RESEARCH

CrossMark

An integrative genomic analysis of transcriptional profiles identifies characteristic genes and patterns in HIV-infected long-term non-progressors and elite controllers

Jiwei Ding[1†], Ling Ma[1†], Jianyuan Zhao[1], Yongli Xie[1], Jinming Zhou[2], Xiaoyu Li[1*] and Shan Cen[1*]

Abstract

Background: Despite that most HIV-infected individuals experience progressive CD4+T cell loss and develop AIDS, a minority of HIV-infected individuals remain asymptomatic and maintain high level CD4+T cell counts several years after seroconversion. Efforts have been made to understand the determinants of the nonprogressive status, exemplified by the clinical course of elite controllers (ECs) who maintain an undetectable viremia and viremic nonprogressors (VNPs) who have a normal CD4+ count in spite of circulating viral load. However, the intrinsic mechanism underlying nonprogression remained elusive. In this study, we performed an integrative analysis of transcriptional profiles to pinpoint the underlying mechanism for a naturally occurring viral control.

Methods: Three microarray datasets, reporting mRNA expression of the LTNPs or ECs in HIV-infected patients, were retrieved from Gene Expression Ominbus (GEO) or Arrayexpress databases. These datasets, profiled on the same type of microarray chip, were selected and merged by a bioinformatic approach to build a meta-analysis derived transcriptome (MADNT). In addition, we investigated the different transcriptional pathways and potential biomarkers in CD4+ and CD8+ cells in ECs and whole blood in VNPs compared to HIV progressors. The combined transcriptome and each subgroup was subject to gene set enrichment analysis and weighted co-expression network analysis to search potential transcription patterns related to the non-progressive status.

Results: 30 up-regulated genes and 83 down-regulated genes were identified in lymphocytes from integrative meta-analysis of expression data. The interferon response and innate immune activation was reduced in both CD4+ and CD8+T cells from ECs. Several characteristic genes including CMPK1, CBX7, EIF3L, EIF4A and ZNF395 were indicated to be highly correlated with viremic control. Besides that, we indicated that the reduction of ribosome components and blockade of translation facilitated AIDS disease progression. Most interestingly, among VNPs who have a relatively high viral load, we detected a two gene-interaction networks which showed a strong correlation to immune control even with a rigorous statistical threshold (p value = 2−e4 and p value = 0.004, respectively) by WGCNA.

*Correspondence: lixiaoyu@imb.pumc.edu.cn;
shancen@imb.pumc.edu.cn
†Jiwei Ding and Ling Ma contributed equally to this work
[1] Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences, Beijing 100050, China
Full list of author information is available at the end of the article

Ding *et al. J Transl Med* (2019) 17:35

Page 2 of 13

Conclusions: We have identified differentially expressed genes and transcriptional patterns in ECs and VNPs compared to normal chronic HIV-infected individuals. Our study provides new insights into the pathogenesis of HIV and AIDS and clues for the therapeutic strategies for anti-retroviral administration.

Keywords: Elite controller, Long-term nonprogressor, Viremic nonprogressor, Meta-analysis, GSEA, WGCNA

Background

Despite the great progress in combination anti-retroviral therapy (cART), AIDS is still a non-curative disease that causes a high death rate every year. It is interesting that there are a small proportion of HIV-infected individuals who maintain a stable CD4+ T cell count within the normal reference range over several years from the diagnosis of HIV infection in the absence of ART and clinical symptoms. These patients with spontaneous and sustained control of HIV disease progression were identified as long term nonprogressors (LTNPs) [1, 2]. Similarly, there are approximately 1% of HIV-infected patients who maintain undetectable viral load for a prolonged period (generally less than 50 copies/ml), who are identified as elite controllers. These two groups of patients classified by the immunologic parameters or viral parameters represent useful models of natural protection against disease progression and may have important implications for prophylactic and therapeutic strategies [3]. Thus far it is still in debate whether viral factor, host or environmental factor contributes to the LTNP status [2, 4]. Association of augmented control with attenuated viruses due to deletions or mutations of regulatory proteins or other special viral polymorphism have been reported by several researches [5–8]. For instance, a F72L mutation in HIV-1 Vpr in a LTNP was reported to confer a significant reduction in Vpr nuclear import and virion incorporation implying a link between efficient Vpr nuclear import and HIV disease progression [9]. Moreover, viruses isolated from some LTNPs were found to have gross deletions in nef genes, suggesting the importance for nef gene in AIDS disease progression [10, 11]. However, studies of nef genes in elite controller yield contradicting results [12, 13]. A phylogenetic analysis revealed that nef sequences from patients with different rates of progression did not form distinct cluster between LTNP and progressors, suggesting the degree of variation in nef is unlikely to be indicative of disease progression [14, 15].

Cellular immune responses have also been reported to contribute to viral control [16, 17]. For instance, a tight association was observed between Gag specific CD8+ T cells and viral control [17–20]. Some elite suppression was linked with higher level of cytolytic granules within HIV-specific CD8+ T cells [21]. Additionally, CD8+ T cells isolated from ECs exhibited more polyfunctional capability in response to HIV specific antigens [22–25].

On the other hand, CD4+ T cells from ECs retained an ability to proliferate and produce interleukin 2 (IL-2) in response to HIV compared with normal patients under ART [26]. In addition to CD4+ and CD8+ T lymphocytes, a correlation of Interleukin 17 secreting T (Th17) cell level and HIV disease progression was observed in LTNPs compared to TPs, supporting a role of this cell subset in HIV pathogenesis [27]. Besides, humoral immune response also plays a role in the context of natural viral control. Some researchers have reported that ECs maintained HIV-1 specific memory B cell response which contributed to neutralizing responses in contrast to treated patients [28].

Moreover, various approaches have been undertaken to uncover the host genetic factors or specific genes involved in virologic control of HIV infection in ECs [29, 30]. A study showed a role of p21 in ECs via indirectly blocking reverse transcription by inhibiting CDK2-dependent phosphorylation [31]. The HLA class I allele HLA-B*57 [32, 33] and HLA-B*27 alleles [34, 35] were overrepresented among elite controllers and viremic controllers (VCs) compared with normal progressors, which underscored the important role of CD8+ T cells in naturally viral control. [25]. Most notably, HLA class I molecules might affect HIV-1 immune control by interacting with their receptors on innate immune cells, such as the killer cell immunoglobulin-like receptors (KIR) on natural killer (NK) cells. Two studies reported that activating KIR3DS1 allele in combination with Bw4-80I and KIR3DL1*004 in the presence of HLA-Bw4 showed a pronounced protection against AIDS progression [36, 37]. Genome-wide association studies have been carried out to decipher the association between naturally occurring single nucleotide polymorphism (SNPs) and viremic control in ECs and LTNPs [38, 39]. An international HIV controllers study identified over 300 SNPs on chromosome 6 involved in viral control [39]. However, only approx. 20% of the protective effects can be attributable to the SNPs discovered, indicating other unknown mechanisms were accountable for the observed control. Further studies are necessary to pinpoint more novel pathways and intrinsic host factors responsible for virological control.

In most studies investigating the host intrinsic factors controlling disease progression, only a few ECs or LTNPs were involved and the conclusions are often

Ding *et al. J Transl Med*   (2019) 17:35

Page 3 of 13

controversial. To overcome this limitation of individual studies, a large-scale transcriptional study was necessary to reduce random error and increase statistical power. In this study, we combined relevant microarray data to increase statistical power to uncover the biological differences between LTNP or ECs and chronic progressors. To decrease the heterogeneity and increase the consistency between different datasets, we only select microarray data from Human Genome U133A or Human Genome U133 plus 2 Array (Affemetrix Company). Furthermore, a new illumina data series GSE87620 was used as a validating set. To be most informative, we conducted our analysis in two steps. First, we combined the three datasets into meta-analysis derived transcriptome, which provided a comprehensive comparison of nonprogressors (NP) with progressors (PP). Second, we split these study subjects into three subgroups, namely CD4+T cell samples from ECs, CD8+T cell samples from ECs and whole blood samples from VNPs, thereby providing an elaborate comparison between aviremic controllers or viremic controllers and progressors. Differential expression analysis, gene set enrichment analysis and WGCNA approach were carried out in each step. Our study have revealed some characteristic biomarkers and transcriptional patterns and highlight several key genes in nonpathogenic individuals, Most strikingly, we identified key transcriptional modules in VNPs which have never been reported before. These findings may better the understanding of HIV-1 viremic and immune control and AIDS progression.

Methods

Selection of studies and datasets

Expression profiling studies including LTNPs or ECs were identified through Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) using search term ("long term no progressor" OR "elite controller"). To ensure the relevant studies were not missed, search in Arrayexpress (http://www.ebi.ac.uk/arrayexpress) was also performed. Nine microarray gene expression datasets, reporting expression data of LTNPs or ECs and normal patients were retrieved from public repositories. Three datasets (GSE24081, GSE6740 and GSE57730) profiling on the same version of microarray chips [Human Genome U133 Plus 2.0 Array (HG-U133_Plus_2) or Human Genome U133A ver2.0 (U133A)], met the inclusion criteria and were included in the integrated analysis to build a MADNT set. The characteristics of these datasets were listed in Table 1. Detailed clinical parameters of each patient involved were listed in Additional file 1: Table S1.

Data processing

Microarray meta-analysis were carried out according to the guidelines described in [40]. Each datasets were log2 transformed and normalized by Agilent Gene-Spring software (Version 11.5, Agilent, USA). Then, gene matching was done for all probes. When multiple probes matched the same gene symbol, the probe presented the greatest inner-quartile range (IQR) was selected to represent the target gene symbol. After matching all the probes to a common gene symbol, "MetaDE" R package [41] was exploited to merge the common gene symbols across multiple studies by p value combination using Fisher methods. Differentially expressed genes were selected with adjusted p value < 0.05, based on false discovery rate (FDR) by the Benjamini–Hochberg procedure and moderated t test.

Enrichment analysis

Enrichment analysis for KEGG pathway and Gene Ontology terms were carried out by David online tool (https://david.ncifcrf.gov). Gene set enrichment analysis (GSEA) [42] was carried out using GSEA version 3.0, downloaded from the Broad Institute (http://www.broadinstitute.org/gsea/downloads.jsp). Expression data sets and phenotype labels were created according to GSEA specifications. Gene set permutations were set to be done 1000 times for each analysis using the weighted enrichment statistic

Table 1 General information of each dataset

First author	GEO number	Platform	Sample source	Sample size (PP/NP)
Martin Hyrcza	GSE6740	GPL96; Affymetrix Human Genome U133A Array	CD4+T cell CD8+T cell	PP = 10; NP = 10 *3 NPs are ECs and 2 with extremely low viral load
W. Nicholas Haining	GSE24081	GPL3921; Affymetrix HT Human Genome U133A Array	CD8+T cell	PP = 18; NP = 24 *All NPs are ECs
Gregory K Tharp	GSE57730	GPL570; Affymetrix Human Genome U133 Plus 2.0 Array	Whole blood	PP = 7; NP = 5 *All NPs are VNPs

PP progressor, *NP* nonprogressor, *EC* elite controllers, *VNP* viremic nonprogressor

Ding *et al. J Transl Med*   (2019) 17:35

Page 4 of 13

and signal to noise metric. Gene sets with FDR lower than 0.05 were considered significant.

WGNCA

Weighted gene coexpression network analysis (WGCNA) is a gene coexpression network-based approach [43, 44]. A gene co-expression network is defined as undirected, weighted gene network, in which the nodes represent expression profiles while edges represent pairwise correlation between gene expressions. Briefly, correlation coefficient Smn between characteristic gene m and gene n is calculated by their expression values between different samples using the formulation: $Smn = |cor(mn)$. The correlation matrix was then transformed into an undirected network by raising the absolute value of each entry to a power β using 6 as correlation coefficient threshold. Genes were clustered into different modules using dynamic tree cutting method.

Protein–protein interactions (PPI) networks in each module were visualized by Cytoscape 3.6.0. The Network Analyzer examined the network for topological parameters, including degree, connectivity, betweenness and closeness.

HIV infection assay and western blotting

To analyze the antiviral activity of several up-regulated genes, HIV infection assay was performed as previously described [45, 46]. Briefly, 200 ng pNL4-3-R-E-Luc vector and 150 ng pCMV-VSV-G packaging vector were transfected into the 293T cells together with 200 ng empty pCMV6-Entry vector, control gene (SLFN11, GADD45G) or pCMV6-Entry-CMPK1, ZNF395, METTL9, GADD45A, PKKAR2B, OAT, CHPT1, SPOCK2 and CBX7 in 12-well plate using Lipofectamine 2000 (Thermo Scientific) according to the manufacturer's instructions. The old culture medium was replaced with fresh culture medium at 6 h post-transfection. Subsequently, the supernatants were collected at 48 h post-transfection and cleared by filtration with the 0.45 μm filter. Cell lysis were harvested for SDS-PAGE and western blotting with α-myc, p24 (Gag) and β-actin antibody, respectively. The amount of (infectious) virus particles in the supernatants was determined via one-cycle infection assay. For infection assays, 200 μl of 1×10^5 SupT1 or Jurkat cells per well in 96 well plate were infected with supernatants collected from 293T cells. At 48 h post-infection, the SupT1 cells were lysed and luciferase activity was determined using a firefly Luciferase Assay System (Promega).

293T cells were maintained at 37 °C in high glucose Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fatal bovine serum (Gibco). SupT1 and Jurkat cells were maintained at RPMI 1640 Medium with 10% fetal bovine serum.
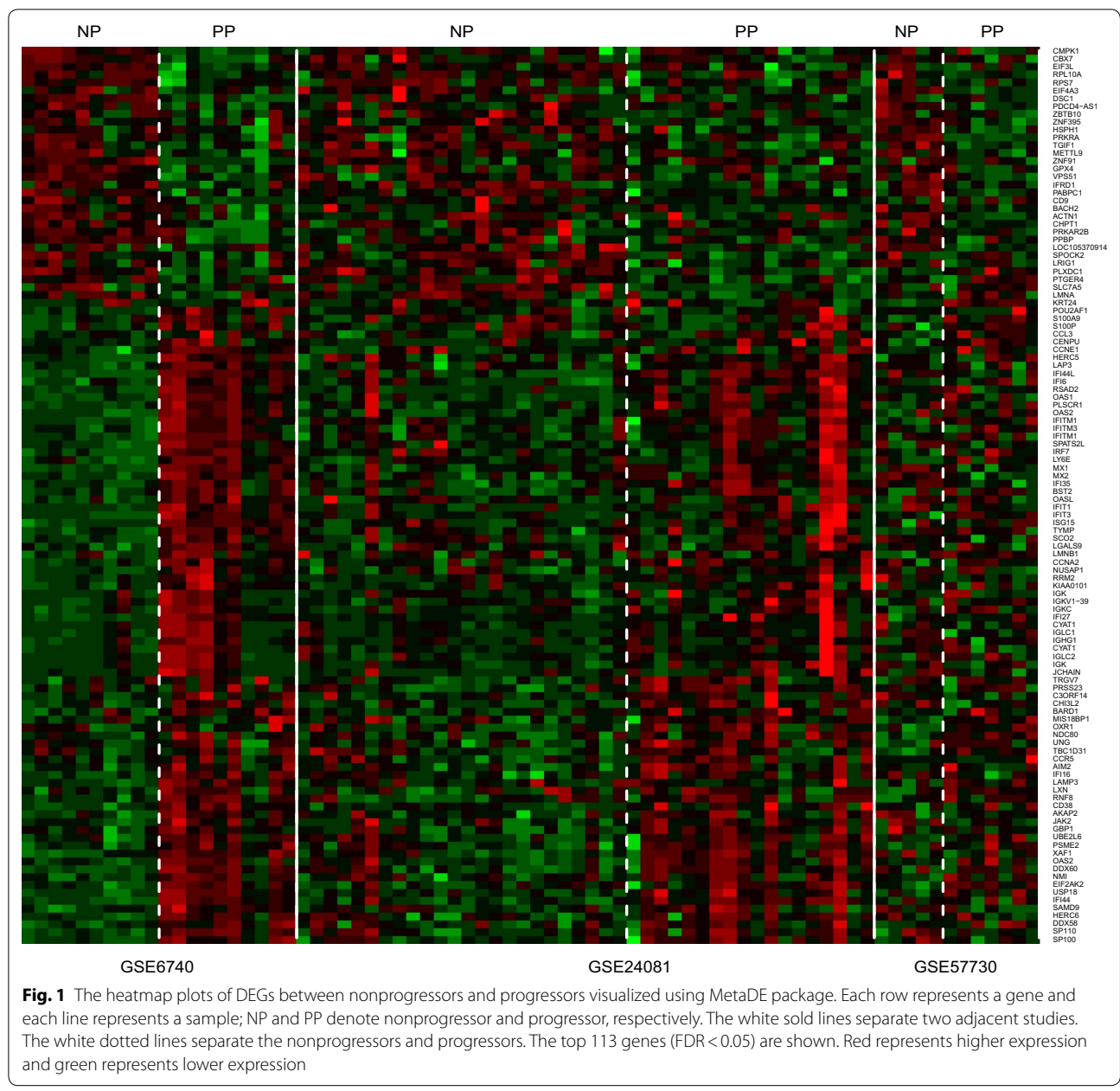
Results

Description of datasets

We searched two public repositories for mRNA transcriptional profile related to viremic control using the search term" "elite controller" [OR] "long-term non-progressors". The inclusion criteria were set as (1) the microarray chips used in the study was from Human Genome U133A or Human Genome U133 plus 2 Array (Affemetrix Company); (2) sample source was CD4+, CD8+ T cells or whole blood. Three studies met the inclusion-criteria were selected for the meta-analysis study. Details about these datasets were outlined as in Table 1. The three series included 35 PPs, 34 ECs (except for two patients with extremely low viremia) and 5 VNPs. The following information was extracted from each data series including GEO number, author name, sample source, platform and the number of progressors and controllers.
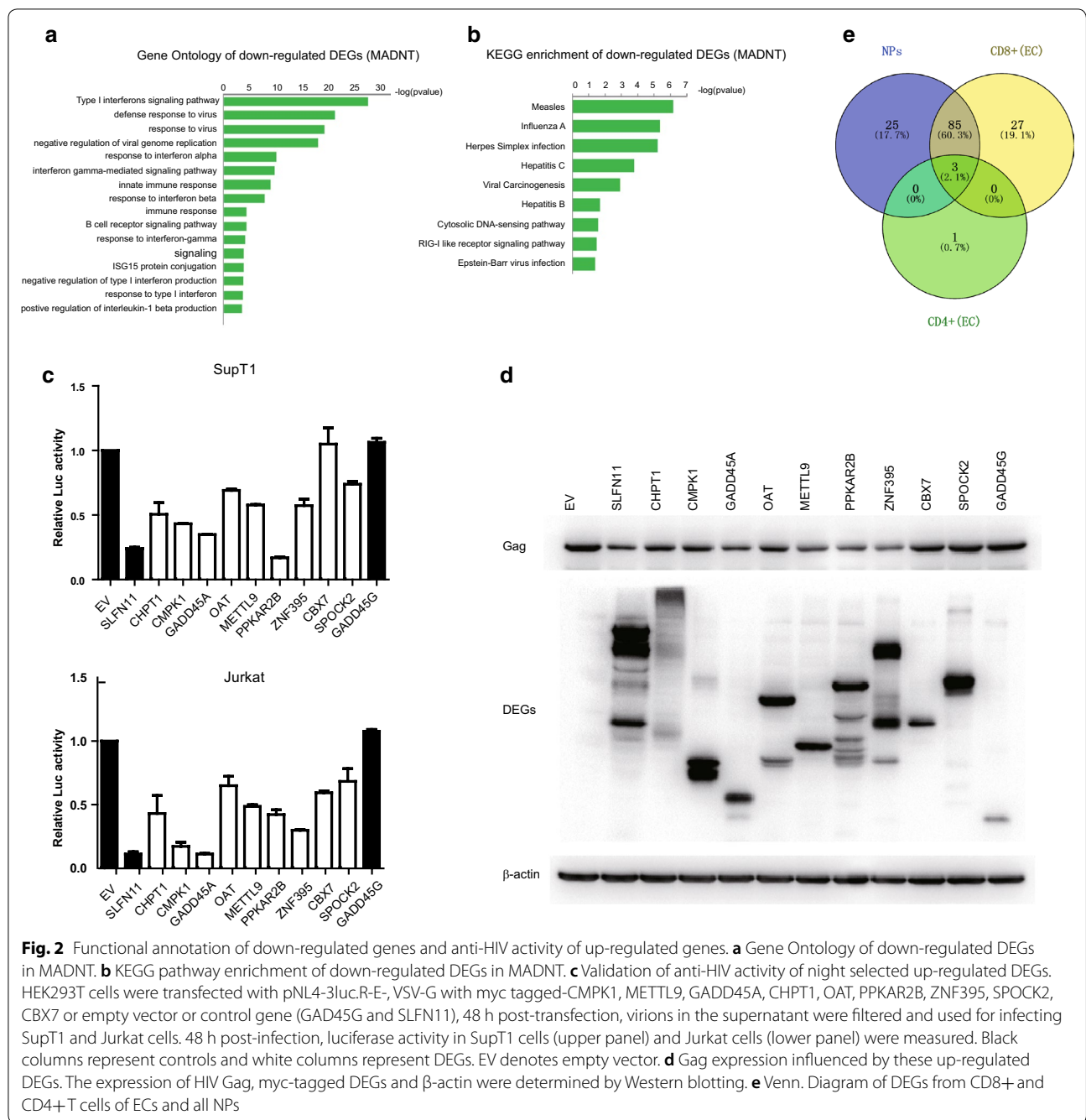
Microarray meta-analysis

To compile expression data for meta-analysis, Cel raw data was firstly preprocessed by GeneSpring 11.5 software (Agilent Technologies, Santa Clara, CA). The background subtraction, normalization, and log base 2 transformation of gene signals were carried out using Robust Multi-Array Analysis (RMA) summarization algorithm. Quality of each dataset was assessed in Genespring and expression data was retrieved from GeneSpring. MetaDE packages were exploited to combine the three datasets and differentially expressed genes (DEG) in NPs compared to normal chronic progressors. DEG p value in each study was calculated by moderated-t statistical analysis. Fisher combined probability test was chosen for meta-analysis statistical method. For statistics, Benjamini and Hochberg method was used for multiple testing correction. Genes with FDR less than 0.05 was accepted as DEGs. We found 30 upregulated DEG and 83 down-regulated DEGs in NPs compared with PPs (Fig. 1). Up-regulated genes included CMPK1, CD9, METTL9, EIF4A3, ZNF395 and so on, while down-regulated genes were overrepresented by CD38, LAG3, some Interferon-stimulated genes (ISGs) and inflammatory genes. Upregulated and down-regulated DEGs in the comparison of NPs and PPs were shown in Additional file 2: Table S2.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed for the functional evaluation of both up-regulated and down-regulated DEGs using DAVID (Database for Annotation, Visualization and Integrated Discovery) online tool. In regard to GO analysis, "type 1 interferon signaling pathway", "defense response to virus", "interferon-gamma mediated-signaling pathway", "innate immune

Ding *et al. J Transl Med*   (2019) 17:35

Page 5 of 13

Fig. 1 The heatmap plots of DEGs between nonprogressors and progressors visualized using MetaDE package. Each row represents a gene and each line represents a sample; NP and PP denote nonprogressor and progressor, respectively. The white sold lines separate two adjacent studies. The white dotted lines separate the nonprogressors and progressors. The top 113 genes (FDR < 0.05) are shown. Red represents higher expression and green represents lower expression

response" and "B cell signaling pathway" were enriched in down-regulated genes (Fig. 2a). In the pathway analysis, several anti-viral response pathways was enriched, including "Measles", "Influenza A", "Hepatitis B" and "Hepatitis C" (Fig. 2b). The analysis failed to enrich GO terms or KEGG pathways with FDR less than 0.05 in up-regulated genes. Unlike down-regulated genes, the biological meanings with up-regulated genes were sporadic and their relevance to AIDS disease progression was not immediately clear. Furthermore, night genes (CMPK1, METTL9, CHPT1, OAT, GADD45A, SPOCK2, CBX7, ZNF395 and PPKAR2B) were selected to validate

the anti-HIV activity. 293T cells were transfected with NL4-3lucR-E- and VSV-G along with these genes, empty vector or control genes (GADD45G, SLFN11). 48 h post-transfection, the supernatant was harvested and used for infecting SupT1 and Jurkat cells. After another 48 h, luciferase activity in SupT1 or Jurkat cells was measured to assess the anti-HIV activity. CMPK1, GADD45A, PPKAR2B and CHPT1 suppressed HIV-1 production by more than twofold. METTL9, ZNF395 SPOCK2 and OAT exerted a moderate inhibitory effect on HIV-1 production. Notably, GADD45A inhibited HIV-1 production by more than eightfold in Jurkat cells (Fig. 2c).

Ding *et al. J Transl Med* (2019) 17:35

Page 6 of 13

Fig. 2 Functional annotation of down-regulated genes and anti-HIV activity of up-regulated genes. **a** Gene Ontology of down-regulated DEGs in MADNT. **b** KEGG pathway enrichment of down-regulated DEGs in MADNT. **c** Validation of anti-HIV activity of night selected up-regulated DEGs. HEK293T cells were transfected with pNL4-3luc.R-E-, VSV-G with myc tagged-CMPK1, METTL9, GADD45A, CHPT1, OAT, PPKAR2B, ZNF395, SPOCK2, CBX7 or empty vector or control gene (GAD45G and SLFN11), 48 h post-transfection, virions in the supernatant were filtered and used for infecting SupT1 and Jurkat cells. 48 h post-infection, luciferase activity in SupT1 cells (upper panel) and Jurkat cells (lower panel) were measured. Black columns represent controls and white columns represent DEGs. EV denotes empty vector. **d** Gag expression influenced by these up-regulated DEGs. The expression of HIV Gag, myc-tagged DEGs and β-actin were determined by Western blotting. **e** Venn. Diagram of DEGs from CD8+ and CD4+ T cells of ECs and all NPs

GADD45G, the expression of which varied little between NPs and PPs was used here as a negative control gene and showed no effect on HIV replication. SLFN11 which was reported to inhibit HIV-1 production in the previous studies was used here as a positive control [47]. To further investigate whether these up-regulated DEGs influences the expression of HIV gene, Gag expression in 293T cells was assessed by Western blotting. CHPT1, CMPK1, GADD45A, OAT, METTL9, PPKAR2B and

ZNF395 significantly decreased the expression of HIV-1 Gag (Fig. 2d). The functional assay confirmed the reliability of our analysis results.

A pilot study reported that serum from LTNPs and ECs have distinct neutralizing capabilities and antibody-dependent cellular cytotoxicity (ADCC) activity [48]. To uncover the divergent transcription pattern between aviremic controllers and VNPs from different cell types, study subjects were grouped into three subsets

Ding *et al. J Transl Med*   (2019) 17:35

Page 7 of 13

aforementioned. Data of CD8+ T cells from two datasets were merged into a combined transcriptome by MetaDE package. 115 DEGs were identified in CD8+ T cells (Additional file 3: Table S3), which highly overlap with those identified in MADNT. Only 4 DEGs (OAS3, IFI44, IFI44L and EIF2AK2) were identified in CD4+ T cells of ECs (Fig. 2e), and no DEGs met the criteria (FDR < 0.05) in VNPs.

Taken together, statistical combination of the relevant datasets has created a meta-analysis derived nonprogressor transcriptome (MADNT) and different subgroups, which can be subject to downstream GSEA and WGCNA analysis.

Gene set enrichment analysis

Given the limited power to detect the transcriptional pattern in up-regulated genes in LTNPs and ECs, GSEA was carried out to investigate the intrinsic common features associated with virological or immunologic control. Rather than setting a cutoff value for single DE genes, GSEA evaluates the whole dataset at biological pathway level by performing unbiased global search for predefined gene sets. MADNT and different subgroups mentioned above were interrogated for pathway enrichment using canonical pathway from MSigDb 2 GO-BP collection and also C2 (KEGG and Reactome) collection. Fourteen pathways were enriched with FDR less than 0.05 in NPs which included "KEGG_RIBOSOME", "3′UTR mediated translational regulation", "Nuclear transcribed mRNA catabolic process nonsense-mediated decay", "SRP-dependent cotranslational protein targeting to membrane" and "translational initiation" (Fig. 3a). Besides, we noticed that KEGG pathway "Graft verse host disease" was also enriched at moderately significant level (FDR around 0.19 and nominal p < 0.03), which included HLA-DQB1, HLA-DQA1, HLA-DOB, HLA-DMA and HLA-DOA. More than 30 pathways were enriched in progressors, which included "responses to type I interferon", "mitotic cell cycle", "defense response to virus", "innate immune response" and "regulation of interleukin_1 beta production". Furthermore, to validate the reliability of these datasets from Affemetrix microarray, we selected another dataset (GSE87620) published recently, using illumina Human HT-12 V4 microarrays. Using GSE87620 as a validating set, we selected the up-regulated genes with FDR < 0.05 using online tool GEO2R and performed Gene Ontology annotation (Fig. 3b). The enriched pathways were highly consistent with the results from MADNT. "translational initiation", "SRP-dependent cotranslational protein targeting to membrane" and "nuclear-transcribed-mRNA catabolic process, nonsense-mRNA decay" was enriched in NPs.

In the aspect of stratification subgroup, 6 pathways were enriched with FDR less than 0.25 in CD4+ T cells of ECs which included "multi organism metabolic process", "translational initiation", "Nuclear transcribed mRNA catabolic process nonsense-mediated decay", "protein targeting to membrane"; Similarly, 32 pathways were enriched with FDR less than 0.01 in CD8+ T cells of ECs, most of which were associated with pathways related to translation and protein-targeting to membrane (Additional file 4: Table S4).

Weighted gene co-expression network analysis

WGCNA provides insights to disease pathogenesis by studying co-expression genes between clinical samples based on gene expression correlation coefficients. The reliability of this algorithm depends on a large sample size, which we are able to apply the MADNT and two subgroups (CD8+ T cells (EC) and VNP) expression set into WGCNA analysis.

Using WGCNA, 15 distinct modules was found (Fig. 4a). The expression data from different genes within each calculated module was used to determine the module epigengenes (ME, the first component of the respectively module), which was correlated to the clinical parameter (AIDS progression). Several modules showed significant positive correlation with lose of viremic control, including the tan modules (p = 4e−07, correlation coefficient = 0.55) and purple modules (p = 0.001, correlation coefficient = 0.37) and grey60 module (p = 0.003, correlation coefficient = 0.34) (Fig. 4b). However, no modules were obviously correlated to viremic control. For the three modules related to disease progression, 39 characteristic genes were subject to functional annotation and pathway enrichment. They were totally enriched in 19 GO terms. The genes were largely related to type I interferon signaling pathway (IFIT1, IFIT3, IFI6, OAS1, OAS2, MX1), complement activation (IGHA1, IGHD, IGHG1, IGHM, IGKV1-17, IGKV1-37, IGKV1-39) and positive regulation of B cell activation (IGHA, IGHD, IGKC, IGLC1). Additionally, PPI networks of the three modules were visualized by Cytoscape 3.6 (Fig. 4c). RRM2, IGLC2 and RSAD2 were the hub genes of grey60, purple and tan modules, respectively.

As to the stratification subgroup, WGCNA approach was applied to CD8+ T cells of ECs and VNPs, respectively. We identified 24 distinct coexpression modules in CD8+ T cells containing 28 to 1174 genes per module. Salmon module showed a strong positive correlation with AIDS progression including CD38 and ISGs which is consistent with DEG and GSEA analysis. Magenta and purple module (Additional file 5: Table S5) showed a positive correlation with viral control in ECs including gene involved in regulation of
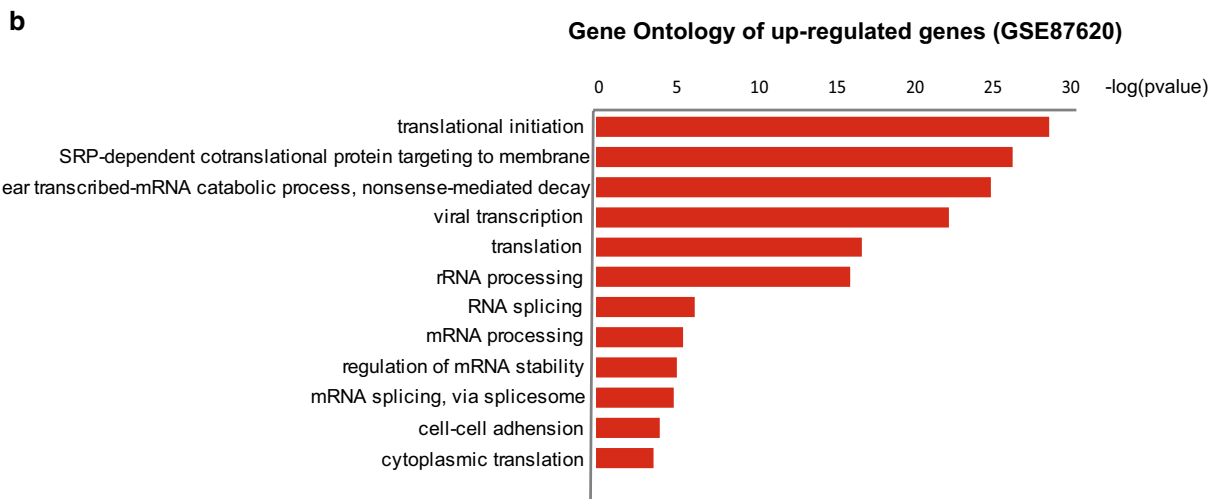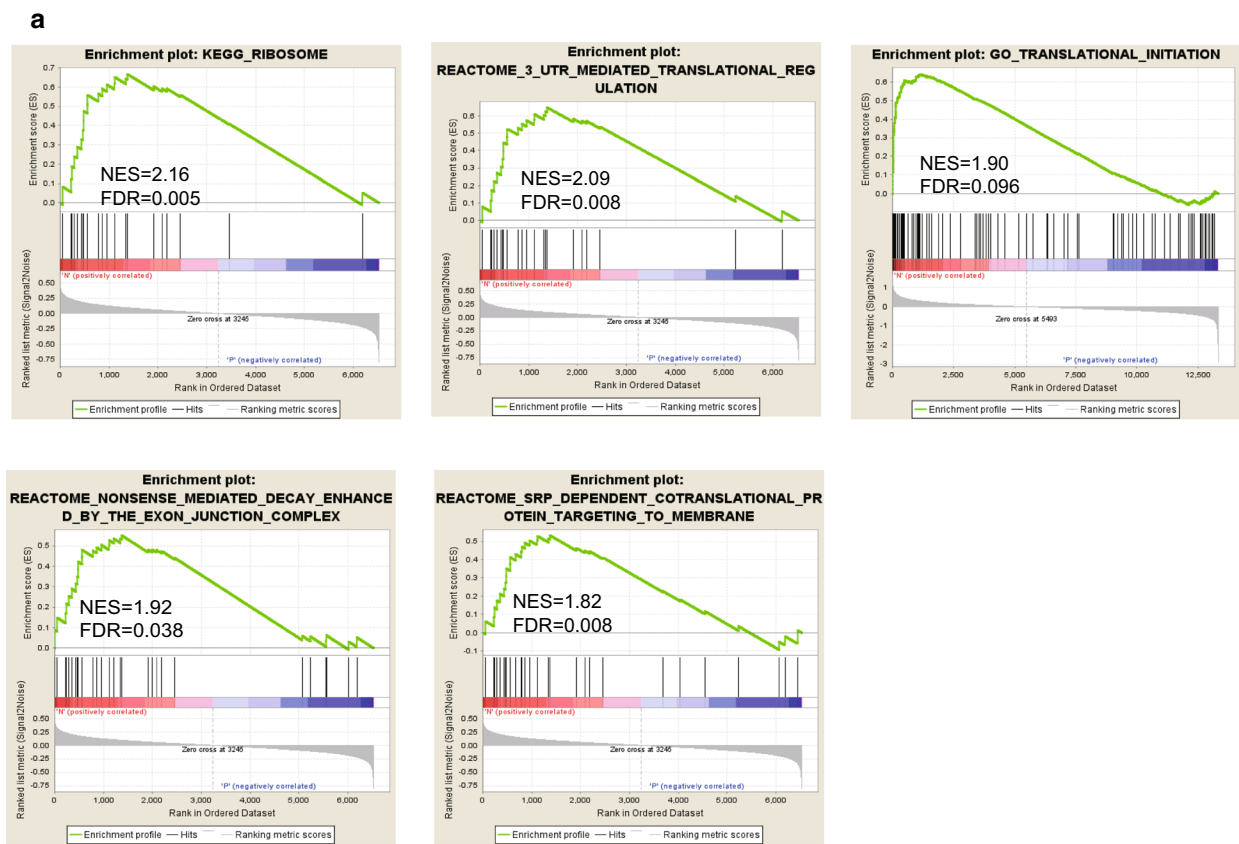
Ding *et al. J Transl Med*   (2019) 17:35

Page 8 of 13

Fig. 3 GSEA analysis of MADNT expression data. **a** Prominent transcriptional patterns in NPs. The diagram showed GSEA plots for 5 pathways recapitulating the main difference between NPs and PPs (NPs were shown at left-hand side in GSEA plot, PPs right-hand side). The vertical axis in the upper graph indicates Enrichment Score (ES) for genes in each gene set. The barcode plot indicates the position of genes in each gene set. Red and blue colors represent positive and negative Pearson correlation with HIV progression. **b** Functional annotation of up regulated genes with FDR less than 0.05 in the validating set (GSE87620)

cellular response to stress (DNAJA1, HSPA5, PIK3R1 and PMAIP1), leukocyte migration (SCL7A5, YES1, B4GALT1, NRAS, PDE4B, PIK3R1 and PODXL) and response to cytokines (REL, SKIL, MAPKAPK2 and NFKB2). Interestingly, two modules were significantly associated with immune control in VNPs (darkturquoise,

Ding *et al. J Transl Med*   (2019) 17:35

Page 9 of 13
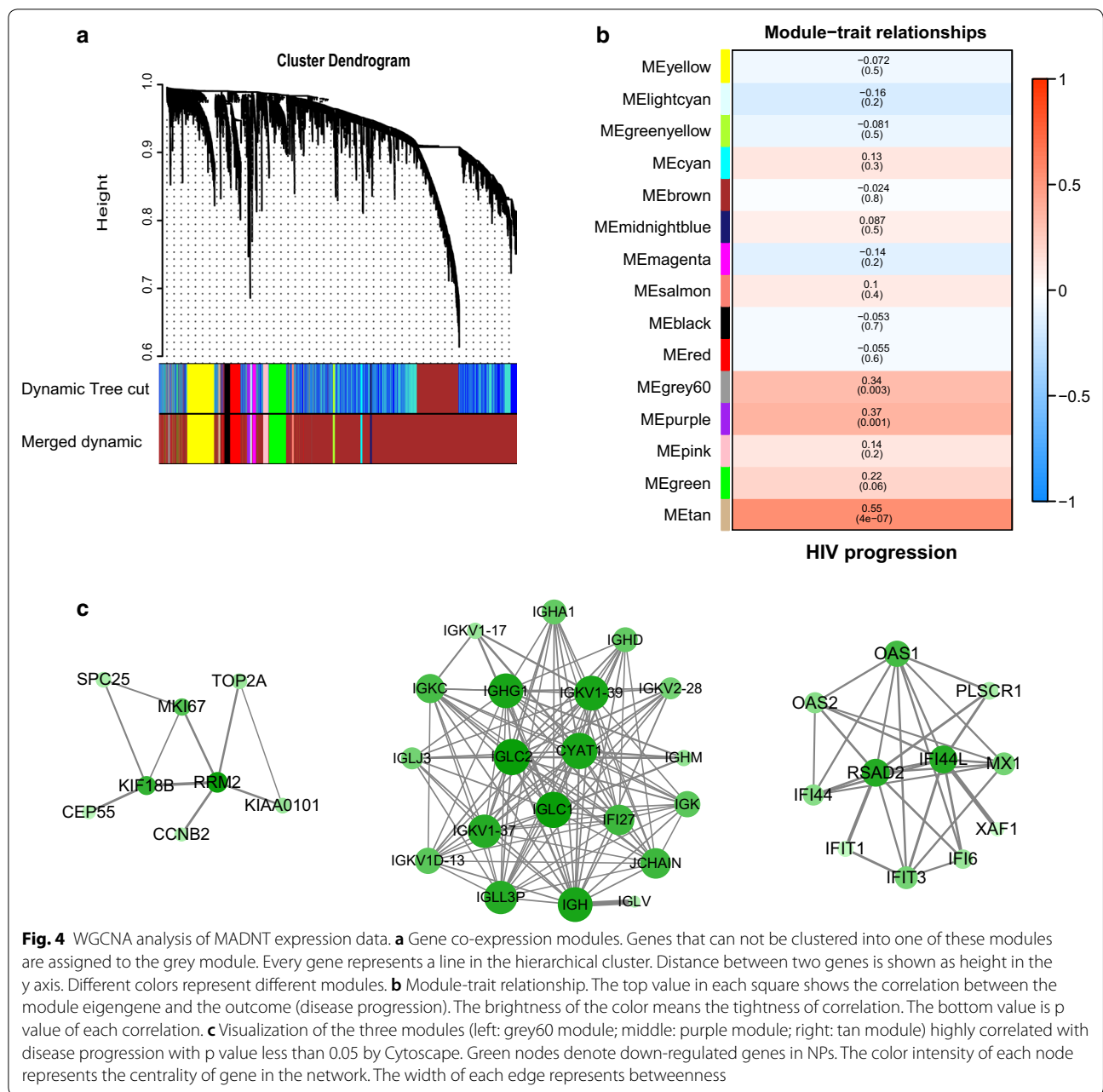
Fig. 4 WGCNA analysis of MADNT expression data. **a** Gene co-expression modules. Genes that can not be clustered into one of these modules are assigned to the grey module. Every gene represents a line in the hierarchical cluster. Distance between two genes is shown as height in the y axis. Different colors represent different modules. **b** Module-trait relationship. The top value in each square shows the correlation between the module eigengene and the outcome (disease progression). The brightness of the color means the tightness of correlation. The bottom value is p value of each correlation. **c** Visualization of the three modules (left: grey60 module; middle: purple module; right: tan module) highly correlated with disease progression with p value less than 0.05 by Cytoscape. Green nodes denote down-regulated genes in NPs. The color intensity of each node represents the centrality of gene in the network. The width of each edge represents betweenness

ME correlation = 0.87, p-value = 2e−4; red, ME correlation = 0.77, p-value = 0.004) (Fig. 5a). ZNF395, PDCD4-AS1 and EIF4A3 that are among the 30 up-regulated genes were involved in the darkturquoise module, Notably, KIR2DS3 and KIR3DL3 which were reported to affect immune control of HIV by interacting with HLA-B molecules in a specific manner also appeared in the darkturquoise module. We failed to enrich GO or KEGG terms of these modules, probably indicating unknown important pathways or interactions between genes in these two modules accounting for the immune control in

VNPs. PPI networks of the two modules were visualized by Cytoscape 3.6 (Fig. 5b, c).

Discussion

Increasing transcriptional profiles allowed us to investigate the significant genes and functional pathways relevant to viral control in ECs and VNPs in a large scale level. In this study, we adopted an integrative approach to uncover the transcriptional patterns in LTNPs and ECs and identified several key genes and functional pathways

Ding *et al. J Transl Med*   (2019) 17:35

Page 10 of 13

Fig. 5 WGCNA analysis of VNP transcriptome. **a** Module-trait relationship. The top value in each square represent the correlation between the module eigengene and the outcome (disease progression). The brightness of the color means the tightness of correlation. The bottom value is p value of each correlation. Visualization of darkturquoise module (**b**) and red module (**c**) which was highly correlated with disease progression with p value less than 0.05 by Cytoscape. Orange nodes denote up-regulated genes in VNPs. The color intensity of each node represents the centrality of gene in the network. The width of each edge represents betweenness

which are significantly connected to viral control and disease progression.

We found several common genes up-regulated in all ECs and VNPs, some of which have never been reported before, including ZBTB10, ZNF395, GPX4, CHPT1, METTL9, SPOCK2, EIF4A3, GADD45A, SLC75A, TGIF1, CMPK1, LMNA, PLXDC1 and BACH2. GADD45A is a stress-induced protein and participate in various cellular pathways. A recent study reported that overexpression of GADD45 proteins reduces HIV-1 production through suppressing transcription from the HIV-1 LTR promoter [49]. We have experimentally examined several up-regulated genes for their anti-HIV

activity in cell lines, most of which were demonstrated to suppress the replication of VSVG-pseudotyped HIV-luc viruses. The biological relevance of these up-regulated genes in the context of HIV pathogenesis needs further investigation.

On the other hand, an arsenal of ISGs were down-regulated in NPs compared to chronic progressors, which further debates the role of interferon in chronic HIV infection. Several researches by other groups observed similar phenomenon. Margalida Rotger showed several ISGs (IF44, MX1, EIF2AK2, IFI6, LY6E, TRIM22) were up-regulated in RPs who got a progressive immunosuppression soon after seroconversion while a lower

Ding *et al. J Transl Med*   (2019) 17:35

Page 11 of 13

expression of ISGs was observed in CD8+ T cells in viremic nonprogressor (VNP) profile [50]. Sankaran et al. reported a significant increase in gene expression regulating immune activation and inflammatory response in intestinal mucosa in HIV-infected patients with high viral load compared to LTNPs [51]. In accordance with this, CD38 and LAG3 which were found to be elevated in RPs have also been identified in our analysis. Consistent with this knowledge, a research reported a significant increase of ISG expression with increased viral load, including genes of intrinsic antiretroviral defense [52]. Currently, the contribution of IFN-Is to the control of viral infection and to the immunopathogenesis of AIDS is still under debate [53, 54]. Supportively, Liang [55] implied that type I interferon contributed to aberrant immune activation, T cell depletion and dysfunction during chronic HIV-1 infection. They found that persistent HIV-1 infection in humanized-mice led to induction of IFNs and ISGs including MX2, IFITM3, Trim22, ISG15, OAS1, and IFN regulatory factor 7 (IRF7) both in peripheral blood mononuclear cells (PBMCs) and in the spleen. They also observed the enhanced expression of CD38 and HLA-DR, similar to our data. Persistent immune activation plays a central role in CD4+ T cell depletion and progression to AIDS and may be considered to be a predictor of disease progression in ART-naïve patients. Taken together, our analysis supports the concept that IFNs and ISGs are a double-edged sword during chronic infection and highlighted the caution of IFNs usage in HIV-1 chronically infected patients.

Given the limited power of classical function annotation, GSEA and WGCNA algorithm was performed on our combined transcriptional profile individually. Pathways including "translational initiation", "3′UTR mediated translational regulation" and "nuclear transcribed-mRNA nonsense-mediated decay" were shown to be positively correlated with viremic control. In contrast, pathways related to interferon and immune response, complement activation and cell cycle seemed to be highly correlated with AIDS disease progression. Notably, ribosomal substitutes including RPS20, RPS28, RPS15A, RPS25, RPS6, RPS21, RPS3, RPL36, RPL9, RPL31, RPL23, RPL27, RPL30, RPL29, RPL35 and translational initiation factors including EIF2S3, EIF1, EIF2C2, EIF4G2, EIF4A3 were elevated in NPs compared with chronic progressors by GSEA analysis. This indicates that defects in ribosome components and blockade of translational initiation probably play a significant deleterious role in HIV progression. In agreement with our findings, a recent study reported several ribosomal formation genes including RPL27, PRS7, RPL24, RPS13 and RPL10L was down regulated in NK cells from HIV infected individuals [56]. Our most interesting findings were the identification of

gene networks which were highly associated with non-progression status in VNPs. These patients who retained a functional immune control of circulating viruses represented a very perplexing question pertinent to the LTNP status. Our findings hopefully provide a clue to resolve the mystery.

Nevertheless, we can not rule out the possibility that the changed transcription level of some gene found herein may reflect a consequence rather than a cause of low viral replication in LTNPs. The role of these candidate genes reported herein in controlling HIV-1 infection awaits further investigation.

Conclusions

In summary, our integrative analysis of microarray data relevant to ECs and VNPs provides an overview of the biomarkers and transcriptional patterns in nonpathogen individuals compared with chronic progressors. We observed that HIV controllers had reduced immune activation and interferon response, which collaborates the concept that interferons might play a deleterious role in AIDS disease progression. Moreover, we have identified several key genes responsible for the non-progressive status, some of which were validated in cell lines. We also detected several pathways that may be related to the exacerbation of immunosuppression in AIDS patients, i.e., blockade of translation and dysfunction of T cell homeostasis. Altogether, our integrative genome-wide analysis has provided new knowledge for HIV-1 pathogenesis and immune intervention of the disease progression of AIDS.

Additional files

Additional file 1. Clinical parameters of patients involved.

Additional file 2. DEGs in NPs compared with PPs in MADNT.

Additional file 3. DEGs in NPs compared with PPs in CD8+ T cells.

Additional file 4. Pathways enriched in CD8+ T cells from ECs analyzed by GSEA.

Additional file 5. Genes in modules correlated with protection in CD8+ T cells (EC).

Ding *et al. J Transl Med*   (2019) 17:35

Page 12 of 13

cell; IRF7: IFN regulatory factor 7; IQR: inner-quartile range; DMEM: Dulbecco's modified Eagle medium.

Authors' contributions
DJW performed the majority of work described here. DJW and ML wrote and edited the manuscript. ZJY and XYL collected and examined the data. SC and XYL supervised the project. All authors read and approved the final manuscript.

Author details
[1] Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences, Beijing 100050, China. [2] Key Laboratory of the Ministry of Education for Advanced Catalysis Materials, Department of Chemistry, Zhejiang Normal University, Jinhua, Zhejiang 321004, China.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Imami N, Westrop SJ, Grageda N, Herasimtschuk AA. Long-term nonprogression and broad HIV-1-specific proliferative T-cell responses. Front Immunol. 2013;4:58.
2. Saksena NK, Rodes B, Wang B, Soriano V. Elite HIV controllers: myth or reality? AIDS Rev. 2007;9(4):195–207.
3. Grabar S, Selinger-Leneman H, Abgrall S, Pialoux G, Weiss L, Costagliola D. Prevalence and comparative characteristics of long-term nonprogressors and HIV controller patients in the French Hospital Database on HIV. Aids. 2009;23(9):1163–9.
4. Gonzalo-Gil E, Ikediobi U, Sutton RE. Mechanisms of virologic control and clinical characteristics of HIV+ elite/viremic controllers. Yale J Biol Med. 2017;90(2):245–59.
5. Blankson JN, Bailey JR, Thayil S, Yang HC, Lassen K, Lai J, et al. Isolation and characterization of replication-competent human immunodeficiency virus type 1 from a subset of elite suppressors. J Virol. 2007;81(5):2508–18.
6. Wang B, Dyer WB, Zaunders JJ, Mikhail M, Sullivan JS, Williams L, et al. Comprehensive analyses of a unique HIV-1-infected nonprogressor reveal a complex association of immunobiological mechanisms in the context of replication-incompetent infection. Virology. 2002;304(2):246–64.
7. Lum JJ, Cohen OJ, Nie Z, Weaver JG, Gomez TS, Yao XJ, et al. Vpr R77Q is associated with long-term nonprogressive HIV infection and impaired induction of apoptosis. J Clin Investig. 2003;111(10):1547–54.
8. Yamada T, Iwamoto A. Comparison of proviral accessory genes between long-term nonprogressors and progressors of human immunodeficiency virus type 1 infection. Adv Virol. 2000;145(5):1021–7.
9. Caly L, Saksena NK, Piller SC, Jans DA. Impaired nuclear import and viral incorporation of Vpr derived from a HIV long-term non-progressor. Retrovirology. 2008;5:67.
10. Kirchhoff F, Greenough TC, Brettler DB, Sullivan JL, Desrosiers RC. Brief report: absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection. N Engl J Med. 1995;332(4):228–32.
11. Salvi R, Garbuglia AR, Di Caro A, Pulciani S, Montella F, Benedetto A. Grossly defective nef gene sequences in a human immunodeficiency virus type 1-seropositive long-term nonprogressor. J Virol. 1998;72(5):3646–57.
12. Pushker R, Jacque JM, Shields DC. Meta-analysis to test the association of HIV-1 nef amino acid differences and deletions with disease progression. J Virol. 2010;84(7):3644–53.
13. Brambilla A, Turchetto L, Gatti A, Bovolenta C, Veglia F, Santagostino E, et al. Defective nef alleles in a cohort of hemophiliacs with progressing and nonprogressing HIV-1 infection. Virology. 1999;259(2):349–68.
14. Michael NL, Chang G, d'Arcy LA, Tseng CJ, Birx DL, Sheppard HW. Functional characterization of human immunodeficiency virus type 1 nef genes in patients with divergent rates of disease progression. J Virol. 1995;69(11):6758–69.
15. Mourich DV, Lee S, Reyes-Teran G, Mackewicz CE, Levy JA. Lack of differences in nef alleles among HIV-infected asymptomatic long-term survivors and those who progressed to disease. AIDS Res Hum Retroviruses. 1999;15(17):1573–5.
16. Pernas M, Casado C, Arcones C, Llano A, Sanchez-Merino V, Mothe B, et al. Low-replicating viruses and strong anti-viral immune response associated with prolonged disease control in a superinfected HIV-1 LTNP elite controller. PLoS ONE. 2012;7(2):e31928.
17. Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, Goepfert PA. Magnitude of functional CD8+T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma. J Virol. 2002;76(5):2298–305.
18. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov VV, et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. J Exp Med. 2009;206(6):1253–72.
19. Berger CT, Frahm N, Price DA, Mothe B, Ghebremichael M, Hartman KL, et al. High-functional-avidity cytotoxic T lymphocyte responses to HLA-B-restricted Gag-derived epitopes associated with relative HIV control. J Virol. 2011;85(18):9334–45.
20. Ferrando-Martinez S, Casazza JP, Leal M, Machmach K, Munoz-Fernandez MA, Viciana P, et al. Differential Gag-specific polyfunctional T cell maturation patterns in HIV-1 elite controllers. J Virol. 2012;86(7):3667–74.
21. Blankson JN, Siliciano RF. Elite suppression of HIV-1 replication. Immunity. 2008;29(6):845–7.
22. Betts MR, Nason MC, West SM, De Rosa SC, Migueles SA, Abraham J, et al. HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+T cells. Blood. 2006;107(12):4781–9.
23. Ferre AL, Hunt PW, Critchfield JW, Young DH, Morris MM, Garcia JC, et al. Mucosal immune responses to HIV-1 in elite controllers: a potential correlate of immune control. Blood. 2009;113(17):3978–89.
24. Almeida JR, Price DA, Papagno L, Arkoub ZA, Sauce D, Bornstein E, et al. Superior control of HIV-1 replication by CD8+T cells is reflected by their avidity, polyfunctionality, and clonal turnover. J Exp Med. 2007;204(10):2473–85.
25. Walker BD, Yu XG. Unravelling the mechanisms of durable control of HIV-1. Nat Rev Immunol. 2013;13(7):487–98.
26. Emu B, Sinclair E, Favre D, Moretto WJ, Hsue P, Hoh R, et al. Phenotypic, functional, and kinetic parameters associated with apparent T-cell control of human immunodeficiency virus replication in individuals with and without antiretroviral treatment. J Virol. 2005;79(22):14169–78.

Ding *et al. J Transl Med*   (2019) 17:35

Page 13 of 13

27. Salgado M, Rallon NI, Rodes B, Lopez M, Soriano V, Benito JM. Long-term non-progressors display a greater number of Th17 cells than HIV-infected typical progressors. Clinical immunology. 2011;139(2):110–4.

28. Rouers A, Klingler J, Su B, Samri A, Laumond G, Even S, et al. HIV-specific B cell frequency correlates with neutralization breadth in patients naturally controlling HIV-infection. EBioMedicine. 2017;21:158–69.

29. Wu JQ, Sasse TR, Wolkenstein G, Conceicao V, Saksena MM, Soedjono M, et al. 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.

30. O'Connor GM, Holmes A, Mulcahy F, Gardiner CM. Natural Killer cells from long-term non-progressor HIV patients are characterized by altered phenotype and function. Clin Immunol. 2007;124(3):277–83.

31. Leng J, Ho HP, Buzon MJ, Pereyra F, Walker BD, Yu XG, et al. A cell-intrinsic inhibitor of HIV-1 reverse transcription in CD4(+) T cells from elite con-trollers. Cell Host Microbe. 2014;15(6):717–28.

32. Leitman EM, Willberg CB, De Burgh-Thomas A, Streeck H, Goulder PJ, Mat-thews PC. Subdominant Gag-specific anti-HIV efficacy in an HLA-B*57-positive elite controller. Aids. 2016;30(6):972–4.

33. Moosa Y, Tanko RF, Ramsuran V, Singh R, Madzivhandila M, Yende-Zuma N, et al. Case report: mechanisms of HIV elite control in two African women. BMC Infect Dis. 2018;18(1):54.

34. Magierowska M, Theodorou I, Debre P, Sanson F, Autran B, Riviere Y, et al. Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term nonprogressor status in human immunodeficiency virus-1-infected individuals. Blood. 1999;93(3):936–41.

35. Horton H, Frank I, Baydo R, Jalbert E, Penn J, Wilson S, et al. Preservation of T cell proliferation restricted by protective HLA alleles is critical for immune control of HIV-1 infection. J Immunol. 2006;177(10):7406–15.

36. Martin MP, Gao X, Lee JH, Nelson GW, Detels R, Goedert JJ, et al. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nat Genet. 2002;31(4):429–34.

37. Martin MP, Qi Y, Gao X, Yamada E, Martin JN, Pereyra F, et al. Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. Nat Genet. 2007;39(6):733–40.

38. Guergnon J, Dalmasso C, Broet P, Meyer L, Westrop SJ, Imami N, et al. Single-nucleotide polymorphism-defined class I and class III major his-tocompatibility complex genetic subregions contribute to natural long-term nonprogression in HIV infection. J Infect Dis. 2012;205(5):718–24.

39. International HIVCS, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science. 2010;330(6010):1551–7.

40. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conduct-ing a meta-analysis of gene expression microarray datasets. PLoS Med. 2008;5(9):e184.

41. Wang X, Kang DD, Shen K, Song C, Lu S, Chang LC, et al. An R pack-age suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. Bioinfor-matics. 2012;28(19):2534–6.

42. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545–50.

43. Zhang B, Horvath S. A general framework for weighted gene co-expres-sion network analysis. Stati Appl Genet Mol Biol. 2005;4:17.

44. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9:559.

45. Ma L, Zhang Z, Liu Z, Pan Q, Wang J, Li X, et al. Identification of small molecule compounds targeting the interaction of HIV-1 Vif and human APOBEC3G by virtual screening and biological evaluation. Sci Rep. 2018;8(1):8067.

46. Ding J, Zhao J, Yang Z, Ma L, Mi Z, Wu Y, et al. Microbial natural product alternariol 5-*O*-methyl ether inhibits HIV-1 integration by blocking nuclear import of the pre-integration complex. Viruses. 2017;9:5.

47. Li M, Kao E, Gao X, Sandig H, Limmer K, Pavon-Eternod M, et al. Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. Nature. 2012;491(7422):125–8.

48. Brombin C, Diomede L, Tudor D, Drillet AS, Pastori C, Poli E, et al. A non-parametric procedure for defining a new humoral immunologic profile in a pilot study on HIV infected patients. PLoS ONE. 2013;8(3):e58768.

49. Liang Z, Liu R, Zhang H, Zhang S, Hu X, Tan J, et al. GADD45 proteins inhibit HIV-1 replication through specific suppression of HIV-1 transcrip-tion. Virology. 2016;493:1–11.

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

51. Sankaran S, Guadalupe M, Reay E, George MD, Flamm J, Prindiville T, et al. Gut mucosal T cell responses and gene expression correlate with protec-tion against disease in long-term HIV-1-infected nonprogressors. Proc Natl Acad Sci USA. 2005;102(28):9860–5.

52. Rotger M, Dang KK, Fellay J, Heinzen EL, Feng S, Descombes P, 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.

53. Doyle T, Goujon C, Malim MH. HIV-1 and interferons: who's interfering with whom? Nat Rev Microbiol. 2015;13(7):403–13.

54. Bosinger SE, Utay NS. Type I interferon: understanding its role in HIV pathogenesis and therapy. Current HIV/AIDS reports. 2015;12(1):41–53.

55. Cheng L, Yu H, Li G, Li F, Ma J, Li J, et al. Type I interferons suppress viral replication but contribute to T cell depletion and dysfunction during chronic HIV-1 infection. JCI Insight. 2017;2:12.

56. Boeijen LL, Hou J, de Groen RA, Verbon A, Boonstra A. Persistent replica-tion of HIV, HCV and HBV results in distinct gene expression profiles by human NK cells. J Virol. 2018. https://doi.org/10.1128/JVI.00575-18.