Research Article

# Thrombospondin Regulation Differentiates HIV Patients with Natural and Therapy-Mediated Control of Plasma Viremia: A Genome-Wide Transcriptomic Analysis

**Viviane Nascimento da Conceicao[1,2], Wayne B. Dyer[3], Monica Miranda-Saksena[4], Kaushal Gandhi[5], Nitin Saksena[2]**

1. Dept of Medicine, School of Medicine, The University of Texas Health Science Center at San Antonio, San Antonio, United States; 2. ARNA Pharma and Arna Lifesciences, Australia; 3. Research & Development, Australian Red Cross Lifeblood, Sydney, Australia; 4. Herpes Neuropathogenesis group, Westmead Institute for Medical Research, University of Sydney, Australia; 5. Integrated Sciences, Australia

**The genomic mechanisms underlying natural and effective control of viremia in HIV+ long-term non-progressors (LTNP) and elite controllers (ECs) remain obscured and poorly understood. We performed genome-wide transcriptomic expression analysis (Bead-Studio; 25,000 genes) on peripheral blood mononuclear cells from nine therapy-naïve LTNPs compared with 15 newly diagnosed HIV+ patients before and after > one year of control of viremia by HAART. Only significantly differentially expressed (DE) genes with p-value <0.01 and FDR (false discovery rate (FDR) of <1% were considered for further analysis. Pathway analysis was performed using MetaCoreTM to derive the functional annotations. Functionally significant genes were validated using quantitative real-time PCR, flow cytometry, and confocal and deltavision microscopy. Although LTNP had naturally controlled viremia (low to undetectable), gene expression levels in these LTNP were distinct from those in HIV+ patients with viremia controlled by HAART (below detection), highlighting the critical role and uniqueness of enriched pathways in the natural control of viremia in the LTNPs and ECs. Thrombospondin (THBS1) ($R^2$ = 0.942) was identified as a potential biomarker in our study, discriminating between viremic patients and LTNPs at the genomic ($R^2$ = 0.942, p = 2.654e-08) and proteomic (p = 0.003761) levels. The expression levels of THBS1 were correlated with plasma viremia ($R^2$ = 0.81557; p = 0.0003761). Our results suggest a significant distinction between immune pathways in patients with therapy-mediated control of HIV replication compared to those with natural control of viremia in LTNPs, as identified by**

 doi.org/10.32388/WGKALR 

**enriched genomic expression in immune activation, cytoskeletal remodeling, apoptosis, and T-cell signaling pathways. Thrombospondin plays an essential role in apoptosis; therefore, the downregulation of this marker in viremic patients may offer potential as a biomarker for characterizing pathways of immunological control in untreated LTNP, in addition to predicting optimal treatment response in newly diagnosed HIV+ patients.**

Viviane Nascimento da Conceição and Wayne B. Dyer contributed equally to this work and should be recognized as co–first authors.

**Corresponding authors:** Nitin K. Saksena, nitin.saksena@bigpond.com; Monica Miranda-Saksena, monica.miranda@sydney.edu.au

## 1. Introduction

HIV can manipulate and subvert human gene machinery through complex interactions with the host immune system. The virus depends on the host cell for replication, but it is constantly exposed to the host's antiviral immune response. Manipulation of the host response by HIV forms the basis for host-virus interactions, during which the expression of host genes and proteins is altered to establish persistent infection [1].

The clinical progression of HIV-1 disease is generally slow, taking several years from the initial infection to the development of AIDS. In the absence of antiretroviral therapy, a subset of persistently infected individuals showed no signs of disease progression beyond 15-20 years, with normal CD4+ and CD8+ T cell counts and undetectable plasma viremia (<50 copies of viral RNA/ml). These individuals are called Elite Controllers (ECs). Cell-mediated and humoral immune responses and the production of neutralizing antibodies are usually robust in the absence of antiretroviral therapy in ECs [2][3][4][5][6]. HIV disease progression varies among individuals, whereby host genetic and immune factors interact with viral genetic factors to determine the rate of disease progression to AIDS [2][3][7][8][9][10][11]. Despite numerous investigations into the underlying factors contributing to elite control, the genomic basis for natural control of viral replication in the absence of antiretroviral treatment remains unclear. Thus, identifying the determinants of protection against HIV-1 disease progression is of great importance, as this may facilitate the development of new treatment strategies and biomarkers for prognosis and diagnosis [12].

In HIV1+ individuals with a virological response to HAART, plasma viremia remains below detectable limits (<50 copies/ml) but depends on sustained therapy, comparable to the viral loads seen in ECs in the absence of ART. However, the critical difference is that despite effective HAART, the immune system of HIV individuals never recovers to normal levels, a feature different from that of ECs. This finding suggests a qualitative distinction between the two groups.

The underlying genomic basis of elite viral control remains unclear. Previous gene microarray studies on HIV have used specific cell types such as CD8+T cells, macrophages, cell lines, and dendritic cells to understand gene regulation during HIV infection [13][14]. Our study is the first to dissect the genomic basis of drug-mediated control of HIV as opposed to natural control in LTNPs. This was achieved by comparing genome-wide transcriptomes, encompassing all 25,000 human genes (48,000 transcripts or probes) to define critical molecular differences between these two groups at various immune functional levels. Current microarray platforms allow high-throughput genome-wide transcriptome analysis, making it possible to understand the fundamental cellular, molecular, and genetic mechanisms underlying non-progressive HIV diseases [14]. This study used whole primary uncultured peripheral blood mononuclear cells (PBMCs) to represent a holistic genome-wide expression. Second, since we used two time points and archival samples from chronically infected LTNPs and other HIV groups, our work was only possible on whole frozen PBMCs, from which the separation of individual cell types was experimentally not feasible owing to their low quantities of cells and loss of RNA.

Our previous study showed a distinct genomic distinction between HIV patients before and after highly active antiretroviral therapy [15]. Additionally, our studies have shown an increased immune activation and inflammation role in host gene expression in HIV + individuals with viremia, while HAART downregulated genes associated with immune activation [13].

Here, we show genome-wide primary PBMC transcriptome analysis from nine long-term non-progressors (LTNPs), who had low to undetectable plasma viremia and CD4+ T cell counts >500 after 10 to 16 years infection in the absence of antiretroviral therapy. These genome-wide datasets were then compared with 15 patients with recently acquired HIV infection, before and after aggressive antiviral therapy. Our purpose was to differentiate the underlying basis of natural versus drug-mediated control of viremia and whether LTNPs are truly a unique group compared to HIV patients responding successfully to HAART.

## 2. Results

### *Clinical profiling of study participants*

We analysed genomic and proteomic expression in PBMCs collected from 15 HIV+ patients before HAART (TP1) and after long-term suppression of viremia by HAART (TP2), compared with nine therapy-naïve chronically infected HIV+ patients identified in the 1990s and classified as long-term non-progressors (LTNPs). All patients experienced an increase in CD4+ T cells and decreased viremia to below-detectable levels of HIV RNA during treatment (**Table 1**).

| Participant group | *Time HIV+ | ART (months) | CD4+ (cells/ µL) | CD8+ (cells/ µL) | †Viral load (copies/ml) | ‡ART regimen |
|---|---|---|---|---|---|---|
| **LTNP:** | | | | | | |
| **S6** | 10.7 yr | naive | 882 | 672 | †>2000 | - |
| **S12** | 10.1 yr | naive | 805 | 1012 | †<1200 | - |
| **S16** | 10.4 yr | naive | 825 | 975 | 940 | - |
| **S23** | 12.3 yr | naive | 592 | 736 | 1700 | - |
| **S24** | 11.9 yr | naive | 507 | 494 | 530 | - |
| **C13** | 11.1 yr | naive | 1012 | 484 | <400 | - |
| **C53** | 13.4 yr | naive | 666 | 720 | <400 | - |
| **C64** | 11.9 yr | naive | 1088 | 1024 | <400 | - |
| **C122** | 16.1 yr | naive | 726 | 2145 | †<400 | - |
| **ART treated:** | | | | | | |
| **SC18-pre** | ≥6 mo | - | 510 | 980 | 112000 | |
| **-post** | | 25 | 810 | 740 | <4500 | RTV, SQV, d4T, 3TC |
| **SC19-pre** | <3 mo | - | 610 | 1408 | 108634 | |
| **-post** | | 36 | 720 | 720 | <50 | SQV, d4T, 3TC |
| **SC27-pre** | <3 mo | - | 510 | 1960 | 169841 | |
| **-post** | | 19 | 620 | 1200 | <400 | SQV, d4T, 3TC |
| **SC35-pre** | <1 mo | - | 750 | 750 | 605236 | |
| **-post** | | 25 | 1010 | 840 | <400 | RTV, SQV, d4T, 3TC |
| **SC39-pre** | <6 mo | - | 1390 | 840 | 1602 | |
| **-post** | | 36 | 1730 | 930 | <50 | RTV, SQV, d4T, 3TC |
| **SC40-pre** | 0 mo | - | 600 | 680 | 3678 | |
| **-post** | | 23 | 800 | 840 | <400 | NVP, d4T, 3TC |
| **SC51-pre** | <6 mo | - | 400 | 680 | 10117 | |

| Participant group | *Time HIV+ | ART (months) | CD4+ (cells/ µL) | CD8+ (cells/ µL) | †Viral load (copies/ml) | ‡ART regimen |
|---|---|---|---|---|---|---|
| **-post** | | 12 | 810 | 1130 | <400 | RTV, IDV, d4T, 3TC |
| **SC61-pre** | <6 mo | - | 510 | 930 | 6264 | |
| **-post** | | 26 | 990 | 870 | <50 | RTV, IDV, d4T, 3TC |
| **SC64-pre** | 0 mo | - | 1050 | 950 | >750000 | |
| **-post** | | 22 | 1160 | 1160 | <50 | SQV, NLF, NVP, ddI, d4T, 3TC |
| **SC65-pre** | ≥6 mo | - | 360 | 1360 | 647095 | |
| **-post** | | 24 | 620 | 1217 | <50 | RTV, IDV, d4T, 3TC |
| **SC70-pre** | <6 mo | - | 650 | 1069 | 334088 | |
| **-post** | | 33 | 1222 | 624 | <50 | RTV, IDV, d4T, 3TC |
| **SC71-pre** | <6 mo | - | 830 | 1050 | 5208 | |
| **-post** | | 28 | 1073 | 1073 | <50 | RTV, IDV, d4T, 3TC |
| **SC75-pre** | <3 mo | - | 300 | 1420 | 92002 | |
| **-post** | | 12 | 610 | 861 | <50 | RTV, IDV, d4T, 3TC |
| **DT64-pre** | >6 mo | - | 440 | 1200 | 24835 | |
| **-post** | | 24 | 880 | 968 | <50 | NVP, ddI, d4T, 3TC |
| **DT73-pre** | 4 yr | - | 510 | 831 | 142900 | |
| **-post** | | 28 | 722 | 399 | <50 | RTV, IDV, d4T, 3TC |

**Table 1.** Clinical status of study participants.

**Time HIV+ (years or months).* †*Viral load reported from closest test date; below detection is reported as <400 or <50.* ‡*Drug abbreviations: 3TC (lamivudine); d4T (stavudine); ddI (didanosine); RTV (ritonavir); SQV (saquinavir); NVP (nevirapine); IDV (indinavir); NLF (nelfinavir); NVP (nevirapine).*


doi.org/10.32388/WGKALR

### *Microarray, clustering analysis and identification of differently expressed genes (DEGs)*

To ascertain the validity of each group, hierarchical clustering analysis was performed on the whole normalized dataset using BRB Array Tools software. There was a clear separation between LTNPs and TP1 (**Figure 1a**) and LTNPs and TP2 (**Figure 1b**), with each group forming an independent cluster. In the LTNP versus TP1 comparison, only one patient did not show a clear separation and clustered with the LTNPs (**Figure 1a**); however, this did not affect our differential expression analysis.

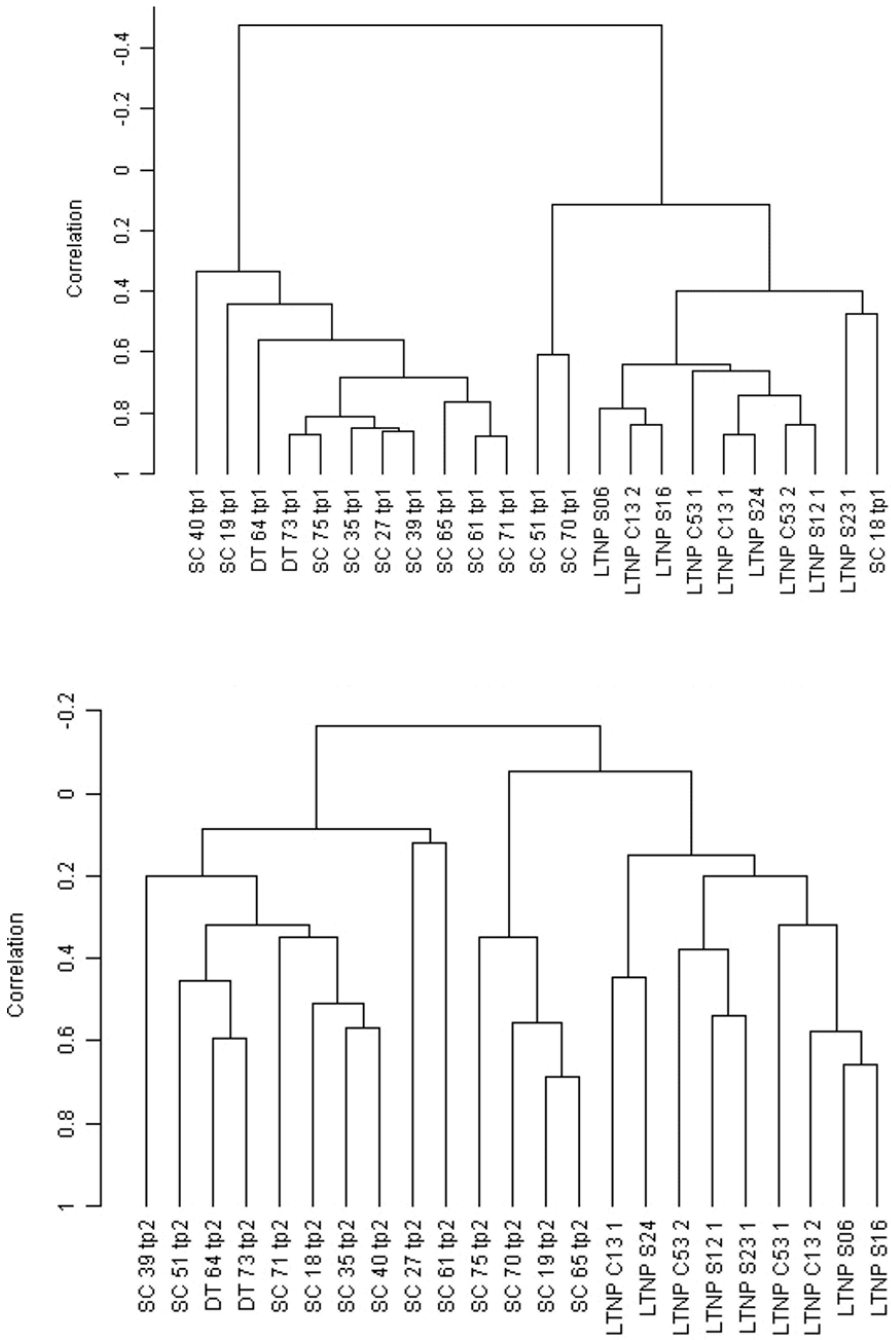



**Figure 1.** Hierarchical clustering analysis of global gene expression profiles between LTNP vs TP1 (**a**) and LTNP vs TP2 (**b**), the difference between the groups is very distinct and clear. Similarities in the gene expression patterns among individuals were evaluated and visualized with BRB Array Tools Analysis. The algorithm used is named Correlation, which computes the Pearson correlation using a 1-r distance measure. The distance on the X axis represents the similarity relationships among samples.


doi.org/10.32388/WGKALR

The analysis of DE genes between LTNP and TP1 resulted in 965 DE genes (**Supplementary data file-1**) for this comparison. Of the 965 DE genes, 706 were upregulated, and 259 were downregulated (according to their fold-change calculated using the MeV software). In contrast, when LTNPs were compared against the TP2 group, 1181 DE genes (**Supplementary data file-2**) were differentially expressed, with 727 genes upregulated and 454 genes downregulated.

### *MetaCoreTM from GeneGo, Inc Pathway Analysis*

To further ascertain the functional attributes of the DE genes between each contrast, we performed a more detailed gene analysis in MetaCoreTM to analyse the significance of DE genes at the pathway level for each contrast. We used the MetaCore website to cross-examine the biological context of the DE lists generated from the microarray analysis. MetaCoreTM analyses high-throughput data in the context of pathways, networks, and maps to identify and rank important pathways, networks, and disease terms for lists of genes, proteins, transcripts, or compounds. The DE gene lists were integrated into MetaCoreTM from GeneGo, Inc. for detailed pathway analysis for each contrast. We further analysed therapy-naïve LTNPs against the two time points before and after HAART therapy (TP1 and TP2) in HIV+ individuals.

## *Comparison between LTNP vs. TP1: Natural control in LTNP versus viremic patients prior to HAART initiation*

### *Analysis of up-regulated genes*

Genetic pathways were evaluated using the MetaCore Analytical Suite (GeneGo Inc.). The data shown using the Map Folder tool (a label that MetaCore provides to a set of genes with an overarching function of pathway participation) demonstrated that the immune response is the most significant when comparing LTNP and TP1. In these experiments, we also determined the differences in gene expression caused by viral infection in patients who naturally control HIV disease. Our data showed (**Figure 2**) that the most significant folder for this comparison was the immune response (13%) with 11 of 54 DE genes enriched in this pathway) (p=1.03e-07), followed by vascular development and tissue remodelling, wound repair and mitogenic signalling (12%), inflammatory response (11%), cell differentiation (9%), apoptosis (9%), DNA damage response, and calcium signalling (7%).

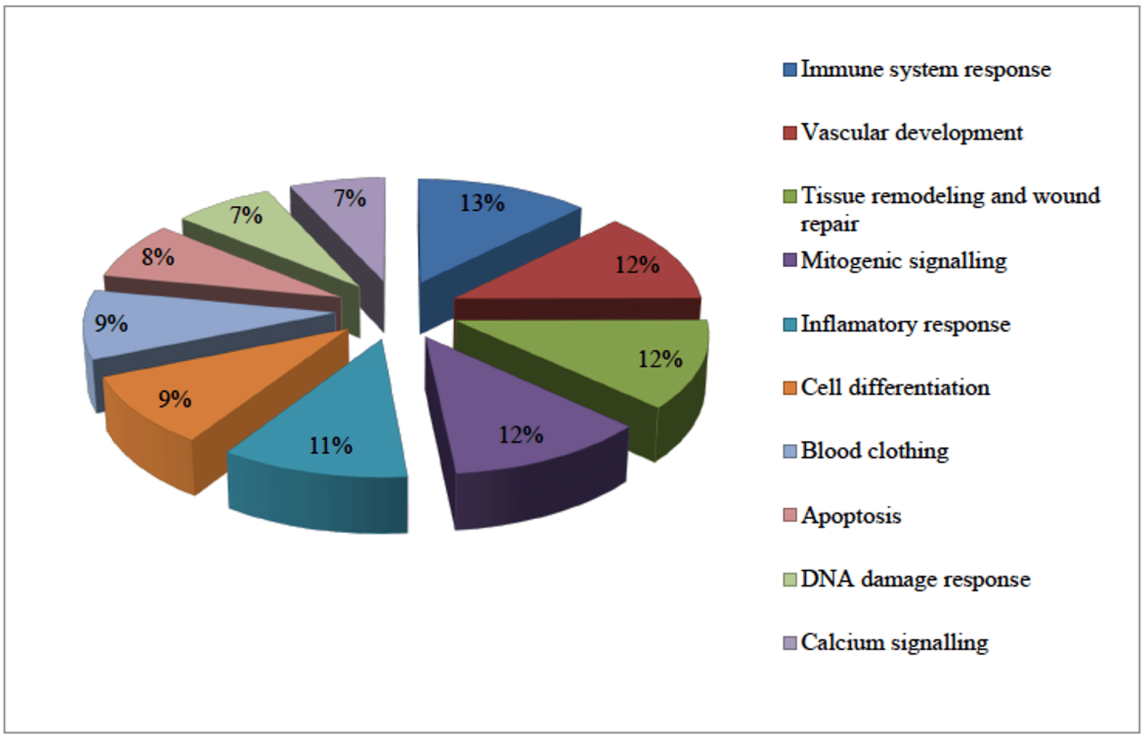



**Figure 2.** Statistically significant map folders, showing the 10 most significant map folders up-regulated in the group LTNP in comparison to the TP1. Percentage was calculated by-LOG (base 10) in Excel from the p-value calculated by MetaCore$^{TM}$.

**Supplementary Figure 1** shows the hypergeometric distribution of the map folders according to the log (p-value) (FDR<0.01 and p-value<0.01). Most of the pathways involved in these folders are related to the immune response and are upregulated in the LTNPs. All map folders showed significant p-values (**Supplementary Table 2**). As can be seen, each map folder comprises a series of pathways maps related to it. **Supplementary Table 2** shows only the top ten most significant pathways with their respective p-values and ratios (the number of genes appearing in the pathway that belong to the uploaded DE list versus the total number of genes in the pathway).

These results confirm that the immune system response was most significantly upregulated in LTNP (p=1.12e-06) than in HIV+ patients with detectable viremia. We have discussed only the top two pathways in **Supplementary Table 2**, but **Supplementary Data File 3** contains the full list of enriched pathway maps in the order of significance.

*Gene set enrichment analysis of consistently present probe sets up-regulated in the LTNP as compared to the TP1 group*

To evaluate the potential of expression profiles generated from the comparison of LTNP to the TP1 group, all probe sets detected as present across two sample groups (LTNP and TP1) were imported into MetaCore for the gene set enrichment analysis. Several significant GeneGO pathways were identified, and the genes were enriched in functional ontologies largely related to the immune response in the GeneGo maps. Furthermore, under the process network, confirmation of the involvement of immune response and B-cell receptor pathway was evident, with it being the top-ranked network process (p=4.390e-8), with 26 of 137 genes being enriched in this network. In addition, the role of B cells in LTNPs was also evident from their high enrichment in GeneGo disease (p=2.414e-9), with 51 of 393 genes enriched in B cells and lymphoma. The GeneGo maps, GeneGo process network, and GenGo Disease underpin the importance of the immune response and B-cell/B-cell-receptor pathway, being significantly enriched and correlated positively with the LTNP group and negatively with the TP1 group (**Supplementary Figure 2**).

Furthermore, we found enrichment of genes positively correlated with LTNPs in metabolic networks, with the 1,2-didocosahexaenoyl-sn-glycerol 3-phosphate pathway (**Supplementary figure 2d**) being the top-ranked. In GeneGO processes, cellular metabolism was the most prominent (p=3.035e-20), with 553 of 9400 genes enriched (**Supplementary figure 2e**) with functions in various signalling pathways, positive regulation of biological processes, signal transduction, phosphorylation, etc. **Supplementary Table 3** shows some enriched genes in the cellular metabolic process with their respective shared networks and functions. The associations between the networks were highly significant.

The enrichment analysis of the pathways provided by MetaCoreTM also included a comparative analysis of the overrepresented genes among the pathways. This analysis assesses the overall functional character of the sample set, providing a ranked representation of ontologies that are most saturated or "enriched" with the input data. Each GeneGo Process Network represents a comprehensive biological process with a specific functional theme. **Supplementary Figure 1** shows the top ten most enriched GeneGo processes for the sample set (LTNP vs. TP1) with their respective p-values. Four of the ten enrichments, representing only the genes from our DE list between LTNP and TP1, were in the immune response supported by strong p-values. Interestingly, apart from the top-ranked enrichment in the BCR pathway, which was positively correlated with LTNPs, enrichment of GO processes in the immune response related to protein folding, cell adhesion and cytoskeleton, phagocytosis, inflammation, and signalling were also highly significant (**Supplementary Figure 1**).

*Pathway analysis of up-regulated DE genes between LTNP and TP1*

To shed light on the role of differential modulation of genes between LTNP and TP1, we carried out pathway analysis using the DE gene list obtained from the comparison of LTNP and TP1 contrast. The most significant pathway upregulated in the LTNP group was the B-cell receptor (BCR) signalling pathway (p=1.03e-07), which was also confirmed by enrichment analysis (**Supplementary Figure 2**) and pathway analysis (**Figure 3;** and **Supplementary Figure 3**). This signalling pathway involves many distinct processes, including survival, tolerance or apoptosis, cytoskeletal changes, DNA damage and repair, proliferation, and differentiation into antibody-producing or memory B cells. Eleven of the 54 DE genes were overrepresented in this pathway, with high p-values (Supplementary Table 4). The genes from our DE list for this contrast involved in this pathway are related to the immune response activation of the antigen-presenting cells, such as MAPK1/3, CD79, BtK, Syk, SHIP, and HRAS (see **Supplementary Table 4** for p-values). Notably, the strongest statistical association of CD22 (p=8.25e-05) was with this pathway, which was upregulated in the LTNP group. This has high functional significance because CD22 is found on the surface of mature B cells and, to a lesser extent, in some immature B cells. Moreover, CD22 is a regulatory molecule that prevents the overactivation of the immune system. To emphasize the functional relevance of the BCR pathway in non-progressive HIV disease, a strong association was observed with the CD79 complex (p=0.0001467) (**Supplementary Table 5; Supplementary Figure 6**). CD79 is a transmembrane protein that forms a complex with the B-cell receptor (BCR) and generates a signal following antigen recognition by the BCR. Thus, it is clear that the genes in the BCR network and their upregulation during non-progressive HIV disease, when compared with HIV patients with detectable viremia (TP1), may be functionally important, which has not been previously shown.

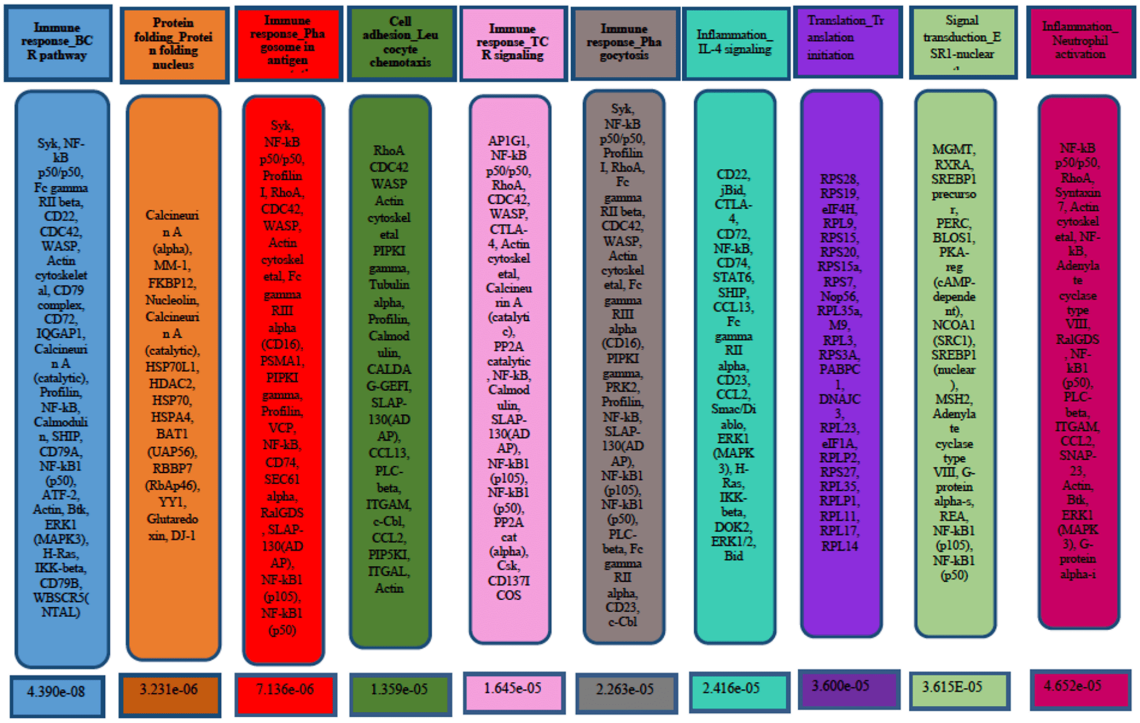



**Figure 3.** Process network derived from the comparison between LTNP x TP1. Each enriched category has its own p value shown at the bottom of the bar for the respective GO process. The genes shown within the bar are the DE genes derived from our microarray analysis.

The second most significant pathway related to the comparison was transport macro-pinocytosis regulation by growth factors (p=5.41e-07), with 11 of 63 DE genes overrepresented in this pathway (**Supplementary Figure 4**). All genes were upregulated in LTNPs. Macro-pinocytosis, an actin-dependent process, is a cell-type-specific receptor-independent endocytic pathway associated with actin-dependent plasma membrane ruffling, used mainly by dendritic cells to engulf whole pathogens. Furthermore, between the macro-pinocytosis pathway (**Supplementary Figure 4**) and the BCR pathway (**Figure 6**), some of the most significant genes, such as HRAS and CDC42, overlapped, suggesting a functional interaction between these two pathways.

Notably, the clustering of the highly significant DE genes in the actin cytoskeletal area of the macro-pinocytosis pathway, which regulates signalling pathways that control diverse cellular functions, including cell morphology, migration, endocytosis, and cell cycle progression (CDC42), genes that regulate actin polymerization in response to extracellular signals (profilin), genes involved in cellular proliferation (CtBP1), actin polymerization (Destrin or DSTN), and the actin cytoskeleton (WASP, Profilin)

(**Supplementary Figure 4**). All these genes showed strong p-values and were upregulated in the LTNPs (**Supplementary Table 5**). This is a significant observation, as genes related to the actin cytoskeleton play an important functional role in HIV disease.

### *Analysis of the down-regulated genes (LTNP versus TP1)*

Next, we determined the relevance of the downregulated genes between the LTNP and TP1 contrasts to determine how gene modulation was related to pathway modulation. As shown in the map folder analysis (**Figure 4**), we found that the most significantly downregulated pathways in the LTNP group were either directly or indirectly involved in apoptosis (19%), followed by the inflammatory response, cell cycle, and its regulation, DNA damage response (12%), vascular development (angiogenesis) (11%), immune system response (10%), mitogenic signalling (10%), cystic fibrosis disease (7%), calcium signalling, and cell differentiation (5%). **Supplementary Figure 5** shows the hypergeometric distribution of the map folders according to their -log(p-value) (FDR<0.01 and p-value<0.01). Most of the pathways involved in these folders were directly or indirectly related to apoptosis, and this pathway was uniquely coordinated and downregulated in the LTNP group.

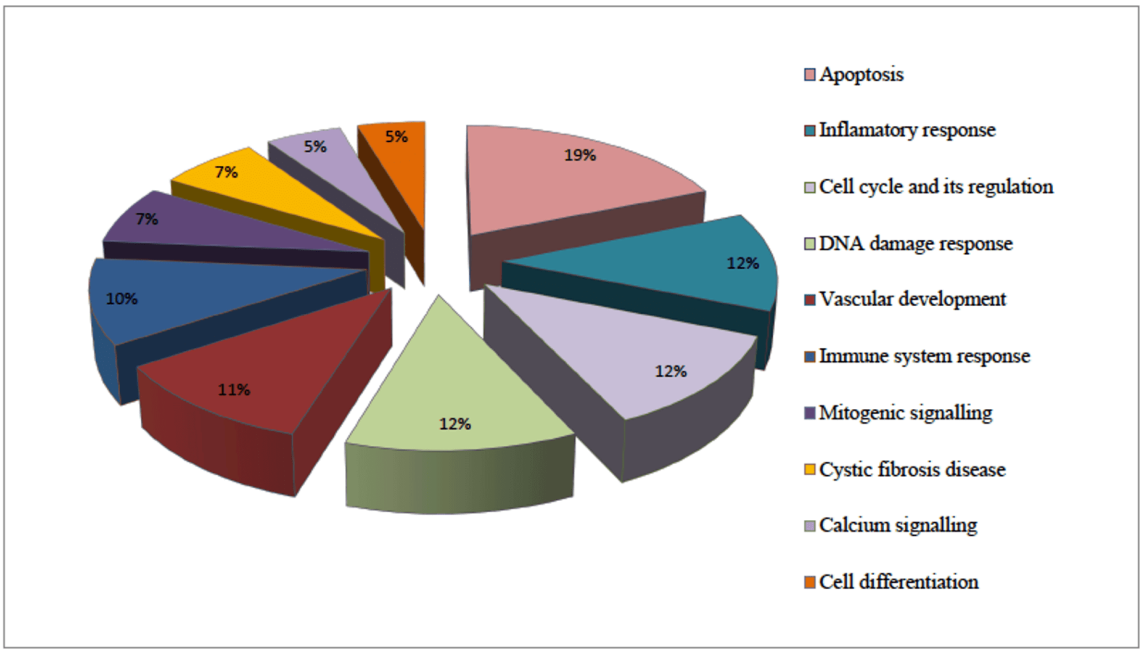



**Figure 4.** Statistically significant map folders, showing the 10 most significant map folders down-regulated in the group LTNP, in comparison to the TP1. Percentage calculated by –LOG (base 10) in Excel from the p-value calculated by MetaCore$^{TM}$.

To visualize a more profound view of the map folders (see Supplementary Table 6), we show only the top ten map folders (taken from **figure 10**) in their order of significance. Notably, 5 of the 10 pathways were significantly related to apoptosis and immune responses (**Supplementary Table 6**). This was consistent with the LTNP versus TP1 contrast for the upregulated genes discussed earlier. The p-values in each case were derived with FDR <0.01 and the cut-off p-value <0.01. Thus, even though there was an overlap between the upregulated and downregulated genes between LTNP and TP1 at the level of the immune response, a stark difference was the coordinated downregulation of apoptosis-related pathways involving genes related to apoptosis, DNA damage, cell cycle regulation, and cell differentiation. (**Supplementary Table 7**).

### *Pathway analysis of the down-regulated DE genes between LTNP and TP1*

Apoptosis was the most significant function (p=7.04e-09) and was systematically downregulated in the LTNP group. **Supplementary figure 4** shows the complete list of pathway maps in the order of significance. In the apoptosis map folder, the most significantly downregulated pathway was the pigment epithelium-derived factor (PEDF) (p=2.30e-06), followed by the apoptosis survival TNRF1 signalling pathway (p=1.45e-05) in the inflammatory response map folder (p=5.33e-06) (**Supplementary Table 6** and **Supplementary Figures 6 and 7**).

PEDF is also known as Serpin F1 (SERPINF1), a multifunctional secreted protein with anti-angiogenic, anti-tumorigenic, and neurotrophic functions. This pathway includes genes related to cellular responses to virus infection, such as NF-kB and anti-apoptotic genes (SOD2, c-IAP-1, and 2), which are very significant in this pathway. A more profound view of the PEDF pathway showed an intense clustering of genes related to the family of anti-apoptotic genes, or the apoptosis inhibitor family of genes (cIAP-1, cIAP-2, SOD2), in the anti-apoptosis pathway, along with functionally associating genes in this network (NFkB p50, p65, and GDNF) (**Supplementary Figure 6**).

The other significant pathway, which was also downregulated in LTNPs, was the Tumor-Necrosis Factor Receptor-1 (TNFR1) signalling pathway, as shown in **Supplementary Figure 7**. This was the second most significant pathway downregulated in the LTNP group. Notably, all the anti-apoptotic genes (SOD2, c-IAP-1, c-IAP-2, and Diablo-Smac) were downregulated in the LTNP group and were also shared in this pathway (**Supplementary Figures 6 and 7**), suggesting a considerable functional interaction between these two pathways. Overall, there was a coordinated downregulation unique to LTNPs, providing possible mechanistic insights into the long-term survival of LTNPs with HIV and disease progression in the absence of HAART. This is a unique observation found in this study, which distinguishes natural

controllers from HIV patients with viremia in the absence of HAART. Moreover, the independent visualization of both up-and downregulated genes has provided a clear view of how gene modulation guides the disease stage in HIV infection, in this case between LTNPs and viremic patients in the absence of HAART (TP1).

## *Natural control of viremia versus drug-mediated control of HIV: LTNP vs. TP2*

### *Up-regulated genes in LTNP vs TP2 groups*

We performed the same analysis for LTNP versus TP2, and the results were consistent with the previous comparison, LTNP versus TP1. Both comparisons shared highly related functional transcriptomic profiles, suggesting, for the first time, that there is little to demarcate transcriptomically between TP1 and TP2 when LTNPs and aviraemic patients on HAART are compared in parallel. Most importantly, these data revealed that LTNPs are a unique group of HIV patients with suppressed viremia on HAART (TP2). Alternatively, these data also imply that, although both LTNPs and TP2 have below detectable levels of HIV (<50 copies/mL), LTNPs are still distinct in their immunological profile, suggesting that despite suppressed viremia with HAART, immunological deficits prevail in the TP2 group.

The analysis of the Map Folders demonstrates that the immune response is the most significant, being up-regulated in the comparison between LTNP and TP2 (p=3.79e-14). Our data (**Figure 5**) showed that the most significant factors for this comparison were the immune response (15%), inflammatory response (14%), DNA damage response (13%), mitogenic signalling (10%), apoptosis, cell cycle and its regulation, cystic fibrosis disease and vascular development (8%), and calcium signalling (7%). **Supplementary Figure 8** shows the hypergeometric distribution of these map folders according to their log (p-value) (FDR<0.01 and p-value<0.01). Most of the pathways involved in these folders were related to the immune response and were upregulated in the LTNPs compared to the TP2. A more detailed view of the map folders, in order of significance, of the top 10 pathways is shown in **Supplementary Table 7**. We observed that three out of ten pathways were related to the immune response pathway in BCR signalling (top-ranked) (p=1.15e-09), CD16-signaling (p=1.44e-08), CCR3 signalling (p=1.80e-08), and VEGF signalling (p=8.79e-08). In addition, 2 of the 10 pathways were involved in cytoskeletal remodelling and regulation, again with strong p-values (**Supplementary Table 8**). In each of these pathways, there was considerable over-representation of genes from our DE list, with strong statistical support, suggesting the validity of the differential expression of the upregulated genes between LTNP and TP2.

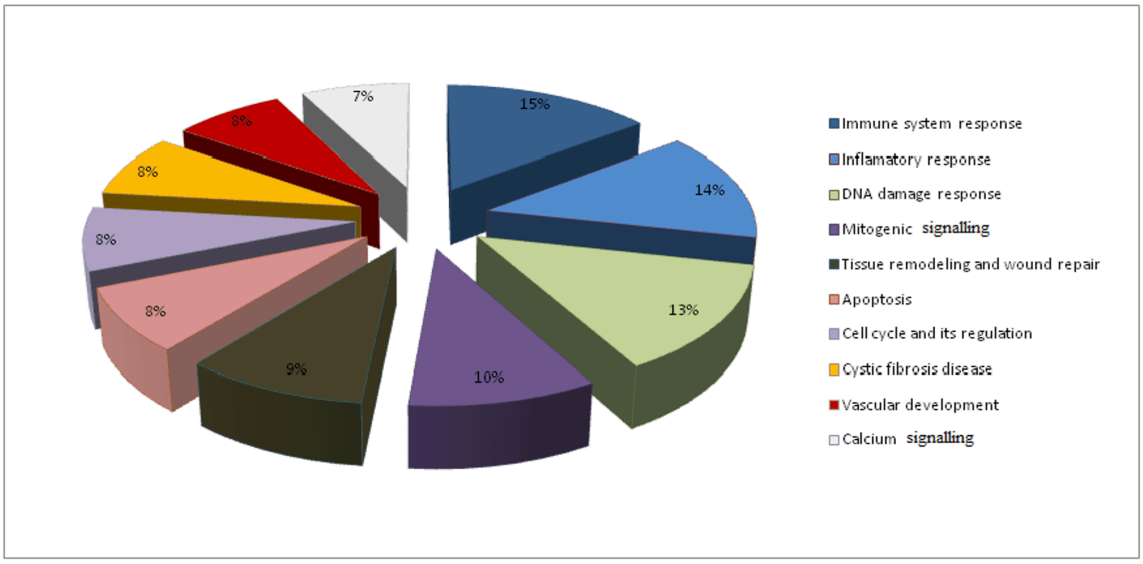



**Figure 5.** Statistically significant map folders showing the 10 most significant map folders up-regulated in the group LTNP group, when compared with the TP2 group. Percentage calculated by -LOG(base 10) in Excel from the p-value calculated by MetaCore$^{TM}$.

## *Pathway analysis of up-regulated DE genes between LTNP and TP2*

Among the upregulated genes between LTNP and TP2, the most significantly upregulated pathway was the BCR pathway (**Supplementary Figure 9**), which is analogous to the comparison between LTNP and TP1. Most of the pathways involved in this comparison were immune response-related and are shown in **Supplementary files 4 and 5**. From this comparison between LTNP and TP2, there are far more genes within the pathway with greater statistical significance. Eighteen significant DE genes were overrepresented in the BCR pathway (**Supplementary Table 8**). A large majority of the genes identified in this pathway play important roles in B-cell signalling and the activation of key leukocyte subsets mediating immune responses to HIV. Several genes identified in the LTNP vs. TP1 analysis (**Supplementary Figure 3**) and the LTNP vs. TP2 analysis (**Supplementary Figure 9**) overlapped, representing 6 of 10 most significantly altered genes (**Supplementary Table 8**). This suggests that there was little separation between the TP1 and TP2 groups compared to the LTNP group, further suggesting that the immune deficits caused by HIV replication prevail despite suppression of viremia during HAART.

### *Analysis of the DE down-regulated genes between the LTNP and TP2 groups*

In this section, we have compared the transcriptomic profiles between the two groups, LTNP and TP2, where the below-detectable levels of plasma viremia are the features. The key difference was the natural control of viremia as opposed to drug-mediated control in HIV patients.

As seen in the LTNP vs. TP1 contrast for the downregulated genes, apoptosis was again the single most discriminatory feature between LTNPs and HIV patients on and off therapy. To continue the investigation, we analyzed the map folders shown in **Supplementary Figure 10**. Only the top 10 map folders are shown in the order of significance for the LTNP vs. TP2 comparison of the downregulated genes. The hypergeometric distribution showed that apoptosis was the most significant (18%), followed by DNA damage response (14%), cell cycle and its regulation (13%), mitogenic signalling (11%), cell differentiation (10%), protein degradation (9%), vascular development (7%), inflammatory response, tissue remodelling and wound repair, and protein synthesis (6%). Supplementary Figure 10 represents these map folders as a hypergeometric distribution for these maps folders.

All map folders showed significant p-values. As can be seen, each map folder comprises a series of pathways maps related to it. **Supplementary Table 9** summarizes the top ten most significant pathways with their respective p-values and ratios. These results confirmed that apoptosis was the most significantly downregulated function in LTNP patients (p=5.97e-18) compared with HIV+ patients after HAART treatment. We have discussed only the top two pathways here, but in **Supplementary figure 6**, we have shown the remaining pathway maps, in order of significance, for all the pathways shown in Supplementary Table 9. Comparing the top 10 pathways between LTNP vs. TP1 and LTNP vs. TP2 for the downregulated genes, it is apparent that at the top-ranked pathway levels, there was considerable overlap between them (**Supplementary Tables 6 and 9**). There was a dominance of apoptosis-related pathways in both contrasts, suggesting that apoptosis is a fundamental cellular process that plays a key role in HIV disease. Moreover, it appears that its downregulation is the key mechanism for the non-progression of HIV.

### *Gene set enrichment analysis (GSEA) of consistently present probe sets down-regulated in LTNP compared to the TP2 group*

To clarify the significance of the downregulated genes between LTNP and TP2, we used enrichment analysis tools to examine the most significant map folder. A summary of this type of MetaCoreTM analysis can be observed in the enrichment analysis process, as shown in **Supplementary Figure 10**. Several

significant GeneGO pathways were identified, and the genes were enriched in functional ontologies largely related to apoptosis in GeneGo maps. In the process network, confirmation of the involvement of apoptosis and protein degradation pathways was evident from the top-ranked network process (p=7.606e-10), with 35 of 166 genes enriched in this network. The GeneGo maps, GeneGo process network, and GenGo Disease underpin the importance of the apoptosis pathway being significantly enriched and positively correlated with the LTNP group and negatively correlated with the TP1 group.

Furthermore, as before, we again observed a significant enrichment of metabolic networks in LTNPs, with the 1-linoleylglycerol 3-phosphate pathway (p=1.533e-5) (**Supplementary Figure 11d**) being the top-ranked and cellular macrometabolic being the most prominent (p=6.257e-55), with 690 of 7016 genes enriched (**Supplementary Figure 11e**). The **Supplementary Table 10** shows some of the enriched genes in the cellular macrometabolic process with the respective networks and functions they share. The functional associations between the networks were highly significant. Again, the enrichment analysis of LTNP vs. TP1 and LTNP vs. TP2 contrasts for the downregulated genes clearly showed a functional involvement of metabolic networks in a variety of cellular functions, such as axon guidance, cellular organization, morphogenesis, cellular metabolic processes, electron transport, and cellular metabolic and catabolic processes. These analyses imply that downregulation is a significant feature of LTNPs.

The enrichment analysis of the pathways provided by the MetaCoreTM system also included a comparative analysis of the over-represented genes among the pathways. To visualize the individual genes represented in the process networks, Supplementary Figure 11 shows the top ten most enriched GeneGo processes for the sample set, according to their p-values for each enriched process within the network. Consistent with our observations of the downregulated genes in the LTNP vs. TP1 contrast, the LTNP vs. TP2 also showed the dominance of apoptosis-related genes enriched in the process network; pathway analysis downregulated DE genes between LTNP and TP2. Pathway analysis (**Figure 6**) showed that the most significant pathway was the role of IAP proteins in apoptosis, with significant downregulation of all the essential pro-apoptotic genes in this pathway, as shown in **Supplementary Table 11**. The TNFR1 pathway is the second most significantly downregulated pathway in LTNP compared to that in TP2 patients. This pathway is primarily involved in inflammation and apoptosis. Furthermore, between IAP apoptosis and the TBFR1 pathway (**Figure 6**), some of the most significant genes (**Supplementary Table 12**), such as caspases 3, 8, and 9; Bid; BIRC2; and FADD, overlapped between both pathways, suggesting a functional interaction between these two pathways. Thus, the downregulation of apoptotic genes, coupled with the downregulation of anti-apoptotic genes uniquely in LTNPs, is the single most significant feature of non-progressive HIV disease.

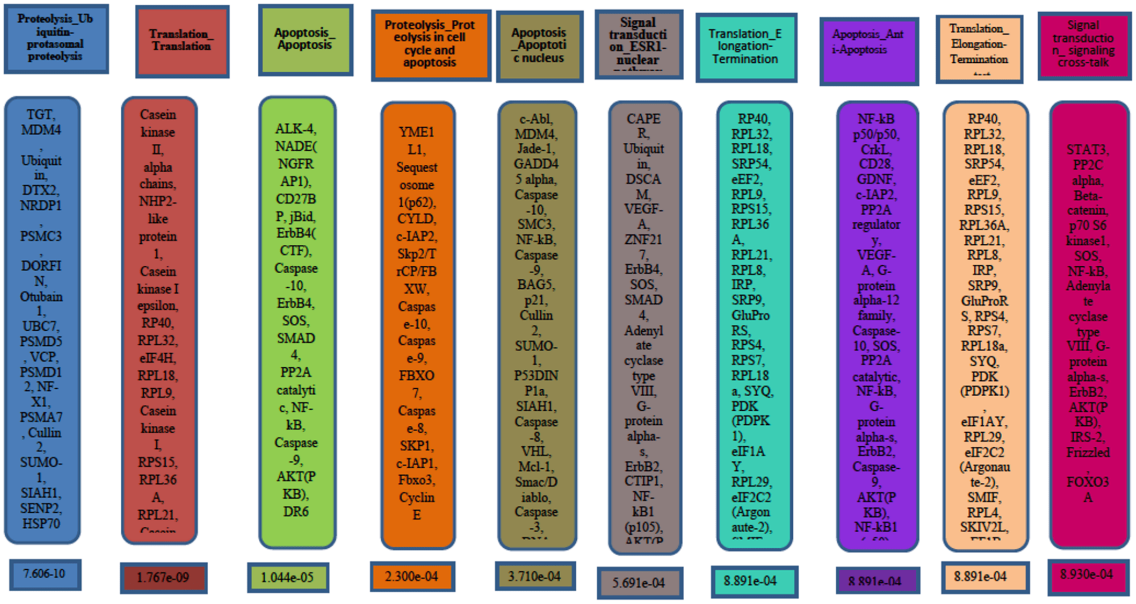



**Figure 6.** Process-network for the comparison between LTNP x TP1. Enriched genes in each process within the network are shown within the bars, along with the p values for each process network shown at the bottom of each bar.

## *Quantitative RT-PCR validation of functionally significant genes*

### *LTNP versus TP1*

To confirm the differential expression of genes from the Illumina microarray, the mRNA expression levels of 16 functionally significant genes in HIV infection were validated by quantitative PCR. To validate these genes, the mRNA from the PBMC of the same patient at both time points (TP1 and TP2), along with the therapy-naïve LTNPs, was used in the qPCR analysis. For the contrast between LTNP and TP1, we validated eight genes (OAS2, IFIT1, STAT1, CD74, FKBP3, THBS1, IFIT3, and MAPK3) shown in **Table 2** with their qPCR fold-change and function. As can be seen, these genes are involved directly or indirectly in the host's immune response to HIV. We also found that all eight genes showed consistent patterns of expression trends between microarray and qRT-PCR (**Figure 7**), which were further analyzed separately using $R^2$ correlation (**Figure 8 a-n**).


doi.org/10.32388/WGKALR

| Gene symbol | Fold change | homology |
|---|---|---|
| FKBP3 | 1.56 | Play a role in immuno-regulation and basic cellular processes involving protein folding and trafficking |
| IFIT3 | -1.59 | Interferon-induced protein with tetratricopeptide repeats 3 |
| THBS1 | -7.22 | Play roles in platelet aggregation, angiogenesis, and tumorigenesis |
| CD74 | 1.79 | Involved in the formation and transport of MHC class II protein |
| STAT1 | -2.28 | Involved in up-regulating genes due to a signal by either type I, type II or type III interferons |
| OAS2 | -2.37 | Involved in the innate immune response to viral infection |
| IFIT1 | -48.11 | Interferon-induced protein with tetratricopeptide repeats 1 |
| MAPK3 | 12.53 | Act in a signalling cascade that regulates various cellular processes such as proliferation, differentiation, and cell cycle progression in response to a variety of extracellular signals. |

**Table 2.** q-RTPCR validation of selected genes comparing LTNP vs newly acquired viraemic HIV infection

**FKBP3 (FK506-binding protein 3); IFIT3 (Interferon-induced protein with tetratricopeptide repeats 3); THBS1 (Thrombospondin 1); CD74 (LA-DR antigens-associated invariant chain or Cluster of Differentiation 74); STAT1 (Signal transducer and activator of transcription 1); OAS2 (2'-5'-oligoadenylate synthetase 2); IFIT1 (Interferon-induced protein with tetratricopeptide repeats 1); MAPK3 (Mitogen-activated protein kinase 3).*

IFN-related genes constituted half of the eight validated genes (OAS2, IFIT1, STAT1, and IFIT3), downregulated in both microarray and qRT-PCR (**Figure 7**). Three of the remaining four non-IFN-induced genes (CD74, FKBP3, and MAPK3) were upregulated, except for THBS1, which showed consistent downregulation.

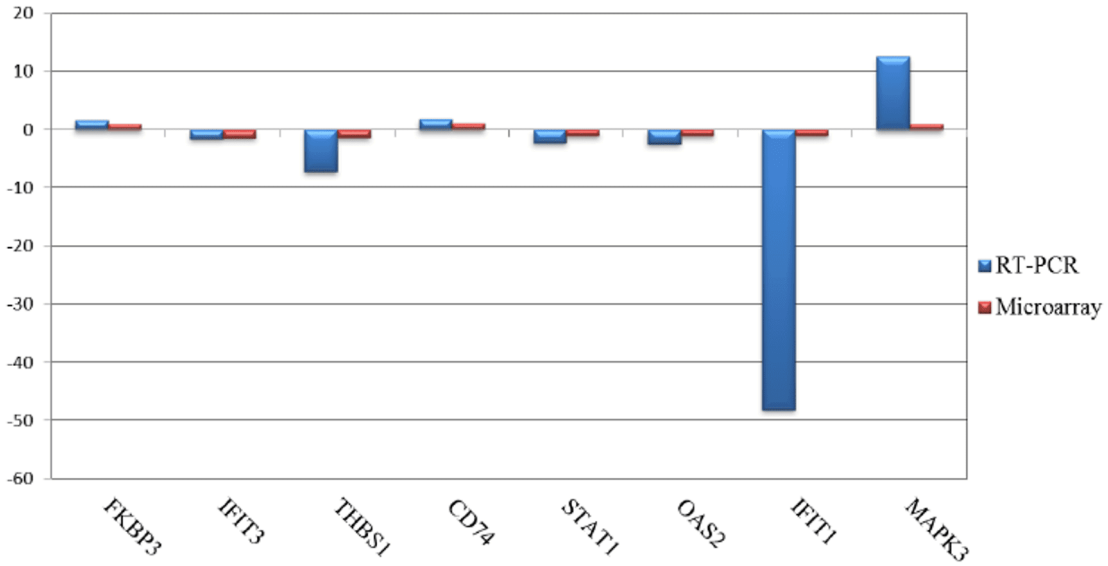
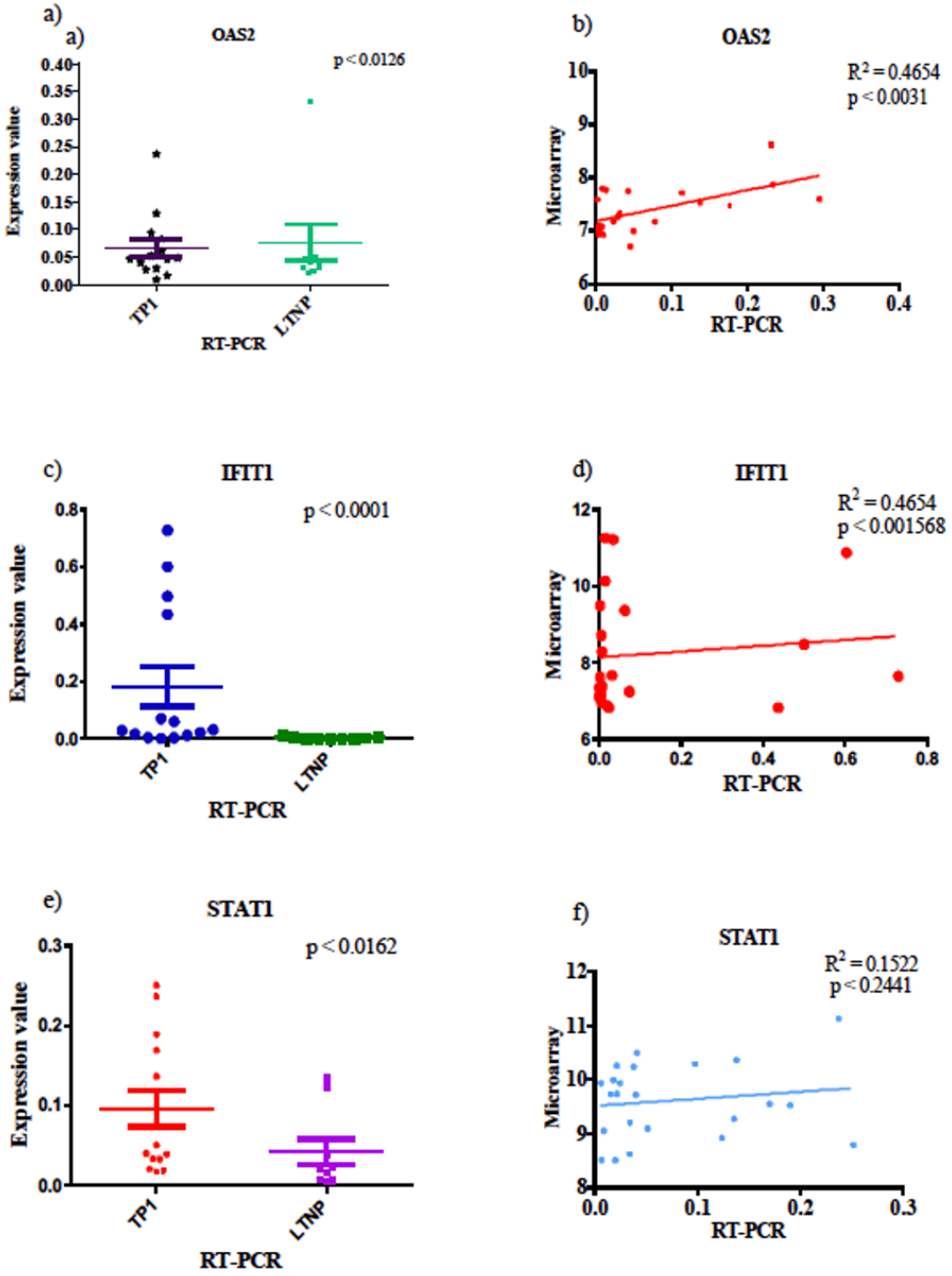



**Figure 7.** Correlation between RT-PCR and Microarray fold-changes of the 8 DE genes significant for the LTNP and TP1 comparison, involving a combination of both up- and -down-regulated genes. Fold-change is calculated simply as the ratio of final value to the initial value.

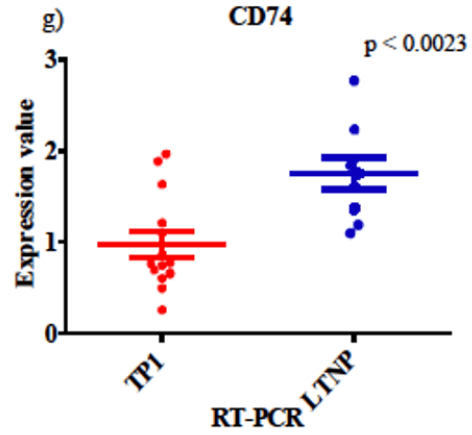
Although five of the eight genes validated by qRT-PCR showed good statistical significance, three did not achieve the same level of significance (Figures 8 a, c,e, g, i, k,m, n). However, when the same genes were analyzed for correlation between qRT-PCR and microarray, there were also five of eight genes significant (**Figures 8 b, d,f, h,j, l,n, o**), but these genes were different between qRT-PCR and correlation analysis. For instance, IFIT1 was highly significant in q-RTPCR but not in the correlation analysis (**Figures 8 c and d**). Similar results were observed for STAT1 and FKBP3, which were significant in qRT-PCR, but not in correlation analysis (**Figure 8 e, f, i, and j**). Thrombospondin (THBS1) was the most significant of eight genes (Figure 8 k and l), based on qRT-PCR ($p=0.003759$) and correlation analysis ($R^2=0.9361$; $p=0.001$).

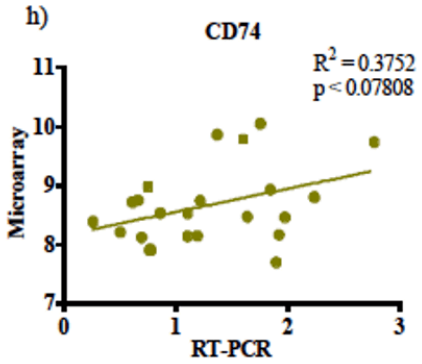

doi.org/10.32388/WGKALR

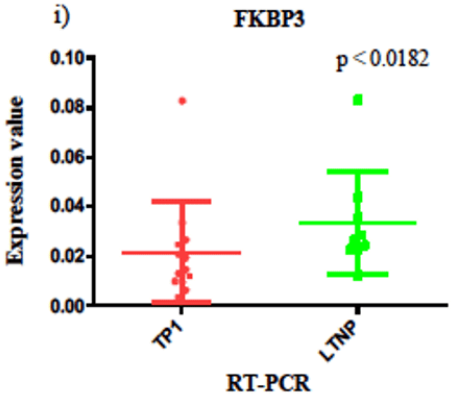


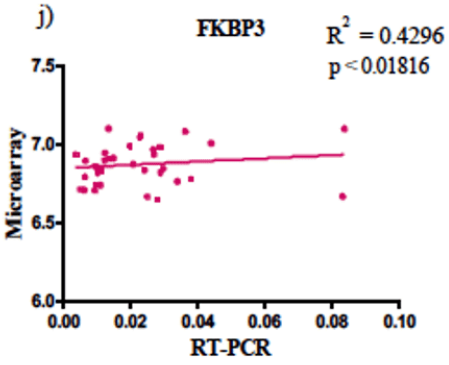


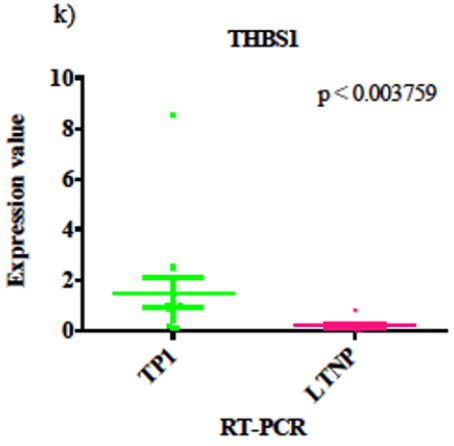
doi.org/10.32388/WGKALR

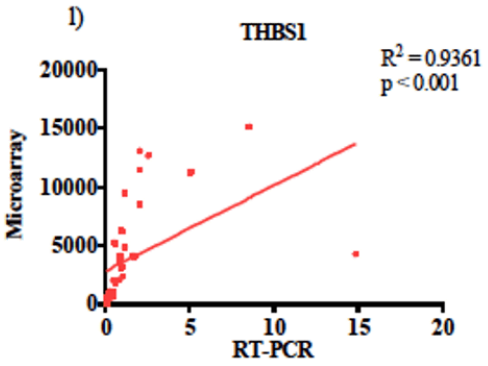




doi.org/10.32388/WGKALR

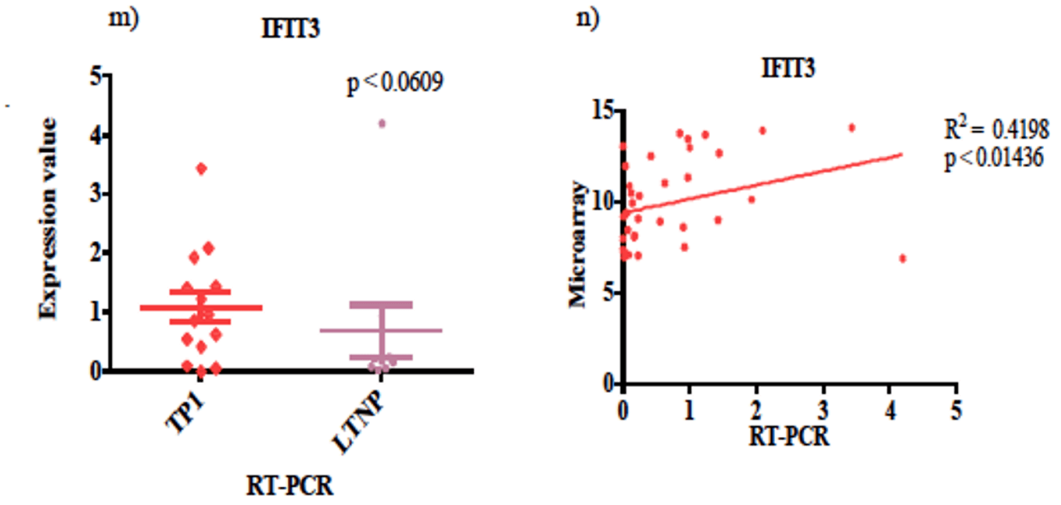



**Figure 8.** Association of gene expression by real-time PCR of the DE found significant in our analysis (a, c, e, g, i, k, m). Representative scatterplots with linear regression analysis confirm the association between expression levels between microarray and q-RTPCR (figures b, d, f, h, j, l, n) p- values for the expression between the different group were measured using the GraphPad Prism®software and also correlation values.

In the following sections, we analyzed the functional and biological significance of THBS1 in HIV infection, comparing viremic (newly diagnosed untreated patients) and aviremic patients (natural viral control in LTNP and HARRT-treated patients).

### *LTNP versus TP2*

Next, for the comparison between TP2 and LTNP, we validated 10 genes by qRT-PCR (FKBP3, IFIT3, THBS1, CD47, CD46, BIRC2, FADD, CASP3, OAS2, and MAPK3) (**Table 3**). **Table 3** also shows the functional significance of each gene included in the qRT-PCR analysis. It should be noted that FKBP3, MAPK3, and THBS1 were significant in both LTNP vs. TP1 and LTNP vs. TP2 comparisons. Therefore, these genes were included in both comparisons for qRT-PCR validation. Of the ten genes selected for qRT-PCR validation, five were upregulated, and five were downregulated in the microarray expression analysis between LTNP and TP2. As explained previously, five of the ten genes were upregulated, and five were downregulated in the microarray gene expression experiments. In qRT-PCR validation, the same occurred, with 5 of 10 downregulated genes (BIRC2, FADD, CASP3, IFIT3, and THBS1) and five upregulated genes (FKBP3, CD46,

CD47, OAS2, and MAPK3) (**Figure 9**). Thus, all ten genes showed consistent trends in expression between microarray and qRT-PCR, underpinning their functional significance.

| Gene symbol | Fold change | Homology |
|---|---|---|
| FKBP3 | 2.17 | Play a role in immunoregulation and basic cellular processes involving protein folding and trafficking |
| IFIT3 | -2.66 | Interferon-induced protein with tetratricopeptide repeats 3 |
| THBS1 | -10.74 | Play roles in platelet aggregation, angiogenesis, and tumorigenesis |
| CD47 | 1.89 | functions as a "don't eat me" signal for phagocytic cells |
| CD46 | 2.58 | An inhibitory complement receptor |
| BIRC2 | -1.20 | A member of the Inhibitor of Apoptosis family that inhibit apoptosis by interfering with the activation of caspases |
| FADD | -1.79 | An adaptor molecule that bridges the Fas-receptor, to caspase-8 through its death domain to form the death-inducing signalling complex (DISC) during apoptosis. |
| CASP3 | -3.02 | Caspase protein that interacts with caspase 8 and caspase 9. |
| OAS2 | 1.68 | Involved in the innate immune response to viral infection |
| MAPK3 | 10.65 | Act in a signalling cascade that regulates various cellular processes such as proliferation, differentiation, and cell cycle progression in response to a variety of extracellular signals. |

**Table 3.** q-RTPCR validation of selected genes comparing LTNP vs. ART-treated patients.

**Full gene annotations: *FKBP3 (FK506-binding protein 3); IFIT3 (Interferon-induced protein with tetratricopeptide repeats 3); THBS1 (Thrombospondin 1); MAPK3 (Mitogen-activated protein kinase 3); CD47 (Cluster of Differentiation 47); CD46 (Cluster of Differentiation 46); BIRC2 (Baculoviral IAP repeat containing 2); FADD (Fas(TNFRSF6)-associated via death domain); CASP3 (Caspase 3, apoptosis-related cysteine peptidase); OAS2 (2′-5′-oligoadenylate synthetase 2); IFIT1 (Interferon-induced protein with tetratricopeptide repeats 1); MAPK3 (Mitogen-activated protein kinase 3).*

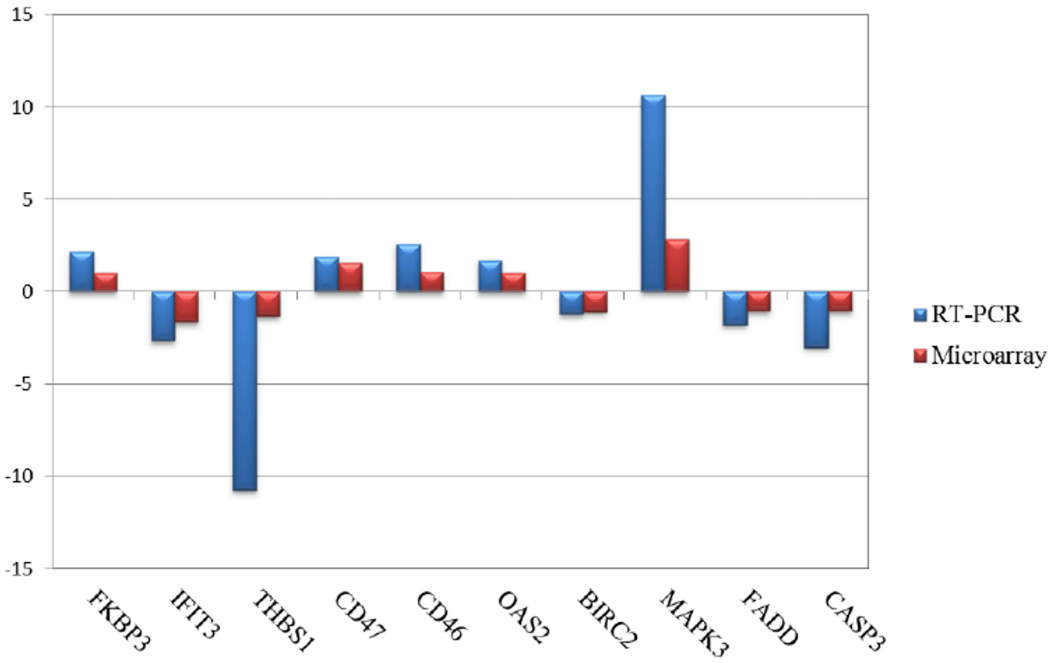
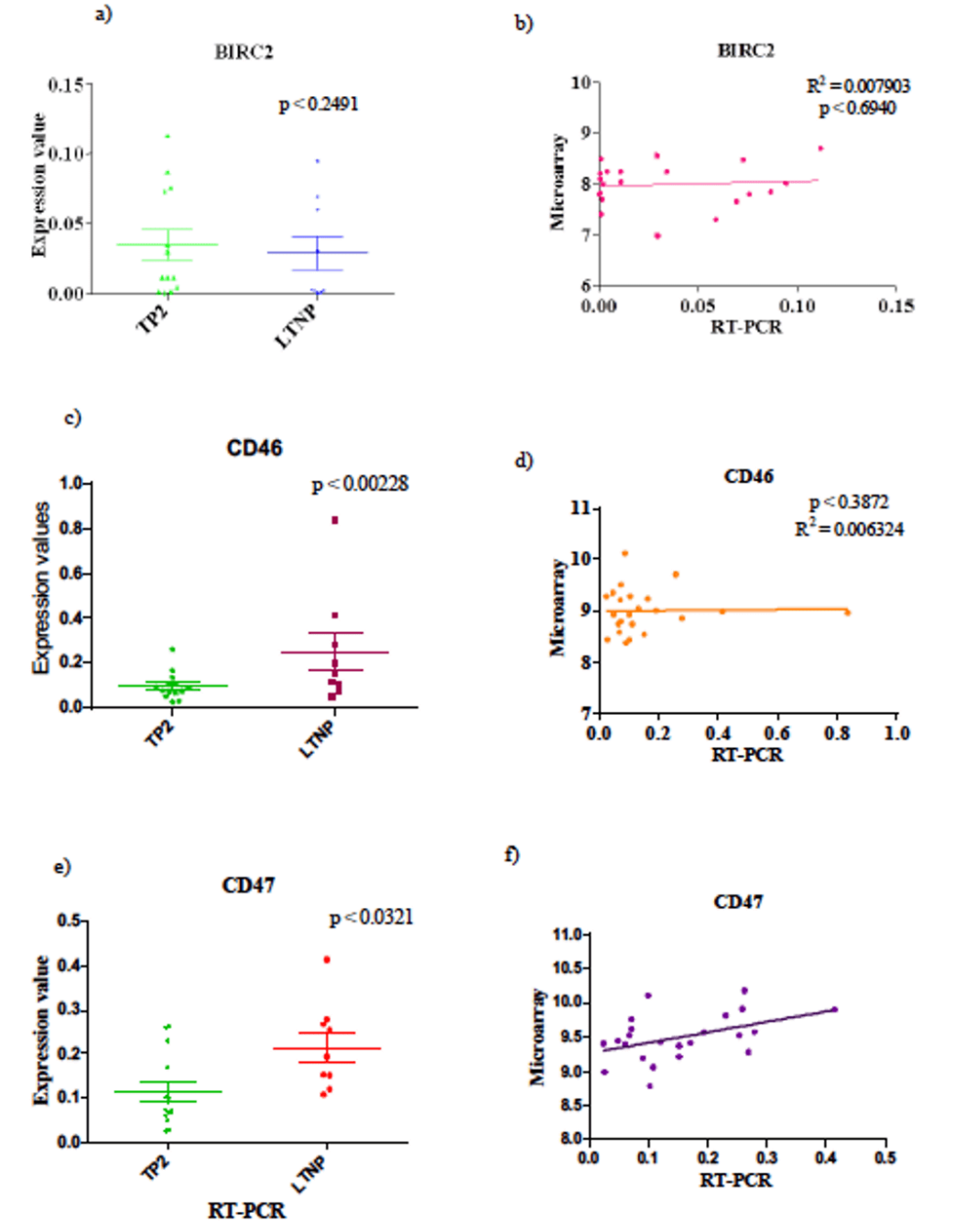



**Figure 9.** Correlation between RT-PCR and Microarray fold-changes. Fold-change is calculated simply as the ratio of the final value to the initial value.

We further visualized this dataset with statistical rigor by first analyzing the significance of the expression of individual genes in the qRT-PCR assay, followed by the calculation of R-squared correlation and correlation values (as explained above) to delineate the correlation between microarray and qRT-PCR data (**Figure 10 a-t).**

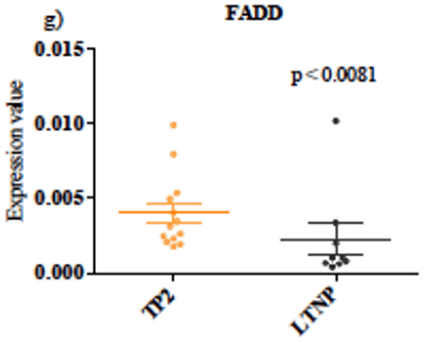
Of the ten genes included in the functional validation, CD46, CD47, FKBP3, THBS1, and CASP3 were statistically significant in both qRT-PCR (**Figure 10 c, e, I, k, and o**), with THBS1 being the most significant in both q-RTPCR (p=0.0011) and correlation analysis ($R^2$=0.942; p=2.65e-08) (**Figure 10 k and l**), along with integrin-associated protein (IAP or CD47), which is a receptor for thrombospondin family members ($R^2$=0.375; p=0.0321) (**Figure 10 e-f**). This was also consistent with our observations of its validity in the LTNP versus TP1 contrast. BIRC2 and MAPK3 (**Figure 10 a, b, m, and n**) did not achieve statistical significance, especially MAPK3 ($R^2$=0.5476). The FADD gene was only significant in qRT-PCR (p=0.0081) but not in the correlation analysis (**Fig 10 g and h**). The genes that showed perfect concordance between qRT-PCR and microarray analysis were CD46, CD47, FKBP3, THBS1, CASP3, IFIT3, and OAS2 (**Figure 10 c, d,**

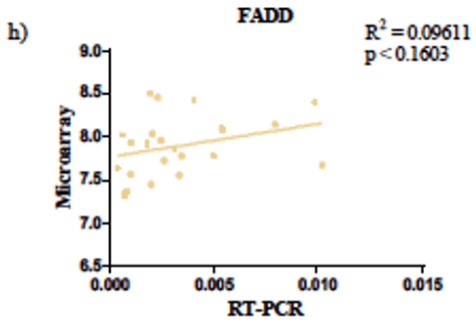
**e, f, i, j, k, l, o, p, q, r, s, and t)**. Thus, seven of the ten genes showed perfect statistical concordance between qRT-PCR and microarray, suggesting their strong functional significance in the LTNP vs. TP2 contrast.

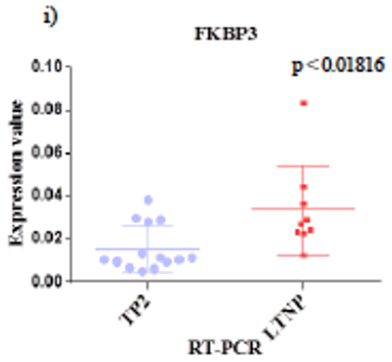


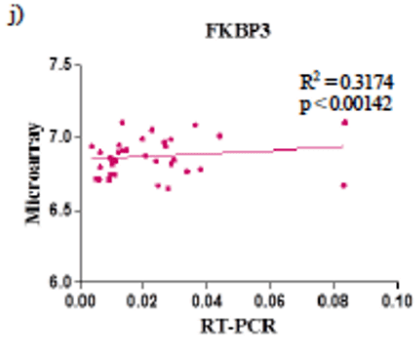


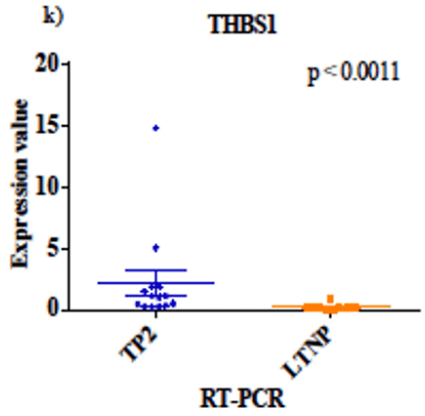
doi.org/10.32388/WGKALR

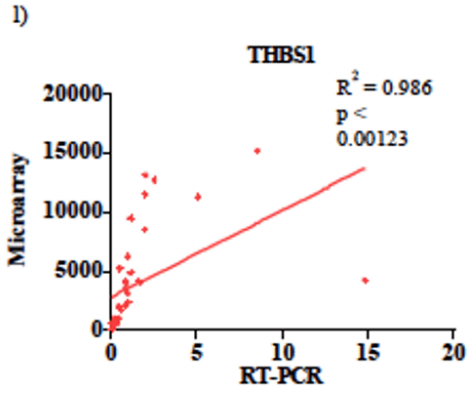




doi.org/10.32388/WGKALR

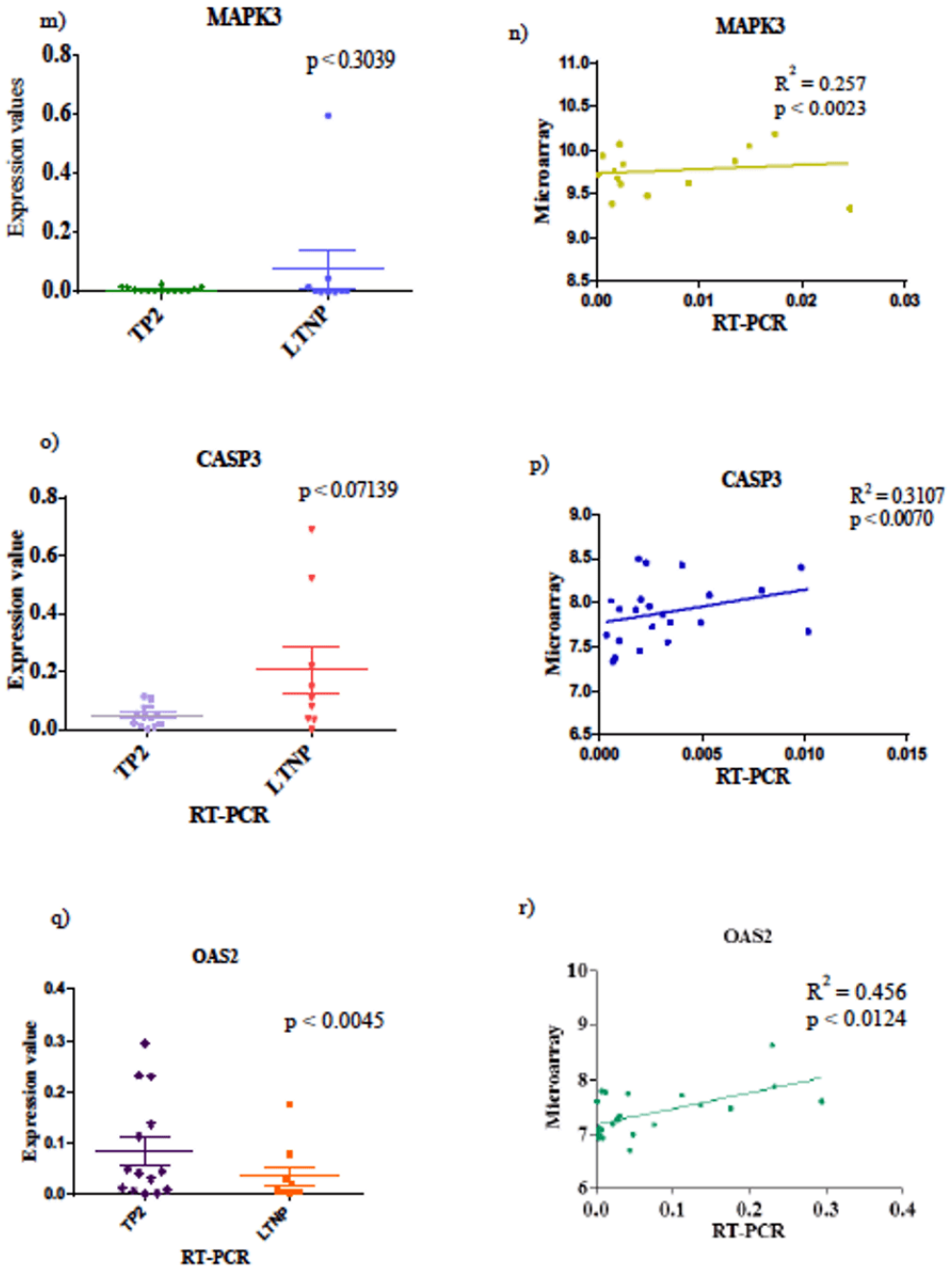



**Figure 10.** Association of gene expression by real-time PCR of the DE genes found significant in our analysis (a, c, e, g, i, k, m, o,). Representative scatter plots with linear regression analysis confirm the association of expression levels between microarray and q-RTPCR (figures b, d, f, h, j, l, n, p,). p-values for the expression between the different group were measured using the GraphPad Prism®software.


doi.org/10.32388/WGKALR

*Functional validation using Confocal and Delta Vision Microscopic analysis for the cellular localization of THBS1or TSP1 protein in PBMC and monocytes*

For biological analyses using Confocal/Delta Vision microscopy and flow cytometry, we examined fresh blood samples from five therapy-naïve patients with clinical profiles resembling the TP1 group, all with detectable plasma viraemia (>400 copies/mL), and five patients on HAART with undetectable plasma viremia (<50 copies/mL) to represent the TP2 group. These were compared to frozen PBMC from LTNP, and HIV-negative healthy controls.

Using confocal and Delta Vision microscopy, our main objective was to analyze the cellular localization of THBS1 in whole PBMC, and bead-separated monocytes derived from healthy donors, HIV viremic patients, and HIV+ LTNPs. Initially, we stained whole PBMCs for different patient groups using Deltavision microscopy following the manufacturer's protocol. The **Figure 11** shows the localization of the THBS1 protein in the membrane and perinuclear space of the cells in patients with high viral load (a), patients with no HIV infection (b), and LTNPs (c).

The protocol was the same as that used for Deltavision microscopy, but only monocytes were used because of the excellent staining of thrombospondin. Thrombospondin is a platelet-associated protein; however, its expression in diverse blood leukocytes remains unclear. As shown in **Figure 11**, we analyzed patients with detectable viral loads: HIV-negative individuals and HIV LTNPs. These data show the staining of monocytes in these three groups (**Figure 11 panels a, d, g**), followed by DAPI staining, which shows evidence of the membrane and perinuclear localization of THBS1 in monocytes (**Figure 11 panels b, e, h**). Staining of the whole PBMC for the same groups was also performed, and the results were similar to those of monocytes (**Figure 12 panels a-i**). Because of the specific staining of monocytes in all three groups, we chose to quantitatively analyze THBS1 in monocytes in all three patient groups using flow cytometry.

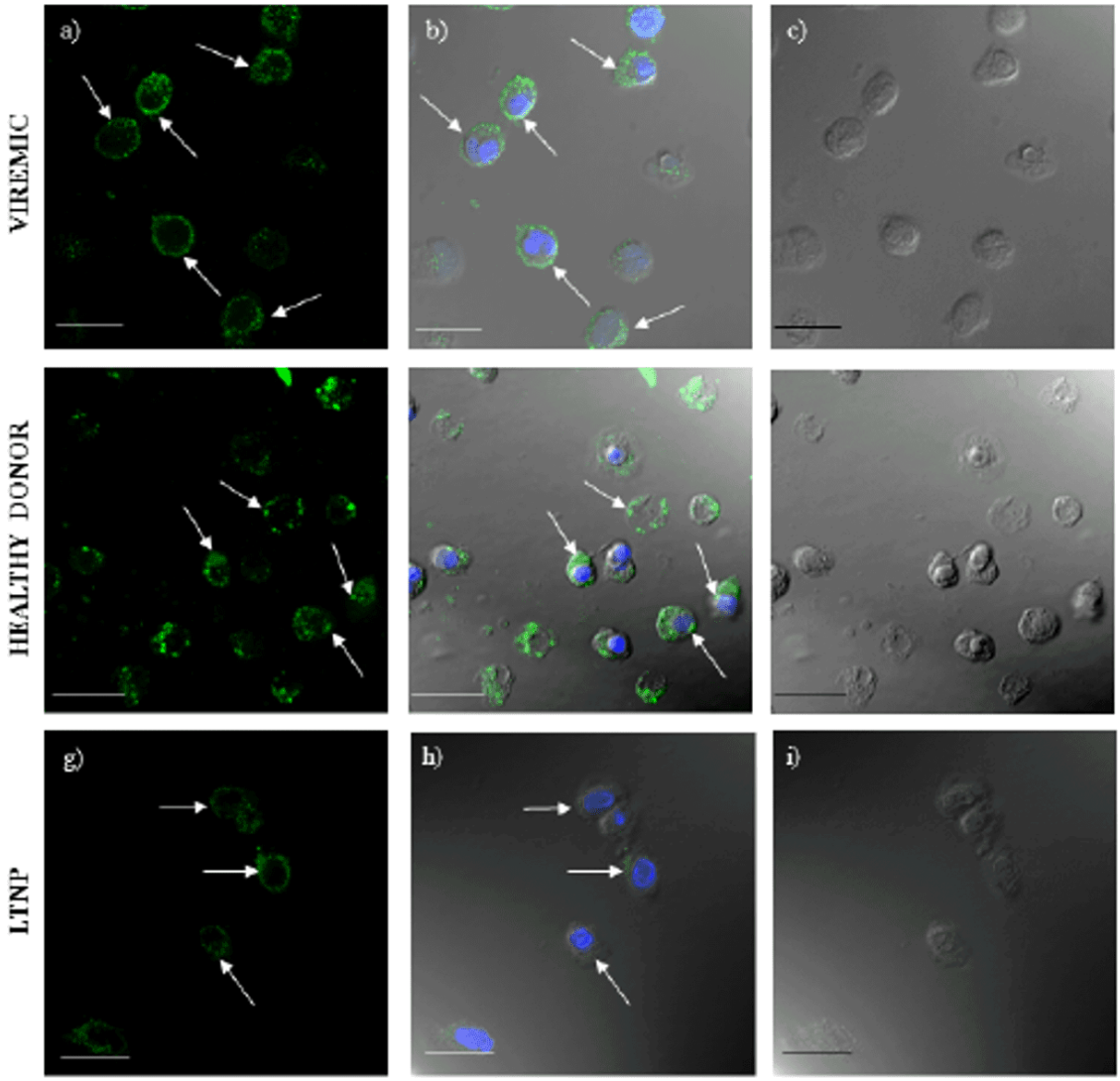



**Figure 11.** Confocal micrographs showing localization of the THBS1 protein (arrows) in CD14 positive monocytes derived from health donor (a-c), HIV positive viremic (d-f) and LTNP (g-i) patients. THBS1 was found distributed in a punctate pattern throughout the perinuclear cytoplasm and plasma membrane in all three patient cells. Panels a, d, g show anti-THBS1 stained with A488 (green). Panels b, e, h show overlay of THBS1 (green) and DAPI (blue) (nucleus) over corresponding DIC image. Panels c, f, i show corresponding DIC images. Bars, 5 μm.

### *Flow cytometry (FACS) analysis of the THBS1 expression between the different patient groups*

Flow cytometry was used to quantitatively analyze thrombospondin in monocytes derived from HIV patients with detectable viral loads, HIV-negative individuals, and HIV+ LTNPs. We used whole PBMCs from the patient groups (HIV+ patients with below detectable virus in plasma, HIV+ individuals with detectable virus and LTNPs) and the gated on CD14 and CD16 monocytes.

The main objective of these experiments was to delineate the functional relevance of THBS1 at the protein level in monocytes and identify its correlation with plasma viremia. Whole frozen PBMCs were used in each group for these experiments. We aimed to determine the cell population with increased protein expression. Therefore, using a flow cytometer (BD FACSCanto II), we gated different cell populations for each group and found that THBS1 was mostly stained in the monocytes (**Supplementary Figure 14**). Each coloured line represents one group, demonstrating THBS1 is highly expressed in monocytes. Therefore, we chose to concentrate our dataset in the monocyte population.

For the remaining experiments, we focused only on CD14+ cells. The cells were frozen and thawed according to the protocol. Initially, the PBMC were stained LIVE/DEAD® Fixable Dead Cell Stain Kit from Invitrogen® to check for cell viability since our samples were frozen. First, we stained the samples with DAPI from the kit to check viability and then stained them with CD14, CD16, and THBS1 antibodies. Using flow cytometry (BD FACSCanto II), the DAPI+ cell population from all groups was gated first (P3 = **supplementary Figure 15 b**) from the monocyte population (P1 = **supplementary Figure 15 a**), and then CD14+ (P4), CD16+, and THBS1+ cell populations were analyzed subsequently. We included five patients with BD viral load, five with detectable VL before HAART treatment, five LTNPs, and five HIV-negative controls (**Supplementary Table 13**). **Supplementary Table 14** shows the percentages of each quadrate gate for the expression of THBS1 and CD16.

The expression of THBS1 on thawed PBMC was determined using flow cytometry. The proportion of cells positive for THBS1 is represented as % PBMC, and the density of THBS1 molecules on cells is represented by the Mean Fluorescence Intensity (MFI) (**Table 4**). THBS1 was highly expressed in patients with high viral loads, significantly different from that in patients with LTNP and healthy controls. After HAART-associated viral suppression, the expression of THBS1 in the LTNP and healthy controls was insignificant. These data suggested that thrombospondin is a marker of uncontrolled viral replication and is not upregulated in patients with long-term suppression of viral replication.

| Participant group | THBS1 expression (% PBMC) | THBS1 cell density (MFI) | p-value (vs. LTNP) | p-value (vs. controls) |
|---|---|---|---|---|
| Pre-ART (viremic) | 48.3 | 1175 | 0.00376 | 0.00376 |
| Post-ART (VL <400) | 18.5 | 250.4 | 0.0734 | 0.298 |
| LTNP (low VL) | 3.75 | 161 | | |
| Healthy controls | 2.35 | 199 | | |

**Table 4.** Flow cytometric analysis of THBS1 expression on PBMC from newly diagnosed HIV-positive patients before and after treatment, compared to HIV-positive LTNP and healthy controls.

*MFI (Mean Fluorescence Intensity).*

## 3. Discussion

The underlying genomic basis for elite viral control remains obscure. Previous gene microarray studies concerning HIV disease have used specific cell types, such as CD8+T cells, macrophages, cell lines, and dendritic cells, to understand gene regulation during HIV infection [15][13]. Our study is the first to dissect the genomic basis of drug-mediated control of HIV, as opposed to natural control in LTNPs. This was achieved by comparing genome-wide transcriptomes encompassing all 25,000 human genes (48,000 transcripts or probes) to define critical molecular differences between these two groups at various immune functional levels. Current microarray platforms allow high-throughput genome-wide transcriptome analysis, making it possible to understand the fundamental cellular, molecular, and genetic mechanisms underlying non-progressive HIV diseases [16]. This study used whole primary uncultured peripheral blood mononuclear cells (PBMCs) to represent a holistic genome-wide expression. Second, since we used two-time points and archival samples from chronically infected LTNPs and other HIV groups, our work was only possible on whole frozen PBMCs, from which separation of individual cell types was experimentally not feasible due to quantity of cells.

Employing the vast scientific knowledge of the immune system in humans and understanding the natural immune control of HIV infection will enable its translation into the development of effective immunotherapies, leading to new preventative strategies and possible cures. This has been the focus of several global studies. Not only is there a renewed interest in HIV persistence in the face of HAART, but there is also a renewed focus on the co-existence of host and virus in the absence of therapeutic drugs, as has been observed in long-term non-progressing individuals or elite controllers, who represent <1% of HIV-1-infected individuals [17][18]. LTNPs maintain CD4+T-cell counts >500, with undetectable levels of HIV RNA (<50 copies/mL plasma) in the absence of therapy for >10 years. However, as time progresses, very few of these can be defined as elite controllers [9][11][19]. Such individuals hold clues for HIV cure and novel immunological factors for potential immunotherapies.

Microarray gene expression profiling has been applied to subcategorize complex diseases and predict treatment outcomes and disease progression using “gene signatures,” representing differentially expressed genes that correlate with the disease phenotype [20]. Together, these technologies, combined with web-based computational platforms for data analysis, not only assist in the visualization of the dataset but also provide a clear understanding of the HIV-host interactions and identify host targets for developing anti-HIV therapeutics [21].

Therefore, in this study, using the Illumina Human-12 v3 Expression BeadChip encompassing all 25,000 human genes (>48,000 gene transcripts or probes), we performed a genome-wide analysis of nine HIV+, therapy-naïve long-term non-progressors to define the genomic basis of natural control of HIV in these individuals. Based on the whole genome transcriptome, we compared HIV+ LTNPs to HIV patients off drugs while experiencing detectable plasma viremia (TP1) and patients on HAART with below detectable viremia (<40 copies of HIV RNA/mL plasma) (TP2). These analyses provide deeper comparative insights into the genes and pathways differentially expressed in each group, thereby explaining the underlying disease stages during HIV infection. This is the first comparative analysis based on the whole human genome, demonstrating the genomic basis of the natural control of HIV as opposed to drug-mediated control.

We detected more than a thousand significant differentially expressed genes between the patient groups, LTNP and TP1, and between LTNP and TP2. Through these comparisons, we demonstrated that the immune system response pathway was downregulated in patients with detectable viremia in the absence of HAART (TP1) and in patients with BDL on HAART (TP2), compared with the therapy naïve-LTNPs. This is in sharp contrast to the findings of the TP1 and TP2 comparisons, where the immune response system

was upregulated in TP1. When LTNPs were compared against TP1, the most significant pathway upregulated in the LTNP group was the B-cell receptor (BCR) signalling pathway (p=1.03e-07), which was also confirmed by enrichment and pathway analyses. This signalling pathway involves many distinct processes, including survival, tolerance or apoptosis, cytoskeletal changes, DNA damage and repair, proliferation, and differentiation into antibody-producing or memory B cells. Eleven of the 54 DE genes were enriched in this pathway, all recording statistical significance, as apparent in the p-values. Genes in this pathway are related to the activation of the immune response of antigen-presenting cells, such as CD22, MAPK1/3, CD79, BtK, Syk, SHIP, and HRAS.

HIV infection induces chronic immune activation, which limits the capacity of the immune system to respond to other opportunistic infections. HIV specific targets HIV-specific CD4+ helper T-cells, resulting in sub-optimal co-stimulation of both HIV-specific B-cell and cytotoxic T-cell response maturation [22]. This explains why the virus causes the deregulation of genes related to the immune response (H-Ras, SHP-1, CDC42, STAT1, CD79 complex, etc.), as observed in our study. Without helper T-cells, it is impossible for the host to produce antibodies effectively, nor does it eliminate the infected cells concealing HIV, thereby resulting in the collapse of the immune system, leading to the progression of HIV disease.

Expression of B cell receptor (BCR) pathway constituents are significantly altered during progression of HIV infection, and we demonstrated similar changes before and after treatment of newly acquired HIV infection. A large majority of BRC pathway permutations are directly associated with high levels of HIV plasma viremia [23]. Clearance of HIV plasma viremia by antiretroviral therapy leads to normalized B cell function, as evidenced by reductions in polyclonal and HIV-specific Ig levels, restoration of APC function, responsiveness to CD4+T cell help, and normalization of expression of B cell markers of activation [24][25]. Our results corroborated these studies, showing that the expression of this receptor is dysregulated in viraemic HIV-infected patients. The aberrant immune hyperactivation induced by HIV replication represents a systemic effect, as demonstrated by the profound effects of HIV on B cells and CD8+T cells, two populations of lymphocytes that are not direct targets for productive HIV replication [23][26][27]. Viral replication-specific activation of the BCR pathway alters antigen processing and presentation, immunoglobulin production, and interferon-mediated antiviral effects, because expression of this receptor is deregulated in HIV-infected patients [28][29]. This finding is consistent with our results. Furthermore, HIV-induced immune activation of B cells is thought to contribute to the increased frequency of B-cell malignancies observed in HIV-infected individuals, which was especially observed before the widespread use of effective ART [25][30], meaning that the virus has a direct effect on B-cell function. More importantly,

the up regulation of components of the BCR pathway in LTNPs may result in a vigorous immune response against HIV, which may play a pivotal role in prolonging the asymptomatic phase of the infection [30][31], as shown by the results of our study. We found significant upregulation of CD22 (p=0.0001) in the BCR pathway in the LTNP group. This is functionally relevant because CD22 acts as a regulatory molecule to prevent over-activation of immune responses [14]. To emphasize the critical role of the BCR pathway in non-progressive HIV disease, we also observed a strong association with the CD79 complex (p=0.00015), which forms a complex with the B-cell receptor (BCR) for signal transduction following antigen recognition. Our study, therefore, provides a new understanding of the extent to which HIV replication may perturb the BCR pathway. In contrast, preserved function may contribute to long-term natural control of HIV.

The comparison of upregulated and downregulated genes has provided functional insights into the roles of these genes in HIV disease progression and non-progression. The comparison of downregulated genes between LTNP and TP1 showed that apoptosis was a standout feature, with evidence for coordinated downregulation in the LTNPs (p=7.04e-09), with six of the top 10 pathways in apoptosis and immune response. The most significantly downregulated pathway was pigment epithelium-derived factor (PEDF) (p=2.30e-06), followed by the apoptosis survival TNRF1 signalling pathway (p=1.45e-05) in the inflammatory response map folder (p=5.33e-06). PEDF is also known as Serpin F1 (SERPINF1), a multifunctional secreted protein with anti-angiogenic, anti-tumorigenic, and neurotrophic functions. This pathway includes genes related to cellular responses to virus infection, such as NF-kB and anti-apoptotic genes (SOD2, c-IAP-1, and 2), which are very significant in this pathway. A more profound view of the PEDF and TNFR1 signalling pathways showed an intense clustering of genes related to the family of anti-apoptotic genes or the apoptosis inhibitor family of genes (cIAP-1 and cIAP-2, SOD2, Diablo-smac) in and around the anti-apoptotic part of the pathway, along with their functionally associated genes in the anti-apoptosis network (NF-κB p50, p65, and GDNF). The similarities between these pathways suggest a considerable functional interaction between them.

Non-progressors are characterised by low level apoptosis and high expression levels of anti-apoptotic genes, resulting in a reduced rate of CD4+ loss [32]. In contrast, our viraemic patients (TP1) had upregulated CD38, which is associated with activation and cell death, and a critical marker on CD8+ T cells that differentiates detectable from undetectable viral replication and progressive from non-progressive disease [33][34]. Genes activated for positive apoptosis regulation in progressors agree with other activated functions in this group, such as the cell cycle and DNA replication, which are implicated in several

pathways linked to cellular death. T-cell apoptosis is thought to be one strategy by which HIV-1 evades host immune supervision, which is intact because of tight regulation by the downregulation of anti-apoptotic genes, as seen in our study. Activation-induced apoptosis is a mechanism for the loss of T cells during continuous stimulation of the immune system and has been proposed as an important factor in the pathogenesis of HIV-1 infection [35]. Apoptosis is linked to several events typical of HIV infection, including cell activation and cytopathic effects of HIV [36]. We found that serpin F1 (SERPINF1) was downregulated in the LTNP group. This gene has anti-angiogenic actions, protects cerebellar granule cells from apoptosis, and induces apoptosis in endothelial cells [37]. Gene profile expression studies in cell lines infected in vitro with HIV showed increased expression of pro-apoptotic genes [38] and serpin 1, which agrees with our findings, with their upregulation during viremia.

AIDS pathogenesis could also be explained by immune dysregulation involving proinflammatory cytokines, especially tumor necrosis factor (TNF) α (TNF). These studies confirm our results that in patients that are in the initial phase of the infection (TP1), apoptosis is upregulated and contributes to HIV immune pathogenesis. In contrast, LTNPs can prevent the course of this rapidly evolving infection, and some believe that it is directly related to the high levels of anti-apoptotic genes in peripheral lymphocytes compared to normal patients who progress to AIDS [39].

HIV-infected individuals undergoing treatment with a combination of antiretroviral drugs show decreased FAS and FAS ligand (FAS-L), inhibition of apoptosis induced by FAS or by TNFR-1 and TNFR-2, and inhibited expression of TNF-related apoptosis-inducing ligand (TRAIL), as well as increased BCL-2 expression, in response to treatment [40]. TNF is secreted by activated macrophages and lymphocytes and induces diverse responses, including inflammation and apoptosis. It can be directly or indirectly involved in the modulation of T cell apoptosis via members of the TNFR superfamily, such as TNFR1, TNFR2, and Fas, or it can be responsible for stimulating HIV-1 replication in infected cells [41][42][43]. HIV-1 proteins target the TNFR signalling pathway, modulating gene expression, especially HIV long terminal repeat (LTR) stimulation and T cell apoptosis, leading to immune suppression and formation of viral reservoirs during HIV infection [44]. LTNPs generally display lower frequencies of apoptotic T cells and lymphocytes expressing FAS and FASL than AIDS patients [39]. These findings are in perfect agreement with those of this study. Our results showed that apoptosis was upregulated in patients receiving therapy compared to those who naturally controlled the disease. It has also been shown that upregulation of the physiological mechanisms controlling peripheral CD4 T-cell deletion depends on the expression of a family of ligands (CD95-ligand [CD95-L], tumor necrosis factor [TNF], tumor necrosis factor-related apoptosis-inducing

ligand [TRAIL]) and death receptors (CD95, TNFR1, and TNFR2) that mediate apoptosis in susceptible cells [45]. These results concur with our findings. Overall, there appears to be a block to apoptosis in LTNPs, and the downregulation of apoptotic pathways may be a significant mechanism that confers durable protection in conjunction with strong immune responses in natural controllers.

Furthermore, the transcriptomic differences that segregated LTNPs from the TP2 group also yielded vital information on the natural control of viremia, as opposed to the drug-mediated control of HIV. Based on the below-detectable levels of viremia in both groups, we hypothesized the possibility of considerable genomic overlap between these two groups. However, genomic comparisons between these groups revealed that natural control of viremia in LTNPs is guided by a distinct set of immunological and genetic factors compared to drug-mediated control in HIV patients. Even more surprising was that although TP1 and TP2 were transcriptomically separated based on the DE genes, the parallel comparison of TP1 and TP2 against the LTNPs showed very few transcriptomic differences separating TP1 and TP2 from each other. This panoramic view of host gene modulation was only visible when LTNPs were included as a comparator, suggesting that1. LTNPs are a distinct group: 2. Immunological deficits in HIV viremic patients off therapy continue to prevail even in the presence of HAART, possibly because of suboptimal immune reconstitution; and 3. There is a genomic basis for viral control in natural controllers, which is entirely different from the suppressed viremia during HAART. Overall, these data reinforce that even though the virus can be controlled to BDL levels with HAART, immune restoration remains a considerable challenge in HAART-treated patients, making LTNPs a distinct group. This is the first study to demonstrate this aspect of HIV pathogenesis through panoramic visualization of host genome expression. These inter-group transcriptomic comparisons have provided a more profound transcriptomic view of the host genes/pathways that guide the natural control of HIV in vivo and the underlying immunological deficits in HIV patients during suppressive HAART.

Retroviruses such as HIV exploit the cytoskeletal network to facilitate viral infection and dissemination. It is activated by growth factors or phorbol esters in specific cell types, such as macrophages and epithelial cells, and operates constitutively in dendritic cells [46][47]. Our results agree with these studies, showing that actin-dependent filopodial bridges and intercellular membrane nanotubes mediate cell-to-cell transmission of retroviruses. Thus, their downregulation in viremic patients and upregulation in LTNPs are directly related to HIV disease staging and possibly plasma viremia [46][47].

It has been determined that a decrease in HIV-1 virus evolution correlates with weaker viral fitness and the inability to evade the host immune system. Some studies have shown that viral strains derived from

LTNPs are less evolved and thus less capable of evading the host immunological response than viral strains derived from HIV+ progressors. However, the components of the host immune system that maintain the virus in check in LTNPs and their genomic basis are unknown. Our results were confirmed by studies that showed that in PBMC from HIV-1-infected subjects, in all stages of the disease, there is a decrease in the STAT gene family, probably as a consequence of chronic activation of certain signal transduction patterns by dysregulated cytokine production, which can, in turn, be exhausted or down-modulated. HIV can impair T-cell and monocyte functions by disrupting cytokine-dependent and antigen-dependent signal transduction [24][48][49].

One of the most notable findings in our work was the identification of a novel biomarker for HIV viremia, thrombospondin-1 (THBS1), which was highly and uniquely significant in LTNPs compared to both TP1 and TP2, with its downregulation observed in natural controllers. THBS1 encodes Thrombospondin 1 (TSP-1). It was first isolated from platelets stimulated with thrombin and was designated as a 'thrombin-sensitive protein' [50]. Since its initial recognition, the functions of TSP-1 are involved in multiple biological processes, including angiogenesis, apoptosis, activation of TGF-β, and immune regulation. Therefore, TSP-1 is a multifunctional protein. Additionally, other studies have demonstrated that THBS1 influences various aspects of the vascular system, including platelet activation, angiogenesis, and wound healing. Therefore, an increase in THBS1 expression has been linked to multiple disease states, including tumor progression, atherosclerosis, and arthritis [51].

In the context of its relevance to HIV infection, several studies have demonstrated that this membrane protein exhibits anti-HIV-1 activity, as it can be considered a physical barrier to mucosal transmission due to its high expression in saliva [52][53]. Similar studies have also shown that TSP1 inhibits HIV-1 infection in peripheral blood mononuclear cells and that THBS1 has functional binding domains in conserved regions flanking the V3 loop of gp120, thus providing a molecular mechanism to inhibit HIV-1 activity. Despite several studies on THBS1 and its relevance in HIV infection, there is a considerable lack of understanding of how this protein, and its family members (TSP 2-5) may influence plasma viremia. According to our data, THBS1 was downregulated uniquely in the LTNP group. When we compared CD14 and CD16 monocytes from patients belonging to the TP1 and TP2 groups for correlation between their expression and plasma viremia, we found no correlation. However, comparing HIV patients with detectable viremia against LTNPs gave highly credible results on the expression levels of thrombospondin and its receptor CD47 and its association with viremia, suggesting that TSP1 may be implicated in the progression of AIDS and its associated pathologies, possibly by modulating the bioavailability and biological activity of

extracellular Tat [52]. This concurs with our findings on the microarray expression of THBS1 and subsequent functional validation by qRT-PCR, which showed that THBS1 was upregulated in the viremic group.

In contrast, in LTNPs, this gene was downregulated. Since our study was based on different sets of HIV patients, which differed between microarray and protein validation using FACS and confocal microscopy, it is apparent that there was no statistical significance despite the different sets of group-specific patients recruited for microarray experiments and that the significance of TSP1 expression in viremic and LTNP groups is universal and may have immense biological, clinical, and diagnostic significance. This has given more confidence to our analysis and has provided a more profound view of its association with viremia in HIV patients, based on which our data suggest that thrombospondin is a promising candidate biomarker for predicting viremia in HIV patients, but a validation through larger cohorts of LTNPs is warranted.

Thrombospondin is a new potential biomarker for predicting viremia in HIV patients.

Overall, these data may be among the first to provide preliminary but statistically strong evidence that thrombospondin could play a role in the response to uncontrolled HIV replication. The comprehensive overview of genomic events has guided its discovery, offering clues to the natural control of viremia in HIV+ LTNPs, as well as revealing some similarities between genomic events that occur during HAART initiation in HIV patients for controlling viremia and natural control in non-progressors. Given thrombospondin's role in regulating apoptosis, strategies aimed at controlling apoptosis or leveraging apoptosis pathways as therapeutic targets may enhance future HIV treatments.

Regarding natural control of HIV by non-progressing elite controllers and after CCR5 Δ32 homozygous cord blood allogeneic transplantation [54][55][56], it is also important to highlight that up to 60% of HIV genetic sequences in elite controllers were found in regions called “gene deserts” – barren areas of chromatin, the DNA-protein complex involved in chromosome condensation during cell division [57]. Despite these findings, the authors cautioned that their study did not confirm that their patients achieved a sterilizing cure for HIV infection through natural immune-mediated mechanisms [57], as recently shown in a non-progressing Sydney patient [58].

Lastly, it is essential to address some limitations of the study. Despite the small sample size (9 LTNPs, 15 ART-treated patients), we obtained statistically significant results that highlight the role of thrombospondin in HIV disease. Since LTNPs make up less than 1% of the HIV+ population, sampling is challenging. A larger sample size and coordinated international collaborations would improve statistical power and better reflect immune and genomic variability. Additionally, the study utilized frozen PBMCs for

genome-wide analysis. Using whole PBMCs was unavoidable due to the archival nature of the collection from a rare group of HIV patients. Isolating individual cell types from frozen samples isn't ideal due to low and inconsistent yields, which can affect gene expression analysis. Whole PBMCs offer a broad view of lymphocyte gene expression. Although sorted cells are necessary to confirm specific signals, that was not the primary focus of our study.

Despite these limitations, the study is important in HIV research as it identifies Thrombospondin-1 (THBS1) as a potential new biomarker that distinguishes natural control of viremia in long-term non-progressors (LTNPs) from therapy-mediated control in HIV patients. It offers insights into various immune and genomic mechanisms that could aid in the development of new therapeutic strategies and biomarkers for prognosis and treatment optimization.

## 4. Materials and Methods

### *Study participants and ethics approval*

The collection of blood samples from HIV-positive LTNP and treated HIV-positive patients was approved by St. Vincent's Hospital Human Research Ethics Committee (approval # 94/043) and the Australian Red Cross Blood Service Human Research Ethics Committee (approval # 8/97) (**Supplementary Table 1**).

### *RNA extraction*

The cells were first lysed for total RNA extraction from frozen PBMCs, and RNA was extracted according to the manufacturer's instructions (Qiagen RNeasy Purification Kit, Germany). The on-column digestion of DNA during RNA purification was performed to purify the samples. DNase was efficiently removed during the subsequent wash steps. Total RNA quality was assessed, and its concentration was measured using the Agilent RNA 6000 series II Nano kit (Agilent Technologies, CA, USA) using an Agilent 2100 Bioanalyzer according to the manufacturer's protocol. All RNA integrity numbers considered suitable for microarray analysis were seven or higher for all samples analyzed.

### *Genome-wide microarray*

Total RNA was reverse-transcribed to synthesize the first strand of cDNA, followed by second-strand synthesis. Double-stranded cDNA was transcribed and amplified in vitro to synthesize biotin-labelled complementary mRNA (cRNA). cRNA amplification and biotin labelling were performed using the Illumina TotalPrep RNA amplification kit (Ambion, Inc., Austin, TX, USA) with 250 ng of total RNA as input material.

cRNA yields were quantified using an Agilent Bioanalyzer. Seven hundred fifty nanograms of cRNA sample were hybridized on a HumanHT12 V3 Expression Bead Chip (Illumina, Inc., CA, USA). The chips were stained with streptavidin-Cye3 conjugate and scanned using an Illumina BeadArray Reader (Illumina Inc.).

### *Differential gene expression analysis*

Preliminary gene expression analysis was performed using Bead Studio version 3, followed by detailed analysis using BRB Array Tools. Data were normalized using a cubic spline function to minimize variation due to non-biological factors. The average signal intensity for each gene was measured using the Beadstudio v3. Around 24 000 genes were selected for differential expression analysis, with a detection p-value of less than 0.01 [50].

Following the normalization of the entire chip data, which covered more than 25,000 annotated human genes and more than 48,000 probes covering RefSeq and UniGene annotated genes, we performed clustering analysis using BRB-Array Tools software. The statistical test used by the BRB-Array software for the 234 DE genes was a Pearson's centred correlation, and the average linkage method was implemented to obtain the figures.

The data acquired from the Illumina Microarray BeadChip were used to obtain the DE list in the Illumina®BeadStudio Data Analysis Software. Furthermore, this DE gene list was analyzed using the BRB-ArrayTools software installed as an Excel package to perform various pre-processing steps, including computing probe-set expression summaries, normalization, filtering, and calculating quality control indices [47][48][49]. Quantile normalization was performed to diminish the biological errors due to false positives, to see more clearly the systematic biological differences between the samples, and to compensate for systematic technical differences between chips. Significance Analysis of Microarrays (SAM) was also performed on the BRB-Array Tools software to identify genes with statistically significant changes in expression by assimilating a set of gene-specific t-tests. During this process, each gene was assigned a score based on its change in gene expression relative to the standard deviation of repeated measurements for that gene. Genes with scores greater than the threshold were deemed to be potentially significant. The percentage of genes identified by chance was the false discovery rate (FDR). To estimate the FDR, nonsense genes were identified by analyzing the permutations of the measurements. The thresholds were adjusted to identify smaller or larger sets of genes, and FDRs were calculated for each set [50]. In this part of the study, we used the FDR of <0.05 with a p-value of <0.01 and $-1.10 \leq$ fold change $\geq 2.85$, so the genes that received a fold-change value higher or equal to -1.10 were appointed down-regulated

and the genes that received a fold-change minor or equal than 2.85 were up-regulated. The cut-off for the fold change was just a consequence of the selection of the FDR and p-value.

Furthermore, differentially expressed genes between TP1 and TP2 were identified using MeV (MultiExperiment Viewer), a desktop application for analyzing, visualizing, and mining large-scale genomic data. We used the false discovery rate (FDR), which defined the expected proportion of false positives among the declared significant results. An FDR of 5% was used for the differential gene expression analysis of TP1 and TP2, whereas significance analysis of microarray (SAM) was used to identify the significant genes for this group. Differentially expressed genes were analyzed as paired samples [52].

### *Pathways enrichment analysis*

The commercial software MetaCoreTM (GeneGo, MI, USA) was used. The differentially expressed genes from the MeV analysis were further analyzed to identify biological networks using GeneGo Maps modules and GeneGo Folders. MetacoreTM conducts a functional analysis of network pathways based on a manually curated database of human protein-protein, protein-DNA, and protein–compound interactions; metabolic and signaling pathways; and the effects of bioactive molecules on gene expression. For each group, the most significant pathways were selected. The p-value was calculated for the common and unique groups. The results were ranked by –log(p-value). By default, the most significant result for the common part is displayed [11].

To further verify these findings, we performed a gene enrichment analysis that mapped gene IDs of the dataset onto gene IDs in entities of built-in functional ontologies represented in MetaCore using pathway maps and networks. The DE list provided by MeV was uploaded to the MetaCoreTM website from Thomson Reuters Systems Biology Solutions to compare our samples and learn which genes are overrepresented in functionally relevant pathways. This analysis assessed the overall functional character of the sample set, providing a ranked representation of ontologies that are most saturated or "enriched" with the input data. Each GeneGo Process Network represents a comprehensive biological process with a specific functional theme.

### *Quantitative Real-time PCR (q-RTPCR)*

RT-qPCR was used to corroborate the relationship between microarray expression trends and RT-qPCR results. Fourteen genes were selected for RT-qPCR because of their high scores and relevance to this study.

5ng of total RNA was reverse transcribed using oligo d(T) and Superscript III, followed by RNase H treatment (Invitrogen Life Technologies), according to the manufacturer's protocol. PCR primers were designed for the genes selected based on the microarray data, as well as for the control genes (GAPDH: Glyceraldehyde 3-phosphate dehydrogenase), using Primer 3 (http://frodo.wi.mit.edu/primer3/) (**Supplementary files 6 and 7**). The cDNA was subjected to RT-PCR with defined primers and SYBR Green (Invitrogen Life Technologies) using an MX3000p Stratagene real-time cycler (Stratagene, La Jolla, CA, USA). The data were analyzed using MxPro™ QPCR software version 4.0.1 (Stratagene, La Jolla, CA, USA). For all experiments, duplicates were used, and relative mRNA expression was calculated using the comparative ΔΔCt method for all data. Values of fold-change represent averages from duplicate experiments. Fold change was calculated simply as the ratio of the changes between final value and the original value over the initial value. Thus, if the original value is X and final value is Y, the fold change is (Y - X)/X or equivalently Y/X - 1. negative values were considered downregulated and positive values were considered upregulated. Data were further analyzed using the Wilcox test to check for statistical significance.

### *Flow cytometric analysis*

Flow cytometry was used to quantitatively analyze thrombospondin in monocytes derived from HIV patients with detectable viral loads, HIV-negative individuals, and HIV+ LTNPs. We used whole PBMCs from the patient groups (HIV+ patients with undetectable virus in the plasma, HIV+ individuals with detectable virus, and LTNPs) and gated them on CD14 and CD16 monocytes.

The main objective of these experiments was to delineate the functional relevance of THBS1 at the protein level in monocytes and identify its correlation with plasma viremia. Whole frozen PBMCs were used in each group for experiments. Initially, we wanted to determine which cell population would have increased protein expression. Therefore, using flow cytometry (BD FACSCanto II), we gated different cell populations for each group and found that THBS1 was mostly stained in the monocytes. Each colored line represents one group, and it is clear that THBS1 is highly expressed in monocytes. Therefore, for a better analysis, we chose to concentrate our dataset only on the monocyte population.

For the remainder of the experiments, we focused only on CD14+ cells. Cells were frozen and thawed according to the manufacturer's instructions. Initially, PBMCs were stained with the LIVE/DEAD® Fixable Dead Cell Stain Kit from Invitrogen® to check for cell viability, as our samples were frozen. First, we stained the samples with DAPI from the kit to check viability and then stained them with CD14, CD16, and THBS1 antibodies. Using flow cytometry (BD FACSCanto II), the DAPI+ cell population from all groups was

gated first from the monocyte population, then the CD14+ (P4), CD16+, and THBS1+ cell populations were analyzed. We used five patients with BD viral load, five with detectable VL before HAART treatment, five LTNPs, and five HIV donors.

To measure the expression value from the FACS results (**Table 4**), we used the Mean Fluorescence Intensity (MFI) to normalize the stained samples of the negative control to report them more accurately. We also used the percentage (%) of cells expressing THBS1, which was used to determine the expression, and as we can observe from our results, THBS1 is highly expressed in patients with high viral loads when compared with any of the groups, especially LTNP. The p-values for the comparisons based on the expression values of the groups were highly significant.

### *Confocal microscopy*

PBMC cultures on coverslips were incubated for 15 min at room temperature in a blocking buffer to block non-specific binding. Cells were rinsed three times in PBS before being incubated for 45 min with primary antibody [Thrombospondin (THBS1) 1, Abcam, Cambridge)] diluted 1:50 in antibody dilution buffer. After rinsing six times in PBS, the cells were incubated for one h with a secondary antibody (fluorescein-conjugated Alexa Fluor®488, goat anti-rabbit, Invitrogen, USA) diluted 1:200 containing 1μL Hoechst (33258, pentahydrate [bis-benzimide], Invitrogen, USA) diluted 1:100 for nuclear visualization. This was followed by six washes with PBS before the cells were mounted on glass slides. The slides were examined using Leica TCS SP5 Confocal Microscope. Excitation was performed with the 405 nm line (DAPI) and 476 nm line (FITC) from the blue diode and multiline argon lasers. The excitation was performed at 20% of the power.

The cells were examined with an HCX PL APO 63.0x1.30 37°C UV Glycerol immersion objective, using a FluoView FV1000 Series Confocal Laser Scanning Microscope. Confocal images were acquired using the Leica Application Suite (Advanced Fluorescence) software. The scan mode was set to XYZ, and the image stack was recorded from the XY sections in the z-direction. The z-position was adjusted until the cell nuclei were visualized. High resolution was achieved by setting the image size to 1024 × 1024 pixels, with approximately 1,321,963 pixels per image. The zoom factor was set between 1 and 3 to allow for a better presentation of the cell.

The image quality was optimized to reduce the photobleaching rate and noise level. Noise reduction was achieved by signal averaging-performing 4-6 scans per image. The gain and offset levels were optimized for high contrast without saturation or loss of detail. Other parameters, such as pinhole size and laser

power, were also adjusted to acquire better-quality images. Image cropping, brightness, and contrast adjustments for presentation purposes were performed using Photoshop CS5.1 image editing software.

### *DeltaVision microscopy*

PBMCs on coverslips were treated and stained, as described above. Images were acquired using a Photometrics CoolSnap QE camera with sequential exposures and deconvolved using the Sedat & Agard algorithms available through the Deltavision SoftWoRx software, version 3.0.0. The background was subtracted post-capture using SoftWoRx through a single adjustment of the level histogram. The images were cropped using Adobe Photoshop CS5.

## Statements and Declarations

### *Funding*

This work was funded by a development grant from the NHMRC to NKS. The Australian governments fund the Australian Red Cross Lifeblood to provide blood, blood products, and services to the Australian community.

### *Conflicts of Interest*

The authors declare no conflicts of interest.

### *Ethics*

The collection of blood samples from HIV-positive LTNP and treated HIV-positive patients was approved by the St. Vincent's Hospital Human Research Ethics Committee (approval # 94/043) and the Australian Red Cross Blood Service Human Research Ethics Committee (approval # 8/97). Written informed consent was obtained from all participants involved in the study.

### *Data Availability*

The microarray data reported in this paper will be deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo, and accession number will be provided.

### *Author Contributions*

As a part of her Ph.D. degree, VC carried out all the experimental work related to molecular and genetic studies and their validation and drafted the manuscript. NKS conceived, conceptualized, designed, coordinated, supervised the study and assisted in end-to-end review of experiments, interpretation, and drafting the manuscript. WBD provided patient samples, clinical details, assisted with the manuscript writing, and co-authored the manuscript. KG helped with the bioinformatics analysis, statistics, and interpretation. MMS assisted directly with confocal and deltavision microscopy and image and data interpretation. All authors have read and approved the final manuscript.

### *Acknowledgements*

VC is thankful to the University of Sydney for the IPRS Fellowship for Ph.D. studies and the WMIRF for the Top-Up grant. We thank all study participants and Dr. Cassy Workman, Dr. Jenny Learmont, and Dr. John S. Sullivan for participant management and specimen collection.

## References

1. ^*Wu JQ, Sassé TR, Wolkenstein G, et al. (2013). "Transcriptome Analysis of Primary Monocytes Shows Global Down-Regulation of Genetic Networks in HIV Viremic Patients Versus Long-Term Non-Progressors." Virology. **435**(2):308-19.*
2. a, b*Mahajan SD, Agosto-Mojica A, Aalinkeel R, et al. (2010). "Role of Chemokine and Cytokine Polymorphisms in the Progression of HIV-1 Disease." Biochem Biophys Res Commun. **396**(2):348-52.*
3. a, b*Pastori C, Weiser B, Barassi C, et al. (2006). "Long-Lasting CCR5 Internalization by Antibodies in a Subset of Long-Term Nonprogressors: A Possible Protective Effect Against Disease Progression." Blood. **107**(12):4825-33.*
4. ^*Dyer WB, Kuipers H, Coolen MW, et al. (2002). "Correlates of Antiviral Immune Restoration in Acute and Chronic HIV Type 1 Infection: Sustained Viral Suppression and Normalization of T Cell Subsets." AIDS Res Hum Retroviruses. **18**(14):999-1010.*
5. ^*Dyer WB, Ogg GS, Demoitie MA, et al. (1999). "Strong Human Immunodeficiency Virus (HIV)-Specific Cytotoxic T-Lymphocyte Activity in Sydney Blood Bank Cohort Patients Infected With Nef-Defective HIV Type 1." J Virol. **73**(1):436-43.*
6. ^*Dyer WB, Geczy AF, Kent SJ, et al. (1997). "Lymphoproliferative Immune Function in the Sydney Blood Bank Cohort, Infected With Natural Nef/Long Terminal Repeat Mutants, and in Other Long-Term Survivors of Tra*

*nsfusion-Acquired HIV-1 Infection." AIDS.* ***11****(13):1565-74.*

7. ^*Canducci F, Marinozzi MC, Sampaolo M, et al. (2009). "Dynamic Features of the Selective Pressure on the H uman Immunodeficiency Virus Type 1 (HIV-1) Gp120 CD4-Binding Site in a Group of Long Term Non Progres sor (LTNP) Subjects." Retrovirology.* ***6****:4.*
8. ^*Learmont JC, Geczy AF, Mills J, et al. (1999). "Immunologic and Virologic Status After 14 to 18 Years of Infecti on With an Attenuated Strain of HIV-1. A Report From the Sydney Blood Bank Cohort." N Engl J Med.* ***340****(2 2):1715-22.*
9. a, b*Dyer WB, Learmont JC, Geczy AF, et al. (2004). "High Frequency of Non-Progression 20 Years After Transf usion-Transmitted Human Immunodeficiency Virus Type-1 Infection." Transfusion.* ***44****(9):1397-8.*
10. ^*Gorry PR, McPhee DA, Verity E, et al. (2007). "Pathogenicity and Immunogenicity of Attenuated, Nef-Delete d HIV-1 Strains in Vivo." Retrovirology.* ***4****:66.*
11. a, b, c*Dyer WB, Zaunders JJ, Yuan FF, et al. (2008). "Mechanisms of HIV Non-Progression; Robust and Sustaine d CD4+ T-Cell Proliferative Responses to P24 Antigen Correlate With Control of Viraemia and Lack of Disease Progression After Long-Term Transfusion-Acquired HIV-1 Infection." Retrovirology.* ***5****:112.*
12. ^*Rodés B, Toro C, Paxinos E, et al. (2004). "Differences in Disease Progression in a Cohort of Long-Term Non-Progressors After More Than 16 Years of HIV-1 Infection." AIDS.* ***18****(8):1109-16.*
13. a, b, c*Wu JQ, Dwyer DE, Dyer WB, et al. (2011). "Genome-Wide Analysis of Primary CD4+ and CD8+ T Cell Tra nscriptomes Shows Evidence for a Network of Enriched Pathways Associated With HIV Disease." Retrovirolo gy.* ***8****:18.*
14. a, b, c*McLaren PJ, Mayne M, Rosser S, et al. (2004). "Antigen-Specific Gene Expression Profiles of Peripheral Bl ood Mononuclear Cells Do Not Reflect Those of T-Lymphocyte Subsets." Clin Diagn Lab Immunol.* ***11****(5):977-8 2.*
15. a, b*da Conceicao VN, Dyer WB, Gandhi K, Gupta P, Saksena NK. (2014). "Genome-Wide Analysis of Primary Pe ripheral Blood Mononuclear Cells From HIV+ Patients-Pre-And Post- HAART Show Immune Activation and I nflammation the Main Drivers of Host Gene Expression." Mol Cell Ther.* ***2****:11. doi:10.1186/2052-8426-2-11. PMI D 26056580; PMCID PMC4451969.*
16. ^*Sturn A, Quackenbush J, Trajanoski Z. (2002). "Genesis: Cluster Analysis of Microarray Data." Bioinformatic s.* ***18****(1):207-8.*
17. ^*Blankson JN. (2010). "Effector Mechanisms in HIV-1 Infected Elite Controllers: Highly Active Immune Respon ses?" Antiviral Res.* ***85****(1):295-302.*

18. ^ *Migueles SA, Connors M. (2010). "Long-Term Nonprogressive Disease Among Untreated HIV-Infected Individuals: Clinical Implications of Understanding Immune Control of HIV." JAMA. **304**(2):194-201.*

19. ^ *Mandalia S, Westrop SJ, Beck EJ, et al. (2012). "Are Long-Term Non-Progressors Very Slow Progressors? Insights From the Chelsea and Westminster HIV Cohort, 1988-2010." PLoS One. **7**(2):e29844.*

20. ^ *Nikolsky Y, Ekins S, Nikolskaya T, et al. (2005). "A Novel Method for Generation of Signature Networks as Biomarkers From Complex High Throughput Data." Toxicol Lett. **158**(1):20-9.*

21. ^ *Ekins S, Andreyev S, Ryabov A, et al. (2006). "A Combined Approach to Drug Metabolism and Toxicity Assessment." Drug Metab Dispos. **34**(3):495-503.*

22. ^ *Zaunders JJ, Munier ML, Kaufmann DE, et al. (2005). "Early Proliferation of CCR5(+) CD38(+++) Antigen-Specific CD4(+) Th1 Effector Cells During Primary HIV-1 Infection." Blood. **106**(5):1660-7.*

23. a, b *Moir S, Malaspina A, Pickeral OK, et al. (2004). "Decreased Survival of B Cells of HIV-Viremic Patients Mediated by Altered Expression of Receptors of the TNF Superfamily." J Exp Med. **200**(7):587-99.*

24. a, b *Abbate I, Dianzani F, Capobianchi MR. (2000). "Activation of Signal Transduction and Apoptosis in Healthy Lymphomonocytes Exposed to Bystander HIV-1-Infected Cells." Clin Exp Immunol. **122**(3):374-80.*

25. a, b *Moir S, Fauci AS. (2009). "B Cells in HIV Infection and Disease." Nat Rev Immunol. **9**(4):235-45.*

26. ^ *Xu H, He XQ, Chen R, et al. (2008). "[Construction and Identification of Antisense C-Jun N-Terminal Kinase 1 Eukaryotic Fluorescent Expressing Plasmids and JNK1-/- Human Embryo Lung Fibroblasts Cell Line]." Zhong hua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi [Chinese Journal of Industrial Hygiene and Occupational Diseases]. **26**(9):538-41.*

27. ^ *Malaspina A, Moir S, Ho J, et al. (2006). "Appearance of Immature/Transitional B Cells in HIV-Infected Individuals With Advanced Disease: Correlation With Increased IL-7." Proc Natl Acad Sci U S A. **103**(7):2262-7.*

28. ^ *Chen KC, Wang TY, Chan CH. (2012). "Associations Between HIV and Human Pathways Revealed by Protein-Protein Interactions and Correlated Gene Expression Profiles." PLoS One. **7**(3):e34240.*

29. ^ *Silva R, Moir S, Kardava L, et al. (2011). "CD300a Is Expressed on Human B Cells, Modulates BCR-Mediated Signaling, and Its Expression Is Down-Regulated in HIV Infection." Blood. **117**(22):5870-80.*

30. a, b *Moir S, Fauci AS. (2008). "Pathogenic Mechanisms of B-Lymphocyte Dysfunction in HIV Disease." J Allergy Clin Immunol. **122**(1):12-9; quiz 20-1.*

31. ^ *Paroli M, Propato A, Accapezzato D, et al. (2001). "The Immunology of HIV-Infected Long-Term Non-Progressors--A Current View." Immunol Lett. **79**(1-2):127-9.*

32. ^ *Kirchhoff F, Greenough TC, Brettler DB, et al. (1995). "Brief Report: Absence of Intact Nef Sequences in a Long-Term Survivor With Nonprogressive HIV-1 Infection." N Engl J Med. **332**(4):228-32.*

33. ^ Liu Z, Cumberland WG, Hultin LE, et al. (1997). "Elevated CD38 Antigen Expression on CD8+ T Cells Is a Stronger Marker for the Risk of Chronic HIV Disease Progression to AIDS and Death in the Multicenter AIDS Cohort Study Than CD4+ Cell Count, Soluble Immune Activation Markers, or Combinations of HLA-DR and CD38 Expression." J Acquir Immune Defic Syndr Hum Retrovirol. **16**(2):83-92.

34. ^ Zaunders JJ, Geczy AF, Dyer WB, et al. (1999). "Effect of Long-Term Infection With Nef-Defective Attenuated HIV Type 1 on CD4+ and CD8+ T Lymphocytes: Increased CD45RO+CD4+ T Lymphocytes and Limited Activation of CD8+ T Lymphocytes." AIDS Res Hum Retroviruses. **15**(17):1519-27.

35. ^ Jiao Y, Fu J, Xing S, et al. (2009). "The Decrease of Regulatory T Cells Correlates With Excessive Activation and Apoptosis of CD8+ T Cells in HIV-1-Infected Typical Progressors, But Not in Long-Term Non-Progressors." Immunology. **128**(1 Suppl):e366-75.

36. ^ Franceschi C, Franceschini MG, Boschini A, et al. (1997). "Phenotypic Characteristics and Tendency to Apoptosis of Peripheral Blood Mononuclear Cells From HIV+ Long Term Non Progressors." Cell Death Differ. **4**(8):815-23.

37. ^ Anguissola S, McCormack WJ, Morrin MA, et al. (2011). "Pigment Epithelium-Derived Factor (PEDF) Interacts With Transportin SR2, and Active Nuclear Import Is Facilitated by a Novel Nuclear Localization Motif." PLoS One. **6**(10):e26234.

38. ^ Badley AD, Parato K, Cameron DW, et al. (1999). "Dynamic Correlation of Apoptosis and Immune Activation During Treatment of HIV Infection." Cell Death Differ. **6**(5):420-32.

39. ^ a, b Moretti S, Marcellini S, Boschini A, et al. (2000). "Apoptosis and Apoptosis-Associated Perturbations of Peripheral Blood Lymphocytes During HIV Infection: Comparison Between AIDS Patients and Asymptomatic Long-Term Non-Progressors." Clin Exp Immunol. **122**(3):364-73.

40. ^ Balestrieri E, Grelli S, Matteucci C, et al. (2007). "Apoptosis-Associated Gene Expression in HIV-Infected Patients in Response to Successful Antiretroviral Therapy." J Med Virol. **79**(2):111-7.

41. ^ Cummins NW, Badley AD. (2010). "Mechanisms of HIV-Associated Lymphocyte Apoptosis: 2010." Cell Death Dis. **1**(11):e99.

42. ^ Garg H, Blumenthal R. (2006). "HIV Gp41-Induced Apoptosis Is Mediated by Caspase-3-Dependent Mitochondrial Depolarization, Which Is Inhibited by HIV Protease Inhibitor Nelfinavir." J Leukoc Biol. **79**(2):351-62.

43. ^ Herbein G, Khan KA. (2008). "Is HIV Infection a TNF Receptor Signalling-Driven Disease?" Trends Immunol. **29**(2):61-7.

44. ^ Herbein G, Gras G, Khan KA, et al. (2010). "Macrophage Signaling in HIV-1 Infection." Retrovirology. **7**:34.

45. ^de Oliveira Pinto LM, Garcia S, Lecoeur H, et al. (2002). "Increased Sensitivity of T Lymphocytes to Tumor Necrosis Factor Receptor 1 (TNFR1)- And TNFR2-Mediated Apoptosis in HIV Infection: Relation to Expression of Bcl-2 and Active Caspase-8 and Caspase-3." Blood. **99**(5):1666-75.

46. a, bWang JH, Wells C, Wu L. (2008). "Macropinocytosis and Cytoskeleton Contribute to Dendritic Cell-Mediated HIV-1 Transmission to CD4+ T Cells." Virology. **381**(1):143-54.

47. a, b, cMaréchal V, Prevost MC, Petit C, et al. (2001). "Human Immunodeficiency Virus Type 1 Entry Into Macrophages Mediated by Macropinocytosis." J Virol. **75**(22):11166-77.

48. a, bZhou H, Xu M, Huang Q, et al. (2008). "Genome-Scale RNAi Screen for Host Factors Required for HIV Replication." Cell Host Microbe. **4**(5):495-504.

49. a, bAlhetheel A, Yakubtsov Y, Abdkader K, et al. (2008). "Amplification of the Signal Transducer and Activator of Transcription I Signaling Pathway and Its Association With Apoptosis in Monocytes From HIV-Infected Patients." AIDS. **22**(10):1137-44.

50. a, b, cBaenziger NL, Brodie GN, Majerus PW. (1971). "A Thrombin-Sensitive Protein of Human Platelet Membranes." Proc Natl Acad Sci U S A. **68**(1):240-3.

51. ^McLaughlin JN, Mazzoni MR, Cleator JH, et al. (2005). "Thrombin Modulates the Expression of a Set of Genes Including Thrombospondin-1 in Human Microvascular Endothelial Cells." J Biol Chem. **280**(23):22172-80.

52. a, b, cCrombie R. (2000). "Mechanism of Thrombospondin-1 Anti-HIV-1 Activity." AIDS Patient Care STDS. **14**(4):211-4.

53. ^Crombie R, Silverstein RL, MacLow C, et al. (1998). "Identification of a CD36-Related Thrombospondin 1-Binding Domain in HIV-1 Envelope Glycoprotein Gp120: Relationship to HIV-1-Specific Inhibitory Factors in Human Saliva." J Exp Med. **187**(1):25-35.

54. ^Duarte RF, Salgado M, Sánchez-Ortega I, et al. (2015). "CCR5 Δ32 Homozygous Cord Blood Allogeneic Transplantation in a Patient With HIV: A Case Report." Lancet HIV. **2**(6):e236-42.

55. ^Gupta RK, Abdul-Jawad S, McCoy LE, et al. (2019). "HIV-1 Remission Following CCR5Δ32/Δ32 Haematopoietic Stem-Cell Transplantation." Nature. **568**(7751):244-8.

56. ^Ding J, Liu Y, Lai Y. (2021). "Knowledge From London and Berlin: Finding Threads to a Functional HIV Cure." Front Immunol. **12**:688747.

57. a, bJiang C, Lian X, Gao C, et al. (2020). "Distinct Viral Reservoirs in Individuals With Spontaneous Control of HIV-1." Nature. **585**(7824):261-7.

58. ^Zaunders J, Dyer WB, Churchill M, et al. (2019). "Possible Clearance of Transfusion-Acquired Nef/LTR-Deleted Attenuated HIV-1 Infection by an Elite Controller With CCR5 Δ32 Heterozygous and HLA-B57 Genotype." J V

*irus Erad. 5(2):73-83.*

**Supplementary data:** available at https://doi.org/10.32388/WGKALR

## Declarations

**Funding:** This work was funded by a development grant from the NHMRC to NKS. The Australian governments fund the Australian Red Cross Lifeblood to provide blood, blood products, and services to the Australian community.

**Potential competing interests:** No potential competing interests to declare.