP	FC	pfp	P.value	log2
10 most significantly up-regulated genes					
GATM	873.6	2.388	5.70.E-04	8.37E-08	1.256
VCAM1	917.8	2.553	8.31.E-04	2.44E-07	1.352
ZCCHC7	961.2	1.632	1.11.E-03	6.5E-07	0.707
HLA-DQB1	973.7	1.321	1.16.E-03	8.52.E-07	0.402
MAT2A	952.6	1.406	1.22.E-03	5.39E-07	0.492
TK2	1018	1.965	1.83.E-03	2.14E-06	0.975
ITPR3	1053	1.458	2.59.E-03	4.18.E-06	0.544
DYNLL1	1078	1.131	3.81.E-03	6.72E-06	0.178
OR5AR1	1101	2.17	4.89.E-03	1E-05	1.118
SH2D3C	1097	1.826	4.97.E-03	9.48E-06	0.869
10 most significantly down-regulated genes					
ACTG1	770.7	0.833	1.12.E-05	4.92.E-09	-0.263
POTEKP	814.8	0.888	2.01.E-05	1.77E-08	-0.171
HLA-DRB6	805	0.714	2.28.E-05	1.34E-08	-0.485
ACTB	834	0.808	2.91.E-05	2.99E-08	-0.308
CCL3L1	900.5	0.789	1.23.E-04	1.62E-07	-0.342
PGAM1	911.3	0.771	1.43.E-04	2.1E-07	-0.376
NAPSB	968.3	0.973	3.97.E-04	7.59E-07	-0.039
ARPC1B	975.4	0.816	4.30.E-04	8.84E-07	-0.293
TRIR	1004	0.809	7.32.E-04	1.61E-06	-0.306
RPS26P11	1013	0.847	8.22.E-04	1.93E-06	-0.240

Fig. 2 The significant genes in the LC, LH, LA, and HC groups. **a** Venn diagram representing the number of DEGs in LTNPs with the others. **b**, **c** and **d** The 10 most significantly up- and down-regulated genes in each group. Star on group name means that the list of the group does not include genes associated with fourth group
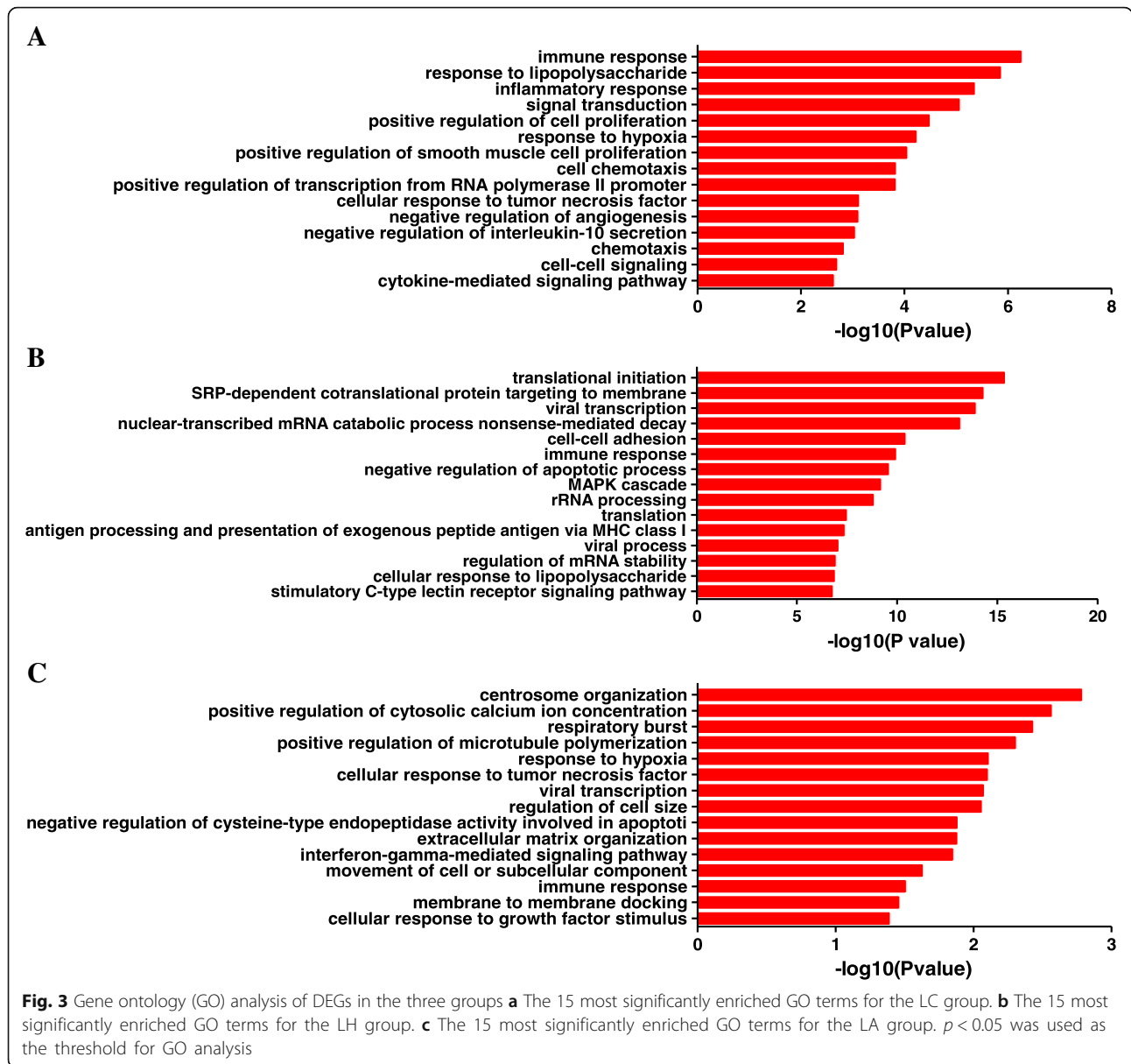
association between interleukin gene polymorphisms and HIV susceptibility. In general, many studies of HIV gene expression have placed emphasis on the comparison in between normal progressors and non-progressors using transcriptional expression profiles [1–5] or meta-analysis [6]. But, in this study, the comparison of gene expression had performed in all the groups, healthy donors, treated HIV-1 patients, non-treated HIV-1 patients and LTNPs.

In our study, we analyzed the gene expression profiles of LTNPs using meta-analysis to understand HIV pathogenic mechanisms. The gene expression profiles of LTNPs were compared with HIV-1 patients, treated HIV-1 patients, and health controls. GO term analysis and KEGG pathway analysis provide understanding of the possible roles of significant genes in the pathogenesis of LTNPs. In all groups, the immune response GO term was significantly enriched. We found that the salmonella infection pathway was significantly enriched in all three groups. Meanwhile, the pathogenic *Escherichia coli* infection pathway was enriched in LH and LA groups. Many researchers have studied the association with bacterial translocation and immune activation of HIV in intestinal microbiome [22, 23]. It was found that microbiota composition is different in between LTNPs and HIV patients [24] and also microbiome dysbiosis is less marked in treated HIV patients and LNTPs than HIV patients [25]. These results suggest that microbiome in LTNPs may have an effect on immune control and HIV disease progression.

We identified the 14 significant common genes. The GO terms of the 14 common genes significantly expressed in LTNPs were associated with immune response, movement of cell, and regulation of cell death. Of them, *THBS1, SERPINB2, NR4A2, DYNLL1, FPR1, JUNB, HLA-DQB1, MS4A6A, ACTB,* and *ACTG1* have been reported to have interactions with HIV-1 proteins [26–32]. However, little is known about their functions

Fig. 3 Gene ontology (GO) analysis of DEGs in the three groups **a** The 15 most significantly enriched GO terms for the LC group. **b** The 15 most significantly enriched GO terms for the LH group. **c** The 15 most significantly enriched GO terms for the LA group. $p < 0.05$ was used as the threshold for GO analysis

in association with LTNPs. In addition, *HK3, RMB38, PHLDA1,* and *HLA-DRB6* had not been previously identified as being related to HIV infection.

HIV-1 gp120 is important for HIV binding to its high-affinity cellular receptor CD4. Thrombospondin 1 (*THBS1*, also known as *TSP-1*), which has biological functions such as in the immune response, activation of MAPK activity and cell adhesion, can directly bind to gp120 in a concentration-dependent fashion. *THBS1* can inhibit HIV infection of PBMCs and transformed T and promonocytic cell lines, presumably through the interaction with the CD36-related TSP-1-binding motif [27, 28]. Consequentially, LTNPs may inhibit HIV entry into host cells and maintain the level of CD4+ T cells

through up-regulated expression of *THBS1*. The phylogenetic cluster of envelopes in LTNP-EC patients revealed ineffective binding to CD4, and this modified the signaling activity for actin/tubulin cytoskeletons to result into low fusion and infection capacity [26]. Cytoskeletal components play a major role in various steps of HIV-1 infection. The cytoskeleton contributes to viral assembly and production [29, 30]. Interestingly, *ACTB* and *ACTG1* were found to be down-regulated in LTNPs, and therefore, they may inhibit HIV-1 production, fusion, or entry into host cells. Jayappa et al. [31] found that the knockdown of Dynelin light chain LC8-type 2 (*DYNLL1*) resulted in significantly lower levels of HIV-1 reverse transcription in cells. The expression of *DYNLL1* was
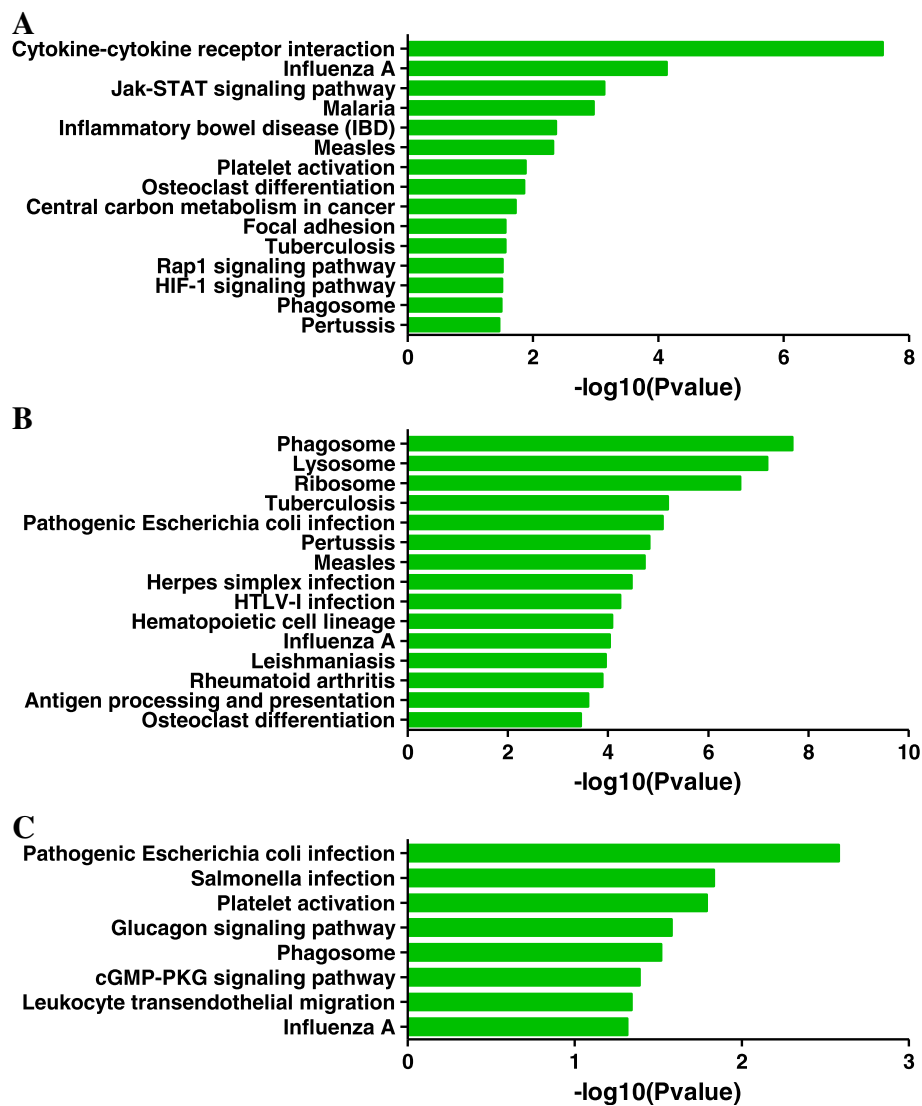
Fig. 4 KEGG pathways of DEGs in the three groups. **a** The 15 most enriched pathways in the LC group. **b** The 15 most enriched pathways in the LH group. **c** The 15 most enriched pathways in the LA group

down-regulated in the LC and LH groups. Thus, the down-regulated *DYNLL1* in LTNPs may inhibit HIV-1 replication and help maintain a low viral load.

We found that the expression of Pleckstrin homology like domain family A member 1 (*PHLDA1*), also known as T-cell Death-associated gene 51 (*TDAG51*), was significantly up-regulated in all the three groups. It has been implicated in the regulation of cell death and suppression of metastasis in cancers [33–35], and its expression is significantly reduced in human cancers [36, 37]. *PHLDA1* overexpression significantly inhibits the phosphorylation and activation of Akt (also known as protein kinase B) [36, 38]. Akt1 may act to enhance HIV replication through promoting the survival of infected cells [39]. Blocking Akt activation triggered by Nef was found

to limit HIV-1 recovery from latently-infected T cells [40], and PI3K/Akt inhibitors were reported to reduce HIV-1 production from infected primary human macrophages [41]. For this reason, the up-regulation of *PHLDA1* in all three groups may inhibit HIV-1 replication through the suppression of the Akt signaling pathway.

RNA binding motif protein (RBM38, also called RNPC1) is known to interact with its target mRNAs and regulate their expression via mRNA stability [42–44]. It is capable of binding and stabilizing the mRNA of p21 [43]. Cyclin-dependent kinase inhibitor 1A (CDKN1A; p21) has been documented for its role in antiretroviral infection [45]. The expression of p21 in human macrophages was induced after HIV-1 infection
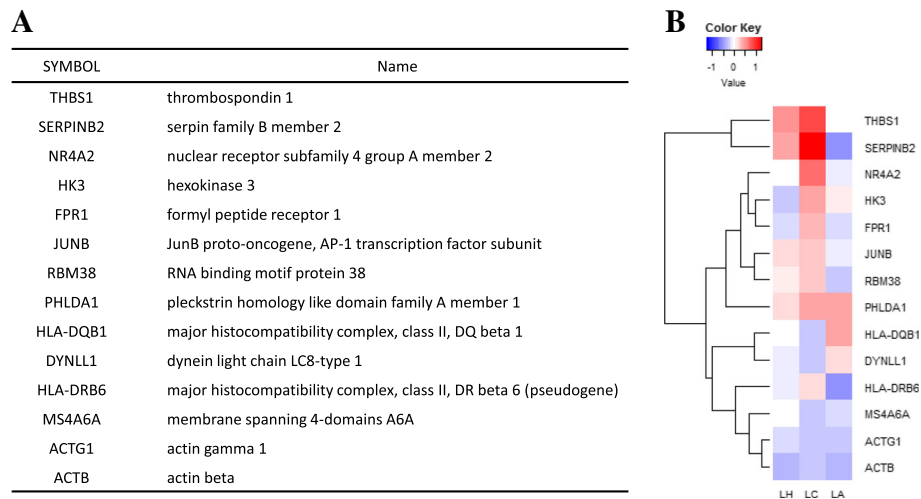
Fig. 5 The 14 common genes in the three groups excluding the fourth group. **a** The list of common DEGs. **b** The heatmap of the 14 common DEGs

[46]. In addition, up-regulation of p21 was found in CD4+ T cells from elite controllers [47]. RBM38 may regulate elite controllers through the stability of CDKN1A, a key regulator of HIV infection. Of other DEGs, membrane spanning 4-domains A6A (MS4A6A) has been reported to associate with Alzheimer's disease [48, 49], and the expression of MS4A6A could be down-regulated by the immunosuppressive domain (gp41) of HIV-1 [50]. Probably, the down-regulation of MS4A6A may reduce new infection of HIV-1 in LTNPs. However, the relationship of these genes with LTNPs should be elucidated in detail.

Our meta-analysis may have some limitations, e.g. limitation of sample size, heterogeneity of microarray platform, and consideration of multifactorial genes interactions.

Despite these limitations, our results provide the candidate biomarkers which play a role in the HIV pathogenesis in LTNPs.

Conclusions

In summary, our meta-analysis identified the 14 significant genes as the biomarkers for LTNPs infected with HIV-1. *THBS1*, *DYNLL1*, *ACTB*, and *ACTG* may control a low viral load. Especially, *PHLDA1* which is unknown to relationship with HIV infection and disease progression may inhibit HIV-1 replication through the suppression of the Akt signaling pathway and delay the disease progression of patients infected with HIV-1. In addition, RBM38 may regulate LTNPs through the stability of CDKN1A and the down-regulation of MS4A6A may
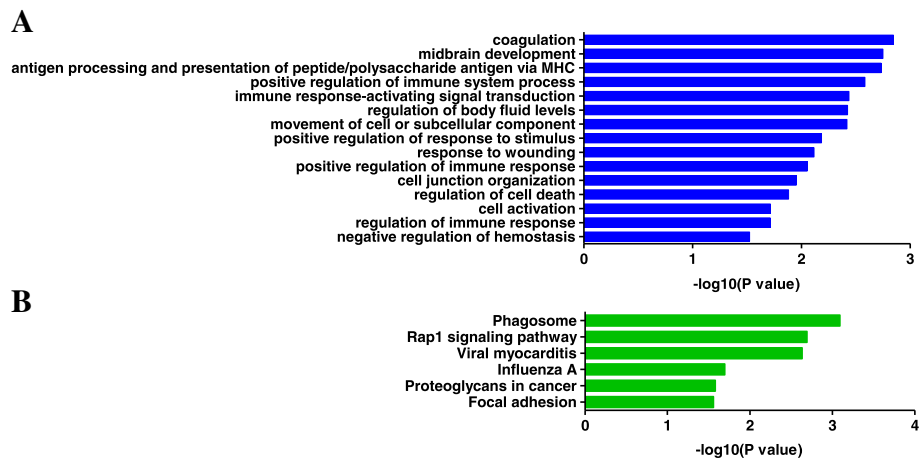
Fig. 6 Functional enrichment analysis of the 14 common DEGs. **a** GO-based enrichment analysis of the 14 common DEGs in terms of biological process. **b** KEGG pathways of the 14 common DEGs with *p*-values < 0.05

modify infection of HIV-1 in LTNPs. Further studies are necessary to elucidate how these genes effects on the mechanisms of delayed disease progression in LTNPs. These results suggest that significant genes identified in a meta-analysis provide clues to the cause of delayed disease progression and give a deeper understanding of HIV pathogenesis in LTNPs.

Additional files

Additional file 1: **Figure S1.** Venn diagram of the differentially expressed genes (DEGs) identified from each individual microarray analysis. (**A**) The DEGs in LTNPs compared with Healthy Controls; (**B**) The DEGs in LTNPs compared with patients infected HIV-1 without ART; (**C**) The DEGs in LTNPs compared with patients with HIV-1 with ART. (**D**) The DEGs in patients infected HIV-1 compared with Healthy Controls. The up-regulated or down-regulated genes had fold changes ($|\log_2 FC| >= 1$) and *adj. p*-values < 0.05. (PPTX 46 kb)

Additional file 2: **Figure S2.** The estimated pfp (percentage of false prediction) versus the number of identified genes in the LC, LH, LA, and HC groups. The identified genes are marked in red with a cutoff = 0.05. (PPTX 292 kb)

Additional file 3: **Table S1A.** All the significantly down-regulated genes in the LC group. **B**. All the significantly up-regulated genes in the LC group. **C**. All the significantly down-regulated genes in the LH group. **D**. All the significantly up-regulated genes in the LH group. **E**. All the significantly down-regulated genes in the LA group. **F**. All the significantly up-regulated genes in the LA group. **G**. All the significantly down-regulated genes in the HC group. **H**. All the significantly up-regulated genes in the HC group. (XLSX 248 kb)

Abbreviations
ART: Antiretroviral treatments; DAVID: Database for annotation, visualization and integrated discovery; DEG: Differentially expressed gene; EC: Elite controller; FC: Fold change; GEO: Gene Expression Omnibus; GO: Gene ontology; HC: Patients infected with HIV without ART versus healthy controls; HIV: Human immunodeficiency virus; KEGG: Kyoto Encyclopedia of Genes and Genomes; LA: LTNPs versus patients infected with HIV with ART; LC: LTNPs versus healthy controls; LH: LTNPs versus patients infected with HIV without ART; LTNP: Long-term non-progressor; PBMC: Peripheral blood mononuclear cell

Availability of data and materials
All data generated or analysed during this study are included in this published article.

Authors' contributions
K-CK and SYL conceived and designed the study. K-CK, SYL, YKP and C-HY analyzed the data. SYL and YKP interpreted the data and wrote the paper. K-CK and KK revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Hyrcza MD, Kovacs C, Loutfy M, Halpenny R, Heisler L, Yang S, Wilkins O, Ostrowski M, Der SD. Distinct transcriptional profiles in ex vivo CD4+ and CD8+ T cells are established early in human immunodeficiency virus type 1 infection and are characterized by a chronic interferon response as well as extensive transcriptional changes in CD8+ T cells. J Virol. 2007;81(7):3477–86.
2. Vigneault F, Woods M, Buzon MJ, Li C, Pereyra F, Crosby SD, Rychert J, Church G, Martinez-Picado J, Rosenberg ES, et al. Transcriptional profiling of CD4 T cells identifies distinct subgroups of HIV-1 elite controllers. J Virol. 2011;85(6):3015–9.
3. 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(2):e55791.
4. Salgado M, López-Romero P, Callejas S, López M, Labarga P, Dopazo A, Soriano V, Rodés B. Characterization of host genetic expression patterns in HIV-infected individuals with divergent disease progression. Virology. 2011; 411(1):103–12.
5. Wu JQ, Sasse TR, Wolkenstein G, Conceicao V, Saksena MM, Soedjono M, Perera SS, Wang B, Dwyer DE, Saksena NK. Transcriptome analysis of primary monocytes shows global down-regulation of genetic networks in HIV viremic patients versus long-term non-progressors. Virology. 2013;435(2): 308–19.
6. 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(1):191.
7. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS Med. 2008;5(9):e184.
8. Tseng GC, Ghosh D, Feingold E. Comprehensive literature review and statistical considerations for microarray meta-analysis. Nucleic Acids Res. 2012;40(9):3785–99.
9. Hong F, Breitling R. A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. Bioinformatics. 2008;24(3):374–82.
10. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002; 30(1):207–10.
11. Del Carratore F, Jankevics A, Eisinga R, Heskes T, Hong F, Breitling R. RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets. Bioinformatics. 2017; 33(17):2774–5.
12. Team RC. R: A Language and Environment for statistical computing. In: R Foundation for Statistical Computing; 2018. http://www.r-project.org/.
13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
14. Jeffery IB, Higgins DG, Culhane AC. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinformatics. 2006;7(1):359.
15. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA. The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 2007;8(9):R183.
16. Ashburner M. Gene ontology: tool for the unification of. Biology. 2000;25(1):25–9.
17. Poropatich K, Sullivan DJ, Jr. Human immunodeficiency virus type 1 long-term non-progressors: the viral, genetic and immunological basis for disease non-progression. J Gen Virol. 2011;92(Pt 2):247-68
18. Fellay J, Ge D, Shianna KV, Colombo S, Ledergerber B, Cirulli ET, Urban TJ, Zhang K, Gumbs CE, Smith JP et al. Common genetic variation and the control of HIV-1 in humans. PLoS Genet. 2009;5(12):e1000791.
19. Limou S, Zagury JF. Immunogenetics: Genome-wide association of non-progressive HIV and viral load control: HLA genes and beyond. Front Immunol. 2013;4:118-118.

20. Limou S, Coulonges C, Herbeck JT, van Manen D, An P, Le Clerc S, Delaneau O, Diop G, Taing L, Montes M et al. Multiple-cohort genetic association study reveals CXCR6 as a new chemokine receptor involved in long-term nonprogression to AIDS. J Infect Dis. 2010;202(6):908-15.

21. Tsiara CG, Nikolopoulos GK, Dimou NL, Pantavou K, Bagos P, Mensah B, Talias M, Braliou G, Paraskeva D, Bonovas S et al. Interleukin gene polymorphisms and susceptibility to HIV-1 infection: a meta-analysis. 2018;97.

22. Williams B, Landay A, Presti RM. Microbiome alterations in HIV infection a review. Cell Microbiol. 2016;18(5):645-51.

23. Brenchley JM, Price DA, Schacker TW, Asher TE, Silvestri G, Rao S, Kazzaz Z, Bornstein E, Lambotte O, Altmann D et al. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. Nat Med. 2006;12:1365.

24. Nowak P, Troseid M, Avershina E, Barqasho B, Neogi U, Holm K, Hov JR, Noyan K, Vesterbacka J, Svard J et al. Gut microbiota diversity predicts immune status in HIV-1 infection. Aids. 2015;29(18):2409-18.

25. Vujkovic-Cvijin I, Dunham RM, Iwai S, Maher MC, Albright RG, Broadhurst MJ, Hernandez RD, Lederman MM, Huang Y, Somsouk M et al. Dysbiosis of the gut microbiota is associated with HIV disease progression and tryptophan catabolism. Sci Transl Med. 2013;5(193):193ra191-193ra191.

26. Casado C, Marrero-Hernandez S, Marquez-Arce D, Pernas M, Marfil S, Borras-Granana F, Olivares I, Cabrera-Rodriguez R, Valera MS, de Armas-Rillo L, et al. Viral characteristics associated with the clinical Nonprogressor phenotype are inherited by viruses from a cluster of HIV-1 elite controllers. mBio. 2018; 9(2):e02338.

27. Chen H, Herndon ME, Lawler J. The cell biology of thrombospondin-1. Matrix Biol. 2000;19(7):597–614.

28. Lopez-Dee Z, Pidcock K, Gutierrez LS. Thrombospondin-1: multiple paths to inflammation. Mediat Inflamm. 2011;2011:296069.

29. Matarrese P, Malorni W. Human immunodeficiency virus (HIV)-1 proteins and cytoskeleton: partners in viral life and host cell death. Cell Death Differ. 2005;12:932.

30. Fontenot DR, den Hollander P, Vela EM, Newman R, Sastry JK, Kumar R. Dynein light chain 1 peptide inhibits human immunodeficiency virus infection in eukaryotic cells. Biochem Biophys Res Commun. 2007; 363(4):901–7.

31. Jayappa KD, Ao Z, Wang X, Mouland AJ, Shekhar S, Yang X, Yao X. Human immunodeficiency virus type 1 employs the cellular dynein light chain 1 protein for reverse transcription through interaction with its integrase protein. J Virol. 2015;89(7):3497–511.

32. Liu M, Zhao J, Chen K, Bian X, Wang C, Shi Y, Wang JM. G protein-coupled receptor FPR1 as a pharmacologic target in inflammation and human glioblastoma. Int Immunopharmacol. 2012;14(3):283–8.

33. Park CG, Lee SY, Kandala G, Lee SY, Choi Y. A novel gene product that couples TCR signaling to Fas(CD95) expression in activation-induced cell death. Immunity. 1996;4(6):583–91.

34. Ren L, Mendoza A, Zhu J, Briggs JW, Halsey C, Hong ES, Burkett SS, Morrow JJ, Lizardo MM, Osborne T, et al. Characterization of the metastatic phenotype of a panel of established osteosarcoma cells. Oncotarget. 2015; 6(30):29469–81.

35. Neef R, Kuske MA, Prols E, Johnson JP. Identification of the human PHLDA1/ TDAG51 gene: down-regulation in metastatic melanoma contributes to apoptosis resistance and growth deregulation. Cancer Res. 2002;62(20): 5920–9.

36. Li G, Wang X, Hibshoosh H, Jin C, Halmos B. Modulation of ErbB2 blockade in ErbB2-positive cancers: the role of ErbB2 mutations and PHLDA1. PLoS One. 2014;9(9):e106349.

37. Nagai MA. Pleckstrin homology-like domain, family A, member 1 (PHLDA1) and cancer. Biomedical Reports. 2016;4(3):275–81.

38. Magi S, Iwamoto K, Yumoto N, Hiroshima M, Nagashima T, Ohki R, Garcia-Munoz A, Volinsky N, Von Kriegsheim A, Sako Y, et al. Transcriptionally inducible Pleckstrin homology-like domain, family A, member 1, attenuates ErbB receptor activity by inhibiting receptor oligomerization. J Biol Chem. 2018;293(6):2206–18.

39. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, Stec E, Ferrer M, Strulovici B, Hazuda DJ, et al. Genome-scale RNAi screen for host factors required for HIV replication. Cell Host Microbe. 2008;4(5):495–504.

40. Kumar A, Abbas W, Colin L, Khan KA, Bouchat S, Varin A, Larbi A, Gatot J-S, Kabeya K, Vanhulle C, et al. Tuning of AKT-pathway by Nef and its blockade by protease inhibitors results in limited recovery in latently HIV infected T-cell line. Sci Rep. 2016;6:24090.

41. Chugh P, Bradel-Tretheway B, Monteiro-Filho CM, Planelles V, Maggirwar SB, Dewhurst S, Kim B. Akt inhibitors as an HIV-1 infected macrophage-specific anti-viral therapy. Retrovirology. 2008;5:11.

42. Cheng G, Ji C, Yang N, Meng L, Ding Y, Wei J. RNA-binding protein RBM38: acting as a tumor suppressor in colorectal cancer, vol. 9; 2016.

43. Shu L, Yan W, Chen X. RNPC1, an RNA-binding protein and a target of the p53 family, is required for maintaining the stability of the basal and stress-induced p21 transcript. Genes Dev. 2006;20(21):2961-72.

44. Zhang J, Cho SJ, Shu L, Yan W, Guerrero T, Kent M, Skorupski K, Chen H, Chen X. Translational repression of p53 by RNPC1, a p53 target overexpressed in lymphomas. Genes Dev. 2011;25(14):1528-43.

45. de Pablo A, Bogoi R, Bejarano I, Toro C, Valencia E, Moreno V, Martin-Carbonero L, Gomez-Hernando C, Rodes B. Short communication: p21/ CDKN1A expression shows broad interindividual diversity in a subset of HIV-1 elite controllers. AIDS Res Hum Retroviruses. 2016;32(3):232-36.

46. Vázquez N, Greenwell-Wild T, Marinos NJ, Swaim WD, Nares S, Ott DE, Schubert U, Henklein P, Orenstein JM, Sporn MB et al. Human immunodeficiency virus type 1-induced macrophage gene expression includes the p21 gene, a target for viral regulation. J Virol. 2005;79(7):4479-91.

47. Chen H, Li C, Huang J, Cung T, Seiss K, Beamon J, Carrington MF, Porter LC, Burke PS, Yang Y et al. CD4+ T cells from elite controllers resist HIV-1 infection by selective upregulation of p21. J Clin Invest. 2011;121(4):1549-60.

48. Proitsi P, Lee SH, Lunnon K, Keohane A, Powell J, Troakes C, Al-Sarraj S, Furney S, Soininen H, Kloszewska I et al. Alzheimer's disease susceptibility variants in the MS4A6A gene are associated with altered levels of MS4A6A expression in blood. Neurobiology of aging. 2014;35(2):279-90.

49. Ma J, Zhang W, Tan L, Wang H-F, Wan Y, Sun F-R, Tan C-C, Yu J-T, Tan L, Alzheimer's Disease Neuroimaging I. MS4A6A genotypes are associated with the atrophy rates of Alzheimer's disease related brain structures. Oncotarget 2016, 7(37):58779-88.

50. Denner J, Eschricht M, Lauck M, Semaan M, Schlaermann P, Ryu H, Akyuz L. Modulation of cytokine release and gene expression by the immunosuppressive domain of gp41 of HIV-1. PLoS One. 2013;8(1):e55199.

51. Rotger M, Dang KK, Fellay J, Heinzen EL, Feng S, Descombes P, Shianna KV, Ge D, Gunthard HF, Goldstein DB, et al. Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. PLoS Pathog. 2010;6(2):e1000781.

52. Rotger M, Dalmau J, Rauch A, McLaren P, Bosinger SE, Martinez R, Sandler NG, Roque A, Liebner J, Battegay M, et al. Comparative transcriptomics of extreme phenotypes of human HIV-1 infection and SIV infection in sooty mangabey and rhesus macaque. J Clin Invest. 2011;121(6):2391–400.

53. Buzon MJ, Martin-Gayo E, Pereyra F, Ouyang Z, Sun H, Li JZ, Piovosa M, Shaw A, Dalmau J, Zangger N, et al. Long-term antiretroviral treatment initiated at primary HIV-1 infection affects the size, composition, and decay kinetics of the reservoir of HIV-1-infected CD4 T cells. J Virol. 2014;88(17): 10056–65.

54. Sedaghat AR, German J, Teslovich TM, Cofrancesco J Jr, Jie CC, Talbot CC Jr, Siliciano RF. Chronic CD4+ T-cell activation and depletion in human immunodeficiency virus type 1 infection: type I interferon-mediated disruption of T-cell dynamics. J Virol. 2008;82(4):1870–83.