HTLV-1 clonality in natural infection

A thesis submitted to Imperial College London

for the degree of Doctor of Philosophy

By

Anat Melamed

Imperial College London

2013

Section of Immunology

Department of Medicine

Wright-Fleming Institute

Imperial College

Norfolk Place

London W2 1PG

Declaration

All work presented in this thesis is the author's own other than where clearly stated in the statement of collaborations.

Signature.....

Date.....

Summary

Human T-Lymphotropic virus type 1 (HTLV-1) maintains persistent infection in the host by driving expansion of HTLV-1-infected T cells. However, the wide variation observed in clonal abundance in vivo remains unexplained. We hypothesize that the site of proviral integration in the host cell genome determines the rate of expression of HTLV-1 genes such as the viral transactivator Tax, which in turn determine the level of clonal expansion in vivo.

Utilising Molecular biology techniques, High-throughput sequencing and computing tools, we have developed a high-throughput method for the amplification, mapping and quantification of proviral integration sites. We have used this technique, in combination with bioinformatics analysis, to identify factors associated with clonal expansion. Using PBMCs sorted for $CD4^+$ and $CD8^+$ T cell subsets, we observed differences in the form of the clonal distribution between the cell types. In addition, using flow-cytometric sorting of Tax expressing $CD4^+$ cells, we identified genomic patterns associated with the presence or absence of spontaneous Tax expression. We have also identified a correlation between the expression of Tax and the size of the infected clones. Finally, we identified genomic factors associated with HTLV-1 integration in vitro, suggesting a possible role of these factors in targeting of the pre-integration complex to particular sites in the host genome.

The thesis will detail and discuss the findings from these analyses, as well as their impact on the current knowledge. These results would play a role in improving our understanding of HTLV-1 persistence in the face of a strong immune response, and may provide excellent basis from a population level to test mechanisms of proviral latency in natural infection.

Acknowledgments

I would first like to thank my supervisor, Charles Bangham, for supervision, guidance, advice and inspiration.

I would like to thank the patients and staff at the National Centre for Human Retrovirology. In particular I would like to thank Graham Taylor and his team, especially Maria Antonietta Demontis, for patient recruitment and helpful advice.

I would like to thank Laurence Game, Adam Giess and Nathalie Lambie at the MRC Clinical Sciences Centre genomics laboratory, as well as Robert Sampson previously at the flow cytometry facility at St Mary's. Special thanks go to our collaborators, Nirav Malani, Frederic Bushman and Charles Berry, as well as Niall Gormley for all their help.

I would like to acknowledge the fantastic support, advice and encouragement by the former and present members of the immunology section, and in particular Nicolas Gillet and fellow students Sonja Tattermuch, Katerina Seich al Basatena, Lucy Cook and Danny Laydon.

I would like to thank Kristian Hood, for his big heart and boundless patience.

Finally, to my beloved ima and aba, Alisa and Izhak Melamed, for the support, encouragement and inspiration throughout, I am truly grateful.

Table of Contents

List of figures

List of Tables

List of abbreviations

AC	Asymptomatic carrier
ATLL	Adult T-cell leukemia/lymphoma
AZT	Azidothymidine (zidovudine)
BLV	Bovine leukemia virus
CA	Capsid (protein)
ChIP-Seq	Chromatin immunoprecipitation-sequencing
CNS	Central nervous system
CRE	cAMP responsive element
CREB/ATF	cAMP-responsive element binding/activating transcription factors
CSF	Cerebrospinal fluid
CTCF	CCCTC-binding factor
CTL	Cytotoxic T lymphocyte
DC	Dendritic cell
DEISA	Data Extraction for Integration Site Analysis
DMSO	Dimethyl-sulfoxide
DNA	Deoxyribonucleic acid
ELAND	Efficient Large-scale Alignment of Nucleotide Databases
ELISA	Enzyme-linked immuno sorbent assay
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
GFP	Green fluorescent protein
GLUT-1	Glucose transporter 1
HAART	Highly-active anti-retroviral therapy
HAM/TSP	HTLV-1-associated myelopathy/tropical spastic paraparesis
HAT	Histone acetyl transferase

HBV	Hepatitis B virus
HBZ	HTLV-1 bZIP factor
HDAC	Histone deacetylase
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HPV	Human papillomavirus
HRP-2	Hepatoma-derived growth factor related protein 2
HSPG	Heparan sulfate proteoglycan
HTLV	Human T-lymphotropic virus
ICAM	Intercellular adhesion molecule
IFN	Interferon
IL	Interleukin
IN	Integrase
INI1	Integrase interactor 1
IPCR	Inverse PCR
KIR	Killer cell immunoglobulin-like receptor
nrLAM-PCR	Non-restrictive linear amplification–mediated PCR
LEDGF	Lens epithelium-derived growth factor
LMPCR	Linker mediated PCR
LTR	Long terminal repeat
MHC	Major histocompatibility complex
MLV	Murine leukemia virus
MTOC	Microtubule organizing center
NEB	New England Biolabs
NFkB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NHP	Nonhuman primate
NLS	nuclear localization signal
NPC	Nuclear pore complex

NRES	National research ethics service
NRP	Neuropilin
NRSF	Neuron-restrictive silencer factor
OCI	Oligoclonality index
ORF	Open reading frame
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
PIC	Pre-integration complex
PVL	Proviral load
QPCR	Quantitative PCR
RNA	Ribonucleic acid
mRNA	Messenger RNA
St	Strongyloides Scoralis
STAT	Signal transducer and activator of transcription
STLV	Simian T-lymphotropic virus
TCR	T-cell receptor
TF	Transcription factor
TFBS	TF binding site
TLR	Toll-like receptor
TNFα	Tumor necrosis factor alpha
TRE	Tax-responsive elements
TSS	Transcription start site
UIS	Unique integration site
WBC	White blood cell
WHO	World health organization

Glossary

- UIS – unique integration site (defined by a specific genomic coordinate of HTLV-1 integration site).
- Sisters – HTLV-1-infected cells, sharing a common UIS, assumed to have arisen from a common progenitor.
- Clone – the entire population of sisters in a given sample.
- Clone abundance – the number of sisters in a clone
- Singleton – a clone observed only once (one sister, clone abundance = 1)
- Relative abundance – the relative proportion of the proviral load represented by a given clone (Equation 3.1).
- Absolute abundance – the number of sisters in a given clone per 10000 PBMCs (Equation 3.2).

Statement of collaboration

- Patient recruitment and diagnosis was carried out by Dr. Graham Taylor, National Centre for Human Retrovirology, St Mary's Hospital.
- Chapter 3:
    - The novel method for the high-throughput analysis of HTLV-1 integration sites was designed by Dr. Nicolas Gillet and developed by Dr Gillet and myself, in collaboration with Illumina R&D (in particular Dr. Niall Gormley, Associate Director, Scientific Research).
    - Early experiments generating initial data for calibration curve were carried out by Dr Gillet (clearly indicated on relevant figures). Calibration curve formula was fitted to experimental result by Dr. Charles Berry (UC San Diego).
    - My main contributions to the development of the method were in carrying out supporting experiments and analysis and designing the bioinformatics tools used by myself (now also used by other lab members) to extract the data generated by the high-throughput system (not used by Dr. Gillet).
- Chapter 4 and Chapter 5:
    - See Table 3.1 for details of datasets used. Integration sites isolated from *in vitro* infection of T cells were isolated in two independent experiments by Nicolas Gillet (Gillet et al., 2011) and myself. *In vivo* integration sites reported in chapter 4 were identified and quantified by Nicolas Gillet (Gillet et al., 2011). Nirav Malani (Group of FD Bushman, U Penn, USA) developed the hiAnnotator R package and generated *in silico* integration sites.
    - Multivariate analysis test was designed with Daniel Laydon.

Chapter 1 - **Introduction**

1.1.Human T-Lymphotropic virus

Human T-Lymphotropic virus Type 1 (HTLV-1) was the first identified human retrovirus. Originally identified from a cell line developed from the cells of a patient with cutaneous T cell leukemia (Poiesz et al., 1980), it is currently estimated to infect more than 10 million individuals worldwide, and while it is often carried as an asymptomatic persistent infection, it is associated with a number of diseases, in particular inflammatory disease and T cell malignancies, for which there is currently no cure or preventive vaccine.

Seven subtypes of HTLV-1 are currently known, distinguished by nucleotide substitutions in the Env and long terminal repeats (LTR) regions (Gessain and Cassar, 2012; Slattery et al., 1999). The different subtypes are often associated with particular geographic regions, while subtype A (termed the cosmopolitan strain) is found in multiple infection hotspots (Gessain and Cassar, 2012).

HTLV-1 is a member of a large group of deltaretroviruses, Primate T-Lymphotropic viruses (PTLVs, Figure 1.1). So far within this group 4 genotypes of human retroviruses have been identified (Kalyanaraman et al., 1982; Calattini et al., 2005; Wolfe et al., 2005) and 4 genotypes infecting exclusively non-human primates (Simian T-Lymphotropic viruses – STLVs; Slattery et al., 1999). The low sequence diversity between STLV-1 and HTLV-1, indicates a relatively recent common ancestor. Compared to HTLV-1, HTLV-2 has not been associated with a specific disease outcome and is usually carried at a lower viral burden than HTLV-1 (Murphy et al., 2004). The more recently identified HTLV-3 and HTLV-4 are distantly related to either HTLV-1 and HTLV-2, and have so far only been found in a small number of individuals in Central Africa (Mahieux and Gessain, 2011).

The viral burden of HTLV-1 is measured by the proviral load (PVL) which is quantified by the number of proviral copies per 100 peripheral blood mononuclear cells (PBMCs). It has

recently been demonstrated by limited dilution cloning of naturally infected cells that, at least for non-malignant infected cells, there is a single proviral copy per cell (Cook et al., 2012), implying that the PVL measurement is equivalent to the percentage of infected PBMCs in each patient. The PVL is remarkably stable over time in a given patient, but can vary by several orders of magnitude between patients (Demontis et al., 2013; Matsuzaki et al., 2001).

The two main questions in HTLV-1 research are:

1) How does the virus maintain lifelong persistent infection in the face of a strong, constitutive immune response?
2) What determines the risk of HTLV-1 associated disease?

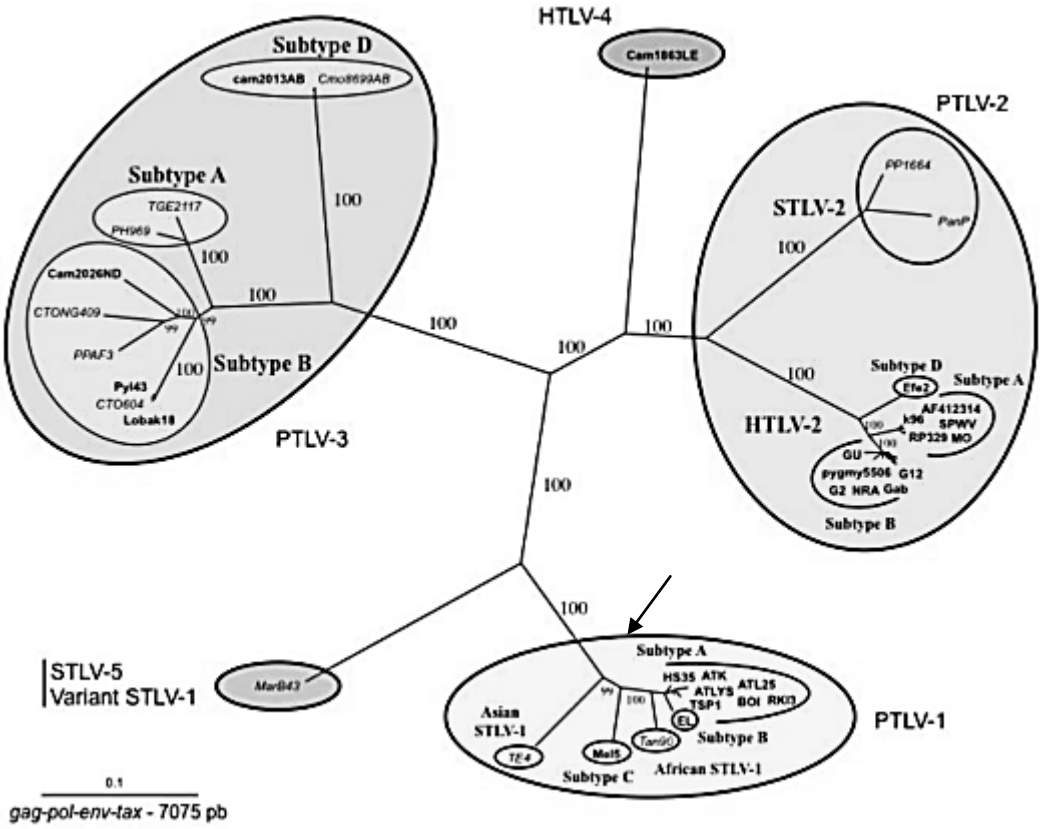

Figure 1.1: HTLV-1 and related viruses.

Phylogenetic tree of HTLV-1 and related viruses, based on nucleotide sequences of viral genes. Primate T-lymphotropic viruses (PTLVs) include both different subtypes of the human T-lymphotropic viruses (HTLVs) and the closely related simian T-lymphotropic viruses (STLVs). The most widespread variant of HTLV-1 is the cosmopolitan strain subtype A (arrow). The distribution of HTLVs and the close sequence homology within each PTLV group suggests multiple zoonotic events. Figure adapted from Mahieux and Gessain (2011).

1.2.HTLV-1 Epidemiology

HTLV-1 prevalence varies greatly world-wide, and appears to be mostly focused in particular endemic hotspots. These include Japan, Africa, the Caribbean islands and South America (Figure 1.2; Proietti et al., 2005). In some endemic regions, such as Kyushu in Japan, seroprevalence can be as high as 8% of blood donors (Maeda et al., 1984).

Transmission of HTLV-1 requires cell-cell contact (Yamamoto et al., 1982b) and cell-free blood products are non-infectious (Jason et al., 1985). HTLV-1 can be transmitted through blood transfusions (Okochi and Sato, 1984) or by sexual contact (Murphy et al., 1989a; Moriuchi and Moriuchi, 2004). Transmission through breast-feeding from mother to child is frequent (Hino et al., 1997), with an increased risk of transmission with increased duration of breast-feeding (Ureta-Vidal et al., 1999). Due to the known methods of virus transmission, in many areas seroprevalence data comes from routine screening of pregnant women and blood donations (Brant et al., 2011; Gessain and Cassar, 2012).

The worldwide seroprevalence rates of HTLV-1 infection are difficult to estimate. Most recent estimate predicts up to 10 million individuals infected in known HTLV-endemic areas. These, however, are thought to be underestimated in some highly populated areas such as China, India and East Africa (Gessain and Cassar, 2012).

Seroprevalence in England and Wales is estimated at about 79 cases per million blood donations (HPA - Health protection agency. See - http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/HTLV/Epidemiology/) and 470 cases per million pregnant women (Taylor et al., 2005). The majority of HTLV-1 infected individuals in the UK originate from HTLV-1-endemic countries (HPA; Tosswill et al., 2000).

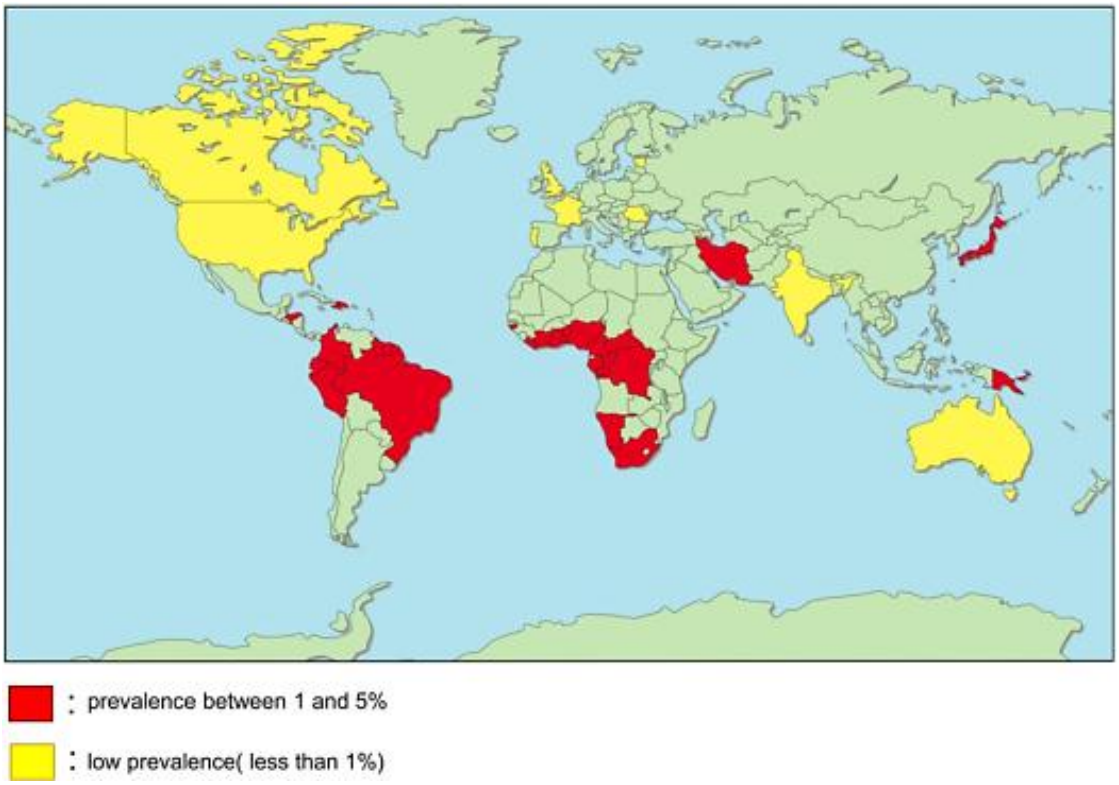

Figure 1.2: HTLV-1 world-wide prevalence.

Map shows countries with endemic HTLV-1, or with a low prevalence of HTLV-1.

Figure from Watanabe, 2011.

1.3.Cellular and Molecular biology of HTLV-1

While HTLV-1 has been shown in vitro to infect multiple cell types (see below), the infection in vivo appears to be restricted to the $CD4^+$ T-cells (the main reservoir of HTLV-1) and $CD8^+$ cells (Richardson et al., 1990; Nagai et al., 2001a).

1.3.1. HTLV-1 receptor

While HTLV-1 predominantly infects $CD4^+$ T cells in vivo it does not use the CD4 molecule as an attachment receptor. HTLV-1 is able to attach in vitro to a variety of cell types (Yamamoto et al., 1984). HTLV-1 or HTLV-1-pseudotyped virus has been shown to infect many different cell types, including T cells, B cells, monocytes (Koyanagi et al., 1993) dendritic cells (Jones et al., 2008), epithelial cells (Setoyama et al., 1998) and astrocytes (Lehky et al., 1995), suggesting that the receptor is present on various cell types. Over the last few years three molecules have been identified as playing an important role in HTLV-1 entry: Glucose transporter 1(GLUT-1; Manel et al., 2003; Jin et al., 2006), Heparan Sulfate Proteoglycans (HSPGs; Pinon et al., 2003; Jones et al., 2005) and Neuropilin-1 (NRP-1; Ghez et al., 2006; Jin et al., 2010).

The interplay between the three molecules involved in HTLV-1 cell entry is not yet fully understood, however current models suggest sequential binding of HSPGs, NRP1 and GLUT-1 leading to the necessary conformation change of Env and viral entry (reviewed in Ghez et al., 2010).

1.3.2. Molecular events leading to proviral integration

Like other viruses, a retrovirus must identify and enter its host cell, be transported to its target intracellular compartment (the nucleus in the case of retroviruses), initiate the expression of its own genes and replication of its genome, followed by the assembly of nascent virions and release from the infected cell. The retrovirus replication cycle also contains certain steps that are characteristic of retroviruses – reverse transcription of its single-stranded RNA genome to double-stranded DNA, and integration of the double-stranded DNA molecule into the genome of the host cell. The retroviral replication cycle is a complex process that involves several viral proteins and recruitment of host factors (reviewed in Goff, 2007).

Viral entry and uncoating

The specific mechanisms identified that HTLV-1 uses in order to infect $CD4^+$ T-cells are detailed in section 1.5.2. HTLV-1 virions are effectively 100% cell-associated and are transmitted through cell-to-cell contact via the virological synapse (Igakura et al., 2003)

The entry of retroviruses to a target cell requires the viral attachment and fusion protein Env, as well as a host cell receptor. Env is processed prior to budding from an infected cell to generate a trimeric complex of two domains, an extracellular surface domain (SU) largely responsible for the binding to the receptor and a transmembrane domain (TM) (Moulard and Decroly, 2000; DeLarco and Todaro, 1976; Einfeld, 1996). HTLV-1 Env is a 62 kDa protein (Pique et al., 1992, reviewed in Jones et al., 2011). The host cell receptor varies between different retroviruses and is one of the chief factors that determine host-cell tropism. The best described retroviral receptors are those for HIV-1: attachment to the primary receptor CD4 is required to enable the subsequent conformational changes and binding to a co-

receptor, which is either the chemokine receptor CCR5 or CXCR4 (Dalgleish et al., 1984; Permanyer et al., 2010).

Viral entry to a host cell requires the fusion of the retroviral and cell membranes following the binding of the retroviral SU protein to the receptor molecules. Two modes of entry have been observed in retroviruses: while most retroviruses are able to mediate entry by direct fusion of the viral and host cell membranes in a pH-independent manner, certain retroviruses require a low pH for efficient entry and release into the host cell cytoplasm, via endocytosis and subsequent release of the viral core from the acidified endosome (McClure et al., 1990; Yamada and Ohnishi, 1986; Barnard et al., 2004).

Trafficking within the cell and reverse transcription

Retroviruses such as HTLV-1 are approximately 2 orders of magnitude smaller than their target lymphocytes. In order to complete their replication cycle they must be transported within the cell to reach the nucleus. HIV has been shown to travel towards the nucleus using both the actin and microtubule components of the host cell cytoskeleton (Arhel et al., 2006; McDonald et al., 2002; Naghavi et al., 2007). HTLV-1 also takes advantage of host cytoskeleton: HTLV-1 infection via the virological synapse requires polarization of the microtubule organizing center (MTOC, see section 1.5.2) towards the site of cell-cell contact (Igakura et al., 2003); this polarization is triggered by HTLV-1 Tax protein (Nejmeddine et al., 2005). The formation of the virological synapse has since been observed also in HIV-1 between $CD4^+$ T cells and between dendritic cells and T cells (Jolly et al., 2004; McDonald et al., 2003).

Reverse transcription, a hallmark of retroviral infection, is the multistep process in which the virus transcribes its single-stranded RNA genome to a linear, double-stranded DNA molecule (Gilboa et al., 1979). This process is mediated by the retroviral RT protein, a product of the *pol* gene (Hill et al., 2005), and contains several distinct steps. DNA synthesis during reverse transcription is primed by a tRNA primer binding to a unique primer binding site in the retroviral genome; HTLV-1 utilizes $tRNA^{pro}$ as the primer for its reverse transcription. The RNA present in the RNA:DNA duplexes formed during this process is degraded by RNase H activity of the RT protein. Two strand-transfer or translocation ('jump') steps during this process result in duplication of the U3, R and U5 domains, generating two direct repeats – the long terminal repeats (LTR). The double-stranded DNA proviruses produced by this process remain associated with viral and host proteins, and this protein DNA complex is termed the pre-integration complex (PIC).

Nuclear entry

Productive infection by retroviruses usually requires the target host cell to undergo mitosis, because the PIC can enter the nucleus only after disassembly of the nuclear membrane. A notable exception to this are the lentiviruses (e.g. HIV-1), which interact with the nuclear pore complex (NPC). Interactions between Integrase protein and TPNO3 / importin α have been shown to play a role in HIV-1 nuclear import (Levin et al., 2010). The HIV-1 capsid protein (CA), which is associated with the PIC, must also play a role, as a single point mutation in this protein (N74D) can alter the interaction between the PIC and the NPC (Lee et al., 2010). In addition, nuclear localization signal domains (NLS) are found on other components of the lentiviral PIC including Integrase and matrix proteins (Bouyac-Bertoia et al., 2001; von Schwedler et al., 1994).

Integration

Integration into an infected cell genome is one of the defining features of retroviral infection, although integration by different mechanisms can also take place in certain other viral infections, such as infection with human papilloma virus (HPV) and hepatitis B virus (HBV). Whereas integration in HBV and HPV is non-essential for expression of viral proteins (Neuveut et al., 2010; Hafner et al., 2008), it is essential for the efficient expression and replication of retroviruses. Integration requires the activity of the retroviral protein Integrase (produced from the *pol* gene), and other host factors. Integration can take place at any site in host genome, but each retrovirus has a specific bias to certain types of integration site (Derse et al., 2007; Wu et al., 2003; Mitchell et al., 2004); this bias is partly attributable to host factors that bind the PIC. The best-described integration-targeting factor is lens epithelium-derived growth factor/p75 (LEDGF/p75), which binds directly to HIV-1 integrase (Cherepanov et al., 2003). LEDGF may play a role in entry to the host nucleus (Llano et al., 2004) and it appears to be used by the virus as a chromatin-tethering factor, increasing integration efficiency and biasing integration towards active transcription units (Maertens et al., 2003; Ciuffi et al., 2005; Shun et al., 2007; De Rijck et al., 2010). The viral protein CA may also play a role in integration site selection, as a point mutation in this protein can alter the integration site bias in the genome (Schaller et al., 2011); the mechanism of this is not clear. The process of integration is further detailed in chapter 4.1.

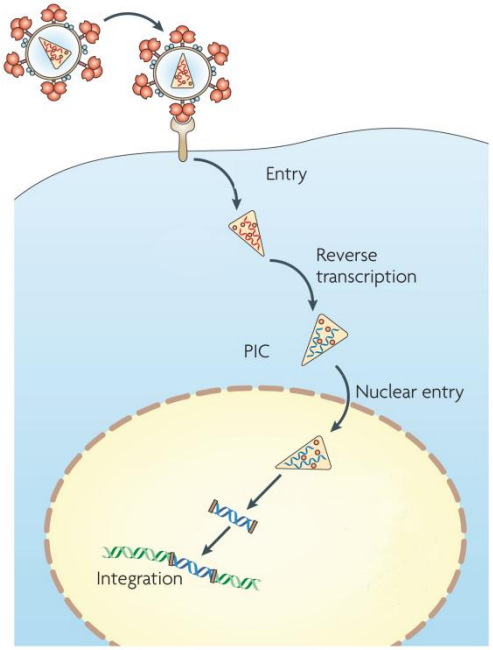

Figure 1.3: Molecular mechanisms leading to retroviral provirus integration

The replication cycle of retroviruses include multiple steps leading up to the integration of the retroviral provirus into the host cell genome. Following entry of the uncoated viral core into the host cell, the single-stranded RNA molecule is reverse transcribed to double-stranded DNA. The protein coated proviral DNA (PIC – pre-integration complex) is then required to enter the host cell nucleus and undergo integration into the host cell DNA. See text for further details. Figure adapted from Goff, 2007.

1.3.3. HTLV-1 gene expression

HTLV-1 carries two RNA copies of its genome in the virion, which are reverse transcribed inside the host cell. This is followed by integration of the double stranded DNA provirus into the host genome, flanked by the LTR on both sides. The HTLV-1 proviral genome is nearly 9 kb long, it encodes the retroviral genes gag, pro, pol and env, as well as the pX region at its 3' end (Figure 1.4; Shuh and Beilke, 2005), which contains non-structural genes (viral accessory and regulatory genes), differentially expressed by alternative splicing. These plus strand transcripts may be unspliced, singly spliced or doubly spliced.

The first gene to be expressed is *tax*. This encodes the viral transactivator Tax, required for expression of other proviral genes (Bex and Gaynor, 1998; Kashanchi and Brady, 2005), as well as a large number of host factors through interaction with transcription factors and chromatin remodellers (Boxus et al., 2008). See Chapter 5.1 for further details, Tax transactivates proviral genes by recruiting host transcription machinery to 3' Tax-responsive elements (TRE) located at the promoter region in the LTR (Brady et al., 1987). Tax is required for activation of other proviral genes and is implicated with many aspects of HTLV-1 mediated host gene deregulation (Ng et al., 2001; Matsuoka and Jeang, 2007).

The regulatory protein Rex, regulates proviral gene expression on a post-transcriptional level. Rex is essential for establishing in vivo infection in the rabbit model (Ye et al., 2003a), and regulates the nuclear export of doubly spliced versus unspliced transcripts (Hidaka et al., 1988). The Rex responsive element is a short sequence present in the U3-R region of the LTR recruiting Rex to the transcript via a stem-loop structure (Seiki et al., 1988; Grone et al., 1994). The resulting Rex-mRNA complex recruits additional host factors which enables it to translocate out of the nucleus (Kashanchi and Brady, 2005). Due to the role of Rex in regulating export of unspliced and singly spliced transcripts, it has been implicated in

determining bi-phasic kinetics of viral gene regulation, where by initially Tax/Rex are expressed, and all other transcripts follow as the levels of Tax and Rex are decreased (Rende et al., 2011; Rende et al., 2012).

The accessory protein p12 is expressed from open reading frame I (ORF I) of the pX. It localizes in the endoplasmic reticulum of the host cell (Ding et al., 2001) and has various functions, including decreased surface expression of MHC-1 (Johnson et al., 2001) and activation of various mitogenic pathways such as activation of STAT-5 (Nicot et al., 2001) which results in reduced dependency on interleukin 2 (IL-2) for cell proliferation.

p13, and p30 accessory proteins are expressed from ORF II of the pX. Though they share part of their sequence, their localization and function differ (Koralnik et al., 1993). Selectively localizing to the inner membrane of the mitochondria (Ciminale et al., 1999), p13 appears to be involved in modulating apoptosis in the infected cell and to sensitize cells to reactive oxygen species (Hiraragi et al., 2005). p30 regulates gene expression by sequestering Tax/Rex mRNA in the nucleus (Nicot et al., 2004) and by binding to the Rex responsive element (Sinha-Datta et al., 2007). In the rabbit model of HTLV-1 infection, translation of p13 and p30 were shown to be required for establishing infection with a high PVL (Bartoe et al., 2000; Silverman et al., 2004), and it has been shown in the nonhuman primate (NHP) model that p12 and p30 are essential in establishing the infection due to their role in sustaining the infection of dendritic cells (DCs; Valeri et al., 2010).

Only one gene is encoded on the reverse strand of the HTLV-1 genome. The recently identified HTLV-1 b-zip factor (HBZ; Larocca et al., 1989; Gaudray et al., 2002), is transcribed from the 3' end of the provirus. HBZ appears to have a role in regulation of Tax mediated transactivation (Gaudray et al., 2002; Basbous et al., 2003) through recruitment of host cofactors including CREB-2, JunB and c-Jun.

HBZ protein is difficult to detect in HTLV-1 infected cells using existing techniques, even in cells expressing a high level of HBZ mRNA (Suemori et al., 2009), although HBZ protein can be readily detected by western blot of lysates from cells transfected with HBZ expression vectors (Suemori et al., 2009). This has raised the question whether HBZ protein is actually translated in vivo. At least one study has demonstrated by altering the HBZ ATG start codon that the effect of HBZ on Tax is exerted at the mRNA level (Satou et al., 2006). In cells infected with an HTLV-1 molecular clone, a large proportion of HBZ mRNA is retained in the nucleus in a Rex-independent manner (Rende et al., 2011). The strongest evidence for the existence of HBZ protein in vivo is the anti-HBZ immune response: approximately 10% of patients carry anti-HBZ antibody responses (Enose-Akahata et al., 2013) and HBZ-specific cytotoxic T lymphocytes (CTLs) are detected in over 30% of infected individuals (Hilburn et al., 2011). Tax-specific CTLs, conversely, are found in all patients (Jacobson et al., 1990; Goon et al., 2004a; Hilburn et al., 2011). Thus, the difficulty in detecting HBZ protein in PBMCs suggests that either protein levels are under the limit of detection of our current methods, or that the export or translation of HBZ is regulated by the virus

.

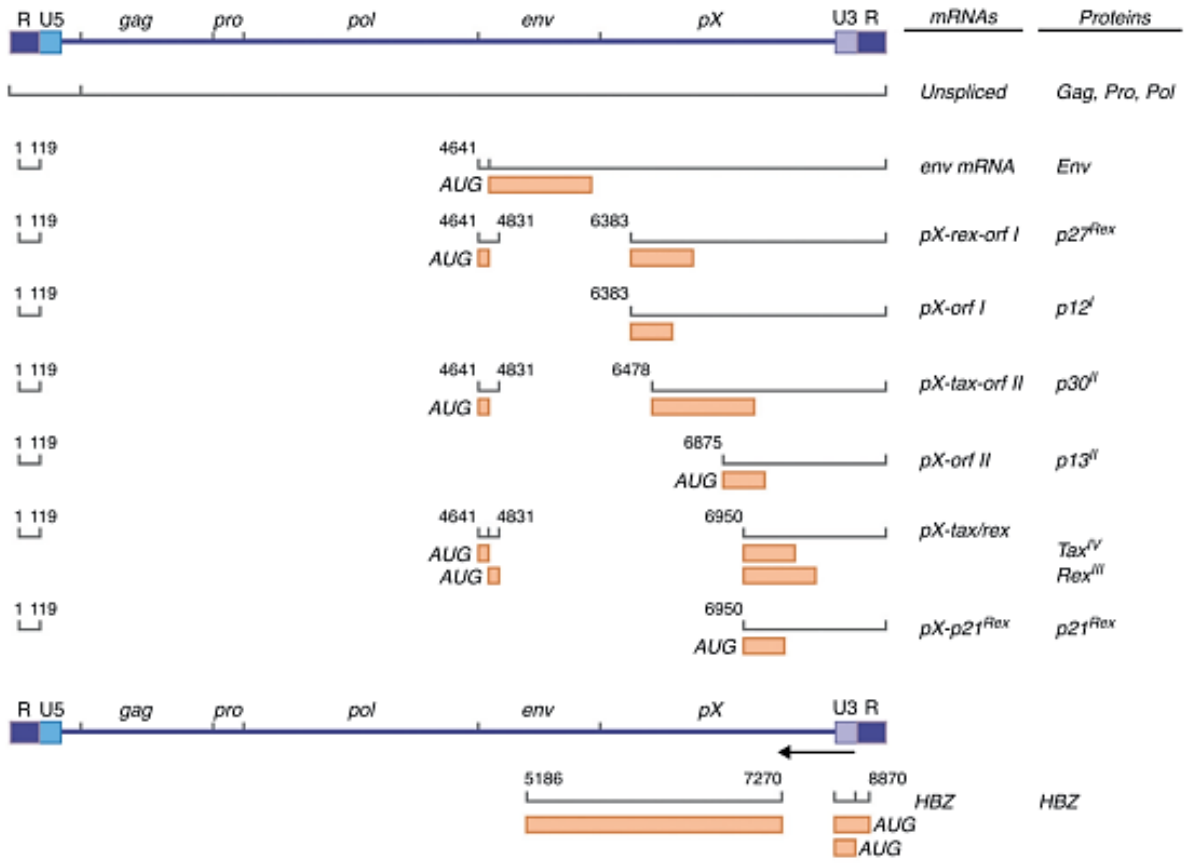

Figure 1.4: Gene expression map of HTLV-1 genome.

HTLV-1 genes are transcribed to either unspliced /spliced mRNAs. Black lines denote mRNA, orange boxes denote protein coding sequence. HTLV-1 sense strand encodes the essential retroviral genes: *gag*, *pro*, *pol* and *env*, as well as additional genes from the pX region. Expression of these genes is initiated from a promoter located on the 5'LTR. The antisense gene *HBZ* is expressed from a promoter at the 3'LTR on the reverse strand. Figure from Lairmore et al., 2012

1.4.HTLV-1 associated clinical outcomes

While over 90% of infected individuals remain life-long asymptomatic carriers (AC), HTLV-1 has been associated with several distinct clinical manifestations, including T cell malignancies, inflammatory diseases, and particular co-infections with other pathogens.

1.4.1. Malignancy

Adult T-cell Leukemia/Lymphoma (ATLL) is a severe malignancy associated with HTLV-1 infection (Gallo et al., 1983). Almost exclusively a $CD4^+$ T cell malignancy, ATLL carries a poor prognosis even on existing treatment (Ratner et al., 2009). ATLL is estimated to affect up to 5% of infected individuals, often in their sixties or older, several decades after infection (Yamaguchi and Watanabe, 2002; Murphy et al., 1989b).

ATLL is classified (Shimoyama, 1991) into four main disease subtypes: Smouldering, chronic, acute and lymphoma type. Diagnosis of ATLL is based on histological evidence of T cell malignancy, the presence of morphologically abnormal cells in the blood ("flower cells") and confirmation of HTLV-1 infection by serology or polymerase chain reaction (PCR). The different subtypes are distinguished by the involvement of different tissues, in particular skin (cutaneous ATLL), lymphocyte count and percentage and different blood markers such as calcium levels.

Acute and lymphoma types carry the most severe prognosis, and median survival stands at 13 months from diagnosis (Ratner et al., 2009). The chronic and smouldering subtypes carry a better prognosis, however both can develop into acute ATLL (Nicot, 2005).

1.4.2. Inflammatory disease

HTLV-1 has been associated with a number of inflammatory diseases. These include uveitis, an inflammation of the eyes (Mochizuki et al., 1992); polymyositis, an inflammatory myopathy (Morgan et al., 1989); infectious dermatitis, an inflammation manifested as skin lesions which usually affects children in HTLV-1-endemic countries (LaGrenade et al., 1990), and arthritis (Nishioka et al., 1989). HTLV-1 has been associated with some neurological manifestations of disease such as polyneuropathy, motor neuron disease and cognitive deficiencies, associated with lymphocyte infiltration of the central nervous system (CNS; Araujo and Silva, 2006). Infected individuals can develop more than one manifestation of HTLV-1 related disease (Nakagawa et al., 1995).

The most common and best studied HTLV-1-associated inflammatory disease is HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP). This is an inflammation of the CNS affecting up to 5% of infected individuals (Gessain et al., 1985). It can affect HTLV-1-infected individuals decades after infection or rapidly (Nakagawa et al., 1995). It is characterized by a T-lymphocyte infiltration through the blood-brain-barrier to the spinal cord, and extensive damage of the axons and myelin, in particular in the lower lumbar region (Yoshioka et al., 1993). Diagnostic criteria for HAM/TSP were determined by the World Health Organization (WHO, 1989). Sixty percent of patients report weakness of the legs as the primary symptom (Araujo et al., 1998), although initial symptoms also include bladder dysfunction or weakness, back pain or sensory symptoms of the legs (Araujo and Silva, 2006). Confirmation of the disease is made by the presence of anti-HTLV-1 antibodies or HTLV-1 antigens in the cerebrospinal fluid (CSF) by ELISA. Abnormal lymphocytes, high white blood cell (WBC) counts or high protein levels may also be found in the CSF.

HAM/TSP is a chronic illness which can progress slowly or acutely and there is no prognostic marker to indicate which individuals will require wheelchair. The severity of disease is measured based on motor disability such as the Kurtzke disability status scale (Kurtzke, 1955)

HAM/TSP currently has no effective cure. Current therapies focus on managing the related symptoms such as spasticity and pain. Recently, attempts have been made to reduce the proviral load (using antiretroviral therapy or by modulating viral expression by histone deacetylase (HDAC) inhibitors) or to modulate lymphocyte proliferation or cytokine production (by the use of interferon or ciclosporin) (Araujo et al., 2008).

1.4.3. Coinfection

HTLV-1 is also associated with an exacerbated form of certain co-infections. The more severe, disseminated form of the helminth *Strongyloides stercoralis* (St) is associated with HTLV-1 infection (Marcos et al., 2008) and there is evidence that the co-infection predisposes to the development of ATLL (Nakada et al., 1987). HTLV-1 carriage is also associated with *Mycobacterium tuberculosis* infection (de Lourdes Bastos et al., 2009) and with more severe disease (Pedral-Sampaio et al., 1997). Co-infection with HTLV-1 and the human immunodeficiency virus type 1(HIV-1) is often associated with a relatively high $CD4^+$ T cell count and symptoms resembling HAM/TSP (Beilke et al., 2005).

1.4.4. Mechanisms of HTLV-1-mediated pathogenesis

It is still not clear what is the mechanism of pathogenesis in HTLV-1-associated disease.

Imaging of the spine of HAM/TSP patients, as well as Investigations of the cell infiltrates in the CSF during various stages of the disease have led to the following three models of pathogenesis (Hollsberg, 1997; Araujo and Silva, 2006):

1) The direct toxicity of T cells – HTLV-1-specific CTL killing of infected glial cells (Lehky et al., 1995).
2) The autoimmunity (mimicry) model – suggests cross reactivity between viral and neuronal antigens. HTLV-1-specific antibodies or CTLs are activated by exposure to viral antigens, travel to the CNS where they cross-react to a "self" antigen presented on glial cells (Nagai et al., 1996), such as the neuronal heterogenous nuclear ribonucleic protein-A1 (Levin et al., 2002)
3) Bystander model – suggests that the immune response within the CNS against infected cells triggers the release of inflammatory cytokines such as TNFα leading to damage of nearby cells (Ijichi et al., 1993; Daenke and Bangham, 1994).

It is poorly understood what are the mechanisms enabling or driving an infected clone to transform into a malignant one, however it is likely to be a combination of both viral and host factors. While Tax protein has been implicated in cell transformation both in vitro (Grassmann et al., 1989) and in vivo (Hasegawa et al., 2006), it is not thought to be necessary for maintenance of the malignant clone: proviruses from the majority of ATLL patients lose Tax expression, either through large deletions of the provirus (Tamiya et al., 1996) - some of which appear to have occurred prior to integration (Miyazaki et al., 2007), through point mutations (both missense or nonsense) in the *tax* gene (Furukawa et al., 2001) or through DNA methylation at the 5' LTR preventing Tax-response element-mediated expression of viral genes (Taniguchi et al., 2005). While Tax mRNA is often absent in ATLL cases, HBZ is

consistently found in ATLL (as well as other infected individuals) (Satou et al., 2006), suggesting that HBZ is necessary for maintenance of the infected clone.

1.4.5. Animal models of HTLV-1 infection and HTLV-1-associated disease

Owing to the rare and insidious nature of HTLV-1 infection, it is necessary to use animal models in order to study the early steps in infection. One of the most common models for the study of early infection and virus dissemination is the rabbit, which can be efficiently infected to a high PVL (Miyoshi et al., 1985, Haynes et al., 2010). However, a recent study comparing the rabbit and NHP models of infection showed that when examining the role of particular proviral genes in establishing infection, the different models do not provide the same results (Valeri et al., 2010). HTLV-1-infected rabbits do not develop HTLV-1 associated disease (Dodon et al., 2012).

Transgenic mouse models are often used to study tumour formation driven by HTLV-1 genes (Hasegawa et al., 2006; Satou et al., 2011). A more recently developed mouse model is the humanized mice, where the human hematopoietic stem cells are used to reconstitute a variety of hematopoietic cells in an immune deficient mouse (Van Duyne et al., 2009), providing a more physiological context for establishment of virus infection in vivo. Two approaches for this have been used – the hematopoietic cells were either infected in vitro prior to engraftment (Banerjee et al., 2010) or a humanized mouse was infected by intraperitoneal inoculation with irradiated HTLV-1-producing cells (Villaudy et al., 2011). The humanized mouse model allows for the study of the early events in infection, as well as development of tumours in the context of human T cells. However, an important drawback with currently used models is the lack of a functional immune response which is predicted to play a major

role in shaping the virus dissemination in human HTLV-1 infection. A recently developed HLA-A2-expressing mouse strain was developed that is able to present antigens to the CTL response. This model is currently being tested for use in HTLV-1 infection research (Dodon et al., 2012).

HTLV-1-related viruses in their natural hosts can also be used as models of HTLV-1 biology in the human. Bovine Leukemia Virus (BLV; Miller et al., 1969) is a B-lymphotropic deltaretrovirus. Natural infection of its bovine host can either remain asymptomatic or cause enzootic bovine leukemia, a B-cell malignant disease manifested as lymphoid tumours. Transmission of the virus appears similar to HTLV-1, through BLV-infected cells present in body fluids such as blood and milk. BLV infection of cattle is used as a model for tumorigenesis, and for the testing of potential therapeutic or vaccine approaches in the context of an effective immune response (Rodriguez et al., 2011). In a similar manner, old-world primates such as baboons, naturally infected with STLV-1, can be used to test potential therapeutics (Afonso et al., 2010).

While these animal models are potentially useful for the study of HTLV-1 molecular biology and proviral load, there is no good model for HTLV-1-induced HAM/TSP, and existing animal models either require using a different, related virus, or they fail to replicate the delicate balance between virus-driven replication and immune selection of infected T-cell clones.

1.4.6. Host and viral determinants of disease

So far the main predictor of inflammatory and malignant disease is the proviral load. The median PVL of asymptomatic carriers is at about 0.3%. The median PVL exceeds 5% for HAM/TSP patients and can exceed 100% in ATLL cases (Nagai et al., 1998). The prevalence of HAM/TSP rises exponentially with the PVL, in loads above 1%, and in a 10 year followup study of 64 HAM/TSP patients (Matsuzaki et al., 2001) a higher proviral load was associated with a faster disease progression, suggesting a role for proviral load in pathogenesis. A high proviral load is also considered to be a prognostic factor in ATLL (Iwanaga et al., 2010), as well as blood markers such as high lactic dehydrogenase and hypercalcemia (Tsukasaki et al., 2009).

Since the proviral load is such an important predictor of pathogenesis and disease progression, efforts have focused on improving our understanding of what determines the proviral load, and in seeking therapeutic approaches for reducing the proviral load.

Expression of Tax in $CD4^+$ T-cells may play a role in disease: at a given proviral load, HAM/TSP patients have a higher proportion of Tax^+CD4^+ T cells than asymptomatic HTLV-1 carriers (Asquith et al., 2005b).

HTLV-1-specific CTLs can kill Tax-expressing cells in vitro (Asquith et al., 2005a; Hanon et al., 2000a). The viral suppression assay (Asquith et al., 2005a) is a flow-cytometry based assay which quantifies the efficiency (Tax cells killed per day per $CD8^+$ cell) of the CTL response to Tax. Fresh unstimulated PBMCs are incubated overnight in the presence of different numbers of autologous $CD8^+$ T cells, and surviving Tax-expressing cells are quantified by flow cytometry. The strength of this assay is that naturally-processed viral antigens are presented to autologous $CD8^+$ T-cells, with minimal in vitro manipulation. The viral suppression efficiency has been found to be inversely correlated with the proviral load

and with proportion of Tax-expressing cells (Kattan et al., 2009); however, at a given viral suppression efficiency, the proviral load was still higher in HAM/TSP patients than in ACs (Asquith et al., 2005a). Tax expression, therefore, varies between individuals, but also within individuals. This variable level of Tax expression may have a direct impact on HTLV-1 dynamics, since cells that express Tax after overnight ex vivo culture turn over faster in vivo (Asquith et al., 2007) and $CD4^+$ T cells that express a greater amount of Tax are killed more efficiently in vitro by autologous CTLs (Kattan et al., 2009).

It is not known what determines the rate, amount and timing of Tax expression by an individual cell.

The role of host genetics

The risk of developing HTLV-1-related disease varies between different endemic countries. HAM/TSP is more frequent in Africa and the Caribbean than in Japan (Maloney et al., 1998; Kaplan et al., 1990). The incidence of disease cannot be explained by viral sequence (Daenke et al., 1990; Niewiesk et al., 1994). This suggests that the genetic background could play a role in determining disease risk. In a genetic association study carried out on an ethnically homogenous group in the Kagoshima prefecture of Kyushu, Japan, more than 40 candidate genes were examined. HLA-A*02 and HLA-Cw*08 were found to be protective (associated with a lower load in ACs and with remaining asymptomatic) and HLA-B*54 was found to be associated with increased risk of HAM/TSP (Jeffery et al., 1999; Jeffery et al., 2000; Vine et al., 2002). In particular, HLA-I alleles which are predicted to strongly bind HBZ are associated with a better outcome of HTLV-1 infection (Macnamara et al., 2010). More recently, further analysis of the same cohort has demonstrated that the protective effect of HLA-Cw*08 and the detrimental effect of HLA-B*54 was limited to those individuals which

also carry a particular killer cell immunoglobulin-like receptor (KIR) allele, KIR2DL2 (Seich Al Basatena et al., 2011). Class II alelles HLA-DRB1*0101 has been found to be associated with increased risk of HAM/TSP (Kitze et al., 1998). The 191C allele of IL-15 has been found to be associated with reduced proviral load (Vine et al., 2002). This is of particular interest because IL-15, which is upregulated by Tax, is thought to increase survival of $CD8^+$ T cells in HTLV-1 infection (Azimi et al., 2001).

While the genetic polymorphisms give insight into mechanisms of HTLV-1 infection (such as the role of MHC genes in controlling the proviral load), these factors only explain less than 10% of the observed variation in proviral load. (Vine et al., 2002; Bangham, 2008)

Gender appears to play a part in infection and disease. HTLV-1 is transmitted from male to female with higher frequency than female to male (Murphy et al., 1989a), and it is suspected that girls are at higher risk of infection through breast-feeding compared with boys (Ureta-Vidal et al., 1999). As a result, more women than men are infected with HTLV-1. Infected men have an increased lifetime risk of ATLL (incidence in Japan – 3-5% in males compared to 1-2% of females; Yamaguchi and Watanabe, 2002), while women have a higher risk of HAM/TSP (Maloney et al., 1998). Deterioration of HAM/TSP symptoms appears to be more rapid in women with disease onset pre-menopause, which suggests a role for hormones in disease progression (Lima et al., 2005).

1.5. Dynamics of HTLV-1 infection

HTLV-1 causes a persistent life-long infection in infected individuals. This persistent infection remains at a constant level despite a strong specific constitutively active immune response, and can lead to severe or fatal disease. It is still not clear how the virus regulates its expression and replication in order to escape this immune control and maintain a constant viral burden. The proviral load, in turn, is our best prognostic marker for HTLV-1 associated disease. This suggests that in order to better understand and monitor HTLV-1 dynamics and pathogenesis, we must take a closer look at the proviral load itself (Figure 1.6), and what determines the proviral load.

1.5.1. The immune response to HTLV-1

Innate immunity

The interferon pathway appears to play a role in infection and pathogenesis. In vitro infection of pDCs appears to induce an interferon-α (IFNα) response, mediated by the Toll-like receptor (TLR) 7 (Colisson et al., 2010), which is known to respond to single stranded RNA (Kawai and Akira, 2011). HTLV-1, in response, appears to impair the IFN response in infected pDC as these show decreased IFN production ex vivo (Hishizawa et al., 2004). It is not clear, however, whether the IFN response is beneficial in HTLV-1 infection. Tattermusch and colleagues have recently shown that interferon-stimulated genes (including genes that are inducible by type 1 or type 2 IFNs) are upregulated in HAM/TSP compared with high-load ACs (Tattermusch et al., 2012) and this upregulation was found to be correlated with disease severity. Furthermore this study has shown that neither Tax nor HBZ mRNA were downregulated by IFNα treatment of infected cells. While IFNα treatment has shown some

promise in ATLL when used in conjunction with AZT (azidothymidine; Hodson et al., 2011; Kchour et al., 2007), it is not been shown to be very effective in the treatment of HAM/TSP (Arimura et al., 2007).

Adaptive immunity

HTLV-1 infection elicits a strong antibody response. In the serum of HTLV-1infected individuals antibodies against Gag, Env, Tax, Rex and HBZ can be found (Lal et al., 1994; Burbelo et al., 2008; Enose-Akahata et al., 2013). While in the sera of the majority of patients antibodies against Tax were found (de Souza et al., 2011), anti-HBZ antibodies are only found in about 10% of cases (Enose-Akahata et al., 2013).

There is typically a higher titre of anti-Env and anti-Tax antibodies in HAM/TSP patients than in ACs, but a positive correlation between the proviral load and antibody titre is not always found (Kira et al., 1992; de Souza et al., 2011; Burbelo et al., 2008). Anti-HTLV-1 antibody titres vary between patients but appear to be stable in a patient over time (Burbelo et al., 2008).

It is not clear what role, if any, anti-HTLV-1 antibodies may have in controlling the infection: because of the highly cell-associated nature of the virus, it is thought that antibodies have limited impact on the infection.

The arm of the immune system that has the most significant impact on HTLV-1 infection is the HTLV-1-specific CTL response. The association of particular MHC-I alleles with protection from HTLV-1-associated disease and low proviral load (Bangham, 2008), as well as the existence of escape mutations in known CTL epitopes (Niewiesk et al., 1995) suggest a significant role of the CTL response in exerting selection upon HTLV-1. CTLs recognize

short linear peptides processed through the endogenous pathway and presented on MHC-1 molecules. There is a high frequency of activated anti-HTLV-1 CTLs in the blood of HTLV-1 infected individuals, both patients with HAM/TSP (Jacobson et al., 1990) and asymptomatic carriers (Parker et al., 1992). The immunodominant viral protein is Tax, followed by Gag and Pol (Goon et al., 2004a). In the context of the common HLA-A*0201 allelle, one epitope, Tax11-19, has been shown to elicit a particularly strong CTL response (Elovaara et al., 1993; Nagai et al., 2001b; Kozako et al., 2011). In recent years, more attention has been given to the "quality" of the CTL response (Bangham, 2009), rather than focusing on CTL frequency. There is now strong evidence that the CTL frequency in response to a given antigen concentration is not a useful measure of the effectiveness of the immune response. A better index of CTL efficiency is the functional avidity, defined as the reciprocal value of the peptide concentration at which 50% of specific CTLs respond to antigen (Kattan et al., 2009).

The HBZ protein is not immunodominant (Suemori et al., 2009), however recent evidence has emerged which suggests that the anti-HBZ CTL response has significant clinical importance. MacNamara et al used predictions of binding strength of HTLV-1-derived peptides to the MHC I allelles present in a large HTLV-1 infected individual cohort from Kagoshima, Japan (Jeffery et al., 1999). They found that possession of HLA alleles that are predicted to bind strongly to HBZ peptides is protective in HTLV-1 infection, and the number of strong binding HBZ peptide-binding HLA alleles correlated negatively with HTLV-1 proviral load (Macnamara et al., 2010). Further work has demonstrated that patients with IL-2-secreting HBZ-specific CTLs were more frequent in ACs and individuals with a low PVL (Hilburn et al., 2011).

The CTL response mounted against HTLV-1 may be repressed by Tregs: There is an inverse correlation between CTL lysis efficiency (measured using the viral suppression assay described above) and the percentage of $CD4^{+}FoxP3^{+}Tax^{-}$ cells (Toulza et al., 2008). The high frequency of $FoxP3^{+}$ cells could be explained (Toulza et al., 2010) by Tax-driven upregulation of CCL22, which is a ligand for the CCR4 molecule expressed on $FoxP3^{+}$ cells.

1.5.2. HTLV-1 proliferation

Proliferation of a retrovirus can take place by one of two mechanisms. Mitotic spread is the passive replication of the provirus that accompanies mitosis of infected cells. This will result in expansion of existing clones, in each of which all daughter cells share the same proviral integration site. Infectious spread is the de novo infection of a new cell, generating a new infected clone with a different integration site. Mitotic and infectious spread can also be distinguished by the replication mechanism involved – while mitotic spread utilizes the host cell machinery and DNA polymerase to replicate the DNA, infectious spread uses the error prone viral reverse transcriptase in order to complete the viral replication cycle.

Mechanism of cell-cell transmission for de-novo infection

HTLV-1 is extremely highly cell-associated, and cell-cell contact is required for transmission of the virus (Yamamoto et al., 1982b). The virological synapse (**Error! Reference source not found.**A) is a virus-induced area of cell-cell contact (Igakura et al., 2003) which allows transmission of the virus between infected and target cell, by polarization of the cytoskeleton towards the uninfected cell. This polarization is triggered by the viral gene Tax as well as by stimulation of the intercellular adhesion molecule-1(ICAM-1; Nejmeddine et al., 2005;

Nejmeddine et al., 2009). Tax has been shown to increase the efficiency of HTLV-1 transmission by over 10 fold (Mazurov et al., 2010). The virological synapse is similar to the immunological synapse, however while in the virological synapse the microtubule organizing center (MTOC) is polarized in the HTLV-1 infected cell towards the cell-cell junction (Igakura et al., 2003), triggered by adhesion molecules, the MTOC in the case of the immunological synapse is polarized in the responding T cell towards the antigen presenting cell, triggered by the T cell receptor (Grakoui et al., 1999). The use of electron tomography to visualize the virological synapse demonstrated that the virus particles are present in multiple clefts within the synapse, surrounded by tightly bound plasma membranes forming confined intracellular sites (Majorovits et al., 2008). This is consistent with the observation that cell free virus particles are not found in plasma, and suggests a mechanism of escape from immune surveillance by HTLV-1-specific antibodies and complement (reviewed in Nejmeddine and Bangham, 2010).

While not produced by primary infected lymphocytes, cell free HTLV-1 virions can be recovered from some chronically infected cell lines such as MT2, and these can inefficiently infect other cells in vitro (Fan et al., 1992; Mazurov et al., 2010). Cell free infection of dendritic cells (DCs) has been described (Macatonia et al., 1992; Jones et al., 2008), and these infected DCs were able, in turn, to productively infect primary T cells ex vivo by cell-to-cell contact. It is not clear what role this mechanism may play in vivo (**Error! Reference source not found.**B), however the fact that HTLV-1-specific $CD4^+$ cells are more likely to be infected with HTLV-1 themselves (Goon et al., 2004b), may be explained by presentation of HTLV-1 peptides on infected DCs in the context of MHC-II, allowing preferential infection of HTLV-1 specific Th cells.

Another observation recently reported was the visualization by electron microscopy of extracellular viral clusters on the surface of infected cells in vitro(**Error! Reference source not found.**C), reported to form bio-film like carbohydrate rich assemblies (Pais-Correia et al., 2010) . These are suggested to allow a protective environment for transmission of virus particle to target cells on the surface of both infected and target cell.

It is difficult to quantify the role played by each mechanism in vivo.

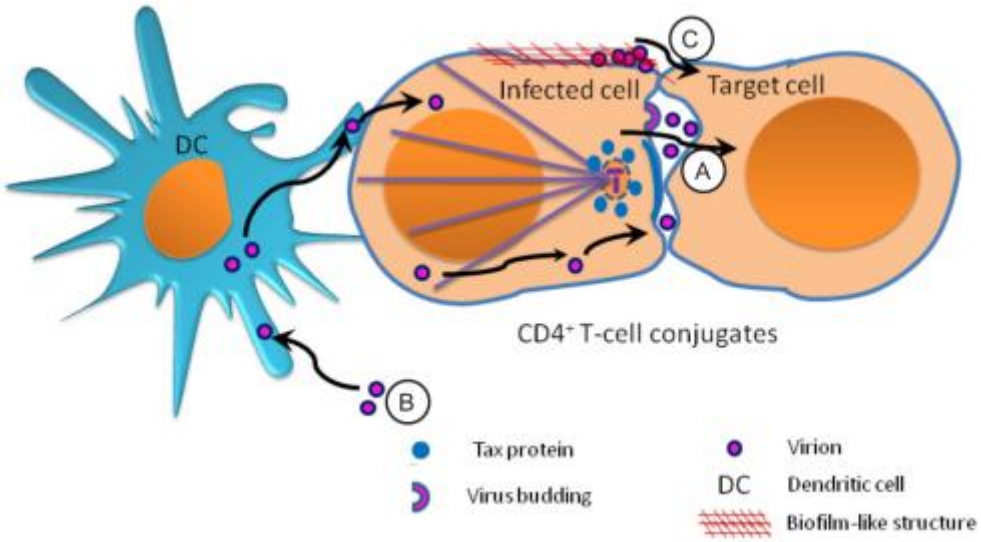

Figure 1.5: Mechanisms of HTLV-1 cell-cell transmission

Three proposed mechanisms of viral transmission to susceptible cell. (A) Viral transmission via the virological synapse, an isolated area of cell-cell contact which allows for the transmission of HTLV-1 particles to a target cell avoiding exposure to the humoral immune response. (B) Cell free infection of dendritic cells leads to cell-cell transmission to susceptible T cells. (C) Virus accumulation at cell surface in biofilm like structures. Figure adapted from Nejmeddine and Bangham, 2010

Mitotic expansion of infected cell clones

It is not clear what determines infected cell proliferation in HTLV-1 infection. HTLV-1 infection has been shown to drive through cell cycle checkpoints (Matsuoka and Jeang, 2005) and regulate apoptosis of infected cells, both could result in a higher proliferation of infected cells. (Kasai and Jeang, 2004). HTLV-1-infected cells turn over faster in vivo (Asquith et al., 2007). Two viral genes have been implicated. As discussed above, Tax has been shown both in vitro (Grassmann et al., 1989) and in vivo (Hasegawa et al., 2006) to be sufficient for efficient cell transformation. More recently HBZ has been shown to be able to promote cell proliferation and to drive cell transformation in a mouse model (Satou et al., 2011).

The ratio of infectious to mitotic spread

Two lines of evidence led to the conclusion that mitotic spread plays a significant role in maintenance of the proviral load: First, integration site analysis, demonstrated the existence of long-lived expanded clones in vivo (Furukawa et al., 1992; Cavrois et al., 1996; Wattel et al., 1995; Etoh et al., 1997; Gillet et al., 2011). Second, the low observed frequency of sequence diversity between HTLV-1 isolates compared with that of other retroviruses (Ina and Gojobori, 1990; Overbaugh and Bangham, 2001) is consistent with the idea that the majority of proviral replication takes place using the proofreading host cell DNA polymerase rather than the error-prone reverse transcriptase.

The ratio of mitotic to infectious spread of HTLV-1 is not currently known. Moreover, there is little known about the role that expression of proviral genes play in determining this ratio. Both HBZ and Tax have been shown to be able to drive cell proliferation, therefore both may

have a role in driving mitotic spread in vivo. Tax protein has also been shown to play a role in driving cell-to-cell transmission of HTLV-1, and indeed it is possible that a central function of Tax is to drive infectious spread (Nejmeddine et al., 2005; Mazurov et al., 2010). At the same time, Tax is a major immunogen, exposing an infected cell clone to the immune response.

The proviral load, therefore, is composed of the sum of all infected cell clones. We postulate two main selection forces determining the clone size: the first is the ability of a cell clone to proliferate, and the second, is its susceptibility to be killed. If these selection forces acted equally across cell clones we would expect to see an equal survival advantage in all clones, and each clone would reach the same abundance in the steady state.

This work focused on developing experimental and bioinformatics methods for the investigation of the clonal distribution of the load, in order to test specific hypotheses regarding the role of the genomic environment and proviral expression in in vivo clonal expansion.

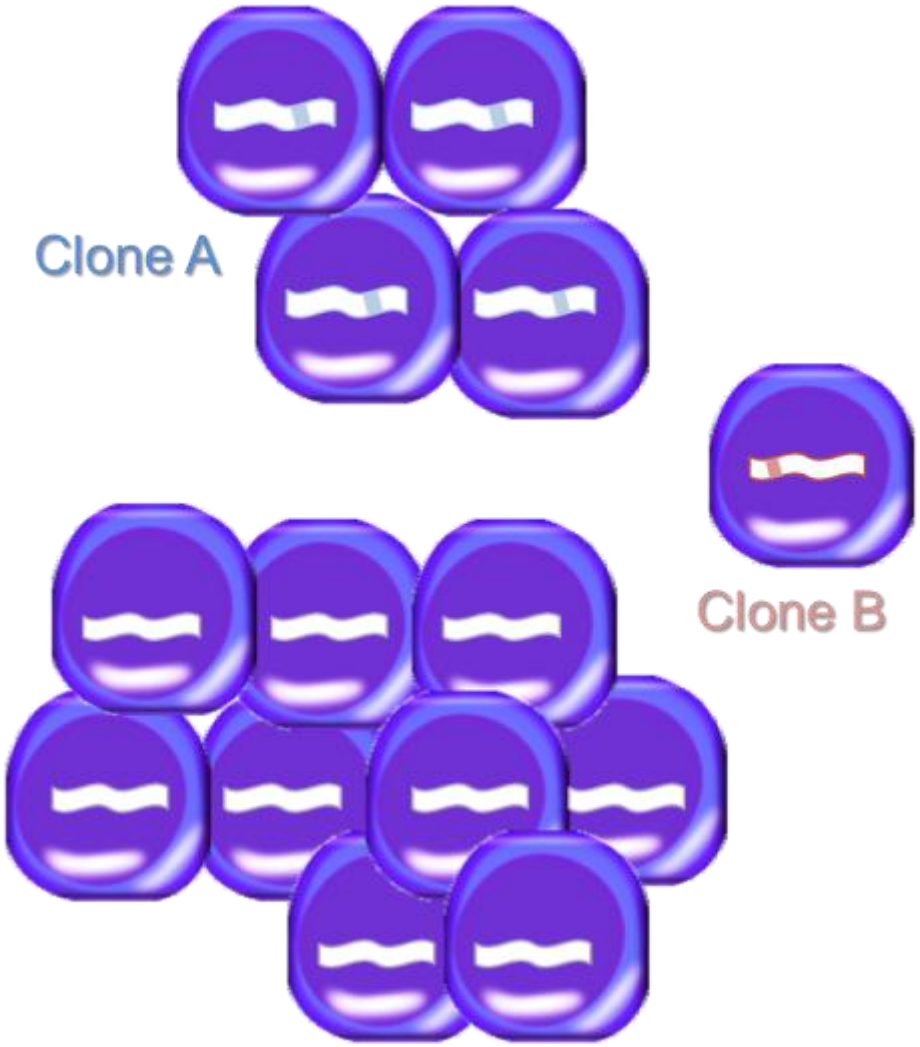

Figure 1.6: The clonal distribution of the proviral load.

Infected cell clones can be distinguished by their unique integration site (UIS). Virus proliferation can take place either by mitotic spread (expansion of infected clones, all daughter cells sharing the same integration site) or by infectious spread (generating a new infected clone with a novel integration site). This takes place on a background of mostly uninfected cells. The proviral load is the sum of all infected T cell clones (% infected PBMCs).

1.6.Aims

The work aims to develop a high-throughput sequencing based method to amplify, map and quantify HTLV-1 proviral integration sites from HTLV-1 infected cells, and then use the method to test the following hypotheses:

1) The genomic environment flanking the proviral integration site is associated with targeting of the provirus to the genome, and with the long-term clonal expansion of infected cells.
2) The genomic environment flanking the integration site is associated with proviral gene expression from the 5'LTR, which in turn has an effect on clonal expansion of infected cells.
3) The infected $CD8^+$ cells shape the infected cell clonal distribution in vivo.

Chapter 2 - **Materials and Methods**

2.1.Primary cells and cell lines

2.1.1. Patient samples

Blood samples were donated by HTLV-1-infected individuals attending the HTLV-1 clinic at the National Centre for Human Retrovirology (Imperial College Healthcare NHS trust) at St Mary's Hospital, London UK, with fully informed written consent. This study was approved by the UK National Research Ethics Service (NRES reference 09/H0606/106). Diagnosis of HAM/TSP was done according to WHO criteria (WHO, 1989).

PBMCs were isolated from blood using Histopaque-1077 (Sigma-Aldrich) and cryo-preserved in fetal bovine serum (FBS; Gibco – Life Technologies) containing 10% dimethyl-sulfoxide (DMSO; Sigma-Aldrich).

2.1.2. Cell lines used

HTLV-1 negative Jurkat cells (Schneider et al., 1977) were used as negative control and for in vitro infection (see below). Tarl2 (Tateno et al., 1984) is an HTLV-1-infected rat leukocyte cell line containing one copy of the HTLV-1 provirus per cell. MT2 (Miyoshi et al., 1980) is an HTLV-1-producing cell line. All cell lines were cultured at containment level 2 (Jurkat) or 3 (MT2 and Tarl2) lab in complete medium: RPMI-1640 medium (Sigma-Aldrich) supplemented with 1% l-Glutamine, 1% penicillin/streptomycin and 10% heat-inactivated FBS. Cells were maintained at a concentration of 10^5-10^6 cells/ml.

2.1.3. In vitro infection of Jurkat cells

Two independent assays of in vitro infection were carried out (one by Dr. Nicolas Gillet, see Table 3.1). Jurkat cells were co-cultured for 3h with γ-irradiated (^{137} Cs, 40,000 cGy) MT2 cells pre-labelled with anti-CD4 MicroBeads (Miltenyi). MT2 cells were then depleted from the co-culture using magnetic separation (Miltenyi), and infected Jurkat cells were maintained in culture for 14 days in complete medium at 37^{0}C with 5% CO2. Genomic DNA was extracted and the proviral integration sites analysed as described below. Integration sites from MT2 were also analysed to exclude possible contamination of the Jurkat integration sites. No contaminating MT2 integration sites were found after 14 days.

2.2.Cell sorting

2.2.1. Tax sorting

See also Figure 2.1. PBMCs from 10 patients with HAM/TSP with a high proviral load (range 12.2-50.6 copies per 100 PBMCs) were depleted of $CD8^+$ cells using anti-CD8 antibody-coated magnetic beads (Miltenyi) and incubated in complete medium for 18 hours at 37^0C with 5% CO_2. After 18h culture, the cells were fixed and permeabilized (eBioscience), and then stained for intracellular expression of Tax (anti-Tax mAb LT4, Tanaka et al 1991), as well as surface expression of CD4 (clone RPA-T4, eBioScience), and CD8 (clone SK1, eBioScience). Live/Dead cell marker (Invitrogen) was used to exclude possible autofluorescence by dead cells. Cells were sorted by Fluorescence-activated cell sorting (FACS; FACSAria IIIU, BD Biosciences) to isolate two populations of live $CD4^+$ cells based on Tax expression.

The flow gating strategy (FACSDiva, BD Biosciences; Figure 2.2A) was as follows: first dead cells and doublets were excluded. The gate was then set on $CD4^+CD8^-$ cells (the vast majority of $CD8^+$ cells had been depleted) and then on Tax-expressing or non-expressing cells. Gates were set to ensure a clear demarcation between the Tax^+ and Tax^- populations.

DNA was extracted from whole unsorted PBMCs from each patient and analysed separately to identify the patient of origin of each clone; 46% of the clones were attributed in this way.

Recovered cells from all 10 patients were combined in two pools, respectively $CD4^+Tax^+$ cells and $CD4^+Tax^-$ cells. The DNA from each pool was extracted, and integration sites content analysed (see 2.4 below).

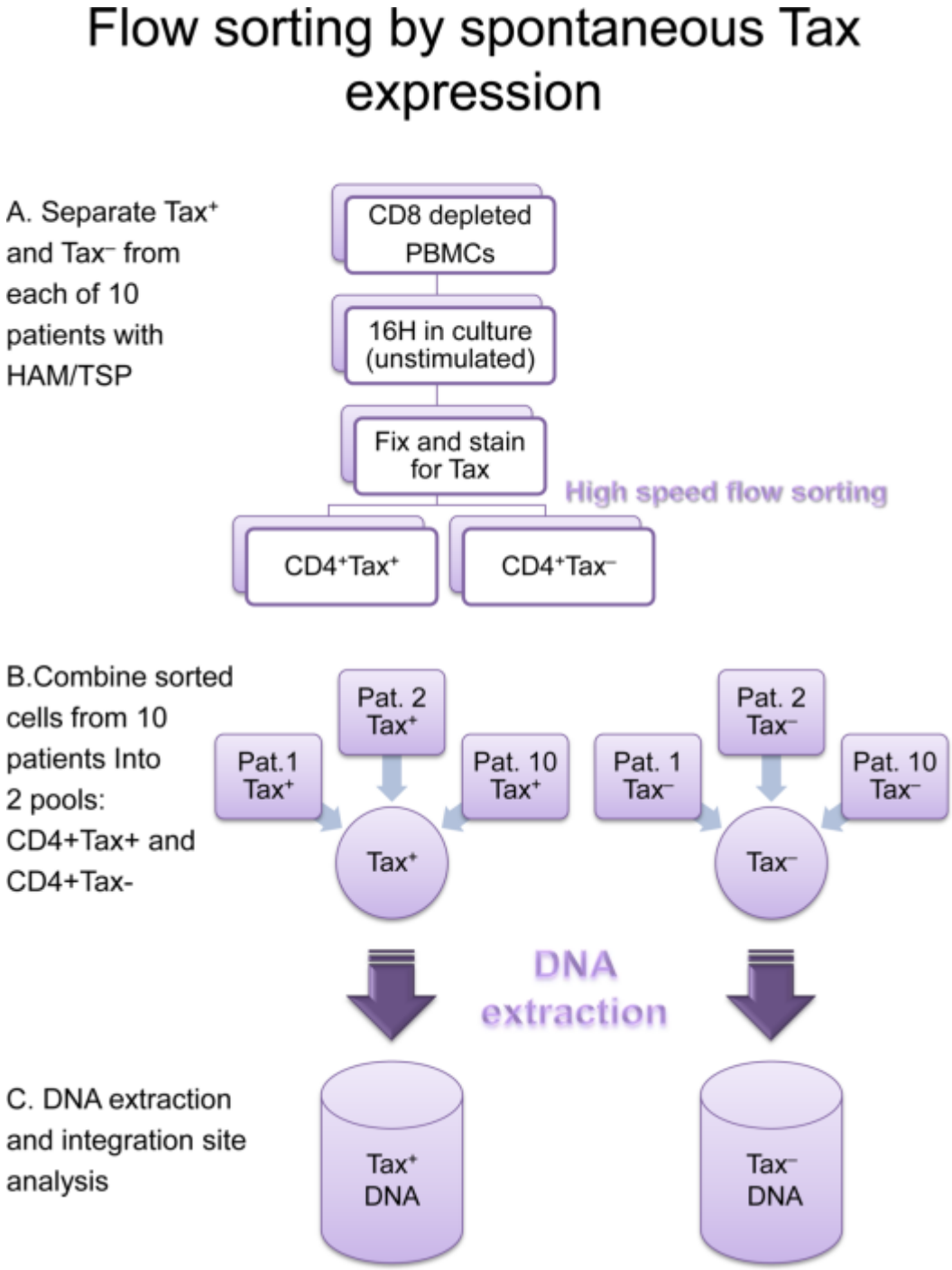

Figure 2.1: Protocol for flow-sorting of Tax-expressing cells.

(A) $CD8^+$ depleted PBMCs from 10 HAM/TSP patients were studied. The cells were incubated overnight, and then sorted for Tax expressing or non-expressing cells by FACS (Figure 2.2). (B) Recovered cells from all 10 patients were combined into two cell pools of $CD4^+Tax^+$ cells and $CD4^+Tax^-$ cells. (C) Genomic DNA was extracted from each cell pool separately, and integration site analysis was carried out.

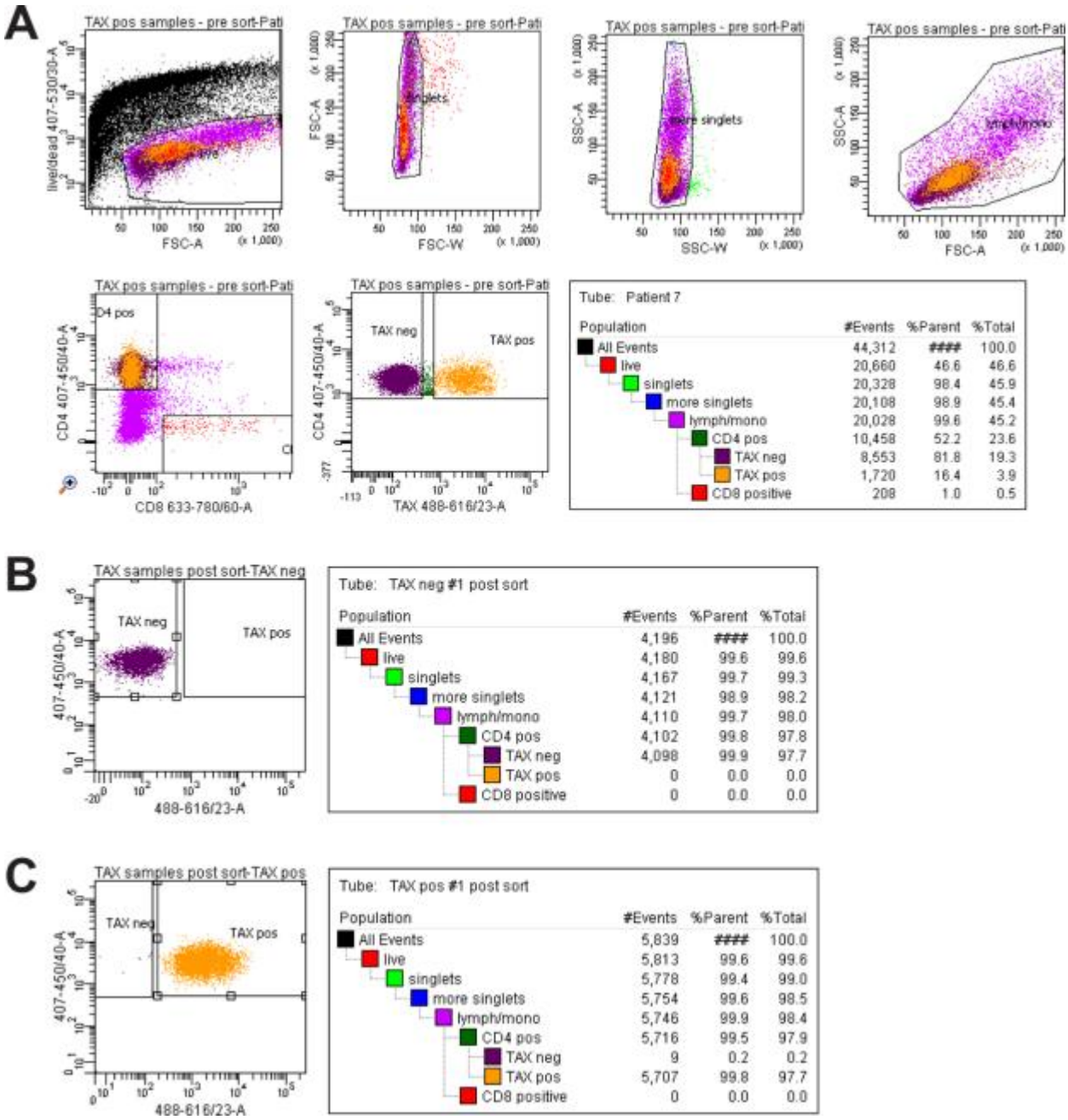

Figure 2.2: Flow cytometry sorting by Tax expression.

(A) Representative FACS plots of the gating procedure used for sorting $CD4^{+}Tax^{+}$ ('TAX pos') and $CD4^{+}Tax^{-}$ ('TAX neg') cells. Gate setting was used to unequivocally distinguish between negative and positive populations. (B) Purity testing of Tax-sorted cells, Tax^{+} cells not detected. (C) Purity testing of Tax^{+} sorted cells, 0.2% were Tax^{-}.

2.2.2. CD4+ and CD8+ T-cell sorting

See also Figure 2.3. Uncultured, cryopreserved PBMCs from 12 HTLV-1 infected individuals (6 patients with HAM/TSP and 6 asymptomatic carriers) were sorted for $CD4^+$ and $CD8^+$ expression, using positive selection magnetic sorting (Miltenyi). In order to maximize purity and recovery, cells were sorted 4 times (repeating each selection), first positively selecting $CD4^+$ cells, and then positively selecting $CD8^+$ cells from the $CD4^-$ fraction.

Unsorted PBMCs, as well as each sorted fraction was stained for CD3 (clone UCHT1, eBioscience, CD4 (clone RPA-T4, eBioscience) and CD8 expression (clone SFCI21Thy2D3, Beckman Coulter). FACS (CyAn™, Beckman Coulter) was used to analyse the purity of the resulting populations and to quantify the cell populations in the unsorted samples.

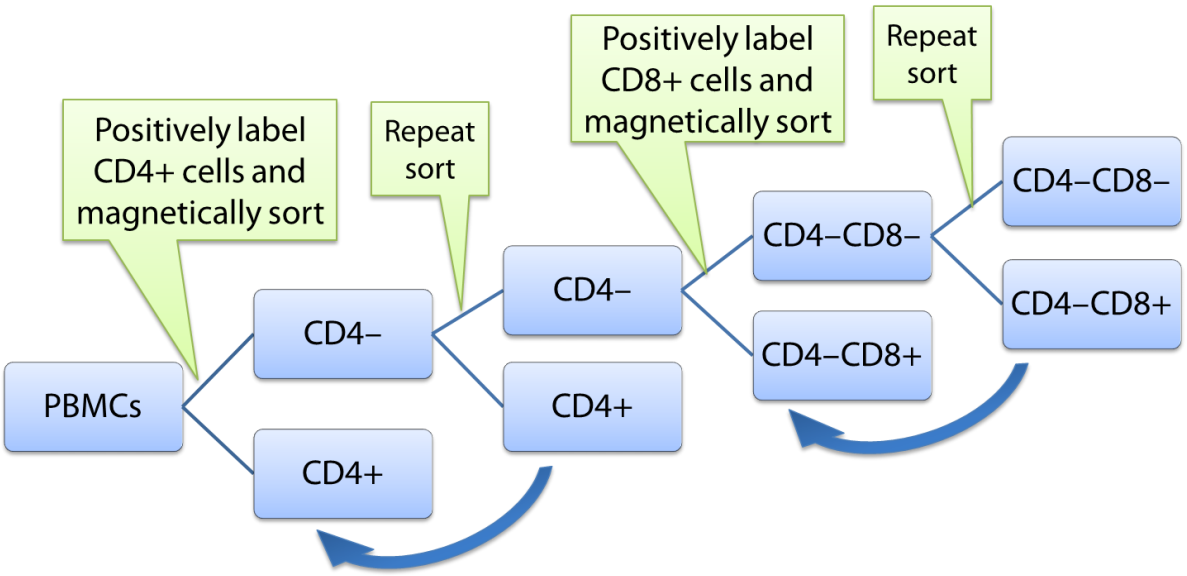

Figure 2.3: Cell sorting approach for the purification of $CD8^+$ and $CD4^+$ cell populations.

2.3.Molecular biology methods

2.3.1. DNA extraction

DNA was extracted from primary cells and cell lines using the DNeasy Blood & Tissue kit (Qiagen) according to manufacturer's instructions and eluted and stored in EB buffer (Qiagen).

DNA concentration was quantified using Nanodrop 1000 (Thermo Scientific).

2.3.2. Proviral load measurements

All primers described in this section are detailed in Appendix I. The proviral load was measured by quantitative PCR (QPCR) method using Light cycler®2.0 (Roche). PCR was carried out using the LightCycler® FastStart DNA Master^PLUS reagent (Roche) according to manufacturer's instructions. Proviral copies were quantified using Tax specific primers (SK43, SK44, Kwok et al., 1988) and normalized to number of cells using β-Actin specific primers (Actin-fw, Actin-rev). Tax amplification was done over 45 cycles of 10 seconds at 95^{o}C, 5 seconds at 58^{o}C, 8 seconds at 72^{o}C and 10 seconds at 87^{o}C. Amplification of actin was done over 45 cycles of 10 seconds at 95^{o}C, 5 seconds at 60^{o}C, 12 seconds at 72^{o}C and 2 seconds at 85^{o}C.

Standard curves were generated from Tarl2 DNA at 6 different concentrations (5 to 0.02 ng/µl) and DNA samples were tested at 3 different concentrations (5 to 0.56 ng/µl). The proviral load was determined as the mean number of Tax copies per 100 cells.

2.3.3. LTR sequencing

Because the HTLV-1 LTR varies in sequence among isolates, the LTR was sequenced using Sanger sequencing (MRC Clinical Sciences Centre Genomics Laboratory) before amplification of the integration site using patient-specific primers. PCR conditions were: 3 minutes at 98^0C; 35/45 cycles of 10 seconds at 98^0C, 20 seconds at 64^0C, 20 seconds at 72^0C followed by 10 min at 72^0C, using primers flanking the 5'LTR-Gag junction. We assumed that the sequence of the primer binding sites in the 5'LTR and 3'LTR were identical.

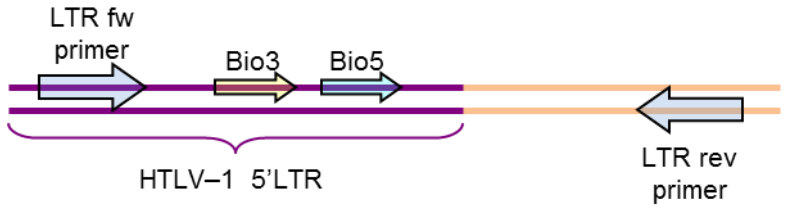

Figure 2.4: Location of the primers used to amplify the LTR region.
The forward primer (5LTRFW) and the sequencing primer (LTRseq) are upstream of the LM-PCR primers binding sites (Bio3, Bio5, see Figure 3.2B); reverse primer (5LTRRV) is downstream of the LTR-Gag junction in order to enable the complete amplification of the desired region.

2.4.Analysis and quantification of proviral integration sites

2.4.1. Preparation of integration site libraries for sequencing

Analysis of HTLV-1 proviral integration sites is detailed in Chapter 3. See also figure 3.1.

Up to 10 μg of DNA is sheared using focused-ultrasonicator (Covaris S2) using the following conditions: Initial quick burst step - 20% duty cycle, level 5 intensity, 200 cycles per burst, 5 seconds followed by 5% duty cycle, level 3 intensity, 200 cycles per burst, 90 seconds.

This is followed by enzymatic end repair to blunt the using 15U T4 DNA polymerase (New England Biolabs – NEB), 50U T4 Polynucleotide Kinase (NEB) and 5U DNA Polymerase I, Large (Klenow) Fragment (NEB) for 30 minutes at 20^0C. This is followed by adding a base on 3' ends using 30U Klenow Fragment (3'-5' exo-) for 30 minutes at 37^0C.

The DNA fragments are ligated to partially double stranded linkers (figure 3.2) containing a sample-identifying 6 base barcode using NEB Quick Ligation Kit.

Two-step PCR is used to selectively amplify the HTLV-human genome junction at the 3'LTR. The first PCR is carried out between the two primers Bio3 (at the LTR) and Bio4 (at the linker) and the second (nested) PCR between Bio5 (at the LTR) and P7 (at the linker). The second PCR step is also used to prepare the required ends for high-throughput sequencing by adding P5 to the Bio5 primer. PCR conditions for both PCR steps are 30 seconds at 96^0C; 7 cycles of 5 seconds at 94^0C, 60 seconds at 72^0C; 23 cycles of 5 seconds at 94^0C, 60 seconds at 68^0C; and 9 minutes at 68^0C.

2.4.2. Library quantification prior to sequencing.

Prepared libraries of PCR-amplified products were combined based on DNA concentration. Libraries were then quantified using QPCR which more specifically quantifies DNA amplicons which contain the correct structure for high-throughput sequencing. A standard curve was based on previously quantified and sequenced libraries.

2.4.3. Flow cell design

HTLV-1 integration site DNA libraries were sequenced in the Illumina sequencing platform. These sequencing systems use flow cells devided into 8 lanes, one of which is always used as control and contains only DNA libraries containing bacteriophage phiX DNA. The samples on each flow cell were sorted according to the following criteria:

1) Samples were randomly mixed between the lanes.
2) Where more than one sample from a particular patient was sequenced, the samples were represented only once on each lane, to enable the barcode sorting analysis.
3) Where more than one sample from a particular patient was sequenced, the same barcode was not used more than once with samples from the same patient to avoid bias due to the use of a particular barcode.

2.4.4. High-throughput sequencing using the Illumina pipeline

Prepared libraries were sequenced by Dr. Laurence Game and Adam Giess at the MRC Clinical Sciences Centre Genomics lab at the Hammersmith Hospital, London on an Illumina GAIIx or HiSeq 2000 sequencer. Sequencing was paired-end (50 bp each read) with a 6 bp barcode. PhiX DNA was used on one of the lanes as a control. The Illumina analysis pipeline was used for image processing, base-calling and alignments, with default filter and quality settings. Alignment of the reads was done using the Eland Paired algorithm of CASAVA against the UCSC human genome build 18 (excluding haplotypes, unplaced contigs and mitochondrial sequences) and HTLV-1 sequences. The last cycle of each read was excluded from the alignment, for increased accuracy.

Three sequencing reads were performed (figure 3.3): **Read 1** uses the HTLVseq sequencing primer, generating a sequence at the integration site; **Read 2** uses the SBS8 sequencing primer, generating a sequence at the shear site and **Read 3** uses the SBS8rev sequencing primer, defining the sample-identifying 6-base barcode.

2.5.Bioinformatics

2.5.1. Random sites

As a matched control for the experimentally derived HTLV-1 integration sites, a random list of integration sites generated *in silico* was produced. A list of 192,171 genomic coordinates was generated *in silico* by Nirav Malani (Group of FD Bushman, U Penn, USA) based on genomic coordinates in hg18 chromosomes (excluding gaps).

In order to eliminate any *in silico* sites that originate from genomic regions where true integration sites would not be identified (gaps, repeated regions), 50 bp DNA sequences at the genomic sites (as well as "paired end" matches within 200 bp) were generated using the Galaxy tool (Blankenberg et al., 2010) and back-aligned to the human genome reference through the Illumina pipeline in an identical manner to the "true" integration sites. Resulting export files were extracted using the same filters (see chapter 3 for details) as "true" integration sites resulting in a total of 176,514 sites.

A "single read" sample (not including the "paired" sites within 200 bp, total of 170,433 sites) was also generated but was found not to differ significantly from the paired list with respect to the genomic environment.

2.5.2. Annotation of genomic environment

Genomic environment flanking Integration sites were compared with specific annotations to the human genome. See Table 4.3 for detailed information and references for each dataset used.

Transcription unit and CpG islands data were retrieved from publicly available datasets on the NCBI ftp site (ftp.ncbi.nih.gov/gene/) and UCSC tables (Karolchik et al., 2004), respectively. Transcription factor and chromatin remodeler chromatin immunoprecipitation - sequencing (ChIP-seq) datasets were retrieved from published datasets (see Table 4.3). Where possible, data from primary $CD4^+$ cells were used; otherwise, data from human $CD4^+$ cell lines or other available cell line. Where raw ChIP-seq data were available we used the SISSRs algorithm (Jothi et al., 2008) to identify the position of a putative transcription factor binding site. Comparisons of genomic coordinates of integration sites and genomic annotations were carried out using the R package hiAnnotator (http://malnirav.github.com/hiAnnotator), kindly provided by Nirav Malani and Frederic Bushman (University of Pennsylvania, USA)

2.5.3. Quantification of Tax expressing/non-expressing clones

To calculate the fraction of Tax^+ cells in a given clone, the frequency of Tax^+ and Tax^- cells were normalized to the mass of genomic DNA per cell from each respective cell population, to correct for experimental variation in efficiency of genomic DNA isolation (Table 5.1). As proviral load data was not available for Tax-sorted samples, clonal abundance was quantified by the number of sisters per clone.

2.5.4. Calculation of CD8+ T-cell contribution to the proviral load

The contribution of infected $CD8^+$ cells to the proviral load of each infected patient was calculated by using two approaches:

1) Calculation based on proviral load in either $CD4^+$ or $CD8^+$ population. If $\left(\frac{CD8^+PV^+}{CD8^+}\right)$ is the proviral load within $CD8^+$ cells, and $\frac{CD8^+}{PBMC}$ is the proportion of PBMCs which are $CD8^+$, the contribution of infected $CD8^+$ cells to the proviral load is calculated in the following manner:

Equation 2.1

$$\%\ \text{of load in } CD8^+ \text{ cells} = 100 \times \left(\left(\frac{CD8^+PV^+}{CD8^+}\right) \Big/ \left(\frac{CD8^+PV^+}{CD8^+} \times \frac{CD8^+}{PBMC}\right) + \left(\frac{CD4^+PV^+}{CD4^+} \times \frac{CD4^+}{PBMC}\right) \right)$$

2) The proportion of cells identified within the unsorted PBMC integration site data as belonging to $CD8^+$ clones.

2.5.5. Statistics

Statistical analysis was carried out using R version 2.13.0 (http://www.R-project.org/). See text for details of particular statistical tests. Where appropriate, non-parametric tests were used. Results were considered statistically significant when $p < 0.05$.

Two separate logistic regression analyses were carried out, respectively, to identify independent predictors of HTLV-1 integration targeting (Chapter 4) and independent

predictors of Tax positivity (Chapter 5). Input variables used were published ChIP-seq datasets (see 2.5.2 above, Table 4.3). For integration targeting (Chapter 4), the binary outcome measure was a "true" integration site (from 4521 identified in vitro integration sites) or a "false" integration site (45210 random genomic locations). For spontaneous Tax expression (Chapter 5), the binary outcome was either Tax positivity (20813 Tax^+ cells) or Tax negativity (10326 Tax^- cells). For each outcome variable, two separate analyses were carried out, respectively at two distances from the integration site - 100 bases and 1kb. Each transcription factor binding site (TFBS) was tested (presence or absence of the TFBS within a given distance of the integration site) as an independent predictor in each analysis.

The multivariate analysis approach is detailed in Figure 2.5. First, for each TFBS and at each distance, we tested whether the relative position (upstream/downstream) of the integration site and the TFBS determined the outcome by using a likelihood ratio test to compare two competing models: 1) presence or absence of TFBS upstream or (separately) downstream; 2) presence or absence of TFBS, regardless of relative position. Next, we carried out univariate analysis of each individual TFBS, based on the model chosen by the likelihood ratio test. Only TFBS that were significant (p-value < 0.05) after correction for multiple comparisons (Benjamini-Hochberg) were used in the multivariate analysis. Multivariate analysis was carried out using a step-down logistic regression method.

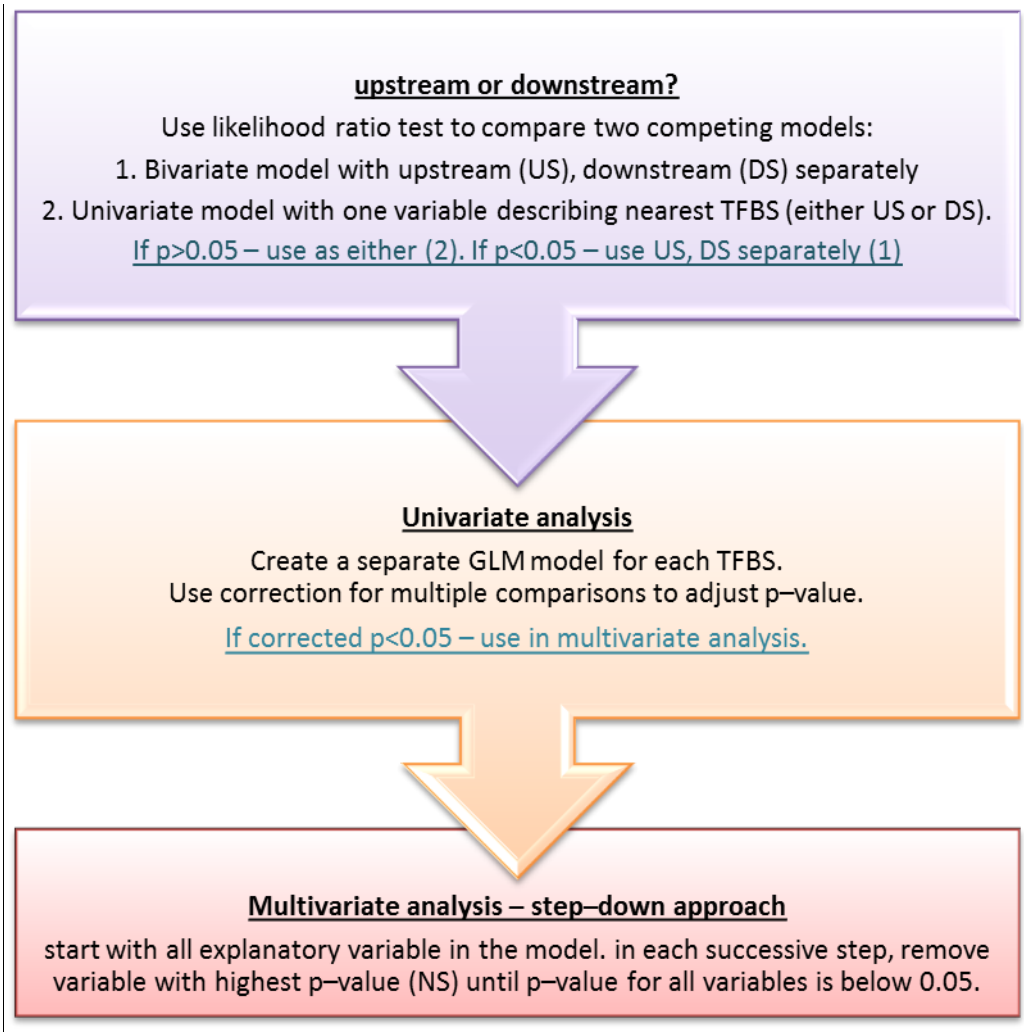

Figure 2.5: Analytical approach for multivariate analysis.

Multivariate analysis was used in order to find independent correlates of proviral integration and proviral expression. For integration targeting, the binary outcome measure was a "true" integration site (from 4521 identified in vitro integration sites) or a "false" integration site (45210 random genomic locations). For spontaneous Tax expression, the binary outcome was Tax positivity (20813 Tax^+ cells) or Tax negativity (10326 Tax^- cells). Analytical steps for building multivariate logistic model are detailed here (see also section 2.5.5).

Chapter 3 - High-throughput method for mapping, identification and quantification of retroviral integration sites

3.1 Introduction

3.1.1 Previous methods used for analysis of integration sites

The importance of studying the clonal distribution and proviral integration sites in HTLV-1 infection was realized shortly after the identification of the virus. Yoshida et al. (1984) demonstrated that HTLV-1 infection preceded the clonal expansion of a putative malignant clone in ATLL patients utilizing a southern blot assay which reveals a strong band when a large (i.e. abundant) clone is present. Owing to the low sensitivity of this assay, newer methods were developed based on inverse-PCR (IPCR; Takemoto et al., 1994) and linker-mediated PCR (LMPCR; Wattel et al., 1995). The LMPCR technique has also been used to identify genomic integration sites of other retroviruses (e.g. analysis of MLV integration in Wu et al., 2003 or of HIV integration in Schroder et al., 2002) and by this group (Meekings et al., 2008) and others (Derse et al., 2007) to identify a small number of clones from HTLV-1-infected cells. Though effective, the LMPCR method was affected by 3 major limitations: First, the use of restriction enzymes limited the ability to identify integration sites which were not within a specific proximity range from the nearest restriction site (Wang et al., 2008). Second, the techniques were limited with respect to the number of integration sites reasonably identified in a single experiment from each patient, leading to a severe under-estimation of the true number of infected clones in a patient. Third, those techniques were not quantifiable, and any estimation of the proportion of infected cells in a given clones would have required further laborious and inefficient quantitative techniques.

3.1.2 The use of high-throughput sequencing in retroviral integration research

The Illumina (previously Solexa) high-throughput sequencing technology (Bentley et al., 2008) is based on sequencing-by-synthesis mechanism. This technology employs bridge amplification of flow cell-tethered templates followed by sequencing using fluorescently labelled reversible chain terminators, to deliver a large number of short reads (as of 2012, up to 3 billion reads per flow cell on the HiSeq 1000/2000, Liu et al., 2012).

The development of high-throughput sequencing techniques (Holt and Jones, 2008) has opened the door to developing methods for deeper analysis of integration site of retroviruses, retroviral vectors and other retroelements (Wang et al., 2008; Gabriel et al., 2009; Williams-Carrier et al., 2010). However the challenge to improve the recovery and accurately quantify the relative frequency of the integration sites remained. Approaches to tackle these problems included the use of a combination of restriction enzymes (Wang et al., 2010), use of phage Mu retrotransposition (Brady et al., 2011) and non-restrictive linear amplification–mediated PCR (nrLAM-PCR; Gabriel et al., 2009). However, even a combination of restriction enzymes, and to a lesser extent the use of transposons, still require recognition motifs that could limit recovery and may skew quantification (Brady et al., 2011).

3.1.3 Aim

We aimed to develop a new approach, based on the Illumina high-throughput sequencing platform, which would be able to address the limitations of the previous methods and simultaneously map and quantify an unprecedented (in HTLV-1) number of distinct integration sites from infected PBMCs. This chapter will detail the main steps in this approach, as well as its strength and limitations.

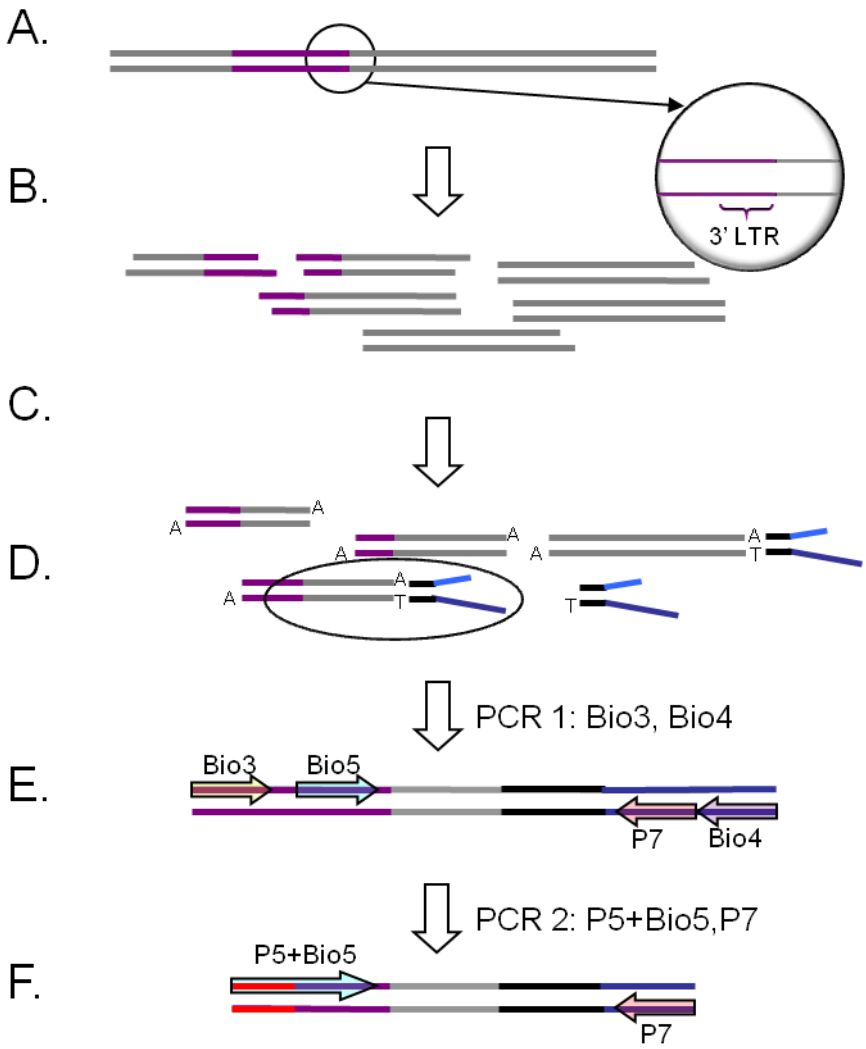

Figure 3.1: Process of LMPCR – amplification of HTLV-1 integration sites

(A) The region of interest is the junction between the 3'LTR and the human genome. (B) DNA is randomly sheared by sonication. (C)-(D) DNA ends are blunt ended and A base is added at 3' ends to allow ligation to linker (Figure 3.2). (E) PCR step between primers Bio3 (LTR specific, patient specific primers are used) and Bio4 (linker specific). (F) Nested PCR step between primers Bio5 (LTR specific) and P7 (linker specific). P7 and P5 oligonucleotides are required for attachment to Illumina flow cell. P5 oligonucleotide is added by including it in the PCR primer (P5-Bio5). P7 is designed into the linker.

3.2 Method description

3.2.1 Preparation of libraries for sequencing

Based on the classical LMPCR, our method was modified to address the limitations of the previous techniques and is illustrated in Figure 3.1. The target site of interest (Figure 3.1A) is the junction between the 3'LTR of HTLV-1 and the human genome. First, the use of restriction enzymes to digest the genomic DNA is replaced by random shearing of the DNA using sonication (Figure 3.1B). This eliminates the possibility of a systematic loss of integration sites that lie far from the nearest restriction site, and also allows us to quantify the frequency of the integration sites (see below). The sheared DNA is blunt-ended (Figure 3.1C) and ligated to a partially double-stranded adaptor (Figure 3.1D, Figure 3.2), which is then followed by a nested PCR (Figure 3.1E/F). This nested PCR strategy enhances the specificity and sensitivity of the assay. The use of a partially double-stranded linker used (similar to splinkerette, see Devon et al., 1995) improves specificity of the assay by allowing the selective amplification (during first PCR) of amplicons containing HTLV-1 LTR only.

The linkers used also contain a 6-base barcode sequence, unique to each sample. This barcode enables the multiplexing of multiple samples per lane (usually up to 40 on the HiSeq).

Lastly, the resulting DNA libraries are sequenced on the Illumina platform, generating a remarkable number of reads originating from a high number of individual clones.

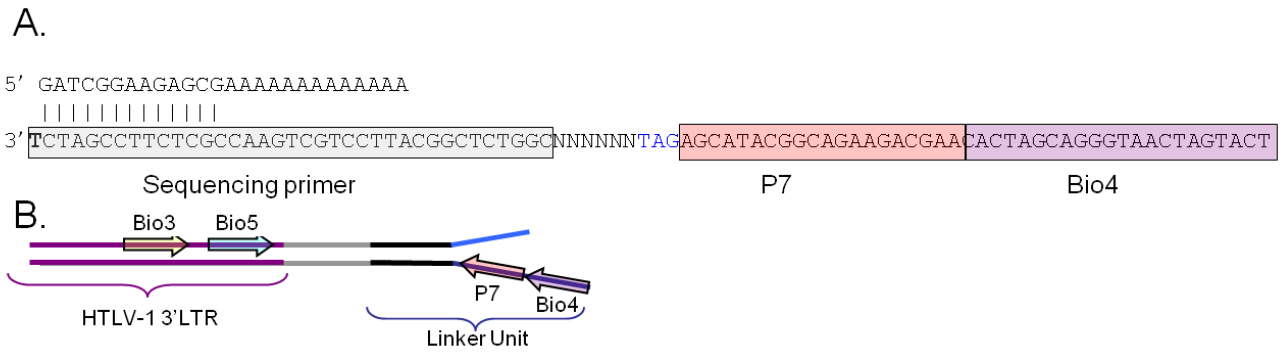

Figure 3.2: linker structure.

(A) Structure of the partially double stranded linker used for analysis of proviral integration sites. The nested PCR primers (P7, Bio4), as well as the paired sequencing primer sequences are highlighted. The linker does not contain a complementary sequence for the Bio4 primer, thus increasing PCR specificity (DNA containing proviral sequences only will be amplified). P7 is also used as grafting primer for attachment of resulting DNA amplicon to the flow cell. 6 base sample-specific barcode (here denoted by NNNNNN) is located between the sequencing primer binding site and P7 binding site. (B) Basic structure of ligated DNA. The Bio5, Bio3 primer binding sites (Wattel et al., 1995) are HTLV-1-specific, used for nested PCR of HTLV-1 integration sites.

3.2.2 High-throughput sequencing

The specificity of the technique is further improved by using a sequencing primer which binds a site, 5 bases short of the LTR terminus (Figure 3.3). This ensures that reads from all true HTLV-1 integration sites will begin with the sequence ACACA, and enables us to programmatically eliminate any reads arising from mispriming of the sequencing primer and focus only on true integration sites.

Sequencing of integration sites containing amplicons (Figure 3.3) is done on the Illumina platform. Each amplicon is sequenced 3 times:

1) Using a sequencing primer complementary to the HTLV-1 LTR – sequence (50 bases) maps to the integration site
2) Using a sequencing primer complementary to the linker, reading into the genomic sequence – sequence (50 bases) maps to the shear site.
3) Using a sequencing primer complementary to the linker, reading into the 6-base sample-specific barcode.

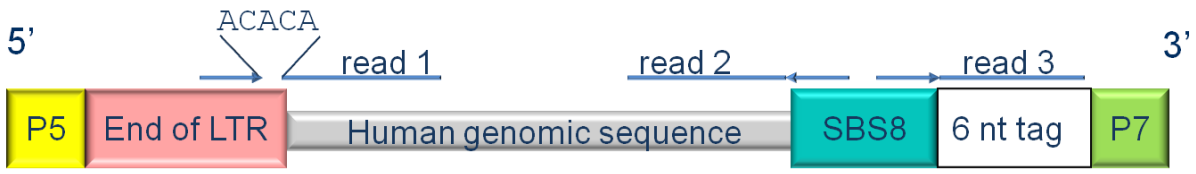

Figure 3.3: Amplicon basic structure for high-throughput sequencing

The LMPCR process to prepare the integration site libraries for sequencing is detailed in Figure 3.1. P5, P7 are required oligonucleotides for attachment to Illumina flow cell. Each amplicon is sequenced three times (detailed above). As the 3'LTR and 5' LTR terminus are identical, we would expect 50% of amplicons to contain HTLV-1 specific sequences instead of human genomic sequence. These are identified by alignment and removed bioinformatically (see below).

3.3 Data extraction –DEISA

3.3.1 Main steps in pipeline

In order to extract the data generated by the Illumina sequencing, I developed a bespoke application on the platform of the Microsoft ACCESS system – DEISA (Data Extraction for Integration Site Analysis). The application (Figure 3.4) provides a user-friendly interface controlling a large number of SQL tables and queries, ensuring consistency between experiments in the manner by which data is extracted, and allows other users to extract data for their experiments without any previous knowledge or expertise in using SQL or ACCESS.

The input file to DEISA is the Illumina export file, which contains all sequenced reads, their mapping (as performed using the efficient large-scale alignment of nucleotide databases (ELAND) algorithm in the Illumina GERALD/CASAVA pipeline) and several quality details. DEISA is able to auto-import the files, and then process them in turn. The process is composed of the following steps:

1) Data filtering – the sequencing reads are screened for specificity (must begin with ACACA), the quality of sequencing and the quality of mapping.
2) Lane data extraction (Figure 3.5) – identification of the integration sites by distinct read 1 genomic coordinates, calculating the number of shear sites by the number of distinct read 2 genomic coordinates for each integration site. Sorting of integration sites to the different barcodes on a multiplexed lane.
3) Data refinement – removal of artifact UIS (UIS which were mapped to a particular genomic locations but are in fact a PCR duplicate of another UIS with a small PCR/sequencing error causing a mismapping), calibration of the number of shear-sites (see below) and generation of the final list of integration sites.

3.3.2 Criteria used

Three main criteria are used for initial assessment of each read:

1) Quality of sequence cluster – remove reads which come from clusters overlapping on the flow cell, which do not produce reliable sequences.
2) Specificity – Read1 sequence (Figure 3.3) must start with ACACA (for HTLV-1 integration sites).
3) Quality of mapping to the human genome. The Sequence must be uniquely mapped to the human genome reference in both reads, with a mapping score >10 (based on the Illumina Gerald pipeline and is required to be of reasonable length (measured by the genomic coordinates difference between the integration site and the shear site). Sequences mapped uniquely to the upstream LTR are also quantified separately.

3.3.3 Quantification

PCR bias is a known complexity when using a variable template (Polz and Cavanaugh, 1998), in particular because of preferential amplification of short fragments in each successive round of PCR (Figure 3.9). This PCR bias eliminates the ability to rely solely on the number of reads for determining the frequency of each integration site (the clonal abundance). The random shearing of the DNA prior to ligation to the adaptor allows an alternative form of quantification (Figure 3.6) – as the shearing site of the DNA will occur in a different position for DNA originating from each infected cell, by counting the number of different shear sites we are able to count the number of cells in each clone. We define an infected **clone** as all the cells which share an integration site. We define the different cells sharing an integration site as **sisters** (thus, the clone abundance is the total number of sisters

in a given clone, n). We define a clone which was only identified once as a **singleton** (i.e. clones for which $n = 1$).

We define three main measures of abundance of a given clone:

$S = total\ number\ of\ clones\ identified\ in\ sample$

The clone abundance - $n_i =$ calibrated number of sisters in the i^{th} clone

The relative abundance – the share of the load taken up by each clone:

Equation 3.1 $$\text{Relative abundance} = \frac{\boldsymbol{n_i}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{S}} \boldsymbol{n_i}}$$

The Absolute abundance – the number of sisters of each clone per 10000 PBMC:

Equation 3.2 $$\text{Absolute abundance} = \text{PVL} \times \frac{\mathrm{n_i}}{\sum_{\mathrm{i=1}}^{\mathrm{S}} \mathrm{n_i}}$$

3.3.4 Barcode errors

We assume that the probability that HTLV-1 integrates at the same genomic location in two infected individuals is negligibly small, because previous studies (Meekings et al., 2008; Derse et al., 2007) showed no evidence of hot-spots of HTLV-1 integration. Therefore any UIS identified in more than one sample on the same sequencing lane is assumed to be a barcode read error and resolved by assigning the integration site to the sample where more unique shear sites were identified for that clone (or in the case of two singletons, to the sample with the greater number of PCR duplicates for that clone). If both samples have the same number of PCR duplicates for the clone (normally if both = 1) the integration site cannot be reliably assigned, and is discarded.

3.3.5 Data purification

Prior to generating the final integration site list for each sample, a step of data purification is carried out:

1) Integration sites mapped to chromosome 'Y' in a sample coming from a female subject are removed.
2) Removal of "twins" – on certain occasions, a change of one or a few bases (due to misincorporation during PCR or ambiguity in sequencing) could result in a mismapping of some duplicates of an integration site, suggesting a new integration site. We eliminate those by disregarding smaller 'clones' where the integration site lies within 2 bases of the site in a larger (more abundant) clone in the same sample, and by sequence similarity as quantified by the ClustalW algorithm (http://www.ebi.ac.uk/Tools/msa/clustalw2/)

Figure 3.4: DEISA start page

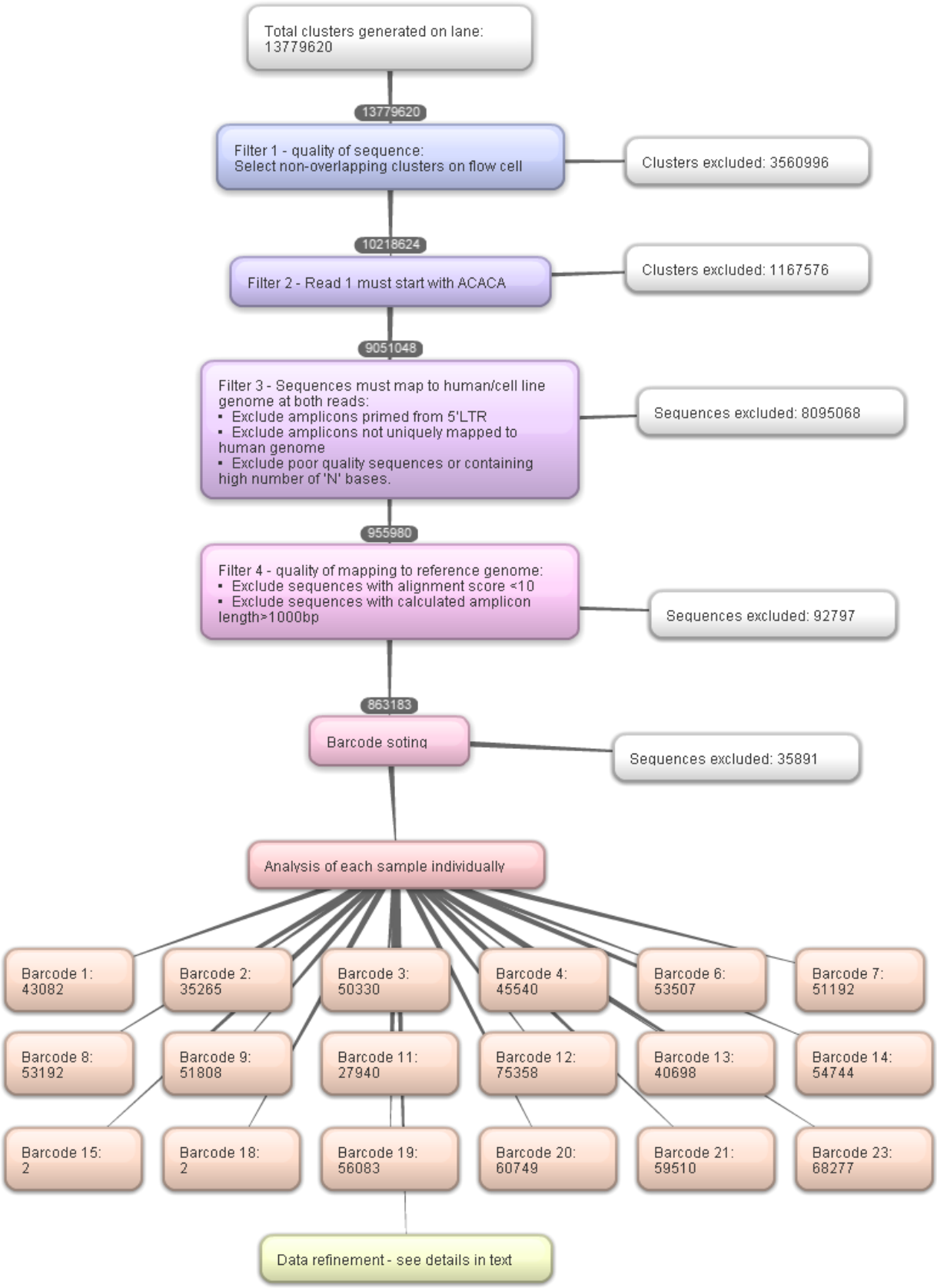

Figure 3.5: Example of data extraction of a complete flow cell

Illustration of the process of filtering high-throughput sequences for analysis of HTLV-1 integration sites. Figure from a lane on a typical sequencing run, using Ilumina Genome Analyser II.

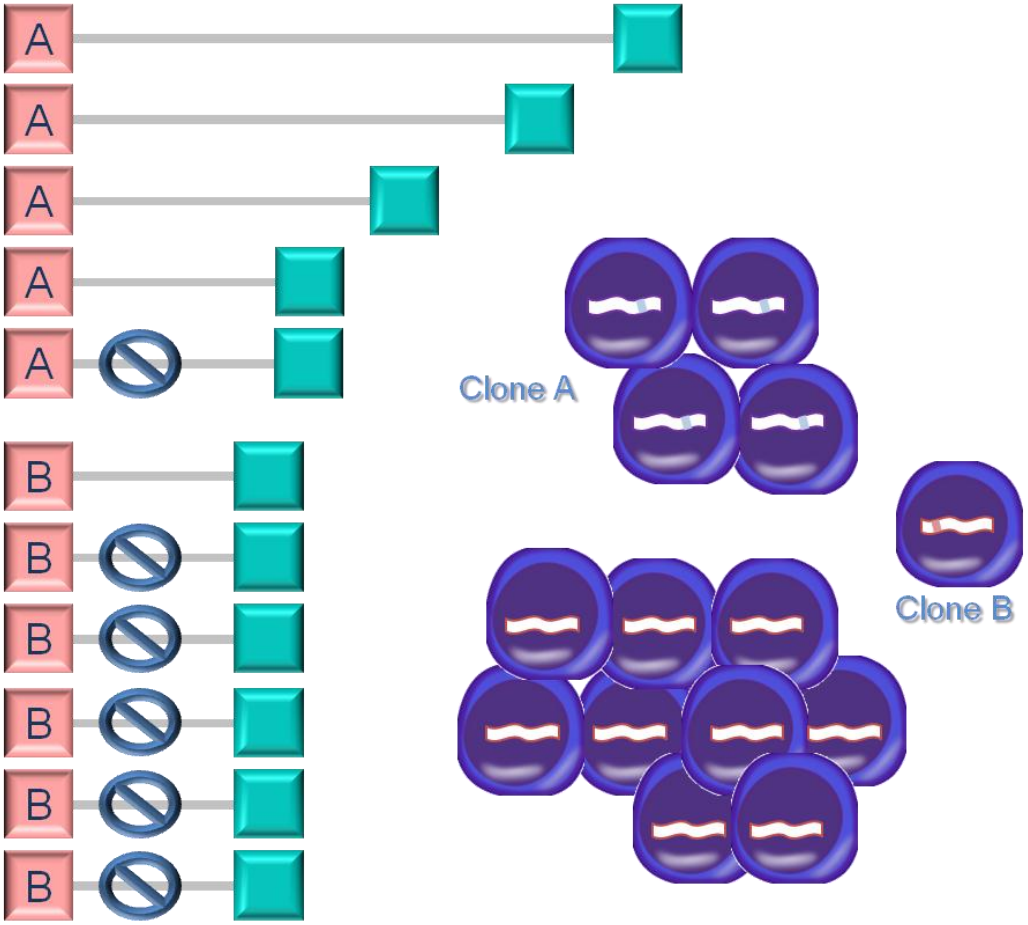

Figure 3.6: Quantification of proviral integration sites.

Two clones are shown for illustration. The clones are identified based on shared integration site (pink). The clonal abundance of each given clone is based upon the number of unique shear sites (green) for that clone. Identical shear sites (PCR duplicates) are excluded.

3.3.6 Oligoclonality index

The ability to quantify the integration sites has provided an unprecedented view of the clonal distribution in HTLV-1 infection (Figure 3.11A). This has allowed us to define the oligoclonality index (OCI, Figure 3.7), a parameter which describes the clonal distribution. The OCI varies between the two extremes of 0 (polyclonal, all clones share an equal share of the proviral load) and 1 (monoclonal, all infected cells belong to a single clone), and is calculated in a similar manner to the Gini index (Gini, 1912). The OCI removes the need for arbitrary categories of mono/oligo/polyclonality and is an objective, rigorous parameter for comparison of clonality within and between infected individuals and cell populations.

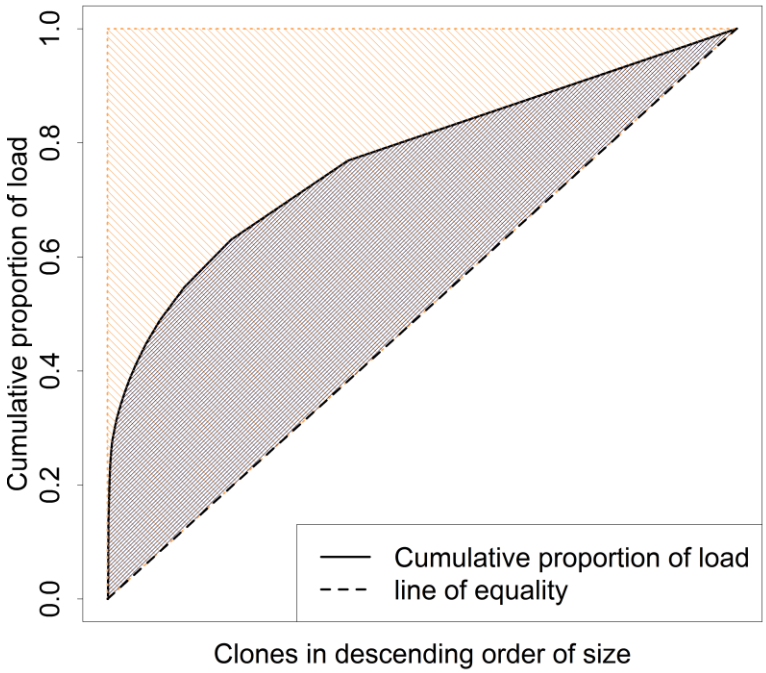

Figure 3.7: Oligoclonality index (OCI) - a parameter of clonal distribution
The oligoclonality index is a measure of clonal distribution inequality. Ranking the clones in descending order of clonal abundance (X axis) and calculating the relative proportion of the load (Y axis), its value can be described as the ratio between the dark shaded area between the cumulative abundance curve and the line of equality, and the lightly shaded area above the line of equality.

3.3.7 Correction for multiple hits at same sites in large clones.

We assume, for the purpose of quantification of clonal abundance, that due to the random shearing, DNA coming from each cell will be sheared at different sites. The likelihood that two cells belonging to the same clone have the same shear site by chance increases with the number of cells in that clone. In order to quantify this likelihood, Nicolas Gillet carried out an experiment where patient DNA was used in successively lower concentrations. "True" sister count for each clone was estimated by multiplying the dilution factor by the observed number of shear sites in the lowest concentration for the same clone.

A spline function (Equation 3.3), fitted by Dr. Charles Berry to this curve, can be used to estimate the clonal abundance of highly abundant clones with a large number of shear sites.

Equation 3.3

$$y = \exp\begin{pmatrix}\log(\min(50, x)) + 1.18 \times \max(0, \log(x) - \log(50)) \\ + 0.707 \times \max(0, \log(x) - \log(50))^2\end{pmatrix}$$

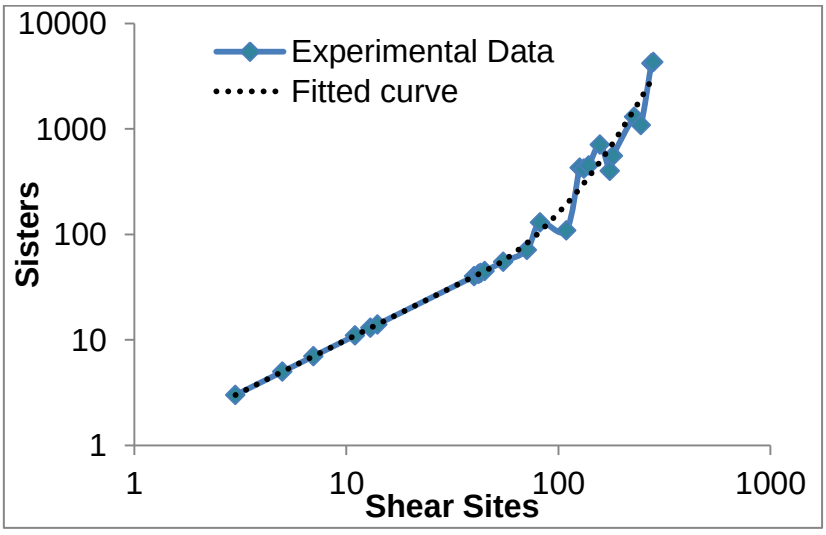

Figure 3.8: Calibration curve for correction of clonal abundance.
Experiments performed by Nicolas Gillet. Calibration curve was fitted by Charles Berry. See text above for details.

3.3.8 Calculation of sensitivity

Loss of proviral integrations can take place during each successive step of the process, by loss of DNA during purifications, by incomplete enzymatic processes or during sequencing and analysis steps. This requires careful monitoring of ratio of proviral copies recovered to the expected number of copies calculated from the independent measure of the proviral load and the amount of DNA available. This ratio is a useful measure of sensitivity, and an indicator of the success of each sample on a particular flow cell.

The calculation of sensitivity is carried out in the following manner:

$S = total\ number\ of\ clones\ identified\ in\ sample$

$n_i =$ calibrated number of sisters in the i^{th} clone

Equation 3.4
$$Observed\ number\ of\ proviral\ copies = N = \sum_{i=1}^{S} n_i$$

Equation 3.5
$$Input\ number\ of\ proviral\ copies = PVL\left[\frac{copies}{100\ cells}\right] \times DNA[\mu g] \times \frac{150000\ cells}{1\ \mu g\ DNA}$$

Equation 3.6
$$Proportion\ of\ proviruses\ recovered = \frac{Observed\ number\ of\ proviral\ copies}{Input\ number\ of\ proviral\ copies}$$

3.4 Results

3.4.1 Yield

The number of proviral integrations (sisters) recovered varies between infected individuals, and largely depends on the input (DNA + proviral load of the sample). See Table 3.1 for details of the datasets used in the experiments described in this work.

Table 3.1: Integration site datasets used

Dataset	Cell type	total UIS	total infected individuals	reference	Chapter
in vitro (1)	Jurkat	4521	N/A	Gillet et al., 2011	4
in vitro (2)	Jurkat	1805	N/A	This work	4
In vivo (1) [a]	PBMC	78563	63	Gillet et al., 2011	4
In vivo (2)	PBMC	20202	10	This work	5
Tax Negative	$CD4^+$	6700	10 (pooled)	This work	5
Tax Positive	$CD4^+$	13054	10 (pooled)	This work	5
In vivo (3)	PBMC	22510	12	This work	6
$CD4^+$ cells	$CD4^+$	33073	12	This work	6
$CD8^+$ cells	$CD8^+$	2399	12	This work	6
Random UIS	N/A	176505	N/A	This work	control

[a] Where multiple time points were available for any single patient, samples from only the most recent time point were used for this work. In addition, only patients with a single clinical diagnosis were included (e.g, patients with both ATLL and HAM/TSP were excluded).

3.4.2 Confirmation of PCR bias

The distribution of the mean number of PCR duplicates by amplicon length in a representative sample is shown in Figure 3.9. The results confirm the previously reported bias towards short fragments: shorter fragments are more often found, and have a higher number of PCR duplicates than longer fragments. This observation therefore confirms the need for a PCR duplicate-independent method of quantification.

Amplicon length varies typically mostly between ~100 and ~500 bases, because of the preferential amplification of short fragments and the more efficient generation of clusters with amplicons under 500 bases (Quail et al., 2008).

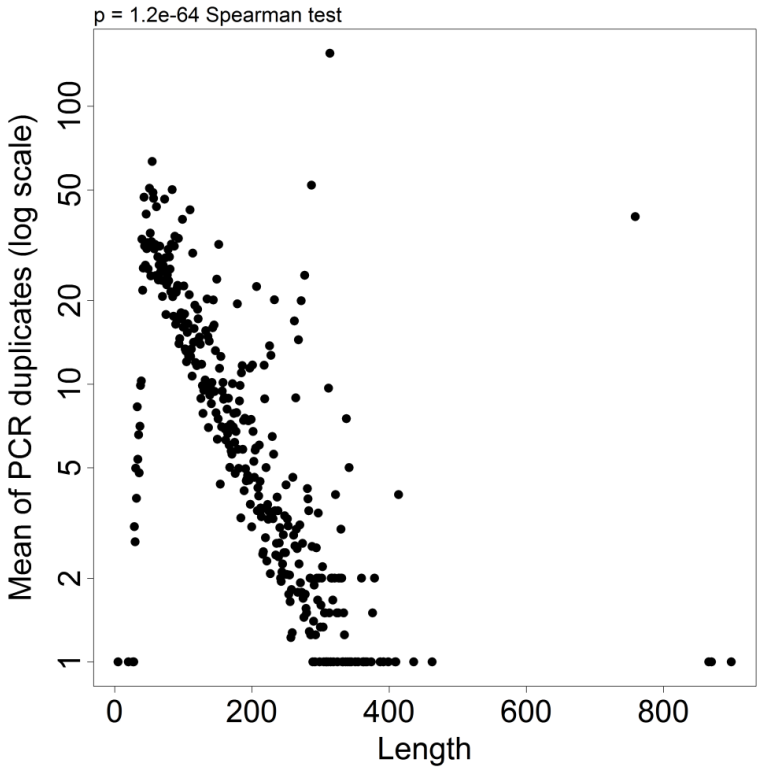

Figure 3.9: Shorter fragments are amplified more efficiently by PCR.
The mean of number of PCR duplicates is inversely correlated with the amplicon length, confirming a bias towards a more efficient amplification of short fragments by PCR. Figure from a typical sequencing run, using Illumina Genome Analyser II.

3.4.3 Recovery of proviral integrations

The recovery plot of the data described in chapter 6 is shown in Figure 3.10, using data from the Illumina HiSeq platform. The majority of the samples were of relatively high input (owing to the high proviral load) and the median recovery was 4.2%. There was a weak (non-statistically significant) negative correlation between the input proviral copies and the recovery, which may reflect the limited ability of PCR to amplify a high number of different amplicons simultaneously.

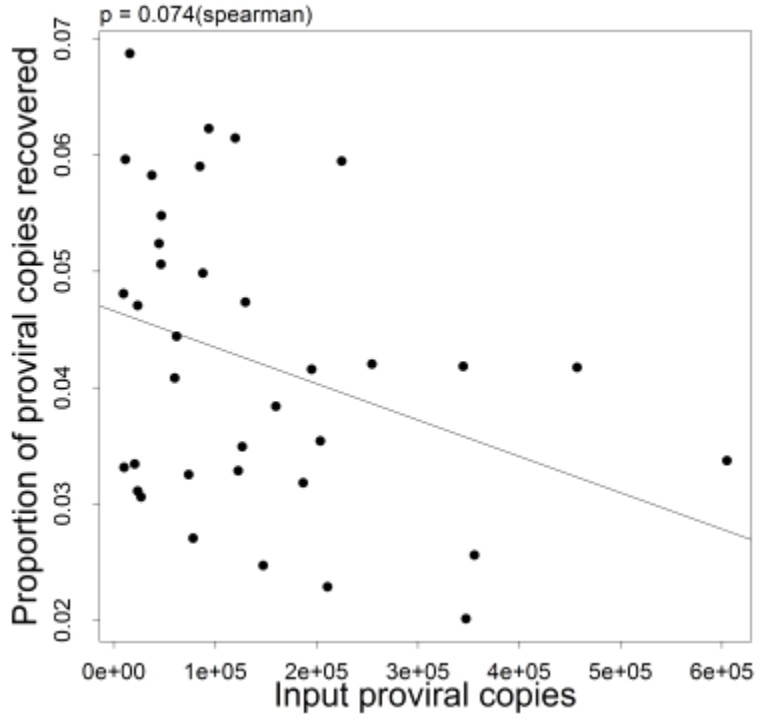

Figure 3.10: Recovery versus input proviral copies.
The median recovery is 4.2% of input sisters (proviral copies). There is a weak inverse correlation between proportion of recovery and number of input copies. Data from experiment conducted on the HiSeq 2000.

3.4.4 OligoClonality index

Representative clonal frequency distributions from two asymptomatic patients identified by this method are shown in Figure 3.11. The experimental protocol enables us to rank the integration sites by the clonal abundance in vivo (Figure 3.11A) and interrogate the genomic environment flanking the integration site. We are then able to quantify the shape of the clonal distribution (Figure 3.11B) and use that to compare the distribution between different infected individuals, between cell populations, before and after treatment and over time.

In 2011, Nicolas Gillet and colleagues reported that as predicted, in the malignant state the OCI is higher, reflecting selective proliferation of the putative malignant clone. There was no difference between the OCI of HAM/TSP patients and that of ACs (Figure 3.12).

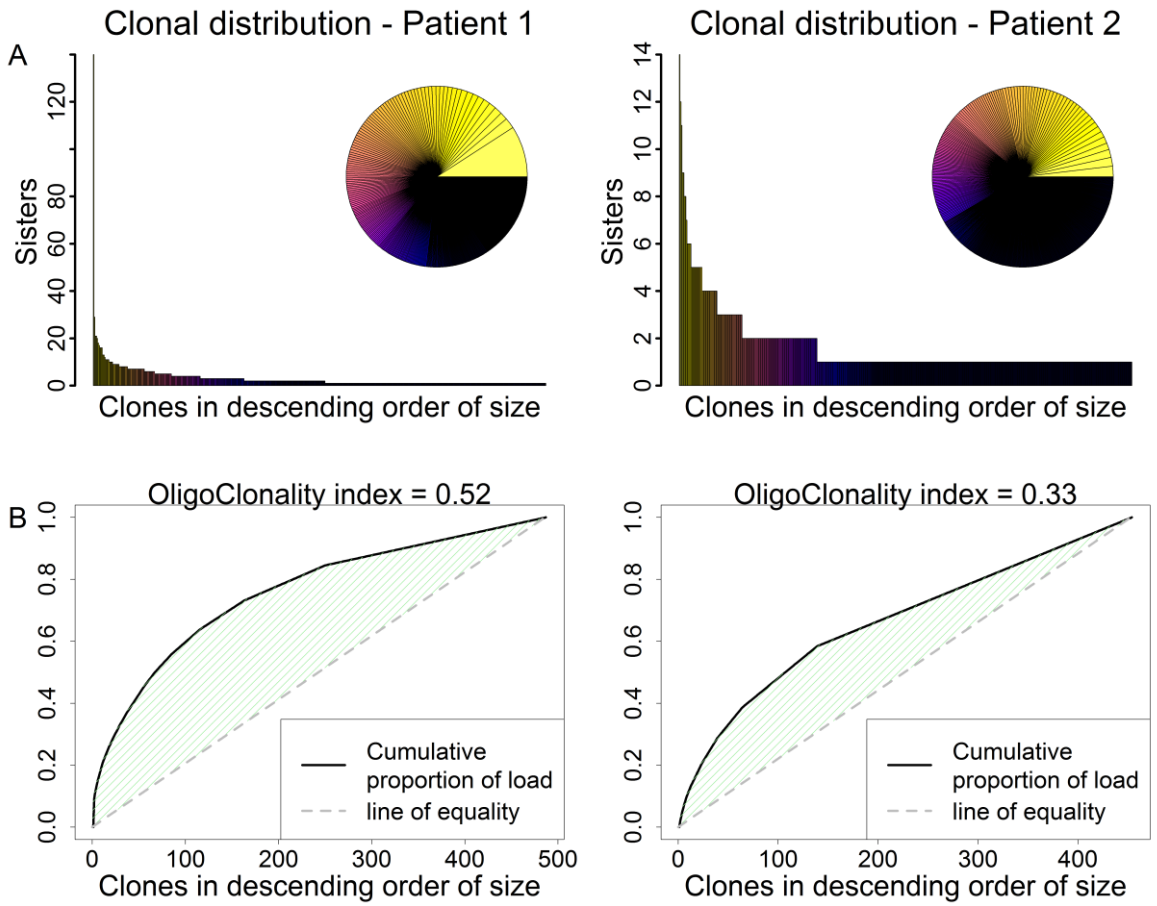

Figure 3.11: Example of the clonal distribution of two typical asymptomatic patients.
(A) Two alternative depictions of the clonal distribution in order of clone size from the largest to the smallest. The histogram shows for each clone the number of sisters identified (note difference in scale between patients). The inset pie chart shows the relative proportion of the load occupied by each clone. Patient 1 is more oligoclonal than patient 2, with the largest clone representing approximately 9% of the proviral load. Patient 2 has a more uniform clone frequency distribution, in which no clone represents more than 2% of the load. (B) Cumulative proportion of the clone across all clones in decreasing order of size allows the calculation of the OCI (see Figure 3.7)

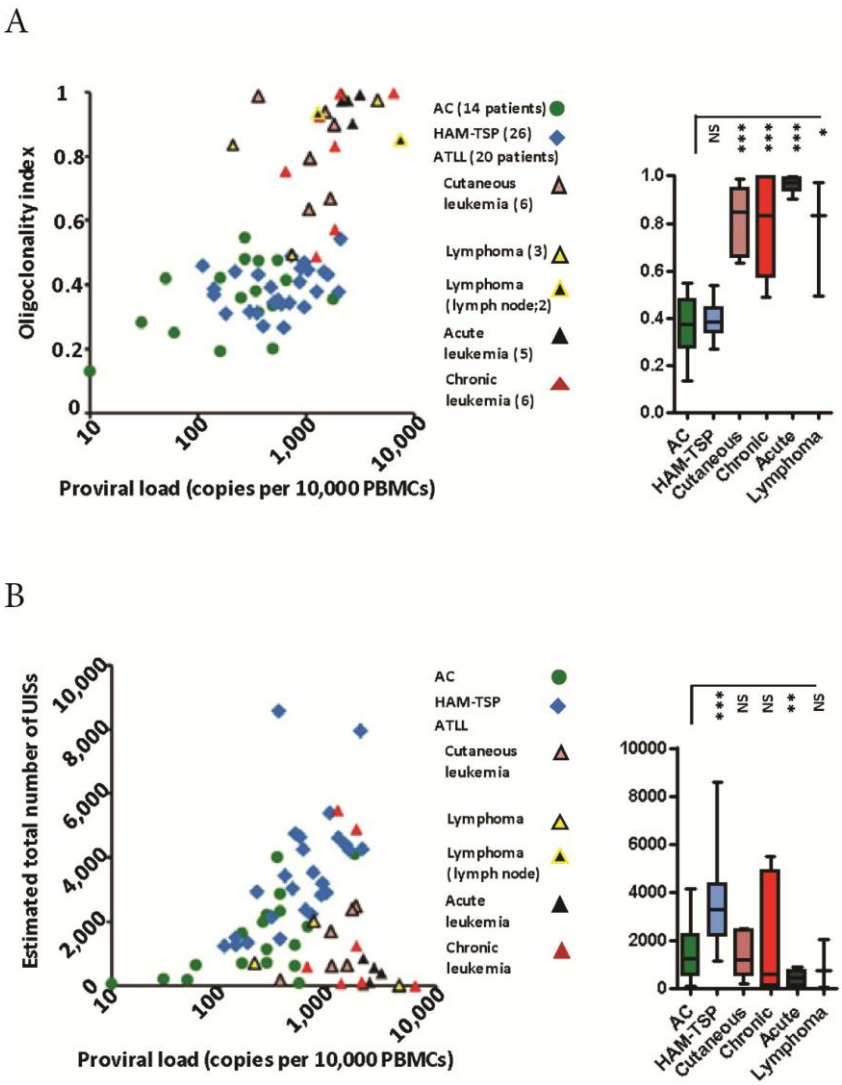

Figure 3.12: Oligoclonality and estimated total number of clones.

Figure adapted from Gillet et al., 2011. (A) OCI is significantly higher for ATLL patients than non-malignant cases. There is no significant difference between asymptomatic carriers and HAM/TSP patients. (B) The high PVL of HAM/TSP appears to be explained by a higher clone count.

3.1 Discussion

In order to more profoundly understand and monitor the proviral load of HTLV-1-infected individuals, it is necessary to monitor and understand the different clones that make up the proviral load. We have developed a method that is able to simultaneously amplify, map and quantify the proviral integration sites. This method has two main advantages over other techniques. The first is its high sensitivity. The random shearing prevents the loss of integration sites (including potentially major clones) that lie far from the nearest restriction site, and the high-throughput sequencing allows the identification of an unprecedented (in HTLV-1) number of clones from each patient. The second main advantage is the ability to quantify each integration site. The random shearing results in a unique shear site for each infected cell, and the simultaneous sequencing of both the integration site and the shear site allows the relative quantification of recovered clones.

Not all proviruses in the sample are sequenced. The stochastic nature of the detection of a given integration site by LMPCR has been described by Cavrois et al. (1995) and high-throughput sequencing is also subject to sampling error. Calculation of the predicted number of proviruses input allows us to calculate the percentage of expected proviruses that are observed. However, it should be noted that any sample represents a small proportion of all infected cells in the blood. We assume equal mixing of the infected cells in the circulation (this assumption may be unjustified in solid lymphoid tissues such as the lymph node), and therefore that there should be no particular bias towards certain infected T-cell clones in the blood sample used. Similarly, the random shearing of DNA using this method for identification of integration site effectively eliminates restriction site dependent bias in recovery of any particular provirus. While we do not anticipate any systematic loss of

particular clones, the probability of re-recovery of any one particular clone may depend on its abundance.

This method can be used for two parallel avenues of research. The mapping of the genomic integration site allows us to study the genomic environment of the integrated provirus in great detail, using the available annotations to the human genome. This allows us to interrogate the genomic environment in order to test for whether the genomic environment determines the proviral integration targeting, clonal expansion or expression of proviral genes (Gillet et al., 2011, Chapter 4, Chapter 5 of this work). At the same time, this method provides a marker for each HTLV-1 infected clone. This can be used to track an infected clone over time, in response to therapy, to quantify the appearance of new mutations and to test hypotheses on the possible role of that clone in pathogenesis. This can be done on a population level, in the context of all other clones, without the need for limited dilution cloning of T cells, which requires a lengthy period of cell culture, which may be further from the physiological state of the cell or may select for particular subtypes of clones that are able to grow in culture.

There are two main limitations to this method, which are shared with any other method based on high-throughput sequencing. The sequencing of samples using the Illumina method may not be as efficient in extremely high GC content sites (Aird et al., 2011). The second limitation is its dependency on the available human genome reference. The human genome reference still contains some minor gaps (Kidd et al., 2010), as well as repeats and highly polymorphic sites which are not possible to map sequence data to with confidence. This should be considered in particular when comparing sequencing data to random expectation, and to make sure that the random expectation is drawn from a more "realistic" human reference (see section 2.5.1). When comparing two datasets of integration sites to each other,

there is no a priori reason to assume that one will be systematically less efficiently mapped than the other.

Lastly, similar to most studies of HTLV-1 infection, there is a sensitivity limit, dependent on the proviral load, determined by the number of input copies of the provirus in the protocol. While the proviral load can vary greatly between different individuals (Nagai et al., 1998 Demontis et al., 2013), a very low copy number would result in a small number of sisters (integrations) detected, significantly reducing the power of this technique. We do not expect basic mechanisms of selection on infected clones to differ between high- and low-load patients.

Our method of analysis and quantification of retroviral integration sites is flexible. With minor adjustments, both the experimental and the analytical aspects of the method can be used for other retroviruses, retroelements or gene therapy vectors. It can be used for monitoring response to anti-retroviral treatment, for the analysis of mechanisms contributing to HIV latency, for tracking positive selection of gene therapy clones in treated patients and for identification of novel insertions due to reactivation of endogenous retroviruses in humans and animals.

For HTLV-1 research, this method has opened the door to multiple lines of research. It has recently been used to demonstrate there is a single provirus copy per cell in non-malignant infected $CD4^+$ cells (Cook et al., 2012), and there is a currently on-going project using high-throughput sequencing data of HTLV-1 integration site aimed at estimating the total number of infected clones in the blood, and to use that in order to model the ratio of infectious to mitotic spread in HTLV-1. Deep integration site analysis can be used as a guide in defining pre-malignant clones and the molecular biology mechanisms that may be involved in clonal transformation, as well as investigating the role of infected T-cell clones in the HTLV-1-

associated inflammatory diseases. Finally, this method will also be an indispensable tool to investigate the determinants of the rate of proviral gene expression and the balance of selection of infected clones during chronic infection, using a genome-wide approach which could lead to suggested mechanisms to be tested using molecular biology techniques.

Chapter 4 - The role of the genomic environment in determining proviral integration targeting and clonal expansion

4.1 introduction

4.1.1 Retroviral integration

Viral DNA integration is the process by which a double-stranded viral DNA molecule is inserted into the host genome, allowing expression of viral genes from the host chromosome. While certain DNA viruses can integrate into host DNA, in the case of retroviral infection, the integration into the host genome is a necessary step in the virus replication cycle, required for expression (Sakai et al., 1993) and persistence of the retroviral genome in the infected cell. The proviral integration is mediated by the integrase (IN) protein, processed from the gag-pro-pol open reading frame (Nam et al., 1988) and carried to the infected cell within the virion.

The reverse transcription step in the retroviral replication cycle results in the formation of a double-stranded linear DNA molecule, a direct copy of the RNA genome carried in the virion flanked by two repeated regions (long terminal repeats, LTR). The DNA is coated by host and viral proteins, composing the PIC (Bowerman et al., 1989). Whether the PIC requires a cell to be dividing in order to enter the nucleus depends on the virus: while lentiviruses can infect both dividing and non-dividing cells, most retroviruses, such as murine leukemia virus (MLV), require cell division for efficient infection (Roe et al., 1993; Lewis and Emerman, 1994; Naldini et al., 1996).

The integration process is composed of three main steps: processing, joining and repair. The 3’ end processing step, results in the removal of two nucleotides from either 3’end of the double stranded molecule by IN, resulting in a 5’ overhang in both sides. This is followed by a strand transfer step, where the IN forms a new phosphodiester bond between the viral 3’OH end and host cell DNA. This creates a gap at both ends of the integrated molecule which is thought to be completed by host DNA repair enzymes. The result is the duplication of a

number of bases at the integration site at either end of the integrated provirus. The integrated HTLV-1 provirus is flanked by a 6-base repeat of duplicated human genome (Seiki et al., 1983).

Different host genes have been found to play a role in integration. The *barrier-to-autointegration factor* gene was identified in Moloney murine leukemia virus as a component of the PIC, increasing the efficiency of integration by preventing auto-integration (integration of the LTR ends to sites within the viral DNA) (Lee and Craigie, 1998). The host proteins Hmga1, INI1 and LEDGF/p75 have also been identified to interact with the HIV PIC (Van Maele et al., 2006). Lens epithelium-derived growth factor (LEDGF/p75) has been well characterized as an integrase co-factor (Cherepanov et al., 2003), targeting the PIC, increasing integration efficiency and determining integration site distribution (Maertens et al., 2003; Ciuffi et al., 2005; Shun et al., 2007).

4.1.2 Integration site selection bias.

The site of retroviral integration can be important for both virus replication and for pathogenesis, because it may affect both viral and host gene expression. The genomic environment at the site of integration may affect the level of expression of proviral genes (Shan et al., 2011) leading to a latent infection (Siliciano and Greene, 2011).

Retroviral integration has the ability to alter host gene expression in cis through various mechanisms jointly referred to as insertional mutagenesis. Integration of the provirus can either induce aberrant expression through the introduction of the viral promoter or enhancer, or inactivate a gene through disruption of induction, expression or correct splicing (Uren et al., 2005). It is also likely that higher-order mechanisms could be affected, whereby the integration alters the 3 dimensional folding of the genome, causing long-range effects, for

example by integration in proximity to binding sites of host factors such as CCCTC-binding factor (CTCF), known to play a role in DNA looping (Hou et al., 2008)

The pattern of integration sites from in vitro and in vivo samples provides insight into the effect of proviral integration site on the survival of the infected clones: deleterious integration (for example through haploinsufficiency) would be subject to negative selection. In contrast, insertions which enable or promote cell proliferation in vivo may be found in expanded clones.

A startling example of the importance of testing for integration site bias was shown in the case of gene therapy vectors (Hacein-Bey-Abina et al., 2008), where the selective expansion of certain clones can result in adverse outcomes (Fehse and Roeder, 2008).

Integration sites can be found across the host cell genome, but they are not random. There is a weak nucleotide consensus target site favoured by retroviruses (Wu et al., 2005). On a broader scale, integration is often favoured within transcriptionally active regions. Retroviral integration site selection has been found to be associated with transcription units (Schroder et al., 2002; Wu et al., 2003), the transcriptional level of the genomic environment (Maxfield et al., 2005) and nucleosomal positioning (Roth et al., 2011).

Integration site selection differs between different retroviruses, thus the integration site choice cannot be simply explained by accessibility of chromatin to the integration machinery. Derse et al have shown that integrases closely related in peptide sequence also share similar integration site bias (Derse et al., 2007). This is consistent with the idea that closely related integrases would utilize similar host factors in order to direct the PIC to the chromatin and facilitate DNA integration. The host protein LEDGF/p75 has been identified as an HIV-1 integrase cofactor (Maertens et al., 2003) which has a significant impact on integration site targeting (Ciuffi et al., 2005; Shun et al., 2007). More recently, a second host factor, HRP-2,

was identified in LEDGF knock-out cells, and found to have an effect on directing integration site selection in those cells (Schrijvers et al., 2012; Wang et al., 2012). More recently, several bromodomain and extraterminal domain (BET) proteins, Brd2, Brd3 and Brd4, have been shown to directly bind MLV IN and knock-down of all three significantly reduced MLV integration at transcription start sites (Sharma et al., 2013). There are no known integrase cofactors in HTLV-1 infection.

4.1.3 HTLV-1 integration sites selection

HTLV-1 was first shown to have a non-specific integration (without a particular preferred site) by Seiki and colleagues (Seiki et al., 1984). With improvements in both the techniques used to identify HTLV-1 integration sites, and the understanding of the human genome, a more refined picture of integration site specificity emerges. The nucleotide sequence in the immediate vicinity of the integration site is not random but is biased towards a loose consensus sequence (Chou et al., 1996; Derse et al., 2007; Meekings et al., 2008). Using hybridization methods and LMPCR, it was shown that integration is biased towards regions of the target host genome with a high GC content (Zoubak et al., 1994; Derse et al., 2007), Work mostly done on integration sites from ATLL cases/cell lines showed a bias towards integration in or near genes (Ozawa et al., 2004; Hanai et al., 2004; Doi et al., 2005). Since these studies lacked the ability to quantify the clonal abundance of observed integration sites, it is difficult to assess how many of the bias differences observed between malignant and non-malignant cases were in fact differences between clones of low and high abundance. In more recent years, it was shown that in integration sites found in patients without malignancy, there is also a bias towards integration near transcription start sites and CpG islands, and that this bias is greater for integration sites found in infected patients (after years

of selection) than in cell lines infected in vitro (Derse et al., 2007; Meekings et al., 2008). The refined understanding of factors determining the bias in targeting of the integration site to the human genome is important; however, in order to understand the effect of long-term selection in the host it is necessary to address the association between the genomic environment and the survival and expansion of the infected clone to identify the factors that determine the success or failure of each clone.

The integration site can be used to identify each infected clone, because each mitotic replication results in two sister cells that share an identical integration site. The association between clonal distribution and disease has been studied in ATLL (Yoshida et al., 1984), HAM/TSP (Furukawa et al., 1992) and co-infected patients with Strongyloides or infectious dermatitis (Gabet et al., 2003; Gabet et al., 2000). Clonality was initially visualized by southern blot (Yoshida et al., 1984), by counting bands on gel electrophoresis following IPCR (Takemoto et al., 1994) which was able to distinguish crudely between patients with a "monoclonal" or "oligoclonal" pattern of integration, in whom one or several bands were visible on the gel, and "polyclonal patients" in whom no discrete band was detected. The use of classical sequencing following LMPCR (Wattel et al., 1995) allowed the isolation of a larger number of integration sites, but the quantification of integration site frequency was probably biased by preferential PCR amplification of short products (see section 3.3.3).

The development of high-throughput sequencing method discussed in this work by Dr. Nicolas Gillet and myself has opened the door to a deeper, more refined understanding of the role of the genomic environment in determining integration site and long term clonal abundance, in malignant and non-malignant cases; in disease versus healthy carriers; and by suggesting further mechanistic avenues of research into the role of the proviral expression, integration site and the immune selection in determining clone size and proviral load.

4.1.4 Aim

In order to understand any association between the genomic environment flanking the integrated provirus and the proviral expression, it is imperative to refine our understanding of the attributes that are associated with de novo proviral integration, as well as the factors which promote or inhibit long-term clonal expansion. In this work I used previously published as well as unpublished integration site data with the aim to test the hypothesis that the HTLV-1 provirus integration site is favoured near particular genomic elements, and whether the effect of the genomic environment on the clonal expansion is symmetrical with respect to the integrated provirus.

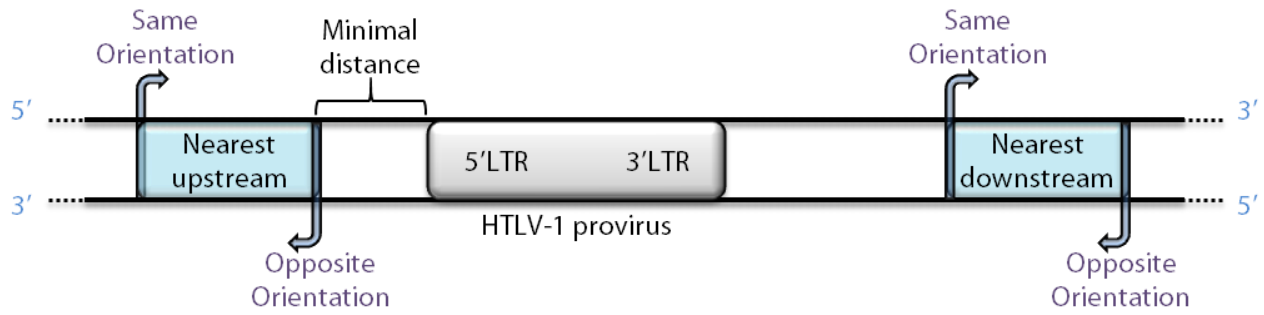

Figure 4.1: Genomic environment flanking the integrated provirus

Blue blocks denote a genomic feature (e.g. transcription start site). 'Upstream', 'Downstream' and relative orientation are defined relative to the sense strand of the HTLV-1 provirus. The distance to the nearest genomic feature is calculated separately for upstream (closer to the 5' LTR) and downstream features, and unless otherwise stated, calculated as the distance between the integration site and the nearest end of the genomic feature.

4.2 Results

4.2.1 The genomic environment flanking the proviral integration site

A schematic description of the genomic environment flanking the integration site is shown in Figure 4.1. We defined ‘upstream’ and ‘downstream’ with respect to the integrated provirus: ‘upstream’ denotes elements nearer to the 5’LTR end of the provirus; ‘downstream’ denotes elements in closer proximity to the 3’ end of the provirus. Therefore – ‘upstream’, ‘downstream’, ‘same’ and ‘opposite’ sense are all defined with respect to the plus strand proviral open reading frames (which include the proviral transactivator Tax). For each genomic feature (e.g. transcription start site) we defined the minimal distance to the integration site as the difference between the genomic coordinates of the nearest respective feature (either upstream or downstream) and the integration site.

4.2.2 In vitro infection as a model of initial integration targeting preferences.

Details of datasets used are summarized in Table 3.1. In order to examine the initial integration site preference we infected Jurkat T-cells by a short co-culture with lethally irradiated MT2 cells (HTLV-1-producing cell line). After 2 weeks (to allow elimination of residual MT2 cells) the integration sites were analysed using our high-throughput method for identification and quantification of retroviral integration sites (Chapter 3). Two independent experiments were carried out, using the same protocol. No significant differences were found between the two datasets (Figure 4.10), and they were therefore pooled to form one combined in vitro dataset.

In order to examine the genomic integration bias during chronic infection (after decades of infection and clone selection) we used recently published integration site data (Gillet et al., 2011) – that were obtained using the identical high-throughput protocol described here: these data included integration sites analysed from PBMC samples from 63 subjects with 3 clinical manifestations of HTLV-1 infection: ACs, HAM/TSP and ATLL (Table 4.1). Where more than one timepoint was available per patient, the most recent timepoint was used. Patients with more than one clinical manifestation (i.e. diagnosed both with ATLL and HAM/TSP) were excluded. The abundance of each clone was quantified as the number of copies of that clone per 10000 PBMCs, and the clones were divided into bins according to their abundance using logarithmic scale delimiters (Table 4.2)

Table 4.1: In vivo integration sites – sample data by patient (Gillet et al., 2011)

Clinical Status[a]	Patient code	Gender	PVL[b]	Sequencing results		
				# sequencing reads	# proviruses	# clones
AC	HAY	F	3.6	88587	8034	2070
AC	HBE	F	3.6	7860	1400	888
AC	HBF	F	7.6	264490	3536	1615
AC	HBK	F	8.9	232565	4284	980
AC	HBT	F	0.5	12183	295	126
AC	HBX	F	2.7	42889	1607	642
AC	HBY	M	1.6	10377	540	422
AC	HCM	F	0.6	134303	216	55
AC	HCS	M	5.4	35281	257	76
AC	HDA	F	0.3	10290	133	89
AC	HDG	M	2.7	90212	2776	1042
AC	HDR	F	2.7	44265	1252	471
AC	HDS	F	2.5	18010	1775	952
AC	HES	F	17.7	31271	3574	1981
AC	HEZ	F	3.4	44252	3753	1935
AC	HFE	F	6.5	12767	1298	702
AC	HFG	F	1.6	44436	940	412
AC	LFP	F	0.6	59035	343	242
ATLL	AN	M	24.3	28545	2324	45
ATLL	C3	F	64	15560	2969	5
ATLL	C4	F	20.4	64408	10315	30
ATLL	HCG-LEY	M	8	71393	3312	2039
ATLL	HDM-LFK	F	6.3	38787	2513	637
ATLL	JH	F	31.2	76432	14639	107
ATLL	KD5	F	21.2	39703	2065	10
ATLL	LEP	M	21.2	78233	10063	276
ATLL	LER	M	52.3	7017	1053	7
ATLL	LEU	F	45.5	19440	6656	6
ATLL	LEZ	M	26.7	86061	2843	243
ATLL	LFA	F	10.6	30600	4293	1025
ATLL	LFC	F	12.3	29217	5661	2426

a AC: Asymptomatic Carriers; ATLL: Adult T-cell Leukemia/Lymphoma; HAM/TSP: HTLV-1-Associated Myelopathy / Tropical Spastic Paraparesis.
b PVL : proviral load – number of copies per 100 cells

Clinical Status[a]	Patient code	Gender	PVL[b]	Sequencing results		
				# sequencing reads	# proviruses	# clones
ATLL	LFE	F	7.4	28028	3768	1258
ATLL	P7	F	18.3	39283	12980	3182
ATLL	S1	F	18.1	27415	9138	829
ATLL	S2	F	10.8	7531	1338	234
ATLL	S4	M	16.7	14621	1982	629
ATLL	S5	F	3.6	24410	4391	53
ATLL	S6	M	15	24823	5880	282
HAM-TSP	TAA	F	2.2	70585	5995	1805
HAM-TSP	TAL	F	8.7	106366	12839	3024
HAM-TSP	TAN	F	4	21024	2083	1388
HAM-TSP	TAS	F	3	13485	1478	901
HAM-TSP	TAT	F	9.6	10525	2204	1281
HAM-TSP	TAW	F	1.4	94529	3528	1006
HAM-TSP	TAY	F	1.4	66941	3582	1069
HAM-TSP	TAZ	M	12.5	13636	3545	1965
HAM-TSP	TBA	F	9.6	27784	2679	1215
HAM-TSP	TBC	F	20	84806	26108	7460
HAM-TSP	TBG	F	14.6	76744	16617	3609
HAM-TSP	TBJ	F	15.8	16907	2791	1442
HAM-TSP	TBO	F	7	13519	1904	1066
HAM-TSP	TBP	M	3.5	36279	5855	3604
HAM-TSP	TBR	F	4.7	47148	2949	1461
HAM-TSP	TBS	F	10.4	21127	5216	2396
HAM-TSP	TBU	F	1.1	70258	1846	749
HAM-TSP	TBW	M	7.3	124980	13240	3136
HAM-TSP	TCG	F	3.6	28031	1678	766
HAM-TSP	TCO	F	1.8	17903	789	497
HAM-TSP	TCQ	M	5	44629	3830	2155
HAM-TSP	TCR	M	5.9	38437	3606	2048
HAM-TSP	TCT	F	5.5	26719	3407	1943
HAM-TSP	TDA	M	6.2	7125	1165	805
HAM-TSP	TW	F	8.9	112958	16420	3749

Table 4.2: In vivo integration sites (Gillet et al., 2011) – sample data by clone abundance

		Clone numbers by absolute abundance bin[a]				
Clinical Status[b]	**No. Patients**	**<0.1**	**0.1-1**	**1-10**	**>10**	**Total**
AC	**18**	3201	10275	1192	32	**14700**
ATLL	**20**	42	11627	1587	67	**13323**
HAM/TSP	**25**	14234	33521	2748	37	**50540**
Grand Total	**63**	**17477**	**55423**	**5527**	**136**	**78563**

4.2.3 Integration sites are favoured within genes at the same orientation as the host gene.

As previously reported (Gillet et al., 2011) 47% of integration sites are found within genes. This indicated that HTLV-1 integration is somewhat favoured within genes (compared to 43% of random sites, $p<10^{-11}$, X^2 test). As expected by chance, ~50% of proviruses integrated within host genes lay in the same transcriptional orientation as the host gene (Figure 4.2).

While at the level of primary infection (in vitro) both integration in the same and the opposite orientation are favoured, integration sites found in infected subjects (in vivo) contained fewer integration events in the opposite orientation to the host gene than would be expected by chance and the bias towards integration in the same orientation increased with clone abundance.

a Absolute abundance – number of proviruses per 10000 PBMC
b AC: Asymptomatic Carriers; ATLL: Adult T-cell Leukaemia/Lymphoma; HAM/TSP: HTLV-1-Associated Myelopathy / Tropical Spastic Paraparesis

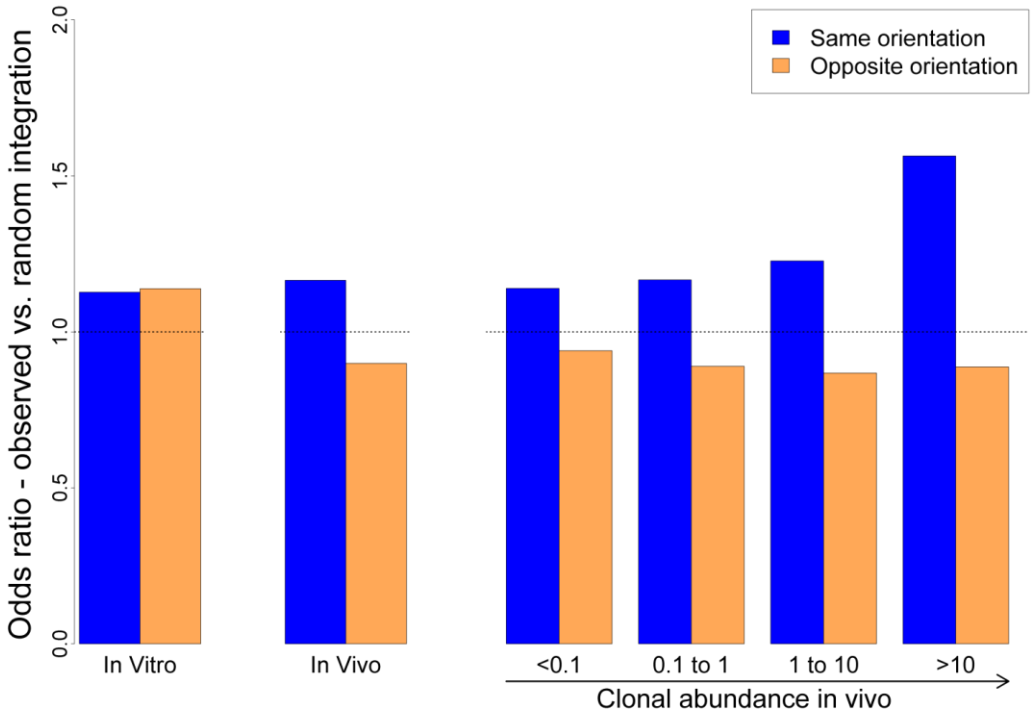

Figure 4.2: HTLV-1 integration within transcription units.

Clonal abundance was quantified by absolute abundance – number of copies per 10000 PBMCs. The proportion of HTLV-1 integration sites within transcription units was compared to random expectation. Integration within genes was favoured for initial infection (in vitro) but during chronic infection was disfavoured for integration in the opposite orientation (in vivo). The increased frequency of integration was higher with increasing clonal abundance. The relative orientation is with respect to the sense strand of HTLV-1.

4.2.4 Bias towards integration in proximity to transcription start site (TSS) is asymmetric

Gillet et al reported (2011) an association between clonal abundance and integration within 10kb of a TSS. In the present work I examined the genomic features in the immediate vicinity of the integration site, either upstream or downstream of the integration site. I found (Figure 4.3) that the strongest bias was at a distance of 1kb between the integration site and the nearest TSS (highest odds ratio compared to random sites). We refined our analysis to test whether the effect observed was directional. The results showed (Figure 4.3) that same-sense transcriptional orientation (both upstream and downstream) was favoured to a similar extent in both in vitro and in vivo integration sites. However in the opposite orientation, this bias was weaker (closer to random expectation) in the in vivo sites compared to the in vitro sites. At the distance in which the integration bias was the highest – 1kb from the integration site – the positive trend towards integration in proximity to TSS with increasing clonal abundance (previously reported) was only maintained for the integrations at the same orientation (i.e., for those proviruses integrated in the same transcriptional orientation as the nearest host TSS); in particular this was true for those integrated with a nearby downstream TSS rather than upstream (Figure 4.4).

While the proportion of integration sites in proximity to TSS was similar between in vitro and in vivo sites in the same orientation, it was reduced in vivo for the opposite orientation.

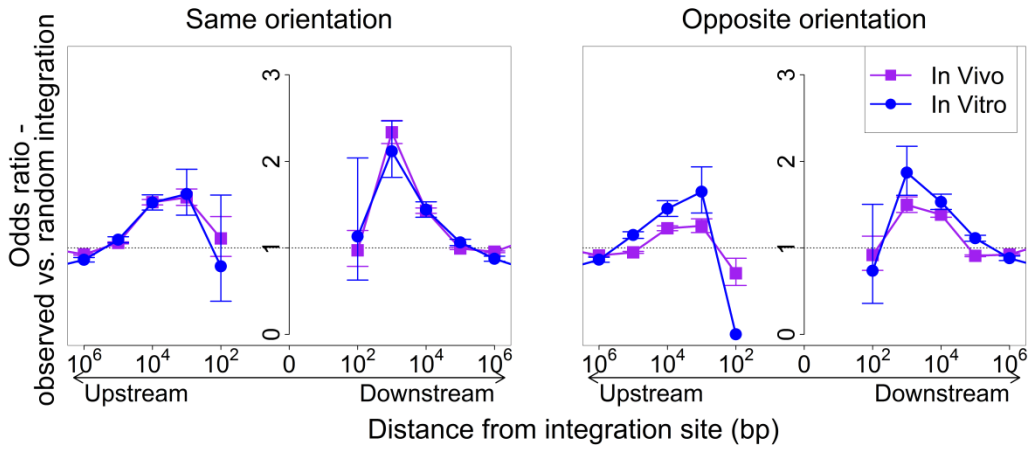

Figure 4.3: HTLV-1 proviral integration in proximity to transcription start sites (TSS).
Mountain plot –the X axis denotes the distance from the integration site to the nearest TSS (on a log scale, positive values represents downstream from the integration site, negative values represent upstream from the integration site). Windows are non-cumulative. Y axis denotes the odds ratio (compared to random sites) of having a TSS within this distance of an integration site. Whiskers denote the standard error. Peak preference (highest odds ratio) both for same and opposite orientation was at 1kb of the integration site. 'Upstream', 'Downstream' and the relative orientation are with respect to the sense strand of HTLV-1.

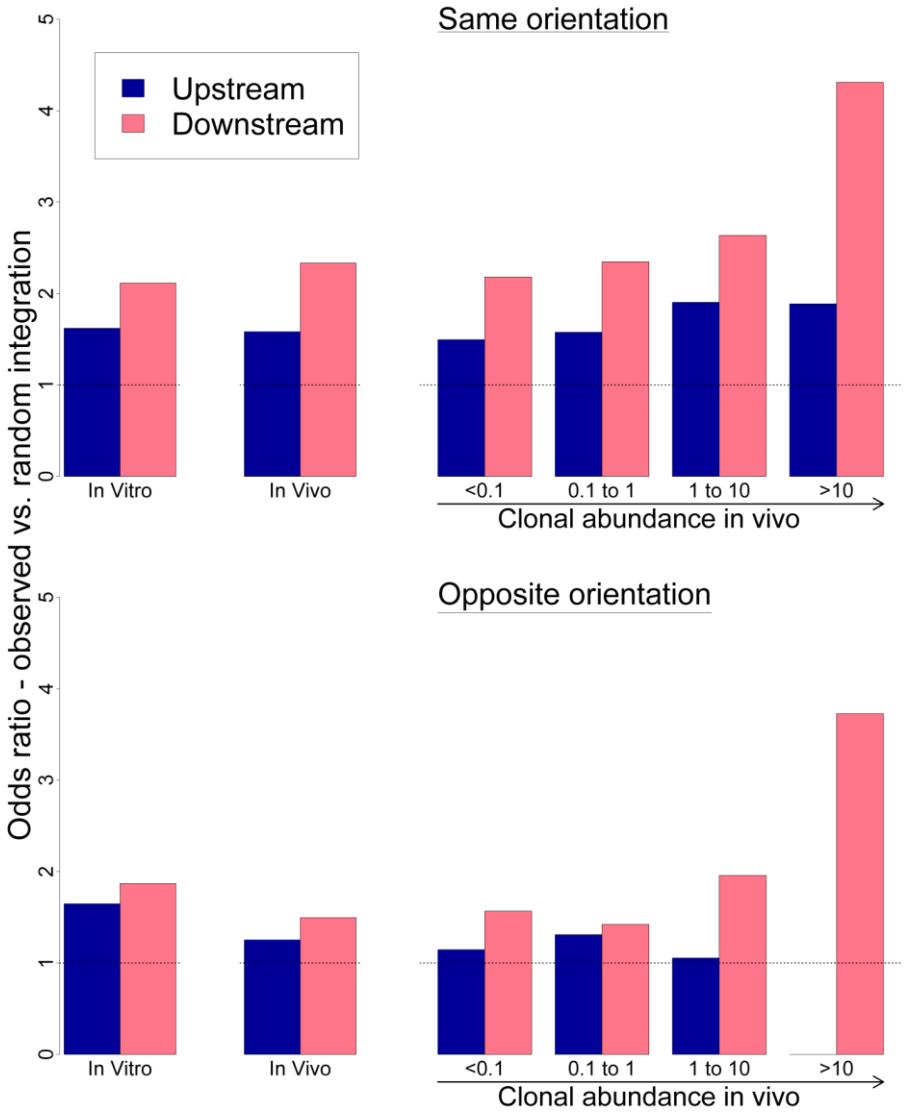

Figure 4.4: Proviral integration within 1kb of a transcription start site (TSS).

Absolute clonal abundance was quantified by absolute abundance – number of copies per 10000 PBMCs. The proportion of proviral integration sites within 1kb of a TSS was compared between integration sites (in vitro and in vivo) and with random expectation. When the nearest TSS was in the same orientation as the integrated provirus, the odds ratio was similar between in vitro and in vivo sites, and increased with clone size. 'Upstream', 'Downstream' and the relative orientation are with respect to the sense strand of HTLV-1.

4.2.5 Integration sites are favoured in proximity to CpG islands in an asymmetric fashion

It was previously reported that integration is favoured within 10 kb of a CpG island (Meekings et al., 2008) and that this bias is greater with increasing clonal abundance (Gillet et al., 2011). Here I extended this analysis and found that the strongest bias (highest odds ratio) both in vivo and in vitro was at a distance of 1kb from the nearest CpG island (Figure 4.5). At this distance, the trend towards stronger bias in larger clones was observed mostly for integration with a CpG island downstream (Figure 4.6).

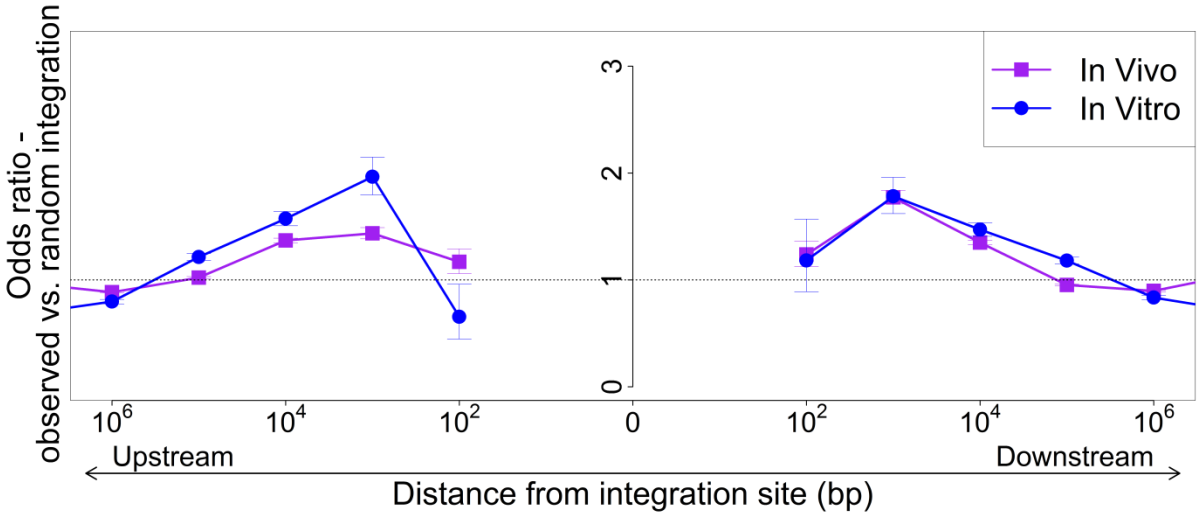

Figure 4.5: HTLV-1 integration in proximity to CpG island.
Integration was favoured in proximity to CpG islands. Peak bias (highest odds ratio) was at within 1kb from integration site, and then decline back to random expectation. 'Upstream', 'Downstream' are with respect to the sense strand of HTLV-1.

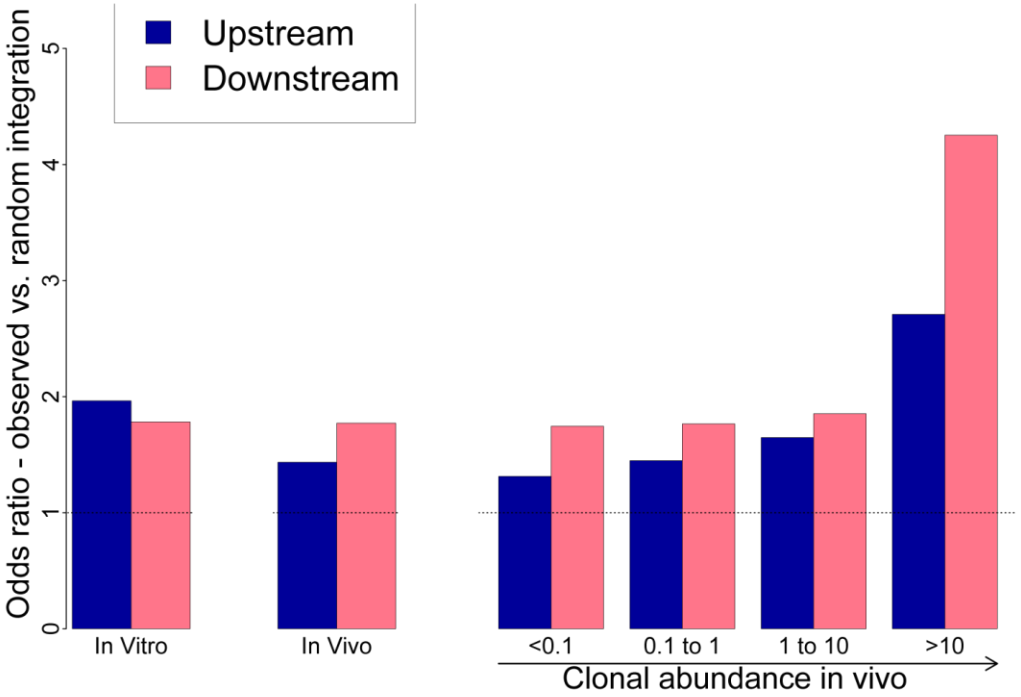

Figure 4.6: Integration within 1kb of a CpG island

Integration with a CpG island downstream was favoured in vivo, with an increase in bias with increased clone size. Clonal abundance was measured by copies per 10000 PBMCs. 'Upstream', 'Downstream' are with respect to the sense strand of HTLV-1.

4.2.6 Integration sites are symmetrically favoured in proximity to TFBS

Since proviral integration is biased towards transcriptionally active regions of the genome, we wished to test the hypothesis that proviral integration in proximity to transcription factor binding sites (TFBS) is associated with integration targeting, proviral expression or clonal abundance. The TFBS data used was publicly available datasets of genome binding positions of transcription factors, chromatin regulators and histone modifiers isolated in ChIP-seq experiments. Where available we used data from primary $CD4^+$ cells; otherwise, we used data from Jurkat or other human cell lines. Where primary data was available, we used the SISSRs Perl script (Jothi et al., 2008) which is designed to predict the site of transcription factor binding; otherwise, reported peak data was used. See Table 4.3 for complete listing of the datasets used.

Both in vitro and in vivo integration sites showed a bias towards integration in proximity to TFBS in general: 28% of in vitro and 25% of in vivo sites (compared with 20% of random sites) lay within 1kb of at least one of the TFBS tested (Figure 4.7). However, when examining the TFBS separately, both in vitro and in vivo there was a remarkably strong bias (high OR) towards integration in proximity to specific TFBS, in particular certain histone acetylases (HATs) , HDACs, STAT-1 and p53 ChIP-seq sites.

Two interesting patterns emerged in this analysis. First, the bias towards integration in proximity to most TFBS (e.g. p300, Figure 4.8B) was largely symmetrical, there was a similar odds ratio upstream and downstream of the integration site. In contrast, the bias towards integration near certain TFBS (e.g. STAT-1, Figure 4.8D) was asymmetrical, showing a higher bias towards one direction (often downstream). Second, for some of the

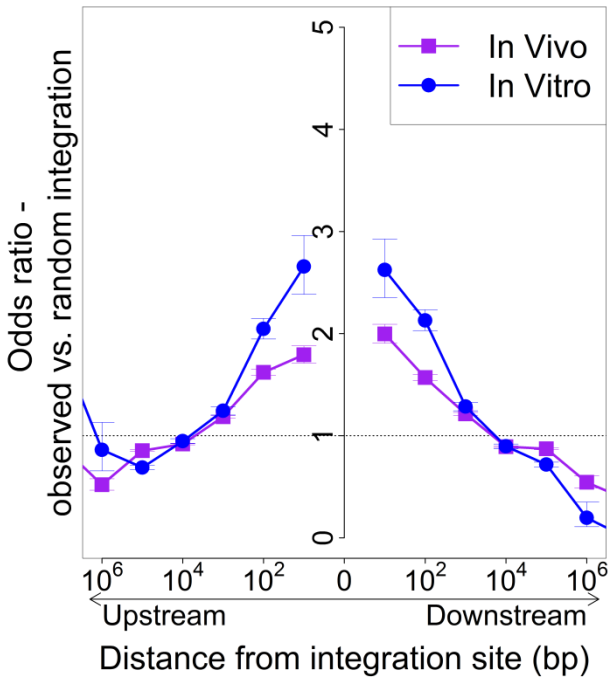

Figure 4.7: HTLV-1 integration in proximity to any TFBS tested.

Integration in proximity to at least one of the TFBS tested (see Table 4.3) was favoured. Higher odds ratio was observed for in vitro sites than in vivo. 'Upstream', 'Downstream' are with respect to the sense strand of HTLV-1.

transcription factors, we observed a sharp decrease at 10 bases from the integration site (e.g. STAT-1, Figure 4.8D). This pattern was consistently observed in multiple in vitro and in vivo datasets (Figure 4.10 and Figure 4.11, respectively), suggesting that this is a real effect rather than simply caused by the smaller number of integration sites immediately adjacent to the TFBS.

TFBS have been observed to co-localize in the human genome (Dunham et al., 2012), therefore we wished to test which TFBS of those studied were associated with integration independently from the other TFBS. A multivariate analysis approach was utilized (Figure 4.9). First, a likelihood ratio test was used to compare between two alternative analyses for each TFBS in order to test whether the integration was selectively associated with either upstream or downstream integration: a univariate test for both upstream and downstream and a bivariate test comparing upstream vs downstream, and subsequently tested each individual annotation separately as an explanatory variable using a univariate model. All factors that were significant in this univariate analysis (p. value < 0.05 after correction for multiple comparisons) were combined in a multivariate model, using a step-down approach, until only independent significant factors ($p < 0.05$) remained. Two separate multivariate models were made, respectively, to identify TFBS within 100 bases and within 1000 bases of the integration site (Figure 4.9). Most factors that were independently associated with integration targeting were symmetrically associated around the integration site. The factors which had the highest odds ratio were the transcription factor p53 and Histone deacetylase 6.

When comparing clones of different clonal abundance, it was observed that TFBS that were associated with integration targeting were often associated with small (low-abundance) clones (Figure 4.12).

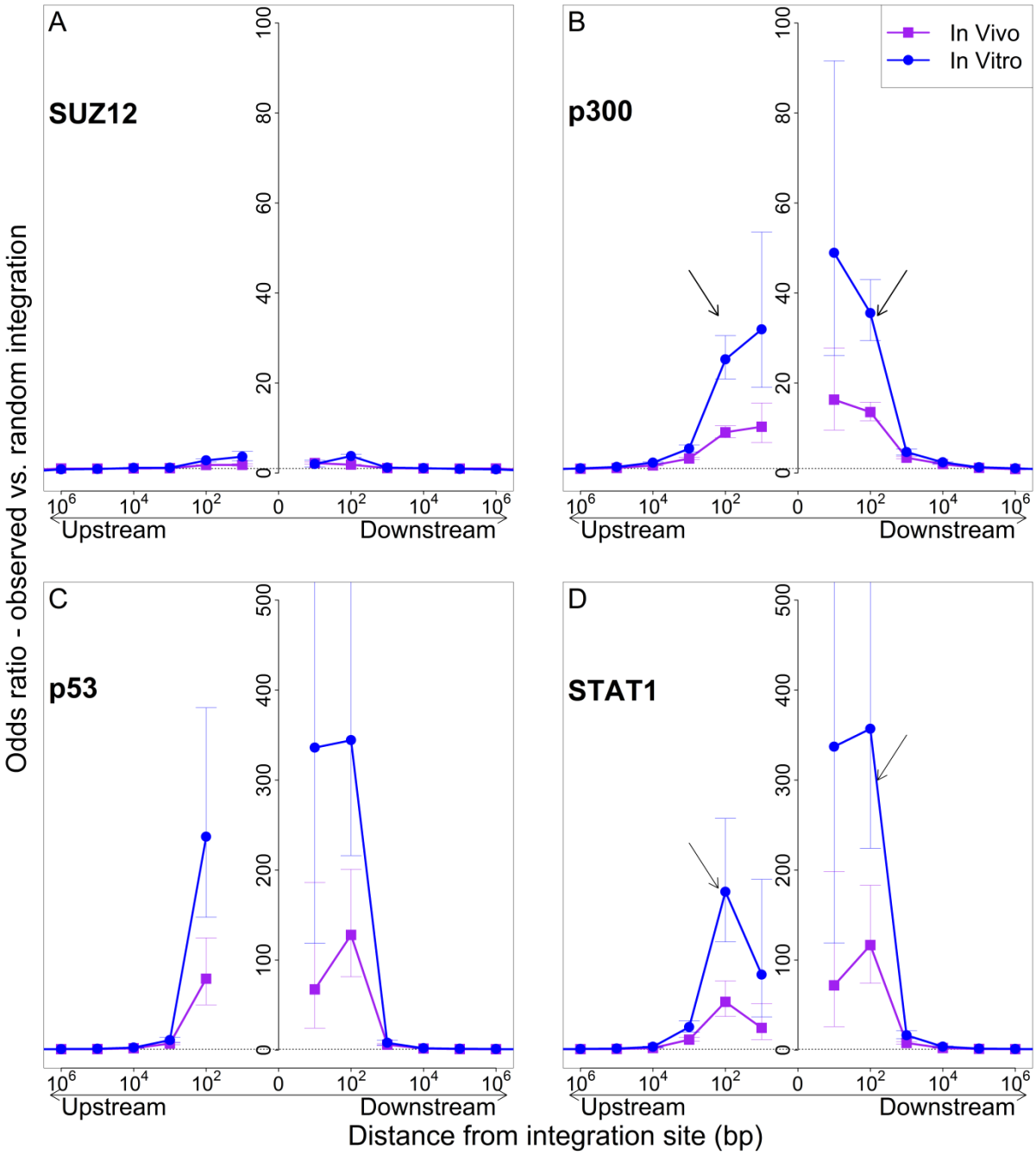

Figure 4.8: HTLV-1 integration in proximity to transcription factor binding sites.

Integration site targeting was associated with close proximity to certain transcription factor binding sites. Notable selected TFBS are shown here, see Table 4.3 for a list of the maximum odds ratio for each TFBS tested. Differences in preference between in vitro and in vivo remained the same even when randomly sampling the same number of in vivo sites as are in vitro sites (not shown), therefore we exclude the possibility that the difference is due to differences in clone numbers. 'Upstream', 'Downstream' are with respect to the sense strand of HTLV-1.

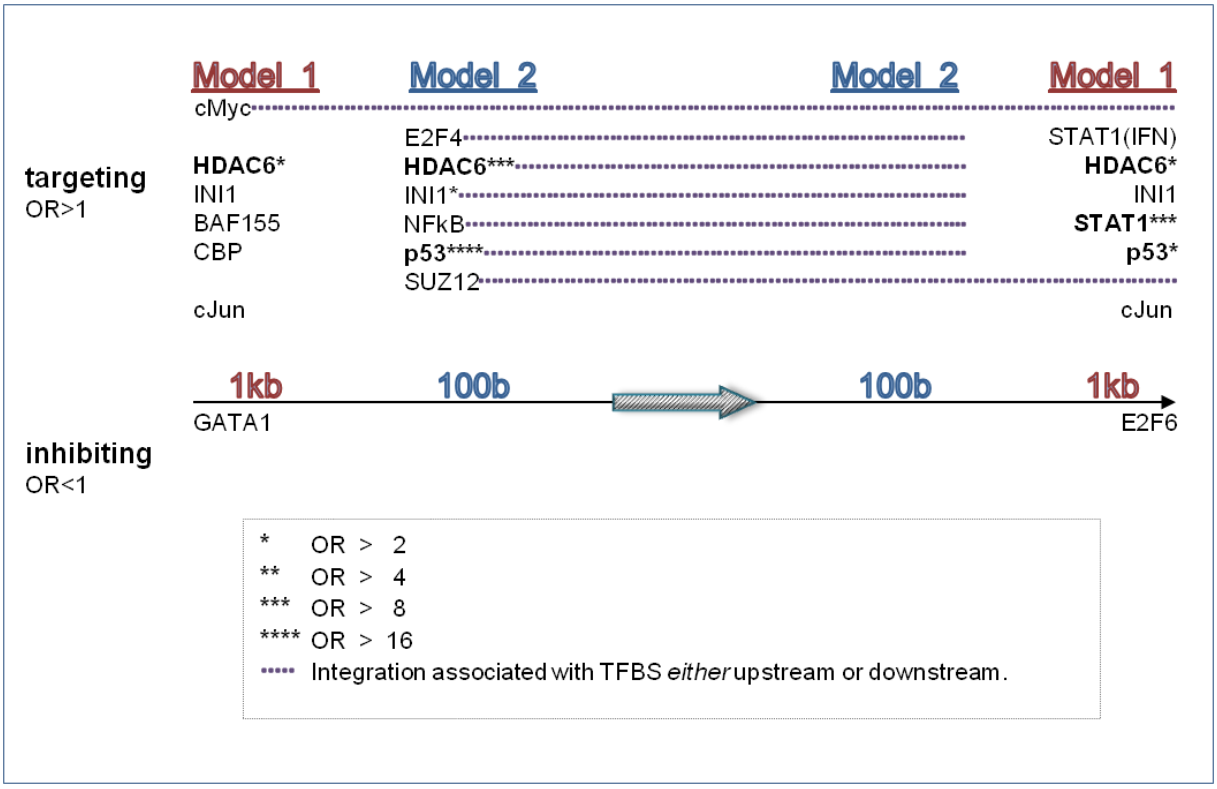

Figure 4.9: HTLV-1 integration in proximity to TFBS - multivariate analysis

TFBS (ChIP-seq identified binding sites for transcription factors, histone modifiers and chromatin regulators) independently associated with integration of HTLV-1 was identified using multivariate analysis. OR – odds ratio. Outcome measure was 'True' integration site. TFBS shown above the line were associated with an excess frequency of integration compared with random (OR>1); TFBS below the line were significantly less likely to lie near the provirus (OR <1). Model 1 and Model 2 (carried out independently) test for TFBS within 1kb and 100bp of integration site, respectively. Arrow denotes orientation of integrated provirus. Dotted line – TFBS were associated with integration either upstream or downstream.

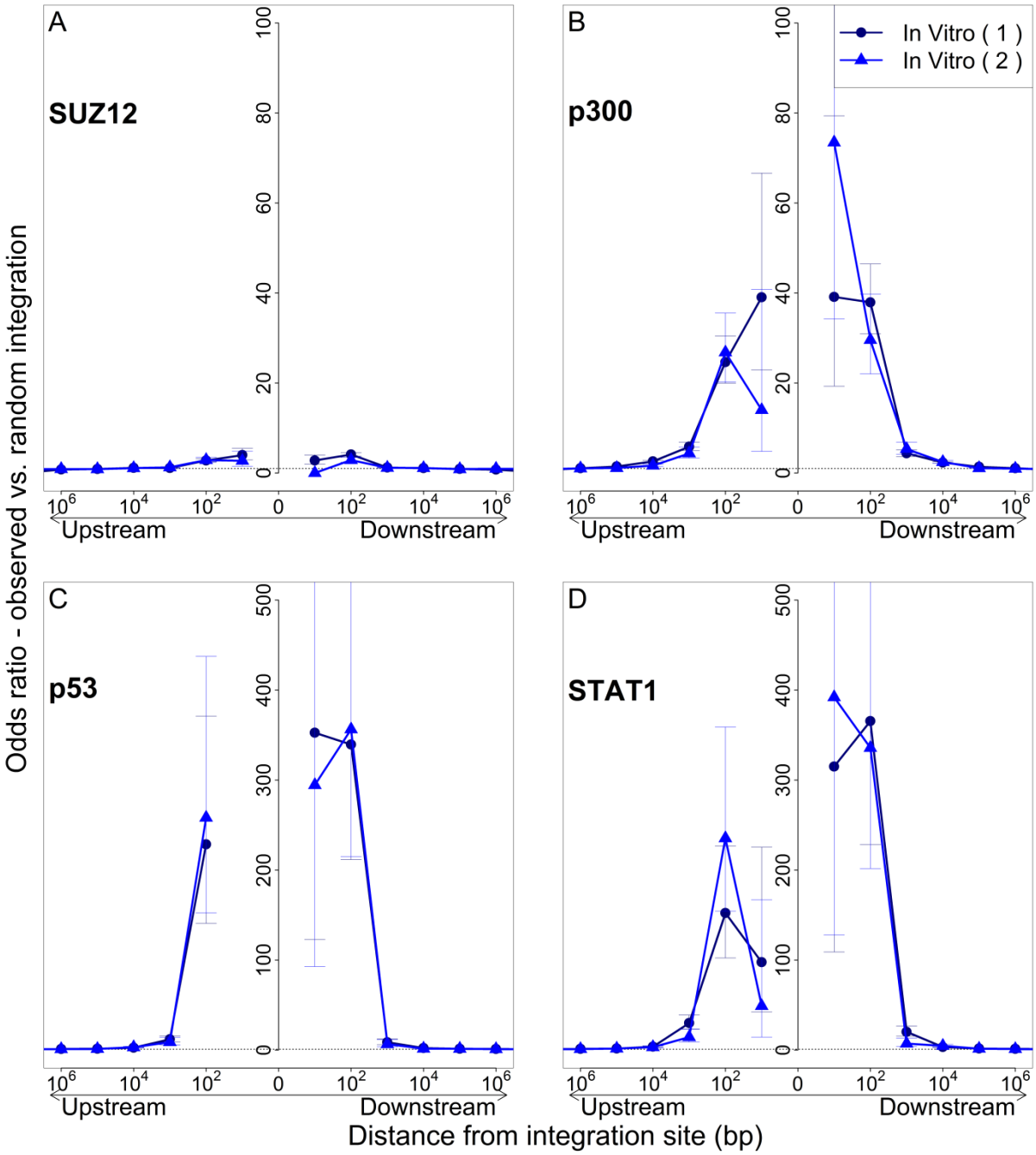

Figure 4.10: Bias towards integration in proximity to TFBS is consistent across datasets.
Pattern of integration preference in proximity to transcription factor binding sites was consistent between different datasets resulting from independent experiments. 'Upstream', 'Downstream' are with respect to the sense strand of HTLV-1.

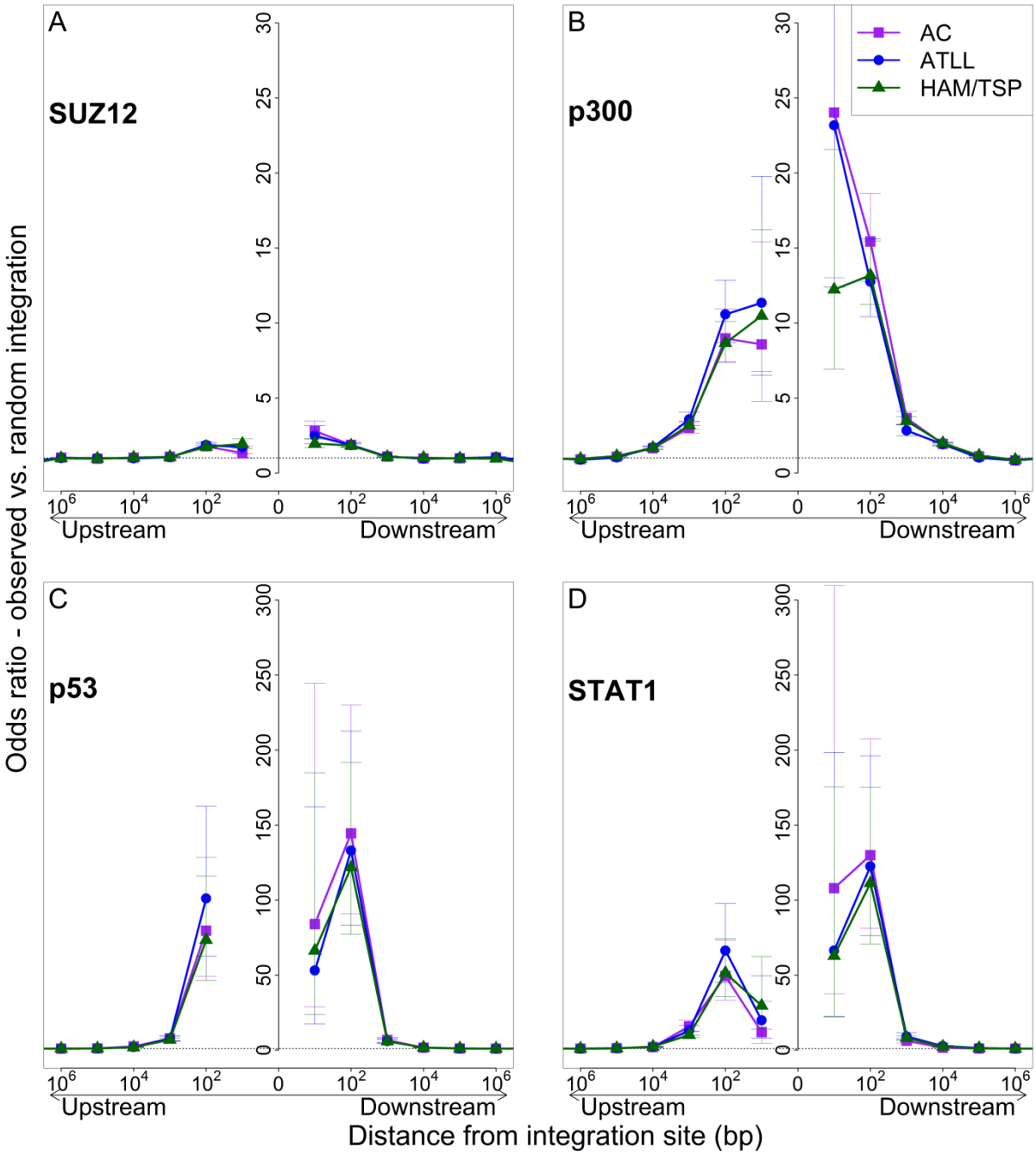

Figure 4.11: Bias towards integration in proximity to TFBS similar between clinical outcomes

Pattern of integration in proximity to transcription factor binding sites was similar between clones identified from patients of different clinical subtypes. 'Upstream', 'Downstream' are with respect to the sense strand of HTLV-1.

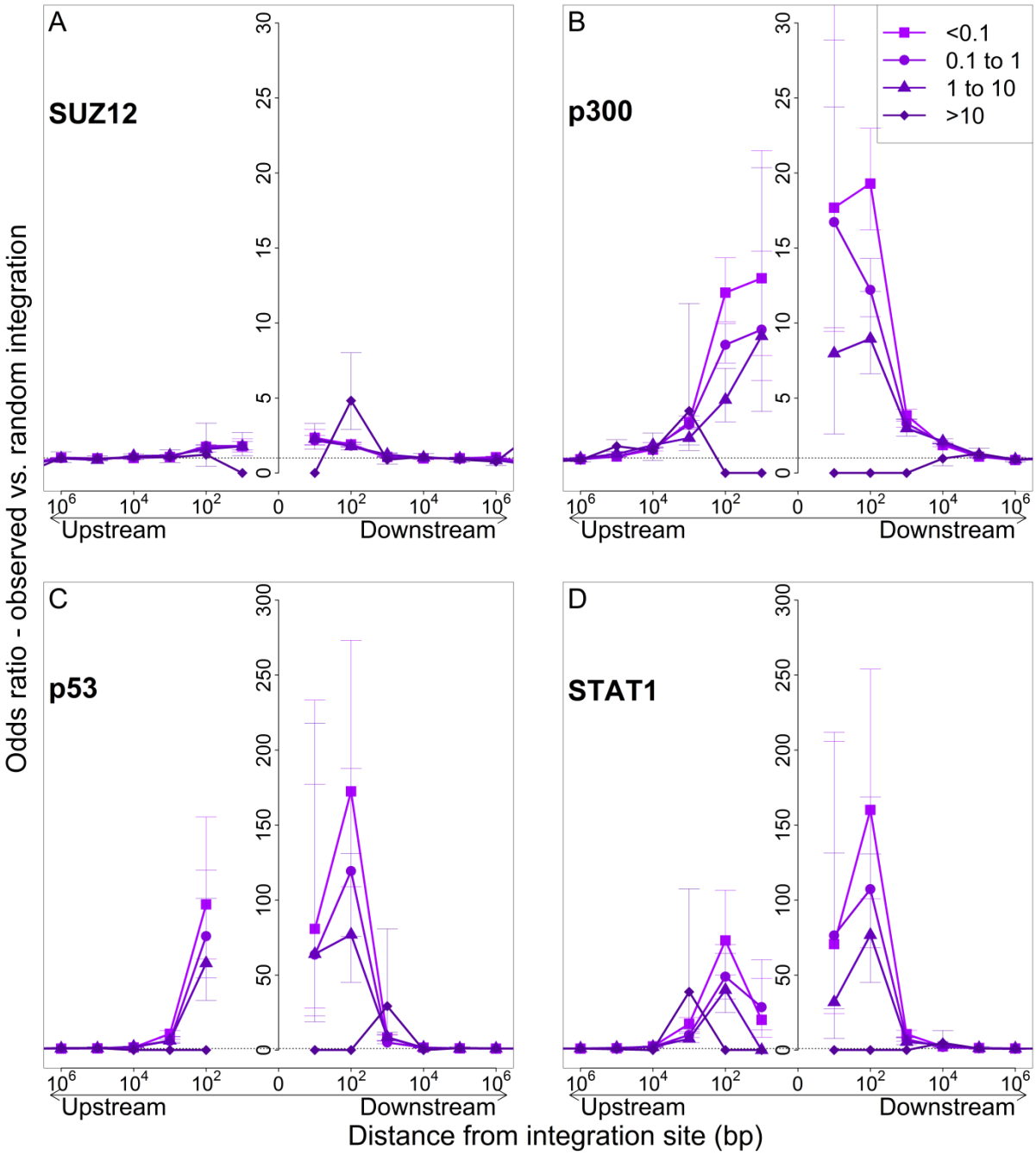

Figure 4.12: Integration in proximity to TFBS is most associated with smallest clones.

Bias towards integration in proximity to TFBS was higher for less expanded clones. 'Upstream', 'Downstream' are with respect to the sense strand of HTLV-1.

Table 4.3 Human genome annotations used in this work

Annotation	Functional category	Cell type	Reference (see below)	peak calling algorithm	Max. OR
RefSeq genes		N/A	Fujita et al., 2011	N/A	
CpG islands					
NRSF (REST)	Transcription factor / silencer	Jurkat	Jothi et al., 2008; Johnson et al., 2007	SISSRs	167.57
CTCF	Insulator / regulation of chromatin architecture	pr CD4	Jothi et al., 2008; Barski et al., 2007	SISSRs	12.89
BRG1	Chromatin regulators	HeLa	Euskirchen et al., 2011	Peak-Seq	2.31
INI1					2.75
BAF155					2.62
BAF170					1.81
cJun	Transcription factors	K562	Raha et al., 2010	Peak-Seq	2.33
cFos		GM12878	Rozowsky et al., 2011	SISSRs	39.15
cMyc					8.25
JunD					30.49
E2f4		GM06990	Lee et al., 2011	See ref	2.99
E2f6		K562	Trojer et al., 2011	SISSRs	7.45
Foxp3		Act. CD4	Birzele et al., 2011	MACS	5.21
		Treg			6.98
GATA1		K562	Fujiwara et al., 2009	SISSRs	19.21
GATA2					17.79
NFkB		GM12878	Kasowski et al., 2010	Peak-Seq	2.62
STAT1		Act. CD4	Liao et al., 2011	SISSRs	335.45
STAT1IFN		IFNγ Stim. CD4			8.18
Yy1		K562	Myers lab for the ENCODE project. GEO Accession code: GSM803470	MACS	2.28

Annotation	Functional category	Cell type	Reference (see below)	peak calling algorithm	Max. OR
Znf263		K562	Frietze et al., 2010	Sole-Search	3.66
CBP	Histone modification enzymes	pr CD4	Wang et al., 2009	SISSRs	11.43
p300					48.88
MOF					31.93
Tip60					33.45
PCAF					34.90
HDAC1					14.30
HDAC2					93.15
HDAC3					223.49
HDAC6					91.80
p53	Transcription factor	IMR90	Botcheva et al., 2011	SISSRs	343.70
Rad21	Cohesin component	GM12878	Myers lab for the ENCODE project. GEO accession code: GSM803416	MACS	1.64
SUZ12 (PRC2)	Chromatin regulators	K562	Ram et al., 2011	See ref	3.76

Act. Activated; Stim. Stimulated; Max. maximum. GEO: Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/)

4.3 Discussion

The mechanisms dictating HTLV-1 proviral targeting and long term survival of infected cell clones are not fully understood. We have developed a method which is able to identify and quantify an unprecedented number of proviral integration sites, giving us for the first time a deep insight into potential correlates of proviral integration, which will lead to further mechanistic testing of proviral integration. In order to elucidate the effect of the genomic environment on initial proviral integration of HTLV-1, we utilized the previously used in vitro infection model for initial infection. Infected cells were kept in culture for no more than two weeks (required for the elimination of any residual cells from MT2, the infecting cell line). In this setting, selection is minimal and should represent the physiological state of infection prior to induction of the adaptive immune response and the long term effects of clonal expansion during chronic infection. We mapped the integration sites from two independent such experiments.

4.3.1 Integration bias within and in proximity to transcription units

The results confirmed the previously observed small bias towards integration within or in proximity to host transcription units (Meekings et al., 2008; Gillet et al., 2011). This bias towards integration within genes is considerably lower than that observed in HIV infection (~70%, Ciuffi et al., 2005), suggesting a fundamental difference in the mechanism determining integration targeting between the two viruses. In HIV-1 infection, LEDGF/p75 is known to have an important role in targeting HIV integrase to the host DNA (Maertens et al., 2003), and in its absence, the preference towards integration within genes is diminished (Ciuffi et al., 2005). There was no bias in the relative transcriptional orientation of the host gene and the provirus in in vitro infection, suggesting that the bias towards same-sense

integration observed in integration sites mapped from infected individuals is the result of long-term selection acting in vivo.

The bias towards integration in proximity to CpG islands, a marker of a transcriptionally active genomic region, was also higher with increasing clonal abundance, in particular when the CpG island lay downstream of the integration site. This is consistent with another recent observation (Gillet et al., 2011) that integration near a high density of activating histone marks was correlated with clonal abundance, suggesting that an overall active genomic environment is conducive to clonal expansion.

4.3.2 Integration bias in proximity to TFBS.

We observed a weak bias towards integration in proximity to any TFBS, however when examining each TFBS in turn, the bias observed was remarkably strong in certain cases (e.g. STAT1, NRSF) in single factor analysis. The bias identified was consistent across several in vitro and in vivo datasets, and was stronger for in vitro integration than was for in vivo sites. In addition, the bias towards integration in proximity to these TFBS was higher for clones of lower absolute abundance, suggesting that these TFBS do not confer an advantage on the infected clones during chronic infection but rather may be more important during initial integration. Since different TFBS can cluster in particular areas of the genome (Dunham et al., 2012), we also carried out a multivariate (logistic regression) analysis, considering all of the TFBS examined, in order to identify TFBS which were independently associated with proviral integration targeting. The results (Figure 4.9) confirm HDAC6, STAT1 and p53 as significant independent correlates of integration, and also included INI1, NFkB and cMyc. STAT1 and p53 are known to play a significant role in HTLV-1 infection. HTLV-1 infection dysregulates p53 related pathway gene expression (Tattermusch et al., 2012). Interferon-

stimulated genes, including STAT1, are known to be upregulated by HTLV-1 in vivo, in particular in HTLV-1 associated disease (Tattermusch et al., 2012). It is not yet known whether insertional mutagenesis may play a role in these processes.

STAT1 has been reported to be highly associated with proviral integration of MLVs (Santoni et al., 2010). The authors attributed this effect to the presence of certain epigenetic histone marks (H3K4me3, H3K4me1 and H3K9ac) in proximity to the integration site.

The association of HTLV-1 integration to TFBS was typically symmetrical and short range (within 100b of integration site), falling to random expectation beyond 1kb from the integration site (consistent with the idea that this is an integration related mechanism which is thought to be symmetrical with respect to the recognized host sequence (Grandgenett, 2005) due to intasome symmetry (Hare et al., 2010). In some instances the integration bias dropped at distances below 100bp. This may be due to steric hindrance between the pre-integration complex and the DNA bound transcription factor. Such hindrance has already been described in vitro (Pryciak and Varmus, 1992).

Only a minority of integration sites were found in proximity to any particular TFBS, therefore the existence of the TFBS is not by itself necessary for integration. p53 in particular may not play a direct role in this effect as in the cell line used for in vitro infection, Jurkat, has been reported to be p53 null due to several mutations (Hainaut et al., 1997). Rather, the TFBS or factors associated with the TFBS may increase the efficiency or frequency of integration but the mechanism is not clear. In the case of HIV-1, host factors which are associated with proviral integration are well described. LEDGF/p75 has been identified as integrase cofactor (Cherepanov et al., 2003) which determines integrase localization (Maertens et al., 2003) and integration site distribution (Ciuffi et al., 2005). In its absence, HIV-1 preference towards integration within genes (and in particular LEGDF regulated

genes) is reduced (Ciuffi et al., 2005; Shun et al., 2007). A study of host factors associated with HTLV-1 integrase is currently underway.

An important limitation of this particular analysis approach is the fact that some of the available ChIP-seq data were not isolated in $CD4^+$ T-cells, but rather in other cell types. While the most appropriate dataset available was used, it is possible that the distribution of these in $CD4^+$ T-cells would be different. In addition, it is not known whether HTLV-1 infection of $CD4^+$ T-cells itself alters the genome-wide distribution of these binding sites.

Another open question which remains, concerns the role of higher-order chromatin structure in targeting the integration. It is known that the conformation of the DNA determines the efficiency of retroviral integration (Pruss et al., 1994). It is not known what role this may play in HTLV-1 infection, but it is possible that the TFBS which were found associated to proviral targeting do not themselves improve the efficiency of HTLV-1 integration, but rather that they represent a nuclear localization or higher-order structure that is more accessible to the HTLV-1 pre-integration complex.

4.3.3 Effect of genomic environment on clonal expansion in vivo

The abundance of an HTLV-1-infected clone in vivo will be determined by the net effect of two opposing forces: the clone's ability to proliferate and the clone's susceptibility to be killed: the strong in vitro activity of CTL suggests that killing by the CTL response is a significant selection force. If these selection forces acted equally upon all clones then we would expect to see all in vivo clones having similar clonal abundance. However, the high-throughput quantification of HTLV-1 proviral integration sites (Gillet et al., 2011, and this work) shows that in fact there is a wide variation in clonal abundance both between infected

individuals and within each infected individual. We hypothesize that the genomic environment of the integrated provirus is a determinant of this variation, by determining the frequency and intensity of proviral gene expression (in particular *tax* and *HBZ*). This gene expression would promote cell proliferation and confer a selective advantage on particular clones. The results shown here demonstrate that integration at the same orientation into or near a host gene favours clonal expansion, as does integration in proximity to CpG islands. However, opposite-sense orientation was disfavoured in vivo compared to random integration, which suggests that such integrated proviruses are selected out during chronic infection.

General transcriptional activity appears to favour clonal expansion, especially when the provirus and the flanking host gene lie in the same transcriptional orientation. What is therefore the role of proviral expression in determining clonal expansion? In the next chapter, we discuss the role of the genomic environment in determining level of Tax expression.

Publication arising from this chapter

Melamed, A., Laydon, D. J., Gillet, N. A., Tanaka, Y., Taylor, G. P. & Bangham, C. R. 2013. Genome-wide Determinants of Proviral Targeting, Clonal Abundance and Expression in Natural HTLV-1 Infection. PLoS Pathog, 9, e1003271.

Chapter 5 - **Association between genomic features, proviral expression and clonal expansion**

5.1 Introduction

5.1.1 Retroviral Latency and the genomic integration site

It is poorly understood how the genomic environment flanking the integrated provirus affects the level or the frequency of retroviral expression. It has been shown with both HIV-1- and HTLV-1-based vectors that expression of proviral genes can differ between otherwise identical clones, where the only difference is the integration site (Jordan et al., 2001; Landry et al., 2009).

Lee et al. (2008) have shown that the expression of murine endogenous retroviruses can vary according to the tissue, suggesting an association with the expression level of the flanking genomic environment. Focusing on the integration sites from HIV-infected individuals under highly-active anti-retroviral therapy, Ikeda et al. (2007) found that almost all integrated HIV proviruses lay within genes. Using green fluorescent protein (GFP)-expressing Rous Sarcoma Virus-based vectors, Plachy et al. (2010) found that stably expressing integrants were found to lie within or near to broadly transcribed genes (i.e., genes which are not tissue-specific), whereas experiments using a similar vector based on HIV revealed that proviruses with inducible expression were more frequently integrated in longer intergenic regions or in highly expressing genes (Lewinski et al., 2005). In HTLV-1 infection, integration sites from provirus-expressing cell lines have been found in sites with a higher GC content (Zoubak et al., 1994) and proviruses expressing Tax were more likely to be found within RefSeq genes, or in active regions of the genome (Meekings et al., 2008).

These results appear at first somewhat contradictory: proviral expression appears to be associated both with transcriptionally active regions of the host genome (i.e. not in gene deserts), but not with highly-expressing genes. This suggests that a more refined view is required to understand the mechanisms that regulate proviral transcription. The role of the

relative orientation with respect to the host gene has been investigated in artificial systems (Han et al., 2008; Lenasi et al., 2008) but conflicting results were obtained, and it remained unclear whether transcriptional interference plays a role in regulating proviral latency.

5.1.2 Does HTLV-1 cause a latent infection?

The role of the proviral expression in the field of HTLV-1 research is highly debated. By several lines of evidence, HTLV-1 appears to cause a latent infection: first, HTLV-1 virion particles are not found in the plasma, and cell free blood products are non-infectious (Okochi and Sato, 1984). Second, HTLV-1 is thought to proliferate predominantly by mitotic replication of the host cell, due to the low sequence diversity (Ina and Gojobori, 1990) and long-lived infected clones (Cavrois et al., 1996; Gillet et al., 2011). Lastly, HTLV-1 mRNA is often not found by real-time PCR immediately ex-vivo (Yamano et al., 2002).

In contrast to the latent infection view, the presence of constitutively activated Tax-specific CTL in the circulation (Daenke et al., 1996) suggests that in fact a certain level of Tax expression does take place in vivo. Similarly, a chronic IgM antibody response in a certain proportion of infected individuals (Kira et al., 1992) suggests a persistent viral protein expression (Asquith and Bangham, 2008) in vivo.

Freshly isolated HTLV-1-infected PBMCs generally do not express Tax mRNA or protein immediately ex vivo (Figure 5.1A); however, depletion of $CD8^+$ cells or culture in the presence of concanamycin A which inhibits perforin-mediated killing by CTLs, (Figure 5.1B) allows a measurable level of Tax protein expression in a subset of infected cells after a few hours in culture, suggesting that Tax expression is suppressed or controlled at least in part by the CTL response (Hanon et al., 2000a).

Seemingly conflicting observations emerge: Tax is found to be the immunodominant protein in HTLV-1 infection (Goon et al., 2004a), and the majority of HTLV-1 infected individuals have both circulating active Tax-specific $CD4^+$ and $CD8^+$ cells, which would suggest a strong selection for CTL escape mutations in the *tax* gene. Though these can be found (Niewiesk et al., 1995; Furukawa et al., 2001), HTLV-1 still is relatively conserved in sequence and constitutively active CTLs can be found in patients throughout the duration of the infection (Daenke et al., 1996), able to kill Tax-expressing cells in vitro (Asquith et al., 2005a), suggesting that CTL escape is not a significant mechanism of persistence (Gould and Bangham, 1998) in HTLV-1.

This would suggest that perhaps rather than modifying the sequence of Tax, HTLV-1 regulates the expression of the protein to its advantage.

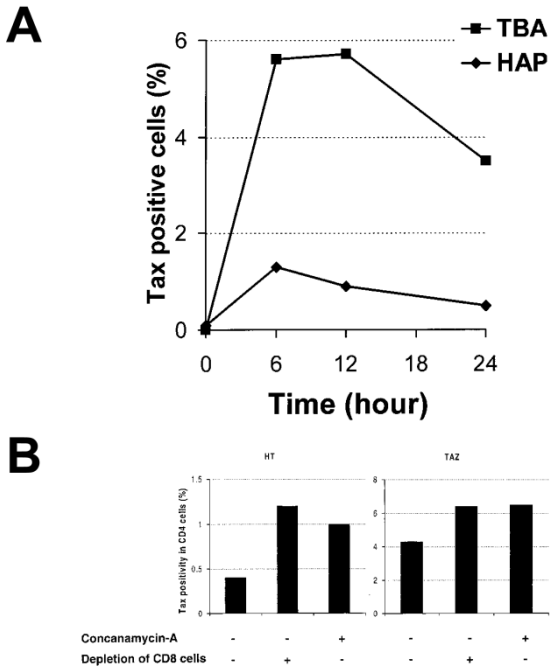

Figure 5.1: Expression of Tax ex vivo is controlled by $CD8^+$ cell activity.

(A) Expression of Tax ex vivo at different time points following isolation. The freshly isolated $CD4^+$ cells do not readily express Tax protein, however following short incubation in the absence of $CD8^+$ cells, a significant percentage of $CD4^+$ cells express Tax protein. Samples from a HAM/TSP patient (TBA) and an asymptomatic carrier (HAP) are shown. (B) Incubation of PBMCs from an asymptomatic carrier (HT, left panel) or a HAM/TSP patient (TAZ, right panel) in the presence of concanamycin A (inhibits perforin mediated killing by CTLs) or in the absence of $CD8^+$ cells allows increased Tax expression. Figure adapted from Hanon et al., 2000a

5.1.3 Expression kinetics in HTLV-1 infection

Proviral gene transcription from the 5'LTR (which includes the expression of Tax itself) is regulated by three imperfect repeat sequences at the LTR termed the Tax responsive elements (TRE) (Brady et al., 1987). These sequences do not allow direct binding of Tax to the LTR DNA, but form a complex with cAMP-responsive element (CRE) binding/activating transcription factors (CREB/ATF; Giebler et al., 1997). Tax has been shown in vitro to interact with multiple components of this group, such as CREB and XBP1 (reviewed in Matsuoka and Jeang, 2007; Boxus et al., 2008) in order to transactivate many viral and host genes. Host genes dysregulated by HTLV-1 Tax include genes involved in cell cycle and apoptosis, cytokines and DNA repair (Ng et al., 2001; Chevalier et al., 2012), as well as genes involved with interferon pathway signalling (Suzuki et al., 2010). This transactivation appears to be made possible by further interactions with histone-modifying enzymes (HDACs, p300/CBP, PCAF and the SWI/SNF complex).

Tax activity itself is regulated by other viral factors. The antisense viral gene HBZ has been shown to dysregulate Tax transactivating activity by recruitment of CREB (Figure 5.2, Gaudray et al., 2002). The viral protein p30 exerts post-transcriptional control on Tax by inhibiting Tax/Rex transcript export from the nucleus (Nicot et al., 2004).

Kinetic studies using real-time PCR have shown that the first transcripts to be expressed in the infected cell are Tax/Rex. The level of Tax rises sharply over the first 4-8 hours, and all other transcripts increase at a slower rate, reaching peak expression only after about 24 hours, suggesting a bi-phasic kinetics of gene transcription (Rende et al., 2011). This is consistent with the known role of Tax in transactivating other proviral genes.

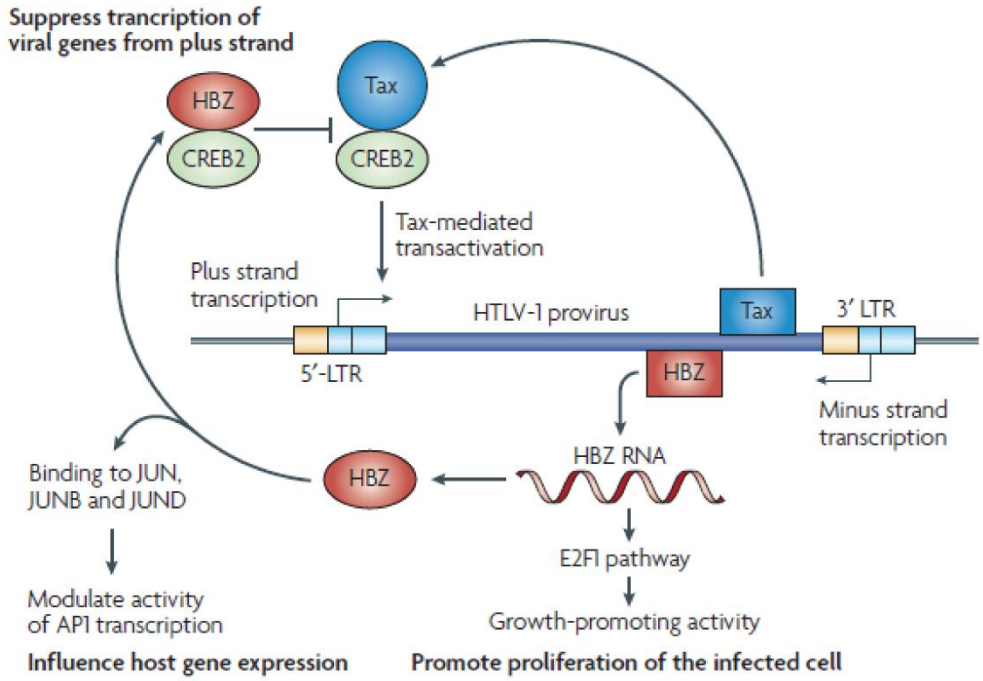

Figure 5.2: Current model of interaction between HBZ and Tax in HTLV-1 infection.
Tax mediated transcription requires recruitment of host cell factors such as CREB2, which is inhibited by HBZ. Figure from Matsuoka and Jeang, 2007

5.1.4 The selection forces acting upon the Tax expressing cell

Two particular viral genes are implicated in driving cell proliferation, thus promoting clonal expansion, which is thought to ultimately lead to malignant transformation. The Tax protein has been shown to promote proliferation both in vitro (Grassmann et al., 1989) and in vivo (Hasegawa et al., 2006) through multiple pathways: Tax cooperates with various cell factors (notably CREB and p300/CBP) to activate various host factors which push the infected cell through cell cycle check points into S phase (reviewed in Marriott and Semmes, 2005). These include binding to cyclins and activating transcription of cdk2 and cdk4, driving Rb degradation and inactivating p53. In addition, Tax upregulates the expression of various

cytokines and their receptors such as IL-15 and IL-2, which can promote proliferation (Matsuoka and Jeang, 2007).

The more recently identified HTLV-1 gene, HBZ (Gaudray et al., 2002) has also been shown to support proliferation of infected cells in vitro (Satou et al., 2006), possibly mediated by the host protein Activating transcription factor 3 (ATF3) which can upregulate CDC2 and cyclin E2. (Hagiya et al., 2011). In a HBZ-transgenic mouse model (Satou et al., 2011), $CD4^+$ T cell proliferation was also observed, as well as development of inflammatory lesions in the skin and lungs. Conversely, infection of rabbits with HBZ-null HTLV-1 mutants revealed that HBZ was not essential for HTLV-1-mediated lymphocyte proliferation (Arnold et al., 2006). HBZ mRNA expression correlates with indicators of HAM/TSP disease severity (Saito et al., 2009).

It has yet to be fully clarified whether Tax or HBZ play a more significant role in chronic infection and in disease: While Tax was found to be necessary for maintenance of transformed phenotype in particular cell lines (Yamaoka et al., 1992), its expression is abrogated in ~60% of ATLL cases either through proviral deletions at the 5'LTR (Tamiya et al., 1996), mutations in *tax* gene (Furukawa et al., 2001) or methylation of the enhancer/promoter region of the HTLV LTR (Taniguchi et al., 2005). HBZ mRNA, however, is consistently found in all cases of ATLL, and HBZ appears to be protected from mutations (Fan et al., 2010).

Alongside its crucial activities for transactivating proviral and host genes, Tax also represents the immunodominant antigen recognized by CTL, and anti-Tax CTL can be found circulating in the blood of the majority of HTLV-1-infected individuals (Jacobson et al., 1990; Goon et al., 2004a; Parker et al., 1992) with or without associated disease. IFN-γ and IL-2 $CD8^+$ T cell responses to Tax are found (using ELIspot) in most infected individuals (Hilburn et al.,

2011), whereas $CD8^+$ T cell responses to HBZ are only found in the minority of infected individuals. The rate of Tax^+ cell killing per day per autologous $CD8^+$ cell is inversely correlated with HTLV-1 proviral load (Asquith et al., 2005a) and cells expressing a higher amount of Tax protein are killed more efficiently (Kattan et al., 2009).

Lastly, Tax protein (Cartier and Ramirez, 2005) and Tax-specific CTL (Jacobson et al., 1992) are found in the CNS in HAM/TSP, consistent with the notion that Tax may play a role in pathogenesis of HTLV-1. However there are conflicting hypotheses on the role of HTLV-1-specific $CD8^+$ cells which may either play a protective role against HTLV-1 disease or an active role in the pathogenesis (Bangham, 2009; Jacobson, 2002) or, perhaps most likely, both (Bangham, 2009).

It appears, therefore, that two opposing forces act on the Tax-expressing clone, determining the survival and expansion potential of the clone. On the one hand, Tax is able to drive clonal expansion. On the other, Tax expression exposes the infected cell to a CTL response.

5.1.5 Aim

We aim to test the hypothesis that the genomic environment flanking the integrated provirus affects the transcriptional state of the provirus, and to test whether the level of proviral expression is correlated with clone abundance.

5.2 Results

In order to investigate the level of expression in different HTLV-1-infected clones, we cultured $CD8^+$ T-cell-depleted PBMCs from 10 HAM/TSP patients overnight to allow spontaneous proviral expression (Hanon et al., 2000a). We then used high-speed cell sorting to separate Tax-expressing and non-expressing $CD4^+$ cells (figure 2.1, 2.2). Sorted cells were combined into Tax positive and Tax negative cell pools to allow efficient DNA extraction, then analysed for integration site content using our high-throughput method. In addition, integration sites from unsorted PBMCs from all 10 patients were analysed separately in order to attribute integration sites from the pooled samples back to the patients. It was possible to attribute 46% of clones in this manner.

Table 5.1: flow cytometry sorting for Tax^+, Tax^- cells

Patient #	**$CD4^+Tax^+$ cells**	**$CD4^+Tax^+$ / $CD4^+$**	**$CD4^+Tax^+$ /total**	**$CD4^+Tax^-$ cells**	**$CD4^+Tax^-$ / $CD4^+$**	**$CD4^+Tax^-$ /total**
1	497,369	38.1%	7.0%	781,154	60.0%	11.1%
2	644,048	41.6%	10.8%	893,925	56.1%	14.5%
3	136,793	29.2%	9.8%	326,983	69.8%	23.4%
4	461,568	21.5%	9.0%	1,866,092	77.9%	32.6%
5	242,228	10.8%	3.8%	2,368,791	88.4%	31.1%
6	242,984	11.9%	6.3%	1,957,727	87.3%	46.2%
7	106,896	16.4%	3.9%	583,892	81.8%	19.3%
8	222,177	19.4%	3.1%	943,268	79.1%	12.5%
9	165,110	23.6%	5.6%	587,502	73.1%	17.2%
10	925,785	53.8%	23.2%	578,717	32.8%	14.2%
Total cells	3,644,958			10,888,051		
Total DNA (µg)	4.067			8.19		
Total proviral copies found	20813			10326		

5.2.1 Majority of clones are either 100% or 0% positive

The integration sites identified in the Tax positive cells and in the Tax negative cells were added and compared in order to identify the proportion of Tax positive cells in each clone. The majority of clones (68%), as expected, were singletons (clones found only in one cell). However, even among clones in the highest abundance bin (over 10 cells), the majority of clones contained either > 90% or < 10% Tax positive cells, suggesting that the Tax positivity (or lack of) is an inherent property of certain clones (Figure 5.3).

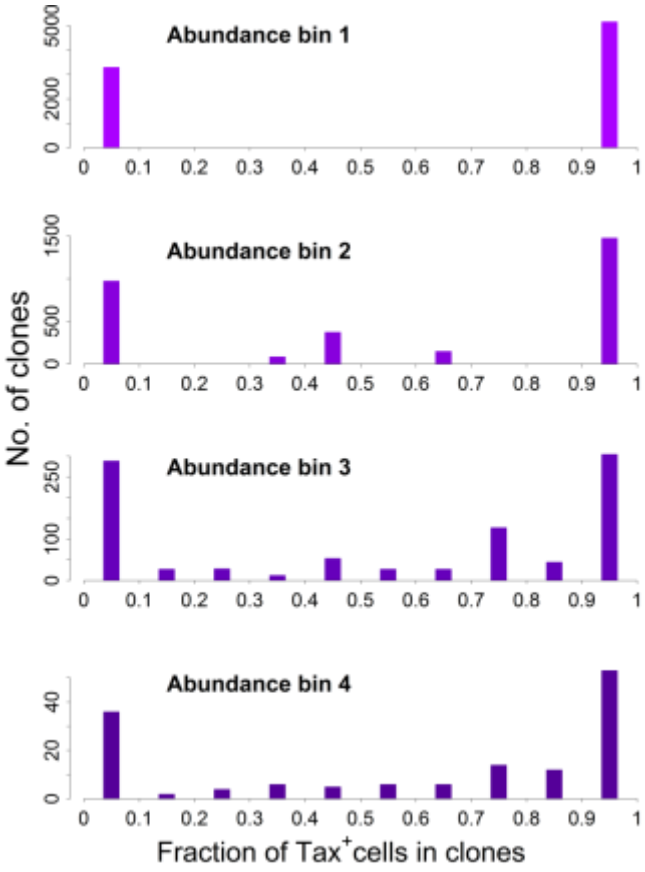

Figure 5.3: The majority of clones are either 100% or 0% Tax positive.
Frequency distribution of clones according to the proportion of Tax^+ cells in each clone, binned according to the number of sisters in each clone. Bin1 – 1 cell detected; bin 2 – 2-3 cells detected; bin3 – 4-10 cells detected; bin 4 – more than 10 cells detected.

5.2.2 Association between spontaneous Tax expression and integration within a gene.

When the provirus was integrated within a gene, we observed a slight but significant excess of frequency of Tax^+ cells Tax^- cells (46% vs 43% respectively, $p < 10^{-3}$, $\chi 2$ test). Consistent with previous observations, the integrated provirus was found more frequently in the same transcriptional orientation as the host gene (Figure 5.4). However, while the proviruses in the Tax^+ cells were found with similar frequency in the same or the opposite transcriptional orientation to the host gene in which they were integrated, the Tax^- cells were significantly more frequently present in the same orientation as the host gene (52% of Tax^+ vs 59% of Tax^- cells, $p < 10^{-15}$, $\chi 2$ test). Thus, T cell clones that were 100% Tax^- were significantly more likely to carry a provirus in the same orientation as the host gene.

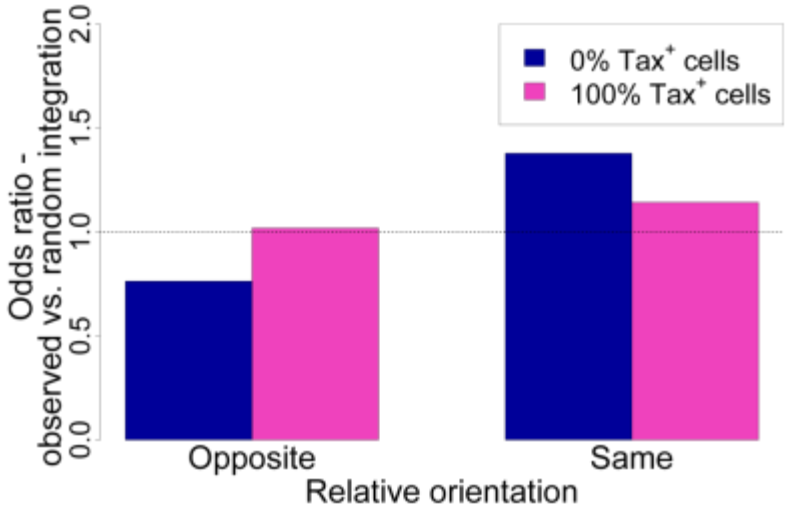

Figure 5.4: Same-sense integration within a host TSS is associated with Tax silencing.
Odds ratio of integration of Tax^+ or Tax^- clones within a gene compared to random expectation. Tax expressing clones were more frequently integrated within a gene but integration in the same sense orientation was strongly associated with Tax silencing. Relative orientation is with respect to the sense strand of HTLV-1.

5.2.3 Integration in proximity to a TSS or CpG island

The pattern of bias towards integration in vivo in proximity to a TSS or CpG island is consistent with previous findings (Figure 5.5A, Figure 5.6), with a peak bias at 1kb from the integration site. The bias towards integration within 1kb of a TSS was higher for Tax^+ cells; however, for proviruses with an upstream nearby TSS, there was an excess frequency of integration in the same orientation as the host gene for Tax negative cells.

The mean proportion of Tax^+ cells per clone (across all clones, including singletons) was 0.6. We wished to test whether integration in proximity to TSS would alter this proportion. We found that a nearby TSS upstream of the integration site was associated with a high proportion of Tax positive cells if the provirus was in the opposite orientation as the upstream TSS, but with a lower proportion of Tax positivity if the provirus was in the same orientation. The closer the TSS to the provirus, the stronger these effects became (Figure 5.5B). This asymmetry was not observed for TSS downstream of the integration site (Figure 5.5B), where proximity to a TSS was associated with increased proportion of Tax positivity.

Similarly, we observed an excess frequency of Tax + cells where the nearest CpG island lay downstream of the integration site, and an excess frequency of Tax^- where the nearest CpG island lay upstream of the integration site. (Figure 5.6)

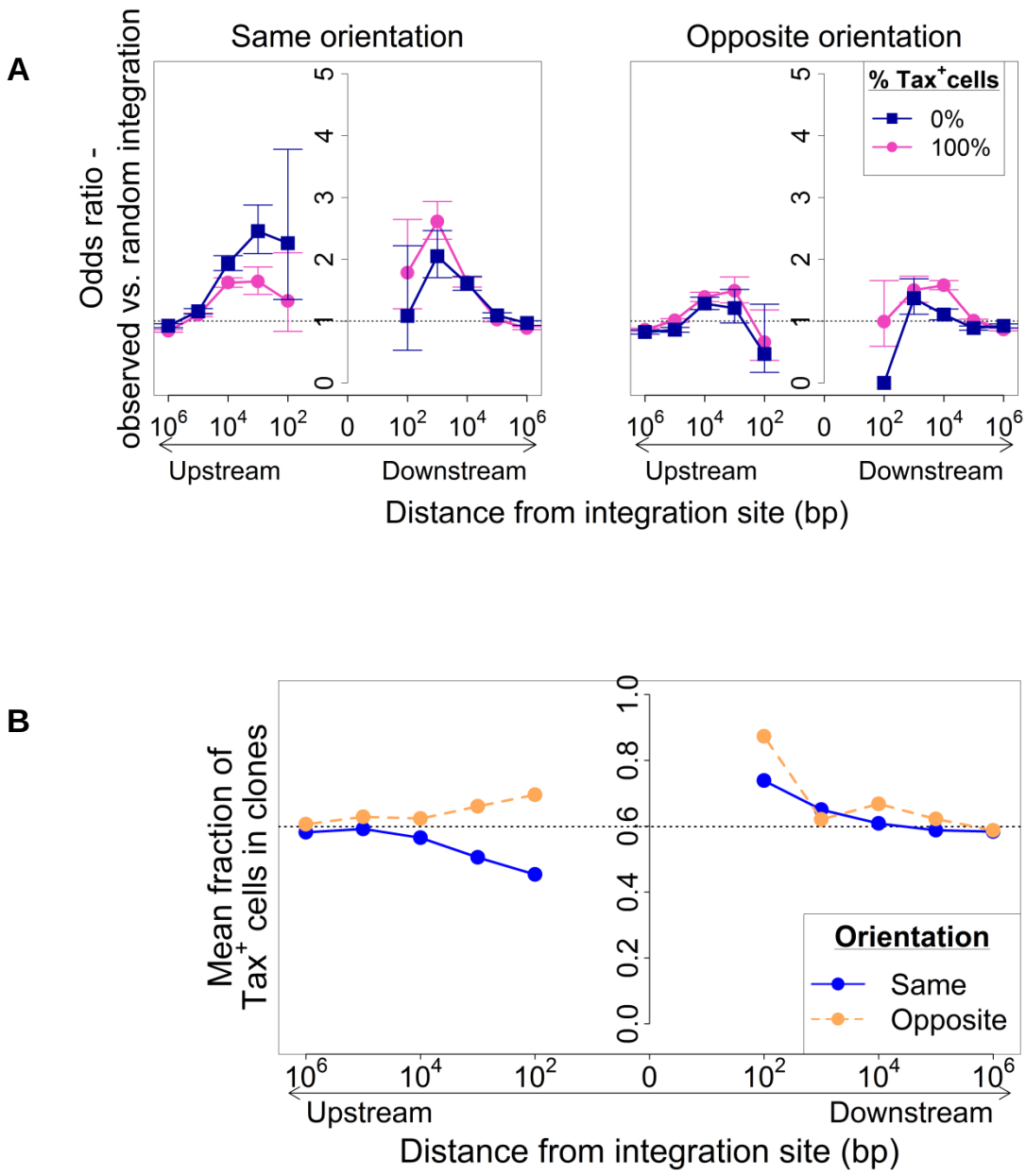

Figure 5.5: Tax expression is favoured in proximity to TSS, but disfavoured in proximity to upstream TSS of the same orientation.

'Upstream', 'Downstream' and relative orientation are with respect to the sense strand of HTLV-1. (A) Mountain plot showing the odds ratio of integration of Tax^+ or Tax^- clones in proximity to TSS compared to random expectation. (B) Clones with upstream TSS had a higher mean positivity if the nearest upstream TSS was in the opposite orientation, and a lower mean positivity if the nearest upstream TSS was in the same orientation. The presence of a downstream TSS was associated with increased Tax positivity, regardless of orientation.

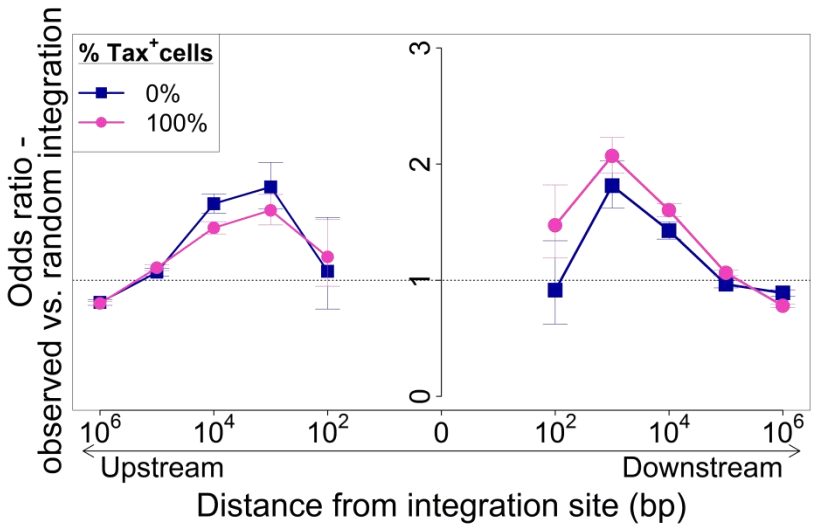

Figure 5.6: Odds ratio of integration in proximity to CpG island.

Bias towards integration in proximity to CpG island was similar between Tax-expressing and non-expressing clones, but was higher in Tax-expressing clones if a CpG island lay downstream of the integration site. 'Upstream', 'Downstream' are with respect to the sense strand of HTLV-1.

5.2.4 Integration in proximity to TFBS

Consistent with previous in vivo results, there was a small preference towards integration in proximity to a TFBS (within those TFBS tested using available ChIP-seq data) for both Tax^+ and Tax^- cells. In very close proximity upstream of the integration site (within 10 bases) there was a slightly greater bias for the Tax^+ cells than for Tax^- cells. (Figure 5.7).

We wished to test whether integration in proximity to the different TFBS would alter the mean frequency of Tax^+ positivity of the infected clone. We observed that for certain TFBS (including STAT1, cJun, NRSF; Figure 5.8) the presence of a TFBS in close proximity upstream of the integration site was associated with increased mean Tax positivity. In stark contrast, a BRG-1 upstream to the integration site was associated with Tax negativity. An infected cell with a BRG-1 downstream to the integration site, however, was more likely to be Tax positive.

In order to identify TFBS which were independently associated with Tax positivity or negativity, a multivariate logistic regression analysis was carried out as previously described, using the outcome measure of Tax^+ cell and integration within all tested TFBS within 100 or 1000 bases as explanatory variables. The results (Figure 5.9) confirmed the asymmetric effects of the BRG-1 binding site, and in addition revealed significant asymmetric associations between Tax expression and several other TFBS, notably STAT1, NRSF, and HDAC1. Thus, an NRSF binding site (Figure 5.8C) 100bp downstream was a significant predictor of Tax negativity, but the closest upstream NRSF binding site was not independently associated with Tax expression.. Conversely, a STAT1 binding site (Figure 5.8D) at 100bp or HDAC1 binding site at 1kb upstream of the provirus strongly favoured Tax expression, but the presence of a downstream STAT1/HDAC1 binding site was not an independent predictor of Tax expression after multivariate analysis.

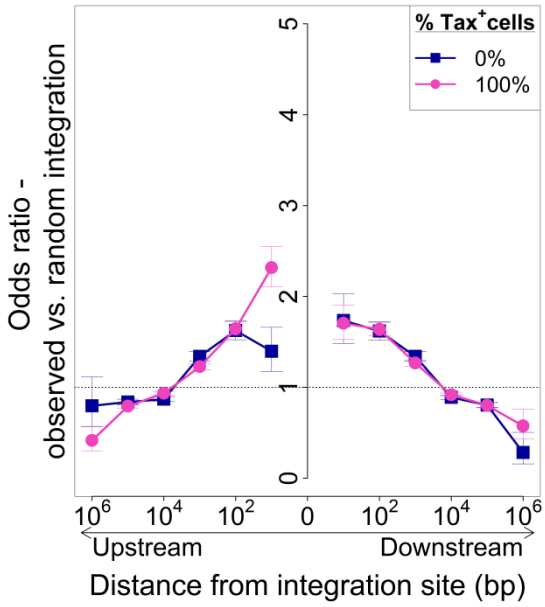

Figure 5.7: Integration in proximity to any TFBS of those tested.

Small preference towards integration in proximity to any TFBS with a small excess preference for Tax expressing clones to have TFBS in close proximity upstream to the integration site. 'Upstream', 'Downstream' are with respect to the sense strand of HTLV-1.

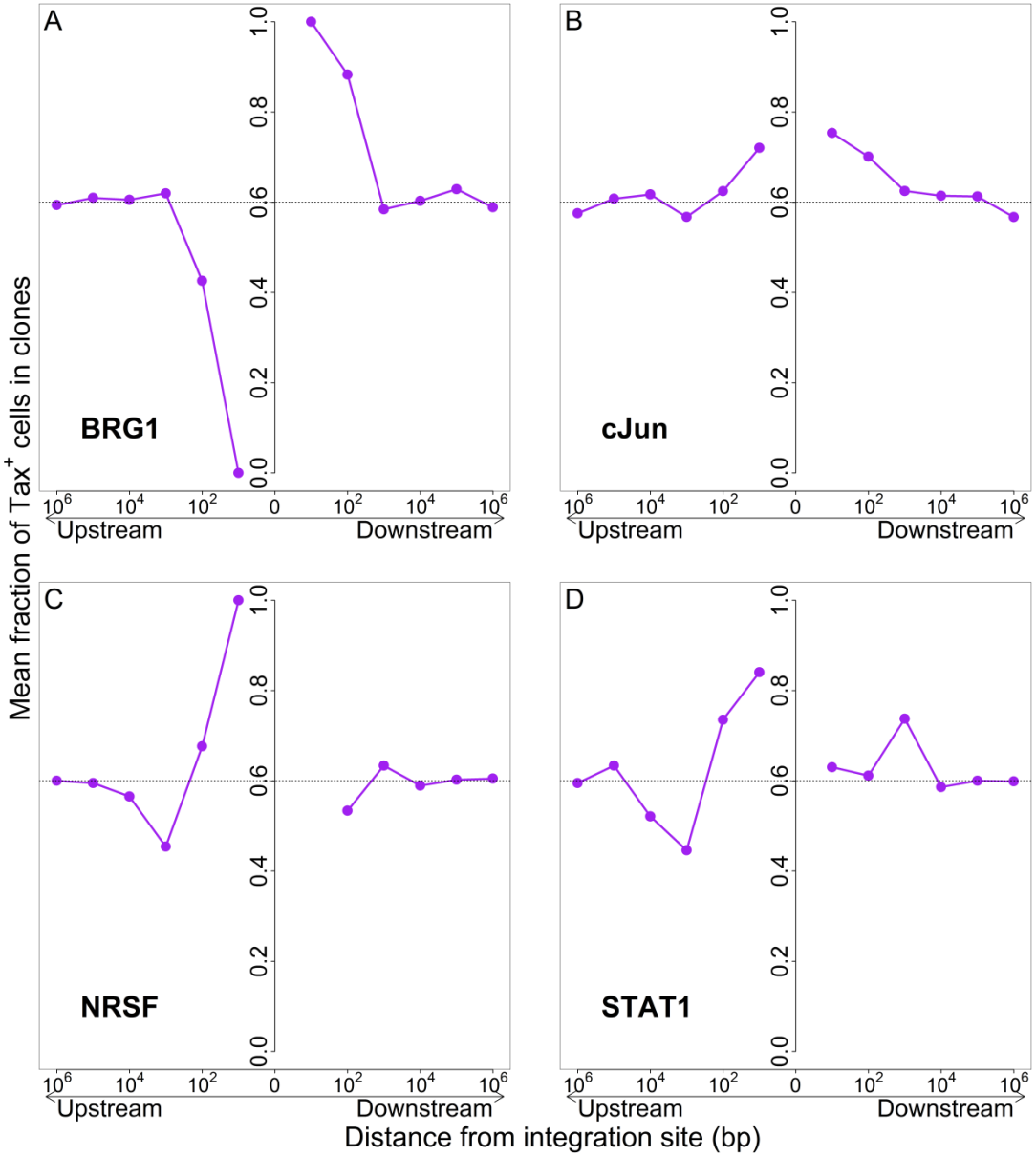

Figure 5.8: Tax expression is altered by integration in proximity to particular TFBS.

The mean proportion of clones within 10b-1000kb of a nearby TFBS was calculated to determine the effect of nearby TFBS on Tax positivity. The effect of nearby TFBS on Tax expression was notably asymmetrical. 'Upstream', 'Downstream' are with respect to the sense strand of HTLV-1.

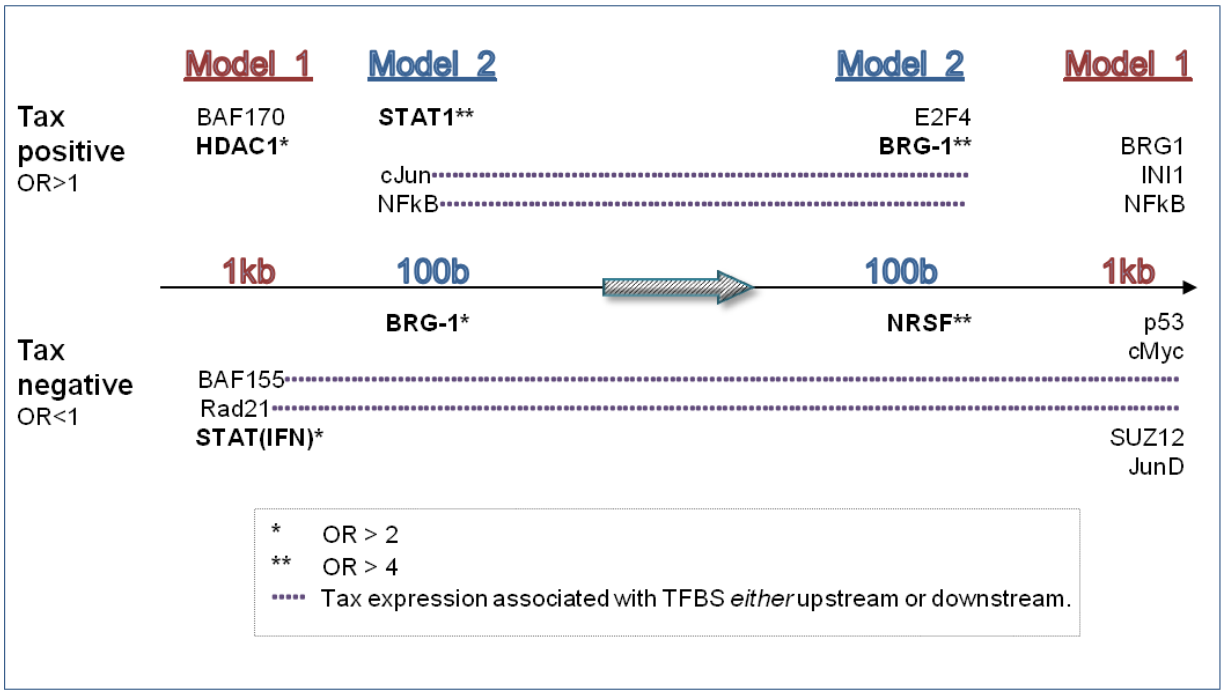

Figure 5.9: Integration in proximity to TFBS: multivariate analysis.

TFBS that were independently associated with Tax expression were identified by multivariate analysis. Outcome measure was Tax expression of a given cell. TFBS shown above the line were associated with Tax expression (OR>1); TFBS below the line were associated with Tax silencing (OR <1). Model 1 and Model 2 (carried out independently) test for TFBS within 1kb and 100bp of integration site, respectively.

5.2.5 Association between Tax expression and clone abundance.

Lastly, we wished to test the distribution of Tax positive cells and Tax negative cells between the different clone abundance bins. We assigned all clones to abundance bins based on the number of sisters in each clone. While small (low abundance) clones were the most frequent both among Tax^+ cells and Tax^- cells (as small clones are remarkably more numerous than larger ones), the slope was significantly ($p = 0.006$, ANCOVA test) higher in Tax^+ cells, because there were more Tax^+ cells in the singletons than in the largest clones (Figure 5.10). Across all clones, the fraction of Tax^+ cells in each clone abundance bin was significantly negatively associated ($p<10^{-16}$, $\chi 2$ test for trend) with clonal abundance. Among the clones which we were able to attribute to each patient, the negative correlation was observed for

most patients; however in a few patients (in particular, those with a high oligoclonality index; Figure 5.11) the largest clone abundance bin included a high number of Tax^+ cells. Therefore, at least for HAM/TSP patients, the majority of the high frequency of Tax-expressing cells observed is due to a large number of small clones, rather than a small number of expanded Tax-expressing clones.

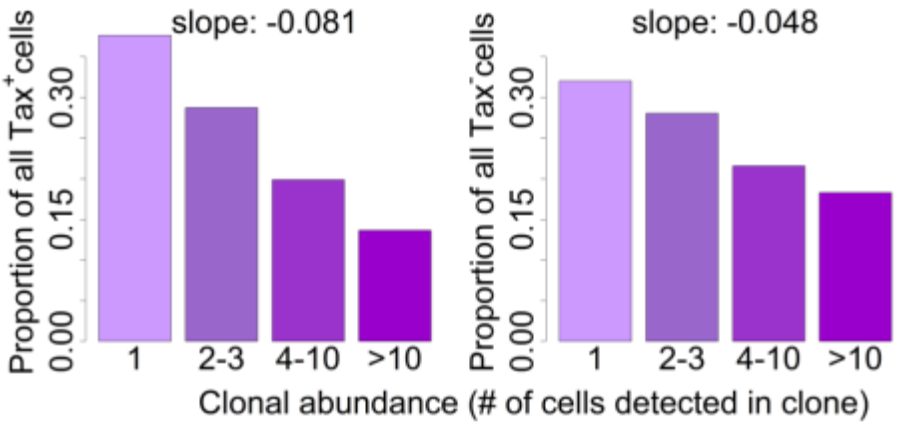

Figure 5.10: Frequency distribution of Tax^+, Tax^- cells between different clone abundance.
In both Tax^+ (left) and Tax^- (right) cell populations, the proviral load was dominated by smaller clones. However the proportion of Tax^+ cells declined significantly with increasing clone abundance (ANCOVA test for difference in slope, $p = 0.006$).

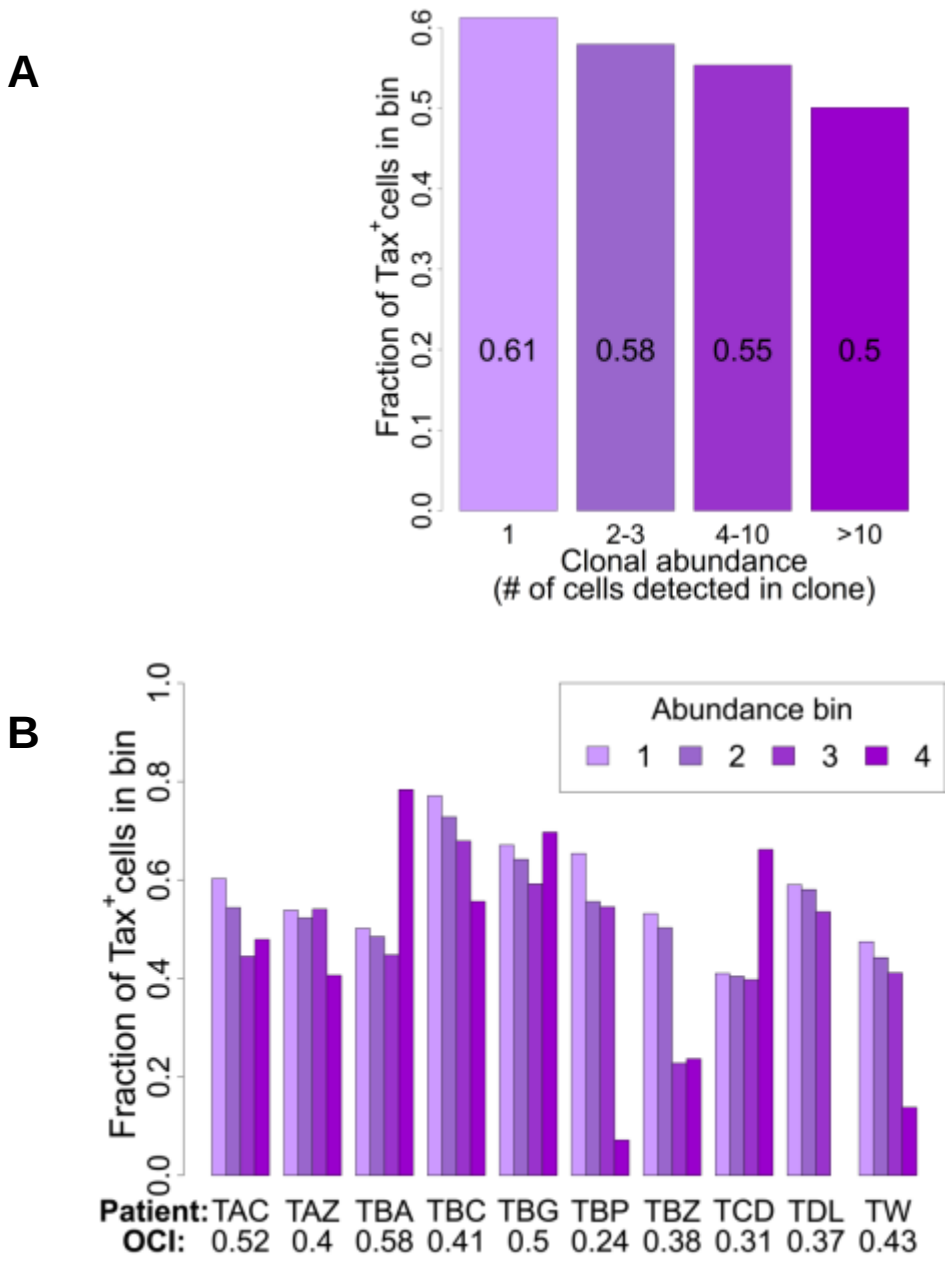

Figure 5.11: Tax$^+$ cells were more frequent in smaller clones.

(A) Mean fraction of Tax$^+$ cells in bins of increasing clonal abundance (total number of cells in each respective clone). The fraction of Tax$^+$ cells was negatively correlated with clonal abundance ($p < 10^{-16}$, $\chi 2$ test for trend). (B) In the majority of patients there was an inverse correlation between clone abundance bin and fraction of Tax$^+$ cells. While this correlation was highly significant in all patients combined (see A), in certain patients (particularly those with a high oligoclonality index) the most abundant clones contained a high proportion of Tax$^+$ cells. Clones were binned according to the number of sisters observed, as above.

5.3 Discussion

The mechanism of proviral latency is of great importance in retroviral infection. HIV-1-infected individuals receiving highly-active anti-retroviral therapy (HAART) maintain a population of transcriptionally latent proviruses which persist in resting $CD4^+$ cells (Finzi et al., 1997; see also Durand et al., 2012) as a reservoir which can reactivate upon interruption of therapy. During HTLV-1 infection it was long debated whether there is proviral expression during in vivo infection. HTLV-1 gene expression is difficult to detect in fresh PBMC; however, the strong, constitutively active anti-HTLV-1 CTL found in infected individuals suggests frequent reactivation of the proviral genes in vivo, though this may be intermittent, regulated by the viral genes and host immune system (Bangham et al., 2009). While it is difficult to detect Tax protein in fresh PBMCs, a short-term culture of $CD8^+$ cell-depleted PBMCs reveals a population of infected cells expressing Tax protein (Hanon et al., 2000a). Here, I used this phenomenon in order to investigate the effect of the genomic environment on the likelihood of Tax expression by an infected cell. It is difficult to say with certainty what is the magnitude or frequency of Tax expression in vivo, and it is impossible to say that the same HTLV-infected cells which spontaneously express Tax in culture would also express Tax in vivo. However, the fact that cells which express Tax ex vivo turn over more rapidly in vivo (Asquith et al., 2007) and the association between the proportion of such ex vivo expressing $CD4^+$ and clinical outcome (Asquith et al., 2005b) suggest there is a clear clinical relevance to cells that spontaneously express Tax^+ ex vivo.

5.3.1 Transcriptional interference is a possible mechanism of proviral latency.

In this work I observed an association between spontaneous expression of Tax and integration within genes or in proximity to transcription start sites. Tax expression was also weakly associated with the presence of a CpG island in proximity (in particular downstream) and with certain TFBS. This confirms previous work (Zoubak et al., 1994; Meekings et al., 2008) and is consistent with the idea that on a large scale, integration in a genomically active region (rather than heterochromatin) is conducive to Tax expression from the integrated provirus. However, refining the analysis to consider the relative position and orientation of the provirus and the host genome has uncovered a more localised effect: where the provirus was integrated just downstream of a transcription start site, the integrated provirus was significantly more likely to be transcriptionally silent, but only if the upstream TSS was in the same orientation as Tax. This observation is consistent with the idea of transcriptional interference, where RNA polymerase II transcribing from the upstream promoter blocks initiation of transcription from 5'LTR promoter (Lenasi et al., 2008). This mechanism has been tested before in single gene in vitro models with conflicting results (Han et al., 2008; Lenasi et al., 2008). More recently, using an HIV model of latency, infecting Bcl-2-transduced primary cells using GFP-expressing HIV-based vector, Shan et al have shown that proviruses with inducible GFP expression were more likely to lie in the same orientation as the host gene, and were more like to be in highly active genes; persistence of GFP expression was associated with opposite sense integration(Shan et al., 2011). The evidence presented here demonstrates that in naturally HTLV-1-infected cells, same-sense orientation of the upstream promoter is associated with silencing of proviral gene expression, consistent with a mechanism of transcriptional interference. In order to validate this mechanism, it would be useful to test for the expression of the upstream genes, for example in HTLV-1 –infected T

cell clones (Cook et al., 2012) and compare that with the level of Tax expression in each clone.

We conclude that regulation of transcription of HTLV-1 Tax takes place at two levels. First at a regional level, the transcriptional activity of the genome favours proviral expression and clonal expansion (Meekings et al., 2008; Gillet et al., 2011). Second, a local effect of upstream promoter up to 1kb from the integration site can silence expression from 5'LTR, overriding the regional effects.

5.3.2 Role of Transcription factors and chromatin remodellers in Tax expression

While at the targeting level the associations observed between integration and nearby transcription factor binding sites were largely symmetrical, the effect of proximity to TFBS on Tax expression was often asymmetrical. The asymmetry of these associations suggests a mechanistic interaction between the proviral expression and transcription of the nearby host genome. A notable example of this asymmetry was BRG1, one of the two ATPase components of the SWI/SNF complex. This is a multi-subunit chromatin remodelling complex (Euskirchen et al., 2012) which controls gene expression by repositioning of nucleosomes in a DNA topology-dependent manner (Gavin et al., 2001). The SWI/SNF complex is heterogeneous, and it is thought that the particular combination of its subunits determines its effect on transcription (Euskirchen et al., 2011). BRG1 directly interacts with Tax, and suppression of BRG1 using siRNA inhibits Tax transactivation (Wu et al., 2004, Easley et al., 2010). Here we found that clones which had a BRG1 binding site downstream of the integration site were more likely to express Tax, while clones which had such a binding site just upstream of the integration site were more likely to be Tax silent.

Several other SWI/SNF subunits are known to have a role in retroviral infection. INI1 colocalizes and interacts with HIV-1 integrase (Kalpana et al., 1994; Turelli et al., 2001), however its role in HIV-1 expression is not clear. Here we found that integration with a downstream INI1 site within 1kb was independently associated with Tax expression, similar to the downstream BRG1 site.

The DNA binding sites of other components of the SWI/SNF complex also seem to have a role in Tax expression. An upstream BAF170 binding site, known to be highly expressed in HTLV-1 infected cells (Van Duyne et al., 2011) was here found to be associated with spontaneous Tax expression, while a proximal binding site of BAF155, which has been reported to be downregulated in HTLV-1 infected cells (Van Duyne et al., 2011) was here found to be associated with Tax silencing (Figure 5.9).

Some of the other transcription factors whose binding site was found to be associated with Tax expression are also known to have a role in transactivation of proviral genes. cJun and cofactors are recruited by Tax for transactivation from the 5'LTR , leading to histone acetylation (Lemasson et al., 2002); the treatment of HTLV-1 infected cells with sodium valproate, an HDAC inhibitor, increases the amount of Tax expression (Mosley et al., 2006). It is not clear what role the proximity to the binding site itself may play in upregulating transcription of Tax, though it may increase the efficiency of recruitment of such factors to the 5'LTR promoter.

5.3.3 Effect of proviral expression on clonal abundance

Tax has been implicated in driving clonal proliferation. As discussed in section 5.1, it is known to push the infected cell through cell cycle progression, and to transform cells both in vitro and in animal models. In addition, cells which express Tax ex vivo have a higher turnover rate in vivo (Asquith et al., 2007). Therefore, we expected that clones which express Tax would be expanded in vivo. However, this work has revealed the opposite, i.e. a highly significant negative correlation (Figure 5.11A): Tax expressing cells were significantly more likely to belong to small clones and there was a significantly higher number of low abundance clones expressing Tax (Figure 5.10). This observation is likely to be due to the immune response. Tax is the immunodominant protein in the CTL response against HTLV-1 (Goon et al., 2004a; Kannagi et al., 1991) and activated autologous Tax-specific cells control the expression of Tax ex vivo (Hanon et al., 2000a). In addition, cells which express a higher level of Tax ex vivo are killed more efficiently by autologous CTLs (Kattan et al., 2009). Our results agree with this – if the initial integration favours Tax expression, these cells will be more readily identifiable to the immune response and will consequently be cleared. We conclude that the small HTLV-1-infected clones express Tax in vivo, and turn over faster. Clones which do not express Tax readily will be able to escape CTL for longer and proliferate, perhaps driven by HBZ. It would be interesting and important to compare these results with a similar test of HBZ expression however since there is currently no available method to sort HBZ-expressing cells; this cannot be tested directly. There is currently a project underway which aims to identify HBZ expressing clones by their ability to be cleared by HBZ-specific CTLs. It is possible that the critical role of Tax in maintaining proviral load is not in driving clonal expansion, but in facilitating cell-to-cell transmission: expression of proviral genes, promoting new virion production and mediating the cell polarization required

for cell-cell virus transmission through the virological synapse (Nejmeddine et al., 2005; Mazurov et al., 2010)

We observed a significant negative correlation between clone abundance and the percentage of Tax^+ cells. While this was highly significant in all patients combined, in a small number of patients (in particular those with a high oligoclonality index, Figure 5.11B) the most abundant clones (bin 4, clones with greater than 10 cells) contained a high proportion of Tax^+ cells. What has enabled these clones to survive the CTL control?

It is possible that antigen-expressing clones may have escaped control by the immune response, for example by CTL escape mutations in the *tax* gene (Niewiesk et al., 1995). Clone-specific sequencing of exon 3 of the *tax* gene of 38 highly abundant clones from 8 patients did not reveal significant differences in the occurrence of Tax mutations between clones with a high or low frequency of Tax^+ cells; and only in one patient was a difference in amino acid sequence found between one clone and the others (data not shown).

It should be noted that since Tax transactivates all other proviral genes, the immune response to Tax-expressing cells could be targeted against viral proteins other than Tax itself, such as Gag, which is a strong target for both the CTL and antibody responses. Therefore we cannot rule out the possibility of CTL escape mutations in another proviral gene.

A number of open questions remain:

1) What drives selective clonal expansion?
2) What is the role of Tax in HTLV-1-related disease?

The results of this study are consistent with the idea that it is not Tax, but HBZ that is responsible for clonal expansion in vivo. Other evidence for this idea comes from ATLL, where the expression of Tax (and consequently other plus-strand proviral genes) is in fact

often abrogated by deletions, mutations and methylations. The finding of the duplicated host sequence still flanking the deleted provirus suggests that at least in some cases the deletions took place prior to proviral integration, and the infected clone expanded in vivo in the absence of Tax expression (Tamiya et al., 1996; Miyazaki et al., 2007). HBZ, which is expressed from the minus strand, appears protected from mutations (Fan et al., 2010), and HBZ mRNA expression is detected in all ATLL cases (Saito et al., 2009).

At a given proviral load, HAM/TSP patients have a higher proportion of $CD4^+$ cells expressing Tax than ACs (Asquith et al., 2005b). This is consistent with our previous finding (Gillet et al., 2011) that the increased proviral load in HAM/TSP is due to a larger number of different clones (rather than expansion of a few clones) and our present finding that the Tax expressing cells are associated with a large number of low-abundance clones rather than a small number of expanded Tax-expressing clones. However, the role of Tax in pathogenesis is not clear. Tax is more readily detectable in cells isolated from the CSF of both patients with HAM/TSP and ACs than in the blood (Moritoyo et al., 1999). This could imply that the Tax-expressing cells in the CSF are less well controlled (and may selectively proliferate) than those in the blood, or that Tax-expressing cells are more likely to infiltrate the CSF. In addition, expression of cytokines in the CSF by other infiltrating cells may promote expression by altering the transcriptional activity of the infected cell.

This work opens the way to more questions. The effects of transcription factors and chromatin remodellers on Tax expression should be investigated in an in vitro model, and the implications of such effects for persistence, latency and pathogenesis should be analysed.

Publication arising from this chapter

Melamed, A., Laydon, D. J., Gillet, N. A., Tanaka, Y., Taylor, G. P. & Bangham, C. R. 2013. Genome-wide Determinants of Proviral Targeting, Clonal Abundance and Expression in Natural HTLV-1 Infection. PLoS Pathog, 9, e1003271.

Chapter 6 - The contribution of infected CD8$^+$ cells to the proviral load

6.1 Introduction

6.1.1 HTLV-1 host cell tropism

HTLV-1 is known to preferentially infect $CD4^+$ cells (Richardson et al., 1990), however $CD8^+$ cells have also been described as a reservoir for infected cells (Richardson et al., 1990; Nagai et al., 2001a; Cho et al., 1995; Hanon et al., 2000b). The HTLV-1-associated malignancy, ATLL, is a predominantly $CD4^+CD8^-$ malignancy (Uchiyama et al., 1977) , although relatively rare $CD4^+CD8^+$ double positive, $CD4^-CD8^-$ double negative and $CD4^-CD8^+$ phenotypes have also been described (Raza et al., 2010; Kim et al., 2006, Ciminale et al., 2000; Kamihira et al., 1992). The related virus HTLV-2 appears to have a complementary tropism, with most of its load found in $CD8^+$ cells (Ijichi et al., 1992), however HTLV-2-infected $CD4^+$ cells can also be found (Lal et al., 1995).

It is debated which is the determining factor in the host cell tropism of HTLV-1 and HTLV-2. Early evidence showed that the cellular receptors for HTLV-1 are found on many cell types (Yamamoto et al., 1984; Krichbaum-Stenger et al., 1987). Experimental in vitro infections by HTLV-1 were successfully done in multiple cell types, including human B cells, fibroblasts, endothelial and Glial cells (Yamamoto et al., 1982a; Yoshikura et al., 1984; Hoxie et al., 1984; Saida et al., 1988), however in vivo the load appears to be present almost entirely in the $CD4^+$ and $CD8^+$ cells (Richardson et al., 1990).

Despite apparent higher Tax expression in $CD4^+$ cells than in $CD8^+$ cells (Newbound et al., 1996), experiments in which Tax/Rex sequences were swapped between recombinant HTLV-1 and HTLV-2 did not alter the host cell tropism of either virus (Ye et al., 2003b).

With the increased understanding of the host molecules involved in cellular entry of HTLV-1 and HTLV-2, Jones et al. (2006) have shown that $CD4^+$ cells express high levels of HSPGs

while $CD8^+$ cells express much higher levels of GLUT-1. By modifying the expression of these molecules, this group was able to increase infection of $CD8^+$ cells by HTLV-1, suggesting that that the host cell tropism is at least in part determined by the concentration or distribution of receptor molecules on the target cell surface. In vivo and in vitro experiments carried out in the rabbit model by Kannian et al. (2012) gave contrasting results. When animals were infected with either HTLV-1 or HTLV-2, both $CD4^+$ cells and $CD8^+$ cells were initially infected by HTLV-1, but the proviral load differed over time in the different cell populations. In addition, when infecting PBMCs in vitro, HTLV-1 preferentially immortalized $CD4^+$ cells and HTLV-2 preferentially immortalized $CD8^+$ cells, suggesting that it is the long-term selection dictating the observed difference in tropism. It should be noted that it is difficult to assess the relevance of these assays as the proviral load measured was remarkably high (up to 80% of cells), which is unlikely to be the case in the early stages of infection in healthy human carriers of HTLV-1.

6.1.2 Clonality of HTLV-1 infected $CD8^+$ cells.

The issue of the clonality of infected $CD8^+$ cells has been overlooked in the HTLV-1 field. As HTLV-1 preferentially infects $CD4^+$ T-cells and ATLL is largely a $CD4^+$ T cell malignancy, the standard model of HTLV-1-driven transformation (Figure 6.1) focuses on HTLV-1-infected $CD4^+$ cells which clonally expand (presumably driven by viral genes and controlled by the immune response); some of the expanded clones are driven to transformation by somatic mutations.

Therefore, previous clonality studies by us (Gillet et al., 2011) and others (e.g. Cavrois et al., 1998) have analysed DNA extracted from total PBMCs, with the underlying assumption that the clonality in PBMCs is equivalent to the clonality in each infected subset of the PBMCs.

Several groups (Eiraku et al., 1998; Ureta-Vidal et al., 2001) have previously reported oligoclonality of both $CD4^+$ and $CD8^+$ in HTLV-1 infected individuals (in all clinical manifestations), however, these studies were done by different methods measuring or even quantifying clonality of CD3 variable β chain, without distinguishing between infected and uninfected clones.

In recent years, a group led by Eric Wattel has carried out a series of experiments to investigate the difference between infected $CD4^+$ and $CD8^+$ cells by generating infected and uninfected $CD4^+$ and $CD8^+$ T cell clones by limiting dilution from a cohort of HAM/TSP patients. These experiments showed a higher rate of proliferation in infected clones than uninfected clones in both $CD4^+$ and $CD8^+$ populations; however the authors concluded that the mechanisms differed between the two cell types: while infected $CD4^+$ clones showed an increase in transition through the cell cycle (thus, increased replication), infected $CD8^+$ clones showed a significant decrease in the proportion of cells undergoing apoptosis (Sibon et al., 2006), a cIAP-2 dependent mechanism (Zane et al., 2010). In addition, by using restriction enzyme IPCR to detect proviral integration sites, they were able to show that expanded clones were present in both the $CD4^+$ and $CD8^+$ cell populations.

Using tetramers containing the immunogenic epitope Tax11-19 in the context of HLA-A*0201, and comparing to a non-HTLV-1 epitope, Hanon et al. (2000b) have previously shown that Tax-specific CTLs were more likely than EBV-specific CTLs to express Tax themselves. This observation suggests that HTLV-1-infected $CD8^+$ cells may often be the same cells that control the viral infection, and that clonal expansion of these cells may have an important role either in the dynamics or pathogenesis of HTLV-1.

6.1.3 Aim

This work aims to test the hypothesis that the shape of the clonal distribution of infected $CD8^+$ cells differs from that of $CD4^+$ cells, and to investigate the role of infected $CD8^+$ cells in determining the clonal distribution of the load.

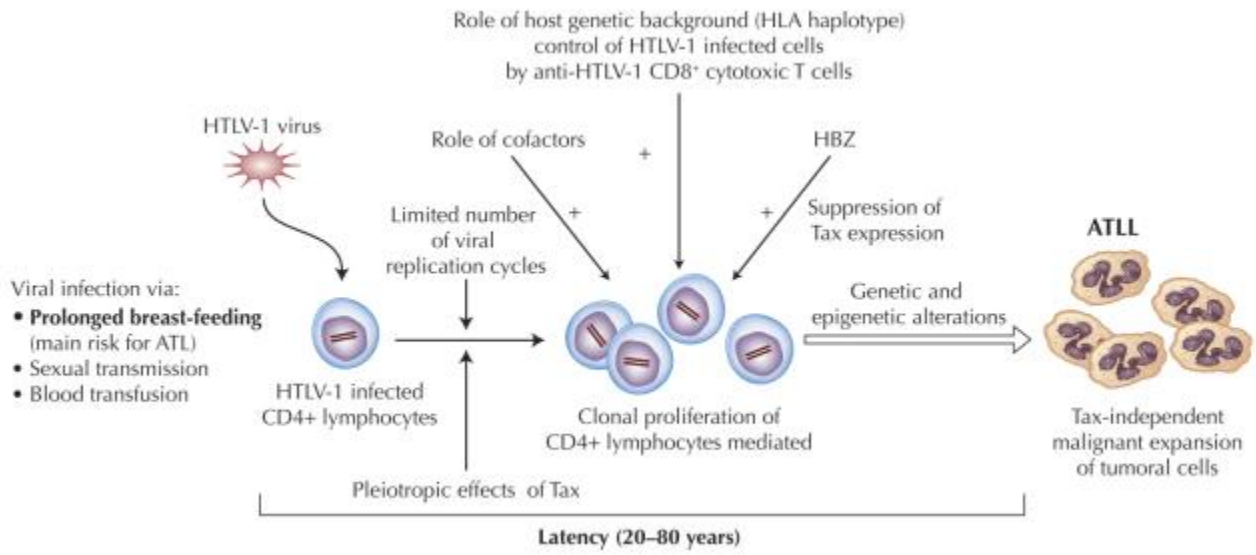

Figure 6.1: Standard model of HTLV-1 latency.
Infection of $CD4^+$ lymphocytes results in long term clonal expansion regulated by viral genes and external cofactors (e.g. co-infections such as Strongyloides stercoralis) and controlled by CTL response. Additional somatic mutations are then thought to allow certain clones further expansion and malignant transformation. Figure adapted from Mahieux and Gessain, 2007.

6.2 Results

6.2.1 Magnetic associated cell sorting for $CD4^+$ and $CD8^+$ cell populations

PBMCs from 12 HTLV-1-infected individuals were sorted by using magnetic associated cell separation for $CD4^+$ and $CD8^+$ cell populations. Details on the patients studied here are shown in Table 6.1. I selected patients with a relatively high proviral load (lowest proviral load in PBMC - 3.6 copies per 100 PBMC; Table 6.2) to ensure sufficient proviruses for analysis. The 12 patients included 6 HAM/TSP patients and 6 healthy carriers. One of the asymptomatic carriers (HGL) was diagnosed with HAM/TSP symptoms approximately 1 year after the sample was taken.

Purity of samples was assessed using flow cytometry, and proviral load of unsorted cells, $CD4^+$ and $CD8^+$ populations was quantified using QPCR. The HTLV-1 proviral integration sites were identified and quantified using the high-throughput method described in chapter 3.

The purity of cell populations isolated from each subject is shown in Table 6.1.The median frequency of contaminating $CD8^+$ cells in the $CD4^+$ fraction was 0.94% (range 0.43-5.44%) and the median frequency of contaminating $CD4^+$ cells in the $CD8^+$ cell fraction was 0.47% (0.07-2.07%). Based on the assumption that HTLV-1 infects differentiated, mature $CD4^+$ or $CD8^+$ single-positive cells, we regarded any integration site found in both $CD4^+$ and $CD8^+$ cells as a contaminating site and allocated it to the sample (either $CD4^+$ or $CD8^+$ cells) in which it was identified with a higher frequency (greater number of sisters). The number of contaminating sisters was inversely correlated with the purity of the cell populations, and there was a highly significant correlation between the proportion of $CD8^+$ cells in the PBMC and the proportion of the load carried in $CD8^+$ cells (Figure 6.2).

The proviral load was measured using quantitative PCR. The median proviral load in $CD4^+$ cells and $CD8^+$ cells was 12.3 copies (6.0-30.2) and 2.0 (1.1-6.2) copies per 100 cells, respectively.

The proportion of the load carried by the $CD8^+$ cells was calculated based on the proviral load measured and the proportion of PBMCs in each cell population (see methods). The median percentage of the proviral load represented by $CD8^+$ cells was 5.02% (2.29%-35.32%). See Table 6.2 for full details.

The proportion of the load carried by the $CD8^+$ cells was then further verified by the proportion of sisters (in the unsorted sample) attributed to the $CD8^+$ cells (across all $CD8^+$ clones found in the unsorted sample). There was a highly significant correlation between the results of the two methods of calculation of the $CD8^+$ cell contribution to the load (Figure 6.4; Pearson linear regression, $p < 10^{-6}$, $r = 0.969$).

Table 6.1: Patients samples used in chapter 6

Patient			Absolute counts[a]		Populations within T cells		Populations in PBMC		Purity test	
code	Clinical diagnosis	Known comorbidities[a]	$CD4^+$	CD8	$CD4^+$ / $CD3^+$	$CD8^+$ / $CD3^+$	$CD4^+$ in total	$CD8^+$ in total	$CD8^+$ in $CD4^+$	$CD4^+$ in $CD8^+$
HBX	AC	None	NA	NA	63.9	33.7	45.2	13.8	1.15	2.07
HBZ	AC	None	948	284	51.8	41.8	58.3	11.2	0.915	0.985
HCP	AC	None	732	335	53.5	33.5	32.7	10.5	1.1	1.46
HEZ	AC	Hepatitis[b]	631	214	53.8	34.3	41.9	13.9	0.801	0.069
HGL	AC[c]	None	835	378	63.7	33.4	47.7	14.1	0.684	0.364
HHD	AC	None	1316	419	67.4	30	44.8	13.7	0.469	0.693
TAN	HAM/TSP	None	588	260	42.9	48.5	47	12.4	1.95	0.11
TAZ	HAM/TSP	None	1560	1058	44.1	44.6	47.8	27.1	0.858	0.573
TBW	HAM/TSP	None	300	1254	12.4	85.6	26.3	58.1	5.44	0.077
TDB	HAM/TSP	None	838	406	43.5	44	53	12.5	1.53	0.29
TDL	HAM/TSP	Shingles	1526[d]	707[d]	57.9	38.4	45.5	18.5	0.43	0.337
TDT	HAM/TSP	None	752	392	50.6	41.5	44.4	15.3	0.959	0.573

NA – not available.

[a] Absolute cell counts and data on comorbidities were supplied by Dr. Maria Antonietta of Dr. Graham Taylor's group.

[b] HEZ – Hepatitis of unknown origin (negative for HCV, HBV)

[c] HGL – considered to be asymptomatic carrier at time of blood sample, but was diagnosed with HAM/TSP about. a year later.

[d] TDL – absolute cell counts from an earlier timepoint (1 month prior)

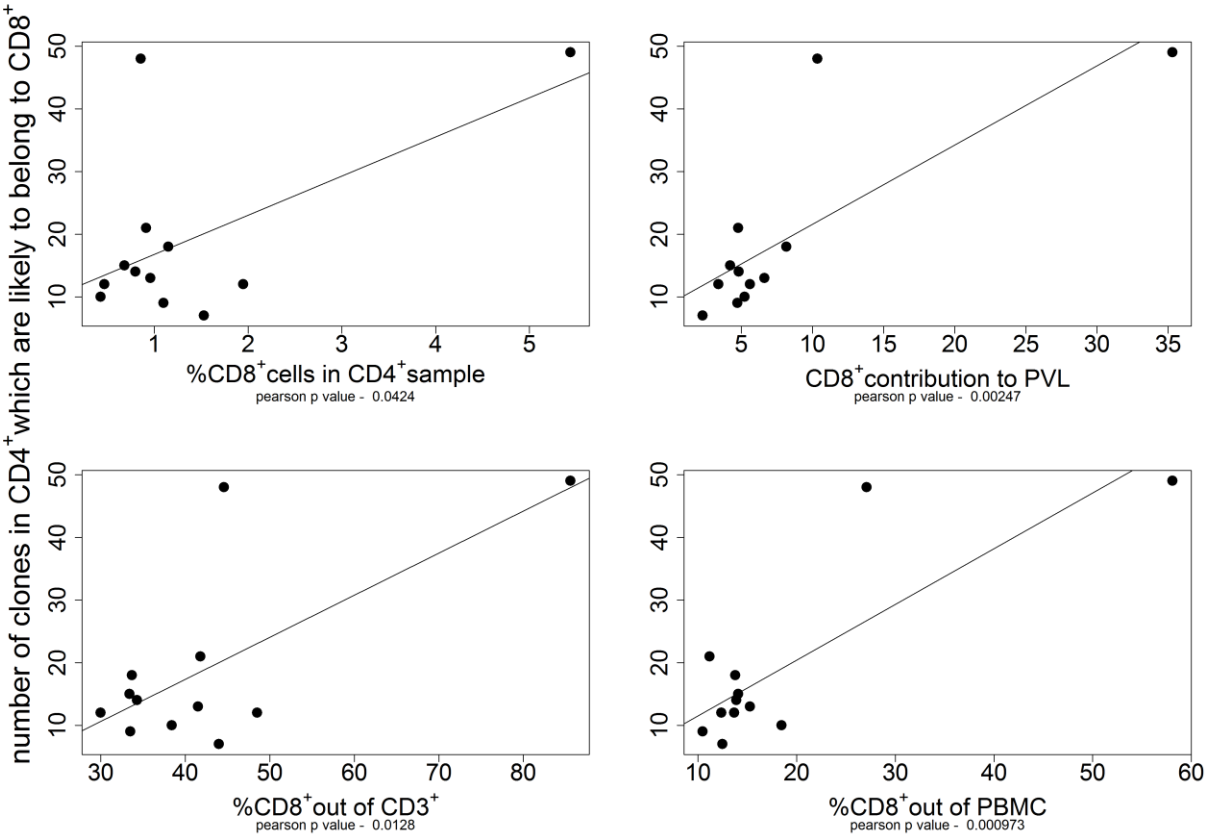

Figure 6.2: Analysis of contaminating integration sites.

Each integration site was attributed to either to $CD4^+$ or $CD8^+$ cells based on frequency in each sample. The number of clones in the $CD4^+$ samples that were attributed to $CD8^+$ cells is plotted here against potential causes of contamination. The strongest correlate was the proportion of $CD8^+$ cells in the PBMCs.

Table 6.2: The viral burden in each T cell subset.

Patient	Proviral Load (copies per 100 cells)			Proportion of load present in each population (%)	
	PVL[$CD4^+$]	PVL[$CD8^+$]	PVL[PBMC]	$CD4^+$	$CD8^+$
HBX	6.0	1.8	3.7	91.8	8.2
HBZ	6.5	1.7	3.8	95.2	4.8
HCP	10.1	1.6	4.6	95.3	4.7
HEZ	11.8	1.8	5.3	95.2	4.8
HGL	12.8	1.9	4.3	95.8	4.2
HHD	17.4	2.0	4.7	96.6	3.4
TAN	10.2	2.3	6.4	94.4	5.6
TAZ	30.2	6.2	11.3	89.7	10.4
TBW	17.8	4.4	9.4	64.7	35.3
TDB	10.6	1.1	3.6	97.7	2.3
TDL	22.9	3.1	8.0	94.8	5.2
TDT	17.3	3.6	8.0	93.4	6.6
Median	**12.3**	**2.0**	**5.0**	**95.0**	**5.0**

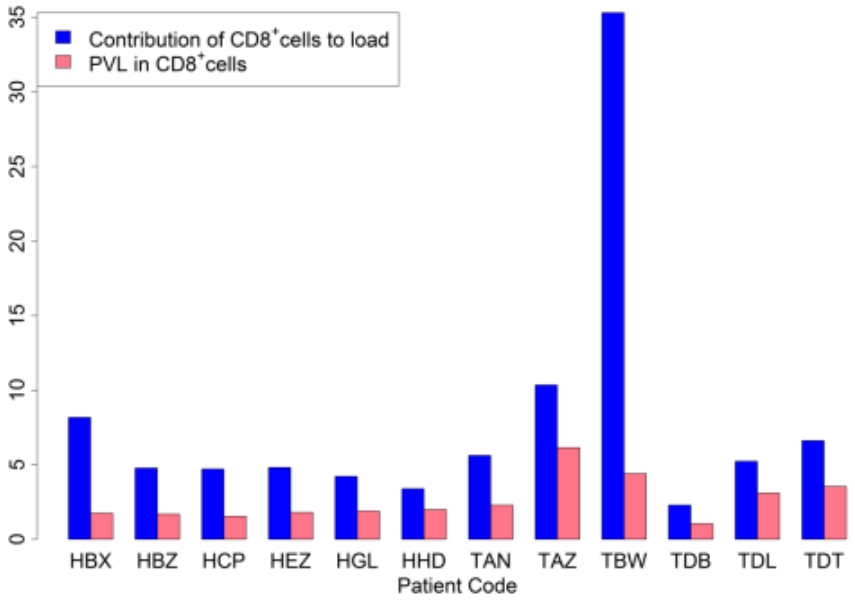

Figure 6.3: The contribution of $CD8^+$ cells to the proviral load.

Pink – the proviral load measured (QPCR) in $CD8^+$ cells (copies per 100 cells). Blue – the percentage of the proviral load carried in the $CD8^+$ cells, estimated based on PVL measured in $CD4^+$ and $CD8^+$ cell population and flow cytometry data.

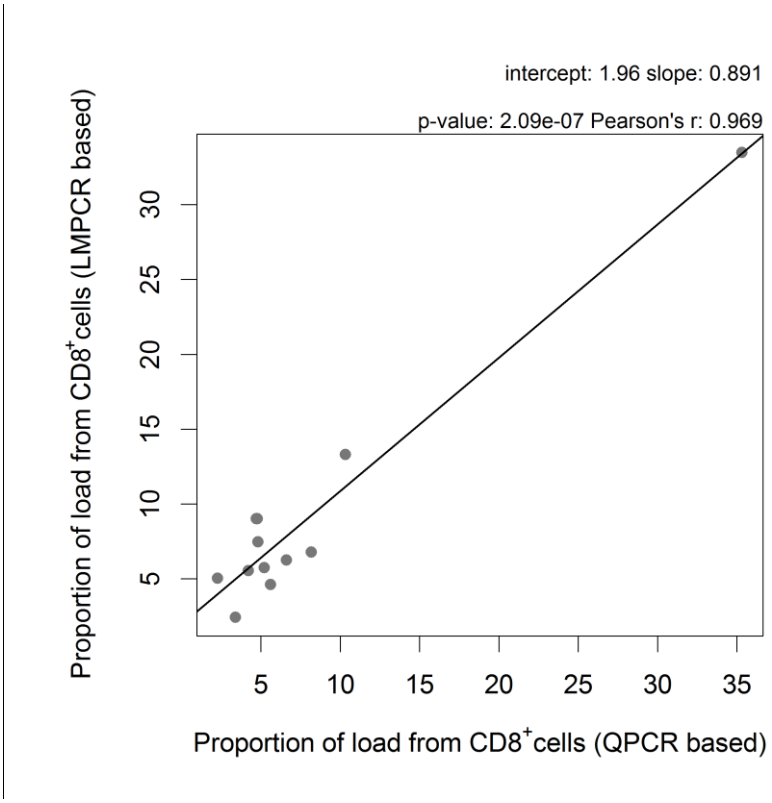

Figure 6.4: Two methods of estimations of the proportion of the load contributed by $CD8^+$ cells. The proportion of the load estimated from the proportion of sisters belonging to $CD8^+$ clones (from the high-throughput sequencing analysis) was very similar to the estimate from PVL measured in each population and the flow-cytometry data.

6.2.2 The role of infected $CD8^+$ cells in shaping the proviral load

We wished to test the role of the infected $CD8^+$ cells in shaping the proviral load in each sample. The proviral load in PBMC was strongly correlated ($p<10^{-6}$, Pearson's $R = 0.963$) with the proviral load in $CD8^+$ cells, and also correlated ($p<10^{-3}$, Pearson's $R = 0.864$) with the proviral load in $CD4^+$ cells (Figure 6.5).

We then examined the proviral load distribution in each fraction. Samples from $CD8^+$ cells had fewer sisters (Figure 6.6; Right panel), which could be predicted from the lower load in these samples; however, the number of distinct clones in the $CD8^+$ samples was even more remarkably lower than in $CD4^+$ samples (Figure 6.6; left panel).

The oligoclonality index was significantly higher in $CD8^+$ samples (median = 0.60) than in $CD4^+$ samples (median = 0.53). That is, there was a higher diversity of distribution within $CD4^+$ clones (Figure 6.7). There was no significant difference in OCI between asymptomatic carriers and HAM/TSP patients, in agreement with a previous report (Gillet et al., 2011). However, the difference in OCI between the $CD8^+$ and $CD4^+$ samples was only statistically significant for HAM/TSP patients (Figure 6.8). See Figure 6.9 for a characteristic distribution found in $CD4^+$ and $CD8^+$ samples from patient HCP. Whereas the oligoclonality index of HTLV-1 in $CD4^+$ cells was often determined by one or few expanded clones, then a sharp decline and a long tail of singletons (clones which we only observe once), the $CD8^+$ distribution was often characterized by a significantly smaller number of distinct clones and a larger number of expanded clones, and a much smaller proportion of the load was made up by singletons.

Finally, we aimed to quantify the relative role of $CD4^+$ cells and $CD8^+$ cells in determining the clone frequency distribution in unsorted PBMCs. In order to do that, we compared the clones identified as $CD4^+$ or $CD8^+$ to the clones found in the unsorted sample. See Figure 6.10 for detailed (colour coded) distribution of the top 50 clones in each patient; only one out of the 600 clones was not initially found in either $CD4^+$ or $CD8^+$, however, after a close inspection of the sequence, we believe that clone to be a $CD8^+$ clone which was mapped differently owing to a sequencing error. Unexpectedly, in 5 out of the 12 infected individuals (4/6 HAM/TSP patients) the largest clone was a $CD8^+$ clone. In 8/12 individuals (including all HAM/TSP patients) a $CD8^+$ clone was present among the 3 largest clones. An extreme case was patient TBW (who is known to have a distorted CD4/CD8 ratio, see Table 6.1), in whom the proviral load was dominated by a large number of $CD8^+$ clones, including the largest clone which represents over 15% of the load.

In order to answer the question whether $CD8^+$ clones are more likely to be major clones than expected by chance, we compared the proportion of clones with an absolute abundance of more than 1 copy per 10000 PBMCs between $CD4^+$ and $CD8^+$ clones (Figure 6.11A). This proportion was significantly greater in $CD8^+$ clones (~10%) than in the $CD4^+$ clones (~5%; $p = 5.36 \times 10^{-10}$, Fisher's exact test). Ranking all clones in descending order of abundance (Figure 6.11B), $CD8^+$ clones were over-represented among the 10 most abundant clones (clone 1-10) compared to the next 10 (clones 11-20; $p = 0.01$, Fisher's exact test).

The contribution of $CD8^+$ cells to the load was correlated with the proviral load in unsorted cells (Figure 6.12A) and with the proviral load in $CD8^+$ cells (Figure 6.12B). This association approached significance when all patients were included ($p = 0.06$ for both tests, Pearson linear regression), and was highly significant when excluding patient TBW, who had an altered CD4/CD8 ratio, and was consequently a significant outlier ($p =0.02$ for association with proviral load, $p = 0.004$ for association with PVL in $CD8^+$ cells). No significant association was found between proviral load in $CD4^+$ cells and the contribution of $CD8^+$ cells to the load.

Lastly, we wished to compare the proviral load to the oligoclonality index, as a measure of non-uniformity in the clone frequency distribution. We found that, as observed before (Gillet et al 2011), there was no correlation between oligoclonality index and proviral load in the unsorted PBMCs (Figure 6.13; top panel), or in the $CD4^+$ cells (Figure 6.13; bottom left panel). In the $CD8^+$ cells, however, we observed a positive correlation approaching significance ($p = 0.06$) between the oligoclonality and proviral load in $CD8^+$ cells (Figure 6.13; bottom right panel).

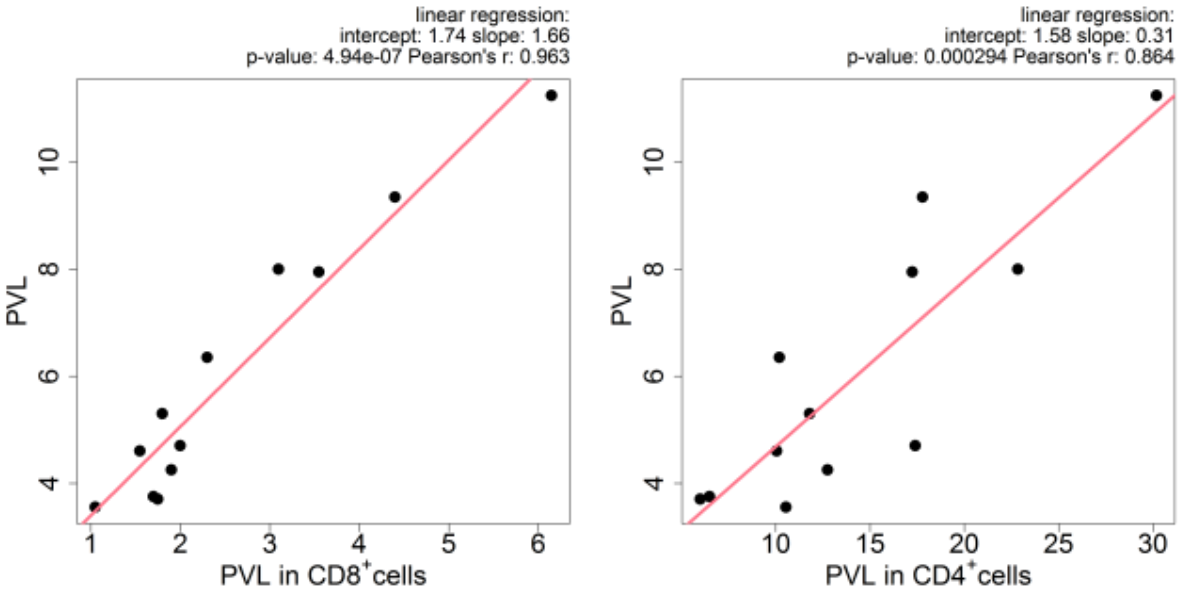

Figure 6.5: Proviral load (PVL) in $CD8^+$ cells (left) and in $CD4^+$ cells(right) positively correlates with the PVL in unsorted PBMCs.

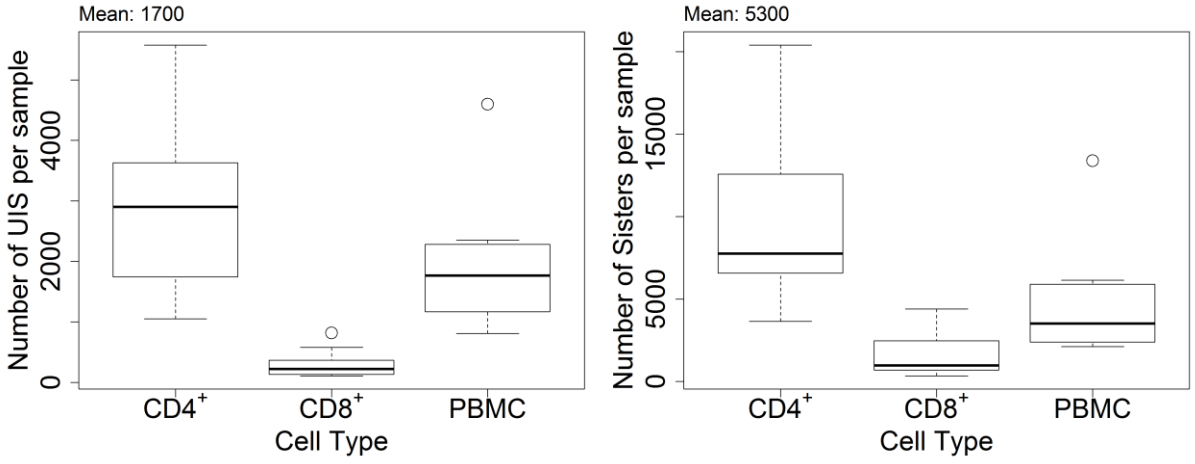

Figure 6.6: The $CD8^+$ infected cells were characterised by a small number of clones.

The $CD8^+$ samples tested were characterized by a lower amount of sisters (proviruses; right), owing to a lower load and a lower amount of available DNA, however the difference in clone numbers was even more remarkable. (Median total sisters per $CD4^+$ sample~7750, per $CD8^+$ sample~960. Median total clones per $CD4^+$ sample ~2900, per $CD8^+$ sample-220).

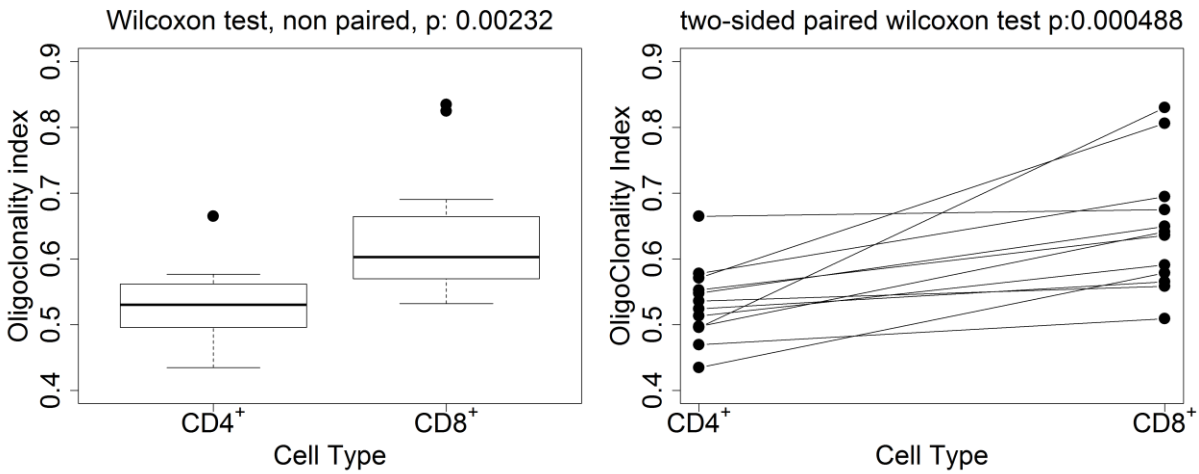

Figure 6.7: Oligoclonality index was significantly higher in infected $CD8^+$ cells than in $CD4^+$ cells.

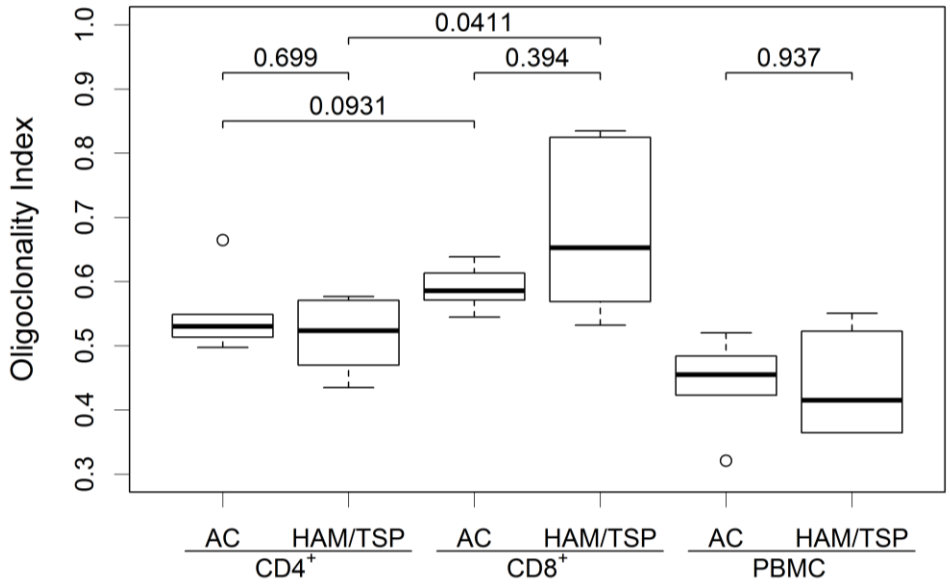

Figure 6.8: Greater difference between OCI[$CD8^+$] and OCI[$CD4^+$] in HAM/TSP patients than ACs.

No significant difference was found (Wilcoxon test, see individual p.values above each comparison) in the unsorted sample between HAM/TSP patients and asymptomatic carriers (AC); however the difference in OCI between $CD4^+$ and $CD8^+$ samples was only statistically significant for the HAM/TSP patients.

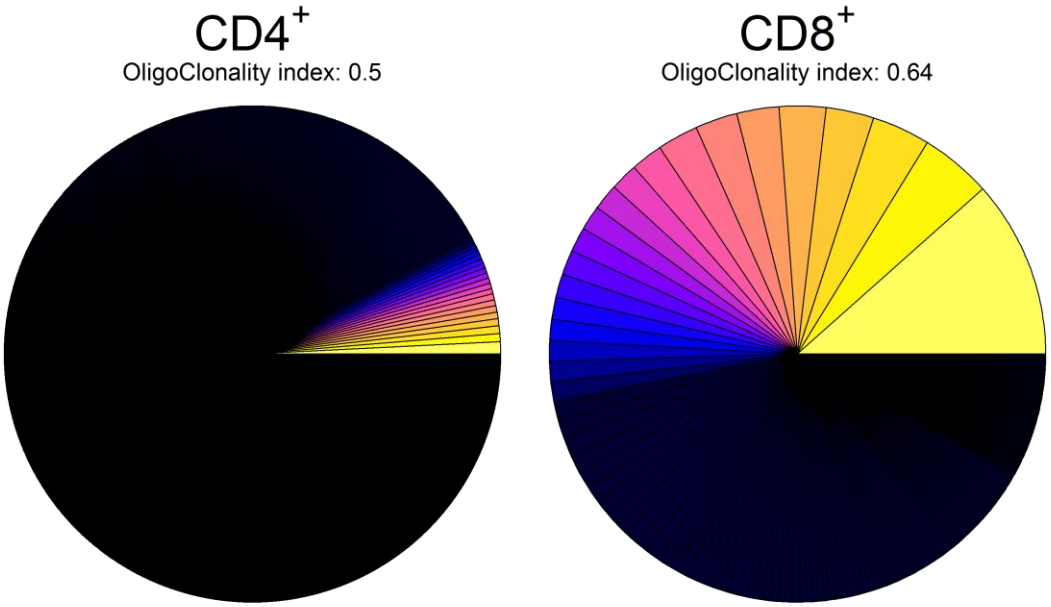

Figure 6.9: Example of clonal distribution in $CD4^+$ and $CD8^+$ infected cell populations.
The top 20 clones from each sample are highlighted in colour, the rest are shown in dark blue. $CD4^+$ (left) cell populations are frequently characterized by a large number of clones, and the OCI is increased by one or few expanded clones. In contrast, $CD8^+$ (right) cell populations are frequently characterized by a small number of relatively large clones, increasing the OCI. Patient shown here is HEZ.

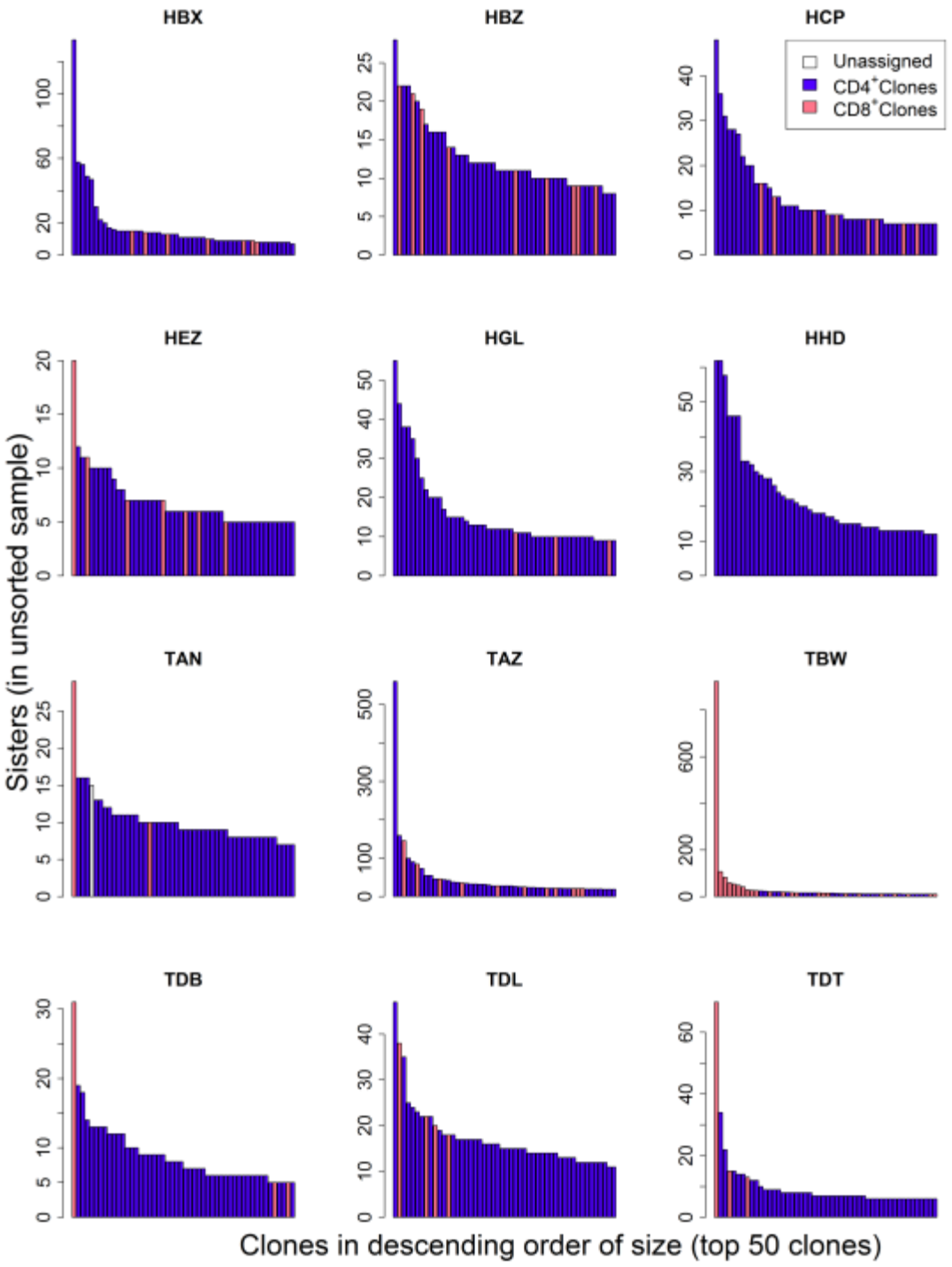

Figure 6.10: The CD8$^+$ clones are over-represented in the top 50 clones from each patient.
The 50 most abundant clones are shown for each patient, using representative colours for each patient based on which cell it is attributed to. The CD8$^+$ clones (peach) represented the largest clone in 5/12 patients, and represented at least one clone within the top three of 8/12 patients (including all HAM/TSP patients).

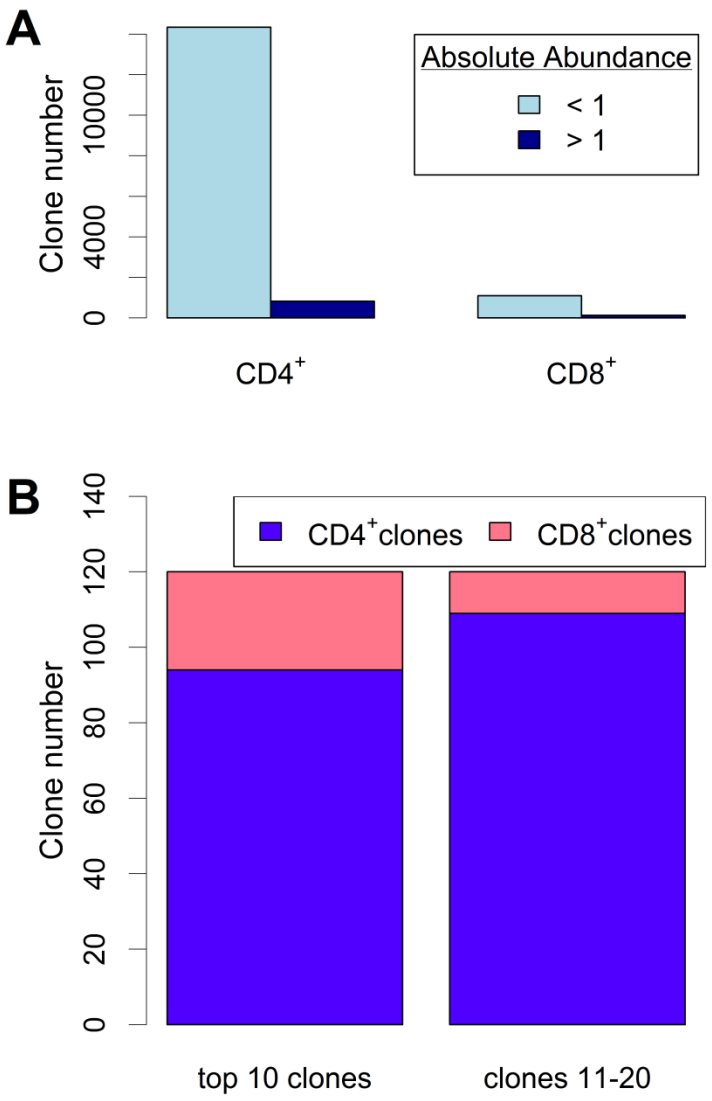

Figure 6.11: CD8$^+$ clones were significantly over-represented in high-abundance clones.

(A) The majority of clones (both CD4$^+$ and CD8$^+$ clones were low abundance (fewer than 1 copy per 10000 PBMCs); however, the CD8$^+$ clones were over-represented (10% of all CD8$^+$ clones vs 5% of all CD4$^+$ clones) in the clones of high abundance (greater than 1 copy per 10000 PBMC; $p = 5.36 \times 10^{-10}$, Fisher's exact test). (B) Ranking all clones for each patient by decreasing order of size, the CD8$^+$ clones were significantly more likely to be found in the top 10 clones than in the following 10 clones ($p = 0.01$, Fisher's exact test).

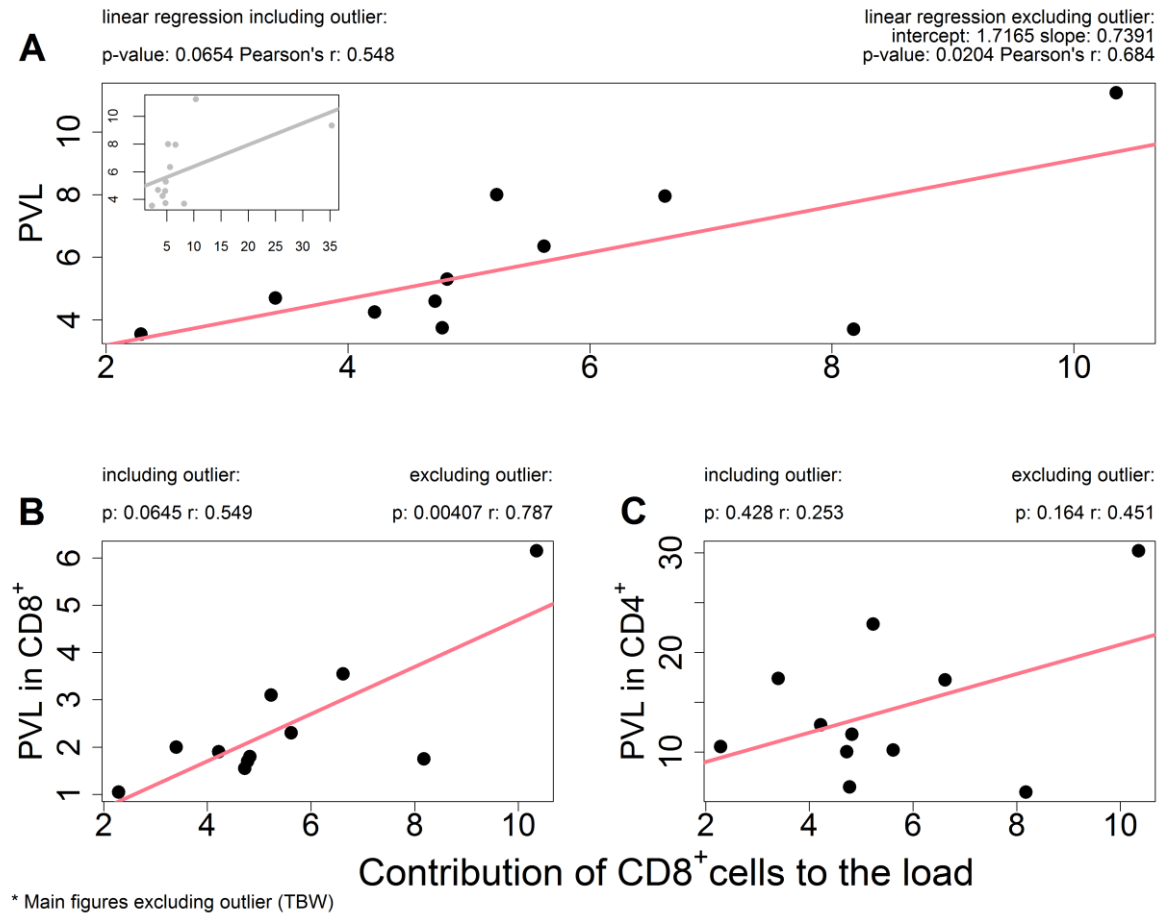

Figure 6.12: The proviral load is associated with the contribution of $CD8^+$ cells to the load. Contribution of $CD8^+$ cells to the load (in percentage) to the (A) proviral load in unsorted PBMC, (B) proviral load in $CD8^+$ cells and (C) proviral load in $CD4^+$ cells. One outlier (patient TBW) was omitted to enable a clear assessment of the other 11 individuals.For comparison, the outlier has been included in the inset shown in panel (A) Statistical tests are shown including (left) or excluding (right) the outlier. (A) The contribution of $CD8^+$ cells to the load correlated with proviral load in unsorted cells. (B) The contribution of $CD8^+$ cells to the load strongly correlated with proviral load in $CD8^+$ cells. (C) The contribution of $CD8^+$ cells to the load did not correlate with proviral load in $CD4^+$ cells.

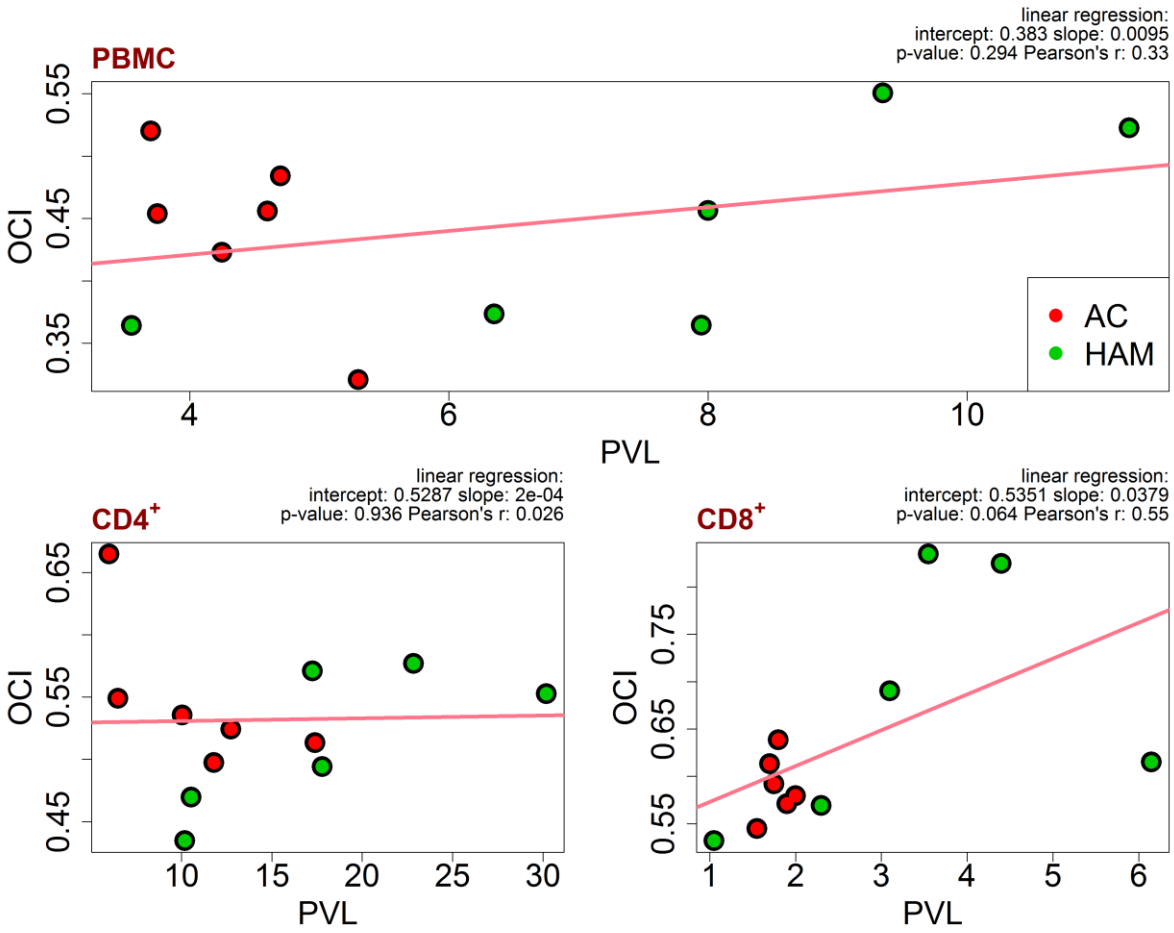

Figure 6.13: Proviral load in CD8$^+$ cells correlated with the oligoclonality index (OCI).
The proviral load (PVL) in unsorted cells (top), CD4$^+$ cells (bottom left) and CD8$^+$ cells (bottom right) was compared to the (OCI) in each cell population. PVL was quantified by PCR as copies per 100 cells. There was no correlation in unsorted cells or in CD4$^+$ cells, but there was a correlation approaching statistical significance in CD8$^+$ cells. Values for Pearson linear regression are shown for each panel. Using a ranked correlation (Spearman), there was no correlation found between PVL and OCI in either unsorted (PBMC) or CD4$^+$ cell samples. There was a significant ($p = 0.02$) correlation between proviral load and OCI in CD8$^+$ cell samples.

6.2.3 The role of infected CD8+ cells in shaping cell populations.

The proportion of load carried by the $CD8^+$ cells is influenced by the load in the $CD8^+$ cells but also by the cell populations in vivo. We wished to test whether there is any association between the $CD4^+$ or $CD8^+$ populations in the peripheral blood and the proviral load in those populations. We found that the proviral load in $CD8^+$ cells, and to a lesser extent the proviral load in $CD4^+$ cells, was positively correlated ($p<10^{-4}$ and $p<10^{-3}$, respectively when excluding TBW, $p = 0.02$ and $p = 0.156$, respectively when including TBW) with the percentage of $CD8^+$ cells in PBMCs (Figure 6.14A,B). No such correlation was observed with the percentage of $CD8^+$ cells within total $CD3^+$ cells (Figure 6.14C,D) or with the percentage of $CD4^+$ cells within PBMC (Figure 6.14E,F), suggesting that this is a virus-driven clonal expansion of $CD8^+$ cells rather than an alteration of the CD4/CD8 T cell ratio.

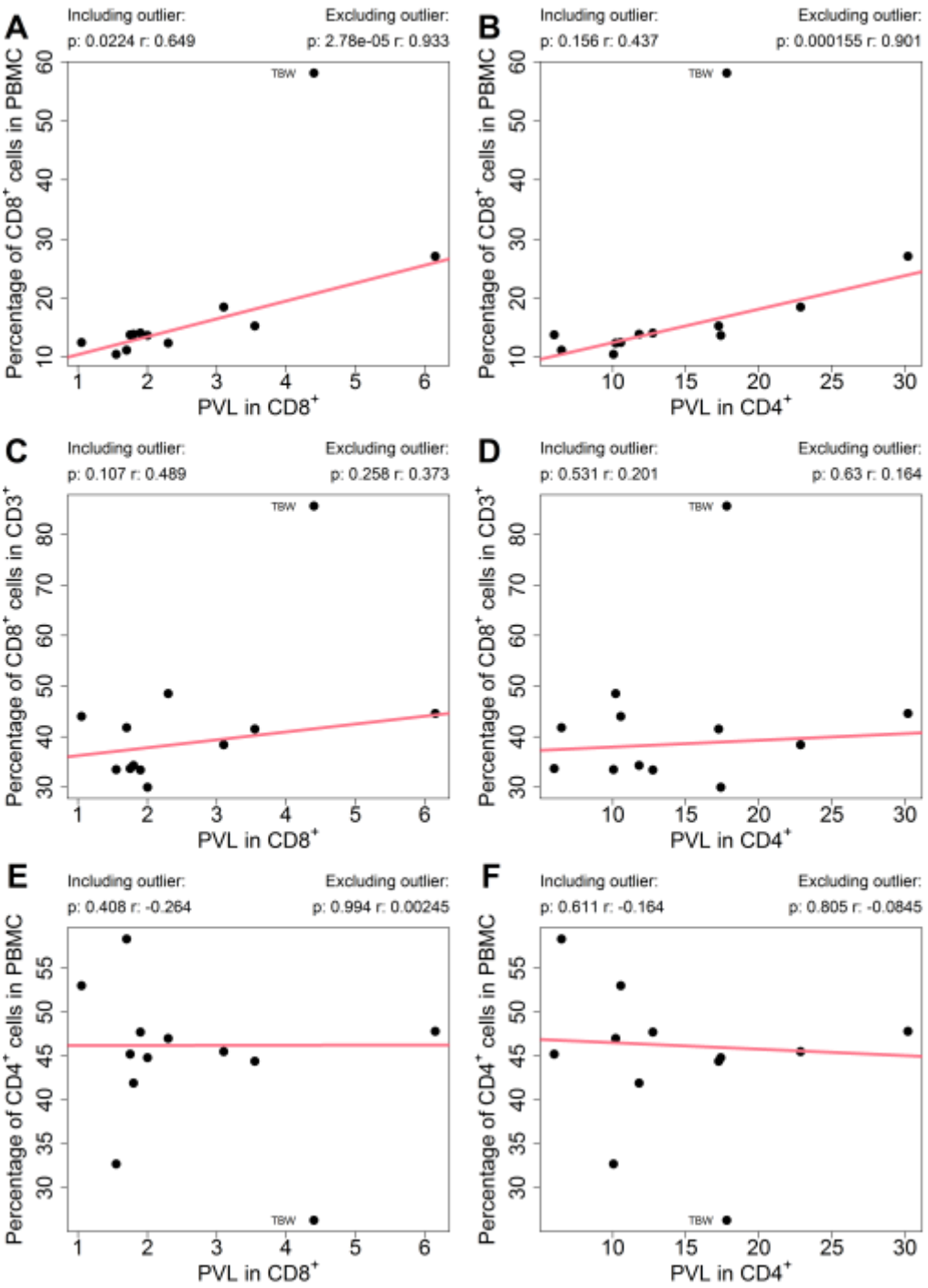

Figure 6.14: The proviral load in $CD8^+$ cells was strongly associated with the proportion of $CD8^+$ cells in PBMCs.

Comparisons of the proviral load in $CD8^+$ cells (A,C,E) and $CD4^+$ cells (B,D,F) with the proportion of $CD8^+$ cells in PBMC (A, B), $CD8^+$ cells in CD3+ (C, D), $CD4^+$ cells in PBMC (E, F). Two p values and Pearson's r are shown for each correlation test, respectively including or excluding the outlier (TBW, highlighted in each plot) – a $CD4^+$ lymphopenic patient. See text and Table 6.1 for details.

6.3 Discussion

HTLV-1 is mostly found in $CD4^+$ cells in vivo, and ATLL is typically a malignancy of $CD4^+$ cells. Therefore, the role of infected $CD8^+$ cells is often not directly investigated, and even overlooked in HTLV-1 research. In this work the aim was to investigate the role that infected $CD8^+$ cells play in shaping the proviral load.

We have observed that the proviral load within $CD8^+$ cells (i.e proportion of provirus-positive $CD8^+$ cells) correlates with a patient's total proviral load (proportion of provirus-positive PBMCs). This could be simply explained by the infected $CD8^+$ cells being a sub-population of all infected cells, carrying the virus to a somewhat lesser extent than in $CD4^+$ cells as previously thought. In order to investigate these infected cells, I carried out a high-throughput analysis of the proviral integration sites in $CD4^+$, $CD8^+$ or unsorted total PBMCs.

We have observed that while the infected $CD8^+$ cells represent the minority of infected cells (median of just over 5% of all infected cells), their clone frequency distribution is significantly different from that of infected $CD4^+$ cells. We quantified this difference in distribution in a number of ways.

First, our measure of distribution, the oligoclonality index, was significantly higher in $CD8^+$ cell samples than in $CD4^+$ cell samples, indicating that the $CD8^+$ have a more unequal frequency distribution of infected clones. The oligoclonality index was significantly higher in $CD8^+$ cells than in $CD4^+$ cells in nearly all patients, but in particular was higher among HAM/TSP patients compared to asymptomatic carriers. It is difficult to say, however, whether the difference in statistical significance is in itself of importance to clinical outcome, or whether it reflects the somewhat lower proviral load in asymptomatic carriers, which also affects the oligoclonality index.

Investigating this difference in distribution, I found that in non-malignant cases, $CD4^+$ cells may have a small number of large clones, and that the proviral load in $CD4^+$ cells is dominated by a long tail of very small clones (most of which were observed only once). In contrast, the infected $CD8^+$ clone distribution is often characterized by a larger proportion of relatively large clones among a lower total number of clones. This observation suggests that infected $CD8^+$ cells are subject to selection forces different from those acting on infected $CD4^+$ cells. The oligoclonality index in $CD8^+$ cells positively correlated with the proviral load in $CD8^+$ cells, while no such correlation was observed in the $CD4^+$ cells. This is consistent with the idea that while the load is increased in $CD4^+$ cells primarily by adding new clones, it is increased in $CD8^+$ by clonal expansion. In order to answer this question it will be required to improve our understanding of the ratio between mitotic and infectious spread at equilibrium of HTLV-1 infection. There is currently a mathematical project underway in our group modelling the infectious to mitotic ratio in order to answer such questions.

Assigning clones identified in the unsorted sample to the $CD4^+$ and $CD8^+$ populations, revealed that the role of the infected $CD8^+$ cells is more important than previously estimated - $CD8^+$ clones are often the largest (in 5/12 patients) or among the largest clones in each patient. In fact, $CD8^+$ clones were twice as likely as $CD4^+$ clones to have an absolute abundance of at least 1 copy per 10000 PBMC. It is clear that clonality analysis previously carried out (by us and others) included many large $CD8^+$ clones in the analysis; selection forces favouring $CD8^+$ clonal expansion in particular may not be apparent when examining clonality of unsorted PBMCs ($CD8^+$ clones represent approximately 10% of all clones over this size limit).

I observed that the total viral burden, as well as the proviral load within $CD8^+$ cells (the proportion of $CD8^+$ cells which are infected) correlated with the contribution of $CD8^+$ cells to the load (i.e. the proportion of infected cells which are $CD8^+$ cells), and that the proviral load within $CD8^+$ cells correlates with the percentage of $CD8^+$ cells in the PBMC ($CD8^+$/PBMC).

This study leads to several important questions:

1) Why are the $CD8^+$ clones more oligoclonal than the infected $CD4^+$ clones?

2) What mechanisms and selection forces shape the distribution of $CD8^+$ infected cells?

3) What is the role of infected $CD8^+$ cells in shaping the proviral load?

4) What are the implications for the immune response against HTLV-1?

5) What is the role of infected $CD8^+$ cells in HTLV-1-associated pathogenesis?

6.3.1 Selection forces and potential mechanisms determining clonal distribution of CD8+ cells

The observed selective expansion of a small number of infected clones suggests one of two (not mutually exclusive) main mechanisms driving the clonal expansion:

1) T-cell receptor (TCR) driven: the exposure to a specific antigen may be driving activation and selective expansion of infected clones. The most likely source of antigen is HTLV-1 itself, with Tax recognized to have the highest immunogenicity. However, other antigens may contribute to TCR-driven clonal expansion, such as cytomegalovirus.

2) Provirus driven: the genomic environment of the provirus (chapter 4) and the tendency of a clone to express proviral genes (chapter 5) appear to be associated with selective clonal expansion. Expression of proviral genes could have an indirect effect on clonal selection (e.g. by exposing the infected cell to the immune response). Expression may also have a more direct effect as Tax is known to drive proliferation (although these experiments are often done on $CD4^+$ cells) and to upregulate various cytokines and cytokine receptors which are known to be involved with regulating cell cycle and cell survival.

Are the large $CD8^+$ clones HTLV-1-specific? Hanon et al. (2000b) have previously demonstrated that Tax11-19-specific CTLs were themselves more likely to be expressing Tax. This is consistent with the idea that that infected $CD8^+$ are also specific to proviral antigens, perhaps being infected as a result of an increased chance of virus transmission during the cell-cell contact required for formation of the immunological synapse (Grakoui et al., 1999). We aim to test in the future whether the HTLV-1-specific cells are more likely to be infected and whether these belong to large $CD8^+$ clones (rather than small ones). Because a limited number of antigen specificities can be tested at one time, this may not be sufficient to conclude whether all large clones can be explained by a TCR-driven mechanism.

Oligoclonality of $CD8^+$ cells in health and disease has been described in the past (Monteiro et al., 1995; Batliwalla et al., 1996) and clonal expansion of HTLV-1 specific CTLs has been extensively described in the literature. As early as 1995, Elovaara et al reported a limited Tax-specific TCR repertoire, based on PCR analysis of TCR Vβ usage (Elovaara et al., 1995). The same group reported in 1996 oligoclonal expansion of Tax11-19-specific CTL in the blood of 3 HAM/TSP patients (Utz et al., 1996) by analysing Vβ and Vα genes and using a similar approach, it was then reported that HAM/TSP patients have a greater number of CTL clones (in blood and CSF) than asymptomatic carriers (Hoger et al., 1997). Advances in methods to quantify TCR clonal diversity allowed Ureta-Vidal and collegues to use an immunoscope method to characterize Vb distribution in CD3 cells from patiens with HAM/TSP and ACs, as well as uninfected controls. They reported that the degree of oligoclonality was higher in $CD8^+$ cells than in $CD4^+$ cells (both in infected and uninfected individuals), but higher (either $CD4^+$ cells or $CD8^+$ cells) in infected than uninfected. For HAM/TSP patients, oligoclonality in $CD8^+$ was significantly higher in HAM/TSP than AC (Ureta-Vidal et al., 2001).

Considering the possible association between infection and specificity, the clonal expansion described by these studies is consistent with that observed in this work; however, as none of these studies considered the proviral status of the Tax-specific cells, it is difficult to compare them to this work with certainty.

Sibon et al. (2006) used a limiting dilution cloning method to compare T cell clones derived from infected and uninfected $CD4^+$ and $CD8^+$ clones. They concluded that while infected $CD4^+$ T cells show an increased progression through the cell cycle (compared to uninfected clones), there was no significant difference infected $CD8^+$ clones and uninfected ones. In contrast, infected $CD8^+$ clones had significantly fewer cells undergoing apoptosis than

uninfected $CD8^+$ clones. They concluded that infected $CD8^+$ clones proliferate not through increased replication but through resistance to cell death. This may be related to a previous observation by another group, focusing on HTLV-1-specific CTLs (which may or may not be infected), demonstrating that Tax-specific CTLs from 4 of 5 patients tested were prevented from undergoing cell death by an IL-15 (and not IL-2) dependent mechanism (Azimi et al., 2001). The same study showed that this selective blockade of cell death by IL-15 was observed in particular in patients where the expression of IL-15R was confined to the Tax-specific CTLs. As Tax is known to upregulate IL-15R (Mariner et al., 2001), it is very likely that the effect observed by Azimi et al was in fact IL-15-mediated survival of HTLV-1-infected $CD8^+$ cells. The source of the IL-15 in vivo may be the same or other HTLV-1-infected cells: HTLV-1 Tax is also known to upregulate IL-15 (Azimi et al., 1998; Azimi et al., 2000). It will be necessary in future studies to assess the expression of IL-15R and other cytokines/receptors (e.g. IL-2, CD25 and others) which may also be implicated in clonal expansion of infected $CD8^+$ cells (Ku et al., 2000).

What is the role of Tax expression of infected $CD8^+$ cells in driving clonal expansion? It is known that in vitro they can express Tax protein (Hanon et al., 2000b); however it is not clear if the amount or frequency of Tax expression differs between infected $CD4^+$ and infected $CD8^+$ cells: Newbound et al. (1996) have shown that in vitro infected $CD4^+$ cells express a greater amount of viral RNA from the 5'LTR than $CD8^+$ cells, despite a similar load in both cell populations; it was later shown by Nagai et al. (2001a) that in cells taken from patients with HAM/TSP, Tax expression per infected cell was variable between patients but similar between $CD4^+$ and $CD8^+$ cell population within patients. My work comparing Tax expressing and non-expressing $CD4^+$ cells from 10 HAM/TSP patients (Chapter 5) revealed that in the $CD4^+$ cell population Tax-expressing cells were significantly associated with smaller clones. This observation is consistent with the idea that expression of Tax protein in $CD4^+$ cells

exposes the infected cell to cell killing by Tax-specific CTLs in vivo. But would infected $CD8^+$ cells show the same trend? While we know that Tax expressing $CD8^+$ cells *can* be killed by Tax-specific CTLs (Hanon et al., 2000b) it is not clear whether $CD8^+$ cells are as susceptible or whether this killing is as efficient in $CD8^+$ cells. It may be that the balance of effects of Tax (promoting proliferation as well as antigen presentation) is different in the case of infected $CD8^+$, which helps them to maintain a viral reservoir in relatively large expanded clones.

It remains to be tested what role the genomic environment plays in determining clonal expansion of $CD8^+$ clones.

6.3.2 Implications for disease pathogenesis

HAM/TSP

One of the main points of controversy in HAM/TSP research is the role of Tax-specific CTLs in inflammatory disease: are they protective or detrimental?

It has also been shown that the proviral load of HAM/TSP patients is correlated with the frequency of HTLV-1-specific CTLs (Kubota et al., 2000; Nagai et al., 2001b). A high frequency of Tax-specific CTLs are found in the T cell infiltrates in CSF of HAM/TSP patients (Jacobson et al., 1992; Hoger et al., 1997), as well as in the lung infiltrates of patients with pulmonary involvement (Kawabata et al., 2012). That, together with the observation that a high HTLV-1 proviral load is often found in T cell infiltrates in the CNS in HAM/TSP patients (Lezin et al., 2005; Demontis et al., 2013), led to the notion that specific infiltration of infected cells across the blood brain barrier attracts HTLV-1-specific CTL which cause the tissue damage by a bystander effect by secretion of cytokines such as IFNγ and TNFα (Ijichi

et al., 1993; Daenke and Bangham, 1994). In contrast, HTLV-1-specific CTLs are also thought to play a protective role, controlling the proviral load throughout the infection, as the efficiency of Tax specific CTLs negatively correlates with proviral load (Asquith et al., 2005a), and genetic determinants of MHC presentation to CTLs (particular HLA-I alleles) have been shown to be associated both with remaining a healthy carrier, and with lower proviral load in healthy carriers (Jeffery et al., 1999; Jeffery et al., 2000).

By shifting the focus onto the expanded infected $CD8^+$ clones, a third possibility emerges, which may help to reconcile seemingly conflicting findings: the high $CD8^+$ T cell frequency is often attributed to the higher proviral load in HAM/TSP, which leads to a higher antigen concentration. If indeed the expanded $CD8^+$ clones observed in this work are identified as being HTLV-1-specific this would lead to a much more refined understanding of the role of infected cells in driving pathogenesis, and raises the important question whether the $CD8^+$ infected cells preferentially infiltrate the CSF, are they cleared by the immune response as efficiently as infected $CD4^+$ T cells and are they as efficient in killing other infected cells? The answer to the last question is not clear, as conflicting data on a small number of infected clones shows that the infection either impairs or does not impair CTL function (Popovic et al., 1984; Mitsuya et al., 1986; Faller et al., 1988).

Our evidence demonstrates that the proviral load (and in particular the proviral load in $CD8^+$ cells) was positively correlated with the proportion of $CD8^+$ cells in the PBMC (Figure 6.14A,B). This correlation results from a selective expansion of particular populations. Since the infected cells represent the minority of $CD8^+$ cells (median of 1.95 copies per 100 $CD8^+$ cells) clonal expansion of the infected cells is not sufficient to explain this positive correlation; it is, however, consistent with the idea of a paracrine effect, for example expansion driven by cytokines secreted by infected cells. If so, this may have a role in

pathogenesis and may shed some light on the link between proviral load and risk of inflammatory disease.

ATLL

$CD8^+CD4^-$ ATLL is very rarely reported in the literature, and the best estimates put the $CD8^+CD4^-$ ATLL at 3.7% of cases (Kamihira et al., 1992). In addition, the phenotype of the infected cells may be altered during malignant transformation, because aberrant expression of certain cytokines (i.e. IL-4) could modify the expression of these factors (Paliard et al., 1988).

Within the standard model of ATLL development (Figure 6.1) the malignant transformation follows years of clonal expansion. Consistent with this, in the BLV-infected sheep model of ATLL development Moules et al. (2005) have shown that the premalignant clone is expanded and distinct in early in infection, and a sharp increase in genetic instability coincided with the onset of malignancy.

Should we have expected $CD8^+$ ATLL to be more frequent? It is still not known what the important features of a pre-malignant clone are. Without a better understanding of these features, it is difficult to assess the transformation potential of infected $CD8^+$ cells. Based on the standard model, we should expect to see more frequent $CD8^+$ malignancy as the $CD8^+$ infected clones are overrepresented within the most expanded clones (both when considered by rank order and by absolute abundance). This reasoning would suggest that if a necessary or sufficient precondition for malignant transformation is clonal expansion, there would be a far greater frequency of $CD8^+$ phenotype ATLL. The low frequency of $CD8^+$ ATLL is also

surprising considering the observed correlation between clonal expansion and increased proviral load, known to be a risk factor for ATLL (Iwanaga et al., 2010; Hodson et al., 2013).

What role does normal $CD8^+$ biology play in determining the incidence of $CD8^+$ ATLL? in other cutaneous lymphomas such as mycosis fungoides, the overwhelming majority of cases have a $CD4^+$ phenotype (Willemze et al., 2005), however $CD4^-CD8^+$ cells are not protected from malignancy and $CD8^+$ leukaemias are known , such as T cell Large granular lymphocytic leukaemia, which is most often a malignancy of $CD8^+$ cells (Chan et al., 1986; Dhodapkar et al., 1994).

The results presented here, demonstrating an over-representation of $CD8^+$ cell clones among the 50 most abundant clones, suggest that long-term clonal expansion is not a necessary or sufficient process for eventual transformation. There are previous reports on malignant transformation occurring on the background of entirely polyclonal population (Iwatsuki et al., 1990); this is also consistent with a recent observation based on high-throughput sequencing of longitudinal samples taken from ATLL patients, where the putative malignant clone was not always among the most expanded clones pre-malignancy (Cook et al, Unpublished).

A potential mechanism for the apparent lack of $CD8^+$ malignancies was proposed by Sibon et al. (2006), who observed an increase in aberrant cell morphologies in infected $CD4^+$ clones compared with $CD8^+$ clones, suggesting that the increased cell division they observed in $CD4^+$ clones (compared with reduced cell death observed in $CD8^+$ clones) results in an increased mutation rate in the $CD4^+$ cells. It is difficult, however, to interpret this confidently as the clones studied by Sibon et al were themselves selected and cultured ex vivo, and it is not clear whether in vivo there is a greater rate of mutations in infected $CD4^+$ cells.

Another potential explanation is that while $CD8^+$ clones can expand, there may be clonal succession over time (Snyder et al., 2008), and the infected subclones themselves do not

persist as long as their $CD4^+$ counterparts. This is, however, unlikely: in Gillet et al, large clones were reported to expand with time, rather than to disappear. As it now seems likely that about 10% of those clones were $CD8^+$, this is likely to have been observed in the PBMC data. In addition, Ureta-Vidal et al have reported observing long-lived $CD8^+$ clones in HTLV-1 infection (not shown), however they did not distinguish infected from non-infected clones(Ureta-Vidal et al., 2001).

A more targeted analysis of $CD8^+$ samples taken at different timepoints will need to be done in order to answer this question rigorously.

Another potential avenue for research is the related virus HTLV-2. HTLV-2 preferentially infects $CD8^+$ cells (Ijichi et al., 1992) and has not been clearly associated with any disease, although it has been linked to a higher mortality than uninfected individuals from all causes (Biswas et al., 2010). A separate study looking into the HTLV-2 proviral integration site distribution in HTLV-2 carriers is currently being carried out.

While the proviral load carried by infected $CD8^+$ cells has not been the subject of a great deal of research over the years, these cells may have great relevance to the field of HTLV-1 research. This present study is limited by the use of a relatively small group (12 patients), and by the intentional selection of patients with a high proviral load, who may not represent the majority of asymptomatic carriers. However, it is not likely that the basic molecular mechanisms of persistence will differ between high and low proviral load carriers. The present study highlights the difference in clonal distribution between infected $CD4^+$ and $CD8^+$ cells, which may reflect differences in the relative importance of the selection forces acting upon the infected cell clones. These may also lead the way to a better understanding of HTLV-1 related pathogenesis, and the selection mechanisms directing infected and non-infected cell expansion in HTLV-1 infected individuals.

Chapter 7 - **Discussion**

The set point of proviral load of HTLV-1 does not vary within an infected subject over the course of the infection, despite a strong, active immune response. Two proposed mechanisms for HTLV-1 escape from immune control are not sufficient to explain this. First, while Tax escape mutants have been found, they are not very frequent (in non-malignant infection) in keeping with an overall stable proviral sequence (Niewiesk et al., 1995). Second, the infection is not latent, and Tax expression does take place in vivo. Evidence from the investigation of the immune response (in particular the CTL response) to Tax and HBZ demonstrates that the CTLs can control Tax expressing cells in vitro and in vivo (Hanon et al., 2000a; Asquith and Bangham, 2008).

The proviral load remains the best prognostic marker of HTLV-1-related disease (Nagai et al., 1998; Matsuzaki et al., 2001; Iwanaga et al., 2010). The approach taken here is to investigate the mechanisms that take place during chronic infection and the selection forces acting on the load, by exploring the infected clones that make up the proviral load.

We have developed a new method, based on the new technology of high-throughput sequencing to analyse and quantify HTLV-1 integration site. This method enabled us two avenues of research. First, the mapping to the human genome, coupled with the growing number of annotations to the human genome (such as the ENCODE project, see Dunham et al., 2012) will allow us an unprecedented view of the characteristics of the genomic environment at the integration site, to better understand the mechanisms that HTLV-1 utilizes for integration; this may even lead to insights into therapeutic or prophylaxis approaches targeting this process. Second, the integration site can be used as an identifying marker, distinguishing an infected T cell clone, allowing us to test for the dynamic changes over the course of the infection not just as changes to the proviral load but as changes on thousands of

clones simultaneously. This can be used to monitor the response to therapy against ATLL or HAM/TSP, to examine the characteristics of cell infiltrates in inflammatory disease, or to track a potentially malignant clone before initiation of disease.

In this work I present the following main findings:

1) The initial targeting of HTLV-1 integration was equally favoured in genes in either relative orientation. There was a strong bias towards integration in proximity to specific TFBS, notably STAT-1, HDAC6 and p53.

2) The effect of the genomic environment on proviral expression and clonal expansion was largely asymmetric. An upstream host gene promoter was associated with silencing of Tax expression. The effect of integration in proximity to ChIP-seq sites of transcription factors and chromatin remodellers was often asymmetric, with the notable example of the SWI/SNF proteins, in particular BRG1.

3) Tax expression was associated with less expanded clones, suggesting that the cells more readily expressing Tax are efficiently controlled by the immune response in vivo.

4) While infected $CD8^+$ cells represent the minority of the proviral load, they are no less important – as shown by a high degree of oligoclonality (which positively correlated with the proviral load in those cells), $CD8^+$ clones were over-represented among the most expanded clones.

The work carried out in this thesis adds significantly to previous work (Gillet et al., 2011; Meekings et al., 2008) to build a growing picture of the factors associated with integration targeting, proviral expression and clonal expansion (Figure 7.1). Initial integration was

favoured in genomically active regions, in particular within genes or in proximity to gene start sites or CpG islands, or in gene-dense regions (Meekings et al., 2008). There was no initial bias for a particular relative orientation of the host genes, suggesting that the bias observed in vivo (favouring same-sense orientation) is entirely driven by long-term selection, rather than integration bias.

These results, in the context of previous findings, suggest two levels of control of the proviral expression. First, a regional effect - integration in genomically active regions favours proviral expression (Meekings et al., 2008, and this work). Second, a more local effect, within 10 kb of the integration site, where the existence of an upstream promoter of the same orientation as Tax, may silence the expression from the 5'LTR. One possible mechanism described which could account for this effect is the transcriptional interference (Shearwin et al., 2005). The observed preferential expansion of clones with same-sense proviral orientation is consistent with the observation that clones which do not express Tax (for example by mutation of Tax) gain a selective advantage and outgrow their counterparts.

Unlike the largely symmetrical effect of the host genome on initial integration, both the relative position of the nearest host gene (upstream or downstream of the provirus) and its relative orientation showed significant associations with clonal abundance and expression. It will be important to take this asymmetry into account in future analysis of the host environment flanking the retroviral genome.

Observations reported here demonstrate that it is not simply access to chromatin which determines proviral integration and expression. First, while integration in proximity to some TFBS showed a remarkable bias, this was not the case for others. Second, the role of relative position and orientation of host genes or TFBS found in regards to proviral expression and

clonal expansion indicate a mechanistic interaction between transcription of the HTLV-1 proviral genes and transcription of the flanking host genes.

What, therefore, determines clone abundance in vivo? While general transcriptional activity appears associated with Tax expression, it is also associated with proviral expansion (Gillet et al., 2011). A potential explanation is that the regional effect also determine the magnitude or rate of HBZ expression. HBZ has been suggested to play a role in cell proliferation, and is an inhibitor of Tax by sequestering host factors (Gaudray et al., 2002; Basbous et al., 2003). A potential role for HBZ in driving clonal expansion is also in line with analysis of host genetics which showed that possessing an HLA genotype predicted to bind strongly to HBZ peptides is protective in HTLV-1 (Macnamara et al., 2010). It appears that most infected individuals have a good immune response against Tax. This response shapes the clonal distribution and causes the Tax-expressing clones to remain small. This controls the virus over the long-term, allowing a subclinical persistent infection, which aids the spread of the virus in the population over several decades. The role of Tax may be in driving cell-to-cell transmission, adding new clones. Over time, large clones become more abundant (Gillet et al., 2011) and the chronic immune response leads to inflammatory disease. It is possible that individuals who also have an MHC genotype which allows them to identify HBZ expressing cells, are able to clear the Tax^{-} clones more efficiently, leading to a lower proviral load and a reduced risk of HTLV-1-related disease.

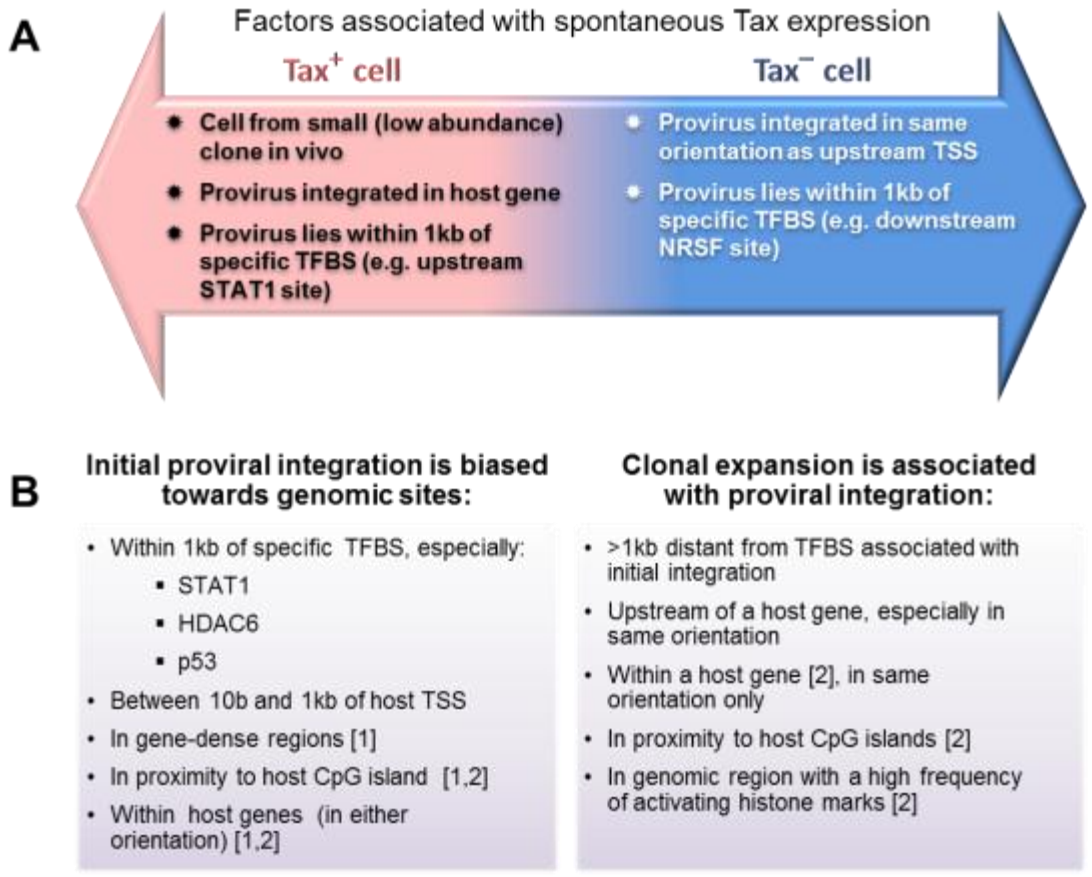

Figure 7.1: Summary: the emerging picture of factors associated with spontaneous Tax expression integration targeting and clonal expansion.

(A) Factors associated with presence or absence of spontaneous Tax expression of a given cell after overnight incubation. (B) Genomic environment features associated either with targeting of the initial integration (left panel) or clonal expansion in vivo (right panel). Findings were made in this work unless otherwise stated. 1 – Meekings et al., 2008, 2 –Gillet et al., 2011.

HTLV-1 is mostly found in $CD4^+$ cells in vivo, however a median of 5% of infected cells are $CD8^+$ cells. The infected $CD8^+$ cells examined here had a distinct clonal frequency distribution from that of infected $CD4^+$ cells, characterized by fewer, relatively expanded clones. The correlation between the degree of oligoclonality and proviral load in $CD8^+$ cells, not observed for $CD4^+$ cells, suggests that clonal expansion (rather than de novo infection) has a more significant role in proliferation of these cells. This will need to be tested further, potentially with the use of mathematical models currently being developed to model the infectious to mitotic spread ratio in HTLV-1 chronic infection. As $CD8^+$ cells (in particular, HTLV-1 specific CTLs) are potentially implicated in HTLV-1-mediated pathogenesis, more consideration should be taken to whether these are infected or not, and the mechanisms that promote their proliferation. The results of this study support a protective role for HTLV-1 specific CTLs regulating expression of proviral proteins and proliferation of infected cells.

This work carries implications both for the field of HTLV-1 research as well as other retroviruses, since mechanisms employed by HTLV-1 to regulate its expression would also be used by other retroviruses, such as latent HIV. The population level findings of this work will lead to further mechanistic testing of mechanisms of HTLV-1 persistence.

References

Afonso, P. V., Mekaouche, M., Mortreux, F., Toulza, F., Moriceau, A., Wattel, E., Gessain, A., Bangham, C. R., Dubreuil, G., Plumelle, Y., Hermine, O., Estaquier, J. & Mahieux, R. 2010. Highly active antiretroviral treatment against STLV-1 infection combining reverse transcriptase and HDAC inhibitors. *Blood,* 116**,** 3802-8.

Aird, D., Ross, M. G., Chen, W. S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D. B., Nusbaum, C. & Gnirke, A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol,* 12**,** R18.

Araujo, A., Lima, M. A. & Silva, M. T. 2008. Human T-lymphotropic virus 1 neurologic disease. *Curr Treat Options Neurol,* 10**,** 193-200.

Araujo, A. Q., Andrade-Filho, A. S., Castro-Costa, C. M., Menna-Barreto, M. & Almeida, S. M. 1998. HTLV-I-associated myelopathy/tropical spastic paraparesis in Brazil: a nationwide survey. HAM/TSP Brazilian Study Group. *J Acquir Immune Defic Syndr Hum Retrovirol,* 19**,** 536-41.

Araujo, A. Q. & Silva, M. T. 2006. The HTLV-1 neurological complex. *Lancet Neurol,* 5**,** 1068-76.

Arhel, N., Genovesio, A., Kim, K. A., Miko, S., Perret, E., Olivo-Marin, J. C., Shorte, S. & Charneau, P. 2006. Quantitative four-dimensional tracking of cytoplasmic and nuclear HIV-1 complexes. *Nat Methods,* 3**,** 817-24.

Arimura, K., Nakagawa, M., Izumo, S., Usuku, K., Itoyama, Y., Kira, J. & Osame, M. 2007. Safety and efficacy of interferon-alpha in 167 patients with human T-cell lymphotropic virus type 1-associated myelopathy. *J Neurovirol,* 13**,** 364-72.

Arnold, J., Yamamoto, B., Li, M., Phipps, A. J., Younis, I., Lairmore, M. D. & Green, P. L. 2006. Enhancement of infectivity and persistence in vivo by HBZ, a natural antisense coded protein of HTLV-1. *Blood,* 107**,** 3976-82.

Asquith, B. & Bangham, C. R. 2008. How does HTLV-I persist despite a strong cell-mediated immune response? *Trends Immunol,* 29**,** 4-11.

Asquith, B., Mosley, A. J., Barfield, A., Marshall, S. E., Heaps, A., Goon, P., Hanon, E., Tanaka, Y., Taylor, G. P. & Bangham, C. R. 2005a. A functional CD8+ cell assay reveals individual variation in CD8+ cell antiviral efficacy and explains differences in human T-lymphotropic virus type 1 proviral load. *J Gen Virol,* 86**,** 1515-23.

Asquith, B., Mosley, A. J., Heaps, A., Tanaka, Y., Taylor, G. P., Mclean, A. R. & Bangham, C. R. 2005b. Quantification of the virus-host interaction in human T lymphotropic virus I infection. *Retrovirology,* 2**,** 75.

Asquith, B., Zhang, Y., Mosley, A. J., De Lara, C. M., Wallace, D. L., Worth, A., Kaftantzi, L., Meekings, K., Griffin, G. E., Tanaka, Y., Tough, D. F., Beverley, P. C., Taylor, G. P., Macallan, D. C. & Bangham, C. R. 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**,** 8035-40.

Azimi, N., Brown, K., Bamford, R. N., Tagaya, Y., Siebenlist, U. & Waldmann, T. A. 1998. Human T cell lymphotropic virus type I Tax protein trans-activates interleukin 15 gene transcription through an NF-kappaB site. *Proc Natl Acad Sci U S A,* 95**,** 2452-7.

Azimi, N., Mariner, J., Jacobson, S. & Waldmann, T. A. 2000. How does interleukin 15 contribute to the pathogenesis of HTLV type 1-associated myelopathy/tropical spastic paraparesis? *AIDS Res Hum Retroviruses,* 16**,** 1717-22.

Azimi, N., Nagai, M., Jacobson, S. & Waldmann, T. A. 2001. IL-15 plays a major role in the persistence of Tax-specific CD8 cells in HAM/TSP patients. *Proc Natl Acad Sci U S A,* 98**,** 14559-64.

Banerjee, P., Tripp, A., Lairmore, M. D., Crawford, L., Sieburg, M., Ramos, J. C., Harrington, W., Jr., Beilke, M. A. & Feuer, G. 2010. Adult T-cell leukemia/lymphoma development in HTLV-1-infected humanized SCID mice. *Blood,* 115**,** 2640-8.

Bangham, C. R. 2009. CTL quality and the control of human retroviral infections. *Eur J Immunol,* 39**,** 1700-12.

Bangham, C. R., Meekings, K., Toulza, F., Nejmeddine, M., Majorovits, E., Asquith, B. & Taylor, G. P. 2009. The immune control of HTLV-1 infection: selection forces and dynamics. *Front Biosci,* 14**,** 2889-903.

Bangham, C. R. M. 2008. Human T-Lymphotropic Virus Type 1 (HTLV-1)-Associated Diseases. *In:* KASLOW, R. A., MCNICHOLL, J.M., HILL, A.V.S (ed.) *Genetic Susceptibility to Infectious Diseases.* New York: Oxford University Press.

Barnard, R. J., Narayan, S., Dornadula, G., Miller, M. D. & Young, J. A. 2004. Low pH is required for avian sarcoma and leukosis virus Env-dependent viral penetration into the cytosol and not for viral uncoating. *J Virol,* 78**,** 10433-41.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. & Zhao, K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell,* 129**,** 823-37.

Bartoe, J. T., Albrecht, B., Collins, N. D., Robek, M. D., Ratner, L., Green, P. L. & Lairmore, M. D. 2000. Functional role of pX open reading frame II of human T-lymphotropic virus type 1 in maintenance of viral loads in vivo. *J Virol,* 74**,** 1094-100.

Basbous, J., Arpin, C., Gaudray, G., Piechaczyk, M., Devaux, C. & Mesnard, J. M. 2003. The HBZ factor of human T-cell leukemia virus type I dimerizes with transcription factors JunB and c-Jun and modulates their transcriptional activity. *J Biol Chem,* 278**,** 43620-7.

Batliwalla, F., Monteiro, J., Serrano, D. & Gregersen, P. K. 1996. Oligoclonality of CD8+ T cells in health and disease: aging, infection, or immune regulation? *Hum Immunol,* 48**,** 68-76.

Beilke, M. A., Japa, S., Moeller-Hadi, C. & Martin-Schild, S. 2005. Tropical spastic paraparesis/human T leukemia virus type 1-associated myelopathy in HIV type 1-coinfected patients. *Clin Infect Dis,* 41**,** e57-63.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A.,

Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Chiara, E. C. M., Chang, S., Neil Cooley, R., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Scott Furey, W., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* 456**,** 53-9.

Bex, F. & Gaynor, R. B. 1998. Regulation of gene expression by HTLV-I Tax protein. *Methods,* 16**,** 83-94.

Birzele, F., Fauti, T., Stahl, H., Lenter, M. C., Simon, E., Knebel, D., Weith, A., Hildebrandt, T. & Mennerich, D. 2011. Next-generation insights into regulatory T cells: expression profiling and FoxP3 occupancy in Human. *Nucleic Acids Res,* 39**,** 7946-60.

Biswas, H. H., Kaidarova, Z., Garratty, G., Gibble, J. W., Newman, B. H., Smith, J. W., Ziman, A., Fridey, J. L., Sacher, R. A., Murphy, E. L. & Study, H. O. 2010. Increased all-cause and cancer mortality in HTLV-II infection. *J Acquir Immune Defic Syndr,* 54**,** 290-6.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A. & Taylor, J. 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol,* Chapter 19**,** Unit 19 10 1-21.

Botcheva, K., Mccorkle, S. R., Mccombie, W. R., Dunn, J. J. & Anderson, C. W. 2011. Distinct p53 genomic binding patterns in normal and cancer-derived human cells. *Cell Cycle,* 10**,** 4237-49.

Bouyac-Bertoia, M., Dvorin, J. D., Fouchier, R. A., Jenkins, Y., Meyer, B. E., Wu, L. I., Emerman, M. & Malim, M. H. 2001. HIV-1 infection requires a functional integrase NLS. *Mol Cell,* 7**,** 1025-35.

Bowerman, B., Brown, P. O., Bishop, J. M. & Varmus, H. E. 1989. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev,* 3**,** 469-78.

Boxus, M., Twizere, J. C., Legros, S., Dewulf, J. F., Kettmann, R. & Willems, L. 2008. The HTLV-1 Tax interactome. *Retrovirology,* 5**,** 76.

Brady, J., Jeang, K. T., Duvall, J. & Khoury, G. 1987. Identification of p40x-responsive regulatory sequences within the human T-cell leukemia virus type I long terminal repeat. *J Virol,* 61**,** 2175-81.

Brady, T., Roth, S. L., Malani, N., Wang, G. P., Berry, C. C., Leboulch, P., Hacein-Bey-Abina, S., Cavazzana-Calvo, M., Papapetrou, E. P., Sadelain, M., Savilahti, H. & Bushman, F. D. 2011. A method to sequence and quantify DNA integration for monitoring outcome in gene therapy. *Nucleic Acids Res,* 39**,** e72.

Brant, L. J., Cawley, C., Davison, K. L., Taylor, G. P. & Group, H. N. R. S. 2011. Recruiting individuals into the HTLV cohort study in the United Kingdom: clinical findings and challenges in the first six years, 2003 to 2009. *Euro Surveill,* 16.

Burbelo, P. D., Meoli, E., Leahy, H. P., Graham, J., Yao, K., Oh, U., Janik, J. E., Mahieux, R., Kashanchi, F., Iadarola, M. J. & Jacobson, S. 2008. Anti-HTLV antibody profiling reveals an antibody signature for HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP). *Retrovirology,* 5**,** 96.

Calattini, S., Chevalier, S. A., Duprez, R., Bassot, S., Froment, A., Mahieux, R. & Gessain, A. 2005. Discovery of a new human T-cell lymphotropic virus (HTLV-3) in Central Africa. *Retrovirology,* 2**,** 30.

Cartier, L. & Ramirez, E. 2005. Presence of HTLV-I Tax protein in cerebrospinal fluid from HAM/TSP patients. *Arch Virol,* 150**,** 743-53.

Cavrois, M., Leclercq, I., Gout, O., Gessain, A., Wain-Hobson, S. & Wattel, E. 1998. Persistent oligoclonal expansion of human T-cell leukemia virus type 1-infected circulating cells in patients with Tropical spastic paraparesis/HTLV-1 associated myelopathy. *Oncogene,* 17**,** 77-82.

Cavrois, M., Wain-Hobson, S., Gessain, A., Plumelle, Y. & Wattel, E. 1996. Adult T-cell leukemia/lymphoma on a background of clonally expanding human T-cell leukemia virus type-1-positive cells. *Blood,* 88**,** 4646-50.

Cavrois, M., Wain-Hobson, S. & Wattel, E. 1995. Stochastic events in the amplification of HTLV-I integration sites by linker-mediated PCR. *Res Virol,* 146**,** 179-84.

Chan, W. C., Link, S., Mawle, A., Check, I., Brynes, R. K. & Winton, E. F. 1986. Heterogeneity of large granular lymphocyte proliferations: delineation of two major subtypes. *Blood,* 68**,** 1142-53.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E. & Debyser, Z. 2003. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem,* 278**,** 372-81.

Chevalier, S. A., Durand, S., Dasgupta, A., Radonovich, M., Cimarelli, A., Brady, J. N., Mahieux, R. & Pise-Masison, C. A. 2012. The transcription profile of Tax-3 is more similar to Tax-1 than Tax-2: insights into HTLV-3 potential leukemogenic properties. *PLoS One,* 7**,** e41003.

Cho, I., Sugimoto, M., Mita, S., Tokunaga, M., Imamura, F. & Ando, M. 1995. In vivo proviral burden and viral RNA expression in T cell subsets of patients with human T lymphotropic virus type-1-associated myelopathy/tropical spastic paraparesis. *Am J Trop Med Hyg,* 53**,** 412-8.

Chou, K. S., Okayama, A., Su, I. J., Lee, T. H. & Essex, M. 1996. Preferred nucleotide sequence at the integration target site of human T-cell leukemia virus type I from patients with adult T-cell leukemia. *International Journal of Cancer,* 65**,** 20-24.

Ciminale, V., Hatziyanni, M., Felber, B. K., Bear, J., Hatzakis, A. & Pavlakis, G. N. 2000. Unusual CD4+CD8+ phenotype in a greek patient diagnosed with adult T-cell leukemia positive for human T-cell leukemia virus type I (HTLV-I). *Leuk Res,* 24**,** 353-8.

Ciminale, V., Zotti, L., D'agostino, D. M., Ferro, T., Casareto, L., Franchini, G., Bernardi, P. & Chieco-Bianchi, L. 1999. Mitochondrial targeting of the p13II protein coded by the x-II ORF of human T-cell leukemia/lymphotropic virus type I (HTLV-I). *Oncogene,* 18**,** 4505-14.

Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J. R. & Bushman, F. 2005. A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med,* 11**,** 1287-9.

Colisson, R., Barblu, L., Gras, C., Raynaud, F., Hadj-Slimane, R., Pique, C., Hermine, O., Lepelletier, Y. & Herbeuval, J. P. 2010. Free HTLV-1 induces TLR7-dependent

innate immune response and TRAIL relocalization in killer plasmacytoid dendritic cells. *Blood,* 115**,** 2177-85.

Cook, L. B., Rowan, A. G., Melamed, A., Taylor, G. P. & Bangham, C. R. 2012. HTLV-1-infected T cells contain a single integrated provirus in natural infection. *Blood,* 120**,** 3488-90.

Daenke, S. & Bangham, C. R. 1994. Do T cells cause HTLV-1-associated disease?: a taxing problem. *Clin Exp Immunol,* 96**,** 179-81.

Daenke, S., Kermode, A. G., Hall, S. E., Taylor, G., Weber, J., Nightingale, S. & Bangham, C. R. 1996. High activated and memory cytotoxic T-cell responses to HTLV-1 in healthy carriers and patients with tropical spastic paraparesis. *Virology,* 217**,** 139-46.

Daenke, S., Nightingale, S., Cruickshank, J. K. & Bangham, C. R. 1990. Sequence variants of human T-cell lymphotropic virus type I from patients with tropical spastic paraparesis and adult T-cell leukemia do not distinguish neurological from leukemic isolates. *J Virol,* 64**,** 1278-82.

Dalgleish, A. G., Beverley, P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F. & Weiss, R. A. 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature,* 312**,** 763-7.

De Lourdes Bastos, M., Osterbauer, B., Mesquita, D. L., Carrera, C. A., Albuquerque, M. J., Silva, L., Pereira, D. N., Riley, L. & Carvalho, E. M. 2009. Prevalence of human T-cell lymphotropic virus type 1 infection in hospitalized patients with tuberculosis. *Int J Tuberc Lung Dis,* 13**,** 1519-23.

De Rijck, J., Bartholomeeusen, K., Ceulemans, H., Debyser, Z. & Gijsbers, R. 2010. High-resolution profiling of the LEDGF/p75 chromatin interaction in the ENCODE region. *Nucleic Acids Res,* 38**,** 6135-47.

De Souza, J. G., Da Fonseca, F. G., Martins, M. L., Martins, C. P., De Carvalho, L. D., Coelho-Dos-Reis, J. G., Carneiro-Proietti, A. B., Martins-Filho, O. A., Barbosa-Stancioli, E. F. & Giph 2011. Anti-Tax antibody levels in asymptomatic carriers, oligosymptomatic carriers, patients with rheumatologic disease or with HAM/TSP do not correlate with HTLV-1 proviral load. *J Clin Virol,* 50**,** 13-8.

Delarco, J. & Todaro, G. J. 1976. Membrane receptors for murine leukemia viruses: characterization using the purified viral envelope glycoprotein, gp71. *Cell,* 8**,** 365-71.

Demontis, M. A., Hilburn, S. & Taylor, G. P. 2013. Human T cell lymphotropic virus type 1 viral load variability and long-term trends in asymptomatic carriers and in patients with human T cell lymphotropic virus type 1-related diseases. *AIDS Res Hum Retroviruses,* 29**,** 359-64.

Derse, D., Crise, B., Li, Y., Princler, G., Lum, N., Stewart, C., Mcgrath, C. F., Hughes, S. H., Munroe, D. J. & Wu, X. 2007. Human T-cell leukemia virus type 1 integration target sites in the human genome: comparison with those of other retroviruses. *J Virol,* 81**,** 6731-41.

Devon, R. S., Porteous, D. J. & Brookes, A. J. 1995. Splinkerettes--improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Res,* 23**,** 1644-5.

Dhodapkar, M. V., Li, C. Y., Lust, J. A., Tefferi, A. & Phyliky, R. L. 1994. Clinical spectrum of clonal proliferations of T-large granular lymphocytes: a T-cell clonopathy of undetermined significance? *Blood,* 84**,** 1620-7.

Ding, W., Albrecht, B., Luo, R., Zhang, W., Stanley, J. R., Newbound, G. C. & Lairmore, M. D. 2001. Endoplasmic reticulum and cis-Golgi localization of human T-lymphotropic virus type 1 p12(I): association with calreticulin and calnexin. *J Virol,* 75**,** 7672-82.

Dodon, M. D., Villaudy, J., Gazzolo, L., Haines, R. & Lairmore, M. 2012. What we are learning on HTLV-1 pathogenesis from animal models. *Front Microbiol,* 3**,** 320.

Doi, K., Wu, X., Taniguchi, Y., Yasunaga, J., Satou, Y., Okayama, A., Nosaka, K. & Matsuoka, M. 2005. Preferential selection of human T-cell leukemia virus type I provirus integration sites in leukemic versus carrier states. *Blood,* 106**,** 1048-53.

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B. K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., Shoresh, N., Simon, J. M., Song, L., Trinklein, N. D., Altshuler, R. C., Birney, E., Brown, J. B., Cheng, C., Djebali, S., Dong, X., Dunham, I., Ernst, J., Furey, T. S., Gerstein, M., Giardine, B., Greven, M., Hardison, R. C., Harris, R. S., Herrero, J., Hoffman, M. M., Iyer, S., Kelllis, M., Khatun, J., Kheradpour, P., Kundaje, A., Lassman, T., Li, Q., Lin, X., Marinov, G. K., Merkel, A., Mortazavi, A., Parker, S. C., Reddy, T. E., Rozowsky, J., Schlesinger, F., Thurman, R. E., Wang, J., Ward, L. D., Whitfield, T. W., Wilder, S. P., Wu, W., Xi, H. S., Yip, K. Y., Zhuang, J., Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., Snyder, M., Pazin, M. J., Lowdon, R. F., Dillon, L. A., Adams, L. B., Kelly, C. J., Zhang, J., Wexler, J. R., Green, E. D., Good, P. J., Feingold, E. A., Bernstein, B. E., Birney, E., Crawford, G. E., Dekker, J., Elinitski, L., Farnham, P. J., Gerstein, M., Giddings, M. C., Gingeras, T. R., Green, E. D., Guigo, R., Hardison, R. C., Hubbard, T. J., Kellis, M., Kent, W. J., Lieb, J. D., Margulies, E. H., Myers, R. M., Snyder, M., Starnatoyannopoulos, J. A., Tennebaum, S. A., et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

Durand, C. M., Blankson, J. N. & Siliciano, R. F. 2012. Developing strategies for HIV-1 eradication. *Trends Immunol,* 33**,** 554-62.

Easley, R., Carpio, L., Guendel, I., Klase, Z., Choi, S., Kehn-Hall, K., Brady, J. N. & Kashanchi, F. 2010. Human T-lymphotropic virus type 1 transcription and chromatin-remodeling complexes. *J Virol,* 84**,** 4755-68.

Einfeld, D. 1996. Maturation and assembly of retroviral glycoproteins. *Curr Top Microbiol Immunol,* 214**,** 133-76.

Eiraku, N., Hingorani, R., Ijichi, S., Machigashira, K., Gregersen, P. K., Monteiro, J., Usuku, K., Yashiki, S., Sonoda, S., Osame, M. & Hall, W. W. 1998. Clonal expansion within CD4+ and CD8+ T cell subsets in human T lymphotropic virus type I-infected individuals. *J Immunol,* 161**,** 6674-80.

Elovaara, I., Koenig, S., Brewah, A. Y., Woods, R. M., Lehky, T. & Jacobson, S. 1993. High human T cell lymphotropic virus type 1 (HTLV-1)-specific precursor cytotoxic T lymphocyte frequencies in patients with HTLV-1-associated neurological disease. *J Exp Med,* 177**,** 1567-73.

Elovaara, I., Utz, U., Smith, S. & Jacobson, S. 1995. Limited T cell receptor usage by HTLV-I tax-specific, HLA class I restricted cytotoxic T lymphocytes from patients with HTLV-I associated neurological disease. *J Neuroimmunol,* 63**,** 47-53.

Enose-Akahata, Y., Abrams, A., Massoud, R., Bialuk, I., Johnson, K. R., Green, P. L., Maloney, E. M. & Jacobson, S. 2013. Humoral immune response to HTLV-1 basic leucine zipper factor (HBZ) in HTLV-1-infected individuals. *Retrovirology,* 10**,** 19.

Etoh, K., Tamiya, S., Yamaguchi, K., Okayama, A., Tsubouchi, H., Ideta, T., Mueller, N., Takatsuki, K. & Matsuoka, M. 1997. Persistent clonal proliferation of human T-lymphotropic virus type I-infected cells in vivo. *Cancer Res,* 57**,** 4862-7.

Euskirchen, G., Auerbach, R. K. & Snyder, M. 2012. SWI/SNF chromatin-remodeling factors: multiscale analyses and diverse functions. *J Biol Chem,* 287**,** 30897-905.

Euskirchen, G. M., Auerbach, R. K., Davidov, E., Gianoulis, T. A., Zhong, G., Rozowsky, J., Bhardwaj, N., Gerstein, M. B. & Snyder, M. 2011. Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. *PLoS Genet,* 7**,** e1002008.

Faller, D. V., Crimmins, M. A. & Mentzer, S. J. 1988. Human T-cell leukemia virus type I infection of CD4+ or CD8+ cytotoxic T-cell clones results in immortalization with retention of antigen specificity. *J Virol,* 62**,** 2942-50.

Fan, J., Ma, G., Nosaka, K., Tanabe, J., Satou, Y., Koito, A., Wain-Hobson, S., Vartanian, J. P. & Matsuoka, M. 2010. APOBEC3G generates nonsense mutations in human T-cell leukemia virus type 1 proviral genomes in vivo. *J Virol,* 84**,** 7278-87.

Fan, N., Gavalchin, J., Paul, B., Wells, K. H., Lane, M. J. & Poiesz, B. J. 1992. Infection of peripheral blood mononuclear cells and cell lines by cell-free human T-cell lymphoma/leukemia virus type I. *J Clin Microbiol,* 30**,** 905-10.

Fehse, B. & Roeder, I. 2008. Insertional mutagenesis and clonal dominance: biological and statistical considerations. *Gene Ther,* 15**,** 143-53.

Finzi, D., Hermankova, M., Pierson, T., Carruth, L. M., Buck, C., Chaisson, R. E., Quinn, T. C., Chadwick, K., Margolick, J., Brookmeyer, R., Gallant, J., Markowitz, M., Ho, D. D., Richman, D. D. & Siliciano, R. F. 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science,* 278**,** 1295-300.

Frietze, S., Lan, X., Jin, V. X. & Farnham, P. J. 2010. Genomic targets of the KRAB and SCAN domain-containing zinc finger protein 263. *J Biol Chem,* 285**,** 1393-403.

Fujita, P. A., Rhead, B., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Cline, M. S., Goldman, M., Barber, G. P., Clawson, H., Coelho, A., Diekhans, M., Dreszer, T. R., Giardine, B. M., Harte, R. A., Hillman-Jackson, J., Hsu, F., Kirkup, V., Kuhn, R. M., Learned, K., Li, C. H., Meyer, L. R., Pohl, A., Raney, B. J., Rosenbloom, K. R., Smith, K. E., Haussler, D. & Kent, W. J. 2011. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res,* 39**,** D876-82.

Fujiwara, T., O'geen, H., Keles, S., Blahnik, K., Linnemann, A. K., Kang, Y. A., Choi, K., Farnham, P. J. & Bresnick, E. H. 2009. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell,* 36**,** 667-81.

Furukawa, Y., Fujisawa, J., Osame, M., Toita, M., Sonoda, S., Kubota, R., Ijichi, S. & Yoshida, M. 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**,** 1012-6.

Furukawa, Y., Kubota, R., Tara, M., Izumo, S. & Osame, M. 2001. Existence of escape mutant in HTLV-I tax during the development of adult T-cell leukemia. *Blood,* 97**,** 987-93.

Gabet, A. S., Kazanji, M., Couppie, P., Clity, E., Pouliquen, J. F., Sainte-Marie, D., Aznar, C. & Wattel, E. 2003. Adult T-cell leukaemia/lymphoma-like human T-cell leukaemia virus-1 replication in infective dermatitis. *Br J Haematol,* 123**,** 406-12.

Gabet, A. S., Mortreux, F., Talarmin, A., Plumelle, Y., Leclercq, I., Leroy, A., Gessain, A., Clity, E., Joubert, M. & Wattel, E. 2000. High circulating proviral load with oligoclonal expansion of HTLV-1 bearing T cells in HTLV-1 carriers with strongyloidiasis. *Oncogene,* 19**,** 4954-60.

Gabriel, R., Eckenberg, R., Paruzynski, A., Bartholomae, C. C., Nowrouzi, A., Arens, A., Howe, S. J., Recchia, A., Cattoglio, C., Wang, W., Faber, K., Schwarzwaelder, K., Kirsten, R., Deichmann, A., Ball, C. R., Balaggan, K. S., Yanez-Munoz, R. J., Ali, R. R., Gaspar, H. B., Biasco, L., Aiuti, A., Cesana, D., Montini, E., Naldini, L., Cohen-Haguenauer, O., Mavilio, F., Thrasher, A. J., Glimm, H., Von Kalle, C., Saurin, W. & Schmidt, M. 2009. Comprehensive genomic access to vector integration in clinical gene therapy. *Nat Med,* 15**,** 1431-6.

Gallo, R. C., Kalyanaraman, V. S., Sarngadharan, M. G., Sliski, A., Vonderheid, E. C., Maeda, M., Nakao, Y., Yamada, K., Ito, Y., Gutensohn, N., Murphy, S., Bunn, P. A., Jr., Catovsky, D., Greaves, M. F., Blayney, D. W., Blattner, W., Jarrett, W. F., Zur Hausen, H., Seligmann, M., Brouet, J. C., Haynes, B. F., Jegasothy, B. V., Jaffe, E., Cossman, J., Broder, S., Fisher, R. I., Golde, D. W. & Robert-Guroff, M. 1983. Association of the human type C retrovirus with a subset of adult T-cell cancers. *Cancer Res,* 43**,** 3892-9.

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. *Journal of Virology,* 76**,** 12813-12822.

Gavin, I., Horn, P. J. & Peterson, C. L. 2001. SWI/SNF chromatin remodeling requires changes in DNA topology. *Mol Cell,* 7**,** 97-104.

Gessain, A., Barin, F., Vernant, J. C., Gout, O., Maurs, L., Calender, A. & De The, G. 1985. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet,* 2**,** 407-10.

Gessain, A. & Cassar, O. 2012. Epidemiological Aspects and World Distribution of HTLV-1 Infection. *Front Microbiol,* 3**,** 388.

Ghez, D., Lepelletier, Y., Jones, K. S., Pique, C. & Hermine, O. 2010. Current concepts regarding the HTLV-1 receptor complex. *Retrovirology,* 7**,** 99.

Ghez, D., Lepelletier, Y., Lambert, S., Fourneau, J. M., Blot, V., Janvier, S., Arnulf, B., Van Endert, P. M., Heveker, N., Pique, C. & Hermine, O. 2006. Neuropilin-1 is involved in human T-cell lymphotropic virus type 1 entry. *J Virol,* 80**,** 6844-54.

Giebler, H. A., Loring, J. E., Van Orden, K., Colgin, M. A., Garrus, J. E., Escudero, K. W., Brauweiler, A. & Nyborg, J. K. 1997. Anchoring of CREB binding protein to the human T-cell leukemia virus type 1 promoter: a molecular mechanism of Tax transactivation. *Mol Cell Biol,* 17**,** 5156-64.

Gilboa, E., Mitra, S. W., Goff, S. & Baltimore, D. 1979. A detailed model of reverse transcription and tests of crucial aspects. *Cell,* 18**,** 93-100.

Gillet, N. A., Malani, N., Melamed, A., Gormley, N., Carter, R., Bentley, D., Berry, C., Bushman, F. D., Taylor, G. P. & Bangham, C. R. 2011. The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. *Blood,* 117**,** 3113-22.

Gini, C. 1912. *Variabilita e Mutabilitd*

Goff, S. P. 2007. Host factors exploited by retroviruses. *Nat Rev Microbiol,* 5**,** 253-63.

Goon, P. K., Biancardi, A., Fast, N., Igakura, T., Hanon, E., Mosley, A. J., Asquith, B., Gould, K. G., Marshall, S., Taylor, G. P. & Bangham, C. R. 2004a. Human T cell lymphotropic virus (HTLV) type-1-specific CD8+ T cells: frequency and immunodominance hierarchy. *J Infect Dis,* 189**,** 2294-8.

Goon, P. K., Igakura, T., Hanon, E., Mosley, A. J., Barfield, A., Barnard, A. L., Kaftantzi, L., Tanaka, Y., Taylor, G. P., Weber, J. N. & Bangham, C. R. 2004b. Human T cell lymphotropic virus type I (HTLV-I)-specific CD4+ T cells: immunodominance hierarchy and preferential infection with HTLV-I. *J Immunol,* 172**,** 1735-43.

Gould, K. G. & Bangham, C. R. 1998. Virus variation, escape from cytotoxic T lymphocytes and human retroviral persistence. *Semin Cell Dev Biol,* 9**,** 321-8.

Grakoui, A., Bromley, S. K., Sumen, C., Davis, M. M., Shaw, A. S., Allen, P. M. & Dustin, M. L. 1999. The immunological synapse: a molecular machine controlling T cell activation. *Science,* 285**,** 221-7.

Grandgenett, D. P. 2005. Symmetrical recognition of cellular DNA target sequences during retroviral integration. *Proc Natl Acad Sci U S A,* 102**,** 5903-4.

Grassmann, R., Dengler, C., Muller-Fleckenstein, I., Fleckenstein, B., Mcguire, K., Dokhelar, M. C., Sodroski, J. G. & Haseltine, W. A. 1989. Transformation to continuous growth of primary human T lymphocytes by human T-cell leukemia virus type I X-region genes transduced by a Herpesvirus saimiri vector. *Proc Natl Acad Sci U S A,* 86**,** 3351-5.

Grone, M., Hoffmann, E., Berchtold, S., Cullen, B. R. & Grassmann, R. 1994. A single stem-loop structure within the HTLV-1 Rex response element is sufficient to mediate Rex activity in vivo. *Virology,* 204**,** 144-52.

Hacein-Bey-Abina, S., Garrigue, A., Wang, G. P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., Asnafi, V., Macintyre, E., Dal Cortivo, L., Radford, I., Brousse, N., Sigaux, F., Moshous, D., Hauer, J., Borkhardt, A., Belohradsky, B. H., Wintergerst, U., Velez, M. C., Leiva, L., Sorensen, R., Wulffraat, N., Blanche, S., Bushman, F. D., Fischer, A. & Cavazzana-Calvo, M. 2008. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest,* 118**,** 3132-42.

Hafner, N., Driesch, C., Gajda, M., Jansen, L., Kirchmayr, R., Runnebaum, I. B. & Durst, M. 2008. Integration of the HPV16 genome does not invariably result in high levels of viral oncogene transcripts. *Oncogene,* 27**,** 1610-7.

Hagiya, K., Yasunaga, J., Satou, Y., Ohshima, K. & Matsuoka, M. 2011. ATF3, an HTLV-1 bZip factor binding protein, promotes proliferation of adult T-cell leukemia cells. *Retrovirology,* 8**,** 19.

Hainaut, P., Soussi, T., Shomer, B., Hollstein, M., Greenblatt, M., Hovig, E., Harris, C. C. & Montesano, R. 1997. Database of p53 gene somatic mutations in human tumors and cell lines: updated compilation and future prospects. *Nucleic Acids Res,* 25**,** 151-7.

Han, Y., Lin, Y. B., An, W., Xu, J., Yang, H. C., O'connell, K., Dordai, D., Boeke, J. D., Siliciano, J. D. & Siliciano, R. F. 2008. Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. *Cell Host Microbe,* 4**,** 134-46.

Hanai, S., Nitta, T., Shoda, M., Tanaka, M., Iso, N., Mizoguchi, I., Yashiki, S., Sonoda, S., Hasegawa, Y., Nagasawa, T. & Miwa, M. 2004. Integration of human T-cell leukemia virus type 1 in genes of leukemia cells of patients with adult T-cell leukemia. *Cancer Sci,* 95**,** 306-10.

Hanon, E., Hall, S., Taylor, G. P., Saito, M., Davis, R., Tanaka, Y., Usuku, K., Osame, M., Weber, J. N. & Bangham, C. R. 2000a. Abundant tax protein expression in CD4+ T cells infected with human T-cell lymphotropic virus type I (HTLV-I) is prevented by cytotoxic T lymphocytes. *Blood,* 95**,** 1386-92.

Hanon, E., Stinchcombe, J. C., Saito, M., Asquith, B. E., Taylor, G. P., Tanaka, Y., Weber, J. N., Griffiths, G. M. & Bangham, C. R. 2000b. Fratricide among CD8(+) T lymphocytes naturally infected with human T cell lymphotropic virus type I. *Immunity,* 13**,** 657-64.

Hare, S., Gupta, S. S., Valkov, E., Engelman, A. & Cherepanov, P. 2010. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature,* 464**,** 232-6.

Hasegawa, H., Sawa, H., Lewis, M. J., Orba, Y., Sheehy, N., Yamamoto, Y., Ichinohe, T., Tsunetsugu-Yokota, Y., Katano, H., Takahashi, H., Matsuda, J., Sata, T., Kurata, T., Nagashima, K. & Hall, W. W. 2006. Thymus-derived leukemia-lymphoma in mice transgenic for the Tax gene of human T-lymphotropic virus type I. *Nat Med,* 12**,** 466-72.

Hidaka, M., Inoue, J., Yoshida, M. & Seiki, M. 1988. Post-transcriptional regulator (rex) of HTLV-1 initiates expression of viral structural proteins but suppresses expression of regulatory proteins. *EMBO J,* 7**,** 519-23.

Hilburn, S., Rowan, A., Demontis, M. A., Macnamara, A., Asquith, B., Bangham, C. R. & Taylor, G. P. 2011. In vivo expression of human T-lymphotropic virus type 1 basic leucine-zipper protein generates specific CD8+ and CD4+ T-lymphocyte responses that correlate with clinical outcome. *J Infect Dis,* 203**,** 529-36.

Hill, M., Tachedjian, G. & Mak, J. 2005. The packaging and maturation of the HIV-1 Pol proteins. *Curr HIV Res,* 3**,** 73-85.

Hino, S., Katamine, S., Miyata, H., Tsuji, Y., Yamabe, T. & Miyamoto, T. 1997. Primary prevention of HTLV-1 in Japan. *Leukemia,* 11 Suppl 3**,** 57-9.

Hiraragi, H., Michael, B., Nair, A., Silic-Benussi, M., Ciminale, V. & Lairmore, M. 2005. Human T-lymphotropic virus type 1 mitochondrion-localizing protein p13II sensitizes Jurkat T cells to Ras-mediated apoptosis. *J Virol,* 79**,** 9449-57.

Hishizawa, M., Imada, K., Kitawaki, T., Ueda, M., Kadowaki, N. & Uchiyama, T. 2004. Depletion and impaired interferon-alpha-producing capacity of blood plasmacytoid dendritic cells in human T-cell leukaemia virus type I-infected individuals. *Br J Haematol,* 125**,** 568-75.

Hodson, A., Crichton, S., Montoto, S., Mir, N., Matutes, E., Cwynarski, K., Kumaran, T., Ardeshna, K. M., Pagliuca, A., Taylor, G. P. & Fields, P. A. 2011. Use of zidovudine and interferon alfa with chemotherapy improves survival in both acute and lymphoma subtypes of adult T-cell leukemia/lymphoma. *J Clin Oncol,* 29**,** 4696-701.

Hodson, A., Laydon, D. J., Bain, B. J., Fields, P. A. & Taylor, G. P. 2013. Pre-morbid human T-lymphotropic virus type I proviral load, rather than percentage of abnormal lymphocytes, is associated with an increased risk of aggressive adult T-cell leukemia/lymphoma. *Haematologica,* 98**,** 385-8.

Hoger, T. A., Jacobson, S., Kawanishi, T., Kato, T., Nishioka, K. & Yamamoto, K. 1997. Accumulation of human T lymphotropic virus (HTLV)-I-specific T cell clones in HTLV-I-associated myelopathy/tropical spastic paraparesis patients. *J Immunol,* 159**,** 2042-8.

Hollsberg, P. 1997. Pathogenesis of chronic progressive myelopathy associated with human T-cell lymphotropic virus type I. *Acta Neurol Scand Suppl,* 169**,** 86-93.

Holt, R. A. & Jones, S. J. 2008. The new paradigm of flow cell sequencing. *Genome Res,* 18**,** 839-46.

Hou, C., Zhao, H., Tanimoto, K. & Dean, A. 2008. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci U S A,* 105**,** 20398-403.

Hoxie, J. A., Matthews, D. M. & Cines, D. B. 1984. Infection of human endothelial cells by human T-cell leukemia virus type I. *Proc Natl Acad Sci U S A,* 81**,** 7591-5.

Igakura, T., Stinchcombe, J. C., Goon, P. K., Taylor, G. P., Weber, J. N., Griffiths, G. M., Tanaka, Y., Osame, M. & Bangham, C. R. 2003. Spread of HTLV-I between lymphocytes by virus-induced polarization of the cytoskeleton. *Science,* 299**,** 1713-6.

Ijichi, S., Izumo, S., Eiraku, N., Machigashira, K., Kubota, R., Nagai, M., Ikegami, N., Kashio, N., Umehara, F., Maruyama, I. & Et Al. 1993. An autoaggressive process against bystander tissues in HTLV-I-infected individuals: a possible pathomechanism of HAM/TSP. *Med Hypotheses,* 41**,** 542-7.

Ijichi, S., Ramundo, M. B., Takahashi, H. & Hall, W. W. 1992. In vivo cellular tropism of human T cell leukemia virus type II (HTLV-II). *J Exp Med,* 176**,** 293-6.

Ikeda, T., Shibata, J., Yoshimura, K., Koito, A. & Matsushita, S. 2007. Recurrent HIV-1 integration at the BACH2 locus in resting CD4+ T cell populations during effective highly active antiretroviral therapy. *J Infect Dis,* 195**,** 716-25.

Ina, Y. & Gojobori, T. 1990. Molecular evolution of human T-cell leukemia virus. *J Mol Evol,* 31**,** 493-9.

Iwanaga, M., Watanabe, T., Utsunomiya, A., Okayama, A., Uchimaru, K., Koh, K. R., Ogata, M., Kikuchi, H., Sagara, Y., Uozumi, K., Mochizuki, M., Tsukasaki, K., Saburi, Y., Yamamura, M., Tanaka, J., Moriuchi, Y., Hino, S., Kamihira, S., Yamaguchi, K. & Joint Study on Predisposing Factors Of, A. T. L. D. I. 2010. Human T-cell leukemia virus type I (HTLV-1) proviral load and disease progression in asymptomatic HTLV-1 carriers: a nationwide prospective study in Japan. *Blood,* 116**,** 1211-9.

Iwatsuki, K., Inoue, F., Takigawa, M., Iemoto, G., Nagatani, T., Nakajima, H. & Yamada, M. 1990. Exchange of dominant lymphoid cell clones in a patient with adult T-cell leukemia/lymphoma. *Acta Derm Venereol,* 70**,** 49-52.

Jacobson, S. 2002. Immunopathogenesis of human T cell lymphotropic virus type I-associated neurologic disease. *J Infect Dis,* 186 Suppl 2**,** S187-92.

Jacobson, S., Mcfarlin, D. E., Robinson, S., Voskuhl, R., Martin, R., Brewah, A., Newell, A. J. & Koenig, S. 1992. HTLV-I-specific cytotoxic T lymphocytes in the cerebrospinal fluid of patients with HTLV-I-associated neurological disease. *Ann Neurol,* 32**,** 651-7.

Jacobson, S., Shida, H., Mcfarlin, D. E., Fauci, A. S. & Koenig, S. 1990. Circulating CD8+ cytotoxic T lymphocytes specific for HTLV-I pX in patients with HTLV-I associated neurological disease. *Nature,* 348**,** 245-8.

Jason, J. M., Mcdougal, J. S., Cabradilla, C., Kalyanaraman, V. S. & Evatt, B. L. 1985. Human T-cell leukemia virus (HTLV-I) p24 antibody in New York City blood product recipients. *Am J Hematol,* 20**,** 129-37.

Jeffery, K. J., Siddiqui, A. A., Bunce, M., Lloyd, A. L., Vine, A. M., Witkover, A. D., Izumo, S., Usuku, K., Welsh, K. I., Osame, M. & Bangham, C. R. 2000. The influence of HLA class I alleles and heterozygosity on the outcome of human T cell lymphotropic virus type I infection. *J Immunol,* 165**,** 7278-84.

Jeffery, K. J., Usuku, K., Hall, S. E., Matsumoto, W., Taylor, G. P., Procter, J., Bunce, M., Ogg, G. S., Welsh, K. I., Weber, J. N., Lloyd, A. L., Nowak, M. A., Nagai, M., Kodama, D., Izumo, S., Osame, M. & Bangham, C. R. 1999. HLA alleles determine human T-lymphotropic virus-I (HTLV-I) proviral load and the risk of HTLV-I-associated myelopathy. *Proc Natl Acad Sci U S A,* 96**,** 3848-53.

Jin, Q., Agrawal, L., Vanhorn-Ali, Z. & Alkhatib, G. 2006. Infection of CD4+ T lymphocytes by the human T cell leukemia virus type 1 is mediated by the glucose transporter GLUT-1: evidence using antibodies specific to the receptor's large extracellular domain. *Virology,* 349**,** 184-96.

Jin, Q., Alkhatib, B., Cornetta, K. & Alkhatib, G. 2010. Alternate receptor usage of neuropilin-1 and glucose transporter protein 1 by the human T cell leukemia virus type 1. *Virology,* 396**,** 203-12.

Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science,* 316**,** 1497-502.

Johnson, J. M., Nicot, C., Fullen, J., Ciminale, V., Casareto, L., Mulloy, J. C., Jacobson, S. & Franchini, G. 2001. Free major histocompatibility complex class I heavy chain is preferentially targeted for degradation by human T-cell leukemia/lymphotropic virus type 1 p12(I) protein. *J Virol,* 75**,** 6086-94.

Jolly, C., Kashefi, K., Hollinshead, M. & Sattentau, Q. J. 2004. HIV-1 cell to cell transfer across an Env-induced, actin-dependent synapse. *J Exp Med,* 199**,** 283-93.

Jones, K. S., Fugo, K., Petrow-Sadowski, C., Huang, Y., Bertolette, D. C., Lisinski, I., Cushman, S. W., Jacobson, S. & Ruscetti, F. W. 2006. Human T-cell leukemia virus type 1 (HTLV-1) and HTLV-2 use different receptor complexes to enter T cells. *J Virol,* 80**,** 8291-302.

Jones, K. S., Lambert, S., Bouttier, M., Benit, L., Ruscetti, F. W., Hermine, O. & Pique, C. 2011. Molecular aspects of HTLV-1 entry: functional domains of the HTLV-1 surface subunit (SU) and their relationships to the entry receptors. *Viruses,* 3**,** 794-810.

Jones, K. S., Petrow-Sadowski, C., Bertolette, D. C., Huang, Y. & Ruscetti, F. W. 2005. Heparan sulfate proteoglycans mediate attachment and entry of human T-cell leukemia virus type 1 virions into CD4+ T cells. *J Virol,* 79**,** 12692-702.

Jones, K. S., Petrow-Sadowski, C., Huang, Y. K., Bertolette, D. C. & Ruscetti, F. W. 2008. Cell-free HTLV-1 infects dendritic cells leading to transmission and transformation of CD4(+) T cells. *Nat Med,* 14**,** 429-36.

Jordan, A., Defechereux, P. & Verdin, E. 2001. The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO J,* 20**,** 1726-38.

Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. 2008. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res,* 36**,** 5221-31.

Kalpana, G. V., Marmon, S., Wang, W., Crabtree, G. R. & Goff, S. P. 1994. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science,* 266**,** 2002-6.

Kalyanaraman, V. S., Sarngadharan, M. G., Robert-Guroff, M., Miyoshi, I., Golde, D. & Gallo, R. C. 1982. A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. *Science,* 218**,** 571-3.

Kamihira, S., Sohda, H., Atogami, S., Toriya, K., Yamada, Y., Tsukazaki, K., Momita, S., Ikeda, S., Kusano, M., Amagasaki, T. & Et Al. 1992. Phenotypic diversity and prognosis of adult T-cell leukemia. *Leuk Res,* 16**,** 435-41.

Kannagi, M., Harada, S., Maruyama, I., Inoko, H., Igarashi, H., Kuwashima, G., Sato, S., Morita, M., Kidokoro, M., Sugimoto, M. & 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**,** 761-7.

Kannian, P., Yin, H., Doueiri, R., Lairmore, M. D., Fernandez, S. & Green, P. L. 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**,** 3757-66.

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**,** 1096-101.

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D. & Kent, W. J. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res,* 32**,** D493-6.

Kasai, T. & Jeang, K. T. 2004. Two discrete events, human T-cell leukemia virus type I Tax oncoprotein expression and a separate stress stimulus, are required for induction of apoptosis in T-cells. *Retrovirology,* 1**,** 7.

Kashanchi, F. & Brady, J. N. 2005. Transcriptional and post-transcriptional gene regulation of HTLV-1. *Oncogene,* 24**,** 5938-51.

Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S. M., Habegger, L., Rozowsky, J., Shi, M., Urban, A. E., Hong, M. Y., Karczewski, K. J.,

Huber, W., Weissman, S. M., Gerstein, M. B., Korbel, J. O. & Snyder, M. 2010. Variation in transcription factor binding among humans. *Science,* 328**,** 232-5.

Kattan, T., Macnamara, A., Rowan, A. G., Nose, H., Mosley, A. J., Tanaka, Y., Taylor, G. P., Asquith, B. & Bangham, C. R. 2009. The avidity and lytic efficiency of the CTL response to HTLV-1. *J Immunol,* 182**,** 5723-9.

Kawabata, T., Higashimoto, I., Takashima, H., Izumo, S. & Kubota, R. 2012. Human T-lymphotropic virus type I (HTLV-I)-specific CD8+ cells accumulate in the lungs of patients infected with HTLV-I with pulmonary involvement. *J Med Virol,* 84**,** 1120-7.

Kawai, T. & Akira, S. 2011. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity,* 34**,** 637-50.

Kchour, G., Makhoul, N. J., Mahmoudi, M., Kooshyar, M. M., Shirdel, A., Rastin, M., Rafatpanah, H., Tarhini, M., Zalloua, P. A., Hermine, O., Farid, R. & Bazarbachi, A. 2007. Zidovudine and interferon-alpha treatment induces a high response rate and reduces HTLV-1 proviral load and VEGF plasma levels in patients with adult T-cell leukemia from North East Iran. *Leuk Lymphoma,* 48**,** 330-6.

Kidd, J. M., Sampas, N., Antonacci, F., Graves, T., Fulton, R., Hayden, H. S., Alkan, C., Malig, M., Ventura, M., Giannuzzi, G., Kallicki, J., Anderson, P., Tsalenko, A., Yamada, N. A., Tsang, P., Kaul, R., Wilson, R. K., Bruhn, L. & Eichler, E. E. 2010. Characterization of missing human genome sequences and copy-number polymorphic insertions. *Nat Methods,* 7**,** 365-71.

Kim, Y. J., Hwang, E. S., Kim, I. H. & Yu, D. S. 2006. CD4/CD8 double-positive, acute type of adult T-cell leukemia/lymphoma with extensive cutaneous involvement. *Int J Dermatol,* 45**,** 1193-5.

Kira, J., Nakamura, M., Sawada, T., Koyanagi, Y., Ohori, N., Itoyama, Y., Yamamoto, N., Sakaki, Y. & Goto, I. 1992. Antibody titers to HTLV-I-p40tax protein and gag-env hybrid protein in HTLV-I-associated myelopathy/tropical spastic paraparesis: correlation with increased HTLV-I proviral DNA load. *J Neurol Sci,* 107**,** 98-104.

Kitze, B., Usuku, K., Yamano, Y., Yashiki, S., Nakamura, M., Fujiyoshi, T., Izumo, S., Osame, M. & Sonoda, S. 1998. Human CD4+ T lymphocytes recognize a highly conserved epitope of human T lymphotropic virus type 1 (HTLV-1) env gp21 restricted by HLA DRB1*0101. *Clin Exp Immunol,* 111**,** 278-85.

Koralnik, I. J., Fullen, J. & Franchini, G. 1993. The p12I, p13II, and p30II proteins encoded by human T-cell leukemia/lymphotropic virus type I open reading frames I and II are localized in three different cellular compartments. *J Virol,* 67**,** 2360-6.

Koyanagi, Y., Itoyama, Y., Nakamura, N., Takamatsu, K., Kira, J., Iwamasa, T., Goto, I. & Yamamoto, N. 1993. In vivo infection of human T-cell leukemia virus type I in non-T cells. *Virology,* 196**,** 25-33.

Kozako, T., Akimoto, M., Toji, S., White, Y., Suzuki, S., Arima, T., Suruga, Y., Matsushita, K., Shimeno, H., Soeda, S., Kubota, R., Izumo, S., Uozumi, K. & Arima, N. 2011. Target epitopes of HTLV-1 recognized by class I MHC-restricted cytotoxic T lymphocytes in patients with myelopathy and spastic paraparesis and infected patients with autoimmune disorders. *J Med Virol,* 83**,** 501-9.

Krichbaum-Stenger, K., Poiesz, B. J., Keller, P., Ehrlich, G., Gavalchin, J., Davis, B. H. & Moore, J. L. 1987. Specific adsorption of HTLV-I to various target human and animal cells. *Blood,* 70**,** 1303-11.

Ku, C. C., Murakami, M., Sakamoto, A., Kappler, J. & Marrack, P. 2000. Control of homeostasis of CD8+ memory T cells by opposing cytokines. *Science,* 288**,** 675-8.

Kubota, R., Kawanishi, T., Matsubara, H., Manns, A. & Jacobson, S. 2000. HTLV-I specific IFN-gamma+ CD8+ lymphocytes correlate with the proviral load in peripheral blood of infected individuals. *J Neuroimmunol,* 102**,** 208-15.

Kurtzke, J. F. 1955. A new scale for evaluating disability in multiple sclerosis. *Neurology,* 5**,** 580-3.

Kwok, S., Ehrlich, G., Poiesz, B., Kalish, R. & Sninsky, J. J. 1988. Enzymatic amplification of HTLV-I viral sequences from peripheral blood mononuclear cells and infected tissues. *Blood,* 72**,** 1117-23.

Lagrenade, L., Hanchard, B., Fletcher, V., Cranston, B. & Blattner, W. 1990. Infective dermatitis of Jamaican children: a marker for HTLV-I infection. *Lancet,* 336**,** 1345-7.

Lairmore, M. D., Haines, R. & Anupam, R. 2012. Mechanisms of human T-lymphotropic virus type 1 transmission and disease. *Curr Opin Virol,* 2**,** 474-81.

Lal, R. B., Giam, C. Z., Coligan, J. E. & Rudolph, D. L. 1994. Differential immune responsiveness to the immunodominant epitopes of regulatory proteins (tax and rex) in human T cell lymphotropic virus type I-associated myelopathy. *J Infect Dis,* 169**,** 496-503.

Lal, R. B., Owen, S. M., Rudolph, D. L., Dawson, C. & Prince, H. 1995. In vivo cellular tropism of human T-lymphotropic virus type II is not restricted to CD8+ cells. *Virology,* 210**,** 441-7.

Landry, S., Halin, M., Vargas, A., Lemasson, I., Mesnard, J. M. & Barbeau, B. 2009. Upregulation of human T-cell leukemia virus type 1 antisense transcription by the viral tax protein. *J Virol,* 83**,** 2048-54.

Larocca, D., Chao, L. A., Seto, M. H. & Brunck, T. K. 1989. Human T-cell leukemia virus minus strand transcription in infected T-cells. *Biochem Biophys Res Commun,* 163**,** 1006-13.

Lee, B. K., Bhinge, A. A. & Iyer, V. R. 2011. Wide-ranging functions of E2F4 in transcriptional activation and repression revealed by genome-wide analysis. *Nucleic Acids Res,* 39**,** 3558-73.

Lee, K., Ambrose, Z., Martin, T. D., Oztop, I., Mulky, A., Julias, J. G., Vandegraaff, N., Baumann, J. G., Wang, R., Yuen, W., Takemura, T., Shelton, K., Taniuchi, I., Li, Y., Sodroski, J., Littman, D. R., Coffin, J. M., Hughes, S. H., Unutmaz, D., Engelman, A. & Kewalramani, V. N. 2010. Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe,* 7**,** 221-33.

Lee, M. S. & Craigie, R. 1998. A previously unidentified host protein protects retroviral DNA from autointegration. *Proc Natl Acad Sci U S A,* 95**,** 1528-33.

Lee, Y. K., Chew, A., Phan, H., Greenhalgh, D. G. & Cho, K. 2008. Genome-wide expression profiles of endogenous retroviruses in lymphoid tissues and their biological properties. *Virology,* 373**,** 263-73.

Lehky, T. J., Fox, C. H., Koenig, S., Levin, M. C., Flerlage, N., Izumo, S., Sato, E., Raine, C. S., Osame, M. & Jacobson, S. 1995. Detection of human T-lymphotropic virus type I (HTLV-I) tax RNA in the central nervous system of HTLV-I-associated

myelopathy/tropical spastic paraparesis patients by in situ hybridization. *Ann Neurol,* 37**,** 167-75.

Lemasson, I., Polakowski, N. J., Laybourn, P. J. & Nyborg, J. K. 2002. Transcription factor binding and histone modifications on the integrated proviral promoter in human T-cell leukemia virus-I-infected T-cells. *J Biol Chem,* 277**,** 49459-65.

Lenasi, T., Contreras, X. & Peterlin, B. M. 2008. Transcriptional interference antagonizes proviral gene expression to promote HIV latency. *Cell Host Microbe,* 4**,** 123-33.

Levin, A., Hayouka, Z., Friedler, A. & Loyter, A. 2010. Transportin 3 and importin alpha are required for effective nuclear import of HIV-1 integrase in virus-infected cells. *Nucleus,* 1**,** 422-31.

Levin, M. C., Lee, S. M., Kalume, F., Morcos, Y., Dohan, F. C., Jr., Hasty, K. A., Callaway, J. C., Zunt, J., Desiderio, D. & Stuart, J. M. 2002. Autoimmunity due to molecular mimicry as a cause of neurological disease. *Nat Med,* 8**,** 509-13.

Lewinski, M. K., Bisgrove, D., Shinn, P., Chen, H., Hoffmann, C., Hannenhalli, S., Verdin, E., Berry, C. C., Ecker, J. R. & Bushman, F. D. 2005. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J Virol,* 79**,** 6610-9.

Lewis, P. F. & Emerman, M. 1994. Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. *J Virol,* 68**,** 510-6.

Lezin, A., Olindo, S., Oliere, S., Varrin-Doyer, M., Marlin, R., Cabre, P., Smadja, D. & Cesaire, R. 2005. Human T lymphotropic virus type I (HTLV-I) proviral load in cerebrospinal fluid: a new criterion for the diagnosis of HTLV-I-associated myelopathy/tropical spastic paraparesis? *J Infect Dis,* 191**,** 1830-4.

Liao, W., Lin, J. X., Wang, L., Li, P. & Leonard, W. J. 2011. Modulation of cytokine receptors by IL-2 broadly regulates differentiation into helper T cell lineages. *Nat Immunol,* 12**,** 551-9.

Lima, M. A., Bica, R. B. & Araujo, A. Q. 2005. Gender influence on the progression of HTLV-I associated myelopathy/tropical spastic paraparesis. *J Neurol Neurosurg Psychiatry,* 76**,** 294-6.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. 2012. Comparison of next-generation sequencing systems. *J Biomed Biotechnol,* 2012**,** 251364.

Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M. & Poeschla, E. M. 2004. LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J Virol,* 78**,** 9524-37.

Macatonia, S. E., Cruickshank, J. K., Rudge, P. & Knight, S. C. 1992. Dendritic cells from patients with tropical spastic paraparesis are infected with HTLV-1 and stimulate autologous lymphocyte proliferation. *AIDS Res Hum Retroviruses,* 8**,** 1699-706.

Macnamara, A., Rowan, A., Hilburn, S., Kadolsky, U., Fujiwara, H., Suemori, K., Yasukawa, M., Taylor, G., Bangham, C. R. & Asquith, B. 2010. HLA class I binding of HBZ determines outcome in HTLV-1 infection. *PLoS Pathog,* 6**,** e1001117.

Maeda, Y., Furukawa, M., Takehara, Y., Yoshimura, K., Miyamoto, K., Matsuura, T., Morishima, Y., Tajima, K., Okochi, K. & Hinuma, Y. 1984. Prevalence of possible

adult T-cell leukemia virus-carriers among volunteer blood donors in Japan: a nation-wide study. *Int J Cancer,* 33**,** 717-20.

Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z. & Engelborghs, Y. 2003. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem,* 278**,** 33528-39.

Mahieux, R. & Gessain, A. 2007. Adult T-cell leukemia/lymphoma and HTLV-1. *Curr Hematol Malig Rep,* 2**,** 257-64.

Mahieux, R. & Gessain, A. 2011. HTLV-3/STLV-3 and HTLV-4 viruses: discovery, epidemiology, serology and molecular aspects. *Viruses,* 3**,** 1074-90.

Majorovits, E., Nejmeddine, M., Tanaka, Y., Taylor, G. P., Fuller, S. D. & Bangham, C. R. 2008. Human T-lymphotropic virus-1 visualized at the virological synapse by electron tomography. *PLoS One,* 3**,** e2251.

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 Retrovirol,* 17**,** 167-70.

Manel, N., Kim, F. J., Kinet, S., Taylor, N., Sitbon, M. & Battini, J. L. 2003. The ubiquitous glucose transporter GLUT-1 is a receptor for HTLV. *Cell,* 115**,** 449-59.

Marcos, L. A., Terashima, A., Dupont, H. L. & Gotuzzo, E. 2008. Strongyloides hyperinfection syndrome: an emerging global infectious disease. *Trans R Soc Trop Med Hyg,* 102**,** 314-8.

Mariner, J. M., Lantz, V., Waldmann, T. A. & Azimi, N. 2001. Human T cell lymphotropic virus type I Tax activates IL-15R alpha gene expression through an NF-kappa B site. *J Immunol,* 166**,** 2602-9.

Marriott, S. J. & Semmes, O. J. 2005. Impact of HTLV-I Tax on cell cycle progression and the cellular DNA damage repair response. *Oncogene,* 24**,** 5986-95.

Matsuoka, M. & Jeang, K. T. 2005. Human T-cell leukemia virus type I at age 25: a progress report. *Cancer Res,* 65**,** 4467-70.

Matsuoka, M. & Jeang, K. T. 2007. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. *Nat Rev Cancer,* 7**,** 270-80.

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 Neurovirol,* 7**,** 228-34.

Maxfield, L. F., Fraize, C. D. & Coffin, J. M. 2005. Relationship between retroviral DNA-integration-site selection and host cell transcription. *Proc Natl Acad Sci U S A,* 102**,** 1436-41.

Mazurov, D., Ilinskaya, A., Heidecker, G., Lloyd, P. & Derse, D. 2010. Quantitative comparison of HTLV-1 and HIV-1 cell-to-cell infection with new replication dependent vectors. *PLoS Pathog,* 6**,** e1000788.

Mcclure, M. O., Sommerfelt, M. A., Marsh, M. & Weiss, R. A. 1990. The pH independence of mammalian retrovirus infection. *J Gen Virol,* 71 (Pt 4)**,** 767-73.

Mcdonald, D., Vodicka, M. A., Lucero, G., Svitkina, T. M., Borisy, G. G., Emerman, M. & Hope, T. J. 2002. Visualization of the intracellular behavior of HIV in living cells. *J Cell Biol,* 159**,** 441-52.

Mcdonald, D., Wu, L., Bohks, S. M., Kewalramani, V. N., Unutmaz, D. & Hope, T. J. 2003. Recruitment of HIV and its receptors to dendritic cell-T cell junctions. *Science,* 300**,** 1295-7.

Meekings, K. N., Leipzig, J., Bushman, F. D., Taylor, G. P. & Bangham, C. R. 2008. HTLV-1 integration into transcriptionally active genomic regions is associated with proviral expression and with HAM/TSP. *PLoS Pathog,* 4**,** e1000027.

Miller, J. M., Miller, L. D., Olson, C. & Gillette, K. G. 1969. Virus-like particles in phytohemagglutinin-stimulated lymphocyte cultures with reference to bovine lymphosarcoma. *J Natl Cancer Inst,* 43**,** 1297-305.

Mitchell, R. S., Beitzel, B. F., Schroder, A. R., Shinn, P., Chen, H., Berry, C. C., Ecker, J. R. & Bushman, F. D. 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol,* 2**,** E234.

Mitsuya, H., Jarrett, R. F., Cossman, J., Cohen, O. J., Kao, C. S., Guo, H. G., Reitz, M. S. & Broder, S. 1986. Infection of human T lymphotropic virus-I-specific immune T cell clones by human T lymphotropic virus-I. *J Clin Invest,* 78**,** 1302-10.

Miyazaki, M., Yasunaga, J., Taniguchi, Y., Tamiya, S., Nakahata, T. & Matsuoka, M. 2007. Preferential selection of human T-cell leukemia virus type 1 provirus lacking the 5' long terminal repeat during oncogenesis. *J Virol,* 81**,** 5714-23.

Miyoshi, I., Kubonishi, I., Sumida, M., Hiraki, S., Tsubota, T., Kimura, I., Miyamoto, K. & Sato, J. 1980. A novel T-cell line derived from adult T-cell leukemia. *Gann,* 71**,** 155-6.

Mochizuki, M., Watanabe, T., Yamaguchi, K., Takatsuki, K., Yoshimura, K., Shirao, M., Nakashima, S., Mori, S., Araki, S. & Miyata, N. 1992. HTLV-I uveitis: a distinct clinical entity caused by HTLV-I. *Jpn J Cancer Res,* 83**,** 236-9.

Monteiro, J., Hingorani, R., Choi, I. H., Silver, J., Pergolizzi, R. & Gregersen, P. K. 1995. Oligoclonality in the human CD8+ T cell repertoire in normal subjects and monozygotic twins: implications for studies of infectious and autoimmune diseases. *Mol Med,* 1**,** 614-24.

Morgan, O. S., Rodgers-Johnson, P., Mora, C. & Char, G. 1989. HTLV-1 and polymyositis in Jamaica. *Lancet,* 2**,** 1184-7.

Moritoyo, T., Izumo, S., Moritoyo, H., Tanaka, Y., Kiyomatsu, Y., Nagai, M., Usuku, K., Sorimachi, M. & Osame, M. 1999. Detection of human T-lymphotropic virus type I p40taxprotein in cerebrospinal fluid cells from patients with human T-lymphotropic virus type I-associated myelopathy/tropical spastic paraparesis. *Journal of Neurovirology,* 5**,** 241-248.

Moriuchi, M. & Moriuchi, H. 2004. Seminal fluid enhances replication of human T-cell leukemia virus type 1: implications for sexual transmission. *J Virol,* 78**,** 12709-11.

Mosley, A. J., Meekings, K. N., Mccarthy, C., Shepherd, D., Cerundolo, V., Mazitschek, R., Tanaka, Y., Taylor, G. P. & Bangham, C. R. 2006. Histone deacetylase inhibitors increase virus gene expression but decrease CD8+ cell antiviral function in HTLV-1 infection. *Blood,* 108**,** 3801-7.

Moulard, M. & Decroly, E. 2000. Maturation of HIV envelope glycoprotein precursors by cellular endoproteases. *Biochim Biophys Acta,* 1469**,** 121-32.

Moules, V., Pomier, C., Sibon, D., Gabet, A. S., Reichert, M., Kerkhofs, P., Willems, L., Mortreux, F. & Wattel, E. 2005. Fate of premalignant clones during the asymptomatic phase preceding lymphoid malignancy. *Cancer Res,* 65**,** 1234-43.

Murphy, E. L., Figueroa, J. P., Gibbs, W. N., Brathwaite, A., Holding-Cobham, M., Waters, D., Cranston, B., Hanchard, B. & Blattner, W. A. 1989a. Sexual transmission of human T-lymphotropic virus type I (HTLV-I). *Ann Intern Med,* 111**,** 555-60.

Murphy, E. L., Hanchard, B., Figueroa, J. P., Gibbs, W. N., Lofters, W. S., Campbell, M., Goedert, J. J. & Blattner, W. A. 1989b. Modelling the risk of adult T-cell leukemia/lymphoma in persons infected with human T-lymphotropic virus type I. *Int J Cancer,* 43**,** 250-3.

Murphy, E. L., Lee, T. H., Chafets, D., Nass, C. C., Wang, B., Loughlin, K., Smith, D. & Investigators, H. O. S. 2004. Higher human T lymphotropic virus (HTLV) provirus load is associated with HTLV-I versus HTLV-II, with HTLV-II subtype A versus B, and with male sex and a history of blood transfusion. *J Infect Dis,* 190**,** 504-10.

Nagai, M., Brennan, M. B., Sakai, J. A., Mora, C. A. & Jacobson, S. 2001a. CD8(+) T cells are an in vivo reservoir for human T-cell lymphotropic virus type I. *Blood,* 98**,** 1858-61.

Nagai, M., Kubota, R., Greten, T. F., Schneck, J. P., Leist, T. P. & Jacobson, S. 2001b. Increased activated human T cell lymphotropic virus type I (HTLV-I) Tax11-19-specific memory and effector CD8+ cells in patients with HTLV-I-associated myelopathy/tropical spastic paraparesis: correlation with HTLV-I provirus load. *J Infect Dis,* 183**,** 197-205.

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 Neurovirol,* 4**,** 586-93.

Nagai, M., Yashiki, S., Fujiyoshi, T., Fujiyama, C., Kitze, B., Izumo, S., Osame, M. & Sonoda, S. 1996. Characterization of a unique T-cell clone established from a patient with HAM/TSP which recognized HTLV-I-infected T-cell antigens as well as spinal cord tissue antigens. *J Neuroimmunol,* 65**,** 97-105.

Naghavi, M. H., Valente, S., Hatziioannou, T., De Los Santos, K., Wen, Y., Mott, C., Gundersen, G. G. & Goff, S. P. 2007. Moesin regulates stable microtubule formation and limits retroviral infection in cultured cells. *EMBO J,* 26**,** 41-52.

Nakada, K., Yamaguchi, K., Furugen, S., Nakasone, T., Nakasone, K., Oshiro, Y., Kohakura, M., Hinuma, Y., Seiki, M., Yoshida, M. & Et Al. 1987. Monoclonal integration of HTLV-I proviral DNA in patients with strongyloidiasis. *Int J Cancer,* 40**,** 145-8.

Nakagawa, M., Izumo, S., Ijichi, S., Kubota, H., Arimura, K., Kawabata, M. & Osame, M. 1995. HTLV-I-associated myelopathy: analysis of 213 patients based on clinical features and laboratory findings. *J Neurovirol,* 1**,** 50-61.

Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F. H., Verma, I. M. & Trono, D. 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science,* 272**,** 263-7.

Nam, S. H., Kidokoro, M., Shida, H. & Hatanaka, M. 1988. Processing of gag precursor polyprotein of human T-cell leukemia virus type I by virus-encoded protease. *J Virol,* 62**,** 3718-28.

Nejmeddine, M. & Bangham, C. R. 2010. The HTLV-1 Virological Synapse. *Viruses,* 2**,** 1427-47.

Nejmeddine, M., Barnard, A. L., Tanaka, Y., Taylor, G. P. & Bangham, C. R. 2005. Human T-lymphotropic virus, type 1, tax protein triggers microtubule reorientation in the virological synapse. *J Biol Chem,* 280**,** 29653-60.

Nejmeddine, M., Negi, V. S., Mukherjee, S., Tanaka, Y., Orth, K., Taylor, G. P. & Bangham, C. R. 2009. HTLV-1-Tax and ICAM-1 act on T-cell signal pathways to polarize the microtubule-organizing center at the virological synapse. *Blood,* 114**,** 1016-25.

Neuveut, C., Wei, Y. & Buendia, M. A. 2010. Mechanisms of HBV-related hepatocarcinogenesis. *J Hepatol,* 52**,** 594-604.

Newbound, G. C., Andrews, J. M., O'rourke, J. P., Brady, J. N. & Lairmore, M. D. 1996. Human T-cell lymphotropic virus type 1 Tax mediates enhanced transcription in CD4+ T lymphocytes. *J Virol,* 70**,** 2101-6.

Ng, P. W., Iha, H., Iwanaga, Y., Bittner, M., Chen, Y., Jiang, Y., Gooden, G., Trent, J. M., Meltzer, P., Jeang, K. T. & Zeichner, S. L. 2001. Genome-wide expression changes induced by HTLV-1 Tax: evidence for MLK-3 mixed lineage kinase involvement in Tax-mediated NF-kappaB activation. *Oncogene,* 20**,** 4484-96.

Nicot, C. 2005. Current views in HTLV-I-associated adult T-cell leukemia/lymphoma. *Am J Hematol,* 78**,** 232-9.

Nicot, C., Dundr, M., Johnson, J. M., Fullen, J. R., Alonzo, N., Fukumoto, R., Princler, G. L., Derse, D., Misteli, T. & Franchini, G. 2004. HTLV-1-encoded p30II is a post-transcriptional negative regulator of viral replication. *Nat Med,* 10**,** 197-201.

Nicot, C., Mulloy, J. C., Ferrari, M. G., Johnson, J. M., Fu, K., Fukumoto, R., Trovato, R., Fullen, J., Leonard, W. J. & Franchini, G. 2001. HTLV-1 p12(I) protein enhances STAT5 activation and decreases the interleukin-2 requirement for proliferation of primary human peripheral blood mononuclear cells. *Blood,* 98**,** 823-9.

Niewiesk, S., Daenke, S., Parker, C. E., Taylor, G., Weber, J., Nightingale, S. & Bangham, C. R. 1994. The transactivator gene of human T-cell leukemia virus type I is more variable within and between healthy carriers than patients with tropical spastic paraparesis. *J Virol,* 68**,** 6778-81.

Niewiesk, S., Daenke, S., Parker, C. E., Taylor, G., Weber, J., Nightingale, S. & Bangham, C. R. 1995. Naturally occurring variants of human T-cell leukemia virus type I Tax protein impair its recognition by cytotoxic T lymphocytes and the transactivation function of Tax. *J Virol,* 69**,** 2649-53.

Nishioka, K., Maruyama, I., Sato, K., Kitajima, I., Nakajima, Y. & Osame, M. 1989. Chronic inflammatory arthropathy associated with HTLV-I. *Lancet,* 1**,** 441.

Okochi, K. & Sato, H. 1984. Transmission of ATLV (HTLV-I) through blood transfusion. *Princess Takamatsu Symp,* 15**,** 129-35.

Overbaugh, J. & Bangham, C. R. 2001. Selection forces and constraints on retroviral sequence variation. *Science,* 292**,** 1106-9.

Ozawa, T., Itoyama, T., Sadamori, N., Yamada, Y., Hata, T., Tomonaga, M. & Isobe, M. 2004. Rapid isolation of viral integration site reveals frequent integration of HTLV-1 into expressed loci. *J Hum Genet,* 49**,** 154-65.

Pais-Correia, A. M., Sachse, M., Guadagnini, S., Robbiati, V., Lasserre, R., Gessain, A., Gout, O., Alcover, A. & Thoulouze, M. I. 2010. Biofilm-like extracellular viral assemblies mediate HTLV-1 cell-to-cell transmission at virological synapses. *Nat Med,* 16**,** 83-9.

Paliard, X., Malefijt, R. W., De Vries, J. E. & Spits, H. 1988. Interleukin-4 mediates CD8 induction on human CD4+ T-cell clones. *Nature,* 335**,** 642-4.

Parker, C. E., Daenke, S., Nightingale, S. & Bangham, C. R. 1992. Activated, HTLV-1-specific cytotoxic T-lymphocytes are found in healthy seropositives as well as in patients with tropical spastic paraparesis. *Virology,* 188**,** 628-36.

Pedral-Sampaio, D. B., Martins Netto, E., Pedrosa, C., Brites, C., Duarte, M. & Harrington, W., Jr. 1997. Co-Infection of Tuberculosis and HIV/HTLV Retroviruses: Frequency and Prognosis Among Patients Admitted in a Brazilian Hospital. *Braz J Infect Dis,* 1**,** 31-35.

Permanyer, M., Ballana, E. & Este, J. A. 2010. Endocytosis of HIV: anything goes. *Trends Microbiol,* 18**,** 543-51.

Pinon, J. D., Klasse, P. J., Jassal, S. R., Welson, S., Weber, J., Brighty, D. W. & Sattentau, Q. J. 2003. Human T-cell leukemia virus type 1 envelope glycoprotein gp46 interacts with cell surface heparan sulfate proteoglycans. *J Virol,* 77**,** 9922-30.

Pique, C., Pham, D., Tursz, T. & Dokhelar, M. C. 1992. Human T-cell leukemia virus type I envelope protein maturation process: requirements for syncytium formation. *J Virol,* 66**,** 906-13.

Plachy, J., Kotab, J., Divina, P., Reinisova, M., Senigl, F. & Hejnar, J. 2010. Proviruses selected for high and stable expression of transduced genes accumulate in broadly transcribed genome areas. *J Virol,* 84**,** 4204-11.

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**,** 7415-9.

Polz, M. F. & Cavanaugh, C. M. 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol,* 64**,** 3724-30.

Popovic, M., Flomenberg, N., Volkman, D. J., Mann, D., Fauci, A. S., Dupont, B. & Gallo, R. C. 1984. Alteration of T-cell functions by infection with HTLV-I or HTLV-II. *Science,* 226**,** 459-62.

Proietti, F. A., Carneiro-Proietti, A. B., Catalan-Soares, B. C. & Murphy, E. L. 2005. Global epidemiology of HTLV-I infection and associated diseases. *Oncogene,* 24**,** 6058-68.

Pruss, D., Reeves, R., Bushman, F. D. & Wolffe, A. P. 1994. The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J Biol Chem,* 269**,** 25031-41.

Pryciak, P. M. & Varmus, H. E. 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell,* 69**,** 769-80.

Quail, M. A., Kozarewa, I., Smith, F., Scally, A., Stephens, P. J., Durbin, R., Swerdlow, H. & Turner, D. J. 2008. A large genome center's improvements to the Illumina sequencing system. *Nat Methods,* 5**,** 1005-10.

Raha, D., Wang, Z., Moqtaderi, Z., Wu, L., Zhong, G., Gerstein, M., Struhl, K. & Snyder, M. 2010. Close association of RNA polymerase II and many transcription factors with Pol III genes. *Proc Natl Acad Sci U S A,* 107**,** 3639-44.

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M., Durham, T., Zhang, X., Donaghey, J., Epstein, C. B., Regev, A. & Bernstein, B. E. 2011. Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell,* 147**,** 1628-39.

Ratner, L., Harrington, W., Feng, X., Grant, C., Jacobson, S., Noy, A., Sparano, J., Lee, J., Ambinder, R., Campbell, N., Lairmore, M. & Consortium, A. M. 2009. Human T cell leukemia virus reactivation with progression of adult T-cell leukemia-lymphoma. *PLoS One,* 4**,** e4420.

Raza, S., Naik, S., Kancharla, V. P., Tafera, F. & Kalavar, M. R. 2010. Dual-Positive (CD4+/CD8+) Acute Adult T-Cell Leukemia/Lymphoma Associated with Complex Karyotype and Refractory Hypercalcemia: Case Report and Literature Review. *Case Rep Oncol,* 3**,** 489-94.

Rende, F., Cavallari, I., Corradin, A., Silic-Benussi, M., Toulza, F., Toffolo, G. M., Tanaka, Y., Jacobson, S., Taylor, G. P., D'agostino, D. M., Bangham, C. R. & Ciminale, V. 2011. Kinetics and intracellular compartmentalization of HTLV-1 gene expression: nuclear retention of HBZ mRNAs. *Blood,* 117**,** 4855-9.

Rende, F., Cavallari, I., Romanelli, M. G., Diani, E., Bertazzoni, U. & Ciminale, V. 2012. Comparison of the Genetic Organization, Expression Strategies and Oncogenic Potential of HTLV-1 and HTLV-2. *Leuk Res Treatment,* 2012**,** 876153.

Richardson, J. H., Edwards, A. J., Cruickshank, J. K., Rudge, P. & Dalgleish, A. G. 1990. In vivo cellular tropism of human T-cell leukemia virus type 1. *J Virol,* 64**,** 5682-7.

Rodriguez, S. M., Florins, A., Gillet, N., De Brogniez, A., Sanchez-Alcaraz, M. T., Boxus, M., Boulanger, F., Gutierrez, G., Trono, K., Alvarez, I., Vagnoni, L. & Willems, L. 2011. Preventive and therapeutic strategies for bovine leukemia virus: lessons for HTLV. *Viruses,* 3**,** 1210-48.

Roe, T., Reynolds, T. C., Yu, G. & Brown, P. O. 1993. Integration of murine leukemia virus DNA depends on mitosis. *EMBO J,* 12**,** 2099-108.

Roth, S. L., Malani, N. & Bushman, F. D. 2011. Gammaretroviral integration into nucleosomal target DNA in vivo. *J Virol,* 85**,** 7393-401.

Rozowsky, J., Abyzov, A., Wang, J., Alves, P., Raha, D., Harmanci, A., Leng, J., Bjornson, R., Kong, Y., Kitabayashi, N., Bhardwaj, N., Rubin, M., Snyder, M. & Gerstein, M. 2011. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol,* 7**,** 522.

Saida, T., Saida, K., Funauchi, M., Nishiguchi, E., Nakajima, M., Matsuda, S., Ohta, M., Ohta, K., Nishitani, H. & Hatanaka, M. 1988. HTLV-I myelitis: isolation of virus, genomic analysis, and infection in neural cell cultures. *Ann N Y Acad Sci,* 540**,** 636-8.

Saito, M., Matsuzaki, T., Satou, Y., Yasunaga, J., Saito, K., Arimura, K., Matsuoka, M. & Ohara, Y. 2009. In vivo expression of the HBZ gene of HTLV-1 correlates with

proviral load, inflammatory markers and disease severity in HTLV-1 associated myelopathy/tropical spastic paraparesis (HAM/TSP). *Retrovirology,* 6**,** 19.

Sakai, H., Kawamura, M., Sakuragi, J., Sakuragi, S., Shibata, R., Ishimoto, A., Ono, N., Ueda, S. & Adachi, A. 1993. Integration is essential for efficient gene expression of human immunodeficiency virus type 1. *J Virol,* 67**,** 1169-74.

Santoni, F. A., Hartley, O. & Luban, J. 2010. Deciphering the code for retroviral integration target site selection. *PLoS Comput Biol,* 6**,** e1001008.

Satou, Y., Yasunaga, J., Yoshida, M. & Matsuoka, M. 2006. HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. *Proc Natl Acad Sci U S A,* 103**,** 720-5.

Satou, Y., Yasunaga, J., Zhao, T., Yoshida, M., Miyazato, P., Takai, K., Shimizu, K., Ohshima, K., Green, P. L., Ohkura, N., Yamaguchi, T., Ono, M., Sakaguchi, S. & Matsuoka, M. 2011. HTLV-1 bZIP factor induces T-cell lymphoma and systemic inflammation in vivo. *PLoS Pathog,* 7**,** e1001274.

Schaller, T., Ocwieja, K. E., Rasaiyaah, J., Price, A. J., Brady, T. L., Roth, S. L., Hue, S., Fletcher, A. J., Lee, K., Kewalramani, V. N., Noursadeghi, M., Jenner, R. G., James, L. C., Bushman, F. D. & Towers, G. J. 2011. HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathog,* 7**,** e1002439.

Schneider, U., Schwenk, H. U. & Bornkamm, G. 1977. Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. *Int J Cancer,* 19**,** 621-6.

Schrijvers, R., Vets, S., De Rijck, J., Malani, N., Bushman, F. D., Debyser, Z. & Gijsbers, R. 2012. HRP-2 determines HIV-1 integration site selection in LEDGF/p75 depleted cells. *Retrovirology,* 9**,** 84.

Schroder, A. R., Shinn, P., Chen, H., Berry, C., Ecker, J. R. & Bushman, F. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell,* 110**,** 521-9.

Seich Al Basatena, N. K., Macnamara, A., Vine, A. M., Thio, C. L., Astemborski, J., Usuku, K., Osame, M., Kirk, G. D., Donfield, S. M., Goedert, J. J., Bangham, C. R., Carrington, M., Khakoo, S. I. & Asquith, B. 2011. KIR2DL2 enhances protective and detrimental HLA class I-mediated immunity in chronic viral infection. *PLoS Pathog,* 7**,** e1002270.

Seiki, M., Eddy, R., Shows, T. B. & Yoshida, M. 1984. Nonspecific integration of the HTLV provirus genome into adult T-cell leukaemia cells. *Nature,* 309**,** 640-2.

Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc Natl Acad Sci U S A,* 80**,** 3618-22.

Seiki, M., Inoue, J., Hidaka, M. & Yoshida, M. 1988. Two cis-acting elements responsible for posttranscriptional trans-regulation of gene expression of human T-cell leukemia virus type I. *Proc Natl Acad Sci U S A,* 85**,** 7124-8.

Setoyama, M., Kerdel, F. A., Elgart, G., Kanzaki, T. & Byrnes, J. J. 1998. Detection of HTLV-1 by polymerase chain reaction in situ hybridization in adult T-cell leukemia/lymphoma. *Am J Pathol,* 152**,** 683-9.

Shan, L., Yang, H. C., Rabi, S. A., Bravo, H. C., Shroff, N. S., Irizarry, R. A., Zhang, H., Margolick, J. B., Siliciano, J. D. & Siliciano, R. F. 2011. Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. *J Virol,* 85**,** 5384-93.

Sharma, A., Larue, R. C., Plumb, M. R., Malani, N., Male, F., Slaughter, A., Kessl, J. J., Shkriabai, N., Coward, E., Aiyer, S. S., Green, P. L., Wu, L., Roth, M. J., Bushman, F. D. & Kvaratskhelia, M. 2013. BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc Natl Acad Sci U S A,* 110**,** 12036-41.

Shearwin, K. E., Callen, B. P. & Egan, J. B. 2005. Transcriptional interference--a crash course. *Trends Genet,* 21**,** 339-45.

Shimoyama, M. 1991. Diagnostic criteria and classification of clinical subtypes of adult T-cell leukaemia-lymphoma. A report from the Lymphoma Study Group (1984-87). *Br J Haematol,* 79**,** 428-37.

Shuh, M. & Beilke, M. 2005. The human T-cell leukemia virus type 1 (HTLV-1): new insights into the clinical aspects and molecular pathogenesis of adult T-cell leukemia/lymphoma (ATLL) and tropical spastic paraparesis/HTLV-associated myelopathy (TSP/HAM). *Microsc Res Tech,* 68**,** 176-96.

Shun, M. C., Raghavendra, N. K., Vandegraaff, N., Daigle, J. E., Hughes, S., Kellam, P., Cherepanov, P. & Engelman, A. 2007. LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev,* 21**,** 1767-78.

Sibon, D., Gabet, A. S., Zandecki, M., Pinatel, C., Thete, J., Delfau-Larue, M. H., Rabaaoui, S., Gessain, A., Gout, O., Jacobson, S., Mortreux, F. & Wattel, E. 2006. HTLV-1 propels untransformed CD4 lymphocytes into the cell cycle while protecting CD8 cells from death. *J Clin Invest,* 116**,** 974-83.

Siliciano, R. F. & Greene, W. C. 2011. HIV latency. *Cold Spring Harb Perspect Med,* 1**,** a007096.

Silverman, L. R., Phipps, A. J., Montgomery, A., Ratner, L. & Lairmore, M. D. 2004. Human T-Cell Lymphotropic Virus Type 1 Open Reading Frame II-Encoded p30II Is Required for In Vivo Replication: Evidence of In Vivo Reversion. *Journal of Virology,* 78**,** 3837-3845.

Sinha-Datta, U., Datta, A., Ghorbel, S., Dodon, M. D. & Nicot, C. 2007. Human T-cell lymphotrophic virus type I rex and p30 interactions govern the switch between virus latency and replication. *J Biol Chem,* 282**,** 14608-15.

Slattery, J. P., Franchini, G. & Gessain, A. 1999. Genomic evolution, patterns of global dissemination, and interspecies transmission of human and simian T-cell leukemia/lymphotropic viruses. *Genome Res,* 9**,** 525-40.

Snyder, C. M., Cho, K. S., Bonnett, E. L., Van Dommelen, S., Shellam, G. R. & Hill, A. B. 2008. Memory inflation during chronic viral infection is maintained by continuous production of short-lived, functional T cells. *Immunity,* 29**,** 650-9.

Suemori, K., Fujiwara, H., Ochi, T., Ogawa, T., Matsuoka, M., Matsumoto, T., Mesnard, J. M. & Yasukawa, M. 2009. HBZ is an immunogenic protein, but not a target antigen for human T-cell leukemia virus type 1-specific cytotoxic T lymphocytes. *J Gen Virol,* 90**,** 1806-11.

Suzuki, S., Zhou, Y., Refaat, A., Takasaki, I., Koizumi, K., Yamaoka, S., Tabuchi, Y., Saiki, I. & Sakurai, H. 2010. Human T cell lymphotropic virus 1 manipulates interferon regulatory signals by controlling the TAK1-IRF3 and IRF4 pathways. *J Biol Chem,* 285**,** 4441-6.

Takemoto, S., Matsuoka, M., Yamaguchi, K. & Takatsuki, K. 1994. A novel diagnostic method of adult T-cell leukemia: monoclonal integration of human T-cell lymphotropic virus type I provirus DNA detected by inverse polymerase chain reaction. *Blood,* 84**,** 3080-5.

Tamiya, S., Matsuoka, M., Etoh, K., Watanabe, T., Kamihira, S., Yamaguchi, K. & Takatsuki, K. 1996. Two types of defective human T-lymphotropic virus type I provirus in adult T-cell leukemia. *Blood,* 88**,** 3065-73.

Taniguchi, Y., Nosaka, K., Yasunaga, J., Maeda, M., Mueller, N., Okayama, A. & Matsuoka, M. 2005. Silencing of human T-cell leukemia virus type I gene transcription by epigenetic mechanisms. *Retrovirology,* 2**,** 64.

Tateno, M., Kondo, N., Itoh, T., Chubachi, T., Togashi, T. & Yoshiki, T. 1984. Rat lymphoid cell lines with human T cell leukemia virus production. I. Biological and serological characterization. *J Exp Med,* 159**,** 1105-16.

Tattermusch, S., Skinner, J. A., Chaussabel, D., Banchereau, J., Berry, M. P., Mcnab, F. W., O'garra, A., Taylor, G. P. & Bangham, C. R. 2012. Systems biology approaches reveal a specific interferon-inducible signature in HTLV-1 associated myelopathy. *PLoS Pathog,* 8**,** e1002480.

Taylor, G. P., Bodeus, M., Courtois, F., Pauli, G., Del Mistro, A., Machuca, A., Padua, E., Andersson, S., Goubau, P., Chieco-Bianchi, L., Soriano, V., Coste, J., Ades, A. E. & Weber, J. N. 2005. The seroepidemiology of human T-lymphotropic viruses: types I and II in Europe: a prospective study of pregnant women. *J Acquir Immune Defic Syndr,* 38**,** 104-9.

Tosswill, J. H., Taylor, G. P., Tedder, R. S. & Mortimer, P. P. 2000. HTLV-I/II associated disease in England and Wales, 1993-7: retrospective review of serology requests. *BMJ,* 320**,** 611-2.

Toulza, F., Heaps, A., Tanaka, Y., Taylor, G. P. & Bangham, C. R. 2008. High frequency of CD4+FoxP3+ cells in HTLV-1 infection: inverse correlation with HTLV-1-specific CTL response. *Blood,* 111**,** 5047-53.

Toulza, F., Nosaka, K., Tanaka, Y., Schioppa, T., Balkwill, F., Taylor, G. P. & Bangham, C. R. 2010. Human T-lymphotropic virus type 1-induced CC chemokine ligand 22 maintains a high frequency of functional FoxP3+ regulatory T cells. *J Immunol,* 185**,** 183-9.

Trojer, P., Cao, A. R., Gao, Z., Li, Y., Zhang, J., Xu, X., Li, G., Losson, R., Erdjument-Bromage, H., Tempst, P., Farnham, P. J. & Reinberg, D. 2011. L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. *Mol Cell,* 42**,** 438-50.

Tsukasaki, K., Hermine, O., Bazarbachi, A., Ratner, L., Ramos, J. C., Harrington, W., Jr., O'mahony, D., Janik, J. E., Bittencourt, A. L., Taylor, G. P., Yamaguchi, K., Utsunomiya, A., Tobinai, K. & Watanabe, T. 2009. Definition, prognostic factors, treatment, and response criteria of adult T-cell leukemia-lymphoma: a proposal from an international consensus meeting. *J Clin Oncol,* 27**,** 453-9.

Turelli, P., Doucas, V., Craig, E., Mangeat, B., Klages, N., Evans, R., Kalpana, G. & Trono, D. 2001. Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. *Mol Cell,* 7**,** 1245-54.

Uchiyama, T., Yodoi, J., Sagawa, K., Takatsuki, K. & Uchino, H. 1977. Adult T-cell leukemia: clinical and hematologic features of 16 cases. *Blood,* 50**,** 481-92.

Uren, A. G., Kool, J., Berns, A. & Van Lohuizen, M. 2005. Retroviral insertional mutagenesis: past, present and future. *Oncogene,* 24**,** 7656-72.

Ureta-Vidal, A., Angelin-Duclos, C., Tortevoye, P., Murphy, E., Lepere, J. F., Buigues, R. P., Jolly, N., Joubert, M., Carles, G., Pouliquen, J. F., De The, G., Moreau, J. P. & Gessain, A. 1999. Mother-to-child transmission of human T-cell-leukemia/lymphoma virus type I: implication of high antiviral antibody titer and high proviral load in carrier mothers. *Int J Cancer,* 82**,** 832-6.

Ureta-Vidal, A., Pique, C., Garcia, Z., Dehee, A., Tortevoye, P., Desire, N., Gessain, A., Chancerel, B., Gout, O., Lemonnier, F. A. & Cochet, M. 2001. Human T cell leukemia virus Type I (HTLV-I) infection induces greater expansions of CD8 T lymphocytes in persons with HTLV-I-associated myelopathy/tropical spastic paraparesis than in asymptomatic carriers. *J Infect Dis,* 183**,** 857-64.

Utz, U., Banks, D., Jacobson, S. & Biddison, W. E. 1996. Analysis of the T-cell receptor repertoire of human T-cell leukemia virus type 1 (HTLV-1) Tax-specific CD8+ cytotoxic T lymphocytes from patients with HTLV-1-associated disease: evidence for oligoclonal expansion. *J Virol,* 70**,** 843-51.

Valeri, V. W., Hryniewicz, A., Andresen, V., Jones, K., Fenizia, C., Bialuk, I., Chung, H. K., Fukumoto, R., Parks, R. W., Ferrari, M. G., Nicot, C., Cecchinato, V., Ruscetti, F. & Franchini, G. 2010. Requirement of the human T-cell leukemia virus p12 and p30 products for infectivity of human dendritic cells and macaques but not rabbits. *Blood,* 116**,** 3809-17.

Van Duyne, R., Guendel, I., Narayanan, A., Gregg, E., Shafagati, N., Tyagi, M., Easley, R., Klase, Z., Nekhai, S., Kehn-Hall, K. & Kashanchi, F. 2011. Varying modulation of HIV-1 LTR activity by Baf complexes. *J Mol Biol,* 411**,** 581-96.

Van Duyne, R., Pedati, C., Guendel, I., Carpio, L., Kehn-Hall, K., Saifuddin, M. & Kashanchi, F. 2009. The utilization of humanized mouse models for the study of human retroviral infections. *Retrovirology,* 6**,** 76.

Van Maele, B., Busschots, K., Vandekerckhove, L., Christ, F. & Debyser, Z. 2006. Cellular co-factors of HIV-1 integration. *Trends Biochem Sci,* 31**,** 98-105.

Villaudy, J., Wencker, M., Gadot, N., Gillet, N. A., Scoazec, J. Y., Gazzolo, L., Manz, M. G., Bangham, C. R. & Dodon, M. D. 2011. HTLV-1 propels thymic human T cell development in "human immune system" Rag2(-)/(-) gamma c(-)/(-) mice. *PLoS Pathog,* 7**,** e1002231.

Vine, A. M., Witkover, A. D., Lloyd, A. L., Jeffery, K. J., Siddiqui, A., Marshall, S. E., Bunce, M., Eiraku, N., Izumo, S., Usuku, K., Osame, M. & Bangham, C. R. 2002. Polygenic control of human T lymphotropic virus type I (HTLV-I) provirus load and the risk of HTLV-I-associated myelopathy/tropical spastic paraparesis. *J Infect Dis,* 186**,** 932-9.

Von Schwedler, U., Kornbluth, R. S. & Trono, D. 1994. The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of

infection in macrophages and quiescent T lymphocytes. *Proc Natl Acad Sci U S A,* 91**,** 6992-6.

Wang, G. P., Berry, C. C., Malani, N., Leboulch, P., Fischer, A., Hacein-Bey-Abina, S., Cavazzana-Calvo, M. & Bushman, F. D. 2010. Dynamics of gene-modified progenitor cells analyzed by tracking retroviral integration sites in a human SCID-X1 gene therapy trial. *Blood,* 115**,** 4356-66.

Wang, G. P., Garrigue, A., Ciuffi, A., Ronen, K., Leipzig, J., Berry, C., Lagresle-Peyrou, C., Benjelloun, F., Hacein-Bey-Abina, S., Fischer, A., Cavazzana-Calvo, M. & Bushman, F. D. 2008. DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. *Nucleic Acids Res,* 36**,** e49.

Wang, H., Jurado, K. A., Wu, X., Shun, M. C., Li, X., Ferris, A. L., Smith, S. J., Patel, P. A., Fuchs, J. R., Cherepanov, P., Kvaratskhelia, M., Hughes, S. H. & Engelman, A. 2012. HRP2 determines the efficiency and specificity of HIV-1 integration in LEDGF/p75 knockout cells but does not contribute to the antiviral activity of a potent LEDGF/p75-binding site integrase inhibitor. *Nucleic Acids Res,* 40**,** 11518-30.

Wang, Z., Zang, C., Cui, K., Schones, D. E., Barski, A., Peng, W. & Zhao, K. 2009. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell,* 138**,** 1019-31.

Watanabe, T. 2011. Current status of HTLV-1 infection. *Int J Hematol,* 94**,** 430-4.

Wattel, E., Vartanian, J. P., 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**,** 2863-8.

Who 1989. Report from the scientific group on HTLV-1 infection and its associated diseases, convened by the regional office for the Western Pacific of the World Health Organisation in Kagoshima, Japan, 10–15 December 1988. *Wkly Epidemiol Rec 49.*

Willemze, R., Jaffe, E. S., Burg, G., Cerroni, L., Berti, E., Swerdlow, S. H., Ralfkiaer, E., Chimenti, S., Diaz-Perez, J. L., Duncan, L. M., Grange, F., Harris, N. L., Kempf, W., Kerl, H., Kurrer, M., Knobler, R., Pimpinelli, N., Sander, C., Santucci, M., Sterry, W., Vermeer, M. H., Wechsler, J., Whittaker, S. & Meijer, C. J. 2005. WHO-EORTC classification for cutaneous lymphomas. *Blood,* 105**,** 3768-85.

Williams-Carrier, R., Stiffler, N., Belcher, S., Kroeger, T., Stern, D. B., Monde, R. A., Coalter, R. & Barkan, A. 2010. Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize. *Plant J,* 63**,** 167-77.

Wolfe, N. D., Heneine, W., Carr, J. K., Garcia, A. D., Shanmugam, V., Tamoufe, U., Torimiro, J. N., Prosser, A. T., Lebreton, M., Mpoudi-Ngole, E., Mccutchan, F. E., Birx, D. L., Folks, T. M., Burke, D. S. & Switzer, W. M. 2005. Emergence of unique primate T-lymphotropic viruses among central African bushmeat hunters. *Proc Natl Acad Sci U S A,* 102**,** 7994-9.

Wu, K., Bottazzi, M. E., De La Fuente, C., Deng, L., Gitlin, S. D., Maddukuri, A., Dadgar, S., Li, H., Vertes, A., Pumfery, A. & Kashanchi, F. 2004. Protein profile of tax-associated complexes. *J Biol Chem,* 279**,** 495-508.

Wu, X., Li, Y., Crise, B. & Burgess, S. M. 2003. Transcription start regions in the human genome are favored targets for MLV integration. *Science,* 300**,** 1749-51.

Wu, X., Li, Y., Crise, B., Burgess, S. M. & Munroe, D. J. 2005. Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J Virol,* 79**,** 5211-4.

Yamada, S. & Ohnishi, S. 1986. Vesicular stomatitis virus binds and fuses with phospholipid domain in target cell membranes. *Biochemistry,* 25**,** 3703-8.

Yamaguchi, K. & Watanabe, T. 2002. Human T lymphotropic virus type-I and adult T-cell leukemia in Japan. *Int J Hematol,* 76 Suppl 2**,** 240-5.

Yamamoto, N., Chosa, T., Koyanagi, Y., Tochikura, T., Schneider, J. & Hinuma, Y. 1984. Binding of Adult T-Cell Leukemia-Virus to Various Hematopoietic-Cells. *Cancer Letters,* 21**,** 261-268.

Yamamoto, N., Matsumoto, T., Koyanagi, Y., Tanaka, Y. & Hinuma, Y. 1982a. Unique Cell-Lines Harboring Both Epstein-Barr Virus and Adult T-Cell Leukemia-Virus, Established from Leukemia Patients. *Nature,* 299**,** 367-369.

Yamamoto, N., Okada, M., Koyanagi, Y., Kannagi, M. & Hinuma, Y. 1982b. Transformation of human leukocytes by cocultivation with an adult T cell leukemia virus producer cell line. *Science,* 217**,** 737-9.

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 HTLV-1-associated myelopathy (HAM/TSP). *Blood,* 99**,** 88-94.

Yamaoka, S., Tobe, T. & Hatanaka, M. 1992. Tax protein of human T-cell leukemia virus type I is required for maintenance of the transformed phenotype. *Oncogene,* 7**,** 433-7.

Ye, J., Silverman, L., Lairmore, M. D. & Green, P. L. 2003a. HTLV-1 Rex is required for viral spread and persistence in vivo but is dispensable for cellular immortalization in vitro. *Blood,* 102**,** 3963-9.

Ye, J., Xie, L. & Green, P. L. 2003b. Tax and overlapping rex sequences do not confer the distinct transformation tropisms of human T-cell leukemia virus types 1 and 2. *J Virol,* 77**,** 7728-7735.

Yoshida, M., Seiki, M., Yamaguchi, K. & Takatsuki, K. 1984. 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,* 81**,** 2534-7.

Yoshikura, H., Nishida, J., Yoshida, M., Kitamura, Y., Takaku, F. & Ikeda, S. 1984. Isolation of HTLV derived from Japanese adult T-cell leukemia patients in human diploid fibroblast strain IMR90 and the biological characters of the infected cells. *Int J Cancer,* 33**,** 745-9.

Yoshioka, A., Hirose, G., Ueda, Y., Nishimura, Y. & Sakai, K. 1993. Neuropathological studies of the spinal cord in early stage HTLV-I-associated myelopathy (HAM). *J Neurol Neurosurg Psychiatry,* 56**,** 1004-7.

Zane, L., Sibon, D., Legras, C., Lachuer, J., Wierinckx, A., Mehlen, P., Delfau-Larue, M. H., Gessain, A., Gout, O., Pinatel, C., Lancon, A., Mortreux, F. & Wattel, E. 2010. Clonal expansion of HTLV-1 positive CD8+ cells relies on cIAP-2 but not on c-FLIP expression. *Virology,* 407**,** 341-51.

Zoubak, S., Richardson, J. H., Rynditch, A., Hollsberg, P., Hafler, D. A., Boeri, E., Lever, A. M. & Bernardi, G. 1994. Regional specificity of HTLV-I proviral integration in the human genome. *Gene,* 143**,** 155-63.

Appendices

Appendix I: Oligonucleotides used in this work

PCR and sequencing primers:

Name	Sequence
Actin fw	TCACCCACACTGTGCCCATCTATGA
Actin rev	CATCGGAACCGCTCATTGCCGATAG
SK43	CGGATACCCAGTCTACGTGT
SK44	GAGCCGATAACGCGTCCATCG
5LTRFW	CTCGCATCTCTCCTTCACG
5LTRRV	CTGGTGGAAATCGTAACTGGA
LTRseq	GGTTGAGTCGCGTTCT
Bio3	CCTTTCATTCACGACTGACTGCCG
Bio3a[a]	CCTTTCGTTCACGTCTGACTGCCG
Bio3b	CTTTTCATTCACGACTGACTGCCG
Bio3c	CCTTTCATTCACGACTAACTGCCG
Bio3d	CCTTTCATTCACGACTGACTGCCA
Bio3e	CCTTTCATTCACAACTGACTGCCG
Bio4	TCATGATCAATGGGACGATCA
P5Bio5	AATGATACGGCGACCACCGAGATCTACACTGGCTCGGAGCCAGCG ACAGCCCAT
P5Bio5a	AATGATACGGCGACCACCGAGATCTACACTGGCTCGGAGCCAGCG ACAGCCCAC
P7	CAAGCAGAAGACGGCATACGA
P5	AATGATACGGCGACCACCGA
VUshort	GATCGGAAGAGCGAAAAAAAAAAAA
VUlong	TCATGATCAATGGGACGATCACAAGCAGAAGACGGCATACGAGAT **NNNNNN**CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
HTLVseq	CAGCCCATCCTATAGCACTCTCCAGGAGAGAAATTTAGT
SBS8	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
SBS8rev	AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG

[a] Bio3a-Bio3e, P5Bio5a are patient specific primer variations of Bio3, P5Bio5.

Mutiplexing Barcodes

See VUlong – the long arm of the VU linker. Underlined sequence is a 6 letter barcode distinguishing between each sample.

Barcode number	Barcode sequence (5’ to 3’)	Barcode number	Barcode sequence (5’ to 3’)
VU1	ATCACG	VU29	TTCCCA
VU2	CAGGAC	VU30	TATCCT
VU3	TGCACT	VU31	TGACAG
VU4	GCTCTA	VU32	TACGCC
VU5	CTAGAG	VU33	TGCCTC
VU6	ATCCGT	VU34	GCCACC
VU7	GGTACG	VU35	TTTGGC
VU8	AATTCC	VU36	TCACAT
VU9	CGTATA	VU37	TGTAAG
VU10	AATCCC	VU38	TTGACG
VU11	TGTGTG	VU39	TTACAC
VU12	CACACA	VU40	TCCACG
VU13	TACTAC	VU41	TTTTAA
VU14	AGACCT	VU42	TATGGG
VU15	GCCTGT	VU43	GAGCGT
VU16	GACTAG	VU44	TCAAAG
VU17	GATCTG	VU45	TTATTC
VU18	TAGCAG	VU46	TAGGAA
VU19	CTGAAC	VU47	TACGTA
VU20	GAACTC	VU48	TTTATA
VU21	CTATGC	VU49	TATCAG
VU22	GTCTGA	VU50	TTGCGC
VU23	CCAAGA	VU51	ATAGCG
VU24	GGACCA	VU52	TCTACA
VU25	CATCAT	VU53	TATCTA
VU26	GCACGA	VU54	TCGTAA
VU27	TTGAGT	VU55	TGGGCG
VU28	ATCACG	VU56	TCGCCA

Appendix II – abstract of publication associated with this thesis

Melamed, A., Laydon, D. J., Gillet, N. A., Tanaka, Y., Taylor, G. P. & Bangham, C. R. M. (2013) Genome-wide Determinants of Proviral Targeting, Clonal Abundance and Expression in Natural HTLV-1 Infection. PLoS Pathog 9(3): e1003271.

The regulation of proviral latency is a central problem in retrovirology. We postulate that the genomic integration site of human T lymphotropic virus type 1 (HTLV-1) determines the pattern of expression of the provirus, which in turn determines the abundance and pathogenic potential of infected T cell clones in vivo.

We recently developed a high-throughput method for the genome-wide amplification, identification and quantification of proviral integration sites. Here, we used this protocol to test two hypotheses. First, that binding sites for transcription factors and chromatin remodelling factors in the genome flanking the proviral integration site of HTLV-1 are associated with integration targeting, spontaneous proviral expression, and in vivo clonal abundance. Second, that the transcriptional orientation of the HTLV-1 provirus relative to that of the nearest host gene determines spontaneous proviral expression and in vivo clonal abundance.

Integration targeting was strongly associated with the presence of a binding site for specific host transcription factors, especially STAT1 and p53. The presence of the chromatin remodelling factors BRG1 and INI1 and certain host transcription factors either upstream or downstream of the provirus was associated respectively with silencing or spontaneous expression of the provirus. Cells expressing HTLV-1 Tax protein were significantly more frequent in clones of low abundance in vivo. We conclude that transcriptional interference and chromatin remodelling are critical determinants of proviral latency in natural HTLV-1 infection.