HIV EVOLUTION IN THE BRAIN

by

TESS Z GRIFFIN

(Under the Direction of Ming Zhang)

ABSTRACT

This work aimed to define the diversity of human immunodeficiency virus in the brain and the consequences of viral compartmentalization in the brain on both viral evolution and on brain processes. A knowledge gap in this area exists due to both limited sample availability and due to presence of the blood brain barrier, an interface between the central nervous system and the rest of the body that limits movement of virus and antiretroviral drugs into the brain. In order to help fill this gap, this dissertation work focused on neuropathogenesis due to viral presence in the brain. In chapters one and two of this dissertation we compiled and analyzed the HIV-1 brain-related repertoire of information available from publically accessible databases in order to understand, first, the mechanisms that shape viral genomic diversity in the brain, and second, viral perturbation of molecular signaling pathways in the brain. In the first chapter we developed and employed a computational pipeline that revealed general viral diversity as shaped primarily by pressures unique to the individual host, and secondarily by diagnosis of HIV-1-associated neurocognitive disorder in the host. In addition, we discovered multiple novel HAND-associated viral amino acid markers. This work also resulted in the development of a comprehensive and carefully curated compilation of viral sequence data

and associated clinical information from HIV-1 individuals assessed for HAND available online for continued use by researchers. In the second chapter of this dissertation, comparison of HIV-1 infected brains to both healthy brains and to brains with the brain tumor, glioblastoma multiforme, led to the identification of signaling pathways both unique to HIV presence in the brain and in common with glioblastoma multiforme. In all, our studies have shown the significant impact of HIV presence in the brain on both evolution of the viral genome and on molecular-level changes in the brain. We hope this work will ultimately help shed light on treatment options for the nearly 50% of the HIV-1 population suffering from neurocognitive impairment due to viral presence in the brain.

INDEX WORDS: HIV, brain, HIV-associated neurocognitive disorder, glioblastoma multiforme

HIV EVOLUTION IN THE BRAIN

by

TESS Z GRIFFIN

BS, University of Massachusetts, 2000

MEd, University of Massachusetts, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2015

HIV EVOLUTION IN THE BRAIN

by

TESS Z GRIFFIN

Major Professor:	Ming Zhang
Committee:	Jeffrey F.D. Dean
	Erin Dolan
	Cheolwoo Park

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
May 2015

TABLE OF CONTENTS

Page

CHAPTER

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction to the field of bioinformatics

The fields of computer science and mathematics have informed and transformed molecular biology for several decades now [1, 2]. Prior to the computational and data-driven idea of bioinformatics we now envision, the term "bioinformatics" was used to describe a much more theoretical investigation of biological information processing systems [3]. Not until relatively recently, however, has this intersection of biology, mathematics, statistics and computer science been recognized as a key component of modern biology, as reflected by a rising demand for biologists proficient in computational methods and for educational programs that meet these demands [4]. This need has been driven, in large part, by an exponential increase in genomic data over the last two decades, requiring not only development of computational tools and algorithms capable of processing and analyzing high-throughput quantities of information, but more importantly, researchers that understand the computational challenges and biological questions that underlie these data toward their accurate, sophisticated, and biologically meaningful interpretation. In addition, continuous and rapid advances in data generating platforms require researchers stay current in their knowledge of such technologies and their statistical implications, at times even mandating a researcher set computational precedent in his or her field of study.

The work presented in this dissertation took advantage of bioinformatics methodologies to understand the neurological implications of infection by human immune deficiency virus (HIV). We capitalized on publically available data to increase our analytical power through the development and analysis of meta-datasets. Finally, we developed a publically available online resource to give back to the HIV scientific community, with the hope that this resource will help facilitate advancement in the field. Ultimately, our goal was not only to successfully employ and create computational tools, but also to capture the initial spirit with which Hesper and Hogeweg coined the term bioinformatics. "In short, under the heading of bioinformatics we wanted to combine pattern analysis and dynamic modeling and apply them to the challenge of unraveling pattern generation and informatic processes in biotic systems at multiple scales" [5].

1.2 Introduction to human immunodeficiency virus

Human immunodeficiency virus (HIV) is an etiological agent of autoimmune deficiency syndrome (AIDS), a disease that has claimed ~39 million lives globally in the last three decades [6]. At only 9.7 thousand nucleotides in length and encoding for 15 proteins, the informational capacity of human immunodeficiency virus type 1 (HIV-1) is a mere fraction of its host human genome with 3 billion nucleotide base pairs [7]. Yet, despite its relative simplicity, this virus has evolved not only to evade an array of host immune responses, but also to use this host defense system for its own replicative purposes [8].

1.3 Effects of highly active antiretroviral therapy

Five years following the discovery of HIV as the driver of AIDS in 1984 [9, 10], the first HIV therapeutic agent, reverse transcriptase inhibitor azidothymidine (AZT), was approved for use [11]. In 1996, three protease inhibitors, saquinavir, indinavir, and ritonavir, were all FDA approved to be used in combination with reverse transcriptase inhibitors, and together, afforded the HIV population the first highly active antiretroviral therapy (HAART) regimen against AIDS onset [12]. By 1998, HAART was in wide use in developed countries, and in response, HIV-related deaths in these countries were rapidly declining [13]. In addition to its ability to slow down clinical progression, HAART has also afforded HIV individuals a level of immune system restoration from viral damage [14]. Our ability to control viral replication within the brain and viral damage to the brain, however, remains highly limited [15].

1.4 HIV-associated neurocognitive impairment mechanism

Dysfunction of the central nervous system in HIV-infected individuals became a source of concern and investigation following early clinical observations of acquired immunodeficiency syndrome (AIDS) [16]. Discovery of the direct effect of HIV on the brain came first from observation of viral RNA and viral protein presence in the brain tissue of infected individuals and second, from evidence of active transcription and translation of viral genome in brain-derived macrophage cells [17-19]. It is now known that the brain serves as a reservoir for latent virus and that virus within the brain can infect several different types of cells, including macrophages, microglia, and neural progenitor cells [20]. Coupled with the presence of the blood brain barrier (BBB), targeting virus located in the brain is a great challenge [21, 22].

Within days following transmission, infected circulating monocytes, a type of white blood cell, enter into the central nervous system (CNS) through the BBB and differentiate into macrophages [23] (**Figure 1.1**). Infected macrophages and, subsequently, recruited microglia, release neurotoxic mediators, viral proteins, and cellular activation products that damage neighboring cells, including neurons [24]. Correspondingly, both macrophage levels in the brain and degree of macrophage activation serve as the best indicators of degree of neurocognitive impairment within the HIV individual. Although to some extent these cells do help protect against infection, they also support the initial establishment of infection within the brain, and therefore play a key role in development of HIV-associated neurocognitive disorders (HAND) in the HIV individual [25].

HAND progression is marked by a number of distinct histopathological events. These events begin with infiltration and accumulation of macrophage cells within the CNS and the subsequent formation of microglial nodules and multi-nucleated giant cells [26]. This is followed by widespread astrogliosis, an abnormal increase in astrocyte level in response to neuronal damage, and eventually leads to pallor of myelin sheaths, neuronal loss, and loss of synaptic connections between neurons [24]. Areas of the brain affected by this chronic inflammation, as evidenced by increased astrogliosis in these regions, include the hippocampus and the entorhinal cortex, areas of the brain involved in memory, and subcortical white matter, an area controlling multiple functionalities, including memory, language, and movement [27].

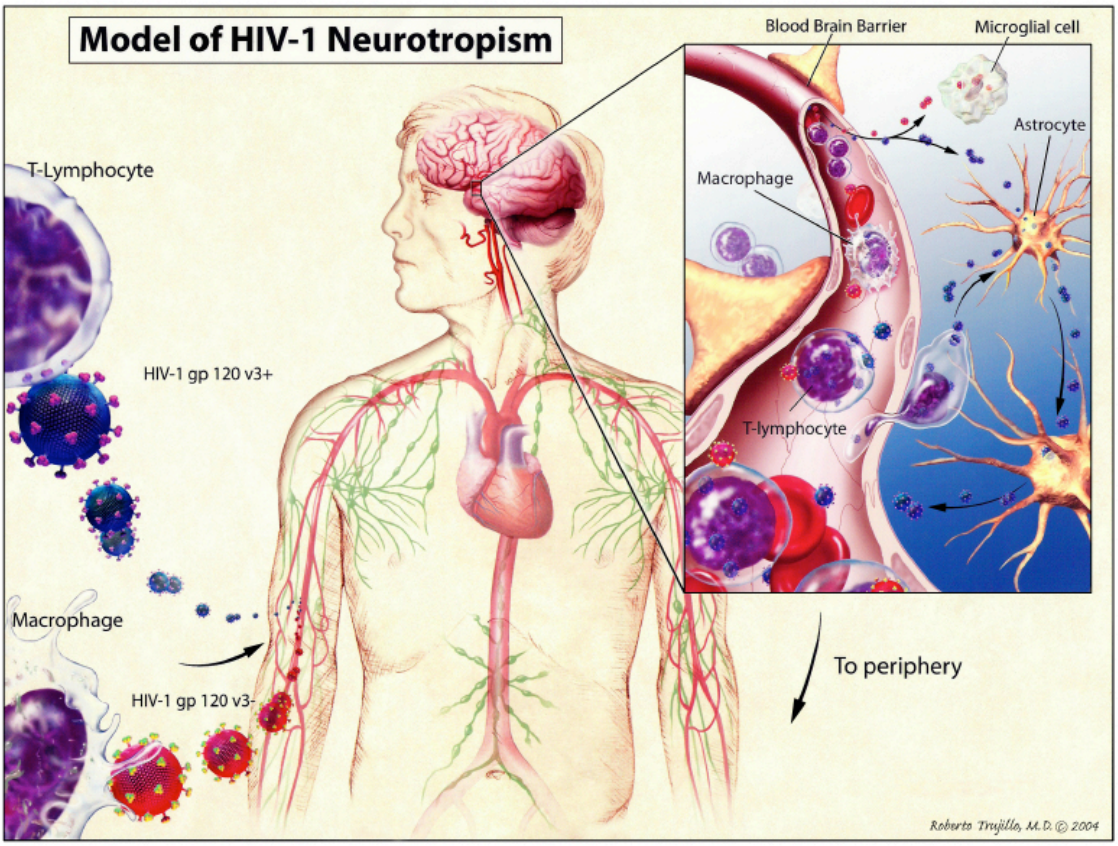


Figure 1.1 HIV-1 entry into the human brain (from Trujillo, 2004) [23]

HIV enters the brain early following initial infection. Infected monocytes traveling through the circulatory system, and in particular, mature monocytes, as the white blood cell population most vulnerable to HIV infection, enter into the CNS through the BBB. In this manner, virus is brought into the brain, and infection established through monocytic differentiation and by subsequent macrophage and microglial recruitment to this site [25, 28].

1.5 HAND in the HIV population

Prior to HAART introduction, HIV-associated neurocognitive disorder (HAND) was severe and would rapidly progress to AIDS dementia complex (ADC) (sometimes referred to as HIV-associated dementia (HAD)). While HAART has greatly reduced the incidence of severe forms of HAND to only 1% of the HIV population, the percentage of HIV individuals suffering from milder forms of HAND continues to rise [29]. In the recent and large-scale CNS HIV Anti-Retroviral Therapy Effects Research (CHARTER) study, nearly half of all HIV-1-infected individuals suffered from a form of HAND (**Figure 1.2**) [29, 30]. A number of reasons potentially contribute to this trend, including: incomplete suppression of virus located within the CNS due to poor penetration of HAART through BBB [21]; extant brain damage prior to HAART use [31]; neurotoxicity of HAART drugs that do cross into the CNS [32]; potential for very low levels of virus in the CNS to promote brain damage through chronic inflammation [33]; and accumulation of damage to the CNS resulting from prolonged patient lives [34]. Ultimately, individuals affected by HAND have difficulty performing even day-to-day activities, and suffer from deficits in learning, motor function, verbal fluency, memory, and attention [35]. Even more crucial, an individual with HAND is at a three-fold increased risk of death as compared to a mentally healthy HIV individual [36].

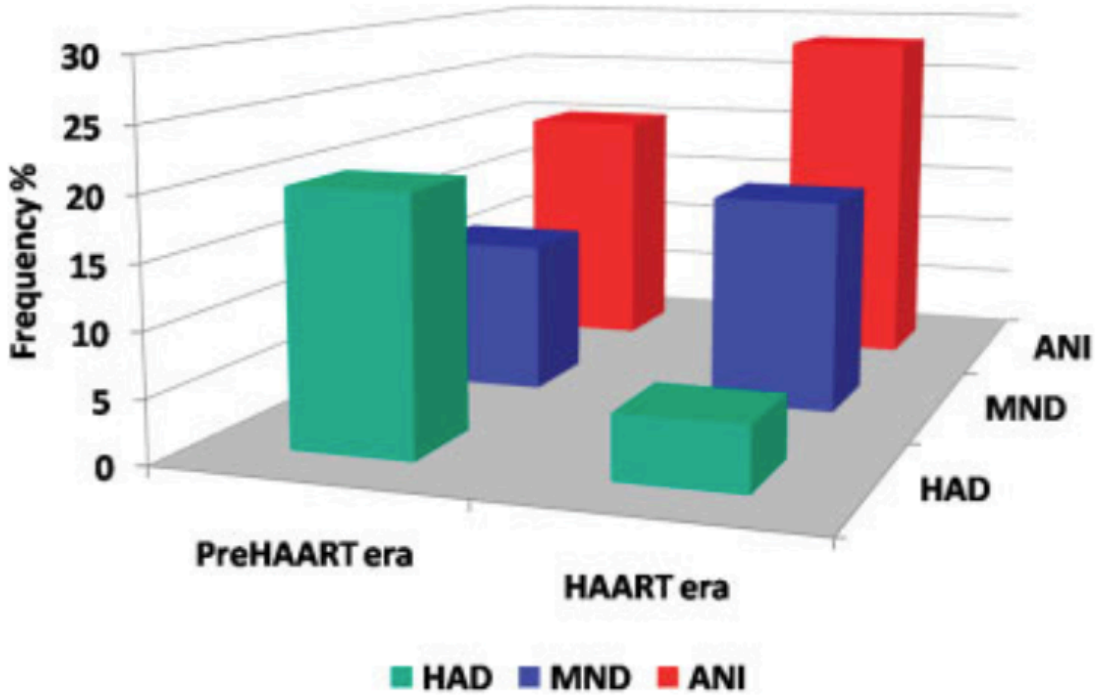


Figure 1.2 Rates of HIV-associated neurocognitive disorder before and after introduction of HAART (from McArthur (2010) and Heaton (2011)) [29, 37]

HAND incidence rates prior to and following HAART introduction. Categories HAD, MND and ANI refer to HAND types HIV-associated dementia, mild neurocognitive disorder, and asymptomatic neurocognitive impairment, respectively. Despite HAART use, rates of HAND along the disease progression spectrum are still at 50% in the HIV population.

1.6 Viral evolution in the human brain

The fact that we have limited control over viral replication in the brain, that the brain acts as a reservoir for latent virus, and that HIV, both directly and indirectly, produces a devastating amount of damage on the brain, demands we gain a better understanding of viral processes within the brain. One of the most rapidly evolving entities known, HIV-1 is characterized by its high genetic variability and an ability to evolve quickly relative to, and in response to, its host environment [38]. Along with processes controlling general intrahost viral diversity, namely recombination and substitutional evolution, viral diversity in the brain is dictated by the founder effect, the initial viral community, by selective pressures unique to the brain environment, including tissue-specific innate immune responses, and by a restricted influx of new viral members into the brain due to presence of the BBB [39].

1.7 Prior research on brain viral evolution

Inefficient entry of HAART into the CNS compartmentalizes viral evolution in the brain relative to treatable tissues that undergo evolutionary pressures imposed by drug treatment. While effect of brain compartmentalization on viral sequence evolution has been studied for two decades now, this body of literature remains highly conflictive and suggests a more sophisticated interaction between tissue and virus than has yet been uncovered.

Due to its role in both macrophage tropism and viral replication in the brain, the *env* gene, and in particular the V3 loop region of this gene, has become the focus of HIV compartmentalization research. Studies by Korber [40], Ball [41], Wong [42], Chen [43],

McCrossan [44], and Pillai [45], on the effect of compartmentalization on the V3 loop of the *env* gene, have all indicated presence of selective pressures specific to the brain environment that shape the viral genome. In addition to the V3 loop, the V1/V2 loop of the *env* gene has also been shown to undergo a brain compartmentalization effect [46]. A comparable amount of research, however, has indicated limited evidence of brain compartmentalization on *env*, including work by Reddy [47], Smith [48], and Smit [49] on the V3 loop, as well as work by Beebe [50], Reddy [47], Brew [51], Hughes [52], Zhang [53], and Strain [54] on regions of the *env* gene beyond the V3 loop. Like HIV compartmentalization research, work on HIV evolution specific to the HAND brain has also led to contradictory findings. As early as 1994, Power discovered two *env* codons that differed between HAND and non-HAND patients [55]. This finding has since been refuted and validated [56, 57]. Additional HAND-associated signatures have since been discovered in both the *env* [45, 58] and *nef* genes [59, 60].

1.8 Our work on brain viral evolution

Toward an enhanced understanding of brain environmental pressures on viral genome evolution, in the first of two research projects presented in this dissertation, we developed a comprehensive viral sequence database to observe and measure variation in the HIV sequence in response to tissue compartmentalization, and to test for genomic variation in response to HAND presence. We anticipated creation of this database to provide both an increase in sample size and in clinical annotation with which to test for genomic variation across a greater region of the viral genome, across a greater range of HAND types, and with more power than possible prior to development of this resource.

1.8.1 Development of the HAND database resource

The HAND Database was developed as a publically available resource of HAND-associated viral sequence data and associated clinical annotations. The largest HAND-specific viral sequence database to date, this resource originally contained 5,783 sequences from 163 individuals. The only other published HAND viral sequence resource at the time, the HIV Brain Sequence Database (HIVBrainSeqDB), focused on brain-derived viral sequences from the *env* region of the viral genome [61]. In contrast, we developed the HAND Database as a more general HAND resource, containing all viral sequences with HAND-associated information, regardless of HAND type, sampling tissue type, or viral genomic region. In addition, the HAND Database included sequences from HIV individuals that had been tested for HAND, but were found to not have the disorder, effectively providing a control group for future HAND research. In addition to viral sequence data, this resource included detailed clinical annotation on each sequence entry, including, for example, patient information (HAND status and type, HIV therapy status and type, and sex), sampling information (tissue, year, and country), and sequence information (genotype and polyprotein). Overall, database sequences were sampled from twenty different tissue types and from five geographical regions, and sequence coverage included the five HIV-1 genes, *gag, pol, env, tat,* and *nef*.

In addition to database breadth, we worked to ensure database quality. Sequence and clinical data were filtered and annotated: clinical data were cross-referenced whenever possible, sequence data were filtered for contamination and for sequencing errors, and finally, sequences were re-genotyped, as HIV genotyping has been shown to be error-prone [62].

We also created a website to host this resource, with a number of features to facilitate database use, including, multiple search and filtering mechanisms, the ability to create custom Boolean searches, multiple data export options, and help and contact pages (**Figure 1.3**). We anticipated this resource would prove useful in HIV compartmentalization research, in HAND research, as well as in the more general investigation of viral evolutionary processes.

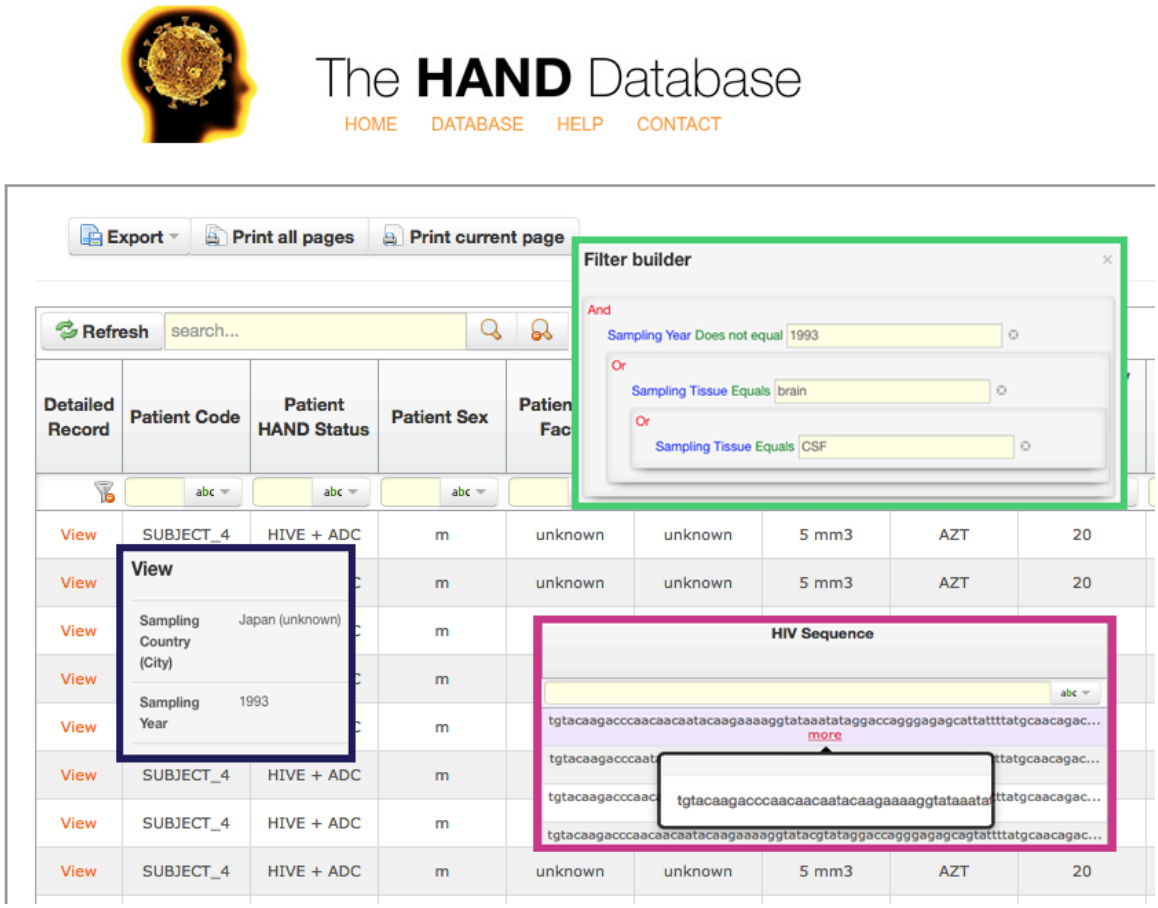


Figure 1.3 The HAND database search interface

The HAND Database provides flexible searching, filtering, and browsing capabilities, including a detailed view option and a Boolean search builder. In addition, viral sequence data and HIV individual annotations of interest can be exported into a variety of file formats for further use. Additional website pages include a home page with background information on HAND, a help page with database information, and a contact page for database support.

1.8.2 Use of the HAND database in the study of brain viral evolution

The first study detailed in this dissertation implemented the HAND Database in the systematic study of HIV-1 genome changes associated with both tissue compartmentalization and with presence of HAND. In our compartmentalization research we leveraged a phylogenetic approach in combination with this database to uncover evidence of viral evolution specific to tissue type. We found no significant evidence of a tissue compartmentalization effect on viral evolution. However, when testing individual HIV genes, we did observe an interaction between the brain environment and the *env* region of the viral genome. In general, the greatest effect on the genome, accounting for nearly 100% of its overall variation, was the HIV individual hosting that particular viral sequence.

In the second part of this study looking for viral markers associated with HAND, we again used the HAND Database, but this time in combination with an information theory approach toward the quantification of an individual HIV gp120 amino acid ability to indicate HAND presence in the HIV individual [63]. More specifically, we used Shannon Entropy to calculate the amount of information provided in a particular message, in this case an individual position in an amino acid alignment, by its reduction of uncertainty with regard to an event, in this case HAND diagnosis [64]. An aligned amino acid position with more Shannon Entropy provided us with more information on the outcome of HAND as a probabilistic event. HAND and non-HAND Shannon Entropy measurements were compared to find amino acid positions with significant differences in variability between the two groups [65]. Using this technique, we validated 11 prior viral genomic markers associated with HAND and HAD diagnoses, and identified 56 novel

HAND-associated HIV gp120 amino acid markers ($p < 0.05$), 26 of which remained significant at the more conservative level of $p < 0.005$ (**Figure 1.4**).

It is important we note that in our search for HAND viral markers, we took measures to account for genomic characteristics specific to HIV. These steps included testing for recombination and hypermutation events, use of alignment algorithms to help maintain N-linked glycosylation site information, as well as more general measures to ensure accuracy of interpretation, including increasing our statistical power through removal of host-specific biases and through an increase in sample size. While HAND research is still limited in many ways, including breadth of viral genome sequenced and geographical regions sampled, we hope this work will help push these limits by consolidating prior sequencing efforts and by providing new regions of interest in HIV gp120 for continued study.

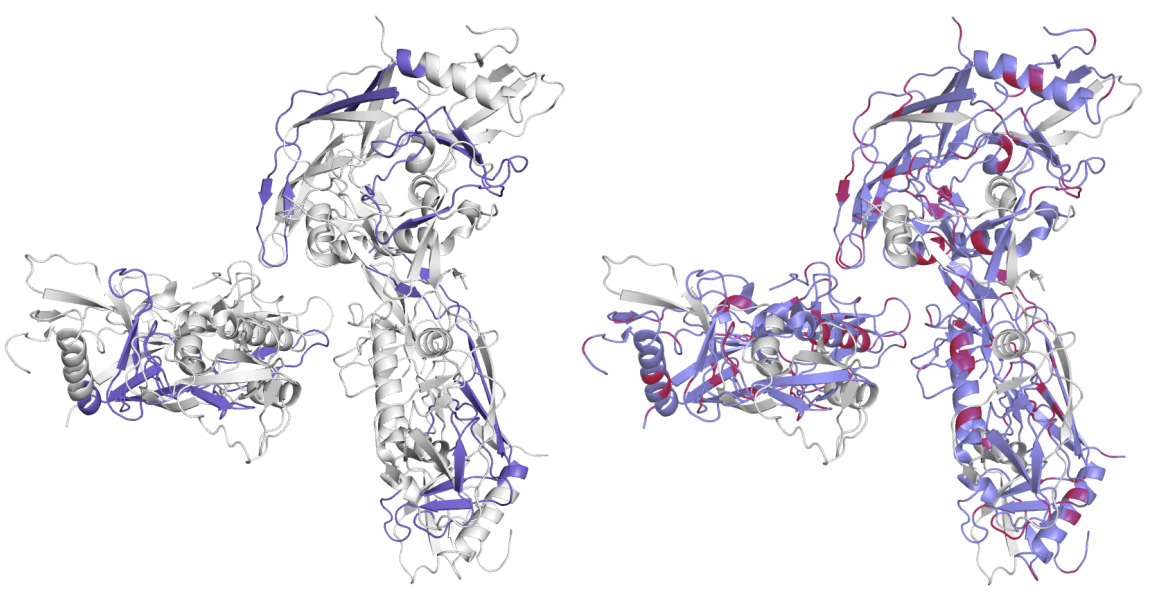


Figure 1.4 Identification of HIV gp120 amino acids associated with HAND

The left-hand model shows the genomic region of the HIV gp120 trimer tested for HIV-associated dementia (HAD)-associated amino acids in the recent high-throughput study by Holman et al. [56], as depicted on the 4NCO PDB HIV gp120 model in light blue [66]. We selected this study to compare against as it made use of the recent HIVBrainSeqDB Database [61]. The right-hand model shows the genomic region of the HIV gp120 trimer we found to be significantly associated with HAND at $p < 0.05$, as shown in red on the 4NCO PDB HIV gp120 model [66]. The light blue region of the right-hand model indicates the region of the HIV gp120 trimer explored within this study. In addition to testing a region of over three times that of prior work, we also observed 56 novel HAND-associated amino acids and validated 11 prior HAND markers.

1.9 Malignancy in the HIV population

The two primary causes of morbidity in treated HIV-1 patients are opportunistic infections and tumors [67, 68]. With HIV disease progression, not only is the HIV individual at an increased risk for neurocognitive damage, but also for formation and proliferation of malignant cells [69]. This propensity is attributable to both a failing of cancer detection systems in the HIV individual and to HIV-specific mechanisms that promote tumor formation, including: oncogene activation, tumor suppressor gene inactivation, immune system impairment, and disease co-infection [70].

Brain tumor primary CNS lymphoma (PCSNL) is well documented in the HIV population, particularly in severely immunocompromised individuals and considered an AIDS-defining malignancy. Glioma brain tumors, however, are not as frequent in the HIV population, and in fact, the most common primary brain tumor in the general population, glioblastoma multiforme (GBM), is rarely observed in HIV individuals. Between 1984 and 2010, only 55 glioma cases, 19 of these GBM cases (**Figures 1.5 and 1.6**) were reported worldwide in the HIV population [69, 71]. While lymphoproliferative cells give rise to PCSNLs, gliomas develop from the malignant transformation of neuroectodermal-derived supporting cells, a process primarily associated with mutations in either the *TP53* tumor suppressor gene, or the *p16/RB/E2F pathway* [71]. In HIV individuals, glial tumors may result from a number of viral-related mechanisms, including viral inactivation of tumor suppressor genes and activation of oncogenes, viral induction of cytokines and cellular growth factors, and viral promotion of malignant transformation of glial and astrocyte cells via over-expression of transforming growth factor-β2 (*TGF-β2*) [72, 73].

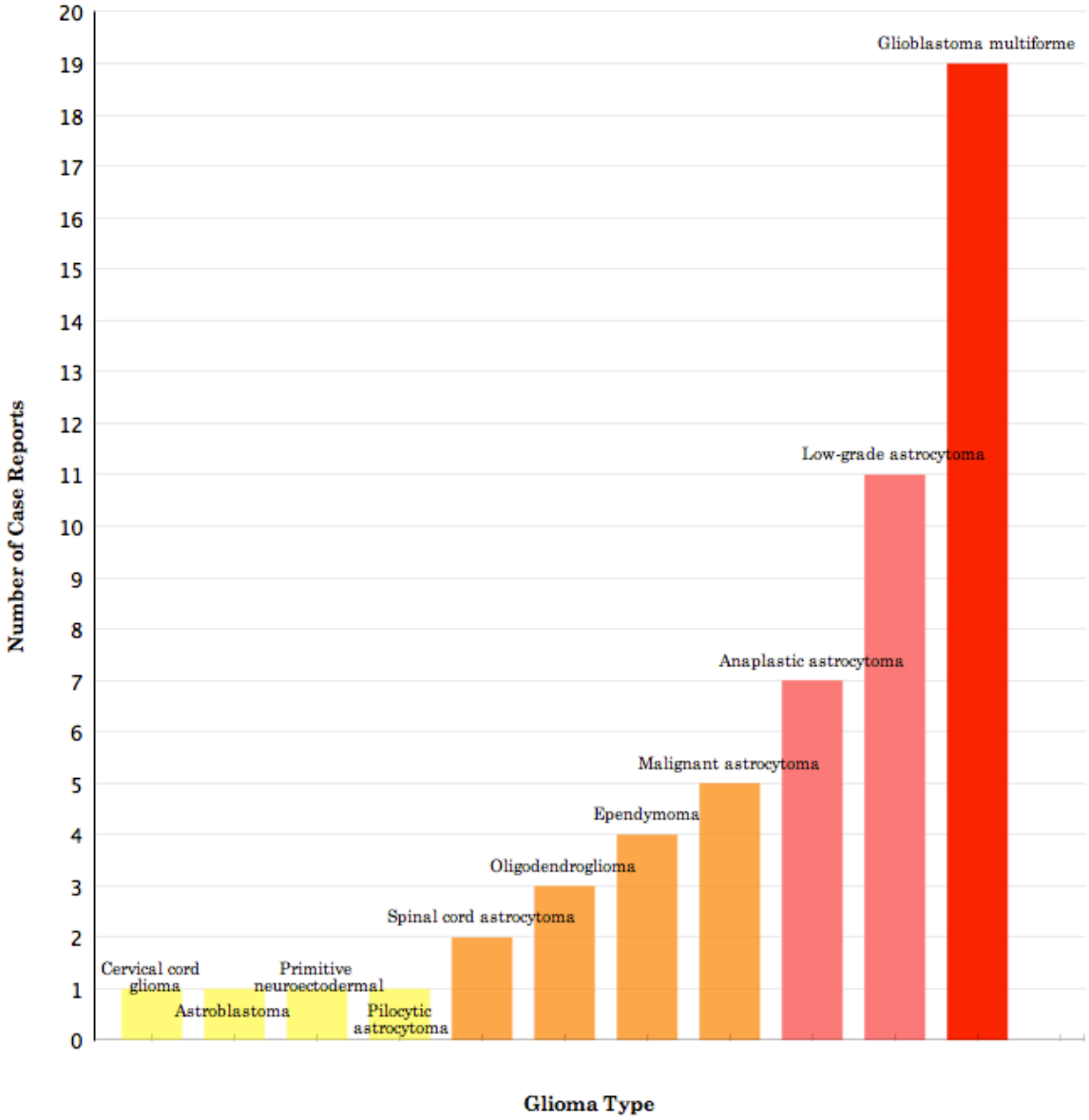


Figure 1.5 Glioma cases in the HIV-1 population by histological type (modified from Cedeno-Laurent & Trujillo, 2011) [71]

Between 1984 and 2010 the most frequently observed glioma in the HIV population was glioblastoma multiforme (GBM), with 19 total cases of GBM reported during that time period.

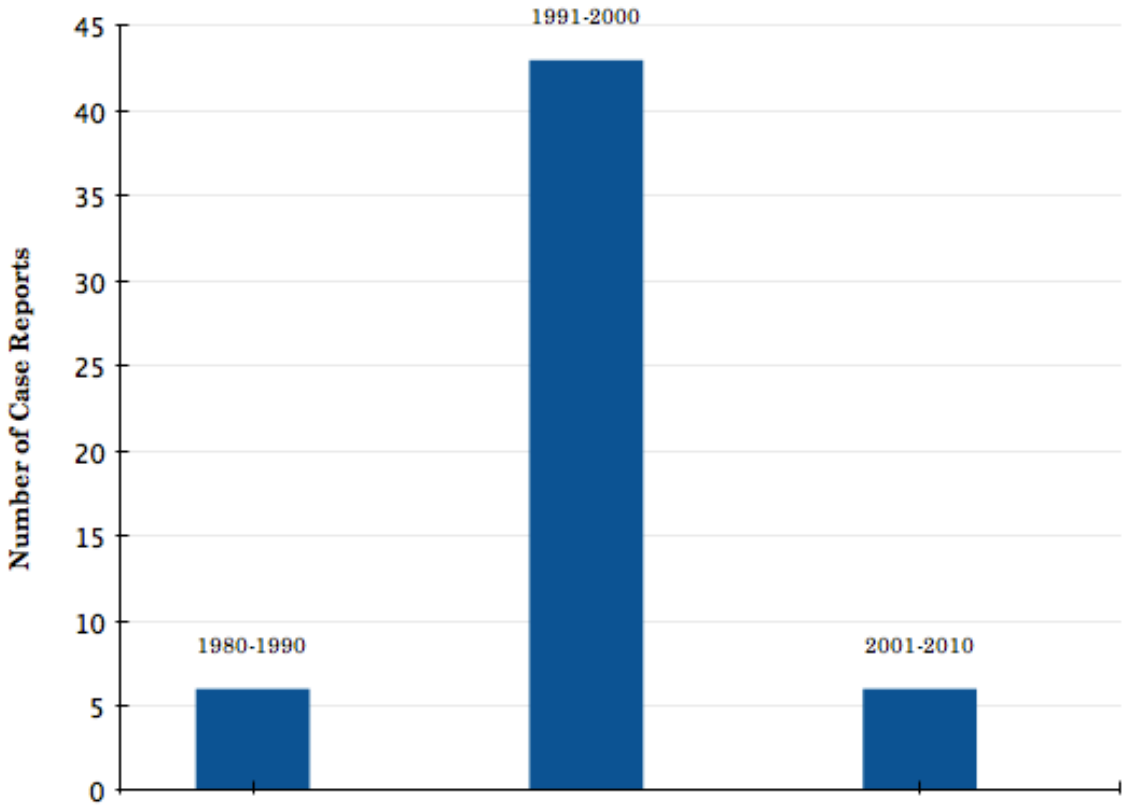


Figure 1.6 Glioma cases in the HIV-1 population by decade (modified from Cedeno-Laurent & Trujillo, 2011) [71]

Within the span of nearly three decades, only 55 glioma cases were reported in the HIV population [71]. While the initially low reporting of gliomas between the years 1980 to 1990 may be attributable to poor documentation of the malignancy and a short life expectancy for HIV individuals, there was a reduction in glioma cases between the decades of 1991-2000 and 2001-2010. While this may in part be due to the introduction of HAART in 1996, an extremely limited sample size to start with lead us to hypothesize a potential effect of the virus itself against GBM development.

1.10 Glioblastoma multiforme development

According to a recent report by the Central Brain Tumor Registry of the United States (CBTRUS), in the general population, GBMs account for 16.3% of all reported brain tumors, and almost 50% of all neuroepithelial tissue tumors [74]. While GBM etiology is still largely unknown, a great body of literature has helped elucidate those molecular mechanisms that underlie both, progression of GBM, and to some extent, initiation of GBM. However, despite advances in GBM research as well as in cancer therapies, prognosis for GBM patients remains extremely poor [75].

One of the primary difficulties involved in GBM treatment lies in its heterogeneity [76], and as a result there has been a move toward therapies that target specific signaling pathway molecules over more generalized cancer treatment options (**Figures 1.7 and 1.8**). Genetic events and signaling pathway alterations central to primary GBM development, the *de novo* form of this tumor, include changes in the signaling *epidermal growth factor receptor (EGFR) pathway*, due to *EGFR* amplification (~40% of primary GBMs) and overexpression (~60% of primary GBMs), shown to impact a number of cancer mechanisms including cellular growth, movement, proliferation and tumor vascularization [77], and to confer chemo-resistance to cells through its regulation of cell death pathways [78]. Additional hallmarks of primary GBM development include loss of cyclin-dependent kinase inhibitor 2A (*CDKN2A* also termed *INK4a/ARF*) (30-40% of primary GBMs) [79], important in cell cycle control, and *PIK3CA* mutation (~7% of primary GBMs) [80, 81]. A characteristic of both primary and secondary GBMs is activation of RAS-mediated signaling, a molecule important in cell survival and cell migration. RAS regulates vascular endothelial growth factor expression (*VEGF*), in this

manner controlling tumor vascularization [82, 83]. Mutation of tumor suppressor gene phosphatase and tensin homolog (*PTEN*) (~30% of primary GBMs), is also common to both primary and secondary GBM development (~5% of secondary GBMs) [84, 85]. Hallmarks of secondary GBM development, a GBM tumor arising from a low grade astrocytoma, include overexpression of *PDGF/PDGFRα* (~60% of secondary GBMs) [85], and inactivating mutations in tumor suppressor pathways*, tumor protein p53 (TP53) pathway* (> 65% of secondary GBMs) and *retinoblastoma protein (RB1) pathway* [86-88].

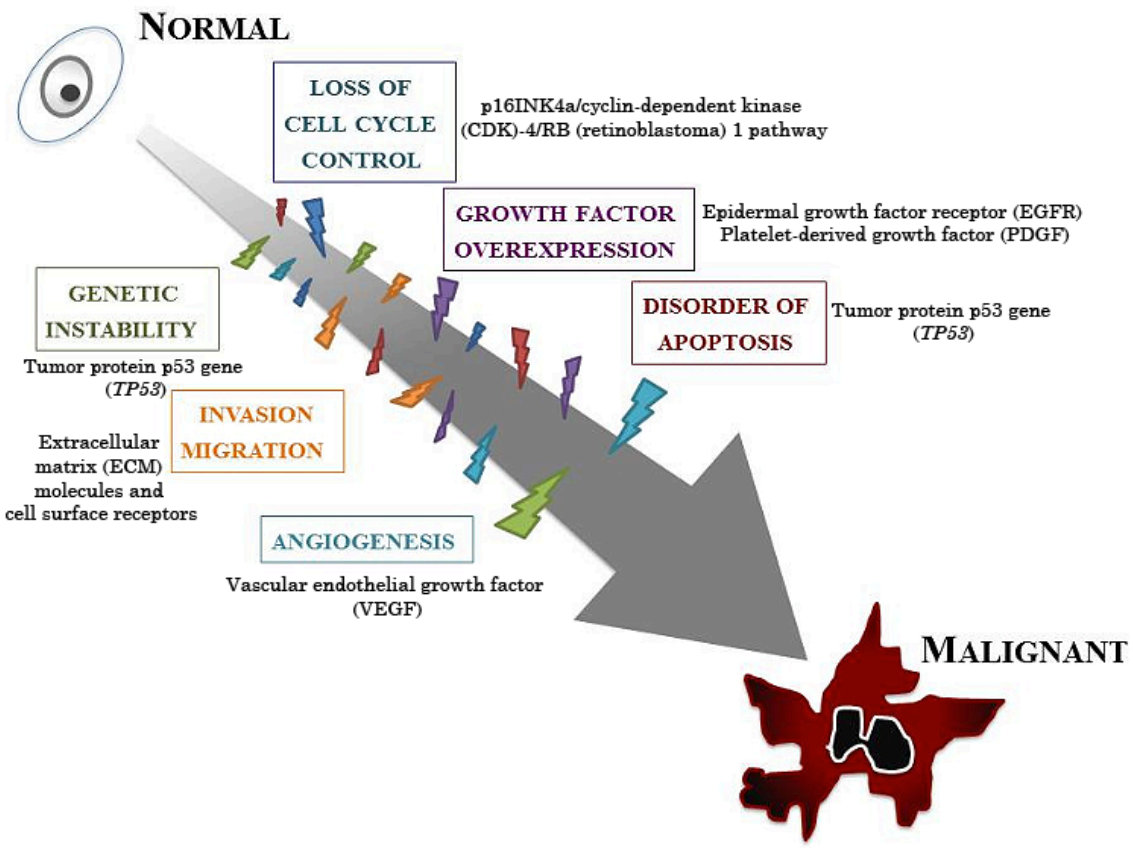


Figure 1.7 Key intracellular events leading to neoplastic transformation of glial cells (modified from Nakada et al., 2011) [80]

Specific changes in intracellular signaling pathways underlie key events in gliomagenesis. These events frequently occur in combination with one another and include: loss of cell cycle control, tumor vascularization, cell invasion and migration, loss of apoptotic control, growth factor overexpression, and genetic instability. Major pathways and molecules involved in these events are listed next to each [80].

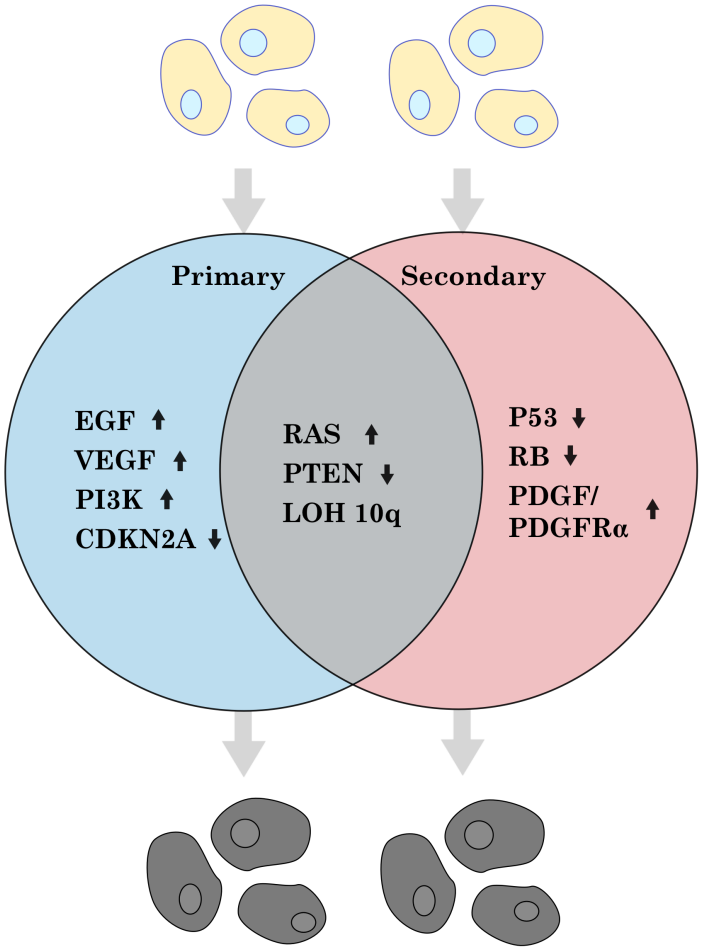


Figure 1.8 Genetic and molecular changes that lead to either secondary or primary GBMs

There are two types of GBM tumors, primary and secondary, each resulting from a unique set of genetic and molecular alterations in glial cells [89]. Up arrows indicate genes that are over-expressed or molecular products that are activated or amplified in GBM cells, and down arrows indicate under-expression or loss of genes. LOH 10q refers to loss of heterozygosity on chromosome 10q, seen in the development of both tumor types [90].

1.11 Immune system use overlap in HIV and GBM

In both HIV infection of the brain and GBM development, cells in the host immune system are hijacked and repurposed by each disease. In both disease states, the body reacts with a heightened immune response and with an abnormally prolonged state of inflammation. In GBM development this can occur, for example, through *NFκB pathway* activation potentially due to *EGFR* amplification, and in HIV chronic inflammation can result from release of pro-inflammatory cytokines due to direct activation of lymphocytes and macrophages by the virus [91, 92]. The inflammatory response is the first response of the immune system to injury, and serves to prevent spread of infection as well as to promote healing. This system uses monocytes, an innate immune system blood cell, to respond quickly to pathological processes. In a matter of hours following immune system activation, monocytes can be found at the source of infection, actively differentiating into macrophages toward elicitation of an immune response including antigen presentation, phagocytosis, and cytokine production. While this is a crucial response, when inflammation becomes chronic, it has been shown to drive a number of maladies including rheumatoid arthritis, diabetes, depression, heart disease, stroke, and more recently, cancer [93].

Infection by HIV and development of GBM both use the innate immune system in highly similar ways, and as the main inflammatory cell of the CNS, microglia are key to neurocognitive impairment in both diseases. In GBM development, activation of microglial receptors triggers the innate immune response and attracts microglial cells, resident macrophages in the brain, to the glioma [94]. Consequently, microglial density in a glioma is positively correlated with tumor malignancy and invasiveness [95]. In HIV-1

infection, circulating monocytes, some of which are infected, directly access the CNS through the BBB and differentiate into resident macrophages. Within a matter of days following initial infection, HIV is detectable in brain tissue. In addition to driving HIV-associated neurocognitive impairment, macrophage/microglia (MØ) also serve as latent viral reservoirs [96], and it follows that as with GBM progression, brain MØ level and brain MØ activation have been shown to also serve as indicators of HAND progression [29, 97].

1.12 Protease inhibitor effects on malignancy formation

Protease inhibitors included in HAART have been found to exhibit tumor suppression effects on a number of tumor cell lines [98]. A recent and large-scale study testing the effects of multiple protease inhibitors on tumors in the HIV-1 population demonstrated protease inhibitor (PI), Nelfinavir, to have the greatest effect on tumor inhibition [99]. Nelfinavir has two anti-cancer pathways. First, as a protease inhibitor, Nelfinavir inhibits proteosome activity and prevents protein degradation. This, in turn, triggers the *unfolded protein response (UPR) pathway*, and if sustained, UPR triggers apoptosis, programmed cell death, in the cell [100]. In mice, Nelfinavir has been shown to inhibit GBM growth through stimulation of UPR leading to apoptosis, as measured by increased expression of *CHOP*, a marker of endoplasmic reticulum (ER) stress, and activation of caspase-4, an enzyme involved in ER-stress induced apoptosis [101], both molecules involved in cell death pathways. A second study found Nelfinavir to act on GBM through the *PI3K/AKT pathway*. Normally activated by GBM through loss of tumor suppressor antagonist, *PTEN*, Nelfinavir works to inhibit AKT phosphorylation,

resulting in the downregulation of vascular endothelial growth factor (*VEGF*) and hypoxia-inducible factor 1-alpha (*HIF-1α*), key molecules in angiogenesis [102].

1.13 Overlap between HIV and GBM diseases

While the first study detailed in this dissertation made use of viral genomic data to understand the effects of brain compartmentalization on the virus, the second study used human transcriptomic data to understand viral effects on the brain from the perspective of the host. While HIV-1 individuals have an increased risk for developing cancer, the most common primary brain tumor in the general population, GBM, is rarely reported in the HIV-1 population [71]. The fact that we have limited data to fully attribute this incidence rate to HAART use, in addition to the well-documented impediment to protease inhibitor entry into the brain due to presence of the BBB, led us to hypothesize a potential role for HIV infection on GBM development in HIV-infected individuals [71, 103]. In an attempt to uncover biological mechanisms that might account for this lowered incidence, the second study explored links between HIV-1 infection of the brain and GBM development. Using a comparative network analysis approach, we examined biological functions that were either significantly enriched or significantly perturbed in both disease states.

As we hypothesized, both diseases had profound effects on the host immune system and on the host nervous system. In particular, the immune system was the most heavily implicated in both, with immune system signaling pathway, *antigen processing and presentation*, significantly enriched and activated in both diseases. While we anticipated activation of cancer pathways in the GBM group, we made the unanticipated discovery of

cancer pathway enrichment and perturbation in the HIV-infected brain. The three cancer-signaling pathways, *transcriptional misregulation in cancer*, *pathways in cancer*, and *glioma*, were activated in both diseases, HIV and GBM. Of particular interest, however, was a difference between the two diseases in their activation of cancer pathway, *cell death of glioma cells* (**Figure 1.9**). The *cell death of glioma cells pathway* has been shown to involve five genes: B-cell CLL/lymphoma 2 (*BCL2*) [104, 105], C-X-C motif chemokine 12 (*CXCL12*) [106], eukaryotic translation initiation factor 2-alpha kinase (*EIF2AK3*) [107], Fas cell surface death (*FAS*) [108, 109], and tumor necrosis factor (ligand) superfamily, member 10 (*TNFSF10*) [110, 111]. These five molecules were all significantly differentially expressed in the HIV+ group as compared to the control group, and this pathway was activated in this group. In the GBM group, however, only three of these molecules exhibited differential expression and appropriate directionality toward pathway activation, and the *glioma cell death pathway* was therefore not activated in this group. Upstream regulator analysis of the two genes differentially expressed between GBM and HIV, *BCL2* and *CXCL12*, showed evidence of differential regulation of the *glioma cell death pathway* by GBM and HIV that might provide insight into differences in GBM signaling pathway regulatory control between the two diseases.

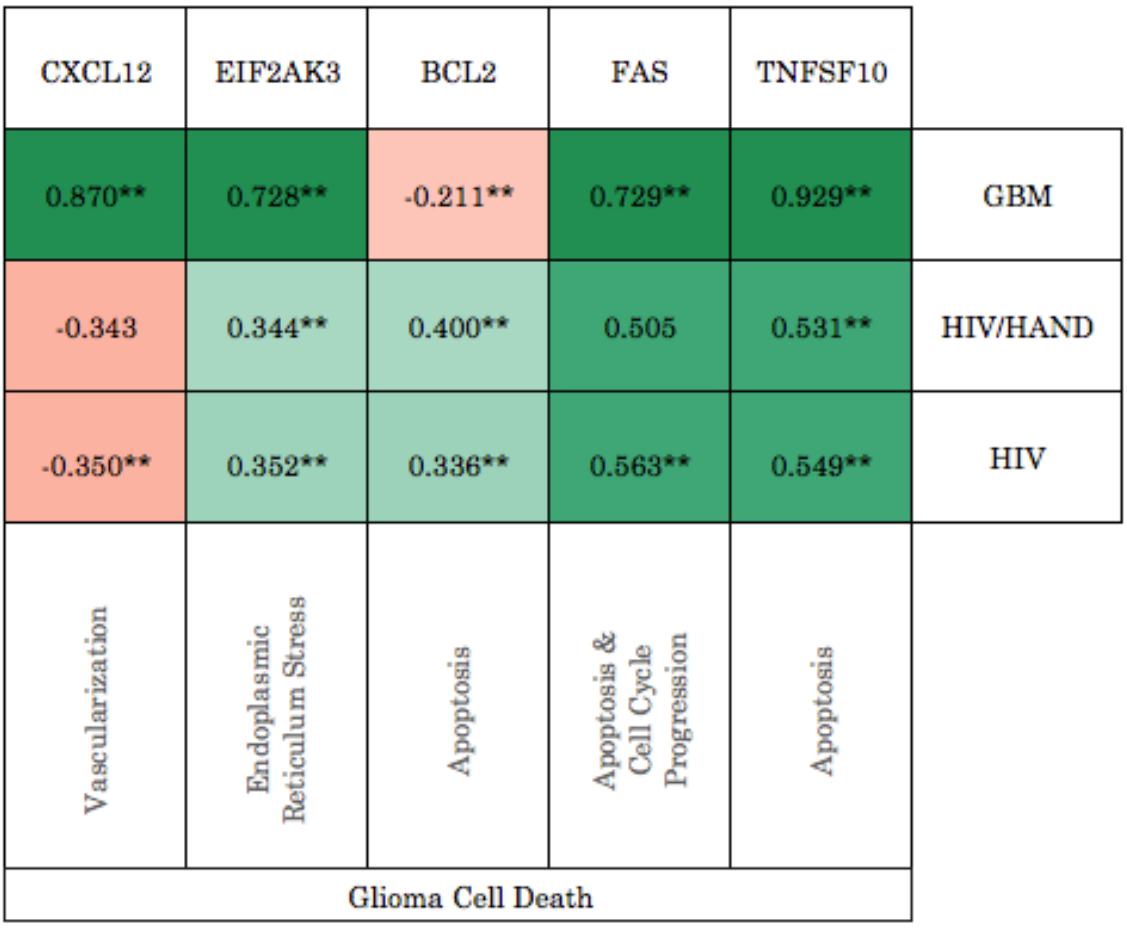


Figure 1.9 Activation of molecules in the cell death of *glioma cell death pathway* across the three groups, HIV+, HIV+/HAND+, and GBM+

Genes *BCL2*, *CXCL12*, *EIF2AK3*, *FAS*, and *TNFSF10* were all significantly differentially expressed in our HIV+ group, and based on both significance and directionality of this differential expression, the *glioma cell death pathway* was predicted as activated in this group. In the GBM+ group, the *glioma cell death pathway* was not predicted as activated, with only three of these genes both significantly differentially expressed and with the appropriate directionality. Red squares represent down-regulated genes as compared to the control group, and green squares represent up-regulated genes as compared to the control group.

Greater color saturation indicates a greater log fold change, and less color saturation indicates a smaller log fold change. Significant log fold changes (B-H adjusted p-value < 0.05) are marked with two asterisks.

1.14 References

1. Hagen JB: **The Origins of Bioinformatics**. *Nature Reviews Genetics* 2000, **1**(3):231-236.
2. Eck, R. V., and M. O. Dayhoff: **Atlas of Protein Sequence and Structure** *National Biomedical Research Foundation, Springs Silver, Maryland* (1966).
3. Hesper B, Hogeweg P: **Bioinformatica: Een Werkconcept**. *Kameleon* 1970, **1**(6):28-29.
4. Rubinstein A, Chor B: **Computational Thinking in Life Science Education**. *PLoS Computational Biology* 2014, **10**(11):e1003897.
5. Hogeweg P: **The Roots of Bioinformatics in Theoretical Biology**. *PLoS Computational Biology* 2011, **7**(3):e1002021.
6. WHO: **Fact Sheet #360**. In: *HIV/AIDS Fact Sheet #360.* vol. 2015. World Health Organization: World Health Organization; 2014.
7. Frankel AD, Young JA: **HIV-1: Fifteen Proteins and an RNA**. *Annual Review of Biochemistry* 1998, **67**(1):1-25.
8. Dickerson JE, Pinney JW, Robertson DL: **The Biological Context of HIV-1 Host Interactions Reveals Subtle Insights into a System Hijack**. *BMC Systems Biology* 2010, **4**(1):80.
9. Barré-Sinoussi F, Chermann J-C, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vézinet-Brun F, Rouzioux C: **Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS)**. *Science* 1983, **220**(4599):868-871.

10. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, Palker TJ, Redfield R, Oleske J, Safai B: **Frequent Detection and Isolation of Cytopathic Retroviruses (HTLV-III) from Patients with AIDS and at Risk for AIDS**. *Science* 1984, **224**(4648):500-503.

11. Fischl MA, Richman DD, Grieco MH, Gottlieb MS, Volberding PA, Laskin OL, Leedom JM, Groopman JE, Mildvan D, Schooley RT: **The Efficacy of Azidothymidine (AZT) in the Treatment of Patients with AIDS and AIDS-Related Complex**. *New England Journal of Medicine* 1987, **317**(4):185-191.

12. Carpenter CC, Fischl MA, Hammer SM, Hirsch MS, Jacobsen DM, Katzenstein DA, Montaner JS, Richman DD, Saag MS, Schooley RT: **Antiretroviral Therapy for HIV Infection in 1998**. *JAMA: The Journal of the American Medical Association* 1998, **280**(1):78-86.

13. Carpenter CC, Cooper DA, Fischl MA, Gatell JM, Gazzard BG, Hammer SM, Hirsch MS, Jacobsen DM, Katzenstein DA, Montaner JS: **Antiretroviral Therapy in Adults: Updated Recommendations of the International AIDS Society–USA Panel**. *JAMA: The Journal of the American Medical Association* 2000, **283**(3):381-390.

14. Al-Harthi L, Siegel J, Spritzler J, Pottage J, Agnoli M, Landay A: **Maximum Suppression of HIV Replication Leads to the Restoration of HIV-Specific Responses in Early HIV Disease**. *AIDS* 2000, **14**(7):761-770.

15. Paul RH, Sacktor NC, Valcour V, Tashima KT: **HIV and the Brain: New Challenges in the Modern Era**: Springer Science & Business Media; 2009.

16. Horowitz S, Benson D, Gottlieb M, Davos I, Bentson J: **Neurological Complications of Gay-Related Immunodeficiency Disorder**. *Annals of Neurology* 1982, 12(1):80.

17. Koenig S, Gendelman HE, Orenstein JM, Dal Canto MC, Pezeshkpour GH, Yungbluth M, Janotta F, Aksamit A, Martin MA, Fauci AS: **Detection of AIDS Virus in Macrophages in Brain Tissue from AIDS Patients with Encephalopathy**. *Science* 1986, **233**(4768):1089-1093.

18. Vazeux R, Brousse N, Jarry A, Henin D, Marche C, Vedrenne C, Mikol J, Wolff M, Michon C, Rozenbaum W: **AIDS Subacute Encephalitis: Identification of HIV-Infected Cells**. *The American Journal of Pathology* 1987, **126**(3):403.

19. Koyanagi Y, Miles S, Mitsuyasu RT, Merrill JE, Vinters HV, Chen I: **Dual Infection of the Central Nervous System by AIDS Viruses with Distinct Cellular Tropisms**. *Science* 1987, **236**(4803):819-822.

20. Lawrence DM, Durham LC, Schwartz L, Seth P, Maric D, Major EO: **Human Immunodeficiency Virus Type 1 Infection of Human Brain-Derived Progenitor Cells**. *Journal of Virology* 2004, **78**(14):7319-7328.

21. Pardridge WM: **The Blood-Brain Barrier: Bottleneck in Brain Drug Development**. *NeuroRx* 2005, **2**(1):3-14.

22. Kandanearatchi A, Williams B, Everall IP: **Assessing the Efficacy of Highly Active Antiretroviral Therapy in the Brain**. *Brain Pathology* 2003, **13**(1):104-110.

23. Trujillo JR: **Model of HIV-1 Neurotropism**. Wikimedia Commons; 2004.

24. Kaul M, Garden GA, Lipton SA: **Pathways to Neuronal Injury and Apoptosis in HIV-Associated Dementia**. *Nature* 2001, **410**(6831):988-994.

25. Strizki JM, Albright AV, Sheng H, O'Connor M, Perrin L, Gonzalez-Scarano F: **Infection of Primary Human Microglia and Monocyte-Derived Macrophages with Human Immunodeficiency Virus Type 1 Isolates: Evidence of Differential Tropism**. *Journal of Virology* 1996, **70**(11):7654-7662.

26. Geleziunas R, Schipper HM, Wainberg MA: **Pathogenesis and Therapy of HIV-1 Infection of the Central Nervous System**. *AIDS* 1992, **6**(12):1411-1426.

27. Vanzani MC, Iacono RF, Caccuri RL, Troncoso AR, Berria MI: **Regional Differences in Astrocyte Activation in HIV-Associated Dementia**. *Medicina-Buenos Aires* 2006, **66**(2):108.

28. Dunfee R, Thomas ER, Gorry PR, Wang J, Ancuta P, Gabuzda D: **Mechanisms of HIV-1 Neurotropism**. *Current HIV Research* 2006, **4**(3):267-278.

29. McArthur JC, Steiner J, Sacktor N, Nath A: **Human Immunodeficiency Virus‐Associated Neurocognitive Disorders: Mind the Gap**. *Annals of Neurology* 2010, **67**(6):699-714.

30. Heaton R, Clifford D, Franklin D, Woods S, Ake C, Vaida F, Ellis R, Letendre S, Marcotte T, Atkinson J: **HIV-Associated Neurocognitive Disorders Persist in the Era of Potent Antiretroviral Therapy Charter Study**. *Neurology* 2010, **75**(23):2087-2096.

31. Chang L, Ernst T, Witt MD, Ames N, Walot I, Jovicich J, DeSilva M, Trivedi N, Speck O, Miller EN: **Persistent Brain Abnormalities in Antiretroviral-Naive HIV Patients 3 Months after HAART**. *Antiviral Therapy* 2003, **8**(1):17-26.

32. Langford T, Letendre S, Larrea G, Masliah E: **Changing Patterns in the Neuropathogenesis of HIV During the HAART Era**. *Brain Pathology* 2003, **13**(2):195-210.

33. Williams DW, Eugenin EA, Calderon TM, Berman JW: **Monocyte Maturation, HIV Susceptibility, and Transmigration across the Blood Brain Barrier Are Critical in HIV Neuropathogenesis**. *Journal of Leukocyte Biology* 2012, **91**(3):401-415.

34. Neuenburg JK, Brodt HR, Herndier BG, Bickel M, Bacchetti P, Price RW, Grant RM, Schlote W: **HIV-Related Neuropathology, 1985 to 1999: Rising Prevalence of HIV Encephalopathy in the Era of Highly Active Antiretroviral Therapy**. *Journal of Acquired Immune Deficiency Syndromes (1999)* 2002, **31**(2):171-177.

35. Heaton RK, Grant I, Butters N, White DA, Kirson D, Atkinson JH, McCutchan JA, Taylor MJ, Kelly MD, Ellis RJ: **The HNRC 500-Neuropsychology of HIV Infection at Different Disease Stages**. *Journal of the International Neuropsychological Society* 1995, **1**(03):231-251.

36. McArthur J: **Update on the Neurological Manifestations of HIV**. *The PRN Notebook* 2005, **10**(3):6.

37. Heaton RK, Franklin DR, Ellis RJ, McCutchan JA, Letendre SL, LeBlanc S, Corkran SH, Duarte NA, Clifford DB, Woods SP: **HIV-Associated Neurocognitive Disorders Before and During the Era of Combination Antiretroviral Therapy: Differences in Rates, Nature, and Predictors**. *Journal of Neurovirology* 2011, **17**(1):3-16.

38. Coffin JM: **HIV Population Dynamics in Vivo: Implications for Genetic Variation, Pathogenesis, and Therapy**. *Science* 1995, **267**(5197):483-489.

39. Brown RJ, Peters PJ, Caron C, Gonzalez-Perez MP, Stones L, Ankghuambom C, Pondei K, McClure CP, Alemnji G, Taylor S: **Intercompartmental Recombination of HIV-1 Contributes to Env Intrahost Diversity and Modulates Viral Tropism and Sensitivity to Entry Inhibitors**. *Journal of Virology* 2011, **85**(12):6024-6037.

40. Korber B, Kunstman KJ, Patterson BK, Furtado M, McEvilly MM, Levy R, Wolinsky SM: **Genetic Differences between Blood-and Brain-Derived Viral Sequences from Human Immunodeficiency Virus Type 1-Infected Patients: Evidence of Conserved Elements in the V3 Region of the Envelope Protein of Brain-Derived Sequences**. *Journal of Virology* 1994, **68**(11):7467-7481.

41. Ball JK, Holmes EC, Whitwell H, Desselberger U: **Genomic Variation of Human Immunodeficiency Virus Type 1 (HIV-1): Molecular Analyses of HIV-1 in Sequential Blood Samples and Various Organs Obtained at Autopsy**. *Journal of General Virology* 1994, **75**(4):867-880.

42. Wong JK, Ignacio CC, Torriani F, Havlir D, Fitch N, Richman DD: **In Vivo Compartmentalization of Human Immunodeficiency Virus: Evidence from the Examination of Pol Sequences from Autopsy Tissues**. *Journal of Virology* 1997, **71**(3):2059-2071.

43. Chen H, Wood C, Petito CK: **Comparisons of HIV-1 Viral Sequences in Brain, Choroid Plexus and Spleen: Potential Role of Choroid Plexus in the Pathogenesis of HIV Encephalitis**. *Journal of Neurovirology* 2000, **6**(6):498-506.

44. McCrossan M, Marsden M, Carnie F, Minnis S, Hansoti B, Anthony I, Brettle R, Bell J, Simmonds P: **An Immune Control Model for Viral Replication in the CNS During Presymptomatic HIV Infection**. *Brain* 2006, **129**(2):503-516.

45. Pillai SK, Pond SLK, Liu Y, Good BM, Strain MC, Ellis RJ, Letendre S, Smith DM, Günthard HF, Grant I: **Genetic Attributes of Cerebrospinal Fluid-Derived HIV-1 Env**. *Brain* 2006, **129**(7):1872-1883.

46. Salemi M, Lamers SL, Yu S, De Oliveira T, Fitch WM, McGrath MS: **Phylodynamic Analysis of Human Immunodeficiency Virus Type 1 in Distinct Brain Compartments Provides a Model for the Neuropathogenesis of AIDS**. *Journal of virology* 2005, **79**(17):11343-11352.

47. Reddy RT, Achim CL, Sirko DA, Tehranchi S, Kraus FG, Wong-Staal F, Wiley CA: **Sequence Analysis of the V3 Loop in Brain and Spleen of Patients with HIV Encephalitis**. *AIDS Research and Human Retroviruses* 1996, **12**(6):477-482.

48. Smith KM, Crandall KA, Kneissl ML, Navia BA: **PCR Detection of Host and HIV -1 Sequences from Archival Brain Tissue**. *Journal of Neurovirology* 2000, **6**(2):164-171.

49. Smit TK, Wang B, Ng T, Osborne R, Brew B, Saksena NK: **Varied Tropism of HIV-1 Isolates Derived from Different Regions of Adult Brain Cortex Discriminate between Patients with and without AIDS Dementia Complex (ADC): Evidence for Neurotropic HIV Variants**. *Virology* 2001, **279**(2):509-526.

50. Beebe AM, Dua N, Faith TG, Moore PF, Pedersen NC, Dandekar S: **Primary Stage of Feline Immunodeficiency Virus Infection: Viral Dissemination and Cellular Targets**. *Journal of Virology* 1994, **68**(5):3080-3091.

51. Brew BJ, Evans L, Byrne C, Pemberton L, Hurren L: **The Relationship between AIDS Dementia Complex and the Presence of Macrophage Tropic and Non Syncytium Inducing Isolates of Human Immunodeficiency Virus Type 1 in the Cerebrospinal Fluid**. *Journal of Neurovirology* 1996, **2**(3):152-157.

52. Hughes E, Bell J, Simmonds P: **Investigation of the Dynamics of the Spread of Human Immunodeficiency Virus to Brain and Other Tissues by Evolutionary Analysis of Sequences from the P17gag and Env Genes**. *Journal of Virology* 1997, **71**(2):1272-1280.

53. Zhang K, Hawken M, Rana F, Welte FJ, Gartner S, Goldsmith MA, Power C: **Human Immunodeficiency Virus Type 1 Clade A and D Neurotropism: Molecular Evolution, Recombination, and Coreceptor Use**. *Virology* 2001, **283**(1):19-30.

54. Strain M, Letendre S, Pillai S, Russell T, Ignacio C, Günthard H, Good B, Smith D, Wolinsky S, Furtado M: **Genetic Composition of Human Immunodeficiency Virus Type 1 in Cerebrospinal Fluid and Blood without Treatment and**

During Failing Antiretroviral Therapy. *Journal of Virology* 2005, **79**(3):1772-1788.

55. Power C, McArthur JC, Johnson RT, Griffin DE, Glass JD, Perryman S, Chesebro B: **Demented and Nondemented Patients with AIDS Differ in Brain-Derived Human Immunodeficiency Virus Type 1 Envelope Sequences**. *Journal of Virology* 1994, **68**(7):4643-4649.

56. Holman AG, Gabuzda D: **A Machine Learning Approach for Identifying Amino Acid Signatures in the HIV Env Gene Predictive of Dementia**. *PloS One* 2012, **7**(11):e49538.

57. Di Stefano M, Gray F, Leitner T, Chiodi F: **Analysis of Env V3 Sequences from HIV‐1‐Infected Brain Indicates Restrained Virus Expression Throughout the Disease**. *Journal of Medical Virology* 1996, **49**(1):41-48.

58. Dunfee RL, Thomas ER, Wang J, Kunstman K, Wolinsky SM, Gabuzda D: **Loss of the N-Linked Glycosylation Site at Position 386 in the HIV Envelope V4 Region Enhances Macrophage Tropism and is Associated with Dementia**. *Virology* 2007, **367**(1):222-234.

59. Lamers SL, Poon AF, McGrath MS: **HIV-1 Nef Protein Structures Associated with Brain Infection and Dementia Pathogenesis**. *PloS One* 2011, **6**(2):e16659.

60. Lamers SL, Salemi M, Galligan DC, Morris A, Gray R, Fogel G, Zhao L, McGrath MS: **Human Immunodeficiency Virus-1 Evolutionary Patterns Associated with Pathogenic Processes in the Brain**. *Journal of Neurovirology* 2010, **16**(3):230-241.

61. Holman AG, Mefford ME, O'Connor N, Gabuzda D: **HIVBrainSeqDB: A Database of Annotated HIV Envelope Sequences from Brain and Other Anatomical Sites**. *AIDS Research and Therapy.* vol. 7; 2010: 43.

62. Zhang M, Foley B, Schultz A-K, Macke JP, Bulla I, Stanke M, Morgenstern B, Korber B, Leitner T: **The Role of Recombination in the Emergence of a Complex and Dynamic HIV Epidemic**. *Retrovirology* 2010, **7**(1):25.

63. Shannon CE: **A Mathematical Theory of Communication**. *ACM SIGMOBILE Mobile Computing and Communications Review* 2001, **5**(1):3-55.

64. Schug J, Schuller W-P, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ: **Promoter Features Related to Tissue Specificity as Measured by Shannon Entropy**. *Genome Biology* 2005, **6**(4):R33.

65. LANL: **Shannon Entropy-Two**. http://www.hiv.lanl.gov.

66. Julien JP, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, Klasse PJ, Burton DR, Sanders RW, Moore JP *et al*: **Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer**. *Science* 2013, **342**(6165):1477-1483.

67. Hasse B, Ledergerber B, Furrer H, Battegay M, Hirschel B, Cavassini M, Bertisch B, Bernasconi E, Weber R: **Morbidity and Aging in HIV-Infected Persons: The Swiss HIV Cohort Study**. *Clinical Infectious Diseases* 2011, **53**(11):1130-1139.

68. Deeks SG, Phillips AN: **HIV Infection, Antiretroviral Treatment, Ageing, and Non-AIDS Related Morbidity**. *British Medical Journal* 2009, **338**(7689):288-292.

69. Kohler BA, Ward E, McCarthy BJ, Schymura MJ, Ries LA, Eheman C, Jemal A, Anderson RN, Ajani UA, Edwards BK: **Annual Report to the Nation on the Status of Cancer, 1975-2007, Featuring Tumors of the Brain and Other Nervous System**. *Journal of the National Cancer Institute* 2011, **103**(9):714-736.

70. Angeletti PC, Zhang L, Wood C: **The Viral Etiology of AIDS-Associated Malignancies**. *Advances in Pharmacology* 2008, **56**:509-557.

71. Cedeno-Laurent F, Trujillo JR: **Gliomas and Brain Lymphomas in HIV-1/AIDS Patients: Reflections from a 20-Year Follow up in Mexico and Brazil**. *Microbiology Research* 2011, **2**(1):e11.

72. Maxwell M, Galanopoulos T, Neville-Golden J, Antoniades HN: **Effect of the Expression of Transforming Growth Factor-B2 in Primary Human Glioblastomas on Immunosuppression and Loss of Immune Surveillance**. *Journal of Neurosurgery* 1992, **76**(5):799-804.

73. Moulignier A, Mikol J, Pialoux G, Eliaszewicz M, Thurel C, Thiebaut JB: **Cerebral Glial Tumors and Human Immunodeficiency Virus-1 Infection: More Than a Coincidental Association**. *Cancer* 1994, **74**(2):686-692.

74. Dolecek TA, Propp JM, Stroup NE, Kruchko C: **CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2005–2009**. *Neuro-oncology* 2012, **14**(suppl 5):v1-v49.

75. Adamson C, Kanu OO, Mehta AI, Di C, Lin N, Mattox AK, Bigner DD: **Glioblastoma Multiforme: A Review of Where We Have Been and Where We Are Going**. *Expert Opin Investig Drugs* 2009, **18**(8):1061-1083.

76. Parsons DW, Jones S, Zhang X, Lin JC-H, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu I-M, Gallia GL: **An Integrated Genomic Analysis of Human Glioblastoma Multiforme**. *Science* 2008, **321**(5897):1807-1812.

77. Rasheed BA, Wiltshire RN, Bigner SH, Bigner DD: **Molecular Pathogenesis of Malignant Gliomas**. *Current Opinion in Oncology* 1999, **11**(3):162.

78. Nagane M, Levitzki A, Gazit A, Cavenee WK, Huang H-JS: **Drug Resistance of Human Glioblastoma Cells Conferred by a Tumor-Specific Mutant Epidermal Growth Factor Receptor through Modulation of BCL-XL and Caspase-3-Like Proteases**. *Proceedings of the National Academy of Sciences* 1998, **95**(10):5724-5729.

79. Fulci G, Labuhn M, Maier D, Lachat Y, Hausmann O, Hegi ME, Janzer RC, Merlo A, Van Meir EG: **P53 Gene Mutation and Ink4a-Arf Deletion Appear to Be Two Mutually Exclusive Events in Human Glioblastoma**. *Oncogene* 2000, **19**(33):3816-3822.

80. Nakada M, Kita D, Watanabe T, Hayashi Y, Teng L, Pyko IV, Hamada J-I: **Aberrant Signaling Pathways in Glioma**. *Cancers* 2011, **3**(3):3242-3278.

81. Hartmann C, Bartels G, Gehlhaar C, Holtkamp N, von Deimling A: **PIK3CA Mutations in Glioblastoma Multiforme**. *Acta Neuropathologica* 2005, **109**(6):639-642.

82. Guha A, Feldkamp MM, Lau N, Boss G, Pawson A: **Proliferation of Human Malignant Astrocytomas is Dependent on Ras Activation**. *Oncogene* 1997, **15**(23):2755-2765.

83. Holmen SL, Williams BO: **Essential Role for Ras Signaling in Glioblastoma Maintenance**. *Cancer Research* 2005, **65**(18):8250-8255.

84. Maehama T, Dixon JE: **PTEN: A Tumour Suppressor That Functions as a Phospholipid Phosphatase**. *Trends in Cell Biology* 1999, **9**(4):125-128.

85. Endersby R, Baker S: **PTEN Signaling in Brain: Neuropathology and Tumorigenesis**. *Oncogene* 2008, **27**(41):5416-5430.

86. Kleihues P, Ohgaki H: **Primary and Secondary Glioblastomas: From Concept to Clinical Diagnosis**. *Neuro-oncology* 1999, **1**(1):44-51.

87. He J, Olson JJ, James CD: **Lack of P16ink4 or Retinoblastoma Protein (PRB), or Amplification-Associated Overexpression of CDK4 is Observed in Distinct Subsets of Malignant Glial Tumors and Cell Lines**. *Cancer Research* 1995, **55**(21):4833-4836.

88. Ishii N, Maier D, Merlo A, Tada M, Sawamura Y, Diserens AC, Meir EG: **Frequent Co‐Alterations of TP53, P16/CDKN2A, P14ARF, PTEN Tumor Suppressor Genes in Human Glioma Cell Lines**. *Brain Pathology* 1999, **9**(3):469-479.

89. Ohgaki H, Dessen P, Jourde B, Horstmann S, Nishikawa T, Di Patre P-L, Burkhard C, Schüler D, Probst-Hensch NM, Maiorka PC: **Genetic Pathways to Glioblastoma a Population-Based Study**. *Cancer Research* 2004, **64**(19):6892-6899.

90. Fujisawa H, Reis RM, Nakamura M, Colella S, Yonekawa Y, Kleihues P, Ohgaki H: **Loss of Heterozygosity on Chromosome 10 is More Extensive in Primary**

(De Novo) than in Secondary Glioblastomas. *Laboratory Investigation* 2000, **80**(1):65-72.

91. Nogueira L, Ruiz-Ontañon P, Vazquez-Barquero A, Moris F, Fernandez-Luna JL: **The Nfκb Pathway: A Therapeutic Target in Glioblastoma**. *Oncotarget* 2011, **2**(8):646.

92. Appay V, Sauce D: **Immune Activation and Inflammation in HIV -1 Infection: Causes and Consequences**. *The Journal of Pathology* 2008, **214**(2):231-241.

93. Dantzer R, O'Connor JC, Freund GG, Johnson RW, Kelley KW: **From Inflammation to Sickness and Depression: When the Immune System Subjugates the Brain**. *Nature Reviews Neuroscience* 2008, **9**(1):46-56.

94. Vinnakota K, Hu F, Ku M-C, Georgieva PB, Szulzewsky F, Pohlmann A, Waiczies S, Waiczies H, Niendorf T, Lehnardt S: **Toll-Like Receptor 2 Mediates Microglia/Brain Macrophage MT1-MMP Expression and Glioma Expansion**. *Neuro-oncology* 2013:not115.

95. Markovic D, Vinnakota K, Chirasani S, Synowitz M, Raguet H, Stock K, Sliwa M, Lehmann S, Kälin R, Van Rooijen N: **Gliomas Induce and Exploit Microglial MT1-MMP Expression for Tumor Expansion**. *Proceedings of the National Academy of Sciences* 2009, **106**(30):12530-12535.

96. Minagar A, Shapshak P, Fujimura R, Ownby R, Heyes M, Eisdorfer C: **The Role of Macrophage/Microglia and Astrocytes in the Pathogenesis of Three Neurologic Disorders: HIV-Associated Dementia, Alzheimer Disease, and Multiple Sclerosis**. *Journal of the Neurological Sciences* 2002, **202**(1):13-23.

97. Ellis R, Langford D, Masliah E: **HIV and Antiretroviral Therapy in the Brain: Neuronal Injury and Repair**. *Nature Reviews Neuroscience* 2007, **8**(1):33-44.

98. Crum-Cianflone NF, Hullsiek KH, Marconi V, Weintrob A, Ganesan A, Barthel RV, Fraser S, Roediger MP, Agan B, Wegner S: **The Impact of Nelfinavir Exposure on Cancer Development among a Large Cohort of HIV-Infected Patients**. *Journal of Acquired Immune Deficiency Syndrome* 2009, **51**(3):305-309.

99. Bernstein WB, Dennis PA: **Repositioning HIV Protease Inhibitors as Cancer Therapeutics**. *Curr Opin HIV AIDS* 2008, **3**(6):666-675.

100. Johnson GG, White MC, Grimaldi M: **Stressed to Death: Targeting Endoplasmic Reticulum Stress Response Induced Apoptosis in Gliomas**. *Current Pharmaceutical Design* 2011, **17**(3):284.

101. Pyrko P, Kardosh A, Wang W, Xiong W, Schönthal AH, Chen TC: **HIV-1 Protease Inhibitors Nelfinavir and Atazanavir Induce Malignant Glioma Death by Triggering Endoplasmic Reticulum Stress**. *Cancer Research* 2007, **67**(22):10920-10928.

102. Pore N, Gupta AK, Cerniglia GJ, Maity A: **HIV Protease Inhibitors Decrease Vegf/Hif-1alpha Expression and Angiogenesis in Glioblastoma Cells**. *Neoplasia* 2006, **8**(11):889-895.

103. Marzolini C, Mueller R, Li-Blatter X, Battegay M, Seelig A: **The Brain Entry of HIV-1 Protease Inhibitors is Facilitated When Used in Combination**. *Molecular Pharmaceutics* 2013, **10**(6):2340-2349.

104. Shinoura N, Yoshida Y, Nishimura M, Muramatsu Y, Asai A, Kirino T, Hamada H: **Expression Level of BCL-2 Determines Anti-or Proapoptotic Function**. *Cancer Research* 1999, **59**(16):4119-4128.

105. Strack PR, Frey MW, Rizzo CJ, Cordova B, George HJ, Meade R, Ho SP, Corman J, Tritch R, Korant BD: **Apoptosis Mediated by HIV Protease is Preceded by Cleavage of BCL-2**. *Proceedings of the National Academy of Sciences* 1996, **93**(18):9571-9576.

106. Ping Yf, Yao Xh, Jiang Jy, Zhao Lt, Yu Sc, Jiang T, Lin M, Chen Jh, Wang B, Zhang R: **The Chemokine CXCL12 and its Receptor CXCR4 Promote Glioma Stem Cell‐Mediated VEGF Production and Tumour Angiogenesis Via PI3K/AKT Signalling**. *The Journal of Pathology* 2011, **224**(3):344-354.

107. Gupta AK, Li B, Cerniglia GJ, Ahmed MS, Hahn SM, Maity A: **The HIV Protease Inhibitor Nelfinavir Downregulates Akt Phosphorylation by Inhibiting Proteasomal Activity and Inducing the Unfolded Protein Response**. *Neoplasia* 2007, **9**(4):271-278.

108. Frankel B, Longo SL, Kyle M, Canute GW, Ryken TC: **Tumor Fas (Apo-1/Cd95) up-Regulation Results in Increased Apoptosis and Survival Times for Rats with Intracranial Malignant Gliomas**. *Neurosurgery* 2001, **49**(1):168-175; discussion 175-166.

109. Ambar BB, Frei K, Malipiero U, Morelli AE, Castro MG, Lowenstein PR, Fontana A: **Treatment of Experimental Glioma by Administration of Adenoviral Vectors Expressing Fas Ligand**. *Human Gene Therapy* 1999, **10**(10):1641-1648.

110. Dixit D, Sharma V, Ghosh S, Mehta V, Sen E: **Inhibition of Casein Kinase-2 Induces P53-Dependent Cell Cycle Arrest and Sensitizes Glioblastoma Cells to Tumor Necrosis Factor (TNFα)-Induced Apoptosis through SIRT1 Inhibition**. *Cell Death & Disease* 2012, **3**(2):e271.

111. Weinmann L, Wischhusen J, Demma M, Naumann U, Roth P, Dasmahapatra B, Weller M: **A Novel P53 Rescue Compound Induces P53-Dependent Growth Arrest and Sensitises Glioma Cells to Apo2L/Trail-Induced Apoptosis**. *Cell Death & Differentiation* 2008, **15**(4):718-729.

CHAPTER 2

NOVEL HIV-1 ENVELOPE AMINO ACID RESIDUES LINKED TO HIV-1-ASSOCIATED NEUROCOGNITIVE DISORDER[1]

[1]Griffin, T.Z. and Zhang, M. Submitted to *Genome Biology*, 08/02/14.

2.1 Abstract

Background: HIV-1 entry into the brain occurs early in infection, and in nearly half of all infected individuals results in a range of HIV-1-associated neurocognitive disorders (HAND). Prior research on HIV-1 evolution in the brain has provided insight into HAND pathogenesis and has shown the potential for advancement in the area. HIV-1 sequences derived from HAND-assessed individuals, however, are extremely limited, making it difficult to study viral evolution in the HAND environment. **Methods:** Here we compiled and analyzed viral sequence and clinical data from prior research on individuals assessed for HAND, providing an enhanced sample to identify both brain- and HAND-specific viral evolutionary events. **Results:** Testing for impact of HAND on viral evolution revealed 67 HAND-associated amino acid residues within the viral envelope (*env*) genomic region ($p < 0.05$), 56 of which had not been previously reported. Testing for impact of tissue type on viral evolution showed intra-individual sequences, regardless of tissue compartmentalization, as homogeneous. While tissue type did not affect clustering of the *env* gene, sequences derived from brain tissue did exhibit increased diversity as compared to both cerebrospinal fluid and blood sequences ($p < 0.005$). **Conclusions:** Our finding of novel HAND-associated *env* signatures helps advance current understanding of viral evolution within the HAND brain, and points to the potential for genomic regions beyond *env* to also exhibit HAND signatures. Most importantly, this study allowed us to systematically and rigorously test over three times the *env* region than had been possible prior to the development of our HAND resource, and to account for biases related to what we found to be the primary viral sequence clustering mechanism, clustering by individual.

2.2 Background

Although great strides have been made in extending and improving the lives of those with human immunodeficiency virus (HIV), our ability to control viral damage to the HIV individual's brain remains highly limited [1]. Poor penetration of anti-retroviral therapy through the blood brain barrier, and into the central nervous system (CNS), leads to limited external control over viral replication in the brain, where macrophages, microglial cells, and astrocytes have been shown to act as long-term HIV reservoirs and as intermediates for viral damage [2-4]. This challenge is coupled with the potential for even low levels of viral protein products to damage both neurons and the synapses between them where communication occurs [5-7]. In the recent and large-scale CNS HIV Anti-Retroviral Therapy Effects Research Study, nearly half of all HIV-1-infected individuals sampled showed some level of HIV-1-associated neurocognitive disorder (HAND) [8]. Individuals with HAND frequently exhibit deficits in learning, verbal fluency, motor function, and memory. In those areas of the world most affected by HIV and least able to afford anti-retroviral therapy, the most severe form of HAND, HIV-1-associated dementia (HAD), continues to devastate the mental health of untreated individuals [9]. Ultimately, HIV-1 individuals with HAND suffer from a greatly reduced quality of life, and compared to mentally healthy HIV-1 individuals, are at a threefold increased risk of death [10, 11].

Following early clinical observations of acquired immune deficiency syndrome (AIDS), high rates of severe neurocognitive disorders in the HIV population became a growing source of concern [12-15]. Research on the effects of virus in the brain began with the discovery of viral RNA and proteins in brain tissue, and of active transcription

and translation of virus within brain-derived macrophage cells [16-19]. The incidence rate of HIV-1-associated dementia (HAD), has since been linked to sequence subtype, with a higher HAD rate seen in individuals with subtype D as compared to individuals with subtypes A and C [20, 21], with genetic distinction between subtypes largely attributable to variations in the HIV-1 Env protein sequence [22-24]. Due to the critical role of HIV-1 Env in both macrophage tropism and viral replication, research on brain-derived HIV-1 has focused on this region, and in particular, on the V3 loop, a region that functions in CD4+ cell entry [25]. Findings in this area, however, have been inconsistent, including evidence for and against brain-specific evolutionary pressures on the V1V2 and V3 regions of the *env* gene [26-40]. Studies that have searched the viral genome for events linked to HAND have also proven mixed. Power et al. discovered two V3 region codons that differed between HAD and non-HAD individuals, a finding that has since been both refuted and validated [41-43]. Following Power et al., additional HAND-associated amino acid residues have been discovered in both the V3 and V4 regions of the *env* gene, as well as in the *nef* gene [36, 44-46]. Research on the relationship between brain viral sequence changes and HAND pathogenesis, however, has remained largely limited. These limitations include small sample sizes, a disparity of methodologies, and biases in sampling and sequencing, and together, substantiate the need for continued study in the area.

To address those research discrepancies noted above, herein we report a systematic study of HIV-1 genome changes associated with tissue compartmentalization and with HAND status. Our study benefits from our well-curated database devoted to understanding HIV-1-associated neurocognitive disorders. This database contains up-to-

date and comprehensive viral sequence information across multiple tissue types, and representing multiple viral genomic regions, with highly detailed clinical information. HIV-1 individuals included in this database encompass the spectrum of HIV-1-associated neurocognitive disorders, allowing for insight into HAND individuals beyond the more frequently investigated dementia cases. This resource, therefore, provides a robust foundation for testing our hypothesis that there exist differences in the HIV-1 genome between HAND and non-HAND individuals, and allows for pinpointing of HAND-associated signatures. Results from this work indicate an interaction between the Env protein and the brain tissue environment, arguing for the need to consider genomic region when looking for tissue compartmentalization effects on viral evolution. Even more importantly, we identified novel signatures in the HIV gp120 region linked to HAND status. We expect these findings will help elucidate those molecular mechanisms underlying promotion and development of neurocognitive disorders in HIV-1 individuals. Finally, development of the publically available HAND Database will also allow for advanced investigation of HIV-1-associated neurocognitive disorders.

2.3 Methods

2.3.1 Data collection and screen

Sequence data were collected from the NIH Genetic Sequence Database (GenBank) (last accessed 3/2013) and the Los Alamos HIV Sequence Database (last accessed 02/2014). Clinical data, including CD4+ count, sampling geographical regions, and treatment status, were collected from the literature, public databases, and through communication with publication authors. HAND diagnosis methodologies varied

according to study. Post-mortem methodologies included evaluation of HIV-1 individual clinical history [26, 47-50] and neuropathological examination of brain tissue at autopsy [26, 32, 46-53], and pre-mortem assessments included use of the American Academy of Neurology Task Force dementia rating scale [47, 54], the Memorial Sloan-Kettering scale [55], and derivation of a mean deficit score from specialized HIV-related neuropsychological deficit tests [56].

Sequence and clinical data were compiled into a MySQL-based database (MySQL v.5.6.17). All sequences were processed through the LANL Quality Control pipeline [57] and checked for vector contamination using BLASTn [58]. A hypermutation analysis [59] was performed to exclude unusual sequences that would bias phylogenetic analyses. We also performed a jpHMM-based recombination study to examine genotyping quality [60]. To normalize sequencing depth by individual, and to remove individual-specific bias, one sequence per tissue per individual was selected for further analysis. For individuals with multiple sequences per tissue, the sequence most representative of each tissue was selected using the following two criteria. First, selected sequences needed to be, at minimum, 300 nucleotides in length. Second, a consensus sequence was developed per tissue per individual, and from those sequences that met the minimum length requirement; the sequence with the smallest Hamming distance to its tissue-specific consensus sequence was selected for further analysis [61].

2.3.2 Phylogenetic analysis

Sequences selected using the procedure above were aligned with well-characterized HIV subtype reference sequences [57]. Alignments were refined through

use of the HIV-1 N-linked Glycosylation Site Analyzer tool [62], allowing us to maintain each sequence's glycosylation signature by keeping variable loop regions as long as possible. Alignments were next manually edited using SeaView alignment software [63]. A maximum likelihood phylogenetic tree [64], using the Generalised Time-Reversible DNA sequence evolution model, was generated for 93 *env* sequences spanning HXB2 [GenBank: K03455] nucleotide positions 6737 to 7640. Selection of this region for phylogenetic analysis allowed a balance between number of individuals sampled and sequencing depth. Evolutionary model selection for this specific subset of sequences was conducted using the FindModel algorithm [57].

2.3.3 Sequence diversity analysis

F84 evolutionary distance between each sequence and its corresponding subtype reference was used as our diversity measure. A Shapiro-Wilks test was performed to determine whether a sequence set was normally distributed, and a Wilcoxon rank sum test with continuity correction was used to test for significance between tissue types, and between HAND and non-HAND individuals. For tissue comparisons involving more than one tissue group, a one-way ANOVA was performed, followed by Tukey's HSD when applicable. All statistical tests were performed using R [65]. Co-receptor usage per *env*-spanning sequence was predicted using Geno2pheno [coreceptor] [66], and predicted counts were tested for significance across individual groups using a chi-squared double classification test.

2.3.4 HAND-associated signature analysis and mapping

Shannon entropy values were calculated for residues in the HXB2 amino acid range 123 to 472. This range provided us the largest amount of sequence coverage allowable by our dataset. Significant differences between HAND and non-HAND sequences were tested using the Entropy-Two program [57], and parameters included use of amino acid class equivalents in entropy calculation and a statistical confidence value derived through a Monte Carlo randomization strategy with replacement, followed by use of the Bonferroni correction methodology to correct for family-wise error rate. Of note, our decision to use amino acid class equivalents provided a more conservative estimate of significant sites through grouping of amino acids by chemical similarities. HAND-associated signatures were selected using two significance level cutoffs, $p < .05$ and $p < .005$. In addition, frequency values for HAND and non-HAND consensus amino acids were calculated using the LANL tool, Viral Epidemiology Signature Pattern Analysis (VESPA) [57].

All significant positions ($p < .05$) obtained from our Shannon Entropy analysis were mapped using the Polyview-3D tool [67], onto the PDB Env structure 4NCO [68] as a reference protein structure. A multiple amino acid sequence alignment between 4NCO, HXB2, our HAND consensus, and our non-HAND consensus was performed using the Clustal Omega tool with default options [69-71]. This alignment provided us a one-to-one mapping between amino acids of interest and the 4NCO Env scaffold.

2.3.5 Drug resistance sites analysis

HAND-associated amino acid residues were checked against known CTL, T-helper, and antibody epitope sites obtained from the Los Alamos HIV Molecular Immunology Database [72]. In addition, all sequences were screened for drug-induced epitopes using the Stanford University Calibrated Population Resistance tool and the HIVdb Genotypic Resistance Interpretation Algorithm [73-75].

2.4 Results

2.4.1 Dataset characteristics

In total, we collected information on 5,783 neurocognitive disorder related HIV-1 sequences, representing a total of 163 unique individuals, including individuals both with and without HAND. All records are maintained and publically accessible in a MySQL database (v.5.6.17). Sample geographical region information was available for 156 individuals, with the top three regions being North America (60%), Europe (25%), and Asia (6.4%) (**Table 2.1**). Database sequences were derived from twenty different tissue types, the top three tissues being brain (47%), lymph node (14%), and CSF (7%). Additionally, five HIV-1 genes were represented in our dataset, *gag*, *pol*, *env*, *tat*, and *nef*, with the majority of coverage in the gp120 region of the *env* gene (**Figure 2.1**).

Table 2.1 Distribution of individuals sampled, across geographical region, risk factor, and HAND status

	Non-HAND	HAND	Unknown HAND
North America			
Unknown Risk Factor	37	46	2
IV Drug Use	1	2	
Sexual Transmission	1	4	
Europe			
Unknown Risk Factor	15	4	
IV Drug Use		11	3
Sexual Transmission	1	2	2
Blood Transfusion		1	
Asia			
Unknown Risk Factor	2	2	
Blood Transfusion	6		
Oceania			
Unknown Risk Factor	1		
Sexual Transmission		5	
Sub-Saharan Africa			
Unknown Risk Factor	8		
Unknown Geographical Region			
Unknown Risk Factor		7	

The majority of sequence data were from samples from individuals in North America with unknown risk factors. HIV-1 individual data were obtained from the LANL HIV sequence database, relevant literature, and through communication with publication authors. Of note, the unknown risk factor category included the literature-derived notations, “Unknown”, “Not Recorded”, and “Not Available”.

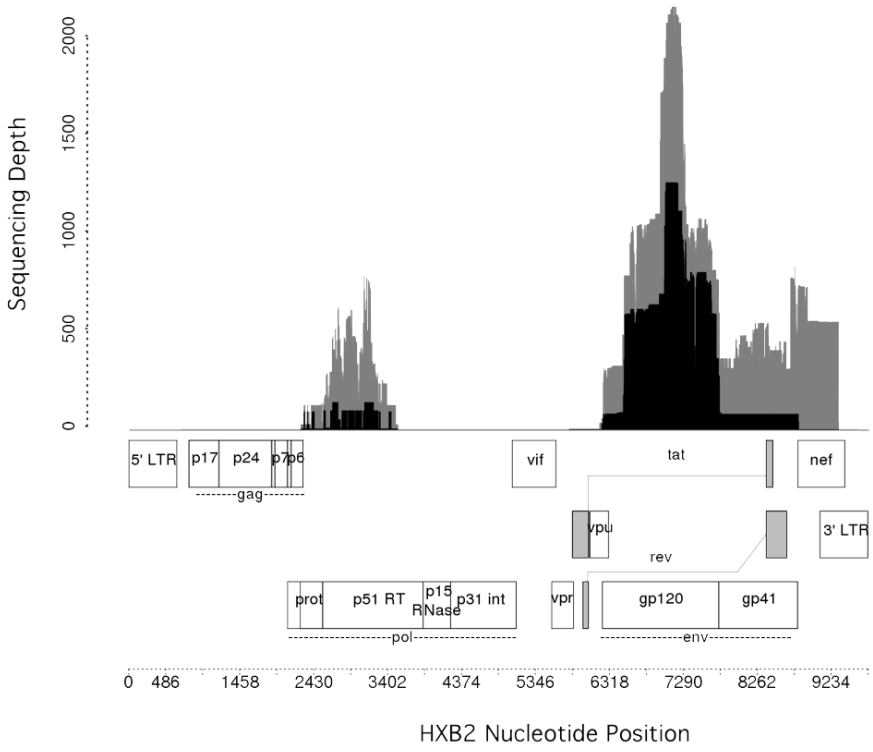


Figure 2.1 Sequencing depth of HAND and non-HAND data

HIV-1 genome coverage as represented by HAND and non-HAND sequence data showed the HIV-1 *env* gene as the genomic region with the greatest amount of sequence data available. The top panel displays sequencing depth across the HXB2 reference sequence. The sequence coverage plot in grey represents HAND data, and the sequence coverage plot in black represents non-HAND data. The bottom panel displays HIV-1 gene location across the HXB2 reference sequence.

We obtained sequence information for 84 unique HAND individuals. Among those, 24 individuals had gender information available, with 12.5% females and 87.5% males. 34 HAND individuals had age information, and their ages ranged from 19 to 63 years, with 56% of individuals between the ages of 40 to 59 years. Information on HAND type, the specific neurocognitive disorder affecting the HAND individual, was also collected, and the top three HAND types were HIV dementia (49%), HIV encephalopathy (42%), and AIDS dementia complex (5%), using literature-reported diagnoses. Non-HAND individuals represented 45% of our database, and served as a key control group. Of the 38 non-HAND individuals with gender information available, 32% were females and 68% were males. Age information was available for 53 non-HAND individuals, and their ages ranged from 23 to 63 years, with 87% of non-HAND individuals between the ages of 30 to 49 years.

To ensure sequence genotyping quality and to exclude a recombinant signal in downstream phylogenetic analyses, database sequences longer than 300 nucleotides in length and with a hypermutation p-value of 0.05 or greater were screened using JPHMM. This analysis revealed recombinant sequences not reported in our data sources. Non-recombinant sequences remained the same between our genotyping result and that of the literature, and 79% of all JPHMM-tested sequences were of the pure B subtype.

In sum, our manually curated HAND database was constructed through a comprehensive search for HAND-related reports, resulting in the largest available HAND sequence dataset as of October 2014. In addition, both sequence and clinical data were validated through data source reconciliation and HIV-1-specific analyses, as detailed above.

2.4.2 Phylogenetic analysis

Our phylogenetic analysis focused on the *env* gene, as it had the greatest sequencing depth among our database genes (**Figure 2.1**). *Env* sequences clustered primarily by individual as revealed by phylogenetic analysis (**Figure 2.2**). Intra-individual sequences, regardless of potential tissue compartmentalization, were homogeneous, as supported by 100% or nearly 100% bootstrap values. *Env* sequences did not cluster by tissue, and HAND status had only a minimal clustering effect, as evidenced by bootstrap values greater than 80% grouping HAND individuals 12, 13, and 14 (**Figure 2.3**).

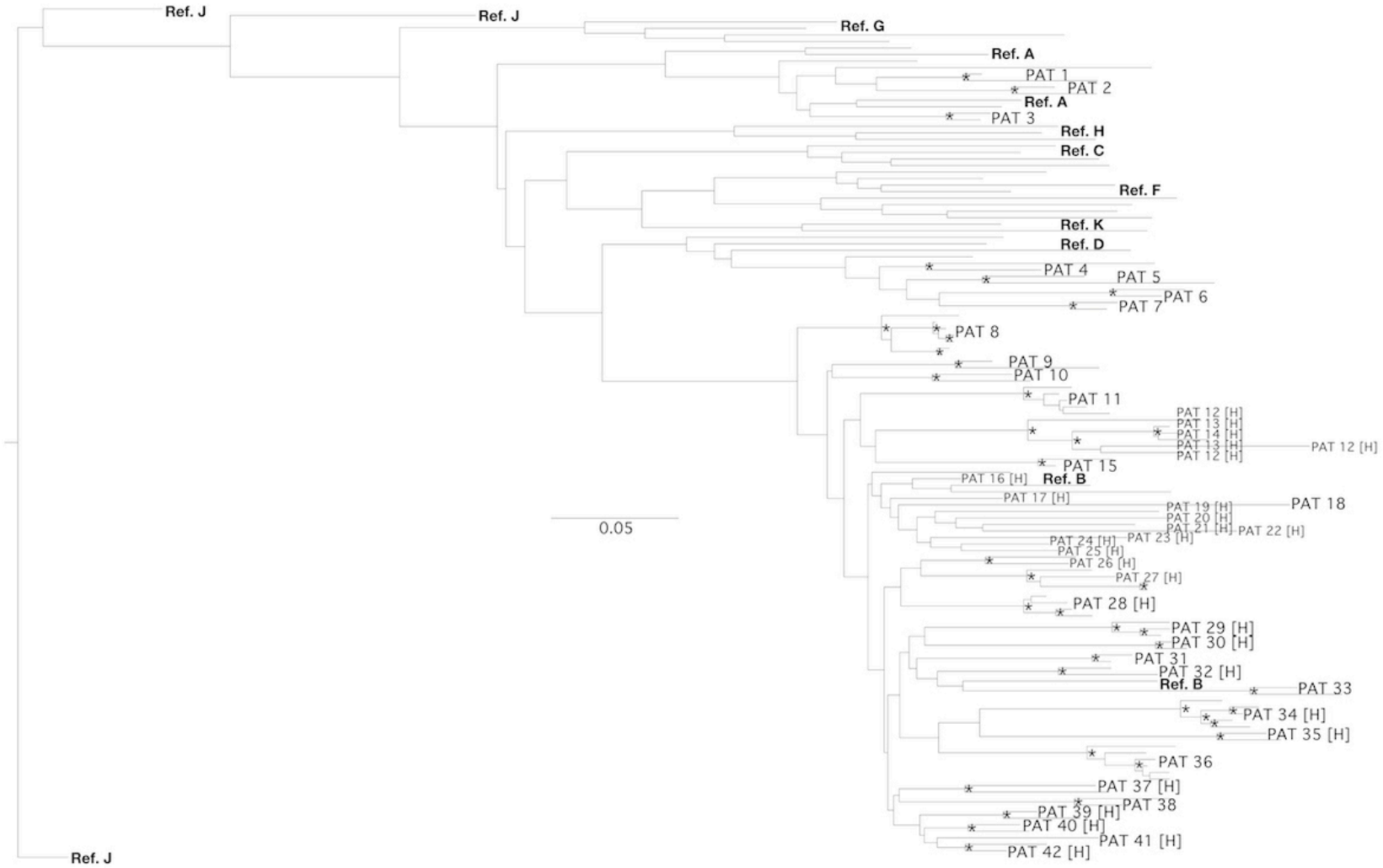

Ref. J
Ref. J
Ref. G
Ref. A
PAT 1
PAT 2
Ref. A
PAT 3
Ref. H
Ref. C
Ref. F
Ref. K
Ref. D
PAT 4
PAT 5
PAT 6
PAT 7
PAT 8
PAT 9
PAT 10
PAT 11
PAT 12 [H]
PAT 13 [H]
PAT 14 [H]
PAT 13 [H]
PAT 12 [H]
PAT 12 [H]
PAT 15
PAT 16 [H]
Ref. B
PAT 17 [H]
PAT 18
PAT 19 [H]
PAT 20 [H]
PAT 21 [H]
PAT 22 [H]
PAT 24 [H]
PAT 23 [H]
PAT 25 [H]
PAT 26 [H]
PAT 27 [H]
PAT 28 [H]
PAT 29 [H]
PAT 30 [H]
PAT 31
PAT 32 [H]
Ref. B
PAT 33
PAT 34 [H]
PAT 35 [H]
PAT 36
PAT 37 [H]
PAT 38
PAT 39 [H]
PAT 40 [H]
PAT 41 [H]
PAT 42 [H]
0.05
Ref. J

Figure 2.2 Phylogenetic tree of *env* region sequences

Phylogenetic analysis of *env* region sequences indicated clustering primarily by individual. Maximum likelihood phylogenetic trees using the Generalised Time-Reversible DNA sequence evolution model were generated for an alignment of *env* sequences spanning the HXB2 nucleic acid coordinates 6737 - 7640. These alignments included 2008 HIV LANL subtype reference sequences, and in addition, alignments were refined through use of the HIV-1 N-linked Glycosylation Site Analyzer tool [62]. Asterisks indicate bootstrap values greater than 75%, and do not apply to reference sequence clustering. Sequences from HAND individuals are labeled with an "[H]" following the patient number.

2.4.3 Sequence diversity analysis

An F84 distance analysis was employed as a means to understand effects of brain compartmentalization and HAND status on viral evolution. Testing of individuals with known treatment history across HAND status revealed no difference in either range or level of diversity (**Figure 2.3a, top**). Testing of treated-only individuals across HAND status, however, indicated lowered diversity in HAND individual sequences ($p = 0.173$) (**Figure 2.3a, bottom**). While non-brain sequences exhibited a smaller range in diversity range as compared to brain sequences, no significant difference in diversity was found between brain and non-brain sequences (**Figure 2.3b**). Inclusion of untreated individuals, however, did result in an increase in diversity range in brain sequences.

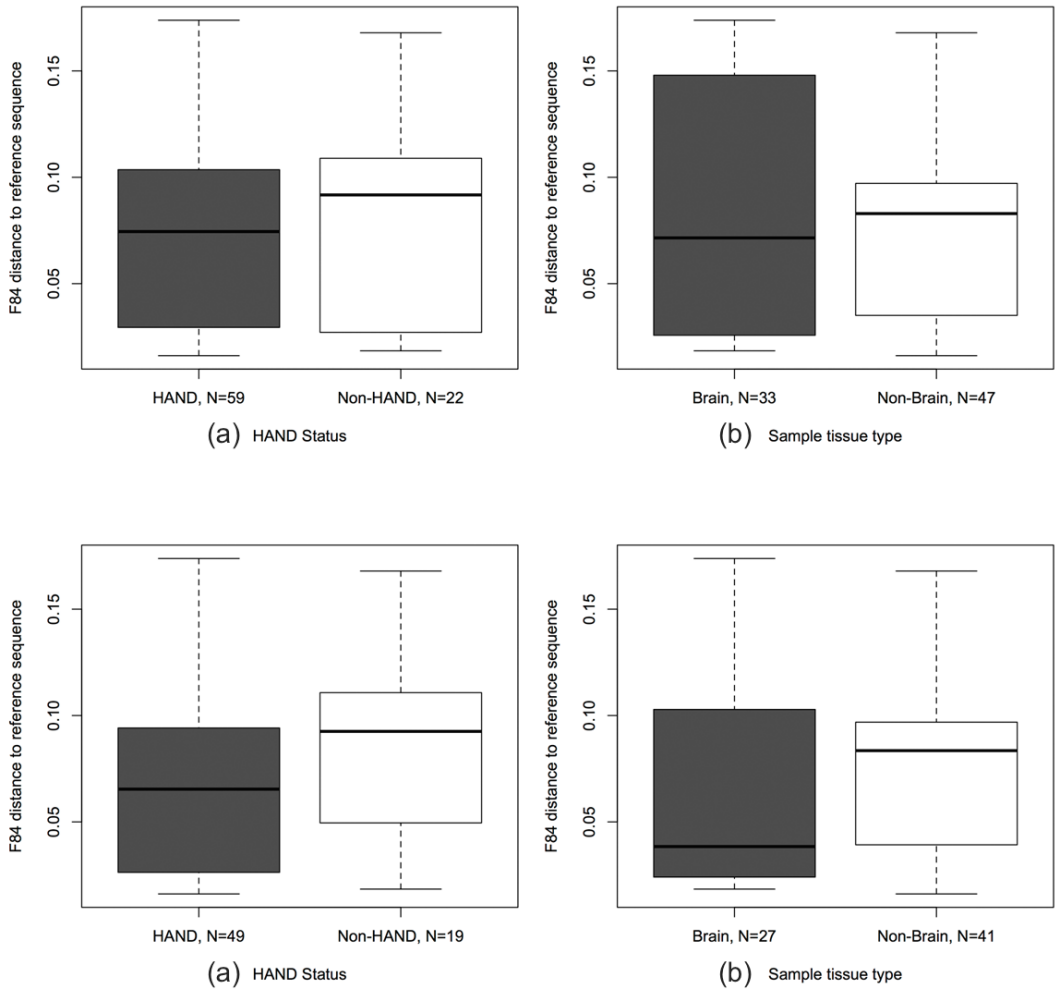


Figure 2.3 F84 *env* region sequence diversity by treatment and across HAND status and sample tissue

F84 diversity for sequences throughout the HIV-1 genome indicated no significant difference between brain sequences and non-brain sequences, as well as no significant difference between HAND and non-HAND sequences. (a) The top two panels compare diversity between individuals for whom we had treatment history available including individuals that had and had not been treated for HIV-1, but not individuals with unknown treatment history. The left-hand plot tests effect of HAND status type (Wilcox

test p-value = 0.5992), and the right-hand plot tests effect of tissue type (Wilcox test p-value = 0.9303). (b) The bottom two panels compare diversity only between individuals that had received HIV-1 treatment. The left-hand plot tests effect of HAND status type (Wilcox test p-value = 0.1689), and the right-hand plot tests effect of tissue type (Wilcox test p-value = 0.173).

We additionally tested the validity in use of cerebrospinal fluid (CSF) viral sequences as proxies for brain-derived sequences [36] through a comparison of diversity range and level across the three tissue types, brain, blood and CSF (**Figure 2.4**). We performed two diversity analyses, one with treated individuals only and one with treated and non-treated individuals combined. In both analyses, CSF sequences were found to have a significantly lower level of diversity as compared to both blood sequences and brain tissue sequences (One-way ANOVA $p = 0.0001$, and $p = 0.0003$ respectively). Brain sequences also had a much larger range of diversity values when compared to both CSF and blood sequences (**Figure 2.4**).

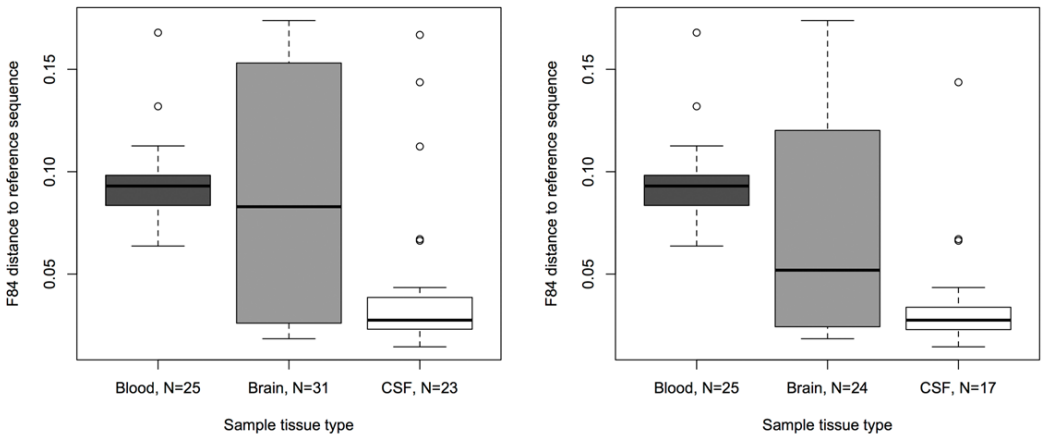


Figure 2.4 F84 *env* region sequence diversity by sample tissue types, blood, brain, and cerebrospinal fluid

Compartmentalization analysis showed significantly lowered diversity in cerebrospinal fluid (CSF) sequences as compared to either brain or blood sequences. Blood and brain sequences did not differ significantly from one another. The left panel compares only individuals for whom we had treatment history available including individuals that had and had not been treated for HIV-1, but not individuals with unknown treatment history (One-way ANOVA p-value = 0.0003, Tukey HSD for blood and CSF: p-value = 0.0005, for brain and CSF: p-value = 0.002, and for brain and blood: p-value = 0.815). The right panel compares only individuals that had received HIV-1 treatment (One-way ANOVA p-value = 0.0001, Tukey HSD for blood and CSF: p-value = 0.00007, for brain and CSF: p-value = 0.021, and for brain and blood: p-value = 0.127).

Overall, we observed no significant difference between non-HAND and HAND sequences and no significant difference between brain and non-brain sequences in terms of diversity. There were, however, indications of HAND-associated differences in both, our phylogenetic analysis and our diversity analysis. When comparing diversity by HAND status in treated patients only, we found diversity to be lower in HAND sequences as compared to non-HAND sequences (**Figure 2.3a, bottom**). In addition we found a highly significant difference in sequence diversity between brain, blood and CSF samples. Compared to both brain and blood sequences, CSF sequences exhibited a significantly lowered level of diversity.

2.4.4 Co-receptor usage analysis

The Geno2pheno coreceptor tool was used to predict viral co-receptor usage, either CCR5 or CXCR4, based on genomic information available in the V3 loop of the *env* gene. In total, we had information on both HAND status and co-receptor usage for 130 *env*-spanning sequences (**Figure 2.5**). We found no significant association between HAND status and co-receptor usage ($X^2 = 0.237$, employing one degree of freedom, less than the tabulated X^2 value for $p = 0.05$), suggesting a fixation of co-receptor usage independent of HAND presence.

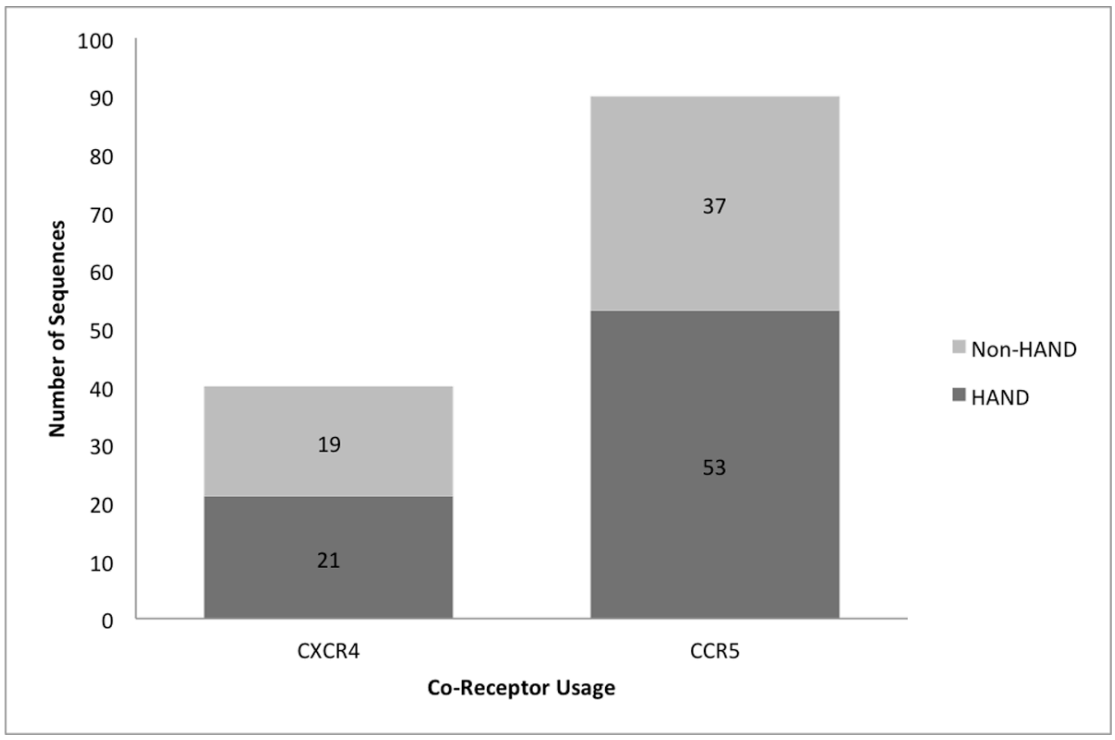


Figure 2.5 Sequence co-receptor usage distributions across HAND status

The majority of sequences exhibited CCR5 co-receptor usage. HAND status, however, was not found to be associated with co-receptor usage, with both CXCR4 and CCR5 usage almost evenly split between HAND and non-HAND sequences.

2.4.5 HAND-associated signature analysis and mapping

We utilized Shannon Entropy to calculate the probability that the presence of a particular amino acid residue at a particular position within the *env* gene was associated with HAND status. The results of this analysis identified amino acid positions in the *env* region, from both brain and non-brain tissues combined, which differed significantly between HAND and non-HAND sequences. Shannon entropy values were calculated for 350 consecutive amino acid residues spanning regions V1 - V5 of the *env* gene (**Figure 2.6**). This area was selected due to its having the greatest sequencing depth in our database. As compared to a recent and high-throughput study locating HIV-1-associated dementia amino acid residues [43], our database allowed us to examine over three times the *env* region for HAND-related residues (**Figure 2.7**). At a significance level of $p < 0.05$, 67 amino acid residues differed significantly between HAND and non-HAND sequences (**Figure 2.8**). Of those, 26 HAND-associated amino acids remained significant at $p < 0.005$. Of note, we found five occurrences of 3 - 4 consecutive amino acid changes significantly associated with HAND at $p < 0.05$. In the V1 loop, four consecutive amino acid residues (HXB2: 149-152) were found to be significantly associated with HAND, and three of them were also significant at $p < 0.005$. Three consecutive amino acid changes (HXB2: 172-174) were found in the V2 loop, one of these at $p < 0.005$. Both, a four amino acid signature (HXB2: 266-269) and a three amino acid signature (HXB2: 281-283) were found between the V2 and V3 loops, each with one residue supported at $p < 0.005$. Finally, a three amino acid signature (HXB2: 335 - 337) was found between the V3 and V4 regions, with one residue also supported at $p < 0.005$.

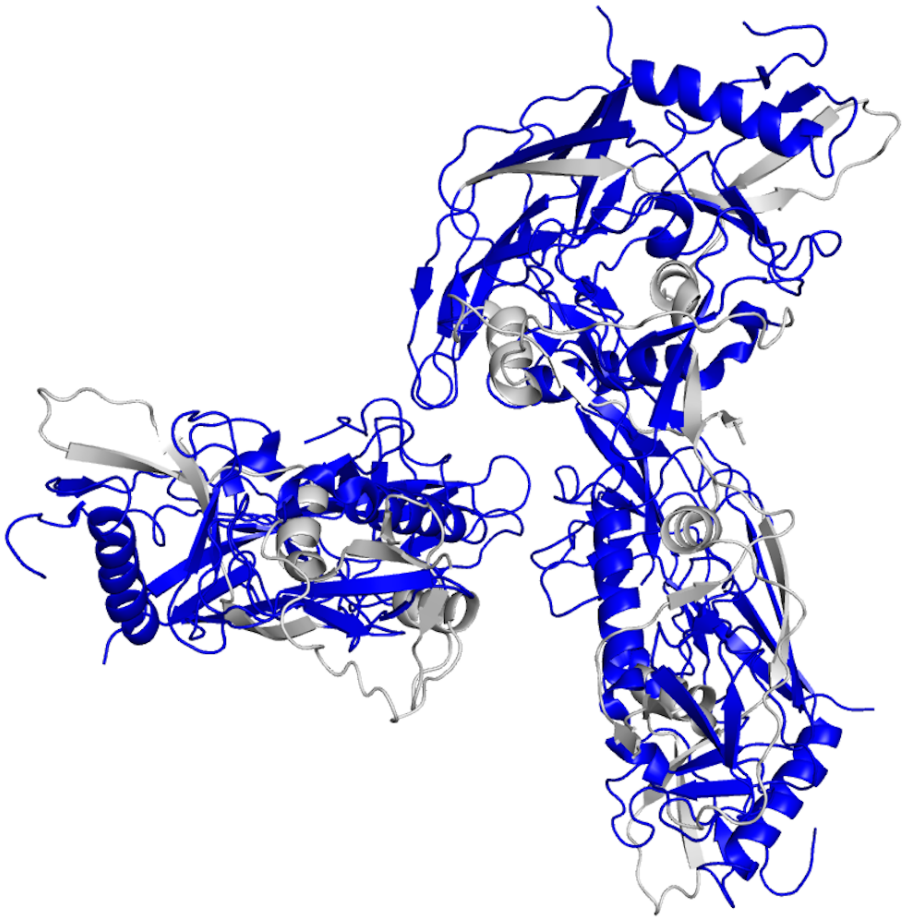


Figure 2.6 Area of the HIV gp120 trimer tested for HAND-associated amino acids in current study

The genomic area tested for HAND-associated amino acids in current study, as shown on the 4NCO PDB HIV gp120 model [68], and indicated in blue.

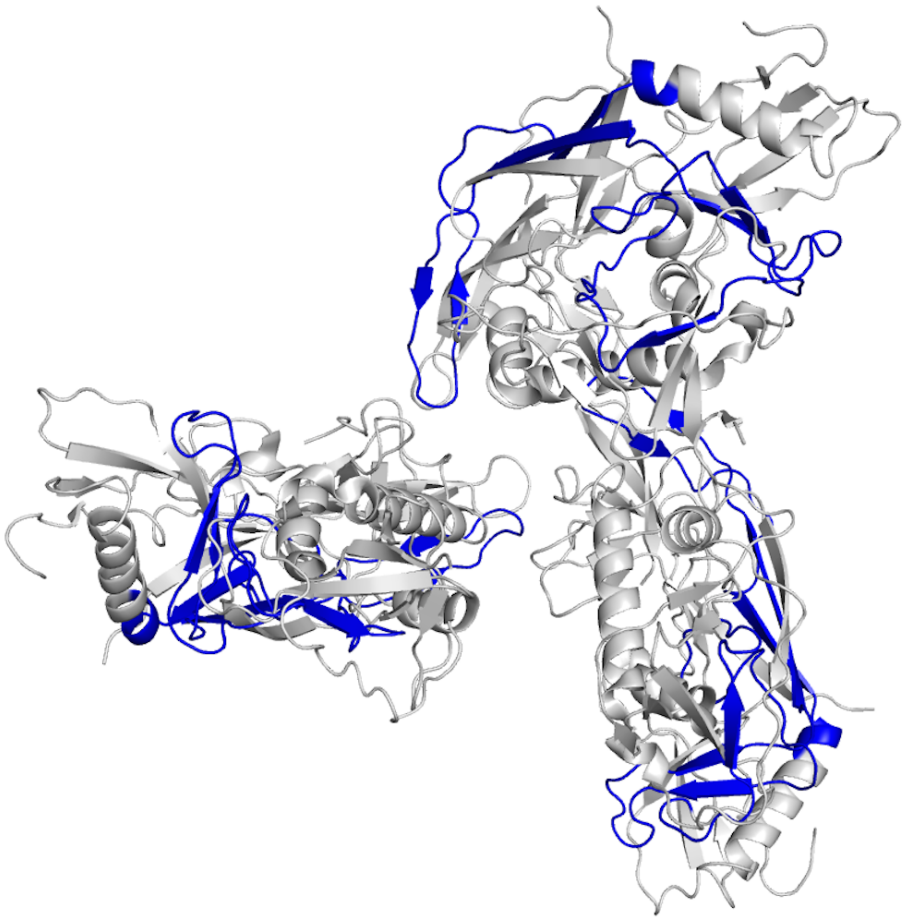


Figure 2.7 Area of the HIV gp120 trimer tested for HAD-associated amino acids in prior high-throughput study

The genomic area tested for HAD-associated amino acids in the recent high-throughput study by Holman et al. [43], as shown on the 4NCO PDB HIV gp120 model [68]. We selected this study as it made use of the recently developed HIVBrainSeqDB Database [80]. Genomic regions tested by Holman for HAD-associated amino acids are shown in blue.

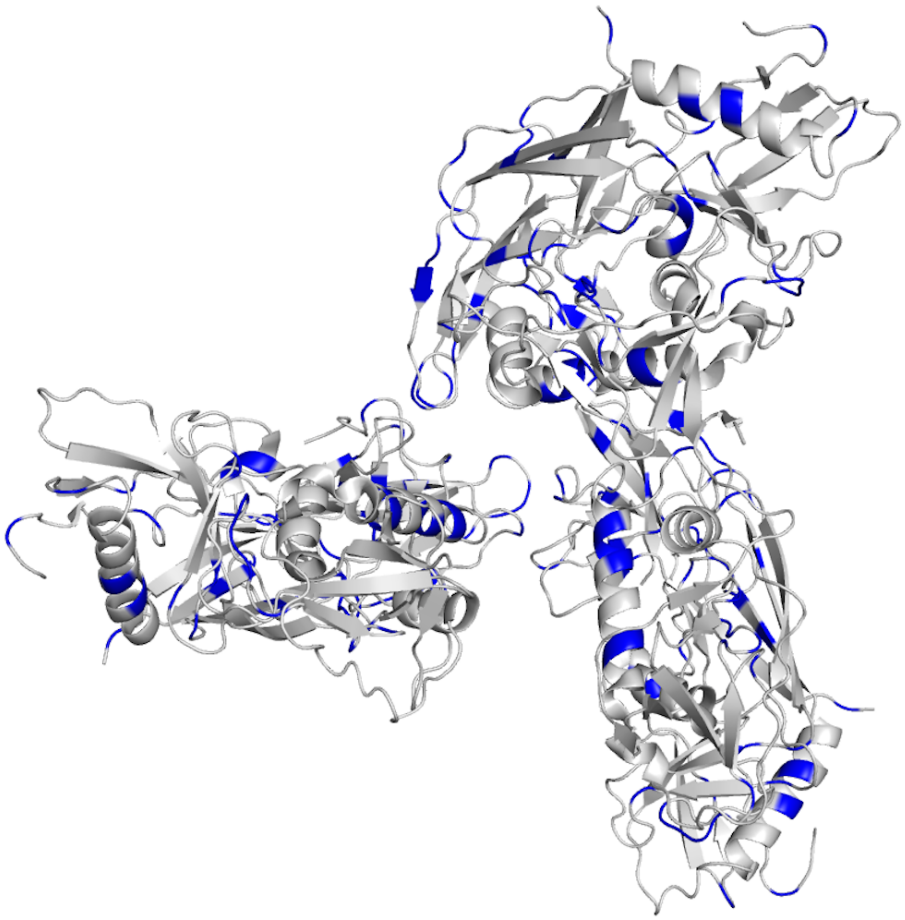


Figure 2.8 Area of the HIV gp120 trimer significantly associated with presence of HAND in current study

Genomic regions found to be significantly associated with HAND at $p < 0.05$ in current study are shown in blue.

Overall, HAND-associated amino acid changes were found in all five V loop regions, as well as in the C2 - C5 regions. In addition, we observed a higher rate of HAND-associated significant amino acid changes in the V loops as compared to the C regions, with rate equal to the number of HAND-associated amino acids found in a specific region over the total number of amino acids in that region. In the V loop region, the incidence rate of HAND-associated amino acids was 0.219, and in the C region this rate was 0.17 (**Table 2.2**).

HAND-associated amino acid residues were checked against known CD8+, CD4+ and antibody epitope positions defined by the Los Alamos HIV Immunology Database. We found all HAND-associated amino acids coincided with at least one epitope type (**Table 2.2**): 10 signatures coincided with CTL/CD8+ epitope sites as reported in the best-defined HIV CTL/CD8+ epitope list [76]; 59 signatures coincided with Los Alamos Database reported T-helper CD4+ epitopes; and 50 HAND signature amino acids with Los Alamos reported antibody epitopes.

Table 2.2 Amino acid residues significantly associated with HAND status

HXB2 Numbering for Significant Positions	HXB2 Amino Acid	Non-HAND Consensus AA	HAND Consensus AA	P-value for Amino Acid Class Equivalent Signature Analysis	P-value for Amino Acid Signature Analysis	CTL/CD8+ Epitope Presence	T-helper CD4 Epitope Presence	Antibody Epitope Presence	V-loop Region
133*	D	0.404 D	0.854 D	0.002	0		✔	✔	V1
134*	L	0.298 A	0.537 L	0	0			✔	V1
149*	M	0.319 M	0.366 E	0.001	0.006			✔	V1
150	E	0.213 E	0.585 E	0.01	0			✔	V1
151*	K	0.191 G	0.537 K	0	0			✔	V1
152*	G	0.511 G	0.805 G	0.001	0.001			✔	V1
161	I	0.660 I	0.585 I	0	0.03		✔	✔	V2
165	I	0.319 I	0.439 I	0.01	0.06		✔	✔	V2
168	K	0.851 K	0.756 K	0.01	0.09		✔	✔	V2
172	E	0.532 E	0.829 E	0.01	0.01		✔	✔	V2
173*	Y	0.553 Y	0.805 Y	0.001	0.003		✔	✔	V2
174	A	0.830 A	0.976 A	0.05	0.05		✔	✔	V2
179*	L	0.723 L	0.976 L	0.001	0		✔	✔	V2
183*	P	0.723 P	0.951 P	0.002	0.002		✔	✔	V2
190	S	0.447 S	0.756 S	0.05	0.04			✔	V2

194	T	1.000 I	0.902 I	0.04	0.01		✔	✔	V2
195	S	0.511 N	0.780 S	0.02	0.02		✔	✔	V2
197	N	0.979 N	0.829 N	0.02	0.02		✔		C2
204	A	1.000 A	0.878 A	0	0	✔	✔		C2
211	E	0.766 E	0.902 E	0.03	0.37	✔	✔	✔	C2
212	P	1.000 P	0.927 P	0.04	0.04	✔	✔	✔	C2
227*	K	1.000 K	0.805 K	0	0		✔	✔	C2
236*	T	1.000 T	0.659 T	0	0		✔	✔	C2
244	T	1.000 T	0.927 T	0.01	0.01		✔	✔	C2
266	A	1.000 A	0.878 A	0.01	0.01		✔	✔	C2
267	E	0.936 E	1.000 E	0.05	0.05		✔	✔	C2
268*	E	0.574 E	0.854 E	0.003	0.003			✔	C2
269	E	0.766 E	0.512 E	0.04	0.06		✔	✔	C2
271*	V	0.426 V	0.756 V	0	0		✔	✔	C2
277	F	0.681 F	0.902 F	0.01	0.01		✔	✔	C2
279*	D	0.894 N	0.512 D	0	0.072		✔	✔	C2
281*	A	0.915 A	0.610 A	0.002	0		✔		C2
282	K	0.915 K	1.000 K	0.05	0.04		✔		C2
283	T	0.404 T	0.707 T	0.01	0.02		✔		C2
293*	E	0.255 K	0.463 E	0.001	0.005		✔	✔	C2
300*	N	0.553 N	0.902 N	0	0.001	✔	✔	✔	V3
305	K	0.617 K	0.976 K	0.01	0	✔	✔	✔	V3
319	T	0.447 A	0.488 A, 0.488 T	0.01	0.01	✔	✔	✔	V3
321*	G	0.511 G	0.902 G	0	0	✔	✔	✔	V3
330	H	0.787 H	0.927 H	0.05	0.05	✔	✔	✔	V3
335	R	0.383 R	0.610 R	0.02	0.01		✔		C3
336	A	0.404 A	0.488 A	0.03	0.03		✔		C3
337*	K	0.277 A, 0.277 E	0.366 K	0	0.054		✔		C3
343*	K	0.319 K	0.610 K	0	0.003		✔		C3

346	A	0.574 V	0.780 V	0.02	0.02		✔		C3
352*	Q	0.553 Q	0.902 Q	0	0		✔		C3
354	G	0.723 N	0.902 N	0.01	0.01		✔		C3
360	I	0.426 I	0.585 V	0	0		✔	✔	C3
362	K	0.404 N	0.634 N	0.04	0			✔	C3
372*	V	0.638 V	0.951 V	0.005	0.005		✔	✔	C3
375	S	0.830 S	0.902 S	0.05	0.27	✔	✔	✔	C3
386	N	0.957 N	0.829 N	0.03	0.03		✔	✔	V4
392*	N	1.000 N	0.829 N	0	0		✔	✔	V4
403	E	0.383 N	0.366 N	0.03	0.18		✔		V4
404*	G	0.277 N	0.268 G	0	0.015		✔		V4
407	N	0.213 S	0.293 G	0.01	0.01		✔		V4
411	S	0.723 N	0.415 N	0.01	0.01		✔		V4
413*	T	0.723 T	0.415 T	0.001	0.001		✔		V4
423	I	0.957 I	0.854 I	0.01	0.08	✔	✔	✔	C4
430	V	0.872 V	1.000 V	0.01	0.01		✔	✔	C4
434	M	0.851 M	1.000 M	0	0		✔	✔	C4
440*	S	0.277 S	0.683 R	0	0		✔	✔	C4
444	R	0.277 S	0.683 R	0.01	0.01		✔	✔	C4
455	T	0.915 T	1.000 T	0.04	0.04		✔	✔	C4
465	S	0.255 G, 0.255 T	0.415 T	0	0.11		✔	✔	V5
467	I	0.468 T	0.415 T	0.02	0.29		✔	✔	V5
471*	G	0.809 G	0.976 G	0.005	0.005		✔	✔	C5

Sixty-seven amino acid positions were found to be significantly associated with HAND status at $p < 0.05$. HXB2 positions labeled with an asterisk were additionally significant at $p < 0.005$. Consensus amino acids are shown along with their frequency. Statistical significance values reported for positions significant at $p < 0.005$ were calculated using 1,000 randomization with replacement. Statistical significance values reported for all

other positions were calculated using 100 randomizations with replacement. CD8+ and CD4+ epitopes and antibody epitope positions were derived from epitope summary tables as provided through the Los Alamos HIV immunology database [70]. For CD8+ T cell epitopes we used best-defined HIV CTL/CD8+ epitopes [76].

2.4.6 Drug resistance sites analysis

All database sequences were screened for potential drug resistance mutations in the reverse transcriptase and protease genomic regions, using the HIVdb Genotypic Resistance Interpretation Algorithm. We found 41 sequences with at least one of three drug resistance mutations tested for: protease inhibitor (PI) surveillance drug-resistance mutations (SDRM), nucleoside reverse transcriptase inhibitor (NRTI) SDRMs, and non-nucleoside reverse transcriptase inhibitor (NNRTI) SDRMs (**Figure 2.9**). In all, we identified 67 HAND-associated amino acid positions, including 56 novel sites that may render drug resistance in different tissues (**Figure 2.8**).

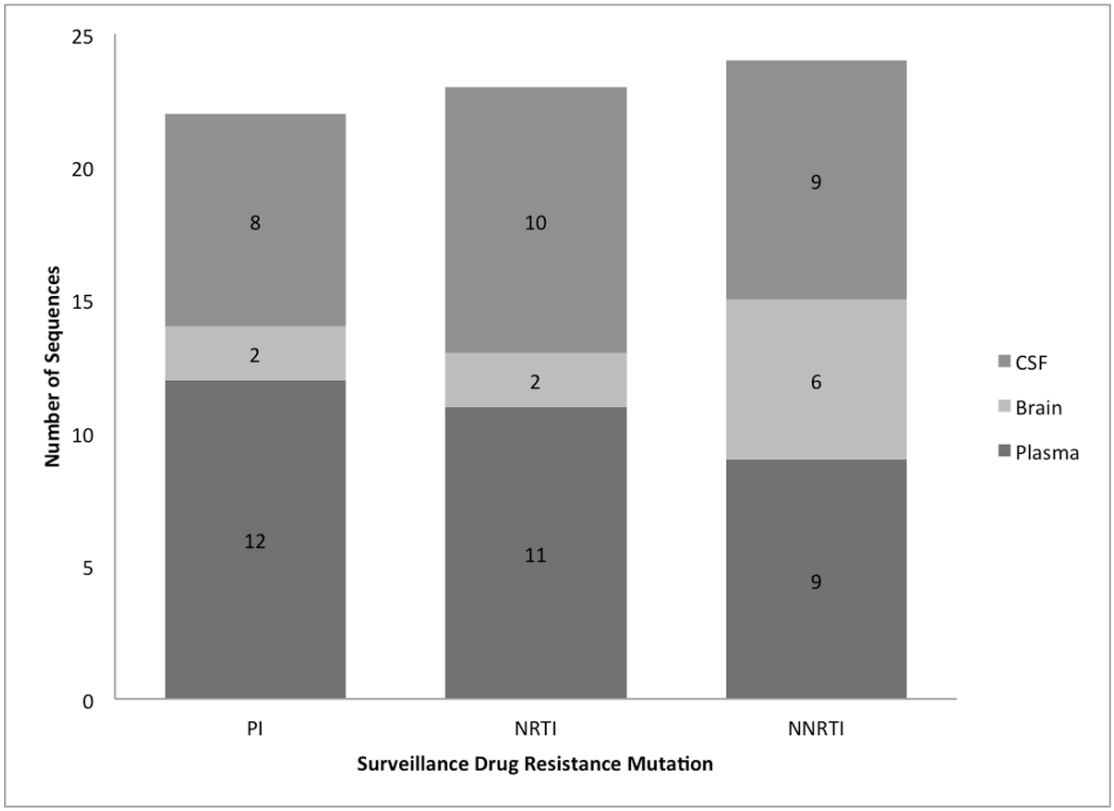


Figure 2.9 Surveillance drug resistance mutation distributions across mutation type and tissue sampled

Surveillance drug resistance mutations were found in sequences from cerebrospinal fluid (CSF), brain, and plasma samples, with the majority from plasma. Drug resistance mutation type was almost evenly split between the three categories, protease inhibitor (PI), nucleoside reverse transcriptase inhibitor (NRTI), and non-nucleoside reverse transcriptase inhibitor (NNRTI).

2.5 Discussion

In this study we developed and analyzed a viral sequence database dedicated to HIV-1-associated neurocognitive disorders (HAND), a common complication occurring in HIV-1 infected individuals [8]. Our well-curated database allowed us to examine viral trajectory under HAND conditions in an unprecedented manner. First, with the biggest sample size to date (5,783 sequences from 163 patients, derived from a comprehensive database and literature search), we were able to address prior HAND research discrepancies with greater statistical power. Second, while the majority of existing HAND research has focused on its most severe manifestation (HIV-1-associated dementia, usually referred to as clinical endpoint), our database enabled tracking of viral changes from individuals across the complete HAND spectrum. As the incidence rate of milder forms of HAND continues to increase, this ability is instrumental toward the identification of potential viral signatures along HAND stages.

In our search for HAND sequence signatures, we also accounted for HIV-1 specific characteristics, including elimination of sequences that have undergone recombination and hypermutation events that would bias phylogenetic analyses, and maintenance of N-linked glycosylation site information, sites critical for viral function but often deleted from variable loop analyses due to alignment difficulty. In addition, we increased the statistical power of our analyses by removing subject-specific biases imposed by unequal sampling of individuals, and across the genome.

Our search for HAND-associated sequence signatures focused on the HIV gp120 region, due to its role in macrophage tropism and viral replication, and as our dataset provided us a robust control group in this region with which to measure our HAND data

against. We identified 67 amino acid residues (56 novel sites) significantly associated with HAND ($p = 0.05$) (**Table 2.2**). A higher rate of HAND-associated changes was observed in the variable regions (V1 – V5) as compared to the conserved regions (C1 – C5), in line with the critical role of variable loops in viral pathogenesis [77].

In addition to our finding of novel positions associated with HAND (**Figure 2.7 and Table 2.2**), our study also confirmed prior HAND and HAD signatures. For example, position HXB2 300, a residue first discovered by Pillai et al. to be a strong predictor of global deficit scores [36], was found to be associated with HAND, in general, in our study. At this position we found the polar and uncharged asparagine as both the non-HAND (0.553 N) and HAND (0.902) consensus. We additionally obtained results that differed from prior research. For instance, two HAD-associated signatures identified by Power et al., HXB2 308 and 333, were not found to be HAND-associated in our study [41]. In line with our results, Holman et al. also failed to find an association between HXB2 333 and HAD [43]. Such differences may be due to increased sample size and control group.

Our diversity analysis of brain-derived, *env*-spanning sequences showed lower levels of viral diversity in HAND sequences as compared to non-HAND sequences. This result, which may point to constraints on viral replication specific to the HAND brain environment, should be interpreted with caution. Prior reports have noted potential interactions between the brain environment and host variables including disease stage and level of neurological impairment [78, 79]. This analysis also revealed sequence diversity to increase in range and amount with the inclusion of non-treated individuals, particularly the case for brain-derived sequences (**Figure 2.3b**). Interestingly, cerebrospinal fluid

(CSF), frequently used as a tissue proxy for surrogating viruses in the brain, was shown in this study as having a significantly lower and narrower amount of genetic diversity as compared to either brain or blood sequences, thus challenging the role of CSF in HAND diagnosis.

The domain of HAND research, including our study, is largely limited by sampling procedures in which certain geographical populations, including those from Sub-Saharan Africa and Asia, are not well represented, and by insufficient clinical information for HAND classification. Furthermore, a still relatively small brain sample size has generated discrepancies between studies in terms of viral genetic change discovery. Finally, using retrospective data, we were also limited in our ability to make comparisons between subtypes and to explore regions beyond the *env* gene. Indeed, a few discoveries of HAND-associated amino acid residues have been made outside of the *env* gene, including HAD signatures in the HIV-1 *nef* gene, and evidence of subtype influence on HAND status has been found, but we could not find sufficient HAND and non-HAND data to test these two additional aspects [20, 21, 45].

2.6 Conclusion

The creation of a well-curated and comprehensive HAND database and research pipeline as described here has expanded our study accuracy and improved the power of our analyses through careful consideration of HIV-1-specific attributes and potential dataset biases, as well as through use of an enlarged dataset. Our findings of novel signatures point to a still incomplete picture of viral evolution within the HAND brain, and there remain genes beyond *env* that may also exhibit HAND signatures, not yet

detectable with current sample sizes. To this end, we have made our database publically available for continued exploration by other researchers of viral evolution in the HAND individual.

Competing Interests

The authors declare that they have no competing interests

Authors' contributions

TZG MZ designed the study, analyzed the data and wrote the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

TZG was funded by the Presidential Scholarship from the University of Georgia. MZ was supported by a UGA Faculty Research Grant

2.7 References

1. Cardenas VA, Meyerhoff DJ, Studholme C, Kornak J, Rothlind J, Lampiris H, Neuhaus J, Grant RM, Chao LL, Truran D *et al*: **Evidence for Ongoing Brain Injury in Human Immunodeficiency Virus-Positive Patients Treated with Antiretroviral Therapy**. *Journal of Neurovirology* 2009, **15**(4):324-333.
2. Kaul M, Garden GA, Lipton SA: **Pathways to Neuronal Injury and Apoptosis in HIV-associated Dementia**. *Nature* 2001, **410**(6831):988-994.
3. Nath A, Clements JE: **Eradication of HIV from the Brain: Reasons For Pause**. *AIDS (London, England)* 2011, **25**(5):577.
4. Rao VR, Ruiz AP, Prasad VR: **Viral and Cellular Factors Underlying Neuropathogenesis in HIV Associated Neurocognitive Disorders (HAND)**. *AIDS Research and Therapy* 2014, **11**:13.
5. Johnson T, Nath A: **Immune Reconstitution Inflammatory Syndrome and the Central Nervous System**. *Current Opinion in Neurology* 2011, **24**(3):284-290.
6. Marcondes MCG, Burudi E, Huitron-Resendiz S, Sanchez-Alavez M, Watry D, Zandonatti M, Henriksen SJ, Fox HS: **Highly Activated CD8+ T Cells in the Brain Correlate with Early Central Nervous System Dysfunction in Simian Immunodeficiency Virus Infection**. *The Journal of Immunology* 2001, **167**(9):5429-5438.
7. Zheng J, Zhuang W, Yan N, Kou G, Peng H, McNally C, Erichsen D, Cheloha A, Herek S, Shi C: **Classification of HIV-I-Mediated Neuronal Dendritic and Synaptic Damage using Multiple Criteria Linear Programming**. *Neuroinformatics* 2004, **2**(3):303-326.

8. Heaton R, Clifford D, Franklin D, Woods S, Ake C, Vaida F, Ellis R, Letendre S, Marcotte T, Atkinson J: **HIV-associated Neurocognitive Disorders Persist in the Era of Potent Antiretroviral Therapy CHARTER Study**. *Neurology* 2010, **75**(23):2087-2096.

9. McArthur J: **Update on the Neurological Manifestations of HIV**. In: *The PRN Notebook.* 2005: 1-6.

10. McArthur JC, Steiner J, Sacktor N, Nath A: **HIV-associated Neurocognitive Disorders: 'Mind the Gap'**. *Annals of Neurology.* 2010, 67(6):699-714.

11. Robertson K, Liner J, Hakim J, Sankalé J-L, Grant I, Letendre S, Clifford D, Diop AG, Jaye A, Kanmogne G *et al*: **NeuroAIDS in Africa**. *Journal of Neurovirology.* vol. 16; 2010: 189-202.

12. Navia BA, Cho ES, Petito CK, Price RW: **The AIDS Dementia Complex: II. Neuropathology**. *Annals of Neurology* 1986, **19**(6):525-535.

13. Navia BA, Jordan BD, Price RW: **The AIDS Dementia Complex: I. Clinical Features**. *Annals of Neurology* 1986, **19**(6):517-524.

14. Belman AL, Ultmann MH, Horoupian D, Novick B, Spiro AJ, Rubinstein A, Kurtzberg D, Cone-Wesson B: **Neurological Complications in Infants and Children with Acquired Immune Deficiency Syndrome**. *Annals of Neurology* 1985, **18**(5):560-566.

15. Snider WD, Simpson DM, Nielsen S, Gold W, Metroka CE, Posner JB: **Neurological Complications of Acquired Immune Deficiency Syndrome: Analysis of 50 Patients**. *Annals of Neurology* 1983, **14**(4):403-418.

16. Epstein LG, Goudsmit J, Paul DA, Morrison SH, Connor EM, Oleske JM, Holland B: **Expression of Human Immunodeficiency Virus in Cerebrospinal Fluid of Children with Progressive Encephalopathy**. *Annals of Neurology* 1987, **21**(4):397-401.

17. Koenig S, Gendelman HE, Orenstein JM, Dal Canto MC, Pezeshkpour GH, Yungbluth M, Janotta F, Aksamit A, Martin MA, Fauci AS: **Detection of AIDS Virus in Macrophages in Brain Tissue from AIDS Patients with Encephalopathy**. *Science* 1986, **233**(4768):1089-1093.

18. Vazeux R, Brousse N, Jarry A, Henin D, Marche C, Vedrenne C, Mikol J, Wolff M, Michon C, Rozenbaum W: **AIDS Subacute Encephalitis: Identification of HIV-infected Cells**. *The American Journal of Pathology* 1987, **126**(3):403.

19. Koyanagi Y, Miles S, Mitsuyasu RT, Merrill JE, Vinters HV, Chen I: **Dual Infection of the Central Nervous System by AIDS Viruses with Distinct Cellular Tropisms**. *Science* 1987, **236**(4803):819-822.

20. Sacktor N, Nakasujja N, Skolasky RL, Rezapour M, Robertson K, Musisi S, Katabira E, Ronald A, Clifford DB, Laeyendecker O: **HIV Subtype D is Associated with Dementia, Compared with Subtype A, in Immunosuppressed Individuals at Risk of Cognitive Impairment in Kampala, Uganda**. *Clinical Infectious Diseases* 2009, **49**(5):780-786.

21. Liner KJ, Hall CD, Robertson KR: **Impact of Human Immunodeficiency Virus (HIV) Subtypes on HIV-Associated Neurological Disease**. *Journal of Neurovirology* 2007, **13**(4):291-304.

22. Spira S, Wainberg MA, Loemba H, Turner D, Brenner BG: **Impact of Clade Diversity on HIV-1 Virulence, Antiretroviral Drug Sensitivity and Drug Resistance**. *Journal of Antimicrobial Chemotherapy* 2003, **51**(2):229-240.

23. Thomson MM, Pérez-Álvarez L, Nájera R: **Molecular Epidemiology of HIV-1 Genetic Forms and its Significance for Vaccine Development and Therapy**. *The Lancet Infectious Diseases* 2002, **2**(8):461-471.

24. Gao F, Morrison SG, Robertson DL, Thornton CL, Craig S, Karlsson G, Sodroski J, Morgado M, Galvao-Castro B, von Briesen H: **Molecular Cloning and Analysis of Functional Envelope Genes from Human Immunodeficiency Virus Type 1 Sequence Subtypes A Through G. The WHO and NIAID Networks for HIV Isolation and Characterization**. *Journal of Virology* 1996, **70**(3):1651-1667.

25. Clapham PR, McKnight Á: **HIV-1 Receptors and Cell Tropism**. *British Medical Bulletin* 2001, **58**(1):43-59.

26. Korber B, Kunstman KJ, Patterson BK, Furtado M, McEvilly MM, Levy R, Wolinsky SM: **Genetic differences Between Blood- and Brain-Derived Viral Sequences from Human Immunodeficiency Virus Type 1-Infected Patients: Evidence of Conserved Elements in the V3 Region of the Envelope Protein of Brain-Derived Sequences**. *Journal of Virology* 1994, **68**(11):7467-7481.

27. Chen H, Wood C, Petito CK: **Comparisons of HIV-1 Viral Sequences in Brain, Choroid Plexus and Spleen: Potential Role of Choroid Plexus in the Pathogenesis of HIV Encephalitis**. *Journal of Neurovirology* 2000, **6**(6):498-506.

28. McCrossan M, Marsden M, Carnie F, Minnis S, Hansoti B, Anthony I, Brettle R, Bell J, Simmonds P: **An Immune Control Model for Viral Replication in the CNS during Presymptomatic HIV Infection**. *Brain* 2006, **129**(2):503-516.

29. Reddy RT, Achim CL, Sirko DA, Tehranchi S, Kraus FG, Wong-Staal F, Wiley CA: **Sequence Analysis of the V3 loop in Brain and Spleen of Patients with HIV Encephalitis**. *AIDS Research and Human Retroviruses* 1996, **12**(6):477-482.

30. Smith KM, Crandall KA, Kneissl ML, Navia BA: **PCR Detection of Host and HIV-1 Sequences from Archival Brain Tissue**. *Journal of Neurovirology* 2000, **6**(2):164-171.

31. Smit TK, Wang B, Ng T, Osborne R, Brew B, Saksena NK: **Varied Tropism of HIV-1 Isolates Derived from Different Regions of Adult Brain Cortex Discriminate between Patients with and without AIDS Dementia Complex (ADC): Evidence for Neurotropic HIV variants**. *Virology* 2001, **279**(2):509-526.

32. Zhang K, Hawken M, Rana F, Welte FJ, Gartner S, Goldsmith MA, Power C: **Human immunodeficiency Virus Type 1 Clade A and D Neurotropism: Molecular Evolution, Recombination, and Coreceptor Use**. *Virology* 2001, **283**(1):19-30.

33. Strain M, Letendre S, Pillai S, Russell T, Ignacio C, Günthard H, Good B, Smith D, Wolinsky S, Furtado M: **Genetic Composition of Human Immunodeficiency Virus Type 1 in Cerebrospinal Fluid and Blood without Treatment and During Failing Antiretroviral Therapy**. *Journal of Virology* 2005, **79**(3):1772-1788.

34. Ball JK, Holmes EC, Whitwell H, Desselberger U: **Genomic Variation of Human Immunodeficiency Virus Type 1 (HIV-1): Molecular Analyses of HIV-1 in Sequential Blood Samples and Various Organs Obtained at Autopsy**. *The Journal of General Virology* 1994, **75**:67-79.

35. Wong JK, Ignacio CC, Torriani F, Havlir D, Fitch N, Richman DD: **In Vivo Compartmentalization of Human Immunodeficiency Virus: Evidence from the Examination of *pol* Sequences from Autopsy Tissues**. *Journal of Virology* 1997, **71**(3):2059-2071.

36. Pillai SK, Pond SLK, Liu Y, Good BM, Strain MC, Ellis RJ, Letendre S, Smith DM, Günthard HF, Grant I: **Genetic Attributes of Cerebrospinal Fluid-Derived HIV-1 *env***. *Brain* 2006, **129**(7):1872-1883.

37. Salemi M, Lamers SL, Yu S, De Oliveira T, Fitch WM, McGrath MS: **Phylodynamic Analysis of Human Immunodeficiency Virus Type 1 in Distinct Brain Compartments Provides a Model for the Neuropathogenesis of AIDS**. *Journal of Virology* 2005, **79**(17):11343-11352.

38. Beebe AM, Dua N, Faith TG, Moore PF, Pedersen NC, Dandekar S: **Primary Stage of Feline Immunodeficiency Virus Infection: Viral Dissemination and Cellular Targets**. *Journal of Virology* 1994, **68**(5):3080-3091.

39. Brew BJ, Evans L, Byrne C, Pemberton L, Hurren L: **The Relationship Between AIDS Dementia Complex and the Presence of Macrophage Tropic and Non Syncytium Inducing Isolates Of Human Immunodeficiency Virus Type 1 in the Cerebrospinal Fluid**. *Journal of Neurovirology* 1996, **2**(3):152-157.

40. Hughes E, Bell J, Simmonds P: **Investigation of the Dynamics of the Spread of Human Immunodeficiency Virus to Brain and other Tissues by Evolutionary Analysis of Sequences from the *p17gag* and *env* Genes**. *Journal of Virology* 1997, **71**(2):1272-1280.

41. Power C, McArthur JC, Johnson RT, Griffin DE, Glass JD, Perryman S, Chesebro B: **Demented and Nondemented Patients with AIDS Differ in Brain-Derived Human Immunodeficiency Virus Type 1 Envelope Sequences**. *Journal of Virology* 1994, **68**(7):4643-4649.

42. Di Stefano M, Gray F, Leitner T, Chiodi F: **Analysis of Env V3 Sequences from HIV‐1‐infected Brain Indicates Restrained Virus Expression Throughout the Disease**. *Journal of Medical Virology* 1996, **49**(1):41-48.

43. Holman AG, Gabuzda D: **A Machine Learning Approach for Identifying Amino Acid Signatures in the HIV *env* gene Predictive of Dementia**. *PloS one* 2012, **7**(11):e49538.

44. Dunfee RL, Thomas ER, Wang J, Kunstman K, Wolinsky SM, Gabuzda D: **Loss of the N-linked Glycosylation Site at Position 386 in the HIV Envelope V4 Region Enhances Macrophage Tropism and is Associated with Dementia**. *Virology* 2007, **367**(1):222-234.

45. Lamers SL, Poon AFY, Mcgrath MS: **HIV-1 Nef Protein Structures Associated with Brain Infection and Dementia Pathogenesis**. In: *PLoS ONE.* vol. 6; 2011: e16659.

46. Lamers SL, Salemi M, Galligan DC, Morris A, Gray R, Fogel G, Zhao L, McGrath MS: **Human Immunodeficiency Virus-1 Evolutionary Patterns Associated with Pathogenic Processes in the Brain**. *Journal of Neurovirology* 2010, **16**(3):230-241.

47. Burkala EJ, He J, West JT, Wood C, Petito CK: **Compartmentalization of HIV-1 in the Central Nervous System: Role of the Choroid Plexus**. *AIDS* 2005, **19**(7):675-684.

48. Caragounis EC, Gisslen M, Lindh M, Nordborg C, Westergren S, Hagberg L, Svennerholm B: **Comparison of HIV-1 *pol* and *env* Sequences of Blood, CSF, Brain and Spleen Isolates Collected Ante-Mortem and Post-Mortem**. *Acta Neurologica Scandinavica* 2008, **117**(2):108-116.

49. Gray L, Roche M, Churchill MJ, Sterjovski J, Ellett A, Poumbourios P, Sherieff S, Wang B, Saksena N, Purcell DF *et al*: **Tissue-specific Sequence Alterations in the Human Immunodeficiency Virus Type 1 Envelope Favoring Ccr5 Usage Contribute to Persistence of Dual-Tropic Virus in the Brain**. *Journal of Virology* 2009, **83**(11):5430-5441.

50. Wang TH, Donaldson YK, Brettle RP, Bell JE, Simmonds P: **Identification of Shared Populations of Human Immunodeficiency Virus Type 1 Infecting Microglia and Tissue Macrophages Outside the Central Nervous System**. *Journal of Virology* 2001, **75**(23):11686-11699.

51. Martin-Garcia J, Cao W, Varela-Rohena A, Plassmeyer ML, Gonzalez-Scarano F: **HIV-1 Tropism for the Central Nervous System: Brain-Derived Envelope Glycoproteins with Lower CD4 Dependence and Reduced Sensitivity to a Fusion Inhibitor**. *Virology* 2006, **346**(1):169-179.

52. Pang S, Vinters HV, Akashi T, O'Brien WA, Chen ISY: **HIV-1 Env Sequence Variation in Brain Tissue of Patients with AIDS-Related Neurologic Disease**. In: *Journal of Acquired Immune Deficiency Syndromes.* vol. 4; 1991: 1082-1092.

53. Lamers SL, Salemi M, Galligan DC, de Oliveira T, Fogel GB, Granier SC, Zhao L, Brown JN, Morris A, Masliah E *et al*: **Extensive HIV-1 Intra-Host Recombination is Common in Tissues with Abnormal Histopathology**. *PLoS One* 2009, **4**(3):e5065.

54. Spudich SS, Huang W, Nilsson AC, Petropoulos CJ, Liegler TJ, Whitcomb JM, Price RW: **HIV-1 Chemokine Coreceptor Utilization in Paired Cerebrospinal Fluid and Plasma Samples: A Survey of Subjects with Viremia**. *The Journal of Infectious Diseases* 2005, **191**(6):890-898.

55. Bratanich AC, Liu C, McArthur JC, Fudyk T, Glass JD, Mittoo S, Klassen GA, Power C: **Brain-derived HIV-1 *tat* Sequences from AIDS patients with Dementia Show Increased Molecular Heterogeneity**. *Journal of Neurovirology* 1998, **4**(4):387-393.

56. Van Marle G, Rourke SB, Zhang K, Silva C, Ethier J, Gill MJ, Power C: **HIV Dementia Patients Exhibit Reduced Viral Neutralization and Increased Envelope Sequence Diversity in Blood and Brain**. *AIDS* 2002, **16**(14):1905-1914.

57. **The Los Alamos HIV Databases**. [http://www.hiv.lanl.gov]

58. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic Local Alignment Search Tool**. *Journal of Molecular Biology* 1990, **215**(3):403-410.

59. Rose PP, Korber BT: **Detecting Hypermutations in Viral Sequences with an Emphasis on G --> A Hypermutation**. *Bioinformatics* 2000, **16**(4):400-401.

60. Zhang M, Schultz AK, Calef C, Kuiken C, Leitner T, Korber B, Morgenstern B, Stanke M: **jpHMM at GOBICS: A Web Server to Detect Genomic Recombinations in HIV-1**. *Nucleic Acids Research* 2006, **34**(Web Server issue):W463-465.

61. Pilcher CD, Wong JK, Pillai SK: **Inferring HIV Transmission Dynamics from Phylogenetic Sequence Relationships**. *PLoS Medicine* 2008, **5**(3):e69.

62. Shaw TI, Zhang M: **HIV N-linked Glycosylation Site Analyzer and its Further Usage in Anchored Alignment**. *Nucleic Acids Research* 2013, **41**(Web Server issue):W454-458.

63. Gouy M, Guindon S, Gascuel O: **SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building**. *Molecular Biology and Evolution* 2010, **27**(2):221-224.

64. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0**. *Systematic Biology* 2010, **59**(3):307-321.

65. Team RDC: **R: A Language and Environment for Statistical Computing**. In: *R Foundation for Statistical Computing.* 2008.

66. Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R: **Bioinformatics Prediction of HIV Coreceptor Usage**. *Nature Biotechnology* 2007, **25**(12):1407-1410.

67. Porollo A, Meller J: **Versatile Annotation and Publication Quality Visualization of Protein Complexes Using POLYVIEW-3D**. *BMC Bioinformatics* 2007, **8**:316.

68. Julien JP, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, Klasse PJ, Burton DR, Sanders RW, Moore JP *et al*: **Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer**. *Science* 2013, **342**(6165):1477-1483.

69. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J *et al*: **Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega**. *Molecular Systems Biology* 2011, **7**:539.

70. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R: **A New Bioinformatics Analysis Tools Framework at EMBL-EBI**. *Nucleic Acids Research* 2010, **38**(Web Server issue):W695-699.

71. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R: **Analysis Tool Web Services from the EMBL-EBI**. *Nucleic Acids Research* 2013, **41**(Web Server issue):W597-600.

72. **The Los Alamos HIV Immunology Database**. [http://www.hiv.lanl.gov/content/immunology]

73. Gifford RJ, Liu TF, Rhee SY, Kiuchi M, Hue S, Pillay D, Shafer RW: **The Calibrated Population Resistance Tool: Standardized Genotypic Estimation of Transmitted HIV-1 Drug Resistance**. *Bioinformatics* 2009, **25**(9):1197-1198.

74. Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW: **Human Immunodeficiency Virus Reverse Transcriptase and Protease Sequence Database**. *Nucleic Acids Research* 2003, **31**(1):298-303.

75. Shafer RW: **Rationale and Uses of a Public HIV Drug-Resistance Database**. *The Journal of Infectious Diseases* 2006, **194 Suppl 1**:S51-58.

76. Llano A W, A, Overa, A, Silva-Arrieta, S, Brander C: **Best-Characterized HIV-1 CTL Epitopes: The 2013 Update**. *HIV Molecular Immunology 2013* 2013:3-19.

77. Yuan T, Li J, Zhang MY: **HIV-1 Envelope Glycoprotein Variable Loops Are Indispensable for Envelope Structural Integrity and Virus Entry**. *PLoS One* 2013, **8**(8):e69789.

78. Keys B, Karis J, Fadeel B, Valentin A, Norkrans G, Hagberg L, Chiodi F: **V3 Sequences of Paired HIV-1 isolates from Blood and Cerebrospinal Fluid Cluster According to Host and show Variation Related to the Clinical Stage of Disease**. *Virology* 1993, **196**(2):475-483.

79. Li S, Juarez J, Alali M, Dwyer D, Collman R, Cunningham A, Naif HM: **Persistent CCR5 Utilization and Enhanced Macrophage Tropism by Primary Blood Human Immunodeficiency Virus Type 1 Isolates from Advanced Stages of Disease and Comparison to Tissue-Derived Isolates**. *Journal of Virology* 1999, **73**(12):9741-9755.

80. Holman AG, Mefford ME, O'connor N, Gabuzda D: **HIVBrainSeqDB: A Database of Annotated HIV Envelope Sequences from Brain and other Anatomical Sites**. In: *AIDS Research and Therapy.* vol. 7; 2010: 43.

CHAPTER 3

THE HAND DATABASE: A CENTRALIZED GATEWAY TO HIV-ASSOCIATED NEUROCOGNITIVE DISORDER RESEARCH[1]

[1]Griffin, T.Z., Ma, Y., and Zhang, M. Submitted to *BMC Medical Genomics*, 10/16/14

3.1 Abstract

Background: Despite an augmented research effort and scale-up of highly active antiretroviral therapy, a high prevalence of HIV-1-associated neurocognitive disorders (HAND) persists in the HIV-infected population. Nearly 50% of all HIV-1-infected individuals suffer from a neurocognitive disorder due to neural and synaptodendritic damage. Challenges in HAND research, including limited availability of brain tissue from HIV patients, variation in HAND study protocols, and virus genotyping inconsistency and errors, however, have resulted in studies with insufficient power to delineate those molecular mechanisms underlying HAND pathogenesis. There exists, therefore, a great need for a reliable and centralized resource specific to HAND research, and in particular, in epidemiological study and surveillance in resource-limited countries where severe forms of HAND persist. **Description:** To address the imperative need mentioned above, here we present the HAND Database, a resource containing well-curated and up-to-date HAND virus information and associated HIV-infected individual data. This database provides information on 5,783 non-redundant HIV-1 sequences from global HAND research published to date, representing a total of 163 unique individuals that have been assessed for HAND. A user-friendly interface allows for flexible searching, filtering, browsing, and downloading of data. The first of its kind in the field, the HAND Database not only bolsters current HAND research by increasing sampling power and reducing study biases due to protocol variation and genotyping inconsistency, it allows for comparison between HAND studies across different dimensions. Development of the HAND Database has also revealed significant knowledge gaps in HIV-driven neuropathology. These gaps include inadequate sequencing of viral genes

beyond *env*, lack of HAND viral data from HIV epidemiologically important regions including Asian and Sub-Saharan African countries, and biased sampling toward the male sex, all factors that impede efforts in improving the quality of life of HIV-infected individuals, and in eliminating virus in the brain. **Conclusion:** Our aim with this database is to provide researchers in the HIV and neuroscience fields a comprehensive and rigorous data source toward both, a better understanding of viral compartmentalization and enhanced design development of strategies against HAND virus. We also expect this resource to be useful as a reliable reference for further HAND epidemiology studies. The HAND Database is freely available and accessible online at http://www.handdatabase.org.

3.2 Background

Human immunodeficiency virus (HIV)-associated neurocognitive disorder (HAND) occurs due to damage to neurons and synapses by viral protein products and by a chemokine/cytokine imbalance in the brain, a pro-inflammatory response to HIV infection of macrophages and microglia [1-3]. HIV entry into the brain is an early event following infection [4], and presence of the blood brain barrier greatly limits entry of antiretroviral therapy into the brain. Our ability to control viral levels within and viral damage to the HIV-infected brain, therefore, remains highly limited. While the introduction of highly active antiretroviral therapy (HAART) brought about a decrease in the incidence of the most severe forms of HAND, i.e., HIV-associated dementia, the prevalence of milder forms has continued to increase [5-7]. In the recent CNS HIV Anti-Retroviral Therapy Effects Research Study, nearly 50% of all HIV-1 individuals

exhibited some form of HAND, including deficits in motor function, verbal fluency, learning, memory, and attention [8]. HAND individuals experience difficulty performing day-to-day tasks, are less likely to adhere to medical treatments and other HIV-1 prevention practices, and ultimately suffer from a threefold increased risk of death as compared to a mentally healthy HIV-1 individual [9]. In addition, in resource-limited countries, the most severe forms of HAND continue to devastate the mental health of HIV individuals [9].

Delineating the underpinning molecular mechanisms of HAND development is critical to providing HIV-infected individuals an elevated quality of life, as well as toward clearance of virus repertoire in the brain. Research in this area, however, has been largely limited by availability of samples from both the brain and from HAND-assessed individuals. In addition, a need to understand HAND progression across an HIV individual's lifespan, coupled with difficulty in obtaining samples, has made cerebrospinal fluid (CSF) sampling a surrogate endpoint for assessing HAND development [10]. Both small sample sizes and indirect CSF inference have made it difficult to uncover a true, yet complex, interaction between viruses and the brain. Furthermore, variations in study methods and result interpretations have also confounded HAND studies, leading to conflicting findings in the field. To address these issues, there exists a great need for a reliable resource, with adequate sample size, for HAND research.

To this end, we have developed such a centralized HAND Database based on all HAND studies published up to date. This resource database is freely accessible at: http://www.handdatabase.org. The HAND Database, the first of its kind, contains well-curated HAND virus information, epidemiology sampling data, patient clinical status,

and therapy treatment information. All information was cross-validated from multiple resources, including the literature, GenBank entry, and author contact. Furthermore, all viral sequences have undergone stringent quality control examination, including genotyping validation, in order to minimize genotyping errors frequently seen in HIV subtype-based studies [11].

The only other published HIV database related to brain tissue, the HIV Brain Sequence Database [12], contains HIV *env* sequences from brain tissue, as well as from other tissues in patients with brain samples. In contrast, our database contains HAND-specific information with regard to virus sequences, genome coverage beyond *env*, and detailed epidemiology sampling information, clinical data, and treatment status, all factors important to the study of HAND pathogenesis. A comprehensive collection of curated HAND HIV information, our HAND Database is a centralized gateway to HIV-associated neurocognitive disorder research.

3.3 Construction and content

3.3.1 Data sources

An extensive literature review was conducted to develop a comprehensive set of HAND-related research articles, from which we were then able to extract sequence data from HAND-assessed individuals. This literature search resulted in the use of data from 41 published studies. Publically available HIV-1 sequence data were collected from GenBank (last accessed 3/2013) and the LANL HIV sequence database (last accessed 2/2014) [13, 14]. HIV-1 individual sampling and clinical information, as well as

treatment status, was collected from the relevant literature, from the two databases mentioned above, and through communication with publication authors.

3.3.2 Sequence and clinical data filtering

All collected sequence data were validated through a series of quality control steps. We first employed the LANL quality control pipeline to check for potential problematic viruses with sequencing errors [13]. Amplification contamination was detected using BLASTn (v. 2.2.26) [15]. In addition, data regarding epidemiology sampling, and clinical and treatment status were cross-referenced whenever available in more than one of the resources listed above.

3.3.3 Genotyping analysis

Genotyping of HIV sequence data is frequently inconsistent and prone to errors [11]. Therefore, all filtered HIV sequences were re-genotyped. Here we applied the jumping profile Hidden Markov Model genotyping program (jpHMM) for which genotyping accuracy is proven [16-18]. In brief, following a hypermutation analysis [13], sequences greater than 300 nucleotides in length and with a hypermutation p-value of 0.05 or greater were subject to genotyping.

3.3.4 Database schema

The HAND Database was constructed using the relational database management system MySQL (v.5.6.17). MySQL was chosen for its ease of use, its high reliability, and as it is freely available. HIV-1 sequence and clinical data were compiled into one flat file,

with annotations divided into three major categories: sequence and sequence descriptor data, HIV-1 patient descriptor data, and sample descriptor data (**Table 3.1**). Sequence data included the HIV-1 nucleotide sequence, sequence accession number, sequence genotype information, and sequence length. Epidemiology data included the geographical location and year at time of sampling, as well as tissue sampled. Patient data at time of sampling included patient age, risk factor, health status, CD4+ count, viral load, HIV treatment information (treatment status, and when applicable, treatment type and duration), and HAND information (HAND status, the presence or absence of HAND, and when applicable, HAND type).

Table 3.1 Overview of database annotations

Category	Annotation	Example	Explanation
Patient	Patient Code	SUBJECT_4	The originating publication ID for this patient was "SUBJECT_4"
Patient	Patient HAND Status	HIVE + ADC	This patient had two forms of HAND: HIVE and ADC
Patient	Patient Sex	m	This patient was male
Patient	Patient Risk Factor	IV Drug User	This patient may have contracted the HIV virus through IV drug use
Patient	Patient Viral Load At Sampling	unknown	Information on this patient's viral load at time of sampling could not be found
Patient	Patient CD4+ Count At Sampling	5 mm3	This patient had a CD4+ count of 5 mm^3 at time of sampling
Patient	Patient HIV Therapy Status	AZT	This patient had received AZT therapy prior to sampling

Patient	Patient HIV Therapy Duration (months)	20	This patient had received HIV therapy for 20 months prior to sampling
Patient	Patient Age At Sampling	29	This patient was 29 years of age at time of sampling
Patient	Patient Health At Sampling	AIDS	This patient had AIDS at time of sampling
Sampling	Sampling Country (City)	Japan (unknown)	This sample was obtained in an unknown city in the country of Japan
Sampling	Sampling Year	1993	This sample was obtained in the year 1993
Sampling	Sampling Tissue	spleen	This sample was obtained from spleen tissue
Sequence	Sequence Polyprotein (Protein)	Tat (gp160)	This sequence segment covers the Tat and Gp160

Sequence	Sequence Genotyping Information[1]	B (B) *	This sequence was listed by the originating data source as a B subtype sequence. JPHMM testing confirmed this subtype. Recombination testing detected recombination events in this sequence.
Sequence	Sequence Accession Number	U82096	The sequence accession number for this sequence is U82096
Sequence	HIV Sequence	cccaaat…	The nucleotide sequence for this sequence entry is "cccaaat…"
Sequence	Sequence Length (nucleotides)	150	This sequence segment is 150 nucleotides in total length

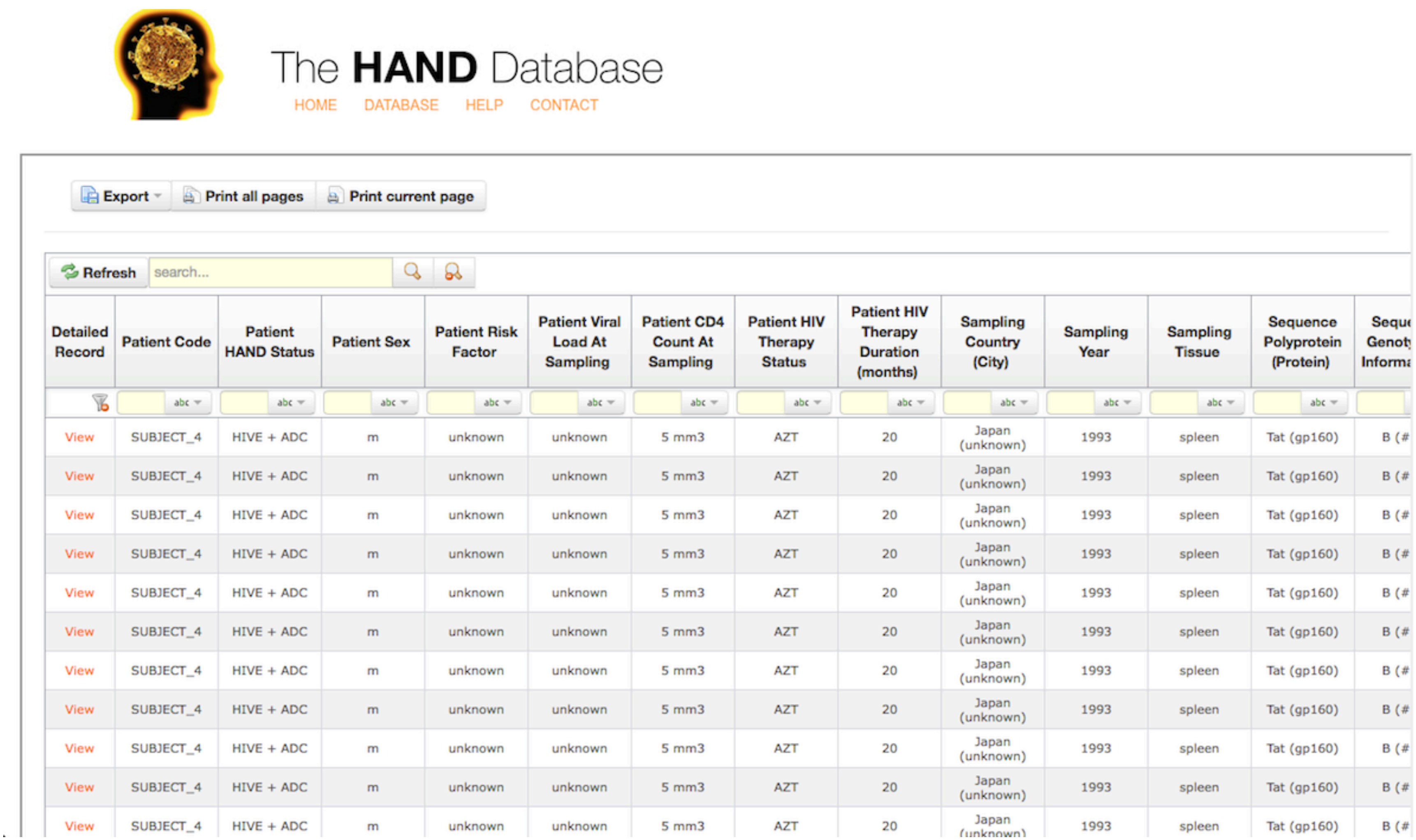
[1]The sequence genotyping annotation provides genotyping and recombination information in the following format: Subtype as reported in the original source material (Subtype as reported by JPHMM testing) Recombination information. Additional

symbols used in this annotation include: "#" = not tested due to short length of sequence, "*" = $p < 0.05$ for recombination test, "(No)" = $p > 0.05$ for recombination test.

3.4 Utility

3.4.1 Database access and web query interface

The HAND Database has been developed into a publically available, web accessible resource. The database website provides a home page with background information on HAND, as well as a help page to assist in database navigation (**Figure 3.1**). The database allows for easy querying and downloading of user-defined data subsets. Researchers can perform a simple search using a keyword, or employ multiple column filters for a custom-made data subset. Selected entries can subsequently be downloaded into a variety of formats at the user's discretion. Additional features include sorting by annotation of interest, as well as an option for viewing the complete record for any given entry.



The **HAND** Database

HOME DATABASE HELP CONTACT

Export | Print all pages | Print current page

Refresh | search...

Detailed Record	Patient Code	Patient HAND Status	Patient Sex	Patient Risk Factor	Patient Viral Load At Sampling	Patient CD4 Count At Sampling	Patient HIV Therapy Status	Patient HIV Therapy Duration (months)	Sampling Country (City)	Sampling Year	Sampling Tissue	Sequence Polyprotein (Protein)	Sequ Genot Inform
	abc	abc	abc	abc	abc	abc	abc	abc	abc	abc	abc	abc	
View	SUBJECT_4	HIVE + ADC	m	unknown	unknown	5 mm3	AZT	20	Japan (unknown)	1993	spleen	Tat (gp160)	B (#
View	SUBJECT_4	HIVE + ADC	m	unknown	unknown	5 mm3	AZT	20	Japan (unknown)	1993	spleen	Tat (gp160)	B (#
View	SUBJECT_4	HIVE + ADC	m	unknown	unknown	5 mm3	AZT	20	Japan (unknown)	1993	spleen	Tat (gp160)	B (#
View	SUBJECT_4	HIVE + ADC	m	unknown	unknown	5 mm3	AZT	20	Japan (unknown)	1993	spleen	Tat (gp160)	B (#
View	SUBJECT_4	HIVE + ADC	m	unknown	unknown	5 mm3	AZT	20	Japan (unknown)	1993	spleen	Tat (gp160)	B (#
View	SUBJECT_4	HIVE + ADC	m	unknown	unknown	5 mm3	AZT	20	Japan (unknown)	1993	spleen	Tat (gp160)	B (#
View	SUBJECT_4	HIVE + ADC	m	unknown	unknown	5 mm3	AZT	20	Japan (unknown)	1993	spleen	Tat (gp160)	B (#
View	SUBJECT_4	HIVE + ADC	m	unknown	unknown	5 mm3	AZT	20	Japan (unknown)	1993	spleen	Tat (gp160)	B (#
View	SUBJECT_4	HIVE + ADC	m	unknown	unknown	5 mm3	AZT	20	Japan (unknown)	1993	spleen	Tat (gp160)	B (#
View	SUBJECT_4	HIVE + ADC	m	unknown	unknown	5 mm3	AZT	20	Japan (unknown)	1993	spleen	Tat (gp160)	B (#
View	SUBJECT_4	HIVE + ADC	m	unknown	unknown	5 mm3	AZT	20	Japan (unknown)	1993	spleen	Tat (gp160)	B (#

Figure 3.1 HAND database search interface

The HAND Database provides flexible searching, filtering, and browsing capabilities. Sequence entries and annotations of interest can be exported into a variety of file formats for further use.

3.4.2 Database content

The HAND Database currently contains 5,783 HIV-1 sequences, representing a total of 163 unique individuals assessed for HAND status. For the 87 individuals with age information available, ages ranged from 19 to 63 years, with the largest proportion of individuals between 30 to 49 years (69%) (**Figure 3.2**). Gender information was available for 64 individuals, the majority of who were males (77%). HAND status, the absence or presence of HAND, was obtained for almost all database individuals (96%), and indicated a close split between non-HAND (44%) and HAND (52%) patients. The top three reported HAND types in HAND-positive individuals were HIV-1 dementia (54%), HIV-1 encephalopathy (35%), and AIDS dementia complex (8%) (**Figure 3.3**). HIV treatment status information, whether or not an individual had received HIV treatment prior to sampling, was available for 67% of individuals, with the majority of individuals with treatment information having received some form of treatment (49%); nearly half of all treated individuals had received HAART (46%) prior to sampling, while the rest had received one or more forms of HIV monotherapy (54%) (**Figure 3.4**).

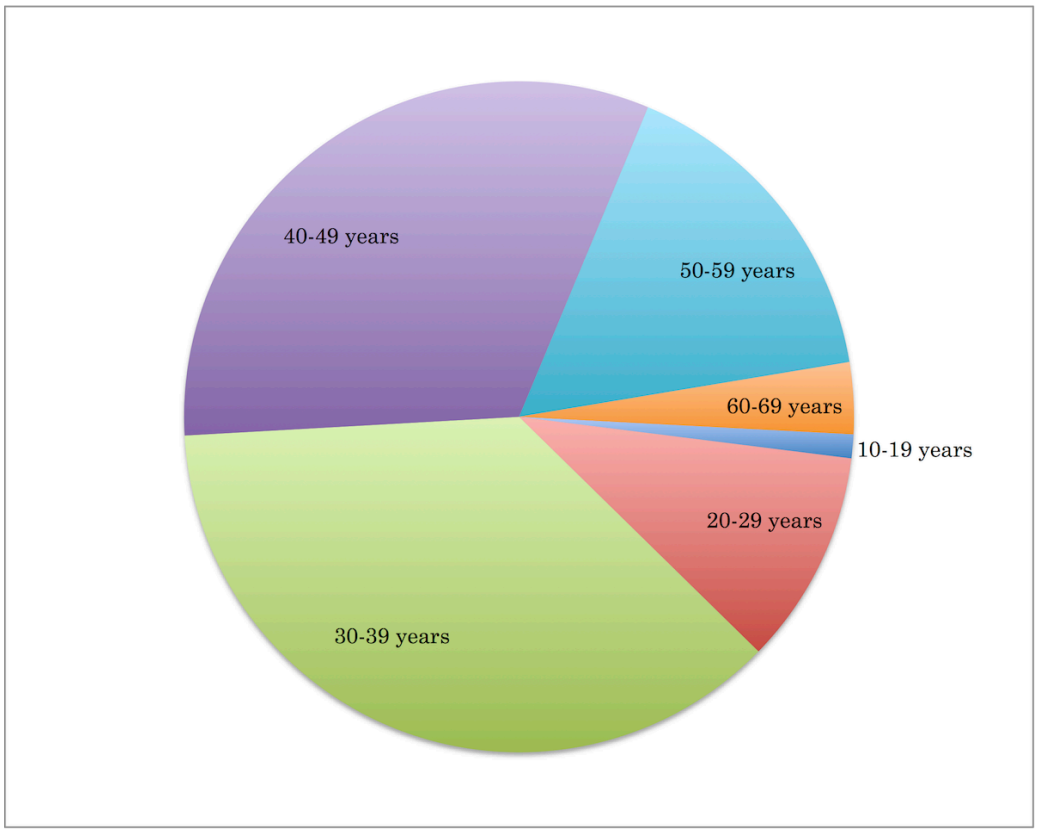


Figure 3.2 Distribution of HAND database entries by age

Chart shows age distribution across all database individuals. Individuals for whom these data were not available were excluded.

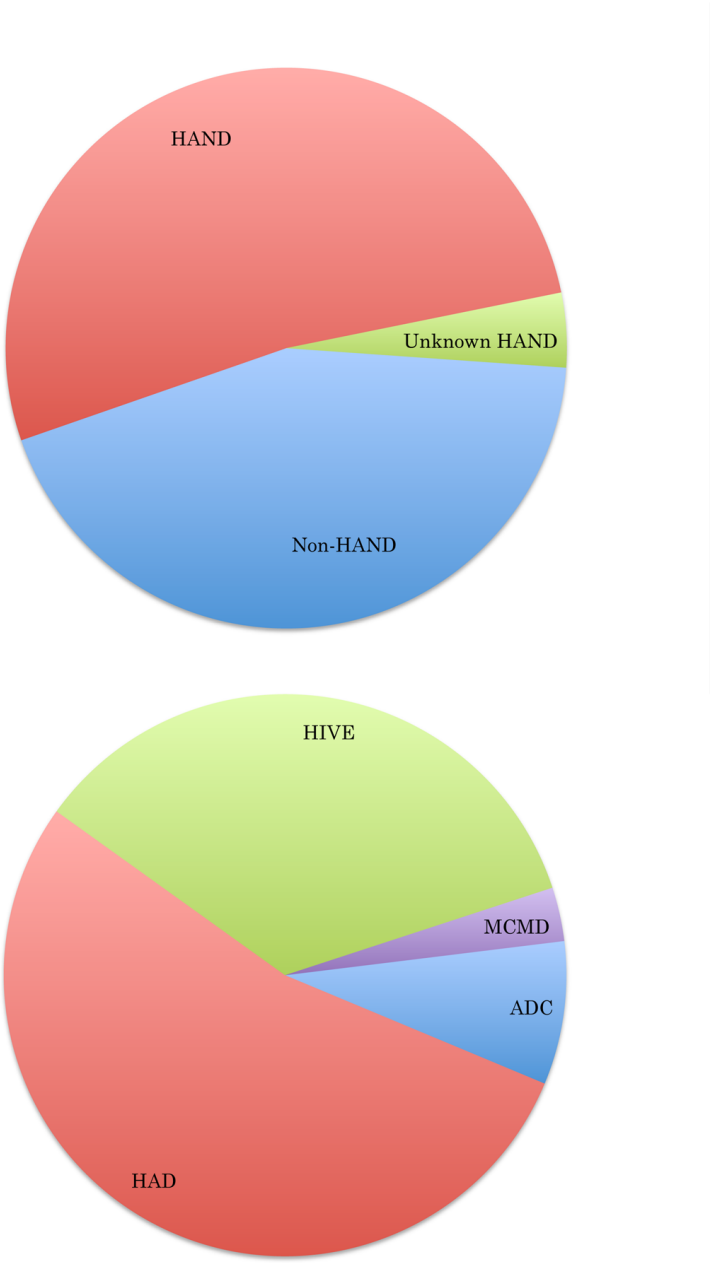
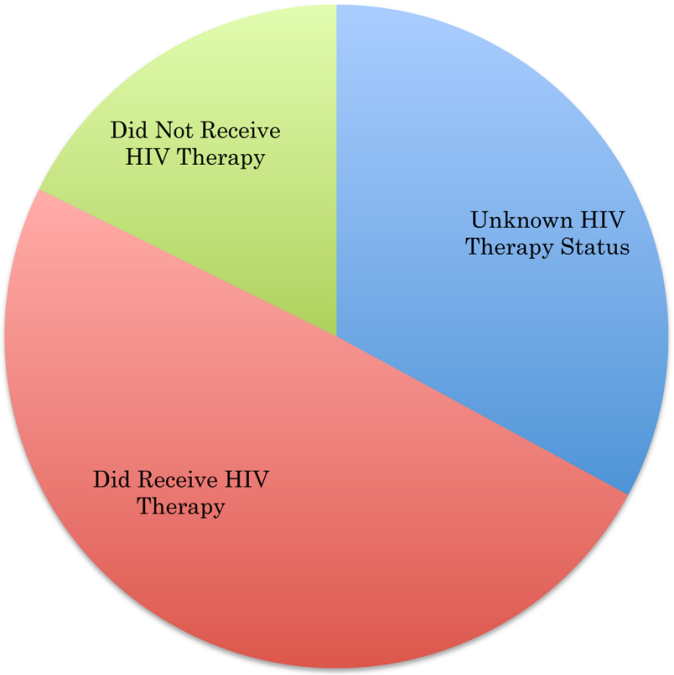


Figure 3.3 Distribution of HAND database entries by HAND status and HAND type

The top chart shows distribution of HAND status across all individuals, and the bottom chart shows distribution of HAND type for those individuals for whom we had this information available. The majority of database individuals with HAND had HIV-associated dementia (HAD), followed by HIV-encephalitis (HIVE), AIDS dementia

complex (ADC), and minor cognitive-motor disorder (MCMD). HAND type designations are as we found to be reported in the literature, and for some individuals, more than one HAND type had been assigned.


Did Not Receive HIV Therapy
Unknown HIV Therapy Status
Did Receive HIV Therapy

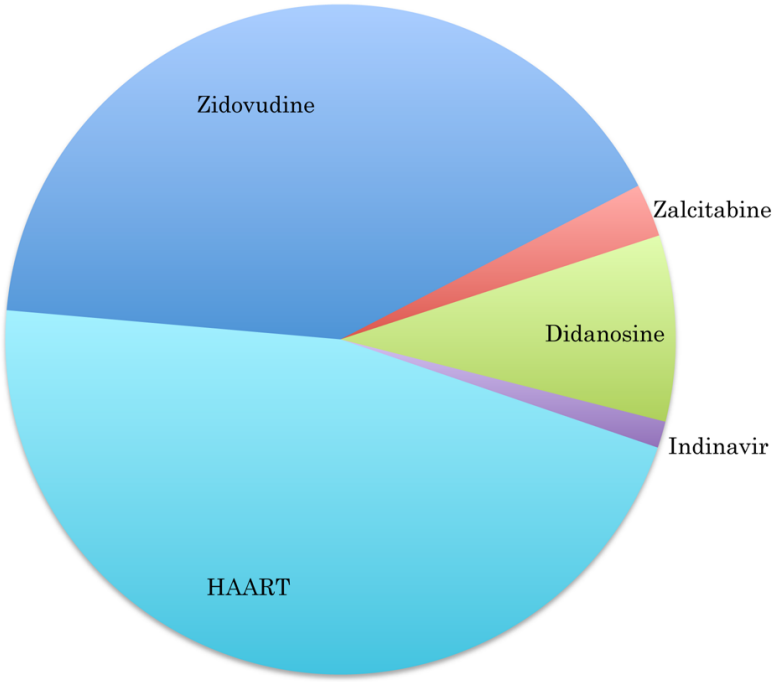

Zidovudine
Zalcitabine
Didanosine
Indinavir
HAART

Figure 3.4 Distribution of HAND database entries by HIV therapy status and HIV therapy type

The top chart shows distribution of HIV therapy status across all individuals, and the bottom chart shows distribution of HIV therapy type for those individuals for whom we had this information available. Therapy type designations are as we found to be reported in the literature, and for some individuals, more than one HIV therapy type had been assigned.

Geographical region sampling information was available for 156 patients, with the top three regions being North America (60%), Europe (25%), and Asia (6.4%) (**Figure 3.5**). Samples were derived from 20 different tissue types, the top three tissues being brain (47%), lymph node (14%), and CSF (7%).

Five HIV-1 genes were represented in our database, *gag*, *pol*, *env*, *tat*, and *nef*, with the majority of sequence coverage in the *env* gene (**Figure 3.6**). This result was expected due the known role of *env* in macrophage tropism, viral replication, and activation of pro-inflammatory responses toward neuronal injury [19-21]. Of all archived sequences, 79% of sequences that underwent genotyping validation were of the pure B subtype, and all non-recombinant sequences were confirmed as correctly reported in the literature. Sixteen sequences were found to have undergone recombination events not reported in either the source literature or databases.

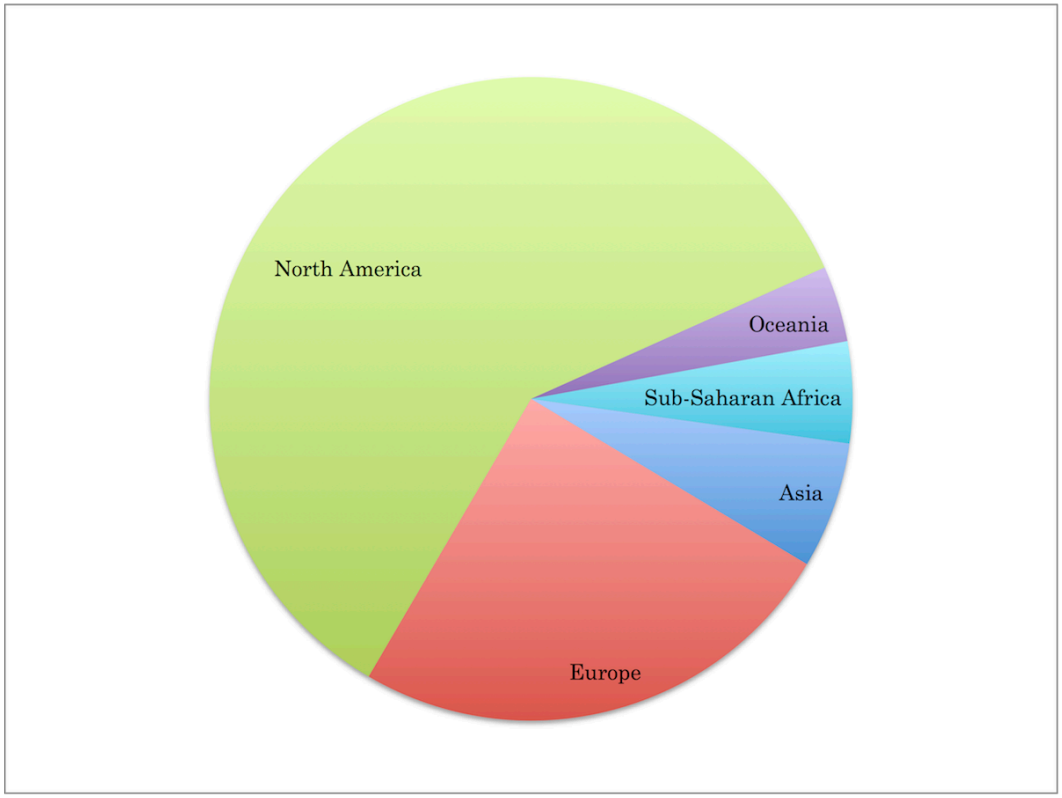


Figure 3.5 Distribution of HAND database entries by sampling geographical region

Chart shows sampling geographical region distribution across all samples. Individuals for whom these data were not available were excluded.

Sequencing Depth
0
1000
2000
3000
5' LTR
p17
p24
p7
p6
gag
vif
tat
vpu
rev
vpr
nef
3' LTR
prot
p51 RT
p15
RNase
p31 int
pol
gp120
gp41
env
0
486
1458
2430
3402
4374
5346
6318
7290
8262
9234
HXB2 Nucleotide Position

Figure 3.6 HAND database HIV-1 genome coverage and sequencing depth

The top panel displays HAND Database sequencing depth across the HXB2 reference sequence, and the bottom panel displays HIV-1 gene location across the HXB2 reference sequence. HXB2 accession number: K03455.

3.5 Discussion

Despite increased HAND research and treatment efforts, the persistent prevalence of HAND continues to pose a great challenge to the HIV research and patient communities. Investigation in this area is limited by small sample sizes, primarily due to difficulty in obtaining tissue samples, and by variation in study protocols and result interpretation. Furthermore, errors and inconsistency in HIV genotyping compound the complexity in delineating viral mechanisms involved in HIV neuropathology. The HAND Database described here serves to narrow these research gaps and addresses the need for a reliable and centralized HAND data source for advanced research purposes.

The HAND database contains up-to-date and well-curated HAND virus and patient information. All sequence data has been subject to stringent quality control examination and genotyping check, therefore laying a solid foundation in the elucidation of viral mechanisms driving neuropathology under various epidemiology settings.

In creating this resource we noted a number of sequencing and sampling biases that limit research in the area, and have developed a set of potential research directions that may greatly benefit the HAND research community. First, although prior studies have indicated the role of multiple HIV proteins, including Nef, Vpr, and Tat [22-29], in HAND development, the majority of research in the area has focused on the gp120 envelope glycoprotein. This sequencing bias is largely due to interest in Env for its roles in conferring viral tropism for microglia and macrophage cells [30-33], in non-neuronal cell replication [34], and for its potential as an HIV therapeutic target [35] (**Figure 3.5**). A shortage of sequence data beyond the *env* gene, however, limits our ability to perform data-driven HAND research on the complete viral genome, and increasing sequencing

efforts in other areas of the genome would provide insight into the role of regulatory and accessory proteins in HAND pathogenesis. Second, there is a distinct lack of sequence data from HIV epidemiologically important regions including many Asian and Sub-Saharan African countries (**Figure 3.2**). Limited access to HAART contributes to an increased vulnerability of HIV individuals in these geographical regions to the most severe forms of HAND. Recent studies indicate HIV-associated dementia (HAD) affects over 25% of HIV individuals in several Sub-Sahara African countries [36-38]. In addition, research on HIV-1-individuals in Thailand has greatly contributed to our understanding of HAND pathogenesis in treatment-naïve individuals [39]. Finally, we noted a bias toward sequencing of male individuals. Research beyond the HIV field has implicated gender as playing a role in determining those genetic processes leading to neurocognitive deficiencies [40, 41]. A lack of information on HAND females, however, will prove an obstacle in determining potential gender differences in HAND pathogenesis.

3.6 Conclusion

Developing a better understanding of mechanisms underlying development of neurocognitive disorders is crucial in providing the HIV patient community with a higher quality of life, and in prevention of enhanced transmission. Through consolidation and validation of data from multiple data sources here we developed the HAND Database, a single, intuitive platform from which researchers can launch their high-throughput HAND sequencing projects. The HAND database contains up-to-date and curated HAND HIV virus and HIV-infected individual information, providing a solid foundation in the

elucidation of viral mechanisms driving HIV neuropathology. In particular, we anticipate this database will be of great use in increasing HAND research efforts in resource-limited countries. We plan to continue expanding the HAND Database as new HAND viral sequence data become publically available.

Availability and requirements

All records are freely available and accessible at www.handdatabase.org.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TZG and MZ designed the study and analyzed the data. YJM contributed to the design stage of the study.

Acknowledgements

TZG was funded by Presidential Scholarship from the University of Georgia. MZ was supported by a University of Georgia Faculty Research Grant 1025GR793002.

3.7 References

1. Johnson T, Nath A: **Immune Reconstitution Inflammatory Syndrome and the Central Nervous System**. *Current Opinion in Neurology* 2011, **24**(3):284-290.
2. Marcondes MCG, Burudi E, Huitron-Resendiz S, Sanchez-Alavez M, Watry D, Zandonatti M, Henriksen SJ, Fox HS: **Highly Activated CD8+ T Cells in the Brain Correlate with Early Central Nervous System Dysfunction in Simian Immunodeficiency Virus Infection**. *The Journal of Immunology* 2001, **167**(9):5429-5438.
3. Zheng J, Zhuang W, Yan N, Kou G, Peng H, McNally C, Erichsen D, Cheloha A, Herek S, Shi C: **Classification of HIV-I-Mediated Neuronal Dendritic and Synaptic Damage Using Multiple Criteria Linear Programming**. *Neuroinformatics* 2004, **2**(3):303-326.
4. Resnick L, Berger JR, Shapshak P, Tourtellotte WW: **Early Penetration of the Blood-Brain-Barrier by HIV**. *Neurology* 1988, **38**(1):9-14.
5. McArthur JC, Steiner J, Sacktor N, Nath A: **HIV-Associated Neurocognitive Disorders: 'Mind the Gap'**. In: *Annals of Neurology*. 2010, 67(6):699-714.
6. Tozzi V, Balestra P, Lorenzini P, Bellagamba R, Galgani S, Corpolongo A, Vlassi C, Larussa D, Zaccarelli M, Noto P *et al*: **Prevalence and Risk Factors for Human Immunodeficiency Virus-Associated Neurocognitive Impairment, 1996 to 2002: Results from an Urban Observational Cohort**. *Journal of Neurovirology* 2005, **11**(3):265-273.
7. Sacktor N, McDermott MP, Marder K, Schifitto G, Selnes OA, McArthur JC, Stern Y, Albert S, Palumbo D, Kieburtz K *et al*: **HIV-Associated Cognitive**

Impairment before and after the Advent of Combination Therapy. *Journal of Neurovirology* 2002, **8**(2):136-142.

8. Heaton RK, Clifford DB, Franklin DR, Jr., Woods SP, Ake C, Vaida F, Ellis RJ, Letendre SL, Marcotte TD, Atkinson JH *et al*: **HIV-Associated Neurocognitive Disorders Persist in the Era of Potent Antiretroviral Therapy: Charter Study**. *Neurology* 2010, **75**(23):2087-2096.
9. McArthur J: **Update on the Neurological Manifestations of HIV**. In: *The PRN Notebook.* 2005: 1-6.
10. Ellis RJ, Moore DJ, Childers ME, Letendre S, McCutchan JA, Wolfson T, Spector SA, Hsia K, Heaton RK, Grant I: **Progression to Neuropsychological Impairment in Human Immunodeficiency Virus Infection Predicted by Elevated Cerebrospinal Fluid Levels of Human Immunodeficiency Virus RNA**. *Archives of Neurology* 2002, **59**(6):923-928.
11. Zhang M, Foley B, Schultz AK, Macke JP, Bulla I, Stanke M, Morgenstern B, Korber B, Leitner T: **The Role of Recombination in the Emergence of a Complex and Dynamic HIV Epidemic**. *Retrovirology* 2010, **7**(25).
12. Holman AG, Mefford ME, O'Connor N, Gabuzda D: **HIVBrainSeqDB: A Database of Annotated HIV Envelope Sequences from Brain and Other Anatomical Sites**. *AIDS Research and Therapy* 2010, **7**(43).
13. Databases LAH: **Los Alamos HIV Databases**. http://www.hiv.lanl.gov/.
14. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: **Genbank**. *Nucleic Acids Research* 2014, **42**(Database issue):D32-37.

15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic Local Alignment Search Tool**. *Journal of Molecular Biology* 1990, **215**(3):403-410.

16. Zhang M, Schultz AK, Calef C, Kuiken C, Leitner T, Korber B, Morgenstern B, Stanke M: **JPHMM at GOBICS: A Web Server to Detect Genomic Recombinations in HIV-1**. *Nucleic Acids Research* 2006, **34**(Web Server issue):W463-465.

17. Schultz AK, Zhang M, Bulla I, Leitner T, Korber B, Morgenstern B, Stanke M: **Jphmm: Improving the Reliability of Recombination Prediction in HIV-1**. *Nucleic Acids Research* 2009, **37**(Web Server issue):W647-651.

18. Schultz AK, Zhang M, Leitner T, Kuiken C, Korber B, Morgenstern B, Stanke M: **A Jumping Profile Hidden Markov Model and Applications to Recombination Sites in HIV and HCV Genomes**. *BMC Bioinformatics* 2006, **7**:265.

19. Toggas SM, Masliah E, Rockenstein EM, Rall GF, Abraham CR, Mucke L: **Central Nervous System Damage Produced by Expression of the HIV-1 Coat Protein Gp120 in Transgenic Mice**. *Nature* 1994, **367**(6459):188-193.

20. Kaul M, Lipton SA: **Experimental and Potential Future Therapeutic Approaches for HIV-1 Associated Dementia Targeting Receptors for Chemokines, Glutamate and Erythropoietin**. *Neurotoxicity Research* 2005, **8**(1-2):167-186.

21. Fontana G, Valenti L, Raiteri M: **Gp120 Can Revert Antagonism at the Glycine Site of NMDA Receptors Mediating GABA Release from Cultured Hippocampal Neurons**. *Journal of Neuroscience Research* 1997, **49**(6):732-738.

22. Lamers SL, Salemi M, Galligan DC, Morris A, Gray R, Fogel G, Zhao L, McGrath MS: **Human Immunodeficiency Virus-1 Evolutionary Patterns Associated with Pathogenic Processes in the Brain**. *Journal of Neurovirology* 2010, **16**(3):230-241.

23. Lamers SL, Poon AF, McGrath MS: **HIV-1 Nef Protein Structures Associated with Brain Infection and Dementia Pathogenesis**. *PloS one* 2011, **6**(2):e16659.

24. Thomas ER, Dunfee RL, Stanton J, Bogdan D, Kunstman K, Wolinsky SM, Gabuzda D: **High Frequency of Defective Vpu Compared with Tat and Rev Genes in Brain from Patients with HIV Type 1-Associated Dementia**. *AIDS Research and Human Retroviruses* 2007, **23**(4):575-580.

25. Gonzalez-Scarano F, Martin-Garcia J: **The Neuropathogenesis of AIDS**. *Nature Reviews Immunology* 2005, **5**(1):69-81.

26. Nath A: **Human Immunodeficiency Virus (HIV) Proteins in Neuropathogenesis of HIV Dementia**. *Journal of Infectious Diseases* 2002, **186 Suppl 2**:S193-198.

27. Kaul M, Lipton SA: **Mechanisms of Neuronal Injury and Death in HIV-1 Associated Dementia**. *Current HIV Research* 2006, **4**(3):307-318.

28. Dunfee R, Thomas ER, Gorry PR, Wang J, Ancuta P, Gabuzda D: **Mechanisms of HIV-1 Neurotropism**. *Current HIV Research* 2006, **4**(3):267-278.

29. King JE, Eugenin EA, Buckner CM, Berman JW: **HIV Tat and Neurotoxicity**. *Microbes and Infection* 2006, **8**(5):1347-1357.

30. Chesebro B, Wehrly K, Nishio J, Perryman S: **Mapping of Independent V3 Envelope Determinants of Human Immunodeficiency Virus Type 1**

Macrophage Tropism and Syncytium Formation in Lymphocytes. *Journal of Virology* 1996, **70**(12):9055-9059.

31. O'Brien WA, Koyanagi Y, Namazie A, Zhao JQ, Diagne A, Idler K, Zack JA, Chen IS: **HIV-1 Tropism for Mononuclear Phagocytes can be Determined by Regions of Gp120 Outside the CD4-Binding Domain**. *Nature* 1990, **348**(6296):69-73.

32. Shioda T, Levy JA, Cheng-Mayer C: **Small Amino Acid Changes in the V3 Hypervariable Region of Gp120 can Affect the T-Cell-Line and Macrophage Tropism of Human Immunodeficiency Virus Type 1**. *Proceedings of the National Academy of Sciences USA* 1992, **89**(20):9434-9438.

33. Jordan CA, Watkins BA, Kufta C, Dubois-Dalcq M: **Infection of Brain Microglial Cells by Human Immunodeficiency Virus Type 1 is CD4 Dependent**. *Journal of Virology* 1991, **65**(2):736-742.

34. Toohey K, Wehrly K, Nishio J, Perryman S, Chesebro B: **Human Immunodeficiency Virus Envelope V1 and V2 Regions Influence Replication Efficiency in Macrophages by Affecting Virus Spread**. *Virology* 1995, **213**(1):70-79.

35. Wu X, Yang ZY, Li Y, Hogerkorp CM, Schief WR, Seaman MS, Zhou T, Schmidt SD, Wu L, Xu L *et al*: **Rational Design of Envelope Identifies Broadly Neutralizing Human Monoclonal Antibodies to HIV-1**. *Science* 2010, **329**(5993):856-861.

36. Joska JA, Westgarth-Taylor J, Myer L, Hoare J, Thomas KG, Combrinck M, Paul RH, Stein DJ, Flisher AJ: **Characterization of HIV-Associated Neurocognitive**

Disorders among Individuals Starting Antiretroviral Therapy in South Africa. *AIDS Behav* 2011, **15**(6):1197-1203.

37. Perriens JH, Mussa M, Luabeya MK, Kayembe K, Kapita B, Brown C, Piot P, Janssen R: **Neurological Complications of HIV-1-Seropositive Internal Medicine Inpatients in Kinshasa, Zaire**. *Journal of Acquired Immune Deficiency Syndrome* 1992, **5**(4):333-340.

38. Howlett WP, Nkya WM, Mmuni KA, Missalek WR: **Neurological Disorders in AIDS and HIV Disease in the Northern Zone of Tanzania**. *AIDS* 1989, **3**(5):289-296.

39. Shiramizu B, Ratto-Kim S, Sithinamsuwan P, Nidhinandana S, Thitivichianlert S, Watt G, deSouza M, Chuenchitra T, Sukwit S, Chitpatima S *et al*: **HIV DNA and Dementia in Treatment-Naive HIV-1-Infected Individuals in Bangkok, Thailand**. *International Journal of Medical Sciences* 2007, **4**(1):13-18.

40. Gatz M, Fiske A, Reynolds CA, Wetherell JL, Johansson B, Pedersen NL: **Sex Differences in Genetic Risk for Dementia**. *Behavior Genetics* 2003, **33**(2):95-105.

41. Artero S, Ancelin ML, Portet F, Dupuy A, Berr C, Dartigues JF, Tzourio C, Rouaud O, Poncet M, Pasquier F *et al*: **Risk Profiles for Mild Cognitive Impairment and Progression to Dementia are Gender Specific**. *Journal of Neurology Neurosurgery and Psychiatry* 2008, **79**(9):979-984.

CHAPTER 4

SHARED MOLECULAR MECHANISMS UNDERLYING BRAIN DISEASE DEVELOPMENT IN HIV AND GBM

4.1 Background

4.1.1 Context

Human immunodeficiency virus (HIV) has claimed ~39 million lives globally in the last three decades [1]. The introduction of highly active antiretroviral therapy (HAART), and its ability to decrease viral load in the body, and in turn increase CD4+ cell count, has drastically impacted clinical progression and mortality rates in the HIV population [2]. Viral suppression by HAART, however, still does little to control viral evolution and replication in the brain, as reflected in the nearly 50% of HIV-infected individuals with some form of HIV-associated neurocognitive impairment (HAND), due to direct damage by HIV to the brain [3]. We are currently challenged by poor penetration of HAART through the blood brain barrier (BBB) and into the central nervous system (CNS), coupled with the potential for even low levels of HIV-1 in the CNS to promote brain damage due to its many indirect neuropathogenic effects [4, 5].

Glioblastoma multiforme (GBM) is the most malignant and most common primary brain tumor, accounting for nearly 50% of all central nervous system tumors [6]. It is also the most fatal, with an average life expectancy of 14 months [7]. GBM's aggressiveness is, in large part, due to its characteristically high degree of tumor cellularity, its rich

angiogenesis [8], and its propensity toward development of chemo-resistance [9]. Despite great advances in cancer treatment, including surgical therapy, radiotherapy, and chemotherapy, treatment in GBM patients is mostly palliative, prognosis is poor, and no current treatment is curative [10].

Insight into GBM pathogenesis from public data resources including, for example, transcriptomic cancer data through the Gene Expression Omnibus database (GEO), and genomic cancer data through The Cancer Genome Atlas (TCGA) [11], has revealed GBM to exhibit extensive intra-tumor heterogeneity (i.e., molecular variation between tumor cells arising from differentiation of cancer stem cells and from accumulation of mutations in malignant cells), with this heterogeneity implicated in treatment failure [12, 13]. Despite its great cellular variability, distinct GBM signaling pathways have been identified as important to its initiation and progression [14] (**Figure 4.1**). For instance, mutation of phosphatase and tensin homolog (*PTEN*) and amplification of epidermal growth factor (*EGF*), both activate the *phosphatidylinositol 3-kinase (PI3K) pathway* leading to an increase in levels of vascular endothelial growth factor (VEGF) and hypoxia-inducible factor (HIF-1α) [15]. Alterations in these pathways are hallmarks of primary GBM, a GBM that arises *de novo*, and these pathways are important in tumor cell proliferation, motility, survival, and angiogenesis [16]. Secondary GBM, a GBM developing from a lower-grade malignancy, is characterized by tumor protein p53 (*TP53*) mutation and retinoblastoma protein (*RB1*) loss, proteins crucial in cell cycle regulation [17, 18]. An increase in protein kinase B (Akt) activity through the *PI3K/Akt pathway* due to *PTEN* mutation is common to both primary and secondary GBMs, and results in inhibition of apoptosis in the tumor cell [19, 20]. Additionally, a characteristic of both

primary and secondary GBM is activation of the *RAS pathway* [21]. As a regulator of cellular signal transduction, RAS proteins are important in cell growth, differentiation and proliferation, and constitutively active RAS signaling cascades have been implicated in a range of cancers [14, 22].

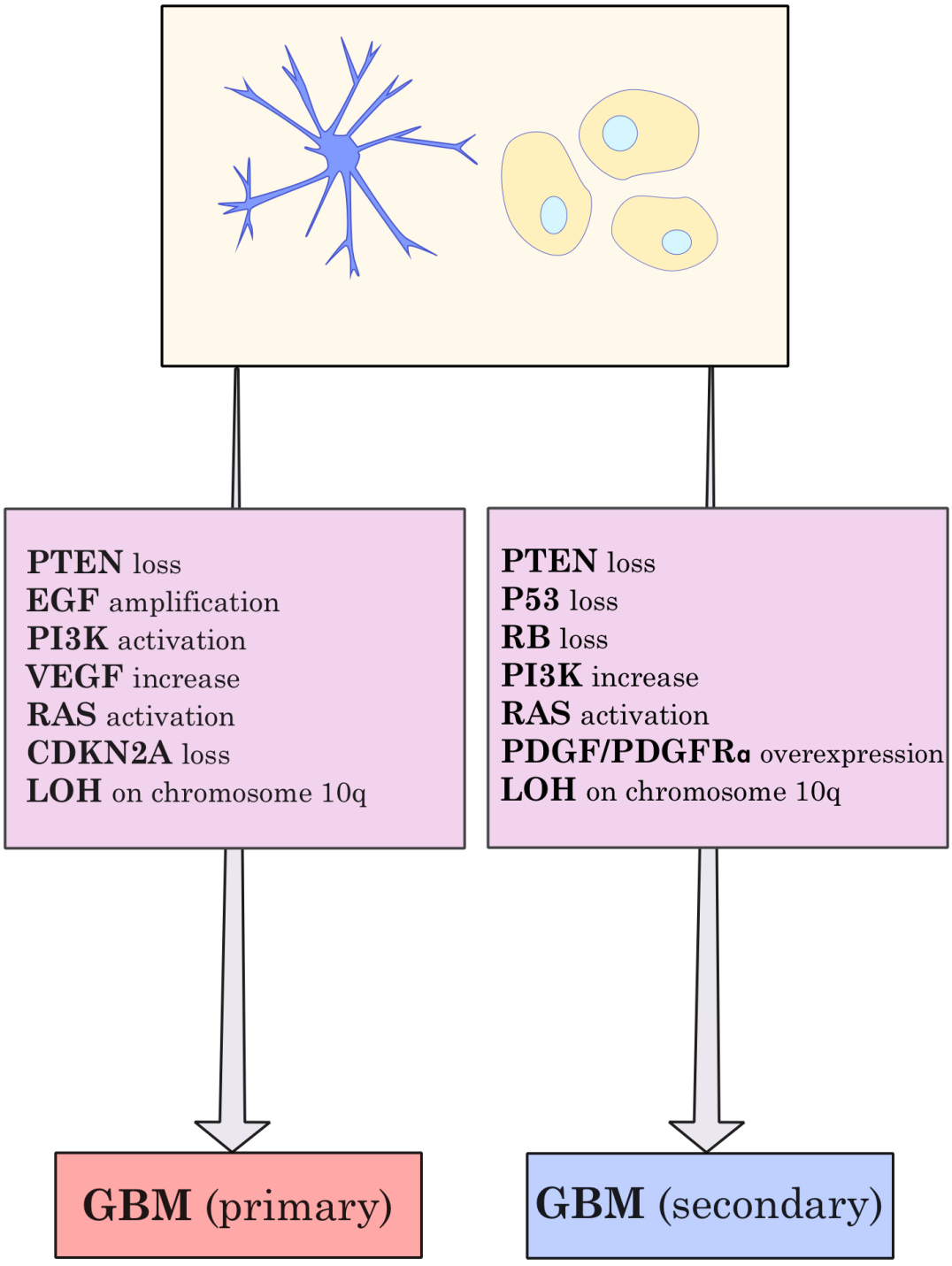

PTEN loss
EGF amplification
PI3K activation
VEGF increase
RAS activation
CDKN2A loss
LOH on chromosome 10q
PTEN loss
P53 loss
RB loss
PI3K increase
RAS activation
PDGF/PDGFRα overexpression
LOH on chromosome 10q
GBM (primary)
GBM (secondary)

Figure 4.1 Key genetic and molecular alterations leading to development of primary and secondary glioblastoma

Malignancy in astrocytes and macrophage/microglia cells involves unique genetic alterations dependent on cellular development into either primary or secondary glioblastoma multiforme (GBM). Key genetic and molecular alterations in primary GBM, a *de novo* malignancy, include increases in epidermal growth factor (EGF), phosphatidylinositol 3-kinase (PI3K), vascular endothelial growth factor (VEGF), and RAS, and decreases in PTEN and CDKN2A. Alterations in secondary GBM, a malignancy developing from a lower-grade lesion, include increases in PI3K, RAS, and PDGF/PDGFRα, and decreases in phosphatase and tensin homolog (PTEN), tumor protein p53 (TP53), and retinoblastoma protein (RB1). Finally, loss of heterozygosity on chromosome 10 is a key genetic alteration observed in both primary and secondary GBMs [14, 23].

Despite a detailed understanding of specific signaling pathways and their roles in cancer promotion and pathogenesis, remarkable levels of inter- and intra- tumor heterogeneity across the GBM regulatory landscape make development of molecular targeted therapies a challenge [24]. For example, use of biopsies to capture and classify tumor cell samples typically undermines the genetic complexity of the entire tumor mass, particularly in highly heterogeneous tumors, and, subsequently, conventional treatments based on this type of sampling only treat or capture a fraction of the tumor mass. In addition, intra-tumor heterogeneity may also allow a tumor the increased capacity to evade cancer treatments, making initially effective therapies no longer useful upon tumor adaptation to treatment [25].

While much knowledge has been gained through the systematic study of the glioblastoma multiforme genome and transcriptome, there is more work to be done with regard to understanding the spectrum of its aberrant pathways. In this study we propose a comparison between two diseases that affect the immune and nervous systems, GBM and HIV-1 infection of the brain, as a means of gaining a better understanding of each. Our hope is that by drawing from these two bodies of knowledge, we can bridge gaps that might not be as obvious through the study of a single disease. We hope that insight gained through this comparison will allow for both, enhanced tumor profiling for more personalized GBM treatment, as well as an enhanced understanding of the effect of HIV-1 on the brain toward HAND prevention and treatment.

In order to overcome challenges associated with GBM heterogeneity, including variable patient response to treatment and an inability of conventional treatment to target all forms of GBM, as well as challenges associated with development of neurocognitive

disorders in HIV individuals, including limitations presented by the BBB and neurotoxicity of both virus and treatment, herein we propose a comparative disease approach between HIV infection and GBM. The aim in this comparison is to develop a better understanding of shared molecular mechanisms that underlie neuropathogenesis in each disease. HIV-acquired immunosuppression results in a significantly higher incidence of AIDS- and non-AIDS- related malignancies [26], and as with GBM development, the inflammatory response is a key driver of neuropathogenesis in HIV-infected individuals [27]. Brain tumor, primary central nervous system lymphoma (PCNSL), has been well documented in the HIV population and is considered the hallmark primary brain tumor in this population; glioma brain tumors, however, are rarely observed in the HIV population. A retrospective analysis in 2011 found that between the years of 1980 and 2010, only 55 glioma cases worldwide in the HIV population had been reported, 19 of those cases of GBM [28]. For 16 out of the 19 HIV+/GBM+ cases reported, the following patient information is available: median age at diagnosis was 38 years of age (range: 19 – 60 years), median CD4+ count at diagnosis was 400 cells/mm3 (range: 80 – 610 cells/mm3), median number of years with HIV, prior to GBM development, was 4 (range: 0 – 11 years), and the gender distribution was 13 males to 3 females [29]. While the age at which these patients presented with GBMs was younger than typical for GBM presentation in the general population, the increased incidence of GBMs in males is in agreement with the GBM gender distribution in the general population [6, 30] (**Table 4.1**).

Table 4.1 Clinical information on HIV+ individuals with GBMs

	Age at Diagnosis* (Years)	Sex	Years with HIV prior to GBM*	CD4+ Count at Diagnosis* (Cells/mm^3)	Survival (Months)*
Range	19 – 60	N/A	0 – 11	80 – 610	0 – (>26)
Median	38	N/A	4	404	3
Total	N/A	13 Male, 3 Female	N/A	N/A	7 (<6 months)
					5 (6 – 12 months)
					4 (>12 months)

Of the 19 documented patients with GBMs, 16 of these cases had information available. From the 21 cases outlined in this reference, two cases were removed due to the detected tumor being a spinal cord tumor, and not a primary brain tumor. Three additional cases were removed from analysis in this reference due to insufficient patient information [29].

**For age at diagnosis, information was available for 15 out of 16 total patients; for years with HIV infection prior to GBM diagnosis, data were available for 11 patients; and for CD4+ count at diagnosis, data were available for 8 patients. For survival, the median was calculated using data for 13 out of 16 total patients, while the range and total were calculated for 15 out of 16 total patients.*

Highly active antiretroviral therapy (HAART) use in HIV individuals has been proposed as an explanation for the unexpectedly low incidence of gliomas in this population [28]. Protease inhibitors used in HAART have been shown to exhibit anti-cancer properties in up to 60 different human cancer cell lines, including promotion of cell death through the endoplasmic reticulum stress response by protease inhibitors Atazanavir and Nelfinavir on malignant glioblastoma cells lines, U251, LN229 and T98G [31], and promotion of cell death through inhibition of NF-κB activation by protease inhibitor Saquinavir on glioblastoma cell line U373 [32-34]. In addition, two Nelfinavir-induced anti-cancer mechanisms have been observed in gliomas, including, activation of the *unfolded protein response pathway* leading to cell death [35], and down-regulation of *VEGF* and *HIF-1α* through the *PI3K pathway* [36]. With 78% of documented glioma cases in HIV individuals having occurred between the years of 1991 and 2000, however, a time during which HAART and Nelfinavir were approved for use [28, 37], it is difficult to directly implicate or dismiss the effect of antiretroviral therapy on GBM incidence in the HIV population. In addition, presence of the BBB has been shown to actively limit entry of protease inhibitors into the brain, thereby limiting Nelfinavir's effect on the brain [38]. We hypothesize, therefore, that while HAART may contribute to a reduced rate of GBMs in the HIV population, this lowered tumor incidence may also suggest a potential mechanism(s) conferred by HIV-1 infection against GBM initiation and development.

Our goal with this work was to elucidate biological pathways involved in HIV-1 and GBM disease development in the brain toward an enhanced understanding of unique and shared molecular mechanisms involved in progression of each disease.

4.1.2 Study description

In this chapter, we used gene expression information to perform a comparative analysis between the two neurological diseases, HIV-1 infection of the brain and glioblastoma multiforme. This comparison was motivated by a similarity in immune system use in both diseases, and by a low incidence rate of GBMs in the HIV population.

Our approach took advantage of publically available brain tissue transcriptional data to construct meta-datasets for comparison between GBM and HIV individuals. Both datasets were initially compared to a control dataset to determine differential expression at both the gene- and pathway- level, unique to each disease. This was followed by a comparison between the two diseases with regard to perturbed and enriched pathways. Results from this work identified commonalities between the two conditions in their use of immune system, nervous system, and cancer system pathways. In addition, we identified a significant difference between the two diseases: Based on both significance and directionality of differential gene expression, cancer pathway, *cell death of glioma cells*, was found to be activated in HIV individuals but not in GBM individuals.

4.2 Methods

4.2.1 Data collection and subject/sample information

Brain tissue sample microarray data were collected from the NCBI Gene Expression Omnibus Databank (GEO) (last accessed 12/2014) [39]. Three GEO datasets comprised the HIV-1-infected (HIV+) sample group (GSE17440, GSE28160, and GSE35864), and six GEO datasets comprised the glioblastoma multiforme (GBM+) sample group (GSE7696, GSE13041, GSE15824, GSE16011, GSE32374, and

GSE36245). The control sample group was also derived from the GEO datasets listed above: twenty-four control samples came from HIV arrays (GSE28160 and GSE35864), and eleven control samples came from GBM arrays (GSE7696, GSE15824, and GSE16011) [40-48]. Matching of control samples within their respective datasets was described for two arrays (GSE16011 and GSE35864) with regard to age in both, and with regard to gender distribution, age at death, and geographical region in the latter.

Due to data availability, in the present study, the HIV+ sample group was collected first, and GBM+ samples were matched to the HIV+ group in terms of subject age (30 – 64 years), and microarray platform used to generate the transcriptional data (Affymetrix® GeneChip Human Genome U133 Plus 2.0 Array). To the best of our ability, and given the limited amount of subject information available, control samples were matched to the HIV+ and GBM+ groups by subject age. Subject data, including age, sex, and treatment status, were collected from the literature, from GEO, and through communication with one publication author (E.T. Tatro, personal communication, December 1, 2014) [46]. The literature described subject diagnoses, including assessment of HIV-1-associated neurocognitive disorders (HAND) in HIV+ subjects and cancer diagnoses in GBM+ subjects, as well as provided sample information including tissue and RNA processing procedures [40-48]. Brain sampling locations were available for HIV+ subjects and for some control subjects, but not for GBM+ subjects. Known sampling locations in the HIV+ group included the three brain regions: white matter from the anterior frontal lobe, grey matter from the frontal neocortex, and basal ganglia tissue.

4.2.2 Dataset filtering

All expression data were generated using the Affymetrix® gene array platform, Human Genome U133 Plus 2.0. Microarray data were obtained in the form of Affymetrix® .CEL files. Raw data files were read into the R software environment [49], and RNA quality was assessed using R packages *affyPLM* (v1.42.0) and *simpleaffy* (v 2.42.0) [50-53]. The *affyPLM* tool provided two quality control measurements, including relative log expression (RLE) and Normalized Unscaled Standard Errors (NUSE) [54]. RLE was used to compare expression for a probe on each array against the median expression for that probe across all arrays. An array that did not center near 0 in an RLE plot indicated its potential as an outlier array. NUSE was used to standardize standard error estimates per probe across all arrays at a median standard error of 1 for each probe across all arrays. An array with an increased standard error as compared to other arrays indicated its reduced quality. The *simpleaffy* tool provided quality control metrics complementary to *affyPLM* including estimation of average background signal, scale factors, and 3' to 5' ratios for β-actin and GAPDH (measures of RNA quality) [55]. Using these array quality indicators, outlier sample arrays and arrays of low RNA quality were detected and removed, including, for example, removal of arrays that did not center near 0 as observed in RLE plots, and removal of arrays with larger standard errors as compared to the average array standard error as observed in the NUSE plots. Arrays were removed from the following GBM datasets: GSE13041 (arrays GSM326904 and GSM326922), GSE7696 (arrays GSM187153, GSM187172, and GSM187187), and GSE16011 (arrays GSM405214, GSM405350, GSM405385, and GSM405389); from the following HIV datasets: GSE28160 (array GSM697506) and GSE35864 (arrays

GSM876848, GSM876849, GSM876861, GSM876872, GSM876873, GSM876885, GSM876896, GSM876897, GSM876902, and GSM876909); and from the following control datasets: GSE28160 (arrays GSM697483 and GSM697485) and GSE35864 (GSM876843) (**Table 4.2**). Arrays removed from dataset GSE35864 also included arrays affected by the associated subject having been diagnosed with a brain-related opportunistic infection, due to the potential for infection-related transcriptomic changes in this sample [48].

Table 4.2 Pre- and post- RNA quality filtration array counts

	# Arrays Pre-Filtration	# Arrays Removed Due to Filtration	# Arrays Removed Due to Opportunistic Infection	# Arrays Post-Filtration	% Arrays Removed Due to Filtration
Control Group	38	3	0	35	7.894%
GBM Group	237	9	0	228	3.797%
HIV Group	88	8	3	77	9.09%
Total	363	20	3	340	5.509%

Individual array quality assessment and removal was based on visual inspection [56] of diagnostic plots produced by RNA quality assessment tools, *affyPLM* and *simpleaffy*, and included evaluation of RLE (relative log expression) and NUSE (normalised unscaled standard error) statistics. A greater percentage of arrays were filtered from the HIV group (9.09%) as compared to either the Control (7.894%) or GBM (3.797%) groups based on RNA quality assessment. The overall percentage of array removal by filtration was 5.5%.

4.2.3 Inter-array normalization

Sample arrays were normalized at the dataset level using the Robust Multiple Array (RMA) algorithm as implemented through the R package, *affy* (v1.44.0) [57]. The *RMA* algorithm was used to perform background correction (to allow for increased sensitivity through removal of non-specific noise), Log2 transformation, and quantile normalization (to adjust for technical variability) of expression data. Prior work has shown this normalization method to produce both consistent down-stream estimates of fold change and differential expression calls with higher sensitivity and higher specificity, as compared to its two largest competitors, *MAS 5.0* and *dChip* [58, 59].

4.2.4 Intra-array normalization

In order to reduce technical variation occurring across datasets, including differences in sample preparation (e.g. RNA concentrations and hybridization efficiencies [60]), following normalization at the within-dataset level, we normalized sample arrays across datasets, at the meta-dataset level. R script, *ComBat.R*, was used to adjust probe values for potential dataset effects [61]. This tool utilized information including both dataset source (individual GEO datasets) and sample type (HIV+, GBM+, or control) to inform probe expression value adjustment by array. Following normalization by *ComBat.R*, R package *arrayQualityMetrics* (v3.22.0) was used to assess data quality at the meta-dataset level, and most importantly to report any remaining batch effects prior to differential expression analysis [62]. To detect outlier arrays, the *arrayQualityMetrics* tool was used to produce MA plots (M = Log2(array intensity) - Log2(pseudo-array intensity), A = 1/2 (Log2(array intensity)+Log2(pseudo-array intensity)), array intensity distribution plots, between-array distance heatmaps (mean absolute difference of the

vector of M-values for each pair of arrays on every probe), and standard deviation versus rank of mean plots. No batch effects were detected and no additional arrays were removed from this analysis.

4.2.5 Probe filtering

Affymetrix® control probes were removed using R package *genefilter* (v1.48.1) [63]. Non-specific filtering, the removal of probes exhibiting consistently small variance across samples was not performed. We chose not to remove small variance probes as such filtering has been shown to interfere with the distributional assumptions of the moderated variance statistic used in our downstream differential expression analyses [64].

4.2.6 Differential expression analyses

Ten differential expression analyses were performed, two differential expression statistical approaches applied on five contrasts: HIV+ group versus Control group; HIV+/HAND+ group versus Control group; GBM+ group versus Control group; HIV+ (treated with HIV therapy) group versus Control group; and HIV+ (not treated with HIV therapy) group versus Control group (**Figure 4.2**). First, R program *limma* (v3.22.4) was used to test for differential expression across probes [65, 66]. The *limma* tool was used to fit a linear model to each probe ID across probe expression values for a given group. This tool utilizes an empirical Bayes approach to borrow information across all probe-wise models toward a differential expression estimate for each probe, and has been shown to produce high confidence measurements, particularly in studies with limited sample sizes. Output from *limma* included a moderated t-statistic per probe, a Benjamini-Hochberg (BH) corrected p-value per probe, and a Log2 fold change per probe [67]. R program

RankProd (v2.38.0) was also used in the identification of differentially expressed probes [68]. *RankProd* uses a non-parametric statistic to detect probes that are consistently ranked high across multiple comparisons. This methodology offers advantages over linear modeling including, use of the biologically motivated fold-change (FC) criterion, fewer assumptions (including no distribution assumption), and increased performance with both noisy data and with data of few replicates [69]. These attributes made *RankProd* well suited to our meta-analysis as it offered a way of overcoming potential heterogeneity across datasets by integrating information across them. Output from *RankProd* included a p-value per probe, and those p-values were adjusted for multiple comparisons using the Benjamini-Hochberg method in R.

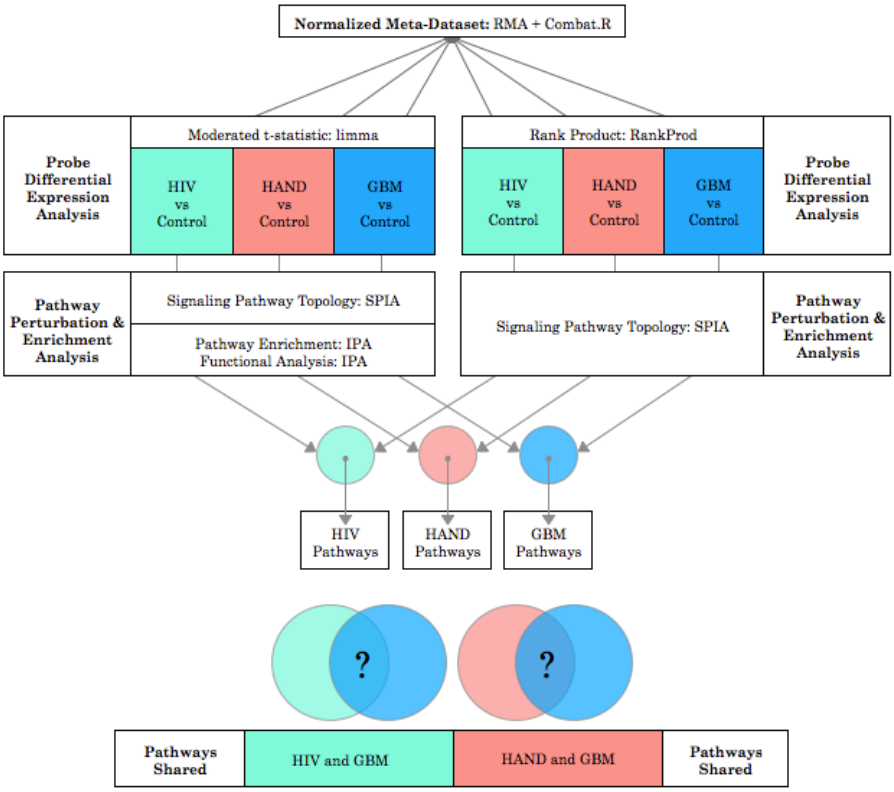


Figure 4.2 Computational pipeline for primary microarray analyses

Microarray meta-datasets (HIV+, GBM+ and control) were first normalized to eliminate both inter- and intra- dataset technical biases. Normalized HIV+ and GBM+ probe values were tested for differential expression (DE) against normalized control probe values using two different DE approaches. Differential expression results from both DE approaches were analyzed for pathway perturbation and pathway enrichment, and significant pathways were compared across groups of interest, including HIV+ versus GBM+ and HIV+/HAND+ versus GBM+. The primary interest of this work, the transcriptional alteration intersection between the two diseases, HIV and GBM, is indicated above by a "?" symbol.

4.2.7 Pathway analyses

Differential expression results were used as input into two pathway analysis programs, Signaling Pathway Impact Analysis (*SPIA*, v2.18.0) [70] and QIAGEN's Ingenuity® Pathway Analysis (*IPA*) (**Figure 4.2**). Prior to each computational pathway analysis, Affymetrix® probe identifiers were collapsed to gene level identifiers, and probe identifiers with no gene level identifiers were removed from further analysis.

Output from both *limma* and *RankProd* were run through separate analyses using *SPIA*. *SPIA* performed a knowledge-based driven pathway analysis using both Log2-fold expression change information and B-H adjusted p-value information to detect over-represented pathways exhibiting aberrant topology perturbation. Pathway information was derived from the Kyoto Encyclopedia for Genes and Genomes Database (KEGG) pathway repository [71]. *SPIA* computed two p-values, pNDE and pPERT, in the measurement of both, pathway enrichment, the over-representation of differentially expressed genes in a pathway, and pathway perturbation, the amount of gene expression change across a pathway topology. In this way, *SPIA* provided a more biologically accurate quantification of gene interactions as compared to a simple pathway enrichment tool. For this study, these p-values pNDE and pPERT were combined using the normal inversion method, the most conservative method available, toward a global pathway significance value, pG. In total, we performed 2,000 bootstrap iterations of the *SPIA* algorithm, the recommended bootstrap iteration value for this tool. Pathway perturbation results summed across both the *limma* and *RankProd* pipelines were used in our final disease comparative analysis (**Figure 4.2**).

A second pathway analysis tool, literature-based QIAGEN's Ingenuity® Pathway Analysis Software (*IPA*), was used to perform a comparative analysis across the three sample groups, HIV+, HIV+/HAND+, and GBM+. Output from *limma* was used as input into *IPA*, including Log2-fold change values, average probe expression values, and B-H adjusted p-values. Datasets were filtered using two cutoffs, a B-H adjusted p-value of less than 0.05 and a Log2-fold change value cutoff dependent on the Log2-fold change range within each dataset. Probe set information specific to the Affymetrix® Human Genome U133 Plus 2.0 microarray platform was used as the background dataset for downstream pathway enrichment analyses, as this microarray platform was used to generate all data included in this study.

4.2.8 Canonical pathway enrichment analysis

Pathway information derived from the literature-based Ingenuity® Knowledge Base, in combination with observed gene changes (as measured by differential expression values), was used to identify, within each sample group, significantly enriched Ingenuity® canonical pathways. Pathway enrichment was calculated by measuring the overlap between the observed (as derived from our datasets) gene sets and the predicted (as derived from the *IPA* Literature Knowledge Base) gene sets. Enrichment significance values were calculated using Fisher's Exact Test, and p-values were adjusted using the Benjamini-Hochberg correction method, using a significance cutoff of B-H adjusted p-value less than 0.05.

4.2.9 Disease and biological function enrichment analysis

Disease and biological function information derived from the literature-based Ingenuity® Knowledge Base, in combination with observed gene changes by sample

group, was used to identify significantly enriched disease and biological functions within each sample group. Reflective of our interest in disease, molecular, and cellular processes, the categories selected for this biological function enrichment analysis included “Diseases and Disorders”, “Molecular and Cellular Functions”, and “Physiological System Development and Function”. Significance values for function enrichment were calculated using Fisher's Exact Test and adjusted using the Benjamini-Hochberg correction method. Those functions and diseases found to be significantly enriched in all three datasets (B-H adjusted $p < 0.05$) were further evaluated using the Ingenuity® Knowledge Base to determine potential activation or inhibition of each function based on the observed directionality of gene expression changes within each dataset. In brief, *IPA* makes a prediction as to a particular biological function’s activation or inhibition based on differential expression information of regulator genes upstream of this function. Statistical significance based on the number and combination of “inhibition” and “activation” predictions is determined using a normally distributed Z-score. A large Z-score, in either direction, indicates a prediction unlikely to occur by chance. In addition, Z-score directionality provides information regarding either inhibition or activation of a significant biological function. Biological functions and diseases that had absolute Z-scores of ≥ 2 were considered either significantly activated or inhibited. This analysis also allowed us to pinpoint differentially activated upstream regulators of these functions.

4.3 Results

4.3.1 Inter-array and intra-array normalization

In contrast to pre-normalization signal intensity distributions, post-normalization intensities were distributed within similar intervals and shared a similar density center, indicating successful adjustment of data (**Figure 4.3**). Post-meta-dataset-normalized signal intensity boxplots were found to distribute in similar intervals and shared a similar density center across all datasets, indicating successful adjustment of data across the meta-dataset. Appropriate data adjustment was also confirmed through MA plots that evaluated the dependency between expression log ratio and mean.

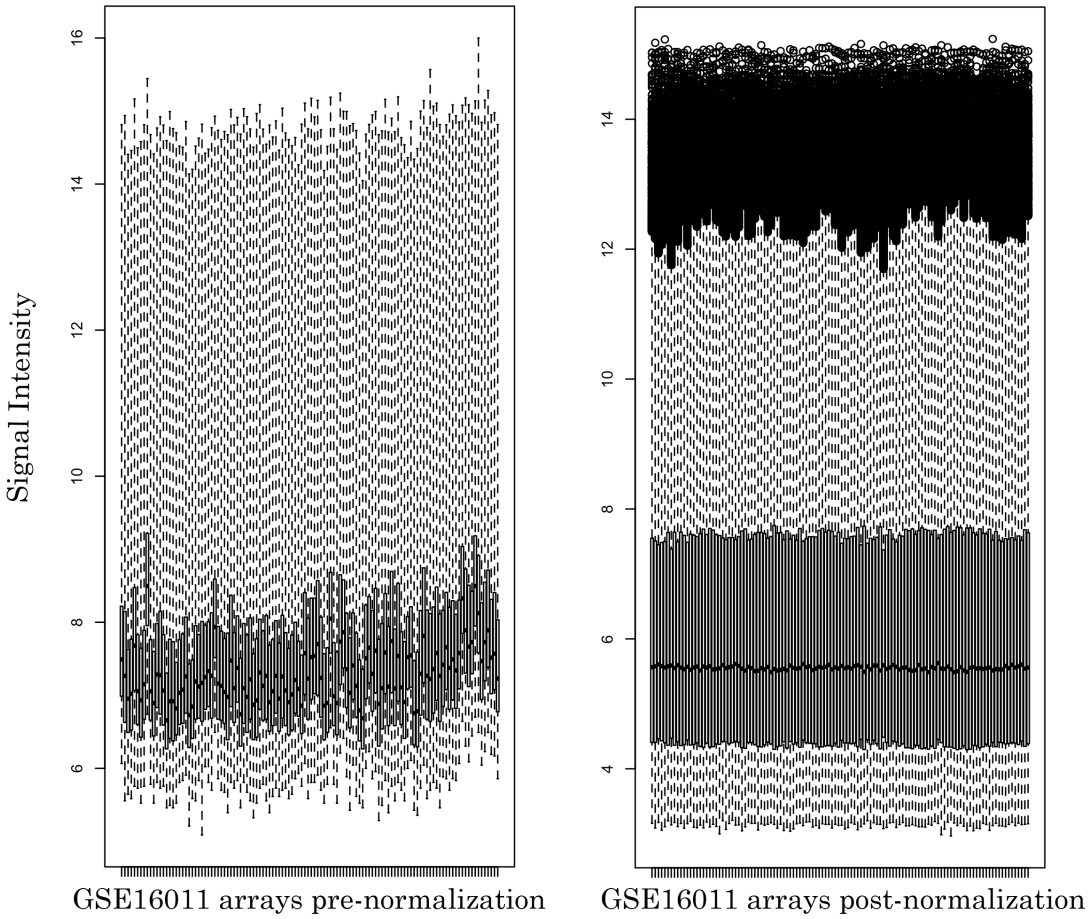


Figure 4.3 Example of a dataset signal intensity range before and after normalization with RMA

Signal intensity boxplots before *RMA* normalization indicated a dataset with differently distributed arrays with regard to both signal intensity median and range. Signal intensity boxplots after *RMA* normalization indicated a dataset with similarly distributed arrays with regard to both signal intensity median and range. Within-dataset normalization by *RMA* was followed by between-dataset normalization by *ComBat.R*. The y-axis measures probe signal intensity and the x-axis shows individual arrays for GBM dataset, GSE16011. Plots were made using the R statistical platform.

4.3.2 Dataset characteristics

In total, we collected brain tissue microarray data and associated clinical information for 77 HIV-1 subject (HIV+, with no GBM) samples, 59 HIV-1-associated neurocognitive disorder subject (HIV+/HAND+, with no GBM) samples, 228 glioblastoma multiforme subject (GBM+, with no HIV) samples, and 35 control subject samples (no HIV or GBM) (**Table 4.3**). Geographical region information was available for all samples, with 29.7% of samples from individuals located in the United States and 70.3% of samples from Europe, namely the Netherlands, Switzerland, and Germany. Subject age information was available for 96.8% of samples, with average ages of 46, 44, 49, and 48 years for HIV+, HIV+/HAND+, GBM+, and control groups, respectively (**Table 4.3**). In the HIV+/HAND+ group, 44.1% of samples were from subjects diagnosed with HIV-associated encephalopathy (HIVE), and remaining samples were from subjects diagnosed with general HIV-1-associated neurocognitive disorder (HAND). HIV treatment in HIV+ subjects included both HIV monotherapy and highly active antiretroviral therapy (HAART). In the HIV+/HAND+ group, the percentage of samples from subjects that had received treatment was 64.4%, and the percentage of samples from subjects that had not received treatment was 35.6%. In the HIV+ group, the percentage of samples from subjects that had received HIV treatment was 71.4%, and the percentage of samples from subjects that had not received HIV treatment was 28.6%. Tumor treatment in the GBM+ group included both radiation and chemotherapy. 78% of samples in this group were from subjects that had received radiation treatment, and 29% were from subjects that had received chemotherapy. No data were filtered out on the basis of the characteristics listed above.

Table 4.3 Clinical information on primary study sample groups

	Total Sample Size	Age Range (Years)	Average Age (Years)	Geographical Region	Gender	Chemotherapy	Radiation	HIV Treatment (ever)	HIV Treatment Type
HIV (No GBM)	77	unknown	46	77 USA	74 male 3 female	N/A	N/A	55 yes 22 no	10 PI 6 non-PI 39 unknown
HIV/ HAND (No GBM)	59	unknown	44	59 USA	56 male 3 female	N/A	N/A	38 yes 21 no	7 PI 5 non-PI 26 unknown
GBM (No HIV)	228	30-64	49	23 USA 15 Germany 109 Netherlands 81 Switzerland	163 male 65 female	66 yes 83 no 79 unknown	177 yes 51 unknown	N/A	N/A
Control (No HIV, No GBM)	35	unknown	48	24 USA 5 Netherlands 6 Switzerland	2 male 5 female 28 unknown	N/A	N/A	N/A	N/A

The HIV+ group included 77 individuals from the United States, primarily male and with an average age of 46 years. The HIV+/HAND+ group included 59 individuals from the United States, primarily male and with an average age of 44 years. The GBM+ group included 228 individuals from both the United States and Europe, primarily male, and with an average age of 49 years. Finally, the control group included 35 individuals from both the United States and Europe, with an unknown gender distribution, and an average age of 48 years.

4.3.3 Differential expression analysis

Differential expression analysis by *limma* showed over four times the number of differentially expressed probes in the Control versus GBM+ contrast as compared to either the Control versus HIV+ contrast or the Control versus HIV+/HAND+ contrast (p-value < 0.05, B-H adjusted p < 0.05) (**Table 4.4**). In addition, the B-H adjusted -log10 p-value range for differential expression calls in the Control versus GBM+ contrast was eight times greater that of either HIV+ contrast (**Figure 4.4**). Finally, the Log2-fold change range for the Control versus GBM+ contrast of -5.74 to 5.57 was nearly three times greater that of either the Control versus HIV+ contrast or the Control versus HIV+/HAND+ contrast, with respective Log2-fold ranges of -1.91 to 1.88 and -2.09 to 1.85 (**Figure 4.4**).

Table 4.4 Significant probe numbers as revealed by differential expression results

DE Contrasts	B-H adjusted p-value <0.05	B-H adjusted p-value <0.005 (*limma*)	Log2 fold change >1 (*limma*)	Log2 fold change >1, and B-H adjusted p-value <0.05 (*limma*)
HIV+ vs. Control	8,189 (*limma*) 9,617 (*RankProd*)	1,629	14 down-regulated 62 up-regulated	10 down-regulated 61 up-regulated
HIV+/HAND+ vs. Control	8,009 (*limma*) 9,050 (*RankProd*)	1,379	15 down-regulated 60 up-regulated	9 down-regulated 58 up-regulated
GBM+ vs. Control	34,905 (*limma*) 21,046 (*RankProd*)	28,104	2,549 down-regulated 2,622 up-regulated	2,549 down-regulated 2,622 up-regulated

limma differential expression (*DE*) results showed a 4-fold increase in the number of differentially expressed probes (B-H adjusted $p < 0.05$) in the Control versus GBM+ contrast as compared to either the Control versus HIV+ contrast or the Control versus HIV+/HAND+ contrast. Differential expression results at the more stringent B-H adjusted $p < 0.005$, showed an even greater difference between GBM+ and HIV+ groups, with a 17-fold increase in the number of differentially expressed probes in the GBM+ group contrast over either of the HIV+ group contrasts. In addition, we observed sensitivity of DE calls based on the DE tool used.

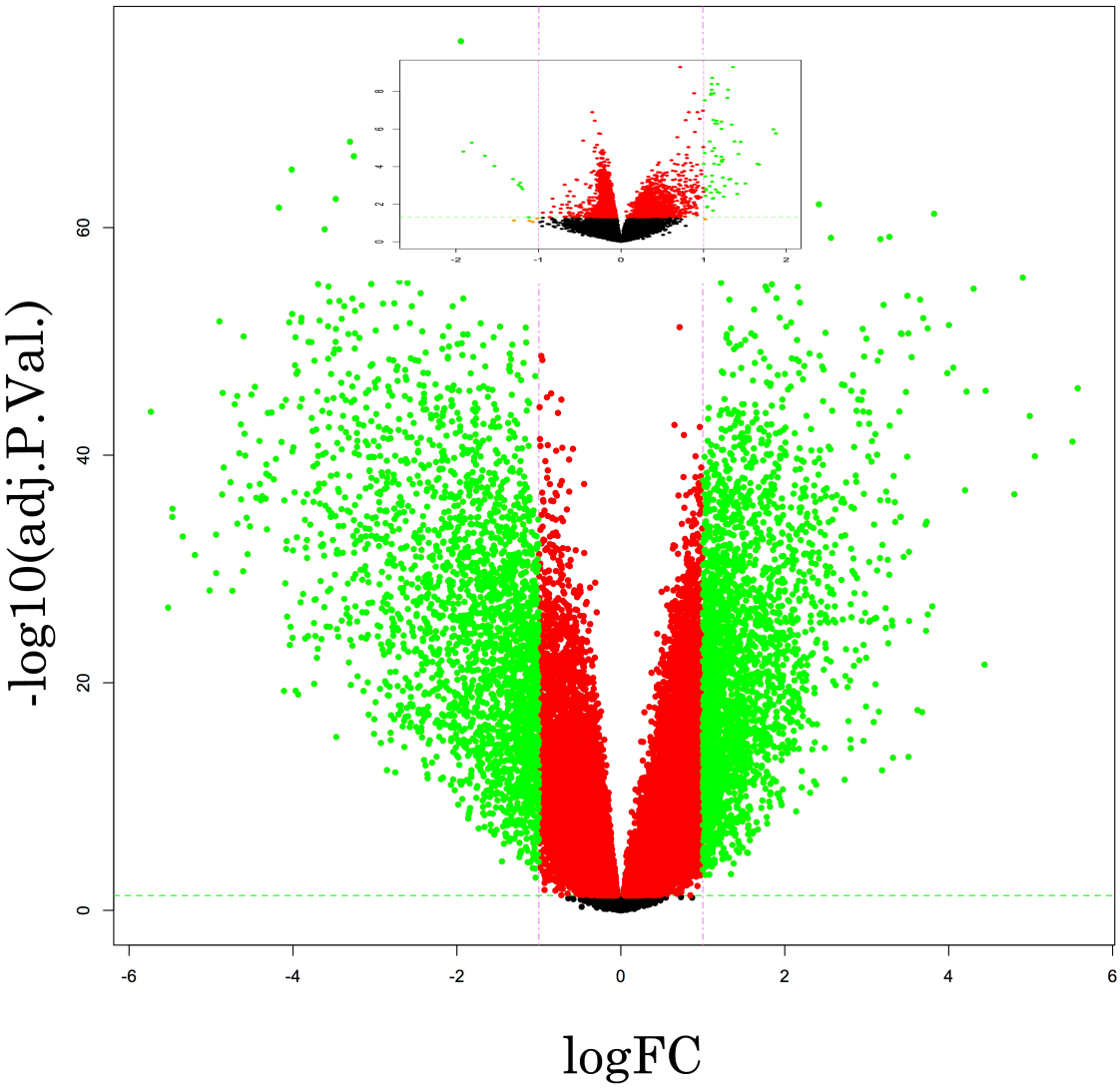

-log10(adj.P.Val.)
0
20
40
60
-6
-4
-2
0
2
4
6
logFC

Figure 4.4 Volcano plots of differential expression calls for control/GBM+ contrast and control/HIV+ contrast

The GBM+ versus Control contrast (outer figure) had differential expression (DE) calls with lower B-H adjusted p-values and greater Log2-fold changes than either the HIV+ (inset figure) or the HIV+/HAND+ versus Control contrasts. The outer figure shows DE results for the GBM+ group, and the inset figure shows DE results for the HIV+ group. The y-axis measures the –log10 of the B-H adjusted p-value for probes, and the x-axis measures the Log2 fold change for probes. Plots were made using the R statistical platform.

4.3.4 Signaling pathway perturbation analysis

Summation of significantly perturbed pathways (B-H adjusted $p < 0.05$) across both the *limma* and *RankProd* differential expression pipelines resulted in 60 KEGG signaling pathways with perturbed topologies in the HIV+/HAND+ group, 50 pathways in the HIV+ group, and 85 pathways in the GBM+ group. Overlap of pathways significant in both *limma* and *RankProd* output resulted in 17 KEGG signaling pathways with perturbed topologies in the HIV+ group, 16 in the HIV+/HAND+ group, and 49 in the GBM+ group. With regard to pathways shared across sample groups, summation of significant pathways across both *limma* and *RankProd* output resulted in a total of 42 pathways shared between HIV+ and GBM+, and a total of 51 pathways shared between HIV+/HAND+ and GBM+ (**Figure 4.5**). The following KEGG signaling pathways were significant in both the HIV+/HAND+ and GBM+ groups but not in the HIV+ group: *cell cycle* (hsa:04110), *focal adhesion* (hsa:04510), *type II diabetes mellitus* (hsa:04930), *prion diseases* (hsa:05020), *prostate cancer* (hsa:05215), *VEGF signaling pathway* (hsa:04370), *progesterone-mediated oocyte maturation* (hsa:04914), *vibrio cholerae infection* (hsa:05110), and *endocrine and other factor-regulated calcium reabsorption* (hsa:04961). All significant signaling pathways shared by both HIV+ and GBM+ were also shared by the HIV+/HAND+ group.

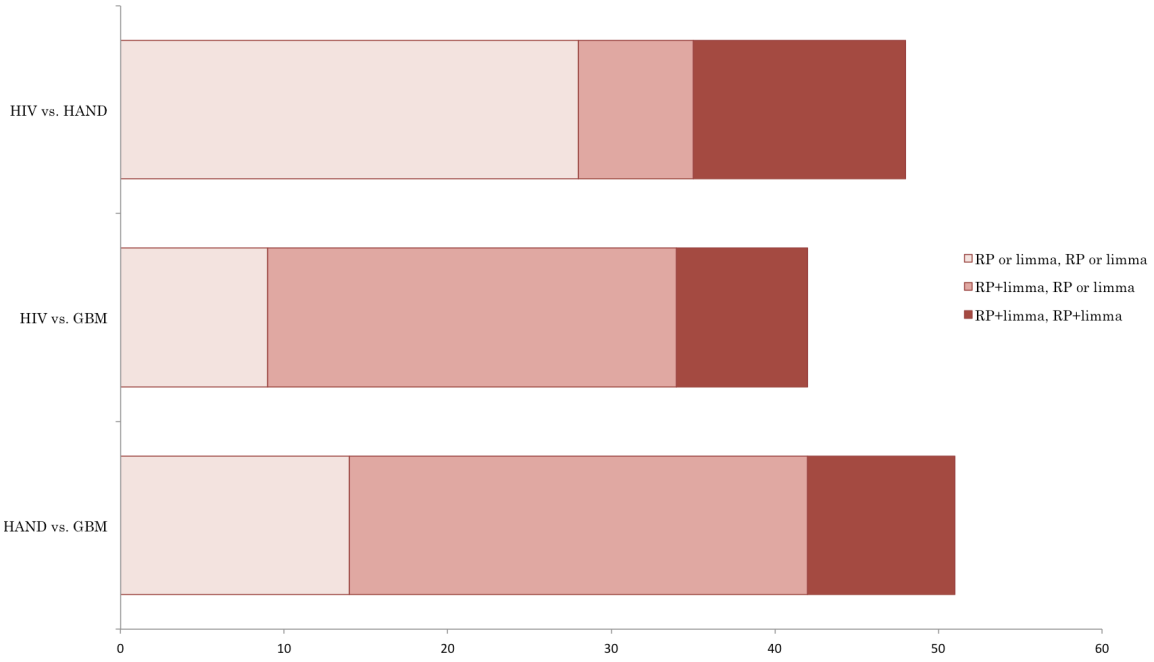
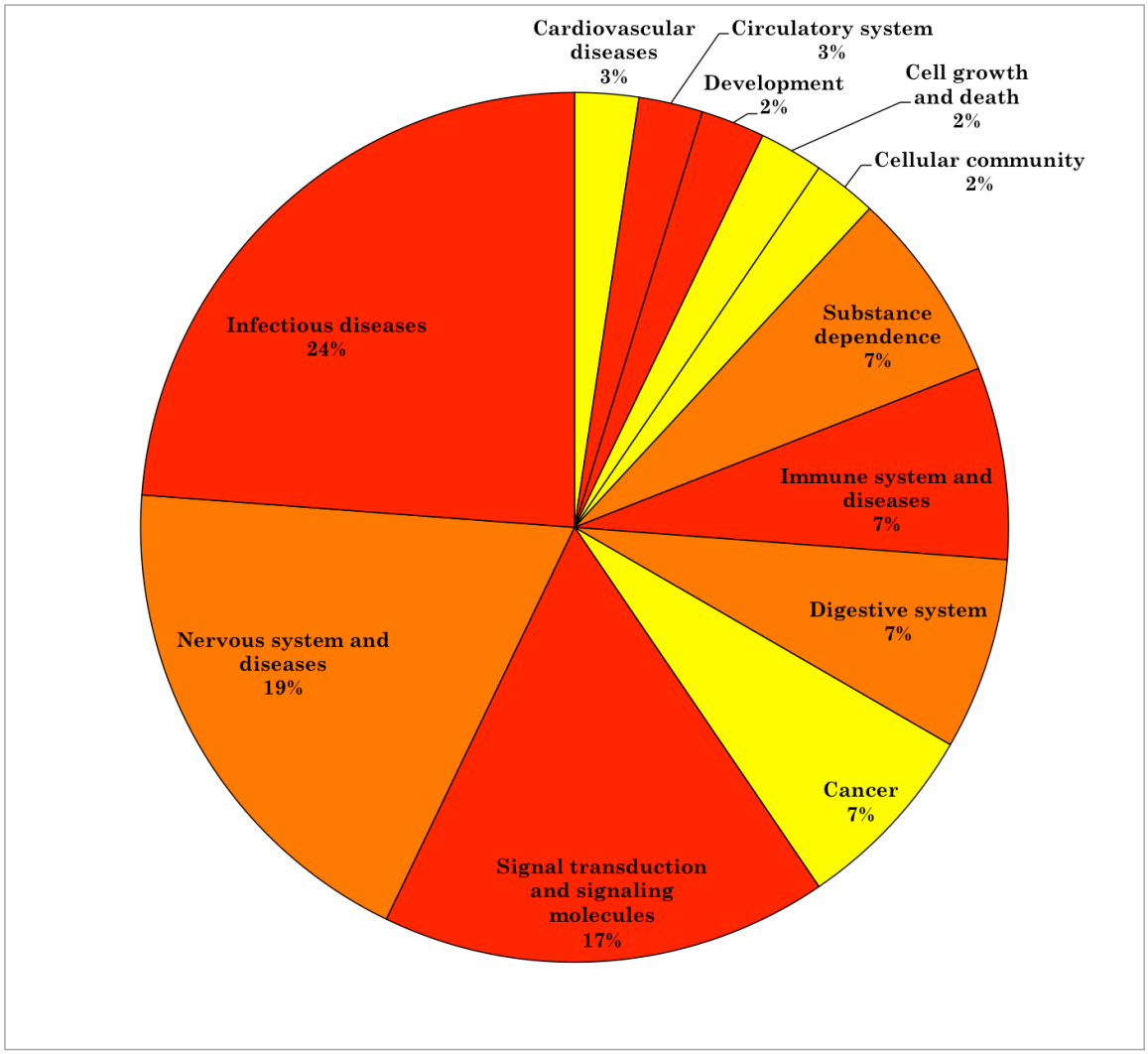


Figure 4.5 Number of shared pathways across contrasts as confirmed by *RankProd* and *limma* output

A total of 42 pathways were shared between the HIV+ and GBM+ group, and a total of 51 pathways were shared between the HIV+/HAND+ and GBM+ group. As expected, due to both groups sharing the same disease, the HIV+ group and the HIV+/HAND+ group shared the greatest number of pathways confirmed by both *RankProd* and *limma*. In addition, the HIV+/HAND+ group and the GBM+ group had more shared pathways confirmed by both DE tools as compared to those pathways shared between HIV+ and GBM+. This finding shows a greater similarity in the extent of altered gene expression between HIV individuals with HAND and GBM individuals, than between HIV individuals without HAND and GBM individuals. The y-axis shows group comparison, and the x-axis shows number of shared pathways.

Pathways shared between HIV+/HAND+ and GBM+ displayed the following distribution, eleven pathways belonged within KEGG category infectious diseases (including bacterial, viral and parasitic diseases), nine pathways within nervous system and neurodegenerative diseases, eight pathways within signal transduction and signaling molecules and interaction, four pathways within cancer, and three pathways within immune system and immune diseases (**Figure 4.6**). Nine additional KEGG categories had between one and three significantly perturbed pathways each, and these categories included cell growth and death, cellular community, circulatory system, and development (**Figure 4.6**).


Cardiovascular diseases 3%
Circulatory system 3%
Development 2%
Cell growth and death 2%
Cellular community 2%
Substance dependence 7%
Immune system and diseases 7%
Digestive system 7%
Cancer 7%
Signal transduction and signaling molecules 17%
Nervous system and diseases 19%
Infectious diseases 24%

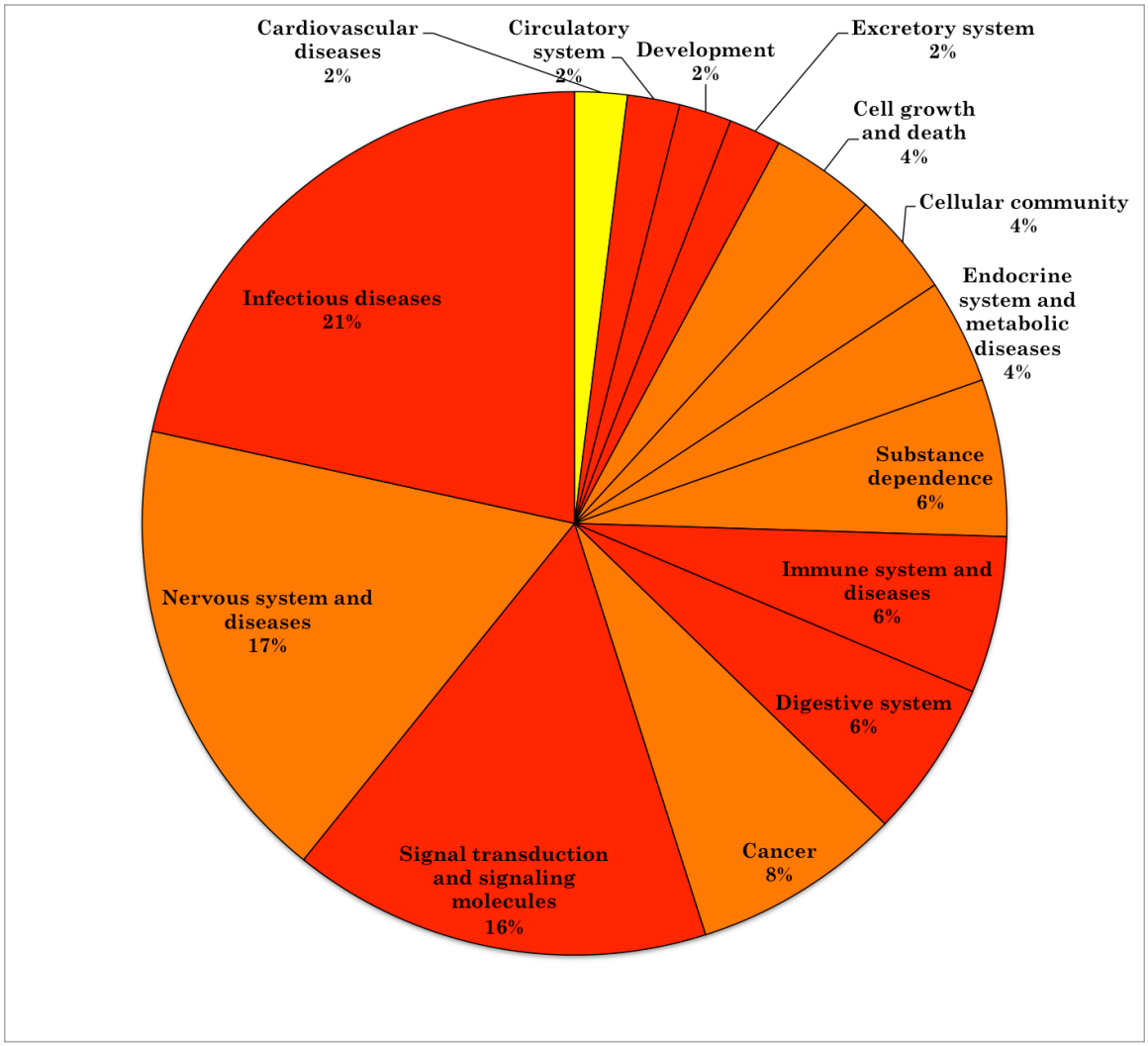


Figure 4.6 High-level KEGG categories with significantly perturbed pathways shared by HIV+ and GBM+ groups (top) and by HIV+/HAND+ and GBM+ groups (bottom)

Those KEGG categories with the greatest number of perturbed signaling pathways shared between the GBM+ and HIV+ groups (top figure) and between the GBM+ and HIV+/HAND+ groups (bottom figure) included infectious diseases, nervous system and diseases, signal transduction and signaling molecules, and cancer. In both figures above, KEGG category names are displayed along with the percentage of pathways within each category shared between the two disease groups (as confirmed by summation across

RankProd and *limma* results). The category color indicates the level of confirmation by DE analyses. Red represents categories with the highest levels of pathway significance confirmation (confirmed by both tools in both diseases in over 50% of shared pathways), and yellow represents categories with the lowest levels of pathway significance confirmation (confirmed by both tools in both diseases in under 25% of shared pathways).

Analysis of signaling pathway perturbation and enrichment in the HIV+ group with treatment as compared to the HIV+ group without treatment revealed pathways unique to each with regard to their overlap with GBM+. Three significant KEGG signaling pathways were shared by both the HIV+ treated group and GBM+, but not with the HIV+ treatment naive group: *calcium signaling pathway* (FDR-adjusted p = 0.004), the *NF-kappa β signaling pathway* (FDR-adjusted p = 0.015), and the *toxoplasmosis disease pathway* (FDR-adjusted p = 0.043). Six significant KEGG signaling pathways, primarily disease pathways, were shared by both the HIV+ treatment naive group and GBM+, but not with the HIV+ treatment group: *retrograde endocannabinoid signaling* (FDR-adjusted p = 0.001), *glutamatergic synapse* (FDR-adjusted p = 0.04), *pertussis* (FDR-adjusted p = 0.0005), *staphylococcus aureus infection* (FDR-adjusted p = 7.60E-07), *tuberculosis* (FDR-adjusted p = 0.002), *systemic lupus erythematosus* (FDR-adjusted p = 8.10E-06), and *viral myocarditis* (FDR-adjusted p = 0.037).

In all, these results demonstrated commonalities in signaling pathway perturbation across both HIV and GBM brain tissue. Those biological processes with the greatest number of significantly perturbed pathways in all three groups (HIV+, HIV+/HAND+, and GBM+) included infectious disease processes, cell signaling processes, nervous system processes, immune system processes, and cancer processes. In addition, more pathways shared between HIV+/HAND+ and GBM+ were confirmed as significant by both differential expression tools than pathways shared between HIV+ and GBM+, and while all pathways significant and shared between HIV+ and GBM+ were also shared between HIV+/HAND+ and GBM+, nine additional pathways were shared only by the HIV+/HAND+ and GBM+ groups.

4.3.5 Glioma pathway analysis

Based on output from differential expression tool *RankProd*, we found KEGG pathway, *glioma* (hsa:05214), as significant in all three groups, HIV+, HIV+/HAND+, and GBM+ (FDR adjusted p-values of 0.0015, 5.70E-06, and 0.006) (**Figure 4.7**). This finding was confirmed by *limma*, employing the empirical Bayes algorithm, for both the HIV+/HAND+ and GBM+ groups (FDR adjusted p-values of 0.022 and 0.01, respectively), but not for the HIV+ group (FDR adjusted p-value = 0.081) (**Figure 4.7**). In all, the results showed all three groups to have a significantly perturbed glioma pathway. In addition, we observed a sensitivity of these results based on the differential expression tool used.

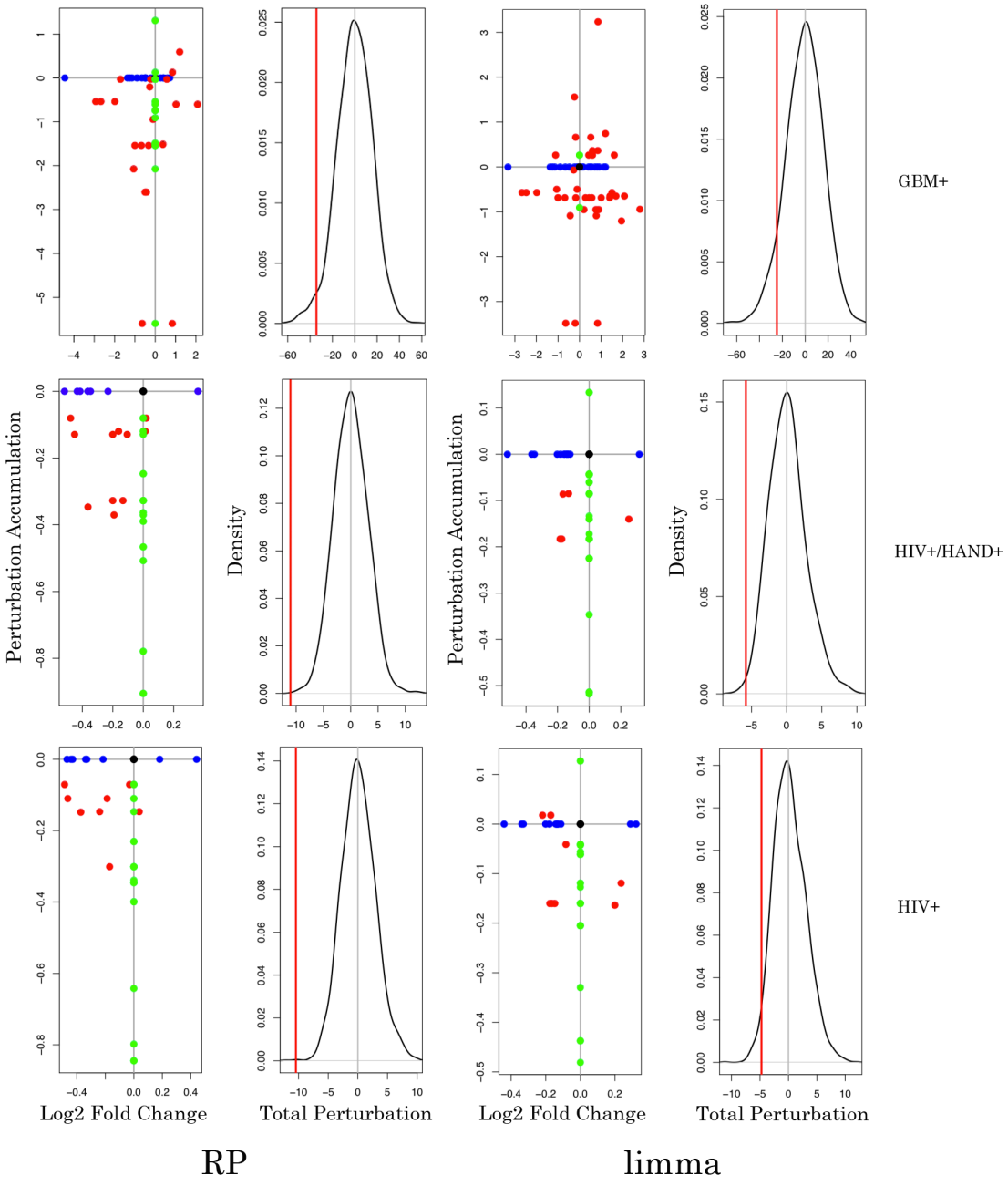

Perturbation Accumulation
Density
Perturbation Accumulation
Density
GBM+
HIV+/HAND+
HIV+
Log2 Fold Change
Total Perturbation
Log2 Fold Change
Total Perturbation
RP
limma

Figure 4.7 Perturbation plots for KEGG pathway hsa:05214 across the three groups HIV+, HIV+/HAND+ and GBM+

The *glioma pathway* was significantly perturbed in all three groups, HIV+, HIV+/HAND+, and GBM+ according to *RankProd* differential expression output, and in HIV+/HAND+ and GBM+ according to *limma* output. The left-hand figure shows the level of perturbation on the *glioma pathway* as contributed by genes in that pathway (the y-axis measures the pathway perturbation accumulation, and the x-axis measures the Log2 fold change per gene). For example, non-differentially expressed genes are indicated by a Log2-fold change of zero on the x-axis, and by a level of perturbation of zero on the y-axis. The perturbation accumulation information in the left-hand figure is tested against a null perturbation distribution in the right-hand figure. The right-hand figure shows the observed net accumulation of perturbation as a red vertical line over the null distribution of the net accumulated perturbation (the y-axis measures the density of net perturbation accumulation, and the x-axis measures the total perturbation accumulation). In this figure, a red vertical line moved away from the null distribution center of zero indicates a potentially significantly perturbed pathway.

4.3.6 Canonical pathway enrichment analysis

Pathway enrichment analysis showed twelve canonical pathways as both significantly enriched (B-H adjusted $p < 0.05$) and shared between all groups (**Figure 4.8**). KEGG categories with enriched canonical pathways overlapped with KEGG categories with significantly perturbed pathways, including, for example, immune system pathway, *antigen presentation* (*SPIA* GBM+, HIV+, and HIV+/HAND+ p-values for this pathway were 2.09E-06, 4.24E-09, 4.71E-10, respectively). Nine of these twelve significant canonical pathways, *antigen presentation*, *allograft rejection signaling*, *OX4O signaling*, *Type I diabetes mellitus signaling*, *autoimmune thyroid disease signaling*, *complement system*, *graft-versus-host disease signaling*, *communication between innate and adaptive immune cells*, and *crosstalk between dendritic cells and natural killer cells*, all belonged to the higher-level immune system and diseases category. Additional enriched pathways included *pathways involved in cancer*, *Alzheimer's disease*, and *cell division* (**Figure 4.8**). In sum, these results revealed twelve canonical pathways, the majority of which are involved in immune system processes, as significantly enriched in all three groups, HIV+, HIV+/HAND+, and GBM+. Of note, of these twelve pathways, those with the greatest differences between the HIV+ and GBM+ groups were the *complement system pathway* (of greater significance in the HIV+ group), the *antigen presentation pathway* (of greater significance in the HIV+ group), and the *molecular mechanisms of cancer pathway* (of greater significance in the GBM+ group).

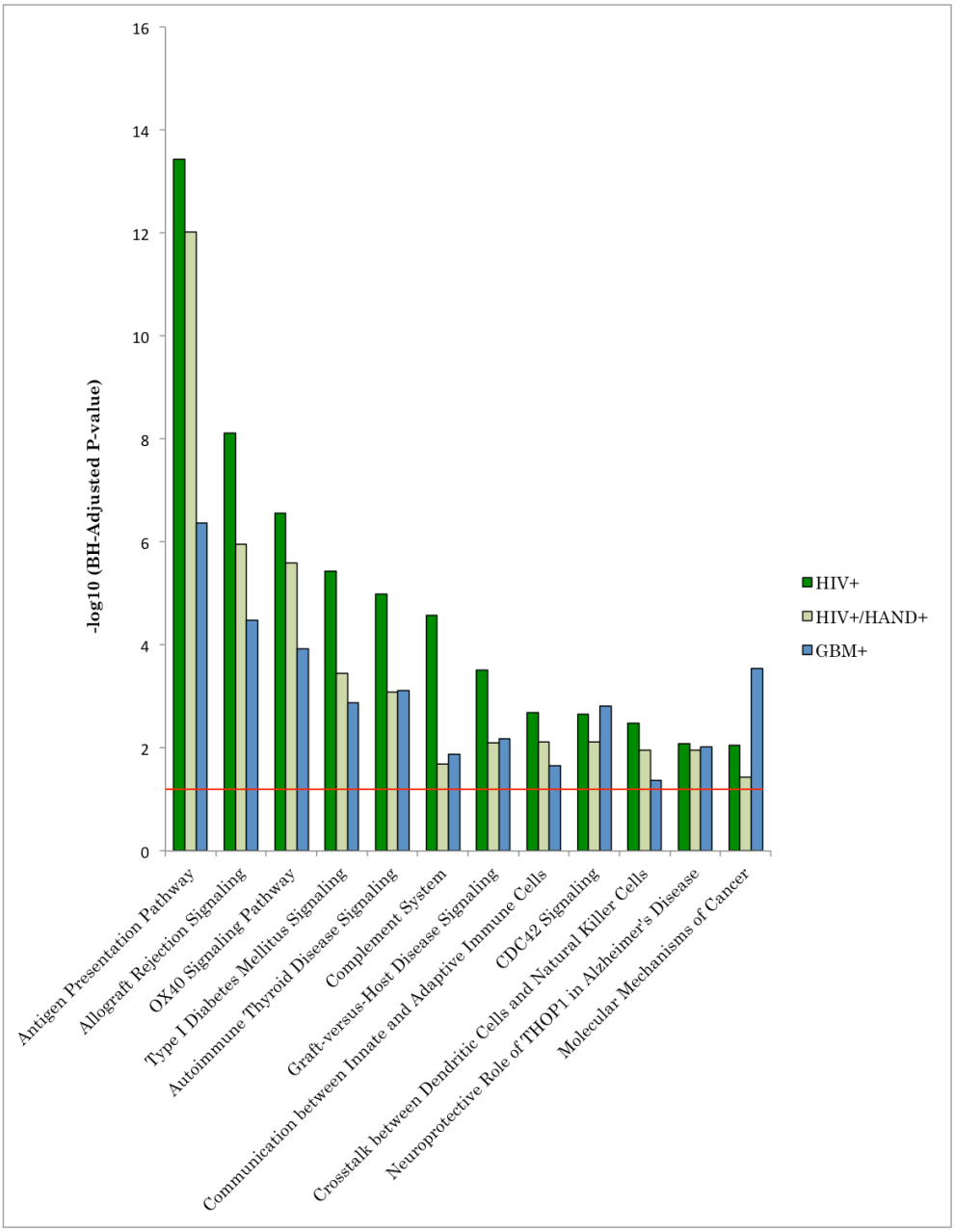


Figure 4.8 Significantly enriched canonical pathways shared across HIV+, HIV+/HAND+, and GBM+ groups

Twelve canonical pathways were significantly enriched in all three groups, HIV+, HIV+/HAND+, and GBM+ (horizontal red line at B-H adjusted $p < 0.05$). Taller bars show increased significance as measured by the –log10 of the B-H adjusted p-value on the y-axis, and canonical pathways are displayed along the x-axis.

4.3.7 Biological function and disease analysis

Functional analysis identified 34 biological functions and diseases as significantly enriched (B-H adjusted $p < 0.05$) and shared across all three groups (**Figure 4.9**). As was found in the canonical pathway enrichment analysis, these functions also overlapped with significant signaling pathways detected by *SPIA*, and included immunological disease, cardiovascular disease, neurological disease, infectious disease, cellular development, cellular movement, cell signaling, and cancer (**Figure 4.9**). Six of these higher-level functions were additionally predicted to have activated pathways in all three groups (Z-score ≥ 2) (**Figures 4.10a, b and c**), and no shared functions were predicted as having inhibited pathways. Functions with activated pathways in all three groups included cardiovascular system development and function, cell death and survival, cellular development, cell morphology, cellular movement, and cell-to-cell signaling and interaction. Functions exhibited different combinations of pathway activation across sample groups, including pathways activated across all sample groups and pathways activated in only one sample group. For example, in function, cell death and survival, the three pathways, *cell survival*, *cell viability* and *cell viability of tumor cell lines* were activated in all three groups with consistently higher activation in the HIV+ (Z-scores = 5.332, 5.255, 4.482) and HIV+/HAND+ groups (Z-scores = 4.706, 4.631, 3.999) as compared to the GBM+ group (Z-scores = 2.69, 2.769, 2.717) (**Figure 4.10b**). The cell death and survival pathway, *cell death of glioma cells*, however, was only activated in the HIV+ group (Z-score = 2.036). Function, cellular development, only had two pathways activated across all three groups, *proliferation of tumor cell lines* and *proliferation of breast cancer cell lines* (**Figure 4.10a**). Shared and activated pathways in the cellular

movement function included *cell movement* and *invasion of cells* (**Figure 4.10a**). Finally, cardiovascular system development and function had the greatest percentage of shared and activated pathways including *angiogenesis*, *vasculogenesis*, *development of blood vessel*, and *development of cardiovascular system* (**Figure 4.10c**). Biological functions that did not share activated pathways across the HIV+ and GBM+ groups included cell morphology and cell-to-cell signaling and interaction (**Figure 4.10b**). Pathways that differed in cell morphology included *autophagy* and *sprouting* activated in the HIV+ (Z-scores = 2.381, 2.356) and HIV+/HAND+ groups (Z-scores = 2.662, 2.375), but not in the GBM+ group (Z-score = 0). On the other hand, three *cell-spreading pathways* in this same function were only activated in the GBM+ group (Z-scores = 2.224, 2.156, 3.22) and not in the HIV+ (Z-scores = 0, 0, 0) or HIV+/HAND+ groups (Z-scores = 0, 0, 0) (**Figure 4.10b**). Finally, activated pathways in function neurological disease were found only in the HIV+ groups and not in the GBM+ group. These pathways included *neuromuscular disease*, *multiple sclerosis*, and *progressive motor neuropathy*. While not significant, there were two pathways, *central nervous system* and *glioma*, that did exhibit activation in the GBM+ group (Z-scores = 1.912, 1.912), but not in either HIV+ group (**Figure 4.10c**). In summary, 34 biological functions were significantly enriched (i.e. exhibited an overrepresentation of differentially expressed genes) and in common to all three groups, HIV+, HIV+/HAND+, and GBM+. In addition, signaling pathways within these functions overlapped with pathways detected by *SPIA* as significant (**Figure 4.6**) and included pathways involved in immune system and disease, cardiovascular system and disease, nervous system and disease, and cancer.

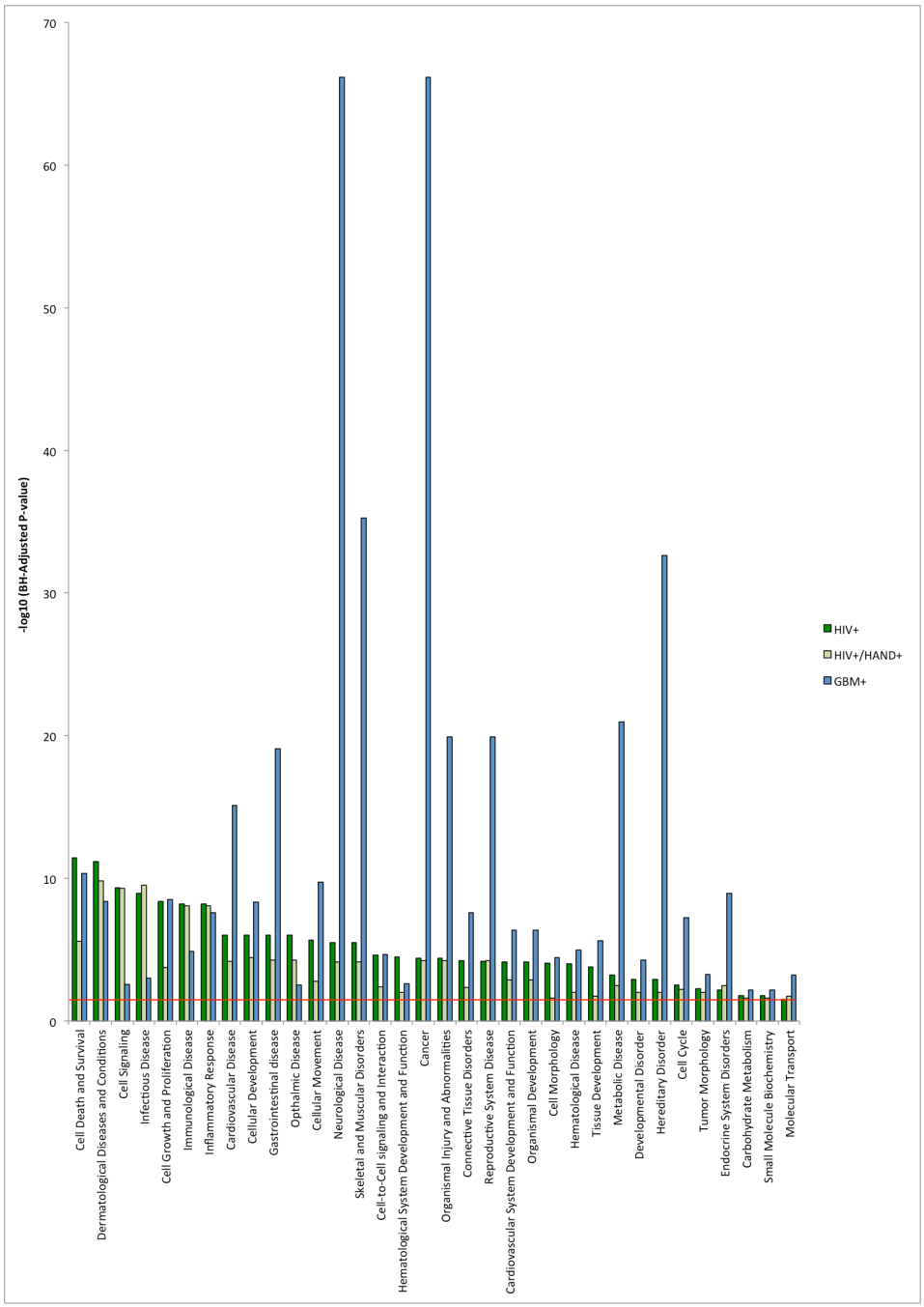
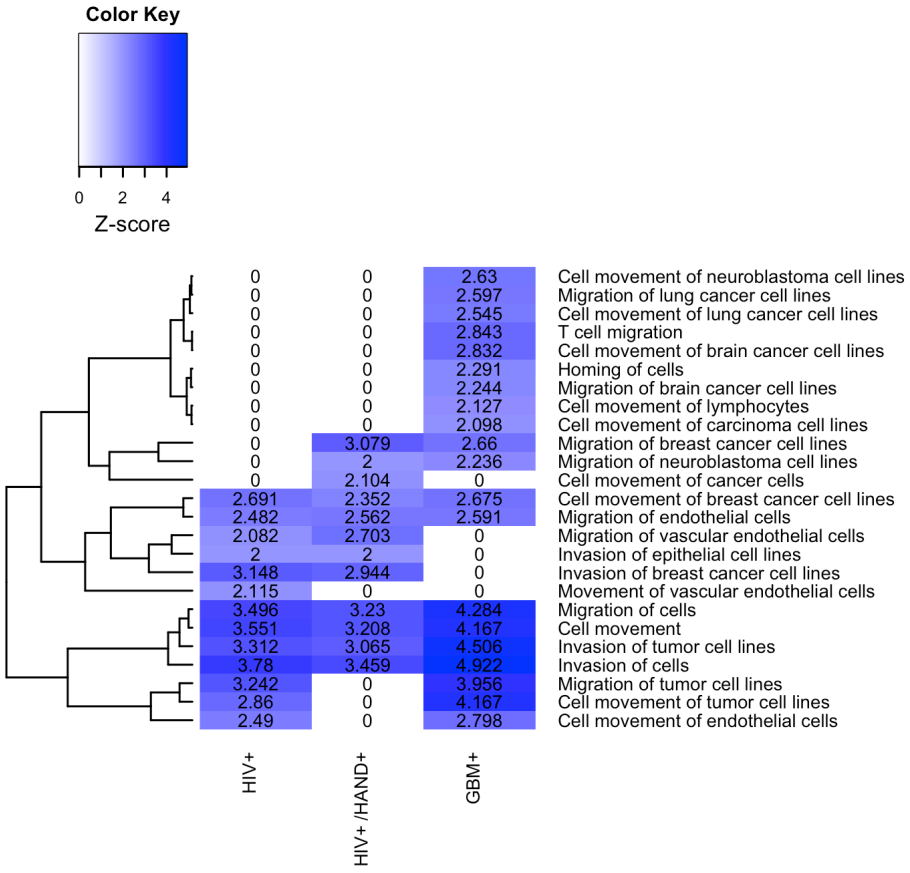


Figure 4.9 Significantly enriched biological functions and diseases shared across HIV+, HIV+/HAND+, and GBM+ groups

Thirty-four biological functions and diseases were significantly enriched in all three groups, HIV+, HIV+/HAND+, and GBM+ (horizontal red line at B-H adjusted $p < 0.05$). Taller bars show increased significance as measured by the –log10 of the B-H adjusted p-value on the y-axis, and biological functions are displayed along the x-axis.


Color Key
0 2 4
Z-score
0 0 2.63 Cell movement of neuroblastoma cell lines
0 0 2.597 Migration of lung cancer cell lines
0 0 2.545 Cell movement of lung cancer cell lines
0 0 2.843 T cell migration
0 0 2.832 Cell movement of brain cancer cell lines
0 0 2.291 Homing of cells
0 0 2.244 Migration of brain cancer cell lines
0 0 2.127 Cell movement of lymphocytes
0 0 2.098 Cell movement of carcinoma cell lines
0 3.079 2.66 Migration of breast cancer cell lines
0 2 2.236 Migration of neuroblastoma cell lines
0 2.104 0 Cell movement of cancer cells
2.691 2.352 2.675 Cell movement of breast cancer cell lines
2.482 2.562 2.591 Migration of endothelial cells
2.082 2.703 0 Migration of vascular endothelial cells
2 2 0 Invasion of epithelial cell lines
3.148 2.944 0 Invasion of breast cancer cell lines
2.115 0 0 Movement of vascular endothelial cells
3.496 3.23 4.284 Migration of cells
3.551 3.208 4.167 Cell movement
3.312 3.065 4.506 Invasion of tumor cell lines
3.78 3.459 4.922 Invasion of cells
3.242 0 3.956 Migration of tumor cell lines
2.86 0 4.167 Cell movement of tumor cell lines
2.49 0 2.798 Cell movement of endothelial cells
HIV+
HIV+ /HAND+
GBM+

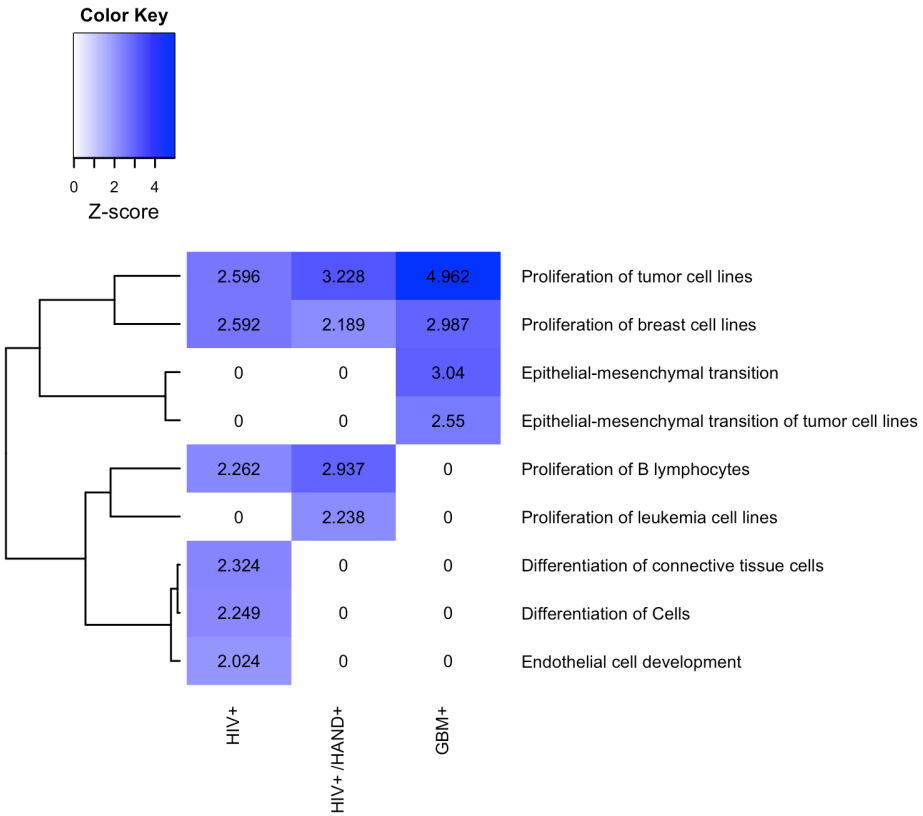
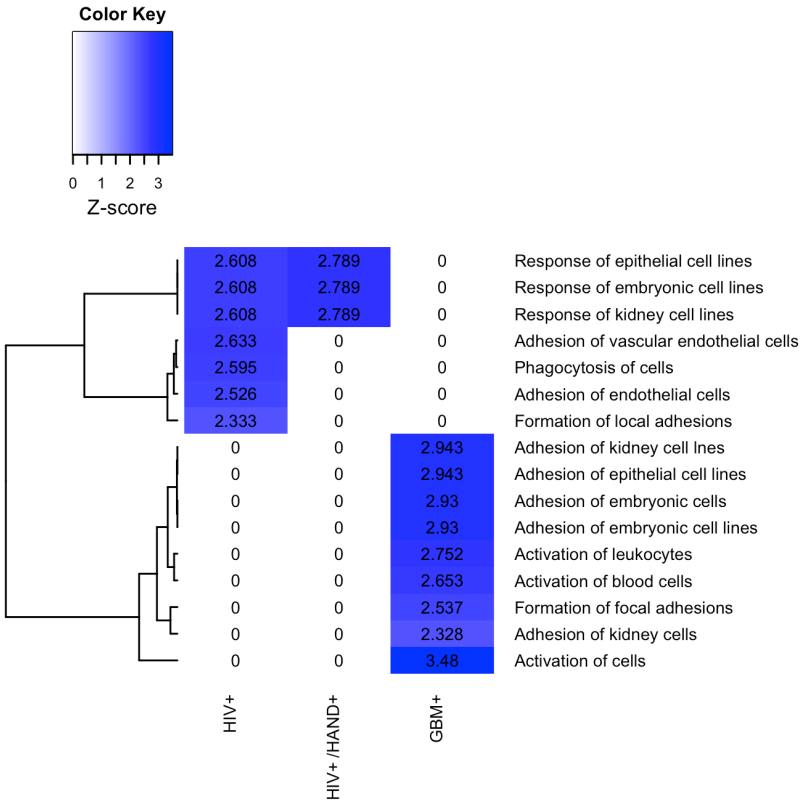
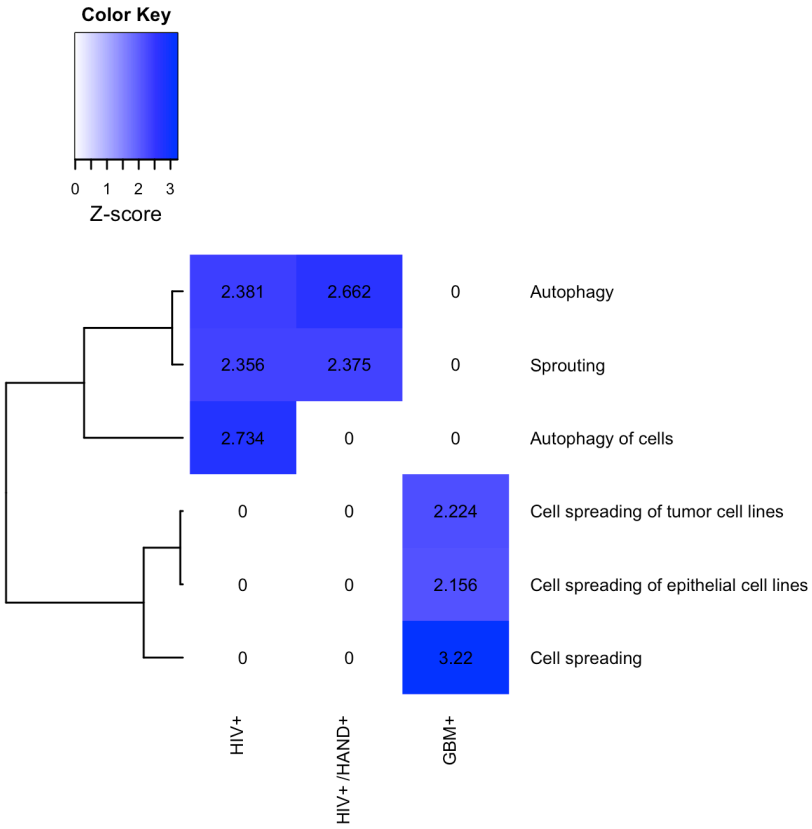

Color Key
0 2 4
Z-score
2.596 3.228 4.962 Proliferation of tumor cell lines
2.592 2.189 2.987 Proliferation of breast cell lines
0 0 3.04 Epithelial-mesenchymal transition
0 0 2.55 Epithelial-mesenchymal transition of tumor cell lines
2.262 2.937 0 Proliferation of B lymphocytes
0 2.238 0 Proliferation of leukemia cell lines
2.324 0 0 Differentiation of connective tissue cells
2.249 0 0 Differentiation of Cells
2.024 0 0 Endothelial cell development
HIV+
HIV+ /HAND+
GBM+

Figure 4.10a Z-score heatmap of pathways involved in cellular movement and cellular development across the three groups, HIV+, HIV+/HAND+, and GBM+

Biological functions, cellular movement (top figure) and cellular development (bottom figure), had activated pathways that overlapped the three groups, HIV+, HIV+/HAND+, and GBM+. Absolute Z-scores of 2 or greater were considered significant. Blue represents pathways with a positive Z-score and reflects pathway activation. Heatmaps were created using the complete linkage clustering method with a Euclidean distance measure through the R *gplots* package (v. 2.12.1).


Color Key
0 1 2 3
Z-score
2.608 2.789 0 Response of epithelial cell lines
2.608 2.789 0 Response of embryonic cell lines
2.608 2.789 0 Response of kidney cell lines
2.633 0 0 Adhesion of vascular endothelial cells
2.595 0 0 Phagocytosis of cells
2.526 0 0 Adhesion of endothelial cells
2.333 0 0 Formation of local adhesions
0 0 2.943 Adhesion of kidney cell lnes
0 0 2.943 Adhesion of epithelial cell lines
0 0 2.93 Adhesion of embryonic cells
0 0 2.93 Adhesion of embryonic cell lines
0 0 2.752 Activation of leukocytes
0 0 2.653 Activation of blood cells
0 0 2.537 Formation of focal adhesions
0 0 2.328 Adhesion of kidney cells
0 0 3.48 Activation of cells
HIV+
HIV+ /HAND+
GBM+


Color Key
0 1 2 3
Z-score
2.381 2.662 0 Autophagy
2.356 2.375 0 Sprouting
2.734 0 0 Autophagy of cells
0 0 2.224 Cell spreading of tumor cell lines
0 0 2.156 Cell spreading of epithelial cell lines
0 0 3.22 Cell spreading
HIV+
HIV+ /HAND+
GBM+

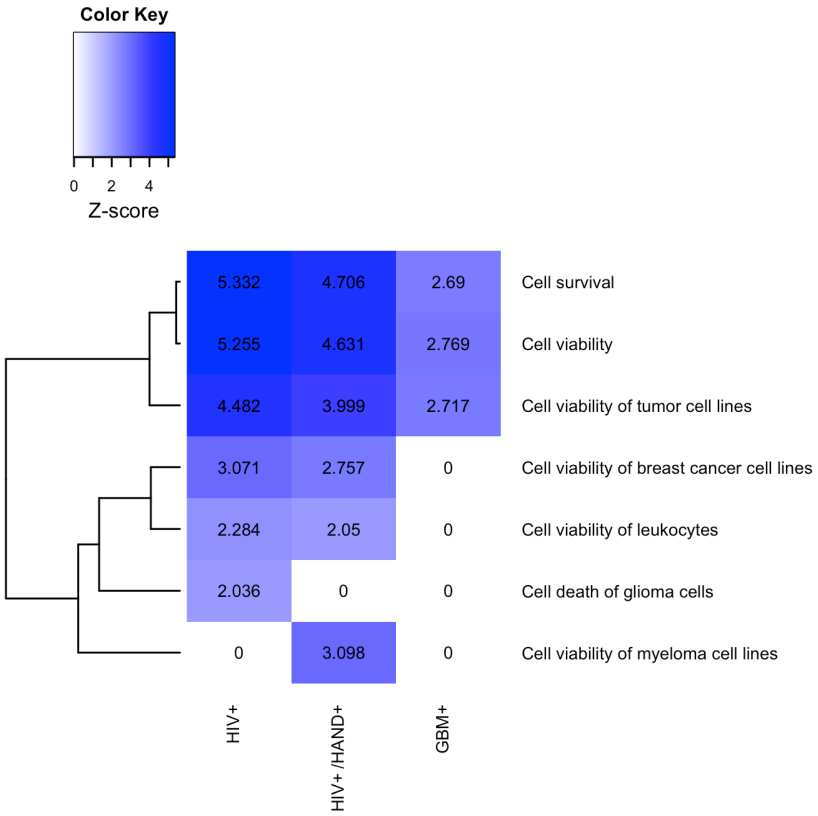
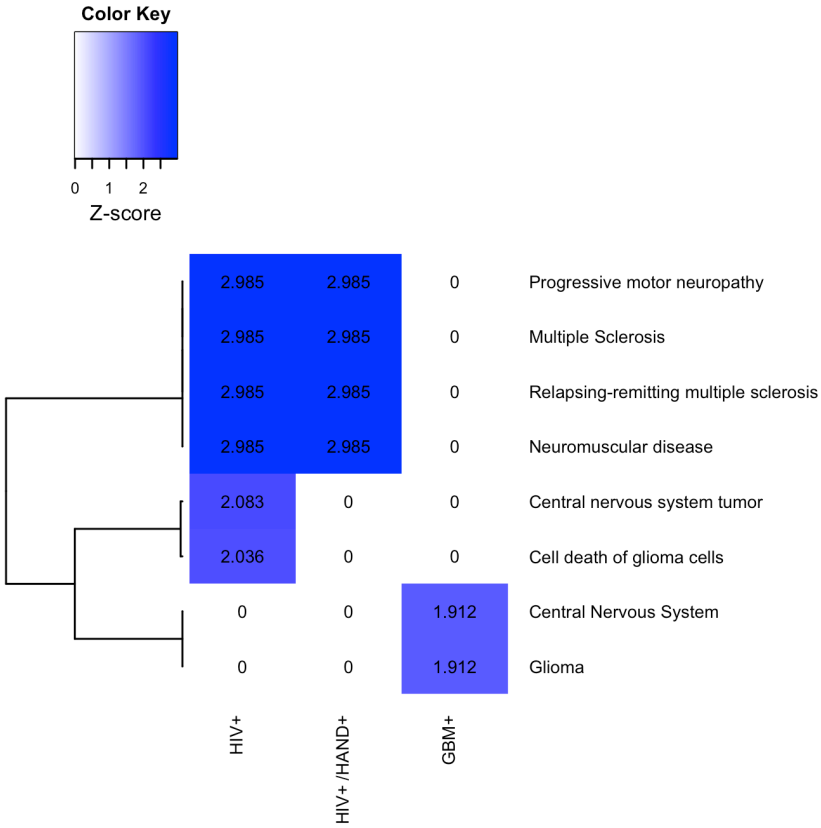


Figure 4.10b Z-score heatmap of pathways involved in cell-to-cell signaling and interaction, cell morphology, and cell death and survival across the three groups, HIV+, HIV+/HAND+, and GBM+

Biological functions, cell-to-cell signaling and interaction (top figure) and cell morphology (middle figure), did not have any activated pathways that overlapped both the HIV+ and GBM+ groups. Biological function, cell death and survival (bottom figure), had activated pathways that overlapped the three groups, HIV+, HIV+/HAND+, and GBM+. Absolute Z-scores of 2 or greater were considered significant. Blue represents pathways with a positive Z-score and reflects pathway activation. Heatmaps were created using the complete linkage clustering method with a Euclidean distance measure through the R *gplots* package (v. 2.12.1).


Color Key
0 1 2
Z-score
2.985 2.985 0 Progressive motor neuropathy
2.985 2.985 0 Multiple Sclerosis
2.985 2.985 0 Relapsing-remitting multiple sclerosis
2.985 2.985 0 Neuromuscular disease
2.083 0 0 Central nervous system tumor
2.036 0 0 Cell death of glioma cells
0 0 1.912 Central Nervous System
0 0 1.912 Glioma
HIV+
HIV+ /HAND+
GBM+

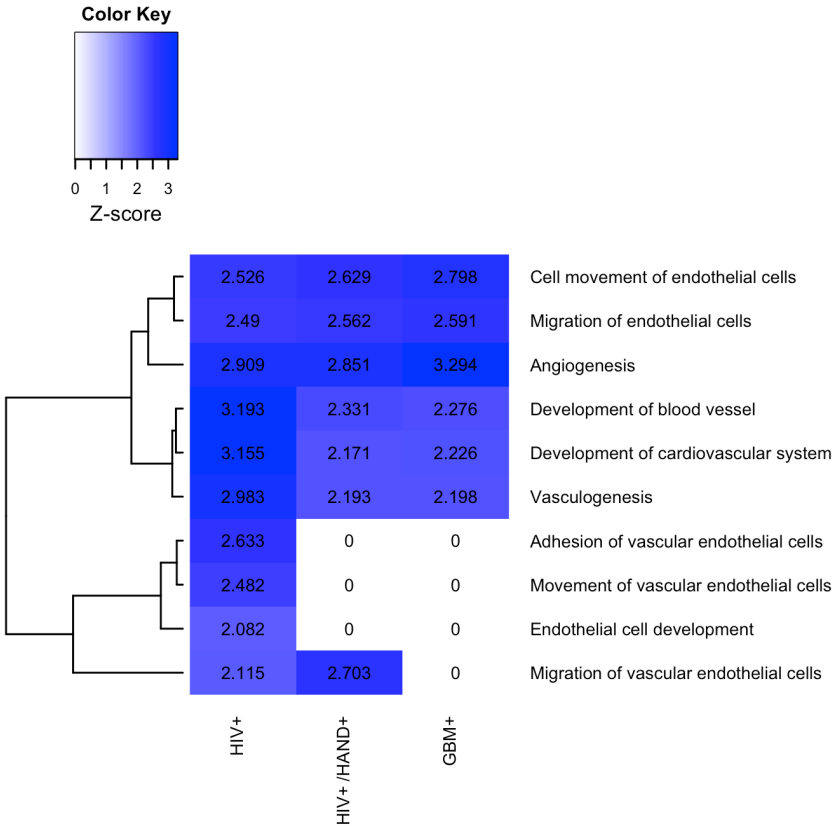
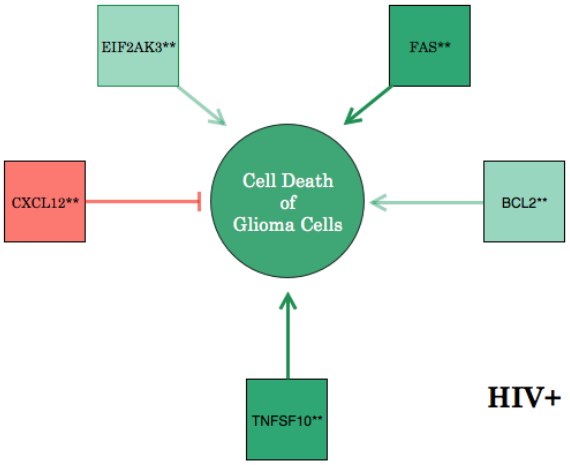


Figure 4.10c Z-score heatmap of pathways involved in neurological disease and in cardiovascular system development and function across the three groups, HIV+, HIV+/HAND+, and GBM+

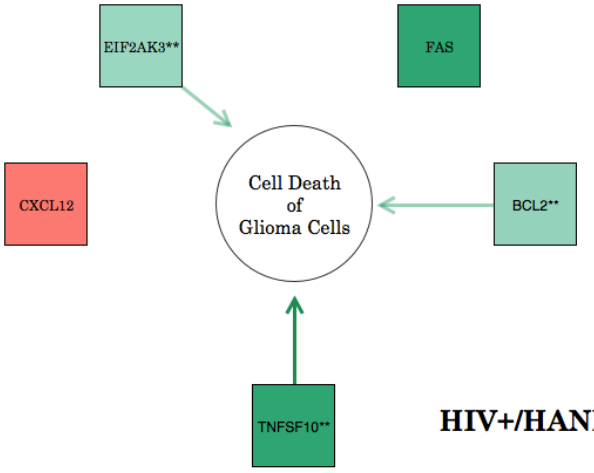
Disease function, neurological disease (top figure), did not have any activated pathways that overlapped the HIV+ and GBM+ groups. Of note is activation of pathway *cell death of glioma cells* in the HIV+ group, but not in the HIV+/HAND+ or GBM+ group. Biological function, cardiovascular system development and function (bottom figure), had activated pathways that overlapped the three groups, HIV+, HIV+/HAND+, and GBM+. Absolute Z-scores of 2 or greater were considered significant. Blue represents pathways with a positive Z-score and reflects pathway activation. Heatmaps were created using the complete linkage clustering method with a Euclidean distance measure through the R *gplots* package (v. 2.12.1).

4.3.9 Cell death of glioma cells pathway analysis

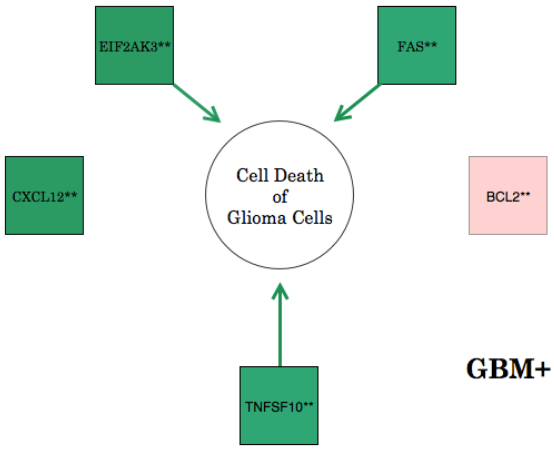
The biological pathway, *cell death of glioma cells*, was predicted as activated only in the HIV+ group, and not in either the HIV+/HAND+ or GBM+ group (**Figure 4.10b**). Five genes are involved in this pathway, namely B-cell CLL/lymphoma 2 (*BCL2*), chemokine (C-X-C motif) (*CXCL12*), eukaryotic translation initiation factor 2-alpha kinase (*EIF2AK3*), Fas cell surface death (*FAS*), and tumor necrosis factor (ligand) superfamily, member 10 (*TNFSF10*) (**Figure 4.11**). The HIV+ group exhibited significant differential expression for all five genes, including up-regulation of *BCL2*, *EIF2AK3*, *FAS*, and *TNFSF10* (Log2 ratios = 0.336, 0.352, 0.563, 0.549; B-H adjusted p-values = 0.02, 0.012, 0.03, 0.008), and down-regulation of *CXCL12* (Log2 ratio = -0.350, B-H adjusted p-value = 0.0367). The HIV+/HAND+ group exhibited significant differential expression for only three of the five genes, *EIF2AK3*, *BCL2*, and *TNFSF10* (Log2 ratios = 0.344, 0.400, 0.531; B-H adjusted p-values = 0.023, 0.008, 0.013), but not for *FAS* or *CXCL12* (Log2 ratios = 0.505, -0.343; B-H adjusted p-values = 0.070, 0.056). In the GBM+ group, while all of these genes were significantly differentially expressed, only three, *EIF2AK3*, *FAS*, and *TNFSF10* (Log2 ratios = 0.728, 0.729, 0.929; B-H adjusted p-values = 2.52E-13, 2.29E-06, 4.78E-07), had the necessary directionality for *glioma cell death pathway* activation (**Figure 4.11**). In sum, our results showed activation of the *glioma cell death pathway* in the HIV+ group but not in either the HIV+/HAND+ or GBM+ groups. All five genes involved in this pathway were significantly differentially expressed and had the necessary directionality toward pathway activation in the HIV+ group, while only three genes met these criteria in the GBM+ group.



HIV+



HIV+/HAND+



GBM+

Figure 4.11 Activation of molecules in the *cell death of glioma cells pathway* across the three groups, HIV+, HIV+/HAND+, and GBM+

Genes *BCL2*, *CXCL12*, *EIF2AK3*, *FAS*, and *TNFSF10* were significantly differentially expressed in the HIV+ group, and the *cell death of glioma cells pathway* was predicted as activated in this group. Red squares represent down-regulated genes, and green squares represent up-regulated genes. Green circles represent activated pathways, and white circles represent non-activated pathway. Differences in color saturation reflect differences in Log2 fold change, with increasing saturation indicating increasing absolute value of Log2 fold change. Asterisks mark those molecules that were significantly differentially expressed as compared to the control group.

4.3.10 Glioma cell death upstream regulator analysis

An upstream regulator analysis was performed to find those upstream regulators that both, were known to control *BCL2* and *CXCL12* genes found in the *glioma cell death pathway*, and that exhibited differing levels of activation between the HIV+ and GBM+ groups (**Figure 4.12**). Seven upstream regulators were predicted as significantly differentially activated between the HIV+ and GBM+ groups. Regulators activated in HIV+ but not in GBM+ included interferon-beta1 (IFNB1), growth hormone, and tumor protein p53 (TP53) (HIV+ Z-scores = 2.833, 2.339, 2.022; GBM+ Z-scores = 0, 0, -1.640). Regulators activated in GBM+ but not in HIV+ included transcriptional activator Myb (MYB) and P38 mitogen-activated protein kinase (P38 MAPK) (HIV+ Z-scores = 0.882, 1.399; GBM+ Z-scores = 2.586, 4.279). Regulators inhibited in HIV+ but not in GBM+ included ubiquitin specific peptidase 18 (USP18) and interferon regulatory factor 4 (IRF4) (HIV+ Z-scores = -2.588, -2.739; GBM+ Z-scores = 0, 0) (**Figure 4.12**). In summary, our results show differences between HIV+ and GBM+ with regard to upstream regulator control of *glioma cell death pathway* genes, potentially explaining the observed activation of this network in the HIV+ group but not in the GBM+ group. An experimental assay to systematically test the effects of each of these upstream regulators on the *glioma cell death pathway* would allow us to better understand the overall effect of HIV cellular environment on GBM development. More specifically, simulation of HIV upstream regulation, as described above, in a human glioblastoma cell line, would allow us to test the ability of each of these regulators to: change *BCL2* and *CXCL12* log fold change directionality; to activate the *glioma cell death pathway*; and to promote glioblastoma cell death.

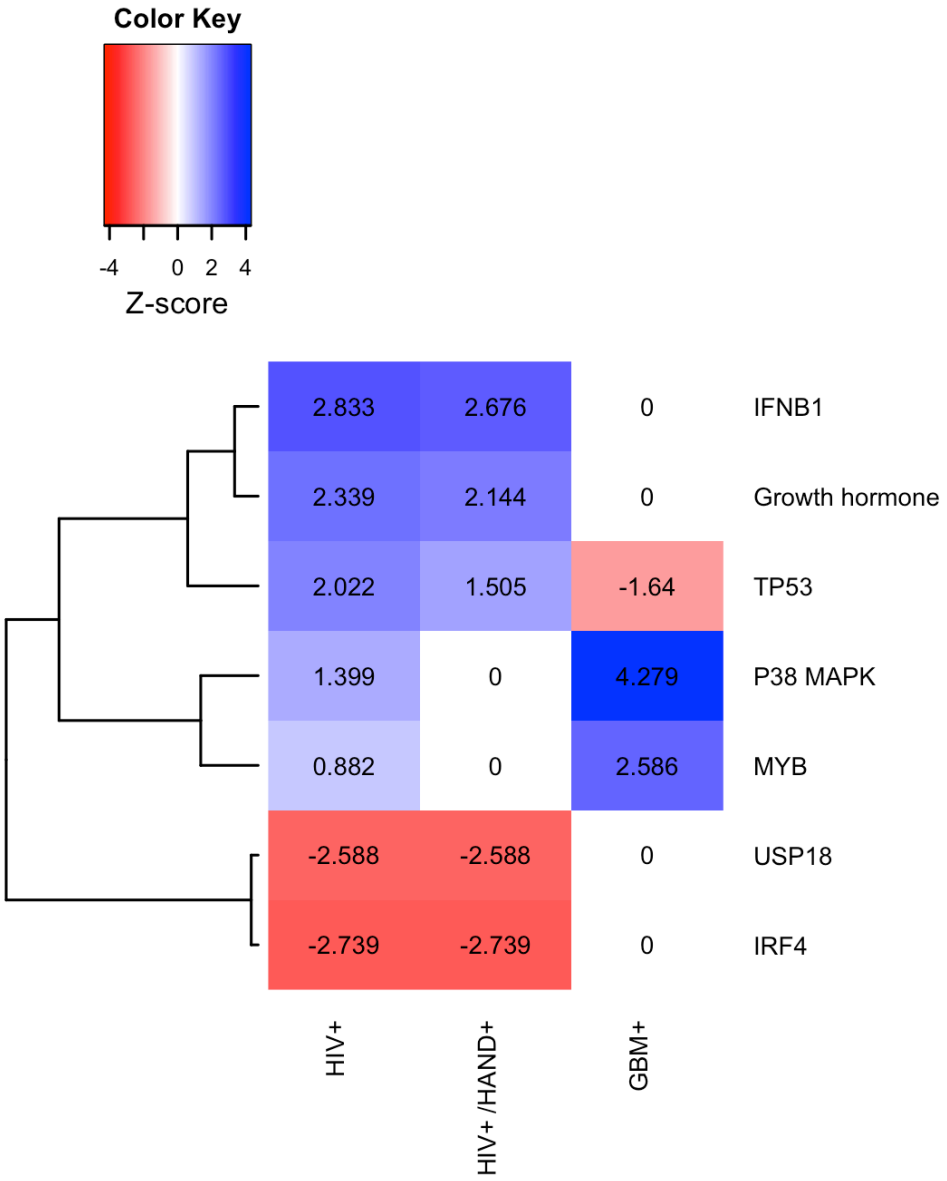


Figure 4.12 Z-score heatmap of upstream regulators involved in pathway *cell death of glioma cells* across the three groups, HIV+, HIV+/HAND+, and GBM+

Seven upstream regulators of *glioma cell death pathway* genes, *BCL2* and *CXCL12*, were differentially activated across the two groups, HIV+ and GBM+. These upstream regulator molecules included IFNB1, growth hormone, TP53, MYB, P38 MAPK, USP18, and IRF4. Absolute Z-scores of 2 or greater were considered significant. Red represents molecules with negative Z-scores (predicted as inhibited), blue represents molecules with

positive Z-scores (predicted as activated), and white represents molecules Z-scores of 0 (no predicted activation or inhibition). Heatmaps were created using the complete linkage clustering method with a Euclidean distance measure through the R *gplots* package (v. 2.12.1).

4.4 Discussion

Of all primary brain tumors, glioblastoma multiforme is the most malignant and most frequent [6]. A paucity of glioblastoma multiforme cases in the malignant-prone HIV-population [28] motivated us to perform a comparative analysis between the two diseases, HIV-infection of the brain and GBM development in the brain, in order to advance our understanding of clinical progression in each. Advantages to this study included: use of a meta-analysis approach in order to increase the power of our analyses; systematic normalization of data both within and between datasets; and implementation of multiple computational tools and statistical approaches in both our differential expression and pathway analyses for a more comprehensive study. By combining multiple methodologies we were able to statistically confirm our results while not limiting them, due to potential individual tool biases.

4.4.1 Overlap between high-level biological pathways

Our findings showed multiple biological processes as shared between the two brain diseases, HIV and GBM, and implicate use of these pathways in a shared neuropathogenesis (**Figures 4.6, 4.8 and 4.9**). **Immune system and disease:** The majority of canonical pathways shared by HIV and GBM belonged to the high-level biological process, immune system and disease. In addition, immune system pathway, *antigen presentation*, was the most significantly enriched canonical pathway for all three groups in our *IPA* pathway analysis, and the most significantly perturbed pathway for two groups (HIV+ and HIV+/HAND+) in our *SPIA* signaling pathway analysis (for GBM this pathway ranked 7 out of 137) (**Figure 4.8**). This result was in line with our predicted key commonalities between the two diseases. As had been reported prior to this work, the

immune system plays an important role in the neuropathogenesis of both diseases: in particular, use of microglia, the primary inflammatory cell of the CNS, in the spread of infection in HIV [72] and for tumor expansion in GBM [73, 74], and chronic immune activation leading to neurocognitive impairment in HIV [75, 76]. **Cancer:** Multiple cancer-associated pathways were also found to overlap the two diseases (**Figures 4.6, 4.8 and 4.9**), including the three KEGG signaling pathways, *transcriptional misregulation in cancer*, *pathways in cancer*, and *glioma*. Our observation of enriched and perturbed cancer pathways in HIV individuals confirmed the prior and unexpected finding by Levine et al. of disproportionate numbers of upregulated genes involved in cancer and cancer suppression correlating with severity of neurocognitive impairment in HIV and Alzheimer's disease (**Table 4.5**) [77]. Cancer gene *CTDSP2* (a gene reported to have a role in transcription [78]), previously found to be positively associated with increasing neurocognitive impairment in both Alzheimer's disease and HIV, was also found to increase in both our HIV+/HAND+ and GBM+ groups. In addition, cancer gene *HINT1* (a gene reported to have a role in apoptotic signaling via p53 [79]), previously found to be negatively associated with increasing neurocognitive impairment in both Alzheimer's disease and HIV, was also found to decrease in this study's HIV+/HAND+ and GBM+ groups. We were not able, however, to confirm a statistically significant increase in additional cancer genes reported to increase with neurocognitive impairment, including *SASH1*, *FBXW12*, *CASC3*, and *CEP350*, and *PGF* [77]. It is important to note that, while our sample size was larger (N = 77 (white matter from the anterior frontal lobe, grey matter from the frontal neocortex, and basal ganglia tissue) as compared to N = 42 samples (grey matter from the frontal neocortex and basal ganglia tissue)) (A.J. Levine,

personal communication, March 25, 2015), a portion of our sample group did overlap with that of Levine et al. [77]. An additional difference between the two studies included Levine et al. testing of two brain regions, basal ganglia and frontal cortex samples, while the current study tested three brain regions, white matter from the anterior frontal lobe, grey matter from the frontal neocortex, and basal ganglia tissue. This finding of altered expression of cancer-associated genes in the HIV individual was not reported in any of the three gene expression studies whose data were utilized in this work [46-48], most likely due to both different research foci and limited sample sizes in prior studies.

Table 4.5 Expression profiles across sample groups for genes found to correlate with neurocognitive impairment in Alzheimer's diseases and HIV in prior research

Gene Symbol	**Directionality with increasing NCI** (Levine)	**Biological Function**	**HIV+** logFC (for gene probe with highest logFC)	**HIV+** adjusted p-value	**HIV+/HAND+** logFC (for gene probe with highest logFC)	**HIV+/HAND+** adjusted p-value	**GBM+** logFC (for gene probe with highest logFC)	**GBM+** adjusted p-value
CTDSP2	Up	*Potential role in gene expression	0.230	0.108	0.290	0.052	0.785	1.34E-12
SASH1	Up	*Potential role in the innate immune system response	0.201	0.184	0.257	0.092	-0.459	3.69E-05
FBXW12	Up	F-box and WD repeat domain containing 12	0.140	0.258	0.149	0.255	0.046	0.627
CASC3	Up	Cancer susceptibility candidate 3 *Potential role in enzyme binding and protein binding	0.098	0.412	0.119	0.344	0.073	0.416
CEP350	Up	*Potential role in microtubule binding and protein binding	0.361	0.003	0.073	0.658	-0.037	0.724
PGF	Up	Placental growth factor	0.047	0.699	0.047	0.713	0.030	0.712
HINT1	Down	*Potential role in apoptotic signaling pathway by p53	-0.199	0.065	-0.235	0.035	-0.303	1.69E-05

**NCI = Neurocognitive impairment. Citations in order as denoted by asterisks in table above include: [78], [80], [81], [82], and [79]*

4.4.2 Cell death of glioma cells network activated in HIV group

Functional analysis revealed cancer-associated pathway, *cell death of glioma cells*, as activated in HIV brains but not in GBM brains (**Figure 4.11**). The five genes, B-cell CLL/lymphoma 2 (*BCL2*), C-X-C motif chemokine 12 (*CXCL12*), eukaryotic translation initiation factor 2-alpha kinase (*EIF2AK3*), Fas cell surface death (*FAS*), and tumor necrosis factor (ligand) superfamily, member 10 (*TNFSF10*), have all been implicated as participants in this network based on the *IPA* Ingenuity® Knowledge Base. In the HIV+ group, we found all five of these genes to be both differentially expressed and with the necessary directionality toward *glioma cell death pathway* activation, while in the GBM+ group, only three *glioma cell death pathway* genes met these criteria, *EIF2AK3*, *FAS*, and *TNFSF10*, with the remaining two genes, *BCL2* and *CXCL12*, while significantly differentially expressed, had the opposite Log2 fold change directionality (**Figure 4.11**).

BCL2: The *BCL2* gene encodes a membrane protein known to block apoptosis in the tumor cell [83]. We found it was significantly upregulated in both our HIV+ and HIV+/HAND+ groups. While classified as an oncogene for its typically anti-apoptotic effects, work by Shinoura et al. has shown that, in glioma cells, BCL2 protein exhibits both pro- and anti- apoptotic effects dependent on its level of expression: anti-apoptotic at low expression levels and pro-apoptotic at high expression levels [84]. In HIV, upregulation of *BCL2* has been shown to serve as a protective mechanism against viral protease apoptosis. In this capacity, BCL2 protein functions by suppressing activation of transcriptional regulator, nuclear factor kappa B (NF-κB), a key molecule in both cellular proliferation and in protection from apoptosis [85]. While our observed upregulation of *BCL2* in the HIV+ group but not in the GBM+ group could provide support for

Shinoura's findings, as well as evidence for BCL2 capacity to protect against glioma cell development, we expect this upregulation to more likely reflect an anti-apoptotic mechanism against viral protease apoptosis. It is also important to note that while protease inhibitor, Nelfinavir, has been shown to downregulate levels of *BCL2* [86], we observed the opposite effect. Limited information on treatment in the HIV+ group (71.43% of the HIV+ sample had been treated, and 13% of the sample had been treated with a protease inhibitor (**Table 4.3**)) made it difficult to statistically assess treatment effect on *BCL2* regulation. We did, however, observe a significant decrease in *BCL2* expression in our treatment group as compared to our non-treatment group (treated Log2 fold change = -0.2001, B-H adjusted p-value = 0.019701; non-treated Log2 fold change = 0.35388, B-H adjusted p-value = 0.10604), in line with the known effects of Nelfinavir on *BCL2*.

Our search for BCL2 regulator molecules revealed the following four regulators as differentially activated across the HIV+ and GBM+ groups: growth hormone, P38 mitogen-activated protein kinase (p38 MAPK), tumor protein 53 (TP53), and ubiquitin specific peptidase 18 (USP18). *BCL2* regulator, growth hormone, has been shown to inhibit apoptosis by upregulating *BCL2* [87]. In our study we found an activation of growth hormone in the HIV+ group as compared to the GBM+ group, in line with *BCL2* upregulation in our HIV+ group and not in our GBM+ group. The effect of phosphorylation of BCL2 by a second regulator, p38 MAPK, remains controversial [88], as p38 MAPK has been demonstrated to both increase [89] and decrease BCL2's anti-apoptotic potential [90]. In the current study, the GBM+ group was predicted as having a higher levels of p38 MAPK activation as compared to the HIV+ group, potentially

serving as evidence for phosphorylation of p38 MAPK as a mechanism for increasing the anti-apoptotic behavior of BCL2 protein. Cellular stress induction of a third BCL2 regulator, TP53, has been shown to promote apoptosis through inhibition of BCL2 functionality, and potentially through activation of apoptosis regulator, BAX, another member of the BCL2 gene family [91]. This is in line with the observed activation of TP53 in our HIV+ group and the observed inhibition of TP53 in our GBM+ group. Finally, regulator USP18 has been shown to play a key role in regulating apoptosis of pancreatic beta cells, a target cell of Type I diabetes. Knockdown of *USP18*, in combination with other molecular signals, has been shown to trigger beta cell apoptosis by decreasing anti-apoptotic signals from molecules BCL2, MCL-1 and BCL-XL to the cell [92]. We observed an inhibition of USP18 in the HIV+ group but not in the GBM+, in line with its knockdown serving to enhance cell recovery from BCL2 protein anti-apoptotic signals.

FAS: The *FAS* gene, involved in Fas-associated death receptor-induced apoptosis [93], was significantly upregulated in both our HIV+ and GBM+ groups (**Figure 4.11**). FAS-mediated caspase activation has been shown to elicit two independent functions in the glioma cell, leading to two opposing processes: cell cycle progression and apoptosis. In its first capacity, FAS activates glioma cell cycle progression and cell proliferation using the downstream, *Ras-Raf-MEK-ERK (MAPK/ERK) pathway* [94]. In its second capacity, FAS induces glioma cell apoptosis through T lymphocyte and natural killer (NK) cell activation, and both cell apoptosis and subject survival rate have been shown to correlate positively with cellular *FAS* expression [95, 96]. In HIV, FAS-mediated apoptosis through activation induced cell death (AICD) has been proposed as a

contributor to CD4+ cell depletion characteristic of HIV disease progression [97]. In addition, a recent study found levels of FAS receptor and FAS ligand to correlate negatively with CD4+ cell count in HIV individuals, *in vivo* support for the role of FAS in T-cell depletion [98]. It is also important to note that while protease inhibitor, Nelfinavir, has been shown to downregulate *FAS* [86], in our study we observed an upregulation of this protein. It is possible, therefore, that a potentially increased level of FAS-mediated apoptosis in our HIV group, as anticipated by upregulated levels of *FAS* observed in this group, both, reflects HIV disease progression, and provides protection against glioma initiation, through the activation of T lymphocyte and natural killer cells. In the GBM+ group, we anticipate FAS may act through its alternative capacity, as an activator of glioma cell proliferation.

TNFSF10: The *TNFSF10* gene encodes a cytokine known to promote apoptosis in tumor cells [99]. We found this gene was upregulated in both the HIV+ and HIV+/HAND+ groups (**Figure 4.11**). A member of the tumor necrosis factor family, TNFSF10 protein has been shown to induce apoptosis in combination with its ligand, tumor necrosis factor-related apoptosis-inducing ligand (TRAIL), and in a range of tumor cell types [100]. While inactivation of *TRAIL* is a hallmark of primary GBM development [101], we observed a significant upregulation of this molecule in our GBM+ group (Log2 fold change = 0.929; B-H adjusted p-value = 4.78E-07) (**Figure 4.11**). Interestingly, GBM cells have been shown to be resistant to TRAIL monotherapy, and more promising results have developed from studies looking to sensitize cells to TRAIL through TP53 activation [102, 103], a molecule we found to be inactivated in our GBM+ group (**Figure 4.12**). In HIV, HIV viral infectivity factor (Vif) protein has been shown to enhance TP53

stability and transcriptional activity toward viral proliferation [104]. In addition, a second viral protein, HIV-1 trans-activator of transcription (Tat), has also been shown to directly affect host levels of TP53 *in vivo*, through its inactivation of the apoptotic molecule [105, 106]. A reciprocal relationship between the two regulates viral replication, with viral latency disrupted when Tat levels are sufficiently high to suppress effective levels of TP53 [107]. Our HIV results fall in line with a Vif-mediated regulation of TP53, but not with a Tat downregulation. In our GBM+ group, inhibition of TP53 may explain failure of TNFSF10 activation in cell death of glioma cells. It is also important to note an alternative explanation for TP53 activation in our HIV+ group, therapy with protease inhibitor Nelfinavir (**Table 4.3**), as this anti-viral drug has been shown to upregulate both *TP53* and TRAIL receptor *DR5* [86].

CXCL12: CXCL12, a signaling molecule produced by cells in response to infection or injury, has been shown to stimulate proliferation of primary GBM stem cells, by activating production of vascular endothelial growth factor (VEGF) through the *PI3K/AKT signaling pathway* [108]. In line with this mechanism, we observed the significant upregulation of *CXCL12* in our GBM+ group, and its downregulation in our HIV+ group. CXCL12 is the unique ligand to chemokine receptor, CXCR4, a receptor hijacked by HIV for viral entry into CD4+ cells, and it follows that CXCR4 antagonists have been proposed as potential therapeutic agents in both cancer and HIV infection [109]. Downregulation of *CXCL12* in our HIV+ group potentially provided the HIV+ group a level of protection from glioma stem cell development.

Our search for CXCL12 regulator molecules revealed the following four regulators as differentially activated between the HIV+ and GBM+ groups: interferon

beta (IFNB1), interferon regulatory factor 4 (IRF4), transcriptional activator Myb (MYB), and tumor protein 53 (TP53). CXCL12 regulator, IFNB1, has been shown to strongly downregulate *CXCR4* (the CXCL12 ligand receptor) [110], vascular endothelial growth factor (*VEGF*), and matrix metallopeptidase 9 (*MMP-9*), all molecules involved in glioma neovascularization [111]. Predicted activation of IFNB1 in the HIV+ group but not in the GBM+ group is in line with both, the observed downregulation of *CXCL12* in the HIV+ group and with our hypothesis of activation of anti-glioma pathways in the HIV+ individual. A second CXCL12 regulator, IRF-4, has been shown to upregulate expression of chemokine receptor *CXCR4*, thereby promoting GBM cell migratory capabilities in response to increased CXCR4 binding with CXCL12 [112]. More specifically, use of quantitative PCR approach has shown a 2.5-fold increase in pre-B cell migration in response to CXCL12 [113]. Inhibition of IRF-4 and downregulation of *CXCL12* in our HIV+ group points to an additional protective mechanism in the HIV brain against glioma development. Overexpression of a third CXCL12 regulator, *MYB*, has been shown to elevate levels of CXCL12 in breast cancer cells [114]. The MYB molecule was predicted as significantly activated in the GBM+ group, but not in the HIV+ group, concordant with the upregulation of *CXCL12* observed in the GBM+ group but not in the HIV+ group. Finally, regulator TP53 has been shown to repress expression of *CXCL12* in human and mouse fibroblast cells [115]. Significant activation of TP53 in the HIV+ group and significant inhibition of TP53 in the GBM+ group, again, provides a potential explanation for *CXCL12* downregulation in the HIV+ brain.

EIF2AK3: The *EIF2AK3* gene, encoding PERK (protein kinase RNA-like ER kinase), is a key effector of the endoplasmic reticulum (ER) stress related unfolded

protein response (UPR) in cells [116]. *EIF2AK3* was upregulated in both our HIV+ and GBM+ groups. In glioma cells, this response has been shown to lead to cell death in otherwise chemo-resistant cancer cells, as evidenced through Nelfinavir-mediated UPR activation [31, 86]. In addition, primary glioma cell survival has been shown to decrease with increased melanoma differentiation associated gene-7 (*mda-7/IL-24*) cytokine-mediated activation of the ER stress response, involving PERK [117, 118]. While the mda-7/IL-24 protein apoptotic mechanism is still unclear, in human melanoma cell lines, it has been shown to decrease levels of B-cell lymphoma-extra large (BCLXL) and to a lesser extent, levels of BCL2 [119]. Therefore, a decrease in *BCL2* expression, as observed in the GBM+ group, may activate EIF2AK3 in stress-related apoptosis in the melanoma cell, a molecule upregulated in the GBM+ group. In HIV, HIV drug, Nelfinavir, has been shown to induce ER stress by blocking the action of proteasomes, leading to EIF2AK3 phosphorylation and protein synthesis inhibition [120]. Significant upregulation of *EIF2AK3* in both our HIV+ and GBM+ groups (Log2 fold change = 0.352 and 0.728, respectively) is not in line with its involvement in the *glioma cell death pathway*, and information concerning its rate of phosphorylation in both groups would help in better understanding its potential contribution in this pathway.

4.5 Limitations

While the findings of this study show key differences and commonalities between HIV- and GBM- associated biological pathways in the brain, it is important we also address limitations to these findings. **GBM tissue sampling information:** Detailed clinical information was limited in both the HIV+ and GBM+ groups, and in particular,

information on brain sampling locations for the GBM+ group was nearly non-existent. It is important to note that work by Levine et al. on HIV brains indicated changes in gene expression as reflective of sampling location [77]. Unavailability of this information for our GBM+ group, however, led to development of this study as a more general analysis of brain changes, and in line with this approach, we constructed an analysis pipeline that would provide an expanded view of potential overlap between the two diseases: we used two differential expression tools and multiple pathway analysis tools in order to fully capture all shared pathways. **HIV sample information:** We were also limited by availability of information on HIV treatment regimens, information that could have potentially allowed for further insight into protease inhibitor effects on tumor suppression in HIV individuals. Furthermore, we were limited by HIV sample availability. The HIV+ group used in this study, while inclusive of all currently available microarray data on HIV brain tissue, was still small (N = 77), and resulted in a considerable difference between the two group sizes (GBM N = 228). We addressed this limitation by matching subject age across the two groups, one of the most important predictors of both brain tumor risk and primary versus secondary GBM development [121, 122]. We also accounted for sample size restrictions by emphasizing trends over statistical significance in our discussion of results. **Meta-analysis limitations:** While most genes have been shown to exhibit reproducible expression across experiments [123], some technical variability is still retained, even after data normalization, and must be accounted for when working with meta-datasets [124]. Several steps were taken in this analysis to help remove technical variability. Two normalization steps were performed, the first to account for within-dataset variability, and the second for between-dataset normalization.

Following normalization, we measured gene differential expression using two approaches: use of the more classically employed moderated variance t-statistic (*limma*), as well as use of a rank product approach (*RankProd*). *Rank Product* has been shown to effectively take into account both low sample size and high noise in data, with its effectiveness having been reported in prior work [124-126]. In this study, we observed sensitivity of results to depend on both group and differential expression tool. For example, use of *limma* DE output in our signaling pathway impact analysis (*SPIA*) resulted in more conservative estimates of HIV+ and HIV+/HAND+ group significant signaling pathways than use of *RankProd* DE output. The opposite effect, however, was seen in the GBM+ group, with *RankProd* output resulting in more conservative *SPIA* significance estimates.

4.6 Conclusion

Despite a relatively high frequency of GBM occurrence in the general population, at around 3.13 cases per 100,000 individuals in the United States [121], coupled with a propensity for glial cell transformation in the HIV individual induced by HIV-specific mechanisms [28], GBMs have been rarely reported in the HIV population. This observation motivated a comparative analysis between the HIV-diseased brain and the GBM-diseased brain toward the elucidation of molecular mechanisms relevant in development of both diseases. As we hypothesized based on work in each field, HIV and GBM brains share pathways involved in the immune system and in the nervous system (**Figures 4.6, 4.8, and 4.9).** While we also anticipated perturbation of cancer pathways in the GBM+ group, we unexpectedly found evidence of cancer pathway enrichment and

perturbation using both pathway analysis tools in the HIV+ group as well (**Figures 4.7 and 4.10c**). This result confirmed an earlier observation by Levine et al. of expression of certain cancer genes in HIV individuals correlating positively with neurocognitive impairment.

Our finding of cancer pathway activation in the HIV+ group included results from a computational analysis of genes involved in cancer pathway, *cell death of glioma cells*, and their regulation: While the HIV brain was found to activate the *glioma cell death pathway* through differential expression of the five genes implicated in it (potentially through upstream regulators growth hormone, IFNB1, TP53, MYB, p38 MAPK, USP18, and IRF4), in GBM brains this protective pathway was not activated.

In addition, within the HIV+ group, we found differences between HAND and non-HAND individuals as well as between treated and non-treated individuals. HAND and non-HAND brains shared different numbers of perturbed signaling pathways with GBM brains: GBM and HAND brains shared more pathways and had more pathways validated by both differential expression tools, as compared to GBM and non-HAND brains (**Figure 4.5**). Differences between the treated and non-treated HIV+ groups included treated HIV individuals having fewer perturbed signaling pathways in common with GBM individuals (# of shared pathways = 7) as compared to non-treated HIV individuals (# of shared pathways = 11). In addition, pathways shared between the GBM+ group and the non-treated HIV+ group included more disease pathways than pathways shared between the GBM+ and treated HIV+ groups. These results indicate a greater similarity in neuropathogenesis between GBM brains and non-treated, HAND-positive, HIV brains.

Ultimately, we hope our finding of key commonalities and differences in the molecular mechanisms underlying both disease states to help shed light on potential treatment options for both HIV and GBM individuals.

4.7 References

1. WHO: **Fact Sheet #360**. In: *HIV/AIDS Fact Sheet #360.* vol. 2015. World Health Organization: World Health Organization; 2014.
2. Thompson MA, Aberg JA, Hoy JF, et al.: **Antiretroviral Treatment of Adult HIV Infection: 2012 Recommendations of the International Antiviral Society–USA Panel**. *JAMA: The Journal of the American Medical Association* 2012, **308**(4):387-402.
3. Heaton R, Clifford D, Franklin D, Woods S, Ake C, Vaida F, Ellis R, Letendre S, Marcotte T, Atkinson J: **HIV-Associated Neurocognitive Disorders Persist in the Era of Potent Antiretroviral Therapy Charter Study**. *Neurology* 2010, **75**(23):2087-2096.
4. Pardridge WM: **The Blood-Brain Barrier: Bottleneck in Brain Drug Development**. *NeuroRx* 2005, **2**(1):3-14.
5. Glass JD, Fedor H, Wesselingh SL, McArthur JC: **Immunocytochemical Quantitation of Human Immunodeficiency Virus in the Brain: Correlations with Dementia**. *Annals of Neurology* 1995, **38**(5):755-762.
6. Dolecek TA, Propp JM, Stroup NE, Kruchko C: **CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2005–2009**. *Neuro-Oncology* 2012, **14**(suppl 5):v1-v49.
7. Adamson C, Kanu OO, Mehta AI, Di C, Lin N, Mattox AK, Bigner DD: **Glioblastoma Multiforme: A Review of Where We Have Been and Where We Are Going**. *Expert Opinion on Investigational Drugs* 2009, **18**(8):1061-1083.

8. Burger PC, Vogel FS, Green SB, Strike TA: **Glioblastoma Multiforme and Anaplastic Astrocytoma Pathologic Criteria and Prognostic Implications**. *Cancer* 1985, **56**(5):1106-1111.

9. Eramo A, Ricci-Vitiani L, Zeuner A, Pallini R, Lotti F, Sette G, Pilozzi E, Larocca L, Peschle C, De Maria R: **Chemotherapy Resistance of Glioblastoma Stem Cells**. *Cell Death & Differentiation* 2006, **13**(7):1238-1241.

10. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA, Marosi C, Bogdahn U *et al*: **Radiotherapy Plus Concomitant and Adjuvant Temozolomide for Glioblastoma**. *New England Journal of Medicine* 2005, **352**(10):987-996.

11. McLendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogianakis GM, Olson JJ, Mikkelsen T, Lehman N, Aldape K: **Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways**. *Nature* 2008, **455**(7216):1061-1068.

12. Parsons DW, Jones S, Zhang X, Lin JC-H, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu I-M, Gallia GL: **An Integrated Genomic Analysis of Human Glioblastoma Multiforme**. *Science* 2008, **321**(5897):1807-1812.

13. Michor F, Polyak K: **The Origins and Implications of Intratumor Heterogeneity**. *Cancer Prevention Research* 2010, **3**(11):1361-1364.

14. Nakada M, Kita D, Watanabe T, Hayashi Y, Teng L, Pyko IV, Hamada J-I: **Aberrant Signaling Pathways in Glioma**. *Cancers* 2011, **3**(3):3242-3278.

15. Endersby R, Baker S: **PTEN Signaling in Brain: Neuropathology and Tumorigenesis**. *Oncogene* 2008, **27**(41):5416-5430.

16. Rao RD, James CD: **Altered Molecular Pathways in Gliomas: An Overview of Clinically Relevant Issues**. *Seminars in Oncology* 2004, **31**(5):595-604.

17. He J, Olson JJ, James CD: **Lack of p16INK4 or Retinoblastoma Protein (pRb), or Amplification-Associated Overexpression of cdk4 Is Observed in Distinct Subsets of Malignant Glial Tumors and Cell Lines**. *Cancer Research* 1995, **55**(21):4833-4836.

18. Ishii N, Maier D, Merlo A, Tada M, Sawamura Y, Diserens AC, Meir EG: **Frequent Co-Alterations of TP53, p16/CDKN2A, p14ARF, PTEN Tumor Suppressor Genes in Human Glioma Cell Lines**. *Brain Pathology* 1999, **9**(3):469-479.

19. Haas-Kogan D, Shalev N, Wong M, Mills G, Yount G, Stokoe D: **Protein Kinase B (PKB/Akt) Activity is Elevated in Glioblastoma Cells Due to Mutation of the Tumor Suppressor PTEN/MMAC**. *Current Biology* 1998, **8**(21):1195-S1191.

20. Chakravarti A, Zhai G, Suzuki Y, Sarkesh S, Black PM, Muzikansky A, Loeffler JS: **The Prognostic Significance of Phosphatidylinositol 3-Kinase Pathway Activation in Human Gliomas**. *Journal of Clinical Oncology* 2004, **22**(10):1926-1933.

21. Holmen SL, Williams BO: **Essential Role for Ras Signaling in Glioblastoma Maintenance**. *Cancer Research* 2005, **65**(18):8250-8255.

22. Rajasekharan S, Raman T: **Ras and Ras Mutations in Cancer**. *Open Life Sciences* 2013, **8**(7):609-624.

23. Fujisawa H, Reis RM, Nakamura M, Colella S, Yonekawa Y, Kleihues P, Ohgaki H: **Loss of Heterozygosity on Chromosome 10 is More Extensive in Primary (De Novo) than in Secondary Glioblastomas**. *Laboratory Investigation* 2000, **80**(1):65-72.

24. Sathornsumetee S, Reardon DA, Desjardins A, Quinn JA, Vredenburgh JJ, Rich JN: **Molecularly Targeted Therapy for Malignant Glioma**. *Cancer* 2007, **110**(1):13-24.

25. Turner NC, Reis-Filho JS: **Genetic Heterogeneity and Cancer Drug Resistance**. *The Lancet Oncology* 2012, **13**(4):e178-e185.

26. Gallagher B, Wang Z, Schymura MJ, Kahn A, Fordyce EJ: **Cancer Incidence in New York State Acquired Immunodeficiency Syndrome Patients**. *American Journal of Epidemiology* 2001, **154**(6):544-556.

27. Appay V, Sauce D: **Immune Activation and Inflammation in HIV-1 Infection: Causes and Consequences**. *The Journal of Pathology* 2008, **214**(2):231-241.

28. Cedeno-Laurent F, Trujillo JR: **Gliomas and Brain Lymphomas in HIV-1/AIDS Patients: Reflections from a 20-Year Follow up in Mexico and Brazil**. *Microbiology Research* 2011, **2**(1):e11.

29. Hall J, Short S: **Management of Glioblastoma Multiforme in HIV Patients: A Case Series and Review of Published Studies**. *Clinical Oncology* 2009, **21**(8):591-597.

30. Dubrow R, Darefsky AS: **Demographic Variation in Incidence of Adult Glioma by Subtype, United States, 1992-2007**. *BMC Cancer* 2011, **11**:325.

31. Pyrko P, Kardosh A, Wang W, Xiong W, Schönthal AH, Chen TC: **HIV-1 Protease Inhibitors Nelfinavir and Atazanavir Induce Malignant Glioma Death by Triggering Endoplasmic Reticulum Stress**. *Cancer Research* 2007, **67**(22):10920-10928.

32. Jiang Z, Pore N, Cerniglia GJ, Mick R, Georgescu MM, Bernhard EJ, Hahn SM, Gupta AK, Maity A: **Phosphatase and Tensin Homologue Deficiency in Glioblastoma Confers Resistance to Radiation and Temozolomide that is Reversed by the Protease Inhibitor Nelfinavir**. *Cancer Research* 2007, **67**(9):4467-4473.

33. Bernstein WB, Dennis PA: **Repositioning HIV Protease Inhibitors as Cancer Therapeutics**. *Current Opinion in HIV and AIDS* 2008, **3**(6):666-675.

34. Pajonk F, Himmelsbach J, Riess K, Sommer A, McBride WH: **The Human Immunodeficiency Virus (HIV)-1 Protease Inhibitor Saquinavir Inhibits Proteasome Function and Causes Apoptosis and Radiosensitization in Non-HIV-Associated Human Cancer Cells**. *Cancer Research* 2002, **62**(18):5230-5235.

35. Johnson GG, White MC, Grimaldi M: **Stressed to Death: Targeting Endoplasmic Reticulum Stress Response Induced Apoptosis in Gliomas**. *Current Pharmaceutical Design* 2011, **17**(3):284.

36. Pore N, Gupta AK, Cerniglia GJ, Maity A: **HIV Protease Inhibitors Decrease VEGF/HIF-1alpha Expression and Angiogenesis in Glioblastoma Cells**. *Neoplasia* 2006, **8**(11):889-895.

37. Paul RH, Sacktor NC, Valcour V, Tashima KT: **HIV and the Brain: New Challenges in the Modern Era**: Springer Science & Business Media; 2009.

38. Gimenez F, Fernandez C, Mabondzo A: **Transport of HIV Protease Inhibitors through the Blood-Brain Barrier and Interactions with the Efflux Proteins, P-Glycoprotein and Multidrug Resistance Proteins**. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2004, **36**(2):649-658.

39. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository**. *Nucleic Acids Research* 2002, **30**(1):207-210.

40. Murat A, Migliavacca E, Gorlia T, Lambiv WL, Shay T, Hamou MF, de Tribolet N, Regli L, Wick W, Kouwenhoven MC *et al*: **Stem Cell-Related "Self-Renewal" Signature and High Epidermal Growth Factor Receptor Expression Associated with Resistance to Concomitant Chemoradiotherapy in Glioblastoma**. *Journal of Clinical Oncology* 2008, **26**(18):3015-3024.

41. Lee Y, Scheck AC, Cloughesy TF, Lai A, Dong J, Farooqi HK, Liau LM, Horvath S, Mischel PS, Nelson SF: **Gene Expression Analysis of Glioblastomas Identifies the Major Molecular Basis for the Prognostic Benefit of Younger Age**. *BMC Medical Genomics* 2008, **1**:52.

42. Grzmil M, Morin P, Jr., Lino MM, Merlo A, Frank S, Wang Y, Moncayo G, Hemmings BA: **MAP Kinase-Interacting Kinase 1 Regulates SMAD2-Dependent TGF-Beta Signaling Pathway in Human Glioblastoma**. *Cancer Research* 2011, **71**(6):2392-2402.

43. Gravendeel LA, Kouwenhoven MC, Gevaert O, de Rooi JJ, Stubbs AP, Duijm JE, Daemen A, Bleeker FE, Bralten LB, Kloosterhof NK *et al*: **Intrinsic Gene Expression Profiles of Gliomas are a Better Predictor of Survival than Histology**. *Cancer Research* 2009, **69**(23):9065-9072.

44. Macy ME, Birks DK, Barton VN, Chan MH, Donson AM, Kleinschmidt-Demasters BK, Bemis LT, Handler MH, Foreman NK: **Clinical and Molecular Characteristics of Congenital Glioblastoma**. *Neuro-Oncology* 2012, **14**(7):931-941.

45. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DT, Konermann C, Pfaff E, Tonjes M, Sill M, Bender S *et al*: **Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma**. *Cancer Cell* 2012, **22**(4):425-437.

46. Tatro ET, Scott ER, Nguyen TB, Salaria S, Banerjee S, Moore DJ, Masliah E, Achim CL, Everall IP: **Evidence for Alteration of Gene Regulatory Networks through Micrornas of the HIV-Infected Brain: Novel Analysis of Retrospective Cases**. *PLoS One* 2010, **5**(4):e10337.

47. Borjabad A, Morgello S, Chao W, Kim SY, Brooks AI, Murray J, Potash MJ, Volsky DJ: **Significant Effects of Antiretroviral Therapy on Global Gene Expression in Brain Tissues of Patients with HIV-1-Associated Neurocognitive Disorders**. *PLoS Pathogens* 2011, **7**(9):e1002213.

48. Gelman BB, Chen T, Lisinicchia JG, Soukup VM, Carmical JR, Starkey JM, Masliah E, Commins DL, Brandt D, Grant I *et al*: **The National NeuroAIDS**

Tissue Consortium Brain Gene Array: Two Types of HIV-Associated Neurocognitive Impairment. *PLoS One* 2012, **7**(9):e46178.

49. Statistical Package R: **R: A Language and Environment for Statistical Computing**. *Vienna, Austria: R Foundation for Statistical Computing* 2009.

50. Bolstad BM: **Low-Level Analysis of High-Density Oligonucleotide Array Data: Background, Normalization and Summarization**. *Dissertation.* University of California, Berkeley; 2004.

51. Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S: **Bioinformatics and Computational Biology Solutions Using R and Bioconductor**: Springer Science & Business Media; 2006.

52. Brettschneider J, Collin F, Bolstad BM, Speed TP: **Quality Assessment for Short Oligonucleotide Microarray Data**. *Technometrics* 2008, **50**(3).

53. Miller, C. J: **Simpleaffy: Very Simple High Level Analysis of Affymetrix Data**. *R package version 2.28* (2007).

54. Bolstad, Ben: **AffyPLM: Model Based QC Assessment of Affymetrix GeneChips**. (2011).

55. Wilson CL, Miller CJ: **Simpleaffy: A Bioconductor Package for Affymetrix Quality Control and Data Analysis**. *Bioinformatics* 2005, **21**(18):3683-3685.

56. Bolstad, Ben: **Preprocessing and Normalization for Affymetrix GeneChip Expression Microarrays**. *Methods in Microarray Normalization* 41 (2008).

57. Gautier L, Cope L, Bolstad BM, Irizarry RA: **Affy--Analysis of Affymetrix Genechip Data at the Probe Level**. *Bioinformatics* 2004, **20**(3):307-315.

58. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data**. *Biostatistics* 2003, **4**(2):249-264.

59. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix Genechip Probe Level Data**. *Nucleic Acids Research* 2003, **31**(4):e15.

60. Eklund AC, Szallasi Z: **Correction of Technical Bias in Clinical Microarray Data Improves Concordance with Known Biological Information**. *Genome Biology* 2008, **9**(2):R26.

61. Johnson WE, Li C, Rabinovic A: **Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods**. *Biostatistics* 2007, **8**(1):118-127.

62. Kauffmann A, Gentleman R, Huber W: **Arrayqualitymetrics--a Bioconductor Package for Quality Assessment of Microarray Data**. *Bioinformatics* 2009, **25**(3):415-416.

63. Gentleman, R., et al: **Genefilter: Methods for Filtering Genes From Microarray Experiments**. *R package version* 1.0 (2011).

64. Bourgon R, Gentleman R, Huber W: **Independent Filtering Increases Detection Power for High-Throughput Experiments**. *Proceedings of the National Academy of Sciences* 2010, **107**(21):9546-9551.

65. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies**. *Nucleic Acids Research* 2015.

66. Smyth GK: **Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments**. *Statistical Applications in Genetics and Molecular Biology* 2004, **3**(1):1-25.

67. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing**. *Journal of the Royal Statistical Society Series B (Methodological)* 1995:289-300.

68. Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J: **Rankprod: A Bioconductor Package for Detecting Differentially Expressed Genes in Meta-Analysis**. *Bioinformatics* 2006, **22**(22):2825-2827.

69. Breitling R, Armengaud P, Amtmann A, Herzyk P: **Rank Products: A Simple, yet Powerful, New Method to Detect Differentially Regulated Genes in Replicated Microarray Experiments**. *FEBS letters* 2004, **573**(1):83-92.

70. Tarca AL, Kathri P, Draghici S: **SPIA: Signaling Pathway Impact Analysis (SPIA) Using Combined Evidence of Pathway Over-Representation and Unusual Signaling Perturbations**. *Bioinformatics* 2011:1-9.

71. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes**. *Nucleic Acids Research* 2000, **28**(1):27-30.

72. Ellis R, Langford D, Masliah E: **HIV and Antiretroviral Therapy in the Brain: Neuronal Injury and Repair**. *Nature Reviews Neuroscience* 2007, **8**(1):33-44.

73. Vinnakota K, Hu F, Ku M-C, Georgieva PB, Szulzewsky F, Pohlmann A, Waiczies S, Waiczies H, Niendorf T, Lehnardt S: **Toll-Like Receptor 2 Mediates Microglia/Brain Macrophage MT1-MMP Expression and Glioma Expansion**. *Neuro-Oncology* 2013:not115.

74. Markovic D, Vinnakota K, Chirasani S, Synowitz M, Raguet H, Stock K, Sliwa M, Lehmann S, Kälin R, Van Rooijen N: **Gliomas Induce and Exploit Microglial MT1-MMP Expression for Tumor Expansion**. *Proceedings of the National Academy of Sciences* 2009, **106**(30):12530-12535.

75. McArthur JC, Steiner J, Sacktor N, Nath A: **Human Immunodeficiency Virus-Associated Neurocognitive Disorders: Mind the Gap**. *Annals of Neurology* 2010, **67**(6):699-714.

76. Minagar A, Shapshak P, Fujimura R, Ownby R, Heyes M, Eisdorfer C: **The Role of Macrophage/Microglia and Astrocytes in the Pathogenesis of Three Neurologic Disorders: HIV-Associated Dementia, Alzheimer Disease, and Multiple Sclerosis**. *Journal of the Neurological Sciences* 2002, **202**(1):13-23.

77. Levine AJ, Miller JA, Shapshak P, Gelman B, Singer EJ, Hinkin CH, Commins D, Morgello S, Grant I, Horvath S: **Systems Analysis of Human Brain Gene Expression: Mechanisms for HIV-Associated Neurocognitive Impairment and Common Pathways with Alzheimer's Disease**. *BMC Medical Genomics* 2013, **6**(1):4.

78. Yeo M, Lin PS, Dahmus ME, Gill GN: **A Novel RNA Polymerase II C-Terminal Domain Phosphatase that Preferentially Dephosphorylates Serine 5**. *Journal of Biological Chemistry* 2003, **278**(28):26078-26085.

79. Weiske J, Huber O: **The Histidine Triad Protein Hint1 Triggers Apoptosis Independent of its Enzymatic Activity**. *Journal of Biological Chemistry* 2006, **281**(37):27356-27366.

80. Dauphinee SM, Clayton A, Hussainkhel A, Yang C, Park YJ, Fuller ME, Blonder J, Veenstra TD, Karsan A: **SASH1 is a Scaffold Molecule in Endothelial TLR4 Signaling**. *Journal of Immunology* 2013, **191**(2):892-901.

81. Zhang Y, Liu S, Mickanin C, Feng Y, Charlat O, Michaud GA, Schirle M, Shi X, Hild M, Bauer A *et al*: **RNF146 Is a Poly(ADP-Ribose)-Directed E3 Ligase that Regulates Axin Degradation and Wnt Signalling**. *Nature Cell Biology* 2011, **13**(5):623-629.

82. Eguether T, Ermolaeva MA, Zhao Y, Bonnet MC, Jain A, Pasparakis M, Courtois G, Tassin AM: **The Deubiquitinating Enzyme CYLD Controls Apical Docking of Basal Bodies in Ciliated Epithelial Cells**. *Nature Communications* 2014, **5**:4585.

83. Ruvolo PP, Deng X, May WS: **Phosphorylation of Bcl2 and Regulation of Apoptosis**. *Leukemia* 2001, **15**(4):515-522.

84. Shinoura N, Yoshida Y, Nishimura M, Muramatsu Y, Asai A, Kirino T, Hamada H: **Expression Level of Bcl-2 Determines Anti-or Proapoptotic Function**. *Cancer Research* 1999, **59**(16):4119-4128.

85. Strack PR, Frey MW, Rizzo CJ, Cordova B, George HJ, Meade R, Ho SP, Corman J, Tritch R, Korant BD: **Apoptosis Mediated by HIV Protease is Preceded by Cleavage of Bcl-2**. *Proceedings of the National Academy of Sciences* 1996, **93**(18):9571-9576.

86. Koltai T: **Nelfinavir and Other Protease Inhibitors in Cancer: Mechanisms Involved in Anticancer Activity**. *F1000Research* 2015, **4**.

87. Haeffner A, Deas O, Mollereau B, Estaquier J, Mignon A, Haeffner-Cavaillon N, Charpentier B, Senik A, Hirsch F: **Growth Hormone Prevents Human Monocytic Cells from Fas-Mediated Apoptosis by Up-Regulating Bcl-2 Expression**. *European Journal of Immunology* 1999, **29**(1):334-344.

88. Kang C, Bharatham N, Chia J, Mu Y, Baek K, Yoon HS: **The Natively Disordered Loop of Bcl-2 Undergoes Phosphorylation-Dependent Conformational Change and Interacts with Pin1**. *PloS One* 2012, **7**(12):e52047.

89. Ito T, Deng X, Carr B, May WS: **Bcl-2 Phosphorylation Required for Anti-Apoptosis Function**. *Journal of Biological Chemistry* 1997, **272**(18):11671-11673.

90. De Chiara G, Marcocci ME, Torcia M, Lucibello M, Rosini P, Bonini P, Higashimoto Y, Damonte G, Armirotti A, Amodei S: **Bcl-2 Phosphorylation by p38 MAPK Identification of Target Sites and Biologic Consequences**. *Journal of Biological Chemistry* 2006, **281**(30):21353-21361.

91. Toshiyuki M, Reed JC: **Tumor Suppressor p53 Is a Direct Transcriptional Activator of the Human Bax Gene**. *Cell* 1995, **80**(2):293-299.

92. Santin I, Moore F, Grieco F, Marchetti P, Brancolini C, Eizirik DL: **USP18 Is a Key Regulator of the Interferon-Driven Gene Network Modulating Pancreatic Beta Cell Inflammation and Apoptosis**. *Cell Death & Disease* 2012, **3**(11):e419.

93. Ashkenazi A, Dixit VM: **Death Receptors: Signaling and Modulation**. *Science* 1998, **281**(5381):1305-1308.

94. Shinohara H, Yagita H, Ikawa Y, Oyaizu N: **Fas Drives Cell Cycle Progression in Glioma Cells Via Extracellular Signal-Regulated Kinase Activation**. *Cancer Research* 2000, **60**(6):1766-1772.

95. Frankel B, Longo SL, Kyle M, Canute GW, Ryken TC: **Tumor Fas (APO-1/CD95) Up-Regulation Results in Increased Apoptosis and Survival Times for Rats with Intracranial Malignant Gliomas**. *Neurosurgery* 2001, **49**(1):168-175; discussion 175-166.

96. Ambar BB, Frei K, Malipiero U, Morelli AE, Castro MG, Lowenstein PR, Fontana A: **Treatment of Experimental Glioma by Administration of Adenoviral Vectors Expressing Fas Ligand**. *Human Gene Therapy* 1999, **10**(10):1641-1648.

97. Poonia B, Pauza CD, Salvato MS: **Role of the Fas/Fasl Pathway in HIV or SIV Disease**. *Retrovirology* 2009, **6**(1):91.

98. Ikomey GM, Okomo-Assoumou M-C, Atashili J, Mesembe MT, Mukwele B, Lyonga E, Eyoh A, Ndumbe PM: **Plasma Concentrations of Soluble Fas Receptors (Fas) and Fas Ligands (Fasl) in Relation to CD4+ Cell Counts in HIV-1 Positive and Negative Patients in Yaounde, Cameroon**. *BMC Research Notes* 2012, **5**(1):322.

99. Kuribayashi K, Krigsfeld G, Wang W, Xu J, Mayes PA, Dicker DT, Wu GS, El-Deiry WS: **TNFSF10 (TRAIL), a p53 Target Gene That Mediates p53-Dependent Cell Death**. *Cancer Biology & Therapy* 2008, **7**(12):2034-2038.

100. Roth W, Wild-Bode C, Platten M, Grimmel C, Melkonyan HS, Dichgans J, Weller M: **Secreted Frizzled-Related Proteins Inhibit Motility and Promote Growth of Human Malignant Glioma Cells**. *Oncogene* 2000, **19**(37):4210-4220.

101. Ohgaki H, Kleihues P: **The Definition of Primary and Secondary Glioblastoma**. *Clinical Cancer Research* 2013, **19**(4):764-772.

102. Dixit D, Sharma V, Ghosh S, Mehta V, Sen E: **Inhibition of Casein Kinase-2 Induces p53-Dependent Cell Cycle Arrest and Sensitizes Glioblastoma Cells to Tumor Necrosis Factor (TNFα)-Induced Apoptosis through SIRT1 Inhibition**. *Cell Death & Disease* 2012, **3**(2):e271.

103. Weinmann L, Wischhusen J, Demma M, Naumann U, Roth P, Dasmahapatra B, Weller M: **A Novel p53 Rescue Compound Induces p53-Dependent Growth Arrest and Sensitises Glioma Cells to Apo2L/TRAIL-Induced Apoptosis**. *Cell Death & Differentiation* 2008, **15**(4):718-729.

104. Izumi T, Io K, Matsui M, Shirakawa K, Shinohara M, Nagai Y, Kawahara M, Kobayashi M, Kondoh H, Misawa N: **HIV-1 Viral Infectivity Factor Interacts with TP53 to Induce G2 Cell Cycle Arrest and Positively Regulate Viral Replication**. *Proceedings of the National Academy of Sciences* 2010, **107**(48):20798-20803.

105. Longo F, Marchetti MA, Castagnoli L, Battaglia PA, Gigliani F: **A Novel Approach to Protein-Protein Interaction: Complex Formation between the p53 Tumor Suppressor and the HIV Tat Proteins**. *Biochemical and Biophysical Research Communications* 1995, **206**(1):326-334.

106. Ariumi Y, Kaida A, Hatanaka M, Shimotohno K: **Functional Cross-Talk of HIV-1 Tat with p53 through its C-Terminal Domain**. *Biochemical and Biophysical Research Communications* 2001, **287**(2):556-561.

107. Li CJ, Wang C, Friedman DJ, Pardee AB: **Reciprocal Modulations between p53 and Tat of Human Immunodeficiency Virus Type 1**. *Proceedings of the National Academy of Sciences* 1995, **92**(12):5461-5464.

108. Ping Yf, Yao Xh, Jiang Jy, Zhao Lt, Yu Sc, Jiang T, Lin M, Chen Jh, Wang B, Zhang R: **The Chemokine CXCL12 and its Receptor CXCR4 Promote Glioma Stem Cell-Mediated VEGF Production and Tumour Angiogenesis via PI3K/AKT Signalling**. *The Journal of Pathology* 2011, **224**(3):344-354.

109. Jähnichen S, Blanchetot C, Maussang D, Gonzalez-Pajuelo M, Chow KY, Bosch L, De Vrieze S, Serruys B, Ulrichts H, Vandevelde W: **CXCR4 Nanobodies (VHH-Based Single Variable Domains) Potently Inhibit Chemotaxis and HIV-1 Replication and Mobilize Stem Cells**. *Proceedings of the National Academy of Sciences* 2010:201012865.

110. Jablonska J, Wu CF, Andzinski L, Leschner S, Weiss S: **CXCR2-Mediated Tumor-Associated Neutrophil Recruitment is Regulated by IFN-B**. *International Journal of Cancer* 2014, **134**(6):1346-1358.

111. Forsyth P, Wong H, Laing TD, Rewcastle N, Morris D, Muzik H, Leco K, Johnston R, Brasher P, Sutherland G: **Gelatinase-A (MMP-2), Gelatinase-B (MMP-9) and Membrane Type Matrix Metalloproteinase-1 (MT1-MMP) are Involved in Different Aspects of the Pathophysiology of Malignant Gliomas**. *British Journal of Cancer* 1999, **79**(11-12):1828.

112. Sciaccaluga M, D'Alessandro G, Pagani F, Ferrara G, Lopez N, Warr T, Gorello P, Porzia A, Mainiero F, Santoro A: **Functional Cross Talk between CXCR4 and PDGFR on Glioblastoma Cells is Essential for Migration**. *PloS One* 2013, **8**(9):e73426.

113. Johnson K, Hashimshony T, Sawai CM, Pongubala JM, Skok JA, Aifantis I, Singh H: **Regulation of Immunoglobulin Light-Chain Recombination by the Transcription Factor IRF-4 and the Attenuation of Interleukin-7 Signaling**. *Immunity* 2008, **28**(3):335-345.

114. Chen L, Xu S, Zeng X, Li J, Yin W, Chen Y, Shao Z, Jin W: **C-myb Activates CXCL12 Transcription in T47D and MCF7 Breast Cancer Cells**. *Acta Biochimica et Biophysica Sinica* 2009:gmp108.

115. Moskovits N, Kalinkovich A, Bar J, Lapidot T, Oren M: **P53 Attenuates Cancer Cell Migration and Invasion through Repression of SDF-1/CXCL12 Expression in Stromal Fibroblasts**. *Cancer Research* 2006, **66**(22):10671-10676.

116. Höglinger GU, Melhem NM, Dickson DW, Sleiman PM, Wang L-S, Klei L, Rademakers R, de Silva R, Litvan I, Riley DE: **Identification of Common Variants Influencing Risk of the Tauopathy Progressive Supranuclear Palsy**. *Nature Genetics* 2011, **43**(7):699-705.

117. Park MA, Yacoub A, Sarkar D, Emdad L, Rahmani M, Spiegel S, Koumenis C, Graf M, Curiel DT, Grant S: **Perk-Dependent Regulation of MDA-7/IL-24-Induced Autophagy in Primary Human Glioma Cells**. *Autophagy* 2008, **4**(4):513-515.

118. Yacoub A, Hamed HA, Allegood J, Mitchell C, Spiegel S, Lesniak MS, Ogretmen B, Dash R, Sarkar D, Broaddus WC: **PERK–Dependent Regulation of Ceramide Synthase 6 and Thioredoxin Play a Key Role in mda-7/IL-24–Induced Killing of Primary Human Glioblastoma Multiforme Cells**. *Cancer Research* 2010, **70**(3):1120-1129.

119. Lebedeva IV, Su Z-Z, Chang Y, Kitada S, Reed JC, Fisher PB: **The Cancer Growth Suppressing Gene mda-7 Induces Apoptosis Selectively in Human Melanoma Cells**. *Oncogene* 2002, **21**(5):708-718.

120. Gupta AK, Li B, Cerniglia GJ, Ahmed MS, Hahn SM, Maity A: **The HIV Protease Inhibitor Nelfinavir Downregulates Akt Phosphorylation by Inhibiting Proteasomal Activity and Inducing the Unfolded Protein Response**. *Neoplasia* 2007, **9**(4):271-278.

121. Johnson DR: **Rising Incidence of Glioblastoma and Meningioma in the United States: Projections through 2050**. *ASCO Annual Meeting: 2012*: *Journal of Clinical Oncology* 2012: 2065.

122. Ohgaki H, Dessen P, Jourde B, Horstmann S, Nishikawa T, Di Patre P-L, Burkhard C, Schüler D, Probst-Hensch NM, Maiorka PC: **Genetic Pathways to Glioblastoma a Population-Based Study**. *Cancer Research* 2004, **64**(19):6892-6899.

123. Consortium M: **The Microarray Quality Control (MAQC) Project Shows Inter-and Intraplatform Reproducibility of Gene Expression Measurements**. *Nature Biotechnology* 2006, **24**(9):1151.

124. Vert G, Nemhauser JL, Geldner N, Hong F, Chory J: **Molecular Mechanisms of Steroid Hormone Signaling in Plants**. *Annual Review of Cell and Developmental Biology* 2005, **21**:177-201.

125. Gurvich N, Berman MG, Wittner BS, Gentleman RC, Klein PS, Green JB: **Association of Valproate-Induced Teratogenesis with Histone Deacetylase Inhibition in Vivo**. *The FASEB Journal* 2005, **19**(9):1166-1168.

126. Wilson CL, Sims AH, Howell A, Miller CJ, Clarke RB: **Effects of Oestrogen on Gene Expression in Epithelium and Stroma of Normal Human Breast Tissue**. *Endocrine-Related Cancer* 2006, **13**(2):617-628.

CHAPTER 5

CONCLUSION

As we come to a better understanding of the two diseases, infectious disease due to HIV, and chronic disease due to GBM, we begin to appreciate how complex each is in its use of the human body in its own advancement. Both diseases have devastating effects on the brain, largely stemming from chronic inflammation in the brain, and in both, those affected suffer from a greatly reduced quality of life. HIV-associated neurocognitive impairment affects nearly 50% of the HIV population, and increases the risk of death in HIV individuals by several fold [1]. Glioblastoma multiforme is the most common brain tumor in the general population, and following detection of the tumor, GBM patients have a median survival rate of 14 months [2].

5.1 Advancing bioinformatics methodologies in epidemiology

It is essential we continue to advance and promote use of bioinformatics methodologies in epidemiology. There are still great steps, however, that need to be taken as a research community in order to ensure the development of accurate and useful computational tools and algorithms that meet the demands of disease research. This includes the creation and maintenance of public databases toward a sharing of high-quality information between scientists. Equally as important as the sharing of biological data, including genomic, transcriptomic, proteomic, and metabolomic data, is the sharing

of rich clinical and experimental information with which to describe and contextualize these data. Meta-analyses, including the two studies detailed in this dissertation, still suffer from a lack of detailed clinical annotation in their efforts to make sense of the abundance of data that already exist. Good models of public databases that help ensure reproducibility include the Gene Expression Omnibus (GEO), a resource made available through the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo/). Researchers depositing data into GEO are asked to include raw and processed data into the database, along with associated experimental and clinical information. For this reason, for the second study of this dissertation our data collection efforts relied heavily on this resource. A second model of quality data sharing is the Los Alamos National Laboratory HIV Database (http://www.hiv.lanl.gov). The result of collaboration between the Los Alamos National Laboratory (LANL), the National Institutes of Health (NIH), and the Department of Health and Human Services (HHS), this publically available and HIV-specific resource is a great asset to the HIV research community. Not only does this database contain viral sequence data and associated clinical information, but also tools specific to viral sequence analysis, and as a resource we made use of in the first study of this dissertation, we would be remiss in not mentioning this as wonderful example of data sharing in the field of epidemiology. In an effort to support this movement, the database resulting from our first study has been made publically available for continued use by other infectious disease researchers. We anticipate this collection of viral sequence data and associated clinical information will be of great use in increasing HAND research efforts.

A second challenge we face in bioinformatics is a lag in statistical methods to accurately assess rapidly changing data types. In the second dissertation study we demonstrated a sensitivity of results dependent on statistical methodology. We used two statistical tests of differential expression, the more classically employed moderated t-statistic (*limma*), and the meta-analysis methodology of rank products (*RankProd*), and found our pathway results to depend on, both, the group assessed and the differential expression tool used. Use of *limma* in the HIV group resulted in more conservative estimates of significant signaling pathways than use of *RankProd* output, and the opposite effect was observed in our GBM group, with more conservative estimates of significant pathways using *RankProd* output. It is important that we continue to test the limits of current statistical tools and report these limitations back to the scientific community.

Finally, a third challenge lies in our ability to effectively and accurately bridge data types. In our second study, for example, both RNAseq and microarray data were available for patients with glioblastoma multiforme. For HIV individuals, however, these data were limited to only microarray data. While these are both forms of transcriptomic information, these two data types are distinct, and the statistical tools used to test these data equally as distinct. In addition, vast differences in experimental protocols used to generate these data types result in the introduction of biases specific to data platforms and to individual laboratories, a challenge in all meta-analysis research. In an effort to keep these biases to a minimum, we ultimately decided to limit our analyses to a single data type across our sample groups. Development of computational protocol and tools that allow for a more feasible and accurate merging of data types than currently possible will

allow investigators to imagine and test new and exciting questions that transcend technology types and periods. This ability, however, also hinges on the promotion of long-term storage options that allow public access to unprocessed data.

5.2 Research challenges and implications for future work

While a meta-analysis approach allowed us to increase our sampling power, particularly in our study of the HIV brain, in both dissertation projects we encountered and worked through a number of restrictions imposed by this type of approach. First, while use of previously generated data allowed us to test our hypotheses with greater power than typical for an HIV research study, this reliance meant we had no control over experimental methodology. In our second study we took this variable into account by only evaluating transcriptomic data in the form of microarrays and only microarrays produced by a single platform. While this facilitated normalization across datasets, it also limited our sample sizes. In addition, in both studies we faced a lack of clinical information availability, again limiting the complexity of our analyses as well as our extrapolation of results. We were also hampered by prior sampling efforts, and in particular, by breadth of geographical sampling, a hurdle that needs to be addressed particularly in the HIV research community. In particular this would allow us to better understand differences in disease progression attributable to huge disparities in treatment options and treatment availability across countries, and even within a country across low and high-income areas, especially as severe forms of HAND continue to devastate the lives of untreated individuals.

And finally, a lack of longitudinal data makes it difficult to precisely identify links between changes in the innate immune system and disease progression in both GBM development and HIV infection.

5.3 Implications of these studies

In our first study, we were able to achieve three primary goals. The first was development of a high-quality and richly annotated database to help advance research on HIV-associated neurocognitive disorders. As the largest publically available resource of its kind, we hope the HAND database will be of great use to the HIV research community. The second aim we were able to achieve was to increase the power and breadth of prior HIV tissue compartmentalization research to help clarify prior findings. Ultimately, we found tissue environment not to play a large role in viral evolution in general, but to potentially interact with particular regions of the genome independently. Finally, and again with increased sampling power and genomic breadth, we tested for viral genomic marker associated with HIV neurocognitive impairment, and in addition to validating prior markers, found novel markers of this disease. We hope this work will help advance our understanding of viral mechanisms that underlie HAND initiation and progression.

In our second study, we were able to achieve two primary goals. The first was validation of anticipated pathway changes shared between the two diseases, infectious disease, HIV, and chronic disease, GBM, as based on prior research in their respective fields. Second, we found the cancer-related pathway, *glioma cell death*, as activated in HIV brains but not in GBM brains. Our additional finding of differential expression of

upregulator molecules controlling this mechanism offers potential molecular targets in GBM treatment development. Perhaps even more importantly, however, by testing beyond a single disease state we hope to have inspired other researchers to also utilize comparative analyses, and to have provided a rigorous and systematic methodology that will allow them to do so.

REFERENCES

1. Heaton R, Clifford D, Franklin D, Woods S, Ake C, Vaida F, Ellis R, Letendre S, Marcotte T, Atkinson J: **HIV-Associated Neurocognitive Disorders Persist in the Era of Potent Antiretroviral Therapy Charter Study**. *Neurology* 2010, **75**(23):2087-2096.
2. Adamson C, Kanu OO, Mehta AI, Di C, Lin N, Mattox AK, Bigner DD: **Glioblastoma Multiforme: A Review of Where We Have Been and Where We Are Going**. *Expert Opinion on Investigational Drugs* 2009, **18**(8):1061-1083.