A SYK-dependent activation of STAT1-IRF1 amplifies the IFN signaling in HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP)

David Felipe Cuéllar Giraldo

Degree Project in Pharmaceutical Modeling, 45 credits, Spring semester 2018

Supervisor: Helder Nakaya
Examiner: Ola Spjuth

Computational Systems Biology Laboratory
Department of Clinical and Toxicological Sciences
University of Sao Paulo, Brazil
Faculty of Pharmacy
Uppsala University
Table of contents

Abstract page 3
Introduction page 4
Purpose page 5
Methods pages 5-6
Results pages 6-10
Discussion pages 10-12
Conclusions pages 13
References page 14-15
Supporting information 16-19
ABSTRACT: Human T-lymphotropic virus 1 (HTLV-1) associated myelopathy/tropical spastic paraparesis (HAM/TSP) is a progressive inflammatory disease of the spinal cord that causes weakness or paralysis of the legs and urinary symptoms. HTLV-1 infection is required to develop the disease and make the diagnosis. However, the differences in the lifetime incidence between asymptomatic carrier populations and the presence of lymphocytic infiltrates in the CNS lesions suggest that additional host immunological factors determine the risk. A signature of interferon-stimulated genes was found to be an important pathway for the disease. However, the pathogenesis still remains enigmatic and there is not an effective disease modifying treatment. This study applied systems approaches to comprehensively integrated data from three GEO series that measured the gene expression level in 39 asymptomatic HTLV-1 carriers and 19 patients with HAM/TSP. This study suggested that IFN signaling is enhanced by a SYK-dependent activation of STAT1 and IRF1, this amplification can be an important mechanism of spinal cord injury in HAM/TSP. The consistently up-regulated differentially expressed genes (DEGs) and the consistently co-expressed module M15 were significantly enriched for the Interferon pathway. STAT1 was found to be a top ranked hub protein and one of the three genes that were DEGs in all of the GEO studies. IRF1 was the upstream top ranked transcription factor and it had a direct protein-protein interaction with STAT1. Finally, the predicted upstream interaction between SYK and the expanded protein interaction network indicated the key regulatory role of this protein. In conclusion, two transcription factors (STAT1 and IRF1) and one protein kinase (SYK) were identified to be key molecular regulators of the IFN signaling amplification found in our patients, this mechanism can lead to an exaggerated neuroinflammatory response to HTLV-1 in the spinal cord and thus, these molecules represent potential drug targets for the chronic debilitating and neglected condition known as HAM/TSP.

Key words: HTLV-1, tropical spastic paraparesis, HTLV-1-associated myelopathy, gene expression, drug target
INTRODUCTION: Human T-lymphotropic virus 1 (HTLV-1) is a positive single stranded RNA oncogenic retrovirus that preferentially infects CD4 T cells. According to previous studies 10-20 million people are infected with HTLV-1 worldwide. It is estimated that 1-3% of carriers develop a debilitating and sometimes fatal chronic inflammatory disease of the central nervous system (CNS), known as HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP) and 5% develop a rapidly progressive malignancy, adult T-cell leukemia (ATL).

Although HTLV-1 high proviral load is the strongest predictor to develop HAM/TSP, it is insufficient to cause alone the disease because only a minority of carriers with high proviral loads develop HAM/TSP. The proviral load that a given patient has is influenced by the genotype of the immune T cells. For instance, the allele HLA-A*02 has been associated with a protective effect in HTLV-1 carriers against developing HAM/TSP in southern Japan, Afro-Caribbean cohort of United States and in Brazilian populations. In contrast the HLA DRB1*0101 and HLA-B*07 have been associated with increased risk of HAM/TSP in a population from Spain. Additionally, some single nucleotide polymorphism (SNPs) have been associated with the neurological condition. SNPs in the promoter of interleukin 6 (IL6-634C), tumor necrosis factor (TNF) and IL28B gene have been associated with an increasing risk of developing HAM/TSP while specific SNPs in IL10, chemokine stromal cell derived factor 1 (SDF1) and IL15 have been associated with a protective role.

A previous study demonstrated a transcriptional signature of interferon-stimulated genes in cases that developed the neurological condition. It has been proposed that this signature is mainly due to IFN type 2 (IFN gamma) because the HTLV-1 Tax protein induces IFN type 2 expression in infected CD4 T cells. It was found that the Tax viral protein of HTLV-1 produces that CD4 CD25 CCR4 Tregs cells lose the expression of the transcription factor FOXP3 and also boosts the expression of the Th1 master regulator T box transcription factor (T-bet) in HTLV-1 infected CCR4 T cells which result in the production of IFN-y and consequently, neuroinflammation.

As the pathogenesis of the disease is not fully elucidated and there is a significant inter-individual and intra-individual variability in the disease progression, the development of an effective treatment has been elusive because the therapeutic window for immunomodulation is within the initial years of the disease where the inflammatory profile is more relevant. Currently, the clinical management of HAM/TSP is guided to control the main symptoms like back pain, spasticity, urinary incontinence and muscle spasm with symptomatic drugs and physical therapy.
PURPOSE: The primary objective of this study is to identify novel drug targets that can potentially modify the disease progression of HAM/TSP.

METHODS:

GEO series selection: Widely accessible gene expression series related to HAM/TSP were obtained from the Gene Expression Omnibus (GEO) database of NCBI (http://www.ncbi.nlm.nih.gov/geo/). The search terms “tropical spastic paraparesis”, “HTLV-1 associated myelopathy”, “HTLV-1 associated myelopathy/tropical spastic paraparesis” and “HTLV-1 infection” were used to mine the GEO series. Those GEO series which compared gene expression of asymptomatic HTLV-1 carriers vs HAM/TSP patients and additionally, used Affymetrix, Illumina or Agilent microarray platforms were included in this study. The exclusion criteria were GEO series that either had less than 4 samples from each group, measured gene expression in vitro or analyzed the expression effect after treatment with a drug.

Datasets acquisition, preprocessing and quality control: Normalized raw gene expression data and its associated metadata were retrieved from the GEO by using the GEOquery R package. The series that used Illumina platform were filtered by removing those probes that had more than 60% of samples with a background p value > 0.05. The probes that did not have annotation were annotated by using the R BiomaRt package and those that were not associated with known genes were excluded. The probes were collapsed by selecting the probe for each gene that had the highest mean expression value across the samples. Quality control was performed on normalized datasets by using the R package ArrayQualityMetrics. Arrays that were called outliers by at least one criterion were removed from the analysis. The outlier’s detection criterion were the distances between arrays, boxplots and MA plots. The 30 percent of genes with lowest expression were removed and the rest of genes expression values were log2 transformed.

Identification of differentially expressed genes (DEGs) and enrichment: The LIMMA (Linear Models for Microarray Analysis) method implemented in R was used to find DEGs. A DEG was defined as a gene with a false discovery rate (FDR) < 0.05 and an absolute fold change (FC) > 1.5. The DEGs analysis was performed independently for each GEO series. Then, a consistently DEG was defined as a DEG that was found in two of the three GEO series. This list of consistently up-regulated and down-regulated DEGs represented the HAM/TSP signature which was analyzed for annotated gene sets enrichment using the online tool ENRICHR. Enrichment terms that had a Benjamini-Hochberg-adjusted p value < 0.05 were considered significant.
Identification of co-expressed gene sets: The modules of co-expressed genes were identified independently for each GEO series by using the Weighted Gene Co-expression Network Analysis (WGCNA) implemented in the CEMiTool package. A co-occurrence weighted network was built in order to integrate the co-expression results of the three GEO series. This was performed by identifying all the pair of genes that appear in the same module and counting their frequency in all the GEO series. This resulting network had nodes which represent genes and their weighted edges are proportional to the number of times that a pair of genes appear in the same co-expression module. This network was used to identify consistently co-expressed modules and an estimate of its mean fold change by using the Community Detection algorithm implemented in Gephi software. Finally, this consistently co-expressed modules were compared with the list of DEGs.

Protein-protein interaction (PPI) Network-Based Enrichment analysis and Hubs identification: The consistently up and down regulated DEGs were mapped on protein-protein interaction using NetworkAnalyst tool. The networks were created by keeping only those nodes that were necessary to connect the seed nodes (minimum interaction network). IMEx interactome was the platform used to build the PPI networks. Degree centrality and betweenness centrality were calculated to assess the topological properties of the PPI networks. The densely connected group of nodes referred as modules in the network were predicted using the “module explorer” panel of NetworkAnalyst which uses a random walk based algorithm for module selection.

Selection of potential drug targets: Upstream transcription factors (TFs) and protein kinases (PKs) were identified by submitting the list of genes of the consistently co-expression module that had the highest percent of DEGs into Expression2Kinases (X2K) web interface. X2K recognized enriched TFs using a ChEA database and then, found transcription complexes by using a ChEA database. Finally, PKs involved in TFs complex formation and functional regulation were identified by using the Kinase Enrichment Analysis module of X2K.

RESULTS:

Five GEO series were available in November 2017 through the GEO public repository. However, only three GEO series fulfilled the inclusion criteria and were selected for the analysis (GSE29312, GSE29332 and GSE38537 which were labeled as WB1, WB2 and T4, respectively). All the selected series had at least 4 individuals for each one of the groups (asymptomatic HTLV-1 carriers (AC) and HAM/TSP) and were collected from whole blood and CD4 cells, respectively. WB1 and WB2 used Illumina platform while T4 used the Agilent platform. The series WB1 and WB2 are part of the SuperSeries GSE 29333, however, each study was analyzed independent because
each study came from an independent cohort with demographic and clinical differences. Three samples were removed from WB1 because they exceeded the outlier detection criteria for the distance between arrays and one sample from WB1 and three samples from WB2 were removed because they were outliers according to the array intensity distribution.

After removing the outliers the total sample size was 58 samples, 39 samples in the AC group and 19 samples in the HAM/TSP group. There were 42 consistently DEGs across the three datasets, 28 genes were up-regulated and 14 genes were down-regulated, these genes composed the HAM/TSP signature (Figure 1 and S1 Figure). STAT1, VAMP5 and GBP5 were present as DEGs in the three GEO series (Figure 1). The highest average log2 fold change from the consistently DEGs was a 1.07 fold up-regulation of the GBP5 gene (Guanylate binding protein 5). The up and down regulated DEGs were tested separately for a common functional or phenotypic basis by conducting a pathway enrichment analysis using the Reactome 2016 public database with EnrichR. It was found that the up regulated DEGs are significantly enriched for the Interferon signaling, cytokine signaling, immune system, interferon gamma signaling and Class I MHC mediated antigen processing and presentation (S1 Table). On the other hand, the down regulated DEGs were enriched for signaling by retinoic acid, arachidonic acid metabolism, metabolism, metabolism of lipids and lipoproteins and synthesis of bile acids and bile salts (S2 Table).

Figure 1. The top panels WB1, WB2 and T4 represent the volcano plots of DEGs for each GEO study. The arrows shows the three consistently up-regulated DEGs that were common in the 3 GEO studies. In blue appear the genes that fulfilled both criterion to be a DEG: a false discovery rate (FDR) < 0.05 and an absolute fold change (FC) > 1.5. b) Venn diagram that shows the intersections of up-regulated DEGs for the three GEO studies and in c) the intersections of down-regulated DEGs for the studies.
The PPI network of the consistently down-regulated DEGs resulted in two subnetworks including one subnetwork that contained most of nodes and edges. This subnetwork resulted in 16 nodes and 16 edges with 10 seed nodes. Three nodes were found to have \( \geq 3 \) of degree centrality. These genes were UBC (degree 6, betweenness centrality 73.5), HIST1H3 (degree 4, betweenness centrality 51.5) and DLAT (degree 3, betweenness centrality 29.5). On the other hand, the PPI network of the consistently up-regulated DEGs resulted in only one subnetwork with 31 nodes and 49 edges with 19 seed nodes. Six nodes were found to have a degree centrality \( \geq 5 \). These genes were UBC (degree 14, betweenness centrality 244.31), STAT1 (degree 10, betweenness centrality 155.03), FBXO6 (degree 6, betweenness centrality 68,36), ISG15 (degree 5, betweenness centrality 13,8), GBP1 (degree 5, betweenness centrality 35,24) and IRF1 (degree 5, betweenness centrality 61,31). Additionally, the analysis of top functional modules in the PPI network built with the consistently up-regulated DEGs allowed to identify the modules of immune system (p-value 5.74e-7), interferon signalling (p-value 5.94e-7) and cytokine signalling (p-value 0.00000129) (S2 Figure). In contrast, the top functional modules identified with the PPI subnetwork of consistently down-regulated DEGs were regulation of pyruvate dehydrogenase complex (p-value 0.0000973), pyruvate metabolism (p-value 0.000261) and pyruvate metabolism and citric acid cycle (p-value 0.001).

Figure 2. a) Network of consistently co-expressed modules and their functional enrichment. The nodes represent genes and their weighted edges are proportional to the number of times that a pair of genes appear in the same co-expression module. b) The network depicts the inferred upstream regulatory network predicted to regulate the list of genes of the consistently up-regulated module M15 (interferon module). IRF1 can be seen as a top TF that has a physical protein-protein interaction with STAT1 and two others top ranked TFs (SOX2 and IRF8). The pink nodes represent the top transcription factors predicted to regulate the expression of the gene list; orange nodes represent intermediate proteins that physically interact with the enriched transcription factors and connect them. Blue nodes represent the top predicted protein kinases known to phosphorylate the intermediate proteins within the predicted subnetwork. Green
network edges represent kinase-substrate phosphorylation interactions between kinases and their substrates. Finally, maroon edges represent physical protein-protein interactions.

The analysis of co-expression modules with CEMiTool identified seven modules for WB1, six modules for WB2 and seven modules for T4. The modules M5 from WB1 and WB2 presented genes related to the interferon pathway and the Gene set enrichment analysis of these two modules revealed that the transcriptional activity is significantly high in the HAM/TSP group compared to the AC group. The network analysis of these modules M5 identified FBX06 and STAT1 as significant common hubs (degree > 200). On the other hand, the module M1 from T4 was enriched for innate immune system and was up-regulated in the HAM/TSP cases compared to the asymptomatic carriers. The analysis of consistently co-expressed modules using Gephi resulted in the identification of different enriched modules (Figure 2a and S3), however only one module (M15) was consistently co-expressed across the three GEO series (Figure 2a), interestingly this module was enriched for the interferon pathway and was found to be consistently up-regulated in the HAM/TSP group compared to AC group (S3 Figure) and it contained 11 from the 34 consistently up-regulated DEGs, including STAT1, VAMP5, WARS and FBX06.

The identification of upstream transcription factors and associated protein kinases was performed for the consistently up-regulated module M15 (Figure 2). The top TF for the analysis was IRF1 (Table 1). SYK was identified as the top upstream PK for the module M15 (Table 2). These top TF and PK previously found with X2K web were not described previously as drug targets for HAM/TSP according to the Open Targets database.

**DISCUSSION**

To the best of our knowledge this is the HAM/TSP gene expression study with the highest population sample size and the first that comprehensively integrate consistently DEGs and co-expression modules with PPI networks and identify upstream key regulatory transcription factors and protein kinases in order to identify potential drug targets.

This study suggested that IFN signaling is enhanced by a SYK-dependent activation of STAT1 and IRF1, this amplification can be an important mechanism of spinal cord injury in HAM/TSP. The IFN- inducible signature found with the up-regulated DEGs, the IFN module found in the PPI network and the enrichment for the IFN pathway in the consistently co-expressed module M15 supported the central role of IFN in this disease. Additionally, as the study of Tattermusch in 2012 (REF 10) included patients with different stages of the disease, our findings suggest that IFN exposure is maintained during the different stages of HAM/TSP. Moreover, STAT1 was found to be
a top ranked hub protein and one of the three genes that were DEGs in all of the GEO studies. IRF1 was the upstream top ranked transcription factor and it had a direct protein-protein interaction with STAT1. Finally, the predicted upstream interaction between SYK and the expanded protein interaction network indicated the key regulatory role of this protein.

This model is consistent with previous studies that claimed that the main mechanism of spinal cord damage by HTLV-1 is bystander damage instead of infection of resident cells in the CNS or mimicry of a host antigen by HTLV-1. Furthermore, it is well known that in several autoimmune diseases, the immune cells are exposed to chronic levels of different types of IFNs causing integration and amplification of IFN signaling. Priming is the process in which this chronic IFN exposure sensitizes immune cells to produce enhanced responses to extracellular signals including IFNs and is characterized by intracellular accumulation of STATs. Subsequently, this chronic exposure of the CNS tissue to inflammation can lead to neurodegeneration which is the typical progression in HAM/TSP.

An IFN-inducible signature was found in patients with HAM/TSP when compared with asymptomatic HTLV-1 carriers. This finding is consistent with previous studies which have reported an IFN signature in HAM/TSP cases that can distinguish between multiple sclerosis and HAM/TSP. Moreover, the IFN signature intensity has been correlated with the clinical severity of the disease. When comparing our HAM/TSP signature with the study published by Tattermusch in 2012, it was observed that there were eight common genes between both studies, these genes includes STAT1, GBP5, GBP1 and WARS.

Our HAM/TSP signature contains three genes which were up-regulated DEGs in the three GEO series, these genes were the Vesicle Associated Membrane Protein 5 (VAMP5), the Guanylate Binding Protein 5 (GBP5) and the Signal Transducer And Activator of Transcription 1 (STAT1), the latter gene was found to be the second highest ranked hub gene in the PPI network constructed with the up regulated DEGs. Additionally, STAT1 was also observed to be the top hub gene in the network built with the co-expression modules M5 and with the consistently co-expressed module M15.

The STAT family regulates the differentiation and response of immune cells and its members have a double function like cytoplasmic signal transducers as well as TFs. When a STAT binds a transmembrane receptor, it becomes a JAK substrate and then can be phosphorylated to create STAT dimers that target the nucleus. Notably, mutations in several members are associated with autoimmunity disorders as well as primary immunodeficiency syndromes. The association between HAM/TSP and STAT1 is supported by several studies including the one that found that a functional
FAS -670 polymorphism in a STAT1 binding site is associated with sporadic and familial aggregation cases of HAM/TSP. It has also been proposed that IFN beta does not produce the expected antiproliferative effect in HTLV-1 infected T cells, partially because there is an increase proportion of phosphorylation of the Tyr 701 in STAT1 beta compared with STAT1 alpha which is the only subunit associated with transcriptional activation. Finally, it is known that Tax is a highly immunogenic HTLV-1 gene that can induce the gene expression of STAT1.

It was found that STAT1 had a direct protein-protein interaction with the Interferon Regulatory Factor 1 (IRF1), this transcription factor was consistently predicted as the most enriched transcription factor in the upstream of the consistently up regulated DEGs and the consistently co-expressed module M15. The interaction between these two TFs is consistent with previous studies that have reported that IRF1 can be either a primary target of STAT1 or it can create a complex with unphosphorylated STAT1. IRF1 binds to interferon stimulated response elements (ISRE) similar to STAT1. Increased expression of IRF1 promotes type I IFN-accelerated lupus nephritis and it has also been associated with other autoimmune diseases like rheumatoid arthritis, juvenile idiopathic arthritis and Sjogren syndrome. Interestingly, STAT1 and IRF1 can interact with each other and bind IFN enhancers together or separately in order to influence the IFN gamma responsiveness. However, most STAT1 binding sites take place at or near to IRF1 sites, but IRF1 often binds isolated from STAT1.

SYK is a non-receptor associated tyrosine kinase activated by binding to receptors containing immunoreceptor tyrosine-based activation motifs (ITAMs) like the Fc gamma receptor. It is expressed in all hematopoietic cells and in several non-hematopoietic cells like neurons present in the hippocampus, cerebellum, nervous tracts and endothelial cells of neurovascular capillaries. Although it has not been previously involved in the pathogenesis of HAM/TSP, it might be a key regulator in the process of IFN priming and amplification as was suggested by our predictions of upstream key regulators of the gene expression found in our patients. Notably, a previous study found that IFN gamma subthreshold concentrations enhanced activation of IFN alpha induced STAT1 and inflammatory STAT1 target genes by a mechanism that is dependent on Syk and on adaptor proteins that activate SYK through ITAMs. Additionally, it was found that increased STAT1 contributed to the enhanced IFN alpha induced activation in IFN gamma primed cells. SYK has been associated with several autoimmune conditions, such as rheumatoid arthritis, systemic lupus erythematosus, systemic sclerosis and thrombocytopenic purpura. The SYK inhibitor fostamatinib was demonstrated to limit tissue damage in a systemic sclerosis mouse model. It has been also found that the application of naked SYK siRNA into joints inhibits the development of arthritis.
Table 1. Transcription factors enrichment analysis shows a ranked list of the top TFs that are predicted to regulate the gene list of the consistently co-expressed module M15 in HAM/TSP.

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>P Value</th>
<th>Z score</th>
<th>Combined score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRF1</td>
<td>1.60E-17</td>
<td>-2.56</td>
<td>98.98</td>
</tr>
<tr>
<td>ATF3</td>
<td>2.46E-08</td>
<td>-1.45</td>
<td>25.4</td>
</tr>
<tr>
<td>SOX2</td>
<td>1.37E-07</td>
<td>-1.6</td>
<td>25.33</td>
</tr>
<tr>
<td>FOXA1</td>
<td>9.74E-06</td>
<td>-1.48</td>
<td>17.1</td>
</tr>
<tr>
<td>IRF8</td>
<td>1.55E-05</td>
<td>-3.27</td>
<td>36.21</td>
</tr>
</tbody>
</table>

Table 2. Kinase enrichment analysis shows a ranked list of the top protein kinases predicted to be the likely regulators of the expanded protein interaction network for the consistently co-expressed module M15 in HAM/TSP.

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>P Value</th>
<th>Z score</th>
<th>Combined score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYK</td>
<td>0.0035</td>
<td>-2.05</td>
<td>11.59</td>
</tr>
<tr>
<td>BMX</td>
<td>0.01643</td>
<td>-1.9</td>
<td>7.81</td>
</tr>
<tr>
<td>FGFR4</td>
<td>0.02077</td>
<td>-1.88</td>
<td>7.27</td>
</tr>
<tr>
<td>MAP3K10</td>
<td>0.02402</td>
<td>-1.75</td>
<td>6.53</td>
</tr>
<tr>
<td>FLT1</td>
<td>0.02618</td>
<td>-1.96</td>
<td>7.12</td>
</tr>
</tbody>
</table>

LIMITATIONS AND FUTURE STUDIES: This study has some important limitations such as the low number of GEO series available for the analysis. Moreover, the three GEO series selected were from two different tissues and used different microarray platforms. In addition, important demographic and clinical information was lacking in order to explore the effect of potential covariates or confounding factors. There was not enough information about the inclusion and exclusion criteria for the cases and controls, including information regarding treatment status, comorbidity and disease severity. Finally and more importantly, future experimental studies will be needed to validate these results.

CONCLUSIONS

In conclusion, two transcription factors (STAT1 and IRF1) and one protein kinase (SYK) were identified to be key molecular regulators of the IFN signaling amplification found in our patients, this mechanism can lead to an exaggerated neuroinflammatory response to HTLV-1 in the spinal cord and thus, these molecules represent potential
drug targets for the chronic debilitating and neglected condition known as HAM/TSP. Although in silico models based on public available transcriptomic data are useful to create new hypothesis, it is fundamental to validate these findings in future experimental studies.

CONFLICTS OF INTEREST:
I declare that I have not conflicts of interest

ACKNOWLEDGMENTS:
I thank to all the members from the Computational Systems Biology Laboratory (CSBL) from the University of Sao Paulo where this project was designed and executed. I specially thank to Dr Helder Nakaya who accepted to be my thesis supervisor and orientated me in all the phases of this project. Cesar Prada helped me in having discussions about the bioinformatics analysis and finally, Mindy Muñoz, Fernando Marcon, Patricia Gonzales, Lucas Cardozo and Pedro Russo who gave important feedback about the results. I also want to thank to Dr Ola Spjuth who accepted to be my thesis examiner. Finally, I thank to much to The Swedish Institute for giving me the scholarship that allowed me to pursue this master thesis and degree.

REFERENCES
A SYK-dependent activation of STAT1-IRF1 amplifies the IFN signaling in HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP)

2018-05-22


Supporting information.

**S1 Table.** Enrichment analysis of consistently up regulated DEGs

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Adjusted P-value</th>
<th>Z-score</th>
<th>Combined Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interferon Signaling</td>
<td>3.596E-06</td>
<td>-2.103E+00</td>
<td>3.764E+01</td>
</tr>
<tr>
<td>Cytokine Signaling in Immune system</td>
<td>2.148E-05</td>
<td>-2.396E+00</td>
<td>3.621E+01</td>
</tr>
<tr>
<td>Immune System</td>
<td>8.281E-05</td>
<td>-2.207E+00</td>
<td>2.952E+01</td>
</tr>
<tr>
<td>Interferon gamma signaling</td>
<td>2.148E-05</td>
<td>-1.769E+00</td>
<td>2.656E+01</td>
</tr>
<tr>
<td>Class I MHC mediated antigen processing &amp; presentation</td>
<td>3.493E-03</td>
<td>-2.158E+00</td>
<td>1.992E+01</td>
</tr>
</tbody>
</table>

**S2 Table.** Enrichment analysis of consistently down regulated DEGs

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Adjusted P-value</th>
<th>Z-score</th>
<th>Combined Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signaling by Retinoic Acid</td>
<td>0.013</td>
<td>-2.020</td>
<td>15.883</td>
</tr>
<tr>
<td>Arachidonic acid metabolism</td>
<td>0.013</td>
<td>-2.009</td>
<td>14.857</td>
</tr>
<tr>
<td>Metabolism</td>
<td>0.045</td>
<td>-2.229</td>
<td>10.894</td>
</tr>
<tr>
<td>Metabolism of lipids and lipoproteins</td>
<td>0.045</td>
<td>-2.166</td>
<td>10.003</td>
</tr>
<tr>
<td>Synthesis of bile acids and bile salts via 24-hydroxycholesterol</td>
<td>0.045</td>
<td>-2.002</td>
<td>9.268</td>
</tr>
</tbody>
</table>
S3 Table. Abbreviators

AC: Asymptomatic HTLV-1 carriers
ATL: Adult T-cell leukemia
CNS: Central nervous system
DEGs: Differentially expressed genes
DLAT: Dihydrolipoamide s-acetyltransferase
GEO: Gene Expression Omnibus
HIST1H3: Histone cluster 1 h3 family member B
IFN: Interferon
IRF1: Interferon Regulatory Factor 1
IRF8: Interferon Regulatory Factor 8
FC: Fold change
FOXP3: Forkhead Box P3
GBP1: Guanylate binding protein 1
GBP5: Guanylate binding protein 5
HAM/TSP: HTLV-1 associated myelopathy/tropical spastic paraparesis
HTLV-1: Human T-lymphotropic virus
ITAMs: Immunoreceptor tyrosine-based activation motifs
ISG15: ISG15 Ubiquitin-like modifier
ISRE: Interferon stimulated response elements
LIMMA: Linear Models for Microarray Analysis
PPI: Protein-protein interaction
PK: Protein kinase
SDF1: Chemokine stromal cell derived factor 1
SNP: Single nucleotide polymorphism
SOX2: SRY-Box 2
STAT1: Signal transducer and activator of transcription 1
SYK: Spleen associated tyrosine kinase
TF: Transcription factor
TNF: tumor necrosis factor
UBC: Ubiquitin C
VAMP5: Vesicle associated membrane protein 5
WGCNA: Weighted gene co-expression network analysis
X2K: Expression2Kinases
S1 Figure. HAM/TSP signature. The heat map shows the HAM/TSP signature for the GEO studies. In blue color are the consistently down-regulated DEGs and in red colors appears the consistently up-regulated DEGs.
**S2 Figure.** PPI network of consistently up regulated DEGs depicting the hubs and the interferon module in blue. The size of the node is proportional to the number of interactions that a certain gene (node) has in the network. The direct PPI between STAT1 and IRF1 can be seen. UBC, STAT1 and FBX06 appears as the most connected hubs.
S3 Figure. Gene expression profile of consistently co-expressed modules across GEO series.