

Comprehensive analysis of small RNA profile by massive parallel sequencing in HTLV-1 asymptomatic subjects with monoclonal and polyclonal rearrangement of the T-cell antigen receptor γ -chain

Daniela Ragner Valadão de Souza^{1§}, Rodrigo Pessôa^{1§}, Youko Nukui², Juliana Pereira², Jorge Casseb¹, Augusto César Penalva de Oliveira³, Patricia Bianca Clissa⁴, Sabri Saeed Sanabani^{1*}

¹ Laboratory of Dermatology and Immunodeficiency, São Paulo University Medical School, Sao Paulo, SP, Brazil.

² Departments of Hematology, Faculty of Medicine, University of São Paulo, São Paulo, Brazil.

³ Institute of Infectious Diseases "Emilio Ribas" (IIER), São Paulo, São Paulo, Brazil.

⁴ Immunopathology Laboratory, Butantan Institute, Sao Paulo-SP, Brazil

* Corresponding author

Sabri Saeed Sanabani, PhD

E-mail: sabyem_63@yahoo.com

Laboratory of Dermatology and Immunodeficiency, LIM56/03.

Instituto de Medicina Tropical de São Paulo

Faculdade de Medicina da Universidade de São Paulo

Av. Dr. Eneas de Carvalho Aguiar, 470 3º andar

São Paulo Brazil 05403 000

Phone: + 5511 3061 7194 ext:218

§These authors contributed equally to this work

Running title: small RNA profile in HTLV-1 asymptomatic subjects

Abbreviations: sRNA: small RNA, miRNA: microRNA, ncRNA: non-coding RNA
tRNA: transfer RNA, rRNA: ribosomal RNA, snRNA : small nuclear RNA, snoRNA :
small nucleolar RNA, scRNA: small cytoplasmic RNA, PBMC: peripheral blood
mononuclear cell, HTLV-I: Human T-lymphotropic virus type I, ASM: asymptomatic
monoclonal, ASP: asymptomatic polyclonal, ATLL: adult T-cell leukemia/lymphoma

Keywords: small RNA, HTLV-1, T-cell antigen receptor, asymptomatic carriers,

massive parallel sequencing

Abstract

Introduction

In this study, we used a massive parallel sequencing technology to investigate the cellular small RNA (sRNA) operating in peripheral blood mononuclear cells (PBMCs) of the Human T-lymphotropic virus type I (HTLV-I) infected asymptomatic subjects with a monoclonal and polyclonal rearrangement of the T-cell antigen receptor γ -chain.

Materials and Methods

Blood samples from 15 HTLV-1 asymptomatic carriers who were tested for clonal *TCR- γ* gene (seven and eight subjects presented monoclonal and polyclonal expansion of HTLV-1 infected cells, respectively), and were submitted to Illumina for small RNA library construction. sRNA libraries were prepared from cryopreserved PBMCs using TrueSeq Small RNA Library Preparation Kit (Illumina). The sRNA-Seq reads were aligned, annotated, and profiled by different bioinformatics tools.

Results

Through bioinformatics analysis, we identified a total of 494 known sRNAs and 120 putative novel sRNAs. Twenty-two known and 15 novel sRNA showed a different expression (>2 -fold) between the asymptomatic monoclonal (ASM) and asymptomatic polyclonal carriers (ASP). The hsa-mir-196a-5p was the most abundantly upregulated micro RNA (miRNA) and the hsa-mir-133a followed by hsa-mir-509-3p were significantly downregulated miRNAs with more than a three-fold difference in the ASM than ASP group. The target genes predicted to be regulated by the differentially expressed miRNAs play essential roles in diverse biological processes including cell proliferation, differentiation, and/or apoptosis.

Discussion

Our results provide an opportunity for a further understanding of sRNA regulation and function in HTLV-1 infected subjects with monoclonality evidence.

Introduction

Human T-lymphotropic virus type I (HTLV-I) is an oncogenic human retrovirus that was first isolated from a T-cell line, HUT102, that had been obtained from a patient with adult T-cell leukemia/lymphoma (ATLL) [1]. Globally, there are an estimated 5-10 million individuals who carry the HTLV-I, with the important caveat that the prevalence remains barely unknown in many areas of the world [2]. The disease burden is unevenly distributed, with a higher incidence of the disease particularly in southwest Japan, the Caribbean islands, South America, and portions of Central Africa [3]. Infection with HTLV-I may result in a spectrum of clinical manifestations, ranging from asymptomatic infection to a number of human disorders, most notably a malignant ATLL and a chronic progressive neuromyelopathy, termed HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP) [4]. The majority of HTLV-I-infected individuals remain asymptomatic for life, while some individuals progress to a pre-leukemic phase that is characterized by small numbers of circulating leukemic cells in the peripheral blood, skin lesions, and a lack of involvement of other organ systems [5]. Only 2.5% to 5% of the virus carriers eventually develop ATLL after a long asymptomatic period [6, 7]. The reason why some people develop the disease, whereas others remain healthy, is likely dependent on both host-related and virus-related factors [8]. Available evidence from molecular studies indicates that the impairment of various cellular functions by viral genes (e.g., *tax* and *HBZ*), genetic and epigenetic changes including DNA methylation, and the host immune system may contribute to the leukemogenesis of ATLL [9-11]. Despite the effective immortalization of the T cells, the markedly prolonged incubation period (>30 years) prior to the onset of ATLL, suggests an additional acquisition of genetic changes beside the viral infection to contribute to the pathogenesis [12].

Infections with HTLV-1 are expected to produce an initial polyclonal T-cell proliferation followed by a monoclonal malignant transformation in asymptomatic HTLV-1 carriers before the diagnosis of Adult T-cell Leukaemia/Lymphoma becomes evident. HTLV-1 carriers who have monoclonal integration of HTLV-1 provirus DNA in their mononuclear cells are suggested to be a high-risk group for development of ATLL, but their prognosis varies from being stable long-term carriers to experiencing the development of ATLL [13-16]. Carvalho and Da Fonseca Porto [17] also found a correlation between monoclonal integration of provirus DNA and abnormal lymphocytes in peripheral blood, with a trend for greater severity of the parasitic infection.

Over the past several years, some studies have provided significant evidence to support transcription of many non-protein-coding regions of the mammalian genome [18], yielding a complex network of transcripts that include tremendous numbers of non-coding RNAs (ncRNA). These molecules play significant roles in normal biological processes and in a variety of human diseases [19]. Within this diverse menagerie of ncRNA, small RNAs (sRNAs), have emerged as potential posttranscriptional regulators of gene expression in both eukaryotes and prokaryotes [20]. Because these molecules are highly complex in terms of structural diversity and function, their families have been grouped into structural and regulatory ncRNA [21]. The structural ncRNA includes transfer RNA (tRNA) and ribosomal RNA (rRNA), as well as other small but stable noncoding RNAs, such as small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), small cytoplasmic RNA (scRNA), Ribonuclease P (RNase P), mitochondrial RNA processing (MRP) RNA, signal recognition particle (SRP) RNA, and telomerase RNA. Regulatory ncRNAs include microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs), and long ncRNAs (lncRNAs) [22]. The miRNAs are

perhaps the single best-understood subclass. They are typically 18-25 nucleotides (nt) in length, single-stranded RNAs that have emerged as principal post-transcriptional regulators of gene expression and play a vital role in several cell processes including cell proliferation, differentiation, apoptosis, embryonic development, and tissue differentiation [23, 24]. The mature molecules of miRNAs are assembled into ribonucleoprotein complexes called miRNA-induced silencing complexes (RISC), which inhibits gene expression by perfect complementary binding, for mRNA degradation, or imperfect binding at the 3'UTR region, to inhibit translation[25]. The rigid control of miRNA expression is critically important to maintain cells in normal physiological states [26], while overexpression of miRNAs has been associated with the occurrence and development of various diseases [25, 27]. It has been proven that oncogenic (onco-miRs) and tumor-suppressor (TS-miRs) miRNAs exert control over important cell cycle components and thus results in either acceleration or deceleration of the cell cycle [28-30].

In theory, dysregulation of host cells miRNAs by HTLV-1 might influence the development of ATLL. Indeed, some reports attest to the critical role of cellular miRNA to proliferation and survival of HTLV-1-infected T-cells. Hybridization-based methodologies, such as microarray and PCR-based assays have been used thus far to identify and profile the cellular miRNAs in HTLV-1 infected cell lines [31, 32]. The study by Pichler *et al.* [33], who first used quantitative PCR to study the interconnections between HTLV-1 and cellular miRNAs, confirmed that HTLV-1 transforms host cells by inducing dysregulation of the expression of specific miRNAs, including miR-146a, which is upregulated by Tax protein (an oncoprotein of HTLV-1). Subsequently, other expression studies revealed both up- and downregulation in a number of miRNAs in HTLV-1/ATLL cell lines and primary ATLL cells [34, 35].

Recently, Ruggero and colleagues [36] employed massive sequencing to identify the repertoire of microRNAs and tRNA fragments (tRFs) expressed in HTLV-1-infected cells compared to normal CD4⁺ T cells. Most of the previous studies mainly focused on expression of specific or diverse miRNA in HTLV-1 infected cell lines, while few studies have focused on determining the expression profiles of sRNAs isolated from human blood samples. Thus, it would be clinically relevant to investigate early gene regulatory mechanisms associated with cell transformation in HTLV-1 asymptomatic patients with monoclonal expansion features.

In this study, we employed Illumina massive parallel sequencing technology to comprehensively characterize the sRNA expression profiles in the peripheral blood mononuclear cells (PBMCs) of HTLV-1 infected asymptomatic subjects with monoclonal (ASM) and polyclonal (ASP) T cell expansions. Our results revealed 494 known sRNAs and 120 putative novel sRNAs. Moreover, 23 known and 15 novel sRNA showed different expressions (>2-fold) between the ASM and the ASP groups. The data obtained in this study provides considerable insight into understanding the expression characteristic of sRNAs in HTLV-1 carriers with monoclonal T cell population. Differentially expressed sRNAs described here may be associated with the transformation of T cells.

Materials and Methods

Patients' characteristics and sample preparations

Peripheral blood samples were initially collected from a relatively large group of asymptomatic HTLV-1 carriers (n = 367). Of these subjects, 15 carriers were selected for this study on the basis of clonality results and classified to ASM (n= 7) and ASP (n=8). The samples were collected after approval by the institutional health research ethic authority and the signing of written informed was obtained from all individual participants included in the study. Isolation of PBMCs was carried out by Ficoll-Hypaque (Amersham, Upsala, Sweden) density centrifugation, washed twice in RPMI 1640 with 10 % fetal calf serum, and stored in liquid nitrogen until use. A summary of the clinical characteristics of study subjects is shown in **Table 1**.

Genomic DNA and RNA extraction

Genomic DNA was extracted from PBMCs using the QIAamp blood kit (QIAGEN, Tokyo, Japan). The miRNeasy Mini Kit (Qiagen, Hilden, Germany) in conjunction with the TRIzol (Life Technologies, USA) procedure were used to extract total RNA and sRNA following the manufacturer's protocols. Briefly, 700 μ L of TRIzol was added to 200 μ L of cryoperserved PBMCs followed by incubation at room temperature for 5 min. Chloroform (Sigma-Aldrich, St. Louis, MO, USA) was added, and the samples were vortexed and incubated at room temperature for 5 min followed by centrifugation at 12,000 \times g for 15 min at 4° C. The aqueous phase containing RNA was transferred to a new tube and isopropanol (Fisher-Scientific, Thermo Fisher Scientific, Waltham, MA, USA) was added. The samples were incubated at room temperature for 10 min followed by centrifugation at 12,000 \times g for 15 min at 4° C. The pellet was washed with 75% Ethanol (Sigma-Aldrich), air-dried, and resuspended in 20 μ L nuclease-free water

(Ambion). For addition of carriers, 2 μL of glycogen (5 $\mu\text{g}/\mu\text{L}$) was added during the isopropanol precipitation step. For the miRNeasy Mini Kit, the aqueous phase was added to 1.5 volumes of 100% ethanol, and the mixture was mixed thoroughly by pipetting. The supernatants were then transferred to the miRNeasy Mini spin column by centrifugation. Thereafter, the column was washed with 700 μl of RWT, 500 μl of RPE, and 500 μl of 80% ethanol, and then dried by centrifugation at 12 000 $\times g$ for 5 min. Finally, the sRNAs on the membrane were eluted in 22 μl of RNase-free water and stored at -80°C until further use.

Concentration measurement of genetic materials

Both DNA and RNA including sRNA concentration measurements by fluorimetry were done on Qubit 2.0 fluorometer using Qubit® DNA or RNA HS Assay Kit (Thermo Fisher Scientific), respectively. Briefly, 10 μL standards or diluted genetic materials were mixed with Qubit® Working Solution and incubated for 2 min. Two standards were prepared, one has 0 $\text{ng}/\mu\text{L}$ and the other has 10 $\text{ng}/\mu\text{L}$ concentrations. To obtain the calibration curve, the relative fluorescence was plotted against the concentrations of the two standards. The fluorometer built-in software was used to compute the concentration values of the genetic materials of each sample.

HTLV-1 proviral load determination

The extracted DNA was used as a template to amplify a 97-bp fragment from the HTLV-1 *tax* region using previously published primers [37] and protocol [38]. Amplification and analysis were performed with the Applied Biosystems 7500 real-time PCR system. The standard curves for HTLV-1 *tax* were generated from MT-2 cells of \log_{10} dilutions (from 10^5 to 10^0 copies). The threshold cycle for each clinical sample was calculated by defining the point at which the fluorescence exceeded a threshold limit. Each sample was assayed in duplicate, and the mean of the two values was

considered the copy number of the sample. The HTLV-1 proviral load was calculated as the copy number of HTLV-1 (*tax*) per 1000 cells = (copy number of HTLV-1 *tax*)/(copy number of *RNase P* gene/2) × 1000 cells. The method could detect one copy per 10³ PBMCs.

Analysis of T cell receptor γ TCR genes

A DNA-based polymerase chain reaction (PCR) of rearranged γ TCR genes was performed according to the previously described protocol [39]. All patients' PCR products were analyzed with the 3130 ABI Prism capillary electrophoresis equipment. 0.5 μ l ROX, 13 μ l Hidi, and 1 μ l template DNA sample were added to each well in a 96-well plate. Data were analyzed using Genescan and Genotyper software (Applied Biosystem, Foster City, CA). T cell clonalities were blindly determined by visual examination of the electropherograms by two analysts and further confirmed by an expert hematology pathologist (coauthor JP).

sRNA construction and massive parallel sequencing (MPS)

For each sample in both groups, sRNA libraries were prepared with the Small RNA v1.5 sample preparation kit as per the manufacturer's instructions (Illumina, San Diego, CA) and previous protocol [40]. Briefly, 5 μ l of purified total RNAs were ligated with 1 μ l RNA 3' Adapter and then with a 5' RNA adapter (Illumina, San Diego, CA). The 5' adapter also included the sequencing primer. After RT-PCR amplification, the resulting products were analyzed using polyacrylamide gel electrophoresis (PAGE) (6% Novex Tris-borate-EDTA [TBE] PAGE; Invitrogen). After gel electrophoresis, sRNA bands at sizes 145-150 bp were excised and purified. Finally, each four libraries were pooled and up to 8-10 pM of the pooled libraries were loaded and sequenced on the MiSeq platform (Illumina) with a 36 base single-end protocol, according to the manufacturer's instructions.

sRNA Data analysis and interpretation

Base calling, demultiplexing, and trimmed FASTQ files were generated using the MiSeq reporter. Only high quality reads with score >30 on the Sanger scale were considered for further analysis. The reads were aligned against the whole genome build: hg19 using Strand NGS v3.1. The later software package was also used for analysis of novel molecule discoveries and interpretations. The distributions of the sRNA data in each clinical condition were conducted according to the quantile normalization algorithm, with a baseline transformation set to the median of all samples. Novel sRNA were discovered and classified by the decision tree method with three-fold validation accuracy using model previously described by Langenberger *et al.* [41]. Additionally, only sRNA sequences meeting the minimum read coverage criterion of >20 were considered as novel or known sRNA and were included in further analyses. sRNAs with fold change more than 2.0 were considered to be differentially expressed. All the sRNA raw data generated by MPS has been deposited to the Zenodo repository.

Functional annotation and pathway analysis of miRNA target genes

The target genes from differentially expressed miRNAs were predicted by PicTar, TarBase microRNAorg databases. Pathway analysis was carried out for a functional analysis of mapping genes to KEGG pathways [42]. Prediction was performed by using DIANA-microT-CDS [43]. The biological functions of these target genes were determined by the enriched GO analysis (false discovery rate p -value <0.05).

Results

Patient Characteristics

The characteristics of the subjects in both groups are shown in **Table 1**. All eight participants within the ASP group were female and the median age was 58 years (range, 31-80 years). Six females and one male were assigned to the ASM group and the median age of the entire group of subjects was 61 years (range, 28-74 years). HTLV-1 proviral load levels varied from one copy to 208 copies/ 10^3 PBMCs in the ASP group and varied from undetectable to 327 copies/ 10^3 PBMCs in the ASM group (two-tailed *P* value = 0.7).

Description of whole genome sRNA-sequencing data from both groups

Illumina MPS revealed a total of 130,415,317 reads. After removal of low quality data and reads that failed vendor's QC, 125,435,316 reads were retained. The filtered sRNA reads on quality metrics for each sample used in the alignment to the human genome sequence dataset are shown in **Table 1**. For instance, sample 131 within the ASP group had 3,779,138 filtered sRNA reads out of 3,944,199 of raw reads and sample 143 in the ASM group displayed 5,472,836 of filtered sRNA reads out of 5,719,645 raw reads. We next investigated the percentage of reads assigned to abundant sRNA categories or sRNA biotypes (**Figure 1**, representative for samples 131 and 143). More than 69% and 59% of the reads in samples 131 and 143, respectively, matched to mature miRNA. The rest of the sequences were matched into other types of sRNAs including intronic RNAs, exonic RNAs, scRNAs, snRNAs, snoRNAs, tRNAs, and others. Next, we evaluated the location of these tags on the chromosome. In the majority of samples from both groups, most of the unique tags were located on chromosome 3, 8, 12, and 19 (**table S1**). The Illumina MPS yielded 614 sRNA molecules, of which 494 sequences were derived from known sRNA, and 120 were novel sRNA. Of the 494 known sRNA, 219 were miRNA,

30 scRNA, 9 scRNA_pseudogene, 162 snoRNA, 23 snRNA, and 51 tRNA (**table S2**). Of the 120 novel genes, 25 were novel miRNA, 50 snoRNA, 3 snRNA, 9 tRNA, and 33 unknown (**table S3**). Considering -3p and -5p mature forms of miRNA, our analysis revealed a total of 2104 mature miRNA, of which 22 had had greater than a two-fold change in expression (**table 2**).

Characteristics of the abundant known and novel sRNAs

Among the 494 known sRNAs identified in this study, 52 sRNAs were found to be differentially expressed (>2-fold) between the ASM group and ASP group, which accounted for 8.5% (52/614) of all the sRNAs investigated (**table S4**). Clustering by group was not noted in the hierarchal combined tree of the 52 significantly differentially expressed sRNA levels as shown in **figure 2**. Of these 52 sRNAs, 50 were upregulated and two were downregulated in the ASM group. The upregulated sRNA includes three miRNA, 10 scRNA, one scRNA_pseudogene, 21 snoRNA, one snRNA, and 14 tRNA. On the other hand, the down-regulated entities include only two miRNA. Of the highly expressed sRNAs, SNORD45, SNORD45B, hsa-mir-196a-1, hsa-mir-196a-2, and SCARNA5 were the top five most abundantly expressed sRNA in PBMCs of the ASM group, among which, SNORD45 had greater than 5.7 fold changes in expression (**table S4**). The most abundantly downregulated sRNAs in PBMCs of the ASM group were hsa-mir-133a-1 and hsa-mir-133a-2, which showed more than 3.8 fold changes in the ASM when compared to the ASP group (**table S4**). Further, of the 22 mature miRNAs that exhibited a >2-fold difference in expression in the PBMCs of the ASM compared to the ASP group, 11 miRNAs were upregulated and 11 were downregulated (**table 2**). The hierarchical clustering analysis of these miRNAs showed distinct patterns of miRNA expression levels between ASM and ASP groups (**Figure 3**). The hsa-mir-196a-5p was the most abundantly upregulated miRNA and the hsa-mir-133a followed

by hsa-mir-509-3p were the significantly downregulated miRNAs with more than a three-fold difference in the ASM than the ASP group. Of the 120 novel sRNA described in this study and shown in **table S3**, 15 were found to be differentially expressed (>2-fold) between the ASM group and ASP group, which accounted for 2.4% of all the sRNAs investigated. Of these, 14 were upregulated and one was downregulated in the PBMCs of the ASM group when compared to the ASP group. Of the upregulated novel sRNAs, four were snoRNA, one miRNA, one tRNA, and eight were unknown while the unique downregulated gene was snoRNA (**Table 3**).

GO Analysis of the Target Genes of the Differentially Expressed sRNAs

The role of a miRNA is ultimately defined by the genes that it targets and by its effect on the expression of those genes. To discover these functions, the target genes of the two groups of the mature significantly expressed miRNAs were predicted using the three programs PicTar, TarBase, and microRNAorg databases. The intersection of these three programs was used to reduce the false-positive rate. As shown in **Table S5**, a *p*-value cutoff of 0.001 was selected to predict nine target genes. The most significantly selected candidate target gene FGD6 is regulated by 13 of the 27 mature significantly expressed miRNAs. The second most selected target gene is the CUGBP Elav-like family member 6 (CELF6) gene, which is regulated by 10 of the 27 mature significantly expressed miRNAs. To identify the actual regulatory functions of the miRNAs, the GO annotation of the target genes was performed on the list of predicted target miRNAs that corresponded with altered miRNAs in the ASM group at *p* value <0.05. The results from this analysis revealed no significant GO enrichments for the differentially expressed genes predicted from the known miRNAs. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of target genes were assigned using mirPath v.3 in DIANA tools webserver. KEGG pathway annotation revealed that 3044 background

genes were annotated for 56 biological functions. The top five pathways were Adherens junction, Proteoglycans in cancer, Axon guidance, Morphine addiction, and Prion diseases.

Discussion

In this study, we used Illumina high-throughput sequencing technology to analyse the global expression of non-coding RNome in PBMCs of HTLV-1 infected asymptomatic subjects with monoclonal and polyclonal rearrangement of the T-cell antigen receptor γ -chain. Our results revealed 52 significantly differentially expressed sRNAs in the ASM group compared with the ASP subjects. Furthermore, we identified 15 novel sRNA exhibited much higher expression levels in ASM patients than reads in ASP. Eleven of the 22 differentially expressed mature miRNA were up-regulated while the other 11 were down-regulated in the ASM group. Out of the 22 most dysregulated miRs, we selected the most significantly up-regulated miRs (miR-196a-5p) and down-regulated miRs (miR-133a, miR-625-3p, and miR-3120-3p) assuming that the measurement of these miRs could discriminate ASM from ASP patients and that the dysregulation of these miRs in ASM may represent early markers for T cell transformation with unknown functional consequences.

We searched for putative target genes associated with enhancement of cellular transformation. The upregulation of miR-196a, located in HOX gene clusters [44], has been suggested to potentially target HOXB8, HOXC8, HOXD8 and HOXA7 [45]. HOX proteins consist of homeodomain-containing transcription factors that are primary determinants of cell fate during embryogenesis, organogenesis, and oncogenesis [46]. Reports from previous studies have proved that upregulation of miR-196a promotes cell proliferation, anchorage-independent growth, and suppressed apoptosis by targeting annexin A1 [47]. Although most studies proved that over expressed miR-196a is associated with oncogenic function in various types of cancers including pancreatic cancer [48], and oesophageal adenocarcinoma [49] but it has been suggested to play a

tumor suppressive role as well. For instance, Li *et al.* [50] have demonstrated a potential role of the up-regulated miR-196a in reducing *in vitro* invasion and *in vivo* spontaneous metastasis of breast cancer cells. Also, miR-196 has been previously demonstrated to play diverse biological functions involving immune response, inflammation, and virus defense [51]. It was also suggested that upregulation of miR-196a may be used in a novel strategy to prevent or treat hepatitis caused by C virus infection, and miR-196a may be valuable in the diagnosis and management of this disease [52, 53]. In relation to T cell leukemia, Coskun and colleagues [54] have shown that miR-196a and miR-196b are regulators of the oncogenic erythroblast transformation-specific (ETS) transcription factors of the ETS related genes (ERG). The ERG plays important physiological and oncogenic roles in hematopoiesis [55]. It is also a prognostic factor in a subset of adult patients with acute T cell leukemia [56]. Coskun and colleagues [54] provided evidence that both miR-196a and miR-196b expression were linked to immature immunophenotype (CD34 positive) in ATLL patients. These findings indicate miR-196a and miR-196b as ERG regulators and have potential implications in ATLL. Based on the above facts, we therefore assumed that the aberrant expression of miR-196a might contribute to the development and transformation of HTLV 1 infected cells in the ASM group.

The miR-133a (known to control HLA-G protein stability) is one of the canonical myomiRs that at the crossroads of the molecular of muscle cells, connecting between pathways for cell differentiation, development and maintenance, but also potentiates aberrant cell growth in a wide range of non-muscle cancers [57]. miR-133 was predicted to bind to both N and G transcripts of rabies viruses and this interaction possibly explains the lengthy dormant state of this virus in skeletal muscles during the early phase of infection [58, 59]. In a recent study, the expression change of miR-133a has

been shown to negatively affect the replication of dengue viruses when overexpressed in mammalian cells [60]. The deregulation of miR-133a has been reported in many tumor types of distinct origin, including ovarian [61], prostate [62], bladder [63], and head and neck squamous cell cancer [64]. In most of these cancers, the miR-133a together with other miRs is associated with negative regulation of Fascin. The latter is a global protein that is strongly upregulated in ATLL-derived cells and in CD4+ T-cells transformed by HTLV-1 or by the viral oncoprotein Tax [65]. The loss of miR-133a expression was also associated with a poor survival of tumor patients [66]. The down-regulated miR-133a together with dysregulation of other documented miRs in this study was positively and significantly correlated with some predicted potential target genes including the CUG-binding protein and embryonic lethal abnormal vision-type RNA-binding protein 3-like factor 6 (CELF6), calcium/calmodulin-dependent serine protein kinase interacting protein 2 (CASKIN2), exportin 1 (XPO1), and the solute carrier family 33 member 1 (SLC33A1). The CELF6 protein is an important regulator of cell-specific alternative splicing during normal development and disease [67]. Altered expression of CELF6 has been associated with renal disorders and body weight changes [68, 69]. CASKIN2 is a neuronal signaling scaffolding protein involved in signaling neuronal synapses. In vitro experiments revealed that overexpression of Caskin2 inhibits serum-induced endothelial cells (EC), proliferation, and DNA synthesis but promotes EC survival during serum starvation, thus suggesting a role of this protein in preventing EC dysfunction in vivo [70]. The human chromosomal maintenance 1 or XPO1 mediates the nuclear export of cellular proteins (cargos bearing a leucine-rich nuclear export signal) and of RNAs. Various viruses, among them HIV-1, HTLV-1 and influenza A use it to export their unspliced or incompletely spliced RNAs out of the nucleus. The XPO1 interacts with, and mediates the nuclear export of HTLV-1 Rex

proteins and critically needs it for its multimerization [71, 72]. The SLC33A1 is an important substrate for a large variety of biochemical reactions occurring in the cell [73].

The search for the putative target RNA molecules in the database showed that miR-625-3p can potentially target the gene that codes for the clathrin heavy chain (CLTC) protein. Clathrin is the major protein component of the coat that surrounds the cytoplasmic face of the organelles (coated vesicles) mediating selective protein transport [74]. Clathrin coats play an essential role in receptor-mediated endocytosis, localization of resident membrane proteins to the compartments within the trans-Golgi network, and transport of proteins to the lysosome/vacuole [75, 76]. The clathrin coats contain both clathrin and four adaptor complexes (APs 1-4), which participate in the selection of protein cargo and their incorporation into clathrin-coated transport vesicles [77]. Previous studies showed that AP-2 and AP-3 interact with HIV-1 Gag and contribute to its trafficking and release [78, 79] and that the disruption of this interaction enhances viral release, while reducing the infectivity of the virions produced [78]. Camus and colleagues, 2007 [80] propose that in HIV-1 and murine leukemia virus, the AP-1 promotes Gag release by transporting it to intracellular sites of active budding, and/or by facilitating its interactions with other cellular partners. The same study concluded that this interaction is conserved among different retroviruses including HTLV and RSV and suggested that AP-1 could also play a role in the assembly of these retroviruses. In human disease, the few available data indicate that the abnormalities and dysregulation of CLTC are associated with large B-cell lymphoma [81], inflammatory myofibroblastic tumor [82], and multinodular goiter pathogenesis [83].

miR-3120-3p was down-regulated (fold change = -3.154) in the ASM group, and was negatively correlated with diverse putative target genes such as the SLC33A1, tripartite

motif containing 27 (TRIM27), XPO1, CASKIN2, and CELF6. The TRIM27 is a member of a large family of proteins that function as Really Interesting New Gene [82] E3 ubiquitin ligases. The TRIM27 has the ability to inhibit the class II phosphatidylinositol 3 kinase C2 β activity, resulting in decreased KCa3.1 channel activity and decreased TCR-stimulated Ca²⁺ influx and cytokine production, thereby identifying TRIM27 as a unique negative regulator of CD4 T cells [84].

There are few microarrays studies that have also characterized the miRNA expression profiles in HTLV-1/ATLL cell lines and ATLL patients [33-35, 85]; however, it is difficult to compare their results with our study due to the different methods employed, different research designs, and different target population.

Our study has several limitations, particularly its retrospective design, and lack of a control non-infected group. Additionally, few samples were included, and no validation study was performed. In spite of these caveats, the current study unravels a number of known and novel subsets of dysregulated sRNA including signature in the PBMCs of the ASM group. The aberrant expression of sRNAs paves the way for assessment of their mechanistic roles and clinical utilities as biomarkers, prognostic indicators, and/or therapeutic targets in early HTLV-1 pathogenesis.

In conclusion, significantly differentially expressed sRNAs including miRNAs and their target genes between ASM and ASP were identified by means of MPS technique. These newly identified miRNAs and target could have unexplored functions in HTLV-1 pathogenesis. Therefore, further investigation of a large population of leukaemia patients using well-matched control groups and sequential samples will be required to elucidate the possible pathogenic impact of these ncRNA molecules in HTLV-1 infection.

Acknowledgements

This study was funded by the Fundação de Amparo à Pesquisa do Estado de São Paulo (grants 2011/12297-2 and 2014/24596-2)

Conflict of Interest

The authors declare that they have no conflict of interest.

Author Contributions

Conceived and designed the experiments: SSS. Performed the experiments: DRVS, RP, SSS. Analyzed the data: DRVS, RP, PBC, SSS. Contributed reagents/materials/analysis tools: YN JP. Wrote the paper: PBC,SSS. Attending physician for sample collection: YN JP JK ACS.

References

1. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC: Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci U S A* 1980, 77(12):7415-7419.
2. Gessain A, Cassar O: Epidemiological Aspects and World Distribution of HTLV-1 Infection. *Front Microbiol* 2012, 3:388.
3. Matsuoka M: Human T-cell leukemia virus type I and adult T-cell leukemia. *Oncogene* 2003, 22(33):5131-5140.
4. Verdonck K, Gonzalez E, Van Dooren S, Vandamme AM, Vanham G, Gotuzzo E: Human T-lymphotropic virus 1: recent knowledge about an ancient infection. *Lancet Infect Dis* 2007, 7(4):266-281.
5. Franchini G, Ambinder RF, Barry M: Viral Disease in Hematology. *Hematology Am Soc Hematol Educ Program* 2000:409-423.
6. Murphy EL, Hanchard B, Figueroa JP, Gibbs WN, Lofters WS, Campbell M, Goedert JJ, Blattner WA: Modelling the risk of adult T-cell leukemia/lymphoma in persons infected with human T-lymphotropic virus type I. *Int J Cancer* 1989, 43(2):250-253.
7. Yamaguchi K, Watanabe T: Human T lymphotropic virus type-I and adult T-cell leukemia in Japan. *Int J Hematol* 2002, 76 Suppl 2:240-245.
8. Matsuoka M, Jeang KT: Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. *Nat Rev Cancer* 2007, 7(4):270-280.
9. Yoshida M, Seiki M, Yamaguchi K, Takatsuki K: Monoclonal integration of human T-cell leukemia provirus in all primary tumors of adult T-cell leukemia suggests causative role of human T-cell leukemia virus in the disease. *Proc Natl Acad Sci U S A* 1984, 81(8):2534-2537.
10. Takeda S, Maeda M, Morikawa S, Taniguchi Y, Yasunaga J, Nosaka K, Tanaka Y, Matsuoka M: Genetic and epigenetic inactivation of tax gene in adult T-cell leukemia cells. *Int J Cancer* 2004, 109(4):559-567.
11. Mesnard JM, Barbeau B, Devaux C: HBZ, a new important player in the mystery of adult T-cell leukemia. *Blood* 2006, 108(13):3979-3982.
12. Nagata Y, Kontani K, Enami T, Kataoka K, Ishii R, Totoki Y, Kataoka TR, Hirata M, Aoki K, Nakano K *et al*: Variegated RHOA mutations in adult T-cell leukemia/lymphoma. *Blood* 2016, 127(5):596-604.
13. Etoh K, Yamaguchi K, Tokudome S, Watanabe T, Okayama A, Stuver S, Mueller N, Takatsuki K, Matsuoka M: Rapid quantification of HTLV-I provirus load: detection of monoclonal proliferation of HTLV-I-infected cells among blood donors. *Int J Cancer* 1999, 81(6):859-864.
14. Ohshima K, Mukai Y, Shiraki H, Suzumiya J, Tashiro K, Kikuchi M: Clonal integration and expression of human T-cell lymphotropic virus type I in carriers detected by polymerase chain reaction and inverse PCR. *Am J Hematol* 1997, 54(4):306-312.
15. Furukawa Y, Fujisawa J, Osame M, Toita M, Sonoda S, Kubota R, Ijichi S, Yoshida M: Frequent clonal proliferation of human T-cell leukemia virus type 1 (HTLV-1)-infected T cells in HTLV-1-associated myelopathy (HAM-TSP). *Blood* 1992, 80(4):1012-1016.
16. Ikeda S, Momita S, Kinoshita K, Kamihira S, Moriuchi Y, Tsukasaki K, Ito M, Kanda T, Moriuchi R, Nakamura T *et al*: Clinical course of human T-lymphotropic virus type I carriers with molecularly detectable monoclonal

- proliferation of T lymphocytes: defining a low- and high-risk population. *Blood* 1993, 82(7):2017-2024.
17. Carvalho EM, Da Fonseca Porto A: Epidemiological and clinical interaction between HTLV-1 and *Strongyloides stercoralis*. *Parasite Immunol* 2004, 26(11-12):487-497.
 18. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S *et al*: Global identification of human transcribed sequences with genome tiling arrays. *Science* 2004, 306(5705):2242-2246.
 19. Esteller M: Non-coding RNAs in human disease. *Nat Rev Genet* 2011, 12(12):861-874.
 20. Farazi TA, Juranek SA, Tuschl T: The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* 2008, 135(7):1201-1214.
 21. Mattick JS, Makunin IV: Non-coding RNA. *Hum Mol Genet* 2006, 15 Spec No 1:R17-29.
 22. Li J, Wu B, Xu J, Liu C: Genome-wide identification and characterization of long intergenic non-coding RNAs in *Ganoderma lucidum*. *PLoS One* 2014, 9(6):e99442.
 23. Cheng AM, Byrom MW, Shelton J, Ford LP: Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res* 2005, 33(4):1290-1297.
 24. Vaz C, Ahmad HM, Sharma P, Gupta R, Kumar L, Kulshreshtha R, Bhattacharya A: Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood. *BMC Genomics* 2010, 11:288.
 25. Bandiera S, Hatem E, Lyonnet S, Henrion-Caude A: microRNAs in diseases: from candidate to modifier genes. *Clin Genet* 2010, 77(4):306-313.
 26. Bueno MJ, Malumbres M: MicroRNAs and the cell cycle. *Biochim Biophys Acta* 2011, 1812(5):592-601.
 27. Szulwach KE, Jin P, Alisch RS: Noncoding RNAs in mental retardation. *Clin Genet* 2009, 75(3):209-219.
 28. Zhang B, Pan X, Cobb GP, Anderson TA: microRNAs as oncogenes and tumor suppressors. *Dev Biol* 2007, 302(1):1-12.
 29. Shenouda SK, Alahari SK: MicroRNA function in cancer: oncogene or a tumor suppressor? *Cancer Metastasis Rev* 2009, 28(3-4):369-378.
 30. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA *et al*: MicroRNA expression profiles classify human cancers. *Nature* 2005, 435(7043):834-838.
 31. Huntzinger E, Izaurralde E: Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* 2011, 12(2):99-110.
 32. Ruggero K, Corradin A, Zanovello P, Amadori A, Bronte V, Ciminale V, D'Agostino DM: Role of microRNAs in HTLV-1 infection and transformation. *Mol Aspects Med* 2010, 31(5):367-382.
 33. Pichler K, Schneider G, Grassmann R: MicroRNA miR-146a and further oncogenesis-related cellular microRNAs are dysregulated in HTLV-1-transformed T lymphocytes. *Retrovirology* 2008, 5:100.
 34. Bellon M, Lepelletier Y, Hermine O, Nicot C: Deregulation of microRNA involved in hematopoiesis and the immune response in HTLV-I adult T-cell leukemia. *Blood* 2009, 113(20):4914-4917.

35. Yeung ML, Yasunaga J, Bennasser Y, Dusetti N, Harris D, Ahmad N, Matsuoka M, Jeang KT: Roles for microRNAs, miR-93 and miR-130b, and tumor protein 53-induced nuclear protein 1 tumor suppressor in cell growth dysregulation by human T-cell lymphotropic virus 1. *Cancer Res* 2008, 68(21):8976-8985.
36. Ruggero K, Guffanti A, Corradin A, Sharma VK, De Bellis G, Corti G, Grassi A, Zanovello P, Bronte V, Ciminale V *et al*: Small noncoding RNAs in cells transformed by human T-cell leukemia virus type 1: a role for a tRNA fragment as a primer for reverse transcriptase. *J Virol* 2014, 88(7):3612-3622.
37. Heneine W, Khabbaz RF, Lal RB, Kaplan JE: Sensitive and specific polymerase chain reaction assays for diagnosis of human T-cell lymphotropic virus type I (HTLV-I) and HTLV-II infections in HTLV-I/II-seropositive individuals. *J Clin Microbiol* 1992, 30(6):1605-1607.
38. Pessoa R, Watanabe JT, Nukui Y, Pereira J, Casseb J, de Oliveira AC, Segurado AC, Sanabani SS: Molecular characterization of human T-cell lymphotropic virus type 1 full and partial genomes by Illumina massively parallel sequencing technology. *PLoS One* 2014, 9(3):e93374.
39. Shadrach B, Warshawsky I: A comparison of multiplex and monoplex T-cell receptor gamma PCR. *Diagn Mol Pathol* 2004, 13(3):127-134.
40. Clissa PB, Pessoa R, Ferraz KF, de Souza DR, Sanabani SS: Data on global expression of non-coding RNAs in mice gastrocnemius muscle exposed to jararhagin, snake venom metalloproteinase. *Data Brief* 2016, 9:685-688.
41. Langenberger D, Bermudez-Santana CI, Stadler PF, Hoffmann S: Identification and classification of small RNAs in transcriptome sequence data. *Pac Symp Biocomput* 2010:80-87.
42. Schulze WX, Deng L, Mann M: Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol Syst Biol* 2005, 1:2005 0008.
43. Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, Dalamagas T, Hatzigeorgiou AG: DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res* 2015, 43(W1):W460-466.
44. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M *et al*: Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A* 2004, 101(9):2999-3004.
45. Yekta S, Shih IH, Bartel DP: MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 2004, 304(5670):594-596.
46. Wynter CV: The dialectics of cancer: A theory of the initiation and development of cancer through errors in RNAi. *Med Hypotheses* 2006, 66(3):612-635.
47. Luthra R, Singh RR, Luthra MG, Li YX, Hannah C, Romans AM, Barkoh BA, Chen SS, Ensor J, Maru DM *et al*: MicroRNA-196a targets annexin A1: a microRNA-mediated mechanism of annexin A1 downregulation in cancers. *Oncogene* 2008, 27(52):6667-6678.
48. Szafranska AE, Doleshal M, Edmunds HS, Gordon S, Luttgies J, Munding JB, Barth RJ, Jr., Gutmann EJ, Suriawinata AA, Marc Pipas J *et al*: Analysis of microRNAs in pancreatic fine-needle aspirates can classify benign and malignant tissues. *Clin Chem* 2008, 54(10):1716-1724.

49. Maru DM, Singh RR, Hannah C, Albarracin CT, Li YX, Abraham R, Romans AM, Yao H, Luthra MG, Anandasabapathy S *et al*: MicroRNA-196a is a potential marker of progression during Barrett's metaplasia-dysplasia-invasive adenocarcinoma sequence in esophagus. *Am J Pathol* 2009, 174(5):1940-1948.
50. Li Y, Zhang M, Chen H, Dong Z, Ganapathy V, Thangaraju M, Huang S: Ratio of miR-196s to HOXC8 messenger RNA correlates with breast cancer cell migration and metastasis. *Cancer Res* 2010, 70(20):7894-7904.
51. Chen C, Zhang Y, Zhang L, Weakley SM, Yao Q: MicroRNA-196: critical roles and clinical applications in development and cancer. *J Cell Mol Med* 2011, 15(1):14-23.
52. Hou W, Tian Q, Zheng J, Bonkovsky HL: MicroRNA-196 represses Bach1 protein and hepatitis C virus gene expression in human hepatoma cells expressing hepatitis C viral proteins. *Hepatology* 2010, 51(5):1494-1504.
53. Liu B, Xiang Y, Zhang HS: Circulating microRNA-196a as a candidate diagnostic biomarker for chronic hepatitis C. *Mol Med Rep* 2015, 12(1):105-110.
54. Coskun E, von der Heide EK, Schlee C, Kuhn A, Gokbuget N, Hoelzer D, Hofmann WK, Thiel E, Baldus CD: The role of microRNA-196a and microRNA-196b as ERG regulators in acute myeloid leukemia and acute T-lymphoblastic leukemia. *Leuk Res* 2011, 35(2):208-213.
55. Baldus CD, Burmeister T, Martus P, Schwartz S, Gokbuget N, Bloomfield CD, Hoelzer D, Thiel E, Hofmann WK: High expression of the ETS transcription factor ERG predicts adverse outcome in acute T-lymphoblastic leukemia in adults. *J Clin Oncol* 2006, 24(29):4714-4720.
56. Loughran SJ, Kruse EA, Hacking DF, de Graaf CA, Hyland CD, Willson TA, Henley KJ, Ellis S, Voss AK, Metcalf D *et al*: The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat Immunol* 2008, 9(7):810-819.
57. Mitchelson KR, Qin WY: Roles of the canonical myomiRs miR-1, -133 and -206 in cell development and disease. *World J Biol Chem* 2015, 6(3):162-208.
58. Jackson AC: Rabies. *Neurol Clin* 2008, 26(3):717-726, ix.
59. Samir M, Vaas LA, Pessler F: MicroRNAs in the Host Response to Viral Infections of Veterinary Importance. *Front Vet Sci* 2016, 3:86.
60. Castillo JA, Castrillon JC, Diosa-Toro M, Betancur JG, St Laurent G, 3rd, Smit JM, Urcuqui-Inchima S: Complex interaction between dengue virus replication and expression of miRNA-133a. *BMC Infect Dis* 2016, 16:29.
61. Miles GD, Seiler M, Rodriguez L, Rajagopal G, Bhanot G: Identifying microRNA/mRNA dysregulations in ovarian cancer. *BMC Res Notes* 2012, 5:164.
62. Tao J, Wu D, Xu B, Qian W, Li P, Lu Q, Yin C, Zhang W: microRNA-133 inhibits cell proliferation, migration and invasion in prostate cancer cells by targeting the epidermal growth factor receptor. *Oncol Rep* 2012, 27(6):1967-1975.
63. Yamasaki T, Yoshino H, Enokida H, Hidaka H, Chiyomaru T, Nohata N, Kinoshita T, Fuse M, Seki N, Nakagawa M: Novel molecular targets regulated by tumor suppressors microRNA-1 and microRNA-133a in bladder cancer. *Int J Oncol* 2012, 40(6):1821-1830.

64. Nohata N, Hanazawa T, Enokida H, Seki N: microRNA-1/133a and microRNA-206/133b clusters: dysregulation and functional roles in human cancers. *Oncotarget* 2012, 3(1):9-21.
65. Kress AK, Kalmer M, Rowan AG, Grassmann R, Fleckenstein B: The tumor marker Fascin is strongly induced by the Tax oncoprotein of HTLV-1 through NF-kappaB signals. *Blood* 2011, 117(13):3609-3612.
66. Wu ZS, Wang CQ, Xiang R, Liu X, Ye S, Yang XQ, Zhang GH, Xu XC, Zhu T, Wu Q: Loss of miR-133a expression associated with poor survival of breast cancer and restoration of miR-133a expression inhibited breast cancer cell growth and invasion. *BMC Cancer* 2012, 12:51.
67. Ladd AN, Nguyen NH, Malhotra K, Cooper TA: CELF6, a member of the CELF family of RNA-binding proteins, regulates muscle-specific splicing enhancer-dependent alternative splicing. *J Biol Chem* 2004, 279(17):17756-17764.
68. Hwang SJ, Yang Q, Meigs JB, Pearce EN, Fox CS: A genome-wide association for kidney function and endocrine-related traits in the NHLBI's Framingham Heart Study. *BMC Med Genet* 2007, 8 Suppl 1:S10.
69. Fox CS, Heard-Costa N, Cupples LA, Dupuis J, Vasan RS, Atwood LD: Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project. *BMC Med Genet* 2007, 8 Suppl 1:S18.
70. Mueller SB, Arora S, Mattocks N, Gurley SB, Kontos CD: Caskin2 is a novel regulator of endothelial cell quiescence. *Arterioscl Throm Vas* 2014, 34: A14.
71. Hakata Y, Umemoto T, Matsushita S, Shida H: Involvement of human CRM1 (exportin 1) in the export and multimerization of the Rex protein of human T-cell leukemia virus type 1. *J Virol* 1998, 72(8):6602-6607.
72. Hakata Y, Yamada M, Shida H: Rat CRM1 is responsible for the poor activity of human T-cell leukemia virus type 1 Rex protein in rat cells. *J Virol* 2001, 75(23):11515-11525.
73. Peng Y, Li M, Clarkson BD, Pehar M, Lao PJ, Hillmer AT, Barnhart TE, Christian BT, Mitchell HA, Bendlin BB *et al*: Deficient import of acetyl-CoA into the ER lumen causes neurodegeneration and propensity to infections, inflammation, and cancer. *J Neurosci* 2014, 34(20):6772-6789.
74. Goldstein JL, Anderson RG, Brown MS: Coated pits, coated vesicles, and receptor-mediated endocytosis. *Nature* 1979, 279(5715):679-685.
75. Schmid SL: Clathrin-coated vesicle formation and protein sorting: an integrated process. *Annu Rev Biochem* 1997, 66:511-548.
76. Molloy SS, Anderson ED, Jean F, Thomas G: Bi-cycling the furin pathway: from TGN localization to pathogen activation and embryogenesis. *Trends Cell Biol* 1999, 9(1):28-35.
77. Ohno H: Clathrin-associated adaptor protein complexes. *J Cell Sci* 2006, 119(Pt 18):3719-3721.
78. Batonick M, Favre M, Boge M, Spearman P, Honing S, Thali M: Interaction of HIV-1 Gag with the clathrin-associated adaptor AP-2. *Virology* 2005, 342(2):190-200.
79. Dong X, Li H, Derdowski A, Ding L, Burnett A, Chen X, Peters TR, Dermody TS, Woodruff E, Wang JJ *et al*: AP-3 directs the intracellular trafficking of HIV-1 Gag and plays a key role in particle assembly. *Cell* 2005, 120(5):663-674.

80. Camus G, Segura-Morales C, Molle D, Lopez-Verges S, Begon-Pescia C, Cazevielle C, Schu P, Bertrand E, Berlioz-Torrent C, Basyuk E: The clathrin adaptor complex AP-1 binds HIV-1 and MLV Gag and facilitates their budding. *Mol Biol Cell* 2007, 18(8):3193-3203.
81. Chikatsu N, Kojima H, Suzukawa K, Shinagawa A, Nagasawa T, Ozawa H, Yamashita Y, Mori N: ALK+, CD30-, CD20- large B-cell lymphoma containing anaplastic lymphoma kinase (ALK) fused to clathrin heavy chain gene (CLTC). *Mod Pathol* 2003, 16(8):828-832.
82. Bridge JA, Kanamori M, Ma Z, Pickering D, Hill DA, Lydiatt W, Lui MY, Colleoni GW, Antonescu CR, Ladanyi M *et al*: Fusion of the ALK gene to the clathrin heavy chain gene, CLTC, in inflammatory myofibroblastic tumor. *Am J Pathol* 2001, 159(2):411-415.
83. Lau ST, Zhou T, Liu JA, Fung EY, Che CM, Lang BH, Ngan ES: Dysregulation of clathrin promotes thyroid cell growth and contributes to multinodular goiter pathogenesis. *Biochim Biophys Acta* 2015, 1852(8):1676-1686.
84. Cai X, Srivastava S, Sun Y, Li Z, Wu H, Zuvela-Jelaska L, Li J, Salamon RS, Backer JM, Skolnik EY: Tripartite motif containing protein 27 negatively regulates CD4 T cells by ubiquitinating and inhibiting the class II PI3K-C2beta. *Proc Natl Acad Sci U S A* 2011, 108(50):20072-20077.
85. Yamagishi M, Nakano K, Miyake A, Yamochi T, Kagami Y, Tsutsumi A, Matsuda Y, Sato-Otsubo A, Muto S, Utsunomiya A *et al*: Polycomb-mediated loss of miR-31 activates NIK-dependent NF-kappaB pathway in adult T cell leukemia and other cancers. *Cancer Cell* 2012, 21(1):121-135.

Figure captions

- 1- Distribution of the total small RNAs on two representative samples. The clean reads were annotated as tRNAs, rRNAs, snoRNAs, snRNAs, scRNA miRNAs, intro, and others based on the Rfam database
- 2- Hierarchical cluster analysis of the 52 known small RNAs differentially expressed comparing ASM vs ASP samples. Rows of the heatmap represent the 52 small RNAs, columns correspond to samples. The cells are colored based on the deviance of the small RNAs expression in the sample from the average expression of the small RNAs, therefore red and blue cells represent respectively expression values higher or lower than mean expression across all samples (white) with color intensity proportional to the difference from the mean, in the regularized logarithmic scale. For clarity purposes, only a subimage of the heatmap has been shown in details.
- 3- Hierarchical cluster analysis of miRNAs differentially expressed comparing ASM vs ASP samples. Hierarchical clusters were generated using the Euclidean distance and Pearson correlations of miRNA expression patterns. The up- (red) and down-regulated (blue) miRNAs were scaled according to the depicted color code.

Supporting Information

Table S1: Distribution of the small RNA massively sequencing reads according to the chromosomes

Table S2: A full list of all known small RNAs identified in this study

Table S3: A full list of all novel small RNAs identified in this study

Table S4: List of small RNAs differentially expressed in asymptomatic HTLV-1 infected subjects with monoclonal (ASM) vs subjects with polyclonal T cell expansion (ASP)

Table S5: List of the potential protein targets for the differentially expressed miRNA in the ASM group

Table 1. Demographic and clinical characteristics of HTLV-1 infected asymptomatic carriers subjected to small RNA analysis

Sample	Sex	Age	Clonality features	Proviral load (copies/1000 PBMCs)	Total input reads	Total filtered reads (%)
131	Female	52	Polyclonal	17	3,944,199	165,061 (4,18%)
146	Female	59	Polyclonal	208	12,538,011	502,457 (4,01%)
151	Female	72	Polyclonal	1	11,568,287	389,210 (3,36%)
152	Female	34	Polyclonal	60	6,323,875	479,958 (7,59%)
167	Female	58	Polyclonal	32	5,828,992	469,333 (8,05%)
172	Female	31	Polyclonal	102	4,307,133	314,845 (7,31%)
182	Female	80	Polyclonal	22	7,195,318	350,231 (4,87%)
188	Female	49	Polyclonal	8	8,913,900	314,436 (3,53%)
54	Female	42	Monoclonal	ND	6,452,823	54,089 (0,84%)
75	Female	45	Monoclonal	ND	11,154,254	70,707 (0,63%)
124	Male	70	Monoclonal	327	8,113,493	142,191 (1,75%)
143	Female	65	Monoclonal	71	5,719,645	246,809 (4,32%)
154	Female	28	Monoclonal	163	11,018,111	425,570 (3,86%)
200	Female	74	Monoclonal	ND	7,918,316	409,349 (5,17%)
212	Male	61	Monoclonal	27	12,777,420	417,351 (3,27%)

Table 2. List of miRNAs differentially expressed in asymptomatic HTLV-1 subjects with monoclonal (ASM) vs subjects with polyclonal T cell expansion (ASP)

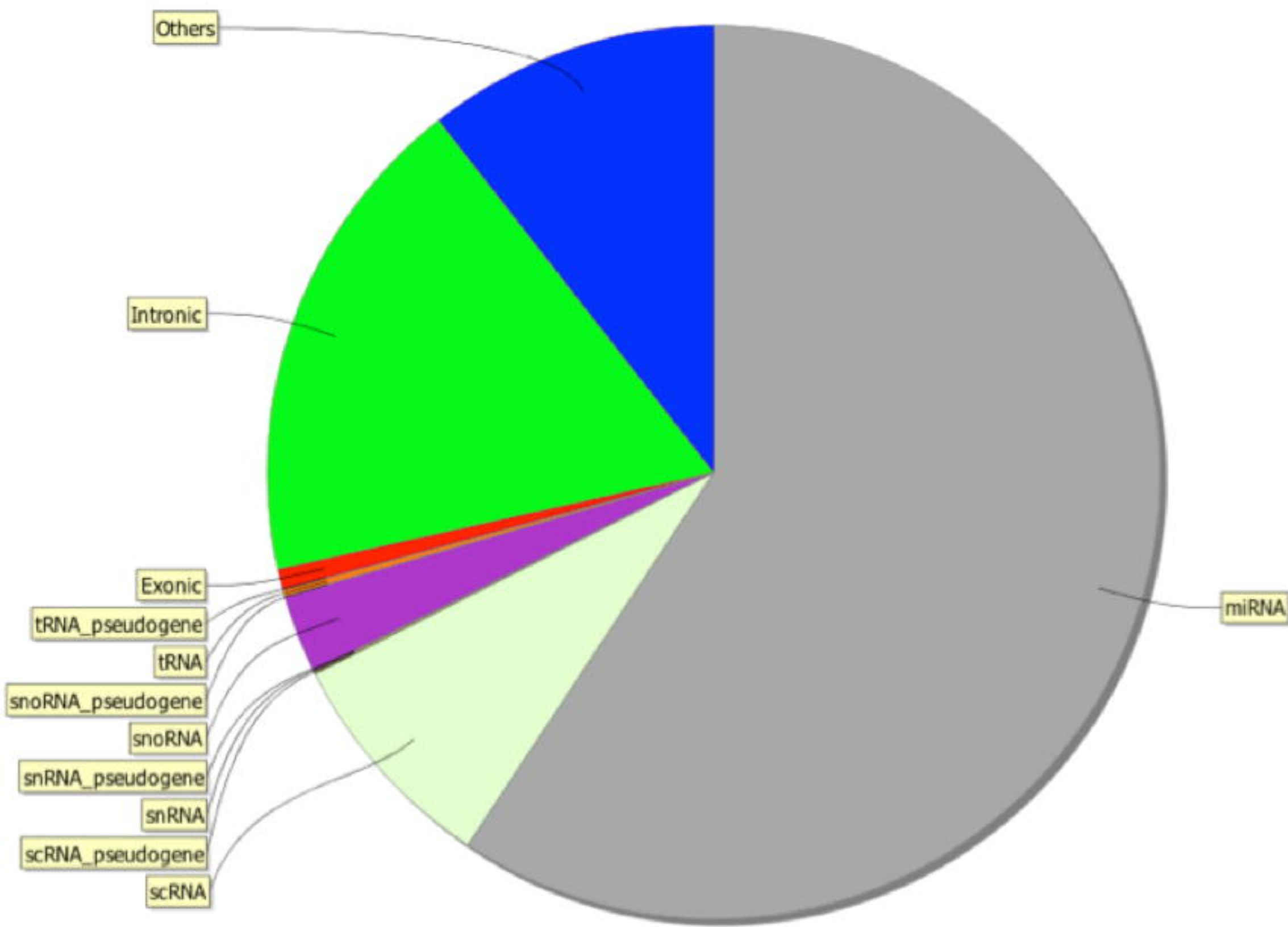
Gene ID	FC ([ASM] vs [ASP])	Log FC ([ASM] vs [ASP])	FC (abs) ([ASM] vs [ASP])	Regulation ([ASM] vs [ASP])	Accession	Entrez ID	DB Refs
MI0000450_1	-4.05	-2.02	4.05	down	hsa-miR-133a	406922	miRBase:MI0000450
MI0003196_1	-3.05	-1.61	3.05	down	hsa-miR-509-3p	574514	miRBase:MI0003196
MI0003639_2	-2.87	-1.52	2.87	down	hsa-miR-625-3p	693210	miRBase:MI0003639
MI0014136_2	-2.86	-1.52	2.86	down	hsa-miR-3120-3p	100422882	miRBase:MI0014136
MI0000088_2	-2.80	-1.48	2.80	down	hsa-miR-30a-5p	407029	miRBase:MI0000088
MI0000750_1	-2.74	-1.46	2.74	down	hsa-miR-26a-2-3p	407016	miRBase:MI0000750
MI0000091_1	-2.38	-1.25	2.38	down	hsa-miR-33a-5p	407039	miRBase:MI0000091
MI0000743_1	-2.14	-1.10	2.14	down	hsa-miR-34c-5p	407042	miRBase:MI0000743
MI0000784_2	-2.13	-1.09	2.13	down	hsa-miR-376a-3p	494325	miRBase:MI0000784
MI0003677_1	-2.04	-1.03	2.04	down	hsa-miR-655	724025	miRBase:MI0003677
MI0000811_2	-2.00	-1.00	2.00	down	hsa-miR-148b-3p	442892	miRBase:MI0000811
MI0005543_2	2.02	1.01	2.02	up	hsa-miR-708-5p	100126333	miRBase:MI0005543
MI0000076_2	2.12	1.09	2.12	up	hsa-miR-20a-3p	406982	miRBase:MI0000076
MI0015995_2	2.13	1.09	2.13	up	hsa-miR-3605-5p	100500853	miRBase:MI0015995
MI0016775_1	2.15	1.10	2.15	up	hsa-miR-4435	100616499	miRBase:MI0016775
MI0003814_1	2.17	1.12	2.17	up	hsa-miR-1271-5p	100302203	miRBase:MI0003814
MI0001735_1	2.18	1.12	2.18	up	hsa-miR-409-5p	574413	miRBase:MI0001735
MI0016777_1	2.18	1.12	2.18	up	hsa-miR-4435	100616341	miRBase:MI0016777
MI0003676_1	2.24	1.16	2.24	up	hsa-miR-654-5p	724024	miRBase:MI0003676
MI0003186_2	2.45	1.29	2.45	up	hsa-miR-502-3p	574504	miRBase:MI0003186
MI0000105_2	2.56	1.36	2.56	up	hsa-miR-29b-1-5p	407024	miRBase:MI0000105 RFAM:RF00074

MI0000279_1	4.27	2.09	4.27	up	hsa-miR-196a-5p	406973	miRBase:MI0000279
-------------	------	------	------	----	-----------------	--------	-------------------

Table 3. List of novel small RNAs differentially expressed in asymptomatic HTLV-1 infected subjects with monoclonal (ASM) vs subjects with polyclonal T cell expansion (ASP)

Gene ID	FC ([ASM] vs [ASP])	Log FC ([ASM] vs [ASP])	FC (abs) ([ASM] vs [ASP])	Regulation ([ASM] vs [ASP])	Gene Type
NEWGENE31	2.76	1.47	2.76	up	Unknown
NEWGENE465	2.42	1.27	2.42	up	snoRNA
NEWGENE24	2.38	1.25	2.38	up	Unknown
NEWGENE53	2.36	1.24	2.36	up	Unknown
NEWGENE75	2.34	1.22	2.34	up	Unknown
NEWGENE96	2.30	1.20	2.30	up	Unknown
NEWGENE38	2.30	1.20	2.30	up	Unknown
NEWGENE95	2.29	1.20	2.29	up	Unknown
NEWGENE467	2.29	1.19	2.29	up	miRNA
NEWGENE203	2.25	1.17	2.25	up	snoRNA
NEWGENE377	2.23	1.16	2.23	up	snoRNA
NEWGENE466	2.22	1.15	2.22	up	tRNA
NEWGENE332	2.12	1.09	2.12	up	snoRNA
NEWGENE48	2.02	1.02	2.02	up	Unknown
NEWGENE186	-2.03	-0.10	2.03	down	snoRNA

154_ASM



131_ASP

