Unravelling the molecular mechanisms of HIV associated neurocognitive disorders through mass spectrometry-based proteomics

Kim Tamara Gurwitz
(GRWKIM001)

Submitted to the University of Cape Town

In fulfilment of the requirements for the degree:

MSc (MED) Medical Biochemistry

Faculty of Health Sciences

University of Cape Town

14 June 2016

Supervisor: Professor Jonathan Blackburn
Co-supervisor: Dr Nelson Soares

DECLARATION

I, *Kim Tamara Gurwitz*, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I have used the Harvard convention for citation and referencing. Each contribution to, and quotation in, this dissertation from the work(s) of other people has been attributed, and has been cited and reference.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 14 June 2016

Acknowledgements

Thank you to my supervisors: Professor Jonathan Blackburn and Dr Nelson Soares, for your guidance and insight throughout the degree.
Thank you to Dr Tariq Ganief and Shaun Garnett, for training and support in cell culture and mass spectrometry activities.
Thank you to Richard Burman and Dr Joseph Raimondo for carrying out the electrophysiological data collection and processing for the current work.
Thank you to Dr Janique Peyper, for proofreading the dissertation manuscript.
Thank you to the entire Blackburn laboratory, for your support.
Thank you to Professor Austin Smith at the Cambridge Stem Cell Institute, UK, for generously donating the cell line used in this current work.
Thank you to Dr Karsten Krug (Proteome Centre Tuebingen, University of Tuebingen, Germany) for sharing your MaxQuant summary statistics R script.
Thank you to the Medical Research Council (MRC) and the National Research Council (NRF) (Free-standing Innovation Masters Scholarship) for financial support of the research.

Abstract

A significant proportion of human immunodeficiency virus type 1 (HIV)-positive individuals are affected by the cognitive, motor and behavioural dysfunction that characterises HIV associated neurocognitive disorders. While the molecular aetiology of this important HIV complication remains largely uncharacterised, HIV transactivator of transcription (HIV-Tat) has been identified as a plausible aetiological cause. Here we have used mass spectrometry-based discovery proteomics to identify the quantitative, cell-wide changes that occur when non-transformed, differentiated human neurons are treated with HIV-Tat over time, as a novel cell culture model representing the initial progression of HIV associated neurocognitive disorders, and as a means to identify putative biomarkers for the illness. We found that our stem cell-based model system displayed morphological and functional neuronal properties and using a Q-Exactive mass spectrometer, we identified over 4000 protein groups (FDR < 0.01) in this system with 131,118 and 45 protein groups differentially expressed at 6, 24 and 48 hours post treatment, respectively. We found changes to the gene expression machinery (*nucleic acid binding* proteins), which suggests that HIV-Tat is involved in preparing the host cell for altered transcriptional and translational activity. We also found cytoskeletal dysregulation in response to HIV-Tat treatment. The 24-hour time point of the time course experiment was largely corroborated with a repeat experiment. A repeat of the entire time course experiment at a lower cell confluence showed that the effect of HIV-Tat treatment to the gene expression machinery was unchanged by cell confluence, while the effect to *cytoskeletal proteins* upon HIV-Tat treatment was present, but less prominent, in lower cell confluence samples. We hypothesise that the gene-expression-machinery effect may be a biphasic response. We further hypothesise that cytoskeletal dysregulation may form part of the molecular mechanism responsible for synaptic injury - as the cytoskeleton is crucial for synapse development and maintenance - and may contribute to memory impairment in HIV associated neurocognitive disorder patients.

Abbreviations

ACN	acetonitrile
ANOVA	analysis of variance
BBB	blood brain barrier
BCA	bicinchoninic acid
BME	beta mercaptoethanol
BSA	bovine serum albumin
buffer A	0.1% FA/dH_2O
buffer B	0.1%FA/ACN
C18	octadecyl bonded silica
CBP	CREB binding protein
ccl2	chemokine ligand 2
CCD	charge couple device
CDK9	cyclin-dependant kinase 9
CHIP-seq	chromatin immunoprecipitation sequencing
CID	collision induced dissociation
CNS	central nervous system
CO_2	carbon dioxide
CSF	cerebrospinal fluid
dH_2O	deionised water
DTT	dithiothreitol
EGF	epidermal growth factor
ELISA	enzyme-linked immunosorbent assay
ESI	electrospray ionisation
FA	formic acid
FASP	filter-aided sample preparation
FDR	false discovery rate
FGF2	fibroblast growth factor 2
fMRI	functional magnetic resonance imaging
GSK3β	glycogen synthase kinase 3 beta
HAART	highly active antiretroviral therapy
HAD	HIV associated dementia
HAND	HIV associated neurocognitive disorders
HCD	higher-energy collision dissociation
HIV	Human immunodeficiency virus type 1
HIV-Tat	HIV transactivator of transcription
HIV-Vpr	HIV viral protein R
HPLC	high pressure liquid chromatography
I	maximum current
IHDS	international HIV dementia scale
IR	input resistance
iTRAQ	isobaric tag for relative and absolute quantitation
K^+	potassium
LC-MS2	liquid chromatography tandem mass spectrometry
LDL	low density lipoprotein
LFQ	label free quantitation
MALDI	matrix assisted laser desorption ionisation

MS	mass spectrometry
MS1	mass spectrometry at peptide ion level
MS2	tandem mass spectrometry (at fragment ion level)
Mtb	*Mycobacterium tuberculosis*
mV	millivolt
MWCO	molecular weight cut-off
m/z	mass-to-charge ratio
Na^+	sodium
NES cell	neuronal epithelial stem cell
NMDAR	N-methyl-D-aspartate receptor
pA	picoamp
PBS	phosphate buffered saline
PEP	posterior error probability
PSM	peptide spectral match
P-TEFb	human positive transcription elongation factor
PTM	post translational modification
RIPA	radioimmunoprecipitation assay
SILAC	stable isotope labelling by amino acids in cell culture
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TAR	viral transactivation response
TFA	trifluoroacetate
TIC	total ion chromatogram
TOF	time of flight
uHPLC	ultra high pressure liquid chromatography
Vm	resting membrane potential

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 HIV associated neurocognitive disorders

South Africa's burden of human immunodeficieny virus type 1 (HIV) infection is well known with a staggering 11.2% of the country's population (corresponding to 6.19 million people) infected with the virus (Department of Health 2015). While the introduction of highly active antiretroviral therapy (HAART) in 1996 has greatly reduced HIV-related mortality and has improved quality of life for HIV-positive individuals, HIV associated comorbidities persist and present a challenge for those treating and living with HIV infection (Brew 2004). One group of complications that affects up to 60% of HIV-positive South Africans, are the HIV associated neurocognitive disorders (HAND) (Joska, Westgarth-Taylor, Hoare, et al. 2011). This is a group of disorders resulting from HIV infection of the central nervous system (CNS), characterised by varying degrees of cognitive, motor, and behavioural dysfunction (Antinori et al. 2007).

While HAART *has* mediated a shift toward less severe forms of HAND, the prevalence of HAND remains high due to a number of factors, including the increased longevity afforded by HAART, incomplete penetrance of antiretroviral drugs into the CNS (allowing for long-term viral replication and subsequent neurotoxicity), and potential antiretroviral-mediated neurotoxicity (Brew 2004; Simioni et al. 2010; Letendre et al. 2008; Valcour et al. 2004; Ances et al. 2010; Marra et al. 2010).

There is some concern that the abovementioned prolonged CNS exposure to HIV and to antiretroviral drugs may compound the long-term adverse effects of HIV infection on neurocognitive functioning, specifically, accelerating neurocognitive aging (Holt et al. 2012). In fact, such premature aging as the result of HIV infection may represent a more pressing global health issue in the HAART era than frank neurocognitive impairment. That said the more severe symptoms of HAND are still relevant, especially in developing countries such as South Africa. Here, 25% of HIV positive patients are diagnosed with the most severe form of HAND (HIV associated dementia (HAD)) upon commencement of HAART therapy (Joska, Westgarth-Taylor, Myer, et al. 2011). The high prevalence of HAD in South Africa upon initiation of HAART may be explained by the national policy on treatment rollout, which stipulates

that HIV positive patients typically qualify for antiretroviral treatment only during late-stage infection (Department of Health 2013), at which point cognitive decline may already be present or may already have progressed (Joska, Westgarth-Taylor, Hoare, et al. 2011).

In all its forms, HAND contributes significantly to the South African healthcare (and ultimately economic) burden. This is largely due to the high prevalence of HIV infection (Department of Health 2015) translating into HAND being the leading cause of dementia in South Africa (Joska, Westgarth-Taylor, Myer et al. 2011; Joska et al. 2011; Nath et al. 2008). The aging HIV-positive population - as a result of the increased longevity afforded by HAART (Brew 2004) and the stabilizing infection rate (UNAIDS 2013) - incentivises the standardization of diagnosis and prognosis of HAND [for example, by identifying predictive markers of disease progression (De Jager et al. 2015)], as well as the identification of drug targets in order to develop novel long-term HAND therapeutic strategies. As HIV enters the CNS early after initial infection (Haase 1986; Fischer-Smith et al. 2001), diagnosis and treatment of HAND are thought to be more beneficial strategies than prevention.

1.2 Current HAND diagnosis and treatment

HAND is largely a diagnosis of exclusion in HIV-positive individuals with impairment in multiple cognitive domains (e.g. attention, working memory) (Antinori et al. 2007). Further, at present, no rapid, reliable diagnosis or specific treatment exists for HAND, and the exact molecular dysregulation underlying neurocognitive symptoms is largely uncharacterised. Current diagnostic tools and treatments offer varying levels of practicality and have been used with mixed success.

For example, while the international HIV dementia scale (IHDS) is a diagnostic test that may be quickly administered and is claimed to be free of cultural bias (Sacktor et al. 2005), it has low accuracy especially for the correct diagnosis of subtle forms of HAND and may be fundamentally flawed due to the absence of a diagnostic gold standard (Haddow et al. 2013). In another HAND diagnostic example, functional magnetic resonance imaging (fMRI) - which measures blood flow as a proxy for brain function - has been suggested as a HAND risk-stratification tool (Holt 2012, Ances et al., 2010). However, given that fMRI is expensive, specialised, and not routinely performed in South Africa, this may not be a viable option in our context. Therefore,

there is a need to identify specific, predictive biomarkers of HAND, which may be detectable by more readily available and inexpensive means. In addition, to facilitate specific biomarker identification (as well as rational drug design), a more detailed understanding of the molecular pathology of HAND is required.

In an example of molecular pathology informing treatment, lithium has been used to treat HAND based on one of its proposed mechanisms of action - the targeting and inhibition of glycogen synthase kinase 3 beta (GSK3β) (an important factor in apoptosis induction) - combined with the observation that *in vitro* treatment of neurons with a known HIV virulence factor, HIV transactivator of transcription (HIV-Tat), leads to phosphorylation and up-regulation of GSK3β's activity (Brown & Tracy 2013; Forde & Dale 2007; Maggirwar et al. 1999; Rao et al. 2013). Furthermore, *in vitro* treatment with lithium has been shown to block HIV-Tat-mediated apoptosis of neuroblastoma cells (Ganief, unpublished). However, while lithium treatment has yielded some success regarding improvement of HAND symptoms during clinical trials (Letendre et al. 2006), as well as improvement of brain network connectivity in humans *in vivo* (Tivarus et al. 2015), apoptosis-inhibited cells may yet be maintained in a neurodegenerative state post-treatment. In addition, lithium has a very narrow therapeutic window, an undesirable adverse effect profile, and long-term exposure to lithium in the context of HIV infection is not well characterised.

Identification of 'upstream' drug targets may facilitate reversion of a neurodegenerative to a normal phenotype, thereby preventing progression of degeneration and improving prognosis. Furthermore, elucidating the exact molecular pathology of the disease, as it progresses over time, may reveal prognostic markers for HAND. Used clinically, such markers may inform at which point after HIV infection HAART should ideally be initiated, and could aid in the identification of possible drug targets for adjuvant therapy for treatment of neurocognitive sequelae.

1.3 Plausible molecular origins of HAND

Of particular importance in elucidating the molecular pathology of HAND is an improved understanding of the actions of HIV-Tat, as mentioned a known neurovirulent factor (Rao et al. 2013). To contextualise HIV-Tat's role in HAND, however, it is also important to consider the wider pathophysiology of HIV entry into the CNS, as well as

all plausible molecular origins of HAND as the true aetiology is most likely multifactorial.

It is well understood that HIV crosses the highly selective blood brain barrier (BBB) early during infection, undetected within peripheral monocytes, which periodically enter the CNS to replenish the brain perivascular macrophage population ('Trojan Horse' model, first proposed by Haase 1986) (**Figure 1.1**). Indeed, *in situ* hybridisation analysis reveals that HIV antigens co-localise with monocytes, suggesting that HIV does target monocytes for infection in the periphery (Fischer-Smith et al. 2001). Further, HIV infected monocytes show greater chemotaxis in response to the chemokine ligand 2 (ccl2) than uninfected monocytes, which suggests that HIV contributes to the movement of HIV-infected monocytes into the CNS (Williams et al. 2013).

Less clear is the degree to which individual HIV-derived molecules and secondary effects contribute to HAND progression. However, the clinical characteristics of HAND (Manji et al. 2013) suggest a subset of likely molecular origins, such as HIV-related or antiretroviral-mediated. Regarding the former, HIV infection of primary CNS cells, and the subsequent inflammatory response, produces products toxic to neurons when present at elevated levels, and HIV proteins themselves are able to enter neurons and affect neuronal functioning (**Figure 1.2** and **1.3**). Regarding the latter, antiretroviral drugs are able to penetrate the BBB or enter the CNS after prolonged inflammation-induced degradation of the BBB, and may have neurotoxic effects (Martín-García et al. 2002) (**Figure 1.4**). It is important to note, however, that the contribution of antiretroviral drugs to HAND progression is the least-studied and most widely contested possible origin of HAND, and is beyond the scope of this thesis.

As neurons do not display the CD4 cell surface receptors required for HIV entry, HIV must affect neurons via the abovementioned indirect mechanisms (Martín-García et al. 2002). While it is important to characterise the effects of all of these etiologies individually as well as together (the contribution of each may be patient-specific and may inform an individualized treatment plan), HIV-Tat has been shown to be particularly important in HAND pathogenesis across most (if not all) patients, as discussed in the following section.

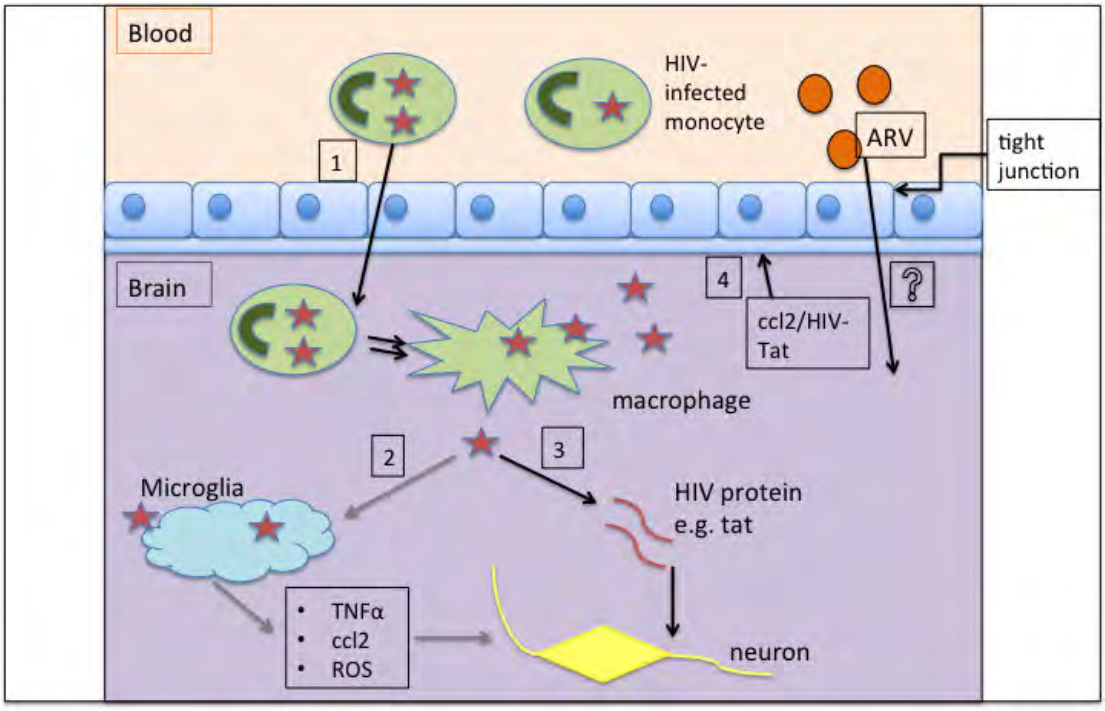


Figure 1. Entry of HIV into the CNS and its subsequent effects HIV crosses the BBB undetected, shuttled within peripheral monocytes, which enter the CNS to differentiate into macrophages ('Trojan Horse' model: Haase 1986; Williams et al. 2013) **(1)**. Once inside the CNS, HIV cannot enter neurons via the canonical 'host CD4 cell surface receptor - HIV-gp120' entry, as neurons do not display CD4 cell surface receptors (Martín-García et al. 2002). HIV must therefore affect neurons by alternate mechanisms either indirectly **(2)**, by triggering and sustaining an inflammatory response within the CNS, the mediators and products of which are toxic to neurons, or directly **(3)**, through the action of HIV proteins on neurons themselves, such as HIV-Tat (Martín-García et al. 2002). Certain antiretroviral drugs are able to penetrate the BBB **(4)** to cause neurotoxicity, or are able to enter the CNS after HIV mediated degradation of the BBB. Degradation of the BBB occurs through the action of HIV proteins (e.g. HIV-Tat) themselves, or through the action of certain chemokines (e.g. ccl2) which are up regulated during the HIV-induced inflammatory response (Robertson et al. 2013; Roberts et al. 2012). Note: HIV particles are depicted as red stars, ROS – reactive oxygen species; TNFα - tumor necrosis factor alpha (a cytokine); tat - HIV-Tat; ARV - antiretroviral drug. Double arrow indicates differentiation. Question mark - unclear exactly how these agents enter the CNS.

1.4 HIV-Tat: structure, function and role in HAND pathogenesis

HIV-Tat is a 86-101 residue protein comprising two exons; the first contains all known functional regions of the protein and is highly conserved (72 amino acid residues) and the second is of variable length (depending on the HIV subtype) and unknown function (**Figure 2a**) (Tahirov et al. 2010). The cysteine rich region of HIV-Tat, located within exon 1, is crucial to its transcription factor activity. These amino acids contribute to the formation of a unique zinc finger motif that recognizes and binds directly to HIV-Tat's crucial host cofactor, human positive transcription elongation factor (P-TEFb), which comprises cyclin-dependant kinase 9 (CDK9) and cyclin T (Ott et al. 2011). The zinc finger motif - made up of helices that coordinate 2 zinc ions - is formed by

intramolecular disulphide linkages between cysteine residues together with the adjacent core region of HIV-Tat (Tahirov et al. 2010). However, cysteine rich regions on *adjacent* HIV-Tat molecules may form *inter*molecular disulphide linkages that disrupt this crucial motif and render the protein inactive **(Figure 2b)** (Tosi et al. 2000) - an important consideration when working with this protein experimentally.

No consensus exists for the mechanism by which free HIV-Tat crosses the cell membrane, but it is accepted that HIV-Tat readily enters cells and accumulates in the nucleus [due to its nuclear localisation signal (**Figure 2a**)] - a feature exploited in the development of a drug delivery system, appropriately named 'cell penetrating peptides' (Richard et al. 2005; Herce & Garcia 2007; Duchardt et al. 2007). In HIV-infected cells, HIV-Tat enhances HIV transcription elongation by recruiting host transcription machinery - such as P-TEFb - to the viral transactivation response (TAR) RNA (Gatignol 2007). P-TEFb bound to HIV-Tat allows viral transcription elongation to proceed from the nascent viral RNA by phosphorylating the C-terminal domain of host RNA polymerase II, which, in turn, allows for further recruitment of host transcription machinery to the viral genome (Ott et al. 2011).

In uninfected cells, HIV-Tat may still affect transcription by binding to accessible host cognate cites (Marban et al. 2011). Indeed, a genome-wide chromatin immunoprecipitation sequencing (ChIP-seq) experiment revealed that HIV-Tat extensively binds the host genome of T lymphocytes (Marban et al. 2011). Further, specific host factors have exhibited altered gene expression as a result of HIV-Tat treatment *in vitro* (Flores et al. 1993; Buonaguro et al. 1994). In addition, due to the flexible nature of HIV-Tat (both structurally and functionally), and the fact that it has many binding partners (Debaisieux et al. 2012), HIV-Tat may additionally affect transcription by recruiting distinct factors to the site of host transcription. Indeed, HIV-Tat is able to recruit chromatin-modifying proteins such as the histone acetylases p300 and CREB binding protein (CBP), thus facilitating relaxation of chromatin structure and subsequent increased transcription (Gatignol 2007). This presumably facilitates increased HIV replication. In this way, HIV-Tat may directly modulate host transcription, and subsequent translation, resulting in global changes to host-cell function.

In support of HIV-Tat's importance as a HAND virulence factor, the level of HIV neurovirulence has been directly linked to an HIV-Tat protein dicysteine motif at position 30-31; this motif correlates HIV clade with the burden of HAND worldwide

(Rao et al. 2013). Even for classically non-neurovirulent clade C HIV [which predominates in Sub-Saharan Africa (Peeters et al. 2003)], the correlation holds. That is to say that while a cysteine-serine motif at position 30-31 is usually associated with clade C virus (and non-neurovirulence), the abovementioned dicysteine motif is present in clade C virus isolated from *HAND patients* (Rao et al. 2013; Kandathil et al. 2009). In addition, *in vitro* assays show that, in comparison to clade C HIV-Tat with the CS motif, clade C HIV-Tat with the CC motif induces greater monocyte chemotaxis and greater neurotoxicity (Rao et al. 2013), again supporting HIV-Tat's role as a neurovirulent factor.

Further to HIV-Tat's neurovirulence effects, similarities between HAND and other dementias may also be explained by the actions of HIV-Tat. For example, HAND patients clinically present with molecular events very similar to those seen in Alzheimer disease. To illustrate, in HAND, disrupted amyloid plaque clearance may result from HIV-Tat's inhibition of neprilysin (a protein responsible for clearance of amyloid plaques), which is also dysregulated in Alzheimer disease (Achim et al. 2011; Rempel & Pulliam 2005). Therefore, despite different aetiologies, at least some similar molecular processes appear to underlie both types of dementia. Indeed, transcriptome analysis reveals an overlap between gene expression networks in brain tissue samples from Alzheimer disease and HAND patients and proteome analysis reveals an overlap in differential expressed proteins in a similar comparison of these dementias (Levine et al. 2013; Zhou et al. 2010; Rempel & Pulliam 2005). Given HIV-Tat's contribution to neurovirulence, this protein may at least in part explain these similarities. While a detailed comparison of the molecular events underlying distinct dementias is beyond the scope of this thesis, the evidence does suggest that HIV-Tat exposure is important in dementia progression during HAND.

Regarding what is known about HIV-Tat's mechanism of neurodegenerative action, it causes N-methyl-D-aspartate receptor (NMDAR)-mediated apoptosis, as well as post-synaptic density loss *in vitro* by binding to low density lipoprotein (LDL) receptor-like protein (Eugenin et al. 2007; King et al. 2010; Kim et al. 2008). While this information might be important in the development of targeted HAND therapies, given that synaptodendritic injury is a better correlate of HAND than frank neuronal loss, and is believed to be a reversible process (Ellis et al. 2007; Everall et al. 1999; Bellizzi et al. 2005), it is perhaps more beneficial to focus on HIV-mediated synaptic damage (rather than induction of apoptosis and frank neural loss) in order to identify damage reversal

strategies. Indeed, HIV-Tat exposure is linked to impaired learning and memory - processes known to involve synaptic dynamics (Goellner & Aberle 2012) - as well as grey matter deficits in mice (Carey et al. 2012; Carey et al. 2013). These observations point to HIV-Tat's direct involvement in the HAND phenotype and suggest a functional consequence of HIV-Tat exposure. Elucidation of the molecular correlates of synaptic damage is therefore important for a better understanding of HAND pathogenesis. While other HIV proteins, such as HIV viral protein R (HIV-Vpr), can negatively affect neuronal physiology, HIV-Tat seems particularly important as it is the only HIV protein actively secreted by infected primary immune cells in the CNS, suggesting that the local CNS concentration of HIV-Tat may be relatively high (Na et al. 2011; Nath 2002; Perry et al. 2010).

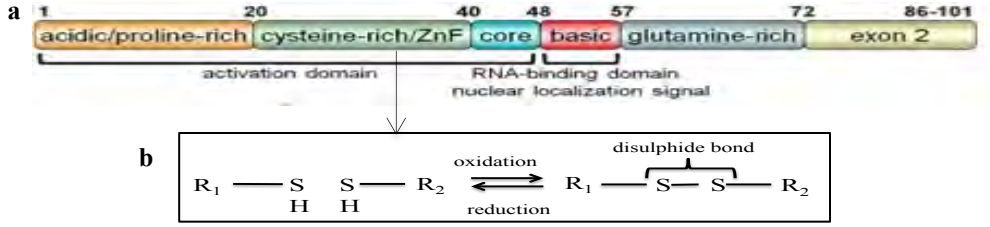


Figure 2. Linear representation of HIV-Tat's structural domains a Domains of HIV-Tat. Exon 1 contains all known functional regions of HIV-Tat while exon 2 is of variable length and unknown function (image from Tahirov et al. 2010). **b** Intermolecular disulphide bonds may form between adjacent HIV-Tat monomers causing multimerisation. R_x-SH: cysteine amino acid with SH (sulphur–hydrogen) functional group.

1.5 Cell culture models in HAND research

Presently, most HAND research utilises rodent models of HIV infection (Carey et al. 2012; Carey et al. 2013) or indirect studies of human cognitive functioning (e.g. via fMRI, Holt et al. 2012). While important insights may be gleaned from these approaches, the applicability of rodent findings to humans, and the sensitivity of indirect methods, is questionable. In the absence of *in vivo* human studies or readily available primary human CNS tissue from patients, a representative human cell culture system is a good practical alternative for HAND research, with the applicability and sensitivity required for rational prognostic marker and drug target discovery.

While the transformed cell line SH-SY5Y (a human neuroblastoma-derived cell line) has been routinely used as a cell culture model of HAND (Cartier et al. 2003), it

may not necessarily be representative of an otherwise healthy human CNS. However, progress in the field of stem cell technology has made possible more 'realistic' cell culture systems, for example *in vitro* maintenance of untransformed normal cells: human embryonic neuronal stem cells have been maintained for up to 100 passages without genetic immortalization, and can be differentiated to form mature human neuronal populations (Sun et al. 2008; Tailor et al. 2013). Such cell culture systems hold much promise for the study of neurological diseases, including HAND. In this current work, we employ such a system in determining the molecular pathology of HAND over time (see section **5.1 NES cell line**).

1.6 Mass spectrometry

Given the lack of reliable diagnostic tests or specific treatments for conditions such as HAND, mass spectrometry (MS)-based discovery proteomics is a valuable technology. Instead of looking at only a small handful of molecules, MS 'widens the net' for putative biomarker and drug target discovery, and provides the scope and sensitivity necessary to quantify cell-wide changes in protein expression in response to stimuli (including disease state) (Walther & Mann 2010). While HAND research has made use of MS to identify appropriate biomarkers, few if any studies have used MS-based discovery proteomics to investigate cell-wide proteomic changes in neurons in response to HIV-derived factors (Donnelly & Ciborowski 2015; Laspiura et al. 2007; Wiederin et al. 2009). Indeed, although certain HIV-Tat molecular targets have been identified in cells, knowledge regarding global cellular changes in response to this known neurovirulent factor is limited, and impedes development of sensitive diagnostic tests and rational drug design.

An MS instrument can be likened to a very sensitive mass scale, which is able to measure molecular mass, or more specifically, ion mass-to-charge ratio (m/z). From the characteristic m/z, one is able to infer analyte identity and quantity using software (discussed in more detail below). The instrument comprises three parts: an ion source to produce charged, gas phase ions; a mass analyser to measure the m/z of resultant ions; and a detector to register the number of ions at each m/z value and transmit this information to a data system to record the mass spectra.

A routine MS-based discovery proteomics workflow for the measurement of ionised peptides (in which information from peptide ions is used to infer the presence of

particular proteins in a sample - 'bottom up approach') can be described as follows: total cell proteome is extracted from cell lysate and digested - typically with trypsin - to generate predictable peptides products. Proteins are digested before MS analysis as peptides ionise more easily than full proteins and sequence information is more readily obtained from short peptides (Duncan et al. 2010). Further, peptides are desired as they may be easily separated into 'fractions', typically by high pressure liquid chromatography (HPLC) based on peptide hydrophobicity (i.e. reversed-phase chromatography) using octadecyl bonded silica (C18) as the solid phase packing material. Peptide fractionation prior to MS ensures that a less complex peptide mixture enters the MS instrument at any given time, thereby allowing the MS instrument to detect as many peptide species as possible (Domon & Aebersold 2010). Ultra high pressure liquid chromatography (uHPLC) may be employed to increase separation speed, decrease sample requirement and increase both sample resolution and subsequent identification rates (Rodriguez-Aller 2010).

Peptides are next ionised to facilitate their entry into, and detection by, the MS instrument. For complex mixtures, separation is typically done 'on-line' with ionisation in which peptides are ionised as they elute from the liquid chromatography column (**Figure 3a**). Electrospray ionisation (ESI) is the preferred ionisation method for 'on-line' fractionation, as ESI requires analytes in the liquid phase. In ESI, the solution of peptides is sprayed in the form of charged droplets from a glass capillary under a strong electrical field. Once exposed to the atmosphere, the solvent evaporates and transfers charge to the peptides thereby creating gaseous peptide ions in a process that is not entirely understood (Konermann et al. 2013). An alternative to ESI is matrix assisted laser desorption ionisation (MALDI), in which sample is spotted onto a target plate with suitable light-absorbing matrix and subsequently desorbed and ionised upon laser-pulse excitation of the matrix (Schwamborn & Caprioli 2010). MALDI is particularly useful for maintaining the spatial distribution of a sample, such as in a tissue section, as the entire section may be placed on the target plate (Schwamborn & Caprioli 2010). As mentioned, as ESI can be coupled to HPLC it is favoured over MALDI for discovery proteomic analysis.

After ionisation, peptide ions enter the mass analyser for measurement and detection. There are many mass analysers in use, each employing a different physical property of the peptide ion to determine the ion's m/z; the most popular three are outlined here. 1) The simplest mass analyser is the time of flight (TOF) analyser, in

which ions are accelerated by an electric field of known strength. The time taken for the peptide ion to reach the detector at the end of a fixed length evacuated tube is measured and allows for the calculation of the peptide ion's mass (Aebersold & Mann 2003). 2) The quadrupole mass analyser comprises four parallel metal rods surrounding a hollow central channel. Oscillating electric fields are applied to the rods such that only ions within a narrow m/z range are able to stably navigate through to the detector (Miller & Denton 1986; Aebersold & Mann 2003). As such, the quadrupole is often used for the targeted identification of molecules where the m/z is known (Miller & Denton 1986). 3) The Orbitrap mass analyser maintains ions in orbital motion around a central spindle shaped electrode and obtains a signal from the lateral motion of ions around this electrode within an electrostatic field (Makarov 2000). The signal is used to create a Fourier transformation, which is, in turn, used to determine the ion's m/z (Makarov 2000). The current work obtained MS data from a Q-Exactive instrument, which has an Orbitrap mass analyser.

Upon detection, the data system connected to the MS instrument outputs MS1 spectra of m/z versus intensity. The intensity of a peptide peak correlates with the relative amount (detection-dependent) of a particular protein in the original experimental sample. If a peptide ion is present at or above a certain intensity (which may or may not be an arbitrary user-defined value) that peptide ion may be selected and fragmented for a second round of MS (tandem MS a.k.a. MS2), to facilitate generation of fragment ion amino acid sequence information (**Figure 3b**) (Steen & Mann 2004).

To elaborate how MS1 and MS2 spectra are obtained, during an MS 'run', MS1 spectra are collected continuously as eluting peptides are ionised and enter the mass analyser. This can be visualised as a total ion chromatogram (TIC), which depicts ion intensity peaks over the entire MS run-time (**Figure 4a**). If one 'zooms in' at a particular time point in the TIC, an individual MS1 spectrum may be visualised (**Figure 4b**). Multiple peptides that co-elute from the HLPC column (and that appear as a summed peak on the TIC) may be detected in a single MS1 spectrum - where each peak potentially indicates a peptide ion. Due to MS2 scan rate limitations, a subset of these peptide ions, for example the 'top 10' (highest intensity), are selected for subsequent fragmentation. Dynamic exclusion is also often set at a particular interval to ensure that peptides selected for MS2 during a full MS scan - according to their peak intensity - are not selected again during that interval period, thereby allowing for wide peptide sequence coverage (a.k.a. data-dependent analysis) (Domon & Aebersold 2010).

Peptide ions selected for MS2 are fragmented to produce fragment ions by collision with an inert gas (e.g. Helium) in a process known as collision-induced dissociation (CID). Higher-energy collision dissociation (HCD) is a modified version of CID specific to the Orbitrap setup in which fragmentation occurs outside of the Orbitrap itself (Olsen et al. 2007). Peptide ions fragment most commonly at the peptide bond to form a series of b/y ions, from which sequence information for the peptide may be obtained through *de novo* sequencing (**Figure 5**) (Steen & Mann 2004). More typically in MS-discovery proteomic experiments, experimental MS2 spectral patterns are mapped to theoretical MS2 spectra generated by *in silico* trypsin digestion of the relevant proteome database (**Figure 3c** and **Figure 4d**). The identification of proteins/protein groups from a high throughput discovery proteomic experiment therefore crucially depends on the availability of a full reference proteome for the species being analysed.

As the MS instrument only *detects* peptide and fragment ion m/z, much downstream computational processing is required to determine protein/protein group identity and quantity for the millions of peptide spectral sets generated (each set comprising an m/z value for the peptide ion and the associated MS2 spectrum). While many softwares exist, MaxQuant open-source software is routinely used by many laboratories to this end as it is able to reliably extract protein identity and quantity information from raw MS2 and MS1 spectra, respectively (Cox et al. 2009). Specifically, MaxQuant's internal search engine, Andromeda, identifies protein groups by probability-based comparison of experimental MS2 spectra to theoretical MS2 spectra generated by *in silico* trypsin digest of the reference proteome provided (Cox et al. 2011) (**Figure 3c**). Andromeda [similar to the Mascot algorithm, which is also probability based (Perkins et al. 1999)] assesses both peptide ion mass and the number and masses of MS2 fragments from the same spectral set in determining its assignments (Cox et al. 2011). SEQUEST, yet another software, makes use of the same probability-based algorithm as Andromeda and Mascot, but first filters the theoretical database, based on a MS1 mass tolerance defined by the experimental MS1 spectrum, before the theoretical spectra for comparison are generated (Eng et al. 1994).

Due to the shear volume of spectra and the absence of certain peaks in experimental samples (among other factors), the possibility of obtaining false positives is high and must be accounted for statistically to ensure confidence in the assignment. In Andromeda, a posterior error probability (PEP) score is given to each assignment,

which can be interpreted as a local false discovery rate (FDR) value; the smaller the PEP, the more confidence that the experimental and *in silico* spectra did not match due to chance alone (Cox et al. 2011). The PEP score is calculated for a particular peptide identification score by comparing the number of peptide spectral matches (PSM) to a decoy database (comprising reverse peptide sequences) for a particular score with the number of PSMs to the true database for that same score (Käll et al. 2008). In addition, the algorithm uses peptide length, where the longer the peptide that matches the true database, the more likely it is that this match did not occur by chance (Cox et al. 2011). This is a *post hoc* score used to verify quality of the identification. In comparison, the global FDR is used to adjust stringency of the matching algorithm (e.g. Andromeda algorithm), again based on the number of hits to decoy and true databases (Käll et al. 2008).

Differences in protein quantity, between treated and control samples, are typically determined in relative terms. This is in contrast to the less common *absolute* quantitation where exact molecular quantities per protein are identified (e.g. AQUA) (Brun et al. 2007). To aid relative quantitation, samples may be labelled ahead of MS analysis. For example, labelling on the protein level may be achieved by stable isotope labelling of amino acids in cell culture (SILAC) (the gold standard), in which 'heavy' or 'light' forms of certain essential amino acids are metabolically incorporated into treated and control cells, respectively (or visa versa) (Ong & Mann 2006). Treated and control samples are combined ahead of proteome processing and run on the MS instrument as a single sample. Due to the presence of heavy and light labels, one may distinguish between the treated and control 'version' of a particular protein by a discriminate mass difference between isotopes on the mass spectrum. Proteins may alternatively be labelled chemically at the peptide level by isobaric tag for relative and absolute quantitation (iTRAQ), in which each experimental sample is labeled with a tag of different composition (but same overall mass) and, upon MS2 fragmentation, unique reporter groups are released from the tags, which can be read in the lower m/z portion of the mass spectrum (Ross et al. 2004). The relative ratio of these reporter ions is compared to give an indication of relative abundance of a particular protein across experimental conditions (Ross et al. 2004).

The ability to combine labelled treated and control conditions (as early as cell harvesting if SILAC is used) reduces any technical variability that may be introduced during proteome processing ahead of MS analysis. However, due to methodological and

technical advancements in MS, label free techniques - in which there is no labelling of proteins or peptides - are gaining popularity. Unlike labelled techniques, no additional (often costly) steps are required during the cell culture/proteome processing stage of sample preparation and there is little limit to the type and number of sample/s that may be analysed (e.g. certain samples, such as human tissue samples, cannot be SILAC-labelled and both SILAC and iTRAQ are limited in the number of experimental conditions that may be compared due to a finite number of labels).

Label free quantitation may be achieved by spectral count and peak intensity. In spectral count, the number of identified MS2 spectra (spectral count) for a particular protein is compared across samples with the assumption that the number of MS2 spectra directly correlates to protein abundance, protein sequence coverage and number of identified unique peptides. This method remains controversial, however, as it relies on counting spectra rather than measuring physical data (Bantscheff et al. 2007). In the peak intensity method, elution into the MS instrument of a particular peptide with a given m/z ratio is monitored over time to establish an elution profile. The cumulative area under the curve of the profile for this monoisotopic mass is then used to determine peptide abundance, which has been shown to correlate linearly with area under the curve (Bondarenko et al. 2002). MaxQuant's label free quantitation (LFQ) value is calculated by a peak intensity approach. Additionally, MaxQuant's 'match between runs' function aids peptide identification in samples with incomplete spectra for particular peptides. Briefly, if a peptide is identified in one sample, its identification information may be transferred to an unidentified spectrum in a second sample through matching up mass with retention time across samples (Cox et al. 2014).

It is important to note that a major caveat of discovery proteomics is the problem of protein inference; proteins can only be uniquely identified if the MS instrument detects a peptide that is unique to the protein. Therefore in many instances, protein groups, and not individual proteins, are identified. This commonly occurs, for example, for isoforms of a protein that differ only slightly in amino acid sequence (Nesvizhskii & Aebersold 2005). All members of a protein group are equally as likely to have been present in the experimental sample and one should be mindful of this during downstream biological significance analysis. To account for the possible ambiguity, gene name (instead of specific protein identity) is often used in order to encompass all the protein isoforms.

Once a list of protein groups, per sample, has been identified, statistical analysis

may be performed to determine those protein groups that show significant differences in quantitation between treated and control conditions. For label free quantitation, the LFQ value may be used as the quantitation measurement for each protein group per sample. Tests of significance typically include 2-sampled independent t-test or analysis of variance (ANOVA) (if more than 2 conditions are being compared at once). These tests take into account the mean expression value (e.g. mean LFQ) for replicates of a particular protein group per condition, as well as the standard deviation of the mean, when determining whether a difference in the means of the treated and control conditions for a particular protein group is significant. Protein groups that meet a user-defined cut-off (usually $p < 0.05$) are deemed significant. It is this list of significantly differentially expressed protein groups that is taken forward for biological significance analysis, such as functional classification (e.g. Panther classification, www.pantherdb.org) and annotation (e.g. GeneCards, www.genecards.org).

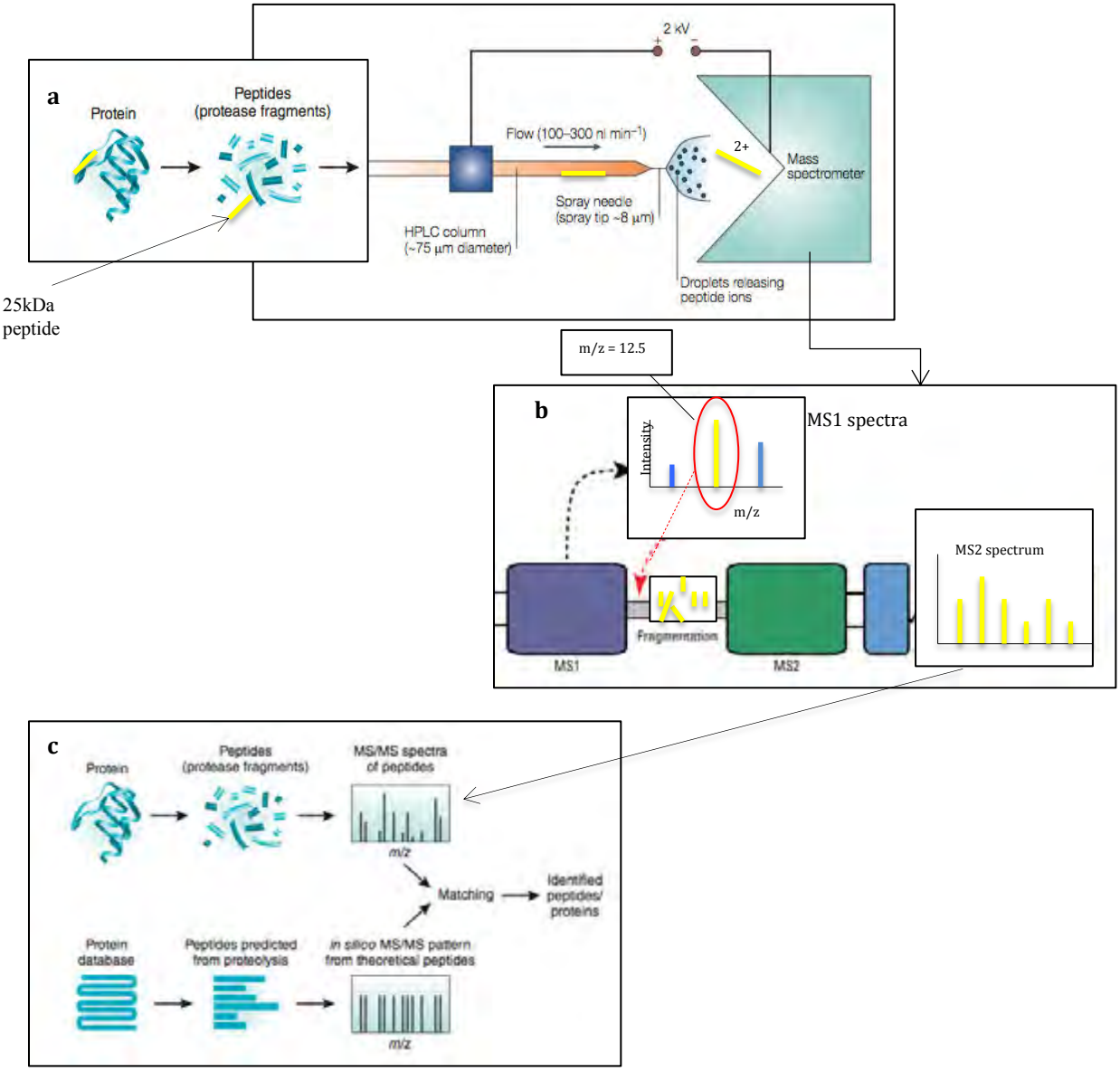


Figure 3. MS-based discovery proteomics workflow used in the current work a Total cell protein is digested, by trypsin, to form a peptide mixture. Peptides are separated 'on-line' by HLPC and are subsequently ionised by ESI, before entering the MS instrument for detection. For example, the 25kDa peptide, indicated by a yellow rectangle, acquires two positive charges during ESI to form a doubly-charged ion (2^+). **b** Peptide ions are detected at the MS1 level. Selected peptide ions are individually fragmented and detected as MS2 spectra. To continue the example of the 'yellow peptide', this 25kDa 2^+ peptide is detected as having a mass-to-charge (m/z) ratio of 12.5 at the MS1 level. This peptide ion is subsequently selected and fragmented into 6 fragments, which are detected as an MS2 spectrum, where each peak corresponds to one fragment (fragment ion). The MS2 spectrum is then used for peptide identification (see **3c**). **c** Experimental and theoretical MS2 spectra are matched *in silico* by a software package such as MaxQuant to allow for protein identification (Figures adapted from: Duncan et al. 2010; Steen & Mann 2004; ThermoScientific, www.piercenet.com/method/overview-mass-spectrometry).

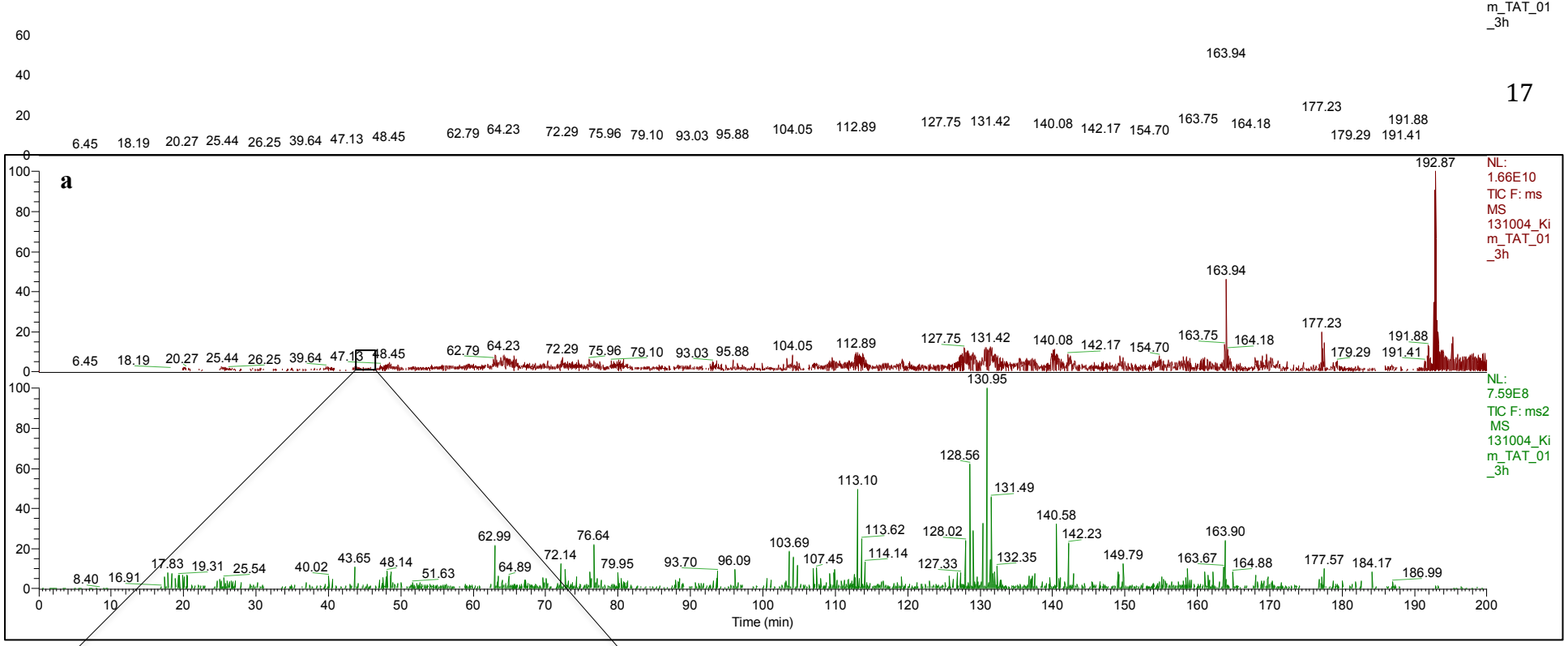


Figure 4. Sequence and protein identification workflow for each spectrum identified a All peptide ions detected during a 200 minute run (i.e. TIC) at the MS1 (red) and MS2 (green) level for a representative experimental sample. **b** An MS1 spectrum captured at 46.214 minutes. **c** A single ion, corresponding to a single peptide, is selected for MS2. Sequence information may be extracted (see **Figure 5**). **d** Annotated experimental MS2 spectrum is matched to *in silico* generated theoretical spectra by Andromeda search engine in MaxQuant for protein identification. (**Figures 4a-c** obtained from current experimental work, **Figure 4d** obtained from Duncan et al. 2010). **Note**: at an intensity of 1.6e10, the TIC was higher than desired and this sample was rerun (See **Results**: optimisation of sample injection volume). This run is included here purely to illustrate the MS workflow using data generated by the current work.

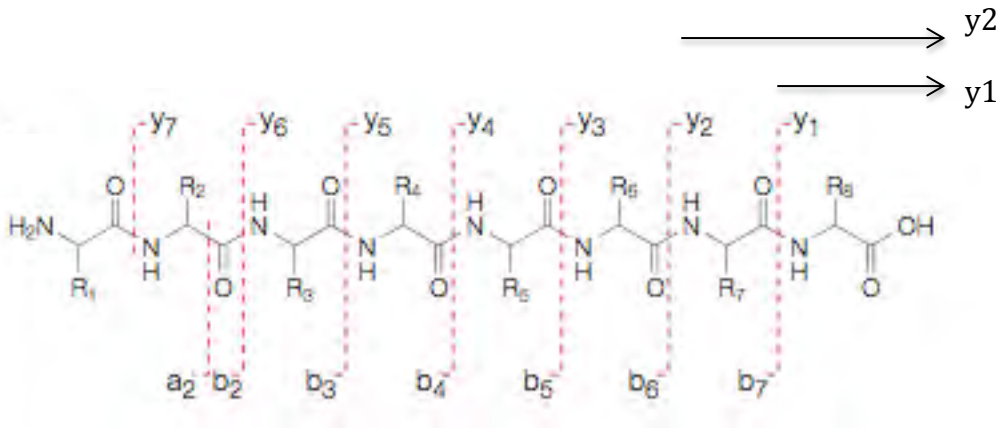


Figure 5. Fragmentation sites along the peptide backbone The most common fragmentation site (upon fragmentation by the MS instrument) is the NH-C=O peptide bond, resulting in formation of a y- or b-ion series. The C-terminal end of the peptide is retained by y-ions. For example, y1 comprises the C-terminal amino acid, while y2 (a di-peptide) comprises the C-terminal amino acid as well as the adjacent amino acid. Conversely, b-ions retain the N-terminal amino acid upon fragmentation. For example, b6 (corresponding to y2) comprises 6 amino acids. Each amino acid has a characteristic molecular mass. The difference in mass between y2 and y1, for example, will reveal the molecular mass of the amino acid at the C-terminus of the peptide, thus enabling identification of this amino acid (provided it is not modified). In this way the peptide fragment can be sequenced *de novo*. The b- and y-ion series determine sequence information from opposite ends of the peptide and can be used to confirm each other (image adapted from Steen & Mann 2004).

2. Aims

The current work aimed to use MS-based discovery proteomic analysis to identify the cell-wide effects of full-length, exogenous, clade B HIV-Tat on non-transformed, differentiated human neuronal stem cells over time, representing an *in vitro* model of early HAND progression. In doing so we hoped to identify putative prognostic and diagnostic biomarkers and/or drug targets for improved diagnosis and treatment of HAND. In addition, we aimed to determine the effect of neuronal cell culture confluence, and (presumably) resultant synaptic density, on HIV-Tat treatment. Further, we aimed to characterize the neuronal properties of the *in vitro* cell culture system used in the current work.

3. Objectives

- Generate and analyse electrophysiology data from control neurons to establish whether the cell line used in this work (at the time of treatment administration) is representative of neurons in *in vitro*.

- Optimise (further to Gurwitz Honours dissertation, 2013) the beta mercaptoethanol (BME) concentration that is able to maintain HIV-Tat in a monomeric state during resuspension of the protein and for long-term storage*.

  *NOTE: HIV-Tat readily forms multimers at high concentrations of the protein due to formation of disulphide bonds between cysteine rich regions of adjacent HIV-Tat monomers in solution (**Figure 2**). As HIV-Tat is required in a monomeric state for activity, HIV-Tat must be kept in a reducing environment (by a reducing agent, such as BME) for short- and long- term storage (Tosi et al. 2000).

- Conduct MS-based discovery proteomic analysis, with label free quantitation, on lysates of cells (differentiated at ~90% cell culture confluence) exposed to the following conditions: HIV-Tat treatment or BME (vehicle only control), for 6, 24 or 48 hours.

- Identify differentially expressed protein groups (from MS data), between treated and control conditions. Determine functional enrichment of protein groups that are significantly differentially expressed in order to identify molecular markers that become dysregulated in these cells in response to HIV-Tat treatment over time. Analyse this information to identify putative HAND biomarkers and to hypothesise mechanisms of early HIV-Tat neurovirulence.

- Repeat time point 2 of the time course experiment (cells harvested after 24 hours of HIV-Tat treatment).

- Repeat entire time course experiment with cells differentiated at lower cell culture confluence (~70%) in order to determine the effect of cell density and (presumably) synaptic density on HIV-Tat treatment.

4. Experimental Procedures

4.1 Cell culture

Stem cells

The cell line used in this work was a non-transformed, neuroepithelial-like stem (NES) cell line derived from 5-week-old (Carnegie stage 15-17) human foetal hindbrain, generously donated by Austin Smith's group at the Cambridge Stem Cell Institute. The NES cells were maintained and differentiated as described in Tailor et al. (2013). Briefly, cells were seeded at approximately 26 000 cells/cm^2 on non-pyrogenic, cell adhesion plates coated with poly-L-ornithine (0.01% w/v in phosphate buffered saline (PBS), Sigma Aldrich) and laminin (0.2% v/v in PBS, Sigma Aldrich) and incubated at 37°C, 95% humidity and 5% carbon dioxide (CO_2). The NES were cultured in DMEM: Hams F12 medium (Sigma Aldrich) supplemented with L-glutamine (2mM, Sigma Aldrich), penicillin/streptomycin (1% v/v, Lonza), N2 (1:100 v/v, Invitrogen), B27 (1:1000 v/v, Invitrogen), and epidermal growth factor (EGF) and fibroblast growth factor 2 (FGF2) (both 0.01ng/µl, Invitrogen) (Tailor et al. 2013).The growth medium was changed daily. For cell culture passage, cells were split 1:3 using tryp-LE™ (Invitrogen) once a density of approximately 70 000 cells/cm^2 (~90% cell culture confluence, determined visually) was reached, which occurred roughly every third day.

Differentiation

NES cells were grown until 90% cell confluence was reached, at which point EGF and FGF2-containing growth medium was removed and replaced with growth medium lacking growth factors and containing a higher concentration of B27 (1:100 v/v) to induce spontaneous NES differentiation (Tailor et al. 2013). The differentiation medium was identical to the stem cell culture medium except for the above-mentioned changes. Cells were washed with PBS before addition of differentiation medium to ensure removal of growth factors. Cell were incubated at 37°C, 95% humidity and 5% CO_2. During the differentiation process and during maintenance of differentiated neuronal cultures, the medium was changed twice weekly.

Visualisation of cells

The NES were visualised using transmitted light, phase contrast setting (20x objective) on a Zeiss Axiovert 200M Fluorescence microscope with a Zeiss AxioCam high-resolution monochrome camera, and images were captured using Zeiss AxioVision software (version 4.8).

4.2 Electrophysiological analysis of differentiated NES cells

Electrophysiological studies of control NES cells, the data processing thereof, were carried out by Richard Burman in a collaboration with Joseph Raimondo's electrophysiology laboratory (University of Cape Town). Experimental procedures were as follows: once day 9 post differentiation was reached, differentiated NES cells - attached to cell adhesion dishes - were removed from incubation and transported rapidly and carefully to the recording chamber of a Zeiss Axioskop Upright Microscope (Zeiss) for electrophysiological recordings. Recordings were made in NES cell differentiation medium (Tailor et al. 2013) at room temperature. Recordings were only obtained during the first 2 hours following NES cell removal from the incubator environment.

A horizontal puller (Model P-1000, Sutter) was used to pull patch pipettes (13-20 MOhm tip resistance) from filamental borosilicate glass capillaries (2.00 mm outer diameter, 1.58 mm inner diameter, Hilgenberg, Germany). The pipettes were filled with an internal solution comprising (in mM): K-gluconate (126); KCl (4); Na 2 ATP (4); NaGTP (0.3); Na 2 -phosphocreatinine (10) and HEPES (10). Osmolarity was adjusted to between 290 and 300 mOsM and the pH was adjusted to between 7.38 and 7.42 with KOH.

NES cells under electrophysiological analysis were visualised with a 40x water-immersion objective (Zeiss) and digital images were obtained with a charge couple device (CCD) camera (VX55, TILL Photonics). Recordings were made in current clamp and voltage clamp mode with an Axopatch 200B amplifier (Molecular Devices). Data acquisition was performed through an ITC-1600 board (Instrutech) connected to a personal computer running custom-written clamp sequences (PulseQ) under IGOR Pro (Wavemetrics). Analysis was performed using custom-written MATLAB (Mathworks) scripts.

4.3 HIV-Tat reconstitution and visualisation

For long-term storage, HIV-Tat (Clade B, 86 amino acids, Diatheva) was suspended in 50mM BME, degassed with N_2 and acidified to pH4 with trifluoroacetate (TFA, Sigma Aldrich), at a storage concentration of 100ng/µl HIV-Tat. For short-term storage, suspended HIV-Tat was buffer exchanged from 50mM BME to 10mM BME (final working concentration of 0.1mM BME) at 4°C using molecular weight cut-off (MWCO) filters (3kDa, Millipore). An assumption of no protein loss was made during buffer exchange and the final eluted volume was adjusted to a concentration of 100ng/µl HIV-Tat in 10mM BME for short-term storage at -80°C.

HIV-Tat was visualized by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with gels comprising 5% bisacrylamide (Sigma Aldrich) stacking, and 15% bisacrylamide separating gels (see **Table A1**, Appendix A). Gels were run at constant 160V in a 1x tris-glycine running buffer. Dithiothreitol (DTT) was omitted from the sample application buffer to showcase the reduction ability of BME at specified concentrations. PageRuler Prestain Ladder (ThermoScientific, SM0671) was used on all gels. Gels were stained with Silver Stain (ThermoScientific) and imaged with an iPhone 5s.

4.4 HIV-Tat treatment

At day 9 post differentiation, differentiated NES cells were treated with a final working concentration of 1ng/µl HIV-Tat in 0.1mM BME. NES cells treated with 0.1mM BME only, were used as a control. All conditions were performed in at least triplicate as far as possible. Samples were incubated at 37°C, 95% humidity and 5% CO_2 for 6, 24 and 48 hours. After incubation as per experimental time point, cells were lifted using tryp-LE (Gibco), pelleted at 600g for 3 minutes, followed by two washes in PBS at 600g for 3 minutes each to remove cell debris. Supernatants were aspirated, cell pellets flash frozen in liquid nitrogen, and pellets subsequently stored at -80°C. Cell counts were performed using the Trypan blue (Sigma Aldrich) dye exclusion method and cell counts were correlated to protein quantity (see below for protein quantitation method).

4.5 Cell lysis

Cell pellets were suspended in deionised water (dH_2O) containing 1x protease inhibitors (Sigma Aldrich) and 1x phosphatase inhibitors (PhosStop, Roche) and subsequently lysed with a 1x Tris/HCl pH7.6 radioimmunoprecipitation assay (RIPA) buffer (adapted from Holden & Horton 2009 buffer 4 - see Appendix A) excluding 10mM DTT, which interferes with bicinchoninic acid (BCA) quantitation (Pierce™ BCA Protein Assay Kit instructions). Cell lysates were each incubated on a rotor for 1 hour at 4°C with 1µl Benzonase nuclease ultrapure (Sigma Aldrich). Cell lysates were finally centrifuged at 14 000g for 10 minutes. Supernatants, containing the cell protein content, were transferred to new 1.5ml flip top tubes.

4.6 Sample preparation for MS analysis - protein quantitation, digestion and cleanup

Protein quantitation, digestion and cleanup were performed in parallel for all samples. Pierce™ BCA Protein Assay Kit microplate procedure (ThermoScientific) was utelised for protein quantitation with Bovine Serum Albumin (BSA) as a standard. Absorbance values were read at 560nm with Biorad iMark™ microplate reader, including path correction for sample volume. BSA standard curves were generated using Microplate Manager software (version 6.1). Total protein present in each sample was then reduced with 100mM DTT for 30 minutes at room temperature.

Peptides for MS analysis were generated on 30kDa MWCO filters (Millipore) using the filter-aided sample preparation (FASP) protocol developed by Matthias Mann's group (Wiśniewski et al. 2009). Briefly: samples were mixed with 8M urea (a non-detergent chaotrope) in tris/HCl pH8.5 in order to remove low molecular weight compounds, present in the lysis buffer (e.g. SDS), that would interfere with downstream MS if not removed (essentially a buffer exchange step); thiol groups were carboamidomethylated by 0.5M iodoacetamide to prevent intra- and inter-molecular disulphide bridge formation in order to facilitate protein digestion with trypsin (incubation with DTT prior to FASP aids this step by reducing the thiols thereby making them available for carboamidomethylation); proteome was digested; and resultant peptides were eluted off the MWCO filters and acidified with 0.1% formic acid (FA) in ultrapure dH_2O (both Sigma Aldrich) (Wiśniewski et al. 2009). An equivalent amount (30µg) of protein from each sample - as determined by BCA

quantitation - was added to each filter, respectively. Samples were incubated with trypsin (1:50 w/w, Promega) for 12 hours.

Samples were desalted using reverse-phase C18 chromatography (C18 columns manufactured in-house from C18 discs, Millipore) in preparation for MS analysis. C18 columns were equilibrated with 80% MS grade acetonitrile (ACN, Sigma Aldrich) acidified with 0.1% FA. 10µg peptide from each sample was desalted on separate C18 columns - one column per sample - by the addition of 2% ACN acidified with 0.1% FA. Finally, peptides were eluted with 60% ACN acidified with 0.1 % FA. Desalted peptides were dried at room temperature using a SpeedyVac (Savant) and resuspended in 0.1% FA (Sigma Aldrich) in ultrapure dH_2O to a final concentration of 250ng/µl (500ng/µl for the 24-hour repeat experiment), which was estimated from BCA quantitation, at the protein level, on the assumption of no protein loss during FASP.

4.7 Liquid chromatography-MS analysis parameters

Liquid chromatography tandem MS (LC-MS2) analysis was performed on a Q Exactive MS instrument coupled to a Dionex uHPLC (R53500nano) (both Thermo Fischer Scientific). For label free quantitation, samples were run individually. Sample volumes for injection were optimized batch-wise. For the 24-hour repeat experiment, 5µl of 0.5µg/µl per sample was injected. Samples were monitored throughout data collection to ensure machine and column stability as well as desired raw data quality. All samples, for all experiments, were injected into the sample loop at a flow rate of 3µl/min in 2% buffer B (0.1%FA/ACN) and 98% buffer A (0.1% FA/dH_2O). Samples were then loaded onto a 50cm, 75µm diameter, C18 reverse phase analytical column (3.6µm Aeris C18 packing material, packed in house, Separations) and subsequently eluted during a 150-minute segmented gradient (4-30% buffer B) at a flow rate of 300nl/min. Eluents were ionised by ESI (~ 2.5kV) before entering the mass analyser. At the end of each sample run, a 30-minute wash cycle was run. The wash comprised a segmented gradient (5-55% buffer B, and final 10 minutes at 80% buffer B). Every four samples, a longer wash (80 minutes) was run.

Mass spectra were collected with settings essentially as those described by (Pirmoradian et al. 2013) with the following amendments: scan range was set at 300-1750 m/z; the ‘top 10’ method was used in cycling between the collection of MS1 and

MS2 spectra with 70 000 resolution at MS1 level; and dynamic exclusion was set at 30 seconds (s). Elaborated scan parameters may be viewed in **Table 1**. Ion fragmentation was performed using the HCD method.

Table 1 Mass spectrometry scan parameters

MS level	Parameter	Setting
MS1	Run time	10-145 minutes
	Polarity	positive
	Default charge state	2
	Scan range	300-1750 m/z
	Resolution	70000
	Automatic gain control target	3e6
	Maximum injection time	250ms
MS2	Resolution	17500
	Automatic gain control target	5e4
	Maximum injection time	80ms
	TopN	10
	Isolation window	2 m/z
	Scan range	200-2000 m/z
	Intensity threshold	6.3e3
	Dynamic exclusion	30s

4.8 Data analysis

Raw mass spectra were processed for peptide peak identification and quantitation by MaxQuant software (version 1.5.03) (Cox et al. 2009). Default MaxQuant settings were used with the following modifications: labeling was set to singlet; LFQ was selected for relative label free quantitation; 'match between runs' was selected to increase the probability of identifying protein groups in all samples and to aid in the comparison of protein group quantities across samples; and only peptides unique to a protein group were used for quantitation. MaxQuant's inbuilt search engine, Andromeda, was used for protein group identification using the UniProt Human proteome database as the reference (FASTA, downloaded on 12/01/15) (Cox et al. 2011). In addition, MS2 spectra were compared to a decoy database, comprising reverse sequences of each protein, in order to statistically account for false positives.

The datasets generated by MaxQuant were filtered to exclude protein groups

that mapped to a sequence in the decoy database or that mapped to a known contaminant. In addition, only protein groups that had an LFQ measurement in each experimental sample, and had a satisfactory PEP score (< 0.01), were taken forward for downstream data analyses.

MS data quality and summary statistics were assessed with the aid of a script developed by Karsten Krug (Proteome Center Tuebingen, University of Tuebingen, Germany) in the R environment for statistical computing. Further data analysis was performed using R scripts developed in-house, for which LFQ values were used as a relative measure of protein group expression. Scatterplots and boxplots of protein group expression values were performed to establish technical variability within replicates and variability across experimental conditions. Hierarchical cluster analysis, using a Euclidean model of absolute distance, was performed in order to determine whether samples in a particular treatment group had similar protein expression profiles. Protein expression profiles were visualised using heat maps, in which the darker the banding pattern, the greater the expression of that particular protein. Differential expression analysis was performed by hypothesis testing, using the independent two-sampled t-test ($p < 0.05$) for each protein group (HIV-Tat treated versus control), in order to identify differentially expressed protein groups. Hypothesis testing results were visualised by volcano plot.

4.9 Biological significance analyses

Functional classification analysis and gene annotation were performed using Panther (www.pantherdb.org; Mi et al. 2013; Mi et al. 2013(1)) and GeneCards (www.genecards.org), respectively, in order to determine the biological meaning of the data. Gene names, corresponding to differentially expressed protein groups (from the comparison of HIV-Tat treated versus control conditions) were used for the analyses. Gene names, and not protein groups, were used in order to reduce false hits due to the 'protein groups problem' (Nesvizhskii & Aebersold 2005).

In Panther Gene List Analysis, the following parameters were selected: 1. 'ID list' selected as list type, 2. Species set to *Homo sapiens*, and 3. functional classification viewed on bar chart. The Panther interface allows one to view bar charts of Panther protein class ontologies at progressively lower Panther protein class terms, thereby corresponding to more and more specific groupings. Bar charts where

recreated, at various hierarchical levels, as line graphs in R (using scripts developed in-house) in order to visualize the change in number of genes that map to a particular Panther classification over time. Each bar chart, per time point, was also viewed as a table of genes (from the input list of differentially expressed protein groups) that mapped to that particular Panther protein class. GeneCards was then consulted to functionally annotate the genes belonging to each grouping. Gene function was correlated to $\log_2$ fold change (HIV-Tat/ control) in order to gain insight as to the effect of HIV-Tat treatment on the neuronal proteome.

Further, statistical over-representation analysis (version 10.0, released 15/05/15) of Panther protein classes per time point was carried out - using the online Panther tool - in order to determine the Panther protein classes significantly over/under represented when compared to the background. The following parameters were user-specified: organism set to *H. sapiens*; list of differentially expressed genes, per time point, was used as the 'analysed list'; full list of genes identified, per time point, was used at the 'reference list' (i.e. background); and Bonferroni correction was set to 'false'.

5. Results

5.1 NES cell line

NES, the cell line used in this work, was derived from 5-week-old (Carnegie stage 15-17) human foetal hindbrain (Tailor et al. 2013). NES cells grew as a monolayer and formed characteristic rosette-like groupings (**Figure 6a**) as has been previously characterised (Tailor et al. 2013). The cells spontaneously differentiated when growth factors, EGF and FGF2, were removed from the growth medium, and after 9 days there were drastic changes to cell morphology, such as decreased cell body size and extensive neuronal network formation (**Figure 6b**). As the cells closely responded to growth cues from the medium (e.g. the removal of growth factors), the assumption that these cells are not genetically immortalized was supported. As such, these non-transformed, differentiated cells cell were preferred over the transformed, neuroblastoma-derived SH-SY5Y cell line routinely used in cell culture models of HAND (Cartier et al. 2003). All HIV-Tat treatments were administered at 9 days post differentiation.

a **b**

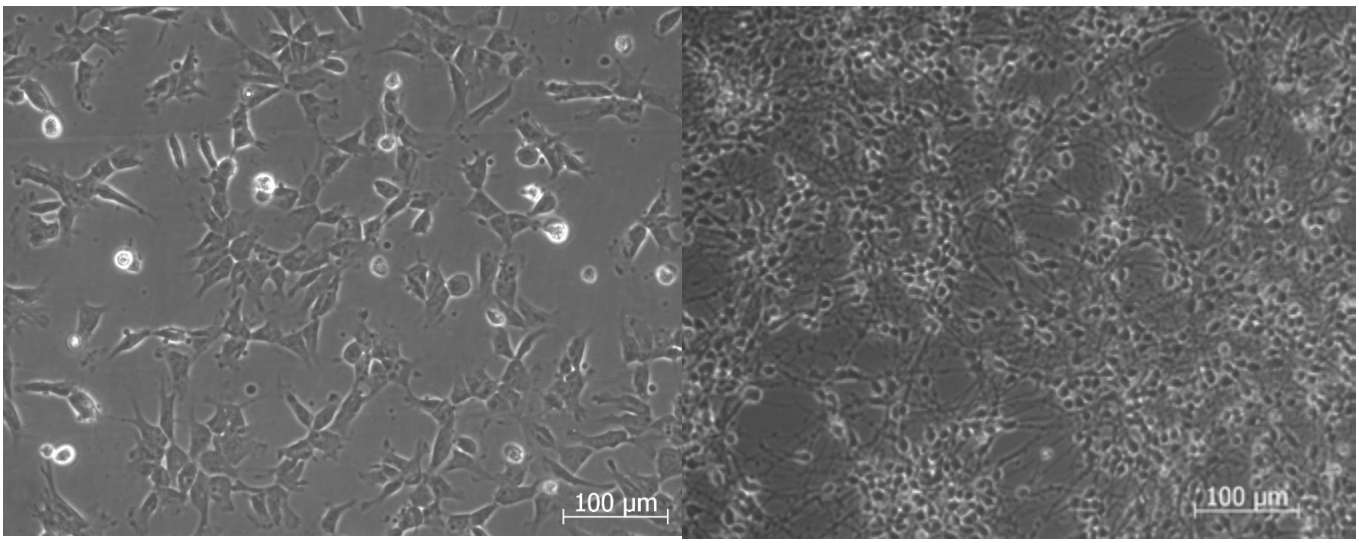


Figure 6. NES cell line a, Proliferating NES (stem cell state). **b**, Differentiated NES, 9 days post differentiation.

5.2 Electrophysiology of differentiated NES cells

Patch-clamp recordings were carried out in a collaboration with Richard Burman from Dr Joseph Raimondo's electrophysiology group (UCT) in pilot study of 10 differentiated, control NES cells (pooled from two experimental sittings) at day 9-post differentiation - the time point at which all HIV-Tat treatments were administered. Electrophysiology data of control NES cells were collected in order to confirm that the cell culture system used in this work had electrical features expected of neuronal tissue. NES cells were targeted for whole-cell recordings based on their morphological properties, namely: a small round or ovoid cell body with diameter between 5-10μm and the presence of two or more extended processes (**Figure 7a**).

Following the attainment of a whole-cell patch, current-clamp was used with the application of current pulses of between 0 and 10 pico amps (pA) (**Figure 7b**). Most cells responded to the injection of current and were capable of generating single spiking activity action potentials (**Figure 7b**), an indication of neuronal maturity (Moody & Bosma 2005). Due to the electrochemical nature of action potential generation, the response of differentiated NES cells to electrical stimulation indicated the presence of voltage-gated ion channels (Hodgkin et al. 1951). Specifically, flow of current through voltage-gated sodium (Na^+) and voltage-gated potassium (K^+) channels were determined in voltage-clamp mode (**Figure 7d** and **c**). To achieve this, NES cells were clamped at 60 milli volts (mV) and then at increasing mV in 20mV steps to elicit a voltage-gated response from both voltage-gated Na^+ and K^+ channels. The maximum current (I) transferred through these channels was then measured with the average maximum INa^+ and IK^+ currents being 249.58 ± 166.91 pA and 181.56 ± 128.18 pA, respectively. These INa^+ and IK^+ values are lower than what has been observed in previous studies (Telias et al. 2014; Song et al. 2013).

Furthermore, average resting membrane voltage (Vm) was determined to be -55.83 ± 17.71mV (**Figure 7e**), which is similar to what has previously been observed in human-induced-pluripotent-stem-cell-derived neurons (Song et al. 2013). The average input resistance (IR) - reflective of ion channel expression - was then determined to be 3853.98 ± 2406.30mΩ (**Figure 7f**), which is slightly higher than observed by previous studies in embryonic-stem-cell-derived neurons (Telias et al. 2014).

The abovementioned findings suggest that the current cell line - in its differentiated state - is broadly similar to the functional neuronal characteristics seen in other stem-cell-derived neuronal lines. It was in this differentiated state that NES cells were treated with exogenous HIV-Tat as an *in vitro* model of an HIV-infected CNS.

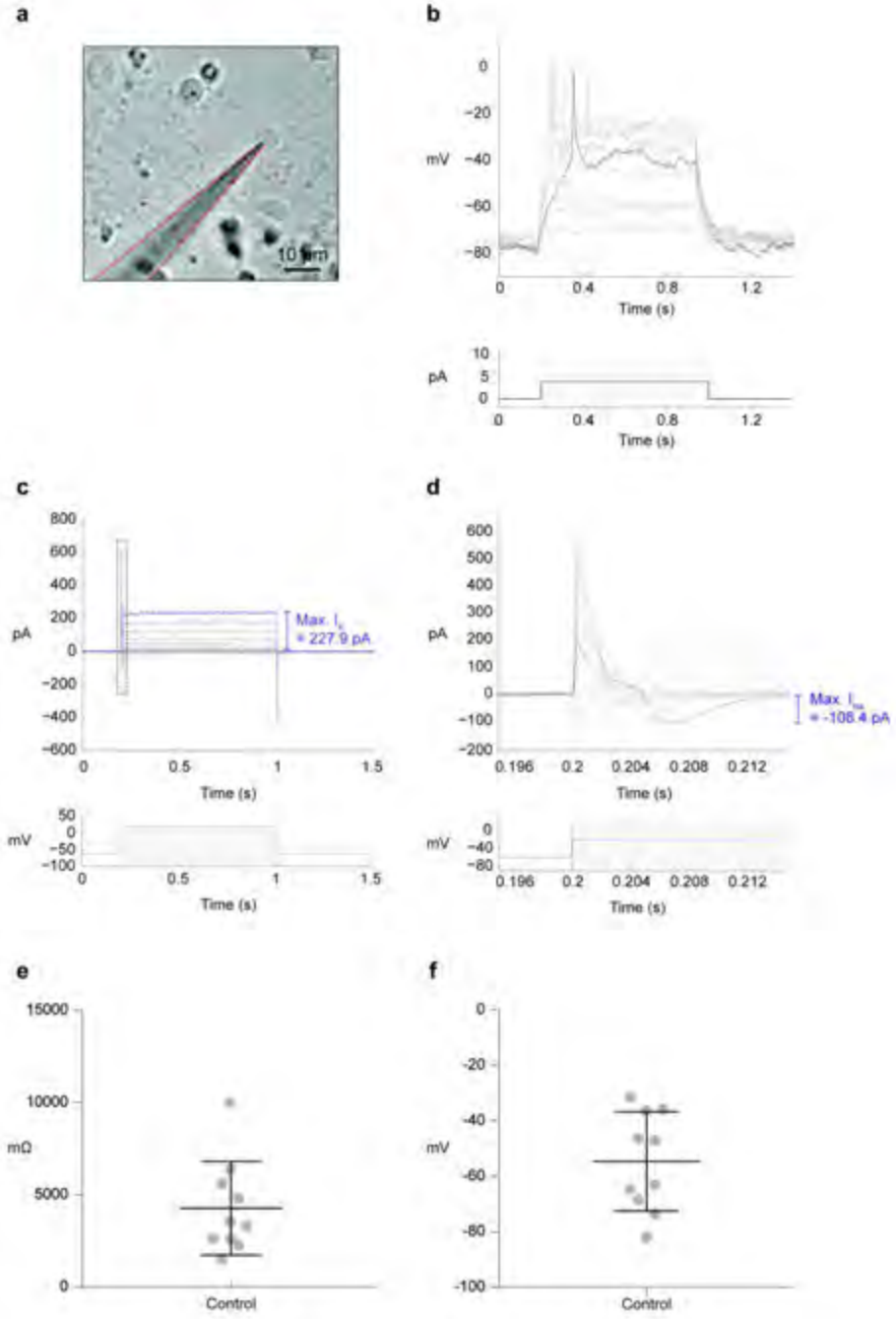


Figure 7. Differentiated NES cells displayed neuronal characteristics at day 9-post differentiation, as indicated by whole-cell patch-clamp a Differentiated NES cells were selected for electrophysiological study based on cell size (5-10µm in diameter) and the presence of two or more extended processes. The red line in this image outlines the micropipette and allows for easy identification of the specific cell recorded in current clamp mode (See **Figure 7b**). **b** Current clamp analysis. Current was injected in increasing steps of 10pA, for 0.8 seconds at a time, resulting in changes to neuronal membrane potential (Vm). Each trace in the upper graph (mV versus time) corresponds to a particular current clamp (pA vs time). A set of corresponding traces from the upper and lower graphs is highlighted to illustrate this coupling. **c and d** Voltage clamp analysis measuring current flowing through voltage-gated potassium (K^+) channels and voltage-gated sodium (Na^+) channels, respectively. Voltage (in mV) was injected in increasing steps of 20mV for 0.8 seconds at a time, resulting in opening of Na^+ channels (**Figure 7d** is an enlargement of the first few milliseconds of the trace in **Figure 7c**) and later opening of K^+ channels (**Figure 7c**), which corresponded to changes in flow of current across the cell membrane (measured in pA). The synchronised opening of these voltage-gated ion channels resulted in an action potential (Hodgkin et al. 1951). **e** Resting membrane potential (Vm) measurements (mV). **f** Input resistance (IR) measurements (mΩ).

5.3 HIV-Tat reconstitution and visualisation

Since HIV-Tat is only active in its monomeric form (Tosi et al. 2000), a reducing agent concentration must be established which both maintains HIV-Tat in a monomeric state and is low enough not to affect cell viability. During previous optimisation of parameters pertaining to HIV-Tat resuspension and buffer exchange, a working concentration of 0.1mM BME was found to meet these criteria (Gurwitz Honours dissertation, 2013). BME at this concentration is in fact often used in cell culture media as a redox scavenger (Danovi et al. 2012).

The current work additionally found that filtered HIV-Tat remained monomeric during long-term storage (≥ 18 months at -80°C) (**Figure 8** lanes 6 and 7) as well as at initial resuspension concentrations of BME as low as 50mM (**Figure 8** lane 3), and can be stored at BME concentrations as low as 10mM (**Figure 8** lane 4). While there was some evidence of filtered HIV-Tat dimerising after 15 months at -80°C (**Figure 8** lane 6), this band was very faint indicating little protein in this conformation. In addition, the current work found that the filtering flow-through contained no protein (Silver Stain lower detection limit - 0.25ng) (**Figure 9**, lane 6 & 7), suggesting that the filtering method - required to achieve a BME treatment concentration of 10mM from a storage concentration of 50mM BME - was robust.

The following observation from SDS-PAGE analysis must be noted: the HIV-Tat isoform used here was an 86 amino acid, 9kDa protein, but it consistently migrated to a gel position corresponding to 12kDa (**Figures 8** and **9)**. Given that the HIV-Tat suspension was pure (as per Diatheva quality control), this observation was interpreted as an effect of SDS-PAGE analysis - for example due to an uncharacteristic binding ratio of SDS (Rath et al. 2009) - as opposed to an effect caused by HIV-Tat itself.

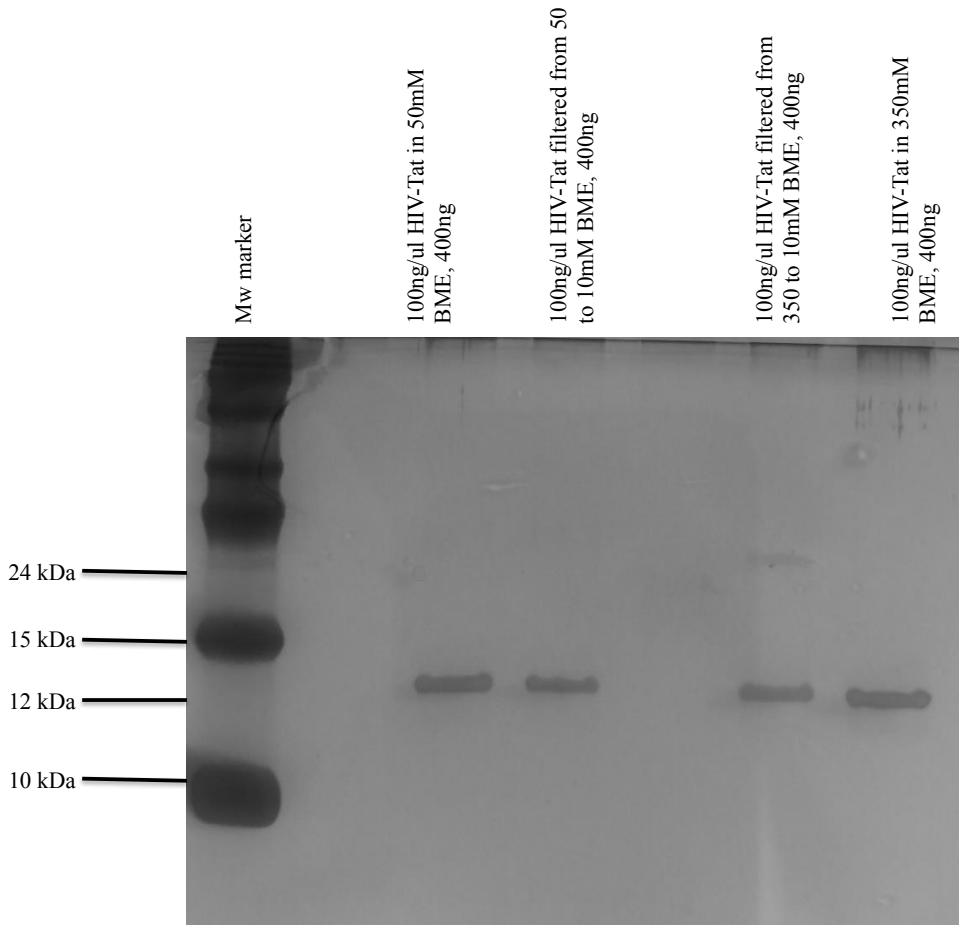


Figure 8. Filtered HIV-Tat was monomeric at a low starting BME concentration (50mM), across filters and over time (lanes from left to right) **Lane 1** molecular weight (Mw) marker. **Lane 2** blank. **Lane 3** HIV-Tat suspended in 50mM BME (100ng/µl, 400ng). **Lane 4** filtered (from 50mM) HIV-Tat in 10mM BME (100ng/µl, 400ng). **Lane 5** blank. **Lane 6** filtered (from 350mM) HIV-Tat in 10mM BME, at -80°C for 15 months (100ng/µl, 400ng). **Lane 7** HIV-Tat suspended in 350mM BME, at -80°C for 18 months (100ng/µl, 400ng). BME - beta mercaptoethanol.

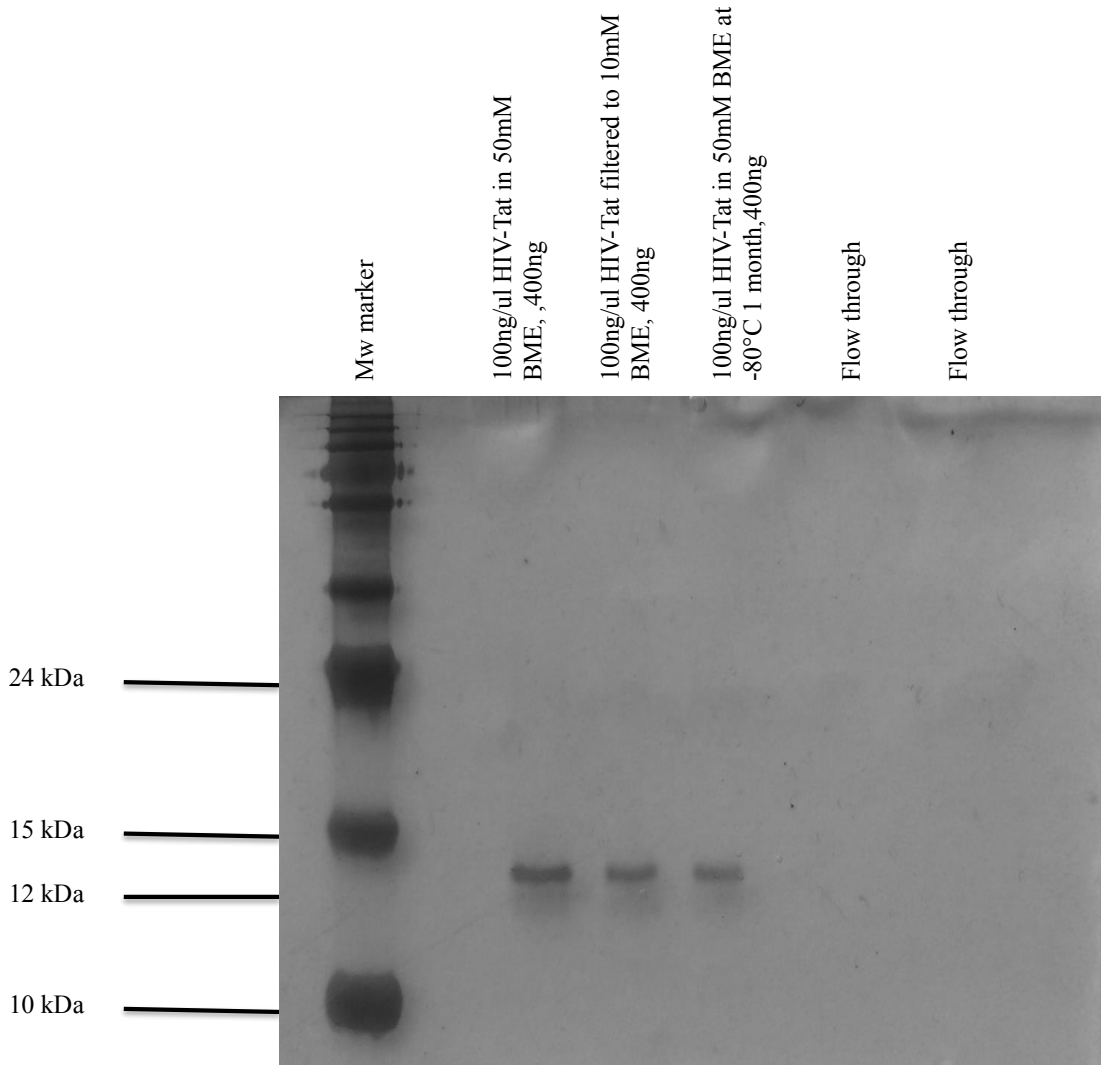


Figure 9. HIV-Tat was undetectable in filtrate flow through (lanes from left to right) **Lane 1** molecular weight (Mw) marker. **Lane 2** blank. **Lane 3** HIV-Tat suspended in 50mM BME (100ng/ µl, 400ng). **Lane 4** filtered (from 50mM) HIV-Tat in 10mM BME (100ng/ µl, 400ng). **Lane 5** filtered (from 50mM) HIV-Tat in 10mM BME, stored at -80°C for 1 month (100ng/ µl, 400ng). **Lane 6** HIV-Tat flow through from filtrate in lane 4. **Lane 7** HIV-Tat flow through from filtrate in lane 5. **Lane 8** blank. BME - beta mercaptoethanol.

5.4 Time course experiment with NES cells differentiated at 90% cell confluence

5.4.1 Experimental layout

Once 90% cell confluence was reached, NES cells were differentiated[1]. At day 9 post-differentiation, cells were treated with 1ng/µl HIV-Tat in 0.1mM BME, and harvested at 6, 24, and 48 hours post-treatment as an *in vitro* model of HIV infection in the CNS over time. This concentration of HIV-Tat has been widely used in cell culture models of HAND and approximates the concentration of HIV-Tat in the CNS of HAND patients (Xiao et al. 2000; Perry et al. 2010). Cells appeared unchanged after treatment and no sign of apoptosis was observed (visually determined, data not shown).

HIV-Tat concentration for treatment was selected based on literature and previous work. Day 9 post differentiation was selected as time point 0, as the proteome of differentiating NES cells stabilizes between 8 and 12 days post differentiation (Garnett, unpublished), and NES cells produce action potentials at this time indicating that they display neuronal tissue characteristics (see above section). Selection of time points for the time course experiments (**Figures 10** and **30**) was based on literature and previous work. For example, in response to HIV-Tat treatment, a large calcium influx has been reported for up to 8 hours after treatment, followed by cessation of this effect (Krogh et al. 2014). Therefore, the 6-hour time point was chosen in an attempt to identify maximal proteomic signatures that may support this finding. The 24-hour time point was chosen to serve as a technical repeat of previous work (Gurwitz Honours dissertation, 2013). The 48-hour time point was selected in order to observe whether the effect of a once-off HIV-Tat dose on gene expression would be sustained over a period of time longer than 24 hours.

The experimental design can be viewed in **Figure 10**. Samples compromised during processing (e.g. culture de-attachment, data not shown) were excluded from analysis, as indicated by dashed out circles in **Figure 10**. Samples were named according to; cell confluence level, harvesting time point, culture condition, and biological replicate number. A sample identifier key can be viewed in **Figure 11**.

[1] Differentiation at 90% cell confluence is specified in order to separate this time course experiment from a second experiment specified later in this work, in which this time course experiment was repeated with cells grown to 70% cell confluence and then differentiated.

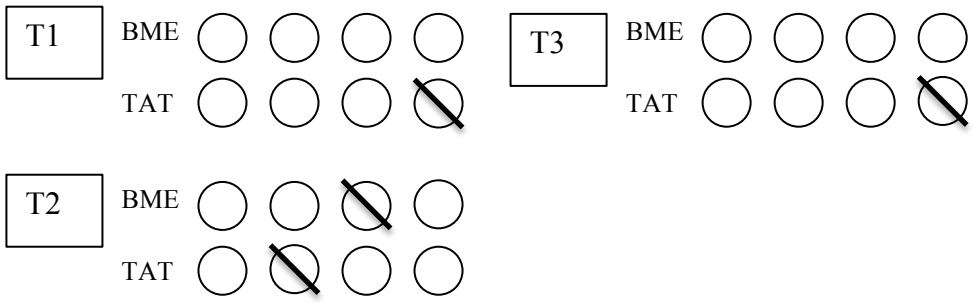


Figure 10. Time course experimental design T1: Time point 1 (harvested at 6 hours post-treatment), T2: Time point 2 (harvested at 24 hours post-treatment), T3: Time point 3 (harvested at 48 hours post-treatment). One circle represents one biological sample (cell culture dish). TAT: HIV-Tat treated samples. BME: control samples. Dashed out circles indicate that the sample was compromised (data not shown) and as the result excluded from final analyses.

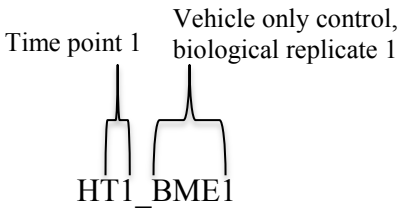


Figure 11. Sample identifier key In this example: sample from time point 1, control sample (BME), and biological replicate 1. The letter 'H' preceding the time point (e.g. HT1) is the dataset identifier for cells differentiated at 90% cell confluence.

5.4.2 Sample preparation for MS analysis

Cell counts were determined from harvested, PBS-washed cell pellets of HIV-Tat treated and control cells. Protein quantitation per sample was then determined by BCA quantitation of lysed cell pellets. Cell count and protein-level BCA quantitation correlated well across samples (**Table 2**; raw BCA quantitation data not shown). In addition, amount of protein recovered per cell was largely in agreement across experimental conditions (average cell count: $2.77 \times 10^6 \pm 0.55 \times 10^6$, average protein yield per cell (in pg): 60.91±17.37). Interestingly, for HT2_BME samples, the protein yield per cell was a bit higher than the rest of the samples across time points, despite similar cell counts across all samples. However, further data quality control measures showed no systematic variation in proteome composition for these samples (see section **5.4.4 Data Analysis**). Protein quantitation facilitated carrying forward the

same amount of protein from each sample (30µg) - as required by label-free MS experiments - for proteome processing by FASP upstream of MS analysis.

Table 2 Cell counts and protein quantitation for time course experiment of cells grown to 90% cell confluence and differentiated

Sample*	**Cell count ** x 10^6**	**Protein Concentration** (µg/µl)	**Protein recovered** (µg)	**Protein yield per cell** (in pg)
HT1_BME1	3.07	2.33	147.48	48.03
HT1_BME2	2.48	2.60	173.32	69.88
HT1_BME3	3.050	2.48	170.62	55.94
HT1_BME4	3.51	2.15	138.48	39.44
HT1_TAT1	2.50	2.55	144.05	57.62
HT1_TAT2	2.39	1.77	137.65	57.59
HT1_TAT3	3.34	2.11	137.87	41.27
HT2_BME1	2.69	2.93	217.99	81.04
HT2_BME2	2.11	3.045	206.15	97.70
HT2_BME4	2.11	2.87	203.70	96.54
HT2_TAT1	2.15	1.56	110.56	51.42
HT2_TAT3	3.58	2.61	185.21	51.73
HT2_TAT4	3.73	2.19	157.76	42.29
HT3_BME1	2.29	2.14	156.49	68.34
HT3_BME2	2.66	2.96	213.71	80.34
HT3_BME3	3.06	2.71	195.30	63.82
HT3_BME4	3.46	2.42	163.49	47.25
HT3_TAT1	2.87	2.72	183.81	64.05
HT3_TAT2	1.90	1.89	119.64	62.97
HT3_TAT3	2.48	1.43	101.32	40.85

* See **Figure 11** sample identifier key; TAT: HIV-Tat treated; BME: control samples.

** Cell counts from 6cm cell culture dishes

5.4.3 LC-MS analysis parameters - optimisation of sample injection volume

While BCA quantitation provided a good protein estimate prior to proteome processing, potential protein loss during processing also had to be accounted for. This was achieved by optimising the volume of sample that was loaded onto the LC-MS instruments, which both facilitated loading of similar amounts of peptide across samples, and prevented peptide over-loading, which may have resulted in HPLC column blockage and/or saturation of the detector. As an initial trial, 1µg peptide (5µl

of a 250ng/µl sample) was loaded onto the reverse phase C18 column and the TIC intensity was noted (**Figure 12c**). A maximum intensity of 5.5e9 at the MS level was the benchmark, as this was high enough so as to increase dynamic range (detection of lower abundant peptide ion peaks), but not so high that it risked blocking the column and/or saturating the detector (Blackburn laboratory, unpublished). If the intensity was significantly below or above 5.5e9 (**Figure 12c**), a new injection volume - aimed at generating the target TIC intensity - was calculated (**Figure 12d**) and the sample was re-injected at this new volume (**Figure 12e**). The process was iterated until the target TIC intensity was achieved.

As all samples in the current work were processed in parallel, it was assumed that any protein loss should be consistent across samples and therefore that the new volume, calculated for the first sample, could be used as the injection volume for all subsequent samples. Selected samples whose TIC's still varied significantly from the desired intensity, were rerun at sample-specific calculated volumes (data not shown). Washes were included after each run (data not shown), as well as a longer wash after every four runs (**Figure 12b**) to ensure no carry over of sample and/or potential contaminants. As can be seen in **Figure 12b** - an example of a wash - the contaminant peak at 51 minutes eluted as a single defined peak, strongly suggesting that the wash removed this contaminant from the column. Further, the column was unchanged by the previous presence of the contaminant, as suggested by the reference sample TIC (**Figure 12a**) and spray stability (**Figure 12**).

The initial liquid chromatography gradient gave good separation of peptides, as judged by the TIC (**Figure 12e**), and so no further optimisation was done.

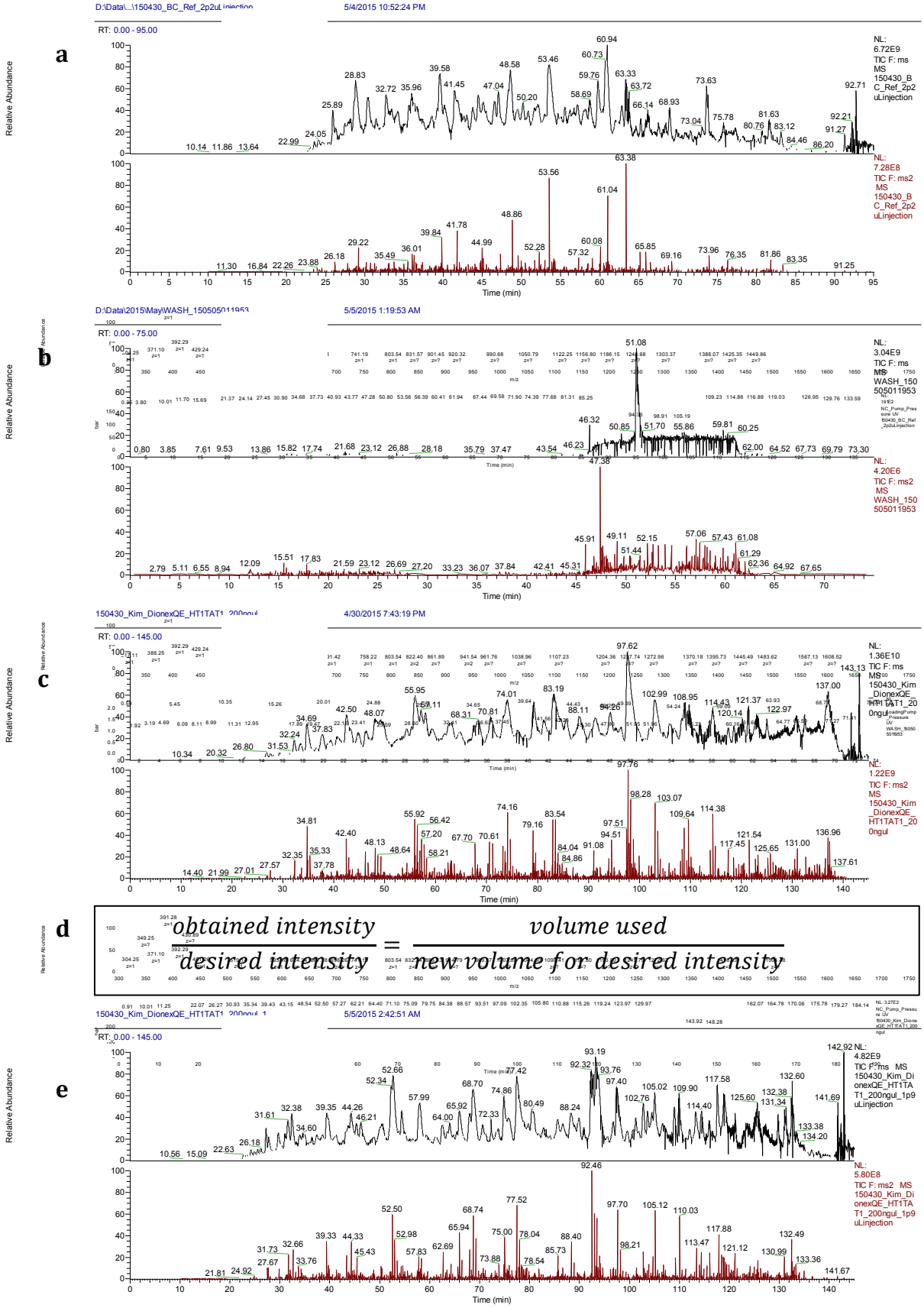


Figure 12. Optimisation of sample injection volume for MS runs a Reference. **b** Wash. **c** Overloading of experimental sample (i.e. too much protein). **d** Equation to calculate new volume of peptide sample to be injected to obtain desired TIC peak intensity (5.5e9 benchmark) **e** New, calculated amount loaded. **Note:** In **a-c** and **e**, trace 1 (black) shows the TIC at the MS1 level; trace 2 (red) shows the TIC at the MS2 level.

5.4.4 Data analysis

MS data quality and raw data processing

MS1 and MS2 data were acquired for all samples on a Q-Exactive MS instrument as described (see section **4.7 LC-MS analysis parameters**). **Figure 4** (in the Introduction) shows an example of TIC, MS1, MS2 and peptide sequence alignment data, generated for the current work by the above approach.

Of the 1 067 211 spectra submitted to MaxQuant software for analysis, 568 656 (53.28%) were mapped to known tryptic peptide sequences in the Uniprot Human protein database corresponding to an identification of 4104 protein groups in total, across the 20 samples analysed (FDR < 0.01), and a minimum of 2314 protein groups identified in each sample. The data were of good quality as indicated by a low percentage of contaminants (0.41% of total protein groups), a low percentage of matches to the decoy database (0.038% of total protein groups), and good tryptic digestion (few missed cleavages, as seen in **Figure 13**). Additionally, average absolute mass error was relatively low and consistent across samples (**Figure 14**).

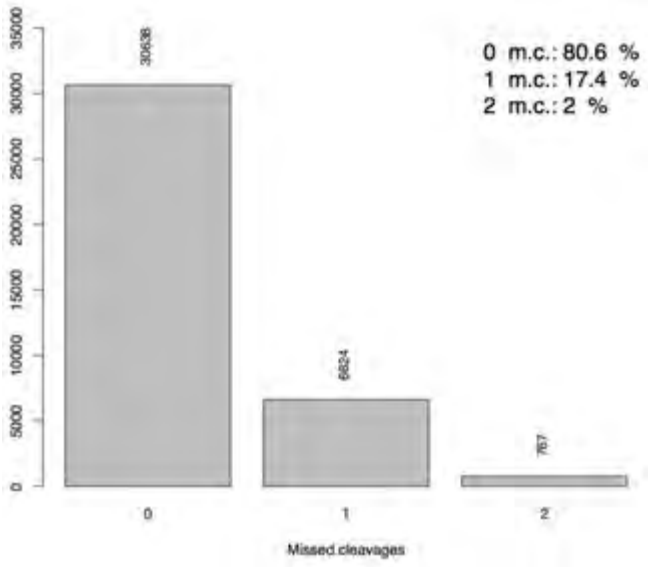


Figure 13. Assessment of tryptic digest completeness m.c. missed cleavages.

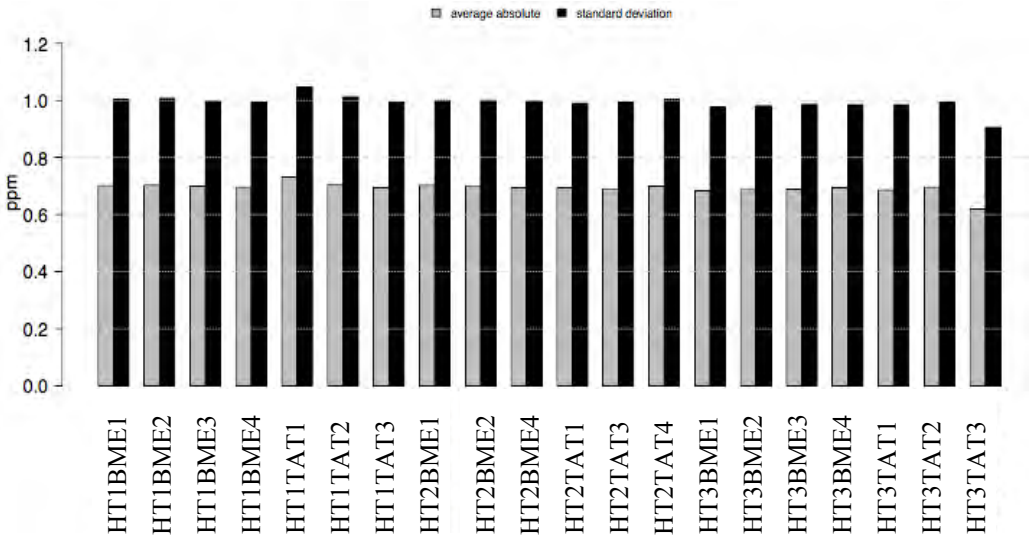


Figure 14. Average absolute mass error of the MS instrument per run ppm – parts per million, TAT - HIV-Tat treated, BME - control.

Descriptive statistical analysis

Technical replicates of proteome-wide quantitation correlated well ($R^2 > 0.95$ for all comparisons; **Figure 15**) and there was no systematic technical variation across experimental conditions or across time points (**Figure 16**). For all time points (**Figure 17a-c**), HIV-Tat and control samples clustered separately, but this clustering was not very strong. Subtle changes to smaller protein sub-groups (representing significantly differentially expressed protein groups) were revealed by hypothesis testing (see next section: *Differential expression analysis*). Additionally, there was evidence of weak clustering by time point when all samples across time points were analysed together (**Figure 17d**). However, as each treated group was only compared to the control group from that time point (i.e. HIV-Tat treated from time point 1 versus control from time point 1 etc.), these effects should not bias downstream analyses.

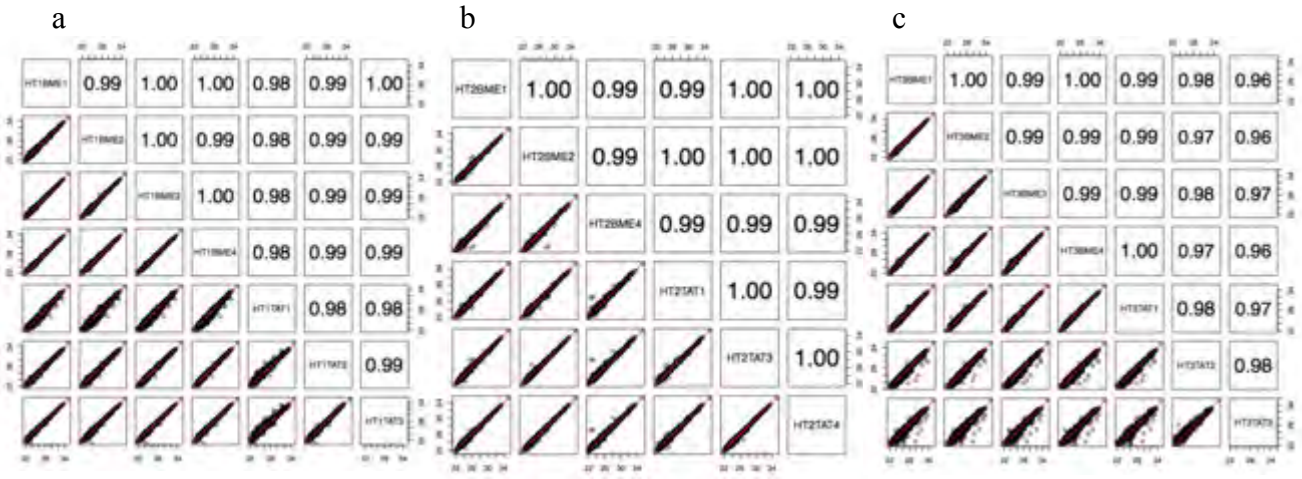


Figure 15. Agreement between technical replicates and across experimental conditions for all three time points log_2LFQ protein expression values from each experimental sample were plotted against those from every other experimental sample, per time point (log_2 of values used for visualisation purposes). **a** HT1: 6 hours post-treatment. **b** HT2: 24 hours post-treatment. **c** HT3: 48 hours post-treatment. Numeric values in the top right half of each graph indicate the R^2 value for each comparison. TAT - HIV-Tat treated, BME - control.

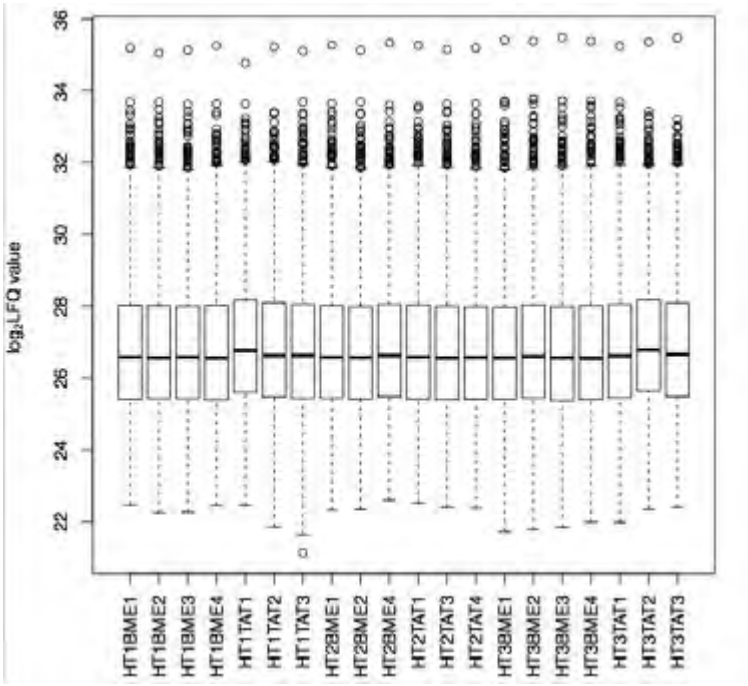


Figure 16. Boxplot representations of log_2LFQ intensity values per replicate per condition per time point. TAT - HIV-Tat treated, BME - control. Log_2 values used for visualisation purposes.

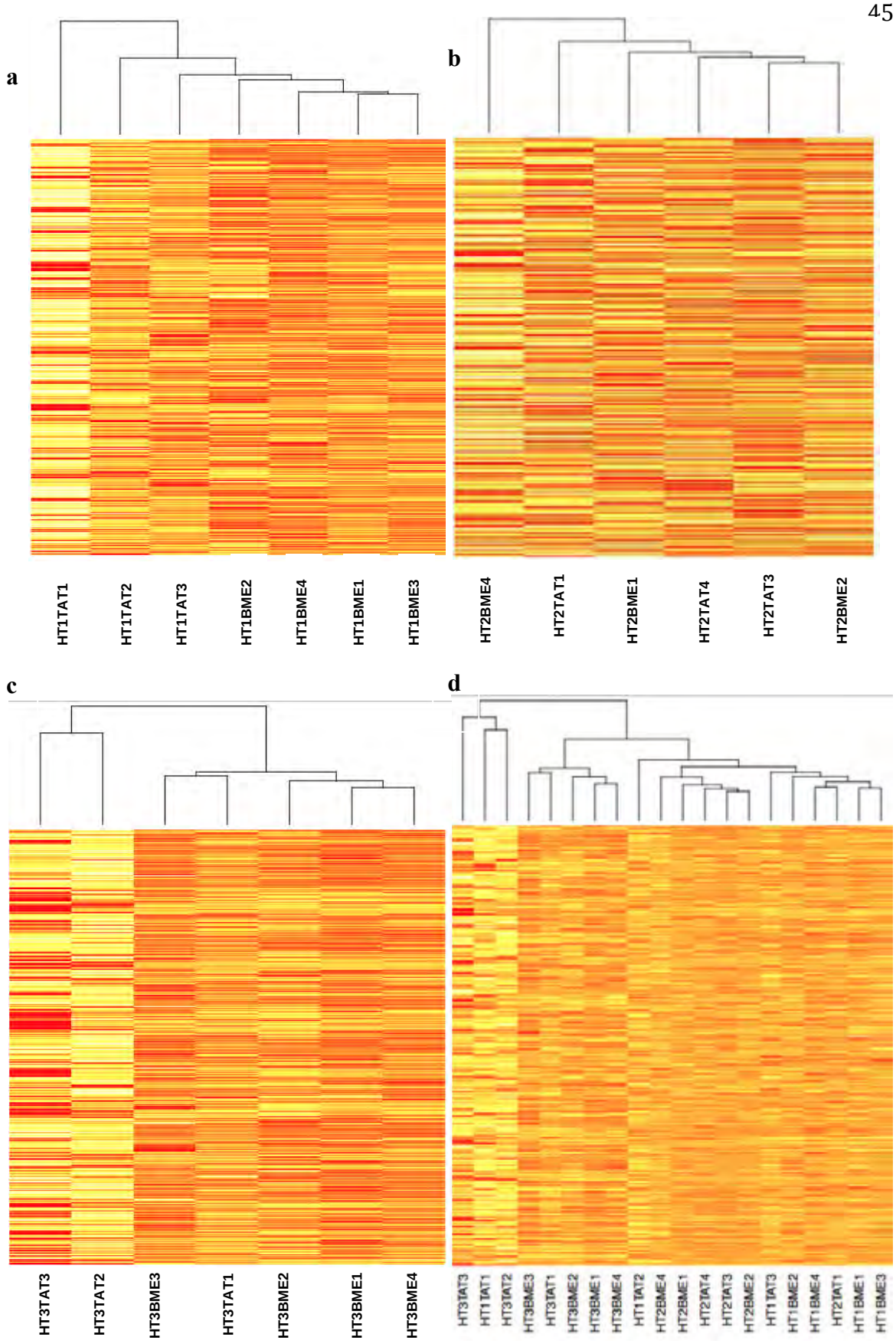


Figure 17. Hierarchical cluster analysis of total proteomic profile (using log_2LFQ values) revealed weak clustering of experimental conditions a HT1: 6 hours post-treatment. **b** HT2: 24 hours post-treatment. **c** HT3: 48 hours post treatment. **d** all time points included. Note: the darker the band, the greater the expression of that particular protein group. TAT - HIV-Tat treated, BME - control. Log_2LFQ values used for visualisation purposes.

Differential expression analysis

Protein groups expressed differentially between HIV-Tat and control samples per time point - as determined by hypothesis testing - were represented as a volcano plot (**Figure 18**). Upon visual inspection of the volcano plots, it was apparent that the magnitude of differential expression was subtle (**Figure 18**, $-1 < \log_2FC < 1$). While only subtle differences in $\log_2$ fold change were expected, as most of the proteome does stay the same in any experiment-control comparison, differences were statistically significant for many protein groups as judged by t test analysis (**Figure 18**, $p < 0.05$; **Table 3**). The top five most up-regulated and top 5 most down-regulated protein groups, per time point, can be viewed in **Tables 4-6**.

Differentially expressed protein groups may be further analysed to determine the temporal changes in protein expression - as well as the significance of such changes - in response to HIV-Tat treatment, since they relate to HIV-Tat's effect on neurons in cell culture in our *in vitro* model of HIV infection of the CNS (see next section: *Biological significance analysis*).

a HT1 (6 hours) **b** HT2 (24 hours) **c** HT3 (48 hours)

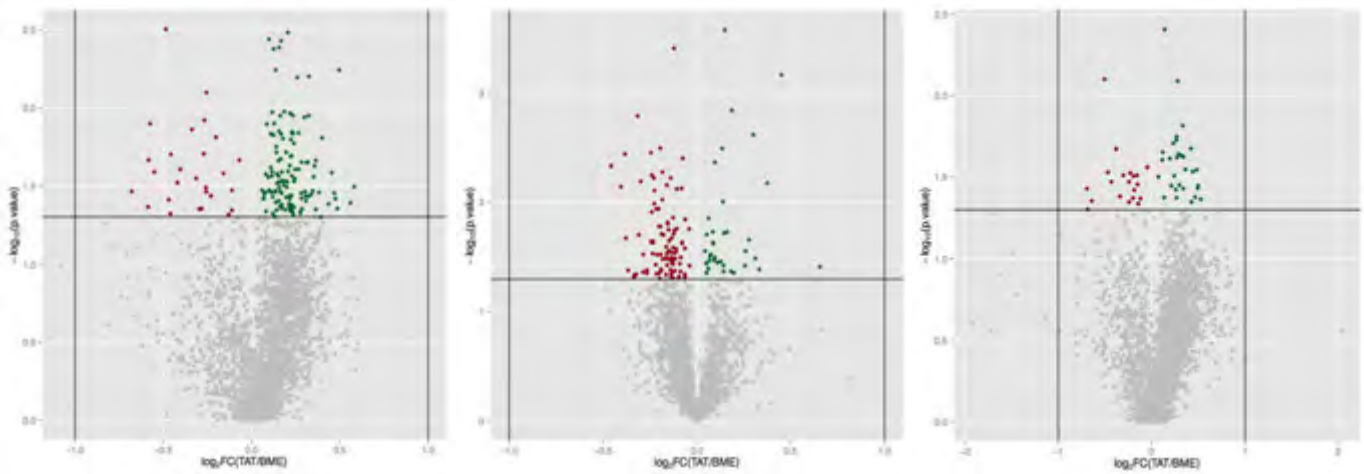


Figure 18. Visual representation of expression changes between HIV-Tat and BME (vehicle-only) treated samples protein groups plotted as a function of significance ($-\log_{10}$ p value) and $\log_2$ fold change ($\log_2FC$). **a** HT1: 6 hours post treatment. **b** HT2: 24 hours post treatment. **c** HT3: 48 hours post treatment. Fold change = TAT/BME. Horizontal line indicates p = 0.05 (level of significance set for the independent t-test analysis per protein group). Vertical lines indicate two fold change in expression ($\pm\log_2 1$). Up-regulated protein groups are coloured green, down-regulated protein groups are coloured red. TAT- HIV-Tat treated. BME - control.

Table 3 Number of genes (gene acting as proxy for 'protein groups') identified at each time point and the subset that were significantly differentially expressed

Time point	Gene IDs: total (full list)	Gene IDs: differentially expressed ($p < 0.05$)
HT1	2420	131
HT2	2669	118
HT3	2314	45

Table 4 Top 5 most up-regulated and top 5 most down-regulated protein groups at **6 hours** post treatment

gene.ID	gene.name	PEP	p.value	log_2FC**
EDF1	Endothelial differentiation-related factor 1	5.56e-30	0.035	0.47
SMRC2	Isoform 2 of SWI/SNF complex subunit SMARCC2	0	0.038	0.47
SNX27	Isoform 3 of Sorting nexin-27	9.84e-18	0.0057	0.5
KHDR2	KH domain-containing, RNA-binding, signal transduction-associated protein 2	2.53e-16	0.041	0.6
CO3A1	Collagen alpha-1(III)	1.36e-33	0.032	0.6
HNRPK	Isoform 2 of Heterogeneous nuclear ribonucleoprotein K	0	0.026	-0.5
PIGU	Isoform 2 of Phosphatidylinositol glycan anchor biosynthesis class U protein	8.91e-23	0.013	-0.57
FYN	Tyrosine-protein kinase	3.06e-64	0.022	-0.58
PTCD3	Pentatricopeptide repeat domain-containing protein 3, mitochondrial	7.97e-46	0.043	-0.58
E41L5	Isoform 3 of Band 4.1-like protein	1.84e-21	0.034	-0.68

***FC- fold change, HIV-Tat/control*

Table 5 Top 5 most up-regulated and top 5 most down-regulated protein groups at **24 hours** post treatment

gene.ID	gene.name	PEP	p.value	log_2FC**
CWC15	Spliceosome-associated protein	1.11e-05	0.032	0.31
LZIC	Protein LZIC	4.34e-25	0.041	0.33
CA123	protein C1orf123	7.64e-41	0.0067	0.38
ERI3	Isoform 3 of ERI1 exoribonuclease 3	2.85e-15	0.00069	0.45
G3V2L5	Ubiquinone biosynthesis monooxygenase	3.06e-06	0.039	0.66
GEMI5	Gem-associated protein 5	3.96e-23	0.042	-0.37
UN13B	Protein unc-13 homolog B	1.26e-15	0.021	-0.38
IGSF3	Immunoglobulin superfamily member 3	1.82e-34	0.0036	-0.38
NUCKS	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	6.80e-24	0.0073	-0.41
FMN2	Formin-2	4.71e-18	0.0047	-0.46

***FC - fold change, HIV-Tat/control*

Table 6 Top 5 most up-regulated and top 5 most down-regulated protein groups at **48 hours** post treatment

gene.ID	gene.name	PEP	p.value	log_2FC**
SNX27	Isoform 3 of Sorting nexin-27	9.84e-18	0.042	0.47
BRE1B	E3 ubiquitin-protein ligase	2.43e-43	0.029	0.49
PHF5A	PHD finger-like domain-containing protein 5A	4.22e-45	0.037	0.49
M0QXF7	protein C19orf10 (Fragment)	2.67e-15	0.036	0.50
RUXF	Small nuclear ribonucleoprotein F	2.18e-80	0.043	0.53
GALE	UDP-glucose 4-epimerase	3.48e-24	0.021	-0.38
KAP3	cAMP-dependent protein kinase type II-beta regulatory subunit	2.82e-123	0.034	-0.43
S4R456	40S ribosomal protein S15 (Fragment)	5.51e-61	0.030	-0.47
STK39	Isoform 2 of STE20/SPS1-related proline-alanine-rich protein kinase	8.86e-115	0.0079	-0.51
HCDH	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	3.46e-101	0.037	-0.69

***FC - fold change, HIV-Tat/control*

5.4.5 Biological significance analysis

Functional classification of protein groups significantly differentially expressed between time points - in accordance with Panther protein class (Mi et al. 2013) - revealed those biological functions affected by HIV-Tat treatment over time[2]. Trends observed for the *absolute* number of protein groups that mapped to a particular protein class at each time point (**Figure 19a**), were in agreement with trends for mapped protein groups normalised to the full list of protein groups identified per time point (**Figure 19b**). These trends were largely maintained when protein groups, per protein class, were normalised to the number of significantly differentially expressed protein groups per time point, except for some protein classes at time point 3 (**Figure 19a-c**). This is perhaps the result of progressively fewer differentially expressed protein groups identified over time (**Table 3**, **Figure 21b**).

For all figures, irrespective of normalisation method described in the above paragraph, the following functional classes were represented by the highest number of protein groups: *nucleic acid binding* (PC00171), *hydrolase* (PC00121), *transferase* (PC00220), *cytoskeletal protein* (PC00085), and *enzyme modulator* (PC00095) (**Figure 19a-c**). Perhaps most informative were *nucleic acid binding* (**Figure 20a**) and *cytoskeletal protein* (**Figure 20b**) classifications, as these in fact *encompassed* many members of the *hydrolase*, *transferase* and *enzyme modulator* classifications (data not shown). Further, more specific protein classes that are functionally linked to these two broader protein classes (i.e. in the same Panther protein class lineage) were significantly over/under represented ($p \leq 0.05$) when compared to the full list (i.e. background) using Panther's over/under representation of protein classes tool (**Tables 7-9**). Trends pertaining to *nucleic acid binding* and *cytoskeletal protein* Panther protein classes were therefore explored in more detail, as described below.

Panther protein class: nucleic acid binding

When *nucleic acid binding* was considered, it was apparent that there were proportionally fewer differentially expressed protein groups that mapped to this

[2] **Please note:** when interpreting functional classification graphs (e.g. **Figures 19** and **20**), it is important observed trends over time are the effects of treatment. For example, if the trend declined, it did not go back to baseline/'normal', but rather that relatively fewer protein groups were implicated for that particular protein class as the result of treatment, in comparison to the control.

classification at 24 hours post treatment, than at 6 and 48 hours post treatment (**Figure 19c**). This trend was mirrored when the proportion of protein groups that map to *RNA binding* and *DNA binding* (subcategories of *nucleic acid binding*) were considered (**Figure 20a**). In addition, most protein groups in the *nucleic acid binding* list were down-regulated at 24 hours after HIV-Tat treatment (i.e. negative $\log_2$ fold change), while the few that were up-regulated in response to treatment are involved in mRNA turnover and RNA degradation (**Table B2**, Appendix B). Taken together, these trends suggest general repression of gene expression by HIV-Tat at 24 hours post treatment.

This was in contrast to what was seen at 6 hours post treatment, in which differentially expressed protein groups that mapped to this protein class were mostly up-regulated (i.e. positive $\log_2$ fold change) in response to HIV-Tat treatment (**Table B1**, Appendix B). These up-regulated protein groups are involved in chromatin remodelling (e.g. SMRC2) and DNA replication (e.g. MCM5 and RFA1) at the DNA level, as well as RNA splicing (e.g. PRP4), mRNA processing (e.g. GMPPA), and translational machinery (e.g. RL10, RL21, RL5, RS2) at the RNA level. This suggests an up-regulation of general host gene expression, by HIV-Tat, at 6 hours post treatment. Most *nucleic acid binding* protein groups were also up regulated at 48 hours post treatment (**Table B3**, Appendix B), and, according to their biological functions, again suggest a general increase in host gene expression machinery, and thus host gene expression, at this time.

Panther protein class: cytoskeletal protein

When *cytoskeletal protein* was considered, *more* protein groups mapped to this classification at 24 hours post treatment than at 6 or 48 hours post treatment (**Figure 19c**), thus displaying the opposite overall trend to *nucleic acid binding* (**Figure 19c**). Although this 'inverted U' trend was observed when all *cytoskeletal protein* groups were considered together, upon inspection of protein groups belonging to *cytoskeletal protein* subclasses, it was apparent that the proportion of *actin* related proteins actually increased over time (**Figure 20b**). As the increase in proportion was greater from 6 to 24 hours than from 24 to 48 hours (**Figure 20b**), one may tentatively hypothesise that this effect plateaus after 48 hours.

Rather than representing cytoskeletal building block components (e.g. actin, microtubules and intermediary filaments), most protein groups that mapped to this protein class are cytoskeletal maintenance/cytoskeleton anchoring proteins (e.g. CTNA1 at 6 and 48 hours, which links the cytoplasmic domain of membrane cadherins to actin at adherens junctions) (GeneCards, www.genecards.org; **Tables B4** and **B6**, Appendix B) or cytoskeletal transport proteins (e.g. KLC2 and ARP10 at 24 hours) (**Table B5**, Appendix B).

In addition, when $\log_2$ fold change in expression was considered, these protein groups were mostly down-regulated across all three time points (**Tables B4-6**, Appendix B), while the few that were up-regulated are involved in cytoskeletal disassembly (e.g. DEST - at 6 hours post treatment - which is an actin depolymerising protein) (GeneCards; **Table B4**, Appendix B). This suggests that HIV-Tat treatment causes instability to cytoskeletal maintenance (as opposed to cytoskeletal dynamics/responsiveness) up to 48 hours after initial treatment, with down-regulation of progressively more protein groups over time.

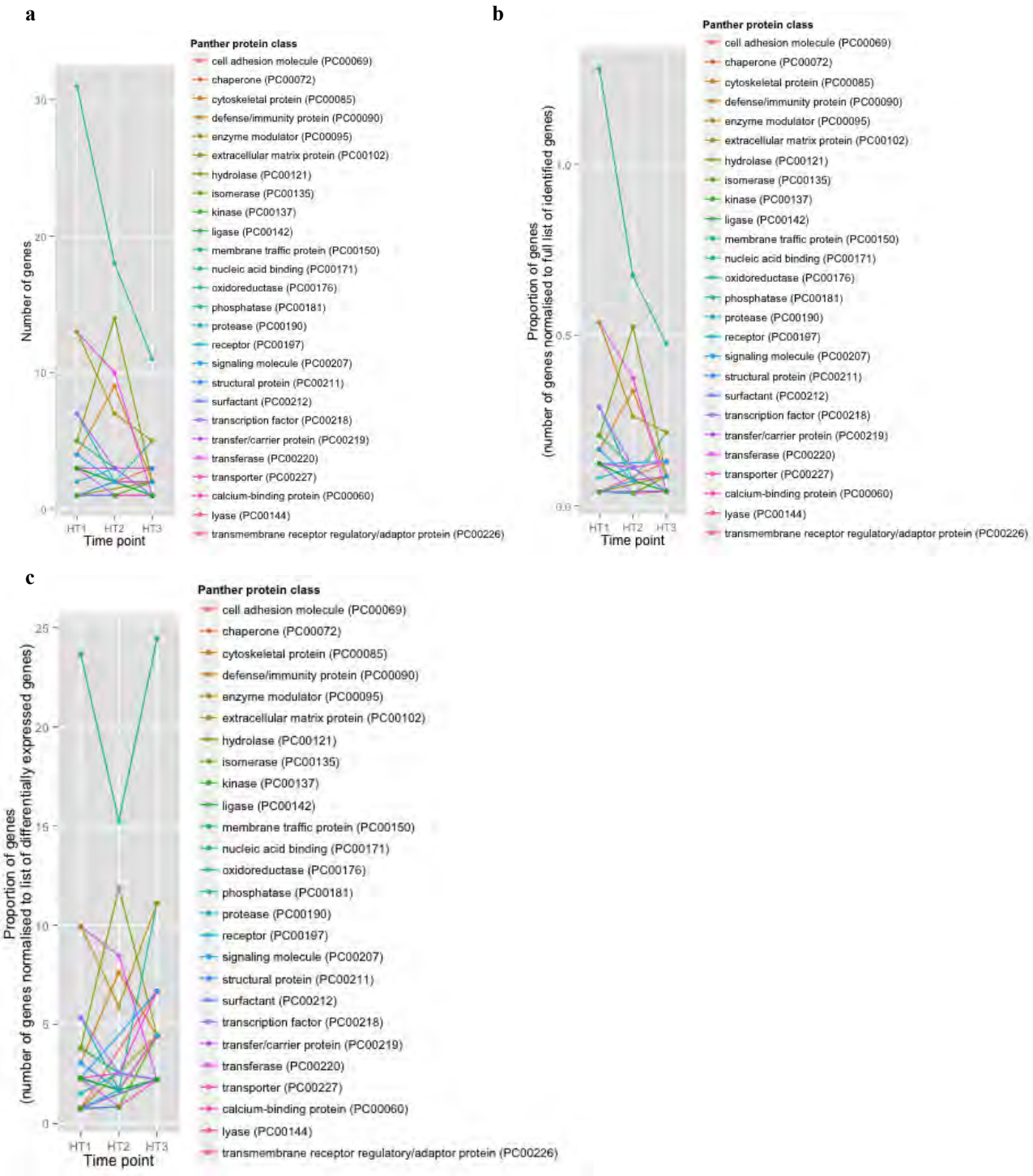


Figure 19. Functional classifications of differentially expressed protein groups over time, according to Panther protein class, for the higher cell confluence experiment a absolute number of genes per time point. **b** number of genes normalised to the full list of identified genes per time point. **c** number of genes normalised to the number of differentially expressed genes per time point. Gene names were used as a proxy for protein groups. HT1: 6 hours post treatment, HT2: 24 hours post treatment, HT3: 48 hours post treatment.

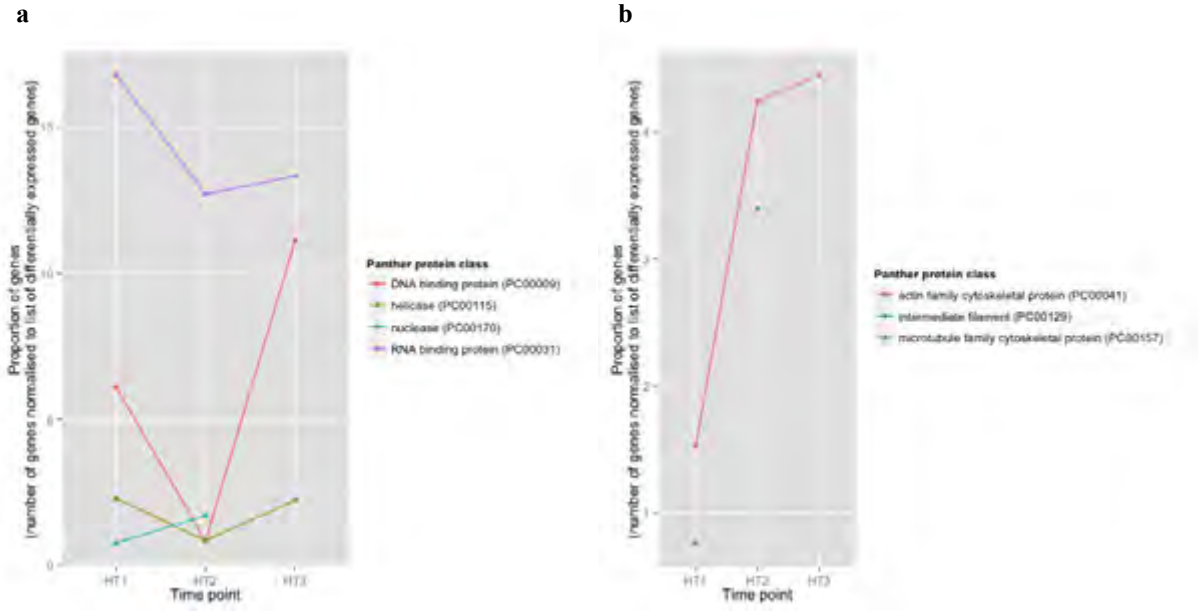


Figure 20. Sub categories per time point of Panther protein classes a *nucleic acid binding* **b** *cytoskeletal protein*. Values were normalised to the number of differentially expressed genes per time point. Gene names were used as a proxy for protein groups. HT1: 6 hours post treatment**,** HT2: 24 hours post treatment**,** HT3: 48 hours post treatment.

Table 7 Top 10** output from over/under representation analysis (i.e. Fisher exact test results) of Panther protein class terms at **6 hours** post treatment

PANTHER Protein Class	Background (out of 2104 protein groups)*	Differentially expressed protein groups - input (out of 115 protein groups)*	Differentially expressed protein groups - expected	over/under representation	fold enrichment	p.value
RNA methyltransferase (PC00033)	2	2	.11	+	> 5	5.52E-03
transcription cofactor (PC00217)	24	5	1.31	+	3.81	1.05E-02
hydrolase (PC00121)	208	5	11.37	-	.44	2.44E-02
microtubule family cytoskeletal protein (PC00157)	54	0	2.95	-	< 0.2	5.03E-02
cytoskeletal protein (PC00085)	163	4	8.91	-	.45	5.14E-02
calcium-binding protein (PC00060)	53	0	2.90	-	< 0.2	5.32E-02
RNA binding protein (PC00031)	290	22	15.85	+	1.39	6.78E-02
chromatin/chromatin-binding protein (PC00077)	42	5	2.30	+	2.18	8.12E-02
small GTPase (PC00208)	30	4	1.64	+	2.44	8.30E-02
nucleic acid binding (PC00171)	447	31	24.43	+	1.27	8.59E-02

** The number of protein groups belonging to background (i.e. full list) and differentially expressed protein group may not correlate with the values in **Table 3** as the Panther algorithm could not map all protein groups from the input lists*

*** **please note** that not all Panther protein classes listed are significantly over/under represented (check p.value column). However, many classes functionally linked to 'nucleic acid binding' and 'cytoskeletal protein' are significantly over/under represented.*

Table 8 Top 10** output from over/under representation analysis (i.e. Fisher exact test results) of Panther protein class terms at **24 hours** post treatment

PANTHER Protein Class	Background (out of 2293 protein groups)*	Differentially expressed protein groups - input (out of 97 protein groups)*	Differentially expressed protein groups - expected	over/under representation	fold enrichment	p.value
DNA binding protein (PC00009)	129	1	5.46	-	< 0.2	2.47E-02
esterase (PC00097)	19	3	.80	+	3.73	4.72E-02
oxidoreductase (PC00176)	147	2	6.22	-	.32	4.77E-02
exoribonuclease (PC00099)	10	2	.42	+	4.73	6.75E-02
reductase (PC00198)	58	0	2.45	-	< 0.2	8.33E-02
signaling molecule (PC00207)	53	0	2.24	-	< 0.2	1.03E-01
hydrolase (PC00121)	234	14	9.90	+	1.41	1.16E-01
tyrosine protein kinase receptor (PC00233)	3	1	.13	+	> 5	1.19E-01
phosphorylase (PC00187)	3	1	.13	+	> 5	1.19E-01
Unclassified (UNCLASSIFIED)	735	37	31.09	+	1.19	1.21E-01

** The number of protein groups belonging to background (i.e. full list) and differentially expressed protein group may not correlate with the values in **Table 3** as the Panther algorithm could not map all protein groups from the input lists*

*** **please note** that not all Panther protein classes listed are significantly over/under represented (check p.value column). However, many classes functionally linked to 'nucleic acid binding' and 'cytoskeletal protein' are significantly over/under represented.*

Table 9 Top 10** output from over/under representation analysis (i.e. Fisher exact test results) of Panther protein class terms at **48 hours** post treatment

PANTHER Protein Class	Background (out of 2022 protein groups)*	Differentially expressed protein groups - input (out of 39 protein groups)*	Differentially expressed protein groups - expected	over/under representation	fold enrichment	p.value
cell adhesion molecule (PC00069)	19	3	.37	+	> 5	5.89E-03
extracellular matrix protein (PC00102)	12	2	.23	+	> 5	2.26E-02
extracellular matrix structural protein (PC00103)	2	1	.04	+	> 5	3.79E-02
peptide hormone (PC00179)	2	1	.04	+	> 5	3.79E-02
epimerase/racemase (PC00096)	16	2	.31	+	> 5	3.82E-02
antibacterial response protein (PC00051)	3	1	.06	+	> 5	5.63E-02
surfactant (PC00212)	3	1	.06	+	> 5	5.63E-02
DNA binding protein (PC00009)	111	5	2.14	+	2.34	6.08E-02
signaling molecule (PC00207)	49	3	.95	+	3.17	6.82E-02
ligand-gated ion channel (PC00141)	4	1	.08	+	> 5	7.43E-02

** The number of protein groups belonging to background (i.e. full list) and differentially expressed protein group may not correlate with the values in **Table 3** as the Panther algorithm could not map all protein groups from the input lists*

*** **please note** that not all Panther protein classes listed are significantly over/under represented (check p.value column). However, many classes functionally linked to 'nucleic acid binding' and 'cytoskeletal protein' are significantly over/under represented.*

Protein groups differentially expressed at more than one time point

While there was little overlap in the protein groups expressed at different time points post treatment (**Figure 21**), the few that did overlap may highlight key points/pathways in the progression of HIV-Tat's effect (**Tables 10-12**). Particularly, those that were up-regulated at one time point, then down-regulated at a second time point - and that correlated with implicated broader biological function classifications - were of interest. For example, MRPL45 (involved in mitochondrial protein translation) (GeneCards) and DDX18 (an RNA helicase) (GeneCards) were both up-regulated at 6 hours post HIV-Tat treatment and subsequently down-regulated at 24 hours post treatment (**Table 10**). These protein groups both mapped to the protein class *nucleic acid binding*.

Also of interest are those protein groups differentially expressed at both 6 and 48 hours, but not 24 hours, post treatment (e.g. CO3A1, CTNA1, SNX27) (**Table 12**). These protein groups had the same 'direction' of expression change (log_2 fold change) at both time points, and were therefore affected in same way at each time point (i.e. both up- or both down- regulated).

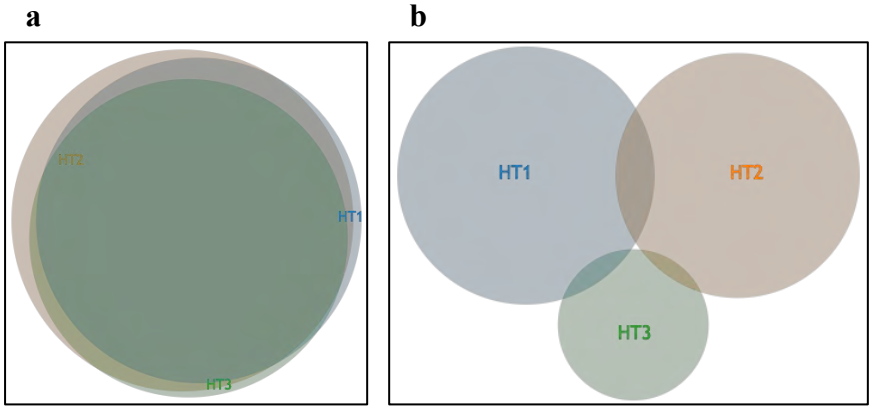


Figure 21. Venn diagrams showing overlap of protein groups between different time points a Overlap of full lists (all protein groups identified) **b** Overlap of lists of differentially expressed protein groups that meet the cut-off criteria for significance after hypothesis testing analysis, $p < 0.05$. HT1: 6 hours post treatment; HT2: 24 hours post treatment; HT3: 48 hours post treatment. Venn diagrams generated using *Venny* online tool: www.stefanjol.nl/venny.

Table 10 Differentially expressed genes at **6 (HT1) and 24 (HT2) hours** post HIV-Tat treatment

geneID	gene.name	Function*	log_2FC**	p.value	PEP
MRPL45	39S ribosomal protein L45, mitochondrial	Mitochondrial protein translation	HT1: 0.23	HT1: 0.027	4.7e-07
			HT2: -0.2	HT2: 0.0032	
ACAD9	Acyl-CoA dehydrogenase family member 9, mitochondrial	Catalyses rate limiting step in β oxidation	HT1: 0.087	HT1: 0.013	2.9e-52
			HT2: -0.16	HT2: 0.025	
CSN3	COP9 signalosome complex subunit 3	Protein degradation	HT1: 0.15	HT1: 0.039	5.6e-18
			HT2: 0.75	HT2: 0.029	
DDX18	ATP-dependent RNA helicase	RNA helicase	HT1: 0.22	HT1: 0.43	8.1e-51
			HT2: -0.28	HT2: 0.029	
GSTA4	Glutathione S-transferase A4	Detoxification of lipid peroxidation products	HT1: 0.16	HT1: 0.011	3.0e-36
			HT2: 0.14	HT2: 0.037	
IGSF3	Immunoglobulin superfamily member 3	Cell surface receptor, immune system	HT1: -0.27	HT1: 0.021	1.8e-34
			HT2: -0.38	HT2: 0.0036	
J3QRU4	Vesicle-associated membrane protein 2	Docking of vesicles with membrane at presynaptic vesicle for neurotransmitter release	HT1: -0.026	HT1: 0.0080	9.9e-14
			HT2: -0.27	HT2: 0.043	
PABP1	Polyadenylate-binding protein 1	Translation initiation	HT1: 0.098	HT1: 0.0036	0
			HT2: 0.052	HT2: 0.020	
VPS16	Vacuolar protein sorting-associated protein 16	Mediation of vesicle trafficking in endosome/ lysosome pathway (by homology)	HT1: 0.11	HT1: 0.024	2.5e-18
			HT2: -0.24	HT2: 0.029	

** Functions from GeneCards www.genecards.org*
***FC- fold change, HIV-Tat/control*

Table 11 Differentially expressed protein groups at both **24 (HT2) and 48 (HT3) hours** post treatment

geneID	gene.name	Function*	log_2FC**	p.value	PEP
BRE1B	E3 ubiquitin-protein ligase	ubiquitination and degradation of syntaxin 1 - essential component of the neurotransmitter release machinery (by homology)	HT2: -0.16	HT2: 0.015	2.43e-43
			HT3: 0.46	HT3: 0.028	
CHTOP	Isoform 3 of Chromatin target of PRMT1	Required for effective mRNA nuclear export	HT2: -0.074	HT2: 0.0039	1.24e-91
			HT3: 0.27	HT3: 0.023	

** Functions from GeneCards www.genecards.org*
***FC- fold change, HIV-Tat/control*

Table 12 Differentially expressed genes at both **6 (HT1) and 48 (HT3) hours** post treatment

geneID	Gene.name	Function*	p.value	Log_2FC**	PEP
CO3A1	Collagen alpha-1(III) chain	Actin component	HT1: 0.032	HT1: 0.58	1.3562e-33
			HT3: 0.029	HT3: 0.44	
CTNA1	Catenin alpha-1	Actin component	HT1: 0.037	HT1: -0.23	0
			HT3: 0.030	HT3: -0.15	
SNX27	Isoform 3 of Sorting nexin-27	Endocytosis of plasma membrane receptors	HT1: 0.0057	HT1: 0.49	9.8375e-18

** Function from GeneCards www.genecards.org*
***FC- fold change, HIV-Tat/control*

5.5 Repeat of 24-hour time point

Differentiated NES cells were treated with HIV-Tat and then harvested 24 hours after treatment in a repeat experiment of time point 2 (HT2) of the time course experiment. While the cell culture and sample preparation work for this repeat was carried out previously and stored at -80°C (Gurwitz Honours dissertation, 2013), MS data were acquired with the most up to date methods in the current work (see section **4.7 LC-MS analysis parameters**) and subsequently reprocessed *in silico*. All experimental procedures outlined in **section 4** are applicable to this dataset. Outlined below are the experimental design, data quality parameters, descriptive statistics and differential expression analyses for this dataset. Further, biological significance data - as well as how this dataset relates to the higher confluence time course experiment - are also included.

5.5.1 Experimental design

The experimental design for this repeat experiment closely follows that of HT2 (**Figure 22**, **Figure 10**).

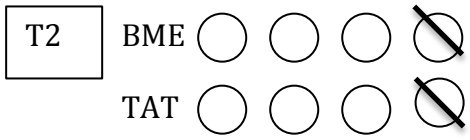


Figure 22. Experimental design T2: cells harvested 24 hours post treatment. One circle: one biological sample (cell culture dish). TAT - HIV-Tat treated samples. BME - control samples. Dashed out circles indicate that that sample was compromised during sample preparation and thus excluded from final analyses.

5.5.2 MS data quality and raw data processing

The data were of good quality as shown by high percentage of MS2 spectra identified (50.77%, **Table 13**). This corresponded to a high number of protein groups identified, low percentage contaminants and few spectra mapping to the decoy database (**Table 13**). Further, tryptic digestion was good (**Figure 23**) with 87.6% total digestion. In addition, low average absolute mass error was evident (0.7 ± 1 ppm, **Figure 24**).

Table 13 Data quality parameters for 24 hour repeat experiment

Parameter	Measurement
MSMS spectra submitted	251285
MSMS spectra identified	127577 (50.77%)
Protein groups identified	3183 (1907 across all 6 samples)
Contaminants (%)	0.5
Mapped to decoy database (%)	1.00

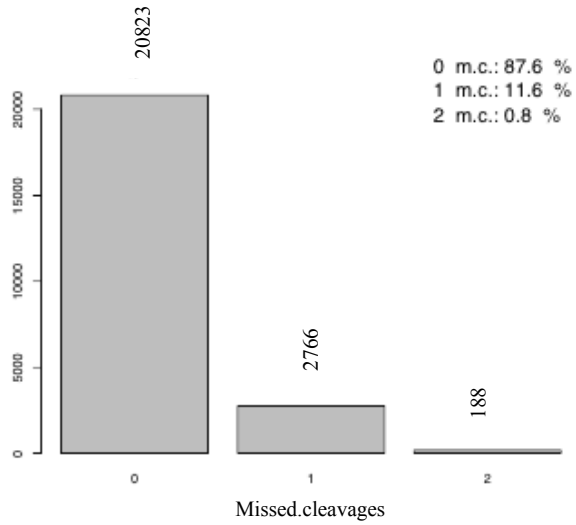


Figure 23. Assessment of tryptic digest completeness m.c.: missed cleavages.

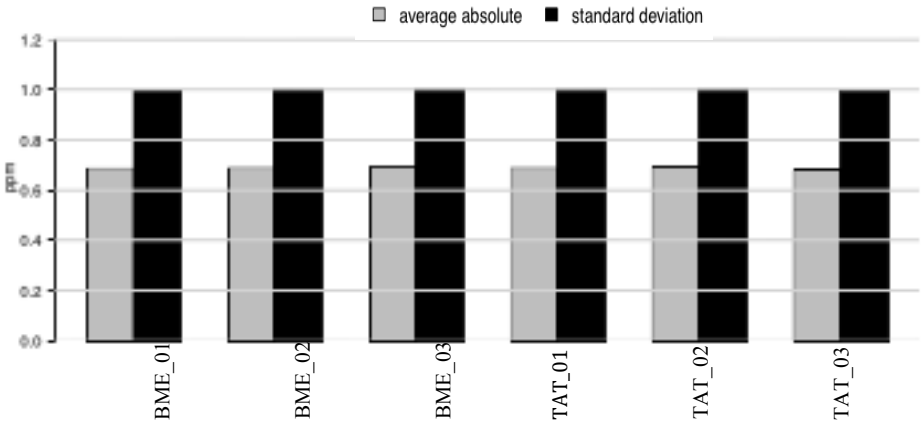


Figure 24. Average absolute mass error of the MS instrument per sample, ppm - parts per million, TAT - HIV-Tat treated, BME - control.

5.5.3 Statistical analysis

Descriptive statistical analysis

Scatterplots of protein expression values, for the 1907 proteins groups identified in each HIV-Tat treated and control sample, revealed linear relationships between technical replicates and across experimental conditions (**Figure 25**). Boxplot analysis (**Figure 26**) showed no systematic technical variation. Taken together, these results indicated high reproducibility. While cluster analysis showed weak clustering of samples by experimental condition (**Figure 27**), more subtle changes to protein expression were revealed by differential expression analysis as described in the following section.

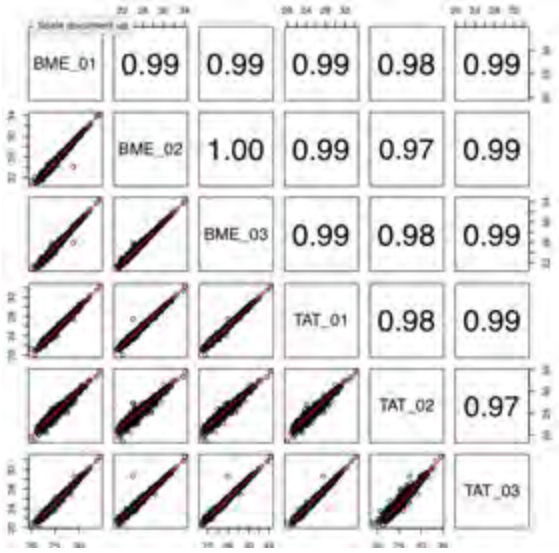


Figure 25. Agreement between technical replicates and across experimental conditions
$\log_2$ protein expression values from each experimental sample were plotted against those from every other experimental sample. Numeric values in the top right section indicate the R^2 value for each comparison. TAT - HIV-Tat treated, BME - control.

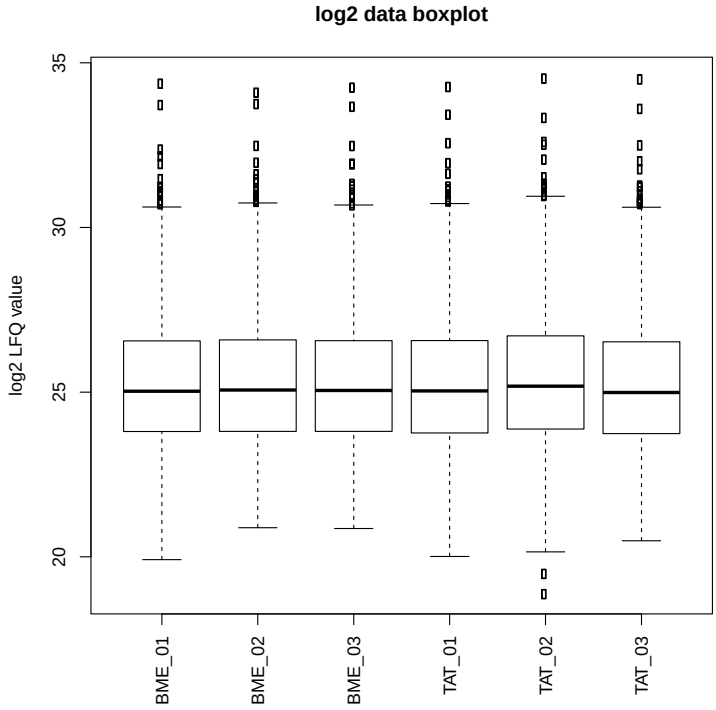


Figure 26. Boxplot representations of log_2LFQ intensity values per replicate per condition. TAT-HIV-Tat treated, BME - control.

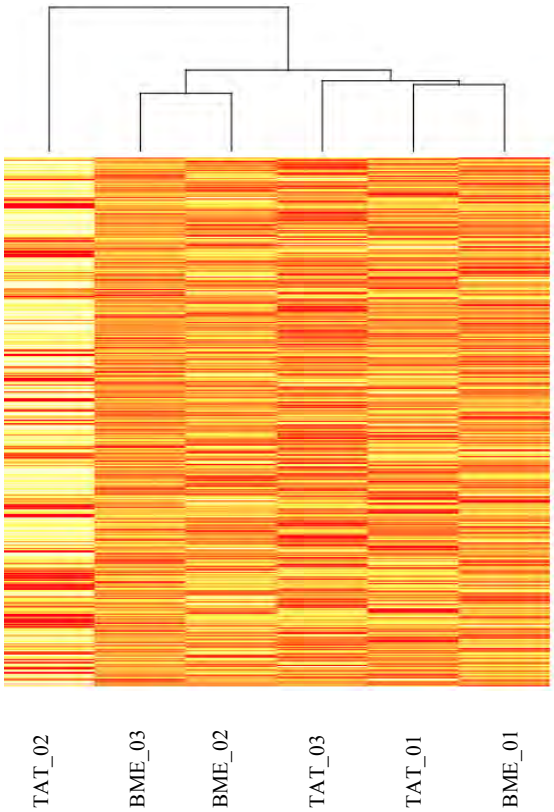


Figure 27. Hierarchical Cluster analysis of total proteomic profile revealed weak clustering of samples by experimental condition. Note: the darker the band, the greater the expression of that particular protein group. TAT - HIV-Tat treated, BME - control.

Differential expression analysis

A total of 68 protein groups were differentially expressed between HIV- Tat treated and control samples ($p < 0.05$) (**Figure 28)**. It is this list of 68 that was taken forward for biological significance analyses and compared to differentially expressed protein groups from time point 2 of the higher cell confluence time course experiment.

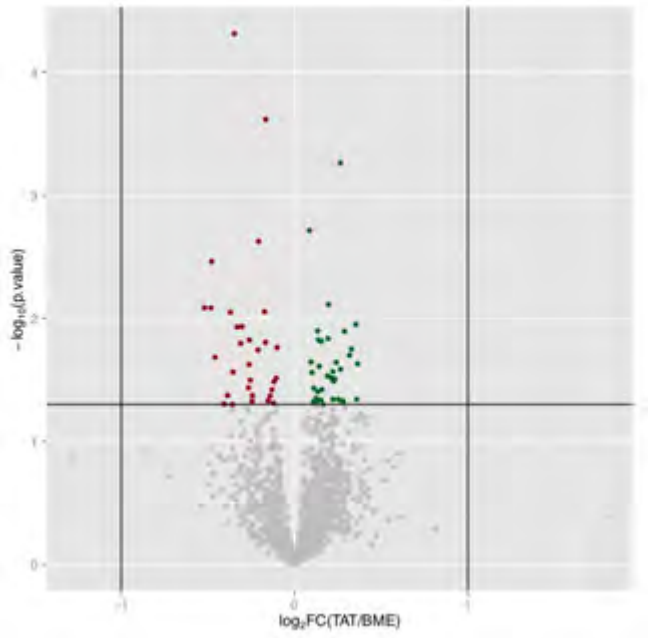


Figure 28. Visual representation of expression changes between HIV-Tat and BME (vehicle) treated samples protein groups plotted as a function of significance ($-\log_{10}$ p value) and $\log_2$ fold change ($\log_2$FC). Foldchange = TAT/BME. TAT- HIV-Tat treated, BME – control. Horizontal line indicates p = 0.05. Vertical lines indicate 2 standard deviations from 0 ($\pm\log_2 1$).

5.5.4 Biological significance analysis

The repeat correlated well with the 24-hour time point (HT2) from the time course experiment (**Figure 29**). It is apparent that both experiments contained many protein groups that mapped to the Panther protein classes *nucleic acid binding* and *cytoskeletal protein* (**Figure 29**). This enrichment was significant for protein classes related to *cytoskeletal protein* ($p < 0.05$) (**Table 14**), such as: class *intermediate filament* (PC00129), which was over-represented greater than 5-fold (p = 2.63e-3); and class *extracellular matrix structural protein* (PC00103), also over-represented greater than 5-fold (p = 3.68e-3). Further, differentially expressed protein groups that mapped to the *cytoskeletal protein* class were all down-regulated in response to HIV-

Tat treatment (**Tables C1**, Appendix C) and related to cytoskeletal maintenance/transport (e.g. CALD1, VINC, TAGLN). This was in agreement with what was seen during the time course experiment at 24 hours post treatment (**Table B5**, Appendix B).

Nucleic acid binding class protein groups - which act at both the DNA and RNA level - appear to have been up-regulated in response to HIV-Tat treatment (**Table C2**, Appendix C). This finding was surprising as it is in opposition to the trend seen at HT2, in which *nucleic acid binding* protein groups are down-regulated (**Table B2**, Appendix B). The protein class *nucleic acid binding* - or other Panther protein classes in the same parent lineage - was/were not significantly over/under represented in this repeat experiment (**Table 14**).

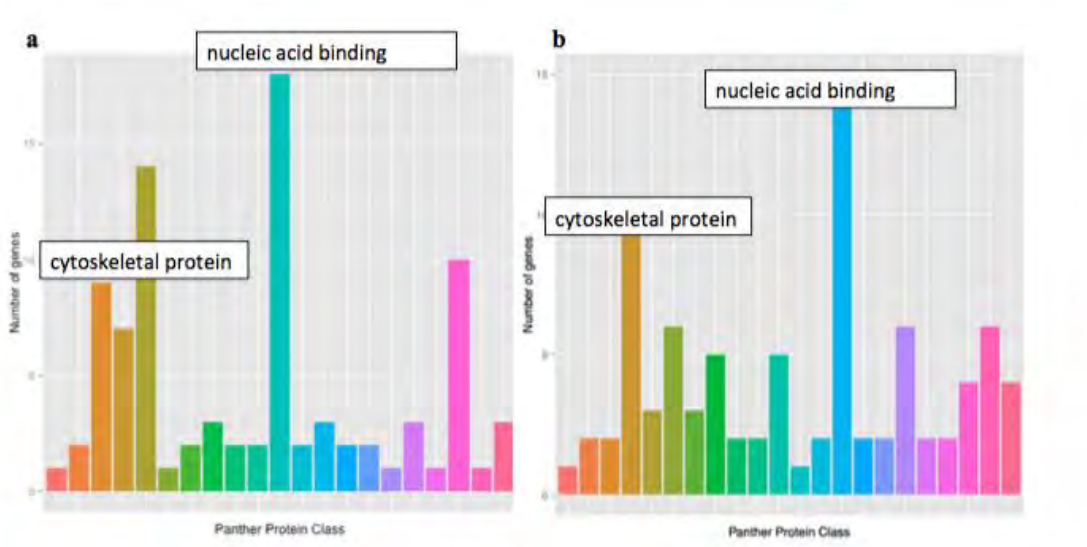


Figure 29. Panther protein class classification for repeated 24 hour time point data correlated with time point two from the time course experiment (HT2) a 24 hour time point for time course experiment (HT2) **b** Repeated 24 hour time point experiment.

Table 14 Top 10 output from over/under representation analysis of Panther protein class terms at **24 hours post** treatment (**repeat** experiment)

PANTHER Protein Class	Background (out of 1612 protein groups)	Differentially expressed protein groups - input (out of 55 protein groups)	Differentially expressed protein groups - expected	over/under represented	fold enrichment	p.value
receptor (PC00197)	50	7	1.71	+	4.10	1.52E-03
intermediate filament (PC00129)	8	3	.27	+	> 5	2.64E-03
defense/immunity protein (PC00090)	17	4	.58	+	> 5	2.75E-03
extracellular matrix structural protein (PC00103)	9	3	.31	+	> 5	3.68E-03
extracellular matrix protein (PC00102)	20	4	.68	+	> 5	4.89E-03
antibacterial response protein (PC00051)	10	3	.34	+	> 5	4.92E-03
surfactant (PC00212)	10	3	.34	+	> 5	4.92E-03
structural protein (PC00211)	11	3	.38	+	> 5	6.40E-03
tyrosine protein kinase receptor (PC00233)	1	1	.03	+	> 5	3.36E-02
non-motor actin binding protein (PC00165)	40	4	1.36	+	2.93	4.76E-02

5.6 Time course experiment with NES cells differentiated at 70% cell confluence

The above time course experiment (cultured cells differentiated at 90% cell confluence and then treated at 9 days post differentiation) was repeated with cells differentiated at lower cell confluence (70%) in order to determine the effect of cell density - and (presumably) resultant synaptic density - on HIV-Tat treatment. All *Experimental Procedures* pertaining to cell culture (except confluence at differentiation), HIV-Tat treatment, and MS runs and data processing applied to this dataset. Please see **Figure 30** for the experimental layout and **Figure 31** for sample identifiers.

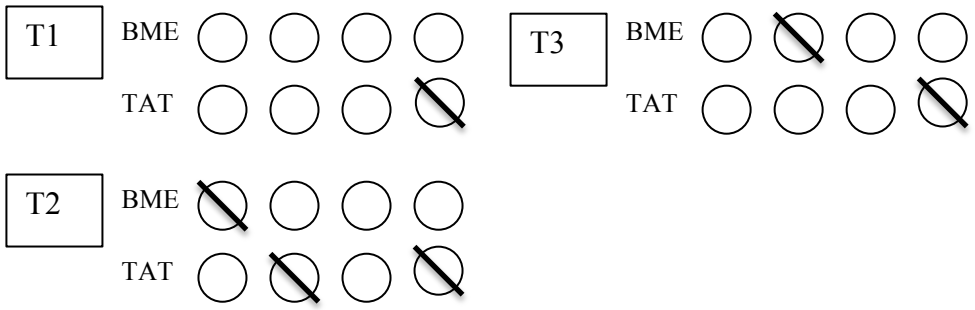


Figure 30. Lower cell confluence (70%) time course experimental design T1: Time point 1 (harvested 6 hours post treatment). T2: Time point 2 (harvested 24 hours post treatment). T3: Time point 3 (harvested 48 hours post treatment). One circle represents one biological sample (cell culture dish). TAT: HIV-Tat treated samples. BME: control samples. Samples with dash through indicate that that sample was compromised (data not shown) and thus excluded for final analyses. Note: T2TAT1 and T2TAT3 were run twice each on the MS instrument and only protein groups common to both runs in each case were taken forward for analysis. This was done in order to account for the shortfall of only having two biological replicates for the T2TAT experimental condition.

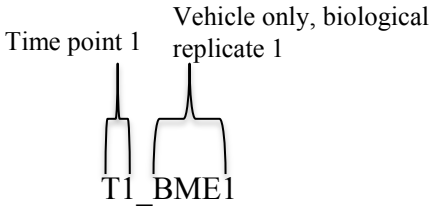


Figure 31. Sample identifier key In this example: sample from time point 1, control sample (BME), and biological replicate 1.

5.6.1 MS data quality and raw data processing

Table 15 details MS data quality parameters from this dataset, which were of good quality as was evident from good percentage MSMS spectra identified, large number of protein groups identified as well as low contaminants and mappings to the decoy database. Further, tryptic digestion was good (**Figure 32**), despite being a bit lower than the expected 80% total digestion (Blackburn laboratory, unpublished). In addition, low average absolute mass error was evident (0.7 ± 1 ppm, **Figure 33**).

Table 15 Data quality parameters for lower cell confluence experiment

Parameter	Measurement
MSMS spectra submitted	1144021
MSMS spectra identified	455737 (39.84%)
Protein groups identified	3692
Contaminants (%)	0.406
Mapped to decoy database (%)	0.01

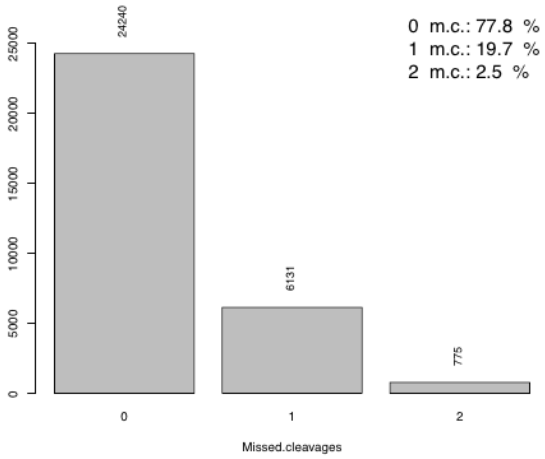


Figure 32. Assessment of tryptic digest completeness m.c. missed cleavages.

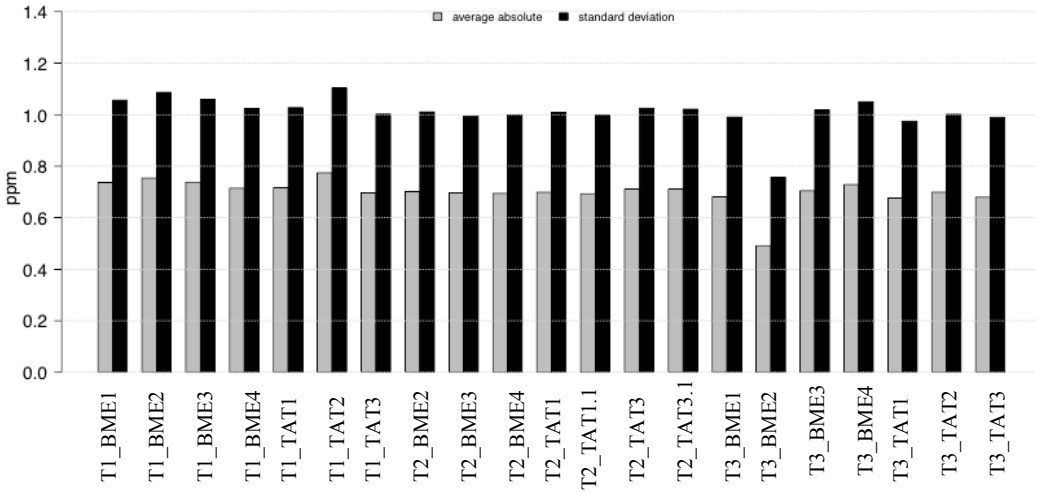


Figure 33. Average absolute mass error of the MS instrument per run ppm - parts per million, BME - control, TAT - HIV-Tat treated.

5.6.2 Statistical analysis

Descriptive statistical analysis

Technical replicates correlated well ($R^2 > 0.9$ in all comparisons, **Figure 34**) - except T2TAT2, which was removed from further analyses (T2TAT1 and T2TAT3 were run twice each and the average taken to account for the loss of T2TAT2). In addition, boxplot analysis (**Figure 35**) showed no systematic technical variation, indicating high reproducibility. Within each time point, treated and control samples did not appear to cluster separately (**Figure 36a-c**). Moreover, when all time points were analysed together by cluster analysis, they did not cluster by time point either (**Figure 36d**) - apart from a few cases of clustering within biological replicates within time points - which suggests that no exogenous time point biases were introduced. More subtle changes, to smaller groups of proteins, were revealed by t test analysis.

T1 (6 hours) T2 (24 hours) T3 (48 hours)

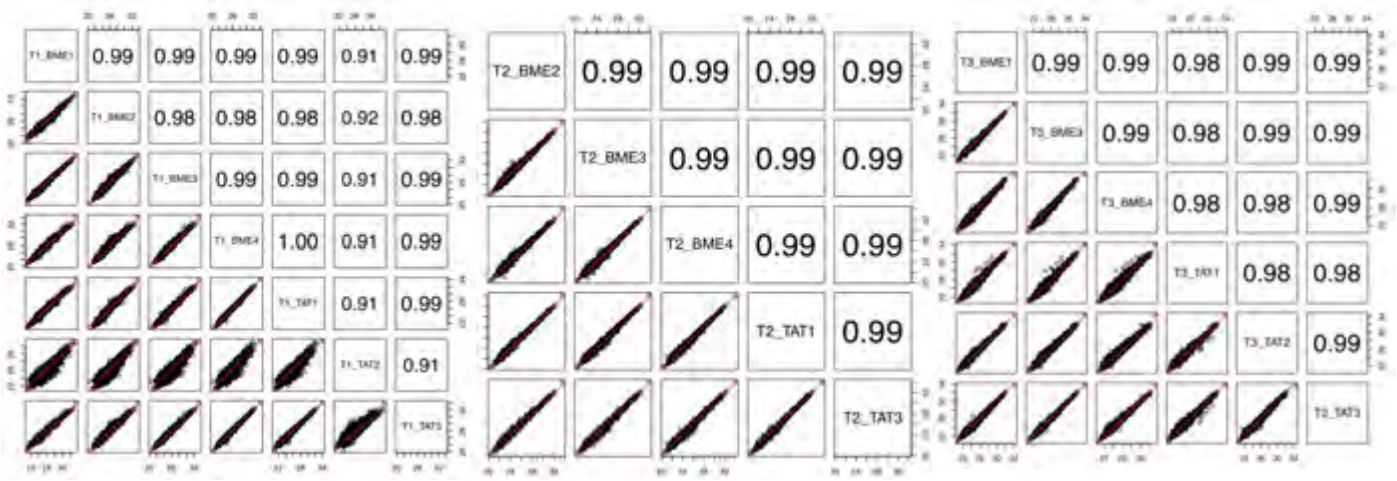


Figure 34. Agreement between technical replicates and across experimental conditions for all three time points $\log_2$ protein expression values from each experimental sample were plotted against those from every other experimental sample, per time point. Numeric values in the top right half of each graph indicate the R^2 value for each comparison. BME - control, TAT - HIV-Tat treated.

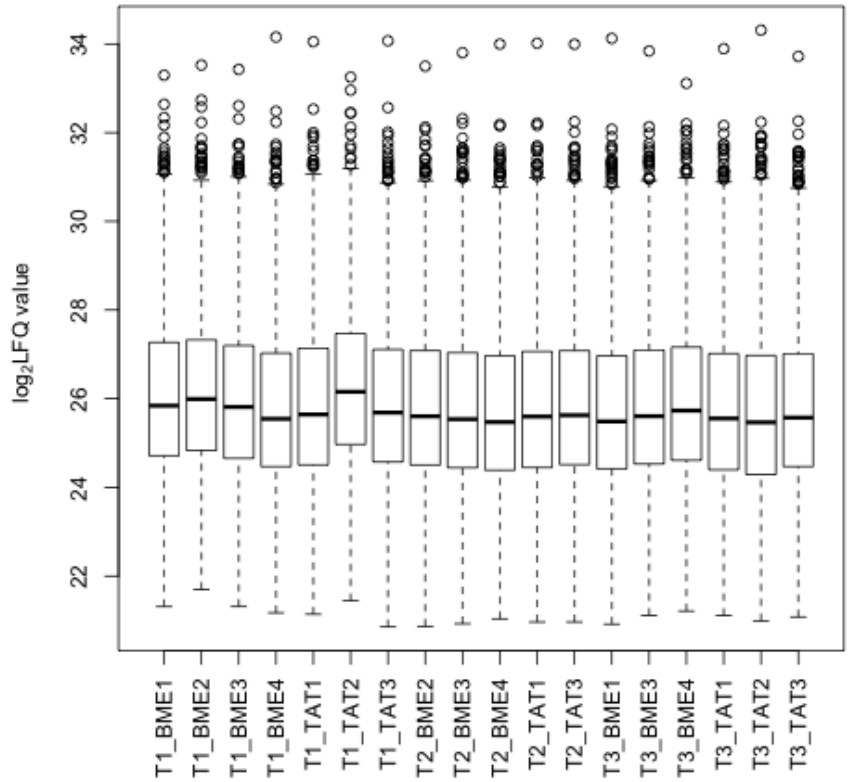


Figure 35. Boxplot representations of $\log_2$LFQ intensity values per replicate per condition per time point

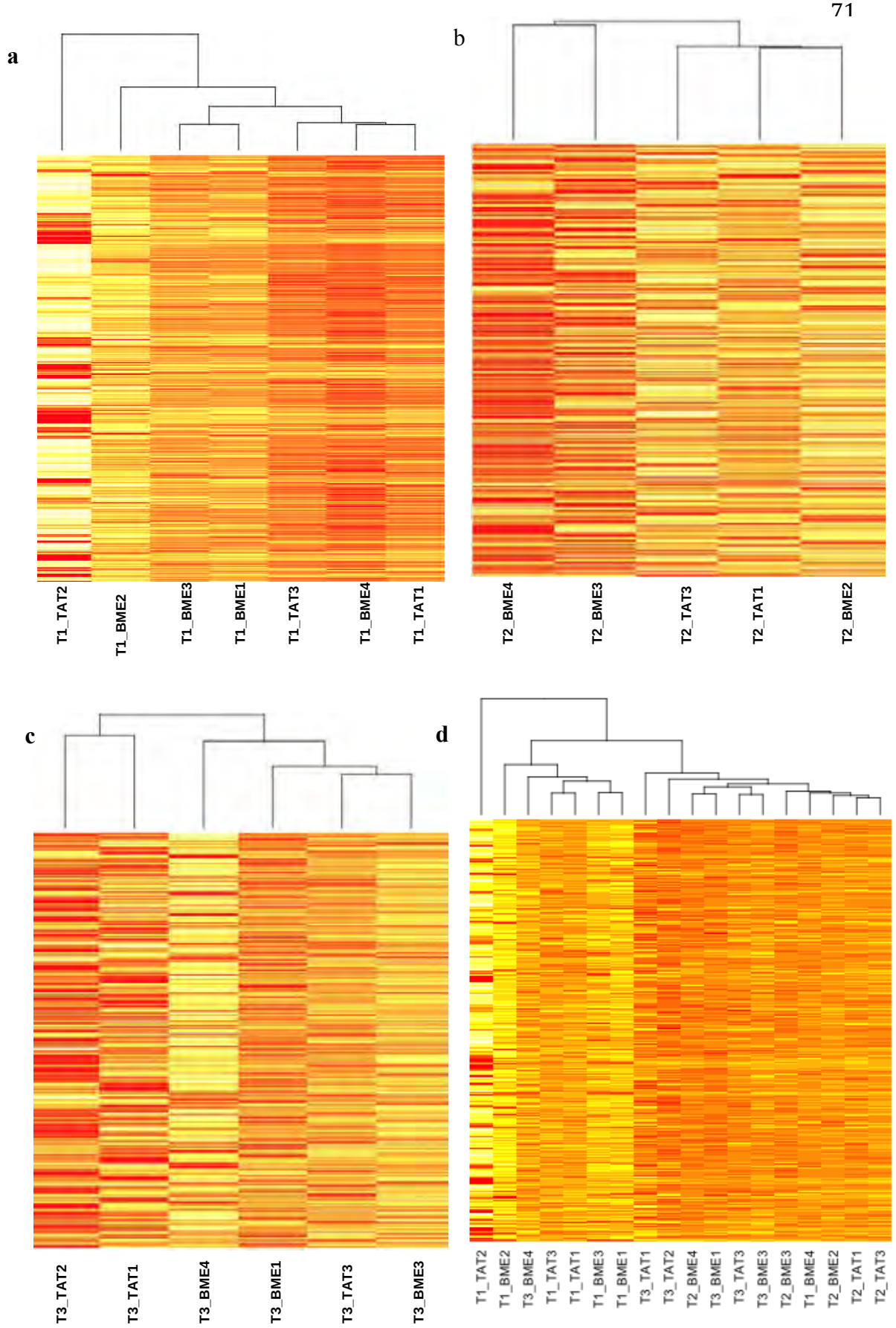


Figure 36. Hierarchical cluster analysis of total proteomic profile revealed that treated and control samples did not cluster separately and that there was no clustering by harvest timepoint **a** T1: 6 hours post treatment. **b** T2: 24 hours post treatment. **c** T3: 48 hours post treatment **d** all time points included. Note: the darker the band, the greater the expression of that particular protein group.

Differential expression analysis

Protein groups differentially expressed between BME and HIV- Tat treated samples ($p < 0.05$), per time point, were visualized by volcano plot (**Figure 37**). Significantly differentially expressed protein groups were taken forward for biological significance analyses and comparison to differentially expressed protein groups from time course data of cells differentiated at 90% cell confluence.

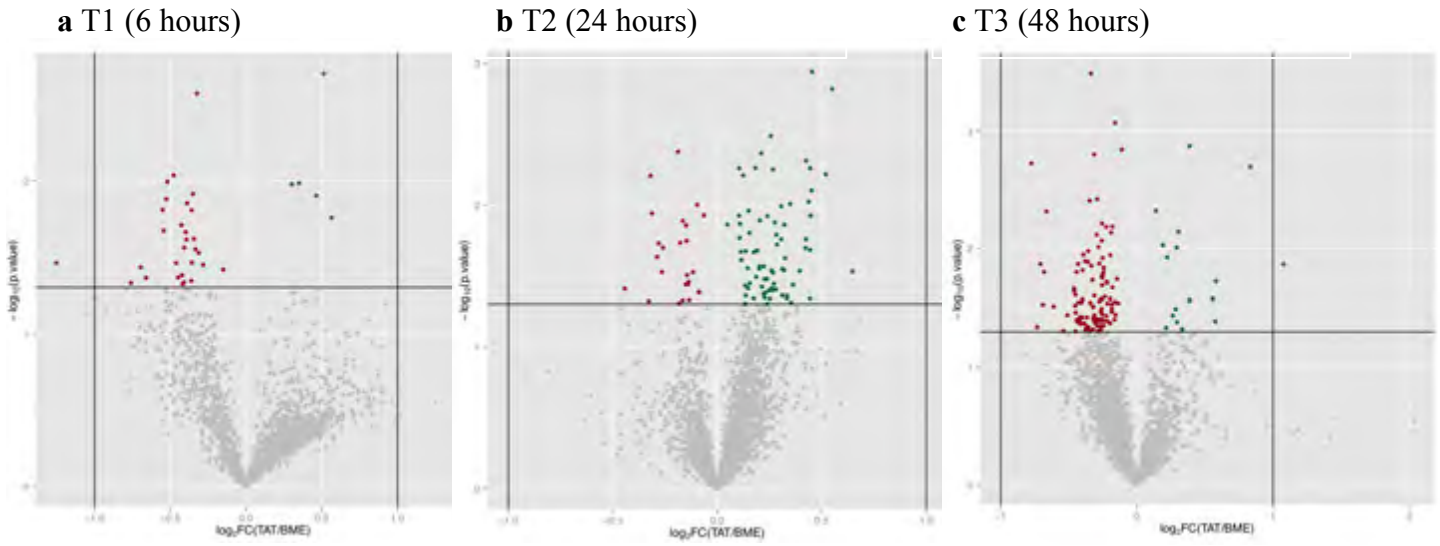


Figure 37. Visual representation of expression changes between HIV-Tat treated and control samples for lower cell confluence time course experiment protein groups plotted as a function of significance ($-\log_{10}$ p value) and $\log_2$ fold change ($\log_2$FC). **a** T1: 6 hours post treatment **b** T2: 24 hours post treatment **c** T3: 48 hours post treatment. FC**:** Fold change = TAT/BME. Horizontal line indicates p = 0.05 (level of significance set for the independent t test analysis per protein group). Vertical lines indicate two fold change in expression ($\pm\log_2 1$). Up-regulated protein groups coloured green, down-regulated protein groups coloured red. BME - control. TAT - HIV-Tat treated.

5.6.3 Comparison of sample preparation across time course experiments

Cell counts were largely uniform across samples (low standard deviation for cell counts, **Table 17**) suggesting robust cell culture plating. BCA quantitation provided additional support that samples differentiated at 70% cell confluence indeed contain lower levels of total protein - seemingly as the result of fewer cells - than cells differentiated at 90% cell confluence (**Table 17**, **Table 16**, **Table 2**). In addition, protein amount per cell correlated between the two experiments (**Table 17**) suggesting uniformity in protein production at the cellular level despite differences in cell confluence. One could hypothesise that fewer cells imply a lower number of

neuronal processes and lower synaptic density, which was visually evident from the cell culture dishes (data not shown).

Table 16 Cell counts and protein quantitation for time course experiment (cells differentiated at 70% cell confluence)

Sample*	**Cell count****	**Protein concentration** (μg/ μl)	**Protein recovered** (μg)	**Protein yield per cell** (pg)
T1_BME1	8.73 x 10^5	0.57	38.59	44.203
T1_BME2	4.2 x 10^5	0.60	37.98	90.42
T1_BME3	1.74 x 10^6	0.69	39.81	22.80
T1_BME4	1.65 x 10^6	1.11	67.82	41.10
T1_TAT1	1.68 x 10^6	1.03	74.01	44.05
T1_TAT2	4.13 x 10^5	0.21	13.99	33.87
T1_TAT3	8.73 x 10^5	0.73	47.012	33.58
T2_BME2	8.3 x 10^5	0.98	58.50	70.48
T2_BME3	2.38 x 10^6	0.78	49.91	24.35
T2_BME4	8.3 x 10^5	0.87	56.028	34.16
T2_TAT1	2.05 x 10^6	0.89	62.92	37.31
T2_TAT3	1.64 x 10^6	0.91	63.70	26.76
T3_BME1	1.046 x 10^6	0.77	55.23	52.80
T3_BME3	1.51 x 10^6	0.57	42.408	28.75
T3_BME4	1.095 x 10^6	0.61	38.962	35.58
T3_TAT1	1.48 x 10^6	0.87	55.071	37.21
T3_TAT2	1.88 x 10^6	0.74	51.8	27.55
T3_TAT3	1.73 x 10^6	0.76	53.68	31.028

* T1_: time point 1; T2_: time point 2; T3_: time point 3; TAT: HIV-Tat treated; BME: control.
**Cell counts from 6cm cell culture dishes

Table 17 Comparison of cell culture parameters for cells differentiated at 70% versus 90% cell confluence

Parameter	**70% cell confluence**	**90% cell confluence**
Average cell count	1.32 x 10^6 ± 0.60 x 10^6	2.77 x10^6 ± 0.55 x 10^6
Average protein yield per cell (pg)	39.77 ± 16.96	60.91 ±17.37

5.6.4 Biological significance analysis

For the lower cell confluence experiment, trends for normalised and un-normalised protein groups per Panther protein class mirrored those seen in the higher cell confluence experiment (**Figure 38**, **Figure 19**). Indeed, the trends of un-normalised protein groups, and protein groups normalised to the full list of protein groups identified per time point, were almost identical (**Figure 38a** and **b**). In addition, the trend change between un-normalised protein groups (**Figure 38a**) and protein groups normalised to the number of differentially expressed protein groups per time point (**Figure 38c**) may be explained by the trend in the number of differentially expressed protein groups over time (**Figure 39**). The protein classes *cytoskeletal protein* and *nucleic acid binding* - which were affected by HIV-Tat treatment in cells differentiated at 90% cell confluence - were also analysed in HIV-Tat treated cells differentiated at 70% cell confluence, as discussed below.

When *cytoskeletal protein* was considered, it was apparent that trends were comparable across the two independent time course experiments (**Figure 40c**), including the observation that most, if not all, of the protein groups were down-regulated in response to HIV-Tat treatment across all time points (**Table D1-3**, Appendix D; **Tables B4-6**, Appendix B). However, the proportion of protein groups that mapped to this protein class was much lower for the lower cell confluence experiment (**Figure 40c**). Further, there was little evidence of over/under-representation (in relation to the 'background list') of protein classes that belonged to the same parent lineage as *cytoskeletal protein* at all three time points for cells differentiated at lower confluence (**Tables 18-20**).

When *nucleic acid binding* was considered, the trend - in terms of the proportion of protein groups that mapped to this protein class per time point - mirrored that of the higher cell confluence experiment (**Figure 40c**). In addition, protein classes in the same parent lineage as *nucleic acid binding* were significantly over/under represented in comparison to background for the lower cell confluence experiment ($p < 0.05$) (**Tables 18-20**). For example, the class *histone (PC00118)* was over represented greater than 5-fold at 6 hours post treatment ($p = 3.16e-2$) (**Table 18**), the class *centromere DNA-binding protein (PC00071)* was over represented greater than 5-fold at 24 hours post treatment ($p = 5.07e-3$) (**Table 19**), and the class *nucleotide phosphatase (PC00173)* was over represented greater than 5-fold at 48

hours post treatment (p = 2.13e-3) (**Table 20**) in relation to the background (i.e. full list of all identified protein groups). Further, when the functions and up-/down-regulation (log_2 fold change in expression values) of the implicated protein groups were considered, the two independent time course experiments were largely in agreement (**Tables D4-6**, Appendix D; **Tables B1-3**, Appendix B). Trends pertaining to *nucleic acid binding* over time for the lower cell confluence experiment are explained in more detail below.

At 6 hours post treatment, most *nucleic acid binding* protein groups were down-regulated (**Table D4**, Appendix D). While this appeared contrary to the higher cell confluence experiment at first glance - where such protein groups were largely up regulated (**Table B1**, Appendix B) - these protein groups (in the lower confluence experiment) were involved in chromatin condensation (**Table D4**, Appendix D) and down-regulation of these protein groups would presumably allow for overall increased gene expression. Therefore, this is in fact consistent with functional changes at 6 hours post treatment in the higher cell confluence experiment. Surprisingly, however, pre-mRNA processing appeared to be down-regulated at this time point in the lower cell confluence experiment (**Table D4**, Appendix D), while mRNA processing was up-regulated at this time point in the higher cell confluence experiment (**Table B1**, Appendix B).

At 24 hours post treatment - when the least number of protein groups mapped to *nucleic acid binding* - there appeared to be an up-regulation of gene expression inhibition genes (e.g. CBX8, SRS10, PRKRA) (**Table D5**, Appendix D) suggesting down-regulation of gene expression at this time point. This was in agreement with the 24-hour time point in the higher cell confluence experiment.

At 48 hours post treatment, most protein groups were down-regulated, but had both gene expression promoting and inhibiting functions (**Table D6**, Appendix D).

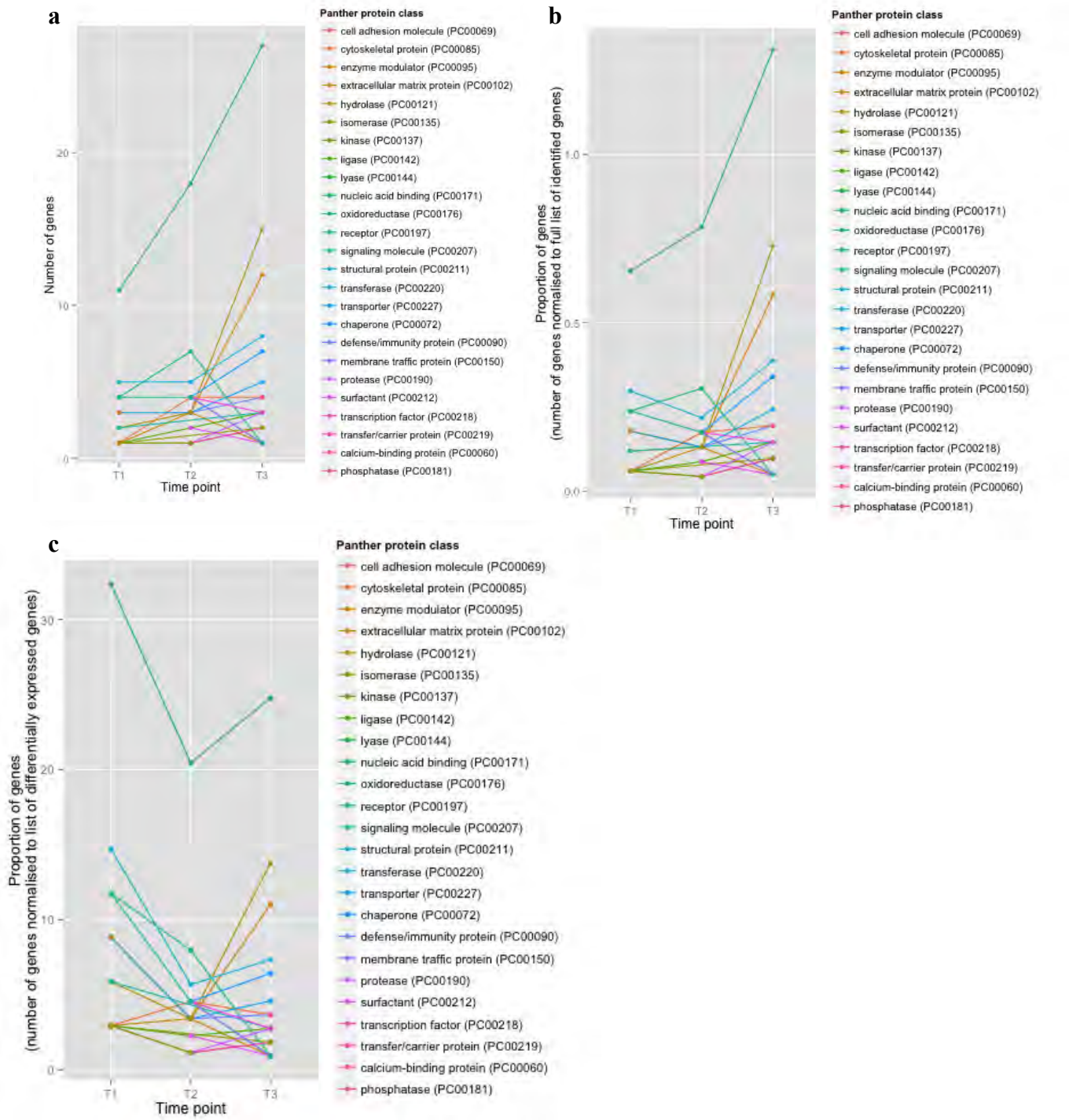
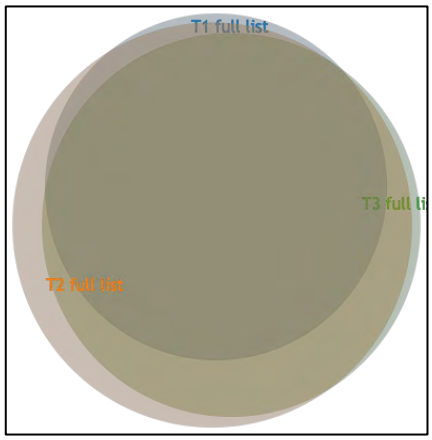


Figure 38. Functional classifications of differentially expressed protein groups over time, according to Panther protein class, for the lower cell confluence experiment a absolute number of genes per time point. **b** number of genes normalised to the full list of identified genes per time point. **c** number of genes normalised to the number of differentially expressed genes per time point. Gene name used as proxy for protein group. T1: 6 hours post treatment, T2: 24 hours post treatment, T3: 48 hours post treatment.

a

b



c

Time point	Gene IDs: total (full list)	Gene IDs: differentially expressed
T1	1682	34
T2	2294	88
T3	2060	109

Figure 39. Venn diagram showing overlap of proteins groups between different time points
a Overlap from differentially expressed protein groups list. Circles are drawn to scale. Time points T1, T2 and T3 (6, 24 and 48 hours, respectively). **b** Overlap of 'full list' – i.e. all protein groups identified in all experimental and all control samples for that time point. Time points T1, T2 and T3 (6, 24 and 48 hours, respectively). **c** Number of protein groups (represented by gene IDs) identified at each time point. Differentially expressed genes meet a cut-off of $p < 0.05$. T1: 6 hours post treatment, T2: 24 hours post treatment, T3: 48 hours post treatment. Venn diagrams generated using *Venny* online tool: www.stefanjol.nl/venny.

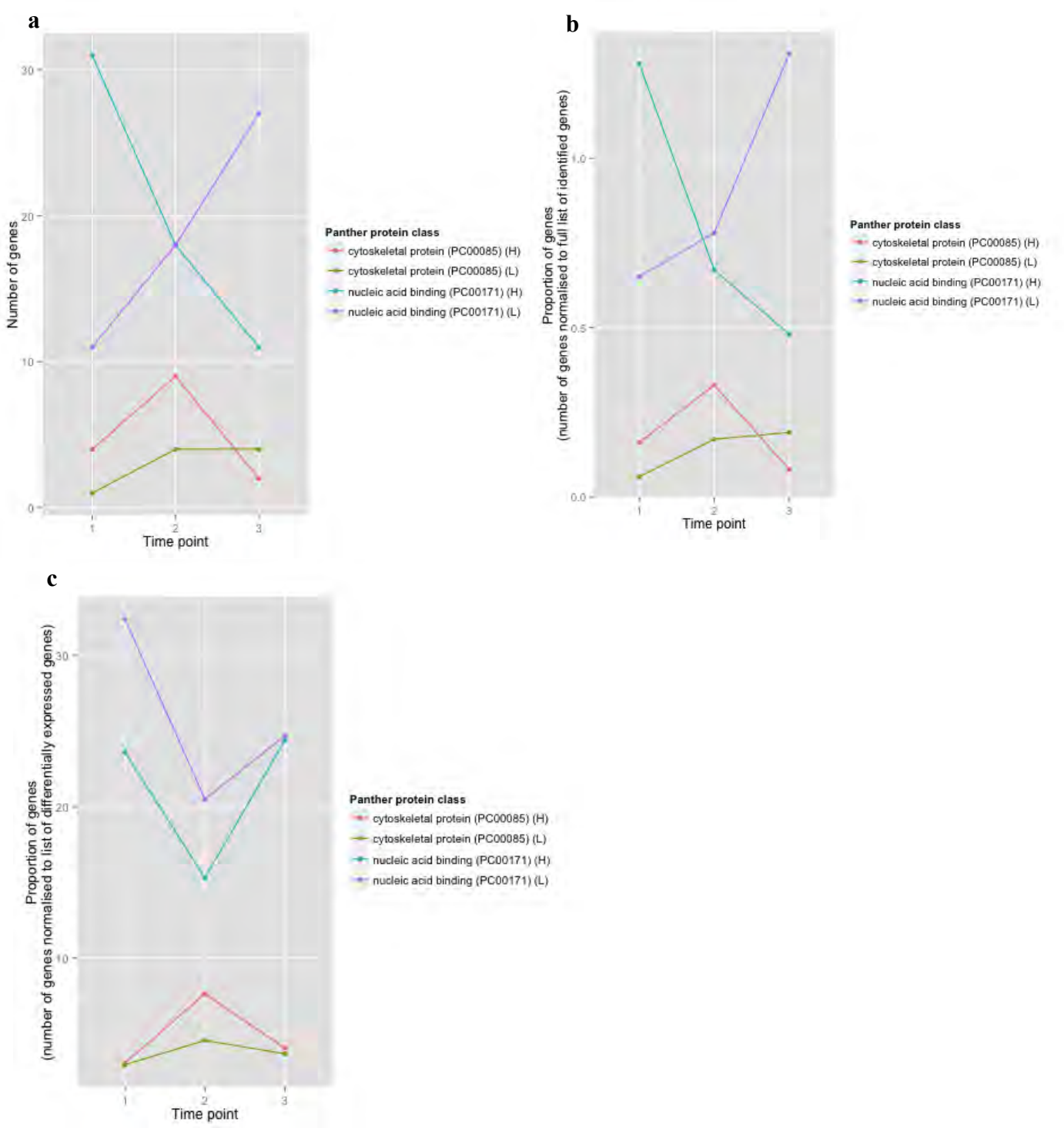


Figure 40. Trends of differentially expressed protein groups that mapped to the protein classes *nucleic acid binding* and *cytoskeletal protein* across time and cell confluence a Absolute number of genes per time point. **b** Number of genes normalised to the full list of identified genes per time point. **c** Number of genes normalised to the number of differentially expressed genes per time point. H: higher cell confluence time course experiment (cells differentiated at 90% cell confluence). L: lower cell confluence time course experiment (cells differentiated at 70% cell confluence). T1: 6 hours post treatment, T2: 24 hours post treatment, T3: 48 hours post treatment. Gene name used as proxy for protein group.

Table 18 Top 11** output from over/ under representation analysis of Panther protein class terms at **6 hours** post treatment (**lower confluence** experiment)

PANTHER Protein Class	Background * (out of 1447 protein groups)	Differentially expressed protein groups – input* (out of 31 protein groups)	Differentially expressed protein groups - expected	over/ under representation	fold enrichment	p.value
acyltransferase (PC00042)	9	3	.19	+	> 5	9.49E-04
ligand-gated ion channel (PC00141)	4	2	.09	+	> 5	3.37E-03
anion channel (PC00049)	6	2	.13	+	> 5	7.38E-03
ATP synthase (PC00002)	7	2	.15	+	> 5	9.91E-03
dehydrogenase (PC00092)	48	4	1.03	+	3.89	1.86E-02
ion channel (PC00133)	11	2	.24	+	> 5	2.32E-02
cation transporter (PC00068)	12	2	.26	+	> 5	2.73E-02
receptor (PC00197)	32	3	.69	+	4.38	3.07E-02
histone (PC00118)	13	2	.28	+	> 5	3.16E-02
decarboxylase (PC00089)	3	1	.06	+	> 5	6.23E-02
DNA binding protein (PC00009)	102	5	2.19	+	2.29	6.38E-02

*** please note** that not all Panther protein classes listed are significantly over/ under represented (check p.value column)*

** The number of protein groups belonging to background (i.e. full list) and differentially expressed protein group may not correlate with the values in **Figure 39c** as the Panther algorithm could not map all protein groups from the input lists*

Table 19 Top 10** output from over/ under representation analysis of Panther protein class terms at **24 hours** post treatment (**lower confluence** experiment)

PANTHER Protein Class	Background * (out of 1945 protein groups)	Differentially expressed protein groups – input* (out of 68 protein groups)	Differentially expressed protein groups - expected	over/ under representa -tion	fold enrichment	p.value
extracellular matrix structural protein (PC00103)	2	2	.07	+	> 5	2.30E-03
centromere DNA-binding protein (PC00071)	3	2	.10	+	> 5	5.07E-03
surfactant (PC00212)	3	2	.10	+	> 5	5.07E-03
extracellular matrix protein (PC00102)	12	3	.42	+	> 5	8.73E-03
antibacterial response protein (PC00051)	4	2	.14	+	> 5	8.81E-03
glycosyltransf erase (PC00111)	17	3	.59	+	> 5	2.20E-02
nuclear hormone receptor (PC00169)	1	1	.03	+	> 5	3.44E-02
ribonucleopr otein (PC00201)	26	3	.91	+	3.30	6.31E-02
DNA ligase (PC00012)	2	1	.07	+	> 5	6.76E-02
receptor (PC00197)	44	4	1.54	+	2.60	6.84E-02

*** please note** that not all Panther protein classes listed are significantly over/ under represented (check p.value column)*

** The number of protein groups belonging to background (i.e. full list) and differentially expressed protein group may not correlate with the values in **Figure 39c** as the Panther algorithm could not map all protein groups from the input lists*

Table 20 Top 10** output from over/ under representation analysis of Panther protein class terms at **48 hours** post treatment (**lower confluence** experiment)

PANTHER Protein Class	Background * (out of 1771 protein groups)	Differentially expressed protein groups – input* (out of 89 protein groups)	Differentially expressed protein groups - expected	over/ under represent -ation	fold enrichment	p.value
nucleotide phosphatase (PC00173)	5	3	.25	+	> 5	2.13E-03
kinase activator (PC00138)	9	3	.45	+	> 5	1.08E-02
oxidoreductase (PC00176)	111	1	5.58	-	< 0.2	2.19E-02
kinase modulator (PC00140)	32	5	1.61	+	3.11	2.29E-02
membrane traffic protein (PC00150)	73	0	3.67	-	< 0.2	2.36E-02
defense/immunity protein (PC00090)	13	3	.65	+	4.59	2.82E-02
chaperone (PC00072)	61	7	3.07	+	2.28	3.42E-02
dehydrogenase (PC00092)	62	0	3.12	-	< 0.2	4.19E-02
deaminase (PC00088)	7	2	.35	+	> 5	4.88E-02
G-protein modulator (PC00022)	42	5	2.11	+	2.37	6.09E-02

*** please note** that not all Panther protein classes listed are significantly over/ under represented (check p.value column)*

** The number of protein groups belonging to background (i.e. full list) and differentially expressed protein group may not correlate with the values in **Figure 39c** as the Panther algorithm could not map all protein groups from the input lists*

6. Discussion

This study aimed to use MS-based discovery proteomic analysis to identify the cell-wide effects of HIV-Tat on non-transformed, differentiated NES cells over time, representing an *in vitro* model of initial HAND progression. Findings from this work, in which clade B HIV-Tat was used, are applicable to the majority of HAND patients from Sub-Saharan Africa who are infected with clade C virus with HIV-Tat CC motif (dicysteine at position 30-31), as measures of neurovirulence are comparable for clade B and clade C HIV-Tat with the CC motif (Rao et al. 2013).

We have shown that differentiated NES cells display morphological and functional neuronal characteristics. We identified protein groups differentially expressed between HIV-Tat treated and control NES cells over time (6, 24 and 48 hours), at two cell confluences, using MS. In addition, we have largely supported findings for the higher (~90%) cell confluence 24-hour time point with a repeat. The cells differentiated at higher cell confluence and then treated show that HIV-Tat affects *nucleic acid binding* and *cytoskeletal* maintenance proteins up to 48 hours post treatment. Results from cells differentiated at lower (~70%) cell confluence and then treated mimic these trends for both *nucleic acid binding* and *cytoskeletal protein* Panther protein classes. However, the effect to *cytoskeletal protein* dysregulation is not statistically over represented (in comparison to the background proteome) at this cell confluence. Relevant sections from the Results are discussed in detail below.

6.1 NES cell line

The NES cell line represents a novel cell culture system for the study of HAND

In this study, we established the differentiated NES cells as a cell culture model representing HAND and showed that these cells display neuronal morphology at 9 days post differentiated (**Figure 6b**).

Further, we provided evidence for the presence of functional neuronal characteristics in these cells using the electrophysiological, patch-clamp technique. This technique, which allows for the measurement of passive electrophysiological properties of a cell by controlling the voltage or current passing through the cell, is described in detail elsewhere (Penner 1995).

While the electrophysiology study performed here was only a pilot (as more cells are typically recorded in order to increase confidence in findings), it is clear that differentiated NES cells displayed properties of neuronal tissue (**Figure 7**). This was apparent from the visual appearance of differentiated NES cells (**Figure 7a** and **Figure 6b**), the ability of NES cells to generate action potentials (**Figure 7b-d**) and the observation of additional membrane parameters characteristic of neurons (**Figure 7e-f**). It must be noted that the IR was a bit higher than expected (Telias et al. 2014), which is relevant because IR is inversely proportional to neurite outgrowth and the increased number of ion channels inserted into the plasma membrane - a process that occurs during the process of neuronal maturation (Moody & Bosma, 2005) - and therefore, one would expect a low IR measurement for mature neurons. Further, average maximum INa^+ and IK^+ were slightly lower than expected (Telias et al. 2014; Song et al. 2013). These observations could therefore indicate that the NES cells were not fully mature in their electrophysiological activity at the time of the experiment. In support, Tailor et al. (2013) show that the NES cells continue to mature up to four months after differentiation.

One may argue that a mono-population of fully mature neurons in culture wouldn't have the *in vivo* pressures (from other cell types etc.) to change. Indeed, cell culture models studying plasticity invariably work with mixed cell populations of neurons and glia (Zemianek et al. 2012). This could be the case with the NES cells, which differentiate into only GABAergic and glutamatergic neurons (i.e. no glia) (Tailor et al. 2013). Therefore by choosing an earlier time point when the proteome of the NES cells are relatively stable (Garnett, unpublished) and when the cells produce action potentials (**Figure 7**) (but the cells are not necessarily fully mature) one may be able to 'construct' plasticity. As neuronal plasticity is crucial in forming synapses, and as the synapse is, in turn, the crucial molecular component for learning and memory (Mayford et al. 2012), plasticity is especially important in the study of HAND given the learning and memory deficits that characterise this illness. The NES cell line used in the current research therefore represents a physiologically relevant and novel system for the study of HAND.

The observation of neuronal characteristics in the differentiated NES cells corroborated expected morphological differences between NES cells in the stem cell and differentiated state (**Figure 6**) (Tailor et al. 2013). Findings were further corroborated by proteomic data, which showed little change in global protein

expression of NES cells between 8 and 12 days post differentiation (Garnett, unpublished), indicating that the cells were no longer in a proliferative state at the chosen time point (post differentiation) for treatment. While electrophysiological properties have been studied previously in differentiated NES cells as mentioned (Tailor et al. 2013), characterisation had not previously been carried out for NES cells as early as 9 days post differentiation.

6.2 Assessment of protein preparation procedures

Cell count and protein quantitation correlated well (**Table 2**), indicating that BCA quantitation was sufficient to estimate total cell protein. While HT2_BME samples seemed to have slightly more protein yield per cell than the other samples (**Table 2**) this did not affect MS analysis as was evident from no observed systematic technical error across samples (**Figure 15** and **16**). A total of 30µg protein was taken from each cell pellet for MS analysis, which was ample protein considering that not more than a hundred ng of protein is required for a MS experiment and that some protein loss may occur during processing (Mann, FASP protocol: lower and upper limits for FASP MWCO filters is 10µg and 200µg, respectively; Wiśniewski et al. 2009). In a label free MS experiment, it is important to ensure that equal quantities of peptide are loaded per sample, as all parameters - other than the parameters explicitly being tested (i.e. experimental conditions) - must be kept constant in order to make meaningful comparisons between MS runs. Further, the amount of peptide loaded must be kept constant across samples in order to compare protein quantity across runs and to make inferences about the individual protein groups that make up the total protein pool. In a label-free MS experiment, each sample is run on the MS instrument separately and, unlike during a labeled method (e.g. SILAC or iTRAQ), there are no internal standards for base value normalisation (Bantscheff et al. 2007). As such, consistency of experimental parameters is of the utmost importance.

FASP was chosen as a proteome processing protocol, as it efficiently removes residual SDS originating from the lysis buffer. It is important to remove SDS, as it would affect tryptic digestion and it ionises more readily than peptides to significantly increase background spectra (Wiśniewski et al. 2009). Peptides were desalted after FASP, as salts - i.e. other ionic compounds - also interfere with ionisation of peptides during ESI (Wiśniewski et al. 2009).

6.3 MS-based proteomic analysis: time course experiment of NES cells differentiated at 90% cell confluence

6.3.1 Assessment of MS data quality and raw data processing

A total of 1 067 211 spectra were recorded, of which 53.28% were mapped to known protein groups by MaxQuant software, corresponding to a total of 4104 protein groups across the 20 samples analysed. In order to increase confidence in findings, a protein group was only taken forward for comparison across experimental conditions if that protein group was observed in all replicates of a given condition (a parameter aided by MaxQuant's LFQ 'match between runs' function during raw spectra processing). Therefore, the more protein groups identified across all samples, the more complete our insights into the effect of HIV-Tat treatment on the neuronal proteome was. However, a snapshot of even just the most abundant (and thus perhaps functionally dominant) proteins in a cell at any given time should give one a good idea of the physiological affects of treatment. Considering a recent estimate that 3000 proteins constitute ~99% of the protein mass of a mammalian cell (Wiśniewski et al. 2012), it is encouraging that this work identified 2314 protein groups (~77% of the 3000 most abundant proteins) common to all HIV-Tat-treated and control samples.

Identity assignment to 53.28% of the mass spectra is in line with what is expected from the Andromeda search algorithm used in this work (Cox et al. 2011), and is among the highest achieved in the Blackburn laboratory, which typically ranges from 30-50% depending on the sample type and origin (unpublished). Spectra not assigned may have low spectral signal-to-noise ratios, which may be introduced if two different peptide precursor ions, that have co-eluted and enter the mass analyser at the same time, are simultaneously selected for MS2 fragmentation. Alternatively, the actual background noise may be high compared to the signal as the result of poor proteome processing - incomplete desalting, for example.

High data quality is also supported by the finding of good tryptic digestion of samples, with an average of 80.6% no missed cleavages (**Figure 13**). It is important that tryptic digestion of the proteome be as complete as possible in order to produce peptides in the optimal mass range for MS sequence identification (Duncan et al. 2010) thereby allowing for appropriate proteome (and individual protein) coverage. In addition, the dataset had a low absolute mass error of 0.7 ± 1 ppm (**Figure 14**) (the acceptable MaxQuant limit is 6ppm, Cox et al. 2009). Mass error is the variance in

average measured peptide mass when compared to theoretical peptide mass generated by *in silico* digestion of the provided protein database (Cox et al. 2009). Consistent and low mass error across samples indicates MS instrument accuracy and stability.

Moreover to data quality, TIC profiles were all of the desired intensity (maximum of ± 5.5e9), indicating correct concentration of the peptide sample for optimal identification of peptide ions (examples of appropriate TICs at the MS and MS2 levels can be viewed in **Figure 12e**). Further, data was collected along the entire length of the chromatogram in defined peaks indicating good separation of the peptide sample during online HPLC (e.g. **Figures 12e**). Good separation of peaks - and thus increased resolution - is important for accurate assignment of mass spectra, and affects the mass error calculation.

6.3.2 Statistical analyses

LFQ expression values per protein group were used to determine statistical differences between samples. To aid visualisation of the data (boxplots, scatterplots, hierarchical clustering and volcano plots), LFQ values were $\log_2$ normalised. For differential expression analysis, LFQ values were not normalised. While HIV-Tat treated and control samples appeared to cluster separately for all 3 time points (**Figure 17a-c**), the samples were actually quite similar, as indicated by long dendrogram branches calculated by the Euclidian model of absolute distance. This was supported by scatterplot representation, which showed linear relationships (i.e. similarity) across experimental conditions (**Figure 15a-c**, $R^2 > 0.95$). The observed clustering by time point across experimental conditions (**Figure 17d**) may be the result of subtle experimental errors during cell harvesting at the different time points. However, this clustering was also weakly correlated, as indicated by imperfect separate clustering of time points (**Figure 17a-c**) and by similarity of LFQ value range across samples as seen by boxplot analysis (**Figure 16**), which revealed no systematic technical error across samples. Further, as mentioned, each treated group was only compared to the control group from that time point, so any experimental effects common to a time point should not bias downstream analyses.

Despite overall similarities in protein expression, more subtle differences on a per-protein-group per time point level were revealed by hypothesis testing (**Figure 18**; **Table 3**). Indeed, differences in protein expression between HIV-Tat treated and

control samples per time point were subtle (**Figure 18**, $-1 < \log_2 FC < 1$), but statistically significant ($p < 0.05$). A total of 131, 118, and 45 protein groups were identified as being significantly differently expressed at 6, 24, and 48 hours post HIV-Tat treatment, respectively (**Table 3**). These lists of differentially expressed protein groups per time point were taken forward for biological significance analysis.

6.3.3 Biological significance analysis

Functional classification into Panther protein classes elucidated the biological significance of protein groups differentially expressed in response to HIV-Tat treatment over time. This represents a protein-centric approach to biological significance analysis, versus a protein families approach, in which Gene Ontology classifications are typically used (Harris et al. 2004). A protein-centric approach was a relevant type of analysis for the current work given the strong statistical evidence on a per peptide basis, as indicated by low PEP scores (**Tables B1**-**B6**, Appendix B). A decision was made to focus on differentially expressed protein groups in the context of functional classifications (versus a focus on the function of individual protein groups in isolation or a focus on the most up-/down- regulated protein groups) as very interesting overarching trends - pertaining to fold change direction (up- or -regulation) of individual protein groups and to broader functions - were observed when the data was analysed in this way. This 'strength in numbers' approach revealed biologically significant phenomena, as detailed below.

The observation that trends over time were almost identical for un-normalised class-mapped protein groups and class-mapped protein groups normalised to the full list of protein groups identified (**Figure 19a** and **b**) indicates that these trends were not merely a function of the number of protein groups identified by the MS instrument per time point (i.e. technical variability). As mentioned in the results section, trends were largely maintained when class-mapped protein groups were normalised to the number of differentially expressed protein groups for a particular time point, except at time point 3 for some protein class trends. For these few, there was a downward trend between time points 2 and 3 before normalisation, and a 'u'-shaped trend after normalisation (**Figure 19c**). This was most probably due to a decrease in the number of differentially expressed protein groups observed over time (**Table 3** and **Figure 21**). This suggests that although there were progressively fewer

protein groups that mapped to some protein classes (in absolute numbers) over time (**Figure 19b**, most notably *nucleic acid binding*), the trend - with respect to the *proportion* of protein groups per protein class compared to the number of differentially expressed protein groups per time point (**Figure 19c**) - represents a *specific* process, in which the proportion is meaningful. This is in opposition to a situation in which there is a decrease in the absolute number of protein groups over time, but the proportion remains constant, which would suggest a non-specific process.

Trends of the abovementioned 'proportion' may then be combined with data concerning up or down-regulated expression (log_2 fold change) of protein groups belonging to the protein classes of interest, in order to further elucidate the biological effects of HIV-Tat treatment on neurons over time, as discussed. In terms of further analyses, the protein classes *nucleic acid binding* and *cytoskeletal protein* were looked at in closer detail (**Figures 19** and **20**). While these protein classes themselves were not necessarily significantly over/under represented (compared to background), many protein classes functionally linked to these classes - via Panther class parent lineage - were over/under represented and occurred in the top 10 Panther protein class mappings for over/under representation analysis ($p \leq 0.06$ - data dependant arbitrary cut-off for significance, individual p values quoted at relevant points in the text below; **Tables 7-9**). Further, as this was a protein-centric approach, log_2 fold changes in expression on a per protein group basis, or per group of protein groups, were important to take into account: if a collection of protein groups, which all map to the same protein class, showed the same effect in terms of up or down-regulation in response to treatment at a particular time point, then this phenomenon was of interest. It points to the effect of HIV-Tat treatment on the 'direction' (up or down) of regulation of a cellular function. As such, trends pertaining to the protein classes *nucleic acid binding* and *cytoskeletal protein* were analysed in greater detail (see below).

HIV-Tat affects differentiated neurons at the level of gene expression machinery

One may argue that, theoretically, the *nucleic acid binding* class would always end up having the largest proportion of protein groups mapped to it [in comparison to other

protein classes (**Figure 19**)], as there are more RNA and DNA binding proteins than all other proteins in the mammalian cell, including brain cells (Gerstberger et al. 2014). Conceptually similarly, one may also argue that there should always be more *RNA binding* than *DNA binding* protein groups in a dataset (**Figure 20a**), since there are more RNA binding proteins than transcription factors in mammalian cells (Gerstberger et al. 2014). Indeed, as mentioned, the protein class *nucleic acid binding* itself was not significantly over/under represented when compared to background (**Tables 7-9**). However, significantly over/under represented protein classes that were functionally linked - and that shared the same parent lineage to this broad class - were apparent. For example: the class *RNA methyltransferase (PO00033)* was over-represented 5-fold at 6 hours post treatment (p = 5.52e-3; **Table 7**); the class *DNA binding protein (PO00009)* was under-represented less than 0.2-fold at 24 hours post treatment (p = 2.47e-2; **Table 8**); and the class *DNA binding protein (PO00009)* was over-represented 2.34-fold at 48 hours post treatment (p = 6.08e-2; **Table B9**).

In addition, over/under representation data correlate well with expression change trends (log_2 fold change). That is to say, an over representation of a protein class functionally linked to *nucleic acid binding* was associated with up-regulation of the specific protein groups that map to *nucleic acid binding* (**Tables B1-3**, Appendix B; **Tables 7-9**). The combined weight of these evidences makes it unlikely that the size of this category biased the trend in *nucleic acid binding* protein groups observed over time. Perhaps the absolute number of protein groups that mapped to this protein class was not as vastly different to the other classes as was suggested by the figures (**Figure 19**), but the *trend*, particularly in terms of the change in 'direction' of expression over time, appeared to be significant.

In further elucidation of the nature of this trend, it was apparent that a higher proportion of protein groups were mapped to *nucleic acid binding* at 6 and 48 hours post HIV-Tat treatment, compared to 24 hours post HIV-Tat treatment (**Figure 19c**). This was maintained at both the DNA and RNA levels (**Figure 20a**), indicating an effect on both transcription and translation machinery. Further, this trend was in agreement with another overall trend in which expression of the differentially expressed protein groups that mapped to *nucleic acid binding* were up-regulated at 6 hours post treatment, followed by down-regulation at 24 hours, and finally up-regulation at 48 hours (**Tables B1-3**, Appendix B). The observation that many protein groups with a similar log_2 fold change sign mapped to a particular protein class at a

particular time point, increases confidence in the observed effects of HIV-Tat treatment.

The observed 'u-shaped' trend in 'nucleic acid binding' over time could be explained in terms of a 'competition' between HIV-Tat and the host

The above hypothesis can be explained as follows:
Initially (as measured at 6 hours post treatment), HIV-Tat directly up-regulates expression of protein groups with *nucleic acid binding* activity either directly, through HIV-Tat's transcription factor activity, or indirectly, by HIV-Tat's activation of other proteins, such as host transcription factors. This is plausible given HIV-Tat's role as a transactivator of transcription for the HIV genome, and its ability to bind many human genome loci and to recruit and bind many host factors (Gatignol 2007; Marban et al. 2011; Gautier et al. 2009; Debaisieux et al. 2012). Moreover, this suggests that HIV-Tat's inherent ability to alter gene expression may extend further than just the HIV genome (i.e. to the human genome), as has previously suggested by the effect of HIV-Tat on expression of specific host factors (Flores et al. 1993; Buonaguro et al. 1994). It would be interesting to use a ChIP-seq approach in the current cell line to determine genome-wide binding sites of HIV-Tat in non-transformed neurons, which may aid in determining the biological processes that HIV-Tat dysregulates - to contribute to HAND pathogenesis - by analysing genes encoded downstream of these binding sites. *Nucleic acid binding* protein groups are presumably targeted by HIV-Tat, as these proteins comprise the transcriptional and translational machinery of the cell, much of which is recruited by HIV-Tat for HIV genome replication.

This initial up-regulation of 'gene expression machinery' protein groups was followed by an overall down-regulation of 'gene expression machinery' protein groups - as well as fewer of these proteins differentially expressed - at 24 hours post treatment. This, to extend the hypothesis, perhaps indicates a host response attempting to curb HIV-Tat's effects and regain host cell homeostasis. Then, as HIV-Tat degrades in the cell as per its half-life and/or as it is actively cleared (both of which are not well documented in the literature), gene expression machinery proteins may increase again to compensate for previous decreased functioning. This can be seen at 48 hours post treatment where there was an increase in the proportion of protein groups mapped to this protein class (**Figure 19c**), as well as an up-regulation in

expression (log_2 fold change) of the corresponding individual protein groups (**Table B3**, Appendix B). One could further hypothesise that gene expression machinery proteins will no longer be differentially expressed and will reach baseline/control levels at time points beyond 48 hours, assuming no additional active HIV-Tat is added to the system.

Alternatively, the up-regulation of gene expression machinery at 48 hours post treatment may be a second, slower response that is perhaps only apparent after 48 hours. This points to the importance of extending the time course shown in the current work in order to understand longer-term effects of HIV-Tat and cell recovery post-exposure (see section **8. Future work**). If this second, related or unrelated, response does continue long-term, one might draw parallels between the trends in *nucleic acid binding* in HIV-Tat treated neurons and other biphasic biological responses in nature, such as the *Mycobacterium tuberculosis* (Mtb) hypoxic response (**Figure 41**). Under oxidative stress, Mtb initially up-regulates Dos regulon (stress response) genes followed by a cessation of this response after 24 hours (Rustad et al. 2008). This is followed by another, longer and enduring response, termed the enduring hypoxic response (Rustad et al. 2008). There is little overlap in the genes involved in the two waves of the response, similarly to what is seen in the current work (**Figures 21**). It is therefore plausible that a similar biphasic response mechanism is occurring within the gene expression machinery of HIV-Tat treated neurons (**Figure 41**).

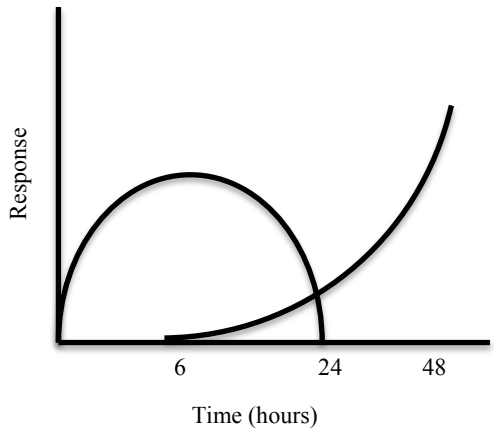


Figure 41. Changes to *nucleic acid binding* protein groups over time, in HIV-Tat treated neurons, may follow a biphasic trend The time points on the diagram (6, 24, and 48 hours) represent time points 1, 2, and 3 for the both the higher and lower cell confluence time course experiment in the current work.

Dysregulation of cytoskeletal maintenance proteins by HIV-Tat may be the molecular mechanism of synaptodendritic injury seen in HAND patients

The observation that proportionally and progressively more actin related proteins were differentially expressed over time (**Figure 20b**) was in agreement with continued repression of these protein groups (i.e. negative $\log_2$ fold change expression values at all 3 time points; **Tables B4-6**, Appendix B) and suggest that the effect on cytoskeletal instability was sustained, perhaps strengthened over time. This observation was supported by over/under representation analysis, in which, for example, the protein class *extracellular matrix protein* (PO00102) - involved in synaptic plasticity (Frischknecht & Gundelfinger 2012) and physically linked to the actin cytoskeleton via cell adhesion molecules (Geiger et al. 2001) - was significantly over represented at 48 hours post treatment ($p = 2.26e\text{-}2$) (**Table 9**). It is worth clarifying that one may observe *down*-regulation (in terms of $\log_2$ fold change) of a particular set of protein groups, but *over* representation of the protein class - to which the set of protein groups belong - in comparison to the background proteome (i.e. Fischer's exact test). This indicates that the implicated protein class (or broader function) was targeted by the treatment - specifically, that the treatment down-regulated many of the individual members of this protein class.

As the cytoskeleton is crucial for synaptic maintenance (Lin & Koleske 2010), and therefore neuronal communication (Goellner & Aberle 2012), dysregulation of cytoskeletal maintenance proteins - as seen in the current work after HIV-Tat treatment - would affect synapses and neuronal processes. The effect of HIV-Tat on the cytoskeleton may therefore explain/contribute to the clinical phenotype of structural brain alterations observed in HAND patients (Ragin et al. 2012). As examples of *in vitro* support for this line of reasoning: Kim et al. (2008), found that rat hippocampal neurons treated with HIV-Tat *in vitro* underwent synaptic degradation; and HIV-Tat treatment has been shown to cause cytoskeletal dysregulation in T lymphocytes (Lopez-Huertas et al. 2010). In addition, experimental evidence showing that HIV-Tat treatment caused grey matter deficits and impaired learning and memory in mice (Carey et al. 2012; Carey et al. 2013) further supports the argument that HIV-Tat detrimentally affects the synapses, which are widely known to be crucial in learning and memory (Mayford et al. 2012). Indeed, many other neurological conditions, such as mental retardation and spastic

paraplegias, have been linked to cytoskeletal damage (Goellner & Aberle 2012). It is therefore plausible that something similar is happening during the molecular pathogenesis of HAND.

Interestingly, cytoskeletal alterations have been associated with HIV-infected-monocyte interactions with the BBB (Woollard et al. 2014) possibly aiding entry of infected monocytes into the central nervous system. It is less clear whether cytoskeletal dysregulation affords any advantages - in terms of viral propagation - in the context of HIV-Tat infected neurons, or whether it is a 'carry over' from an evolved effect in infected monocytes (or other HIV infected cells).

Here, we suggest, for the first time (to our knowledge) that synaptic degradation in HAND may be due to down-regulation of cytoskeletal maintenance proteins, and thus cytoskeleton instability (**Figure 42**). It is important to note that damage to the synapses has been shown to be reversible (Ellis et al. 2007; Ragin et al. 2012); therefore it may be beneficial to focus on prophylactic drug development or early biomarker identification in order to address symptoms prior to the occurrence of irreversible damage.

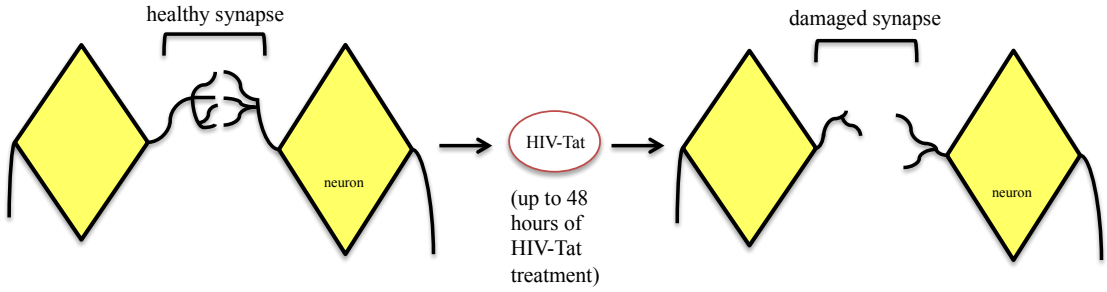


Figure 42. Down-regulation of cytoskeletal maintenance proteins by HIV-Tat may be the molecular correlates of synaptic degradation in HAND patients Schematic diagram of hypothesised HIV-Tat effects on the neuronal cytoskeleton after up to 48 hours of HIV-Tat treatment.

There is no evidence of dysregulated calcium metabolism at 6 hours post HIV-Tat treatment

As mentioned, the 6-hour time point was selected in an effort to identify proteomic signatures that may support literature relating to calcium influx (and therefore neuronal excitotoxicity) as a result of HIV-Tat treatment (Krogh et al. 2014). Although it has previously been shown that there is a large calcium influx in response

to HIV-Tat treatment up to 8 hours after treatment, followed by a cessation of this effect (Krogh et al. 2014), no such evidence was found in the current work. This may be due the time point chosen; perhaps 6 hours post HIV-Tat treatment is not optimal to observe this effect. Alternatively, it may be due to the fact that calcium influx occurs as the result of cellular processes not 'visible' to the current approach, which only measured protein expression changes (see **7. Limitations**). Such a cellular process could, for example, relate to altered protein activity due to post-translational modifications, which this experiment was not designed to consider. It is also plausible that differences exist between the neuronal cell model used in the above-mentioned study and in the current work.

Protein groups differentially expressed at more than one time point may highlight key points/pathways along the progression of HIV-Tat's effects

Little overlap in protein groups identified at progressively later time points post treatment may demonstrate the progression in cell response to HIV-Tat treatment over time (**Figure 21b**). Further, the protein groups that *were* differentially expressed at more than one time point may highlight particular biological functions that were affected by HIV-Tat treatment (**Tables 10-12**). As mentioned, all peptides used to identify a particular protein group in the current work had low PEP scores (< 0.01), indicating confidence in the assignment of a particular mass spectrum to a particular tryptic peptide identity (**Tables B1-6**, Appendix B), and thus confidence in making hypotheses about particular protein groups, or clusters of protein groups, identified in the current experiments.

As an example of protein groups expressed at more than one time point, many *nucleic acid binding* class protein groups were up-regulated at 6 hours post treatment, and subsequently down-regulated at 24 hours post treatment (**Table 10**). This was consistent with functional classification into Panther protein classes, which indicated that protein groups that mapped to *nucleic acid binding* were largely up-regulated at 6 hours post treatment, and subsequently down-regulated at 24 hours post treatment (**Tables B1-2**, Appendix B). The observation that many individual protein groups followed this trend in expression over time further strengthens the hypothesis that HIV-Tat treatment causes an initial up-regulation and subsequent down-regulation of gene expression at 6 and 24 hours post treatment, respectively. Perhaps, therefore,

there is something important about the functions of these particular protein groups. As an example, MRPL45 encodes the 39s mitochondrial ribosome subunit (GeneCards) (**Table 10**). Down-regulation of this gene at 24 hours post treatment would have a significant effect on translation of mitochondrial genes and perhaps cell energy metabolism.

In addition, some protein groups were differentially expressed at both 6 *and* 48 hours post treatment (CO3A1, CTNA1, SNX27). As these protein groups were all present in the 'total list' for the 24-hour time point (raw data not shown), this phenomenon cannot be explained by these protein groups going undetected by the MS instrument at 24 hours post treatment (**Figure 21a**). Rather they were not *significantly* differentially expressed between treated and control conditions at 24 hours post treatment. It is therefore noteworthy that these proteins were significantly differentially expressed at 6 hours and then again at 48 hours post HIV-Tat treatment, implying that this was not a prolonged effect over 48 hours, but that the same gene expression changes occurred *again* at a second time point. This recurrence of the effect was supported by the observation that the same 'direction' (up or down) of expression regulation ($\log_2$ fold change) occurred at the second time point. The genes that encode these protein groups are not located on the same chromosome (GeneCards), so the trend cannot be explained by co-expression due to proximity of genetic loci. It is interesting to note that all protein groups differentially expressed at both 6 and 48 hours post HIV-Tat treatment were cytoskeletal/membrane associated proteins (**Table 12**).

Without careful experimental research into each one of these protein groups' function in a neuron, it is difficult to elucidate how they might be targeted by HIV-Tat treatment specifically, and what the downstream biological effects of this might be. Therefore, the protein groups that are differentially expressed at more than one time point are interesting candidates for follow-up validation work, as possible targets for the diagnosis/treatment of HAND and give further evidence for the molecular pathways affected by HIV-Tat treatment (see **8. Future Work**).

6.4 24-hour time point repeat

The 24-hour time point repeat largely supported findings from the time course experiment

Results from the 24-hour time point repeat experiment further strengthened the evidence for down-regulation of cytoskeletal maintenance processes as the result of HIV-Tat treatment (**Figure 29**, **Table 14**). Surprisingly, though, while many protein groups mapped to the *nucleic acid binding* class in this repeat dataset, the fold change expression differences were not in agreement across the two datasets. In fact, the 24-hour time point repeat data correlated better with the 6-hour than the 24-hour time point (HT2) from the time course experiment.

An explanation for this discrepancy may be as follows: the 24 hour time point repeat was a standalone experiment, whilst HT2 was harvested and processed in the context of additional synchronised time points and therefore, it is conceivable that these two 'identical' time points are in fact out of phase. As the time course experiment was synchronised, it is more likely that experimental conditions were similar across all time points and therefore it appears more meaningful to analyse differences between time points within the time course experiment than to compare across different isolated experiments.

The observation that cytoskeletal maintenance proteins are down-regulated at all three time points of the time course experiment could explain why down-regulation of this cellular process was also observed for the repeat experiment, even though the repeat did not correspond exactly to one of the time points. In fact, this supports the hypothesis that HIV-Tat treatment is associated with sustained dysregulation of cytoskeletal maintenance, at least up to 48 hours post treatment. These findings highlight the importance of looking at a treatment/disease progression profile *over time*, to obtain a better sense of overall trends.

6.5 Time course experiment with NES cells differentiated at 70% cell confluence

NES cells differentiated at 70% cell confluence showed less HIV-Tat mediated cytoskeletal dysregulation than NES cells differentiated at 90% cell confluence

The time course experiment was repeated with cells differentiated at lower confluence (70%), in order to determine the effect of cell density on HIV-Tat treatment effects. We hypothesised that cells differentiated at 70% confluence would have fewer synaptic connections than cells differentiated at higher confluence (90%). Thus, one might expect to observe fewer effects to the cytoskeleton (a major component of synaptic formation and maintenance) after HIV-Tat treatment of cells differentiated at lower confluence. We further hypothesised that cell confluence should not result in alterations to the gene expression machinery effects of HIV-Tat treatment (as seen with cells differentiated at 90% confluence), as this is an *intra*cellular process.

Biological significance analysis

As expected, trends were comparable across the two different cell confluences in terms of the proportion of protein groups that mapped to the protein class *cytoskeletal protein* across time (**Figure 40c**). However, also as expected, cells differentiated at 70% confluence showed no significant over/under representation effects on the cytoskeleton upon HIV-Tat treatment in comparison to the background proteome (**Tables 18-20**). Therefore, it appears that an effect to the cytoskeleton did occur, but was not strong enough to meet threshold requirements for statistically significant over representation (in comparison to the background) at lower synaptic densities. One might hypothesise that if cells are too dispersed, fewer synaptic connections would form, resulting in less synaptic activity and, as a result, a 'smaller' down-regulation of the implicated protein groups in response to HIV-Tat (in comparison to cells with many synaptic connections and high synaptic activity). This provides support for the notion that HIV-Tat affects the synaptic processes of neurons (Ragin et al. 2012; Kim et al. 2008) - as suggested from the statistically significant effect of HIV-Tat on the cytoskeleton observed with cells differentiated at 90% confluence - and implicates the cytoskeleton as a major molecular structure that becomes dysregulated to cause this effect.

Again, as expected, the effect of HIV-Tat on *nucleic acid binding* cellular functioning was comparable across both cell confluences (**Figure 40c**). Furthermore, this effect was *significant* for cells differentiated and treated at lower confluence (**Tables 18-20**). For example, protein groups involved in chromatin condensation were down-regulated at 6 hours post treatment for both cell confluences, which would

presumably facilitate chromatin relaxation and allow for increased viral genome transcription (Gatignol 2007). As mentioned, it is plausible that changes to *nucleic acid binding* upon HIV-Tat treatment would not be affected by cell confluence, as this class of protein groups is largely involved with *intra*cellular processes. This is in comparison to cytoskeletal and synaptic maintenance, which may very well be affected by the number of synapses present and consequent levels of communication between cells, as described above.

One discrepancy between the different cell confluence experiments was that protein groups associated with mRNA processing were down-regulated at 6 hours post treatment for the lower cell confluence experiment (**Table D4**, Appendix D), while they were up regulated at 6 hours post treatment for the higher cell confluence experiment (**Table B1**, Appendix B). This effect - for the lower cell confluence experiment - was not intuitive, especially given the central role of mRNA processing in HIV protein expression (Hryckiewicz et al. 2011). It is unclear why this effect was observed.

One might argue that the HIV-Tat-concentration:cell-count ratio could explain differences between time course experiments at different cell confluences. However, if one calculates the number of HIV-Tat molecules present per cell, it is apparent that HIV-Tat was applied at a concentration in gross excess to the number of cells, even at higher cell confluence. For example, if 2ml medium containing 1ng/µl HIV-Tat (HIV-Tat ± 10kDa) - which corresponds to 200pmol HIV-Tat - was added to 2 million cells, then there would be 1×10^{13} molecules of HIV-Tat in the medium, corresponding to 5×10^{6} HIV-Tat molecules per cell. Further, HIV-Tat's affinity for certain cell surface receptors also ensures that only a certain number of HIV-Tat molecules enter the cell at any one time (Richard et al. 2005; Herce & Garcia 2007). Therefore, it is unlikely that overall cell count will affect the amount of active HIV-Tat able to enter neurons.

7. Limitations

Limitations of every research approach must be acknowledged if future work is to be improved upon.

MS-based proteomics only reveals changes in de novo *protein expression*

The particular experimental technique used in the current work only reveals changes to protein expression and does not take into account other mechanisms that may modulate protein expression and/or function. Importantly, for example, protein activation and function should be understood in terms of post-translational modification (PTM) status. As phosphorylation is the most ubiquitous PTM (Macek et al. 2009), it would be of interest to examine the phosphoproteome of HIV-Tat treated differentiated NES cells. In doing so, another layer of biological meaning would be added to the current data, perhaps revealing affected pathways that would otherwise be overlooked.

The in vivo *concentration of HIV-Tat remains unknown*

While the concentration of HIV-Tat used here is an accepted estimate, determining the true *in vivo* CNS concentration range of HIV-Tat would allow for a more realistic cell culture model of HAND (Xiao et al. 2000; Perry et al. 2010). This limitation may be addressed by determining the HIV-Tat concentration in cerebrospinal fluid (CSF) from HAND patients, rather than the serum measurement currently used. This may be achieved through an enzyme-linked immunosorbent assay (ELISA) or by targeted MS analysis of CSF samples (Perry et al. 2010). Discovery proteomics may also generate interesting data regarding the proteomic composition of CSF from individuals with HAND, and may facilitate identification of candidate diagnostic biomarkers corresponding to markers identified in the secretome of cell culture models (e.g. HIV-Tat treated differentiated NES cells.) As the BBB is compromised during HAND progression, it is conceivable that HIV-Tat stimulated markers may be present in the CSF (Roberts et al. 2012).

HIV-Tat treatment of a mono-neuronal cell population does not reveal the 'full picture' of HIV-Tat's effect in the brain

HIV-Tat is known to affect many brain cell types, and as such, interactions between affected cells must be considered to truly make deductions about the effect of HIV-Tat in the CNS *in vivo*. Notably, HIV infected glia produce a cytokine storm, which produces a significant indirect effect on neurons (González-Scarano & Martín-García 2005). A neuronal-like stem cell line that spontaneously differentiates into mixed

neuronal and glial populations when growth factors are removed from the medium is available in the Blackburn Laboratory (Sun et al. 2008), and may be used in future aspects of this project in order to determine the interactions between different HIV-Tat affected cell types.

8. Future research

Additional experimental time points would allow long-term modelling of the effect of HIV-Tat on neurons

The current time course experiment examined the effect of HIV-Tat on neurons at 6 hours after initial treatment and up to 48 hours post treatment. Findings alluded to a particularly interesting biphasic response progression for protein groups that mapped to Panther protein class *nucleic acid binding* (**Figure 41**). A longer time course would reveal the effect of HIV-Tat on neurons at progressively later time points and would aid in further elucidating the long-term nature of the response. As differentiated NES cell cultures have been maintained for up to four months (Garnett, unpublished), and as neurocognitive symptoms may become apparent within the first year of infection (Ragin et al. 2012), a realistic model of the temporal progression of HAND can potentially now be established.

Validation studies for identified protein groups/pathways of interest

As with all protein groups identified by discovery proteomics, observations should be confirmed by orthogonal approaches, such as Western blot analysis or a targeted MS approach. Such studies increase confidence in the differential expression of a particular protein as the result of treatment as well as its contribution to a particular molecular pathway of interest. Due to the protein groups problem (Nesvizhskii & Aebersold 2005), such studies may also identify which particular protein isoform is differentially expressed as the result of treatment.

Of particular interest are those protein groups that are differentially expressed at more than one time point, as they allude to the temporal progression of the response (**Tables 10-12**). In addition, gene expression machinery proteins and cytoskeleton maintenance proteins are also of interest and should be studies in greater detail in the

context of HIV-Tat treated neurons. Perhaps proteins identified through this detailed approach - or protein groups differentially expressed at more than one time point - could serve as biomarkers for the early progression of HAND (if validated and present in easily accessible clinical samples).

Electrophysiological studies of HIV-Tat treated neurons would correlate molecular pathology and cell function

Further to the validation of particular protein groups/pathways of interest, additional experimental approaches may also increase confidence in findings from proteomic experiments. For example, electrophysiological studies of HIV-Tat treated neurons over time would allow one to correlate the molecular pathology of HAND (at the level of differential protein expression) to cell function changes. In doing so, functional consequences of protein expression and PTM changes may be established. As these cells are able to generate action potentials and display other neuronal characteristics (**Figure 7**), such an experiment is achievable.

9. Conclusions

In the current work, longitudinal proteomic analysis of *in vitro* HIV-Tat treated human neurons (differentiated from an NES cell line) showed that the Panther protein class *nucleic acid binding* and *cytoskeletal protein* were significantly affected by HIV-Tat treatment for up to 48 hours post treatment. This suggests that HIV-Tat treatment affected the gene expression machinery of the cell, as well as cytoskeletal stability, as was suggested from the individual functions of the protein groups that mapped to the *nucleic acid binding* and *cytoskeletal protein* classes, respectively.

Changes to the gene expression machinery suggest that HIV-Tat is involved in preparing the host cell for increased transcriptional and translational activity. One may hypothesise that the observed trend of protein expression (for protein groups that mapped to the Panther protein class *nucleic acid binding*) could be explained in terms of a ‘competition’ between HIV-Tat and the host, in which HIV-Tat is derailing the status quo and the host is trying to maintain/restore the status quo.

As the synapse is an intercellular structure degraded in HAND patients and as the cytoskeleton is a major component of the neuronal synapse, damage to the synapse may be a correlate of cytoskeletal damage upon HIV-Tat treatment.

Therefore, cytoskeletal stability may be linked to synaptic integrity. To our knowledge, this is a novel association in HAND research and may be of interest in determining molecular biomarkers for HAND. The effect to cytoskeletal proteins was evident, but less prominent, in cells at low cell confluence, which suggests that sparsely plated neurons have fewer synaptic connections and provides support for the hypothesis that HIV-Tat is affecting synaptic processes.

Further, protein groups differentially expressed in response to HIV-Tat treatment at more than one time point may represent important molecular switches and are good targets for follow up validation work in the identification of molecular biomarkers for the characterisation and/or treatment of HAND. The cell culture model of HAND established in this work has important implications for HAND research in the diagnosis and treatment of this common HIV comorbidity.

10. References

Achim, C.L. et al., 2011. Increased accumulation of intraneuronal amyloid β in HIV infected patients. *Journal of neuroimmune pharmacology*, 4(2), pp.190–199.

Aebersold, R. & Mann, M., 2003. Mass spectrometry-based proteomics. *Nature*, 422(6928), pp.198–207.

Ances, B.M. et al., 2010. HIV infection and aging independently affect brain function as measured by functional magnetic resonance imaging. *The Journal of infectious diseases*, 201(3), pp.336–340.

Antinori, A. et al., 2007. Updated research nosology for HIV-associated neurocognitive disorders. *Neurology*, 69(18), pp.1789–1799.

Bantscheff, M. et al., 2007. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry*, 389(4), pp.1017–1031.

Bellizzi, M.J. et al., 2005. Synaptic activity becomes excitotoxic in neurons exposed to elevated levels of platelet-activating factor. *The Journal of clinical investigation*, 115(11), pp.3185–3192.

Bondarenko, P.V., Chelius, D. & Shaler, T.A., 2002. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography - tandem mass spectrometry. *Analytical chemistry*, 74(18), pp.4741–4749.

Brew, B.J., 2004. Evidence for a change in AIDS dementia complex in the era of highly active antiretroviral therapy and the possibility of new forms of AIDS dementia complex. *AIDS*, 18(supp 1), pp.S75–S78.

Brown, K.M. & Tracy, D.K., 2013. Lithium: the pharmacodynamic actions of the amazing ion. *Therapeutic advances in psychopharmacology*, 3(3), pp.163–176.

Brun, V. et al., 2007. Isotope-labeled protein standards. *Molecular & cellular proteomics : MCP*, 6(12), pp.2139–2149.

Buonaguro, L. et al., 1994. The human immunodeficiency virus type 1 Tat protein transactivates tumor necrosis factor beta gene expression through a TAR-like structure. *Journal of virology*, 68(4), pp.2677–2682.

Carey, A.N. et al., 2013. Conditional Tat protein expression in the GT-tg bigenic mouse brain induces gray matter density reductions. *Progress in neuro-psychopharmacology & biological psychiatry*, 43, pp.49–54.

Carey, A.N. et al., 2012. Expression of HIV-Tat protein is associated with learning and memory deficits in the mouse. *Behavioural brain research*, 229(1), pp.48–56.

Cartier, L. et al., 2003. Chemokine-induced cell death in CCR5-expressing neuroblastoma cells. *Journal of neuroimmunology*, 145(1-2), pp.27–39.

Cox, J. et al., 2009. A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nature protocols*, 4(5), pp.698–705.

Cox, J. et al., 2014. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction , termed MaxLFQ. *Molecular & cellular proteomics : MCP*, 13(9), pp.2513–2526.

Cox, J. et al., 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research*, 10(4), pp.1794–1805.

Danovi, D. et al., 2012. *High content screening of defined chemical libraries using normal and glioma-derived neural stem cell lines.* 1st ed., Elsevier Inc.

Debaisieux, S. et al., 2012. The ins and outs of HIV-1 Tat. *Traffic*, 13(3), pp.355–363.

Department of Health, 2013. *The South African antiretroviral treatment guidelines*. Pretoria: Department of Health.

Department of Health, 2015. *Mid-year population estimates*. Statistics South Africa. Pretoria: Department of Health.

Domon, B. & Aebersold, R., 2010. Options and considerations when selecting a quantitative proteomics strategy. *Nature biotechnology*, 28(7), pp.710–721.

Donnelly, M.R. & Ciborowski, P., 2015. Proteomics, biomarkers, and HIV-1: A current perspective. *PROTEOMICS - Clinical applications*, 1, pp.1–16.

Duchardt, F. et al., 2007. A comprehensive model for the cellular uptake of cationic cell-penetrating peptides. *Traffic*, 8(18), pp.848–866.

Duncan, M.W., Aebersold, R. & Caprioli, R.M., 2010. The pros and cons of peptide-centric proteomics. *Nature biotechnology*, 28(7), pp.659–664.

Ellis, R., Langford, D. & Masliah, E., 2007. REVIEWS HIV and antiretroviral therapy in the brain : neuronal injury and repair. *Nature reviews neuroscience*, 8, pp.33-44.

Eng, J.K., Mccormack, A.L. & Yates, J.R., 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *American journal for mass spectrometry*, 5, pp.976–989.

Eugenin, E.A. et al., 2007. HIV-tat induces formation of an LRP-PSD-95- NMDAR-nNOS complex that promotes apoptosis in neurons and astrocytes. *Proceedings of the National Academy of Sciences of the United States of America: PNAS*, 104(9), pp.3438–3443.

Everall, I.P. et al., 1999. Cortical synaptic density is reduced in mild to moderate human immunodeficiency virus neurocognitive disorder. *Brain pathology*, 9, pp.209–217.

Fischer-Smith, T. et al., 2001. CNS invasion by CD14+/CD16+ peripheral blood-derived monocytes in HIV dementia: perivascular accumulation and reservoir of HIV infection. *Journal of neurovirology*, 7(6), pp.528–541.

Flores, S.C. et al., 1993. Tat protein of human immunodeficiency virus type 1 represses expression of manganese superoxide dismutase in HeLa cells. *Proceedings of the National Academy of Sciences of the United States of America: PNAS*, 90(August), pp.7632–7636.

Forde, J.E. & Dale, T.C., 2007. Glycogen synthase kinase 3: a key regulator of cellular fate. *Cellular and molecular life sciences : CMLS*, 64(15), pp.1930–1944.

Frischknecht, R. & Gundelfinger, E.D., 2012. The brain's extracellular matrix and its role in synaptic plasticity. *Advanced experimental medical biology*, 970, pp. 153-171.

Gatignol, A., 2007. Transcription of HIV: Tat and cellular chromatin. *Advances in pharmacology*, 55(07), pp.137–159.

Gautier, V.W. et al., 2009. In vitro nuclear interactome of the HIV-1 Tat protein. *Retrovirology*, 6, p.47.

Geiger, B. et al., 2001. Transmembrane extracellular matrix– cytoskeleton crosstalk. *Nature reviews. Molecular cell biology*, 2(November), pp.793–805.

Gerstberger, S., Hafner, M. & Tuschl, T., 2014. A census of human RNA-binding proteins. *Nature reviews genetics*, 15(12), pp.829–845.

Goellner, B. & Aberle, H., 2012. The synaptic cytoskeleton in development and disease. *Developmental neurobiology*, 72, pp.111–125.

González-Scarano, F. & Martín-García, J., 2005. The neuropathogenesis of AIDS. *Nature reviews. Immunology*, 5(1), pp.69–81.

Haase, A.T., 1986. Pathogenesis of lentivirus infections. *Nature*, 332, pp.130–136.

Haddow, L.J. et al., 2013. A systematic review of the screening accuracy of the HIV dementia scale and international HIV dementia scale. *Public library of science*: *PLoS ONE*, 8(4).

Harris, M.A. et al., 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic acids research*, 32(Database issue), pp.D258–D261.

Herce, H.D. & Garcia, A.E., 2007. Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes. *Proceedings of the National Academy of Sciences of the United States of America: PNAS*, 104(52), pp.20805–20810.

Hodgkin, A., Huxley, A. & Katz, B., 1951. Measurement of current-voltage relations in the membrane of the giant axon of Loligo. *Journal of physiology*, 116, pp.424–448.

Holden, P. & Horton, W.A., 2009. Crude subcellular fractionation of cultured mammalian cell lines. *BioMed Central* (*BMC) research notes*, 2, p.243.

Holt, J.L., Kraft-Terry, S.D. & Chang, L., 2012. Neuroimaging studies of the aging HIV-1-infected brain. *Journal of neurovirology*, 18(4), pp.291–302.

Hryckiewicz, K. et al., 2011. HIV RNA splicing. *HIV & AIDS Review*, 10(3), pp.61–64.

De Jager, C.A. et al., 2015. Dementia in rural South Africa: a pressing need for epidemiological studies. *South African medical journal*, 105(3), p.189.

Joska, J., Westgarth-Taylor, J., Myer, L., et al., 2011. Characterization of HIV-associated neurocognitive disorders among individuals starting antiretroviral therapy in South Africa. *AIDS and behavior*, 15(6), pp.1197–1203.

Joska, J. et al., 2011. The neurobiology of HIV dementia: implications for practice in South Africa. *African journal of psychiatry*, 14(1), pp.17–22.

Joska, J., Westgarth-Taylor, J., Hoare, J., et al., 2011. Validity of the international HIV dementia scale in South Africa. *AIDS patient care and STDs*, 25(2), pp.95–101.

Käll, L. et al., 2008. Posterior error probabilities and false discovery rates : two sides of the same coin. *Journal of proteome research*, 7, pp.40–44.

Kandathil, A.J. et al., 2009. Bioinformation amino acid sequence divergence of Tat protein (exon1) of subtype B and C HIV-1 strains : does it have implications for vaccine development? *Bioinformation*, 2063(6), pp.237–241.

Kim, H.J., Martemyanov, K.A. & Thayer, S.A., 2008. Human immunodeficiency virus protein Tat induces synapse loss via a reversible process that is distinct from cell death. *The journal of neuroscience*, 28(48), pp.12604–12613.

King, J.E. et al., 2010. Mechanisms of HIV-Tat-induced phosphorylation of N-methyl-D-aspartate receptor subunit 2A in human primary neurons: implications for neuroAIDS pathogenesis. *The American journal of pathology*, 176(6), pp.2819–2830.

Konermann, L. et al., 2013. Unraveling the mechanism of electrospray ionisation. *Analytical chemistry*, 85, pp.2–9.

Krogh, K.A. et al., 2014. HIV-1 protein Tat produces biphasic changes in NMDA-evoked increases in intracelluar Ca2+ concentration via activation of Src kinase and nitric oxide signaling pathways. *Journal of neurochemistry*, 130(5), pp.642–656.

Laspiura, J.P. et al., 2007. CSF proteomic fingerprints for HIV-associated cognitive impairment. *Journal of neuroimmunology*, 192(1-2), pp.157–170.

Letendre, S. et al., 2008. Validation of the CNS penetration-effectiveness rank for quantifying antiretroviral penetration into the central nervous system. *Archives of neurology*, 65(1), pp.65–70.

Letendre, S.L. et al., 2006. Lithium improves HIV-associated neurocognitive impairment. *AIDS*, 20(14), pp.1885–1888.

Levine, A.J. et al., 2013. Systems analysis of human brain gene expression: mechanisms for HIV-associated neurocognitive impairment and common pathways with Alzheimer's disease. *BioMed central (BMC) medical genomics*, 6, p.4.

Lin, Y.-C. & Koleske, A.J., 2010. Mechanisms of synapse and dendrite maintenance and their disruption in psychiatric and neurodegenerative disorders. *Annunal review of neuroscience*, 33, pp.349–378.

Lopez-Huertas, M.R. et al., 2010. Modifications in host cell cytoskeleton structure and function mediated by intracellular HIV-1 Tat protein are greatly dependent on the second coding exon. *Nucleic acids research*, 38(10), pp.3287–3307.

Macek, B., Mann, M. & Olsen, J. V, 2009. Global and site-specific quantitative phosphoproteomics: principles and applications. *Annual review of pharmacology and toxicology*, 49, pp.199–221.

Maggirwar, S.B. et al., 1999. HIV-1 Tat-mediated activation of glycogen synthase kinase-3beta contributes to Tat-mediated neurotoxicity. *Journal of neurochemistry*, 73(2), pp.578–586.

Makarov, A., 2000. Electrostatic axially harmonic orbital trapping : A High-Performance Technique of Mass Analysis Electrostatic Axially Harmonic Orbital Trapping : A High-Performance Technique of Mass Analysis. *Analytical chemistry*, 72, pp.1156–1162.

Manji, H., Jäger, H.R. & Winston, A., 2013. HIV, dementia and antiretroviral drugs: 30 years of an epidemic. *Journal of neurology, neurosurgery, and psychiatry*, 84(10), pp.1126–1137.

Marban, C. et al., 2011. Genome-wide binding map of the HIV-1 Tat protein to the human genome. *Public library of science: PloS one*, 6(11), p.e26894.

Marra, C.M. et al., 2010. Impact of Combination Antiretroviral Therapy on Cerebrospinal Fluid HIV RNA and Neurocognitive Performance. *AIDS*, 23(11), pp.1359–1366.

Martín-García, J., Kolson, D.L. & González-Scarano, F., 2002. Chemokine receptors in the brain: their role in HIV infection and pathogenesis. *AIDS*, 16(13), pp.1709–1730.

Mayford, M., Siegelbaum, S. a & Kandel, E.R., 2012. Synapses and memory storage. *Cold Spring Harbor perspectives in biology*, 4(6).

Mi, H. et al., 2013. Large-scale gene function analysis with the PANTHER classification system. *Nature protocols*, 8(8), pp.1551–1566.

Mi, H., Muruganujan, A. & Thomas, P.D., 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic acids research*, 41(Database issue), pp.D377–386.

Miller, P.E. & Denton, M.B., 1986. The quadrupole mass filter : basic operating concepts. *Journal of chemical education*, 63(7), pp.617–623.

Moody, W.J. & Bosma, M.M., 2005. Ion channel development, spontaneous activity, and activity-dependent development in nerve and muscle cells. *Physiological reviews*, pp.883–941.

Na, H. et al., 2011. Interactions between human immunodeficiency virus (HIV)-1 Vpr expression and innate immunity influence neurovirulence. *Retrovirology*, 8(1), p.44.

Nath, A. et al., 2008. Evolution of HIV dementia with HIV infection. *International review of psychiatry*, 20(1), pp.25–31.

Nath, A., 2002. Human immunodeficiency virus (HIV) proteins in neuropathogenesis of HIV dementia. *The journal of infectious diseases*, 186 Suppl (Suppl 2), pp.S193–198.

Nesvizhskii, A.I. & Aebersold, R., 2005. Interpretation of shotgun proteomic data: the protein inference problem. *Molecular & cellular proteomics : MCP*, 4(10), pp.1419–1440.

Olsen, J. V et al., 2007. Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods*, 4(9), pp.709–712.

Ong, S.E. & Mann, M., 2006. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature protocols*, 1(6), pp.2650–2660.

Ott, M., Geyer, M. & Zhou, Q., 2011. The control of HIV transcription: keeping RNA polymerase II on track. *cell host and microbe*, 10(5), pp.426–435.

Peeters, M., Toure-Kane, C. & Nkengasong, J.N., 2003. Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. *AIDS*, 17(18), pp.2547–2560.

Penner R. 1995. A Practical Guide to Patch Clamping. In Single-Channel Recording, 2nd edn, ed. Sakmann B & Neher E, pp. 3-30. Springer, Berlin

Perkins, D. et al., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20, pp.3551–3567.

Perry, S.W. et al., 2010. Human immunodeficiency virus-1 Tat activates calpain proteases via the ryanodine receptor to enhance surface dopamine transporter levels and increase transporter-specific uptake and Vmax. *The journal of neuroscience*, 30(42), pp.14153–14164.

Pirmoradian, M. et al., 2013. Rapid and deep human proteome analysis by rapid and deep single-dimension shotgun proteomics. *American society for biochemistry and molecular biology*, pp.1–32.

Ragin, A.B. et al., 2012. Structural brain alterations can be detected early in HIV infection. *Neurology*, 79(24), pp.2328–2334.

Rao, V.R. et al., 2013. Clade C HIV-1 isolates circulating in Southern Africa exhibit a greater frequency of dicysteine motif-containing Tat variants than those in Southeast Asia and cause increased neurovirulence. *Retrovirology*, 10, p.61.

Rath, A. et al., 2009. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America: PNAS*, 106(6), pp.1760–1765.

Rempel, H.C. & Pulliam, L., 2005. HIV-1 Tat inhibits neprilysin and elevates amyloid beta. *AIDS*, 19(2), pp.127–135.

Richard, J.P. et al., 2005. Cellular uptake of unconjugated TAT peptide involves clathrin-dependent endocytosis and heparan sulfate receptors. *Journal of biological chemistry*, 280(15), pp.15300–15306.

Roberts, T.K. et al., 2012. CCL2 disrupts the adherens junction: implications for neuroinflammation. *Laboratory investigation*, 92(8), pp.1213–1233.

Robertson, K., Liner, J. & Meeker, R.B., 2013. Antiretroviral neurotoxicity. *Journal of neurovirology*, 18(5), pp.388–399.

Rodriguez-aller, M., 2010. Coupling ultra high-pressure liquid chromatography with mass spectrometry: constraints and possible applications. *Trends in analytical chemistry*, 29(1), pp.15–27.

Ross, P.L. et al., 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics: MCP*, 3(12), pp.1154–69.

Rustad, T.R. et al., 2008. The enduring hypoxic response of Mycobacterium tuberculosis. *Public library of science: PloS one*, 3(1), p.e1502.

Sacktor, N.C. et al., 2005. The international HIV dementia scale: a new rapid screening test for HIV dementia. *AIDS*, 19(13), pp.1367–1374.

Schwamborn, K. & Caprioli, R.M., 2010. MALDI imaging mass spectrometry - painting molecular pictures. *Molecular oncology*, 4(6), pp.529–538.

Simioni, S. et al., 2010. Cognitive dysfunction in HIV patients despite long-standing suppression of viremia. *AIDS*, 24(9), pp.1243–1250.

Song, M. et al., 2013. Coordinated development of voltage-gated Na and K + currents regulates functional maturation of forebrain neurons derived from Human induced pluripotent stem cells. *Stem cells and development*, 22(10), pp.1551–1563.

Steen, H. & Mann, M., 2004. The ABC's (and XYZ's) of peptide sequencing. *Nature reviews molecular cell biology*, 5(9), pp.699–711.

Sun, Y. et al., 2008. Long-term tripotent differentiation capacity of human neural stem (NS) cells in adherent culture. *Molecular and cellular neurosciences*, 38(2), pp.245–258.

Tahirov, T.H. et al., 2010. Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature*, 465(7299), pp.747–751.

Tailor, J. et al., 2013. Stem cells expanded from the human embryonic hindbrain stably retain regional specification and high neurogenic potency. *The journal of neuroscience*, 33(30), pp.12407–12422.

Telias, M., Segal, M. & Ben-Yosef, D., 2014. Electrical maturation of neurons derived from human embryonic stem cells. *F1000Research*, 196, pp.1–12.

Thomas, P.D. et al., 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome research*, 13(9), pp.2129–2141.

Tivarus, M.E. et al., 2015. Are structural changes induced by lithium in the HIV brain accompanied by changes in functional connectivity? *Public library of science: PloS one*, 10(10), pp.1–16.

Tosi, G. et al., 2000. Highly stable oligomerization forms of HIV-1 Tat detected by monoclonal antibodies and requirement of monomeric forms for the transactivating function on the HIV-1 LTR. *European Journal of Immunology*, pp.1120–1126.

UNAIDS, 2013. *Global Report: UNAIDS Report on the Global AIDS Epidemic 2013*. Geneva: UNAIDS.

Valcour.V, Shikuma C., Shiramizu B., Watters M., Poff P., Selnes O., Holck P, Grove, S.N., 2004. Higher frequence of dementia in older HIV-1 individuals: the Hawaii aging with HIV-1 cohort. *Neurology*, 63(5), pp.822–827.

Walther, T.C. & Mann, M., 2010. Mass spectrometry-based proteomics in cell biology. *The journal of cell biology*, 190(4), pp.491–500.

Wiederin, J. et al., 2009. Biomarkers of HIV-1 associated dementia: proteomic investigation of sera. *Proteome science*, 7(8).

Williams, D.W. et al., 2013. Mechanisms of HIV entry into the CNS: increased sensitivity of HIV infected CD14+CD16+ monocytes to CCL2 and key roles of CCR2, JAM-A, and ALCAM in diapedesis. *Public library of science: PloS one*, 8(7), p.e69270.

Wiśniewski, J.R. et al., 2012. Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Molecular systems biology*, 8, p.611.

Wiśniewski, J.R. et al., 2009. Universal sample preparation method for proteome analysis. *Nature methods*, 6(5), pp.359–362.

Woollard, S.M. et al., 2014. HIV-1 induces cytoskeletal alterations and Rac1 activation during monocyte-blood-brain barrier interactions: modulatory role of CCR5. *Retrovirology*, 11(1), p.20.

Xiao, H. et al., 2000. Selective CXCR4 antagonism by Tat: implications for in vivo expansion of coreceptor use by HIV-1. *Proceedings of the National Academy of Sciences of the United States of America: PNAS*, 97(21), pp.11466–11471.

Zemianek, J.M. et al., 2012. Accelerated establishment of mature signaling patterns following stimulation of developing neuronal networks: "learning" versus "plasticity". *International Journal of Developmental Neuroscience*, 30(7), pp.602–606.

Zhou, L. et al., 2010. First evidence of overlaps between HIV-associated dementia (HAD) and non-viral neurodegenerative diseases: proteomic analysis of the frontal cortex from HIV+ patients with and without dementia. *Molecular neurodegeneration*, 5, p.27.

11. Appendices

11.1 Appendix A

Selected recipes

Table A1. SDS-PAGE gel recipe

	Stacking gel (5%)	Separating gel (15%)
1M Tris/Cl, pH6.8 (ml)	0.625	-
1.5M Tris/Cl, pH8.8 (ml)	-	3.75
30% polyacrylamide (ml)	0.850	7.5
10% Sodium dodecyl sulphate (SDS) (ml)	0.050	0.15
10%Ammonium persulphate (APS) (ml)	0.050	0.15
Tetramethylethylenediamine (TEMED) (ml)	0.010	0.012
dH_2O	3.400	3.4
TOTAL volume (ml)	5	15

RIPA buffer recipe (adapted from Holden and Horton 2009, buffer 4)

150mM NaCl

100mM Tris HCl, pH 7.6

0.5% Sodium deoxycholate

1% SDS

(100mM DTT omitted from this buffer –added to samples after BCA quantitation)

11.2 Appendix B

Time course experiment (cells differentiated at 90% cell confluence)

Classifications of differentially expressed protein groups per time point - according to Panther protein class - can be viewed in **Tables B1-3** (*nucleic acid binding* proteins (PC000171) and **B4-6** (*cytoskeletal protein* PC00085).

Table B1 HT1 (6 hours post treatment), Differentially expressed protein groups mapped to Panther protein class: *nucleic acid binding* (PC000171)

geneID	gene.name	p.value	log_2FC*	subtype	PEP
BAZ1B	Tyrosine-protein kinase BAZ1B	0.021	0.16	DNA	9.10e-274
PARP1	Poly [ADP-ribose] polymerase 1	0.026	0.12	DNA	0
MCM5	DNA replication licensing factor MCM5	0.035	0.28	DNA	6.29e-161
SMRC2	SWI/SNF complex subunit SMARCC2	0.038	0.47	DNA	0
PDS5B	Sister chromatid cohesion protein PDS5 homolog B	0.0041	0.16	DNA	7.86e-164
RFA1	Replication protein A 70 kDa DNA-binding subunit	0.040	0.14	DNA	2.79e-271
BRD4	Bromodomain-containing protein 4	0.038	0.28	DNA	2.36e-59
RCC2	Protein RCC2	0.043	0.16	DNA	0
EDF1	Endothelial differentiation-related factor 1	0.035	0.47	other	5.55e-30
GMPPA	Mannose-1-phosphate guanyltransferase alpha	0.022	0.15	RNA	2.68e-12
MCES	mRNA cap guanine-N7 methyltransferase	0.034	0.30	RNA	3.64e-64
KHDR2	KH domain-containing, RNA-binding, signal transduction-associated protein 2	0.041	0.56	RNA	2.53e-16

PRP4	U4/U6 small nuclear ribonucleoprotein Prp4	0.042	0.23	RNA	5.59e-67
RBM12	RNA-binding protein 12	0.014	0.27	RNA	1.02e-119
MTA70	N6-adenosine-methyltransferase 70 kDa subunit	0.035	0.13	RNA	4.86e-42
DX39A	ATP-dependent RNA helicase DDX39A	0.0064	0.26	RNA	6.11e-210
RL10A	60S ribosomal protein L10a	0.0147	0.11	RNA	1.16e-111
MCA3	Eukaryotic translation elongation factor 1 epsilon-1	0.016	0.18	RNA	4.51e-28
THOC4	THO complex subunit 4	0.028	0.35	RNA	1.23e-269
SYK	Lysine--tRNA ligase	0.013	0.13	RNA	6.43e-154
RS2	40S ribosomal protein S2	0.018	0.14	RNA	8.09e-160
EIF3L	Eukaryotic translation initiation factor 3 subunit L	0.027	0.12	RNA	3.88e-244
RL21	60S ribosomal protein L21	0.028	0.24	RNA	2.68e-71
HNRH2	Heterogeneous nuclear ribonucleoprotein H2	0.014	0.23	RNA	0
EIF3C	Eukaryotic translation initiation factor 3 subunit C	0.034	0.14	RNA	5.57e-146
CPSF3	Cleavage and polyadenylation specificity factor subunit 3	0.017	0.24	RNA	3.37e-101
RL5	60S ribosomal protein L5	0.030	0.14	RNA	2.23e-174
PSPC1	Paraspeckle component 1	0.043	0.21	RNA	5.04e-258
RBM8A	RNA-binding protein 8A	0.022	0.31	RNA	2.03e-240
HNRPK	Heterogeneous nuclear ribonucleoprotein K	0.026	-0.55	RNA	0
DDX18	ATP-dependent RNA helicase DDX18	0.044	0.22	RNA	8.05e-51

**FC- fold change, HIV-Tat/ BME (i.e. treated/control)*

Table B2 HT2 (24 hours post treatment), Differentially expressed protein groups mapped to Panther protein class: *nucleic acid binding* (PC000171)

geneID	gene.name	p.value	log_2FC*	subtype	PEP
SMRD3	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 3	0.017	-0.20	DNA	1.04e-48
TBPL1	TATA box-binding protein-like protein 1	0.019	0.085	DNA	2.97e-58
ZNF24	Zinc finger protein 24	0.001	0.19	DNA	8.49e-05
IF5	Eukaryotic translation initiation factor 5	0.015	-0.062	RNA	1.40e-130
CHTOP	Chromatin target of PRMT1 protein	0.0040	-0.074	RNA	1.24e-91
EIF3G	Eukaryotic translation initiation factor 3 subunit G	0.032	-0.085	RNA	4.00e-133
RS16	40S ribosomal protein S16	0.018	-0.11	RNA	2.23e-47
ZFHX3	Zinc finger homeobox protein 3	0.032	-0.11	RNA	1.66e-13
RU17	U1 small nuclear ribonucleoprotein 70 kDa	0.043	-0.13	RNA	4.91e-130
EFTU	Elongation factor Tu, mitochondrial	0.032	-0.13	RNA	0
SYCC	Cysteine--tRNA ligase, cytoplasmic	0.033	-0.16	RNA	5.80e-85
DDX18	ATP-dependent RNA helicase DDX18	0.030	-0.29	RNA	8.06e-51
WDR7	WD repeat-containing protein 7	0.043	-0.29	RNA	1.43e-15
RBM5	RNA-binding protein 5	0.0016	-0.32	RNA	4.14e-28

LRRF1	Leucine-rich repeat flightless-interacting protein 1	0.014	0.063	RNA	1.37e-13
MRT4	mRNA turnover protein 4 homolog	0.028	0.26	RNA	1.51e-26
EXOS8	Exosome complex component RRP43	0.022	0.28	RNA	8.78e-75
ERI3	ERI1 exoribonuclease 3	0.001	0.45	RNA	2.85e-15

**FC- fold change, HIV-Tat/ BME (i.e. treated/control)*

Table B3 HT3 (48 hours post treatment), Differentially expressed protein groups mapped to Panther protein class: *nucleic acid binding* (PC000171)

geneID	gene.name	p.value	log_2FC*	subtype	PEP
ATPB	ATP synthase subunit beta, mitochondrial	0.028	-0.046	DNA	0
LN28B	Protein lin-28 homolog B	0.0039	0.14	DNA	1.0045e-10
HMGB2	High mobility group protein B2	0.015	0.34	DNA	3.7012e-24
SMRD1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1	0.025	0.12	DNA	3.2527e-191
MSH2	DNA mismatch repair protein Msh2	0.036	0.21	DNA	2.0168e-159
RUXF	Small nuclear ribonucleoprotein F	0.043	0.53	RNA	2.1771e-80
PRP16	Pre-mRNA-splicing factor ATP-dependent RNA helicase PRP16	0.021	0.42	RNA	1.1609e-39

SREK1	Splicing regulatory glutamine/lysine-rich protein 1	0.042	0.1	RNA	1.8535e-49
RL13	60S ribosomal protein L13	0.037	0.34	RNA	1.1597e-115
CHTOP	Chromatin target of PRMT1 protein	0.024	0.27	RNA	1.2434e-91
IF2B1	Insulin-like growth factor 2 mRNA-binding protein 1	0.042	0.27	RNA	0

**FC- fold change, HIV-Tat/ BME (i.e. treated/control)*

Table B4 HT1 (6 hours post treatment), Differentially expressed protein groups mapped to Panther protein class: *cytoskeletal protein* (PC00085)

geneID	gene.name	p.value	log_2FC*	subtype	PEP
MCA3	Eukaryotic translation elongation factor 1 epsilon-1	0.016	0.18	actin	4.51e-28
CTNA1	Catenin alpha-1	0.037	-0.23	actin	0
DEST	Destrin	0.0063	0.32	actin	1.58e-49
LMNA	Prelamin-A/C	0.030	0.20	intermediary filament	1.52e-44

**FC- fold change, HIV-Tat/ BME (i.e. treated/control)*

Table B5 HT2 (24 hours post treatment), Differentially expressed protein groups mapped to Panther protein class: *cytoskeletal protein* (PC00085)

geneID	gene.name	p.value	log_2FC*	subtype	PEP
ARP10	Actin-related protein 10	0.023	-0.079	actin	6.00e-129
FMN2	Formin-2	0.0047	-0.46	actin	4.71e-18

ZFHX3	Zinc finger homeobox protein 3	0.032	-0.11	actin	1.66e-13
TWF1	Twinfilin-1	0.026	-0.11	actin	8.50e-141
ADDA	Alpha-adducin	0.032	-0.16	actin	1.17e-121
OPA1	Dynamin-like 120 kDa protein, mitochondrial	0.022	-0.13	microtubule	8.28e-127
GCP3	Gamma-tubulin complex component 3	0.017	-0.19	microtubule	3.79e-26
KLC2	Kinesin light chain 2	0.0053	-0.18	microtubule	8.14e-106
MAP1S	Microtubule-associated protein 1S	0.033	-0.18	microtubule	1.79e-115

**FC- fold change, HIV-Tat/ BME (i.e. treated/control)*

Table B6 HT3 (48 hours post treatment), Differentially expressed protein groups mapped to Panther protein class: *cytoskeletal protein* (PC00085)

geneID	gene.name	p.value	log_2FC*	subtype	PEP
CTNA1	Catenin alpha-1	0.031	-0.15	actin	0
CTNA2	Catenin alpha-2	0.038	-0.19	actin	0

**FC- fold change, HIV-Tat/ BME (i.e. treated/control)*

11.3 Appendix C

Repeat of 24-hour time point

Tables C1 and **C2** list protein groups and their details for those that mapped to Panther protein class *cytoskeletal protein* (PC00085) and *nucleic acid binding* (PC000171).

Table C1 Differentially expressed protein groups mapped to Panther protein class: *cytoskeletal protein* (PC00085)

geneID	gene.name	log_2FC*	p.value	subtype	PEP
PALLD	Palladin isoform 4	-0.24	0.042	actin	4.3e-35
CALD1	Caldesmon isoform 5	-0.47	0.0030	actin	1.6e-199
VINC	Vinculin isoform 1	-0.16	0.00020	actin	0
VIM	Vimentin	-0.25	0.031	Intermediate filament	0
LMNA	Prelaminin A/C	-0.51	0.0080	Nuclear lamina	8.0e-90
TAGLN	Transgelin	-0.30	0.011	actin	2.3e-154
NES	Nestin	-0.20	0.018	Intermediate filament	0
SPTBN1	Spectrin beta chain, non-erythrocytic	-0.10	0.030	actin	0
PLEC	Plectin	-0.13	0.037	cytolinker	1.4e-109

**FC: fold change. HIV-Tat treated/ control*

Table C2 Differentially expressed protein groups mapped to Panther protein class: *nucleic acid binding* (PC000171)

geneID	gene.name	p.value	Log_2FC	subtype	PEP
ELAV3	ELAV-like protein 3	0.013	0.29	RNA	1.9e-34
H2AW	Core histone macro-H2A.2	0.027	0.22	DNA	5.9e-232
EIF3E	Eukaryotic translation initiation factor 3 subunit E	0.029	0.19	RNA	4.6e-82
PA2G4	Proliferation-associated protein 2G4	0.037	0.11	RNA	1.6e-156
RCC2	SWI/SNF complex subunit SMARCC2	0.015	0.14	DNA	8.8e-152
HNRPM	Isoform 2 of Heterogeneous nuclear	0.048	0.11	RNA	0

	ribonucleoprotein M				
SF3B1	Splicing factor 3B subunit 1	0.044	0.13	RNA	3.6e-215
MCM3	DNA replication licensing factor MCM3	0.046	0.15	DNA	1.8e-222
DDX17	Probable ATP-dependent RNA helicase DDX17	0.039	0.13	RNA	1.0e-259
PRKDC	DNA-dependent protein kinase catalytic subunit	0.030	0.21	DNA	0
RAD23B	UV excision repair protein RAD23	0.016	-0.17	DNA	1.2e-70
TADBP	TAR DNA-binding protein	0.023	0.36	DNA	2.5e-135
MEX3A	RNA-binding protein MEX3A	0.031	0.23	RNA	1.7e-05
HNRPD	Isoform 3 of Heterogeneous nuclear ribonucleoprotein D0	0.031	0.23	RNA	6.1e-221
PRKRA	Interferon-inducible double-stranded RNA-dependent protein kinase activator A	0.046	0.25	RNA	1.5e-34
SP16H	FACT complex subunit SPT16	0.049	0.14	DNA	3.6e-169

**FC: fold change. HIV-Tat/control*

11.4 Appendix D

Time course repeat at lower (70%) cell confluence

Panther protein class outputs, per time point, can be viewed in **Tables D1-4** (*cytoskeletal protein*) and **D5-7** (*nucleic acid binding* proteins).

Table D1 Time point 1, Differentially expressed protein groups mapped to Panther protein class: *cytoskeletal protein* (PC00085)

geneID	gene.name	p.value	log_2FC*	PEP
STML2	Stomatin-like protein 2, mitochondrial	0.010	-0.52	7.91e-58

**FC- fold change, HIV-Tat/ control*

Table D2 Time point 2, Differentially expressed protein groups mapped to Panther protein class: *cytoskeletal protein* (PC00085)

geneID	gene.name	p.value	log_2FC*	PEP
CSRP1	Cysteine and glycine-rich protein 1	0.010	0.31	1.8e-35
DC1L2	Cytoplasmic dynein 1 light intermediate chain 2	0.012	-0.066	4.9e-125
TLN1	Talin-1	0.016	0.16	9.9e-169
KIF5C	Kinesin heavy chain isoform 5C	0.035	-0.14	0

**FC- fold change, HIV-Tat/ control*

Table D3 Time point 3, Differentially expressed protein groups mapped to Panther protein class: *cytoskeletal protein* (PC00085)

geneID	gene.name	p.value	log_2FC*	PEP
COF1	HUMAN Cofilin-1	0.036	-0.22	0
SHLB1	HUMAN Endophilin-B1	0.038	-0.37	5.3e-28
CLIC4	HUMAN Chloride intracellular channel protein 4	0.039	-0.16	1.4e-208
DYL1	HUMAN Dynein light chain 1, cytoplasmic	0.041	0.58	1.4e-182

**FC- fold change, HIV-Tat/ control*

Table D4 Time point 1, Differentially expressed protein groups mapped to Panther protein class: *nucleic acid binding* (PC000171)

geneID	gene.name	p.value	log_2FC*	subtype	PEP
CHD7	Chromodomain-helicase-DNA-binding protein 7	0.011	0.30	DNA	8.47e-24
RAD21	Double-strand-break repair protein rad21 homolog	0.015	-0.55	DNA	4.8e-95
NXF1	Nuclear RNA export factor 1	0.016	-0.36	RNA	5.8e-54
LSM2	U6 snRNA-associated Sm-	0.017	0.57	RNA	3.7e-44

	like protein LSm2				
SUGP2	SURP and G-patch domain-containing protein 2	0.022	-0.40	RNA	2.3e-108
H12	Histone H1,2	0.027	-0.41	DNA	0
ROA2	Heterogeneous nuclear ribonucleoproteins A2/B1	0.028	-0.33	RNA	0
VATB2	V-type proton ATPase subunit B, brain isoform	0.034	-0.36	DNA	3.9e-149
H2A1J	Histone H2A type 1-J	0.034	-1.25	DNA	7.4e-152
ATPA	ATP synthase subunit alpha, mitochondrial	0.045	-0.36	DNA	0
SRSF3	Isoform 2 of Serine/arginine-rich splicing factor 3	0.047	-0.42	RNA	1.2e-113

**FC- fold change, HIV-Tat/ control*

Table D5 Time point 2, Differentially expressed protein groups mapped to Panther protein class: *nucleic acid binding* (PC000171)

geneID	gene.name	log_2FC*	p.value	subtype	PEP
CBX8	Chromobox protein homolog 8	0.26	0.0033	DNA	4.6e-10
SRS10	Isoform 5 of Serine/arginine-rich splicing factor 10	0.43	0.0049	RNA	2.3e-56
DKC1	H/ACA ribonucleoprotein complex subunit 4	0.12	0.0062	RNA/ DNA	1.7e-72
COT1	COUP transcription factor 1	0.44	0.0094	DNA	2.2e-88
THOC4	THO complex subunit 4	0.15	0.011	RNA	2.6e-227
HNRPC	Isoform C1 of Heterogeneous nuclear ribonucleoproteins C1/C2	0.10	0.012	RNA	0
RS17L	40S ribosomal protein S17-like	0.11	0.014	RNA	1.6e-42
DHX9	ATP-dependent RNA helicase A	0.05	0.014	RNA	0
TIAR	Nucleolysin TIAR	0.44	0.021	RNA	1.4e-76
MTA2	Metastasis-associated protein MTA2	0.19	0.021	DNA	1.1e-84
DNLI3	DNA ligase 3	0.33	0.024	DNA	1.7e-61
HNRPK	Isoform 3 of Heterogeneous nuclear ribonucleoprotein K	0.31	0.028	DNA	0
FUS	Isoform Short of RNA-binding protein FUS	0.26	0.029	DNA	0
BAZ1B	Isoform 2 of Tyrosine-protein kinase BAZ1B	0.27	0.036	DNA	3.7e-104
PRKRA	Interferon-inducible double-stranded RNA-dependent protein kinase activator A	0.15	0.039	RNA	1.6e-33

POGZ	Isoform 5 of Pogo transposable element with ZNF domain	0.28	0.039	DNA	4.4e-83
NOP58	Nucleolar protein 58	0.23	0.042	RNA	2.5e-137
YLPM1	YLP motif-containing protein 1	0.23	0.045	DNA	4.5e-125

**FC- fold change, HIV-Tat/ control*

Table D6 Time point 3, Differentially expressed protein groups mapped to Panther protein class: *nucleic acid binding* (PC000171)

geneID	gene.name	log_2FC*	p.value	subtype	PEP
PRP31	U4/U6 small nuclear ribonucleoprotein Prp31	-0.29	0.0038	RNA	4.6e-34
IF4E	Eukaryotic translation initiation factor 4E	-0.35	0.0039	RNA	3.5e-54
LARP7	La-related protein 7	-0.22	0.0065	DNA	9.8e-65
CPSF5	Cleavage and polyadenylation specificity factor subunit 5	-0.19	0.0072	RNA	6.4e-121
IF1AX	Eukaryotic translation initiation factor 1A, X-chromosomal	-0.39	0.011	RNA	1.5e-39
PSPC1	Paraspeckle component 1	-0.25	0.012	DNA	1.2e-151
CRNL1	Isoform 2 of Crooked neck-like protein 1	-0.38	0.013	RNA	1.6e-32
SMRCD	Isoform 3 of SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1	-0.27	0.013	DNA	1.0e-25
CPSF7	Isoform 2 of Cleavage and polyadenylation specificity factor subunit 7	-0.24	0.014	RNA	2.5e-94
RL36L	60S ribosomal protein L36a-like	-0.26	0.017	RNA	3.5e-31
MTA2	Metastasis-associated protein MTA2	-0.14	0.018	DNA	1.1e-84
C2AIL	CDKN2AIP N-terminal-like protein	0.58	0.019	DNA	2.2e-24
SYFB	Phenylalanine--tRNA ligase beta subunit	-0.22	0.019	RNA	2.8e-50
ZCCHL	Zinc finger CCCH-type antiviral protein 1-like	-0.26	0.020	DNA/ RNA	1.8e-32
ZN207	Isoform 2 of BUB3-interacting and GLEBS motif-containing protein ZNF207	-0.45	0.025	DNA	9.0e-74
SMRD3	Isoform 2 of SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 3	-0.23	0.026	DNA	2.7e-40

IF4A3	Eukaryotic initiation factor 4A-III	-0.19	0.026	RNA	2.1e-243	
PRKRA	Interferon-inducible double-stranded RNA-dependent protein kinase activator A	-0.40	0.027	RNA	1.6e-33	
CBX5	Chromobox protein homolog 5	-0.27	0.027	DNA	1.5e-95	
SF3B1	Splicing factor 3B subunit 1	-0.18	0.029	RNA		0
DSRAD	Isoform 5 of Double-stranded RNA-specific adenosine deaminase	0.29	0.032	RNA	1.5e-155	
UBP24	Ubiquitin carboxyl-terminal hydrolase 24	-0.25	0.035		1.2e-27	
EHMT1	Histone-lysine N-methyltransferase EHMT1	-0.16	0.036	DNA	2.0e-24	
CLIC4	Chloride intracellular channel protein 4	-0.16	0.039		1.4e-208	
ZFR	Zinc finger RNA-binding protein	-0.22	0.041	RNA	7.1e-128	
RL3	60S ribosomal protein L3	-0.20	0.042	RNA	3.4e-131	
RBM25	RNA-binding protein 25	-0.26	0.044	RNA	6.0e-80	

**FC- fold change, HIV-Tat/ control*