

QUANTIFICATION OF HTLV-1 EXPRESSION AT THE SINGLE-CELL LEVEL IN NATURALLY-INFECTED T-LYMPHOCYTES

Martin Rafael Billman

A thesis submitted for the degree of PhD

IMPERIAL COLLEGE LONDON
Department of Medicine
Division of Infectious Diseases

September 2017

DECLARATION OF ORIGINALITY

All experiments and analysis described in this thesis were personally carried out by the candidate, unless specifically stated otherwise. In such cases, the responsible person has been duly credited and, if appropriate, referenced.

COPYRIGHT DECLARATION

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

ACKNOWLEDGEMENTS

Over the course of this project, I ended up taking and analysing 2.94 terabytes of images. This works out to 404,046,923,761 voxels, each 64 x 64 x 300 nanometres. This in turn adds up to a volume of 496.5 microliters: the sum of four years' work can be fit into a small Eppendorf tube.

I would like to begin by thanking my supervisors, Professors Charles Bangham and David Rueda, for their guidance encouragement and, perhaps most importantly, their patience over the last four years.

An army of thanks are due to the members of the Bangham and Rueda labs, for providing excellent company and countless pieces of advice during my time here. I'd like to single out a few:

Jocelyn Turpin for help with the probe designs, as well providing me with literally kilos of delicious food. Adam Cawte and Hailey Gahlon for general microscopy advice, good laughs and not letting me eat lunch alone. Sheila Xie for leading me into the fishy world of fluorescence *in situ* hybridisation, please forgive the pun. Aileen Rowan for training me to use the Category 3 facility and *also* providing cheese from time to time. Anat Melamed for sequencing help and providing delicious edible reminders of the Jewish calendar. Michi Miura for his carrying on of my technique, and being so diligent at learning colloquial English that it makes me "over the moon".

A great deal of thanks is due to Chad Whilding and Dirk Dormann of the MRC LMS microscopy facility, who have most likely forgotten more about microscopy and image analysis than I will ever know. Florian Mueller created FISH-Quant, which provided the "automated" portion of my semi-automated analysis pipeline. He has saved years of my life and quite possibly my sanity too.

I owe a great debt to my family for putting up with the first scientist in the family, for encouraging me to explore my intellectual curiosity and having faith in me. Kira: minns du hur du hjälpte mig hitta mitt första praktikjobb som forskare på Karolinska? También es para vos Nona: no creo que ganaré el premio Nobel al final, pero me diste tanto apoyo.

I would also like to thank my fantastic group of friends, who have put up with an increasingly irritable recluse over the last year or so. In particular I am contractually obliged to acknowledge Andreas, who in spite of being half-Danish is far better at mathematics and algebra than I will ever be, and who provided me with the shell radii used in Figure 4-3.

Finally, there are not enough words to describe how much I owe my brilliant, kind, beautiful wife Amy, without whom none of this would have been possible. We made it!

ABSTRACT

The human leukaemia virus HTLV-1 expresses essential accessory genes that manipulate the expression, splicing and transport of viral mRNAs. Two of these genes, *tax* and *hbx*, also promote proliferation of the infected cell, and both genes are thought to contribute to oncogenesis in adult T-cell leukaemia/lymphoma. The regulation of HTLV-1 proviral latency is not understood. *tax*, on the proviral plus strand, is usually silent in freshly-isolated cells, whereas the minus-strand-encoded *hbx* gene is persistently expressed at a low level. However, the persistently activated host immune response to Tax indicates frequent expression of *tax* in vivo. By using single-molecule RNA-FISH, I was able to quantify the expression of HTLV-1 plus-strand and minus-strand transcripts at the single-cell level in naturally-infected T-cell clones isolated from peripheral blood.

While *hbx* is the lone transcript located on the HTLV-1 minus-strand, it is not possible to definitively separate *tax* transcripts from its neighbours on the plus-strand by means of single-molecule RNA-FISH. Detected plus-strand transcripts are referred to as *tax* transcripts throughout the thesis, but it is important to remember that the probe which targets the *tax* sequence will also detect the other plus-strand products, even if they are not thought to be expressed in the same volume as *tax*.

Measuring viral transcription at the single-cell level revealed strong within-clone and between-clone heterogeneity in the expression of *tax* and *hbx*. Both *tax* and *hbx* are transcribed in bursts; *tax* expression is enhanced in the absence of *hbx*, while *hbx* expression increased in cells with high *tax* expression. Unexpectedly, *hbx* expression was found to be strongly associated with the G₂/M phase of the cell cycle, independent of *tax* expression. There is also a link between a cell's level of *hbx* expression and its volume. These results offer an explanation for the paradoxical correlations observed between the host immune response and HTLV-1 transcription.

TABLE OF CONTENTS

DECLARATION OF ORIGINALITY	3
COPYRIGHT DECLARATION.....	4
ACKNOWLEDGEMENTS.....	5
ABSTRACT.....	6
TABLE OF CONTENTS.....	7
LIST OF FIGURES.....	11
LIST OF TABLES.....	13
LIST OF ABBREVIATIONS	14
1 INTRODUCTION.....	19
1.1 HTLV-1.....	20
1.1.1 Epidemiology.....	21
1.1.2 Clinical Significance	23
1.1.3 Infection & Integration.....	24
1.1.4 Infectious vs. mitotic spread	26
1.1.5 Genome Structure.....	27
1.1.6 Regulatory Proteins.....	29
1.1.7 Accessory Proteins	34
1.1.8 Viral Expression and Persistence	35
1.1.9 Tax and HBZ.....	39
1.1.10 Immune determinants and anti-HTLV-1 response.....	41
1.2 Transcriptional bursting and cell-to-cell heterogeneity	43
1.3 Cell Cycle	45
1.4 Fluorescence Microscopy.....	48
1.5 Previous Work.....	49
1.6 Hypothesis & Aims	51
2 MATERIALS & METHODS.....	53
2.1 Cells.....	54
2.1.1 Isolation of PBMCs from infected patients.....	54
2.1.2 Isolation of T-lymphocyte clones from patient PBMCs	54
2.1.3 PCR to ascertain infection status	56
2.1.4 Cell culture	57
2.1.5 Actinomycin D treatment.....	57
2.1.6 Whole cell fixation	57

2.2	Staining and Imaging.....	57
2.2.1	RNA- fluorescence in situ hybridisation (FISH) probe design	57
2.2.2	RNA-FISH	58
2.2.3	Immunofluorescence	58
2.3	Image analysis.....	58
2.3.1	Fluorescence microscopy.....	58
2.3.2	Image processing	59
2.3.3	FISH-Quant analysis	59
2.4	Statistical analysis	61
2.5	DNA-FISH.....	62
2.5.1	Fixation for cryosectioning.....	62
2.5.2	Cryosectioning.....	62
2.5.3	DNA-FISH probe design.....	62
2.5.4	CryoFISH	63
2.5.5	3D DNA-FISH	63
2.5.6	Metaphase spreading and metaphase-FISH	64
2.5.7	Immuno-FISH	64
2.5.8	Imaging.....	65
3	QUANTIFICATION OF INDIVIDUAL VIRAL TRANSCRIPTS IN SINGLE CELLS	67
3.1	Chapter Summary	68
3.2	Introduction	69
3.3	Results.....	70
3.3.1	smFISH reveals stark difference in expression of tax and hbz.....	70
3.3.2	tax expression varies between clones	78
3.3.3	hbz expression is largely consistent between clones	80
3.3.4	hbz count is highly accurate, whereas tax count underestimates the content of high-expressing cells	82
3.4	Conclusions	86
4	VIRAL BURSTS AND mRNA DESTINATION	87
4.1	Chapter Summary	88
4.2	Introduction	89
4.3	Results.....	90
4.3.1	tax and hbz are expressed in transcriptional bursts	90
4.3.2	tax and hbz bursts are distinct from each other and consistent between clones.....	94
4.3.3	Transcriptional bursts have a non-uniform spatial distribution within the nucleus	97
4.3.4	tax RNA is exported to the cytosol, while hbz is retained in the nucleus.....	100

4.3.5	hbz RNA has a half-life of 4.4 hours	104
4.4	Conclusions	106
5	VIRAL EXPRESSION AND THE CELL CYCLE.....	107
5.1	Chapter Summary	108
5.2	Introduction	109
5.3	Results.....	110
5.3.1	tax and hbz expression are not independent of each other.....	110
5.3.2	Cells can be individually cell-cycle staged by microscopy.....	115
5.3.3	Cycling cells have elevated viral expression	118
5.3.4	High tax and hbz is associated with cell cycle progression.....	120
5.3.5	Cell volume is strongly associated with cell cycle stage	125
5.3.6	hbz expression and tax bursts correlate with nuclear volume	127
5.4	Conclusions	131
6	LOCATING THE HTLV-1 PROVIRUS USING DNA-FISH	133
6.1	Chapter Summary	134
6.2	Introduction	135
6.3	Results.....	137
6.3.1	Targeting integration sites by flanking region	137
6.3.2	Targeting the provirus.....	140
6.3.3	Immunofluorescence and Immuno-FISH	142
6.4	Conclusions	144
7	DISCUSSION & SUMMARY OF WORK.....	145
7.1	Chapter Summary	146
7.2	Paradoxes of HTLV-1 persistence.....	147
7.3	Transient tax expression	147
7.4	Consistent hbz expression	151
7.5	Intracellular distribution and half-life of hbz	153
7.6	Viral transcriptional bursts.....	156
7.7	Relationship between tax and hbz expression	159
7.8	Relationship between viral expression and the cell cycle	161
7.9	Intranuclear localisation of the provirus.....	164
7.10	Directions for future work	166
7.11	Final Conclusions.....	169
	REFERENCES	171
8	APPENDICES	193
8.1	Summary of permission for third party copyright works.....	194

8.2	Abstract of publication associated with this thesis	197
8.3	SmRNA-FISH probe libraries.....	198
8.4	Poisson distribution calculations	201
8.5	Analysis workflows.....	203
8.6	tax & hbx expression of each cell.....	205
8.7	Cell cycle data	207

LIST OF FIGURES

Figure 1-1 Geographic distribution of main foci of HTLV-1 infection	22
Figure 1-2 Genome, transcripts and RNA targeting of HTLV-1	28
Figure 1-3 The opposing functions of Tax and HBZ.....	40
Figure 1-4 Cdk/Cyclin complexes and cell cycle regulation	47
Figure 3-1 Representative image of cells stained by smRNA-FISH	71
Figure 3-2 smRNA-FISH reveals heterogeneity of tax and hbz expression	73
Figure 3-3 A minority of cells contribute the bulk of total tax, while hbz expression is more uniform	75
Figure 3-4 Cells with high tax expression are positive for Tax protein	77
Figure 3-5 tax expression in individual clones	79
Figure 3-6 hbz expression in individual clones.....	81
Figure 3-7 hbz counts are highly accurate, but tax is underestimated.....	85
Figure 4-1 tax and hbz are expressed in transcriptional bursts	93
Figure 4-2 The clone-specific characteristics of tax and hbz contribute to the overall differences between clones in the level of expression of the two genes.....	96
Figure 4-3 Burst intranuclear locations are non-random and clone-specific.....	99
Figure 4-4 tax is exported from the nucleus, whereas the majority of hbz is retained.....	102
Figure 4-5 Intracellular distribution of viral transcripts in individual clones	103
Figure 4-6 Decay curve of hbz RNA	105
Figure 5-1 Relationship between tax and hbz expression in individual cells.....	111
Figure 5-2 tax and hbz bursts do not occur independently of each other.....	114
Figure 5-3 Cell-cycle staging of single cells by fluorescence microscopy.....	117
Figure 5-4 Viral expression increases in cycling cells	119
Figure 5-5 Cell cycle distributions of cells according to their level of viral expression.....	122
Figure 5-6 Odds ratios from logistic regression of viral expression and cell cycle stage	124
Figure 5-7 Cell in S and G ₂ /M phase are significantly larger than those in G ₀ /G ₁	126
Figure 5-8 hbz expression per cell is largely consistent as a function of nuclear volume, and tax bursts are larger in cells with larger nuclei.....	129
Figure 6-1 DNA-FISH probes can specifically target the host genome flanking the provirus	139
Figure 6-2 DNA-FISH probes can specifically target the HTLV-1 provirus.....	141
Figure 6-3 Subnuclear compartments and the provirus can be visualised in the same individual cells	143
Figure 7-1 smRNA-FISH reveals cell-to-cell heterogeneity of gapdh expression	155

Figure 7-2 Model of regulation of HTLV-1 transcription and replication in vivo	163
Figure 7-3 smRNA-FISH for tax mRNA validates correlation between expression of Tax protein and Timer Protein	168
Figure 8-1 Workflow for identification of transcriptional bursts.....	203
Figure 8-2 Basis of calculation of transcriptional burst location.....	204
Figure 8-3 tax and hbz spot counts for each individual cell.....	205
Figure 8-4 Histograms of each clone replicate's cells' integrated nuclear intensity	207
Figure 8-5 Frequency of tax expression in cells split according to cell cycle stage in individual clones	Error! Bookmark not defined.
Figure 8-6 Frequency of hbz expression in cells split according to cell cycle stage in individual clones	Error! Bookmark not defined.

LIST OF TABLES

Table 2-1 Patient derived CD4+ T-cell clones studied in project	55
Table 2-2 Number of replicates and cells analysed from each clone in experiments	55
Table 2-3 PCR primer details.....	56
Table 5-1 tax-positive cells are less likely to be hbz-positive.....	113
Table 5-2 Frequency of hbz bursts according to cell-cycle stage and level of tax expression.....	121
Table 5-3 tax expression increases significantly with cell volume	130
Table 8-1 hbz-3'UTR probe oligonucleotide library	198
Table 8-2 tax probe oligonucleotide library	199
Table 8-3 gagpol probe oligonucleotide library	200
Table 8-4 Calculations comparing hbz expression profile of clones to Poisson distribution.....	201
Table 8-5 Expression metrics of each infected clone	206
Table 8-6 Odds ratios and p-values from logistic regression in Fig. 5-6	210
Table 8-7 Expression metrics of each clone in the context of cell cycle stage	211

LIST OF ABBREVIATIONS

6FAM	6-carboxyfluorescein
AC	Asymptomatic carrier
ActD	Actinomycin D
AOTF	Acousto-optic tuneable filter
AP-1	Activator protein 1
Apo	Apochromatic
APOBEC3G	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G
ATF	Activating transcription factor
ATLL	Adult T-cell leukaemia/lymphoma
BIM	Bcl-2-like protein 11
bp	Base pair
bZIP	Basic leucine zipper domain
cAMP	cyclic adenosine monophosphate
CCR4	CC chemokine receptor type 4
CD	Cluster of differentiation
Cdc	Cell division cycle protein
Cdk	Cycling-dependent kinase
Chk	Checkpoint kinase
CFV	Cell-free virion
CNS	Central nervous system
CRE	CREB response element
CREB	cAMP response element-binding protein
CRM-1	Chromosome region maintenance protein 1
CSF	Cerebrospinal fluid
CTL	Cytotoxic T-lymphocyte
DAPI	4',6-diamidino-2-phenylindole
DC	Dendritic cell
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA Pol II	DNA polymerase II
DRB	5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside
ER	Endoplasmic reticulum
FACS	Fluorescence-activated cell sorting
FAM	Fluorescein
FasL	Fas ligand

FISH	Fluorescence <i>in situ</i> hybridisation
FoxO	Forkhead box transcription factor
<i>g</i>	Gravitational force
G ₀ /G ₁	Gap 0 / gap 1 phase of cell cycle
G ₂ /M	Gap 2 / mitotic phase of cell cycle
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GLUT-1	Glucose Transporter 1
GTF	General transcription factor
h	Hour
HAM/TSP	HTLV-1-associated myelopathy / tropical spastic paraparesis
HAT	Histone acetyltransferase
HBZ	HTLV-1 bZIP factor
HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulphonic acid
HDAC	Histone deacetylase
HDI	Histone deacetylase inhibitor
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HSPG	Heparan sulphate proteoglycan
HTLV	Human T-lymphotropic virus
ICAM-1	Intercellular adhesion molecule 1
IEG	Immediate early gene
IFN	Interferon
IκBα	Inhibitor of κB-α
IKK	Inhibitor of κB kinase
IL	Interleukin
IL-2Rα	IL-2 receptor- α
ITIM	Immunoreceptor tyrosine-based inhibition motif
IU	International unit
kb	Kilobase pair
LEF1	Lymphoid-enhancer binding factor 1
LFA-1	Lymphocyte function-associated antigen 1
LM-PCR	Linker-mediated PCR
ln	natural logarithm
lncRNA	long non-coding RNA
LTR	Long terminal repeat
M	Molar concentration, in moles per litre
MCT	Maximum correlation threshold
MHC	Major histocompatibility complex

min	Minute
MTOC	Microtubule organising centre
mTOR	Mammalian target of rapamycin
NA	Numerical aperture
NAP1	Nucleosome assembly protein 1
NF- κ B	Nuclear factor kappa-light-chain-enhancer of activated B cells
NFAT	Nuclear factor of activated T-cells
NHP	Non-human primate
NOR	Nucleolar organising region
NP-1	Neuropilin-1
nt	Nucleotide
ORF	Open reading frame
PBMC	Peripheral blood mononuclear cell
PCAF	p300/CBP-associated factor
PCR	Polymerase chain reaction
PHA	Phytohaemagglutinin
PIC	Pre-initiation complex
PP2A	Protein phosphatase 2A
P-TEFb	positive transcription elongation factor
PTLV	Primate T-lymphotropic virus
PTM	Post-translational modification
PVL	Proviral load
qPCR	quantitative PCR, also known as reverse-transcriptase PCR (RT-PCR)
R	Repetitive region of HTLV-1 LTR
Rb	Retinoblastoma protein
RING	“Really interesting new gene” finger domain
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RNF4	RING finger protein 4
RT	Room temperature
RxRE	Rex-responsive element
σ	standard deviation
s	Second
S	Synthesis phase of cell cycle
smFISH	Single-molecule fluorescence <i>in situ</i> hybridisation
SOCS1	Suppressor of cytokine signalling 1
SRE	SRF response element
SRF	Serum response factor

SSC	Saline sodium citrate
STAT	Signal transducer and activator of transcription
STLV	Simian T-lymphotropic virus
SWI/SNF	Switching-defective-sucrose non-fermenting complex
TBP	TATA-box binding protein
TCF1	T-cell factor 1
TCR	T-cell receptor
TF	Transcription factor
TGF	Transforming growth factor
TIGIT	T-cell immunoreceptor with Ig and ITIM domains
TNF	Tumour necrosis factor
TRE	Tax response element
Treg	Regulatory T-cell
U3	Unique 3' region of HTLV-1 LTR
U5	Unique 5' region of HTLV-1 LTR
UIS	Unique integration site
UTR	Untranslated region
UV	Ultraviolet
VS	Virological synapse
WT	Wild-type

CHAPTER ONE

INTRODUCTION

1.1 HTLV-1

Human T-lymphotropic virus type-1 (HTLV-1) was the first retrovirus to be discovered in humans (Poiesz *et al.*, 1980; Yoshida, Miyoshi and Hinuma, 1982), and is the aetiological agent of a number of diseases including adult T-cell leukaemia/lymphoma (ATLL) and HTLV-1-associated myelopathy, formerly known as tropical spastic paraparesis (HAM/TSP). As a retrovirus, its lifecycle is based on infecting host T-lymphocytes and stably integrating its DNA into the host genome, where it can persist lifelong in the host. HTLV-1 is a member of the *Deltaretrovirus* genus, from the subfamily *Orthoretrovirinae*. Three other related viruses have been discovered so far: HTLV-2 (Kalyanaraman *et al.*, 1982), HTLV-3 (Calattini *et al.*, 2005) and HTLV-4 (Wolfe *et al.*, 2005), but there is no clear evidence that any of them cause disease in humans. HTLV-1 is considered a complex retrovirus by virtue of its regulatory genes, and is distantly related to the other disease-causing complex retroviruses: the *Lentivirus* human immunodeficiency viruses HIV-1 and HIV-2 (McClure *et al.*, 1988).

HTLV-1 is believed to have originated in non-human primates (NHPs), before crossing over to humans as a result of bushmeat hunting approximately 50,000 years ago, although there is evidence of independent, regular cross-over events in regions with regular interaction between humans and NHPs (Verdonck *et al.*, 2007; Filippone *et al.*, 2015; Kazanji *et al.*, 2015). In fact, the genomes of HTLV-1 and simian T-lymphotropic virus type-1 (STLV-1, also known as primate-TLV-1 or PTLV-1), the corresponding virus from non-human primates, are so similar that they can be considered to be a single virus which can infect and cause disease in multiple species (Vandamme *et al.*, 1994; Verdonck *et al.*, 2007; Enose-Akahata *et al.*, 2016).

1.1.1 Epidemiology

At least 10 million people are thought to be infected globally, although adequate data is lacking in several highly populated regions (de Thé and Bomford, 1993; Gessain and Cassar, 2012). Distribution is not uniform, but rather is concentrated in a number of endemic regions such as Brazil or southern Japan (Fig. 1-1), or within particular ethnic groups, where incidence can frequently be over 10% (Steinfort *et al.*, 2008; Gessain and Cassar, 2012). The virus is spread through contact with infected leukocyte-containing bodily fluids – primarily breastfeeding, sexual intercourse, transplant or transfusion of HTLV-1-infected tissue or blood, and intravenous drug use. Transfusion of infected blood carries the highest infection risk, up to 64% depending on how long the blood is stored prior to transfusion (Murphy, 2016), while the risk of infection through intercourse is relatively low – approximately 1% per year in a stable relationship according to one study (Martin, Taylor and Jacobson, 2014). Intrauterine or peripartum mother to child transmission is thought to be rare; vertical transmission is instead thought to occur through breastfeeding (Hino, 2011). HTLV-1-infected cells have been detected in breastmilk and a breastfeeding child is estimated to ingest leukocytes on the order of 10^8 per day (Percher *et al.*, 2016). Furthermore, HTLV-1 has been demonstrated to be capable of crossing epithelial barriers using a transcytosis mechanism (Martin-Latil *et al.*, 2012). The main determinants of transmission appear to be the infectious burden of the mother and the duration of breastfeeding: one study found a third of children of HTLV-1-positive mothers to be HTLV-1-positive themselves after a year of breastfeeding, measured by the presence of anti-HTLV-1 antibodies (Hino, 2011; Percher *et al.*, 2016).

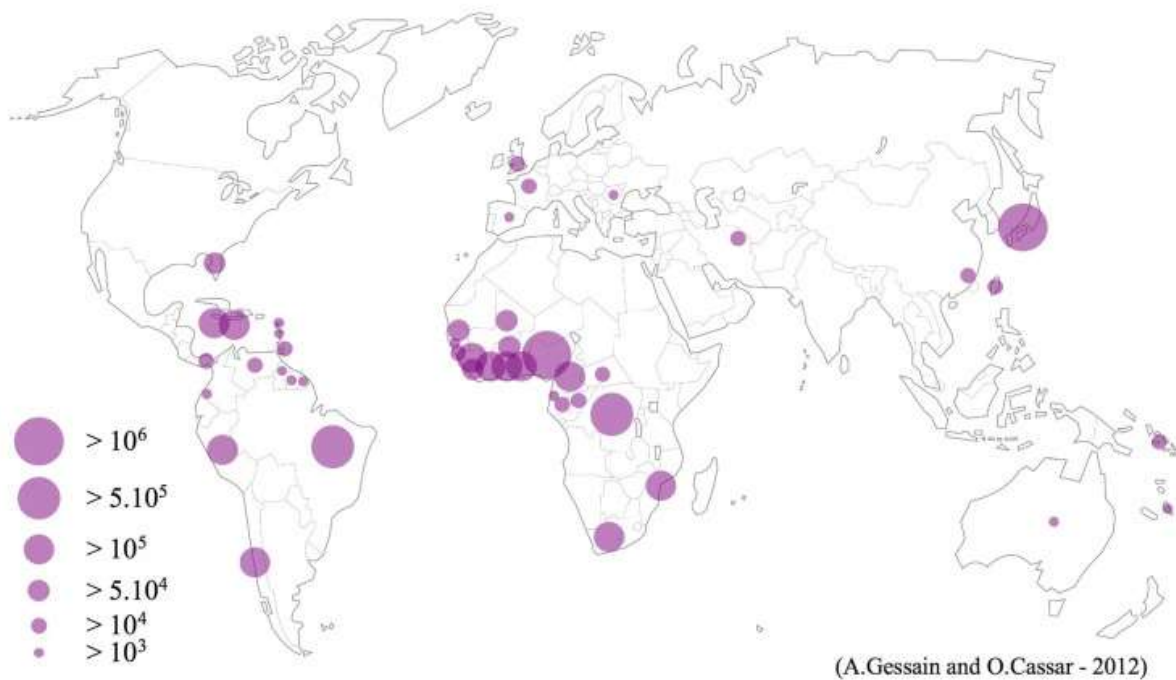


Figure 1-1 | Geographic distribution of main foci of HTLV-1 infection

Estimates based on population of from known endemic areas and epidemiological studies of pregnant woman, blood donors etc. from various adult populations. Significant portions of the world do not track HTLV-1 incidence. Reproduced, unaltered, from Gessain and Cassar, 2012, under the terms of the Creative Commons Attribution License.

1.1.2 Clinical Significance

Unlike HIV, which when untreated causes diseases in >99% of sufferers, only 5-9% of HTLV-1-positive individuals present with symptoms over the course of their infection (Bangham, 2011). The two most common HTLV-1 associated diseases are ATLL and HAM/TSP; there are also a number of less common inflammatory conditions or co-morbidities.

ATLL is characterised by uncontrolled, interleukin 2 (IL-2) independent proliferation of one or more malignant clones carrying an HTLV-1 provirus; transformed cells have an abnormal, multilobed nucleus, and are therefore referred to as “flower cells”. ATLL typically presents in approximately 5% of infected persons, decades after infection, with a median age at diagnosis of 60 in Japan (Matsuoka and Jeang, 2007). Roughly half of patients present with an acute form of the disease, while the remainder suffer from the chronic, smouldering or lymphomatous subtypes (Shimoyama, 1991). Prognoses vary by subtype, but are generally poor, with median survival after diagnosis ranging from 2 years for the chronic subtype to 6 months for the acute form (Bazarbachi *et al.*, 2010). There is no agreed treatment regimen, with potential therapies including chemotherapy, allogeneic haematopoietic stem-cell transplantation, and antiviral therapy (Kato and Akashi, 2015). While chemotherapy can improve prognosis in the chronic subtype, it seems to actually accelerate progression of the acute leukaemia, for which it has instead been found that antiviral therapy (interferon (IFN- α) and zidovudine) can be of benefit. There are also early promising signs of a monoclonal antibody, mogamulizumab, which targets CC-chemokine receptor type 4 (CCR4, higher levels associated with poorer prognosis), in treating the leukaemic subtypes of ATLL (Kato and Akashi, 2015).

Another 1-4% of HTLV-1-infected patients develop HAM/TSP (Martin, Taylor and Jacobson, 2014), an inflammatory condition characterised by elevated levels of inflammatory cytokines such as tumour necrosis factor- α (TNF α) and the formation of lesions in the central nervous system (CNS) (Bangham *et al.*, 2015). Typically, patients have a higher infectious burden in their cerebrospinal fluid (CSF) than in the peripheral blood, with a corresponding, colocalised increased frequency of anti-HTLV-1 CD8+ cytotoxic T-lymphocytes (CTLs).

HAM/TSP has a median age at onset of 46 years, and symptoms include neuralgia, incontinence and the progressive weakening and loss of lower limb function, to the point that patients require canes, wheelchairs or become bedridden. Roughly 15% of HAM patients will eventually succumb to complications directly linked to myelopathy, and they are at increased risk of ATLL if a high proportion, over 4%, of their peripheral blood mononuclear cells (PBMCs) are infected (proviral load, PVL). While

a range of treatment regimens exist, including methotrexate and antiviral treatment, none have been found to provide lasting improvement. As a result, care is primarily palliative (Bangham *et al.*, 2015).

The remainder of HTLV-1 patients who present symptoms have a range of HTLV-associated inflammatory diseases (HAIDs). These include uveitis, associated with opacities in the vitreous body of the eye and retinal vasculitis, and infective dermatitis, a severe recurring eczematous disorder (Goncalves *et al.*, 2010). Patients with acute ATLL suffer from opportunistic infections, as is common with suppressed immune systems, and HTLV-1-positive individuals are far more susceptible to infection by the helminth *Strongyloides stercoralis*, as well as being more vulnerable to associated morbidities once infected (Gotuzzo *et al.*, 1999). The reason behind this is unclear, but is believed to involve skewing of the anti-helminthic immune response by HTLV-1 infection (Hirata *et al.*, 2006).

In summary, while a relatively low proportion of those infected with HTLV-1 develop HTLV-1 associated diseases, progression to disease carries a very high burden of mortality and morbidity, and the search for a vaccine, effective treatment and improved diagnostic tests is ongoing (Willems *et al.*, 2017).

1.1.3 Infection & Integration

How infection is initially established and propagated remains an open question. Martin-Latil *et al.* demonstrated in an *in vitro* system that HTLV-1 cell-free virions (CFVs) can cross an epithelial barrier through transcytosis and subsequently infect human dendritic cells (DCs) (Martin-Latil *et al.*, 2012). While it is possible that DCs can be infected by and produce HTLV-1 CFVs, they require cell contact and what is known as a virological synapse (VS, discussed later in section) in order to infect T-cells, which constitute the overwhelming proportion of HTLV-1-infected cells. Furthermore, cell-free HTLV-1 virions are not detected in the blood of patients, and cell-free blood products such as platelets or plasma cannot transmit infection between individuals (Murphy, 2016). It has also been demonstrated that HTLV-1 can spread between cells through the formation of “viral biofilms”: viral particles enmeshed in a virally induced extracellular matrix constituents such as collagen and tetherin (Pais-Correia *et al.*, 2010). Another possibility is through cell-to cell conduits (Van Prooyen *et al.*, 2010).

Almost all circulating infected cells are T-lymphocytes, with approximately 90% being CD4+ and most of the remainder being CD8+ (Hanon *et al.*, 2000). The virus gains entry through three cell surface markers: heparan sulphate proteoglycans (HSPG) (Jones *et al.*, 2005), Neuropilin-1 (NP-1) (Ghez *et al.*, 2006) and Glucose Transporter 1 (GLUT-1) (Manel *et al.*, 2003). The initial interaction is thought to

involve attachment between the HTLV-1 envelope and HSPG, after which the complex is stabilised by NP-1. Finally, association with GLUT-1 allows fusion and viral entry (Jones *et al.*, 2011). Once inside, the RNA genome is reverse transcribed into DNA by the virus-encoded reverse transcriptase enzyme, and integrated into the human genome using the integrase enzyme (Bangham, 2011), with the host protein phosphatase 2A (PP2A) considered to be a likely co-factor (Maertens, 2016; McCallin *et al.* 2017; manuscript submitted for publication). Once the HTLV-1 genome is stably integrated, the process is irreversible, and the cell will remain infected with what is now termed a provirus until host cell death or detection and destruction by the immune response. *de novo* HTLV-1 integration sites display a clear preference for certain genomic landmarks and epigenetic markers over others, such as transcription factor (TF) binding sites, inside genes and within 1 kilobase (kb) of CpG islands and transcription start sites (Gillet *et al.*, 2011; Melamed *et al.*, 2013). There is evidence that the 3D architecture of the nucleus can influence integration site choice in HIV-1, and it is possible for this to be the case with HTLV-1 as well (Marini *et al.*, 2015).

HTLV-1 replication can proceed along two pathways: the mitotic, where the integrated virus drives host cell proliferation (see section 1.1.4, “Infectious vs. mitotic spread”), and infectious, where virions are assembled and spread to other cells. Unlike HIV, HTLV-1 does not lyse its host cell when expressed, and avoids spreading through CFVs. Instead, it relies on cell-to-cell spread through the use of a VS, which shares several features with the previously characterised T-cell immunological synapse (Igakura *et al.*, 2003). Viral expression leads to the upregulation of adhesive molecules such as intercellular adhesion molecule 1 (ICAM-1) (Igakura *et al.*, 2003; Nejmeddine *et al.*, 2005), which upon contact with the integrin lymphocyte function-associated antigen 1 (LFA-1) on an uninfected cell triggers polarisation of the infected cell microtubule network towards the contact area between cell membranes. HTLV-1 virions are then actively transferred across the VS to the uninfected “target” cell. This may be why anti-HTLV-1 CD4⁺ and CD8⁺ T-cells are more likely to be infected by HTLV-1 (Hanon *et al.*, 2000; Goon *et al.*, 2004) – presumably infected cells can form a VS and complete infection of anti-HTLV-1 target cells faster than the target cells can form an immunological synapse. Aside from potentially allowing HTLV-1 to avoid interference from the immune system in the form of complement or antibodies, the VS is far more efficient: cell-to-cell spread by HTLV-1 has been estimated to be more than efficient than infection with CFVs by a factor of 10,000 (Mazurov *et al.*, 2010).

1.1.4 Infectious vs. mitotic spread

As was mentioned in the previous section, HTLV-1 can follow two strategies to maximise its replication. The first is called the infectious route, where virions are produced in infected cells in order to spread to and be integrated in new cells. The current understanding of the mechanisms of cell-to-cell spread was described section 1.1.3. The other strategy is called the mitotic route, and involves no virions, but rather relies on the provirus-driven host cell proliferation. In so doing, the integrated viral DNA will also be replicated, in this case by the host cell's DNA polymerase II (DNA Pol II), and propagated to daughter cells (Bangham, 2011).

HTLV-1 presents remarkably low sequence diversity between its subtypes, or even within individuals, when compared to other retroviruses (Overbaugh, 2001). This, combined with the fact that DNA Pol II is estimated to be approximately five orders of magnitude more accurate when replicating DNA than the HTLV-1 reverse transcriptase, is strong evidence that the virus primarily makes use of the mitotic route, driving T-cell proliferation, to persist within its host.

Like other retroviruses, only a single provirus appears to be integrated in naturally infected T-cells (Cook *et al.*, 2012; Josefsson *et al.*, 2013), a phenomenon named superinfection resistance, the mechanisms of which remain obscure (Nethe, Berkhout and van der Kuyl, 2005). The result of this is that each infected cell can be characterised by its proviral “unique integration site” (UIS). All future daughter cells will share this UIS, and are referred to as clones. Consequently, infected cells in the blood of patients can be grouped into clones, and the UIS and relative abundance of each clone in a patient can be estimated through a technique known as linker-mediate polymerase chain reaction (LM-PCR). The proportion of cells which are infected can be accurately quantified by the number of copies of proviral DNA per 100 peripheral blood mononuclear cells (PBMCs), or proviral load (PVL), as measured by quantitative PCR (qPCR).

These clones are not equal in abundance – the individual genomic milieu of a clone's UIS plays a significant role in its relative proviral expression and abundance (Melamed *et al.*, 2013). The relative abundance of clones, along with a patients overall PVL, has been shown to be relatively stable over the course of decades (Cavrois *et al.*, 1998; Gillet *et al.*, 2011; Demontis, Hilburn and Taylor, 2012). Interestingly, clones of cells infected with HIV-1 with increased rates of proliferation have also been documented in HIV patients (Murray *et al.*, 2016), suggesting that mitotic proliferation is a unifying trait of retroviral infection.

The estimated number of clones within one person has recently been revised upwards by several orders of magnitude, with the total count now thought to be 10^4 - 10^5 clones per person (Gillet *et al.*, 2011; Laydon *et al.*, 2014). The latter study by Daniel Laydon also estimated that, in the chronic phase of infection, the mitotic route far outstrips the infectious route in terms of generating more infected cells: 10^7 infected daughter cells are estimated to be created by mitotic expansion for every newly infected cell. Nevertheless, this works out to approximately 100 new clones created by infection each day, and combined with the total number of existing infected clones it reminds us that while mitotic spread may dominate over the lifetime of the host, there must be some level of infectious spread, particularly in the early stages of infection.

1.1.5 Genome Structure

The structure of HTLV-1's 9 kb genome and its known transcripts are shown in Fig. 1-2. As with all exogenous replication-competent retroviruses, the HTLV-1 genome contains *gag*, *pol* and *env* genes, flanked by two long terminal repeats (LTRs). These encode, respectively, the structural polyprotein Gag, the viral reverse transcriptase (Pol), protease (Pro) and integrase (Int), and finally the receptor binding and fusion proteins in the polyprotein Env (Bangham, 2011). Additionally, the genome contains an extra region, called *pX*. This region consists of multiple overlapping open reading frames (ORFs), which by alternate splicing encode an array of different transcripts. Promoters responsible for plus- and minus-strand expression are located in the 5'-LTR and 3'-LTR respectively (Kwok *et al.*, 1996).

The LTRs are necessary for integration, and each is split into three regions: the unique 3' (U3), repetitive (R) and unique 5' (U5). The U3 region contains the Tax response element I (TRE-I), which is crucial to plus-strand viral expression from the 5'-LTR (Seiki *et al.*, 1983). It consists of three discontinuous 21-base pair (bp) repeats, each with the core conserved nucleotide sequence TGACG, which is homologous to the cyclic adenosine monophosphate (cAMP) response element binding factor (CREB) response elements (CREs) (Jeang *et al.*, 1988).

The plus-strand of the *pX* region encodes the regulatory proteins Tax and Rex (transcribed together in a single mRNA), along with the less well understood accessory proteins p30, p12 and p13 (Koralnik, Fullen and Franchini, 1993). Rex can also be expressed without its N-terminal domain as protein p21. The HTLV-1 genome also contains an ORF on its minus-strand, a first for human retroviruses (Gaudray *et al.*, 2002). This encodes the regulatory protein HTLV-1 bZIP factor, known as HBZ.

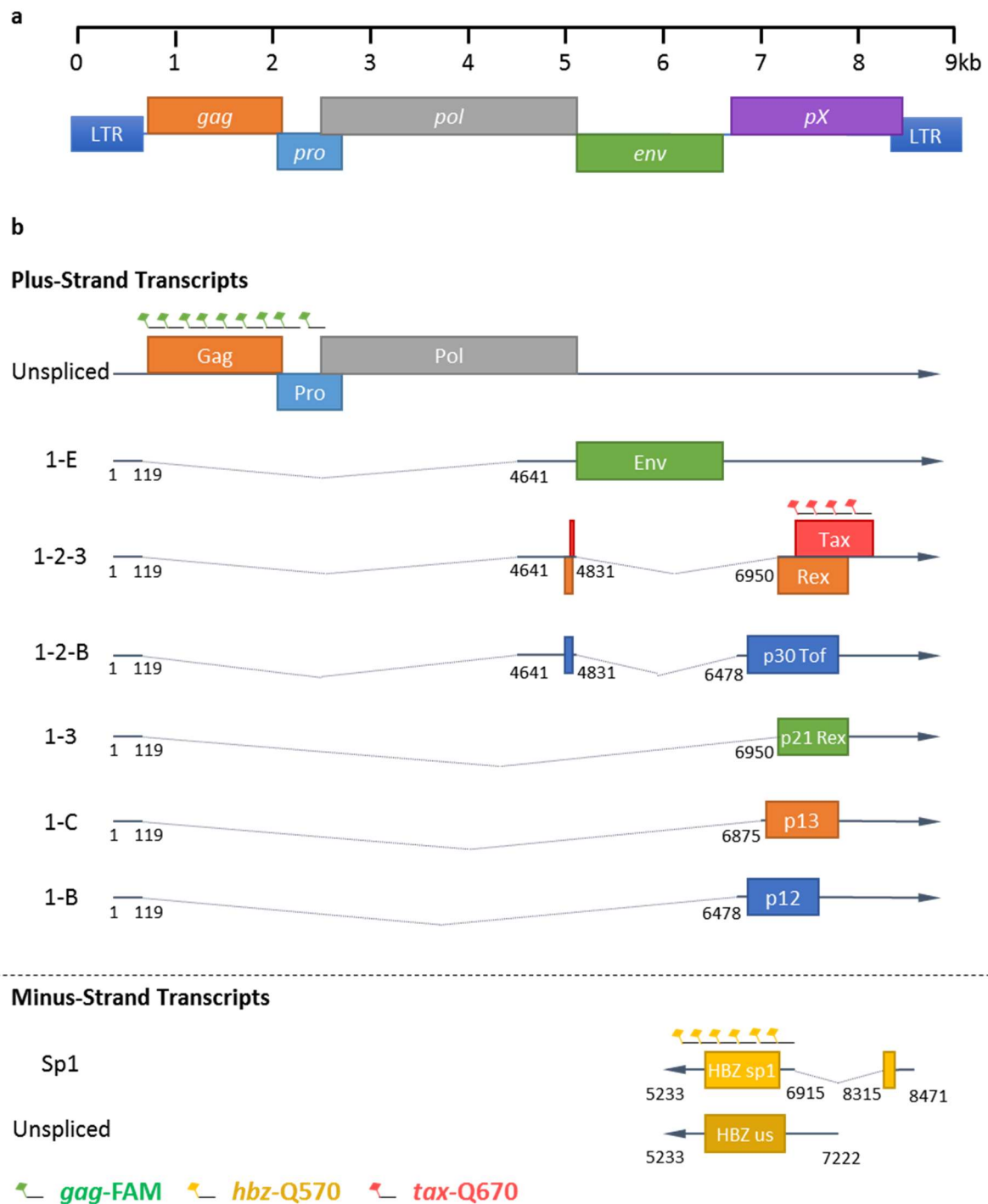


Figure 1-2 | Genome, transcripts and RNA targeting of HTLV-1

a, The HTLV-1 genome is approximately 9 kb in length. Along with *gal*, *pol* and *env*, the genome contains a collection of ORFs in the “*pX*” region which code for a variety of accessory products. **b**, the plus- and minus-strand transcripts produced from the HTLV-1 genome. Protein products are noted for each transcript. The transcripts targeted by RNA-FISH probe libraries are marked with colour coordinated symbols. As the probes are simply sequence specific, it is important to note that the *tax* probe does not differentiate between plus-strand transcripts while the *hbz* probe is insensitive to *hbz* splicing. The *gag* probes should only recognise unspliced plus-strand transcripts.

1.1.6 Regulatory Proteins

1.1.6.1 Rex

Rex is a functional analogue of Rev from HIV-1 and HIV-2 – it stabilises unspliced viral RNAs and controls the splicing of viral mRNAs, as well as the export of doubly spliced viral RNAs from the nucleus via the chromosome region maintenance 1 (CRM-1) pathway (Seiki *et al.*, 1988; Gröne, Koch and Grassmann, 1996; Nakano and Watanabe, 2012). This is achieved by interaction with a Rex-responsive element (RxRE), a sequence found between the U3 and R segments of the LTR, which forms a stem-loop in the untranslated regions (UTR) of plus-strand viral RNAs (Gröne *et al.*, 1994). Rex's control of nuclear export is thought to play a role in viral latency, and is responsible for a two-phase dynamic of viral expression *ex vivo*, where certain transcripts are exported (and therefore translated) before others (Rende *et al.*, 2011; Cavallari *et al.*, 2016).

1.1.6.2 Tax

Tax is translated from the same mRNA transcript as Rex, and is the most thoroughly characterised of the HTLV-1 proteins. It is a potent activator of viral expression, and has been implicated in the dysregulation of a vast array of host cellular transcriptional pathways in order to increase infectivity, drive proliferation and possibly transformation of the host cell (Curren *et al.*, 2012). Perhaps its most important function is as a transactivator, enhancing viral expression from the 5'-LTR through interactions with the Tax response element (TRE). This is discussed in depth in section 1.1.8 ("Viral Expression and Persistence").

Other than driving viral expression, Tax has been linked to an ever-increasing number of host pathways. The presence of Tax has been shown to upregulate expression of serum response factor (SRF) by binding to SRF response elements (SREs) in the promoter of the SRF gene, as well as leading to increased transcription of a number of SRF-controlled genes, including members of the activator protein 1 (AP-1) superfamily c-fos, c-Jun and JunD, which are involved in cellular proliferation and transformation. There may be some reciprocity as well, as the second TRE, TRE-II, also contains an SRE (Fujii *et al.*, 1994; Curren *et al.*, 2012).

Tax has also been demonstrated to hyperactivate both the canonical and non-canonical pathways of nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B), a family of TFs heavily involved in inflammation and apoptosis (canonical pathway) and lymphoid organogenesis (non-canonical), and

which are constitutively activated in many cancers and HTLV-1-infected cell lines. Various studies have shown that Tax interacts with subunits of the inhibitor of κ B (I κ B) kinase (IKK) complex and activates it, leading to the continuous phosphorylation, ubiquitination and proteasomal degradation of the NF- κ B inhibitor I κ B α , allowing TFs RelA and p50 to translocate to the nucleus and activate genes controlled by the NF- κ B response (Chu *et al.*, 1999; Harhaj and Sun, 1999; Carter *et al.*, 2001). Tax can also act upstream of IKK for the same effect, for example by recruiting TGF- β activating protein kinase 1 (TAK1) to activate the complex (Yu *et al.*, 2008). For the non-canonical pathway, cytosolic Tax directly induces the processing of the precursor peptide p100 to p52, allowing it and RelB to translocate to the nucleus and upregulate their target genes (Murakami *et al.*, 1995). Once in the nucleus, these TFs can also be recruited by Tax to nuclear bodies which drive heightened transcription (Bex *et al.*, 1997).

Along with directly acting upon members of the NF- κ B family, various studies have found interactions between Tax and components which are both up- and downstream in the signalling pathway. The presence of Tax leads to upregulation of the inflammatory cytokine IL-2 and its receptor, IL-2 receptor- α (IL-2R α), in a manner that involves NF- κ B and members of the nuclear factor of activated T-cells (NFAT) family (Ballard *et al.*, 1988; Good, Maggirwar and Sun, 1996). It has also been linked, with varying degrees of credibility, with upregulation of other cytokines, including IL-21 and its receptor, which play important roles in a number of T-cell proliferation pathways, and IL-13, whose effects include the suppression of the antiviral immune response (Silbermann, Schneider and Grassmann, 2008; Mizuguchi *et al.*, 2009)

Tax expression is also essential for infectious spread. It has been shown to upregulate adhesive molecules on the cell surface such as ICAM-1 (Tanaka *et al.*, 1995) and cellular adhesion molecule 1 (CADM1) (Manivannan *et al.*, 2016), increasing the odds of interacting with an uninfected cell for long enough for a VS to form and the virus to spread. Tax is also crucial to VS formation and infection as it drives reorganisation of the microtubule network towards the forming VS (Nejmeddine *et al.*, 2005).

Perhaps unsurprisingly, given the number of proliferative pathways Tax dysregulates, it has also been noted to interfere with the host cell's cell-cycle progression. Cells expressing Tax are less susceptible to a number of cell cycle checkpoints, for example entry into synthesis (S) phase, and progression from gap 2 (G₂) to the mitotic (M) phase. Tax can upregulate and interact directly with cyclin-dependent kinases (Cdk) 4 and 6, causing them to interact with Cyclin D, with the end result being that retinoblastoma protein (Rb) is phosphorylated, and E2F2, a TF involved in cell cycle progression from gap 1 (G₁) to S phase, is released and activated (Neuveut *et al.*, 1998; Schmitt *et al.*, 1998; Santiago *et al.*, 1999; Haller *et al.*, 2002). Tax can also directly interact with Rb and target it for degradation via the proteasome (Kehn *et al.*, 2005). Furthermore, studies have also shown a direct interaction

between Tax and the checkpoints kinases (Chk) 1 and 2, inactivating them and thus allowing uncontrolled interaction between cell division cycle protein 25 (Cdc25) and the Cyclin B/Cdk1 complex, which has the consequence of allowing accelerated progression to M-phase (Santiago *et al.*, 1999; Park *et al.*, 2004, 2006). This particular interaction has been shown to be counteracted by the HTLV-1 accessory protein p30 (see section 1.1.7, “Accessory proteins”) (Datta *et al.*, 2007). Finally, evidence from HTLV-1-infected cell lines has suggested that Tax activates the Akt/mammalian target of rapamycin (mTOR) signalling pathway, which is involved in a host of regulatory systems including cell cycle control. Inhibiting mTOR function in these cells with rapamycin induced cell cycle arrest (Ikezoe *et al.*, 2007).

In order to interact with these various pathways, Tax’s localisation and activity is controlled by post-translational modifications, or PTMs. These include SUMOylation and ubiquitination to localise Tax in the cytosol by the “really interesting new gene” (RING) finger protein RNF4 (Fryrear *et al.*, 2012), and allow its activation of the IKK complex. Deubiquitination has been shown to disrupt this interaction – the responsible deubiquitinase can be inactivated by Tax-directed phosphorylation, and is typically found in this phosphorylated state in HTLV-1-positive cell lines (Wu, Zhang and Sun, 2011). Tax can also be acetylated by the transcriptional co-activator p300, a modification linked to increased NF- κ B-dependent transcription (Curren *et al.*, 2012). While Tax is predominantly a nuclear protein, it contains both a nuclear-localisation signal (NLS) and nuclear export signal (NES) in its sequence to allow it to shuttle between cellular compartments (Smith and Greene, 1992; Alefantis *et al.*, 2003).

While Tax’s list of *in vitro* interactions is impressive, it is unclear which of them occur in naturally infected cells *in vivo*. For all its proliferative properties which could contribute to transformation and leukaemogenesis, many leukaemic clones have either silenced Tax through methylation of the 5’-LTR, or deleted it entirely (Takeda *et al.*, 2004; Miyazaki *et al.*, 2007). It is also interesting to note that, while expression of Tax can drive transformation of rat fibroblasts which can then form tumours in mice, and Tax expression can lead to IL-2 independent cell growth in a cell line (a hallmark of ATLL), Tax expression alone has not been found to be sufficient to transform CD4⁺ T-cells (Pozzatti, Vogel and Jay, 1990; Matsumoto *et al.*, 1997; Bellon *et al.*, 2010; Yoshita *et al.*, 2012). Therefore, Tax is unlikely to be the sole agent in leukaemic transformation during HTLV-1 infection.

1.1.6.3 HBZ

HBZ is the lone product of HTLV-1 minus-strand transcription (Gaudray *et al.*, 2002), and together with Tax it is considered the most important viral product in terms of viral persistence, pathogenesis and transformation. It promotes proliferation and prevents apoptosis through multiple avenues using both protein and RNA, and a significant portion of its effects also seem designed to counteract those of Tax (Ma, Yasunaga and Matsuoka, 2016).

The transcript occurs in two isoforms: spliced, in which case transcription begins in the 3'-LTR, and unspliced, when transcription begins just upstream (seen from the plus-strand) of the *tax/rex* ORF. Transcription of both isoforms finishes at a classical polyadenylation site, and polyadenylated *hbz* mRNA has been detected *in vitro* (Cavanagh *et al.*, 2006; Satou *et al.*, 2006). The resulting HBZ proteins are identical except for a small difference in their N-termini, but the spliced RNA and protein have uniformly been detected at higher levels in published studies (Satou *et al.*, 2006; Yoshida *et al.*, 2008; Rowan *et al.*, 2014). Only the spliced HBZ isoform has been found to have a proliferative effect on cells (Cavanagh *et al.*, 2006; Sugata *et al.*, 2015). It is worth noting, however, that in a recent transcriptomics study using 57 ATLL patients, minus-strand transcription was found to run past the polyadenylation site and even the 5'-LTR in most cases, up to a distance of 50kb downstream (Kataoka *et al.*, 2015).

A large number of the HBZ protein's interactions with the host cell seem to be in direct opposition to Tax – this is expanded upon in section 1.1.9 ("Tax and HBZ"). It is capable of binding CBP/p300 (Clerc *et al.*, 2008), and able to activate the transforming growth factor (TGF- β) / Smad. This in turn allows Foxp3 expression, inducing infected T-cells to become regulatory T-cells (Tregs), which can suppress the anti-HTLV-1 immune response and are overrepresented in ATLL leukaemic clones (Satou *et al.*, 2011; Zhao *et al.*, 2011). HBZ-transduced T-cells have also been shown to express increased levels of a co-inhibitory molecule called T-cell immunoglobulin and ITIM domain (TIGIT), with a concomitant increase in IL-10 secretion which can also suppress the immune response (Yasuma *et al.*, 2016). Additionally, there is limited evidence from cell lines that HBZ suppresses mTOR-regulated autophagy, a feature seen in other viral infections and cancers (Yoshita *et al.*, 2012).

HBZ has also been shown to dimerise with members of the AP-1 superfamily (which are also basic leucine zipper domain (bZIP) transcription factors) *in vitro*, including CREB, c-Jun, and JunD (Thébault *et al.*, 2004; Matsumoto *et al.*, 2005; Lemasson *et al.*, 2007). In most cases the effect is to repress expression, either through inhibition of DNA binding activity or sequestration in nuclear bodies (Clerc *et al.*, 2009; Ohshima *et al.*, 2010). It can, however, also lead to activation and overexpression, for

example with JunD, leading to the upregulation of proliferative genes (Thébaud *et al.*, 2004). HBZ has also been found to completely inhibit activation of the canonical NF- κ B pathway, as well as up- and downregulating the non-canonical and canonical Wnt pathways respectively, which seems to boost growth of ATLL cells (Zhao *et al.*, 2009; Ma *et al.*, 2013).

As with Tax, HBZ has a number of proliferative and anti-apoptotic effects. HBZ has also been demonstrated to repress the expression of apoptotic factors such as Bcl-2-like protein 11 (BIM) and Fas ligand (FasL) by binding and inactivating their upregulator, Forkhead box O3a (FoxO3a). Tax also downregulates this pathway by recruiting the upstream regulator Akt (Olagnier *et al.*, 2014; Tanaka-Nakanishi *et al.*, 2014). HBZ also seems to target the same Rb/E2F regulatory system to increase proliferation (Kawatsuki *et al.*, 2016). Inhibiting or knocking-out HBZ has repeatedly been shown to suppress growth of ATLL cell lines and reduce tumorigenesis in *in vivo* models (Satou *et al.*, 2006; Valeri *et al.*, 2010).

HBZ is primarily a nuclear protein (Hivin *et al.*, 2005), but intriguingly only a minority of *hbz* mRNA seems to be exported, even though its export is Rex-independent (Rende *et al.*, 2011; Cavallari *et al.*, 2016), suggesting that HBZ has regulatory effects at both protein and RNA levels. Indeed, *hbz* RNA has been found to have proliferative effects by upregulating the TF E2F1, thereby accelerating the G₁-S transition (Satou *et al.*, 2006). An excellent recent study by Mitobe *et al.* attempted to separate the effects of HBZ protein and RNA by comparing the effects of *hbz* in infected mouse CD4⁺ T-cells, according to whether the provirus contained wild-type (WT), translatable *hbz*, *hbz* lacking a start codon, or *hbz* with silent mutations designed to disrupt its secondary structure (Mitobe *et al.*, 2015). While both RNA and protein displayed the same proliferative effect, increasing the frequency of cells in S-phase, they diverged after that. Without RNA structure, expression of the HBZ protein seemed to in fact have an apoptotic effect on cells compared to the WT, while they were more resistant to apoptosis when only RNA was expressed. Overall, they found HBZ protein effects focused on TFs related to lymphocyte differentiation (Foxp3) and cytokine production (IL-2, IL-4, IL-10 etc.), while RNA-only expression was linked to downregulation of cell-cycle control genes (CDKN1A, RB1; the genes which encode protein Cdk1 and Rb, also affected by Tax) and upregulation of proliferative (E2F1, Cyclin A2, MCM5) and anti-apoptotic (survivin) genes. Unlike Tax, transgenic expression of HBZ alone can transform CD4⁺ T-cells, and HBZ-transgenic mice spontaneously suffer from inflammatory lesions and lymphomas (Satou *et al.*, 2011).

It is universally accepted that HBZ expression is vanishingly low *in vivo*. A monoclonal antibody has recently been developed which it is claimed can detect and visualise HBZ expressed at physiological levels, something previously impossible (Raval *et al.*, 2015; Baratella *et al.*, 2017). Their quantification

concludes that HBZ expression is around 50-fold lower in naturally infected PBMCs than in the HBZ-transduced cell lines which have historically been used to investigate HBZ expression and function, and found cytoplasmic HBZ to be an exclusive feature of the infected PBMCs of HAM/TSP patients. Another study by Shiohama *et al.* was only able to detect antibodies against HBZ in approximately 10-15% of HTLV-1-infected persons, and some of their results are in direct contrast those using the monoclonal antibody (Shiohama *et al.*, 2016). It is therefore worth stressing that conclusions about physiologically relevant protein-protein interactions by HBZ, and their consequences for the virus and cell, have to be made with caution.

1.1.7 Accessory Proteins

The p12 protein localises to the endoplasmic reticulum (ER) and seems to play a role in increasing viral infectivity, as well as downregulating the cell surface presentation of major histocompatibility complex class I (MHC I) molecules (Albrecht *et al.*, 2000; Johnson *et al.*, 2001; Kim *et al.*, 2006). It has also been found to upregulate a number of signalling pathways, including signal transducer and activator of transcription 5 (STAT5), with the putative effect of reducing the host cell's dependence on IL-2 for proliferation (Nicot, 2001). The ER-localisation signal on p12 can be cleaved to produce the protein p8, which localises to the cell surface and may play a role in infectivity through cellular conduits and increasing cell clustering by upregulating LFA-1. It has also shown the potential to induce T-cell anergy through abrogation of T-cell receptor (TCR) signalling (Van Prooyen *et al.*, 2010).

p13 and p30 share an ORF and a portion of their sequences, but have differing cellular locations and roles. p13 has been found in the inner mitochondrial membrane, with effects on mitochondrial morphology and membrane permeability observed at high concentrations, leading to apoptosis through release of cytochrome c (Silic-Benussi *et al.*, 2009; Biasiotto *et al.*, 2010). p30, on the other hand, seems to localise in the nucleolus, and play a role in regulating viral expression by binding tax/rex mRNA and sequestering it in the nucleus (Nicot *et al.*, 2004). At low concentrations, it may also have a role as a TF by interacting with CREB-responsive promoters (Zhang *et al.*, 2000). Finally, p30 may counteract some of Tax's effects regarding pushing cells through cell cycle checkpoints, potentially protecting the host cell from DNA damage-related apoptosis (Datta *et al.*, 2007; Baydoun *et al.*, 2010).

It is worth noting that these proteins are difficult to study because they are expressed at extremely low levels in physiological conditions (Rende *et al.*, 2011), and many of the published reports have

used overexpression systems, either *in vitro* or in animal models. Furthermore, given their overlapping ORFs, and their overlap with the *hbz* ORF, discerning the effects of individual transcripts through the use of knock-outs is a complicated exercise.

1.1.8 Viral Expression and Persistence

HTLV-1 expression begins in and is driven by promoters in the 5'- and 3'-LTRs. For the plus-strand, a complete, unspliced genomic RNA is transcribed, and the various transcripts are produced through splicing (Fig. 1-2). The factors and processes involved in one transcript being produced over another are unclear, but there is a marked difference in the relative expression of the various plus-strand transcripts produced in naturally infected cells, and they appear to be expressed in a particular order. This is partially a result of differential export and translation than transcription (Li *et al.*, 2009; Rende *et al.*, 2011). It has traditionally not been possible to detect HTLV-1 virions, proteins or mRNA in the blood of infected individuals, although it is in rare cases possible with *tax/rex* RNA, and is beginning to occur for *hbz* RNA and possibly protein with increasingly sensitive techniques (Richardson *et al.*, 1997; Kataoka *et al.*, 2015; Baratella *et al.*, 2017). In spite of this paucity of detectable viral markers, infected persons maintain a chronically activated anti-HTLV-1 CTL response. The short-lived nature of the activation state of these cells in the absence of antigenic stimulation indicates that, in spite of an almost latent virus from a detection standpoint, viral antigen is regularly being presented to the immune system (Daenke *et al.*, 1996; Bangham, 2011). Once infected PBMCs are cultured *in vitro*, viral expression quickly emerges – *tax/rex* is detected almost immediately and increases rapidly, followed by the other transcripts after a lag of 1-4 h (Hanon *et al.*, 2000; Rende *et al.*, 2011). *tax/rex* and the structural transcripts *gagpol* and *env* are expressed at the highest level, although this fluctuates over time in multiple systems, while *hbz* and the accessory genes are expressed at a far lower level. *hbz* is expressed at a rate which is lower than *tax/rex* by up to 10,000-fold, but stays relatively stable (Satou *et al.*, 2006; Li *et al.*, 2009).

It remains unclear precisely how viral expression from 5'-LTR and 3'-LTR is controlled *in vivo*, but it is worth noting that CD4+ T-cells, the principal viral reservoir, do not exclusively circulate in the peripheral blood. In fact, they spend a far larger proportion of their time in the lymphatic system, an environment which is far more hypoxic than venous blood. In support of this hypothesis, a recent study by Kulkarni *et al.* has provided evidence that hypoxia and glycolysis may play a role in reactivation of plus-strand expression from latency *in vivo* (Kulkarni *et al.*, 2017). Increased levels of

tax mRNA have also been detected in the bone marrow, another hypoxic compartment (Yasunaga *et al.*, 2016).

Tax has long been known to be able to transactivate transcription from the 5'-LTR promoter, and the fundamental understanding of its mechanisms of action has not changed in the last 10 years (Curren *et al.*, 2012; Giam and Semmes, 2016). Since all structural genes are encoded on the plus-strand, it is crucial to viral expression and replication (Curren *et al.*, 2012). 5'-LTR transcription is achieved through a series of interactions with the TRE. Tax associates with the TF CREB, forming a complex which binds to the 21bp-repeats in the TREs, a step which is essential for high-level transcription (Wagner and Green, 1993; Anderson and Dynan, 1994; Tie *et al.*, 1996). It then recruits a host of cellular co-activators including CBP/p300, which are involved in virtually every eukaryotic gene expression pathway (Kwok *et al.*, 1996; Iyer, Özdag and Caldas, 2004), as well as activating transcription factor (ATF), p300/CBP-associated factor (PCAF) and NFAT (Harrod *et al.*, 2000; Curren *et al.*, 2012). CBP and p300 are histone acetyltransferases (HATs), and chromatin remodelling around the promoter to an 'open' state through hyperacetylation has been shown to be stimulate transcription (Lu *et al.*, 2002). Along with other chromatin remodellers such as the "switching-defective-sucrose non-fermenting" (SWI/SNF) complex and nucleosome assembly protein 1 (NAP1), nucleosomes can be evicted by the Tax complex to make way for transcriptional machinery (Lemasson *et al.*, 2002; Sharma and Nyborg, 2008; Easley *et al.*, 2010). In support of this, histone deacetylases (HDACs) can inhibit plus-strand expression, while HDAC inhibitors (HDIs) such as valproate have been shown to boost expression from the 5'-LTR (Ego, Ariumi and Shimotohno, 2002; Lemasson *et al.*, 2004; Belrose *et al.*, 2011).

As well as influencing integration site choice, the epigenetic landscape around the provirus affects transcription. As could be expected from the need for "open" histones for high plus-strand transcription, association of activating epigenetic marks such as acetylation with the genome around UIS is associated with higher viral expression and more frequent Tax expression (Gillet *et al.*, 2011). An extra 10^{12} lymphocytes have been estimated to be created per year in HTLV-1 patients compared to healthy controls as a result of virus-driven proliferation, and the rate of cellular proliferation was found to correlate with frequency of viral expression *ex vivo* (Asquith *et al.*, 2007).

Anti-Tax CTLs, however, have proven to be highly efficient at killing Tax-positive cells *in vitro* and *ex vivo*, and this may explain both why Tax is undetectable *in vivo*, why clones more likely to express Tax tend to have lower abundance *in vivo*, and why over time the surviving clones are found to be increasingly integrated in silent stretches of DNA, favouring latency (Hanon *et al.*, 2000; Asquith *et al.*, 2005; Gillet *et al.*, 2011; Melamed *et al.*, 2013). Over time, plus-strand expression has also been found to be downregulated or silenced through methylation of the 5'-LTR, silencing mutations of the plus-

strand genes or even deletion of a portion of the proviral genome. This can either involve the loss of just the 5'-LTR, termed a type-I defective provirus, or a deletion up to and including the first *tax* exon, a type-II defective provirus. Either way, plus-strand transcription is totally abrogated in these circumstances, and one of these deletions is found in approximately half of the leukaemic clones isolated from ATLL patients (Takeda *et al.*, 2004; Miyazaki *et al.*, 2007; Kataoka *et al.*, 2015).

It is not clear how this high level of plus-strand expression ends, aside from detection and destruction of the infected cell by anti-Tax CTLs. HBZ protein has been shown to be capable of breaking the cycle of Tax self-upregulation of its own expression by competing for the CREB protein *in vitro*, which is discussed more fully in section 1.1.9 ("Tax and HBZ"). The TF c-Maf can achieve the same effect, also by competing for CREB binding (Hieshima *et al.*, 2011). HBZ protein can also block Rex-dependent export of unspliced RNAs from the nucleus *in vitro*, although this would not affect Tax expression (Philip *et al.*, 2014). Finally, as has been mentioned previously, it is also possible that the accessory protein p30 may play a part in controlling plus-strand expression through retention of *tax/rex* mRNA in the nucleus, although this effect was observed with level of p30 expression which may not occur *in vivo* (Baydoun, Bellon and Nicot, 2012).

The promoter in the 3'-LTR is responsible for minus-strand transcription. It is located in the R-UTS region of the LTR, is a TATA-less promoter, and contains three Sp1 binding elements along with the three CREs present in the 5'-LTR-promoter (Cavanagh *et al.*, 2006; Yoshida *et al.*, 2008). Transcription of the spliced isoform of *hbz* on the minus-strand is dependent on Sp1 binding elements in the 3'-LTR, and It has been reported that HBZ forms a complex with JunD and Sp1 to enhance 3'-LTR promoter activity (Gazon *et al.*, 2012). *hbz* expression can also be induced by Tax through interaction with a TRE in the 3'-LTR (Yoshida *et al.*, 2008).

Intriguingly, while plus-strand expression is often silenced *in vivo*, the 3'-LTR has never been found to be deleted or silenced through epigenetic repression, and all ATLL cell lines and HTLV-1-infected cell lines are positive for *hbz* RNA *in vitro*, through qPCR, and *in vivo* through RNA-Seq (Satou *et al.*, 2006; Kataoka *et al.*, 2015). This may be in part thanks to a functional, conserved binding site for CCCTC-binding factor (CTCF) which has recently been discovered to be located at the beginning of the *pX* region in the plus-strand (Satou *et al.*, 2016). Among its many roles in the regulation of gene expression, CTCF functions as an insulator, demarcating boundaries between different histone modifications (Phillips and Corces, 2009). This may explain the stark difference in histone modifications seen across the proviral genome, and could permit the maintenance of expression from the 3'-LTR. The *hbz* gene also typically displays drastically fewer silencing mutations than genes on the plus-strand (Kataoka *et al.*, 2015). This observation could be due to the necessity of *hbz* expression for

viral persistence: rabbits infected with an *hbx*-negative virus displayed a significantly reduced PVL, and when the same virus was used to infect macaques, reversion mutants with restored *hbx* expression were discovered over time (Valeri *et al.*, 2010). It is also possible that, as a complementary strand gene, *hbx* is protected from mutation by the cytidine deaminase enzyme, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G (APOBEC3G), during the initial infection of a cell by HTLV-1 (Fan *et al.*, 2010).

1.1.9 Tax and HBZ

The regulatory proteins Tax and HBZ are considered the most important in HTLV-1 pathogenesis, and are fascinating for the depth of their interactions, many of which are mutually antagonistic (Fig. 1-3, Matsuoka and Yasunaga, 2013).

In terms of viral expression, they appear to act combinatorially to regulate viral transcription: Tax drives plus-strand expression, but subsequently HBZ (potentially upregulated by Tax) competes for CREB, disrupting the transcription complex. *hbz* RNA alone (mutated to lack a start codon) was not able to reduce plus-strand expression (Satou *et al.*, 2006). Intriguingly, certain factors also have opposing effects on their promoters. While the HDI valproate increases plus-strand transcription, expression of *hbz* falls (Belrose *et al.*, 2011). Similarly, the two Wnt TFs T-cell factor 1 (TCF1) and lymphoid-enhancer binding factor 1 (LEF1) directly interact with Tax to suppress plus-strand transcription, but also cause a modest increase in *hbz* expression by an unknown mechanism (Ma *et al.*, 2015).

Along with expression, many of HBZ's cell interactions seem specifically aimed at counteracting Tax (Fig. 1-3). Tax activates both of the NF- κ B pathways, while HBZ can entirely negate Tax's activation of the canonical pathway by binding and inhibiting RelA/p65, preventing cellular senescence through NF- κ B hyperactivation. Tax upregulates expression of the AP-1 superfamily, but HBZ either downregulates, sequesters or inactivates many of its members (Hivin *et al.*, 2007; Clerc *et al.*, 2009; Ohshima *et al.*, 2010). Tax inhibits TGF- β signalling but HBZ activates it (Zhao *et al.*, 2011), and Tax activates the canonical Wnt pathway, which is blocked by direct HBZ interaction with the TFs mentioned earlier, TCF1 and LEF1, and suppressing their DNA binding ability. Instead, HBZ upregulates the non-canonical Wnt ligand Wnt5a, which boosts ATLL cell growth (Ma *et al.*, 2013).

With all their counterbalanced interactions, it is not surprising that Tax and HBZ have been described as the "Yin and Yang" of the virus (Matsuoka and Yasunaga, 2013). Their seeming crosstalk and mutual antagonism may allow a finer modulation of the host T-cell, balancing the need for proliferation and cell-to-cell infection with the avoidance of senescence, apoptosis and detection followed by destruction by the anti-HTLV-1 immune response.

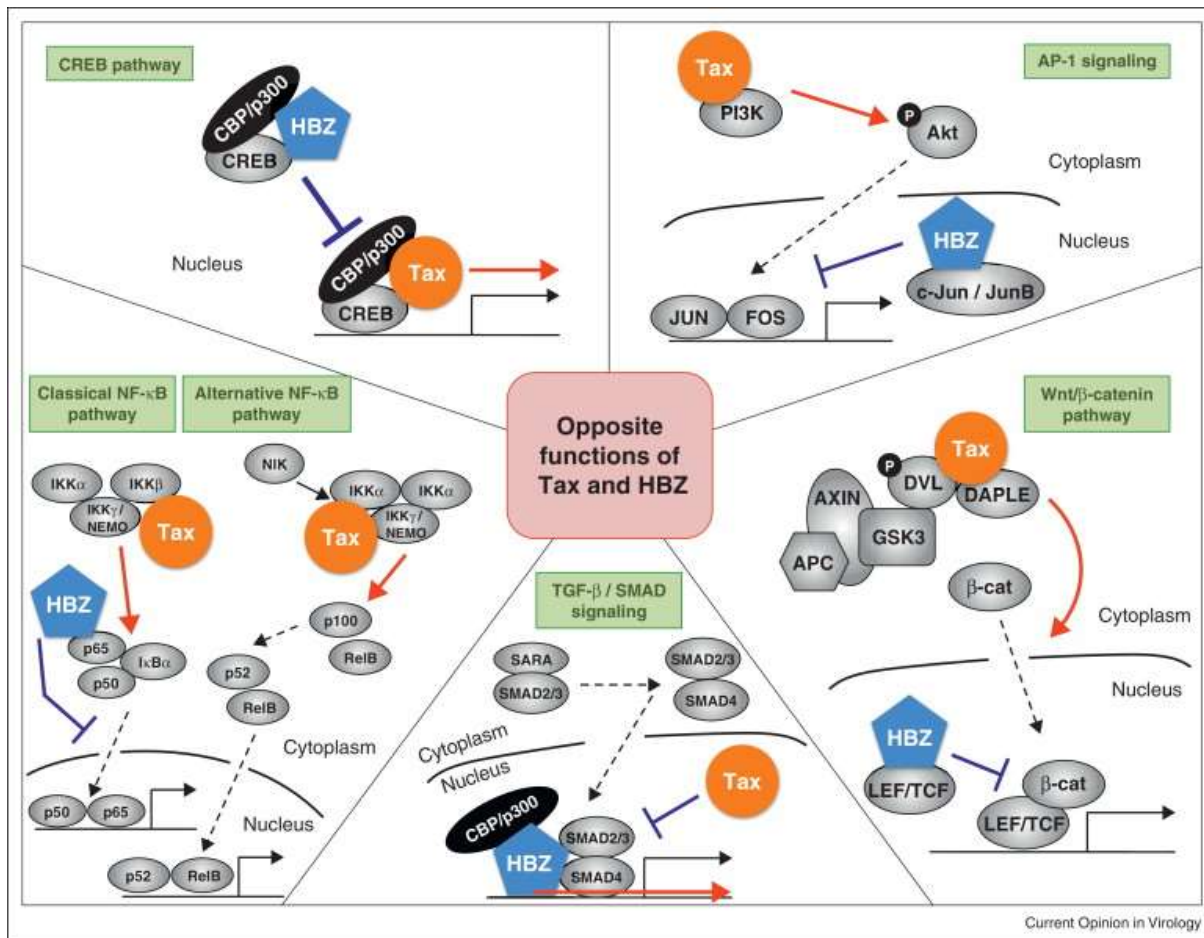


Figure 1-3 | The opposing functions of Tax and HBZ

Tax activates expression of CREB pathway genes through the recruitment of CREB and CBP/p300 to their transcription sites, but HBZ can block this interaction through direct competition for CREB binding. Tax activates the canonical Wnt pathway, but HBZ interacts with LEF and TCF further downstream in the same pathway to suppress target gene expression. HBZ increase TGF-β/SMAD signalling by recruitment of CBP/p300 and other TFs to target promoter sites, but Tax inhibits this pathway. Tax activates both the canonical and non-canonical NF-κB pathways, while HBZ can inhibit activation of the canonical pathway through interaction with RelA/p65. Tax activates the kinase Akt to upregulate AP-1 target gene expression, but HBZ can directly bind and inhibit c-Jun and JunB via their bZIP domains to downregulate the same genes.

Reproduced unaltered from "Current Opinion in Virology, Matsuoka & Yasunaga 2013. 'Human T-cell leukemia virus type 1: replication, proliferation and propagation by Tax and HTLV-1 bZIP factor'. 3(6): 684-691", with permission from Elsevier.

1.1.10 Immune determinants and anti-HTLV-1 response

Once HTLV-1 infection is established, it persists throughout the life of the patient (Cook *et al.*, 2013). This means that over the course of several decades, the virus must be able to evade a host immune response which is strong and constitutively active (Daenke *et al.*, 1996; Bangham, 2011).

The single most important predictor of HTLV-1 associated disease is a patient's PVL: a greater proportion of infected PBMCs correlates with an increased likelihood of disease and disease severity. The PVL stabilises shortly after infection and remains largely stable over decades (Gillet *et al.*, 2011; Demontis, Hilburn and Taylor, 2012). While it is largely consistent within an individual, however, it does vary by a considerably larger degree between individuals – from less than 0.001% of PBMCs being infected, to some cases of ATLL with over 100% PBMC infection (some clones are superinfected) (Demontis, Hilburn and Taylor, 2012; Cook *et al.*, 2014). For reasons which will be explored in this section, an infected individual's particular PVL is thought to be a result of a dynamic equilibrium between virus-driven proliferation and cell-to-cell spread, and infected-cell killing primarily by CD8+ CTLs (Bangham *et al.*, 2009). The host immune response clearly plays a role in the extended incubation period before onset of ATLL, as incidence of the disease increases significantly in individuals with suppressed immune systems (Hoshida *et al.*, 2001).

A PVL of more than ~1% is considered a significant risk-factor for developing HTLV-1 associated disease (Nagai *et al.*, 1998), the relative risk of developing ATLL increases above a PVL of 4%. While there is a clear overall increase in the average PVL of ATLL patients compared to HAM/TSP patients, who in turn have a higher average PVL than asymptomatic carriers (ACs), there is significant overlap between the groups (Gillet *et al.*, 2011; Demontis, Hilburn and Taylor, 2012). Particular class I human leukocyte antigen (HLA) genotypes have a significant role in determining HAM/TSP risk and a person's PVL "set point", at which point virus-driven proliferation and infection is at equilibrium with CTL-led destruction of infected cells (Jeffery *et al.*, 1999; Macnamara *et al.*, 2010). Only half of ACs and HAM/TSP patients have detectable anti-HBZ CD4+ or CD8+ T-cells, and the abundance of IL-2 secreting anti-HBZ CD8+ T-cells is inversely correlated with PVL (Hilburn *et al.*, 2011).

What seems to be the determining factor in the set-point of the dynamic equilibrium between HTLV-1 and the immune response is not the quantity of anti-HTLV-1 cells, but rather their quality. Tax is the immunodominant HTLV-1 antigen, while HBZ is a remarkably poor immunogen (Cook *et al.*, 2013; Sugata *et al.*, 2015). In fact, while anti-Tax CTLs are by far the most abundant anti-HTLV-1 T-cells, and they are far more efficient at killing infected cells than anti-HBZ CTLs, their abundance or functional avidity for Tax epitopes does not correlate with protection from HTLV-1 in the form of a lowered PVL

or risk of disease (Hanon *et al.*, 2000; Goon *et al.*, 2004; Rowan *et al.*, 2014). In contrast, the comparatively poor ability of anti-HBZ CTLs to kill infected cells has been linked to poor presentation of HBZ epitopes by MHC I, which is of course determined by a person's HLA I genotype. The protective genotypes have been found to bind HBZ epitopes more tightly, increasing CTLs functional avidity and lytic efficiency with a concomitant reduction in PVL and disease risk (Kattan *et al.*, 2009; Macnamara *et al.*, 2010; Rowan *et al.*, 2014). In support of this, unlike for Tax expression, the level of *hbz* expression detected *in vivo* correlates with their PVL and risk of deleterious symptoms (Andrade *et al.*, 2013). Interestingly, the study which found this also found *hbz* expression to stay roughly consistent as a factor of the PVL (Saito *et al.*, 2009).

The immunodominance but seeming superfluousness of Tax in the face of the immune response contrasts starkly with the much lower expressed, far less immunogenic but seeming necessity of HBZ. Expression of the minus-strand protein has also been shown to upregulate the FoxP3 TF, driving the differentiation of infected T-cells to FoxP3+ Tregs. These have an anti-inflammatory effect on the immune response, and their frequency (which is heightened in ATLL patients) is anticorrelated with CTL lytic efficiency (Bangham *et al.*, 2009; Satou *et al.*, 2011).

HTLV-1 patients often present a robust antibody response to a number of viral proteins, particularly Gag, Env and Tax. Approximately 50% of HTLV-1-positive individuals possess anti-Tax antibodies, and they can reach very high titres in HAM/TSP patients. Only 10% of ACs and HAM/TSP patients had detectable plasma anti-HBZ antibodies, a figure that rose to 15% in ATLL patients. Overall, antibodies do not seem to offer protection, whether through abundance or affinity (Enose-Akahata *et al.*, 2013; Shiohama *et al.*, 2016).

CADM1 presentation, typically not expressed in lymphocytes, is strongly correlated with infection (65% of the HTLV-1 load is typically found in CADM1+ lymphocytes), and makes cells more susceptible to lysis by CTLs (Manivannan *et al.*, 2016).

Type 1 IFN suppresses HTLV-1 expression and proliferation, which may be mitigated by HTLV-1 by upregulating suppressor of cytokine signalling 1 (SOCS1), or through competition between STAT2 and Tax for the CBP/p300 co-activators (Kinpara *et al.*, 2009, 2013). Another indicator of IFN's effect on HTLV-1 is its efficacy as a treatment for certain ATLL subtypes (Bazarbachi *et al.*, 2010).

In theory, one would expect HTLV-1 to be a good vaccine candidate. Its genome is relatively stable, reducing the probability of escape mutations, and patients readily develop antibodies against surface molecules such as Env (Mahieux, 2015). A very promising vaccine against the related bovine leukaemia virus (BLV) has also recently been developed (Gutiérrez *et al.*, 2014). So far, however, attempts have

fallen short, although a recent HBZ-based vaccine showed promising results in mice (Sugata *et al.*, 2015).

1.2 Transcriptional bursting and cell-to-cell heterogeneity

HTLV-1's latency and expression is necessarily a product of transcription. All cellular life is based on the central dogma that DNA is converted to RNA by transcription, before being translated into protein which carries out the cell's functions. Historically, genes were thought to be expressed constitutively, resulting in a Poisson-like distribution of gene products. In 1957, however, Novick and Weiner observed a high degree of variability in the level of beta-galactosidase production of individual cells, and that induction increased the proportion of cells expressing the enzyme as opposed to an equal increase in every cell's expression (Novick and Weiner, 1957). Over time, it has become increasingly clear that transcription can occur in episodic "bursts", characterised by periods of high mRNA creation interspersed with periods of transcriptional silence. In fact, this manner of expression has come to be seen as almost universal in higher eukaryotes, including humans, and plays a significant role in cells' ability to respond to stimuli (Raj *et al.*, 2006; Raj and van Oudenaarden, 2008; Dar *et al.*, 2012; Nicolas, Phillips and Naef, 2017).

Mammalian transcription is predominantly carried out by RNA Pol II, and is split into three overall processes termed initiation, elongation and termination. RNA Pol II requires a number of general transcription factors (GTFs) to begin transcription, and together they form the PIC (Conaway and Conaway, 1993; Kornberg, 2007; Murakami *et al.*, 2013). Initiation begins with the factor TFIID being recruited to, and binding, the promoter sequence of a gene. This factor includes Tata box-binding protein (TBP) and TAF1/2, which recognise the DNA initiation element and are essential for transcription and its regulation. TFIID then recruits a host of other factors, which interact with the DNA strand and the polymerase to begin initiation. These factors include TFIIB, TFIIE, TFIIF and TFIIH, which is responsible for melting the DNA double helix and allowing the polymerase to clear the promoter sequence by phosphorylating its Serine-5 residue. The polymerase subsequently often does what is called a "promoter-proximal pause", and is only released by the protein Cdk9 (the catalytic subunit of the positive transcription elongation factor, or P-TEFb) phosphorylating the polymerase's Serine-2 residue. Another key factor is a complex of proteins collectively termed Mediator, which recruits a number of the previously mentioned factors, increases the kinase activity of TFIIH, and increases the rate of initiation by retaining the GTFs after a polymerase is released for elongation. Genes which require rapid transcription when activated, such as the IEGs, have paused polymerases

occupying their promoters in this manner, waiting for release by Cdk9 (Senecal *et al.*, 2014). Both HIV-1's and HTLV-1's transactivators Tat and Tax can accelerate this step by directly binding Cdk9 and recruiting P-TEFb to the proviral 5'-LTR (Wei *et al.*, 1998; Zhou *et al.*, 2006; Tantale *et al.*, 2016).

As it is based on the biochemical reactions between polymerases, RNA and DNA, sometimes limited to one or two copies per cell, transcriptional bursting is best understood as stochastic or probabilistic, resulting in significant cell-to-cell heterogeneity. The same stimulus within an isogenic population of cells will not result in a uniform increase in expression, but rather increase the frequency of expression in the form of transcriptional bursts, or the number of transcripts produced from each burst. A particular gene product's absolute abundance within a given cell is determined by its degradation rate, the frequency of active transcription and the amplitude of that transcription (the "size" of the burst). This amplitude in turn is determined by the duration of the burst, and the rate of polymerase initiation and elongation during the burst. Dar *et al* found that both of these parameters varied across 8000 human genes whose expression was tracked *in vitro* (Dar *et al.*, 2012). Factors which contribute to this variation between genes include the characteristics of their promoters, the epigenetic landscape they are situated in and the availability of TFs (Senecal *et al.*, 2014). For example, genes in denser chromatin regions as measured by reduced DNase hypersensitivity tend to have more variation in the frequency of their bursts, and nucleosome occupancy is a deciding factor of a promoter's rate of activation (Dey *et al.*, 2015).

Interestingly, one study has demonstrated a strong link between cell volume and total RNA content for many genes, indicating that for many genes the important consideration is mRNA concentration as opposed to abundance (Padovan-Merhar *et al.*, 2015). Another used multivariate modelling of 183 parameters, including cell size, TF concentration and mitochondrial content, and demonstrated that a significant proportion of the seemingly stochastic heterogeneity between individual cells could in fact be explained by deterministic differences in those cells' stimuli and current condition. It also demonstrated that for many genes, stochastic noise in transcription is passively "filtered" by control of nuclear export, resulting in a stable rate of protein translation (Bahar Halpern *et al.*, 2015; Battich, Stoeger and Pelkmans, 2015; Stoeger, Battich and Pelkmans, 2016). The rate of protein transport back into the nucleus is also a significant contributor to heterogeneity (Sturrock, Li and Shahrezaei, 2017).

An increasing number of studies are investigating viral expression at the single-cell level, the majority of which involve sorting of individual cells by fluorescence-activated cell sorting (FACS) followed by qPCR (Ciuffi, Rato and Telenti, 2016).

A significant number of studies, including the aforementioned Dar *et al.*, have used the HIV-1 promoter as a tool to investigate cell-to-cell heterogeneity and the drivers of transcriptional burst

characteristics. Of particular interest given its viral nature, this promoter also exhibits an all-or-nothing expression pattern, with reactivation of viral expression in latent, isogenic cells appearing to be stochastic and probabilistic (Weinberger *et al.*, 2005; Weinberger and Weinberger, 2013; Murray *et al.*, 2016). Intriguingly, the promoter's expression profile varies depending on the chromatin state surrounding its integration site, and the length of pulses of Tat (the HIV transactivator protein) expression can be modulated by altering the abundance of Tat or its activity, in some cases leading to restored latency (Weinberger, Dar and Simpson, 2008; Skupsky *et al.*, 2010). A recent study by Tantale *et al.* used a similar approach in live cells to demonstrate that transcriptional bursts from the HIV-1 promoter consist of "convoys" of RNA polymerase II (RNA Pol II) processing along the template strand, with a few 100 nt spacing between them; the same was observed with the mammalian promoter for the large RNA Pol II subunit. (Tantale *et al.*, 2016). The activity of the HIV-1 promoter seemed to be driven by two stochastic events working on different timescales. Fluctuations on the scale of minutes seemed to be dependent on binding of the Mediator complex, which is essential for loading of the pre-initiation complex (PIC) and allowing reinitiation. The promoter also underwent longer periods, measured in hours, of either being permissive or refractive to any expression at all, which was linked to binding of the TBP to the promoter, and could be dependent on chromatin state.

1.3 Cell Cycle

The cell cycle is the series of precisely regulated biochemical and morphological steps which cells pass through in the process of dividing and creating daughter cells. Naturally, it is crucial to life and often becomes dysregulated in the context of infection and cancer. Somatic cells undergo mitosis, whose cycle in broad terms involves the phases G_0 , G_1 , S, G_2 and M, with checkpoints controlling the cell's progression from one phase to the next (Fig. 1-4) (Vermeulen, Van Bockstaele and Berneman, 2003). Cells can be resting or quiescent in G_0 , but must grow and increase protein and organelle biogenesis in the first active phase of the cycle, G_1 . After passing through the G_1 checkpoint, also known as the restriction point, the cell becomes committed to undergoing mitosis. It then enters S-phase, so named for the DNA synthesis which replicates the cell's genome: once complete, the cell will have double its normal complement of DNA. It then undergoes a second period of growth (G_2) in preparation for mitosis, expressing mitotic proteins and generally increasing in size so that the resulting daughter cells will have sufficient cytosol, organelles etc. There is then a final checkpoint, the G_2 -M DNA damage checkpoint. True to its name, the cell uses it to ensure that replication has finished, and its DNA is

intact, before allowing mitosis to begin, which if it went awry would conclude in genomic instability or cell death.

The cell cycle is primarily regulated by two sets of molecules: the Cdk family of serine/threonine kinases, and their binding partners, known as cyclins. More than 20 Cdk proteins have been discovered, and each can bind to a number of cyclins to achieve a variety of complex regulatory functions (Satyanarayana and Kaldis, 2009). There is also extensive redundancy among the Cdk and cyclin family members: individual or even multiple gene knock-outs have limited obvious effect on cell fitness. Some of the more important Cdk/Cyclin complexes are shown in Fig. 1-4. Cdk4 and Cdk6 phosphorylate proteins from the Rb family after being activated by Cyclin D. This releases the E2F TFs, resulting in upregulation of E2F-responsive genes (many of which are other cyclins) necessary for cell cycle progression. Cdk2 is activated by Cyclin E and completes Rb protein phosphorylation, further increasing E2F-controlled transcription. This allows passage through the G₁-S checkpoint and the beginning of DNA replication in S-phase. Cdk2 continues to play an important role in S-phase, complexed with Cyclin A which phosphorylates a number of proteins involved in DNA replication. Cdk1 is necessary for progression through the G₂-M checkpoint and the beginning of prophase, as well as having a number of important roles throughout mitosis (Vermeulen, Van Bockstaele and Berneman, 2003; Satyanarayana and Kaldis, 2009).

How the expression of genes which are not functionally related to the cell cycle changes over the course of the cycle remains an area of investigation. When cells are in G₂ they will have twice as many copies of genes, promoters and enhancers as they normally do. What's more, genes which are replicated early in S-phase will double in copy number before late-S genes. How the overall abundance of particular mRNAs is kept relatively constant during this time presents cells with a conundrum. Skinner *et al.* characterised the bursty expression of two genes critical to haematopoietic stem cell pluripotency, *Nanog* and *Oct4*. The promoters from each allele were found to act independently of each other, but their burst frequency fell in G₂/M, somehow "buffering" the abundance of their transcripts in the cell while the gene copy number was doubled (Skinner *et al.*, 2016). Another study has found that, though they act independently, allelic promoters' activity can be correlated through responding to the same stimulus (Kalo *et al.*, 2015). Yet another study reported that for many genes, transcript abundance as a factor of cell size remained constant throughout the cell cycle (Padovan-Merhar *et al.*, 2015).

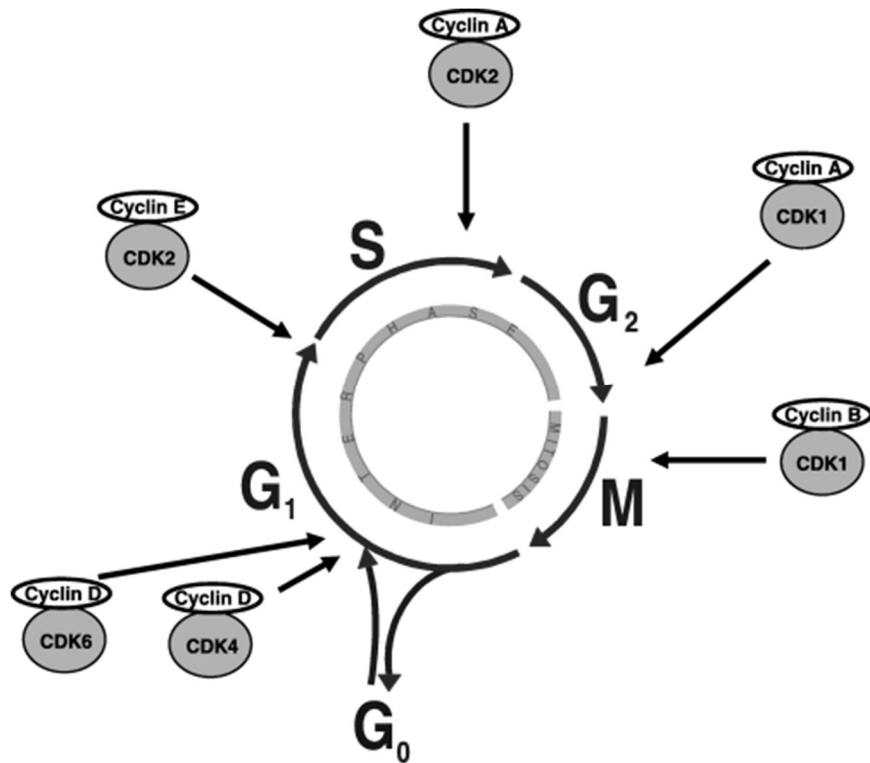


Figure 1-4 | Cdk/Cyclin complexes and cell cycle regulation

The stages at which Cdk/Cyclin complexes regulate cell cycle progression are indicated.

Reproduced, unaltered, from "Vermeulen, K., Van Bockstaele, D. R. and Berneman, Z. N. (2003), The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Proliferation*, 36: 131–149. doi:10.1046/j.1365-2184.2003.00266.x", with permission from John Wiley & Sons.

1.4 Fluorescence Microscopy

Fluorescence microscopy is based upon the excitation of fluorophores using a specific wavelength of light, and measuring the light which these excited fluorophores subsequently emit (Lichtman and Conchello, 2005). These fluorophores can be attached to biological molecules through a variety of processes, including genetically encoding the naturally occurring fluorescent proteins like green fluorescent protein (GFP). Organic chemists have developed thousands of fluorophores with specific excitation and emission spectra, allowing the simultaneous imaging of multiple fluorophores using specific wavelengths of visible light, which is relatively harmless to biological molecules.

The approach chosen in this project is called single-molecule RNA-fluorescence *in situ* hybridisation (smRNA-FISH). It uses up to 48 oligonucleotides, each labelled with a single fluorophore, to hybridise with their complementary sequence on a targeted mRNA transcript (Raj *et al.*, 2008). By targeting unique sections of the transcript sequence, it is possible to specifically label it such that individual transcript molecules can be reliably visualised against the background. This allows the transcripts to be counted, meaning that expression within individual cells can be quantified absolutely. More sophisticated analyses such as characterising transcriptional bursts through visualisation of nascent, unspliced RNA, and tracking mRNA location is also possible (Mueller *et al.*, 2013). The counting itself is based on the principle that each transcript will create a diffraction-limited spot. In the instance where two transcripts are closer to each other than the optical resolution of the microscope, their resulting signal will not be diffraction limited, as well as being brighter than spots from a single transcript. For conventional widefield microscopy, the optical resolution of a microscope objective is defined by the following equations, named after the German physicist Ernst Abbe:

$$\text{Abbe Resolution}_{xy} = \lambda / 2NA$$

$$\text{Abbe Resolution}_z = 2\lambda / NA^2$$

λ corresponds to the wavelength of light used, while NA is the numerical aperture of the objective, which in turn is defined by the sine of the aperture angle (θ) and refractive index (n) of the medium through which imaging occurs (glass, water etc.).

$$NA = n \times \sin \theta$$

The effects of these equations are that, in the widefield microscope used for this project, the resolution limit in the xy-axis for the emission of “green” fluorophores (~550 nm) would be approximately 200 nm, while for a far-red fluorophore (~650nm) the corresponding limit is 240 nm. A fluorescing sub-diffraction molecule would appear as spot of those dimensions, regardless of how much smaller than it truly is. The resolution limit in the z-axis would be approximately 600 nm and 710 nm respectively. In the recent past a number of techniques have been developed which allow the “bending” of this rule through sophisticated optics and computation, such as deconvolution and the fitting of gaussian profiles (Mueller *et al.*, 2013; Verdaasdonk *et al.*, 2014).

As fluorescence involves the detection of specific signals, autofluorescence is a concern that has to be addressed. In most cells, it tapers off rapidly as wavelength increases, being highest around 400-500 nm and significantly lower above 600 nm. Fixatives can also drastically affect autofluorescence; this is one of the reasons why the otherwise excellent fixative glutaraldehyde is unusable in a fluorescence microscopy context, with the weaker formaldehyde producing far less autofluorescence (Lichtman and Conchello, 2005).

1.5 Previous Work

HTLV-1 expression has been studied extensively, especially its two gene products *tax* and *hbx*. However, the majority of these studies have used animals models (Matsumoto *et al.*, 1997; Arnold *et al.*, 2006; Valeri *et al.*, 2010; Satou *et al.*, 2011), or induced expression of viral constructs in immortalised cell lines (Satou *et al.*, 2006; Curren *et al.*, 2012; Mitobe *et al.*, 2015). There is no guarantee that their expression patterns will correspond to the state of affairs *in vivo*, particularly in the case of *hbx* and the accessory proteins, which are expressed at far lower levels than in most *in vitro* systems.

Furthermore, many studies have made use of bulk, population-based techniques such as Western blotting and qPCR. As has become clear, there is no guarantee that the averaged measurements of all cells from these populations will correspond to the individual cells which make up that population, as there may be significant levels of cell-to-cell heterogeneity which are being masked. There are a number of increasingly powerful approaches which may be used to address this knowledge gap, such as single-cell RNA-Seq or high-resolution fluorescence microscopy. Some studies have used flow cytometry to measure protein expression in naturally infected cells (Hanon *et al.*, 2000). While this

may be appropriate for Tax, HBZ expression is not detectable using this method, and *hbx* RNA has been shown to have its own regulatory properties, independent of its protein.

A recent study used RNA-Seq to measure viral expression in ATLL patients (Kataoka *et al.*, 2015). Such global approaches allow a snapshot of the entire host cell's transcriptome, as opposed to smFISH's limitation of a handful of different gene products at a time. What imaging based approaches do offer, however, is both temporal and spatial resolution at the intracellular level, the ability to find interesting subpopulations of cells in a larger, mixed population, and the possibility of correlating expression with a number of other factors such as cell size or cell-cycle stage, which are lost upon cell lysis (Coleman *et al.*, 2015; Vera *et al.*, 2016). It also offers a higher detection rate of the target transcripts than RNA-Seq, in which detection rates typically range from 5% to 20% (Deng *et al.*, 2014; Grün, Kester and van Oudenaarden, 2014). This is especially important when targeting a gene whose expression is as low as *hbx*'s. Finally, smRNA-FISH is by its nature captures information at the single-cell level.

smRNA-FISH is based upon a library of labelled oligonucleotides binding a target RNA transcript, together providing a sufficiently strong fluorescent signal which can be reliably and specifically detected (Raj *et al.*, 2008). This is dependent on the target transcript's sequence being sufficiently unique in order that enough oligonucleotides can hybridise with it, without risking non-specific binding to stretches of DNA which are shared by other transcripts. As a retrovirus, HTLV-1's genome is drastically different from that of its host, making this risk negligible. HTLV-1's genomic stability also plays a crucial role, as the probes are designed using a sequenced genome which can be relied upon to be highly similar to the proviral sequences found in patients. This problem becomes significant in other retroviruses such as HIV, which displays an incredible sequence diversity even within a single patient (Overbaugh, 2001; Lusic *et al.*, 2013).

1.6 Hypothesis & Aims

My hypothesis is that, at the single-cell level, endogenous viral expression displays a level of heterogeneity that has been masked by the population-based approaches of previous studies. In order to address this question, the project had a number of aims:

1. Make use of smRNA-FISH to visualise and thereby quantify the transcripts *tax* and *hbx*, as proxies for expression of the viral plus- and minus-strands.
2. Develop an analysis pipeline to interpret the images created by this technique (described in chapter two)
3. Investigate and quantify the occurrence of viral transcriptional bursts
4. Investigate the intracellular distribution and half-life of viral mRNAs
5. Investigate the relationship between the expression of the plus- and minus-strand of the provirus, and the relationship between viral expression and the cell cycle

CHAPTER TWO

MATERIALS & METHODS

2.1 Cells

2.1.1 Isolation of PBMCs from infected patients

Blood samples were provided from patients attending the National Centre for Human Retrovirology (Imperial College Healthcare NHS Trust, St Mary's Hospital, London), with written informed consent in accordance with the Declaration of Helsinki. Blood samples were stored in the Communicable Diseases Research Tissue Bank, approved by the UK National Research Ethics Service (15/SC/0089).

PBMCs were separated from whole blood using density-gradient centrifugation with Histopaque-1077 (Sigma), before being stored long term in 10% dimethyl sulfoxide (DMSO, Sigma) / foetal calf serum (FCS, Invitrogen) in liquid nitrogen, or used immediately for isolation of individual T-cell clones.

2.1.2 Isolation of T-lymphocyte clones from patient PBMCs

Cell surface markers indicative of infection (CD4, CD25 (Toulza *et al.*, 2008), ICAM-1 (Tanaka *et al.*, 1995)) were used to isolate cells by magnetic activated cell sorting (e.g. CD4+CD25+ cell sorting kit, Miltenyi Biotec). These cells were then plated on 96-well plates using limiting dilution, at ~1 cell / well, in RPMI-1640 medium (Sigma) with 10% human AB serum (Invitrogen), 1 µg/ml phytohaemagglutinin (PHA, Sigma), 50 IU/ml IL-2 (International units, Promokine), 10 µM raltegravir (Selleck) and 0.5×10^6 /ml irradiated feeder cells (PBMCs, mixed from three uninfected donors). The PHA and feeder cells were replaced every two weeks, while IL-2 was replaced twice-a-week. Raltegravir was added whenever medium was replaced to maintain a constant concentration to inhibit secondary spread of HTLV-1. Clones which demonstrated sufficient growth after several weeks had DNA extracted using DNEasy kits (Qiagen), and were analysed by PCR for infection (see section 2.1.3), and LM-PCR for proviral integrity and integration site (Gillet *et al.*, 2011), as well as to ensure that each clone contained a single integration site.

The clones used in this project had been isolated previously by the Bangham group. Their integration sites and the patients they came from are summarised in tables 2-1. Clones A and D came from the same patient, as did clones C and F.

All clones were analysed in triplicate, with cells cultured separately for at least 24 hours prior to fixation.

Table 2-1 | Patient derived CD4+ T-cell clones studied in project

Clone	Lab Code	Patient Code	HTLV-1 UIS (hg19)	Notes
A	3.60	TBJ	Chr 4: 70,567,285	In intron of SULTB1, hepatocyte enzyme
B	TBX4B	TBX	Chr 22: 44,323,168	In intron of PNPLA3/SAMM50, component of mitochondrial outer membrane
C	11.63	TBW	Chr 19: 33,829,548	~30kb downstream of CEBPA and upstream of CEBPG, bZIP TFs involved in growth and proliferation of many cell types
D	3.83	TBJ	Chr 14: 46,204,950	~20kb upstream of RPL10L, ribosome L10-like protein Same patient as clone A
E	8.13	TCX	Chr 16: 53,601,059	Type-II defective provirus, ~30kb upstream of RPGRIP1L, protein found in the centrosome of ciliated cells
F	13.5	TBW	Uninfected	Same patient as clone C

Table 2-2 | Number of replicates and cells analysed from each clone in experiments

General Expression Profiling						
Clone	A	B	C	D	E	F
Replicates	3	3	3 (2 for cell cycle analysis)	3	3	3
Total Cells	3726	3519	2595 (1961)	5691	3946	2592
Actinomycin D Experiments						
Clone	A		B			
Replicates	2		2			
Total Cells (ActD)	3323		3797			
Total Cells (DMSO)	5001		7316			

2.1.3 PCR to ascertain infection status

Three separate PCR reactions were used to characterise clones isolated from patients. They targeted i) the *tax* gene, ii) the 3'-LTR and iii) the 5-LTR-*pX* segment of the provirus. Primers targeting *actin* were also used as a cellular positive control. All PCR products were separated using 1-2% agarose gels (Sigma) and visualised under ultraviolet (UV) light. The primer sequences and reaction times are summarised in table 2-3. All clones used in this project were characterised in this manner by myself, as well as being previously analysed by LM-PCR by Aviva Witkover and Anat Melamed.

Table 2-3 | PCR primer details

Target	Forward Primer	Reverse Primer	PCR Cycle	Notes
<i>tax</i> (SK43/44)	CGGATACCC AGTCTACGT GT	GAGCCGATA ACGCGTCCA TCG	98°C 1min 98°C 10s 64°C 10s } x35 72°C 10s 72°C 10min	Detect <i>tax</i> (159bp) Proof of infection Uses Phusion HF polymerase (NEB) Primers designed by Kwok <i>et al.</i> , 1988
3'-LTR	CCAAGGCCT ACCACCCCT CA	CCCGGGGG ATATTTGGG GCT	98°C 1min 98°C 10s 64°C 10s } x35 72°C 30s 72°C 10min	Detect 3'-LTR (167bp) Proof of infection even if provirus is defective Uses Phusion HF polymerase (NEB)
Provirus-A (5'-LTR- <i>pX</i>)	CTTAGAGCC TCCCAAGTGA AAAACATTT CC	GATGCATGG TCCTGCAAG GATAACA	95°C 2min 95°C 20s 66°C 2min 55s } x30 72°C 15min	Detect proviral deletions Intact provirus = 6.5kb Type-I defective (- <i>gag/env</i> = <6.5kb Type-II defective (-5'-LTR) = PCR fails Uses KOD Hot Start Polymerase (Novagen) Primers from Cook, 2013
Provirus-B (<i>pX</i> -3'-LTR)	CTCTCACAG TGGGCTCGA GA	CAAAGACGT AGAGTTGAG CAAGC	95°C 2min 95°C 10s 59°C 10s } x30 70°C 48s 72°C 10min	Internal proviral control (<i>pX</i> -3'-LTR, 2.85kb) Uses KOD Hot Start Polymerase (Novagen) Primers from Cook, 2013
<i>actin</i>	TCACCCACA CTGTGCCCA TCTATGA	CATCGGAAC CGCTCATTG CCGATAG	98°C 1min 98°C 10s 64°C 10s } x35 72°C 30s 72°C 10min	DNA loading control (295bp) Uses Phusion HF polymerase (NEB) Primers from Gillet <i>et al.</i> , 2013

2.1.4 Cell culture

Clones were cultured in RPMI-1640 medium with added 2 millimolar (mM) L-glutamine (Invitrogen), 100U penicillin and streptomycin (Invitrogen) and 10% AB human serum (Appleton Woods) at 37°C, 5% CO₂. 50 IU/ml IL-2 was added to the culture twice a week, and raltegravir was maintained at 10 µM throughout the culture period. Activation beads with antibodies against CD2, CD3 and CD28 (Miltenyi-Biotech) were replaced every two weeks. All experiments were carried out using cells harvested Day 8 of this cycle, after the addition of fresh medium, IL-2 etc. on Day 7.

2.1.5 Actinomycin D treatment

To quantify transcript half-lives and intracellular distribution, as well as to investigate the link between bursts and transcription, cells were treated with actinomycin D (ActD) to block transcription. ActD (Sigma) was dissolved in DMSO at 1.5 mg/ml before being added to RPMI medium to a final concentration of 1.5 µg/ml (final DMSO concentration 0.1% v/v). Clones A and B were tested, each in two replicates. DMSO alone was used as a vehicle control. After addition of ActD or DMSO, cells were fixed after 0, 2, 4, 6 and 22 hours in otherwise standard culture conditions (see section 2.1.4) before being stained, imaged and analysed using the standard protocol described in sections 2.2-2.4.

2.1.6 Whole cell fixation

Cells were added to poly-L-lysine (Sigma) coated, 12 mm, no. 1 glass coverslips (SLS), before being fixed in 2% formaldehyde (Life Technologies, in PBS) for 15 minutes. Approximately 1x10⁵ cells in 15µl PBS was added to each coverslip. They were then transferred to 70% ethanol for permeabilization or long-term storage at -20°C.

2.2 Staining and Imaging

2.2.1 RNA- fluorescence *in situ* hybridisation (FISH) probe design

RNA-FISH probes were designed using Stellaris's online probe design tool (version 4.1), with the reference genome AB513134 (Fan *et al.*, 2010) used as input for the *gag* probe (targets sequence 804-2954 of genome), *tax* probe (7315-8362, post-splice coding sequence) and *hbx* (7263-5233, minus-strand; post-splice coding sequence and 3'-UTR as reported in Cavanagh *et al.*, 2006). Each probe consists of a library of 18-20 nucleotide (nt) oligonucleotides, each with a fluorophore label (Raj *et al.*, 2008). Probes targeting *gag* were labelled with a fluorescein (FAM) or Quasar-570 fluorophore, while

those targeting *hbz* and *tax* were labelled with a Quasar-570 fluorophore and Quasar-670 fluorophore respectively. Precise sequences of all oligonucleotides used can be found in Appendix section 8.3.

2.2.2 RNA-FISH

Carried out in accordance with Stellaris protocols (Biosearch Technologies). In brief, cells were permeabilised in 70% ethanol before being washed and incubated overnight at 37°C with the desired RNA-FISH probes (Stellaris). They were subsequently washed with DAPI (Sigma) and mounted in Vectashield mounting medium (Vectashield) on glass slides (Vector) and sealed with clear nail varnish for imaging.

2.2.3 Immunofluorescence

Also carried out in accordance with Stellaris protocols, and integrated with the RNA-FISH method. The anti-Tax mAb LT4 (Lee, Tanaka and Tozawa, 1989), conjugated with Alexa-488, was incubated with cells and RNA-FISH probes during the overnight incubation step, before being washed in the same manner as the RNA-FISH probes. Other unlabelled primary antibodies were incubated in the overnight hybridisation step, while secondary antibodies were added in the first of the two subsequent 30 min wash steps.

2.3 Image analysis

2.3.1 Fluorescence microscopy

All coverslips were imaged using an Olympus IX70 inverted widefield microscope with a 100x/1.35 NA UPlan apochromatic (Apo) oil objective and a Spectra Light Engine illumination source (Lumencor), with a resolution of 64 nm/pixel and a spacing of 300 nm between optical slices. Z-offset was applied to correct for chromatic aberration, and exposures were adjusted on a wavelength by wavelength basis to optimise the balance between signal-noise ratio and photobleaching. Excitation filters used were 390/18 (DAPI), 475/28 (*gag*-FAM or Alexa-488 labelled antibodies), 575/25 (*hbz*-Quasar 570) and 632/22 nm (*tax*-Quasar 670), with corresponding emission filters of 457/20, 514/30, 632/60 and 692/40 nm. A Coolsnap HQ camera (Roper Scientific) was used for initial images, before being replaced with an ORCA-Flash 4.0 V2 digital CMOS camera (Hamamatsu).

2.3.2 Image processing

After acquisition, all images were passed through the following pipeline:

- A. Visual inspection of image-stacks to ensure cells were captured in their entirety in the z-axis. Individual 'cut-off' cells were excluded from the analysis. If more than a handful of cells were 'cut-off', the image-stack was discarded.
- B. Discarding of unneeded optical slices above and below cells in the image-stack, reducing image size to optimise subsequent processing and computing time.
- C. Correction for uneven illumination. An illumination function was calculated on a coverslip-by-coverslip basis using pooled maximum projections and the "Fit Polynomial function" (Lindblad and Bengtsson, 2001) from "CorrectIlluminationCalculate" in CellProfiler v2.2.0 (Kamentsky *et al.*, 2011), with a pipeline adapted from Roukos *et al.*, 2015. The images were then corrected using their corresponding function in ImageJ, with a custom macro created by Chad Whilding of the MRC LMS Microscopy Facility.

2.3.3 FISH-Quant analysis

Images were then analysed with an adapted version of FISH-Quant (Mueller *et al.*, 2013; Tsanov *et al.*, 2016), run on MATLAB. The adaptation (by Florian Mueller, creator of FISH-Quant) allowed step no. 5 in the following pipeline.

1. Outlines marking cells and nuclei were automatically generated by CellProfiler and then manually double-checked.
 - a. Nuclei were identified with "IdentifyPrimaryObjects", with the maximum correlation threshold (MCT) thresholding method (Padmanabhan, Eddy and Crowley, 2010).
 - i. The threshold correction factor was 1.0, lower and upper threshold bounds were 0 and 1 respectively, clumped objects were distinguished using intensity and shape was used to draw dividing lines.
 - b. Cell boundaries were identified with "IdentifySecondaryObjects", using Distance-B (Vincent and Soille, 1991; Jones, Carpenter and Golland, 2005) to expand the nuclear boundary outwards by up to 15 pixels unless it reached another cell.
 - i. Otsu, global, 3-stage thresholding was used at this stage, with priority given to minimising weighted variance over entropy. A threshold smoothing scale of 5 and correction factor of 0.9 were used, with lower and upper threshold boundaries of 0 and 1 respectively.

- ii. Given the extremely limited cytosol and near spherical shape of CD4+ T-lymphocytes, this was sufficient for roughly 90% of all cells.
 - c. Cells with poor outlines, high autofluorescence, or objects misidentified as cells were adjusted or removed by hand.
 - i. The number of pixels contained within each cell's cell and nuclear boundary constituted its area. Under the assumption that nuclei are approximately spherical in shape, this area was converted to a radius ($\text{radius} = \sqrt{\text{area} / \pi}$), which could then be used to calculate the volume of the nucleus or cell if it were a sphere ($\text{volume} = 4/3 \times \pi \times \text{radius}^3$).
2. All 'above-background' FISH signals were characterised with an extremely low intensity threshold, to determine the signal intensity threshold which will best separate true and false positives, and calculate an optimal intensity threshold to define 'transcriptional bursts'.
- a. Images are first passed through a double Laplacian of Gaussian filter (Raj *et al.*, 2008; Tsanov *et al.*, 2016) by FISH-Quant to reduce background and improve signal-noise of individual spots.
 - i. For *tax*, the first filter had a kernel of 7 pixels, while the second had a 1-pixel kernel.
 - ii. For *hbz*, the respective kernel sizes were 5 and 1 pixels.
 - b. Local maxima are then identified by passing an intensity threshold and scored to pass a minimum threshold of standard deviation above the background. They are then fitted with 3D Gaussian profiles.
 - i. For *tax* and *hbz*, the sub-region used for Gaussian fitting was 2 pixels around the spot centre in the xy-axis, and 4 pixels around each spot centre in the z-axis.
 - c. The shape parameters of each spot's 3D Gaussian fit are used to exclude signals which are not diffraction limited.
 - i. For *tax*, a PSF σ of 100-220 nm in the x-axis, and 250-550 nm in the z-axis.
 - ii. For *hbz*, 100-135 nm in the x-axis, and 275-420 nm in the z-axis.
3. Spots which passed intensity and shape thresholds were counted in each individual cell. The number of spots inside the nucleus was also noted.

4. All thresholded spots were averaged to estimate the number of nascent RNAs in 'transcriptional bursts' by point-spread-function (PSF) superimposition. The estimated size is in terms of how many complete transcripts the shape and size of the burst most likely corresponds to.
5. Cells were cell-cycle staged according to the integrated intensity of their nuclear DAPI signal. While carried out in FISH-Quant, this process was adapted from the method described in Roukos *et al.*, 2015. Cell cycle gates were determined by a combination of visual inspection of the histograms, and following the rule that cells in G₂/M have double the DNA content (and therefore integrated intensity) of cells in G₀/G₁. That is to say, the lower bound, middle and upper bound of the G₂/M gate should be as close as possible to double that of the respective sections of the G₀/G₁ gate (Fig. 5-3). Cells with an integrated intensity between the upper G₀/G₁ and lower G₂/M gate were considered to be "S". Varying the width of the cell-cycle gates by 10% did not materially affect the conclusions.
6. The individual measurements – mature RNA counts, nascent 'burst' counts, cell and nuclear size and cell cycle stage were collated and interpreted.

Finally, cells containing putative bursts were analysed individually using Imaris 3D-analysis software (Bitplane) to identify the 3D location of the burst relative to the 'centre-of-gravity' of the respective cell nucleus (explained in Appendix, Fig. 8-2); the intranuclear location was then normalised by nuclear volume.

2.4 Statistical analysis

Statistical analysis was carried out with R (version 3.3.3; R Core Team, 2017) and Graphpad Prism 6, which was also used to create all graphs. The use of a Poisson distribution and chi-squared tests to measure the significance of the proportion of *hbz*-negative cells in each clone was carried out by Charles Bangham.

Tests are described in figure legends. All chi-squared, Mann-Whitney and Student's t-tests were two-tailed, and a difference was considered significant if $p < 0.05$. Significance was symbolised in the following manner: ns ($p > 0.05$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$) and **** ($p < 0.0001$).

The relationship between *hbz* and *tax* expression and cell cycle stage was examined using binary and multinomial logistic regression with the “glm” and “multinom” functions respectively in R (Venables and Ripley, 2003; Sing *et al.*, 2005; Sarkar, 2008; Wickham, 2009).

2.5 DNA-FISH

For whole-cell DNA-FISH and immune-FISH, the fixation protocol described in section 2.1.6 was used. For cryo-FISH, however, an alternate protocol had to be followed; see below.

2.5.1 Fixation for cryosectioning

Approximately 1×10^7 cells were pelleted by 5 min centrifugation at 300 *g* in hydrophobic ‘non-stick’ Eppendorf tubes (ThermoFisher Scientific), before being resuspended in 4% formaldehyde, 250 mM HEPES-NaOH (2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulphonic acid, pH 7.6, Alfa Aesar) and incubated for 10 min at room temperature (RT). The cells were then centrifuged again at 300 *g* for 2 min, before having the fixative replaced with 8% formaldehyde, 250 mM HEPES-NaOH and being incubated at 4°C for 30 min. They were subsequently compacted through centrifugation at 4°C, in increasing *g*-forces: 300 *g* for 5 min, 500 *g* for 5 min, 1000 *g* for 2 min, 2000 *g* for 2 min, 4000 *g* for 2 min, 8000 *g* for 2 min, 10000 *g* for 10 min; in total, the cell pellet was incubated in 8% formaldehyde for 2 hours. This was then replaced with 1% formaldehyde, 250 mM HEPES-NaOH for long term storage in liquid nitrogen.

2.5.2 Cryosectioning

150 nm-thick Tokuyasu sections (Xie, Lavitas and Pombo, 2010) were cut from the cell pellets created in section 2.5.1 using a Ultracut UCT 52 ultracryomicrotome (Leica). Once 4-8 adequate sections were assembled, they were picked up from the blade using a 2 mm metal bacteriological loop dipped in 2.1 M D-sucrose, PBS (VWR Chemicals). They were then removed and transferred to 12 mm glass coverslips once the sucrose had thawed. The coverslips were stored at -20°C.

2.5.3 DNA-FISH probe design

Libraries of 45 nt oligonucleotides, each labelled with three 6-carboxyfluorescein (6FAM) or ATTO-550 fluorophores, were designed and manufactured by MYcroarray to target the HTLV-1 genome (166 oligos) and UIS-flanking sequences of clones A (395 oligos complementary to Chr4: 70,568,285-

70,598,285 in genome construct hg19) and C (932 oligos complementary to Chr19: 33779548-33879548). Precise sequences of all oligonucleotides used are available from CRM Bangham.

2.5.4 CryoFISH

Coverslips were washed twice in 2X saline sodium citrate (SSC, VWR Chemicals) buffer at RT, before being incubated with 250 µg/ml RNase A (Sigma), 2X SSC for 2 h at 37°C to remove RNA which could compete for the probes. They were subsequently permeabilised by washing in 0.1% Triton-X 100 (Sigma) in 2X SSC and 0.1 M HCl in water for 10min each. They were then dehydrated by washing in increasing percentages of molecular biology-grade ethanol (VWR Chemicals) before being air dried. An oven set to 80°C was used for denaturation, with 70% deionised formamide (MP Biomedicals), 50 mM phosphate buffer, pH 7 (Acros Organics), in 2X SSC forming the denaturation buffer. The coverslips were then transferred to ice-cold 30% ethanol, with the percentage increasing again before they were air dried prior to hybridisation.

The DNA-FISH probes were precipitated, denatured and reannealed in parallel. First, they were mixed with human Cot1 DNA (Sigma) and salmon sperm DNA (Sigma), along with 3 M sodium acetate and 100% molecular biology grade ethanol before being incubated at -80°C for an hour. They were then centrifuged in 100% ethanol at 15,000 *g* for 15 min before being air dried and resuspended in hybridisation buffer (2X SSC, 0.1 M phosphate buffer, pH 7, and 50% formamide and 20% dextran sulphate, Sigma), and chromosome paint (Applied Spectral Imaging) if one were required for the experiment. Finally, the probe mixture was denatured by incubation at 70°C for 10 min, and subsequently reannealed by incubation at 37°C for 1 hour.

Once coverslips and probes were ready, they were hybridised for 64-72 h at 42°C. After hybridisation, the coverslips were washed three times in 50% formamide, 2X SSC at 42°C, then PBS at 4°C, at which point the coverslips could be mounted on glass slides with DAPI-containing Vectashield mounting medium (Vectashield).

2.5.5 3D DNA-FISH

Coverslips with fixed whole cells were permeabilised by washing in ice-cold 0.4% Triton-X 100 in PBS for 10 min, before being washed in ice-cold 0.7% Triton-X 100 in 0.1 M HCl for a further 10 min. RNA was then removed by incubation with 0.1 mg/ml RNase A in 2X SSC at 37°C for 2 h, before the samples could be denatured by incubation in 50% formamide, 2X SSC at 80°C for 30min. They were then kept in ice-cold PBS until the FISH probes were ready, by the protocol described in the section 2.5.4.

Coverslip hybridisation, washing and mounting was then occurred out in the manner described in section 2.5.4.

2.5.6 Metaphase spreading and metaphase-FISH

In order to lock cells in metaphase, they were incubated in culture with 100 ng/ml colcemid (Sigma) for 2 h at 37°C. They were then centrifuged at 600 *g* for 15min at RT, before being resuspended in 75 mM KCl (Merck), HEPES, pH 7.4 (VWR Chemicals). The cells were then repeatedly centrifuged at 300 *g* before being resuspended through dropwise addition of ice-cold 3:1 methanol/glacial acetic acid (Sigma). At this point the cells could either be used immediately or stored at -20°C for long periods.

To achieve a metaphase spread, freshly resuspended cells were dropped from a height of 50 cm onto dry 12 mm coverslips. The pressure of impact cracks the cell membranes and allows the condensed chromosomes to stick to the coverslip. The coverslips were then air dried for an hour at RT, before being dehydrated in increasing percentages of ethanol and air dried again.

Metaphase spread coverslips could be hybridised with DNA-FISH probes or chromosome paints using the same protocol described for CryoFISH (section 2.5.4), but with a shortened RNA-removal step (30 min instead of 2 h) and no need for permeabilisation. This protocol was derived and adapted from Theunissen and Petrini, 2006.

2.5.7 Immuno-FISH

Antibody staining for proteins could be combined with either 3D or cryoFISH by insertion of a few steps between the two permeabilisation steps in either protocol. After permeabilisation in 0.4% or 0.1% Triton-X 100 respectively, coverslips were washed in 2% bovine serum albumin (BSA, Sigma) in PBS as a blocking solution at RT for 30 min, before being incubated with the primary antibody at an appropriate concentration in the same blocking solution at RT for 2 h. If a secondary antibody was required it would then be incubated in the same blocking buffer for 1 h at RT, with the coverslip being washed in blocking buffer before and after. The FISH protocol would then continue (0.7% Triton-X 100 or 0.1 M HCl, respectively). If solely staining cells with antibodies, the coverslip could also be fixed again in 2% formaldehyde, PBS for 10 min before mounting and imaging. This protocol was derived and adapted from Chaumeil, Micsinai and Skok, 2013.

2.5.8 Imaging

Coverslips were imaged using a TCS SP5 confocal microscope (Leica) with a 63x/1.40NA oil HCX Apo oil objective, run with Leica LAS AF software. The respective laser lines and emission filters (selected by acousto-optical tuneable filter, AOTF) used were: 405nm (DAPI, 413-482 nm emission), Argon/488 nm (Alexa-488 or 6FAM, 496-555 nm emission) and 561 nm (ATTO-500, 570-700 nm emission). Images were taken with a resolution of 96 nm/pixel in the xy-axis, with z-spacing of 500 nm between optical slices if z-stacks were performed.

CHAPTER THREE

QUANTIFICATION OF INDIVIDUAL VIRAL
TRANSCRIPTS IN SINGLE CELLS

3.1 Chapter Summary

This chapter summarises the use of smFISH to visualise and quantify *tax* and *hbx* in individual cells of five naturally-infected clones derived from patients, along with one uninfected control clone. This is the first report of HTLV-1 visualised quantified at the single-molecule, single-cell level. In brief, *tax*, which is a proxy of all plus-strand transcripts, displays highly bimodal expression within a clonal population of cells. A particular clone's level of global *tax* expression is determined by the proportion of its cells which are expressing *tax* very highly at a given time, rather than a change in average expression which is present in all cells. *hbx* (the lone minus-strand transcript), on the other hand, is expressed at extremely low levels. Although the proportion of cells across all clones which have detectable *hbx* is far higher than the proportion with detectable *tax*, *hbx* is nonetheless not present in all cells at all times, and the exact proportion varies between clones. While I have high confidence in the absolute count of *hbx* transcripts per cell, the *tax* count becomes an underestimate in high-expressing cells. The high number of *tax* transcripts in these cells increases crowding, reducing the probability of them being optically resolved and counted individually. Significant portions of the work described here have been submitted for publication (Billman, Rueda and Bangham, 2017).

3.2 Introduction

Understanding the level of viral expression in naturally infected cells is a key step towards understanding viral expression and latency *in vivo*, with the end goal being to find ways to exploit this knowledge to develop more effective therapies. My initial goal was to quantify HTLV-1 expression at the single-cell level, to identify any cell-to-cell heterogeneity that is lost in population-based approaches such as qPCR. To this end, I designed fluorescent-labelled probes targeting *tax*, as a proxy for plus-strand expression, and *hbx* on the minus-strand (Fig. 1-2). These probes were used to stain cells from five different HTLV-1-infected T-cell clones, isolated from patients' PBMCs by limiting dilution (see section 2.1.2, "Isolation of T-lymphocyte clones from patient PBMCs"). Biological triplicates of each clone were studied in independent experiments. All the cells in each clone share a single, unique viral integration site (Cook *et al.*, 2012); the provirus is complete in four clones (A-D), while in the fifth clone (E) the 5'-LTR and *tax* gene are deleted, a phenomenon common in leukaemic clones (Takeda *et al.*, 2004). The diffraction-limited signal of a single spot in smFISH corresponds to a single mRNA molecule (Vargas *et al.*, 2005; Raj *et al.*, 2008; Mueller *et al.*, 2013). In total, 19,477 individual cells were analysed.

3.3 Results

3.3.1 smFISH reveals stark difference in expression of *tax* and *hbz*

Measuring viral expression of individual cells using smFISH, it rapidly became apparent that *tax* expression was not homogeneous within the population of cells (Fig. 3-1). Rather, *tax* presented a bimodal expression profile: the majority of cells were *tax*-negative, while a minority expressed *tax* at a very high level, with hundreds of spots of detected (Fig. 3-2a). This is all the more surprising considering that the cells which make up each clone are isogenic, and have been cultured in identical conditions.

The *hbz* expression profile, in contrast, formed a unimodal population of low-expressing cells, most containing between 0 and 5 spots (Fig. 3-2b). Although there was some variation in the range of expression between different clones, the frequency of cells with successively higher expression levels of *hbz* diminished rapidly: out of ~20,000 cells, the highest observed *hbz* spot count was 25. For comparison, the cell with the highest observed *tax* count had 1015 spots. On average, approximately 75% of all cells had less than 5 *hbz* spots, this population of cells contributed two thirds of total *hbz* expression (Fig. 3.-3b). This contrasts strongly with *tax*, where 4.9% all of cells contained 100 or more *tax* spots: this minority was responsible for 80% of all *tax* transcripts (Fig. 3-3a). While far more commonly expressed than *tax*, *hbz* was not universally present in the cells of any clone.

The *tax* probes had a low false-positive rate: roughly 6% of uninfected cells had a single *tax* signal bright enough to be considered a transcript, with a similar false-positive rate in the *tax*-deleted clone E (Fig. 3-5a). Beyond that the false positive rate fell precipitously: 0.6% of cells with 2 spots and 0.04% with 3 spots. This would have negligible influence on high-expressing cells, where spot counts are measured in the hundreds. The false-positive rate could, however, have a larger effect on estimates of the binary question of what proportion of cells were *tax*-positive. To minimise the impact of false-positive single spots, I therefore defined any cell with 0 or 1 *tax* spots as *tax*-negative. The histograms in Figs. 3-2 and 3-5a reflect this (i.e. “0” represents cell with 0-1 *tax* spots, and “50” represents cell with 2-50 *tax* spots), while those showing exclusively “low” *tax* expression in Fig. 3-5b show the exact counts.

hbz spots had an even lower false-positive rate, 1.3% of uninfected cells were considered to have a single *hbz* spot, and none had more than one. Consequently, the *hbz* count was taken at face value.

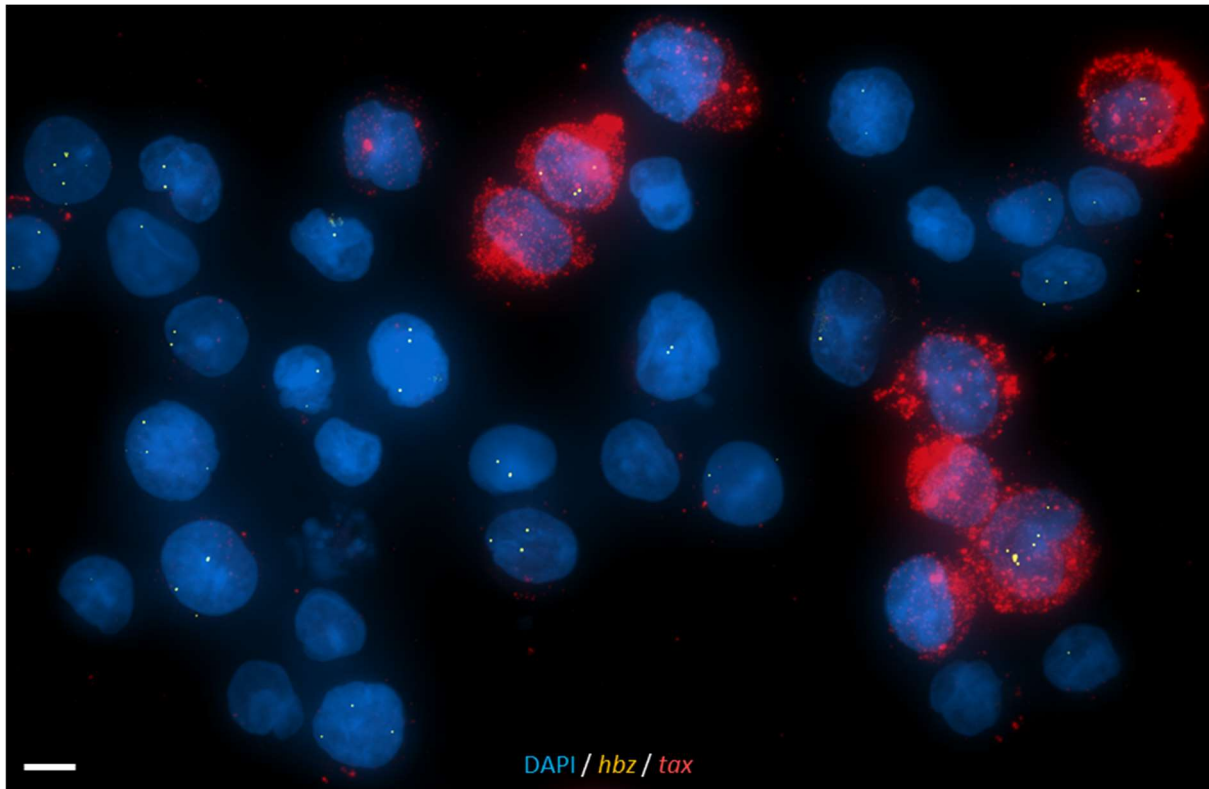


Figure 3-1 | Representative image of cells stained by smRNA-FISH

Image is from clone C, in which the highest overall level of *tax* expression was measured. With good cell density, approximately 20 cells could be captured in a single 89.1 x 66.6 μm frame using the Coolsnap camera at diffraction-limited resolution (<100nm per pixel), while 80-100 could be captured with the 131 x 131 μm frame of the later Hamamatsu camera. The goal was to capture a minimum of 1000 cells per clone replicate. Scale bar is 5 μm .

a, Composite fluorescence image with representative examples of cells expressing different levels of *tax*: i) silent (0-1 spots), ii) low (2-99 spots) and iii) high (≥ 100 spots), along with a frequency distribution of *tax* expression in an uninfected and one HTLV-1-infected clone. The same information is shown for *hbx* expression in panel **b**, where cells are considered i) silent if no *hbx* spots are detected, ii) low if they have 1-4 *hbx* spots, and iii) high if they have ≥ 5 spots. The roman numerals show respective cells' positions in the frequency distribution. Scale bar is 5 μm . n = 3,726 cells of clone A, 2,592 cells of clone F, each from three replicates.

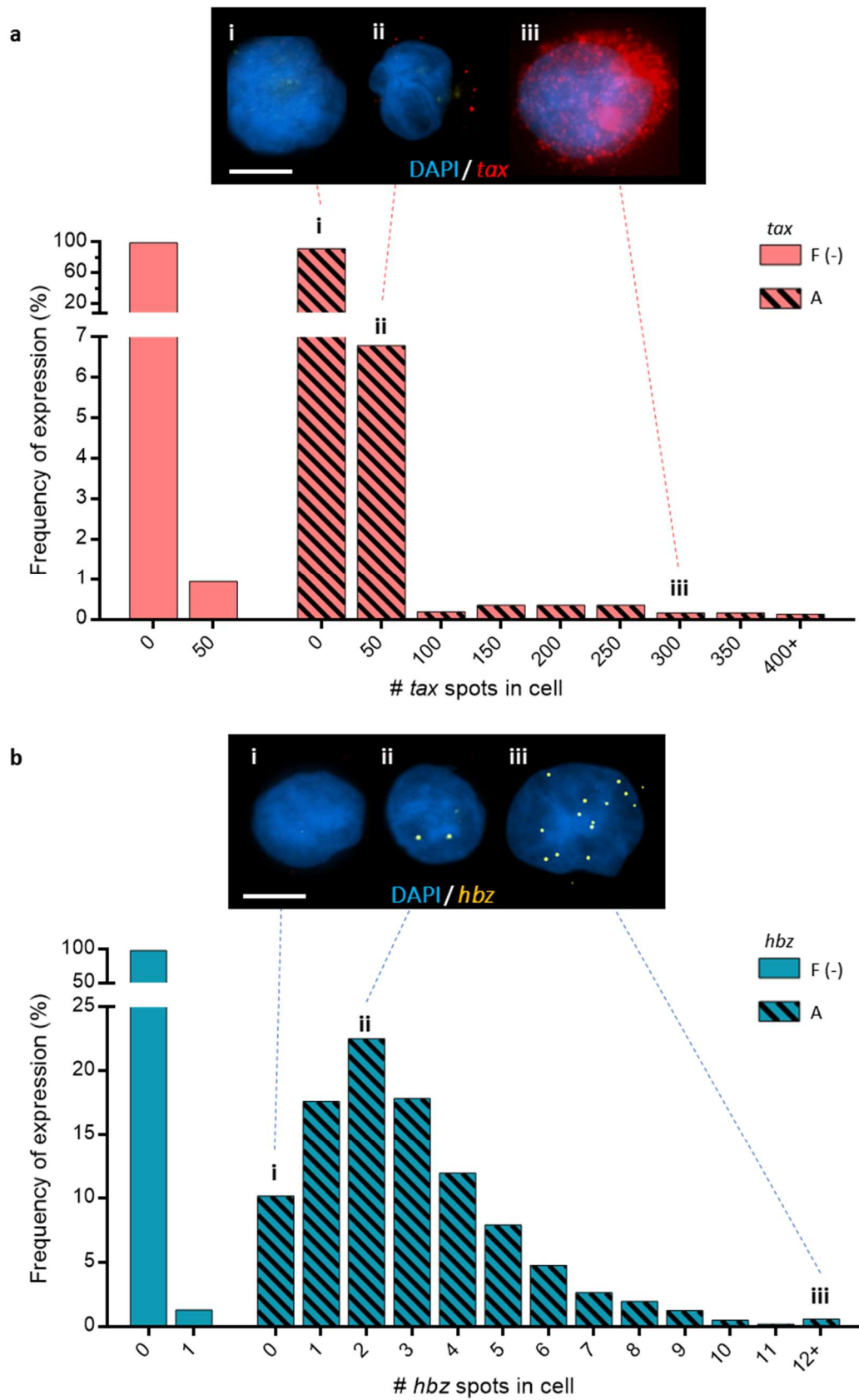


Figure 3-2 | smRNA-FISH reveals heterogeneity of *tax* and *hbz* expression

a, tax-negative cells are by far the most common, while the bulk of total *tax* comes from the minority of cells with very high expression, beginning at approximately 100 spots or more. **b**, *hbz* is far more uniformly expressed, with two thirds of all *hbz* found in cells with 4 spots or less. Data from pooled *tax*-competent clones A-D. n = 15,531 total cells, from three replicates per clone.

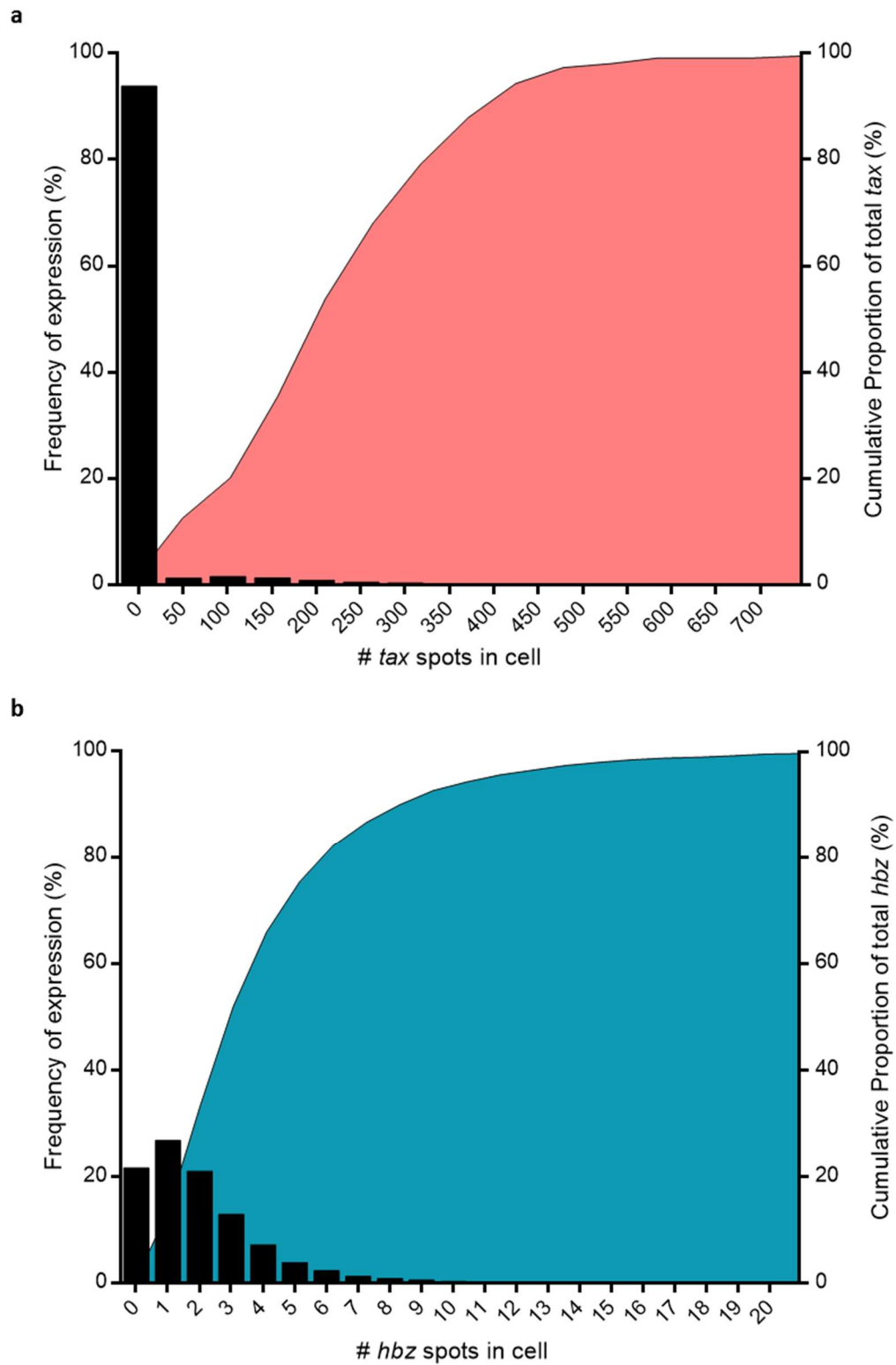


Figure 3-3 | A minority of cells contribute the bulk of total *tax*, while *hbz* expression is more uniform

3.3.1.1 *tax* expression corresponds to presence of Tax protein

To validate the specificity of the *tax* probe, cells from clones A and B were co-stained for *tax* mRNA along with the anti-Tax antibody LT4, labelled with Alexa-488. Virtually all cells with high *tax* also stained for Tax protein (Fig. 3-4). A negligible number of *tax*-negative cells also had positive signals in the Alexa-488 channel, most likely due to autofluorescence. As the *tax* probe does not discriminate between plus-strand transcripts, it was important to see that high *tax* RNA transcription resulted in detectable translation of Tax protein. This work was solely carried out to validate the specificity of the *tax* probe, so no quantitative analysis was carried out.

Previous studies have focused on the implications of Tax protein's intracellular location (Bex *et al.*, 1997; Nejmeddine *et al.*, 2005). Tax is thought to be most abundant in the nuclei of infected cells, with a limited amount near the microtubule organising centre (MTOC), from where it affects polarisation of the cell's microtubule network (Nejmeddine *et al.*, 2005). While no attempt was made to confirm this with staining of an appropriate cellular body, the most common location at which Tax signal was detected was near the "cleft" outside the nucleus, where the MTOC is often found. The lack of discernible intranuclear Tax signal is common when staining naturally infected cells with the LT4 antibody, and is most likely indicative of a diffuse intranuclear localisation, with associated low signal-noise, or the difficulty of accessing the nucleus with this antibody.

No commercially available antibody capable of detecting physiological levels of HBZ exists, and the proportion of *hbz* which is translated remains in question (Rende *et al.*, 2011; Raval *et al.*, 2015), so a similar validation was not possible or appropriate for *hbz*.

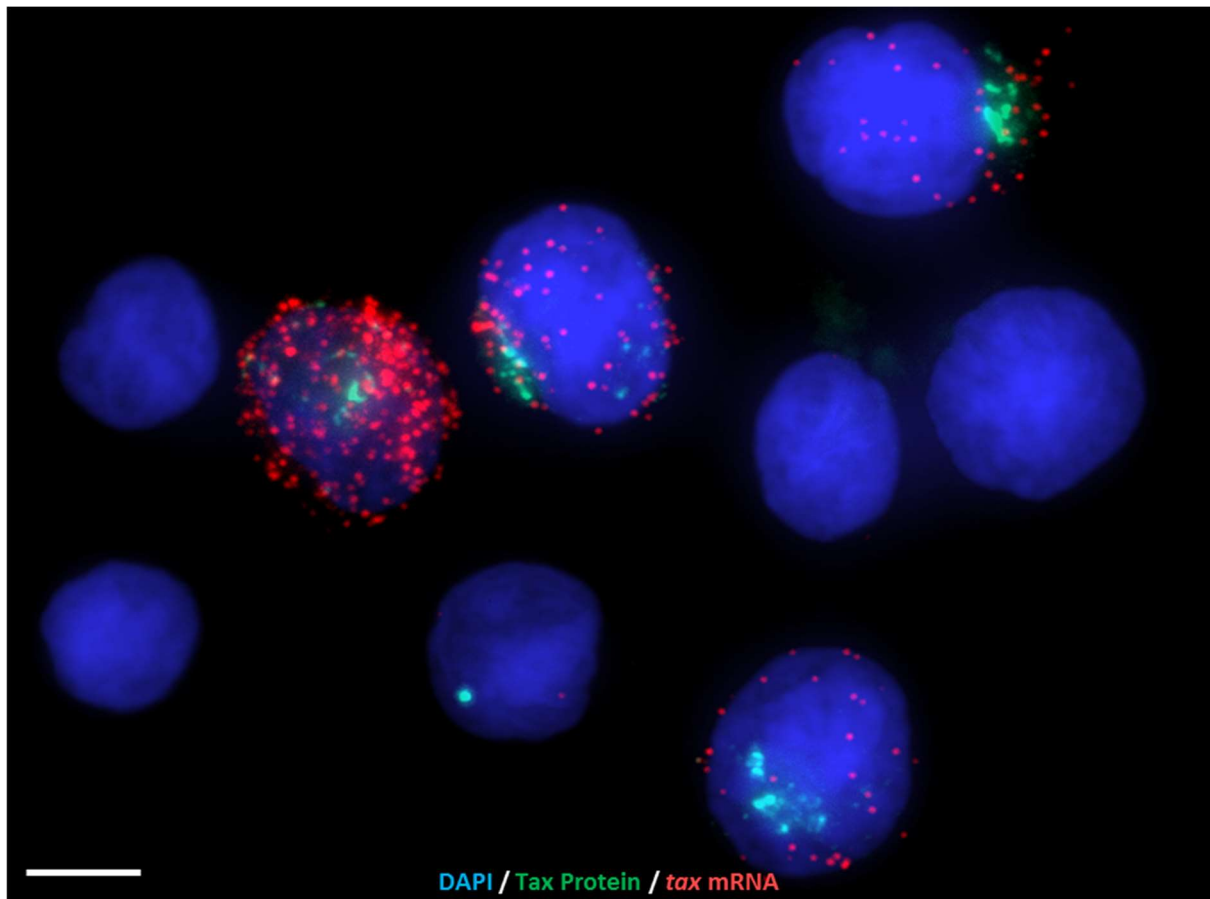


Figure 3-4 | Cells with high *tax* expression are positive for Tax protein

The vast majority of cells which were visibly high-*tax* in terms of RNA were also positive for Tax protein, while a negligible quantity of *tax*-negative cells had similar intensity signals in the Tax protein fluorescence wavelength. Scale bar is 5 μ m.

3.3.2 *tax* expression varies between clones

Tax expression has previously been found to vary between different clones (Melamed *et al.*, 2013). One question was whether this variation was attributable to a difference in the mean expression per cell, a difference in the expression of a proportion of cells, or simply a difference in the proportion of cells which expressed *tax* at all. As has been shown in Fig. 3-5a, *tax* expression is dominated by a minority of cells expressing at a very high level. In three out of the four *tax*-competent clones (A-C), there were significant differences in total *tax* expression, which could be entirely explained by the size of the population of cells with high levels of expression of *tax*.

These same clones (A-C) each had a similar range of intensity of *tax* expression – all cells with the maximum observed number of *tax* spots, or close to it. The maximum level of *tax* observed in clone D was much lower; however, all four clones had a bimodal distribution of *tax* expression.

There was also a notable difference in between clones in the proportion of cell classified as *tax*-negative: 92% and 97% in clones A and D had no detectable *tax*, but in clone B the proportion was 53%, while clone C had the lowest with only 9.4% of cells appearing *tax*-silent. The vast majority of this difference comes from what are considered “*tax*-low” cells (see section 3.3.4, “*hbz* count is highly accurate, while *tax* count underestimates high-expressing cells”): 39% of clone B cells had 2-99 *tax* spots, while clone C had 74% of cells with that level of *tax* expression. Despite this, the lion’s share of each clone’s total *tax* remains in the minority of cells with higher levels of expression (>100 spots, Fig. 3-3a), and it is my conclusion that the differences in proportions of cells which are *tax*-positive or *tax*-negative can largely be explained by differences in the frequency of *tax* expression, as measured by transcriptional bursts (see sections 4.3.1, “*tax* and *hbz* are expressed in transcriptional bursts”; and 4.3.2, “*tax* and *hbz* bursts are distinct from each and consistent between clones”). That is to say, in clones where *tax* transcriptional bursts are more frequent, it is less likely that a cell becomes entirely *tax*-negative before *tax* expression begins anew.

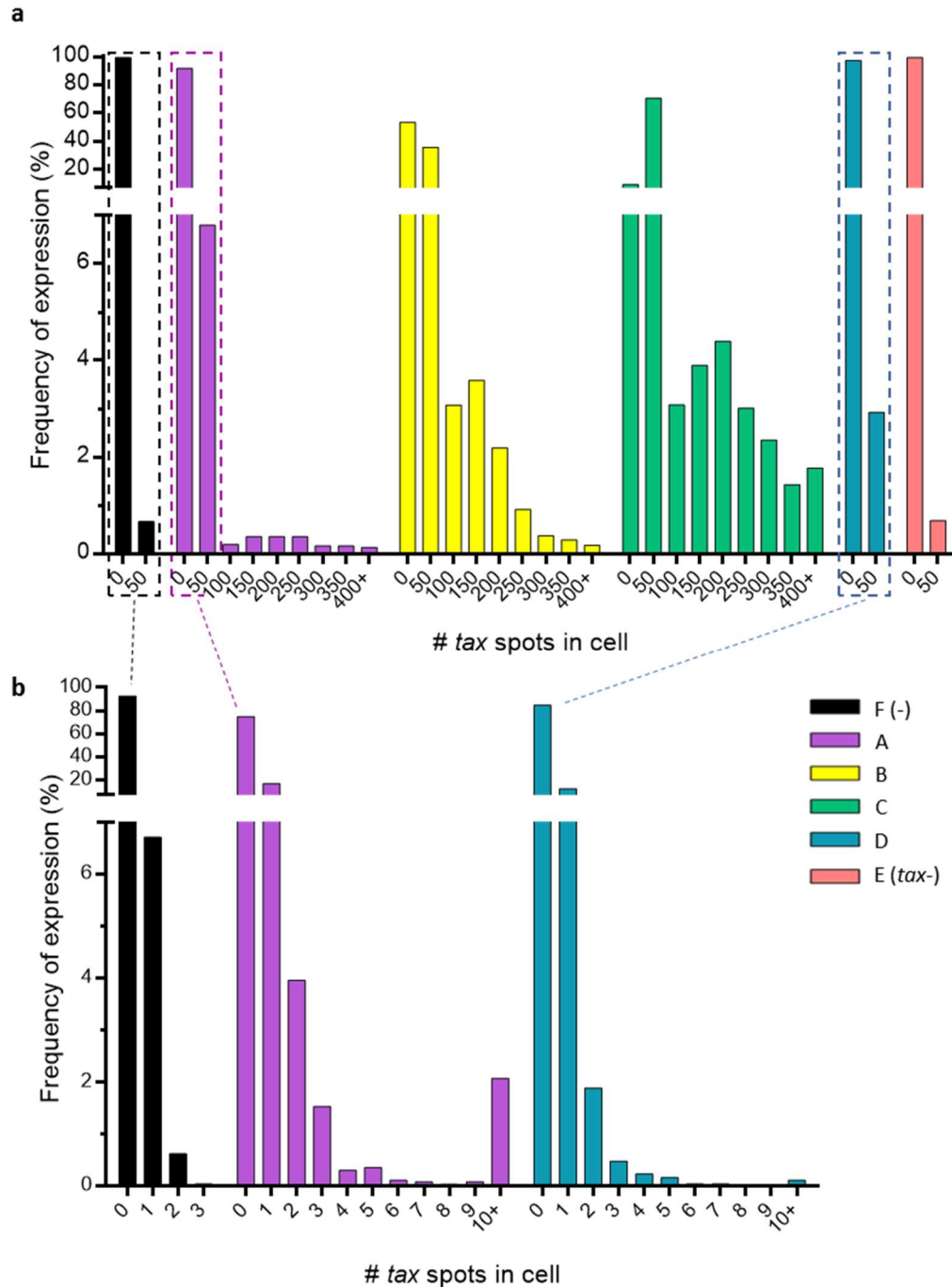


Figure 3-5 | *tax* expression in individual clones

Panel **a** compares the *tax* expression of the clones in the same manner as Fig. 3-2. Clone E is effectively a second negative control for *tax* expression, as *tax* has been deleted from its provirus. Clones A-C have a similar expression dynamic, with clone D having a lower maximum level of *tax*; its range of expression is shown more accurately in the next panel, focusing on low expression. **b**, the false positive rate of *tax* in uninfected cells is roughly 6%, but falls by an order of magnitude when cells with only 1 *tax* spot are also considered to be negative. For this purpose, all other *tax* histograms, and calculations of *tax* incidence, consider 1 or 0 *tax* spots to be equivalent. $n = 19,477$ total HTLV-1-infected cells and 2,592 uninfected cells, in three replicates per clone.

3.3.3 *hbx* expression is largely consistent between clones

hbx expression is far more reproducible between clones than *tax*. The expression profile is unimodal, with the range of each clone shifting slightly “leftwards”, with a lower modal expression and greater proportion of *hbx*-negative cells, or “rightwards”, with a higher modal expression and lower proportion of *hbx*-negative cells. As can be seen in Fig. 3-6, the overall pattern of low, consistent expression was present in all clones. However, the proportion of cells with detectable *hbx* had surprising variation between clones, from a high of 95% in clone A to 65% in clone C. The absence of universal *hbx* was also notably the case in clone E (29% *hbx*-negative), which is 5'-LTR-negative and thus solely reliant on minus-strand expression.

Given the weight of literature stating that *hbx* is constitutively expressed in HTLV-1-infected cells (Satou *et al.*, 2006; Kataoka *et al.*, 2015; Ma, Yasunaga and Matsuoka, 2016), as well as *hbx*'s very low level of expression, it is a fair question to ask if the unexpectedly high frequency of *hbx*-silent cells could be a result of a technical failure of the *hbx* probe to bind and detect RNA in all cells. Putting aside the high signal-noise ratio of the *hbx* probe, the proportion of *hbx*-silent cells in each clone was consistent between replicates. Furthermore, *hbx* was not distributed randomly among the cells of each clone. If it had been, the observed average frequency of ~2-3 *hbx* spots per cell would result in ~5% to ~18% of cells in each clone having zero *hbx* spots (Poisson distribution, Appendix Table 8.4). The observed frequency of *hbx*-negative cells was close to this random expectation (observed/expected = 1.1 and 1.2 respectively) in Clones D and E, although this overrepresentation of *hbx*-negative cells was still significant ($p < 0.0001$ for both). However, *hbx*-negative cells occurred at twice the expected frequency in Clones A, B and C (observed/expected = 1.9, 2.0 and 1.9 respectively, $p < 10^{-38}$ in each case; Appendix Table 8.4). It is notable that *tax* is expressed at a high level in Clones A, B and C, but is low or absent in Clones D and E. These results show that the observed *hbx* spots were not randomly distributed among the cells, implying that some cells in each clone were not expressing *hbx* at a given moment.

There was also some variation in the range of expression within clones: the clones with the lowest proportion of *hbx*-positive cells (B and C) seem to have a higher maximum of *hbx* expression. It is worth repeating that these clones are isogenic and have been cultured in identical conditions. Furthermore, clones A and D have come from the same patient, so the only significant difference between them is likely to be the location of the HTLV-1 proviral integration site.

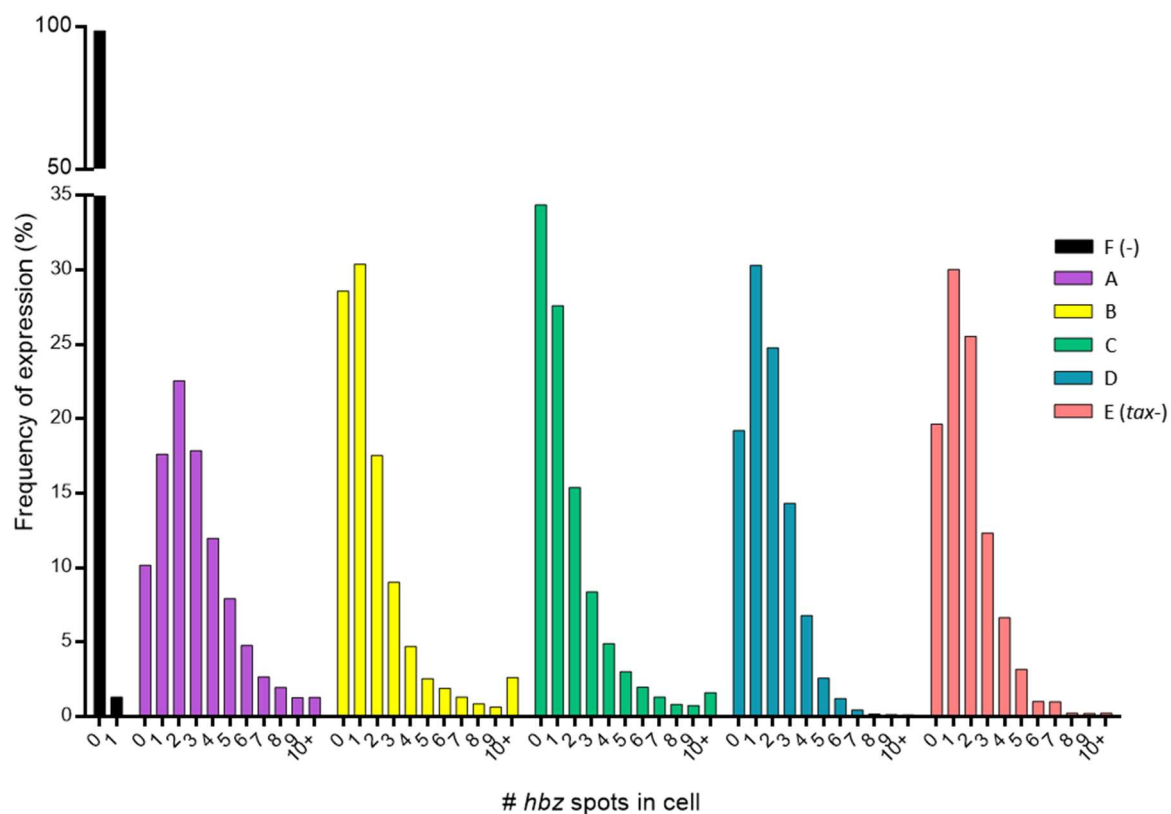


Figure 3-6 | *hbz* expression in individual clones

The overall pattern of expression is consistent between clones – a single population of low expressing cells – with slight shifts to higher or lower average expression, along with lower or higher proportions of *hbz*-positive cells. Clone F shows the extremely low false-positive rate of *hbz* spots. n = 19,477 total HTLV-1-infected cells and 2,592 uninfected cells, in three replicates per clone.

3.3.4 *hbz* count is highly accurate, whereas *tax* count underestimates the content of high-expressing cells

As was mentioned previously, a diffraction-limited spot can be expected to be the result of a single RNA transcript (Raj *et al.*, 2008), forming the basis of absolute quantification of expression by smRNA-FISH. The standard deviation (σ) of a signal's 3D gaussian profile (its point spread function, or PSF) is directly linked to its 'size', for example its full width at half maximum (FWHM). Shape thresholds to select diffraction limited spots were defined using sub-diffraction sized, 175 nm diameter fluorescent beads. For *hbz*, this was the case for the vast majority of detected signals which were bright enough to not be false positives – 98% of all spots detected in a representative replicate passed shape thresholding based on the sigma of a spot's PSF in the xy- and z-planes (Fig. 3-7a). Conversely, only 67% of detected *tax* spots in the same cells passed similar thresholds, adjusted for the difference in emission wavelength of their fluorophores (Fig. 3-7b).

If one focuses on cells with 100 spots or more of *tax*, the disparity between the total number of *tax* spots detected and the number which pass shape thresholds becomes increasingly apparent as the *tax* count increases (Fig. 3-7c). My hypothesis was that this was a result of crowding. More transcripts squeezed into a similar cellular volume increases the possibility that two or more transcripts will be so close to each other that they cannot be resolved microscopically, resulting in a non-diffraction-limited spot. This was borne out by comparing cells with similar numbers of detected spots, but far larger differences in spot count once shape thresholds were applied (Fig. 3-7d). Fig. 3-7e shows that of the three selected cells, only 23.8% of detected *tax* in cell C (the smallest of the three) passed shape thresholds, while 85.7% of detected spots passed the same thresholds in cell A (the largest). In cells with potentially extremely high *tax* expression (≥ 700 detected *tax* spots), larger cells on average had a higher proportion of spots pass shape thresholds than smaller cells ($p < 0.0001$ that linear regression slope deviated from 0).

A large majority of those spots are specific *tax* signals – very few signals which pass the *tax* intensity threshold are non-specific, for example through autofluorescence, by virtue of the fact that the *tax* fluorophore is in the far-red section of the visual spectrum, where autofluorescence is very low. These “too-large” spots are therefore mainly the result of multiple, co-localised transcripts. This means, for example, that instead of cell C having its *tax* count underestimated by approximately 841, it has approximately 841 spots which are discounted on account of resulting from multiple optically superposed transcripts, a number greater than $841 \times 2 = 1682$.

Based on this, an attempt was made to use the more absolute *tax* counts in cells with <100 spots to estimate the expected count in cells with higher levels of expression based on total fluorescence. This method had been demonstrated previously (Raj *et al.*, 2006), but there was too much variation in total, raw fluorescence between individual cells to confidently apply an estimate of the most likely count within individual high-*tax* cells (results not shown).

For this reason, all cells with ≥ 100 thresholded *tax* spots were analysed together as “*tax-high*” cells, with no importance given to the exact count of spots beyond that. This is also the level of expression beyond which cells account for the lion’s share of total *tax* expression (Fig. 3-3a). Altering this definition to 50 or 200 spots did not meaningfully affect any subsequent analyses.

a, *hbz* spots are almost entirely diffraction limited, with tight grouping of PSFs according to their shape in the xy- and z-axis (shown). The mode of each histogram is the σ of the 3D gaussian profiles of diffraction limited spots, as compared to 175 nm fluorescent beads (not shown). **b**, A much higher proportion of *tax* spots have non-diffraction-limited PSFs, failing shape thresholds. **c**, As the number of signals bright enough to qualify as *tax* spots in a cell increases, the proportion which fails shape thresholds increases. **d & e**, On average, a lower proportion of spots in *tax*-high cells pass shape thresholds if the cells have smaller volume. Scale bar is 5 μm . Representative histograms of spots from a replicate from clone C are shown in panels a and b, while all *tax*-competent cells are included in panel c (a handful are beyond the range of the axes); $n = 15,531$ cells.

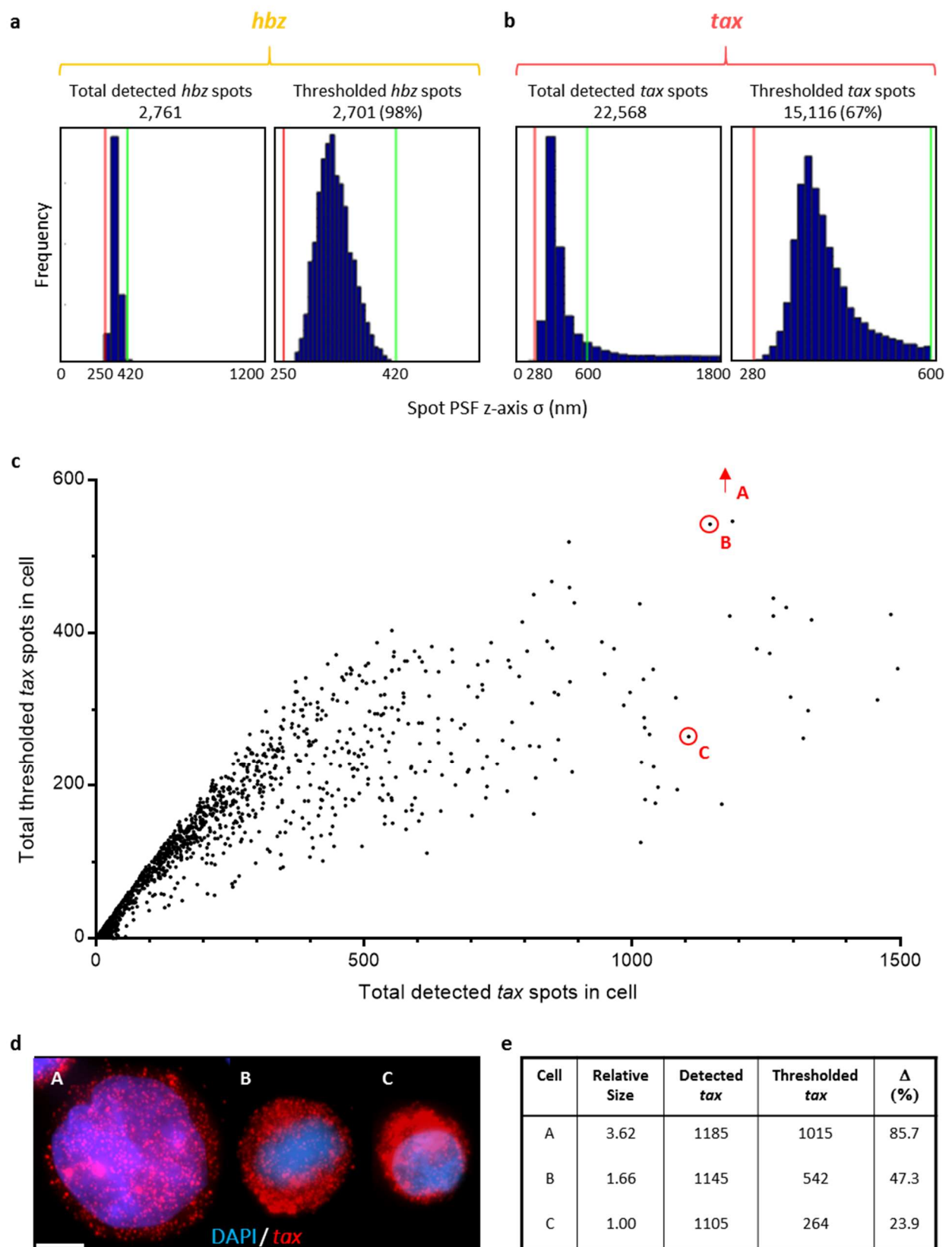


Figure 3-7 | *hbz* counts are highly accurate, but *tax* is underestimated

3.4 Conclusions

The work described in this chapter focused on quantifying viral expression through *tax* and *hbx* mRNA at the single-cell level. These were the overall conclusions:

- It was possible to use smFISH to visualise individual mRNA transcripts of *tax* and *hbx*
- This allowed the quantification of expression of both strands of the HTLV-1 provirus, in naturally infected T-cells, at the single-cell level
- Single-cell quantification reveals a startling degree of heterogeneity in *tax* expression among isogenic cells
- The bulk of *tax* in a population of cells comes from a minority of cells expressing at a very high level
 - The size of this minority is responsible for differences in total *tax* between clones
- *hbx* expression is much lower and far more uniform than *tax*, both within a clone and between clones
- However, there is a degree of variation, both in the range of expression and the proportion of cells which are *hbx*-silent
- Contrary to current literature, *hbx* is not present in all HTLV-1-infected cells at all times
- Absolute quantification by smFISH is based on the principle of detecting diffraction-limited spots
- This was possible in the vast majority of cases for *hbx*, but became increasingly difficult in cells with high levels of *tax* expression, possibly due to spatial crowding of transcripts

CHAPTER FOUR

VIRAL BURSTS AND mRNA DESTINATION

4.1 Chapter Summary

This chapter focuses on two observations made alongside those outlined in the previous chapter: HTLV-1 expression appears to follow a burst pattern, and the intracellular distribution of the *tax* and *hbx* transcripts is not the same. Using actinomycin D (ActD) to block transcription, it was demonstrated that “transcriptional bursts” are linked with ongoing transcription, and that *tax* and *hbx* RNA do not have similar fates once they have been created: the former is exported to the cytosol; the latter is retained in the nucleus. The ActD experiments also allowed me to chart the rate of disappearance of *hbx* mRNA in order to calculate its half-life in a much more rigorous manner than that reported by previous studies. Significant portions of the work described here have been submitted for publication (Billman, Rueda and Bangham, 2017).

4.2 Introduction

The previous chapter described the heterogeneity in levels of both *tax* and *hbx* in isogenic cell populations in identical environments. One of the likely factors behind this is the frequency with which the HTLV-1 genes are expressed. Some genes are not expressed continuously, but rather in bursts. This phenomenon has been found in an increasing number of mammalian genes over the last decade; there is also evidence that the HIV promoter follows what is known as a “bursty” pattern of expression (Raj *et al.*, 2006, 2008; Skupsky *et al.*, 2010; Dar *et al.*, 2012; Tantale *et al.*, 2016; Nicolas, Phillips and Naef, 2017).

Newly created and processed transcripts can have one of several fates: nuclear retention or export to the cytosol for translation, before eventually undergoing degradation to control protein expression or to correct one of various errors (Houseley and Tollervey, 2009). For *tax*, it seems reasonable to assume that its mRNA transcripts follow the classical pathway of export to the cytosol before being translated to Tax protein. Antibodies and cytotoxic T-lymphocytes (CTLs) against Tax typically make up a significant portion of the anti-HTLV-1 response in infected patients, and robust protein expression has previously been demonstrated *in vitro* (Daenke *et al.*, 1996; Hanon *et al.*, 2000; Enose-Akahata *et al.*, 2013). *hbx*, on the other hand, is more enigmatic. *hbx* RNA has been found to be retained in the nucleus at the cell population level (Rende *et al.*, 2011), and minus-strand transcription in ATLL leukaemic clones has been observed by RNA-Seq to run through into the flanking host genome, well past the classical *hbx* polyadenylation point, raising the question whether HBZ protein is expressed in ATLL cells at all (Kataoka *et al.*, 2015). On the other hand, polyadenylated *hbx* mRNAs have been detected *in vitro*, and while HBZ protein is on the very edge of detection in naturally infected cells *ex vivo* (Cavanagh *et al.*, 2006; Raval *et al.*, 2015; Shiohama *et al.*, 2016), the immune response against HBZ epitopes is a determining factor of clinical outcome (Macnamara *et al.*, 2010; Hilburn *et al.*, 2011), suggesting that there must be some amount of HBZ protein expression. In addition, *hbx* RNA has been found to have regulatory properties of its own (Satou *et al.*, 2006; Mitobe *et al.*, 2015).

smRNA-FISH presented the opportunity to investigate the possibility of viral transcriptional bursts, and track the intracellular distribution and final destination of viral mRNAs at the single-cell level.

4.3 Results

4.3.1 *tax* and *hbx* are expressed in transcriptional bursts

In addition to the small proportion of cells with very high levels of *tax* expression, a characteristic feature of *tax* was that many *tax*-positive cells also had a large nuclear spot which was typically far larger and brighter than the average spot (Fig. 4-1a, b). Many of these spots also co-localised with a strong *gag* signal (Fig. 4-1c). The *gag* probe targeted a section of the viral genome which would be spliced out if any viral products were being expressed, remain if the unspliced genome were being transcribed for the purpose of producing virions. Since the majority of splicing is thought to occur co-transcriptionally (Brugiolo, Herzel and Neugebauer, 2013), it seemed likely that these bright spots therefore corresponded to transcriptional bursts, which are increasingly recognized as a feature of both mammalian and retroviral genes (Raj *et al.*, 2006; Raj and van Oudenaarden, 2008; Tantale *et al.*, 2016; Nicolas, Phillips and Naef, 2017). In the putative transcriptional burst, nascent transcripts are created faster than they can diffuse away from the site of transcription, and as a result they cannot be individually resolved by microscopy. The resulting spot is larger and brighter than the average spot (Fig. 4-1b). The same characteristic spots were also identified for *hbx*, but with a smaller difference from average spots in intensity or size.

To test the hypothesis that the bright spots represent transcriptional bursts, we treated two of the clones (A and B) with ActD, to inhibit transcription and allow the putative bursts to disperse and disappear. Bursts were defined as intranuclear spots whose intensity significantly exceeded the main distribution of spot intensities (workflow of burst identification, verification and quantification described in section 2.3.3 “FISH-Quant Analysis” of material and methods, as well as Appendix Fig. 8-1). ActD treatment strongly reduced the frequency of these bright spots of both *tax* and *hbx*, when compared to the DMSO control, and their frequency and size diminished progressively during continuous exposure to ActD (Fig. 4-1d, e). These observations are consistent with the view that the bright spots do indeed represent transcriptional bursts. A small proportion of the bright spots may be chance superpositions of multiple transcripts, but I concluded that a majority represent ongoing or recent transcription. Hereafter, I will refer to these bright spots as bursts.

a, Representative *tax* and *hbz* bursts. Dashed lines show the positions of the intensity profiles in panel b. Scale bar is 5 μ m. **b**, Bursts are defined as specifically stained intranuclear spots that are significantly brighter than the main population of diffraction-limited spots. **c**, Many *tax* bursts are also strongly positive for *gag* which, if spliced out co-transcriptionally, should only be found during ongoing transcription. **d & e**, Upon blocking of transcription with actinomycin D, signals qualifying as bursts become increasingly rare and smaller in estimated size. n = 12,317 DMSO-treated cells, 7,120 actinomycin-D-treated cells, from two replicates each of clones A and B.

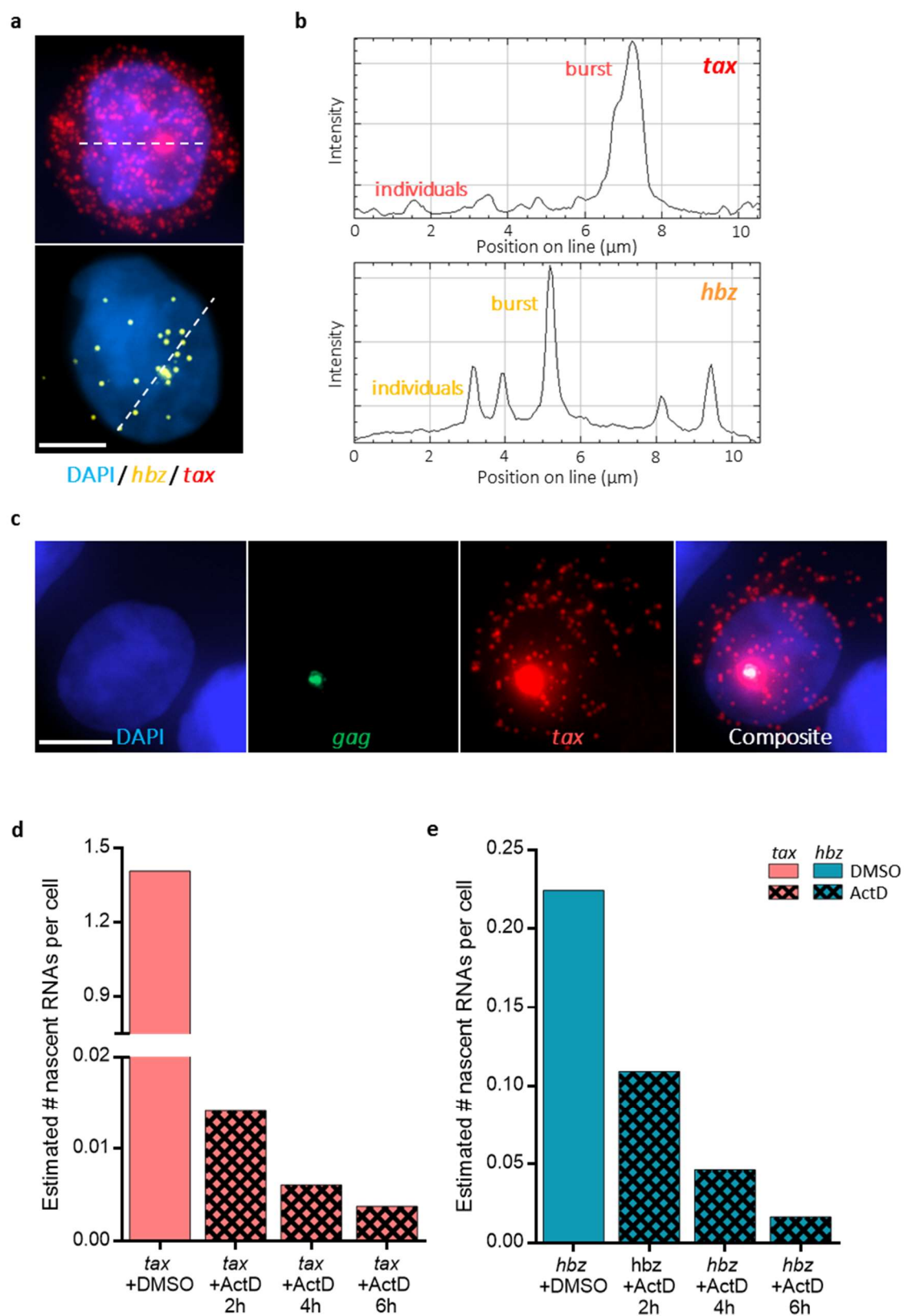


Figure 4-1 | *tax* and *hbz* are expressed in transcriptional bursts

4.3.2 *tax* and *hbz* bursts are distinct from each other and consistent between clones

The size of *tax* and *hbz* bursts was estimated by averaging thousands of diffraction-limited ‘single-transcript’ spots, and superimposing their PSF on that of the burst to estimate the number of transcripts that would be most consistent with a particular burst’s 3D shape and intensity. Nascent RNAs would not necessarily be long enough to bind the full complement of a probe’s oligonucleotide library to bind, potentially meaning they would not be as bright as mature RNAs (Mueller *et al.*, 2013). For direct comparison, burst sizes are estimated as the number of full length RNA transcripts which would best correspond to the bursts’ size and intensity.

Using this technique, we observed that *hbz* bursts were small in size, typically containing 3 to 4 transcripts, with low variation (Fig. 4-2d), although there were some notable exceptions in *tax*-high cells, especially evident in clones B and C which had the highest proportion of these cells.

Conversely, *tax* bursts were much larger: the largest bursts were estimated to contain hundreds of transcripts. *tax* bursts were also more variable in size (spot count), but were roughly comparable between the three clones with similar high-*tax* cells (clones A-C, Fig. 4-2c), while clone D’s lower maximum *tax* expression was accompanied by smaller *tax* bursts.

The frequency and size of *tax* bursts in the clones closely matched their overall levels of *tax*: Clone C had a significantly higher *tax* burst incidence than the other clones to go with its much higher total *tax* and proportion of cells which were *tax*-positive, followed by clones B, A and D in descending order (Fig. 3-5 and section 3.3.2, “*tax* expression varies between clones”).

While the same was largely true for *hbz*, there was more variation: clone A had the most total *hbz*, most frequent *hbz* bursts and highest proportion of *hbz*-positive cells (Fig. 4-2b). However, while clones B-E all had comparable *hbz* burst frequencies (Fig. 4-2b), clones B and C had markedly more *hbz*-negative cells as well as a seeming higher range of expression of *hbz* (Fig. 3-6 and section 3.3.3, “*hbz* expression is largely consistent between clones”). The *tax*-negative clone E appeared to have the lowest burst frequency of all clones, but the difference between it and clones B-D was not found to be significant.

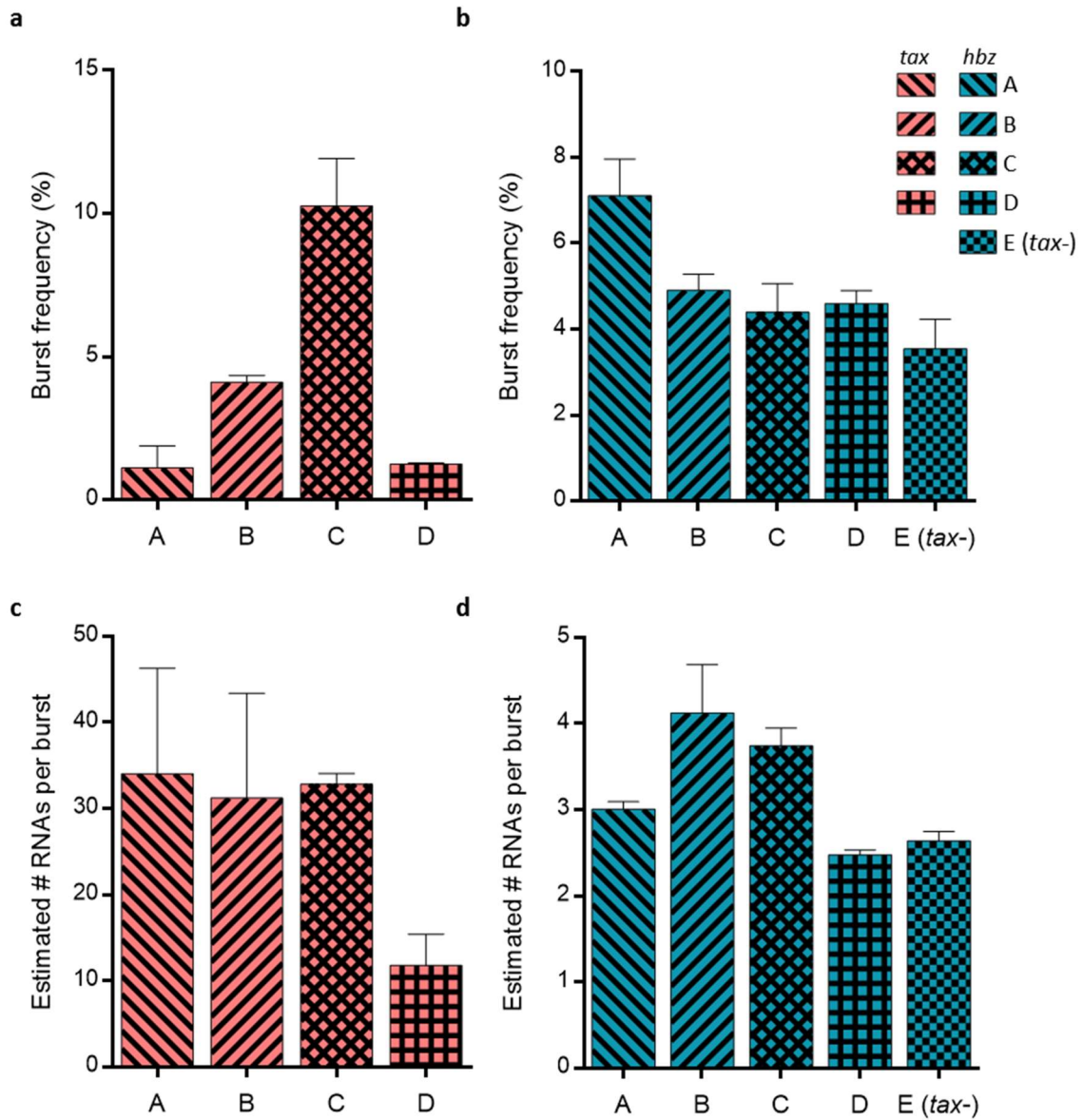


Figure 4-2 | The clone-specific characteristics of *tax* and *hbz* contribute to the overall differences between clones in the level of expression of the two genes

a & b, The frequency of *tax* and *hbz* bursts varies significantly between different clones, and to a degree reflect the clones' overall differences in viral expression. **c,** *tax* bursts are on average far larger and far more variable in size than *hbz* bursts. The average size is comparable between clones A-C but smaller in clone D, reflecting the difference in overall expression. **d,** *hbz* bursts are small and uniform in size, albeit with some variation between clones. Bars denote SEM, based on mean value of three replicates from each clone. $n = 514$ *tax* bursts and 943 *hbz* bursts in 19,477 total cells.

4.3.3 Transcriptional bursts have a non-uniform spatial distribution within the nucleus

If the observed bursts are indeed tied to transcription, then their nuclear location is a proxy for the location of the HTLV-1 provirus. I therefore measured the distance between the centre of a burst and the “centre-of-gravity” of the nucleus (explained in Appendix Fig. 8-2). This distance was then normalised to the volume of the nucleus, such that 100% denotes a burst at the very edge of the nucleus, while a burst with 0% would lie in the centre of the nucleus. In total, 469 *tax* and 818 *hbx* bursts were analysed in this manner. To more simply observe their distribution, three “shells” of equal volume were created within the nucleus (with a caveat that the nuclei were assumed to be perfectly spherical): starting from the centre, the “core” (the inner 69.3% of the nuclear radius), “middle” (between 69.3% and 87.4% of the nuclear radius) and “outer” (beyond 87.4% of the nuclear radius) nucleus (Fig. 4-3a). If the nuclear distribution of viral bursts was more or less uniform, they could be expected to be approximately evenly distributed between the three shells.

The results showed a non-uniform distribution of bursts within the nucleus (Fig. 4-3b), and significant differences between certain clones in the nuclear location of bursts (Fig. 4-3c).

Given the differences in viral expression between these same clones, a natural question was whether there was any correlation between expression and the putative intranuclear location of the provirus. For example, clones A and D have comparable spatial burst distributions, and are significantly different in location from the bursts of clones B and C. The latter clones also have significantly higher rates of *tax* expression and total *tax* (Fig. 3-5a), as well as containing *hbx* bursts which can be far larger than seen in other clones (typically in *tax*-high cells, see chapter 5 and Fig. 4-2d).

However, there was no clear and significant association between the nuclear location of a burst and the state of viral expression in that cell, for example the estimated size of the burst or total count of viral transcripts.

a, Schematic of nucleus split into equal volume “shells”. They are based on the assumption that T-cell nuclei are spherical, and are defined by normalised distance from the nuclear “centre-of-gravity” – 69.3% for core-middle boundary, 87.4% for the middle-outer boundary. **b**, If bursts were uniformly distributed in the nucleus, one would expect a roughly equal distribution, approximately 33% (dashed line), between the different shells. **c**, Intranuclear location of *tax* and *hbx* bursts in different clones, in terms of their distance from the nuclear centre-of-gravity. Boxes denote interquartile range while whiskers denote maxima and minima. The average nuclear radius is 4.2 μm so every 10% from the centre represents 400-500 nm in the average nucleus. *tax* n = 43 Clone A bursts, 126 Clone B bursts, 232 clone C bursts, 68 clone D bursts; *hbx* n = 258 clone A bursts, 126 clone B bursts, 77 clone C bursts, 241 clone D bursts, 116 clone E bursts. Distribution differences between clones compared using Mann-Whitney tests.

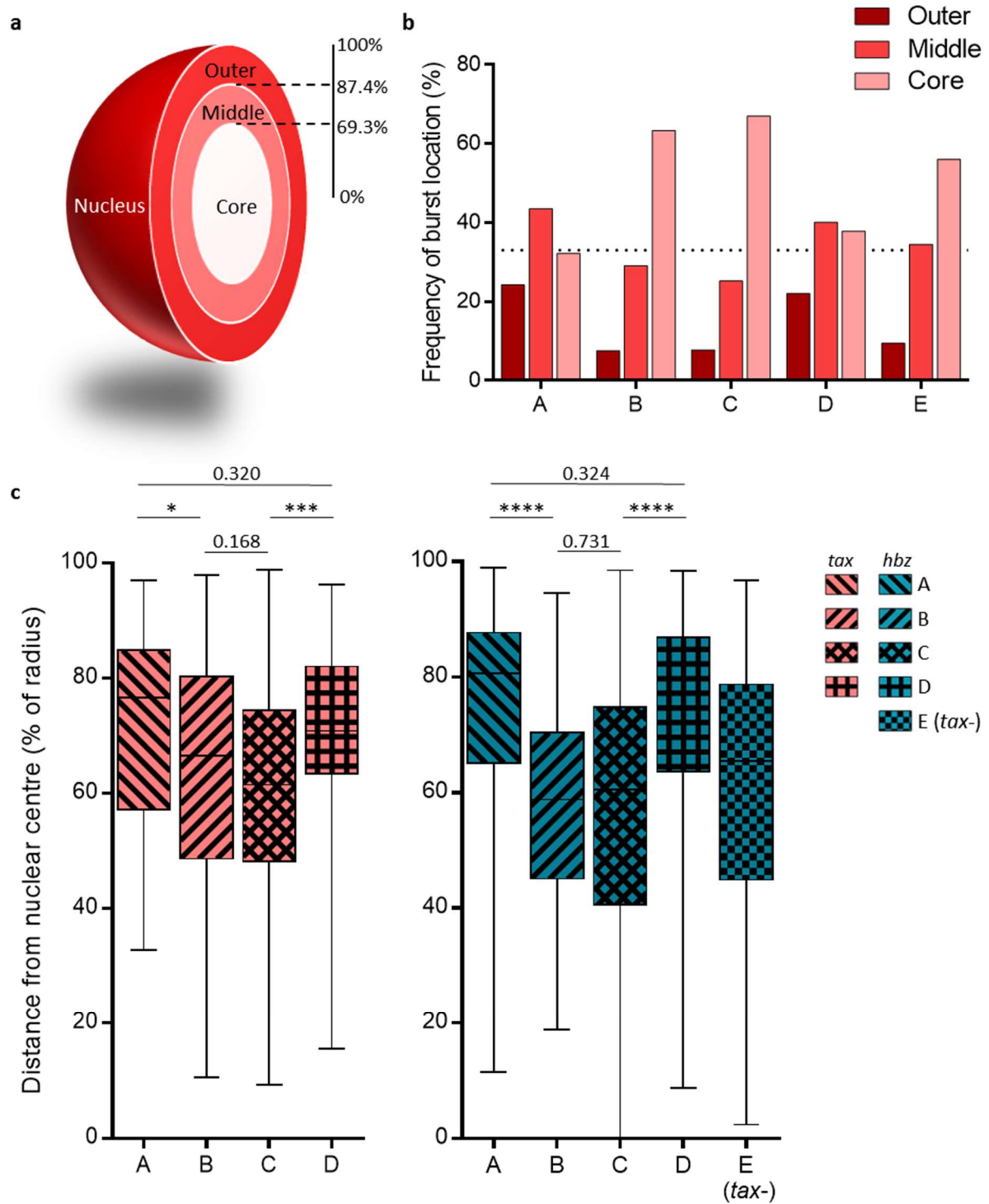


Figure 4-3 | Burst intranuclear locations are non-random and clone-specific

4.3.4 *tax* RNA is exported to the cytosol, while *hbz* is retained in the nucleus

Tax protein generates a strong, persistently activated T-lymphocyte response in infected patients, and the protein can readily be detected after short-term culture (<4 hours) *in vitro* (Hanon *et al.*, 2000; Macnamara *et al.*, 2010). In contrast, previous studies reported conflicting results regarding the expression of HBZ in protein form. Several studies indicate it may primarily be expressed as RNA, which itself promotes proliferation of the infected cell (Rende *et al.*, 2011; Kataoka *et al.*, 2015; Mitobe *et al.*, 2015; Cavallari *et al.*, 2016). However, the *hbz* mRNA has also been shown to be polyadenylated (Cavanagh *et al.*, 2006), and the avidity of CTLs for HBZ epitopes is one of the few significant predictors of lower infectious burden and disease risk *in vivo* (Macnamara *et al.*, 2010; Hilburn *et al.*, 2011; Rowan *et al.*, 2014). Therefore, I decided to investigate the proportion of both transcripts that could be found in the nucleus or cytosol of cells. As well as the individual clones in normal culture conditions, I also examined the same replicates of clones A and B which were treated with ActD, in order to track any changes over time in the intracellular distribution of transcripts once transcription was blocked. It is possible for mRNAs to degrade in the nucleus before being exported, but it is considered a rare event once they have been processed, and the degradation pathways are so efficient that any intermediates are very difficult to detect (Houseley and Tollervey, 2009).

While FISH-Quant analyses spots in 3D, cell and nuclear boundaries are defined using the 2D maximum projections and extended throughout the stack of optical slices. As a result, a spot is considered “nuclear” if it falls within the nuclear boundary in the xy-plane, regardless of its z-axis position. “Cytosolic” spots are those who fall outside the nuclear boundary in the xy-plane. While this results in an overestimation of the proportion of spots inside the nucleus (it includes spots which are above or below it), it is not possible to define the cells and nuclei on a slice-by-slice or 3D basis within the confines of FISH-Quant; it would also be enormously more computationally intensive. The intracellular distribution of transcripts in each cell was calculated as the percentage of total spots which were “nuclear”. In cells with few transcripts, particularly for *hbz*, there was a risk of sampling error. For example, in cells with one transcript, their nuclear proportion would have to be either 0% or 100%, potentially skewing the overall proportion. To avoid this, all cells from each biological replicate were considered as a single cell, so that thousands of nuclear and cytosolic spots could be considered.

With this definition of intracellular distribution, the difference between the localisation of *tax* and *hbz* was stark. About 90% of all *hbz* spots were found to have a “nuclear” localisation, a figure consistent across all clones (Fig. 4-4b and 4-5b). This proportion did not change significantly after transcription

had been blocked for 22 hours with ActD; the proportion of *hbz* spots found within the nucleus remained unchanged (Fig. 4-4a, b).

Conversely, ~60% of *tax* transcripts were found in the cytosol in 3 of 4 *tax*-expressing clones. This percentage fell to 40% within 22 hours after transcription was blocked (Fig. 4-4a, b). As discussed previously, the definition of the nucleus using a 2D maximum projection explains why the proportion of *tax* within the nucleus did not fall below 40% after 22 hours treatment with ActD, when 3D images indicate that the number should be far lower. (Figs. 4-5a). When the 3D images were further filtered to exclude spots in the upper and lower 20% of optical slices (which should be above and below, respectively, most nuclei in the z-axis), the “nuclear” proportion of *tax* fell by an average of 10%, while *hbz* distribution remained unchanged (data not shown). *tax* distribution in the fourth clone studied (clone D) was again the exception – the proportion of *tax* in this clone which was “nuclear” was comparable to the cytosolic-nuclear distribution of *hbz*.

While it can be said with confidence that *tax* mRNAs are rapidly exported from the nucleus once they have been transcribed, it was not possible using this pipeline to quantitatively measure this rate of export, given the imprecise measure of nuclear vs. cytosolic location, and the uncertainty as to exactly when *tax* expression began in any given cell. It is, however, possible to confidently conclude that almost all *hbz* transcripts remain nuclear in localisation.

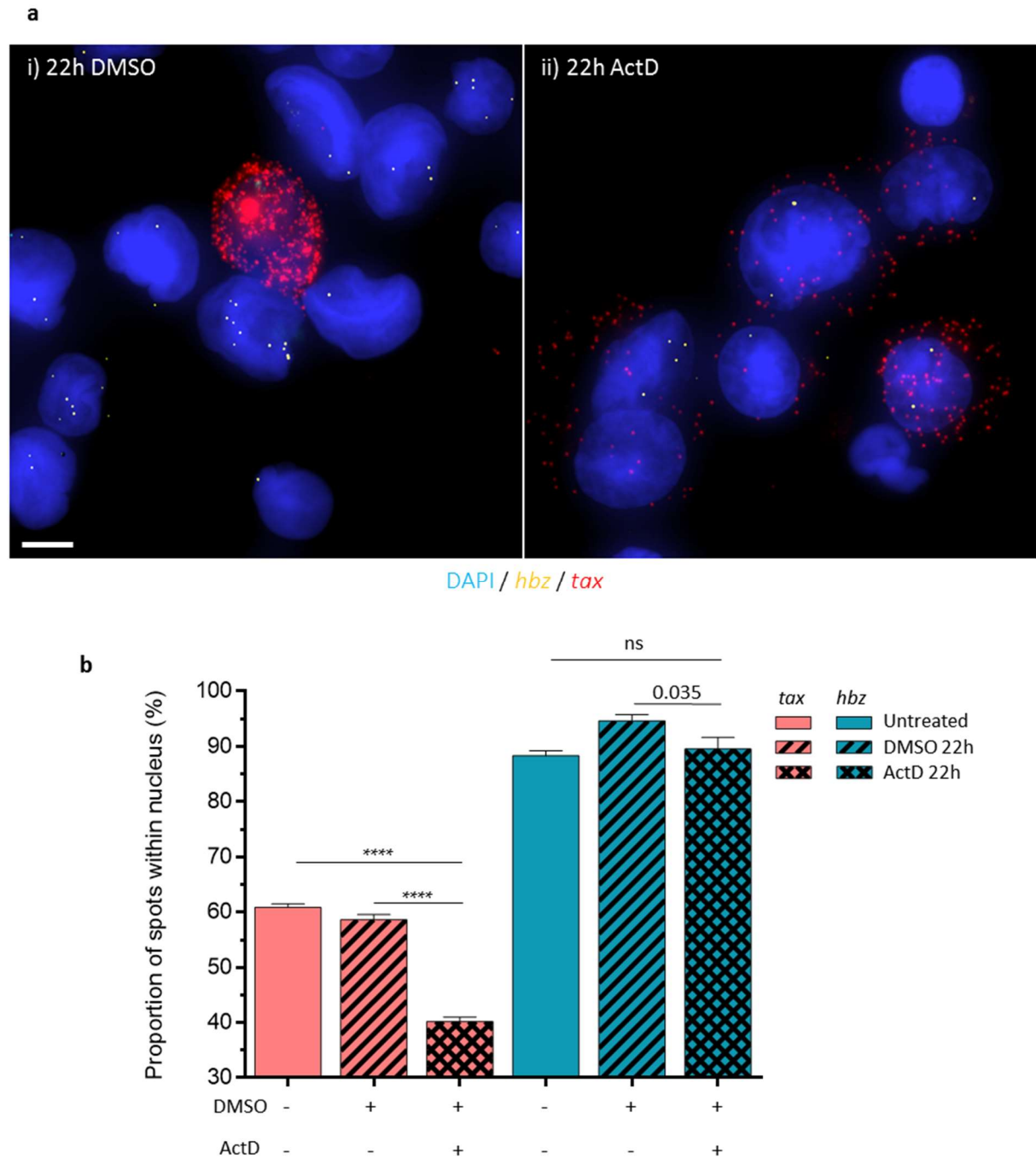


Figure 4-4 | *tax* is exported from the nucleus, whereas the majority of *hbz* is retained

a, Representative images showing i) typical cells after 22 hours in DMSO and ii) typical cells after 22 hours in actinomycin D. Scale bar is 5 μ m. **b**, *hbz* mRNA remains largely intranuclear, whereas *tax* mRNA is mostly cytoplasmic; this difference is accentuated by blocking transcription with actinomycin D. Significance of differences measured using unpaired t-tests with Welch's correction. $n = 936$ untreated (16,192 *tax*, 3,796 *hbz* spots), 390 DMSO-treated (6,147 *tax*, 2,344 *hbz* spots) and 195 ActD-treated (1,969 *tax*, 258 *hbz* spots) cells, each from two pooled biological replicates of clone A and B.

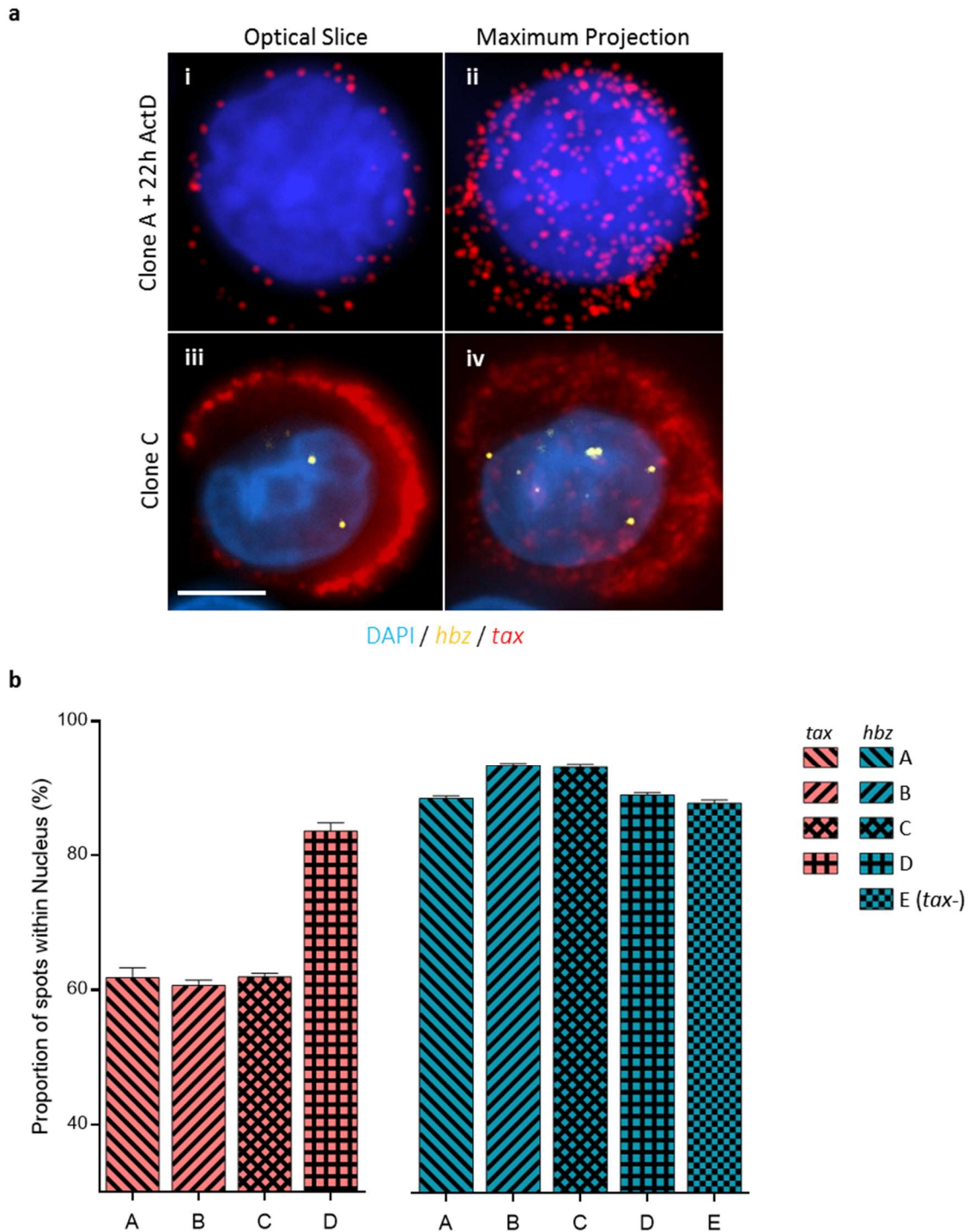


Figure 4-5 | Intracellular distribution of viral transcripts in individual clones

a, Representative images showing i,ii) a *tax*-positive cell which has been exposed to ActD for 22h, and iii,iv) A normally cultured cell in which *tax* transcription has ended. Images i and iii are individual optical slices from the maximum projections shown in ii and iv respectively. As nuclear boundaries are determined using the 2D maximum projection, transcripts which in reality are above or below the nucleus, will erroneously be marked as intranuclear. Scale bar is 5 μ m. **b**, *hbz* mRNA remains largely intranuclear in all clones, whereas *tax* mRNA is mostly cytoplasmic in all clones bar one, where it has a similar distribution to *hbz*. Mean values are shown, and bars denote SEM. $n = 196,086$ *tax* spots and 39,258 *hbz* spots in 19,477 cells, from three replicates per clone.

4.3.5 *hbz* RNA has a half-life of 4.4 hours

I estimated the half-life of *hbz* mRNA from the decline in *hbz* spot count following the start of transcription inhibition with ActD. The density of *tax* spots in many *tax*-expressing cells precluded a precise spot count of *tax* (see section 3.3.4, “*hbz* count is highly accurate, whereas *tax* count underestimates the content of high-expressing cells”); as a result, it was not possible to make a reliable estimate of *tax* mRNA half-life.

Two biological replicates each from clones A and B were exposed to ActD for 0, 2, 4 and 6 hours (one of the replicates failed at 4 hours, leaving 3 measurements for that timepoint); control cells were treated with the DMSO vehicle. A total of 15,524 cells were studied. Initially this time course included a timepoint of 22 hours after exposure to ActD. However, after 22h there was a noticeable increase in the frequency of apoptotic cells, a known effect of ActD treatment (Der *et al.*, 1997). This created a risk of *tax* and *hbz* conferring a survival bias upon *tax*-positive or *hbz*-positive cells, thanks to their anti-apoptotic effects (Mühleisen *et al.*, 2014; Tanaka-Nakanishi *et al.*, 2014; Mitobe *et al.*, 2015), increasing the apparent number of transcripts per cell as more silent cells apoptosed. For this reason, the 22 h timepoint was discarded.

The total counts of *hbz* spots, including the estimated sizes of *hbz* bursts, were combined and divided by the number of cells, to estimate the mean total *hbz* content per cell; this mean value decreased over time (Fig. 4-6), with a half-life of 4.4 hours (95% confidence intervals 4.1 hours to 5.1 hours).

This measurement was the result of a linear regression of the natural log (ln) of the total *hbz* per cell; the reciprocal of the slope corresponds to the mean lifetime τ , which allows calculation of the half-life by:

$$t_{1/2} = \tau \times \ln(2)$$

Fitting the untransformed data with a one-phase exponential decay curve returned similar results – an estimated half-life of 4.2 hours with 95% confidence intervals of 3.8-4.8 hours. This was estimate significantly longer than found in previous studies, which reported an *hbz* mRNA half-life of approximately 1 h by qPCR.

Of potential interest, there was a slight difference in the estimated half-life between the two clones: 5.0 hours for clone A and 4.0 hours for clone B. Aside from the difference in *hbz* burst frequencies (Fig. 4-2b), this could be another potential reason for the varying proportion of cells in each clone which are *hbz*-positive - ~90% in clone A, ~70% in clone C (Fig. 3-6).

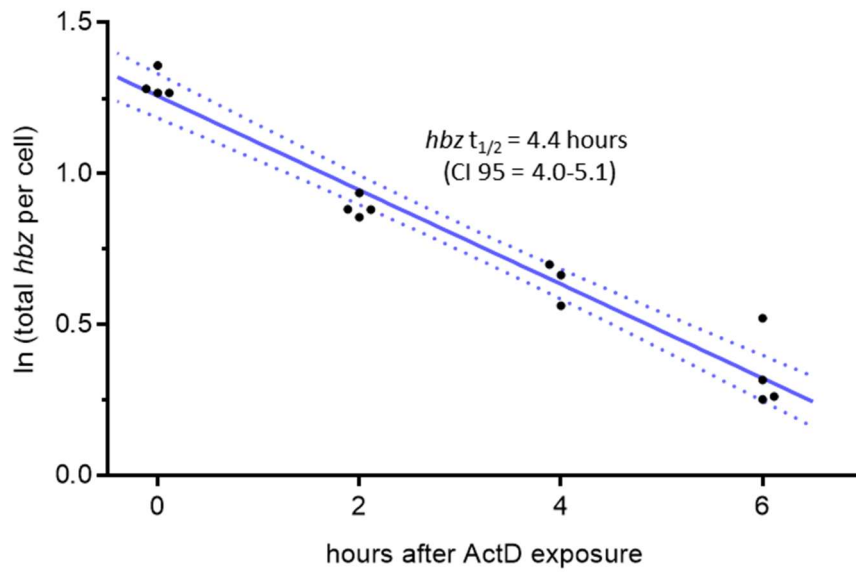


Figure 4-6 | Decay curve of *hbz* RNA

The reciprocal of the slope of the log-transformed data is the mean mRNA lifetime, which can be converted into mRNA half-life. Individual measurements from replicates are shown, along with a linear regression and 95% confidence intervals (dashed lines); one of the four replicates failed at the 4-hour time point. $n = 8126$ total cells, from two replicates each of clones A and B.

4.4 Conclusions

The work in this chapter focused on the beginning and end of viral expression, namely transcriptional bursts, mRNA intracellular distribution and half-life in the context of inhibition of transcription. The broad findings were:

- Both plus- and minus-strand transcription of HTLV-1 appear to occur in “transcriptional bursts”
- These bursts are characterised by being significantly brighter and larger than the main population of diffraction-limited spots
 - *tax* bursts also contain *gag* signal, indicative of unspliced, nascent RNA
- The frequency and size of these putative bursts decreases progressively after inhibition of transcription by actinomycin D
- These bursts, as transcriptional proxies of proviral location, are not uniformly distributed throughout the nucleus
- Each clone has a significantly different average intranuclear burst location, presumably because of the chromosomal location of the respective proviruses
- *tax* is rapidly exported from the nucleus in most cases
- A significant majority of *hbz* is uniformly retained within the nucleus
- *hbz* mRNA is estimated to have a half-life of approximately 4.4 hours, significantly longer than found in previous studies (see Discussion, section 7.5, “Intracellular distribution and half-life of *hbz*”)

CHAPTER FIVE

VIRAL EXPRESSION AND THE CELL CYCLE

5.1 Chapter Summary

The earlier chapters have focused on *tax* and *hbz* in isolation, when in fact they are often expressed in the same cell. This chapter describes the search for correlations between the expression of the plus- and minus-strands, represented by *tax* and *hbz*. It was also recently shown that one could ascertain the cell cycle stage of individual cells using fluorescence microscopy. In this work, the relationship I observed between *tax* and *hbz* within individual cells was consistent with a previously described feedback mechanism to control *tax* expression, and the expression each gene was found to be correlated with cell cycle progression, although more consistently in the case of *hbz*. This culminated in the formulation of a basic model of viral expression based on the interpretation of the work in this chapter, along with observations made earlier in the thesis. Significant portions of the work described here have been submitted for publication (Billman, Rueda and Bangham, 2017).

5.2 Introduction

The previous chapters have considered *tax* and *hbz* transcript abundance, intracellular distribution and transcription independently of each other, with no consideration given to the possibility that HTLV-1 plus- and minus-strand expression could be interrelated. *tax* expression is thought to establish a positive-feedback loop through the recruitment of TFs, including CREB and CBP/p300 by Tax protein (Anderson and Dynan, 1994), and a longstanding unanswered question is how that loop is broken, as well as how transcription is initiated in the first place. One possibility is that HBZ could play a role, as *in vitro* studies have indicated that *hbz* expression can be stimulated by Tax (Yoshida *et al.*, 2008), and that HBZ can compete with Tax for the transcription factors which bind the 5'-LTR promoter (Lemasson *et al.*, 2007). However, previous work has typically been carried out in transfected cell lines, with non-physiological levels of transcription, and not been single-cell based.

Tax and both HBZ protein and *hbz* RNA have a plethora of effects on cell proliferation (Satou *et al.*, 2006; Currer *et al.*, 2012; Matsuoka and Yasunaga, 2013; Mitobe *et al.*, 2015). It has also been demonstrated that genes which have no functional role in the cell cycle can exhibit a changed expression profile while the cell is cycling (Padovan-Merhar *et al.*, 2015; Skinner *et al.*, 2016). Understanding the nature of the relationship between viral expression and cell cycle, if it exists, is therefore crucial. A recent study (Roukos *et al.*, 2015) demonstrated the possibility of cell-cycle staging individual cells using fluorescence microscopy. As a result, it became possible to quantify viral expression in single cells in the context of their cell cycle stage.

5.3 Results

5.3.1 *tax* and *hbz* expression are not independent of each other

Tax protein drives transcription of its own gene by recruiting transcription factors CREB and CBP/p300 to “Tax-Response Elements” in the 5’-LTR promoter (Curren *et al.*, 2012), creating a positive feedback loop that can be inhibited by competition in the form of HBZ protein (Lemasson *et al.*, 2007; Yoshida *et al.*, 2008). I therefore examined the four *tax*-expressing clones to test for a correlation between *tax* and *hbz* expression.

tax expression was strongly anti-correlated with the presence of *hbz*: cells were ~4x more likely to contain a *tax* burst if they were *hbz*-negative (Fig. 5-1b), and the frequency of *tax* bursts rose from 2% to over 8% in *hbz*-negative cells. *hbz*-negative cells also had a higher average amount of *tax* (not shown), and cells with *tax* bursts had a far lower proportion of cells with detectable *hbz* than did cells which lacked a burst or were *tax*-silent (Table 5-1).

The count of *tax* spots was strongly positively correlated with the fraction of cells containing a *tax* burst (Fig. 5-1c) in all clones studied, including clone D, although this clone differed from the other clones studied in having a lower peak *tax* expression and a higher proportion of *tax* mRNA retained in the nucleus. Overall, *tax*-negative cells had a *tax* burst frequency of 0.37%, while *tax*-low cells had 2.6%, and almost half (49.6%) of all *tax*-high cells contained a *tax* burst.

There were many examples of *hbz* bursts in *tax*-negative cells, and plenty of cells with high *tax* expression were *hbz*-negative (Table 5-1). However, high *tax* cells were almost three times more likely to have a burst of *hbz* expression than cells with no or low levels of *tax*, increasing from approximately 4.5% to 14.5% (Fig. 5-1d). This effect was observed in all three clones containing cells with high *tax* expression. *hbz* bursts also tended to be significantly larger in *tax*-high cells than their counterparts in *tax*-low or *tax*-negative cells: on average, they were estimated to contain 4 *hbz* transcripts, while *hbz* bursts in *tax*-low or silent cells were estimated to contain 3 transcripts on average ($p < 0.0001$, Mann-Whitney). This difference accounts for the larger average *hbz* burst size in clones B and C, which contained the most *tax*-high cells (Fig. 4-2d).

These data suggest a model for the temporal progression of HTLV-1 expression (Fig. 5-1a): cells are more likely to express *tax* in the absence of *hbz*, which will then establish a feedback loop, leading to very high levels of *tax*. High levels of *tax* are associated with an increased probability of *hbz* expression, and once *hbz* is again present in a cell, *tax* expression in the form of bursts is more likely to cease.

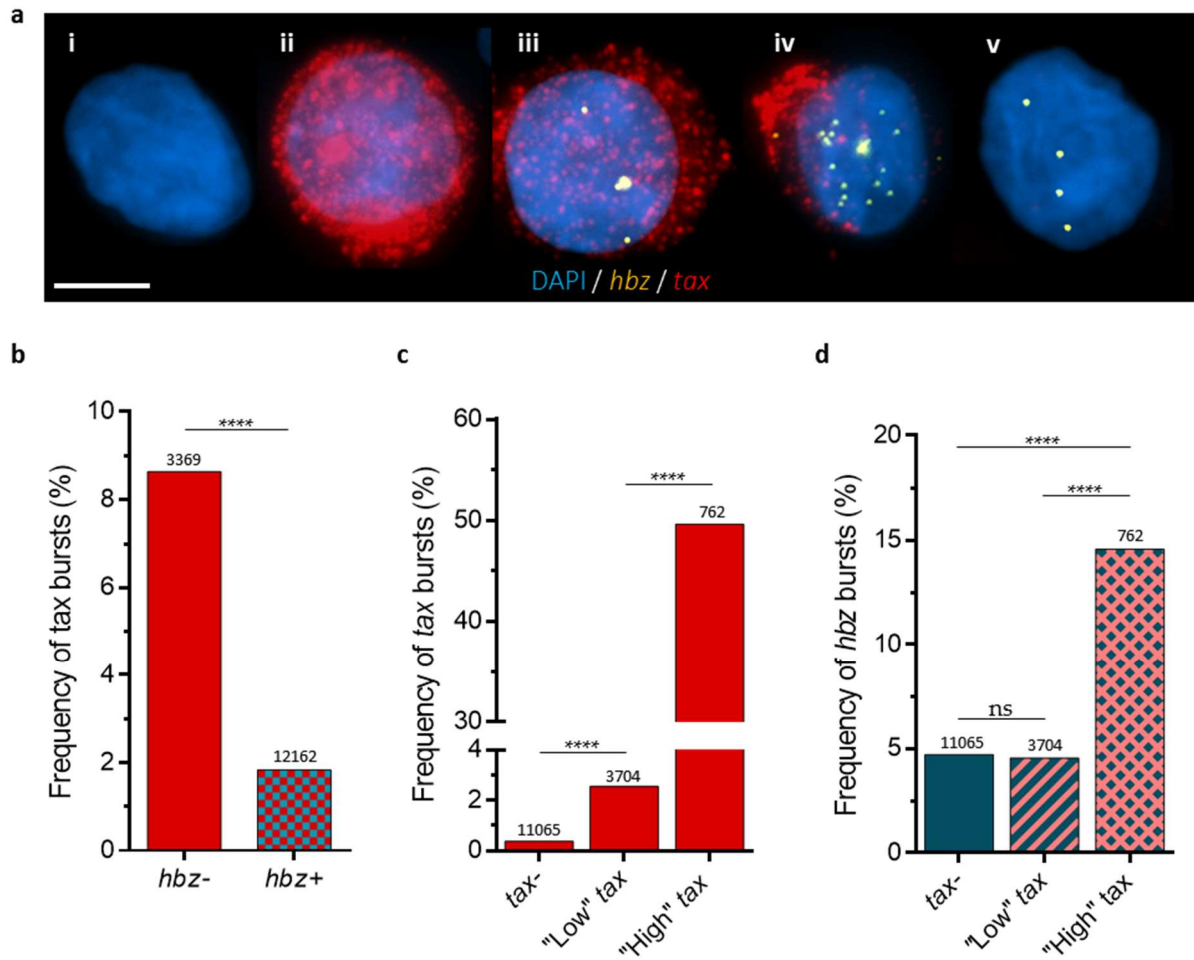


Figure 5-1 | Relationship between *tax* and *hbz* expression in individual cells

a, Representative images of a i) *tax*-silent cell, ii) *tax*-high cell (≥ 100 spots) with a *tax* burst, iii) *tax*-high cell without a *tax* burst (note the presence of an *hbz* burst), iv) *tax*-low cell (< 100 spots) and v) *tax*-silent cell, with low-level *hbz*. Scale bar is 5 μ m. **b**, *tax* bursts occur more frequently in *hbz*-negative cells. **c**, *tax* bursts are indicative of very high *tax* expression, with the proportion of cells which have a burst increasing as the level of *tax* increases. **d**, *hbz* bursts occur more frequently in high-*tax* cells than in *tax*-negative or *tax*-low cells. Significance of differences measured with chi-squared tests. Numbers above columns denote the number of cells in the corresponding population from pooled triplicates of all *tax*-competent clones.

Of the 689 *tax*-low or *tax*-silent cells with *hbz* bursts which were observed, not a single one contained a *tax* burst (Fig. 5-2b). *tax*-high cells which contained a *tax* burst were also approximately half as likely to be *hbz*-positive as cells with similar *tax* spot counts, but lacking a *tax* burst. Furthermore, *hbz* bursts were not evenly distributed throughout *tax*-high cells, instead being found disproportionately in the half of these cells which lacked a *tax* burst: 5% of *tax*-high, *tax* bursting cells also contained an *hbz* burst, while almost 25% of *tax*-high cells lacking a *tax* burst had an *hbz* burst (Fig. 5-2b).

A minority of *tax*-high cells were considered to have “double bursts”, or co-localised spots of both *tax* and *hbz* which were each bright and large enough to pass the transcriptional burst definition, so that these cells could be considered to be actively transcribing both *tax* and *hbz* simultaneously. Given the respective *tax* and *hbz* burst rates in *tax*-high cells of 49% and 15%, one would expect approximately 7-8% of those cells to have both bursts if they occurred independently of each other. However, the observed frequency of so-called double bursts was 2.9% (Fig. 5-2a, $p < 0.0001$ for the difference in *hbz* burst frequency in *tax*-high cells with or without a *tax* burst; chi-squared test). Furthermore, the *tax* bursts in cells which also had an *hbz* burst were estimated to be smaller than those in cells without ($p = 0.0456$), and they had dimmer signals of *gag*, the theoretical indicator of ongoing transcription (data not shown). This strongly suggests the conclusion that, whatever triggers the *hbz* burst, it is rapidly followed by the end of the *tax* burst.

Table 5-1 | *tax*-positive cells are less likely to be *hbz*-positive

		# cells	<i>hbz</i> -positive (%)
All		15531	78.3
<i>tax</i> −		11065	81.6
<i>tax</i> +		4466	70.2
# <i>tax</i>	<i>tax</i> burst		
+	+	136	61.8
++	+	378	36.8
++	−	384	70.6
+	−	3704	74.1

Cells which have *tax* bursts and high-*tax* are significantly less likely to be *hbz*-positive. However, *tax*-high cells without a burst show a distinct increase *hbz* expression.

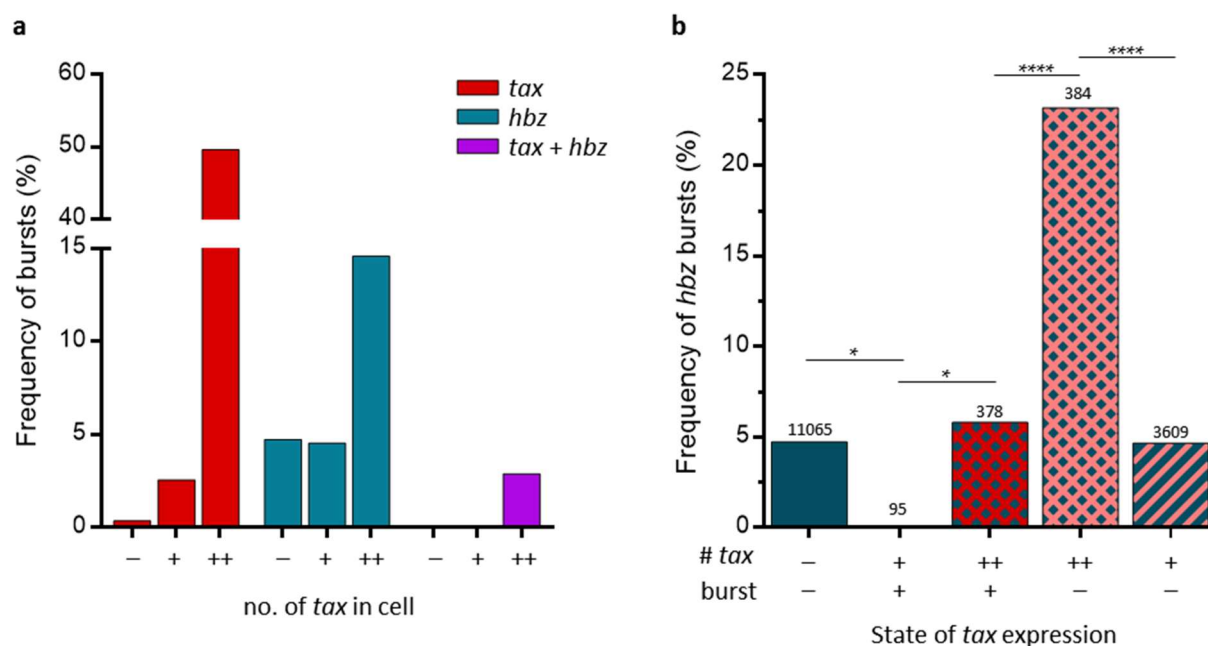


Figure 5-2 | *tax* and *hbx* bursts do not occur independently of each other

a, Co-occurrence of *tax* and *hbx* bursts in *tax*-high cells is far lower than would be expected if they occurred independently of each other. Significance of difference evaluated by chi-squared test. Symbols represent number of *tax* spots detected in a cell: – (0-1 spots, *tax*-silent), + (2-99 spots, *tax*-low) and ++ (>99 spots, *tax*-high). $n = 11,065$ *tax*-silent cells, 3,704 *tax*-low cells and 762 *tax*-high cells from four pooled *tax*-competent clones, with three replicates per clone. **b**, *hbx* bursts are significantly less likely to be observed in cells with a *tax* burst, regardless of the total count of *tax* spots. Significance of difference calculated using chi-squared tests. Number of *tax* per cell represented in the same manner as panel a, with an extra split according to the presence (+) or absence (–) of *tax* bursts. Numbers above each bar show the number of cells in each population, taken from the same clones as panel a.

5.3.2 Cells can be individually cell-cycle staged by microscopy

The cell cycle stage of an individual cell imaged by microscopy can be identified by quantifying the integrated intensity of its DAPI-stained nucleus, which is linearly correlated with DNA content (Roukos *et al.*, 2015). Using the nuclear outlines designed for determining the intracellular distribution of transcripts, it was possible to ascertain the integrated nuclear intensity of each cell whose viral expression was measured (Fig. 5-3a), and convert it into a likely cell cycle stage. The final step was carried out via visual inspection of the resulting histogram of nuclear integrated intensity, based on the principle that cells in G₂/M have double the DNA content, and thus double the integrated intensity of cells in G₀/G₁ – an example of this process is shown in Fig. 5-3b. A similar approach was used in a recent study by Skinner *et al.* (Skinner *et al.*, 2016).

5.3.2.1 Cell cycle profiles are reproducible within clone replicates, and distinct between clones

In total, 18,691 cells were analysed. Three replicates of each clone were cell-cycle staged in this manner, with the exception of clone C, where one replicate was discarded from the analysis as its cells could not be staged with confidence due to technical failure (its presence or absence did not materially affect the results). The cell cycle profiles were reproducible between replicate samples of each respective clone but there were marked differences between clones, with perhaps the most notable being the significantly higher proportion of cells in G₂/M in clone A (Fig. 5-3c).

a, The integrated intensity of nuclei in the DAPI channel is measured and collated, based on the principle that cells in G_2/M (e.g. cell C) have double the DNA content of cells in G_0/G_1 (cells A & B). Scale bar is 5 μm . **b**, Histograms of integrated nuclear intensity from 100s-1000s of cells are visually inspected for each replicate to determine appropriate " G_0/G_1 " boundaries. If there are sufficient cells in " G_2/M " for a readily discernible peak (as above) then the process is repeated, although the lower and upper boundaries of " G_2/M " must be as close as possible to being double the values of the respective " G_0/G_1 " boundaries. If the " G_2/M " peak is not clearly pronounced then the boundaries are set relative to " G_0/G_1 ", using the principle of $2n:4n$. Cells which fall between these two gates are considered to be "S". The histogram shown is from a single replicate of clone A, with 1119 cells. **c**, Using the definitions described above, the proportion of cells in a given cell cycle stage is reproducible between replicates of clones, and different between clones. Mean and SEM of three biological replicates per clone are shown, except for Clone C which only has two replicates due technical failure. $n = 3726$ cells (A), 3519 cells (B), 1961 cells (C), 5691 cells (D), 3946 cells (E).

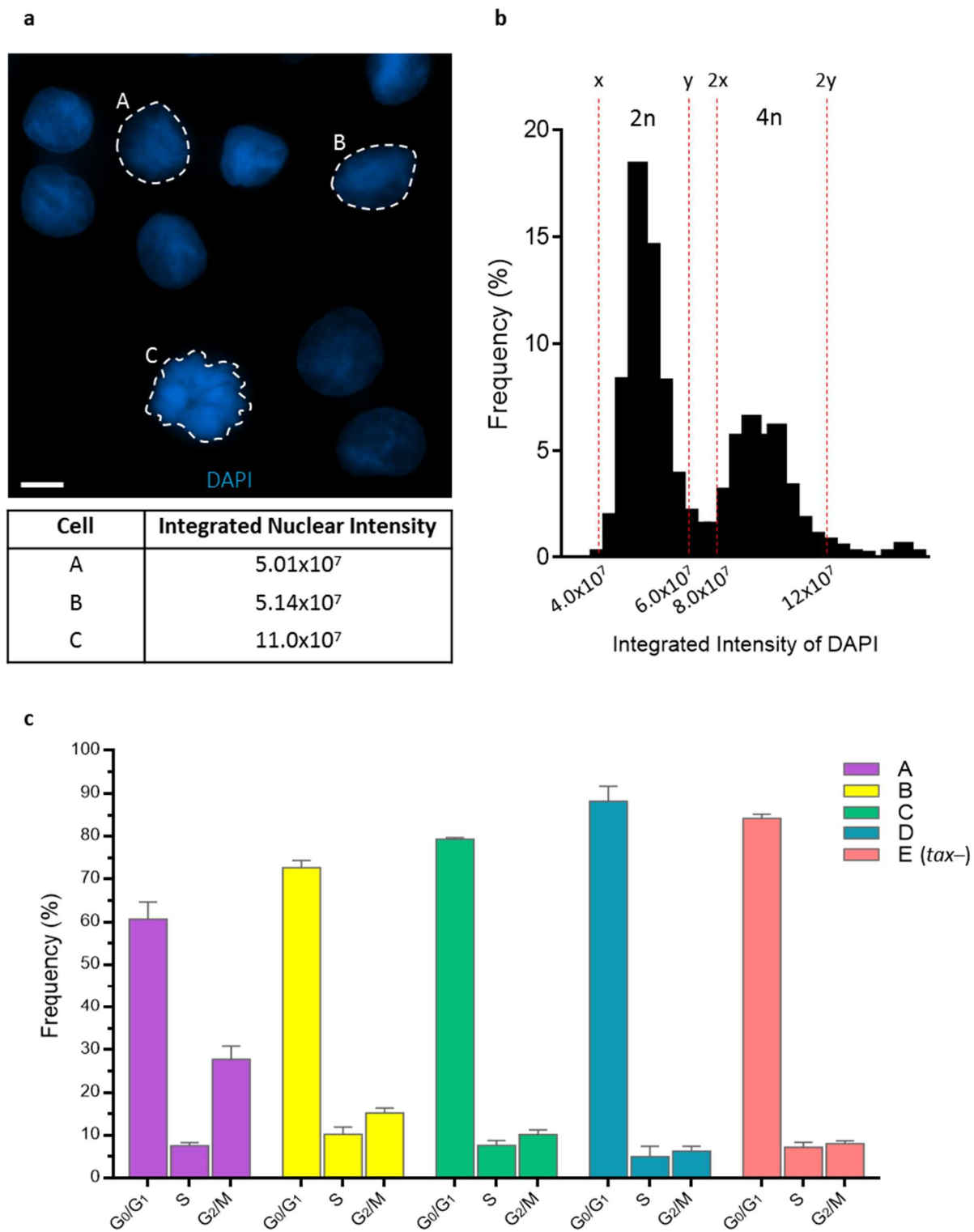


Figure 5-3 | Cell-cycle staging of single cells by fluorescence microscopy

5.3.3 Cycling cells have elevated viral expression

Compared with cells in G_0/G_1 , cells in G_2/M showed a higher intensity of expression (spot counts) for both *tax* and *hbx*. Figs. 5-4a and 5-4c show how the expression profile for both transcripts has shifted to higher expression. The effect is less pronounced in *tax*: the proportion of cells in G_2/M which were *tax*-high was still low. However, since *tax*-high cells contribute over 80% of total *tax*, an increase in *tax*-high cells from 2.6% in G_0/G_1 to 5.6% in G_2/M resulted in a significant increase in total *tax* expression. *tax*-silent and *tax*-low cells were more likely to be in G_0/G_1 , whereas *tax*-high cells were more likely to be in S or G_2/M (Fig. 5-4a), regardless of whether they contained *tax* bursts. This effect was consistent in three of the four *tax*-expressing clones (B-D); in the fourth (clone A) the difference was not significant, most likely due to sampling error from the extremely low frequency of high-*tax* cells. The positive correlation between higher *hbx* and cell-cycle progression was consistent in all clones, including the *tax*-negative clone E (Appendix Fig. 8-5).

While both transcripts were more plentiful in G_2/M in almost all clones, there was more nuance when it came to frequency of expression, in the form of *tax* and *hbx* bursts. The frequency of *hbx* bursts increased in every clone to a significant degree, in some cases almost tripling (Fig. 5-4d). *tax* bursts, on the other hand, were not consistent between clones (Fig. 5-4b). While most increased, this increase was only found to be significant in one clone (clone B), while in clone A there was in fact a decrease in the rate of *tax* bursts in cycling cells which was considered significant, although it is worth repeating that this clone had the lowest frequency of *tax* bursts (47 out of 429 total *tax* bursts).

One confounding factor of *tax* or *hbx* burst incidence in cells in S or G_2/M is the fact that in the former there may be, while in the latter there definitely are two copies of the provirus in the cell. The analysis pipeline involved the definition of bursts as, along with other parameters, the single brightest intranuclear “spot” in a given cell; multiple bursts in a nucleus were not accepted (section 2.3.3 in Materials & Methods, “FISH-Quant analysis”; Appendix Fig. 8-1). This would not affect the burst frequency as it is presented here, as the proportion of cells which have a *tax* or *hbx* burst would not change even if some of them in fact had two bursts. It may, however, affect the frequency of transcription from the perspective of the provirus; however, the ambiguity involved means that it could be either under- or overestimated (see Discussion, section 7.6, “Viral transcriptional bursts”).

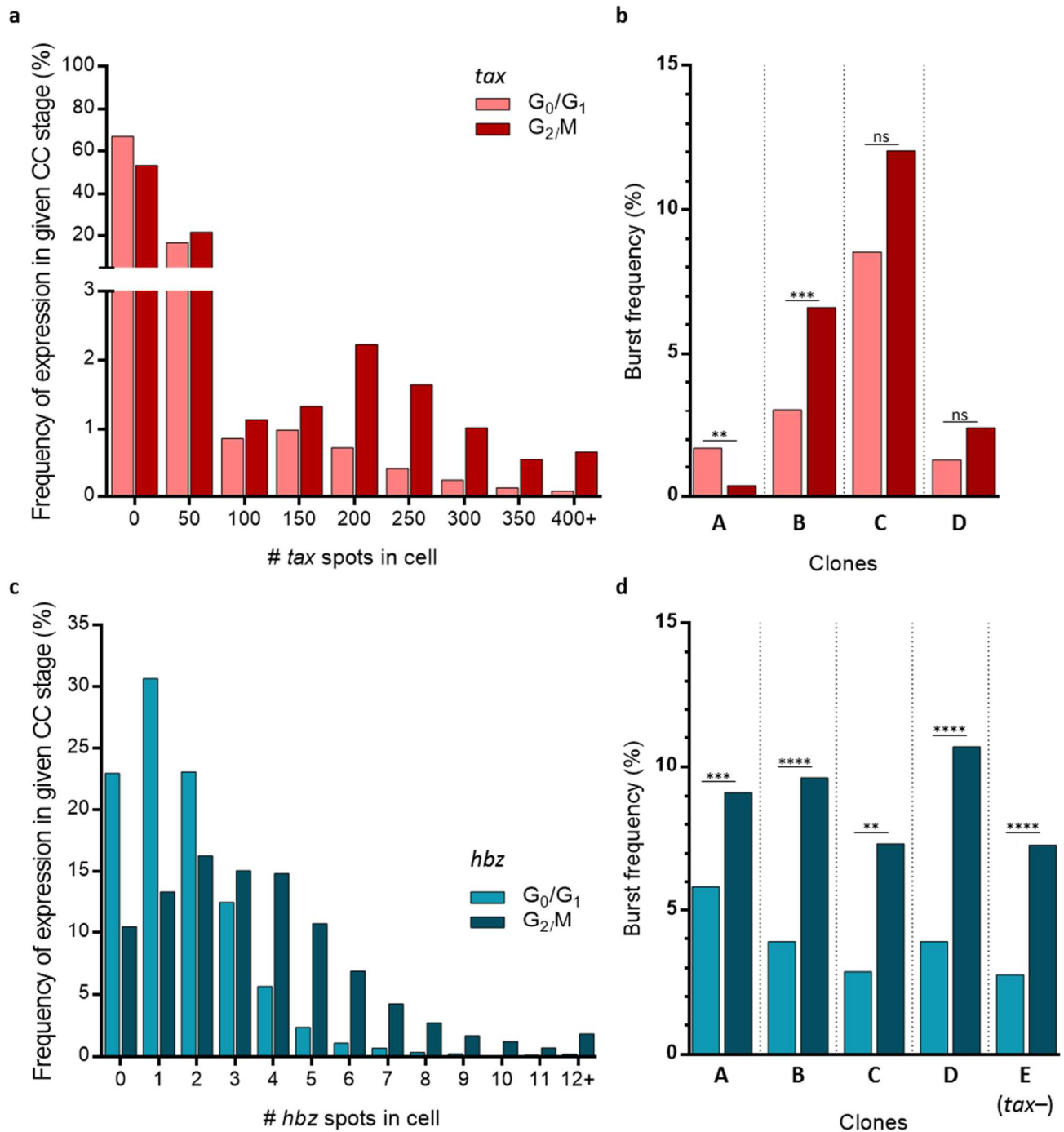


Figure 5-4 | Viral expression increases in cycling cells

a, Cells in G_2/M are on average more frequently *tax*-positive and express higher levels of *tax* than do cells in G_0/G_1 . Histogram bars represent the proportion of cells with a given cell cycle stage which have a given level of viral expression. **b**, The relationship between cell cycle stage and *tax* burst frequency is not consistent between clones. **c**, Cells in G_2/M are more frequently *hbz*+ and express higher levels of *hbz* than do cells in G_0/G_1 . **d**, *hbz* burst frequency consistently increases in cycling cells in all clones. Five clones were pooled for this analysis, of which four had three biological replicate samples each, and one (C) had two biological replicates; total $n = 18,843$ cells.

5.3.4 High *tax* and *hbz* is associated with cell cycle progression

Given that cells showed differences in viral expression based on their cell cycle stage, the question arose of what differences could be seen in the cell cycle distributions of cells dependent on their level of viral expression. Cells were split into “silent”, “low” and “high” bins for both *tax* and *hbz*, along with the presence or absence of a corresponding transcriptional burst (Fig. 5-5).

Depending on their level of *tax* or *hbz* expression, there were remarkable differences in the cell cycle distribution of cells. Approximately 35% of cells with high *tax* were in S or G₂/M, compared to 19% in *tax*-silent cells. The proportion of cells which were cycling in *tax*-low cells was comparable to them being silent, while it was slightly raised in *tax*-low cells which lacked a burst (Fig. 5-5a). The correlation with between *tax*-high cells and G₂/M was strongly consistent in two of three clones capable of expressing *tax* to the level defined as “high” in this thesis, while in the third (clone A) the effect was again not significant, possibly due to sampling error.

hbz showed an even stronger association with the cell cycle (Fig. 5-5b). 84% of *hbz*-silent cells were in G₀/G₁, compared to 67% in *hbz*-low, *hbz* burst+ cells, falling further to 40% in cells with high *hbz*. Cells which were *hbz*-low, *hbz* burst-, however, had a much more similar cell cycle distribution to *hbz*-negative cells, with 81% of cells in G₀/G₁. Put another way, the average proportion of cells in S or G₂/M in all clones, regardless of viral expression, was 13%. In *hbz*-high cells, however, more than half were in S or G₂/M, and therefore progressing through the cell cycle.

Given that high *Tax* can lead to expression of *HBZ* (Yoshida *et al.*, 2008), as well as the relationship between high *tax* and *hbz* expression described earlier in this chapter, the question arose whether *hbz* expression was independently correlated with the cell cycle, or correlated with *tax*. To address this, a logistic model was created to quantify the association between *tax* and *hbz* count, *tax* and *hbz* bursts and cell cycle stage; how robustly one variable could be predicted in an individual cell, given the state of the other variables.

According to this model, both high *tax* and the cell cycle stage were independently correlated with *hbz* bursts (Fig. 5-6). Between cells with the same cell cycle stage, the odds of an *hbz* burst increased by an average factor of 1.008 for every *tax* spot detected, and by 5.8-fold if they had high *tax*. This effect was repeated in clone D, even though it had a lower maximum *tax* expression. Similarly, the probability of any cell with a given level of *tax* (high, low and silent) containing an *hbz* burst increased by an average of 1.5-fold if it was in G₂/M rather than G₀/G₁ (Table 5-2 and Fig. 5-6). Crucially, this effect was repeated in clone E, which is of course incapable of expressing *tax*.

Table 5-2 | Frequency of *hbz* bursts according to cell-cycle stage and level of *tax* expression

# <i>tax</i>	<i>hbz</i> burst %		
	All	G ₀ /G ₁	G ₂ /M
–	4.42	3.53	8.40
2-99	4.56	3.46	7.99
>99	14.7	14.1	18.4

Cells in a given cell-cycle stage have more *hbz* bursts if they are *tax*-high, and cells with comparable levels of *tax* have more *hbz* bursts if they are in G₂/M.

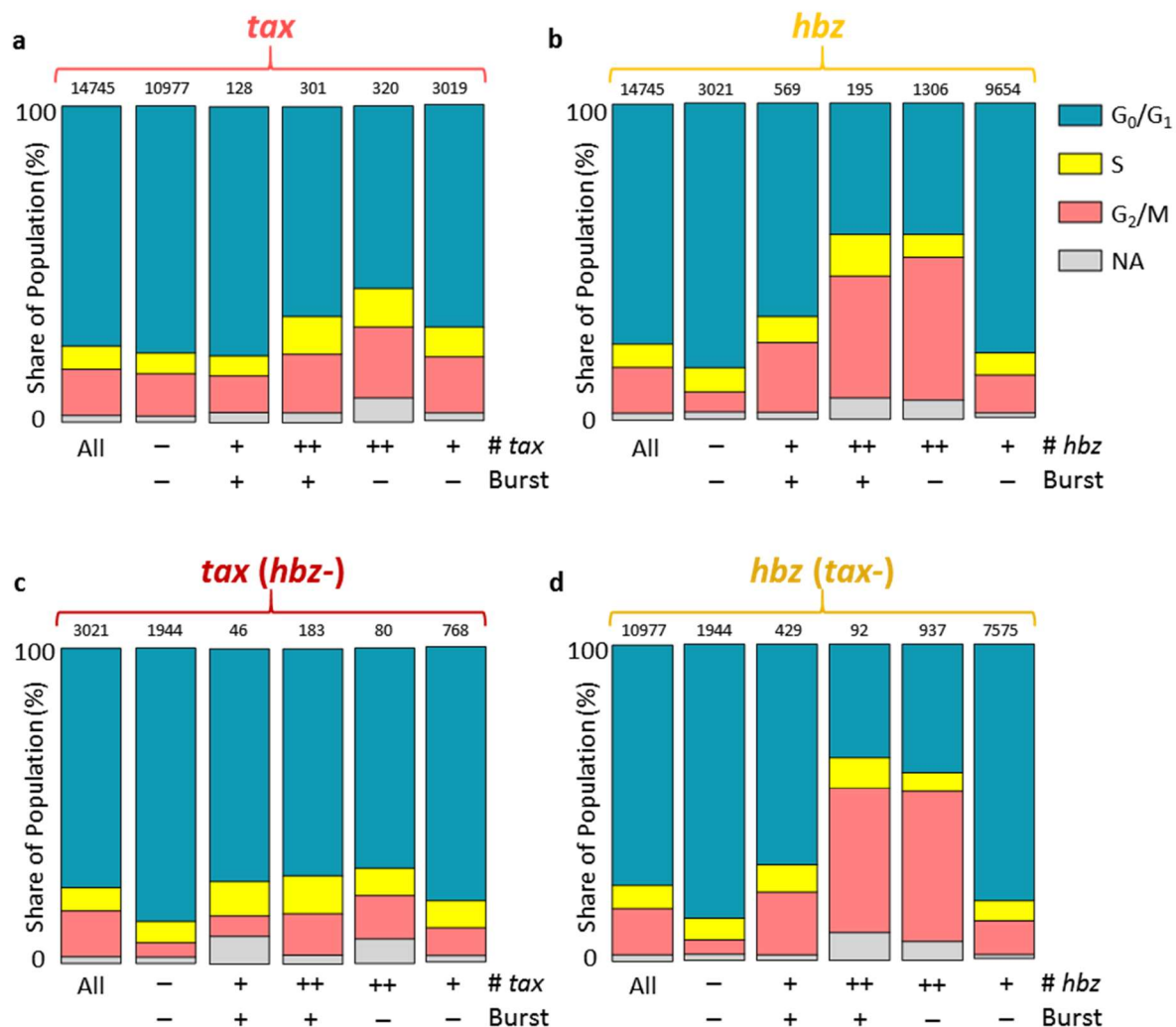


Figure 5-5 | Cell cycle distributions of cells according to their level of viral expression

The cell cycle distributions of cells with different levels of viral expression are represented by the proportion of the columns they occupy. For example, across all cells (top left bar) the cell cycle distribution is 77.5% G_0/G_1 , 7.55% S, 13.2% G_2/M and 1.75% NA. NA (not applicable) refers to cells whose nuclear integrated intensity was too bright to be binned as G_2/M , or too dim to be binned as G_0/G_1 . **a**, Cells with high-*tax* have a higher frequency of cells in S and G_2/M , regardless of whether or not a burst is present. Cells with low *tax* and no burst also have a slightly higher rate of cells in G_2/M . **b**, *hbz* expression is correlated with cell cycle progression, both *hbz* bursts and high-*hbz*. **c**, The increased frequency of G_2/M in *tax*-high cells is lost when cells which are also *hbz*-positive are excluded. **d**, The association between *hbz* and cell cycle progression is largely unchanged in the absence of *tax*, although there is a fall in the proportion of high-*hbz* cells in S-phase. Numbers above each column denote the total number of cells with that level of expression. $n = 14,475$ cells from all *tax*-competent clones, from three replicates each with the exception of clone C, where one replicate was discarded due to technical failure.

Using the same model to investigate the importance of viral expression to the cell cycle, *hbz* was again found to have a much stronger link to G₂/M than *tax*. Each *hbz* spot present in a cell increased its odds of being in G₂/M as opposed to G₀/G₁ by a factor of 1.45, while the same effect for *tax* was only 1.004. Of course, *tax* expression has a far higher ceiling than *hbz*, but even so *tax*-high cells were only 1.58-fold more likely than *tax*-low or silent cells to be in G₂/M, compared to a 7.3-fold increase in the odds if the cells were *hbz*-high (Fig. 5-6). Similar effects were seen with regards to increasing the probability of S-phase, but given the firmer definition of G₂/M in this study, and the fact that cells in S-phase can be expected to reach G₂/M eventually anyway, the conclusion which can be drawn from G₂/M seem more concrete.

One of the key objectives of a logistic model is to quantify the individual contribution of the multitude of input variables on the output variable. This is crucial in the context of *tax* and *hbz* given that Tax protein, and both *hbz* RNA and HBZ protein have been noted for their proliferative, transformative effects, and that they do not exist independently of each other. Another way to ask this question is to observe the cell cycle distributions of cells which have exclusively one of the two transcripts. In other words, seeing how cell cycle distributions change with *tax* levels in *hbz*-silent cells, and how they change with *hbz* levels in *tax*-silent cells (Fig. 5-5c, d).

The first observation worth making is that cells which are entirely devoid of viral transcripts are the least likely to be cycling, with 87% in G₀/G₁. Beyond that, a very interesting effect is that the increased proportion of cells in S-phase visible in both *tax*-high and *hbz*-high cells in Figs. 5-5a and 5-5b appear to be limited largely to *tax+hbz*- cells, whereas the increase in G₂/M seen in *tax*-high cells appears to vanish when *hbz*-positive cells are excluded. In contrast, the strong association between *hbz* and G₂/M remains even if *tax*-positive cells are excluded, although there may be a weakening of the association with S-phase. The positive correlation between *hbz* bursts or high *hbz* and G₂/M was maintained even in cells which had high *tax* (Fig. 5-5c, d).

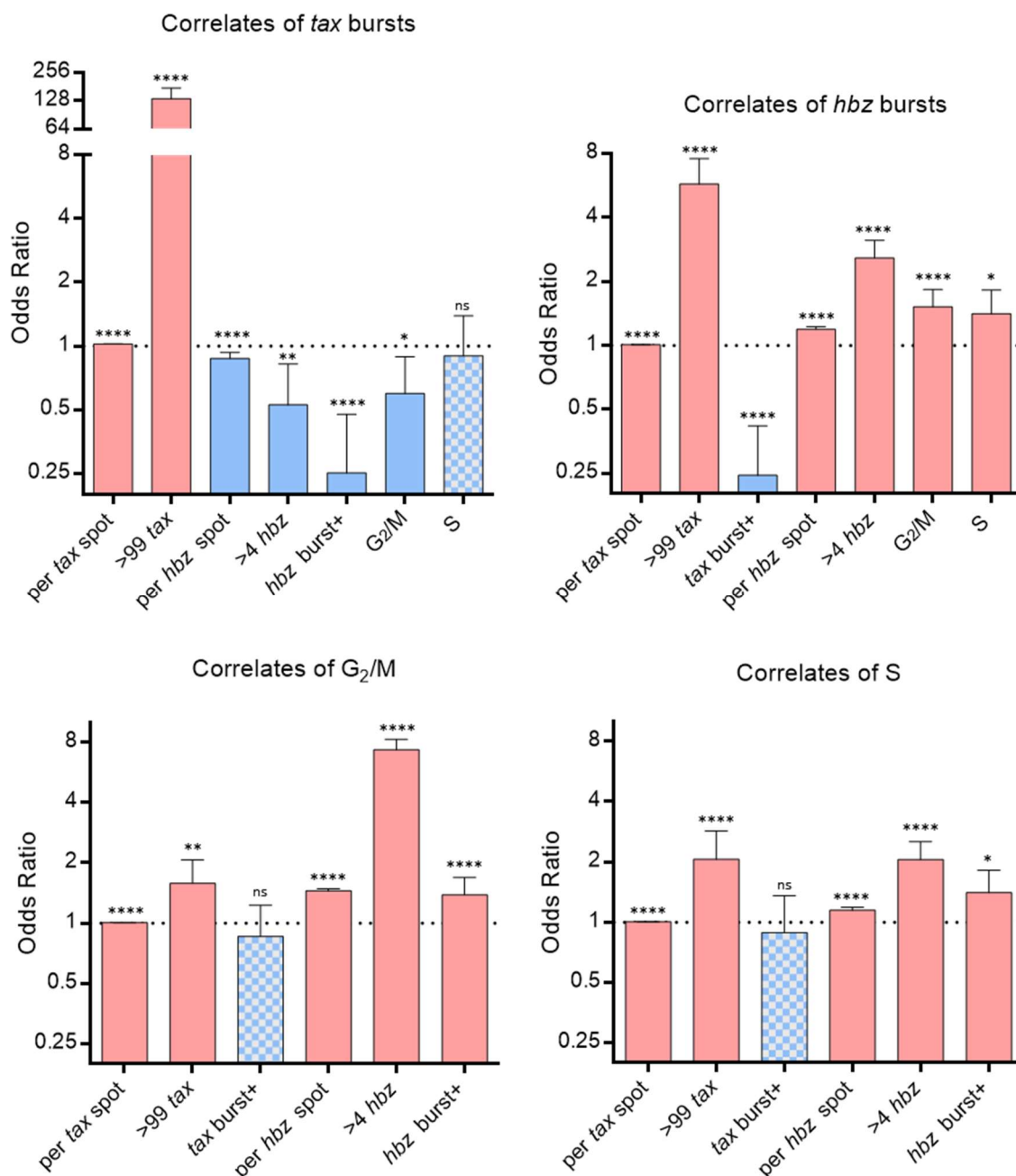


Figure 5-6 | Odds ratios from logistic regression of viral expression and cell cycle stage

An odds ratio above 1 indicates that the presence or increase of that variable has a positive predictive effect on the likelihood of the outcome variable. For example, each *tax* spot in a cell increases the odds of that cell also having a *tax* burst by a factor of 1.024, or 2.4%, while a cell being *tax*-high increases the odds of it containing a *tax* burst by a factor of 135. Columns signify average odds ratios, while bars mark 95% confidence intervals – if they cross the 1.0 line, the effect is not considered significant. Positive (>1) odds ratios are shaded red, while negative (<1) ratios are shaded blue, and non-significant effects are chequered blue. Odds ratios shown are from the same cells as in Fig. 5-5; n = 14,475 cells.

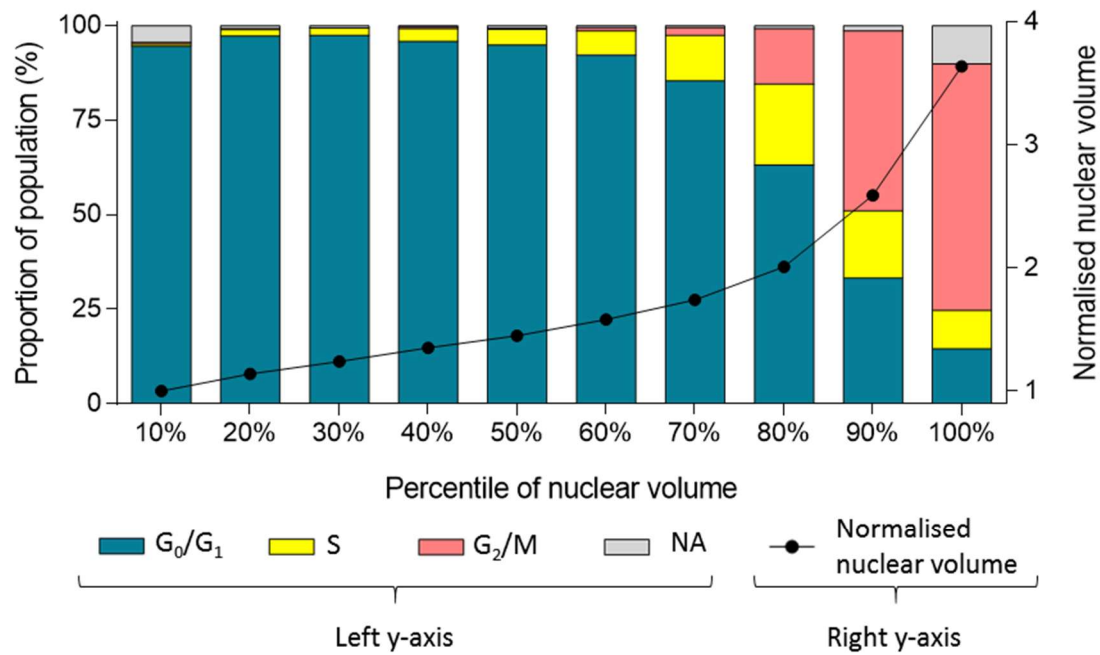
While fixation precludes any kind of robust temporal analysis, an approximation can be made based on transcript numbers and the presence or absence of transcriptional bursts for the purpose of inferring temporal correlation with the cell cycle. For example, a cell with an *hbz* burst and two *hbz* spots is likely to have started expressing *hbz* more recently than a cell with a burst and eight spots, which in turn is likely to have begun expressing *hbz* more recently than a cell with eight *hbz* spots and no *hbz* burst. In this context, and given that *hbz* expression is more likely in cells with high *tax*, an interpretation appears for the cause of abrogation of the increased frequency of G₂/M in *tax*-high cells disappears once *hbz*-positive cells are excluded. Namely, that by the time these *tax*-high cells are in G₂/M, they are also *hbz*-positive. This suggests another, cell-cycle related aspect which can be added to the model of viral expression described earlier.

5.3.5 Cell volume is strongly associated with cell cycle stage

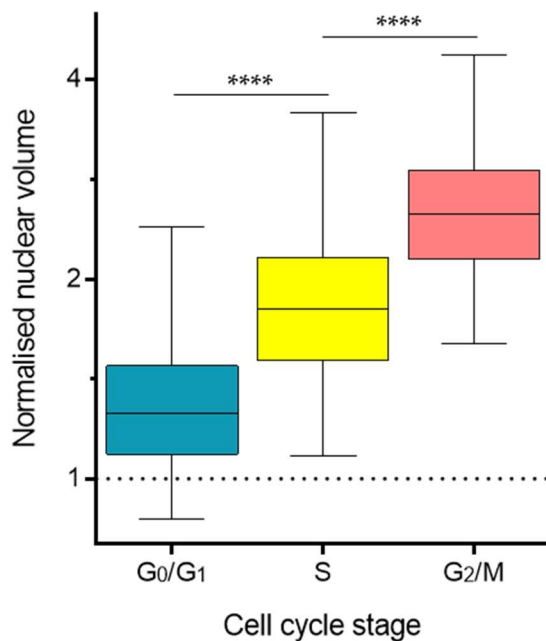
A number of studies have attempted to account for cell-to-cell heterogeneity of expression by accounting for cell cycle stage and cell volume (Battich, Stoeger and Pelkmans, 2015; Padovan-Merhar *et al.*, 2015). Using my analysis pipeline, cell volume can be approximated through the area of each cell's outline as drawn by FISH-Quant, with an assumption that cells and their nuclei are approximately spherical (see section 2.3.3, "FISH-Quant analysis", for a description of the outline drawing process). These measurements are not wholly reliable, as no cell surface marker was used, meaning cell boundaries had to be defined using the cells' autofluorescence. They were also often defined simply as an extension of the nucleus, which did have clearly visible dimensions (Tsanov *et al.*, 2016). I therefore decided to rank cells by their nuclear volume; once they were broken into deciles, a remarkable pattern emerged.

Nuclear volume increased at a constant rate among cells with the lowest 70% of volumes (Fig. 5-7a), before increasing more rapidly in the top 30% of cells. The heterogeneity in nuclear sizes was remarkable: the 10% largest cells were approximately 3.6-fold larger than the smallest 10%. A number of measurements correlated with nuclear volume, but perhaps the most noticeable was the fact that cycling cells were almost entirely absent in the lower 70% of nuclear volumes, while cells in S or G₂/M constituted three quarters of cells with the 10% largest nuclei. As Fig. 5-7b demonstrates, cells in G₂/M are on average significantly larger than cells classified as being in S-phase, who in turn are on average significantly larger than cells deemed to be in G₀/G₁.

a



b



a, Ranking cells by their estimated nuclear volume and splitting them into ten bins, size increases at a constant rate in the lower 70% of volumes, before increasing further in the upper 30%. Cells in S and G₂/M are significantly overrepresented in cells with the 30% highest volumes. n = 18,843 total cells (1843-1884 per decile) from pooled clones. 3 replicates from each clone except clone C, which has two. Volumes normalised to the average volume of the smallest 10% of nuclei, estimated to be 170 μm^3 .

b, Cells in S and G₂/M are significantly larger than cells in G₀/G₁. Boxes show mean and interquartile range while whiskers denote the 2.5-97.5% confidence intervals. Significance measured by Mann-Whitney test. Same cells used as in panel a.

Figure 5-7 | Cell in S and G₂/M phase are significantly larger than those in G₀/G₁

5.3.6 *hbz* expression and *tax* bursts correlate with nuclear volume

A number of measurements of viral expression were also found to increase with nuclear volume. *tax* counts per cell correlates with cellular volume, but not with nuclear volume (Fig. 5.8a). However, this cannot be taken at face value given that cell volume appears to be the determining factor in what proportion of detected *tax* spots are considered in the final count (see section 3.3.4, “*hbz* count is highly accurate, whereas *tax* count underestimates the content of high-expressing cells”). That being said, there was a noticeable increase in the proportion of cells with high *tax* or *tax* bursts in cells with the largest nuclei, and both were heavily skewed toward the diminishing number of G₀/G₁ cells in those high-volume bins (Table 5-3). Of note, *tax* bursts were on average larger in cells with the largest nuclei (Fig. 5.8b), but once cell size was accounted for, *tax* burst frequency actually decreased from cells in G₀/G₁ to G₂/M (Table 5-3).

In contrast, the accuracy of *hbz* measurement in larger or smaller cells is not in question, and the *hbz* count per *hbz*-positive cell did increase in step with increasing nuclear volume (Fig. 5-8a). This correlation seemed to follow a constant, suggesting that absolute *hbz* abundance mattered less than the concentration of *hbz* in a cell. It is also possible that, given *hbz*'s retention in the nucleus, nuclear concentration is the most important metric. This association was clear in all clones, including the *tax*-negative clone E. This increasing nuclear volume in larger cells was in contrast to a recent study on the subject (Padovan-Merhar *et al.*, 2015), so *hbz* abundance per nuclear volume also remained largely constant.

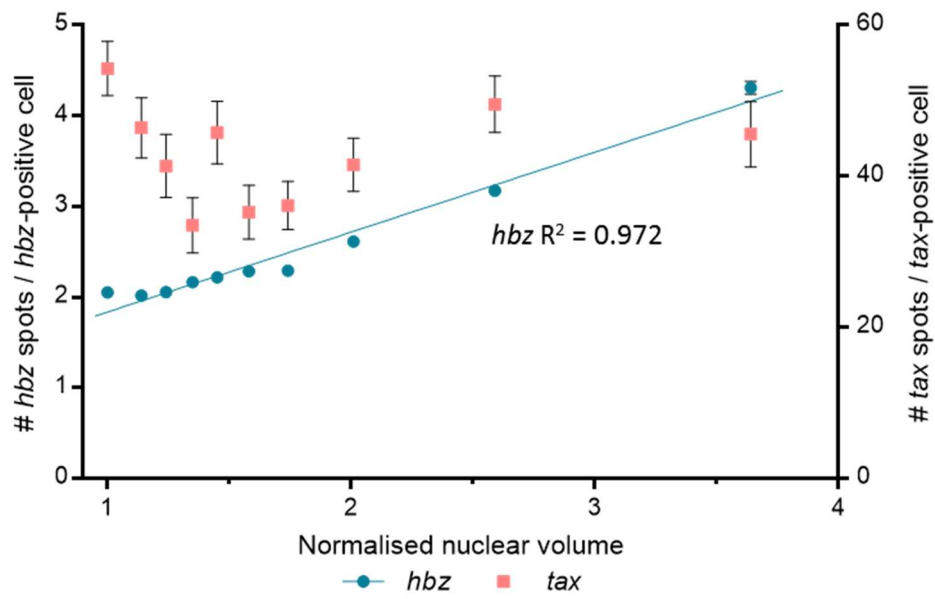
Given this new data, it became necessary to reanalyse the relationship between HTLV-1 plus- and minus-strand expression and the cell cycle by logistic regression, with cell volume as an added variable. While cell volume was found to be significantly positively correlated with a cell's *hbz* abundance, it was not significantly correlated with the likelihood of an *hbz* burst. The significance of other variables in predicting a cell's cell cycle stage or state of HTLV-1 expression remained unchanged when cell volume was factored in, although the magnitude of some correlations was reduced. *hbz* burst frequency and transcript abundance per cell were both higher in cells in G₂/M as opposed to G₀/G₁. However, in neither case did they increase by a factor of two, as might be expected if the two proviruses present in G₂ cells expressed at equal rates, independently of each other: for *hbz* bursts, the frequency increase had a ratio of 1.67, while for overall *hbz* abundance the ratio was 1.34 (Table 5-3). This reinforces the view of HTLV-1 expression as being determined by a complex array of factors, which will require further investigation (see Discussion, section 7.8, “Relationship between viral expression and the cell cycle”).

a, Ranking cells by their estimated nuclear volume and splitting them into ten bins, *hbz* abundance per *hbz*-positive cell increases at a constant rate dependent on nuclear volume. No similar trend is clear in the case of *tax*.

b, The average size of *tax* bursts increases in with nuclear volume. There is no such change in the size of *hbz* bursts in cells with larger nuclei.

Mean and SEM are shown. R^2 of each linear regression is shown, along with the probability that the regression's gradient's deviation from zero is significant. Nuclear volume binning based on same definitions as Fig. 5-7, with all *hbz*, *tax* or burst-positive cells selected from those bins as appropriate. $n = 15,536$ *hbz*-positive cells; 4,490 *tax*-positive cells, 942 cells with *hbz* bursts and 514 cells with *tax* bursts.

a Number of transcripts per cell by volume



b Size of bursts by volume

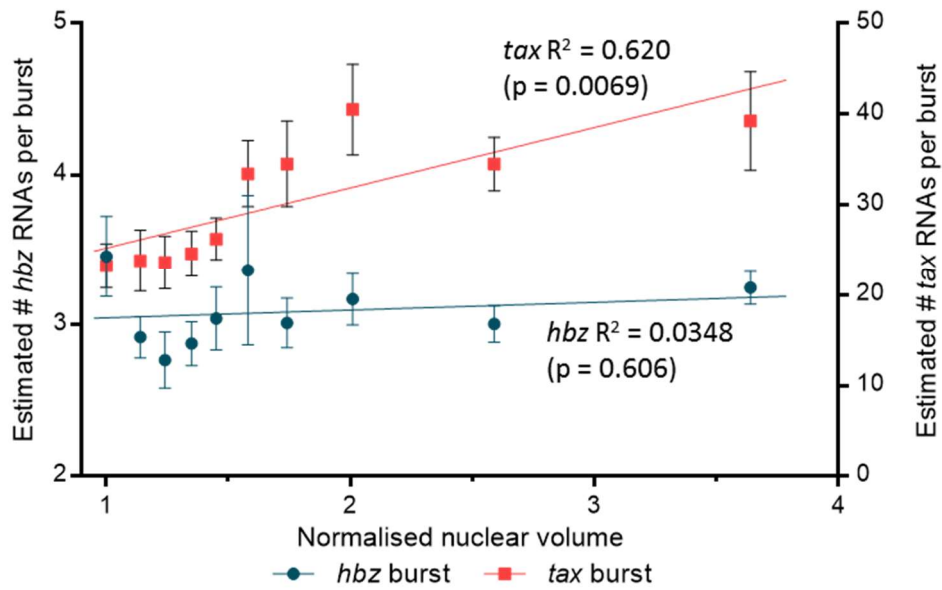


Figure 5-8 | *hbz* expression per cell is largely consistent as a function of nuclear volume, and *tax* bursts are larger in cells with larger nuclei

Table 5-3 | *tax* expression increases significantly with cell volume

Top 20% cells by nuclear volume	G ₀ /G ₁	S	G ₂ /M
<i>tax burst</i> %	7.60	3.87	3.50
<i>tax</i> - high %	13.1	7.93	6.51
<i>hbz burst</i> %	5.67	10.1	9.46
<i>hbz per hbz</i> + cell	3.09	3.38	4.16

While overall *tax* expression and *tax* burst frequency increases markedly in the largest cells, its profile is affected significantly by *hbz* and cell cycle stage.

5.4 Conclusions

The work presented in this chapter focused on the relationship between expression of the viral plus- and minus-strand transcripts, as well as their relationship with the cell cycle. The main conclusions are outlined below:

- *tax* and *hbz* expression are not independent of each other
 - *tax* expression in the form of bursts are significantly more likely to be found in *hbz*-silent cells, a minority of the total cell population
 - *tax* bursts are associated with very high total *tax* spot counts, indicative of a positive feedback loop of expression
 - *hbz* expression in the form of bursts is significantly more likely in *tax*-high cells, suggesting that the positive feedback loop of *tax* expression may be broken by *hbz*
 - I infer that *tax* bursts end very soon after the appearance of *hbz* bursts
- It has become possible to discern the cell cycle stage of individual cells through fluorescence microscopy
- *tax* and *hbz* expression display different degrees of correlation with cell cycle progression
 - Cells in G₂/M tend to have higher levels of viral expression of both transcripts, when compared to cells in G₀/G₁
 - *hbz* bursts are more frequent in G₂/M compared to G₀/G₁ in each clone examined, while the relationship between *tax* bursts and the cell cycle varies between clones
 - Cells with high *hbz* expression or *hbz* bursts are highly significantly more likely to be in G₂/M or S than G₀/G₁
 - While high *tax* is significantly associated with cell cycle progression (i.e. not being in G₀/G₁), *tax* bursts are not
 - The lack of *tax*-high, *hbz*-silent cells in G₂/M suggests that *hbz* has greater influence on the cell cycle than *tax*
- Nuclei of cells in S and G₂/M are significantly larger than those in G₀/G₁
 - In *hbz*-positive cells, *hbz* expression as a function of cellular or nuclear volume remains constant
 - *hbz* burst frequency is not correlated with nuclear volume, but is associated with cell-cycle stage progression
 - *tax* bursts are on average larger in cells with larger nuclei

CHAPTER SIX

LOCATING THE HTLV-1 PROVIRUS

USING DNA-FISH

6.1 Chapter Summary

The previous chapters have all centred around RNA-FISH. The work described in this chapter is centred around visualising and quantifying the spatial distribution of the HTLV-1 provirus itself within the nuclei of infected cells. This was based on using DNA-FISH to target the integrated HTLV-1 genome, or stretches of host genome flanking the provirus of a clone whose UIS had been mapped by LM-PCR. It could then be combined with immunofluorescence in order to quantify the 3D localisation of the provirus relative to intranuclear compartments, such as the nuclear lamina or nucleolus, with which the provirus might interact. While initially successful, ultimately this approach was put aside due to technical setbacks and the rapid progress of the RNA-FISH work described in the previous chapters. In the long run, however, there remains value in mapping the 3D localisation of the HTLV-1 provirus, particularly to investigate how the localisation correlates with proviral expression.

6.2 Introduction

It has recently become apparent that DNA is not randomly distributed in the interphase nucleus, but rather is organised into largely discrete “chromosome territories”, as well as associating with particular intranuclear structures, for example the nuclear lamina, dependent on whether genes in a given genomic neighbourhood are being expressed or repressed (Bickmore and van Steensel, 2013; Gibcus and Dekker, 2013). As a majority of naturally infected T-cells contain a single provirus integrated at a unique genomic location (Cook *et al.*, 2012), the HTLV-1 provirus can find itself in very different genomic environments.

While HTLV-1 integration does have some association with a selection of genomic landmarks (Gillet *et al.*, 2011; Melamed *et al.*, 2013), there is no particular integration preference between chromosomes: the number of UIS found in a given chromosome in cells infected *in vitro* scales with the size of the chromosome (Cook *et al.*, 2014). However, the same paper which demonstrated this also reported a significant *in vivo* preference for chromosomes 13, 14, 15 and 21. In other words, while there was no bias in the number of proviruses which were integrated into these chromosomes in the absence of immune pressure, cells whose provirus was integrated in one of these chromosomes were more likely to survive. Along with chromosome 22, these are the acrocentric chromosomes, whose very short p-arms contain the nucleolar organising regions (NORs), arrays of rDNA codons coding for the ribosomal machinery (Németh *et al.*, 2010; Németh and Längst, 2011). None of the proviruses were mapped to the p-arms and centromeres block long-range interactions between the arms of chromosomes (Bickmore and van Steensel, 2013), indicating that direct interactions with the rDNA arrays were not responsible for the observed survival bias. Another possibility was that there could be an interaction with the nucleolus of the infected cells which increased the probability of survival of the virus.

There is also previous evidence that interactions between proviruses with nuclear bodies can result in modulation of proviral expression. For example, Lusic *et al.* discovered that the HIV-1 provirus spatially associates with PML bodies (named after the promyelocytic leukaemia factor), and that this association favours proviral latency (Lusic *et al.*, 2013).

The nucleolus is the centre of ribosome biogenesis, and is bounded by a perimeter of perinucleolar heterochromatin, associated with the epigenetic mark H3K9me3 (Németh *et al.*, 2010; Németh and Längst, 2011). This mark is strongly associated with constitutive repression, and is also found in the nuclear lamina (Bickmore and van Steensel, 2013; Wendt and Grosveld, 2014). A provirus in such an environment might therefore remain latent and avoid immune detection. However, the nucleolus disperses during mitosis, and can also be disrupted by cellular stressors such as hypoxia or nutrient

deprivation (Boisvert *et al.*, 2007; Boulon *et al.*, 2010). This nucleolar dispersal could in turn allow intermittent viral expression, linked to cell cycle progression or cellular stress, decreasing the probability of immune detection and destruction. I therefore set out to use DNA-FISH to find the intranuclear location of HTLV-1 proviruses in order to test the hypothesis that preferential survival of infected cells through controlled proviral expression could be brought about through its association with the nucleolus, as well as the more general investigation of associations with other intranuclear compartments.

The initial goal was to work with infected clones before moving on to unsorted PBMCs. Starting with the former allowed simpler validation as the proviral integration site would be known, and since all cells in a clone are infected it would be possible to calculate a probe efficiency rate. Conversely, beginning with PBMCs, where not all cells are infected, and the proportion of PVL varies between patients, would always leave open the question of whether any particular cell lacked a proviral signal because the probe had failed to hybridise, or because the cell was uninfected.

6.3 Results

6.3.1 Targeting integration sites by flanking region

The principle of RNA- and DNA-FISH is the same: design probes whose sequence is complementary and specific to a particular sequence which you wish to visualise, and label those probes with a fluorophore. However, DNA-FISH is a much more challenging technique than RNA-FISH. With RNA-FISH, one must be exceedingly careful to avoid RNase contamination, but thermodynamically speaking it is much easier to melt a short single-stranded RNA secondary structure than the base pairing of an entire double-stranded DNA genome. Similarly, the FISH probes are only competing with the transcript's self-hybridisation in the case of RNA-FISH, whereas for DNA-FISH they are competing against the entire complementary strand. To overcome this, DNA-FISH is by necessity a much harsher technique than RNA-FISH, but DNA-FISH probes also generally need to target a longer sequence in order to guarantee sufficient signal-to-noise.

At the beginning of this project, a sequence as short as 9 kb (the length of the HTLV-1 genome) had not been successfully imaged before using DNA-FISH (Boyle *et al.*, 2011). However, since the integration site of the clones used in this study had been mapped by LM-PCR (Table 2.1.2.1), it was possible to target the human genome flanking each UIS. Whilst initially an acceptable workaround, as target length would no longer be an issue, this approach would not be sustainable as it would require a new, prohibitively expensive probe design for every clone to be studied. The location data from these probes would also not be as robust as one which directly targeted the proviral sequence, since each chromosome would produce a signal but only one would contain the provirus.

100 kb of the flanking genomes of two clones (A and C) were targeted as a proof-of-principle. A common method to validate DNA-FISH probe specificity involves hybridising cells with both the probe and what is known as chromosome “paint”, a commercially available probe library which hybridises with entire chromosomes. If the probe is specific, its signal will consistently co-localise with that of the chromosome in which its target is found. A further validation can be carried out using metaphase- or M-FISH. In this case, cells are synchronised in mitosis by colcemid treatment before fixation. They are then dropped onto glass slides, and the pressure of impact cracks the cells, allowing the condensed chromosomes to stick to the glass (see section 2.5.6, “Metaphase spreading and metaphase-FISH”). The probes and chromosome paints are then added without any need to permeabilise the cell, and colocalisation can be observed with a readily recognisable condensed chromosome instead of a diffuse interphase chromosome territory.

Figure 6-1 shows the results of both of these approaches. While metaphase spreads do produce lovely images, the FISH probes exhibited extensive non-specific signals in this context (Fig. 6-1b). It was however also possible with this approach to judge visually whether the position of the probe signal on the chromosome corresponded to that of the known integration site. Simply co-hybridising with chromosome paint with normal cells demonstrated that the probes were successful, with an excellent signal-to-noise ratio (Fig. 6-1a). While not all cells contained two distinct, specific signals, once the protocol was optimised roughly 90% could be expected to contain at least one signal which was readily identifiable. Once this had been demonstrated, these probes could also be used as further validation for the successful binding of the provirus-specific probe which was being developed in parallel (see section 6.3.2, "Targeting the provirus").

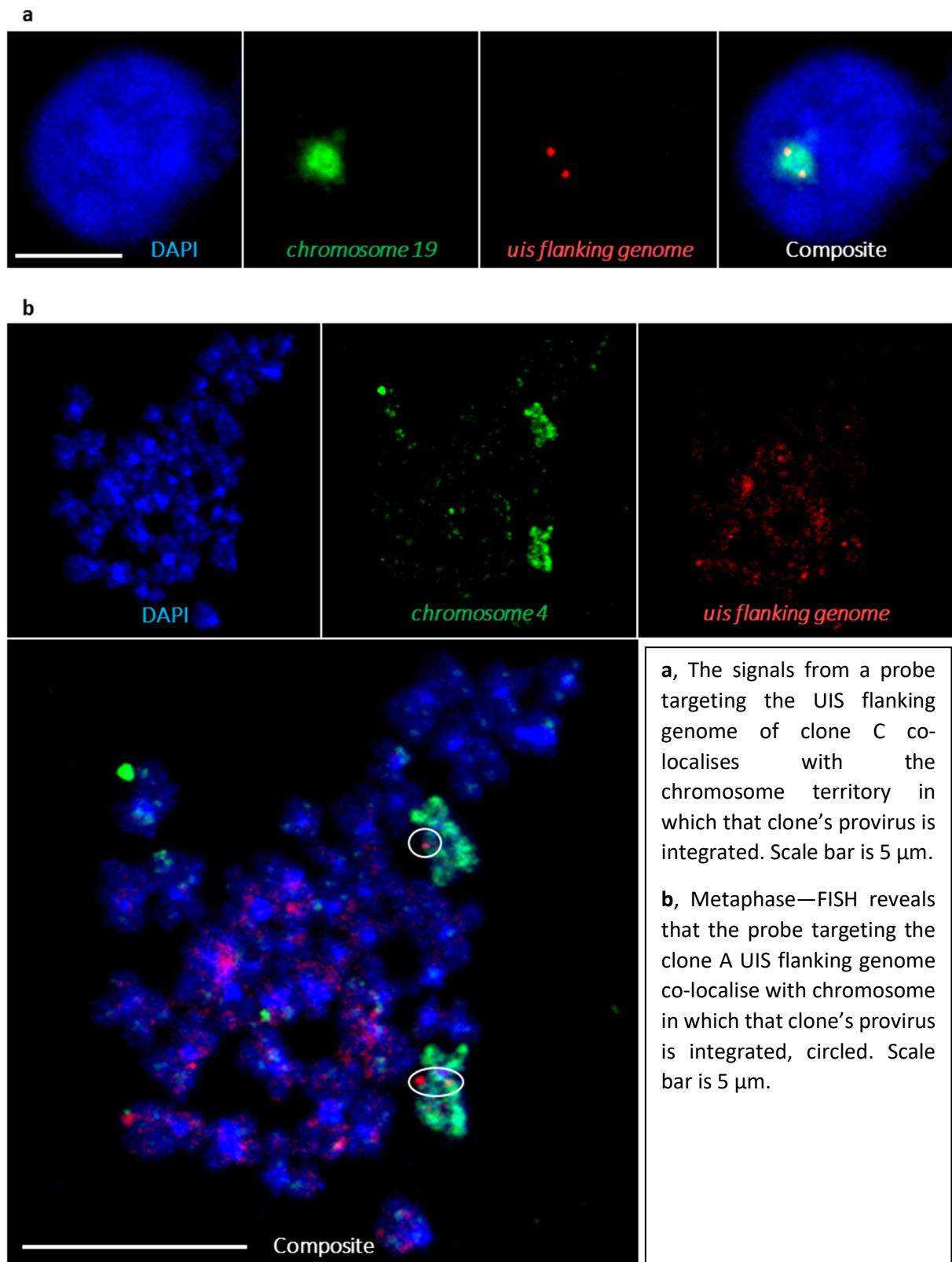


Figure 6-1 | DNA-FISH probes can specifically target the host genome flanking the provirus

6.3.2 Targeting the provirus

Being able to visualise the integrated HTLV-1 provirus represented the holy grail of this part of the project, as it could then have been used in any clone, or even in *ex vivo* infected PBMCs where the proviral integration sites would be mixed and unknown.

The specificity and efficacy of this probe were measured in a similar manner to the flanking genome probes described in the previous section. First, cells were co-hybridised with the proviral probe and the chromosome paint for the chromosome in which the clone's provirus was known to be integrated (by LM-PCR). This was a success (Fig. 6-2a), and it was then possible to combine the proviral probe with the flanking genome. As Fig. 6-2b illustrates, it was possible to see a specific proviral signal colocalised with only one of the two flanking genome signals, as expected. Reflecting the differences in the length of their target sequences, the proviral probe library contained 166 oligonucleotides, while the probe targeting clone C's UIS flanking genome contained 932 oligonucleotides. It is perhaps not surprising that the proviral probe therefore exhibited a lower efficiency: after optimisation, approximately 50% of infected cells contained a specific, robust proviral DNA-FISH signal.

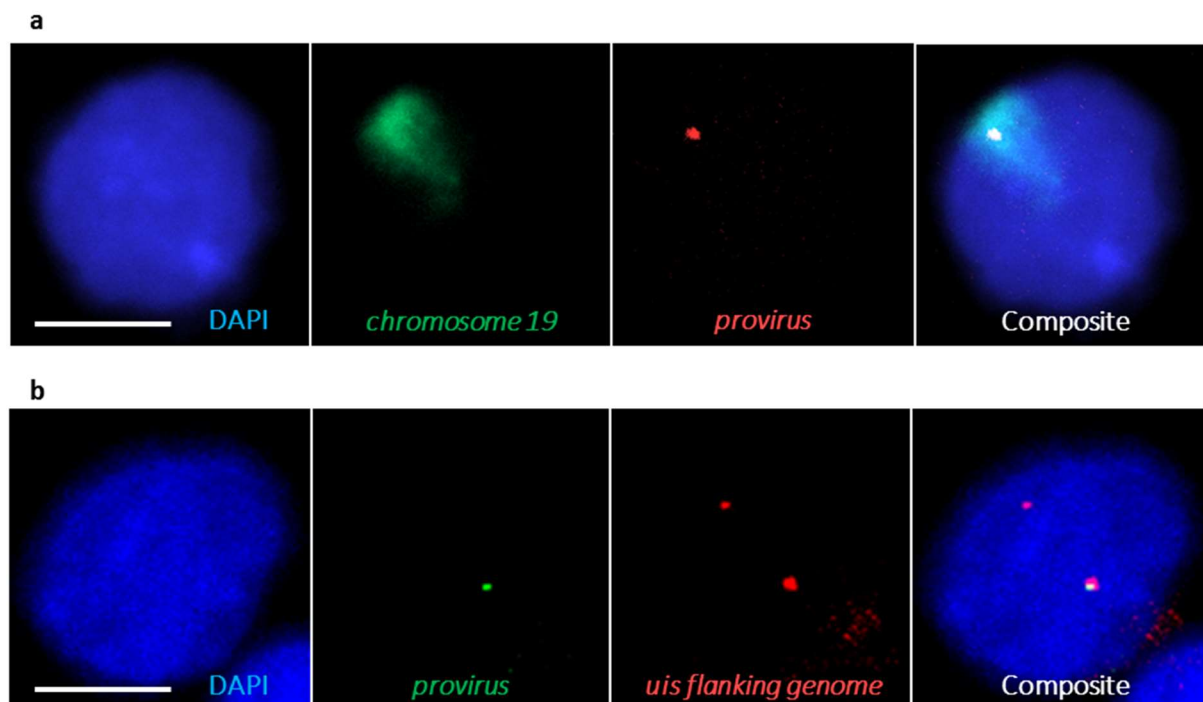


Figure 6-2 | DNA-FISH probes can specifically target the HTLV-1 provirus

a, The signal from a probe targeting the proviral genome co-localises with the chromosome territory in which clone C's provirus is integrated. Image is of a cell from clone C, taken with the widefield IX70 microscope. Scale bar is 5 μ m.

b, The signal from a probe targeting the proviral genome co-localises with only one of the two signals from the probe targeting the UIS flanking genome, as would be expected. Image is of a cell from clone C, taken with the confocal SP5 microscope. Scale bar is 5 μ m.

6.3.3 Immunofluorescence and Immuno-FISH

With functioning probes to target the provirus or its flanking genome, the next step was to combine this with immunofluorescence. This would allow the spatial localisation of the provirus to be resolved relative to subnuclear structures, such as the nuclear lamina (Lamin B1) or nucleolus (Nucleolin), with which it might associate. Fig. 6-3 shows how antibodies were used to visualise these two subnuclear structures, and it was possible to do this in conjunction with detection of the provirus (Fig. 6-3).

Unfortunately, before efforts to systematically capture and quantify the proviral location could begin in earnest, the approach suffered a succession of technical setbacks, and did not progress further. This is discussed further in the Discussion (section 7.9, “Intranuclear localisation of the provirus”).

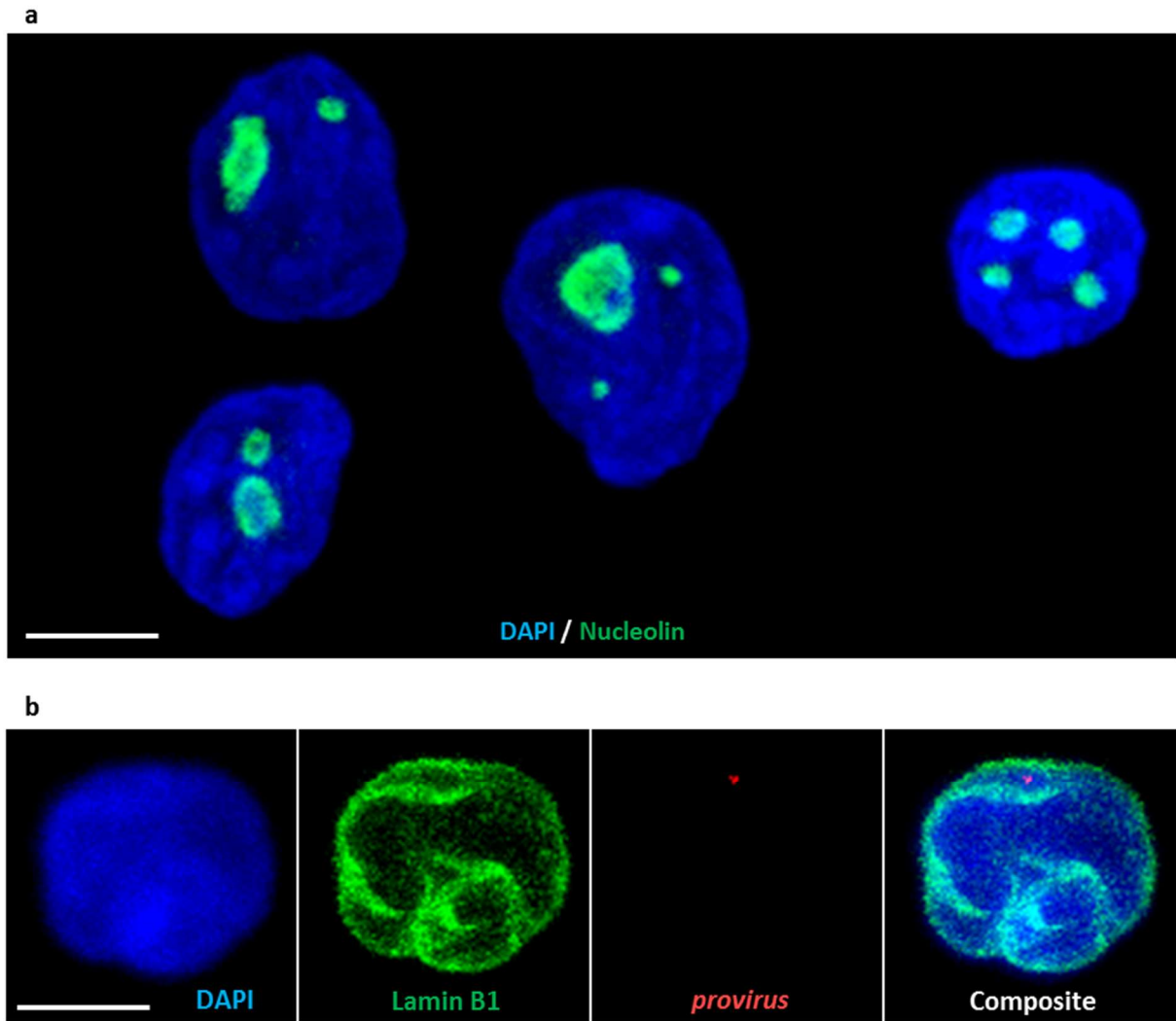


Figure 6-3 | Subnuclear compartments and the provirus can be visualised in the same individual cells

a, Demonstration of successful immunofluorescence stain for Nucleolin, highlighting the nucleoli of infected cells. Image is of cells from clone A taken with the confocal SP5 microscope. Scale bar is 5 μm .

b, Demonstration of successful IF-DNA-FISH, for the purpose of finding the provirus's spatial location in relation to subnuclear structures, such as the nuclear lamina. Image is of a cell from clone A, taken with the confocal SP5 microscope. Scale bar is 5 μm .

6.4 Conclusions

The goal of the work presented in this chapter was to measure the intranuclear location of the HTLV-1 provirus. Below is a broad summary of the results:

- Probes visualising the HTLV-1 provirus and its flanking host genome were designed and shown to be specific to their target sequences, with varying levels of detection efficiency
- These probes could be combined with antibodies, allowing relative position of the provirus and potential colocalisation to be measured
- This avenue of investigation was eventually put aside due to technical setbacks and inconsistency results, and in order to focus on viral expression through smRNA-FISH

CHAPTER SEVEN

DISCUSSION & SUMMARY OF WORK

7.1 Chapter Summary

In this thesis, I have described the use of single-molecule RNA-fluorescence *in situ* hybridisation to investigate the expression profile of integrated HTLV-1 proviruses at the single cell level, something which, to my knowledge, has never been done before. Through this technique, evidence is presented that viral expression of both strands presents a degree of cell-to-cell variation which was hitherto unreported, whether in naturally infected cells, animal models or transformed cell lines. HTLV-1 transcription was found to follow a burst pattern, and the intracellular distributions of plus- and minus-strand transcripts in the form of *tax* and *hbx* were measured. For the first time, the relationship between *tax* and *hbx* expression was investigated within individual cells, as well as the relationship between proviral expression and the cell cycle. A constant theme throughout these observations was the stark contrast between the expression of *tax* and *hbx*. While it ultimately did not progress significantly, work demonstrating the proof-of-principle of locating the HTLV-1 provirus using DNA-FISH was also presented, another first for the field.

This final chapter discusses these findings in the context of the current literature, and asks what they may mean for the broader question of HTLV-1 transcriptional control and latency. These conclusions and inferences are also outlined in the paper submitted for publication (Billman, Rueda and Bangham, 2017). Potential future avenues of investigations are also explored.

7.2 Paradoxes of HTLV-1 persistence

In order for HTLV-1 to survive and propagate within the human host, regulation of proviral latency is crucial. Three paradoxes remain unexplained regarding HTLV-1 gene expression and the host immune response:

1. A strong cytotoxic T-lymphocyte (CTL) response to Tax protein is seen in all immunocompetent hosts, but the magnitude of this response does not explain the observed wide variation between individuals in proviral load (Hanon *et al.*, 2000; Kattan *et al.*, 2009; Demontis, Hilburn and Taylor, 2012).
2. The CTL response to Tax is chronically activated, implying frequent exposure to newly-synthesized Tax protein *in vivo*, but *tax* expression is usually undetectable in fresh PBMCs (Daenke *et al.*, 1996; Hanon *et al.*, 2000; Rende *et al.*, 2011).
3. The CTL response to HBZ is typically weak or undetectable, and *hbz* expression is low in polyclonal PBMCs, but the presence of a detectable anti-HBZ CTL response is associated with a lower PVL and a lower prevalence of inflammatory disease (Saito *et al.*, 2009; Macnamara *et al.*, 2010; Hilburn *et al.*, 2011; Andrade *et al.*, 2013; Rowan *et al.*, 2014; Kataoka *et al.*, 2015).

The results reported in this thesis offer an explanation for each of these paradoxes. By quantifying the frequency and intensity of *tax* and *hbz* expression, I have shown that in clones of naturally-infected, CD4⁺ T-cells, both the plus-strand and minus-strand genes of HTLV-1 are expressed in bursts, which differ strongly in intensity and frequency between the two strands. If this pattern of HTLV-1 gene expression accurately represents the pattern of expression *in vivo*, the paradoxes may be explained as outlined in the following sections.

7.3 Transient *tax* expression

At the single-cell level, expression of *tax* (as a surrogate for the proviral plus strand) is not uniform, but rather exhibits a bimodal expression profile (Fig. 3-1a). The minority of cells expressing *tax* at any one time express it at a very high level (Fig. 3-1aiii). The *tax* transcripts are then rapidly exported to the cytosol (Fig. 4-4a, b) for translation. This intermittent but intense expression of the highly immunogenic Tax protein (Hanon *et al.*, 2000; Goon *et al.*, 2004) is sufficient to account for the observed preponderance of anti-Tax CTLs. However, since the proportion of cells expressing Tax at

any instant is low (of the order of 1% to 10%), the anti-Tax CTL response has a limited impact on the proviral load set point.

Two separate studies have shown that the stochastic “noise” of cell-to-cell transcriptional heterogeneity resulting from bursts is buffered by the nucleus. Controlled nuclear export thus allows a much more constant rate of translation, resulting in significantly reduced heterogeneity at the protein level (Bahar Halpern *et al.*, 2015; Battich, Stoeger and Pelkmans, 2015; Stoeger, Battich and Pelkmans, 2016). The expression profile of *tax* stands in stark contrast to this idea. Single-cell heterogeneity at the transcript level is reflected at the protein level, as few if any *tax*-negative cells stained positive for Tax protein (Fig. 3-4). *tax* mRNA transcripts are also exported rapidly from the nucleus, with seemingly no attempt to “buffer” expression. Within 2 hours of transcriptional inhibition by ActD, the earliest time point, virtually all *tax* had already been exported from nuclei which, with their high *tax* counts, could be expected to have been actively expressing the gene in the recent past (Fig. 4-1d). As has been discussed previously, the limitation of the FISH-Quant nuclear-cytosolic definition of spots to 2D rather than 3D means that the proportion of *tax* spots reported in this thesis are an overestimate (Fig.4-5a). The rapid, almost explosive expression of *tax* followed by its rapid export and seemingly just as fast return to inactivity is reminiscent of the immediate early genes (IEGs), such as the TFs c-fos and c-jun, which are involved in a number of rapid cellular responses to stimuli (Senecal *et al.*, 2014; Battich, Stoeger and Pelkmans, 2015; Bahrami and Drabløs, 2016). Tax has many of the same properties as these proteins, such as upregulating expression through recruitment of CREB, and their promoters are both capable of responding to SRF and CREB through SREs and CREs (Bahrami and Drabløs, 2016). The fact that Tax protein itself has been demonstrated to upregulate a number of these genes is consistent with the notion that the immediate-early stress response enhances proviral plus-strand transcription (Fujii *et al.*, 1994).

The transient nature of Tax expression, and its fleeting appearance in a population of cells are of clear evolutionary and immunological benefit to the virus. The Tax protein is well established as the immunodominant HTLV-1 antigen in the CTL response against the virus (Goon *et al.*, 2004), and anti-Tax CTLs are highly efficient at lysing infected cells which present Tax antigens (Hanon *et al.*, 2000; Cook *et al.*, 2012). Much as “the nail which sticks out shall be hammered down”, there is a strong immunological pressure from CTLs on the virus to minimise presentation of Tax antigens. This is reflected *in vivo*: the most abundant infected T-cell clones in patients have generally been found to have a lower average Tax expression, while clones with a higher rate of Tax expression tend to be far smaller (Melamed *et al.*, 2013). Over time, HTLV-1 patient’s provirus integration profiles have also tended to shift to increasingly silent regions of the genome, pointing more towards latency (Asquith *et al.*, 2005; Melamed *et al.*, 2013). On the other hand, Tax protein expression is clearly of benefit to

the virus. It is necessary for establishing HTLV-1 infection in the tens of thousands of clones found in HTLV-1 infected individuals. Tax expression increases cellular clustering through upregulation of adhesion molecules such as ICAM-1 and CADM1 (Tanaka *et al.*, 1995; Manivannan *et al.*, 2016), and drives the polarisation of the microtubule network upon cell contact to form a virological synapse (Igakura *et al.*, 2003; Nejmeddine *et al.*, 2005). With the development of an LM-PCR and high-throughput sequencing approach, Gillet *et al.* demonstrated that HTLV-1 was far more polyclonal than previously expected, and that the increased PVL seen in HAM/TSP patients was not made up of a small number of high-abundance clones, but instead a larger total number of clones, regardless of their abundance (Gillet *et al.*, 2011). Further to this, despite the overwhelming preference of the provirus for the mitotic over the infectious pathway in chronic infection, new clones are being created through infection on a daily basis (Laydon *et al.*, 2014). All of this demonstrates the importance of Tax expression to HTLV-1 infection *in vivo*. In this light, the episodic, “bursty” pattern of expression of Tax observed in experiments presented in this thesis (Fig. 4-1) presents a viable strategy by the virus to balance its need for proliferation and infection with its avoidance of the immune response. Furthermore, this fleeting but very high level of expression is consistent with the inability to detect Tax *in vivo*, as well as the chronically activated CTL response against it.

A limitation of the approach to study plus-strand expression as outlined in this thesis is the fact that it is not possible to be certain that the transcripts bound by the *tax* probe are in fact *tax*, as opposed to any of the other 6 potential plus-strand transcripts. The extreme genetic efficiency of the HTLV-1 genome, combined with the fact that every non-*tax* plus-strand transcript will contain the *tax* sequence in its 3'-UTR means that there is no clear solution to this problem through the use of smRNA-FISH. qPCR studies have indicated that, along with the structural transcripts, *tax*/*rex* mRNA is expressed at a far higher rate than the accessory genes (Li *et al.*, 2009; Rende *et al.*, 2011), and abundant Tax protein expression has been observed *ex vivo* in multiple studies using naturally infected cells (Hanon *et al.*, 2000; Asquith *et al.*, 2007; Kulkarni *et al.*, 2017). This does not, however, guarantee that every HTLV-1 infected clone will exhibit the same pattern of expression, which bring us to the conundrum of clone D.

Clones A-C, while varying in the proportion of cells observed expressing *tax*, all exhibit the same broad pattern: *tax* tends to be expressed at a very high level or not at all (Fig. 3-5a), and once transcribed is rapidly exported (Fig. 4-5b). While clone D displayed the same binomial expression pattern, its *tax* “ceiling” was far lower, with the vast majority of its *tax*-positive cells containing only a handful of *tax* spots (Fig. 3-5b), most of which displayed an intracellular distribution pattern which was more similar to *hbx* (Fig. 4-5b). The low level of *tax* observed in clone D suggests that the normal positive feedback

loop that drives *tax* mRNA expression is absent or severely attenuated in this clone. There are a number of possible scenarios which could explain these observations:

- 1) Clone D is genuinely expressing *tax* in a pattern which is different from clones A-C
- 2) The clone D provirus contains a mutation which affects *tax* expression
- 3) Nuclear export of *tax* transcripts impaired in clone D

Any of the alternative scenarios has to be reconciled with the fact that clones D and A come from the same patient, so they should be genetically identical except for the location of their proviral integration sites. Since Tax protein has been shown to be necessary to drive high-level transcription from the 5'-LTR, anything which prevented *tax* export and translation could put a significant brake on HTLV-1 plus-strand transcription (Anderson and Dynan, 1994; Tie *et al.*, 1996; Currer *et al.*, 2012). p30 has been reported to sequester *tax/rex* mRNA in the nucleus (Nicot *et al.*, 2004). However, that particular study used a p30 expression construct in a transformed cell line, so it is debatable whether the protein would have the same effect when expressed at the significantly lower levels observed in naturally infected cells (Rende *et al.*, 2011). Furthermore, very little supposed *tax* transcript is leaving the nucleus at all in clone D, so it's not clear that any plus-strand protein would be expressed to any significant degree (Fig. 4-5b). Another possibility is mutation, which although relatively rare in HTLV-1 (Overbaugh, 2001), is thought to occur during initial cell infection through APOBEC3G activity (Fan *et al.*, 2010). If a mutation were to affect Tax protein function, plus-strand expression could still continue, but at much lower level, potentially like that observed in clone D (Fig. 3-5b). Mutations in promoters have also been shown to affect burst size and frequency, which ultimately determine absolute transcript abundance (Tantale *et al.*, 2016). Answering this conundrum would require further investigation, for example through the use of splice-site specific primers in qPCR to identify specific plus-strand transcripts. It is, however, worth stressing that when the human body may contain over 10^5 different clones of HTLV-1-infected T-cells, each with a different proviral integration site (Gillet *et al.*, 2011; Laydon *et al.*, 2014), some degree of variation between clones is to be expected.

The study by Battich *et al.* demonstrated that by multivariate modelling, it was possible to deterministically account for the majority of the seemingly stochastic "noise" of cell-to-cell heterogeneity of expression (Battich, Stoeger and Pelkmans, 2015). However, their study involved genes in the same genomic location in each cell. A significant factor behind the variation in *tax* expression between clones must be related to each clone's UIS. This will be explored further in section 7.6 ("Viral transcriptional bursts").

7.4 Consistent *hbz* expression

The variability in *tax* expression, both within and between clones, is all the more stark when compared with the expression profile of *hbz*. The proviral minus-strand in the form of *hbz* is expressed at a much lower level than the proviral plus-strand (*tax*) (Fig. 3-2b), consistent with previous observations (Satou *et al.*, 2006; Saito *et al.*, 2009). Its range of expression is also several orders of magnitude lower, with a far higher proportion of cells expressing *hbz* at very similar levels in each clone (Fig.3-6). Another shared feature of all the clones is the uniform retention of 80-90% of *hbz* transcripts in the nucleus (Fig. 4-5).

All of these observations are consistent with those made *in vivo*. HBZ protein has consistently been found to be expressed at levels beyond the range of detection at the single-cell level, while always being detected in bulk assays of both patients samples and *in vitro* systems (Satou *et al.*, 2006; Li *et al.*, 2009; Rende *et al.*, 2011; Kataoka *et al.*, 2015). This low expression combined with a low rate of translation explains why HBZ protein is effectively undetectable in physiological conditions. Furthermore, a study which found that increased *hbz* expression was associated with higher PVL and disease risk found that the total detected *hbz* in a patient could be entirely explained by the increased proportion of infected PBMCs (Saito *et al.*, 2009). In other words, total *hbz in vivo* was determined by an increase in the number of cells expressing at a largely constant level, consistent with my findings that 80% of all *hbz* expressing cells contained less than 4 transcripts, accounting for two thirds of total *hbz* expression (Fig. 3-3).

These observations go a long way towards explaining the third paradox posited in section 7.2: the weak or non-existent anti-HBZ response, which nonetheless is one of the only known correlates of protection against HTLV-1-associated inflammatory diseases (Macnamara *et al.*, 2010; Hilburn *et al.*, 2011). There is a significant body of literature indicating that HBZ is essential for HTLV-1 persistence in the host. HTLV-1 viruses that do not express *hbz* establish a lower set-point PVL in animal models, or exhibit reversion mutants which restore *hbz* expression (Arnold *et al.*, 2006; Valeri *et al.*, 2010). The 3'-LTR has never been found to be silenced by epigenetic changes (Miyazaki *et al.*, 2007; Fan *et al.*, 2010), mutated or deleted in ATLL lines or patients, and knocking down *hbz* expression through RNA interference suppresses the growth of ATLL cell lines *in vitro* (Satou *et al.*, 2006). The RNA-Seq study by Kataoka *et al.* also detected *hbz* expression in all ATLL patients (Kataoka *et al.*, 2015).

Given these observations, it is perhaps not surprising that the presence of CD4+ and CD8+ T-cells which can target HBZ epitopes would correlate with a lower PVL, and lower disease risk (Hilburn *et al.*, 2011). In addition, the functional avidity and lytic efficiency of anti-HBZ CTLs (Kattan *et al.*, 2009), partially

determined by the strength of HBZ antigen presentation by MHC I (Macnamara *et al.*, 2010; Rowan *et al.*, 2014), would play a significant part in determining the effectiveness of the anti-HTLV-1 immune response. HBZ appears to be the virus's "Achilles' heel," so it is in the virus's interest to hide it from the immune response to the greatest extent possible. HBZ is remarkably poorly immunogenic (Goon *et al.*, 2004; Enose-Akahata *et al.*, 2013; Rowan *et al.*, 2014; Shiohama *et al.*, 2016), and HBZ peptides bind very poorly to MHC I (Macnamara *et al.*, 2010). These findings, combined with the fact that *hbz* is so rarely exported from the nucleus (Fig. 4-4) for translation, explains why a detectable anti-HBZ T-cell response is low and can be difficult to detect, in spite of being so effective.

A study by Padovan-Merhar *et al.* demonstrated through smRNA-FISH that cells were able to maintain constant transcript abundance as a function of cell volume (Padovan-Merhar *et al.*, 2015). *hbz* expression appears to follow a similar pattern, with differences in average *hbz* abundance per cell largely disappearing once corrected for nuclear volume (Fig. 5-8). Cell size measurements in this analysis pipeline are estimates, since cell boundaries were not specifically stained, so instead had to be defined using the relatively small difference in intensity between cell autofluorescence and background signal. The definition of nuclear volumes, however, can be used with much more confidence (thanks to DAPI staining), and the same count / volume relationship was observed. The fact that there can be so much variation in the size of T-cell nuclei (the largest 10% are 3.6-fold larger than the smallest 10%) stands in contrast to the Padovan-Merhar study, which only observed a limited increase in nuclear volume in larger cells. It should also be noted that the great majority of the genes which they tracked, for example *gapdh*, appeared to be expressed to some degree in all cells; none appeared to be entirely silent (Padovan-Merhar *et al.*, 2015). If true, this fine control of *hbz* expression to maintain a constant RNA "concentration" based on volume is remarkable, and could reflect an adaptation by HTLV-1 that confines *hbz* expression to the absolute minimum necessary, in order to avoid immune detection.

It has long been held that *hbz* is constitutively expressed in infected cells (Ma, Yasunaga and Matsuoka, 2016). However, all previous studies quantified HBZ expression in cell populations, using bulk techniques such as RNA-Seq qPCR or antibody titres (Satou *et al.*, 2006; Kataoka *et al.*, 2015; Mitobe *et al.*, 2015; Shiohama *et al.*, 2016). The single-cell based approach reported here suggests that, while expressed in a significant majority of cells, *hbz* is not present in all cells at all times (Fig. 3-6). While this finding could be a result of probe inefficiency, the size of the *hbz* oligonucleotide library contains significant redundancy (48 in total, while the published detection limit is somewhere in the region of 25 oligonucleotides) (Stellaris protocols, Raj *et al.*, 2008), such that a significant number of oligonucleotides can fail to bind and the transcript should be still be detectable. The proportion of *hbz*-negative cells in each clone was also reproducible, and varied from 10% to 35% between clones

(Fig. 3-6). Finally, these proportions were all found to vary significantly from what would be expected in a Poisson-like distribution of the frequency of *hbz* spots (Appendix, Table 8-4).

The fact that, while much less variable than *tax*, *hbz*'s expression profile was demonstrably non-Poissonian (Appendix, Table 8-4) is a strong indicator that its expression follows a burst pattern. It is also possible that it may benefit the virus to silence *hbz* for brief intervals, given its relationship with *tax* expression (see section 7.7, "Relationship between *tax* and *hbz* expression").

7.5 Intracellular distribution and half-life of *hbz*

As the proportion of *hbz* in the nucleus remained relatively constant once transcription was blocked, a few questions are raised if one trusts the accuracy of *hbz* detection and location definition. Are the nuclear and cytosolic degradation rates of *hbz* RNAs comparable? Is the export of *hbz* constant, or is it triggered by other factors? These questions cannot be addressed by this study, but may be worth investigating in the future. However, given *hbz*'s preference for nuclear localisation, and its potential to affect the cell by RNA alone, it may be that the important functions of *hbz* are exerted by the RNA form, analogous to a long non-coding (lnc) RNA (Batista and Chang, 2013).

The half-life of *hbz* mRNA was estimated to be 4.4 hours, significantly longer than two previous studies, which produced estimates of the half-life closer to 1 hour, although one study did find this increased to almost 3 hours in the presence of Tax (Mocquet *et al.*, 2012; Cavallari *et al.*, 2016). The similarities between these previous studies and the present approach are limited, and the differences are sufficient to explain the disparity, as well as demonstrating the higher credibility of my technique.

The Cavallari study used the same method of transcriptional inhibition, ActD, at comparable concentrations to the one used in this thesis. The Mocquet study instead used the transcriptional inhibitor DRB (5,6-Dichlorobenzimidazole 1- β -D-ribofuranoside), which prematurely terminates elongation through inhibition of Cdk2, but may not block transcription of all genes (Yankulov *et al.*, 1995; Bensaude, 2011). While the experiments described in this thesis involved direct, single-cell level counting of *hbz* transcripts in naturally infected T-cells, both the earlier studies used cell lines transfected with a molecular clone of the provirus or simply with transfected plasmids. In other words, in these studies *hbz* was being expressed at an artificially high rate, in a foreign environment.

Furthermore, both studies used qPCR, which requires a "housekeeping" gene, expressed at constant rate in cells, in order to measure the relative abundance of the target transcript. Each study elected

to use RNA from the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene, a commonly used housekeeping gene in qPCR. However, *gapdh* expression has been shown to vary significantly in activated T-cells (Bas *et al.*, 2004), bringing into question its appropriateness as a standard for quantification. In fact, one of the very first images captured during the optimisation of my smRNA-FISH protocol was of infected T-cells stained for *gapdh* mRNA. Figure 7-1 illustrates that *gapdh* also displays a wide degree of cell-to-cell heterogeneity, as has recently been reported (Padovan-Merhar *et al.*, 2015). Finally, the Cavallari study also separated nuclear and cytosolic fractions of cells to measure mRNA abundance, compartmentalisation and degradation. There is no guarantee that *gapdh* levels would be comparable in both compartments (Fig. 7-1).

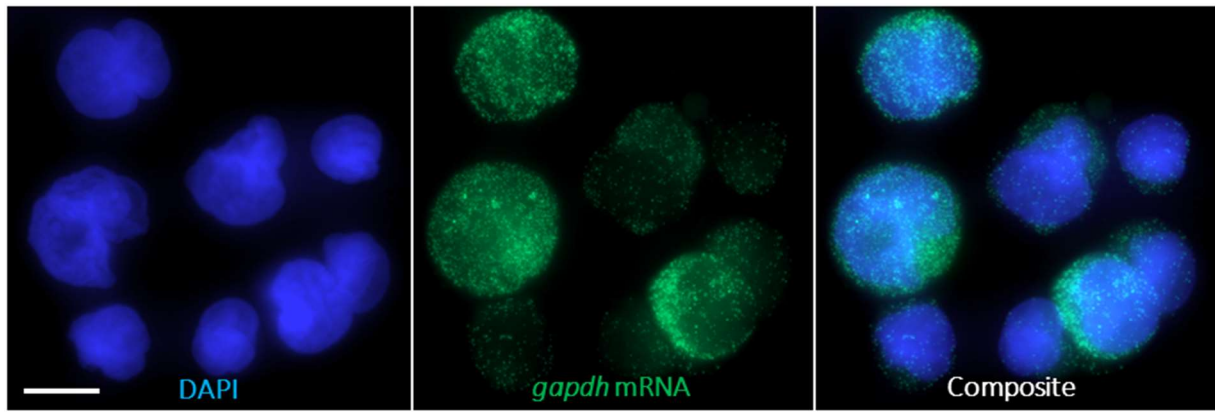


Figure 7-1 | smRNA-FISH reveals cell-to-cell heterogeneity of *gapdh* expression

Putative transcription bursts are visible in the cell with the highest *gapdh* abundance.

7.6 Viral transcriptional bursts

Episodic expression of gene products as a result of transcriptional bursts have been reported with increasing frequency, especially in higher eukaryotes (Dar *et al.*, 2012; Nicolas, Phillips and Naef, 2017), as well as seeming to play a role in reactivation from latency and cell fate in HIV-1 (Weinberger, Dar and Simpson, 2008; Weinberger and Weinberger, 2013). With their significantly non-Poissonian transcript expression patterns (Figs. 3-5 and 3-6), it is clear that HTLV-1 *tax* and *hbz* expression also display this form of transcriptional activity. This is perhaps not surprising when one considers that while its DNA sequence and proteins are significantly different from those of its human host, the HTLV-1 provirus still finds itself in a given human chromatin environment whose state is known to affect viral transcription (Lemasson *et al.*, 2002; Lu *et al.*, 2002; Easley *et al.*, 2010; Belrose *et al.*, 2011), and its transcription is dependent on manipulation of host proteins and TFs in order to express itself (Anderson and Dynan, 1994; Kwok *et al.*, 1996; Harrod *et al.*, 2000; Zhou *et al.*, 2006; Kuhlmann *et al.*, 2007).

A transcript's overall abundance in the cell is a function of its degradation rate, the frequency at which it is expressed, and how many transcripts are created in each burst of expression. Measuring the size of transcriptional bursts using smRNA-FISH is dependent on a number of factors:

1. The burst's duration
2. The initiation rate, or rate at which polymerases are loaded onto and released from the transcription start site
3. The relevant polymerase's elongation rate, and the length of sequence being transcribed
4. The length of time for mRNA to be processed and released from the polymerase, also known as its "retention time"

mRNA retention time is particularly important from the perspective of smRNA-FISH, as once transcripts are released by the polymerase, they will most likely quickly diffuse away from the transcription site, decreasing the apparent size of the burst. If a burst's duration is shorter than this retention time, it is possible to accurately quantify its size. However, if the processing and release time is shorter than the burst duration, one will only ever be able to capture a portion of the burst through this technique.

By inserting different promoters in the vicinity of 8000 different genes in the human genome, Dar *et al.* showed that cellular transcript abundance can be modulated either through the size of transcriptional bursts, or the frequency at which they occur (Dar *et al.*, 2012). A number of studies

have shown that, along with local chromatin environment, burst size and frequency are affected by global factors such as TF availability, depending on the promoter. Given the aforementioned similarities between *tax* and the IEGs, it is serendipitous that a recent study characterised the burst properties of the IEG *c-fos* (Senecal *et al.*, 2014). As an IEG, *c-fos* is transcribed within minutes of a stimulus through activating histone modifications around the promoter, including acetylation, which displaces a nucleosome to reveal the transcriptional start site (Bahrami and Drabløs, 2016). Like the 5'-LTR, the *c-fos* promoter contains SRF, CREB and NF- κ B response elements (Healy, Khan and Davie, 2013) – this is why Tax can upregulate *c-fos* itself – and seemingly like *tax*, *c-fos* is expressed transiently: expression of the protein peaks after approximately an hour, and returns to basal levels after 90 minutes (Greenberg and Ziff, 1984). Another similarity is the variability of *c-fos* transcription, even in the context of serum induction within an isogenic clone (Senecal *et al.*, 2014). Senecal *et al.* found that increasing TF concentration had a positive correlation with overall *c-fos* expression, and that this was explained by more frequent, rather than larger bursts (Senecal *et al.*, 2014). It would make sense that a gene which must be expressed as quickly as possible would be able to achieve the maximum rate of polymerase initiation and transcription, and this may be the case for HTLV-1 plus strand expression as well.

Tantale *et al.* demonstrated in live cells that bursts from the HIV-1 promoter consist of “convoys” of polymerases processing along the DNA strand, transcribing in tandem; they observed the same behaviour from the human POLR2A gene, which encodes the major subunit of RNA Pol II (Tantale *et al.*, 2016) (see section 1.2, “Transcriptional bursting and cell-to-cell heterogeneity”, for the introduction to the factors and interactions involved in transcription). The size of a burst was determined by the length of the gene that was being transcribed, the polymerase elongation rate and the rate with polymerases they could be loaded and released from the PIC. This final step seems to be determined by the Mediator complex, and in particular the association of P-TEFb with the complex so that Cdk9 can phosphorylate the RNA Pol II to commence elongation (Tantale *et al.*, 2016).

Tax, and its HIV-1 counterpart Tat, can accelerate the initiation of transcription by recruiting P-TEFb itself (Zhou *et al.*, 2006), as well as being able to directly assemble essentially all the other necessary TFs such as CREB and CBP/p300 (Kwok *et al.*, 1996; Harrod *et al.*, 2000; Currer *et al.*, 2012). It is clear that, once *tax* is expressed and translated, Tax protein will quickly amplify its own expression. Could IEG expression control offer clues as to how *tax* expression begins and ends? As has been mentioned, they share a number of response elements, so it's possible that plus-strand expression could begin through non-specific action of the stress response. Kulkarni *et al.* have recently shown that glucose depletion and hypoxia can stimulate *tax* expression, so this seems to be a promising avenue of investigation (Kulkarni *et al.*, 2017). How are IEGs suppressed when they are no longer needed? Two

studies have found them to be repressed through the expression of microRNAs (Avraham *et al.*, 2010; Aitken *et al.*, 2015). To my knowledge no similar mechanism has been found in HTLV-1.

In the case of HIV-1 and POL2RA polymerase convoys, burst size increases as more polymerases initiate transcription, before falling again once initiation ceases and the same polymerases begin to terminate transcription (Tantale *et al.*, 2016). *tax* is the last gene on the plus-strand before the polyadenylation site (Fig. 1-2), and as Tax can accelerate the polymerase loading and initiation rate, this partly explains why *tax* bursts appear to be particularly large. However, with an assumed 200 nt spacing between polymerases (the highest polymerase density observed with convoys from the HIV-1 promoter activated by Tat), that would mean a maximum convoy of approximately 45 polymerases producing transcripts along the length of the provirus. While the average *tax* burst is smaller than this (Fig. 4-2c), some bursts have been estimated to be considerably larger, with the largest estimated to contain the equivalent of approximately 190 full-length transcripts. This suggests that a) the intensity of transcription (i.e. polymerase initiation and elongation rates) is operating at a very high level, and b) mRNAs are being retained at the site of transcription for an unspecified time after elongation, perhaps to complete splicing. It is unclear at what rate a burst can be expected to dissipate once mRNAs are released and are free to either diffuse or be transported away.

An impressive number of studies have investigated the effects of genomic location and chromatin environment on transcription through the use of lentiviral vectors to insert an HIV-1 promoter into a variety of genomic sites, before using limiting dilution to isolate and study “clones” of cells who share the same unique integration site (Weinberger, Dar and Simpson, 2008; Skupsky *et al.*, 2010; Dar *et al.*, 2012; Tantale *et al.*, 2016). In the absence of a transcriptional transactivator, this promoter and others are significantly affected by inhibitory epigenetic marks or nucleosome occupancy (Bai and Morozov, 2010; Dey *et al.*, 2015). HTLV-1 carries out this experiment every time it infects a new cell, and the literature bears out the effect of the UIS chromatin environment on proviral transcription (Gillet *et al.*, 2011). So many of Tax’s effects which are linked to upregulating plus-strand transcription are tied to chromatin remodelling, such as expelling nucleosomes and hyperacetylating the promoter site through the recruitment of host co-factors (Harrod *et al.*, 2000; Lu *et al.*, 2002, 2004; Sharma and Nyborg, 2008; Easley *et al.*, 2010), and they can be counteracted through the action of HDACs (Ego, Ariumi and Shimotohno, 2002; Lemasson *et al.*, 2004). The influence of the flanking host genome is also suggested in an anecdotal fashion in this thesis: clone C has the most frequent *tax* bursts by far, approximately 10% (Fig. 4-2a). This particular clone’s UIS happens to be located between the *cebpa* and *cepbg* genes, two TFs associated with cell proliferation in a number of cell types including lymphocytes (Zhang *et al.*, 2004). It seems likely that this particular genomic environment will be highly activated and open, and that it would contribute to higher proviral activity.

Interestingly, Padovan-Merhar *et al.* found that transcriptional bursts increase in size in step with cellular volume, allowing the cell to maintain a constant concentration of mRNA (Padovan-Merhar *et al.*, 2015). While this was not observably the case for *hbz*, the principle did seem to apply to *tax* (Fig. 5.8b)

Only a proportion of all transcription events in the form of bursts will be captured by smRNA-FISH, as there are some limitations. Early or late bursts, which are not significantly brighter than individual full-length transcripts, will not be captured. It is also possible that transcription, particularly in the case of *hbz*, in some cases occurs without polymerase convoying, or in other words one transcript at a time. Such a burst would simply look like any other transcript, if it were detectable at all.

The nascent *gag* probe is also imperfect as it is not will not be removed in the transcription of *gag-pol* transcripts, which some studies have estimated to occur at a comparable rate to *tax/rex* (Li *et al.*, 2009; Belrose *et al.*, 2011). The overlapping ORFs of the plus-strand mean that there are no true introns which can be targeted. This is however not the case for *hbz*, and a nascent RNA probe could be used in a future study to improve minus-strand burst characterisation. This would be especially helpful in G₂ cells, where there may in fact be two bursts at the same time.

7.7 Relationship between *tax* and *hbz* expression

Earlier investigations have pointed to an antagonistic relationship between Tax and HBZ: Tax can drive HBZ expression, which can then compete with Tax for the host factors which are necessary for expression driven from the promoter in the 5'-LTR (Lemasson *et al.*, 2007; Yoshida *et al.*, 2008). Transfected HBZ was also found to be able to block Rex-dependent export of viral mRNAs from the nucleus in a transformed cell line, although it is unclear if this would affect Tax itself (Philip *et al.*, 2014), and with the possible exception of clone D no *tax* burst was ever observed to end in cells where *tax* spots appeared to be sequestered in the nucleus.

These previous observations, in combination with our observations of the correlation between levels of *tax* and *hbz* within individual cells (Figs. 5-1 and 5-2), suggest the following model of the interaction between *tax* and *hbz*. *tax* is more likely to be expressed in an *hbz*-silent cell (Fig. 5-1ai) and, once expressed, *tax* typically reaches very high levels (Fig. 5-1a_{ii}). A high level of *tax*, however, increases the probability of *hbz* expression (Yoshida *et al.*, 2008) (Fig. 5-1a_{ii}), which in turn leads to cessation of *tax* expression. Subsequently, both *tax* and *hbz* transcripts decay (Fig. 5-1a_{iv}, av), and the provirus remains latent until another factor upregulates expression, or the cell becomes silent, restarting the

potential cycle. The factors that regulate the onset of *tax* expression include glucose metabolism and the availability of molecular oxygen (Kulkarni *et al.*, 2017); cellular activation and stress may also contribute.

My single-cell approach indicates that *hbx* is not expressed constantly in all cells, even within a single clone. If the pattern of proviral expression we observed is also found *in vivo*, a temporary lapse in *hbx* expression could, allow a *tax* burst to develop, before *tax* declines either under the influence of *hbx*, or immune detection and destruction. The pattern of frequent, low-level expression of *hbx* is consistent with the notion that the primary function of HBZ – in both its protein and mRNA forms – is to maintain clonal persistence (Satou *et al.*, 2006; Boxus and Willems, 2009; Saito *et al.*, 2009; Matsuoka and Yasunaga, 2013; Ma, Yasunaga and Matsuoka, 2016). Regardless of the cause, limiting expression of the highly immunogenic Tax protein to intermittent bursts is an appropriate adaptation by the virus to optimise the protein's effects in manipulating the host cell, to drive viral replication and clonal proliferation, while minimising its exposure to the intense anti-Tax immune response.

The cause of the observed cell-to-cell heterogeneity in *hbx* expression is unknown. It is also unclear precisely how the relationship between Tax and HBZ is mediated. The previously described mechanisms of interaction between Tax and HBZ act at the protein level (Basbous *et al.*, 2003; Clerc *et al.*, 2008). However, given the very low proportion of *hbx* transcripts exported from the nucleus in any given cell, the low abundance of HBZ protein found in naturally-infected cells and the far lower than expected co-occurrence of *tax* and *hbx* bursts, it is possible that the transcription of *hbx* itself has a negative effect on *tax* expression (Fig. 5-2). At first, this would seem to be contradicted by the Satou *et al.* study, in which an *hbx* construct with a mutated start codon did not reduce Tax expression, while a construct with silent mutations, to disrupt RNA structure but allow translation, was able to do so (Satou *et al.*, 2006). However, the Satou *et al.* study involved transfection of cells with individual *tax* and *hbx* constructs, so they were incorporated into different genomic locations. In the WT provirus, the very process of minus-strand expression could disrupt transcription proceeding on the opposite strand. Opposite strand transcription has been shown to be self-disruptive through two mechanisms: the generation of complementary microRNAs which can target the cellular RNA-interference system to the forward strand transcript, and the physical collision of RNA-polymerases proceeding in opposing directions (Hobson *et al.*, 2012; Proudfoot, 2016). While collision has been shown to be sufficient to terminate elongation, it is not the whole story. It is not clear why *hbx* should always seem to “win” these encounters, or how a collision would disrupt further initiation events at the 5'-LTR. It is also not clear whether any potential microRNAs would affect expression at the level of transcription, as opposed to degrading transcripts after they have been transcribed.

As *hbz* bursts are only defined by their intensity, an alternative explanation would be that instead of bursts corresponding to transcripts accumulating at site of ongoing transcription; they could result from already mature *hbz* transcripts accumulating at a site in order to effect a direct regulation event. This would be consistent with the observation that *hbz* RNA has regulatory properties independent of its protein form. However, it is unlikely for two reasons. First, it seems unlikely that this form of spatial association would be so strongly affected by inhibition of transcription (Fig. 4-1e). Second, no cells were observed with a high number of dispersed *hbz* transcripts and a *tax* burst.

It has been observed that genes which are relatively close to each other, within 100 kb but in some cases up to several megabases apart, typically exhibit very similar expression patterns (Raj *et al.*, 2006; Lieberman-Aiden *et al.*, 2009; Wang *et al.*, 2016). *tax* and *hbz* are located extremely close to each other on opposing strands, which makes their contrasting expression patterns all the more remarkable. However, this similarity of expression between genes is typically based in self-similar chromatin, while the 5'- and 3'-LTRs have been demonstrated to be enriched for differing epigenetic marks. This difference may be maintained by CTCF, acting as an insulator-binding protein, and its main role could be to preserve the 3'-LTR and *hbz* ORF from silencing by hypermethylation (Satou *et al.*, 2016).

7.8 Relationship between viral expression and the cell cycle

The observed longevity of HTLV-1-infected T-cell clones *in vivo* indicates that sustained cell proliferation plays a significant part in viral persistence (Gillet *et al.*, 2011). The relationship between *tax* and *hbz* expression and the cell cycle is therefore of critical importance. I found that cells in G₂/M had on average a significantly higher amount of *tax*, and the minority of cells with high *tax* were found significantly more frequently in S or G₂/M phase than in G₀/G₁ phase (Fig.5-4a). In contrast, *tax* bursts were not significantly associated with S or G₂/M phase (Fig. 5-4b), perhaps because of the low frequency and transient nature of *tax* bursts. The strength of this relationship between *tax* and the cell cycle varied between clones; in contrast, the relationship between *hbz* and the cell cycle was remarkably strong and consistent between clones. The logistic regression analysis indicated that each additional *hbz* transcript detected increased the odds of a cell being in G₂/M (rather than G₀/G₁) by an average of 1.45-fold (7.3-fold in *hbz*-high cells), whereas each additional *tax* transcript increased the odds by 1.004-fold (1.58-fold in *tax*-high cells). This correlation with *hbz* was maintained in the clone incapable of expressing *tax* (clone E), showing that the relationship between *hbz* and the cell cycle is independent of *tax*.

Although several mechanisms have been described by which Tax protein and both HBZ protein and RNA promote cell proliferation (Satou *et al.*, 2006; Matsuoka and Yasunaga, 2013; Mitobe *et al.*, 2015), it remains possible that cell division itself enhances proviral expression: that is, proviral expression may be both cause and consequence of cell division. Just by virtue of the fact there are twice as many proviruses in a cell in G2 or M phase than other cells, expression could increase. Two studies have shown that allelic promoters can activate independently of each other, although their action can become correlated through response to the same stimulus (Kalo *et al.*, 2015; Skinner *et al.*, 2016). Padovan-Merhar *et al.*, meanwhile, found that burst frequency *per allele* decreases in diploid cells in order to equalise expression, and a seemingly similar dosage-compensation response in burst frequency was observed for two crucial genes in haematopoietic stem cells (Padovan-Merhar *et al.*, 2015; Skinner *et al.*, 2016). It is not a simple endeavour to disentangle these relationships with HTLV-1 expression, as the two strands of virus so clearly affect each other (Fig. 5-1 and Table 5-3).

There was a very strong linear correlation between *hbz* spot count and cellular or nuclear volume (Fig. 8). Certain cellular mRNAs show a similar dependence on nuclear volume or cell volume (Battich, Stoeger and Pelkmans, 2015; Padovan-Merhar *et al.*, 2015); the mechanisms responsible are not yet known. However, the frequency of *hbz* bursts was independent of cellular or nuclear volume, but was strongly associated with S/G₂/M. This burst timing contrasts with the transcription rate of many cellular genes, which is reduced during S/G₂/M, probably to compensate for the doubling of the gene copy number (Skinner *et al.*, 2016). The occurrence of an *hbz* burst in S/G₂/M might confer two possible advantages on the virus. First, *hbz* promotes progression through the cell cycle (Satou *et al.*, 2006; Mitobe *et al.*, 2015). Second, the burst will increase the abundance of both HBZ mRNA and protein molecules, which are normally present at limiting frequency, and thereby ensure efficient partitioning of the molecules between the two daughter cells (Soltani and Singh, 2016).

Since HTLV-1 persists in the host by driving proliferation and avoiding the immune response, it is a logical strategy to align its proviral expression with the host cell cycle, particularly for such an immunogenic product as Tax. My current interpretation of HTLV-1 transcription and its regulation is summarised in Fig. 7-2 (Overbaugh, 2001; Yoshida, 2001; Matsuoka and Jeang, 2011).

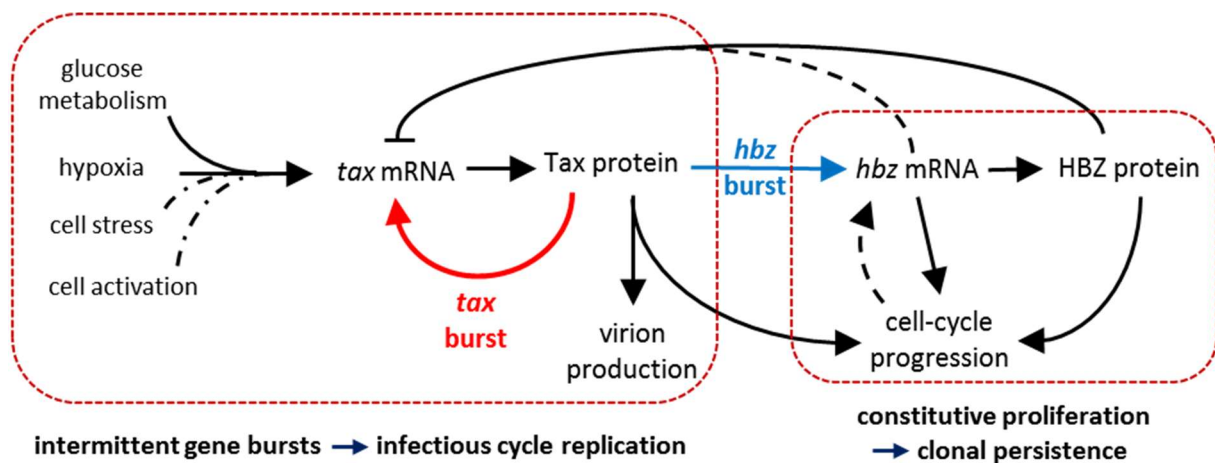


Figure 7-2 | Model of regulation of HTLV-1 transcription and replication in vivo

Constitutive expression of *hbz* mRNA maintains clonal longevity by promoting cellular proliferation. Intermittent bursts of *tax* expression have three chief consequences: production of HTLV-1 virions and resulting infectious spread; accelerated cell-cycle progression; and stimulation of *hbz*, of which the RNA and protein both have proliferative effects. HBZ protein can also inhibit *tax* transcription, terminating the *tax* burst. It is not known whether *hbz* mRNA also counteracts *tax* transcription, but it is suggested by our results. The factors that regulate the frequency of *tax* bursts include glucose metabolism and hypoxia; cell stress and cell activation may also trigger *tax* expression. Aside from Tax, other triggers for *hbz* expression remain to be discovered. The effect of cell volume is unclear, but may set the level to which *hbz* is expressed once a burst has begun. Dashed line represents a link indicated by results in the present study; dotted lines represent hypothetical links.

7.9 Intranuclear localisation of the provirus

The original scope of this project was centred on the spatial localisation of the provirus, and what role that could play in viral transcription and latency. The 3D organisation of the interphase genome has been shown to be highly structured, with particular nuclear locations and intranuclear structures associated with transcriptional activation or repression (Bickmore and van Steensel, 2013; Gibcus and Dekker, 2013). Bickmore *et al.* demonstrated that the intergenic regions of a chromosome, or its inactive genes, were located significantly closer to the nuclear lamina, which is associated with repressive histone modifications and transcriptional repression, than were the gene-rich, transcriptionally active regions (Boyle *et al.*, 2011). It is difficult to discern a direction of causality between a locus's location and genetic activity, but the same group has also shown that by physically tethering a locus to a nuclear structure, or changing its intranuclear location, transcriptional activity can be increased or decreased in a number of genes (Finlan *et al.*, 2008). Another salient example is genomic regions spatially associated with the nucleolar periphery, which is enriched with heterochromatin (Németh and Längst, 2011). As the nucleolus disperses when a cell is subjected to stress, such as hypoxia or nutrient deprivation (Boulon *et al.*, 2010), as well as during the cell cycle, an HTLV-1 provirus integrated in this region could have its expression controlled by these factors. If this potentially more efficient timing of proviral expression reduced its exposure to the immune system, it could go a way towards explaining the *in vivo* preferential survival of clones whose UIS was found in acrocentric, nucleolus-associated chromosomes (Cook *et al.*, 2014), as well as the latency and reactivation of HTLV-1 in the blood of infected patients (Hanon *et al.*, 2000). HTLV-1 has also recently been shown to be capable of forming stable, reproducible long-range spatial associations with other genomic loci through chromatin looping by CTCF (Satou *et al.*, 2016).

Following the lack of progress in DNA-FISH (described later in this section) and the discovery of HTLV-1 transcriptional bursts, it seemed eminently reasonable that these bursts could serve as proxies for the location of the provirus, as a transcription-dependent accumulation of transcripts at a site other than the location of transcription seemed unlikely (Figs. 4-1d, e). Given that interphase chromosomes are arranged in spatially discrete “territories”, it was perhaps not surprising that clones whose proviruses were integrated in different chromosomes could have significantly different intranuclear burst locations. It is worth remembering however, that while chromosome territories are discrete, their precise spatial localisation within the nucleus is subject to significant cell-to-cell heterogeneity, yet still there were measurable differences using as crude a metric as normalised distance from the nuclear centre (Fig. 4-3).

These differences in burst localisation between clones aligned intriguingly with some of the clones' differences in HTLV-1 expression. For example, bursts from clones B and C tended to be closer to the nuclear centre, which is generally associated with transcriptional upregulation. *tax* bursts were significantly more common in these clones, and they also had a higher average *hbx* burst size. However, when analysing the data from the perspective of burst size, or overall transcript counts, no significant association was found: for example, larger *hbx* bursts were just as likely to be found close to the nuclear periphery as in the centre. There may well be an association between the provirus's 3D location and its transcriptional activity, but it would require the co-staining of subnuclear structures in order to elicit more insightful data than normalised distance from the nuclear centre.

In order to properly investigate the link between expression and intranuclear location, it is necessary to visualise the provirus itself. This would serve both as a definitive validation for the transcriptional bursts, but also allow the tracking of the provirus in transcriptionally silent cells. There could, for example, be meaningful differences in the spatial associations and interactions of the provirus depending on its transcriptional state, much as an association with PML has been linked to HIV-1 latency (Lusic *et al.*, 2013).

So why did the DNA-FISH approach stall, after a significant amount of optimisation and progress? I am not able to provide a satisfactory answer. Probe efficiencies suddenly dropped, and could not be improved again. Subsequent optimisation included the replacement of all buffers and materials, the replacement of all probes and cells, and the retesting of fixation, permeabilisation, denaturation and hybridisation conditions. Probe function was eventually restored by switching from heat denaturation with 10 min on an 80°C hot plate to the use of an 80°C oven for 30 min. As the protocol had previously been optimised and worked perfectly well on the hot plate, I cannot adequately explain this. The mechanics of DNA-FISH as compared to RNA-FISH, discussed previously, are such that it can be more difficult to obtain a useable signal-to-noise ratio from a 50 kb stretch of DNA than from 1 kb or even 500 bp of RNA. More recent studies have shown considerable success at labelling and imaging comparatively short stretches of DNA (Beliveau *et al.*, 2015). A final concern, which is simply a more extreme version of the criticism levelled at RNA-FISH, is that the harshness of the fixation, permeabilisation and denaturation treatments necessary for successful hybridisation may affect the cells in a multitude of unknown ways, and allow one to question the extent to which their reported results correspond to the reality of the cell.

7.10 Directions for future work

smRNA-FISH has for the first time revealed the heterogeneity of HTLV-1 expression in naturally-infected T-cells. Naturally, this has presented as many new questions as it has answered old ones.

While it has been of immense value to use isogenic clones in the experiments presented here, it will be important to display to the fullest extent possible that the trends observed here are reproduced in cells that have not been selected for by culture, and in an environment as similar as possible to that found *in vivo*. Realistically, this will most likely involve patient samples observed *ex vivo*. Working with PBMCs means that one can no longer be certain of a particular cell's integration site, or even whether or not it is infected. If one wishes to compare cells from different disease states, sample size is also an issue as most ACs and many HAM/TSP patients have relatively low PVLs. This could however be addressed by staining or enriching cells based on cell surface markers indicative of infection, such as ICAM-1 or CADM1.

It will be necessary to properly model plus- and minus-strand transcription at the single-cell level, accounting for factors such as cell cycle stage and cellular volume. The construction of a robust model would be possible with the data presented in this thesis, but an important extra consideration would involve accurate measurements of Tax protein expression, as the protein's presence in particular is so important to plus-strand expression. It would also be of value to include a cell surface marker in order to more accurately measure total cell volume as opposed to nuclear volume.

The correlation between cell cycle progression and viral expression, particularly in the case of *hbx*, is likely to be of crucial importance to HTLV-1. A further step could involve inhibitors such as aphidicolin to force cells into a particular cell cycle stage, and observing the effect on viral expression. Another avenue for investigation would involve exploiting the natural spike in *tax* expression during short-term culture of infected cells *in vitro* (Hanon *et al.*, 2000; Li *et al.*, 2009; Rende *et al.*, 2011). Using the methods described in this thesis, one could track the rise in viral expression, and correlate it with cell cycle stage and cell volume. The presence of uninfected cells would in this case be a bonus as, if a stain such as CADM1 were included, they would constitute a natural control experiment, experiencing the same change in conditions and stimuli, but lacking HTLV-1 expression. Combining RNA- and DNA-FISH in this context would also permit the inclusion of proviral location as a parameter. This form of combined FISH is challenging, but has been done before, for example in the study of X-chromosome inactivation (Giorgetti *et al.*, 2014).

A number of features of *hbx* expression remain unclear. The Kataoka study found consistent read-through of the minus-strand, suggesting that *hbx* may not be polyadenylated and translated in ATLL,

or that its transcription could be dysregulated (Kataoka *et al.*, 2015). It would be relatively simple to use a second probe complementary to the minus-strand after the *hbx* classical polyadenylation site. *hbx* transcription and potential processing could then be observed in single cells. Similarly, the intron of the *hbx* sp1 mRNA is sufficiently long that it would be possible to target it, allowing a more robust detection of nascent *hbx* RNAs and therefore *hbx* bursts than simply using signal intensity and size.

As has been mentioned previously, one of the limitations of using PBMCs as opposed to individual clones is the problem of observing sufficient numbers of infected cells to be able to draw robust conclusions of HTLV-1 expression. Instead of detecting smRNA-FISH through microscopy, it can also be done by flow cytometry, termed Flow-FISH. This was recently used to detect HIV-1 infected cells with a sensitivity of 1.5 infected cells per 1 million in HIV-1 patients. While it sacrifices the precision of being able to count individual transcripts, and it is debatable whether it would be able to detect *hbx*, one would be able to observe millions of cells via this method instead of thousands; this would offer significantly improved statistical power.

Finally, one built-in limitation of FISH is that it requires fixed cells. Tantale *et al.* demonstrated that it is possible to thoughtfully and elegantly characterise expression and transcriptional bursts in live cells (Tantale *et al.*, 2016). Tax bursts are currently being investigated in our lab using “timer proteins”, fluorescent proteins whose emission wavelength switch from blue to red at a set half-life. This approach has been validated by smRNA-FISH carried out by myself (Fig. 7.3), and shows great promise in investigating the kinetics of HTLV-1 transcriptional bursts in live cells.

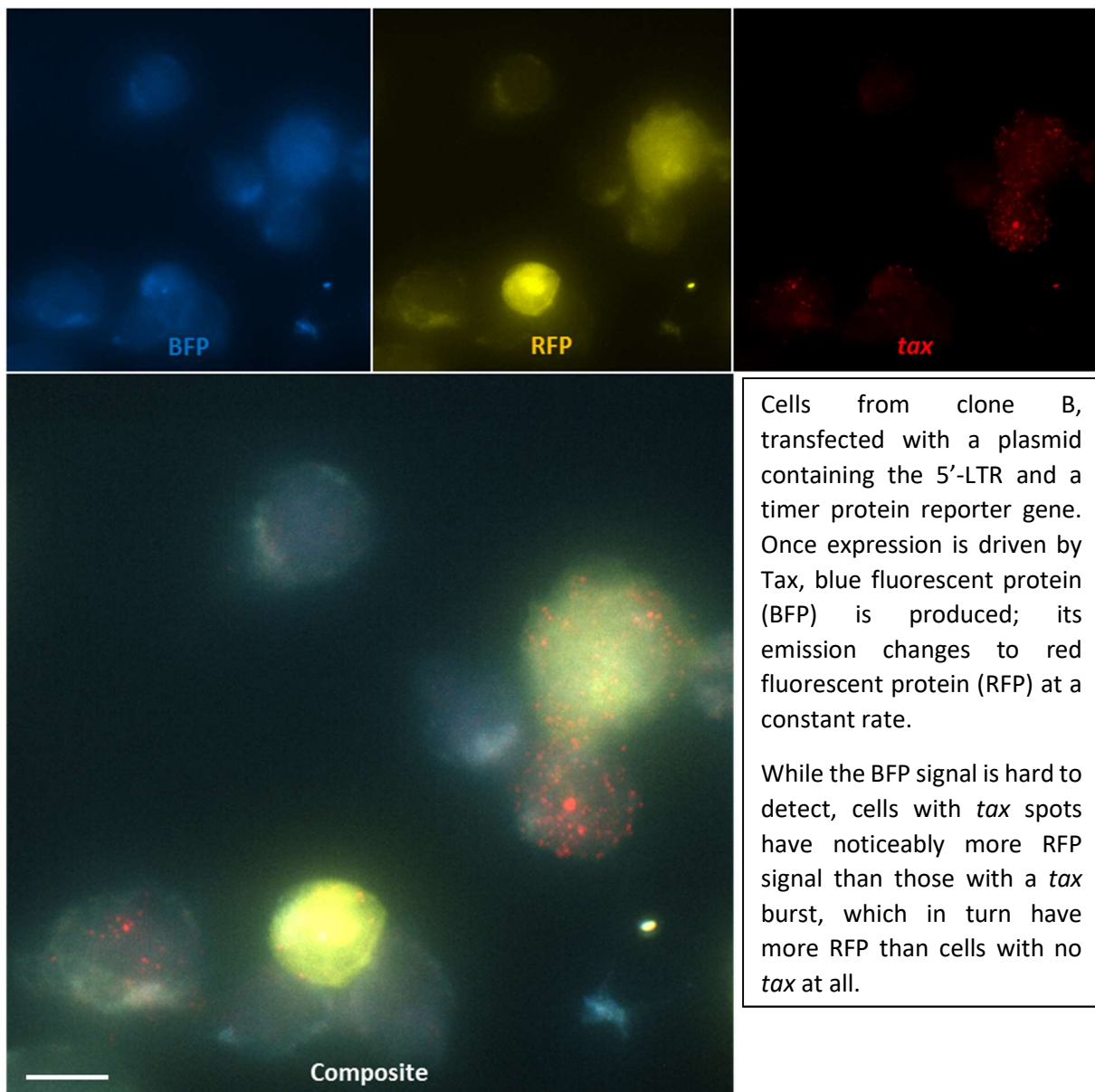


Figure 7-3 | smRNA-FISH for *tax* mRNA validates correlation between expression of Tax protein and Timer Protein

7.11 Final Conclusions

In summary, I have shown that plus-strand expression from the HTLV-1 provirus exhibits high heterogeneity between individual cells, even within the same clone, and that minus-strand expression, while much more homogeneous than the plus-strand, is not present in all cells at all times. The contrast between the levels of expression of the two strands is further demonstrated by differences between the frequency and intensity of their respective transcriptional bursts. Once expressed, the majority of *tax* and *hbx* transcripts have different fates, with the majority of *hbx* retained in the nucleus while *tax* is exported to the cytosol. My results also show that plus- and minus-strand expression do not occur independently of each other, but rather that, at the single-cell level, *tax* expression is more likely to occur in the absence of *hbx*, whose expression is in turn more likely in cells with high *tax*. Finally, proviral expression is correlated with the phase of the cell cycle, with clear implications for driving proliferation and evasion of immunosurveillance. The ability to quantify these relationships at the single-cell level is a first in the HTLV-1 field and opens exciting avenues to elucidate the phenomenon of HTLV-1 latency.

REFERENCES

- Aitken, S., Magi, S., Alhendi, A. M. N., Itoh, M., Kawaji, H., Lassmann, T., Daub, C. O., Arner, E., Carninci, P., Forrest, A. R. R., Hayashizaki, Y., Khachigian, L. M., Okada-Hatakeyama, M. and Semple, C. A. (2015) 'Transcriptional Dynamics Reveal Critical Roles for Non-coding RNAs in the Immediate-Early Response', *PLOS Computational Biology*. Edited by D. Thieffry, 11(4), p. e1004217. doi: 10.1371/journal.pcbi.1004217.
- Albrecht, B., Collins, N. D., Burniston, M. T., Nisbet, J. W., Ratner, L., Green, P. L. and Lairmore, M. D. (2000) 'Human T-Lymphotropic Virus Type 1 Open Reading Frame I p12I Is Required for Efficient Viral Infectivity in Primary Lymphocytes', *Journal of Virology*, 74(21), pp. 9828–9835. doi: 10.1128/JVI.74.21.9828-9835.2000.
- Alefantis, T., Barmak, K., Harhaj, E. W., Grant, C. and Wigdahl, B. (2003) 'Characterization of a Nuclear Export Signal within the Human T Cell Leukemia Virus Type I Transactivator Protein Tax', *Journal of Biological Chemistry*, 278(24), pp. 21814–21822. doi: 10.1074/jbc.M211576200.
- Anderson, M. G. and Dynan, W. S. (1994) 'Quantitative studies of the effect of HTLV-I Tax protein on CREB protein--DNA binding.', *Nucleic Acids Research*, 22(15), pp. 3194–201.
- Andrade, R. G., Gonçalves, P. de C., Ribeiro, M. A., Romanelli, L. C. F., Ribas, J. G., Torres, E. B., Carneiro-Proietti, A. B. de F., Barbosa-Stancioli, E. F. and Martins, M. L. (2013) 'Strong correlation between tax and HBZ mRNA expression in HAM/TSP patients: Distinct markers for the neurologic disease', *Journal of Clinical Virology*, 56(2), pp. 135–140. doi: 10.1016/j.jcv.2012.10.003.
- Arnold, J., Yamamoto, B., Li, M., Phipps, A. J., Younis, I., Lairmore, M. D. and Green, P. L. (2006) 'Enhancement of infectivity and persistence in vivo by HBZ, a natural antisense coded protein of HTLV-1', *Blood*, 107(10), pp. 3976–3982. doi: 10.1182/blood-2005-11-4551.
- Asquith, B., Mosley, A. J., Heaps, A., Tanaka, Y., Taylor, G. P., McLean, A. R. and Bangham, C. R. M. (2005) 'Quantification of the virus-host interaction in human T lymphotropic virus I infection.', *Retrovirology*, 2, p. 75. doi: 10.1186/1742-4690-2-75.
- 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. and Bangham, C. R. M. (2007) 'In vivo T lymphocyte dynamics in humans and the impact of human T-lymphotropic virus 1 infection.', *Proceedings of the National Academy of Sciences of the United States of America*, 104(19), pp. 8035–40. doi: 10.1073/pnas.0608832104.
- Avraham, R., Sas-Chen, A., Manor, O., Steinfeld, I., Shalgi, R., Tarcic, G., Bossel, N., Zeisel, A., Amit, I., Zwang, Y., Enerly, E., Russnes, H. G., Biagioni, F., Mottolese, M., Strano, S., Blandino, G., Borresen-Dale, A.-L., Pilpel, Y., Yakhini, Z., Segal, E. and Yarden, Y. (2010) 'EGF Decreases the Abundance of MicroRNAs That Restrain Oncogenic Transcription Factors', *Science Signaling*, 3(124), p. ra43-ra43. doi: 10.1126/scisignal.2000876.
- Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I. and Itzkovitz, S. (2015) 'Nuclear Retention of mRNA in Mammalian Tissues', *Cell Reports*. The Authors, 13(12), pp. 2653–2662. doi: 10.1016/j.celrep.2015.11.036.
- Bahrami, S. and Drabløs, F. (2016) 'Gene regulation in the immediate-early response process', *Advances in Biological Regulation*, 62(7491), pp. 37–49. doi: 10.1016/j.jbior.2016.05.001.
- Bai, L. and Morozov, A. V (2010) 'Gene regulation by nucleosome positioning.', *Trends in genetics* :

TIG. Elsevier Ltd, 26(11), pp. 476–83. doi: 10.1016/j.tig.2010.08.003.

Ballard, D. W., Böhnelein, E., Lowenthal, J. W., Wano, Y., Franza, B. R. and Greene, W. C. (1988) 'HTLV-I tax induces cellular proteins that activate the kappa B element in the IL-2 receptor alpha gene.', *Science*, 241(4873), pp. 1652–5.

Bangham, C. R. M. (2011) 'HTLV-1', *eLS*, (December), pp. 1–11. doi: 10.1002/9780470015902.a0023613.

Bangham, C. R. M., Araujo, A., Yamano, Y. and Taylor, G. P. (2015) 'HTLV-1-associated myelopathy/tropical spastic paraparesis', *Nature Reviews Disease Primers*, 1(June), p. 15012. doi: 10.1038/nrdp.2015.12.

Bangham, C. R. M., Meekings, K., Toulza, F., Nejmeddine, M., Majorovits, E., Asquith, B. and Taylor, G. P. (2009) 'The immune control of HTLV-1 infection: selection forces and dynamics', *Frontiers in Bioscience*, 14, pp. 2889–2903.

Baratella, M., Forlani, G., Raval, G. U., Tedeschi, A., Gout, O., Gessain, A., Tosi, G. and Accolla, R. S. (2017) 'Cytoplasmic Localization of HTLV-1 HBZ Protein: A Biomarker of HTLV-1-Associated Myelopathy/Tropical Spastic Paraparesis (HAM/TSP)', *PLOS Neglected Tropical Diseases*. Edited by F. Kashanchi, 11(1), p. e0005285. doi: 10.1371/journal.pntd.0005285.

Bas, a, Forsberg, G., Hammarstrom, S. and Hammarstrom, M.-L. (2004) 'Utility of the Housekeeping Genes 18S rRNA, beta-Actin and Glyceraldehyde-3-Phosphate-Dehydrogenase for Normalization in Real-Time Quantitative Reverse Transcriptase-Polymerase Chain Reaction Analysis of Gene Expression in Human T Lymphocytes', *Scandinavian Journal of Immunology*, 59(6), pp. 566–573. doi: 10.1111/j.0300-9475.2004.01440.x.

Basbous, J., Arpin, C., Gaudray, G., Piechaczyk, M., Devaux, C. and 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', *Journal of Biological Chemistry*, 278(44), pp. 43620–43627. doi: 10.1074/jbc.M307275200.

Batista, P. and Chang, H. (2013) 'Long Noncoding RNAs: Cellular Address Codes in Development and Disease', *Cell*. Elsevier Inc., 152(6), pp. 1298–1307. doi: 10.1016/j.cell.2013.02.012.

Battich, N., Stoeger, T. and Pelkmans, L. (2015) 'Control of Transcript Variability in Single Mammalian Cells', *Cell*, 163(7), pp. 1596–1610. doi: 10.1016/j.cell.2015.11.018.

Baydoun, H. H., Bellon, M. and Nicot, C. (2012) 'HTLV-1 Yin and Yang: Rex and p30 master regulators of viral mRNA trafficking.', *AIDS Reviews*, 10(4), pp. 195–204. doi: 10.1007/s11103-011-9767-z.Plastid.

Baydoun, H. H., Pancewicz, J., Bai, X. and Nicot, C. (2010) 'HTLV-I p30 inhibits multiple S phase entry checkpoints, decreases cyclin E-CDK2 interactions and delays cell cycle progression', *Molecular Cancer*, 9(1), p. 302. doi: 10.1186/1476-4598-9-302.

Bazarbachi, A., Plumelle, Y., Carlos Ramos, J., Tortevoeye, P., Otroock, Z., Taylor, G., Gessain, A., Harrington, W., Panelatti, G. and Hermine, O. (2010) 'Meta-Analysis on the Use of Zidovudine and Interferon-Alfa in Adult T-Cell Leukemia/Lymphoma Showing Improved Survival in the Leukemic Subtypes', *Journal of Clinical Oncology*, 28(27), pp. 4177–4183. doi: 10.1200/JCO.2010.28.0669.

Beliveau, B. J., Boettiger, A. N., Avendaño, M. S., Jungmann, R., McCole, R. B., Joyce, E. F., Kim-Kiselak, C., Bantignies, F., Fonseka, C. Y., Erceg, J., Hannan, M. A., Hoang, H. G., Colognori, D., Lee, J. T., Shih, W. M., Yin, P., Zhuang, X. and Wu, C. (2015) 'Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes', *Nature Communications*, 6(May), p. 7147. doi: 10.1038/ncomms8147.

- Bellon, M., Baydoun, H. H., Yao, Y. and Nicot, C. (2010) 'HTLV-I Tax-dependent and -independent events associated with immortalization of human primary T lymphocytes', *Blood*, 115(12), pp. 2441–2448. doi: 10.1182/blood-2009-08-241117.
- Belrose, G., Gross, A., Olindo, S., L??zin, A., Dueymes, M., Komla-Soukha, I., Smadja, D., Tanaka, Y., Willems, L., Mesnard, J. M., Peloponese, J. M. and C??saire, R. (2011) 'Effects of valproate on Tax and HBZ expression in HTLV-1 and HAM/TSP T lymphocytes', *Blood*, 118(9), pp. 2483–2491. doi: 10.1182/blood-2010-11-321364.
- Bensaude, O. (2011) 'Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity?', *Transcription*, 2(3), pp. 103–108. doi: 10.4161/trns.2.3.16172.
- Bex, F., McDowall, A., Burny, A. and Gaynor, R. (1997) 'The human T-cell leukemia virus type 1 transactivator protein Tax colocalizes in unique nuclear structures with NF-kappaB proteins.', *Journal of Virology*, 71(5), pp. 3484–97.
- Biasiotto, R., Aguiari, P., Rizzuto, R., Pinton, P., D'Agostino, D. M. and Ciminale, V. (2010) 'The p13 protein of human T cell leukemia virus type 1 (HTLV-1) modulates mitochondrial membrane potential and calcium uptake', *Biochimica et Biophysica Acta*. Elsevier B.V., 1797(6–7), pp. 945–951. doi: 10.1016/j.bbabi.2010.02.023.
- Bickmore, W. a and van Steensel, B. (2013) 'Genome architecture: domain organization of interphase chromosomes.', *Cell*. Elsevier Inc., 152(6), pp. 1270–84. doi: 10.1016/j.cell.2013.02.001.
- Billman, M. R., Rueda, D. and Bangham, C. R. M. (2017) 'Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1', *Wellcome Open Research*, 2(0), p. 87. doi: 10.12688/wellcomeopenres.12469.1.
- Boisvert, F.-M., van Koningsbruggen, S., Navascués, J. and Lamond, A. I. (2007) 'The multifunctional nucleolus.', *Nature reviews. Molecular cell biology*, 8(7), pp. 574–85. doi: 10.1038/nrm2184.
- Boulon, S., Westman, B. J., Hutten, S., Boisvert, F.-M. and Lamond, A. I. (2010) 'The nucleolus under stress.', *Molecular cell*, 40(2), pp. 216–27. doi: 10.1016/j.molcel.2010.09.024.
- Boxus, M. and Willems, L. (2009) 'Mechanisms of HTLV-1 persistence and transformation.', *British journal of cancer*. Nature Publishing Group, 101(9), pp. 1497–501. doi: 10.1038/sj.bjc.6605345.
- Boyle, S., Rodesch, M. J., Halvensleben, H. a., Jeddelloh, J. a. and Bickmore, W. a. (2011) 'Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis', *Chromosome Research*, 19(7), pp. 901–909. doi: 10.1007/s10577-011-9245-0.
- Brugiolo, M., Herzel, L. and Neugebauer, K. M. (2013) 'Counting on co-transcriptional splicing', *F1000Prime Reports*, 5(April), pp. 1–6. doi: 10.12703/P5-9.
- Calattini, S., Chevalier, S. A., Duprez, R., Bassot, S., Froment, A., Mahieux, R. and Gessain, A. (2005) 'Discovery of a new human T-cell lymphotropic virus (HTLV-3) in Central Africa.', *Retrovirology*, 2(1), p. 30. doi: 10.1186/1742-4690-2-30.
- Carter, R. S., Geyer, B. C., Xie, M., Acevedo-Suárez, C. A. and Ballard, D. W. (2001) 'Persistent Activation of NF-κB by the Tax Transforming Protein Involves Chronic Phosphorylation of IκB Kinase Subunits IKKβ and IKKγ', *Journal of Biological Chemistry*, 276(27), pp. 24445–24448. doi: 10.1074/jbc.C000777200.
- Cavallari, I., Rende, F., Bona, M. K., Sztuba-Solinska, J., Silic-Benussi, M., Tognon, M., LeGrice, S. F. J., Franchini, G., D'Agostino, D. M. and Ciminale, V. (2016) 'Expression of Alternatively Spliced Human T-Cell Leukemia Virus Type 1 mRNAs Is Influenced by Mitosis and by a Novel cis -Acting Regulatory

Sequence', *Journal of Virology*. Edited by S. R. Ross, 90(3), pp. 1486–1498. doi: 10.1128/JVI.02298-15.

Cavanagh, M.-H., Landry, S., Audet, B., Arpin-André, C., Hivin, P., Paré, M.-E., Thête, J., Wattel, E., Marriott, S. J., Mesnard, J.-M. and Barbeau, B. (2006) 'HTLV-I antisense transcripts initiating in the 3'LTR are alternatively spliced and polyadenylated.', *Retrovirology*, 3, p. 15. doi: 10.1186/1742-4690-3-15.

Cavrois, M., Leclercq, I., Gout, O., Gessain, A., Wain-Hobson, S. and 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(1), pp. 77–82. doi: 10.1038/sj.onc.1201906.

Chaumeil, J., Micsinai, M. and Skok, J. A. (2013) 'Combined Immunofluorescence and DNA FISH on 3D-preserved Interphase Nuclei to Study Changes in 3D Nuclear Organization', *Journal of Visualized Experiments*, (72), p. e50087. doi: 10.3791/50087.

Chu, Z.-L., Shin, Y.-A., Yang, J.-M., DiDonato, J. a and Ballard, D. W. (1999) 'IKK γ Mediates the Interaction of Cellular I κ B Kinases with the Tax Transforming Protein of Human T Cell Leukemia Virus Type 1', *Journal of Biological Chemistry*, 274(22), pp. 15297–15300. doi: 10.1074/jbc.274.22.15297.

Ciuffi, A., Rato, S. and Telenti, A. (2016) 'Single-Cell Genomics for Virology', *Viruses*, 8(5), p. 123. doi: 10.3390/v8050123.

Clerc, I., Hivin, P., Rubbo, P.-A., Lemasson, I., Barbeau, B. and Mesnard, J.-M. (2009) 'Propensity for HBZ-SP1 isoform of HTLV-I to inhibit c-Jun activity correlates with sequestration of c-Jun into nuclear bodies rather than inhibition of its DNA-binding activity', *Virology*. Elsevier Inc., 391(2), pp. 195–202. doi: 10.1016/j.virol.2009.06.027.

Clerc, I., Polakowski, N., André-Arpin, C., Cook, P., Barbeau, B., Mesnard, J.-M. and Lemasson, I. (2008) 'An interaction between the human T cell leukemia virus type 1 basic leucine zipper factor (HBZ) and the KIX domain of p300/CBP contributes to the down-regulation of tax-dependent viral transcription by HBZ.', *The Journal of biological chemistry*, 283(35), pp. 23903–13. doi: 10.1074/jbc.M803116200.

Coleman, R. A., Liu, Z., Darzacq, X., Tjian, R., Singer, R. H. and Lionnet, T. (2015) 'Imaging Transcription: Past, Present, and Future', *Cold Spring Harbor Symposia on Quantitative Biology*, 80, pp. 1–8. doi: 10.1101/sqb.2015.80.027201.

Conaway, R. C. and Conaway, J. W. (1993) 'General Initiation Factors for RNA Polymerase II', *Annual Review of Biochemistry*, 62(1), pp. 161–190. doi: 10.1146/annurev.bi.62.070193.001113.

Cook, L. B., Elemans, M., Rowan, A. G. and Asquith, B. (2013) 'HTLV-1: persistence and pathogenesis.', *Virology*, 435(1), pp. 131–40. doi: 10.1016/j.virol.2012.09.028.

Cook, L. B. M. (2013) 'Clonality in adult T-cell leukaemia / lymphoma', *PhD Thesis*.

Cook, L. B., Melamed, A., Niederer, H., Valganon, M., Laydon, D., Foroni, L., Taylor, G. P., Matsuoka, M. and Bangham, C. R. M. (2014) 'The role of HTLV-1 clonality, proviral structure and genomic integration site in adult T cell leukemia/lymphoma.', *Blood*. doi: 10.1182/blood-2014-02-553602.

Cook, L. B., Rowan, A. G., Melamed, A., Taylor, G. P. and Bangham, C. R. M. (2012) 'HTLV-1-infected T cells contain a single integrated provirus in natural infection', *Blood*, 120(17), pp. 3488–3490. doi: 10.1182/blood-2012-07-445593.

Currer, R., Van Duyne, R., Jaworski, E., Guendel, I., Sampey, G., Das, R., Narayanan, A. and Kashanchi, F. (2012) 'HTLV Tax: A Fascinating Multifunctional Co-Regulator of Viral and Cellular Pathways',

Frontiers in Microbiology, 3(NOV), pp. 1–24. doi: 10.3389/fmicb.2012.00406.

Daenke, S., Kermode, A. G., Hall, S. E., Taylor, G., Weber, J., Nightingale, S. and 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(1), pp. 139–46. doi: 10.1006/viro.1996.0101.

Dar, R. D., Razooky, B. S., Singh, A., Trimeloni, T. V., McCollum, J. M., Cox, C. D., Simpson, M. L. and Weinberger, L. S. (2012) 'Transcriptional burst frequency and burst size are equally modulated across the human genome.', *Proceedings of the National Academy of Sciences of the United States of America*, 109(43), pp. 17454–9. doi: 10.1073/pnas.1213530109.

Datta, A., Silverman, L., Phipps, A. J., Hiraragi, H., Ratner, L. and Lairmore, M. D. (2007) 'Human T-lymphotropic virus type-1 p30 alters cell cycle G2 regulation of T lymphocytes to enhance cell survival', *Retrovirology*, 4(1), p. 49. doi: 10.1186/1742-4690-4-49.

Demontis, M. A., Hilburn, S. and Taylor, G. P. (2012) '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 Research and Human Retroviruses*, 28(0), p. 120925062932004. doi: 10.1089/aid.2012.0132.

Deng, Q., Ramskold, D., Reinius, B. and Sandberg, R. (2014) 'Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells', *Science*, 343(6167), pp. 193–196. doi: 10.1126/science.1245316.

Der, S. D., Yang, Y.-L., Weissmann, C. and Williams, B. R. G. (1997) 'A double-stranded RNA-activated protein kinase-dependent pathway mediating stress-induced apoptosis', *Proceedings of the National Academy of Sciences*, 94(7), pp. 3279–3283. doi: 10.1073/pnas.94.7.3279.

Dey, S. S., Foley, J. E., Limsirichai, P., Schaffer, D. V. and Arkin, A. P. (2015) 'Orthogonal control of expression mean and variance by epigenetic features at different genomic loci', *Molecular Systems Biology*, 11(5), pp. 806–806. doi: 10.15252/msb.20145704.

Easley, R., Carpio, L., Guendel, I., Klase, Z., Choi, S., Kehn-Hall, K., Brady, J. N. and Kashanchi, F. (2010) 'Human T-Lymphotropic Virus Type 1 Transcription and Chromatin-Remodeling Complexes', *Journal of Virology*, 84(9), pp. 4755–4768. doi: 10.1128/JVI.00851-09.

Ego, T., Ariumi, Y. and Shimotohno, K. (2002) 'The interaction of HTLV-1 Tax with HDAC1 negatively regulates the viral gene expression', *Oncogene*, 21(47), pp. 7241–7246. doi: 10.1038/sj.onc.1205701.

Enose-Akahata, Y., Abrams, A., Massoud, R., Bialuk, I., Johnson, K. R., Green, P. L., Maloney, E. M. and Jacobson, S. (2013) 'Humoral immune response to HTLV-1 basic leucine zipper factor (HBZ) in HTLV-1-infected individuals', *Retrovirology*, 10(1), p. 19. doi: 10.1186/1742-4690-10-19.

Enose-Akahata, Y., Caruso, B., Haner, B., Charlip, E., Nair, G., Massoud, R., Billioux, B. J., Ohayon, J., Switzer, W. M. and Jacobson, S. (2016) 'Development of neurologic diseases in a patient with primate T lymphotropic virus type 1 (PTLV-1)', *Retrovirology*. BioMed Central, 13(1), p. 56. doi: 10.1186/s12977-016-0290-9.

Fan, J., Ma, G., Nosaka, K., Tanabe, J., Satou, Y., Koito, A., Wain-Hobson, S., Vartanian, J.-P. and Matsuoka, M. (2010) 'APOBEC3G Generates Nonsense Mutations in Human T-Cell Leukemia Virus Type 1 Proviral Genomes In Vivo', *Journal of Virology*, 84(14), pp. 7278–7287. doi: 10.1128/JVI.02239-09.

Filippone, C., Betsem, E., Tortevoeye, P., Cassar, O., Bassot, S., Froment, A., Fontanet, A. and Gessain, A. (2015) 'A Severe Bite From a Nonhuman Primate Is a Major Risk Factor for HTLV-1 Infection in Hunters From Central Africa.', *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 60(11), pp. 1667–76. doi: 10.1093/cid/civ145.

- Finlan, L. E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J. R. and Bickmore, W. A. (2008) 'Recruitment to the Nuclear Periphery Can Alter Expression of Genes in Human Cells', *PLoS Genetics*. Edited by W. Reik, 4(3), p. e1000039. doi: 10.1371/journal.pgen.1000039.
- Fryrear, K. A., Guo, X., Kerscher, O. and Semmes, O. J. (2012) 'The Sumo-targeted ubiquitin ligase RNF4 regulates the localization and function of the HTLV-1 oncoprotein Tax', *Blood*, 119(5), pp. 1173–1181. doi: 10.1182/blood-2011-06-358564.
- Fujii, M., Tsuchiya, H., Chuhjo, T., Minamino, T., Miyamoto, K. and Seiki, M. (1994) 'Serum response factor has functional roles both in indirect binding to the CA₁G box and in the transcriptional activation function of human T-cell leukemia virus type I Tax.', *Journal of Virology*, 68(11), pp. 7275–83. doi: papers3://publication/uuid/FC1538D2-57ED-42D2-BDFF-D03891258D84.
- Gaudray, G., Gachon, F., Basbous, J., Biard-Piechaczyk, M., Devaux, C. and 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(24), pp. 12813–12822. doi: 10.1128/JVI.76.24.12813-12822.2002.
- Gazon, H., Lemasson, I., Polakowski, N., Cesaire, R., Matsuoka, M., Barbeau, B., Mesnard, J.-M. and Peloponese, J.-M. (2012) 'Human T-Cell Leukemia Virus Type 1 (HTLV-1) bZIP Factor Requires Cellular Transcription Factor JunD To Upregulate HTLV-1 Antisense Transcription from the 3' Long Terminal Repeat', *Journal of Virology*, 86(17), pp. 9070–9078. doi: 10.1128/JVI.00661-12.
- Gessain, A. and Cassar, O. (2012) 'Epidemiological Aspects and World Distribution of HTLV-1 Infection.', *Frontiers in microbiology*, 3(NOV), p. 388. doi: 10.3389/fmicb.2012.00388.
- Ghez, D., Lepelletier, Y., Lambert, S., Fourneau, J.-M., Blot, V., Janvier, S., Arnulf, B., van Endert, P. M., Heveker, N., Pique, C. and Hermine, O. (2006) 'Neuropilin-1 Is Involved in Human T-Cell Lymphotropic Virus Type 1 Entry', *Journal of Virology*, 80(14), pp. 6844–6854. doi: 10.1128/JVI.02719-05.
- Giam, C.-Z. and Semmes, O. (2016) 'HTLV-1 Infection and Adult T-Cell Leukemia/Lymphoma—A Tale of Two Proteins: Tax and HBZ', *Viruses*, 8(6), p. 161. doi: 10.3390/v8060161.
- Gibcus, J. H. and Dekker, J. (2013) 'The hierarchy of the 3D genome.', *Molecular cell*. Elsevier Inc., 49(5), pp. 773–82. doi: 10.1016/j.molcel.2013.02.011.
- Gillet, N. A., Cook, L., Laydon, D. J., Hlela, C., Verdonck, K., Alvarez, C., Gotuzzo, E., Clark, D., Farré, L., Bittencourt, A., Asquith, B., Taylor, G. P. and Bangham, C. R. M. (2013) 'Strongyloidiasis and Infective Dermatitis Alter Human T Lymphotropic Virus-1 Clonality in vivo', *PLoS Pathogens*. Edited by W. C. Gause, 9(4), p. e1003263. doi: 10.1371/journal.ppat.1003263.
- Gillet, N. a, Malani, N., Melamed, A., Gormley, N., Carter, R., Bentley, D., Berry, C., Bushman, F. D., Taylor, G. P. and Bangham, C. R. M. (2011) 'The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones.', *Blood*, 117(11), pp. 3113–22. doi: 10.1182/blood-2010-10-312926.
- Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., Tiana, G. and Heard, E. (2014) 'Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription', *Cell*, 157(4), pp. 950–963. doi: 10.1016/j.cell.2014.03.025.
- Goncalves, D. U., Proietti, F. A., Ribas, J. G. R., Araujo, M. G., Pinheiro, S. R., Guedes, A. C. and Carneiro-Proietti, A. B. F. (2010) 'Epidemiology, Treatment, and Prevention of Human T-Cell Leukemia Virus Type 1-Associated Diseases', *Clinical Microbiology Reviews*, 23(3), pp. 577–589. doi: 10.1128/CMR.00063-09.
- Good, L., Maggirwar, S. B. and Sun, S. C. (1996) 'Activation of the IL-2 gene promoter by HTLV-I tax

involves induction of NF-AT complexes bound to the CD28-responsive element.’, *EMBO*, 15(14), pp. 3744–50.

Goon, P. K. C., Biancardi, A., Fast, N., Igakura, T., Hanon, E., Mosley, A. J., Asquith, B., Gould, K. G., Marshall, S., Taylor, G. P. and Bangham, C. R. M. (2004) ‘Human T cell lymphotropic virus (HTLV) type-1-specific CD8+ T cells: frequency and immunodominance hierarchy.’, *The Journal of infectious diseases*, 189(12), pp. 2294–8. doi: 10.1086/420832.

Gotuzzo, E., Terashima, a, Alvarez, H., Tello, R., Infante, R., Watts, D. M. and Freedman, D. O. (1999) ‘Strongyloides stercoralis hyperinfection associated with human T cell lymphotropic virus type-1 infection in Peru.’, *The American journal of tropical medicine and hygiene*, 60(1), pp. 146–9.

Greenberg, M. E. and Ziff, E. B. (1984) ‘Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene’, *Nature*, 311(5985), pp. 433–438. doi: 10.1038/311433a0.

Gröne, M., Hoffmann, E., Berchtold, S., Cullen, B. R. and 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(1), pp. 144–152. doi: 10.1006/viro.1994.1518.

Gröne, M., Koch, C. and Grassmann, R. (1996) ‘The HTLV-1 Rex protein induces nuclear accumulation of unspliced viral RNA by avoiding intron excision and degradation.’, *Virology*, 218(2), pp. 316–25. doi: 10.1006/viro.1996.0200.

Grün, D., Kester, L. and van Oudenaarden, A. (2014) ‘Validation of noise models for single-cell transcriptomics’, *Nature Methods*, 11(6), pp. 637–640. doi: 10.1038/nmeth.2930.

Gutiérrez, G., Rodríguez, S., de Brogniez, A., Gillet, N., Golime, R., Burny, A., Jaworski, J.-P., Alvarez, I., Vagnoni, L., Trono, K. and Willems, L. (2014) ‘Vaccination against δ -Retroviruses: The Bovine Leukemia Virus Paradigm’, *Viruses*, 6(6), pp. 2416–2427. doi: 10.3390/v6062416.

Haller, K., Wu, Y., Derow, E., Schmitt, I., Jeang, K.-T. and Grassmann, R. (2002) ‘Physical Interaction of Human T-Cell Leukemia Virus Type 1 Tax with Cyclin-Dependent Kinase 4 Stimulates the Phosphorylation of Retinoblastoma Protein’, *Molecular and Cellular Biology*, 22(10), pp. 3327–3338. doi: 10.1128/MCB.22.10.3327-3338.2002.

Hanon, E., Hall, S., Taylor, G. P., Saito, M., Davis, R., Tanaka, Y., Usuku, K., Osame, M., Weber, J. N. and Bangham, C. R. (2000) ‘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(4), pp. 1386–92.

Harhaj, E. W. and Sun, S.-C. (1999) ‘IKK γ Serves as a Docking Subunit of the I κ B Kinase (IKK) and Mediates Interaction of IKK with the Human T-cell Leukemia Virus Tax Protein’, *Journal of Biological Chemistry*, 274(33), pp. 22911–22914. doi: 10.1074/jbc.274.33.22911.

Harrod, R., Kuo, Y.-L., Tang, Y., Yao, Y., Vassilev, A., Nakatani, Y. and Giam, C.-Z. (2000) ‘p300 and p300/cAMP-responsive Element-binding Protein Associated Factor Interact with Human T-cell Lymphotropic Virus Type-1 Tax in a Multi-histone Acetyltransferase/Activator-Enhancer Complex’, *Journal of Biological Chemistry*, 275(16), pp. 11852–11857. doi: 10.1074/jbc.275.16.11852.

Healy, S., Khan, P. and Davie, J. R. (2013) ‘Immediate early response genes and cell transformation’, *Pharmacology & Therapeutics*. Elsevier Inc., 137(1), pp. 64–77. doi: 10.1016/j.pharmthera.2012.09.001.

Hieshima, K., Nagakubo, D., Shigeta, A., Tanaka, Y., Hoshino, H., Tsukasaki, K., Yamada, Y. and Yoshie, O. (2011) ‘c-Maf suppresses human T-cell leukemia virus type 1 Tax by competing for CREB-binding protein’, *Cancer Science*, 102(4), pp. 890–894. doi: 10.1111/j.1349-7006.2011.01873.x.

- Hilburn, S., Rowan, A., Demontis, M.-A., MacNamara, A., Asquith, B., Bangham, C. R. and 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', *Journal of Infectious Diseases*, 203(4), pp. 529–536. doi: 10.1093/infdis/jiq078.
- Hino, S. (2011) 'Establishment of the milk-borne transmission as a key factor for the peculiar endemicity of human T-lymphotropic virus type 1 (HTLV-1): the ATL Prevention Program Nagasaki.', *Proceedings of the Japan Academy. Series B, Physical and biological sciences*, 87(4), pp. 152–66. doi: 10.2183/pjab.87.152.
- Hirata, T., Uchima, N., Kishimoto, K., Zaha, O., Kinjo, N., Hokama, A., Sakugawa, H., Kinjo, F. and Fujita, J. (2006) 'Impairment of host immune response against *strongyloides stercoralis* by human T cell lymphotropic virus type 1 infection.', *The American journal of tropical medicine and hygiene*, 74(2), pp. 246–9.
- Hivin, P., Basbous, J., Raymond, F., Henaff, D., Arpin-André, C., Robert-Hebmann, V., Barbeau, B. and Mesnard, J.-M. (2007) 'The HBZ-SP1 isoform of human T-cell leukemia virus type I represses JunB activity by sequestration into nuclear bodies.', *Retrovirology*, 4, p. 14. doi: 10.1186/1742-4690-4-14.
- Hivin, P., Frédéric, M., Arpin-André, C., Basbous, J., Gay, B., Thébault, S. and Mesnard, J.-M. (2005) 'Nuclear localization of HTLV-I bZIP factor (HBZ) is mediated by three distinct motifs.', *Journal of Cell Science*, 118(Pt 7), pp. 1355–62. doi: 10.1242/jcs.01727.
- Hobson, D. J., Wei, W., Steinmetz, L. M. and Svejstrup, J. Q. (2012) 'RNA Polymerase II Collision Interrupts Convergent Transcription', *Molecular Cell*. Elsevier Inc., 48(3), pp. 365–374. doi: 10.1016/j.molcel.2012.08.027.
- Hoshida, Y., Li, T., Dong, Z., Tomita, Y., Yamauchi, A., Hanai, J. and Aozasa, K. (2001) 'Lymphoproliferative disorders in renal transplant patients in Japan.', *International Journal of Cancer*, 91(6), pp. 869–75.
- Houseley, J. and Tollervey, D. (2009) 'The Many Pathways of RNA Degradation', *Cell*. Elsevier Inc., 136(4), pp. 763–776. doi: 10.1016/j.cell.2009.01.019.
- Igakura, T., Stinchcombe, J. C., Goon, P. K., Taylor, G. P., Weber, J. N., Griffiths, G. M., Tanaka, Y., Osame, M. and Bangham, C. R. M. (2003) 'Spread of HTLV-1 Between Lymphocytes by Virus-Induced Polarization of the Cytoskeleton', *Science*, 299(2003), pp. 1713–1716. doi: 10.1029/2001PA-000633.
- Ikezoe, T., Nishioka, C., Bandobashi, K., Yang, Y., Kuwayama, Y., Adachi, Y., Takeuchi, T., Koeffler, H. P. and Taguchi, H. (2007) 'Longitudinal inhibition of PI3K/Akt/mTOR signaling by LY294002 and rapamycin induces growth arrest of adult T-cell leukemia cells', *Leukemia Research*, 31(5), pp. 673–682. doi: 10.1016/j.leukres.2006.08.001.
- Iyer, N. G., Özdag, H. and Caldas, C. (2004) 'p300/CBP and cancer', *Oncogene*, 23(24), pp. 4225–4231. doi: 10.1038/sj.onc.1207118.
- Jeang, K. T., Boros, I., Brady, J., Radonovich, M. and Khoury, G. (1988) 'Characterization of cellular factors that interact with the human T-cell leukemia virus type I p40x-responsive 21-base-pair sequence.', *Journal of Virology*, 62(12), pp. 4499–509.
- 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. and Bangham, C. R. (1999) 'HLA alleles determine human T-lymphotropic virus-I (HTLV-I) proviral load and the risk of HTLV-I-associated myelopathy', *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), pp. 3848–53.
- Johnson, J. M., Nicot, C., Fullen, J., Ciminale, V., Casareto, L., Mulloy, J. C., Jacobson, S. and 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', *Journal of Virology*, 75(13), pp. 6086–6094. doi: 10.1128/JVI.75.13.6086.
- Jones, K. S., Lambert, S., Bouttier, M., Bénit, L., Ruscetti, F. W., Hermine, O. and 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(12), pp. 794–810. doi: 10.3390/v3060794.
- Jones, K. S., Petrow-Sadowski, C., Bertolette, D. C., Huang, Y. and Ruscetti, F. W. (2005) 'Heparan Sulfate Proteoglycans Mediate Attachment and Entry of Human T-Cell Leukemia Virus Type 1 Virions into CD4+ T Cells', *Journal of Virology*, 79(20), pp. 12692–12702. doi: 10.1128/JVI.79.20.12692-12702.2005.
- Jones, T. R., Carpenter, A. and Golland, P. (2005) 'Voronoi-Based Segmentation of Cells on Image Manifolds', in *Lecture Notes in Computer Science*, pp. 535–543. doi: 10.1007/11569541_54.
- Josefsson, L., Palmer, S., Faria, N. R., Lemey, P., Casazza, J., Ambrozak, D., Kearney, M., Shao, W., Kottlil, S., Sneller, M., Mellors, J., Coffin, J. M. and Maldarelli, F. (2013) 'Single Cell Analysis of Lymph Node Tissue from HIV-1 Infected Patients Reveals that the Majority of CD4+ T-cells Contain One HIV-1 DNA Molecule', *PLoS Pathogens*, 9(6). doi: 10.1371/journal.ppat.1003432.
- Kalo, A., Kanter, I., Shraga, A., Sheinberger, J., Tzemach, H., Kinor, N., Singer, R. H., Lionnet, T. and Shav-Tal, Y. (2015) 'Cellular Levels of Signaling Factors Are Sensed by β -actin Alleles to Modulate Transcriptional Pulse Intensity', *Cell Reports*, 11(3), pp. 419–432. doi: 10.1016/j.celrep.2015.03.039.
- Kalyanaraman, V. S., Sarngadharan, M. G., Robert-Guroff, M., Miyoshi, I., Golde, D. and 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(4572), pp. 571–3.
- Kamentsky, L., Jones, T. R., Fraser, A., Bray, M. A., Logan, D. J., Madden, K. L., Ljosa, V., Rueden, C., Eliceiri, K. W. and Carpenter, A. E. (2011) 'Improved structure, function and compatibility for cellprofiler: Modular high-throughput image analysis software', *Bioinformatics*, 27(8), pp. 1179–1180. doi: 10.1093/bioinformatics/btr095.
- Kataoka, K., Nagata, Y., Kitanaka, A., Shiraishi, Y., Shimamura, T., Yasunaga, J., Totoki, Y., Chiba, K., Sato-Otsubo, A., Nagae, G., Ishii, R., Muto, S., Kotani, S., Watatani, Y., Takeda, J., Sanada, M., Tanaka, H., Suzuki, H., Sato, Y., Shiozawa, Y., Yoshizato, T., Yoshida, K., Makishima, H., Iwanaga, M., Ma, G., Nosaka, K., Hishizawa, M., Itonaga, H., Imaizumi, Y., Munakata, W., Ogasawara, H., Sato, T., Sasai, K., Muramoto, K., Penova, M., Kawaguchi, T., Nakamura, H., Hama, N., Shide, K., Kubuki, Y., Hidaka, T., Kameda, T., Nakamaki, T., Ishiyama, K., Miyawaki, S., Yoon, S.-S., Tobinai, K., Miyazaki, Y., Takaori-Kondo, A., Matsuda, F., Takeuchi, K., Nureki, O., Aburatani, H., Watanabe, T., Shibata, T., Matsuoka, M., Miyano, S., Shimoda, K. and Ogawa, S. (2015) 'Integrated molecular analysis of adult T cell leukemia/lymphoma', *Nature Genetics*, 47(11), pp. 1304–1315. doi: 10.1038/ng.3415.
- Kato, K. and Akashi, K. (2015) 'Recent Advances in Therapeutic Approaches for Adult T-cell Leukemia/Lymphoma', *Viruses*, 7(12), pp. 6604–6612. doi: 10.3390/v7122960.
- Kattan, T., MacNamara, A., Rowan, A. G., Nose, H., Mosley, A. J., Tanaka, Y., Taylor, G. P., Asquith, B. and Bangham, C. R. M. (2009) 'The avidity and lytic efficiency of the CTL response to HTLV-1.', *Journal of immunology (Baltimore, Md. : 1950)*, 182(9), pp. 5723–9. doi: 10.4049/jimmunol.0900069.
- Kawatsuki, A., Yasunaga, J., Mitobe, Y., Green, P. L. and Matsuoka, M. (2016) 'HTLV-1 bZIP factor protein targets the Rb/E2F-1 pathway to promote proliferation and apoptosis of primary CD4+ T cells', *Oncogene*, 35(34), pp. 4509–4517. doi: 10.1038/onc.2015.510.
- Kazanji, M., Mouinga-Ondémé, A., Lekana-Douki-Etenna, S., Caron, M., Makuwa, M., Mahieux, R.

- and Gessain, A. (2015) 'Origin of HTLV-1 in hunters of nonhuman primates in Central Africa.', *The Journal of infectious diseases*, 211(3), pp. 361–5. doi: 10.1093/infdis/jiu464.
- Kehn, K., Fuente, C. D. La, Strouss, K., Berro, R., Jiang, H., Brady, J., Mahieux, R., Pumfery, A., Bottazzi, M. E. and Kashanchi, F. (2005) 'The HTLV-I Tax oncoprotein targets the retinoblastoma protein for proteasomal degradation', *Oncogene*, 24(4), pp. 525–540. doi: 10.1038/sj.onc.1208105.
- Kim, S.-J., Nair, A. M., Fernandez, S., Mathes, L. and Lairmore, M. D. (2006) 'Enhancement of LFA-1-mediated T cell adhesion by human T lymphotropic virus type 1 p12I.', *Journal of Immunology*, 176(9), pp. 5463–70. doi: 10.4049/jimmunol.176.9.5463.
- Kinpara, S., Hasegawa, A., Utsunomiya, A., Nishitsuji, H., Furukawa, H., Masuda, T. and Kannagi, M. (2009) 'Stromal cell-mediated suppression of human T-cell leukemia virus type 1 expression in vitro and in vivo by type I interferon.', *Journal of virology*, 83(10), pp. 5101–8. doi: 10.1128/JVI.02564-08.
- Kinpara, S., Kijiyama, M., Takamori, A., Hasegawa, A., Sasada, A., Masuda, T., Tanaka, Y., Utsunomiya, A. and Kannagi, M. (2013) 'Interferon- α (IFN- α) suppresses HTLV-1 gene expression and cell cycling, while IFN- α combined with zidovudin induces p53 signaling and apoptosis in HTLV-1-infected cells.', *Retrovirology*. *Retrovirology*, 10(1), p. 52. doi: 10.1186/1742-4690-10-52.
- Koralnik, I. J., Fullen, J. and 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.', *Journal of Virology*, 67(4), pp. 2360–6.
- Kornberg, R. D. (2007) 'The molecular basis of eukaryotic transcription', *Proceedings of the National Academy of Sciences of the United States of America*, 104(32), pp. 12955–61. doi: 10.1073/pnas.0704138104.
- Kuhlmann, A.-S., Villaudy, J., Gazzolo, L., Castellazzi, M., Mesnard, J.-M. and Duc Dodon, M. (2007) 'HTLV-1 HBZ cooperates with JunD to enhance transcription of the human telomerase reverse transcriptase gene (hTERT)', *Retrovirology*, 4(1), p. 92. doi: 10.1186/1742-4690-4-92.
- Kulkarni, A., Mateus, M., Thinnes, C. C., McCullagh, J. S., Schofield, C. J., Taylor, G. P. and Bangham, C. R. M. (2017) 'Glucose metabolism and oxygen availability govern reactivation from latency of the human retrovirus HTLV-1', *Cell Chemical Biology*. doi: 10.1016/j.chembiol.2017.08.016.
- Kwok, R. P. S., Laurance, M. E., Lundblad, J. R., Goldman, P. S., Shih, H., Connor, L. M., Marriott, S. J. and Goodman, R. H. (1996) 'Control of cAMP-regulated enhancers by the viral transactivator Tax through CREB and the co-activator CBP', *Nature*, 380(6575), pp. 642–646. doi: 10.1038/380642a0.
- Kwok, S., Ehrlich, G., Poiesz, B., Kalish, R. and Sninsky, J. J. (1988) 'Enzymatic amplification of HTLV-1 viral sequences from peripheral blood mononuclear cells and infected tissues.', *Blood*, 72(4), pp. 1117–23.
- Laydon, D. J., Melamed, A., Sim, A., Gillet, N. A., Sim, K., Darko, S., Kroll, J. S., Douek, D. C., Price, D. A., Bangham, C. R. M. and Asquith, B. (2014) 'Quantification of HTLV-1 Clonality and TCR Diversity', *PLoS Computational Biology*. Edited by B. Peters, 10(6), p. e1003646. doi: 10.1371/journal.pcbi.1003646.
- Lee, B., Tanaka, Y. and Tozawa, H. (1989) 'Monoclonal antibody defining tax protein of human T-cell leukemia virus type-I.', *The Tohoku journal of experimental medicine*, 157(1), pp. 1–11.
- Lemasson, I., Lewis, M. R., Polakowski, N., Hivin, P., Cavanagh, M.-H., Thebault, S., Barbeau, B., Nyborg, J. K. and Mesnard, J.-M. (2007) 'Human T-Cell Leukemia Virus Type 1 (HTLV-1) bZIP Protein Interacts with the Cellular Transcription Factor CREB To Inhibit HTLV-1 Transcription', *Journal of Virology*, 81(4), pp. 1543–1553. doi: 10.1128/JVI.00480-06.

- Lemasson, I., Polakowski, N. J., Laybourn, P. J. and 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', *Journal of Biological Chemistry*, 277(51), pp. 49459–49465. doi: 10.1074/jbc.M209566200.
- Lemasson, I., Polakowski, N. J., Laybourn, P. J. and Nyborg, J. K. (2004) 'Transcription Regulatory Complexes Bind the Human T-Cell Leukemia Virus 5' and 3' Long Terminal Repeats To Control Gene Expression', *Molecular and Cellular Biology*, 24(14), pp. 6117–6126. doi: 10.1128/MCB.24.14.6117-6126.2004.
- Li, M., Kesic, M., Yin, H., Yu, L. and Green, P. L. (2009) 'Kinetic Analysis of Human T-Cell Leukemia Virus Type 1 Gene Expression in Cell Culture and Infected Animals', *Journal of Virology*, 83(8), pp. 3788–3797. doi: 10.1128/JVI.02315-08.
- Lichtman, J. W. and Conchello, J.-A. (2005) 'Fluorescence microscopy', *Nature Methods*, 2(12), pp. 910–919. doi: 10.1038/nmeth817.
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S. and Dekker, J. (2009) 'Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome', *Science*, 326(5950), pp. 289–293. doi: 10.1126/science.1181369.
- Lindblad, J. and Bengtsson, E. (2001) 'A Comparison of Methods for Estimation of Intensity Non uniformities in 2D and 3D Microscope Images of Fluorescence Stained Cells', *Proceedings of the 12th Scandinavian Conference on Image Analysis (SCIA)*, (January), pp. 264–271.
- Lu, H., Pise-Masison, C. A., Fletcher, T. M., Schiltz, R. L., Nagaich, A. K., Radonovich, M., Hager, G., Cole, P. A. and Brady, J. N. (2002) 'Acetylation of Nucleosomal Histones by p300 Facilitates Transcription from Tax-Responsive Human T-Cell Leukemia Virus Type 1 Chromatin Template', *Molecular and Cellular Biology*, 22(13), pp. 4450–4462. doi: 10.1128/MCB.22.13.4450-4462.2002.
- Lu, H., Pise-Masison, C. A., Linton, R., Park, H. U., Schiltz, R. L., Sartorelli, V. and Brady, J. N. (2004) 'Tax Relieves Transcriptional Repression by Promoting Histone Deacetylase 1 Release from the Human T-Cell Leukemia Virus Type 1 Long Terminal Repeat', *Journal of Virology*, 78(13), pp. 6735–6743. doi: 10.1128/JVI.78.13.6735-6743.2004.
- Lusic, M., Marini, B., Ali, H., Lucic, B., Luzzati, R. and Giacca, M. (2013) 'Proximity to PML nuclear bodies regulates HIV-1 latency in CD4+ T cells', *Cell Host and Microbe*. Elsevier Inc., 13(6), pp. 665–677. doi: 10.1016/j.chom.2013.05.006.
- Ma, G., Yasunaga, J., Akari, H. and Matsuoka, M. (2015) 'TCF1 and LEF1 act as T-cell intrinsic HTLV-1 antagonists by targeting Tax', *Proceedings of the National Academy of Sciences*, 112(7), pp. 2216–2221. doi: 10.1073/pnas.1419198112.
- Ma, G., Yasunaga, J., Fan, J., Yanagawa, S. and Matsuoka, M. (2013) 'HTLV-1 bZIP factor dysregulates the Wnt pathways to support proliferation and migration of adult T-cell leukemia cells', *Oncogene*, 32(36), pp. 4222–4230. doi: 10.1038/onc.2012.450.
- Ma, G., Yasunaga, J. and Matsuoka, M. (2016) 'Multifaceted functions and roles of HBZ in HTLV-1 pathogenesis', *Retrovirology*. BioMed Central, pp. 1–9. doi: 10.1186/s12977-016-0249-x.
- Macnamara, A., Rowan, A., Hilburn, S., Kadolsky, U., Fujiwara, H., Suemori, K., Yasukawa, M., Taylor, G., Bangham, C. R. M. and Asquith, B. (2010) 'HLA class I binding of HBZ determines outcome in HTLV-1 infection.', *PLoS pathogens*, 6(9), p. e1001117. doi: 10.1371/journal.ppat.1001117.
- Maertens, G. N. (2016) 'B'-protein phosphatase 2A is a functional binding partner of delta-retroviral

- integrase', *Nucleic Acids Research*, 44(1), pp. 364–376. doi: 10.1093/nar/gkv1347.
- Mahieux, R. (2015) 'A vaccine against HTLV-1 HBZ makes sense', *Blood*, 126(9), pp. 1052–1053. doi: 10.1182/blood-2015-06-652040.
- Manel, N., Kim, F. J., Kinet, S., Taylor, N., Sitbon, M. and Battini, J.-L. (2003) 'The ubiquitous glucose transporter GLUT-1 is a receptor for HTLV.', *Cell*, 115(4), pp. 449–59.
- Manivannan, K., Rowan, A. G., Tanaka, Y., Taylor, G. P. and Bangham, C. R. M. (2016) 'CADM1/TSLC1 Identifies HTLV-1-Infected Cells and Determines Their Susceptibility to CTL-Mediated Lysis', *PLoS Pathogens*, 12(4). doi: 10.1371/journal.ppat.1005560.
- Marini, B., Kertesz-Farkas, A., Ali, H., Lucic, B., Lisek, K., Manganaro, L., Pongor, S., Luzzati, R., Recchia, A., Mavilio, F., Giacca, M. and Lusic, M. (2015) 'Nuclear architecture dictates HIV-1 integration site selection', *Nature*. Nature Publishing Group, 521(7551), pp. 227–231. doi: 10.1038/nature14226.
- Martin-Latil, S., Gnadig, N. F., Mallet, A., Desdouits, M., Guivel-Benhassine, F., Jeannin, P., Prevost, M.-C., Schwartz, O., Gessain, A., Ozden, S. and Ceccaldi, P.-E. (2012) 'Transcytosis of HTLV-1 across a tight human epithelial barrier and infection of subepithelial dendritic cells', *Blood*, 120(3), pp. 572–580. doi: 10.1182/blood-2011-08-374637.
- Martin, F., Taylor, G. P. and Jacobson, S. (2014) 'Inflammatory manifestations of HTLV-1 and their therapeutic options', *Expert Review of Clinical Immunology*, 10(11), pp. 1531–1546. doi: 10.1586/1744666X.2014.966690.
- Matsumoto, J., Ohshima, T., Isono, O. and Shimotohno, K. (2005) 'HTLV-1 HBZ suppresses AP-1 activity by impairing both the DNA-binding ability and the stability of c-Jun protein.', *Oncogene*, 24(6), pp. 1001–10. doi: 10.1038/sj.onc.1208297.
- Matsumoto, K., Shibata, H., Fujisawa, J. I., Inoue, H., Hakura, A., Tsukahara, T. and Fujii, M. (1997) 'Human T-cell leukemia virus type 1 Tax protein transforms rat fibroblasts via two distinct pathways.', *Journal of Virology*, 71(6), pp. 4445–51.
- Matsuoka, M. and Jeang, K.-T. (2007) 'Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation.', *Nature reviews. Cancer*, 7(4), pp. 270–80. doi: 10.1038/nrc2111.
- Matsuoka, M. and Jeang, K.-T. (2011) 'Human T-cell leukemia virus type 1 (HTLV-1) and leukemic transformation: viral infectivity, Tax, HBZ and therapy.', *Oncogene*. Nature Publishing Group, 30(12), pp. 1379–89. doi: 10.1038/onc.2010.537.
- Matsuoka, M. and Yasunaga, J. (2013) 'Human T-cell leukemia virus type 1: replication, proliferation and propagation by Tax and HTLV-1 bZIP factor', *Current Opinion in Virology*, 3(6), pp. 684–691. doi: 10.1016/j.coviro.2013.08.010.
- Mazurov, D., Ilinskaya, A., Heidecker, G., Lloyd, P. and Derse, D. (2010) 'Quantitative Comparison of HTLV-1 and HIV-1 Cell-to-Cell Infection with New Replication Dependent Vectors', *PLoS Pathogens*. Edited by T. J. Hope, 6(2), p. e1000788. doi: 10.1371/journal.ppat.1000788.
- McClure, M. a, Johnson, M. S., Feng, D. F. and Doolittle, R. F. (1988) 'Sequence comparisons of retroviral proteins: relative rates of change and general phylogeny.', *Proceedings of the National Academy of Sciences*, 85(8), pp. 2469–2473. doi: 10.1073/pnas.85.8.2469.
- Melamed, A., Laydon, D. J., Gillet, N. a, Tanaka, Y., Taylor, G. P. and Bangham, C. R. M. (2013) 'Genome-wide determinants of proviral targeting, clonal abundance and expression in natural HTLV-1 infection.', *PLoS pathogens*, 9(3), p. e1003271. doi: 10.1371/journal.ppat.1003271.
- Mitobe, Y., Yasunaga, J. -i., Furuta, R. and Matsuoka, M. (2015) 'HTLV-1 bZIP Factor RNA and Protein

Impart Distinct Functions on T-cell Proliferation and Survival', *Cancer Research*, 75(19), pp. 4143–4152. doi: 10.1158/0008-5472.CAN-15-0942.

Miyazaki, M., Yasunaga, J.-I., Taniguchi, Y., Tamiya, S., Nakahata, T. and Matsuoka, M. (2007) 'Preferential Selection of Human T-Cell Leukemia Virus Type 1 Provirus Lacking the 5' Long Terminal Repeat during Oncogenesis', *Journal of Virology*, 81(11), pp. 5714–5723. doi: 10.1128/JVI.02511-06.

Mizuguchi, M., Asao, H., Hara, T., Higuchi, M., Fujii, M. and Nakamura, M. (2009) 'Transcriptional Activation of the Interleukin-21 Gene and Its Receptor Gene by Human T-cell Leukemia Virus Type 1 Tax in Human T-cells', *Journal of Biological Chemistry*, 284(38), pp. 25501–25511. doi: 10.1074/jbc.M109.010959.

Mocquet, V., Neusiedler, J., Rende, F., Cluet, D., Robin, J.-P., Terme, J.-M., Duc Dodon, M., Wittmann, J., Morris, C., Le Hir, H., Ciminale, V. and Jalinot, P. (2012) 'The Human T-Lymphotropic Virus Type 1 Tax Protein Inhibits Nonsense-Mediated mRNA Decay by Interacting with INT6/EIF3E and UPF1', *Journal of Virology*, 86(14), pp. 7530–7543. doi: 10.1128/JVI.07021-11.

Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E., Bertrand, E., Darzacq, X. and Zimmer, C. (2013) 'FISH-quant: automatic counting of transcripts in 3D FISH images', *Nature Methods*. Nature Publishing Group, 10(4), pp. 277–278. doi: 10.1038/nmeth.2406.

Mühleisen, A., Giaisi, M., Köhler, R., Krammer, P. H. and Li-Weber, M. (2014) 'Tax contributes apoptosis resistance to HTLV-1-infected T cells via suppression of Bid and Bim expression', *Cell Death and Disease*, 5(12), p. e1575. doi: 10.1038/cddis.2014.536.

Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D. A., Adams, C. M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B. J., Levitt, M. and Kornberg, R. D. (2013) 'Architecture of an RNA Polymerase II Transcription Pre-Initiation Complex', *Science*, 342(6159), pp. 1238724–1238724. doi: 10.1126/science.1238724.

Murakami, T., Hirai, H., Suzuki, T., Fujisawa, J. and Yoshida, M. (1995) 'HTLV-1 Tax Enhances NF- κ B Expression and Binds to the Products p52 and p100, but Does Not Suppress the Inhibitory Function of p100', *Virology*, 206(2), pp. 1066–1074. doi: 10.1006/viro.1995.1029.

Murphy, E. L. (2016) 'Infection with human T-lymphotropic virus types-1 and -2 (HTLV-1 and -2): Implications for blood transfusion safety', *Transfusion Clinique et Biologique*. Elsevier Masson SAS, 23(1), pp. 13–19. doi: 10.1016/j.tracli.2015.12.001.

Murray, A. J., Kwon, K. J., Farber, D. L. and Siliciano, R. F. (2016) 'The Latent Reservoir for HIV-1: How Immunologic Memory and Clonal Expansion Contribute to HIV-1 Persistence', *The Journal of Immunology*, 197(2), pp. 407–417. doi: 10.4049/jimmunol.1600343.

Nagai, M., Usuku, K., Matsumoto, W., Kodama, D., Takenouchi, N., Moritoyo, T., Hashiguchi, S., Ichinose, M., Bangham, C. R., Izumo, S. and 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.', *Journal of Neurovirology*, 4(6), pp. 586–93.

Nakano, K. and Watanabe, T. (2012) 'HTLV-1 Rex: the courier of viral messages making use of the host vehicle', *Frontiers in Microbiology*, 3(SEP), pp. 1–13. doi: 10.3389/fmicb.2012.00330.

Nejmeddine, M., Barnard, A. L., Tanaka, Y., Taylor, G. P. and Bangham, C. R. M. (2005) 'Human T-lymphotropic Virus, Type 1, Tax Protein Triggers Microtubule Reorientation in the Virological Synapse', *Journal of Biological Chemistry*, 280(33), pp. 29653–29660. doi: 10.1074/jbc.M502639200.

Németh, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Péterfia, B., Solovei, I., Cremer, T., Dopazo, J. and Längst, G. (2010) 'Initial genomics of the human nucleolus.', *PLoS genetics*, 6(3), p. e1000889. doi: 10.1371/journal.pgen.1000889.

- Németh, A. and Längst, G. (2011) 'Genome organization in and around the nucleolus.', *Trends in genetics : TIG*, 27(4), pp. 149–56. doi: 10.1016/j.tig.2011.01.002.
- Nethe, M., Berkhout, B. and van der Kuyl, A. C. (2005) 'Retroviral superinfection resistance', *Retrovirology*, 2(1), p. 52. doi: 10.1186/1742-4690-2-52.
- Neuveut, C., Low, K. G., Maldarelli, F., Schmitt, I., Majone, F., Grassmann, R. and Jeang, K. T. (1998) 'Human T-cell leukemia virus type 1 Tax and cell cycle progression: role of cyclin D-cdk and p110Rb.', *Molecular and cellular biology*, 18(6), pp. 3620–32.
- Nicolas, D., Phillips, N. E. and Naef, F. (2017) 'What shapes eukaryotic transcriptional bursting?', *Mol. BioSyst.* Royal Society of Chemistry. doi: 10.1039/C7MB00154A.
- Nicot, C. (2001) 'HTLV-1 p12I protein enhances STAT5 activation and decreases the interleukin-2 requirement for proliferation of primary human peripheral blood mononuclear cells', *Blood*, 98(3), pp. 823–829. doi: 10.1182/blood.V98.3.823.
- Nicot, C., Dunder, M., Johnson, J. M., Fullen, J. R., Alonzo, N., Fukumoto, R., Princler, G. L., Derse, D., Misteli, T. and Franchini, G. (2004) 'HTLV-1-encoded p30II is a post-transcriptional negative regulator of viral replication', *Nature Medicine*, 10(2), pp. 197–201. doi: 10.1038/nm984.
- Novick, A. and Weiner, M. (1957) 'Enzyme Induction as an All-or-None Phenomenon', *Proceedings of the National Academy of Sciences*, 43(7), pp. 553–66. doi: 10.1073/pnas.43.7.553.
- Ohshima, T., Mukai, R., Nakahara, N., Matsumoto, J., Isono, O., Kobayashi, Y., Takahashi, S. and Shimotohno, K. (2010) 'HTLV-1 basic leucine-zipper factor, HBZ, interacts with MafB and suppresses transcription through a Maf recognition element', *Journal of Cellular Biochemistry*, 111(1), pp. 187–194. doi: 10.1002/jcb.22687.
- Olagnier, D., Sze, A., Bel Hadj, S., Chiang, C., Steel, C., Han, X., Routy, J.-P., Lin, R., Hiscott, J. and van Grevenynghe, J. (2014) 'HTLV-1 Tax-Mediated Inhibition of FOXO3a Activity Is Critical for the Persistence of Terminally Differentiated CD4+ T Cells', *PLoS Pathogens*. Edited by S. R. Ross, 10(12), p. e1004575. doi: 10.1371/journal.ppat.1004575.
- Overbaugh, J. (2001) 'Selection Forces and Constraints on Retroviral Sequence Variation', *Science*, 292(5519), pp. 1106–1109. doi: 10.1126/science.1059128.
- Padmanabhan, K., Eddy, W. F. and Crowley, J. C. (2010) 'A novel algorithm for optimal image thresholding of biological data', *Journal of Neuroscience Methods*. Elsevier B.V., 193(2), pp. 380–384. doi: 10.1016/j.jneumeth.2010.08.031.
- Padovan-Merhar, O., Nair, G. P., Biaesch, A. G., Mayer, A., Scarfone, S., Foley, S. W., Wu, A. R., Churchman, L. S., Singh, A. and Raj, A. (2015) 'Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms', *Molecular Cell*. Elsevier Inc., 58(2), pp. 339–352. doi: 10.1016/j.molcel.2015.03.005.
- Pais-Correia, A.-M., Sachse, M., Guadagnini, S., Robbiati, V., Lasserre, R., Gessain, A., Gout, O., Alcover, A. and Thoulouze, M.-I. (2010) 'Biofilm-like extracellular viral assemblies mediate HTLV-1 cell-to-cell transmission at virological synapses', *Nature Medicine*. Nature Publishing Group, 16(1), pp. 83–89. doi: 10.1038/nm.2065.
- Park, H. U., Jeong, J.-H., Chung, J. H. and Brady, J. N. (2004) 'Human T-cell leukemia virus type 1 Tax interacts with Chk1 and attenuates DNA-damage induced G2 arrest mediated by Chk1.', *Oncogene*, 23(29), pp. 4966–4974. doi: 10.1038/sj.onc.1207644.
- Park, H. U., Jeong, S.-J., Jeong, J.-H., Chung, J. H. and Brady, J. N. (2006) 'Human T-cell leukemia virus type 1 Tax attenuates gamma-irradiation-induced apoptosis through physical interaction with Chk2.',

Oncogene, 25(3), pp. 438–447. doi: 10.1038/sj.onc.1209059.

Percher, F., Jeannin, P., Martin-Latil, S., Gessain, A., Afonso, P., Vidy-Roche, A. and Ceccaldi, P.-E. (2016) 'Mother-to-Child Transmission of HTLV-1 Epidemiological Aspects, Mechanisms and Determinants of Mother-to-Child Transmission', *Viruses*, 8(2), p. 40. doi: 10.3390/v8020040.

Philip, S., Zahoor, M. A., Zhi, H., Ho, Y.-K. and Giam, C.-Z. (2014) 'Regulation of Human T-Lymphotropic Virus Type I Latency and Reactivation by HBZ and Rex', *PLoS Pathogens*. Edited by S. R. Ross, 10(4), p. e1004040. doi: 10.1371/journal.ppat.1004040.

Phillips, J. E. and Corces, V. G. (2009) 'CTCF: Master Weaver of the Genome', *Cell*, 137(7), pp. 1194–1211. doi: 10.1016/j.cell.2009.06.001.

Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D. and 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.', *Proceedings of the National Academy of Sciences of the United States of America*, 77(12), pp. 7415–9. doi: 10.1073/pnas.77.12.7415.

Pozzatti, R., Vogel, J. and Jay, G. (1990) 'The human T-lymphotropic virus type I tax gene can cooperate with the ras oncogene to induce neoplastic transformation of cells.', *Molecular and Cellular Biology*, 10(1), pp. 413–417. doi: 10.1128/MCB.10.1.413.

Van Prooyen, N., Gold, H., Andresen, V., Schwartz, O., Jones, K., Ruscetti, F., Lockett, S., Gudla, P., Venzon, D. and Franchini, G. (2010) 'Human T-cell leukemia virus type 1 p8 protein increases cellular conduits and virus transmission', *Proceedings of the National Academy of Sciences*, 107(48), pp. 20738–20743. doi: 10.1073/pnas.1009635107.

Proudfoot, N. J. (2016) 'Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut', *Science*, 352(6291), p. aad9926. doi: 10.1126/science.aad9926.

Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. and Tyagi, S. (2008) 'Imaging individual mRNA molecules using multiple singly labeled probes', *Nature Methods*, 5(10), pp. 877–879. doi: 10.1038/nmeth.1253.

Raj, A. and van Oudenaarden, A. (2008) 'Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences', *Cell*, 135(2), pp. 216–226. doi: 10.1016/j.cell.2008.09.050.

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. and Tyagi, S. (2006) 'Stochastic mRNA synthesis in mammalian cells', *PLoS Biology*, 4(10), pp. 1707–1719. doi: 10.1371/journal.pbio.0040309.

Raval, G. U., Bidoia, C., Forlani, G., Tosi, G., Gessain, A. and Accolla, R. S. (2015) 'Localization, quantification and interaction with host factors of endogenous HTLV-1 HBZ protein in infected cells and ATL', *Retrovirology*. BioMed Central, 12(1), p. 59. doi: 10.1186/s12977-015-0186-0.

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. M. and Ciminale, V. (2011) 'Kinetics and intracellular compartmentalization of HTLV-1 gene expression: nuclear retention of HBZ mRNAs.', *Blood*, 117(18), pp. 4855–9. doi: 10.1182/blood-2010-11-316463.

Richardson, J. H., Höllsberg, P., Windhagen, A., Child, L. A., Hafler, D. A. and Lever, A. M. (1997) 'Variable immortalizing potential and frequent virus latency in blood-derived T-cell clones infected with human T-cell leukemia virus type I.', *Blood*, 89(9), pp. 3303–14.

Roukos, V., Pegoraro, G., Voss, T. C. and Misteli, T. (2015) 'Cell cycle staging of individual cells by fluorescence microscopy', *Nature Protocols*, 10(2), pp. 334–348. doi: 10.1038/nprot.2015.016.

Rowan, A. G., Suemori, K., Fujiwara, H., Yasukawa, M., Tanaka, Y., Taylor, G. P. and Bangham, C. (2014) 'Cytotoxic T lymphocyte lysis of HTLV-1 infected cells is limited by weak HBZ protein

expression, but non-specifically enhanced on induction of Tax expression.’, *Retrovirology*, 11(1), p. 116. doi: 10.1186/s12977-014-0116-6.

Saito, M., Matsuzaki, T., Satou, Y., Yasunaga, J.-I., Saito, K., Arimura, K., Matsuoka, M. and 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, p. 19. doi: 10.1186/1742-4690-6-19.

Santiago, F., Clark, E., Chong, S., Molina, C., Mozafari, F., Mahieux, R., Fujii, M., Azimi, N. and Kashanchi, F. (1999) ‘Transcriptional up-regulation of the cyclin D2 gene and acquisition of new cyclin-dependent kinase partners in human T-cell leukemia virus type 1-infected cells.’, *Journal of Virology*, 73(12), pp. 9917–27.

Sarkar, D. (2008) *Lattice: multivariate data visualization with R, Use R!* doi: 10.1007/978-0-387-75969-2.

Satou, Y., Miyazato, P., Ishihara, K., Yaguchi, H., Melamed, A., Miura, M., Fukuda, A., Nosaka, K., Watanabe, T., Rowan, A. G., Nakao, M. and Bangham, C. R. M. (2016) ‘The retrovirus HTLV-1 inserts an ectopic CTCF-binding site into the human genome’, *Proceedings of the National Academy of Sciences*, 113(1), p. 201423199. doi: 10.1073/pnas.1423199113.

Satou, Y., Yasunaga, J., Yoshida, M. and Matsuoka, M. (2006) ‘HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells.’, *Proceedings of the National Academy of Sciences of the United States of America*, 103(3), pp. 720–5. doi: 10.1073/pnas.0507631103.

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. and Matsuoka, M. (2011) ‘HTLV-1 bZIP Factor Induces T-Cell Lymphoma and Systemic Inflammation In Vivo’, *PLoS Pathogens*. Edited by J. Luban, 7(2), p. e1001274. doi: 10.1371/journal.ppat.1001274.

Satyanarayana, A. and Kaldis, P. (2009) ‘Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms’, *Oncogene*. Nature Publishing Group, 28(33), pp. 2925–2939. doi: 10.1038/onc.2009.170.

Schmitt, I., Rosin, O., Rohwer, P., Gossen, M. and Grassmann, R. (1998) ‘Stimulation of cyclin-dependent kinase activity and G1- to S-phase transition in human lymphocytes by the human T-cell leukemia/lymphotropic virus type 1 Tax protein.’, *Journal of Virology*, 72(1), pp. 633–40.

Seiki, M., Hattori, S., Hirayama, Y. and Yoshida, M. (1983) ‘Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA.’, *Proceedings of the National Academy of Sciences of the United States of America*, 80(12), pp. 3618–22.

Seiki, M., Inoue, J., Hidaka, M. and Yoshida, M. (1988) ‘Two cis-acting elements responsible for posttranscriptional trans-regulation of gene expression of human T-cell leukemia virus type I.’, *Proceedings of the National Academy of Sciences of the United States of America*, 85(19), pp. 7124–8.

Senecal, A., Munsky, B., Proux, F., Ly, N., Braye, F. E., Zimmer, C., Mueller, F. and Darzacq, X. (2014) ‘Transcription Factors Modulate c-Fos Transcriptional Bursts’, *Cell Reports*. The Authors, 8(1), pp. 75–83. doi: 10.1016/j.celrep.2014.05.053.

Sharma, N. and Nyborg, J. K. (2008) ‘The coactivators CBP/p300 and the histone chaperone NAP1 promote transcription-independent nucleosome eviction at the HTLV-1 promoter’, *Proceedings of the National Academy of Sciences*, 105(23), pp. 7959–7963. doi: 10.1073/pnas.0800534105.

Shimoyama, M. (1991) ‘Diagnostic criteria and classification of clinical subtypes of adult T-cell

leukaemia-lymphoma.', *British Journal of Haematology*, 79(3), pp. 428–437. doi: 10.1111/j.1365-2141.1991.tb08051.x.

Shiohama, Y., Naito, T., Matsuzaki, T., Tanaka, R., Tomoyose, T., Takashima, H., Fukushima, T., Tanaka, Y. and Saito, M. (2016) 'Absolute quantification of HTLV-1 basic leucine zipper factor (HBZ) protein and its plasma antibody in HTLV-1 infected individuals with different clinical status', *Retrovirology*, 13(1). doi: 10.1186/s12977-016-0263-z.

Silbermann, K., Schneider, G. and Grassmann, R. (2008) 'Stimulation of interleukin-13 expression by human T-cell leukemia virus type 1 oncoprotein Tax via a dually active promoter element responsive to NF- κ B and NFAT', *Journal of General Virology*, 89(11), pp. 2788–2798. doi: 10.1099/vir.0.2008/003699-0.

Silic-Benussi, M., Cannizzaro, E., Venerando, A., Cavallari, I., Petronilli, V., La Rocca, N., Marin, O., Chieco-Bianchi, L., Di Lisa, F., D'Agostino, D. M., Bernardi, P. and Ciminale, V. (2009) 'Modulation of mitochondrial K(+) permeability and reactive oxygen species production by the p13 protein of human T-cell leukemia virus type 1.', *Biochimica et Biophysica Acta*. Elsevier B.V., 1787(7), pp. 947–54. doi: 10.1016/j.bbabi.2009.02.001.

Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T. (2005) 'ROCR: visualizing classifier performance in R', *Bioinformatics*, 21(20), pp. 3940–3941. doi: 10.1093/bioinformatics/bti623.

Skinner, S. O., Xu, H., Nagarkar-Jaiswal, S., Freire, P. R., Zwaka, T. P. and Golding, I. (2016) 'Single-cell analysis of transcription kinetics across the cell cycle', *eLife*, 5(JANUARY2016), pp. 1–24. doi: 10.7554/eLife.12175.

Skupsky, R., Burnett, J. C., Foley, J. E., Schaffer, D. V. and Arkin, A. P. (2010) 'HIV Promoter Integration Site Primarily Modulates Transcriptional Burst Size Rather Than Frequency', *PLoS Computational Biology*. Edited by N. Friedman, 6(9), p. e1000952. doi: 10.1371/journal.pcbi.1000952.

Smith, M. R. and Greene, W. C. (1992) 'Characterization of a novel nuclear localization signal in the HTLV-I tax transactivator protein', *Virology*, 187(1), pp. 316–320. doi: 10.1016/0042-6822(92)90320-O.

Soltani, M. and Singh, A. (2016) 'Effects of cell-cycle-dependent expression on random fluctuations in protein levels', *Royal Society Open Science*, 3(12), p. 160578. doi: 10.1098/rsos.160578.

Steinfort, D. P., Brady, S., Weisinger, H. S. and Einsiedel, L. (2008) 'Bronchiectasis in Central Australia: A young face to an old disease', *Respiratory Medicine*, 102(4), pp. 574–578. doi: 10.1016/j.rmed.2007.11.007.

Stoeger, T., Battich, N. and Pelkmans, L. (2016) 'Passive Noise Filtering by Cellular Compartmentalization', *Cell*. Elsevier Inc., 164(6), pp. 1151–1161. doi: 10.1016/j.cell.2016.02.005.

Sturrock, M., Li, S. and Shahrezaei, V. (2017) 'The influence of nuclear compartmentalisation on stochastic dynamics of self-repressing gene expression', *Journal of Theoretical Biology*. Elsevier Ltd, 424, pp. 55–72. doi: 10.1016/j.jtbi.2017.05.003.

Sugata, K., Yasunaga, J. I., Mitobe, Y., Miura, M., Miyazato, P., Kohara, M. and Matsuoka, M. (2015) 'Protective effect of cytotoxic T lymphocytes targeting HTLV-1 bZIP factor', *Blood*, 126(9), pp. 1095–1105. doi: 10.1182/blood-2015-04-641118.

Takeda, S., Maeda, M., Morikawa, S., Taniguchi, Y., Yasunaga, J. I., Nosaka, K., Tanaka, Y. and Matsuoka, M. (2004) 'Genetic and epigenetic inactivation of TAX gene in adult t-cell leukemia cells', *International Journal of Cancer*, 109(4), pp. 559–567. doi: 10.1002/ijc.20007.

- Tanaka-Nakanishi, A., Yasunaga, J. -i., Takai, K. and Matsuoka, M. (2014) 'HTLV-1 bZIP Factor Suppresses Apoptosis by Attenuating the Function of FoxO3a and Altering Its Localization', *Cancer Research*, 74(1), pp. 188–200. doi: 10.1158/0008-5472.CAN-13-0436.
- Tanaka, Y., Fukudome, K., Hayashi, M., Takagi, S. and Yoshie, O. (1995) 'Induction of ICAM-1 and LFA-3 by Tax1 of human T-cell leukemia virus type 1 and mechanism of down-regulation of ICAM-1 or LFA-1 in adult-T-cell-leukemia cell lines', *International Journal of Cancer*, 60(4), pp. 554–561. doi: 10.1002/ijc.2910600421.
- Tantale, K., Mueller, F., Kozulic-Pirher, A., Lesne, A., Victor, J.-M., Robert, M.-C., Capozzi, S., Chouaib, R., Bäcker, V., Mateos-Langerak, J., Darzacq, X., Zimmer, C., Basyuk, E. and Bertrand, E. (2016) 'A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting.', *Nature communications*, 7, p. 12248. doi: 10.1038/ncomms12248.
- Team, R. C. (2017) 'R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing'.
- de Thé, G. and Bomford, R. (1993) 'An HTLV-I vaccine: why, how, for whom?', *AIDS research and human retroviruses*, 9(5), pp. 381–6. doi: 10.1089/aid.1993.9.381.
- Thébault, S., Basbous, J., Hivin, P., Devaux, C. and Mesnard, J.-M. (2004) 'HBZ interacts with JunD and stimulates its transcriptional activity', *FEBS Letters*, 562(1–3), pp. 165–170. doi: 10.1016/S0014-5793(04)00225-X.
- Theunissen, J. F. and Petrini, J. H. J. (2006) 'Methods for Studying the Cellular Response to DNA Damage: Influence of the Mre11 Complex on Chromosome Metabolism', in *Methods in Enzymology*, pp. 251–284. doi: 10.1016/S0076-6879(05)09015-4.
- Tie, F., Adya, N., Greene, W. C. and Giam, C. Z. (1996) 'Interaction of the human T-lymphotropic virus type 1 Tax dimer with CREB and the viral 21-base-pair repeat.', *Journal of Virology*, 70(12), pp. 8368–74.
- Toulza, F., Heaps, A., Tanaka, Y., Taylor, G. P. and Bangham, C. R. M. (2008) 'High frequency of CD4+FoxP3+ cells in HTLV-1 infection: inverse correlation with HTLV-1-specific CTL response', *Blood*, 111(10), pp. 5047–5053. doi: 10.1182/blood-2007-10-118539.
- Tsanov, N., Samacoits, A., Chouaib, R., Traboulsi, A.-M., Gostan, T., Weber, C., Zimmer, C., Zibara, K., Walter, T., Peter, M., Bertrand, E. and Mueller, F. (2016) 'smiFISH and FISH-quant – a flexible single RNA detection approach with super-resolution capability', *Nucleic Acids Research*, p. gkw784. doi: 10.1093/NAR/GKW784.
- 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. and 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(19), pp. 3809–3817. doi: 10.1182/blood-2010-05-284141.
- Vandamme, A. M., Liu, H. F., Goubau, P. and Desmyter, J. (1994) 'Primate T-lymphotropic virus type I LTR sequence variation and its phylogenetic analysis: compatibility with an African origin of PTLV-I.', *Virology*, 202(1), pp. 212–23. doi: 10.1006/viro.1994.1337.
- Vargas, D. Y., Raj, A., Marras, S. a E., Kramer, F. R. and Tyagi, S. (2005) 'Mechanism of mRNA transport in the nucleus', *Proceedings of the National Academy of Sciences*, 102(47), pp. 17008–17013. doi: 10.1073/pnas.0505580102.
- Venables, W. N. and Ripley, B. D. (2003) 'Modern Applied Statistics With S', *Technometrics*, 45(1), pp. 111–111. doi: 10.1198/tech.2003.s33.

- Vera, M., Biswas, J., Senecal, A., Singer, R. H. and Park, H. Y. (2016) 'Single-Cell and Single-Molecule Analysis of Gene Expression Regulation', *Annual Review of Genetics*, 50(1), pp. 267–291. doi: 10.1146/annurev-genet-120215-034854.
- Verdaasdonk, J. S., Stephens, A. D., Haase, J. and Bloom, K. (2014) 'Bending the Rules: Widefield Microscopy and the Abbe Limit of Resolution', *Journal of Cellular Physiology*, 229(July), pp. 132–138. doi: 10.1002/jcp.24439.
- Verdonck, K., González, E., Van Dooren, S., Vandamme, A.-M., Vanham, G. and Gotuzzo, E. (2007) 'Human T-lymphotropic virus 1: recent knowledge about an ancient infection', *The Lancet Infectious Diseases*, 7(4), pp. 266–281. doi: 10.1016/S1473-3099(07)70081-6.
- Vermeulen, K., Van Bockstaele, D. R. and Berneman, Z. N. (2003) 'The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer', *Cell Proliferation*, 36(3), pp. 131–149. doi: 10.1046/j.1365-2184.2003.00266.x.
- Vincent, L. and Soille, P. (1991) 'Watersheds in digital spaces: an efficient algorithm based on immersion simulations', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6), pp. 583–598. doi: 10.1109/34.87344.
- Wagner, S. and Green, M. R. (1993) 'HTLV-I Tax protein stimulation of DNA binding of bZIP proteins by enhancing dimerization.', *Science*, 262(5132), pp. 395–9.
- Wang, S., Su, J.-H., Beliveau, B. J., Bintu, B., Moffitt, J. R., Wu, C. and Zhuang, X. (2016) 'Spatial organization of chromatin domains and compartments in single chromosomes', *Science*, 353(6299), pp. 598–602. doi: 10.1126/science.aaf8084.
- Wei, P., Garber, M. E., Fang, S.-M., Fischer, W. H. and Jones, K. A. (1998) 'A Novel CDK9-Associated C-Type Cyclin Interacts Directly with HIV-1 Tat and Mediates Its High-Affinity, Loop-Specific Binding to TAR RNA', *Cell*, 92(4), pp. 451–462. doi: 10.1016/S0092-8674(00)80939-3.
- Weinberger, A. D. and Weinberger, L. S. (2013) 'Stochastic Fate Selection in HIV-Infected Patients', *Cell*. Elsevier, 155(3), pp. 497–499. doi: 10.1016/j.cell.2013.09.039.
- Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P. and Schaffer, D. V. (2005) 'Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity', *Cell*, 122(2), pp. 169–182. doi: 10.1016/j.cell.2005.06.006.
- Weinberger, L. S., Dar, R. D. and Simpson, M. L. (2008) 'Transient-mediated fate determination in a transcriptional circuit of HIV', *Nature Genetics*, 40(4), pp. 466–470. doi: 10.1038/ng.116.
- Wendt, K. S. and Grosveld, F. G. (2014) 'Transcription in the context of the 3D nucleus', *Current Opinion in Genetics & Development*. Elsevier Ltd, 25(1), pp. 62–67. doi: 10.1016/j.gde.2013.11.020.
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis, Media*. New York, NY: Springer New York. doi: 10.1007/978-0-387-98141-3.
- Willems, L., Hasegawa, H., Accolla, R., Bangham, C., Bazarbachi, A., Bertazzoni, U., Carneiro-Proietti, A. B. de F., Cheng, H., Chieco-Bianchi, L., Ciminale, V., Coelho-dos-Reis, J., Esparza, J., Gallo, R. C., Gessain, A., Gotuzzo, E., Hall, W., Harford, J., Hermine, O., Jacobson, S., Macchi, B., Macpherson, C., Mahieux, R., Matsuoka, M., Murphy, E., Peloponese, J.-M., Simon, V., Tagaya, Y., Taylor, G. P., Watanabe, T. and Yamano, Y. (2017) 'Reducing the global burden of HTLV-1 infection: An agenda for research and action', *Antiviral Research*, 137, pp. 41–48. doi: 10.1016/j.antiviral.2016.10.015.
- 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. and Switzer, W. M. (2005) 'Emergence of unique primate T-lymphotropic viruses among central

African bushmeat hunters.’, *Proceedings of the National Academy of Sciences of the United States of America*, 102(22), pp. 7994–9. doi: 10.1073/pnas.0501734102.

Wu, X., Zhang, M. and Sun, S.-C. (2011) ‘Mutual regulation between deubiquitinase CYLD and retroviral oncoprotein Tax.’, *Cell & bioscience*, 1(1), p. 27. doi: 10.1186/2045-3701-1-27.

Xie, S. Q., Lavitas, L. and Pombo, A. (2010) ‘CryoFISH: fluorescence in situ hybridization on ultrathin cryosections.’, *Methods in molecular biology (Clifton, N.J.)*. Edited by J. M. Bridger and E. V. Volpi. Totowa, NJ: Humana Press (Methods in Molecular Biology), 659, pp. 219–30. doi: 10.1007/978-1-60761-789-1_15.

Yankulov, K., Yamashita, K., Roy, R., Egly, J. and Bentley, D. L. (1995) ‘The Transcriptional Elongation Inhibitor 5,6-Dichloro-1- β -D-ribofuranosylbenzimidazole Inhibits Transcription Factor IIH-associated Protein Kinase’, *Journal of Biological Chemistry*, 270(41), pp. 23922–23925. doi: 10.1074/jbc.270.41.23922.

Yasuma, K., Yasunaga, J., Takemoto, K., Sugata, K., Mitobe, Y., Takenouchi, N., Nakagawa, M., Suzuki, Y. and Matsuoka, M. (2016) ‘HTLV-1 bZIP Factor Impairs Anti-viral Immunity by Inducing Co-inhibitory Molecule, T Cell Immunoglobulin and ITIM Domain (TIGIT)’, *PLOS Pathogens*. Edited by S. R. Ross, 12(1), p. e1005372. doi: 10.1371/journal.ppat.1005372.

Yasunaga, J., Furuta, R., Miura, M., Sugata, K., Saito, A., Hirofumi, A., Ueno, T., Takenouchi, N., Fujisawa, J., Shimizu, M., Matsuda, F., Melamed, A., Bangham, C. R. M. . and Matsuoka, M. (2016) ‘Hematopoietic Stem Cell Infected with HTLV-1 Functions As a Viral Reservoir In Vivo’, *Blood*, 128(22), p. 1343.

Yoshida, M. (2001) ‘Multiple Viral Strategies of HTLV-1 for Dysregulation of Cell Growth Control’, *Annual Reviews Immunology*, 19, pp. 475–96.

Yoshida, M., Miyoshi, I. and Hinuma, Y. (1982) ‘Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease.’, *Proceedings of the National Academy of Sciences*, 79(6), pp. 2031–2035. doi: 10.1073/pnas.79.6.2031.

Yoshida, M., Satou, Y., Yasunaga, J.-I., Fujisawa, J.-I. and Matsuoka, M. (2008) ‘Transcriptional control of spliced and unspliced human T-cell leukemia virus type 1 bZIP factor (HBZ) gene.’, *Journal of virology*, 82(19), pp. 9359–68. doi: 10.1128/JVI.00242-08.

Yoshita, M., Higuchi, M., Takahashi, M., Oie, M., Tanaka, Y. and Fujii, M. (2012) ‘Activation of mTOR by human T-cell leukemia virus type 1 Tax is important for the transformation of mouse T cells to interleukin-2-independent growth.’, *Cancer science*, 103(2), pp. 369–74. doi: 10.1111/j.1349-7006.2011.02123.x.

Yu, Q., Minoda, Y., Yoshida, R., Yoshida, H., Iha, H., Kobayashi, T., Yoshimura, A. and Takaesu, G. (2008) ‘HTLV-1 Tax-mediated TAK1 activation involves TAB2 adapter protein’, *Biochemical and Biophysical Research Communications*, 365(1), pp. 189–194. doi: 10.1016/j.bbrc.2007.10.172.

Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M. L., Dayaram, T., Owens, B. M., Shigematsu, H., Levantini, E., Huettner, C. S., Lekstrom-Himes, J. A., Akashi, K. and Tenen, D. G. (2004) ‘Enhancement of Hematopoietic Stem Cell Repopulating Capacity and Self-Renewal in the Absence of the Transcription Factor C/EBP α ’, *Immunity*, 21(6), pp. 853–863. doi: 10.1016/j.immuni.2004.11.006.

Zhang, W., Nisbet, J. W., Bartoe, J. T., Ding, W. and Lairmore, M. D. (2000) ‘Human T-Lymphotropic Virus Type 1 p30II Functions as a Transcription Factor and Differentially Modulates CREB-Responsive Promoters’, *Journal of Virology*, 74(23), pp. 11270–11277. doi: 10.1128/JVI.74.23.11270-11277.2000.

Zhao, T., Satou, Y., Sugata, K., Miyazato, P., Green, P. L., Imamura, T. and Matsuoka, M. (2011) ‘HTLV-

1 bZIP factor enhances TGF- signaling through p300 coactivator', *Blood*. BioMed Central Ltd, 118(7), pp. 1865–1876. doi: 10.1182/blood-2010-12-326199.

Zhao, T., Yasunaga, J. -i., Satou, Y., Nakao, M., Takahashi, M., Fujii, M. and Matsuoka, M. (2009) 'Human T-cell leukemia virus type 1 bZIP factor selectively suppresses the classical pathway of NF- κ B', *Blood*, 113(12), pp. 2755–2764. doi: 10.1182/blood-2008-06-161729.

Zhou, M., Lu, H., Park, H., Wilson-Chiru, J., Linton, R. and Brady, J. N. (2006) 'Tax interacts with P-TEFb in a novel manner to stimulate human T-lymphotropic virus type 1 transcription.', *Journal of Virology*, 80(10), pp. 4781–91. doi: 10.1128/JVI.80.10.4781-4791.2006.

APPENDICES

8.1 Summary of permission for third party copyright works

Page number	Use	Source work	Copyright holder	Permission
22	Fig. 1-1	Figure 2 2012 <i>Front. Microbio.</i> 3:388. doi: 10.3389/fmicb.2012.00388	© 2012 Gessain and Cassar	Under Creative Commons Attribution License
40	Fig. 1-3	Figure 2 2013 Current Opinion in Virology Volume 3, Issue 6, Pages 684-691 doi: 10.1016/j.coviro.2013.08.010	© 2013 Elsevier B.V	Permission granted (see license below)
47	Fig. 1-4	Figure 1 2003 Cell Proliferation Volume 36, Issue 3, Pages 131–149 doi: 10.1046/j.1365-2184.2003.00266.x	© 2003, John Wiley and Sons	Permission granted (see license below)

ELSEVIER LICENSE TERMS AND CONDITIONS

This Agreement between Mr. Martin Billman ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4164820765168
License date	Aug 09, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	Current Opinion in Virology
Licensed Content Title	Human T-cell leukemia virus type 1: replication, proliferation and propagation by Tax and HTLV-1 bZIP factor
Licensed Content Author	Masao Matsuoka, Jun-ichirou Yasunaga
Licensed Content Date	Dec 1, 2013
Licensed Content Volume	3
Licensed Content Issue	6
Licensed Content Pages	8
Start Page	684
End Page	691
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 2
Title of your thesis/dissertation	QUANTIFICATION OF HTLV-1 EXPRESSION AT THE SINGLE-CELL LEVEL IN NATURALLY-INFECTED T-LYMPHOCYTES
Expected completion date	Sep 2017
Estimated size (number of pages)	200

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

This Agreement between Mr. Martin Billman ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4173791130048
License date	Aug 21, 2017
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Cell Proliferation
Licensed Content Title	The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer
Licensed Content Author	Katrien Vermeulen, Dirk R. Van Bockstaele, Zwi N. Berneman
Licensed Content Date	Jun 17, 2003
Licensed Content Pages	19
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 1
Will you be translating?	No
Title of your thesis / dissertation	QUANTIFICATION OF HTLV-1 EXPRESSION AT THE SINGLE-CELL LEVEL IN NATURALLY-INFECTED T-LYMPHOCYTES
Expected completion date	Sep 2017
Expected size (number of pages)	200

8.2 Abstract of publication associated with this thesis

Title:

Single-cell heterogeneity and cell-cycle-related viral gene bursts in the human leukaemia virus HTLV-1

Abstract:

Background:

The human leukaemia virus HTLV-1 expresses essential accessory genes that manipulate the expression, splicing and transport of viral mRNAs. Two of these genes, tax and hbz, also promote proliferation of the infected cell, and both genes are thought to contribute to oncogenesis in adult T-cell leukaemia/lymphoma. The regulation of HTLV-1 proviral latency is not understood. tax, on the proviral plus strand, is usually silent in freshly-isolated cells, whereas the minus-strand-encoded hbz gene is persistently expressed at a low level. However, the persistently activated host immune response to Tax indicates frequent expression of tax in vivo.

Methods:

We used single-molecule RNA-FISH to quantify the expression of HTLV-1 transcripts at the single-cell level in a total of >19,000 cells from five T-cell clones, naturally infected with HTLV-1, isolated by limiting dilution from peripheral blood of HTLV-1-infected subjects.

Results:

We found strong heterogeneity both within and between clones in the expression of the proviral plus-strand (detected by hybridization to the tax gene) and the minus-strand (hbz gene). Both genes are transcribed in bursts; tax expression is enhanced in the absence of hbz, while hbz expression increased in cells with high tax expression. Surprisingly, we found that hbz expression is strongly associated with the S and G2/M phases of the cell cycle, independent of tax expression. Contrary to current belief, hbz is not expressed in all cells at all times, even within one clone. In hbz-positive cells, the abundance of hbz transcripts showed a very strong positive linear correlation with nuclear volume.

Conclusions:

The occurrence of intense, intermittent plus-strand gene bursts in independent primary HTLV-1-infected T-cell clones from unrelated individuals strongly suggests that the HTLV-1 plus-strand is expressed in bursts in vivo. Our results offer an explanation for the paradoxical correlations observed between the host immune response and HTLV-1 transcription.

8.3 SmRNA-FISH probe libraries

Table 8-1 | *hbz*-3'UTR probe oligonucleotide library

<i>Probe #</i>	<i>Sequence 5'→3'</i>	<i>Pos</i>	<i>HTLV-1 Pos</i>	<i>GC%</i>
1	TGACACAGGCAAGCATCGA	8	7263	53.0%
2	AATAGCCCGTCCACCAATT	52	7219	47.0%
3	TTAACTCTTCCTCCAAGGA	72	7199	42.0%
4	TAGCAAACCGTCAAGCACA	113	7158	47.0%
5	CGGCTTTCCTCTTAAGG	133	7138	53.0%
6	TGCCGATCACGATGCGTTT	202	7069	53.0%
7	TTTTTTTCGCTTCCTTTC	236	7035	37.0%
8	CCAACTGTCTAGTATAGCC	474	6797	47.0%
9	CCCTGCCAATAATTAACCT	562	6709	42.0%
10	TTATTGCAACCACATCGCC	590	6681	47.0%
11	CCCATTACTCTCTTATAA	642	6629	37.0%
12	CTTGTTATCCTTGCAAGGAC	708	6563	47.0%
13	TGGAATCACCCCTGTGCGG	736	6535	58.0%
14	GCTCGAGAGGCCTTACAAA	756	6515	53.0%
15	AGAATCGAGTCCTGACTGG	812	6459	53.0%
16	ACTACAAGAAAGACCCCCC	835	6436	53.0%
17	TACTAATCCCATGTCTCA	856	6415	37.0%
18	TTCCAGTTAACTCAAGCA	1009	6262	42.0%
19	TCAGGAAAGAGCCTCCTAC	1047	6224	53.0%
20	GAGTGGCTGGCGGGATTAC	1085	6186	63.0%
21	TTTCTTGTACACTGTTC	1179	6092	42.0%
22	CTGTCATAACTCCCTCATC	1207	6064	47.0%
23	CTGGACCCACTGCTTTGAC	1255	6016	58.0%
24	CACCTGACGTACCATTTA	1275	5996	42.0%
25	TTCCATCCTCTTCTTAC	1331	5940	42.0%
26	TATACTCTCCCAACGTCTC	1352	5919	47.0%
27	TTGTCTGTATCGATCGTGC	1394	5877	47.0%
28	TGTCCAGTTAACCCTACAA	1432	5839	42.0%
29	GCCCTCTATACCATGGAAA	1468	5803	47.0%
30	TAACTAGACCACATCCTC	1489	5782	47.0%
31	TCCTTAATACCGAACCCAG	1547	5724	47.0%
32	CGACGCTCCAGGATATGAC	1576	5695	58.0%
33	CTCAAGAAGTTTCACGCCT	1646	5625	47.0%
34	CCTACTGGAAGTTTCAGCA	1679	5592	47.0%
35	TATACAGGAGCCGTCTCCA	1701	5570	53.0%
36	TGGGGTGCCAATCATGGAC	1727	5544	58.0%
37	GACCCCTGTTCTTAAAGT	1755	5516	42.0%
38	GCTATTATTCAGCCTCTTA	1778	5493	37.0%
39	AGCCAAACCGAAATGGCGG	1799	5472	58.0%
40	CTATTCCCTCATTGGATTA	1821	5450	37.0%
41	ATGCCACCTATTCCCTATA	1841	5430	42.0%
42	CCTAGTAAGTTACTCCAGC	1864	5407	47.0%
43	AGCAGATCAGGCCCTACAG	1897	5374	58.0%
44	AGCCAGTTTGTTCGTGGAC	1937	5334	53.0%
45	CTAAACCCTGCAATCCTGC	1958	5313	53.0%
46	TGGAGTCTCCTCATACCAC	1978	5293	53.0%
47	AGCTGCTGTACTCTCAAA	1998	5273	47.0%
48	ATCTTCGGTGATTACAGCC	2019	5252	47.0%

Table 8-2 | *tax* probe oligonucleotide library

<i>Probe #</i>	<i>Sequence 5'→3'</i>	<i>Pos</i>	<i>HTLV-1 Pos</i>	<i>GC%</i>
1	AGACTCTGTCCAAACCT	15	7315	50.0%
2	CGTAGACTGGGTATCCGA	38	7338	55.6%
3	CAGTCGCCTTGACACAG	66	7366	55.6%
4	AACATAGTCCCCAGAGA	92	7392	50.0%
5	CGTGACGATGTAGGCGGG	113	7413	66.7%
6	TGGACAGGTGGCCAGTAG	133	7433	61.1%
7	TCCCAGGTGATCTGATGC	153	7453	55.6%
8	AACGCGTCCATCGATGGG	172	7472	61.1%
9	ACTGTAGAGCTGAGCCGA	191	7491	55.6%
10	GGGTCTTAGAGGTTCTCT	242	7542	50.0%
11	ATTGGCGGGGTAAGGACC	264	7564	61.1%
12	GTTGGGGGTTGTATGAGT	283	7583	50.0%
13	GGAGTATTTCGCATGGC	322	7622	55.6%
14	TGGGTTCCATGTATCCAT	350	7650	44.4%
15	GACAGGGTTGGGAGGTGC	378	7678	66.7%
16	GAGTCCGGGTCTGAAAA	397	7697	61.1%
17	GTGTACAGGTTTGGGGC	417	7717	55.6%
18	ATGCAGACAACGGAGCCT	444	7744	55.6%
19	GATGGGGGGGAAAGCTG	472	7772	66.7%
20	TCACATGGGGCAGGAGGG	497	7797	66.7%
21	TGGCCGGGTGGCAAAAA	516	7816	61.1%
22	ACATTGGTGAGGAAGGCC	540	7840	55.6%
23	CCCCTGTGGTAAGGGAAA	593	7893	55.6%
24	ACAGTCTCGGTTAGAAT	619	7919	50.0%
25	GGAAAAGGTGGTGGGCA	638	7938	61.1%
26	CTGTCAGCGTGACGGGTG	668	7968	66.7%
27	GAAGGAGGCCGTTTGGCC	689	7989	61.1%
28	GTGAGGGTTGAGTGGAAC	708	8008	55.6%
29	CCAAATAAGGCCTGGAGT	727	8027	50.0%
30	GCGTGCCATCGGTAAATG	746	8046	55.6%
31	CAGGGCCCGGAAATCATA	765	8065	55.6%
32	ATGGCTGGCCATCTTTAG	785	8085	50.0%
33	GGAGGAGGACTGTAGTAC	808	8108	55.6%
34	GGCCTTGGTTTGAAATTT	838	8138	38.9%
35	GTAGAAATGAGGGGTGGT	857	8157	50.0%
36	ACTGTATGAGGCCGTGTG	878	8178	55.6%
37	GGGGATGTTGGTGTATTC	931	8231	50.0%
38	GTCATCTGCCTCTTTTTC	967	8267	44.4%
39	TTTGGGGCTCATGGTCAT	986	8286	50.0%
40	ACTGAGAGGCTCTAAGCC	1015	8315	55.6%
41	CAGACTTCTGTTTCACGG	1044	8344	50.0%

Table 8-3 | *gagpol* probe oligonucleotide library

Probe #	Sequence 5'→3'	Pos	HTLV-1 Pos	GC%
1	GGGAAAAGATTGGCCCAT	2	804	47.0%
2	CGGAATAGGGCTAGCGCTA	22	824	58.0%
3	TTAAGCCAGTGATGAGCGG	60	862	53.0%
4	ATCGTAACTGGAGGGACCG	109	911	58.0%
5	CAGATCCAGACTGGTGTTT	165	967	47.0%
6	GGCTAGGAGGGAGTAGTTA	190	992	53.0%
7	GTATCCTTTTGGGAGTAGG	211	1013	47.0%
8	AAATTCATTACCCCGGCC	233	1035	47.0%
9	GCTTGGGTTTGGATGAGTA	258	1060	47.0%
10	TTGTGGATCAGAATCCGGG	331	1133	53.0%
11	CGTAGGCTCAACATAGGGA	358	1160	53.0%
12	GTGGGTGCATGACTGGAAG	389	1191	58.0%
13	TGTAGGTCTTTCATTTGCC	435	1237	42.0%
14	GGTCTGCATAAACTGGGGG	493	1295	58.0%
15	CAGTGGGGTCAAACGCTG	527	1329	58.0%
16	GAGGTCTTGGAGGTCTTTG	547	1349	53.0%
17	ACGAGGGAGGAGCAAAGGT	573	1375	58.0%
18	GCCTCTGATATAAGGCTAT	618	1420	42.0%
19	CTGTAATACCTCGGGTTTC	638	1440	47.0%
20	ACCGGCTAAGGGGTTATAA	658	1460	47.0%
21	CCTTGTTGTGTGGATTGT	696	1498	42.0%
22	CAGGAAGGGTCTTTGCAC	771	1573	58.0%
23	CAGGCCTTGGAGGATAGAG	793	1595	58.0%
24	TACGAAGGCGTGGTAAGGC	817	1619	58.0%
25	AGAGCTATGTTGAGGCGTT	837	1639	47.0%
26	TGTTTGATTGGAGTAGGC	908	1710	47.0%
27	CCTAGAGGGCTATTAGTGT	960	1762	47.0%
28	TGACAAGCCCGCAACATAT	981	1783	47.0%
29	GGGTTTTTATAGGCTGGACA	1036	1838	47.0%
30	CGGAAGCACGGCTGATTTG	1059	1861	58.0%
31	TCTTGACATAGGGGGCATG	1137	1939	53.0%
32	TCTCGCTTCCAGTGAGTTG	1158	1960	53.0%
33	CTCTGGTTCTGGGATAGTG	1195	1997	53.0%
34	CAGCGGGGAGGTCTAATAG	1232	2034	58.0%
35	CCTATGGAGTTTTTGGGT	1260	2062	42.0%
36	TGTGGGGGGGAGGTTAAA	1287	2089	58.0%
37	GGTTAGGAAGGACTTGCTG	1309	2111	53.0%
38	GGCAGAATAGATGCTGGGT	1331	2133	53.0%
39	GGCGGGATCTAACGGTATA	1353	2155	53.0%
40	AGCTTCGATAGTCTTTGGG	1416	2218	47.0%
41	GCTATCGGAAGGACTGTCA	1457	2259	53.0%
42	CCCCTAATACGGATGTATT	1504	2306	42.0%
43	AAGTGATCTTGGGTTTGGC	1529	2331	47.0%
44	AACAATAGGCGTTGCCGG	1587	2389	53.0%
45	CCCAGTTGTTTTGGTATC	1624	2426	42.0%
46	TTGTAAGGCATCACGACCT	1650	2452	47.0%
47	CTCAGGGAGGTACAGGACG	1680	2482	63.0%
48	AGGGAAGTGGCTGATTTCG	1773	2575	53.0%

8.4 Poisson distribution calculations

Carried out by Charles Bangham.

Table 8-4 | Calculations comparing *hbz* expression profile of clones to Poisson distribution

1. Data													
Clone	Code	Total cells	hbz Burst %	hbz - cells	hbz + %	hbz - %	hbz /+Cell	hbz /total cells	predicted hbz - %	obs/predicted hbz - %	Var(hbz spots/cell)		
A	360	3726	3.26	379	89.83	10.17	3.26	2.93	5.3	1.9	5.092		
B	TBX4B	3519	2.69	1005	71.44	28.56	2.69	1.92	14.6	2.0	7.247		
C	1163	2595	2.60	892	65.63	34.37	2.60	1.71	18.1	1.9	5.065		
D	383	5691	2.19	1093	80.79	19.21	2.19	1.77	17.1	1.1	2.153		
E	813	3946	2.23	775	80.36	19.64	2.23	1.79	16.6	1.2	2.472		
Key													
	Hbz+ %	observed percentage of hbz +ve cells, i.e. cells containing one or more hbz spot											
	Hbz- %	similarly, observed % of hbz negative cells in that clone											
	Hbz/total cells	observed mean number of hbz spots per (total) cell in that clone											
2. Chi-squared test (5 x 2) of hypothesis that each clone has a similar % of HBZ-negative cells													
		observed		expected		chi-squared (for non-uniformity of proportions across clones)							
Clone	hbz +	hbz -	total	hbz +	hbz -	total	chi-squared						
A	3347	379	3726	2933.2	792.8	3726.0	58.4	215.9	274.3				
B	2514	1005	3519	2770.3	748.7	3519.0	23.7	87.7	111.4				
C	1703	892	2595	2042.9	552.1	2595.0	56.5	209.2	265.8				
D	4598	1093	5691	4480.2	1210.8	5691.0	3.1	11.5	14.6				
E	3171	775	3946	3106.4	839.6	3946.0	1.3	5.0	6.3				
total	15333	4144	19477	15333	4144.0	19477.0	143.1	529.3	672.4				
Result	chi-squared = 672.4; p < 1E-145			Conclusion: there is significant heterogeneity between clones in the % of HBZ negative c									

3. Test of deviation from the frequency of HBZ-negative cells, assuming that the spot counts are distributed among cells according to a Poisson distribution.												
Clone	hbz +	observed		predicted		chi-squared (each clone vs. Poisson expectation)		individual chi-square		z	p	see note
A	3347	hbz -	total	hbz +	hbz -	total	9.2	163.2	172.4	13.1	3.30E-39	a
B	2514	1005	3519	3005	513.8	3519	80.3	469.6	549.9	23.5	4.10E-122	
C	1703	892	2595	2125	469.6	2595	83.9	379.9	463.9	21.5	1.60E-102	
D	4598	1093	5691	4718	973.5	5691	3.0	14.7	17.7	4.2	2.70E-05	b
E	3171	775	3946	3290	655.6	3946	4.3	21.8	26.1	5.1	3.40E-07	b
	15333	4144	19477	16666	2811	19477						
Notes: a. the p level is estimated using the normal (Gaussian) approximation to the chi-squared distribution.												
b. Clones D and E also depart significantly from expected frequency, the low p value is due to the large number of observations.												
Conclusion: The observed hbz spots do not follow a Poisson distribution among all cells: the frequency of hbz-negative cells is greater than expected from a Poisson distribution. This implies either that a fraction of cells in each clone do not express hbz or, more likely, that each cell passes through an hbz-negative phase.												
4. Test of deviation from Poisson distribution (as in 3. above), by cell-cycle (CC) stage.												
Clone	CC Stage	Total Cells	hbz + Cells	hbz - Cells	obs/ predicted hbz - %	chi^2	p					
A	G0/G1	2232	1955	277	1.2	7.2	0.008					
B	G0/G1	2566	1762	804	1.5	192.5	6.0E-44					
C	G0/G1	1432	899	533	1.5	122.1	4.0E-28					
D	G0/G1	4979	3985	994	1.1	4.4	0.04					
E	G0/G1	3306	2611	695	1.1	3.6	0.06					
Clone	CC Stage	Total Cells	hbz + Cells	hbz - Cells	obs/ predicted H	chi^2	p					
A	G2/M	1054	1012	42	2.8	49.5	2.0E-12					
B	G2/M	530	454	76	4	180.6	6.0E-41					
C	G2/M	191	146	45	3.9	104.1	2.0E-24					
D	G2/M	374	334	40	1.9	19.5	1.1E-05					
E	G2/M	316	291	25	2.2	16.4	4.1E-05					
Note: The p level is estimated using the normal (Gaussian) approximation to the chi-squared distribution, as in 3. above.												
Conclusion: The departure from a Poisson distribution of the frequency of hbz-negative cells is greater in G2/M phase than in G0/G1. The departure remains significant in G0/G1, but in this phase the significance is marginal in the two clones (D and E) that are low or negative in tax.												

8.5 Analysis workflows

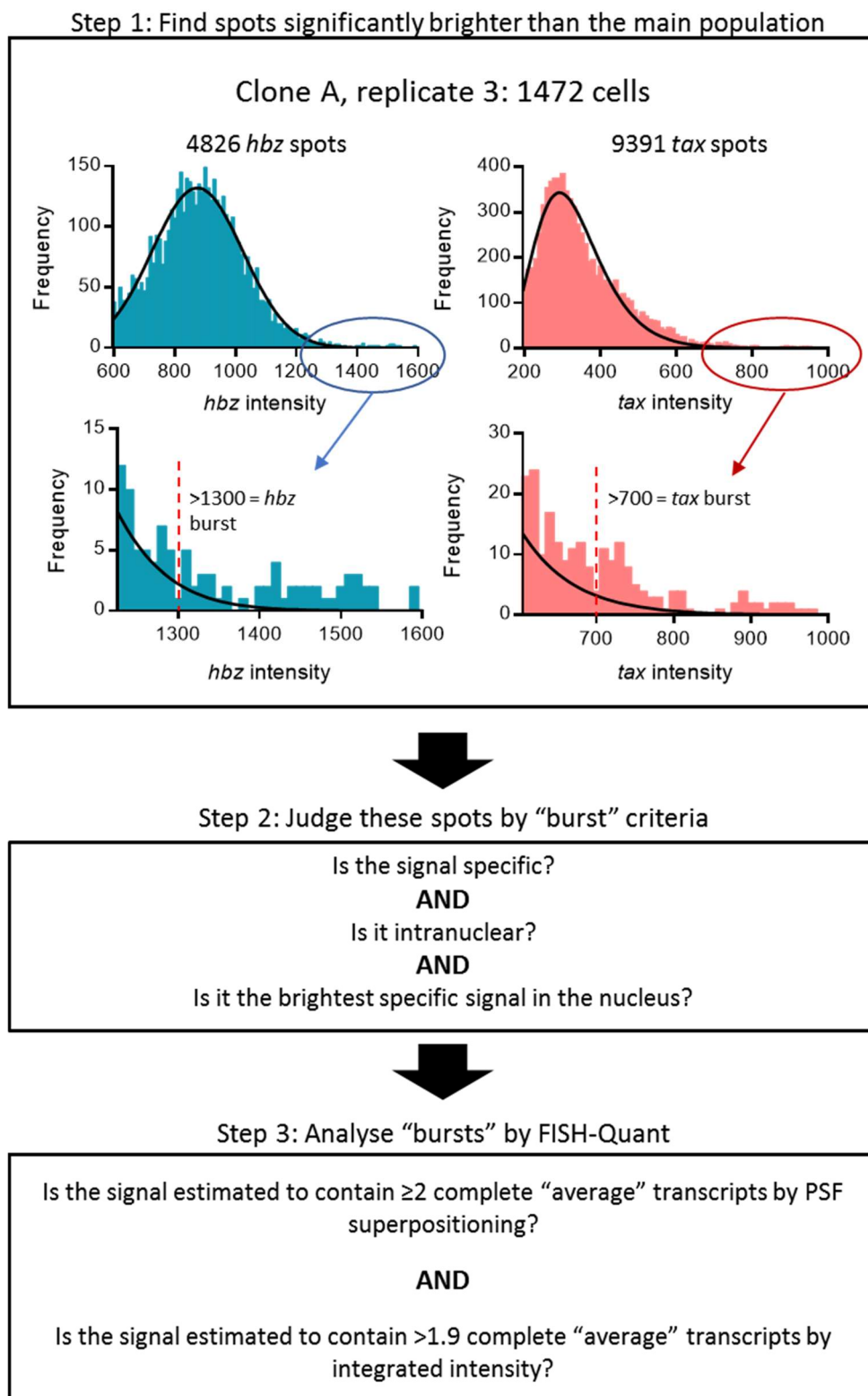


Figure 8-1 | Workflow for identification of transcriptional bursts

- 1) The nuclear “centre of gravity” is calculated by factoring the intensity of each voxel within the nucleus as its mass and the shape of the nucleus
- 2) Burst location is calculated as a proportion of the shortest distance between the nuclear centre of gravity to the nuclear boundary, passing through the burst. The majority of T-cell nuclei are far closer to being spherical than the example below, drawn by Chad Whilding.

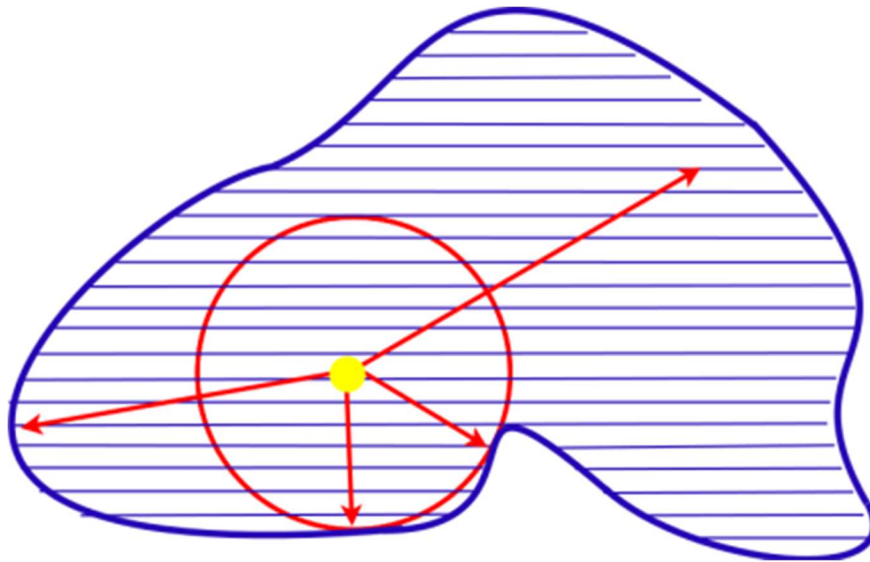


Figure 8-2 | Basis of calculation of transcriptional burst location

8.6 *tax* & *hbz* expression of each cell

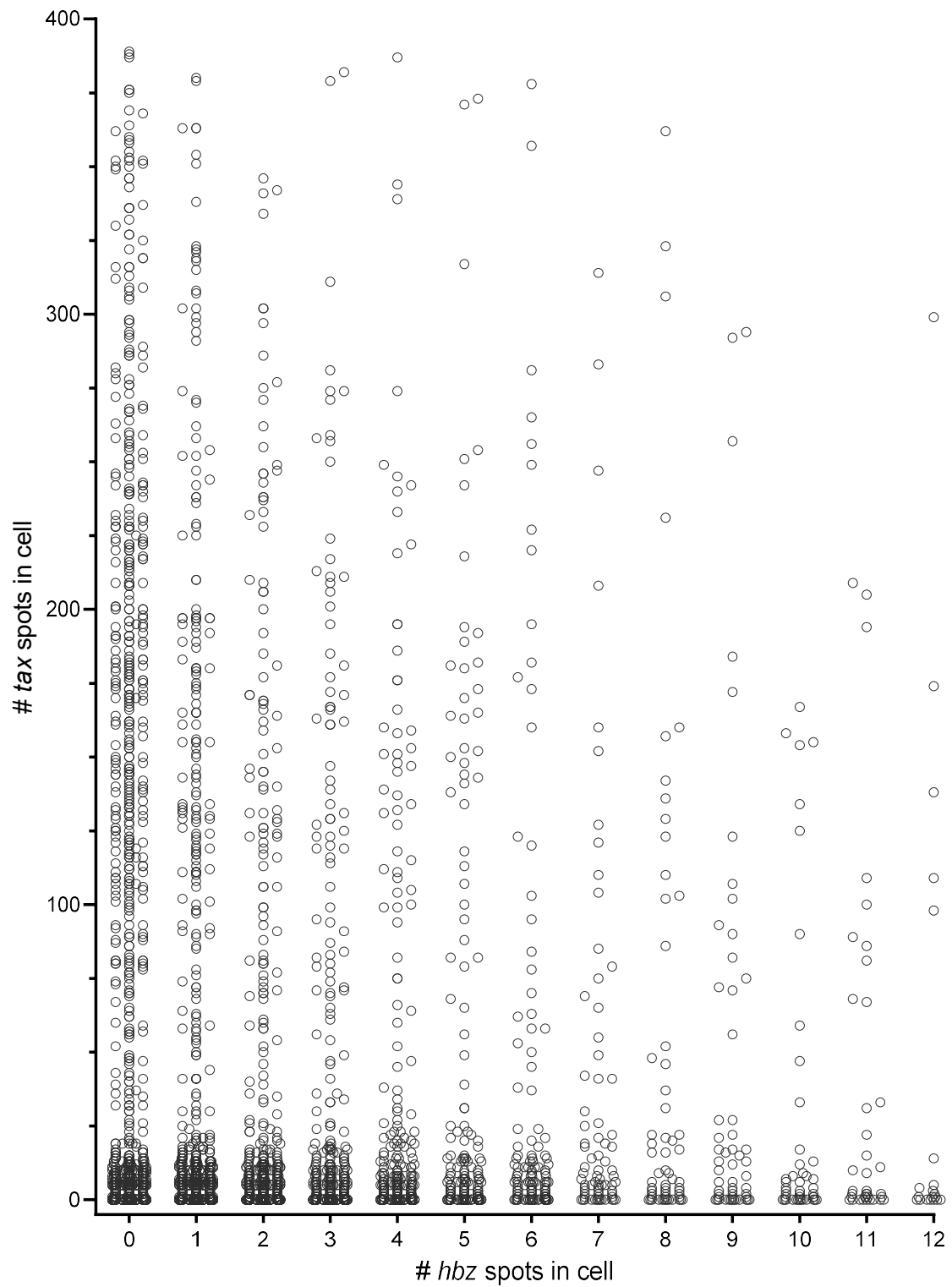


Figure 8-3 | *tax* and *hbz* spot counts for each individual cell

Table 8-5 | Expression metrics of each infected clone

Clone	A	B	C	D	E
Cells	3726	3519	2595	5691	3946
<i>tax</i> - positive (%)	8.48	46.4	90.6	2.92	0.684
# <i>tax</i> / positive cell	44.6	36.8	50.5	3.17	2.04
<i>tax</i> - burst (%)	1.26	4.15	9.60	1.27	0.00
# <i>tax</i> / burst	50.8	27.2	32.6	11.7	n/a
<i>hbz</i> - positive (%)	89.8	71.4	65.6	80.8	80.4
# <i>hbz</i> / positive cell	3.26	2.69	2.60	2.19	2.23
<i>hbz</i> -positive (%)	7.09	4.86	4.16	4.52	3.62
# <i>hbz</i> / burst	3.00	4.22	3.77	2.47	2.64

8.7 Cell cycle data

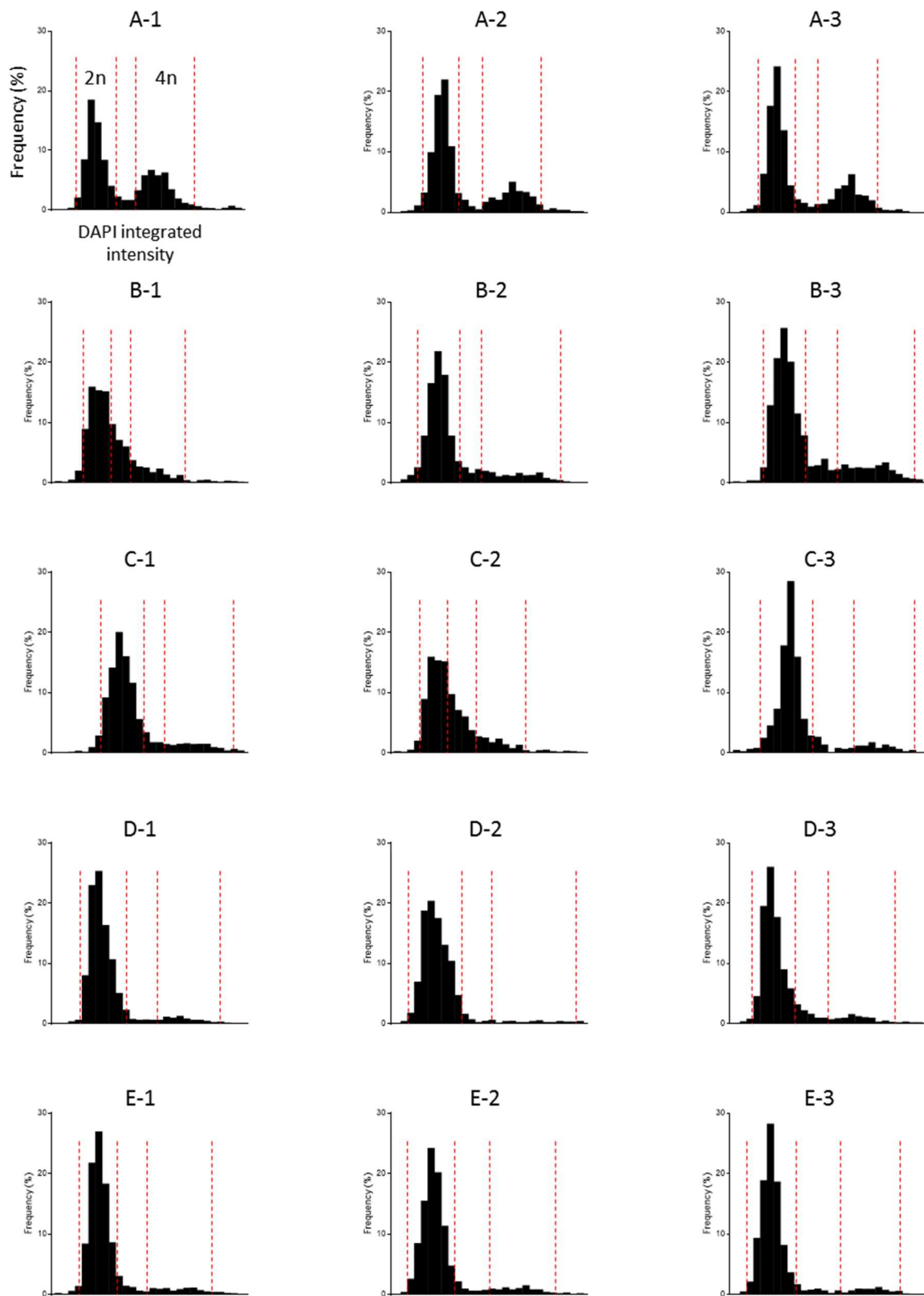


Figure 8-4 | Histograms of each clone replicate's cells' integrated nuclear intensity

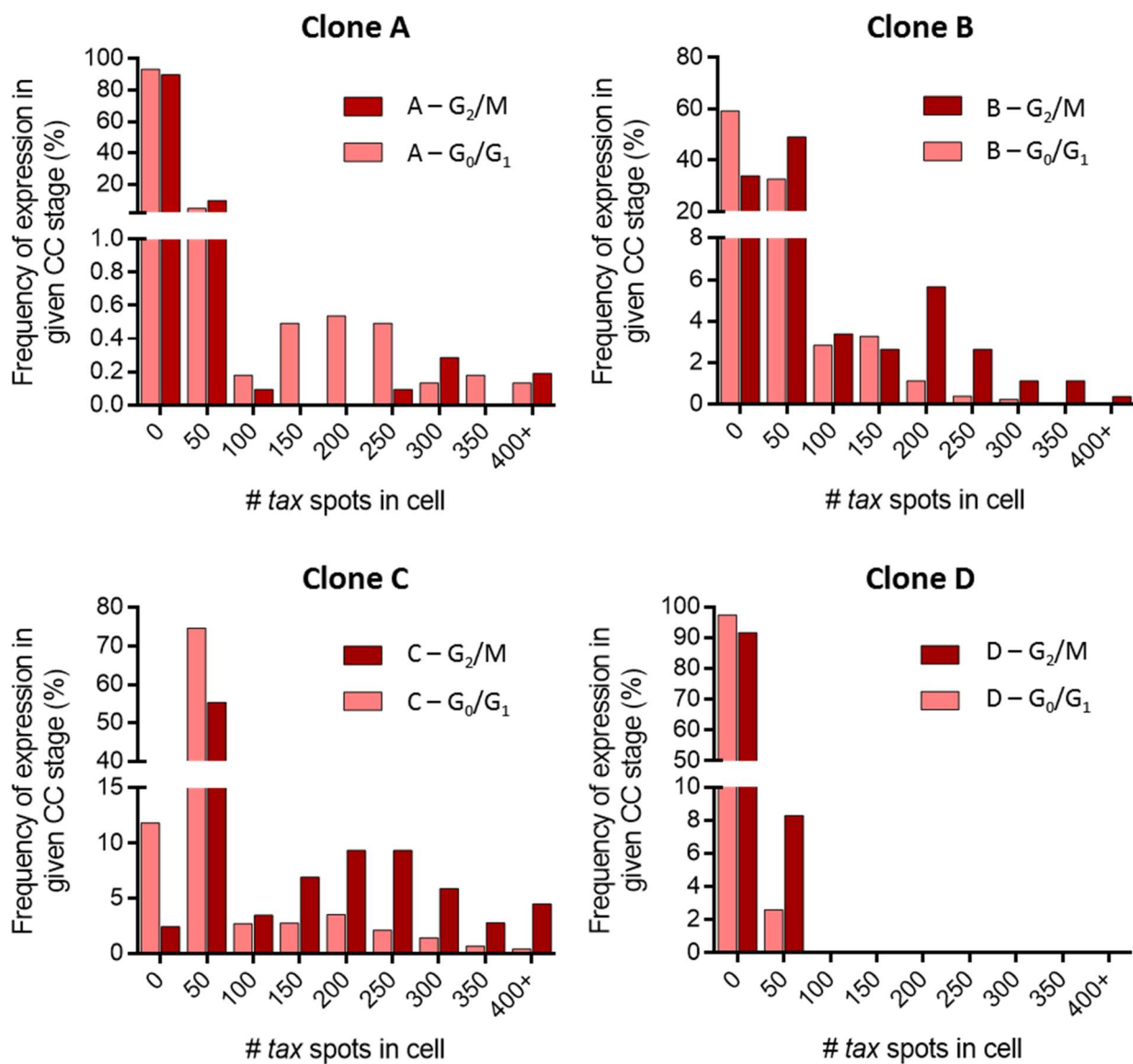


Figure 8-5 | Frequency of *tax* expression in cells split according to cell cycle stage in individual clones

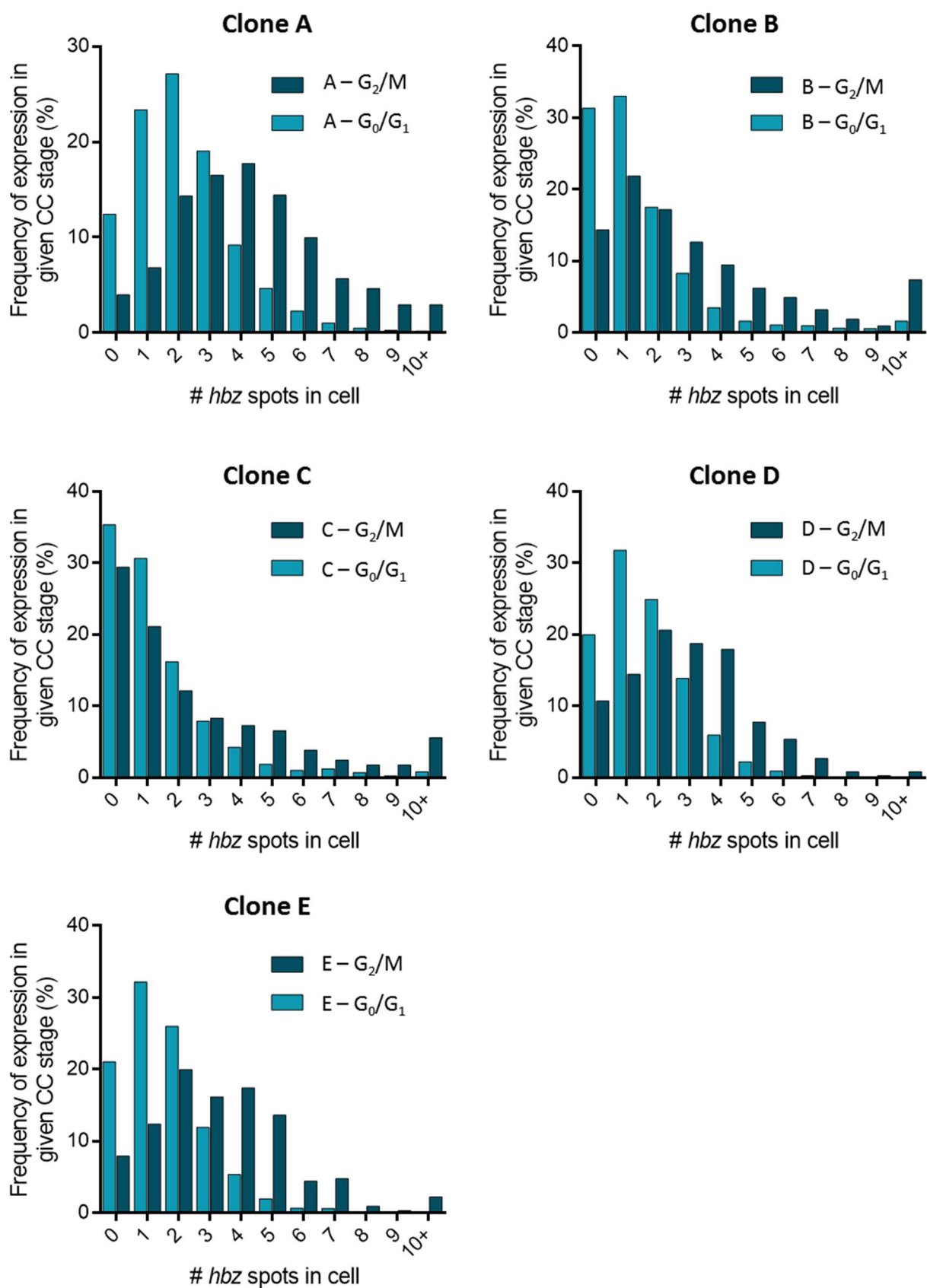


Figure 8-6 | Frequency of *hbz* expression in cells split according to cell cycle stage in individual clones

	Dependent Variables							
Independent Variables	<i>tax</i> bursts		<i>hbz</i> bursts		G_2/M		<i>S</i>	
	Mean OR	p-value	Mean OR	p-value	Mean OR	p-value	Mean OR	p-value
per <i>tax</i> spot	1.024	3.62E-230	1.008	3.54E-30	1.004	5.59E-11	1.004	1.32E-07
>99 <i>tax</i>	135.4	0	5.757	5.68E-35	1.576	0.00102	2.064	1.24E-05
<i>tax</i> burst+			0.2448	7.84E-07	0.8580	0.399	0.8874	0.580
per <i>hbz</i> spot	0.8747	9.91E-05	1.190	1.97E-37	1.448	0	1.147	2.22E-16
>4 <i>hbz</i>	0.5274	0.00646	2.582	7.48E-22	7.304	0	2.055	9.85E-12
<i>hbz</i> burst+	0.2521	4.91E-05			1.381	1.06E-07	1.404	0.016
G_2/M	0.5962	0.0138	1.520	1.34E-05				
<i>S</i>	0.9000	0.646	1.409	0.0104				

Table 8-6 | Odds ratios and p-values from logistic regression in Fig. 5-6

Table 8-7 | Expression metrics of each clone in the context of cell cycle stage

Stage	All			A		B		C		D		E		
	All	G ₀ /G ₁ (%)	S (%)	G ₀ /G ₁ (%)	G ₂ /M (%)	All	G ₀ /G ₁ (%)	G ₂ /M (%)	All	G ₀ /G ₁ (%)	G ₂ /M (%)	All	G ₀ /G ₁ (%)	G ₂ /M (%)
Cells	19477	76.9	7.94	13.2	28.3	3519	72.9	15.1	2595	73.3	11.1	5691	87.5	6.6
<i>tax</i> >1	4493	67.2	11.8	17.3	34.8	1634	64.0	21.4	2350	71.3	12.0	166	77.7	18.7
<i>tax</i> >99	762	50.8	12.5	24.9	10.7	267	50.2	27.0	439	47.6	25.5			
<i>tax</i> burst+	514	59.3	10.3	19.8	8.5	146	53.4	24.0	249	50.6	21.7	72	87.5	12.5
<i>hbz</i> >0	15333	75.3	7.97	15.0	30.2	2514	70.1	18.1	1703	72.2	12.0	4598	86.7	7.3
<i>hbz</i> >4	1822	42.3	9.88	42.3	57.8	346	47.4	37.6	244	44.7	25.8	262	68.3	25.2
<i>hbz</i> burst+	943	60.1	12.1	24.6	36.4	171	58.5	29.8	108	48.1	20.4	257	75.5	15.6
# <i>tax</i> / positive cell	43.1	32.0	42.5	62.5	22.1	36.8	28.5	49.5	50.5	33.6	102.0	3.17	3.25	3.00
# <i>tax</i> / burst	29.8	27.3	36.1	30.2	41.8	27.2	24.6	27.8	32.6	31.6	34.3	11.7	12.0	9.3
# <i>hbz</i> / positive cell	2.56	2.21	2.78	4.00	4.43	2.69	2.32	3.90	2.60	2.22	3.84	2.19	2.09	3.25
# <i>hbz</i> / burst	3.11	3.00	3.16	3.29	3.09	4.22	4.37	3.92	3.77	3.69	4.05	2.47	2.42	2.65
												2.04	2.06	2.00
													n/a	
												2.23	2.06	3.60
												2.64	2.45	3.09