

Clonality of HTLV-2 in Natural Infection

Anat Melamed¹, Aviva D. Witkover¹, Daniel J. Laydon¹, Rachael Brown¹, Kristin Ladell², Kelly Miners², Aileen G. Rowan¹, Niall Gormley³, David A. Price², Graham P. Taylor⁴, Edward L. Murphy⁵, Charles R. M. Bangham^{1*}

1 Section of Immunology, Imperial College London, Wright-Fleming Institute, London, United Kingdom, **2** Institute of Infection and Immunity, Cardiff University School of Medicine, Cardiff, United Kingdom, **3** Illumina, Little Chesterford, Essex, United Kingdom, **4** Section of Infectious Diseases, Imperial College London, Wright-Fleming Institute, London, United Kingdom, **5** Departments of Laboratory Medicine and Epidemiology/Biostatistics, University of California San Francisco and Blood Systems Research Institute, San Francisco, California, United States of America

Abstract

Human T-lymphotropic virus type 1 (HTLV-1) and type 2 (HTLV-2) both cause lifelong persistent infections, but differ in their clinical outcomes. HTLV-1 infection causes a chronic or acute T-lymphocytic malignancy in up to 5% of infected individuals whereas HTLV-2 has not been unequivocally linked to a T-cell malignancy. Virus-driven clonal proliferation of infected cells both in vitro and in vivo has been demonstrated in HTLV-1 infection. However, T-cell clonality in HTLV-2 infection has not been rigorously characterized. In this study we used a high-throughput approach in conjunction with flow cytometric sorting to identify and quantify HTLV-2-infected T-cell clones in 28 individuals with natural infection. We show that while genome-wide integration site preferences in vivo were similar to those found in HTLV-1 infection, expansion of HTLV-2-infected clones did not demonstrate the same significant association with the genomic environment of the integrated provirus. The proviral load in HTLV-2 is almost confined to CD8⁺ T-cells and is composed of a small number of often highly expanded clones. The HTLV-2 load correlated significantly with the degree of dispersion of the clone frequency distribution, which was highly stable over ~8 years. These results suggest that there are significant differences in the selection forces that control the clonal expansion of virus-infected cells in HTLV-1 and HTLV-2 infection. In addition, our data demonstrate that strong virus-driven proliferation *per se* does not predispose to malignant transformation in oncoretroviral infections.

Citation: Melamed A, Witkover AD, Laydon DJ, Brown R, Ladell K, et al. (2014) Clonality of HTLV-2 in Natural Infection. PLoS Pathog 10(3): e1004006. doi:10.1371/journal.ppat.1004006

Editor: Susan R. Ross, University of Pennsylvania School of Medicine, United States of America

Received: December 21, 2013; **Accepted:** February 2, 2014; **Published:** March 13, 2014

Copyright: © 2014 Melamed et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the Wellcome Trust. CRMB and DAP are Wellcome Trust Senior Investigators. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Niall Gormley is an employee of Illumina Inc, a public company that develops and markets systems for genetic analysis. This does not alter our adherence to all PLOS policies on sharing data and materials. The remaining authors have declared that no competing interests exist.

* E-mail: c.bangham@imperial.ac.uk

Introduction

The retroviruses HTLV-1 and HTLV-2 diverged from each other more than one million years ago [1] before becoming established in humans. They are similar in several crucial respects, with homologous genome structures that encode a number of regulatory proteins, including the pro-proliferative gene *tax* [2,3]. Both viruses are transmitted by transfer of infected lymphocytes via breast feeding, blood transfusion and sexual contact [4]. However, their geographical distributions are quite different. HTLV-1 is endemic in particular regions of Japan, sub-Saharan Africa, the Caribbean and South America [5], whereas HTLV-2 is primarily endemic in indigenous populations in Africa and the Americas, although it can also be found among injection drug users in Europe and the United States [4].

HTLV-1 causes both inflammatory and lymphoproliferative diseases. In contrast, HTLV-2 causes little disease. By following a large cohort of HTLV-1/2-infected and seronegative individuals for almost two decades, the HTLV Outcomes study (HOST) detected myelopathy and other neurologic abnormalities among HTLV-2-infected subjects [6,7], a finding supported by other studies [8]. HTLV-2 was also associated with an increase in both all-cause and cancer-related mortality [9], as well as persistently

elevated lymphocyte and platelet counts, suggesting chronic low-level inflammation [10]. However, no mechanistic inferences can yet be drawn.

An important distinction between HTLV-1 and HTLV-2 lies in their host cell predilection. Although they use the same cellular receptors [11], HTLV-1 preferentially infects CD4⁺ T-cells, whereas HTLV-2 favours CD8⁺ T-cells [12,13]. The biological basis for this difference is not clear. In vitro evidence suggests that the relative cell surface density of two host receptors, heparan sulphate proteoglycans and glucose-transporter 1 [14], determines host cell tropism. However, in vivo studies suggest that both T-cell lineages are efficiently infected by both viruses, and that subsequent proliferation of CD4⁺ or CD8⁺ T-cells driven by HTLV-1 or HTLV-2, respectively, leads to differential expansion of the two T-cell subsets [15].

It is known that HTLV-2, like HTLV-1, can immortalize human lymphocytes in culture [16,17]. Both HTLV-1 [18,19] and HTLV-2 [20,21] have also been shown to cause selective proliferation of certain infected T-cell clones in vivo. Although the molecular pathways by which the viral proteins drive cellular proliferation are well described [22], the mechanistic basis of selective clonal proliferation is not understood. We have recently shown that the genomic integration site and transcriptional

In each case the frequency of integration sites detected in certain chromosomes in vivo was remarkably greater (e.g. chromosome 13) or lower (e.g. chromosome 10) than expected by chance.

HTLV-1 and HTLV-2 were also similar with respect to features of the genomic environment flanking the provirus. In particular, activating and repressive histone marks were similarly enriched at integration sites compared to random expectation for both viruses (Figure 2B). Previously, we found that HTLV-1 integration was significantly more frequent than expected on a random basis in proximity to ChIP-seq-verified binding sites for certain transcription factors and chromatin modifying proteins, most notably STAT-1, p53 and HDACs [24]. In vivo, proviruses also lie near these binding sites more frequently than expected by chance. In the present study, we reproduced this observation in an independent cohort of HTLV-1-infected individuals, and observed a similar integration targeting preference for HTLV-2 (Figure 2C). Although the magnitude of bias toward these genomic sites was very similar for both viruses, statistical significance was lower in the case of HTLV-2, most likely owing to the lower number of total integration sites.

HTLV-2 integration is characterized by small numbers of expanded clones

In an earlier study, we showed that HTLV-1 infection is characterized by very large numbers of clones (over 4,000 unique integration sites [UIS] have been observed in 10 µg of PBMC-derived DNA) and that much of the load (in non-malignant cases) is composed of low abundance clones [23]. We confirmed this observation here for HTLV-1, but the clonal distribution in HTLV-2 infection showed several marked differences (Figure 3A). Significantly lower numbers of unique integration sites were identified in samples from HTLV-2-infected individuals (median 16 UIS) than those from HTLV-1-infected individuals (median 766 UIS; Figure 3B). Using the recently developed biodiversity estimator '*DivE*' (Laydon et al., manuscript submitted), HTLV-1-infected subjects were estimated to carry a median of 31,710 distinct clones in the blood (consistent with previous estimates for HTLV-1), whereas HTLV-2-infected subjects were estimated to carry a median of only 976 clones ($p < 0.001$, Mann-Whitney test; Figure 3C).

The distribution of proviral load across identified clones also differed significantly between HTLV-1 and HTLV-2. To compare the two viruses, we used the oligoclonality index [23], a parameter based on the Gini index, as a measure of dispersion describing the magnitude of unevenness of a frequency distribution. The oligoclonality index ranges between 0 and 1, where a value of 0 represents a distribution in which each clone constitutes an equal share of the proviral load, and 1 represents an upper bound where the load is effectively made up by a single clone. A median oligoclonality index of 0.34 was observed in non-malignant HTLV-1 carriers, consistent with our previous findings. In contrast, the oligoclonality index was remarkably variable between HTLV-2-infected individuals, and on average significantly higher than in HTLV-1-infected individuals (median 0.73, $p < 0.001$, Mann-Whitney test; Figure 3D). Furthermore, the proportion of singletons (clones identified only once) was significantly lower in HTLV-2 infection (median = 1.96%) than in HTLV-1 infection (median = 37.17%, $p < 0.001$, Mann-Whitney test; Figure 3E).

Features of clonal expansion and proviral load in HTLV-2 infection

The difference in clonal distribution between HTLV-1 and HTLV-2 was also apparent when measuring the absolute

abundance of each clone as copies per 10,000 PBMCs (Figure 4A). In particular, highly expanded clones (i.e. those that each made up more than 0.1% of PBMCs) represented 20% of all HTLV-2 clones but only a fraction of all HTLV-1 clones (0.18%).

We reported previously that the genomic environment flanking the integration site appears to play a role in determining the equilibrium abundance of a given clone in vivo [23,24]. The positive effect of integration within 10 kb of a RefSeq transcription start site on the abundance of HTLV-1 clones was observed here again ($p = 0.04$, chi-squared test for trend), but there was no correlation between the abundance of HTLV-2 clones and the proximity of a RefSeq gene (Figure 4B).

Whereas there was no significant correlation between the oligoclonality index and the proviral load in HTLV-1 infection ($p = 0.681$, $\rho = 0.112$, Spearman's test), we observed a highly significant positive correlation ($p = 0.0015$, $\rho = 0.599$, Spearman's test) between these parameters in HTLV-2 infection (Figure 4C). In contrast, the proviral load in HTLV-1 infection was more strongly correlated with the total number of clones ($p < 0.001$, $\rho = 0.785$, Spearman's test) compared to that in HTLV-2 ($p = 0.0019$, $\rho = 0.578$, Spearman's test; Figure 4D).

Clonal distribution in HTLV-2 infection does not change over time

To test whether the expanded clones observed in HTLV-2 infection were long-lived, we analysed integration site content in samples collected at an earlier time point (range = 7.5 to 14.4 years, median = 9.9 years) from 10 of the HTLV-2-infected individuals in our cohort. The vast majority of the load (median = 96%) at the later time-point was represented by clones already present at the earlier time-point (Figure 5A). These clones did not change significantly in terms of their relative abundance; those representing >1% of the load at the later time-point were significantly more likely to make up >1% of the load at the earlier time-point and vice versa ($p < 0.001$, OR = 7.73, Fisher's exact test; Figure 5B).

Discussion

Both HTLV-1 and HTLV-2 infect the susceptible host by the same routes, and propagate within the host by the same two non-mutually-exclusive routes: the infectious route, which results in proviral integration at a new genomic site; and the mitotic route, where the provirus is replicated passively when the infected cell undergoes DNA replication and mitosis. It therefore benefits these viruses to drive proliferation of the infected cell. Indeed, the Tax proteins from both HTLV-1 and HTLV-2 have been shown in vitro to accelerate progression through the cell cycle, inhibit apoptosis and transform cells [3,32,33]. Consistent with these in vitro observations, we showed previously using metabolic labelling that cells spontaneously expressing the HTLV-1 Tax protein ex vivo proliferate faster in vivo [34].

HTLV-1 infection causes a chronic or acute T-lymphocytic malignancy in up to 5% of infected individuals [35,36]; however, HTLV-2 is not unequivocally linked to a T-cell malignancy. Compared with the HTLV-1 Tax protein, HTLV-2 Tax is more dependent on IL-2 for the transformation of cells in culture [37]. This observation led to the suggestion that HTLV-2 would cause less in vivo proliferation of infected cells than HTLV-1, which in turn would decrease the oncogenic potential of the virus [26]. However, two lines of evidence go against this model. The first is the recent finding that HTLV-2 Tax has a greater in vitro immortalization capacity than HTLV-1 Tax in primary human T cells [38,39]. The second is the finding reported here that

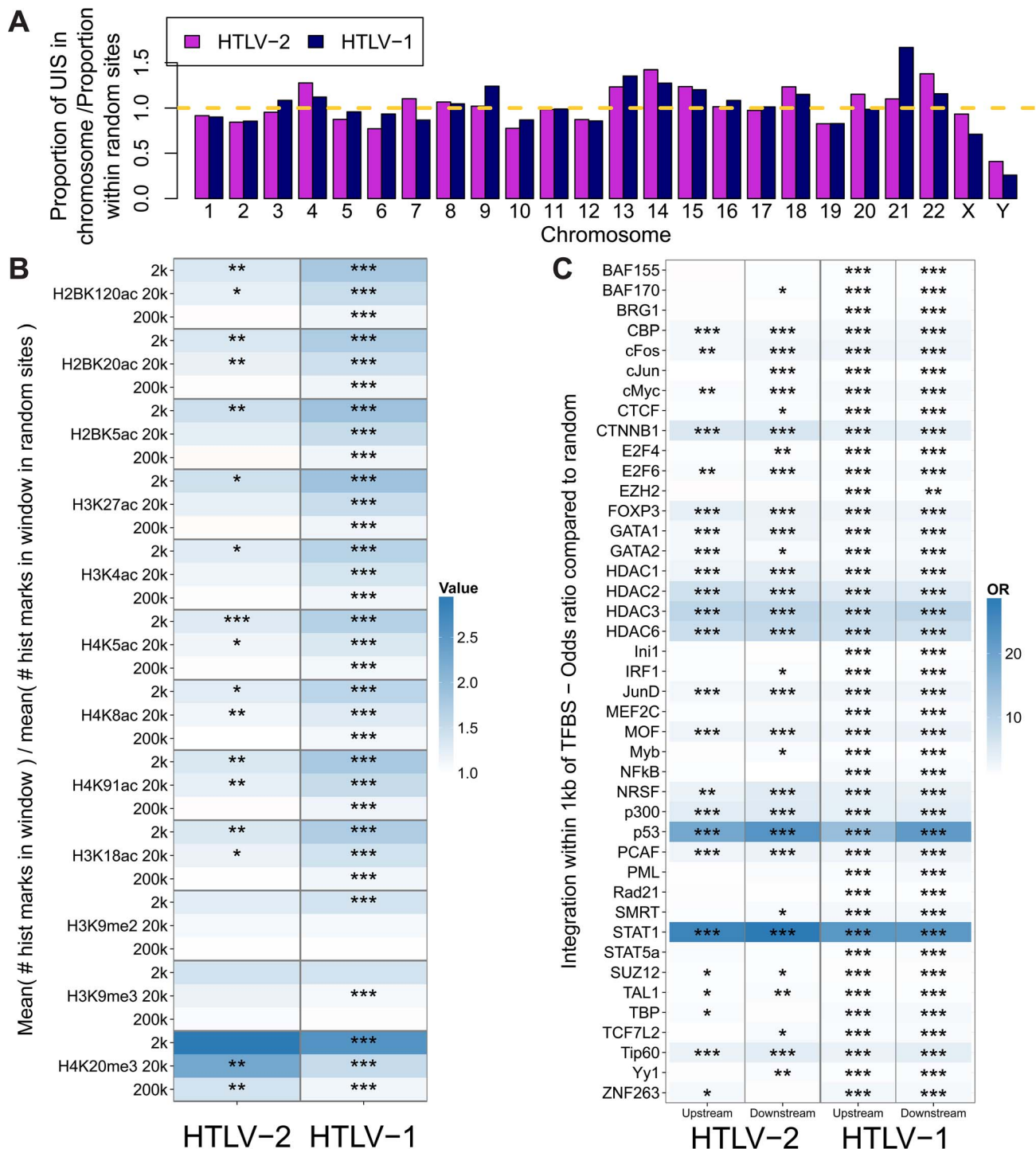


Figure 2. HTLV-1 and HTLV-2 integrate in similar genomic environments. Cells from 28 HTLV-2-infected and 16 HTLV-1-infected subjects were tested for genomic integration site preferences. (A) The ratio of the proportion of sites found in each chromosome (out of the total integration sites found for each virus) to the proportion of randomly generated (in silico) sites in the same chromosome. The yellow dashed line represents random sites (ratio = 1). (B) The number of histone marks (post-translational modifications) in three given windows across integration sites (for example, the 2k window incorporates 1,000 bases on either side of an integration site) compared to the number of histone marks in the same window across random sites. Statistical significance was assessed using the two-tailed Mann-Whitney test (* < 0.05, ** < 0.01, *** < 0.001). (C) The odds ratio of integration within 1 kb of given ChIP-seq sites compared to random sites. The terms upstream and downstream here refer to the 5' and 3' sides of the integrated provirus, respectively. Statistical significance was assessed using Fisher's exact test (* < 0.05, ** < 0.01, *** < 0.001). doi:10.1371/journal.ppat.1004006.g002

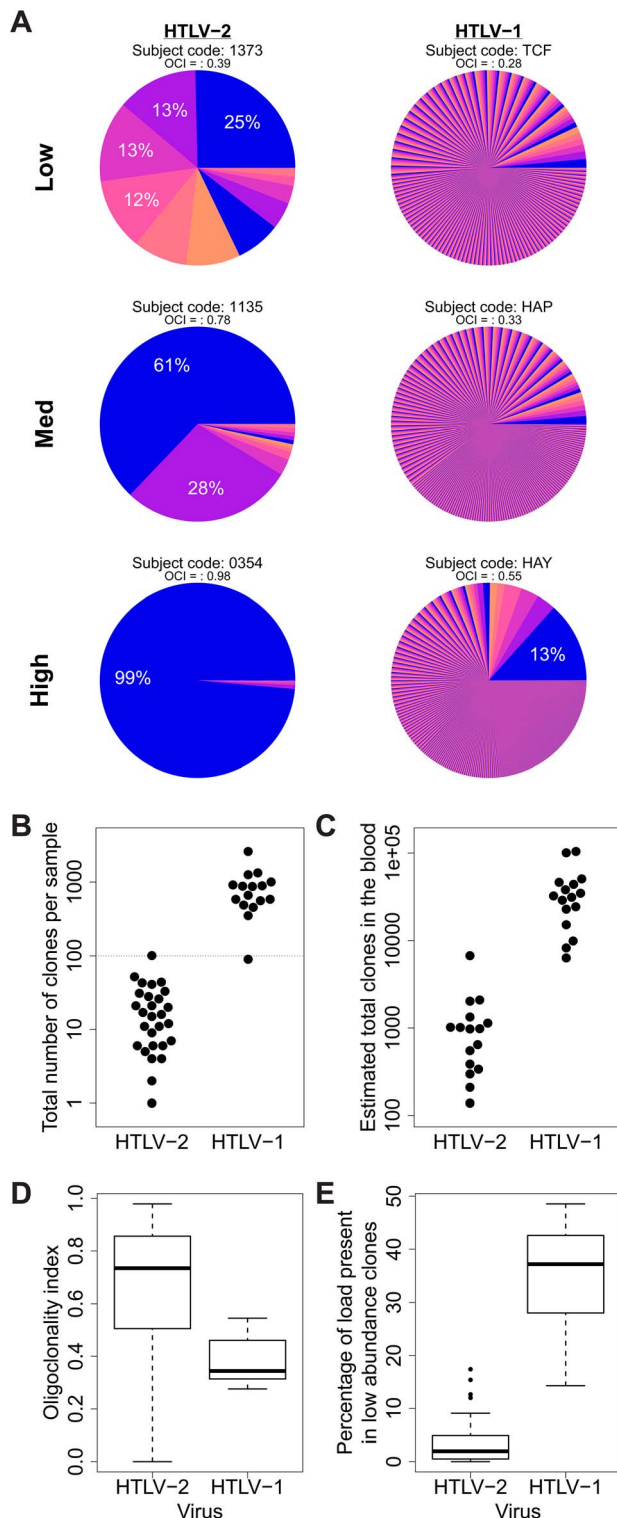


Figure 3. HTLV-2 integration is highly oligoclonal, characterized by small numbers of expanded clones. (A) Clonal distribution in representative subjects with HTLV-1 or HTLV-2 infection. The lowest observed, median and highest observed oligoclonality index values are shown. Each pie slice represents a single clone, proportional to relative abundance. Subjects with >100 proviruses identified are shown. OCI = oligoclonality index. (B) The observed number of clones in each subject with HTLV-1 or HTLV-2 infection ($p < 0.001$, Mann-Whitney test). (C) The total number of clones in the blood was estimated using the *DivE* estimator (Laydon et al., manuscript submitted). Only samples

containing sufficient information are shown. For each subject, the population size of infected cells in the blood was estimated based on the proviral load and average PBMC count. The estimated total number of clones in the blood was between 1 and 2 orders of magnitude lower in HTLV-2-infected subjects than in HTLV-1-infected subjects ($p < 0.001$, Mann-Whitney test). (D) The oligoclonality index across all HTLV-1-infected subjects compared to HTLV-2-infected subjects ($p < 0.001$, Mann-Whitney test). (E) The percentage of the load maintained by clones observed only once compared between HTLV-1 and HTLV-2 ($p < 0.001$, Mann-Whitney test). doi:10.1371/journal.ppat.1004006.g003

HTLV-2 infection in vivo results in a small number of highly expanded T-cell clones (Figure 3). Although non-malignant HTLV-1 infection can result in the preferential expansion of certain clones, including clones that contain the provirus at genomic sites with particular characteristics [23,24], HTLV-2 infection is capable of driving infected T-cells to proliferate selectively, generating clones which are often more highly expanded than those observed in non-malignant HTLV-1 infection (Figure 4A). The resulting clone frequency distribution in HTLV-2 infection is more similar to that observed in Adult T-cell Leukemia/Lymphoma (ATLL) patients than in non-malignant HTLV-1 infection (compare, for example, Figure 3D here with Figure 2B in [23]). That is, HTLV-1 infection is characterized by a large number of distinct clones in the circulation, while HTLV-2 infection is confined to a small number of highly expanded clones (Figure 3).

The host genomic environment flanking HTLV-2 integration sites in vivo closely resembles that of HTLV-1 integration sites. Similarities are evident at the nucleotide and the chromosome levels, and when examining particular genomic features known to be more frequent in proximity to HTLV-1 integration sites than expected by chance [24] (Figure 2). This observation is likely to result from the similarity between the HTLV-1 and HTLV-2 integrases, leading to shared targeting preferences during initial infection and integration. Since the samples analysed here (for both viruses) were drawn from patients infected for many years, this result suggests that there are also similar selection forces acting upon the infected cells in vivo in the two respective infections; the major selection force that differs between infected individuals is likely to be the acquired immune response, in particular cytotoxic CD8⁺ T-lymphocyte (CTL) activity [40].

The selection forces that act upon HTLV-1-infected clones have been the subject of many previous studies [41]. Two principal opposing forces govern the abundance of each clone in vivo: the ability of the clone to proliferate (e.g. through Tax-mediated cell proliferation or through antigen-mediated activation), and the susceptibility of the clone to elimination by CTL-mediated cell killing. We recently demonstrated that the genomic environment at the proviral integration site is associated with the clonal expansion in HTLV-1-infected individuals and with the tendency of a given clone to express the HTLV-1 Tax protein [23,24]. Further, cells that spontaneously expressed the HTLV-1 Tax protein belonged more frequently to low-abundance clones in vivo compared with non-Tax-expressing cells, suggesting that the expression of this dominant T-cell immunogen [42,43] limits proliferation in vivo. We suggest that this limited proliferation of Tax-expressing cells is due to counter-selection by the abundant, chronically activated Tax-specific CTLs.

The immune response to HTLV-2 proteins is less well understood. Oliveira and colleagues showed that high frequencies of CTLs specific for HTLV-2 Tax can be found in the circulation of HTLV-2 carriers [44]. Thus, while the genomic site preferences for HTLV-1 were mirrored in HTLV-2 infection, it is surprising

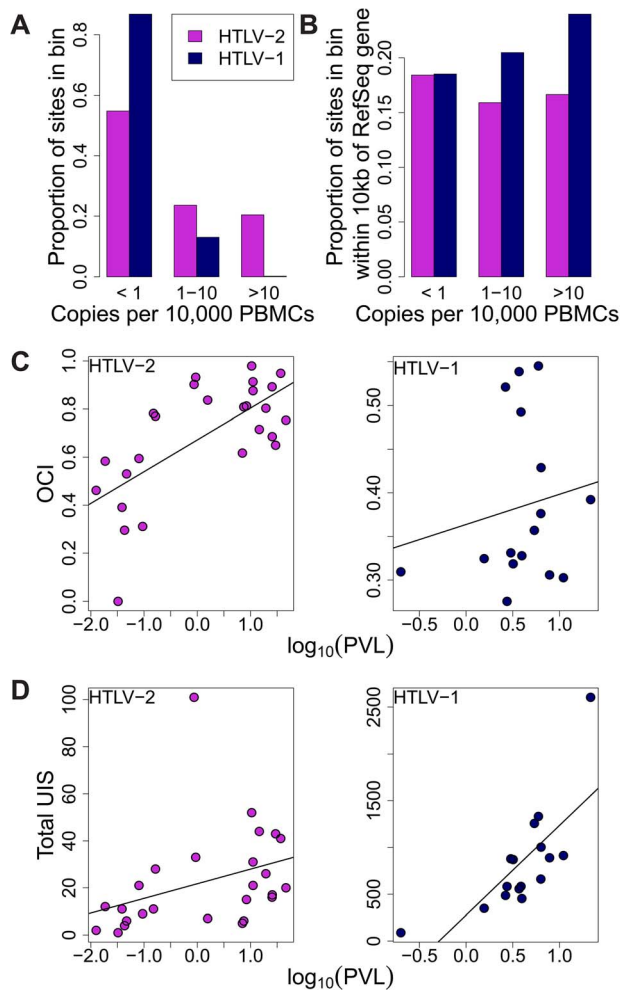


Figure 4. HTLV-2 integration site and clonal expansion. (A) The distribution of integration sites according to clonal abundance. Abundance was quantified by the number of copies estimated in a clone per 10,000 PBMCs (based on relative abundance and proviral load). Abundance bins are defined on a logarithmic scale. (B) The proportion of sites within 10 kb of a RefSeq gene for each abundance bin. A significant positive trend ($p = 0.04$, chi-squared test for trend) was detected for HTLV-1 but not for HTLV-2. (C) Oligoclonality index (OCI) versus $\log_{10}$ (proviral load) for each virus. A strong positive correlation ($p = 0.0015$, Spearman's test) was detected between these parameters for HTLV-2 but not for HTLV-1 ($p = 0.681$, Spearman's correlation). (D) The total number of unique integration sites (UIS) identified in each PBMC sample versus $\log_{10}$ (proviral load) for each virus ($p < 0.001$ for HTLV-1, $p = 0.0019$ for HTLV-2, Spearman's test). doi:10.1371/journal.ppat.1004006.g004

that the integration site plays a lesser role as a determinant of clonal expansion in HTLV-2-infected individuals (Figure 4B).

Although abundant clones (absolute abundance >10 cells per 10,000 PBMCs) represent only a small fraction of all infected clones in HTLV-1 infection, they represented approximately 20% of all HTLV-2 clones in this study (Figure 4A). There are two possible explanations for this discrepancy: either HTLV-2 clones are not controlled as efficiently as HTLV-1 clones by the immune response, or there is an unidentified driver (in addition to the virus itself) that determines the proliferation of HTLV-2 clones. One potential additional driver is antigenic stimulation of the infected cells.

Regardless of the forces that drive this vigorous clonal proliferation of HTLV-2-infected cells, the observed correlation

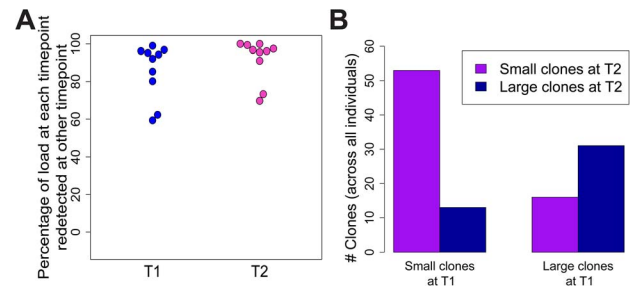


Figure 5. HTLV-2 expanded clones are long-lived and stable. Integration sites identified in PBMCs from an early time-point (T1, median = 9.9 years) were compared to those identified in the same HTLV-2-infected subjects at the present time-point (T2). (A) The percentage of proviral load (cumulative relative abundance) at each time-point represented by clones also present at the other time-point. (B) For clones found at both time-points, expanded clones ($>1\%$ of load) at any one time-point were significantly more likely to be expanded at both time-points ($p < 0.001$, Fisher's exact test). doi:10.1371/journal.ppat.1004006.g005

between the oligoclonality index and proviral load in HTLV-2 infection (a correlation not observed in non-malignant HTLV-1 infection, see Figure 4C) suggests that clonal proliferation plays a greater role in determining the viral burden of HTLV-2 than it does in HTLV-1. Conversely, in HTLV-1 infection, the total number of clones is more important as a determinant of viral burden (Figure 4D), consistent with previous observations [23,28]. We conclude that the proviral load in HTLV-1 infection – and therefore the risk of both inflammatory and malignant disease – is determined primarily by the extent of infectious spread of the virus, and that oligoclonal proliferation per se, contrary to previous belief, does not contribute to HTLV-1-associated diseases. It remains an important question whether infectious spread is mainly confined to the early stages of infection or whether it persists indefinitely, with continual formation and destruction of many low-abundance clones. Work is now in progress to quantify the ratio of infectious spread to mitotic spread in these two infections.

Given the observations that HTLV-2-infected clones proliferate to a greater extent than many HTLV-1-infected clones in vivo, and that HTLV-2 shows transformation potential in vitro [3,45], it is puzzling that HTLV-2 is not associated with a T-cell malignancy. One potential explanation is that the expansion of HTLV-2 clones in vivo is short-lived, and that major clones succeed each other over time. To test this possibility, integration sites and clonal distribution in PBMCs taken at an earlier time-point in the infection were compared to those identified ~ 10 years later in 10 HTLV-2-infected individuals. Although there was no significant difference in the oligoclonality index or the total number of clones after correcting for the different numbers of proviruses detected (not shown), the bulk of the load at the later time-point (median = 96%, Figure 5A) belonged to clones already present at the earlier time-point, and that these clones principally maintained their expanded state over that period of time (Figure 5B). A similar observation was made in HTLV-1 infection by Gillet et al [23]. These observations reinforce the conclusion reached above that oligoclonal proliferation of infected T-cells in vivo does not in itself predispose to malignant disease in these retroviral infections.

To determine the proportion of the proviral load carried by $CD4^+$ and $CD8^+$ T-cells, we analysed the relative contribution of each identified clone to the load to calculate the cumulative contribution of each cell type. Using this method we found that HTLV-2 was primarily restricted to $CD8^+$ T-cell clones

(mean = 96.3%; Figure 1). The main limitation of this method is the sampling probability - i.e. the chance of redetection. If a clone is detected in unsorted PBMCs but not in the sorted sample it is not possible to attribute the clone to either CD4⁺ or CD8⁺ cells, or to distinguish the lack of redetection by chance from the possibility that the load is present in a different cell type (e.g. B cells [46]). However, since more abundant clones are more likely to be redetected in repeated experiments (N. Gillet and H. Niederer, unpublished observations), the fact that 70% of high-abundance clones (each constituting >1% of the load) were redetected compared with only 40% of low-abundance clones (each constituting <1% of load) in one of the cell-sorted populations suggests that low clone abundance, rather than an untested cell type, was responsible for the small fraction of the load not identified within either the CD4⁺ or CD8⁺ T-cell compartments.

It remains unclear what controls the proliferation of HTLV-2 clones in vivo, and what mechanisms underlie the difference in oncogenic potential of the two viruses. Possible factors include differences between HTLV-1 and HTLV-2 in the actions of the respective Tax protein [47] or the antisense proteins HBZ (HTLV-1) and APH-2 (HTLV-2) [48]. Also, CD4⁺ and CD8⁺ T cells may differ in their susceptibility to malignant transformation. A useful insight may be found by examining the clonal distribution of CD8⁺ cells infected with HTLV-1. A comparison between the clonal distribution of HTLV-1 and HTLV-2 in CD8⁺ cells may enable a distinction between effects due to infected cell phenotype and effects due to the differences in viral genome. This project is currently underway.

In summary, we report a comprehensive analysis of integration site preferences and clonal distribution in HTLV-2 infection. By comparison with similar data from HTLV-1-infected individuals, our results suggest an important distinction between virus-driven cell proliferation and virus-driven malignancy, and strengthen the conclusion that oligoclonal proliferation per se does not predispose to malignant transformation.

Methods

Ethics statement

UK blood samples were obtained through the Communicable Diseases Tissue Bank at Imperial College, approved by the UK National Research Ethics Service (NRES reference 09/H0606/106). Samples, with data linkage, were donated by HTLV-1 or HTLV-2-infected subjects attending the National Centre for Human Retrovirology, St Mary's Hospital, Imperial College Healthcare NHS Trust, London after giving written informed consent. HOST Study approved by the University of California San Francisco Committee on Human Research.

Patients and cells

Cryopreserved PBMCs from 16 HTLV-1-infected and 28 HTLV-2-infected individuals were used in this study (Supplementary Table S1). Twenty-six HTLV-2-infected subjects were recruited to the HOST cohort, a long-term study of outcomes of HTLV-1 and HTLV-2 infection [49]. Two HTLV-2 and all 16 HTLV-1-infected subjects were Communicable Diseases Tissue Bank donors. Proviral load data on the HTLV-1-infected individuals were reported previously [50]. Genomic integration sites in 6 of the 16 HTLV-1-infected individuals were also studied previously, albeit at a distinct time-point in each case [23].

All DNA extractions were carried out using a DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's protocol.

Quantification of proviral load

HTLV-2 proviral load was quantified as reported elsewhere [51]. We used the proviral load for each patient at the nearest available time-point, because the proviral load was not always known for each at the same time-point at which clonality was analysed, and because HTLV-2 proviral load is reported to be stable over time [52]. Consistent with this assumption, our findings were not qualitatively altered by using a proviral load measurement from a different time-point.

Cell sorting

Cells were thawed and washed, then surface-stained with directly-conjugated monoclonal antibodies specific for CD3, CD4 and CD8. Flow cytometric sorting was conducted to high purity (>98%) using a custom-modified FACSaria II (BD Biosciences). Lymphocytes were pre-gated on CD3, then sorted as CD4⁺CD8⁻ and CD4⁻CD8⁺ populations. Data were analysed with FACSDiva v6 software (BD Biosciences). For each sample, DNA was extracted from both sorted populations and from a separate aliquot of unsorted PBMCs.

Integration site analysis

Ligation-mediated polymerase chain reaction (LM-PCR) primer binding site sequences were determined from Sanger-sequenced PCR amplicons (HTLV-1 primers: 5LTRfw -CTCGCATCT-CTCCTTCACG, 5LTRrev - CTGGTGGAATCGTAACCTGGA; HTLV-2 primers: H2LTRfw - GACTCACCTTGGGGA-TCCAT, H2LTRrev - TTAGCCAAATGGGCGTTTTTA). Identification of integration sites was performed via LM-PCR followed by high-throughput sequencing as described previously [23], using the HTLV-2-specific forward primers: H2B3 - AAGGGCTAG-GGCTTCCTGAACCTC and H2B5 - CTATAGGCAGGCC-CGCCCCAGGAG (or variants thereof according to defined LTR polymorphisms).

Prepared libraries were mixed and sequenced using either a Genome Analyzer II or a HiSeq System (Illumina). The resulting sequences were aligned to the human genome reference (hg18, excluding haplotypes, randoms and mitochondrial DNA) and HTLV-1/2 upstream sequence using the *cland_pair* implementation of CASAVA 1.8.2 (Illumina).

Bioinformatics

Integration sites were quantified by enumerating unique shear sites as described previously [23]. Bins of absolute clonal abundance were determined by the number of copies per 10,000 PBMCs (i.e. relative clonal abundance multiplied by proviral load). In flow-sorted samples, clones were attributed exclusively either to the CD4⁺ or CD8⁺ cell population. Six clones were initially detected in both CD4⁺ and CD8⁺ cells; in these cases the clone was ascribed to the cell type with the greater number of proviruses of that clone.

In silico sites were derived by random selection of 190,000 sites from the human genome (hg18). To eliminate any potential bias due to alignment limitation, the DNA sequences at those sites were generated using Galaxy [53,54] and back-aligned to the human genome using the same pipeline.

Annotations to the human genome (hg18) were retrieved as described previously (see Table S3 in [24]) and compared to integration sites using the R package *hiAnnotator* (<http://malnirav.github.com/hiAnnotator>), kindly provided by Nirav Malani and Frederic Bushman (University of Pennsylvania, Philadelphia, USA).

The *DivE* estimator was used to estimate the total number of clones, in addition to those observed. *DivE* involves fitting many

mathematical models to individual-based rarefaction curves (Laydon et al., manuscript submitted). Here, individual-based rarefaction curves depict the expected number of clones against the number of infected cells sampled. Four numerical criteria are used to score each model in terms of how consistently it reproduces the total observed rarefaction curve from nested subsamples thereof. Using the geometric mean, estimates from the best-performing models are aggregated to produce the final estimate. Samples with near-linear rarefaction curve (i.e. where the curvature was less than 0.1 (Laydon et al., manuscript submitted)) were excluded from the analysis. Such curves imply a (biologically impossible) linear relationship between the number of infected cells and the number of clones, which is indicative of severe under-sampling. Prohibitively small sample sizes of less than 150 proviruses were also excluded from the analysis. *DivE* requires an estimate of the number of cells in the blood N_{blood} for which we assume: (i) a circulating blood volume of 5L; (ii) a PBMC count of $3 \times 10^9 \text{ L}^{-1}$; and (iii) each infected T-cell carries a single copy of the provirus [55]. The number of PBMCs is thus assumed to be $5 \times 3 \times 10^9$. Proviral load (PVL) is defined as the number of viral copies per 100 PBMCs. Therefore, N_{blood} is given by $\text{PVL} \times 5 \times 3 \times 10^9$.

Statistics

Statistical analysis was carried out using R version 2.15.2 (<http://www.R-project.org/>). The Gini coefficient was calculated using the *reldist* R package ([56]; <http://CRAN.R-project.org/package=reldist>). Two-tailed non-parametric tests including the Mann-Whitney and Fisher's exact test were used for all comparisons.

Supporting Information

Figure S1 Experimental output per tissue type. High-throughput sequencing analyses of integrated HTLV-2 proviruses are shown for each tissue type. $\text{CD4}^+/\text{CD8}^+$ cells – PBMCs from 28 HTLV-2-infected subjects were sorted by flow-cytometry into distinct $\text{CD3}^+\text{CD4}^+\text{CD8}^-$ and $\text{CD3}^+\text{CD4}^-\text{CD8}^+$ populations. PBMCs – unsorted PBMCs from 28 HTLV-2-infected individuals. PBMC early – unsorted PBMCs isolated previously from 10 of the

HTLV-2-infected subjects. (A) The total number of sequencing reads for each HTLV-2 sample. (B) The total number of infected cells (distinct proviruses) identified in each HTLV-2 sample. (C) The total number of unique integration sites identified in each HTLV-2 sample.

(TIF)

Figure S2 Integration bias at the sequence level.

Nucleotide sequences at the genomic sites of HTLV-2 integration are summarized by position and base. The arrow denotes the position of proviral integration. Base 1 is the first nucleotide following the integrated provirus. For each position, the relative proportion of sites containing each base is noted. Remarkably favoured (red) or disfavoured (blue) bases compared to random (in silico) sites are highlighted.

(TIF)

Table S1 Details of subjects in study.

(DOCX)

Acknowledgments

The authors thank Laurence Game, Adam Giess and Marian Dore (Genomics Laboratory, Medical Research Council Clinical Sciences Centre, Hammersmith Hospital, London, UK). The authors wish to thank the High Performance Computing service staff at Imperial College (<http://www.imperial.ac.uk/ict/services/teachingandresearchservices/highperformancecomputing>). We thank Maria-Antonieta Demontis (Section of Infectious Diseases, Imperial College London, London, UK) and the staff and donors at the National Centre for Human Retrovirology (Imperial College Healthcare NHS Trust, St Mary's Hospital, London, UK) and in the US. A list of random integration sites and essential software packages were kindly provided by Nirav Malani and Frederic Bushman (Department of Microbiology, University of Pennsylvania, Philadelphia, PA, USA).

Author Contributions

Conceived and designed the experiments: AM NG DAP GPT CRMB. Performed the experiments: AM ADW RB KL KM AGR. Analyzed the data: AM DJL. Contributed reagents/materials/analysis tools: ELM GPT. Wrote the paper: AM DAP CRMB.

References

- Salemi M, Desmyter J, Vandamme AM (2000) Tempo and mode of human and simian T-lymphotropic virus (HTLV/STLV) evolution revealed by analyses of full-genome sequences. *Mol Biol Evol* 17: 374–386.
- Akagi T, Shimotohno K (1993) Proliferative response of Tax1-transduced primary human T cells to anti-CD3 antibody stimulation by an interleukin-2-independent pathway. *J Virol* 67: 1211–1217.
- Ross TM, Pettiford SM, Green PL (1996) The tax gene of human T-cell leukemia virus type 2 is essential for transformation of human T lymphocytes. *J Virol* 70: 5194–5202.
- Roux DF, Murphy EL (2004) The epidemiology and disease outcomes of human T-lymphotropic virus type II. *AIDS Rev* 6: 144–154.
- Gessain A, Cassar O (2012) Epidemiological Aspects and World Distribution of HTLV-1 Infection. *Front Microbiol* 3: 388.
- Orland JR, Engstrom J, Frider J, Sacher RA, Smith JW, et al. (2003) Prevalence and clinical features of HTLV neurologic disease in the HTLV Outcomes Study. *Neurology* 61: 1588–1594.
- Biswas HH, Engstrom JW, Kaidarova Z, Garratty G, Gibble JW, et al. (2009) Neurologic abnormalities in HTLV-I- and HTLV-II-infected individuals without overt myelopathy. *Neurology* 73: 781–789.
- Araujo A, Hall WW (2004) Human T-lymphotropic virus type II and neurological disease. *Ann Neurol* 56: 10–19.
- Biswas HH, Kaidarova Z, Garratty G, Gibble JW, Newman BH, et al. (2010) Increased all-cause and cancer mortality in HTLV-II infection. *J Acquir Immune Defic Syndr* 54: 290–296.
- Bartman MT, Kaidarova Z, Hirschhorn D, Sacher RA, Frider J, et al. (2008) Long-term increases in lymphocytes and platelets in human T-lymphotropic virus type II infection. *Blood* 112: 3995–4002.
- Manel N, Kim FJ, Kinet S, Taylor N, Sitbon M, et al. (2003) The ubiquitous glucose transporter GLUT-1 is a receptor for HTLV. *Cell* 115: 449–459.
- Ijichi S, Ramundo MB, Takahashi H, Hall WW (1992) In vivo cellular tropism of human T cell leukemia virus type II (HTLV-II). *J Exp Med* 176: 293–296.
- Lai RB, Owen SM, Rudolph DL, Dawson C, Prince H (1995) In vivo cellular tropism of human T-lymphotropic virus type II is not restricted to CD8+ cells. *Virology* 210: 441–447.
- Jones KS, Fugo K, Petrow-Sadowski C, Huang Y, Bertolette DC, et al. (2006) Human T-cell leukemia virus type 1 (HTLV-1) and HTLV-2 use different receptor complexes to enter T cells. *J Virol* 80: 8291–8302.
- Kannian P, Yin H, Doueiri R, Lairmore MD, Fernandez S, et al. (2012) Distinct transformation tropism exhibited by human T lymphotropic virus type 1 (HTLV-1) and HTLV-2 is the result of postinfection T cell clonal expansion. *J Virol* 86: 3757–3766.
- Chen IS, Quan SG, Golde DW (1983) Human T-cell leukemia virus type II transforms normal human lymphocytes. *Proc Natl Acad Sci U S A* 80: 7006–7009.
- Tarsis SL, Yu MT, Parks ES, Persaud D, Munoz JL, et al. (1998) Human T-lymphocyte transformation with human T-cell lymphotropic virus type 2. *J Virol* 72: 841–846.
- Furukawa Y, Fujisawa J, Osame M, Toita M, Sonoda S, et al. (1992) Frequent clonal proliferation of human T-cell leukemia virus type 1 (HTLV-1)-infected T cells in HTLV-1-associated myelopathy (HAM-TSP). *Blood* 80: 1012–1016.
- Wattel E, Vartanian JP, Pannetier C, Wain-Hobson S (1995) Clonal expansion of human T-cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers without malignancy. *J Virol* 69: 2863–2868.

20. Cimorelli A, Duclos CA, Gessain A, Casoli C, Bertazzoni U (1996) Clonal expansion of human T-cell leukemia virus type II in patients with high proviral load. *Virology* 223: 362–364.
21. Gabet AS, Moules V, Sibon D, Nass CC, Mortreux F, et al. (2006) Endemic versus epidemic viral spreads display distinct patterns of HTLV-2b replication. *Virology* 345: 13–21.
22. Matsuoka M, Jeang KT (2011) Human T-cell leukemia virus type 1 (HTLV-1) and leukemic transformation: viral infectivity, Tax, HBZ and therapy. *Oncogene* 30: 1379–1389.
23. Gillet NA, Malani N, Melamed A, Gormley N, Carter R, et al. (2011) The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. *Blood* 117: 3113–3122.
24. Melamed A, Laydon DJ, Gillet NA, Tanaka Y, Taylor GP, et al. (2013) Genome-wide Determinants of Proviral Targeting, Clonal Abundance and Expression in Natural HTLV-1 Infection. *PLoS Pathog* 9: e1003271.
25. Wattel E, Cavrois M, Gessain A, Wain-Hobson S (1996) Clonal expansion of infected cells: a way of life for HTLV-I. *J Acquir Immune Defic Syndr Hum Retrovirol* 13 Suppl 1: S92–99.
26. Higuchi M, Fujii M (2009) Distinct functions of HTLV-1 Tax1 from HTLV-2 Tax2 contribute key roles to viral pathogenesis. *Retrovirology* 6: 117.
27. Gillet NA, Gutierrez G, Rodriguez SM, de Brogniez A, Renotte N, et al. (2013) Massive Depletion of Bovine Leukemia Virus Proviral Clones Located in Genomic Transcriptionally Active Sites during Primary Infection. *PLoS Pathog* 9: e1003687.
28. Gillet NA, Cook L, Laydon DJ, Hlela C, Verdonck K, et al. (2013) Strongyloidiasis and Infective Dermatitis Alter Human T Lymphotropic Virus-1 Clonality in vivo. *PLoS Pathog* 9: e1003263.
29. Cook LB, Elemans M, Rowan AG, Asquith B (2013) HTLV-1: persistence and pathogenesis. *Virology* 435: 131–140.
30. Gini C (1912) Variabilità e Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche. C. Cuppini, Bologna.
31. Meekings KN, Leipzig J, Bushman FD, Taylor GP, Bangham CR (2008) HTLV-1 integration into transcriptionally active genomic regions is associated with proviral expression and with HAM/TSP. *PLoS Pathog* 4: e1000027.
32. Sieburg M, Tripp A, Ma JW, Feuer G (2004) Human T-cell leukemia virus type 1 (HTLV-1) and HTLV-2 tax oncoproteins modulate cell cycle progression and apoptosis. *J Virol* 78: 10399–10409.
33. Robek MD, Ratner L (1999) immortalization of CD4(+) and CD8(+) T lymphocytes by human T-cell leukemia virus type 1 Tax mutants expressed in a functional molecular clone. *J Virol* 73: 4856–4865.
34. Asquith B, Zhang Y, Mosley AJ, de Lara CM, Wallace DL, et al. (2007) In vivo T lymphocyte dynamics in humans and the impact of human T-lymphotropic virus 1 infection. *Proc Natl Acad Sci U S A* 104: 8035–8040.
35. Murphy EL, Hanchard B, Figueroa JP, Gibbs WN, Lofters WS, et al. (1989) Modelling the risk of adult T-cell leukemia/lymphoma in persons infected with human T-lymphotropic virus type I. *Int J Cancer* 43: 250–253.
36. Yamaguchi K, Watanabe T (2002) Human T lymphotropic virus type-I and adult T-cell leukemia in Japan. *Int J Hematol* 76 Suppl 2: 240–245.
37. Tsubata C, Higuchi M, Takahashi M, Oie M, Tanaka Y, et al. (2005) PDZ domain-binding motif of human T-cell leukemia virus type 1 Tax oncoprotein is essential for the interleukin 2 independent growth induction of a T-cell line. *Retrovirology* 2: 46.
38. Ren T, Dong W, Takahashi Y, Xiang D, Yuan Y, et al. (2012) HTLV-2 Tax immortalizes human CD4+ memory T lymphocytes by oncogenic activation and dysregulation of autophagy. *J Biol Chem* 287: 34683–34693.
39. Imai M, Higuchi M, Kawamura H, Yoshita M, Takahashi M, et al. (2013) Human T cell leukemia virus type 2 (HTLV-2) Tax2 has a dominant activity over HTLV-1 Tax1 to immortalize human CD4+ T cells. *Virus Genes* 46: 39–46.
40. Bangham CR (2009) CTL quality and the control of human retroviral infections. *Eur J Immunol* 39: 1700–1712.
41. Bangham CR, Meekings K, Toulza F, Nejmeddine M, Majorovits E, et al. (2009) The immune control of HTLV-1 infection: selection forces and dynamics. *Front Biosci* 14: 2889–2903.
42. Kannagi M, Harada S, Maruyama I, Inoko H, Igarashi H, et al. (1991) Predominant recognition of human T cell leukemia virus type I (HTLV-I) pX gene products by human CD8+ cytotoxic T cells directed against HTLV-I-infected cells. *Int Immunol* 3: 761–767.
43. Goon PK, Biancardi A, Fast N, Igakura T, Hanon E, et al. (2004) Human T cell lymphotropic virus (HTLV) type-1-specific CD8+ T cells: frequency and immunodominance hierarchy. *J Infect Dis* 189: 2294–2298.
44. Oliveira AL, Hayakawa H, Schor D, Leite AC, Espindola OM, et al. (2009) High frequencies of functionally competent circulating Tax-specific CD8+ T cells in human T lymphotropic virus type 2 infection. *J Immunol* 183: 2957–2965.
45. Wang TG, Ye J, Lairmore MD, Green PL (2000) In vitro cellular tropism of human T cell leukemia virus type 2. *AIDS Res Hum Retroviruses* 16: 1661–1668.
46. Casoli C, Cimorelli A, Bertazzoni U (1995) Cellular tropism of human T-cell leukemia virus type II is enlarged to B lymphocytes in patients with high proviral load. *Virology* 206: 1126–1128.
47. Ren T, Cheng H (2013) Differential transforming activity of the retroviral Tax oncoproteins in human T lymphocytes. *Front Microbiol* 4: 287.
48. Barbeau B, Peloponese JM, Mesnard JM (2013) Functional comparison of antisense proteins of HTLV-1 and HTLV-2 in viral pathogenesis. *Front Microbiol* 4: 226.
49. Murphy EL, Watanabe K, Nass CC, Ownby H, Williams A, et al. (1999) Evidence among blood donors for a 30-year-old epidemic of human T lymphotropic virus type II infection in the United States. *J Infect Dis* 180: 1777–1783.
50. Kattan T, MacNamara A, Rowan AG, Nose H, Mosley AJ, et al. (2009) The avidity and lytic efficiency of the CTL response to HTLV-1. *J Immunol* 182: 5723–5729.
51. Lee TH, Chafets DM, Busch MP, Murphy EL (2004) Quantitation of HTLV-I and II proviral load using real-time quantitative PCR with SYBR Green chemistry. *J Clin Virol* 31: 275–282.
52. Kwaan N, Lee TH, Chafets DM, Nass C, Newman B, et al. (2006) Long-term variations in human T lymphotropic virus (HTLV)-I and HTLV-II proviral loads and association with clinical data. *J Infect Dis* 194: 1557–1564.
53. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, et al. (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* Chapter 19: Unit 19 10 11–21.
54. Goecks J, Nekrutenko A, Taylor J, Galaxy T (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11: R86.
55. Cook LB, Rowan AG, Melamed A, Taylor GP, Bangham CR (2012) HTLV-1-infected T cells contain a single integrated provirus in natural infection. *Blood* 120: 3488–3490.
56. Handcock MS, Morris M (1999) Relative distribution methods in the social sciences. New York: Springer. xiii, 265 p.