Research paper

Analyses of HTLV-1 sequences suggest interaction between ORF-I mutations and HAM/TSP outcome

Fernanda Khouri Barreto [a,1], Ricardo Khouri [b,1], Filipe Ferreira de Almeida Rego [a], Luciane Amorim Santos [a], Maria Fernanda de Castro-Amarante [c], Izabela Bialuk [d], Cynthia A. Pise-Masison [c], Bernardo Galvão-Castro [a,e], Antoine Gessain [f], Steven Jacobson [g], Genoveffa Franchini [c], Luiz Carlos Alcantara Jr [a,*]

[a] Centro de Pesquisa Gonçalo Moniz-Fiocruz, Laboratório de Hematologia Genética e Biologia Computacional, Salvador, Bahia, Brazil
[b] KU Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Laboratory of Clinical and Epidemiological Virology, Leuven, Belgium
[c] Animal Models and Retroviral Vaccines Section, National Cancer Institute, Bethesda, USA
[d] Department of General and Experimental Pathology, Medical University in Białystok, Poland
[e] Escola Bahiana de Medicina e Saúde Publica, Salvador, Bahia, Brazil
[f] Unité d'Epidémiologie et Physiopathologie des Virus Oncogènes, Département de Virologie, Batiment Lwoff, Institut Pasteur, Paris, France
[g] Viral Immunology Section, Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, Bethesda, USA

ARTICLE INFO

ABSTRACT

The region known as pX in the 3′ end of the human T-cell lymphotropic virus type 1 (HTLV-1) genome contains four overlapping open reading frames (ORF) that encode regulatory proteins. HTLV-1 ORF-I produces the protein p12 and its cleavage product p8. The functions of these proteins have been linked to immune evasion and viral infectivity and persistence. It is known that the HTLV-1 infection does not necessarily imply the development of pathological processes and here we evaluated whether natural mutations in HTLV-1 ORF-I can influence the proviral load and clinical manifestation of HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP). For that, we performed molecular characterization, datamining and phylogenetic analysis with HTLV-1 ORF-I sequences from 156 patients with negative or positive diagnosis for HAM/TSP. Our analyses demonstrated that some mutations may be associated with the outcome of HAM/TSP (C39R, L40F, P45L, S69G and R88K) or with proviral load (P34L and F61L). We further examined the presence of mutations in motifs of HBZ and observed that P45L mutation is located within the HBZ nuclear localization signal and was found more frequently between patients with HAM/TSP and high proviral load. These results indicate that some natural mutations are located in functional domains of ORF-I and suggests a potential association between these mutations and the proviral loads and development of HAM/TSP. Therefore it is necessary to conduct functional studies aimed at evaluating the impact of these mutations on the virus persistence and immune evasion.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The human T-cell lymphotropic virus type 1 (HTLV-1) was the first human retrovirus identified that is associated with disease, including adult T cell leukemia/lymphoma (ATLL), tropical spastic paraparesis/ HTLV-associated myelopathy (HAM/TSP), and infective dermatitis, polymyositis (Gessain et al., 1985; Goncalves et al., 2003; Poiesz et al., 1980; Yoshida et al., 1982). A recent research based on individuals originated from known HTLV-1 endemic areas estimates 5–10 million infected individuals (Gessain and Cassar, 2012). Indeed, HTLV-1 infection does not necessarily imply the development of pathological processes and the majority of infected individuals remain asymptomatic, with only a low percentage of individuals developing HAM/TSP (0.3– 2%) (Kaplan et al., 1990; Maloney et al., 1998). Studies indicate that high proviral load is typically seen in individuals with a positive diagnosis of HAM/TSP, when compared with infected individuals without the diagnosis for this disease (Nagai et al., 1998).

A region near the 3′ end known as pX encodes important regulatory and accessory proteins shown to effect viral replication mechanisms (Edwards et al., 2011; Lairmore et al., 2012, 2011). Transcription of this region results in alternative forms of messenger RNA (mRNA), which contains four partially overlapping open reading frames (ORFs) and an antisense mRNA that generates the basic leucine zipper (HBZ) protein (Berneman et al., 1992; Ciminale et al., 1992; Gaudray et al., 2002; Koralnik et al., 1992). Interestingly the promoter of the four ORFs is 5′ LTR, but the HBZ promoter is in 3′ LTR (Berneman et al., 1992). Singly

spliced mRNA from ORF-I produces the protein p12 that can be further cleaved into the p8 protein. Differential splicing of mRNA from ORF-II encodes the proteins, p30 and p13. The ORF-III and ORF-IV produce the proteins Rex and Tax, respectively (Ciminale et al., 1992; Johnson et al., 2001; Kiyokawa et al., 1985, 1984; Koralnik et al., 1992).

The expression of ORF-I in infected cells has been shown to regulate the biologically activity of HTLV-1, influencing the virus infectivity and persistence (Pise-Masison et al., 2014). The p12 protein resides in the endoplasmic reticulum and its expression increases the intracellular calcium concentration from the ER through inositol trisphosphate receptors, consequently promoting the dephosphorylation of activate nuclear factor of activated T-cells (NFAT) (Albrecht et al., 2002; Ding et al., 2002; Fukumoto et al., 2009). In addition, p12 can bind newly synthesized major histocompatibility complex class I (MHC I) molecules, reducing its expression on the cell surface. Although downregulation of MHC class I triggers NK cell activation, HTLV-1 infected cells are resistant to NK cell killing through p12's reduction of cell surface ICAM-1 and ICAM-2 (Banerjee et al., 2007).

The p12 protein undergoes double proteolytic cleavage. The first cleavage occurs between amino acid positions 9 and 10 and removes the ER retention/retrieval signal of p12. The second cleavage occurs between amino acids 29 and 30 and generates the p8 protein (Edwards et al., 2011). The p8 protein resides at the cell surface and its expression induces cellular conduits and enhances virus transmission (Van Prooyen et al., 2010).

Trovato et al. demonstrated two natural allelic variants of p12: a lysine-to-arginine change at position 88, which was found more frequently in patients with HAM/TSP and could be important for HTLV-1 outcome (Trovato et al., 1999). More recent studies on ORF-I suggest the co-dependence of p12 and p8 proteins to virus persistence and demonstrated that natural mutations within ORF-I sequences can affect the relative amounts of these proteins (Pise-Masison et al., 2014). Here, we demonstrate that natural mutations in HTLV-1 ORF-I might influence the proviral load and clinical manifestation of HAM/TSP.

2. Materials and methods

To perform the molecular analysis of ORF-I HTLV-1 we analysed 1530 HTLV-1 ORF-I sequences of different isolated cell clones from a total of 156 HTLV-1-infected subjects. These samples were obtained from the Centre Hospitalier Universitaire de Fort-de-France in Martinique, the Institut Pasteur de Cayenne in French Guyana, the Bahia School of Medicine and Public Health and the National Institutes of Health Clinical Center. Blood samples from these individuals were obtained and the DNA extracted from PBMCs was used to determine the proviral load using Real-time PCR of Tax. The same DNA was used for amplification of ORF-I through PCR reactions and these PCR products was purified and cloned into specific vector. Five to twenty clones per patient were isolated and sequenced. These methods were described in detail by Pise-Masison et al.(Pise-Masison et al., 2014).

The final dataset was composed of 879 HTLV-1 ORF-I sequences from 86 patients with not HAM/TSP, referred to as Health Carrier (HC), and 651 HTLV-1 ORF-I sequences from 70 patients with HAM/TSP. These sequences were originated from different geographic regions (Caribbean, France, North America, Africa, and Brazil) and classified according to a negative or positive diagnostic for HAM/TSP. All samples were anonymized and the clinical classification was carried out by medical experts according to World Health Organization (WHO). The research was conformed to the guidelines of the ethics review board of the National Cancer Institute and informed consent was written and obtained from each subject.

2.1. Molecular characterization

The genetic distances among the sequences from patient with negative or positive diagnosis for HAM/TSP were measured within and between the different datasets using the MEGA 5.05 program and the analyses about the proviral load was conducted using the STATA program. The statistical analyses of the proviral load were performed using Mann-Whitney test and a P value lower than 0.05 was considered statistically significant.

Since we get differing numbers of clones from each patient, the characterization of mutations was determined by a qualitative analysis considering the presence or absence of variants among the sequences of each individual. If one of the patient sequences has the variant we consider that the patient has the mutation (Geneious 5.6.5 program). We selected the most frequent natural mutations within ORF-I sequences (F3L, S23P, D26N, G29S, P34L, C39R, L40F, P45L, F61L, S63P, L66P, S69G, R83C, R88K and P91S) to evaluate their possible association with the development of HAM/TSP using the Excel program. The statistical analyses were performed using Fisher's exact test and a P value lower than 0.05 was considered statistically significant.

2.2. Datamining

We used a two-step approach, using an unsupervised method for attribute selection (naïve Bayesian networks analysis), followed by a supervised method for disease progression classification (Decision tree J48). A Bayesian network (BN) is a probabilistic model that describes statistical conditional dependencies between multiple variables. In this study, we learn Bayesian networks from observation of the attributes. Dependencies are visualized in a directed acyclic graph and form the qualitative component of the BN. In this graph, each node corresponds to an attribute, and a direct arc between nodes represents a direct influence. Mathematically, a Bayesian network provides a refactoring of the Joint Probability Distribution (JPD) of the data, using Bayes' rules. As a BN simplifies the JPD, it provides an effective model that summarizes statistical properties of the data. In this way, the best Bayesian network is searched that explains a maximum of the observed associations in the data using a minimum number of direct influences. Bayesian network learning was performed using the B-course software adapted by Deforche et al. (Deforche et al., 2006). In this non-linear model, the conditional dependency was assessed with a nonparametric bootstrap (100× replicates) (Friedman et al., 2013). The Decision tree J48 strategy is based on divide-and-conquer approach, sometimes called top-down induction of decision trees. It was developed and refined by J. Ross Quinlan at the University of Sydney in Australia. Divide-and-conquer algorithms operate by adding tests to the tree that is under construction, always striving to maximize the separation between the classes. Each of these involves finding an attribute to split on. The state of the art for error evaluation, independently from the loss functions, is to execute the 10-fold cross validation when a test set is unavailable. To execute and validate J48 decision tree approach we used the data mining suite Weka (E. WiaF, 2011).

2.3. Phylogenetic analysis

The phylogenetic analysis was performed using consensus sequences of patients from Brazil. Clustal-X was used for the alignment and the TN93 model of nucleotide substitution was selected using jModelTest. The maximum likelihood tree was inferred using PhyML online tool and bootstrap analysis (1000 replicates) was used to calculate the statistical support of the tree branches. Tree visualization and editing was done using FigTree v.1.2.2.

3. Results

3.1. Descriptive characteristics of the study samples

In this study, we analysed HTLV-1 ORF-I sequences from a total of 156 HTLV-1-infected subjects. Our final dataset was composed of 879

HTLV-1 ORF-I sequences from independently isolated clones of 86 non-HAM/TSP individuals (HC) and 651 HTLV-1 ORF-I sequences from independently isolated clones of 70 patients diagnosed with HAM/TSP (Pise-Masison et al., 2014). The median proviral load of patients with HAM/TSP was 22.9 copies/100 cells (IQR 11.9–47.6) which was significantly higher than HC, whose median proviral load was 8.7 copies/100 cells (IQR 0,5–17,8) (p < 0,0001).

Next, we measured the genetic distance between sequences using the MEGA 5.05 program. Interestingly, we found that the diversity between sequences from patients with HAM/TSP (0.014) was lower than that found between sequences from heath carrier (0.017).

3.2. Association of mutations in ORF-I sequences from HTLV-1-infected patients with proviral load and clinical definition

We analysed the presence or absence of fifteen amino acid changes (F3L, S23P, D26N, G29S, P34L, C39R, L40F, P45L, F61L, S63P, L66P, S69G, R83C, R88K and P91S) that influence the expression profile of the HTLV-1 ORF-I protein product. As previously described (Pise-Masison et al., 2014), these mutations resulted in three patterns of ORF-I protein expression: higher levels of p8 expression, higher levels of p12 expression or an equivalent level of p8 and p12 expression. We then compared the frequency of each mutation in non-HAM/TSP
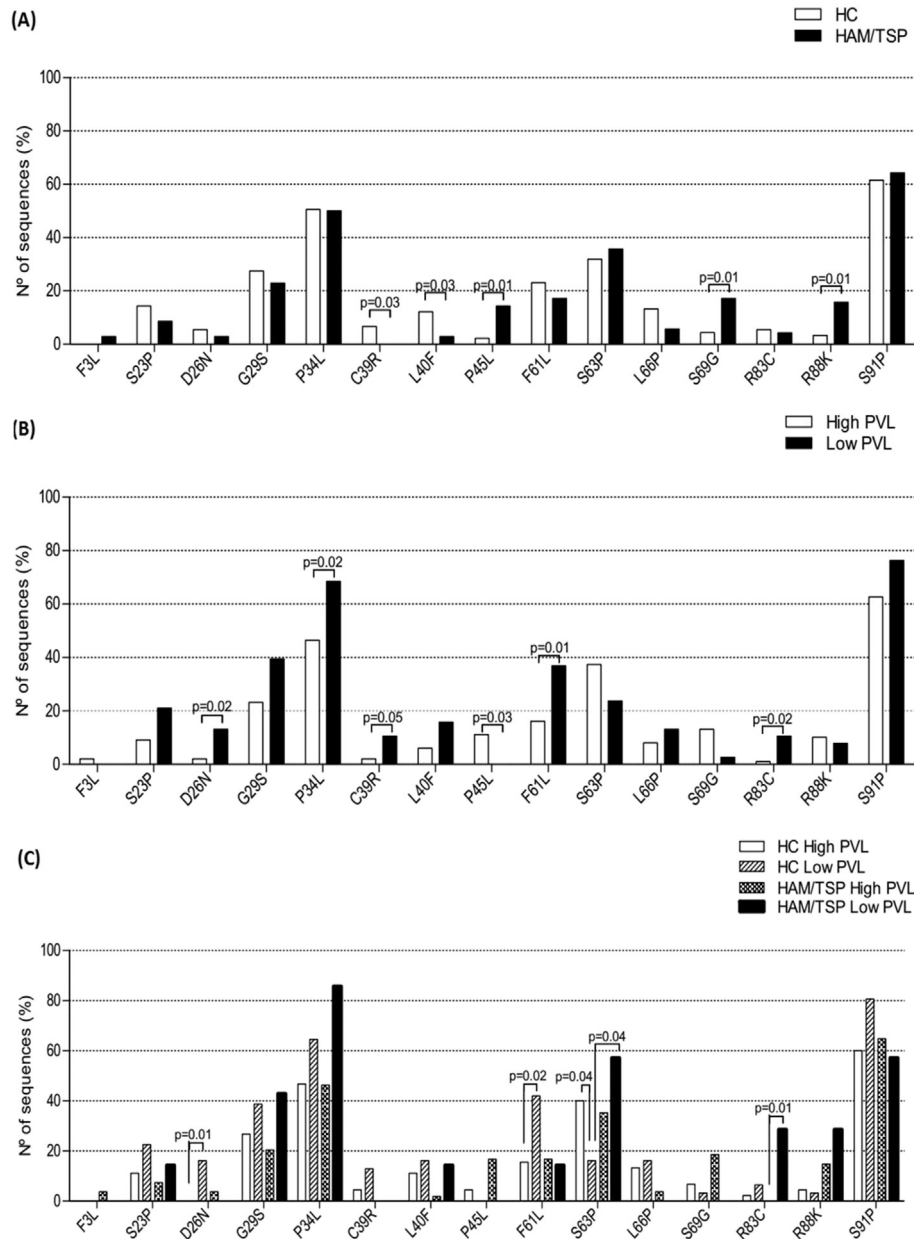

Fig. 1. Frequency of natural mutations within ORF-I sequences. The characterization of mutations was determined by a qualitative analysis considering the presence or absence of variants among the clone sequences of each individual. (A) The graph represents a comparison between patients with positive or negative diagnosis of HAM/TSP for a given mutation. (B) Comparison between patients with high and low proviral loads (PVL). High PVL are indicated by white bars and the patients with low PVL are indicated by black bars. (C) Mutation frequency was graphed for HAM/TSP patients with high PVL (black bar), HAM/TSP with low PVL (dotted bar), no HAM/TSP (HC) with high PVL (white bar) and HC with low PVL (striped bar). The phenotype of ORF-I protein expression is: Mostly expression of p12 (F3L, S23P, G29S, P34L, C39R,L40F, P45L, F61L, S63P, L66P); Mostly expression of p8 (D26N); Same expression of p12 and p8 (S69G, R83C, R88K, S91P). P values were calculated by the Fisher's exact test.

patients (HC) to HAM/TSP patients. In addition, a comparison between amino acid sequence and proviral load was performed (Fig. 1).

As shown in Fig. 1A, mutations P45L, S69G and R88K were found more frequently in patients with a positive diagnosis for HAM/TSP, whereas mutations C39R and L40F were found more frequently in HC. Of note, Trovato et al. showed that a lysine residue at position 88 destabilized the p12 protein by proteasomal degradation and that this mutation could be associated with HAM/TSP (Trovato et al., 1999). While this mutation was found more frequently in HAM/TSP patients than HC, we did not find a correlation between R88K and proviral load (Fig. 1B).

In contrast, mutations D26N, P34L, C39R, F61L and R83C were associated with low proviral load and P45L with high proviral load (Fig. 1B). A combined classification of proviral load and clinical diagnosis reveals mutations D26N and F61L are associated with low proviral load and HC, whereas mutations S63P and R83C are associated with high proviral load and HAM/TSP (Fig. 1C). The mutations P34L and S91P appear most frequently (>than 40% of the sequences analysed) but are not associated with the development of HAM/TSP. However, as seen in Fig. 1B, there was a higher frequency of P34L sequences in individuals with low proviral load.

Similarly, mutations C39R and F61L were more frequently found in patients with low proviral load (Fig. 1B). These three mutations (P34L, C39R and F61L) that result in predominant expression of p12 are associated with low proviral load. Whereas, the mutation D26N, the only mutation that resulted in predominant p8 expression, was not associated with the development of HAM/TSP but was associated with low proviral load and HC (Fig. 1B and C). These results are consistent with that of Pise-Masison et al. (Pise-Masison et al., 2014), where predominant expression of either p12 or p8 was associated with low proviral load.

Low proviral load in some individuals might also be influenced by alterations in remaining open reading frames in which other accessory proteins are encoded, such as viral protein HTLV-1 bZIP factor (HBZ). As described in previous studies, HBZ is encoded by the complementary strand of the HTLV-1 genome (7292 to 6666) and seems to play a role in cellular transformation, survival and proliferation (GenBank: U19949) (Matsuoka, 2010; Matsuoka and Green, 2009). Here, we further examined the presence of these mutations in motifs of HBZ and only four mutations in ORF-I resulted in changes in the HBZ ORF: P34L, P45L, L66P and R88K. Among them, only the P45L mutation is located within the motif previously described (HBZ

nuclear localization signal) and was found more frequently between patients with HAM/TSP and high proviral load.

3.3. Data mining approach reveals that mutation interactions are necessary to define HAM/TSP pathogenesis

To reveal possible interactions among specific mutations in ORF-I, proviral load, and disease definition, we employed a two-step approach, using an unsupervised method for attribute selection (naive Bayesian networks analysis, BN), followed by a supervised method for disease progression classification (Decision tree J48). First, Bayesian network analysis with all sequences showed a central association among disease progression, proviral load and patient origin, that interacted directly with mutations in ORF-I region: D26N, G29S, P34L, L40F, P45L, S63P, L66P, S69G, R83C. This first BN analysis revealed a possible biased geographical distribution of mutation prevalence given that in our dataset we do not have sequences from patients HC and HAM/TSP from all the geographic regions (data not shown). Due to this, we decided to use only samples from Brazil (27 HAM/TSP patients [median of 9 samples/patient, IQR = 7–17] and 48 HC [median of 10.5 samples/patient, IQR = 8–19]). No statistical significance was observed between the medians of the number of samples per HAM/TSP patients and HC (Fig. 2A). We then applied Decision tree J48 to predict HAM/TSP samples, confirming univariate analysis and revealing hidden conditional dependencies among mutations and proviral load (Fig. 2B). We observed several interactions associated with a true positive rate of 0.63 for HAM/TSP, as among D26N mutation and high proviral load or interactions between high proviral load, 26D, 29G and 34P. The predicted results above were plotted in a matrix 2 × 2 and Sensitivity, Specificity, Positive Predictive Values and Negative Predictive Values were described in Table 1.

Since sequences "states" are not statistically independent observations and these findings between mutations and disease manifestation could be associated with shared ancestry inherent, we performed a phylogenetic analysis using sequences of patients from Brazil (n = 75 consensus sequences) but the phylogenetic tree presented poor phylogenetic signal with no bootstrap support. This result might be explained due to the HTLV-1 ORF-I being a short region (300 bp) and highly conserved and it is known that the indicated region for subtyping HTLV-1 is LTR (Alcantara et al., 2009).
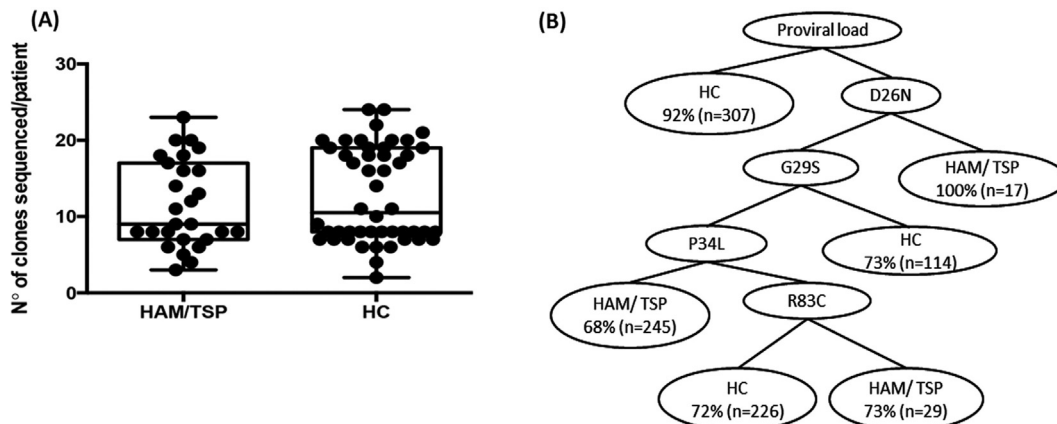

Fig. 2. Analysis of interactions among mutations in ORF-I, proviral load, and development of HAM/TSP. (A) Comparison of number of samples sequenced between HAM/TSP patients and HC from Brazil. This dataset is composed of 27 HAM/TSP patients [median of 9 samples/patient, IQR = 7–17] and 48 HC [median of 10.5 samples/patient, IQR = 8–19] (Mann Whitney test, p = 0.43). (B) Discovery of hidden conditional dependencies between proviral load and ORF-1 mutations associated with HAM/TSP. To reduce noise in the statistical analysis, proviral load measurements were discretized in 25% quartile and the results were represented as binary decision tree graphs. Nodes outline the chosen attributes and the number (n =) of instances and the accuracy defined per decision branch (%). This dataset is composed of 938 samples (Overall Accuracy = 78.25%, Detailed accuracy per HAM/TSP: True positive rate = 0.60, f-measure = 0.65 and AUC = 0.8).

Table 1

Results of conditional dependencies between proviral load, ORF-1 mutations and clinical status. The ORF-I sequences were originated from different isolated cell clones of HTLV-1-infected patients from Brazil. PPV means positive predictive value and NPV means negative predictive value.

Conditional dependency: Proviral load (PVL) and mutation (AA)	Prediction	Sequences		Parameter (%)			
		HC (n = 616)	HAM/TSP (n = 322)	Sensitivity	Specificity	PPV	NPV
Low PVL	HC	283	24	86	63	82	70
High PVL + 26N	HAM/TSP	0	17				
High PVL + 26D + 29S	HC	83	31				
High PVL + 26D + 29G + 34P	HAM/TSP	79	166				
High PVL + 26D + 29G + 34L + 83C	HAM/TSP	8	21				
High PVL + 26D + 29G + 34L + 83R	HC	163	63				

4. Discussion

HTLV-1 infection does not necessarily imply the development of pathological processes and it is not known what determines the manifestation of the disease in an infected individual. To date, proviral DNA levels in the blood are the best predictor of risk for the development of HAM/TSP (Matsuzaki et al., 2001; Yamano et al., 2002). Here, we described HTLV-1 ORF-I mutations that are associated with disease phenotype and may influence the outcome of infection and the development of HAM/TSP.

Computational analysis of the amino acid sequence of p12 predicts the existence of several functional domains, including a noncanonical endoplasmic reticulum (ER) retention signal, two leucine zipper (LZ) motifs, two transmembrane domains, a calcineurin-binding motif and four proline-rich Src homology 3 (SH3)-binding domains (Ding et al., 2002; Fukumoto et al., 2009). Some mutations analysed here are located in functional domains of ORF-I and may be associated with clonal expansion of HTLV-1-infected cells. The D26N and F61L mutations are located in the first and second transmembrane domains, respectively, and are found more frequently in patients with low proviral load, as shown in Fig. 1B. Our data support studies that suggest co-expression of p12 and p8 are important for efficient viral persistence (Pise-Masison et al., 2014).

Only R83C mutation is located in a calcineurin-binding motif and is associated with low proviral load. Calcineurin plays a crucial role in T-cell activation, in part, by dephosphorylating the nuclear factors of activated T cells (NFATs). NFAT is essential for activating cytokine gene expression and, thus, the immune response (Kim et al., 2003). p12 is able to mediate an increase in cytosolic calcium in T-cells by increasing calcium release from the ER (Kim et al., 2003). By depleting ER calcium stores and increasing cytosolic calcium, p12 can modulate cellular processes including T-cell proliferation, viral replication, and viral spread. Early studies on ORF-I showed that it activates NFAT (Albrecht and Lairmore, 2002; Kim et al., 2003). Therefore, mutations in the calcineurin binding region of ORF-I could affect NFAT activity and impact the immune response and viral replication.

Low proviral load might also be influenced by alterations in other accessory proteins, such as HBZ protein, that seems to play a role in cellular transformation, survival and proliferation (Albrecht and Lairmore, 2002; Gaudray et al., 2002; Matsuoka and Green, 2009). Here, we demonstrated only the P45L mutation is located within the HBZ nuclear localization signal and was found more frequently between patients with HAM/TSP and high proviral load. This finding is consistent with results showing that HBZ down-regulates Tax-induced HTLV-1 transcription and thus viral replication (Matsuoka, 2010; Matsuoka and Green, 2009).

Many different machine-learning tools have been used to predict therapy failure and conditional dependency among drug resistance mutations (Pineda-Pena et al., 2014; Prosperi and De Luca, 2012). Here, our data mining approach confirmed the association among HAM/TSP, proviral load and specific single mutations found with univariate analysis. Moreover, it revealed unexpected conditional dependencies among proviral load and mutations in ORF-1 associated with HAM/TSP. Despite the high prevalence of P34L mutation and its association with proviral load, this mutation alone could not determine disease. Only the data mining approach revealed the coordinate dependency between proviral load, P34L and P45L to be associated with HAM/TSP.

5. Conclusions

The results presented in this study suggest that some ORF-I natural mutations may be associated to HAM/TSP development and to the proviral load of HTLV-1-infected individuals. Is important to note that HTLV-1 have four open reading frames (ORFs) in the pX region and all regulatory proteins play a key role in viral pathogenesis. Therefore to investigate better the contribution of mutations found in HTLV-1 ORF-I to the development of HAM/TSP, it is necessary to conduct functional studies aimed at evaluating the impact of these mutations on the virus persistence and immune evasion.

Availability of supporting data

All sequences are available from the NCBI database (accession numbers KM436104-KM437632).

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

References

Albrecht, B., Lairmore, M.D., 2002. Critical role of human T-lymphotropic virus type 1 accessory proteins in viral replication and pathogenesis. Microbiol. Mol. Biol. Rev. MMBR 66, 396–406 (table of contents).

Albrecht, B., D'Souza, C.D., Ding, W., Tridandapani, S., Coggeshall, K.M., Lairmore, M.D., 2002. Activation of nuclear factor of activated T cells by human T-lymphotropic virus type 1 accessory protein p12(I). J. Virol. 76, 3493–3501.

Alcantara, L.C.J., Cassol, S., Libin, P., Deforche, K., Pybus, O.G., Van Ranst, M., Galvao-Castro, B., Vandamme, A.-M., de Oliveira, T., 2009. A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. Nucleic Acids Res. 37, W634–W642. http://dx.doi.org/10.1093/nar/gkp455.

Banerjee, P., Feuer, G., Barker, E., 2007. Human T-cell leukemia virus type 1 (HTLV-1) p12I down-modulates ICAM-1 and − 2 and reduces adherence of natural killer cells, thereby protecting HTLV-1-infected primary CD4 + T cells from autologous natural killer cell-mediated cytotoxicity despite the reduction of major histocompatibility complex class I molecules on infected cells. J. Virol. 81, 9707–9717. http://dx.doi.org/10.1128/JVI.00887-07.

Berneman, Z.N., Gartenhaus, R.B., Reitz, M.S., Blattner, W.A., Manns, A., Hanchard, B., Ikehara, O., Gallo, R.C., Klotman, M.E., 1992. Expression of alternatively spliced human T-lymphotropic virus type I pX mRNA in infected cell lines and in primary uncultured cells from patients with adult T-cell leukemia/lymphoma and healthy carriers. Proc. Natl. Acad. Sci. U. S. A. 89, 3005–3009.

Ciminale, V., Pavlakis, G.N., Derse, D., Cunningham, C.P., Felber, B.K., 1992. Complex splicing in the human T-cell leukemia virus (HTLV) family of retroviruses: novel mRNAs and proteins produced by HTLV type I. J. Virol. 66, 1737–1745.

Deforche, K., Silander, T., Camacho, R., Grossman, Z., Soares, M.A., Van Laethem, K., Kantor, R., Moreau, Y., Vandamme, A.-M., 2006. Analysis of HIV-1 pol sequences using

Bayesian Networks: implications for drug resistance. Bioinforma. Oxf. Engl. 22, 2975–2979. http://dx.doi.org/10.1093/bioinformatics/btl508.

Ding, W., Albrecht, B., Kelley, R.E., Muthusamy, N., Kim, S.-J., Altschuld, R.A., Lairmore, M.D., 2002. Human T-cell lymphotropic virus type 1 p12(I) expression increases cytoplasmic calcium to enhance the activation of nuclear factor of activated T cells. J. Virol. 76, 10374–10382.

Edwards, D., Fenizia, C., Gold, H., de Castro-Amarante, M.F., Buchmann, C., Pise-Masison, C.A., Franchini, G., 2011. Orf-I and Orf-II-encoded proteins in HTLV-1 infection and persistence. Virologie 3, 861–885. http://dx.doi.org/10.3390/v3060861.

Friedman, N., Goldszmidt, M., Wyner, A., 2013. Data Analysis with Bayesian Networks: A Bootstrap Approach (ArXiv13016695 Cs Stat).

Fukumoto, R., Andresen, V., Bialuk, I., Cecchinato, V., Walser, J.-C., Valeri, V.W., Nauroth, J.M., Gessain, A., Nicot, C., Franchini, G., 2009. In vivo genetic mutations define predominant functions of the human T-cell leukemia/lymphoma virus p12I protein. Blood 113, 3726–3734. http://dx.doi.org/10.1182/blood-2008-04-146928.

Gaudray, G., Gachon, F., Basbous, J., Biard-Piechaczyk, M., Devaux, C., Mesnard, J.-M., 2002. The complementary strand of the human T-cell leukemia virus type 1 RNA genome encodes a bZIP transcription factor that down-regulates viral transcription. J. Virol. 76, 12813–12822.

Gessain, A., Cassar, O., 2012. Epidemiological aspects and world distribution of HTLV-1 infection. Front. Microbiol. 3, 388. http://dx.doi.org/10.3389/fmicb.2012.00388.

Gessain, A., Barin, F., Vernant, J.C., Gout, O., Maurs, L., Calender, A., de Thé, G., 1985. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet Lond. Engl. 2, 407–410.

Goncalves, D.U., Guedes, A.C.M., Proietti, A.B., de F.C., Martins, M.L., Proietti, F.A., Lambertucci, J.R., 2003. Dermatologic lesions in asymptomatic blood donors seropositive for human T cell lymphotropic virus type-1. Am. J. Trop. Med. Hyg. 68, 562–565.

Johnson, J.M., Harrod, R., Franchini, G., 2001. Molecular biology and pathogenesis of the human T-cell leukaemia/lymphotropic virus Type-1 (HTLV-1). Int. J. Exp. Pathol. 82, 135–147.

Kaplan, J.E., Osame, M., Kubota, H., Igata, A., Nishitani, H., Maeda, Y., Khabbaz, R.F., Janssen, R.S., 1990. The risk of development of HTLV-I-associated myelopathy/tropical spastic paraparesis among persons infected with HTLV-I. J. Acquir. Immune Defic. Syndr. 3, 1096–1101.

Kim, S., Ding, W., Albrecht, B., Green, P.L., Lairmore, M.D., 2003. A conserved calcineurin-binding motif in human T lymphotropic virus type 1 p12I functions to modulate nuclear factor of activated T cell activation. J. Biol. Chem. 278, 15550–15557. http://dx.doi.org/10.1074/jbc.M210210200.

Kiyokawa, T., Seiki, M., Imagawa, K., Shimizu, F., Yoshida, M., 1984. Identification of a protein (p40x) encoded by a unique sequence pX of human. Gan 75, 747–751.

Kiyokawa, T., Seiki, M., Iwashita, S., Imagawa, K., Shimizu, F., Yoshida, M., 1985. p27x-III and p21x-III, proteins encoded by the pX sequence of human T-cell leukemia virus type I. Proc. Natl. Acad. Sci. U. S. A. 82, 8359–8363.

Koralnik, I.J., Gessain, A., Klotman, M.E., Lo Monico, A., Berneman, Z.N., Franchini, G., 1992. Protein isoforms encoded by the pX region of human T-cell leukemia/lymphotropic virus type I. Proc. Natl. Acad. Sci. U. S. A. 89, 8813–8817.

Lairmore, M.D., Anupam, R., Bowden, N., Haines, R., Haynes 2nd, R.A.H., Ratner, L., Green, P.L., 2011. Molecular determinants of human T-lymphotropic virus type 1 transmission and spread. Virologie 3, 1131–1165. http://dx.doi.org/10.3390/v3071131.

Lairmore, M.D., Haines, R., Anupam, R., 2012. Mechanisms of human T-lymphotropic virus type 1 transmission and disease. Curr. Opin. Virol. 2, 474–481. http://dx.doi.org/10.1016/j.coviro.2012.06.007.

Maloney, E.M., Cleghorn, F.R., Morgan, O.S., Rodgers-Johnson, P., Cranston, B., Jack, N., Blattner, W.A., Bartholomew, C., Manns, A., 1998. Incidence of HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP) in Jamaica and Trinidad. J. Acquir. Immune Defic. Syndr. Hum. Retrovirology Off. Publ. Int. Retrovirology Assoc. 17, 167–170.

Matsuoka, M., 2010. HTLV-1 bZIP factor gene: Its roles in HTLV-1 pathogenesis. Mol. Aspects Med. Role of human retroviruses (HIVs and HTLVs) non structural proteins in virus host interaction 31, pp. 359–366. http://dx.doi.org/10.1016/j.mam.2010.06.002.

Matsuoka, M., Green, P.L., 2009. The HBZ gene, a key player in HTLV-1 pathogenesis. Retrovirology 6, 71. http://dx.doi.org/10.1186/1742-4690-6-71.

Matsuzaki, T., Nakagawa, M., Nagai, M., Usuku, K., Higuchi, I., Arimura, K., Kubota, H., Izumo, S., Akiba, S., Osame, M., 2001. HTLV-I proviral load correlates with progression of motor disability in HAM/TSP: analysis of 239 HAM/TSP patients including 64 patients followed up for 10 years. J. Neurobiol. 7, 228–234. http://dx.doi.org/10.1080/13550280152403272.

Nagai, M., Usuku, K., Matsumoto, W., Kodama, D., Takenouchi, N., Moritoyo, T., Hashiguchi, S., Ichinose, M., Bangham, C.R., Izumo, S., Osame, M., 1998. Analysis of HTLV-I proviral load in 202 HAM/TSP patients and 243 asymptomatic HTLV-I carriers: high proviral load strongly predisposes to HAM/TSP. J. Neurobiol. 4, 586–593.

Pineda-Pena, A.-C., Schrooten, Y., Vinken, L., Ferreira, F., Li, G., Trovao, N.S., Khouri, R., Derdelinckx, I., De Munter, P., Kucherer, C., Kostrikis, L.G., Nielsen, C., Littsola, K., Wensing, A., Stanojevic, M., Paredes, R., Balotta, C., Albert, J., Boucher, C., Gomez-Lopez, A., Van Wijngaerden, E., Van Ranst, M., Vercauteren, J., Vandamme, A.-M., Van Laethem, K., 2014. Trends and predictors of transmitted drug resistance (TDR) and clusters with TDR in a local Belgian HIV-1 epidemic. PLoS One 9, e101738. http://dx.doi.org/10.1371/journal.pone.0101738.

Pise-Masison, C.A., de Castro-Amarante, M.F., Enose-Akahata, Y., Buchmann, R.C., Fenizia, C., Washington Parks, R., Edwards, D., Fiocchi, M., Alcantara, L.C., Bialuk, I., Graham, J., Walser, J.-C., McKinnon, K., Galvão-Castro, B., Gessain, A., Venzon, D., Jacobson, S., Franchini, G., 2014. Co-dependence of HTLV-1 p12 and p8 functions in virus persistence. PLoS Pathog. 10. http://dx.doi.org/10.1371/journal.ppat.1004454.

Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D., Gallo, R.C., 1980. 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. 77, 7415–7419.

Prosperi, M.C.F., De Luca, A., 2012. Computational models for prediction of response to antiretroviral therapies. AIDS Rev. 14, 145–153.

Trovato, R., Mulloy, J.C., Johnson, J.M., Takemoto, S., de Oliveira, M.P., Franchini, G., 1999. A lysine-to-arginine change found in natural alleles of the human T-cell lymphotropic/leukemia virus type 1 p12(I) protein greatly influences its stability. J. Virol. 73, 6460–6467.

Van Prooyen, N., Gold, H., Andresen, V., Schwartz, O., Jones, K., Ruscetti, F., Lockett, S., Gudla, P., Venzon, D., Franchini, G., 2010. Human T-cell leukemia virus type 1 p8 protein increases cellular conduits and virus transmission. Proc. Natl. Acad. Sci. U. S. A. 107, 20738–20743. http://dx.doi.org/10.1073/pnas.1009635107.

WiaF, E., 2011. Data Mining: Practical Machine Learning Tools and Techniques.

Yamano, Y., Nagai, M., Brennan, M., Mora, C.A., Soldan, S.S., Tomaru, U., Takenouchi, N., Izumo, S., Osame, M., Jacobson, S., 2002. Correlation of human T-cell lymphotropic virus type 1 (HTLV-1) mRNA with proviral DNA load, virus-specific CD8(+) T cells, and disease severity in. Blood 99, 88–94.

Yoshida, M., Miyoshi, I., Hinuma, Y., 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proc. Natl. Acad. Sci. U. S. A. 79, 2031–2035.