Tissue Compartmentalization and Tropism of HIV-1: A Dissertation

Robin L. Brese

University of Massachusetts Medical School

Follow this and additional works at: http://escholarship.umassmed.edu/gsbs_diss

Part of the Immunology and Infectious Disease Commons, Tissues Commons, Virology Commons, and the Virus Diseases Commons

Recommended Citation


This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Dissertations and Theses by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
TISSUE COMPARTMENTALIZATION AND TROPISM OF HIV-1

A Dissertation Presented

By

ROBIN LEIGH BRESE

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 10, 2016

IMMUNOLOGY AND VIROLOGY PROGRAM
TISSUE COMPARTMENTALIZATION AND TROPISM OF HIV-1

A Dissertation Presented By

Robin Leigh Brese

This work was undertaken in the Graduate School of Biomedical Sciences

Immunology and Virology Program

The signature of the Thesis Advisor signifies validation of Dissertation content

__________________________
Paul Clapham, Ph.D., Thesis Advisor

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

__________________________
Heinrich Gottlinger, M.D., Ph.D., Member of Committee

__________________________
Timothy Kowalik, Ph.D., Member of Committee

__________________________
Maria Zapp, Ph.D., Member of Committee

__________________________
Manish Sagar, M.D., External Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

__________________________
Trudy Morrison, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the School.

__________________________
Anthony Carruthers, Ph.D., Dean of the Graduate School of Biomedical Sciences

August 10, 2016
ACKNOWLEDGEMENTS

I would like to thank Paul Clapham and the members of the Clapham Lab. I am especially thankful for the friendship of Olivia O’Connell.

I am sincerely grateful for the members of my thesis and dissertation committees: Heinrich Gottlinger, Tim Kowalik, Trudy Morrison, Mohan Somasundaran, and Maria Zapp. Thank you for all of the mentorship, advice, and scientific discussions over the past few years. I truly appreciate it. I would also like to thank my external committee member, Manish Sagar.

I would like to thank my family, friends, and fellow graduate students for their support.

Thank you to my husband, Joseph Libutti, for everything.
Despite the development of effective antiretroviral treatments, there is still no cure for HIV-1. Major barriers to HIV-1 eradication include the diversity of intrapatient viral quasispecies and the establishment of reservoirs in tissue sanctuary sites. A better understanding of these populations is required for targeted treatments. While previous studies have examined the relationship between brain and blood or immune tissues, few have looked at and compared the properties of viruses from other tissue compartments. In this study, 75 full length HIV-1 envelopes were isolated from the frontal lobe, occipital lobe, parietal lobe, colon, lung, and lymph node of an HIV-1 infected subject. No envelopes could be amplified from the plasma or serum. Envelopes were subjected to genotypic and phenotypic characterization. Of the 75 envelopes, 53 were able to infect HeLa TZM-bl cells. The greatest proportion of non-functional envelopes was from the lung, a result of APOBEC-induced hypermutation. Lower frequencies of hypermutation were also observed in the occipital lobe and colon. Envelopes from regions of the brain were almost all macrophage tropic, while those from the body were predominantly non-macrophage tropic. All envelopes used CCR5 as a coreceptor. Phylogenetic analyses showed that sequences were compartmentalized inside the brain. These findings were also observed using PacBio next generation sequencing to examine 32,152 full length sequences. Envelopes from tissues of the body displayed greater variation in sequence length,
charge, and number of potential N-linked glycosylation sites in comparison to envelopes from tissues of the brain. Increased variation was also observed in IC50s for inhibition and neutralization assays using sCD4, maraviroc, b12, PG16, 17b, and 447-52D. The increased variation observed in envelopes from tissues outside the brain suggests that different pressures may be influencing the evolution of these viruses and emphasizes the importance of further studies in these tissue sites.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ III
ABSTRACT ........................................................................................................................... IV
TABLE OF CONTENTS ......................................................................................................... VI
LIST OF FIGURES ................................................................................................................ X
LIST OF TABLES ................................................................................................................... XIII
LIST OF THIRD PARTY COPYRIGHTED MATERIAL ....................................................... XIV
LIST OF ABBREVIATIONS .................................................................................................. XV

CHAPTER I: INTRODUCTION ................................................................................................. 1

1.1 HIV and AIDS ................................................................................................................. 1

1.1.1 Natural History and Discovery ................................................................................. 1

1.1.2 Transmission and Disease Progression ................................................................. 3

1.1.3 Global Health ........................................................................................................... 4

1.2 Genome and Viral Proteins ......................................................................................... 10

1.2.1 Genome Organization ............................................................................................ 10

1.2.2 Enzymatic Proteins ............................................................................................... 10

1.2.3 Structural Proteins ............................................................................................... 12

1.2.4 Regulatory Proteins .............................................................................................. 14

1.2.5 Accessory Proteins ............................................................................................... 15

1.3 HIV-1 Replication Cycle ............................................................................................... 16

1.3.1 Replication ............................................................................................................. 16
1.3.2 Evolution and Diversity ................................................................. 20
1.4 HIV-1 Envelope ............................................................................. 21
1.5 Tropism ......................................................................................... 22
1.6 Compartmentalization ................................................................. 26
1.7 Scope of Thesis .............................................................................. 28

CHAPTER II: MATERIALS AND METHODS ........................................... 30
2.1 HIV-1 Positive Subject and Tissues .............................................. 30
2.2 Nucleic Acid Extraction ............................................................... 30
2.3 PCR ............................................................................................. 30
2.4 Cloning ........................................................................................ 35
2.5 Sanger Sequencing ...................................................................... 36
2.6 Sequence Analyses ...................................................................... 36
2.7 Phylogenetics .............................................................................. 37
2.8 Cell Culture ................................................................................ 38
2.9 Pseudovirus Production ................................................................ 39
2.10 Viral Titer and Tropism ............................................................. 40
2.11 Inhibition and Neutralization Assays ......................................... 42
2.12 Next Generation Sequencing ..................................................... 43

CHAPTER III: GENETIC CHARACTERIZATION OF HIV-1 ENVELOPES FROM TISSUES OF AN HIV-1 INFECTED SUBJECT ........................................... 46
3.1 Introduction .................................................................................. 46
3.2 Results ........................................................................................ 48

3.2.1 Envelopes were isolated from all tissues except
for plasma and serum ..............................................................48

3.2.2 No mutations were observed that have previously been associated with macrophage tropism or CXCR4 use .......................................................50

3.2.3 Sequences were compartmentalized in brain tissue .....51

3.2.4 Length of envelope regions did not differ significantly between different tissue sites ......................................................53

3.2.5 V1-V2 and V1-V5 net charges had greater variability in tissues from the body than in brain tissues ...................54

3.2.6 Average number of predicted N-linked glycosylation sites was more variable in tissue compartments outside the brain ........................................64

3.2.7 The lung compartment had a higher frequency of hypermutated envelopes than other tissues ...............64

3.2.8 Recombination was extensive in all tissue compartments ..................................................................................65

3.2.9 PacBio sequencing ..................................................................................65

3.3 Discussion ..........................................................................................66

CHAPTER IV: FUNCTIONAL CHARACTERIZATION OF HIV-1 ENVELOPES FROM TISSUES OF AN HIV-1 INFECTED SUBJECT ........79

4.1 Introduction ..........................................................................................79

4.2 Results ..................................................................................................80

4.2.1 Most envelopes were able to infect HeLa TZM-bl cells .................................................................80

4.2.2 All envelopes amplified used CCR5 and could not use CXCR4 ..........................................................85

4.2.3 Brain derived envelopes were macrophage tropic while envelopes from tissues from the body were not ...85
4.2.4 Sensitivity to entry inhibitors and neutralizing monoclonal antibodies .........................................................86

4.3 Discussion .........................................................................................................................................................89

CHAPTER V: GENERAL DISCUSSION...................................................................................................................108

5.1 Summary .......................................................................................................................................................108

5.2 Future Studies ..............................................................................................................................................109

5.3 Conclusions ...............................................................................................................................................111

APPENDIX A: PHYLOGENETIC TREES FOR CHAPTER III.................112

APPENDIX B: ANALYSIS OF AN HIV-1 INFECTED SUBJECT WITH TWO DISTINCT VIRAL POPULATIONS.........................................................127

APPENDIX C: FUNDING SOURCES..................................................................................................................142

REFERENCES ......................................................................................................................................................143
LIST OF FIGURES

Figure 1.1 Global distribution of HIV-1 subtypes and circulating recombinant forms (CRFs) ..........................................................6
Figure 1.2 Course of typical HIV-1 infection ..............................................7
Figure 1.3 Number of HIV cases per region ..............................................8
Figure 1.4 HIV-1 genome and virion structure ........................................11
Figure 1.5 HIV-1 replication cycle .............................................................18
Figure 3.1 Phylogenetic analysis of Subject 162 gp160 sequences ..........52
Figure 3.2 Evolutionary divergence of HIV-1 envs by region ...............55
Figure 3.3 V1-V2 lengths of HIV-1 envs from different tissues ...............56
Figure 3.4 V1-V2 lengths of HIV-1 envs from the brain and body ..........57
Figure 3.5 V1-V5 lengths of HIV-1 envs from different tissues ...............58
Figure 3.6 V1-V5 lengths of HIV-1 envs from the brain and body ..........59
Figure 3.7 V1-V2 net charges of HIV-1 envs from different tissues ..........60
Figure 3.8 V1-V2 net charges of HIV-1 envs from the brain and body ......61
Figure 3.9 V1-V5 net charges of HIV-1 envs from different tissues ..........62
Figure 3.10 V1-V5 net charges of HIV-1 envs from the brain and body ......63
Figure 3.11 V1-V5 potential N-linked glycosylation sites (PNGS) of HIV-1 envs from different tissues ..............................................67
Figure 3.12 V1-V5 potential N-linked glycosylation sites (PNGS) of HIV-1 envs from the brain and body ..............................................68
Figure 3.13 Env potential N-linked glycosylation sites of HIV-1 envs from different tissues ............................................................69
Figure 3.14 Env potential N-linked glycosylation sites of HIV-1 envs from the brain and body.................................70
Figure 3.15 Hypermutation and recombination of gp160 sequences........71
Figure 3.16 Phylogenetic analysis of Subject 162 PacBio sequences.........73
Figure 4.1 TZM-bl and macrophage titers of Envs from the brain............81
Figure 4.2 TZM-bl and macrophage titers of Envs from the colon............82
Figure 4.3 TZM-bl and macrophage titers of Envs from the lung.............83
Figure 4.4 TZM-bl and macrophage titers of Envs from the lymph node.................................................................84
Figure 4.5 sCD4 inhibition of HIV-1 Envs from different tissues ..........90
Figure 4.6 sCD4 inhibition of HIV-1 Envs from the brain and body........91
Figure 4.7 Maraviroc inhibition of HIV-1 Envs from different tissues .......92
Figure 4.8 Maraviroc inhibition of HIV-1 Envs from the brain and body......93
Figure 4.9 b12 neutralization of HIV-1 Envs from different tissues ........94
Figure 4.10 b12 neutralization of HIV-1 Envs from the brain and body ....95
Figure 4.11 PG16 neutralization of HIV-1 Envs from different tissues .......96
Figure 4.12 PG16 neutralization of HIV-1 Envs from the brain and body .....97
Figure 4.13 17b neutralization of HIV-1 Envs from different tissues .........98
Figure 4.14 17b neutralization of HIV-1 Envs from the brain and body .......99
Figure 4.15 447-52D neutralization of HIV-1 Envs from different tissues .....100
Figure 4.16 447-52D neutralization of HIV-1 Envs from the brain and body..................................................................101
Figure A.1 Phylogenetic analysis of env gp120 region..........................113
Figure A.2 Phylogenetic analysis of env gp41 region...........................114
Figure A.3 Phylogenetic analysis of env V1-V2 region ........................................ 115
Figure A.4 Phylogenetic analysis of env V1-V5 region ........................................ 116
Figure A.5 Phylogenetic analysis of env C1 region ............................................. 117
Figure A.6 Phylogenetic analysis of env C2 region ............................................. 118
Figure A.7 Phylogenetic analysis of env C3 region ............................................. 119
Figure A.8 Phylogenetic analysis of env C4 region ............................................. 120
Figure A.9 Phylogenetic analysis of env C5 region ............................................. 121
Figure A.10 Phylogenetic analysis of env V1 region ........................................... 122
Figure A.11 Phylogenetic analysis of env V2 region ........................................... 123
Figure A.12 Phylogenetic analysis of env V3 region ........................................... 124
Figure A.13 Phylogenetic analysis of env V4 region ........................................... 125
Figure A.14 Phylogenetic analysis of env V5 region ........................................... 126
Figure A.15 Phylogenetic analysis of Subject 123 gp160 sequences .................... 132
Figure A.16 Alignment of Subject 123 V3 loop sequences .................................. 133
Figure A.17 Phylogenetic analysis of Subject 123 PacBio sequences ................. 139
# LIST OF TABLES

Table 2.1 Subject 162 data..........................................................................................31
Table 2.2 HIV-1 env primers......................................................................................32
Table 2.3 Thermocycler programs used for PCR .......................................................34
Table 2.4 Inhibitors and antibodies...........................................................................45
Table 3.1 Subject 162 HIV-1 envelopes isolated from different tissues............49
Table 3.2 Subject 162 PacBio sequencing results.................................................72
Table A.1 Subject 123 data .......................................................................................130
Table A.2 Subject 123 HIV-1 envelopes isolated from different tissues ......131
Table A.3 Predicted coreceptor use for Subject 123 Envs .................................135
Table A.4 Short Tandem Repeat (STR) Analysis ............................................137
Table A.5 Subject 123 PacBio sequencing results ..............................................138
**LIST OF THIRD PARTY COPYRIGHTED MATERIAL**

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Publisher</th>
<th>License Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Springer</td>
<td>3897640362491</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Elsevier</td>
<td>3897340639413</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>Nature Publishing Group</td>
<td>3897341478479</td>
</tr>
</tbody>
</table>

The following figures were reproduced from journals: No permission required

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.3</td>
<td>AVERT</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Annual Reviews</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

AIDS Acquired Immune Deficiency Syndrome
amfAR The Foundation for AIDS Research
APOBEC apolipoprotein B mRNA editing enzyme catalytic polypeptide-like
AZT azidothymidine
BBB blood brain barrier
BC blood cells
bnAb broadly neutralizing antibody
bp base pairs
C colon
CA capsid
CCR5 C-C chemokine receptor type 5
CD4bs CD4 binding site
CDC Centers for Disease Control
CRF circulating recombinant form
CSF cerebrospinal fluid
CXCR4 C-X-C chemokine receptor type 4
DMEM Dulbecco’s Modified Eagle Medium
DNA deoxyribonucleic acid
EMBOSS European Molecular Biology Open Software Suite
Env envelope protein
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>env</td>
<td>envelope gene</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>ESCRT</td>
<td>endosomal sorting complex required for transport</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>FFU</td>
<td>focus forming units</td>
</tr>
<tr>
<td>FL</td>
<td>frontal lobe</td>
</tr>
<tr>
<td>Gag</td>
<td>group specific antigen</td>
</tr>
<tr>
<td>GALT</td>
<td>gut-associated lymphoid tissue</td>
</tr>
<tr>
<td>gp120</td>
<td>envelope surface glycoprotein</td>
</tr>
<tr>
<td>gp160</td>
<td>envelope polyprotein</td>
</tr>
<tr>
<td>gp41</td>
<td>envelope transmembrane glycoprotein</td>
</tr>
<tr>
<td>GPCR</td>
<td>G protein–coupled receptors</td>
</tr>
<tr>
<td>HAD</td>
<td>HIV-Associated Dementia</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>HR1</td>
<td>heptad repeat 1</td>
</tr>
<tr>
<td>HR2</td>
<td>heptad repeat 2</td>
</tr>
<tr>
<td>HTLV-III</td>
<td>Human T-Lymphotrophic Virus type III</td>
</tr>
<tr>
<td>IC50</td>
<td>concentration of inhibitor or antibody at which infection is reduced by 50%</td>
</tr>
<tr>
<td>IN</td>
<td>integrase</td>
</tr>
<tr>
<td>kb</td>
<td>kilobases</td>
</tr>
<tr>
<td>kD</td>
<td>kilodalton</td>
</tr>
<tr>
<td>L</td>
<td>lung</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>LANL</td>
<td>Los Alamos National Laboratory</td>
</tr>
<tr>
<td>LAV</td>
<td>Lymphadenopathy-Associated Virus</td>
</tr>
<tr>
<td>LN</td>
<td>lymph node</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>MA</td>
<td>matrix</td>
</tr>
<tr>
<td>MEGA</td>
<td>Molecular Evolutionary Genetics Analysis</td>
</tr>
<tr>
<td>MPER</td>
<td>membrane proximal external region</td>
</tr>
<tr>
<td>MSM</td>
<td>men who have sex with men</td>
</tr>
<tr>
<td>M-Tropic</td>
<td>macrophage tropic</td>
</tr>
<tr>
<td>NC</td>
<td>nucleocapsid</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NDRI</td>
<td>National Disease Research Interchange</td>
</tr>
<tr>
<td>Nef</td>
<td>negative factor</td>
</tr>
<tr>
<td>NIH</td>
<td>National Institutes of Health</td>
</tr>
<tr>
<td>NNRTI</td>
<td>non-nucleoside reverse transcriptase inhibitor</td>
</tr>
<tr>
<td>NRTI</td>
<td>nucleoside reverse transcriptase inhibitor</td>
</tr>
<tr>
<td>NSI</td>
<td>non-syncytia inducing</td>
</tr>
<tr>
<td>OL</td>
<td>occipital lobe</td>
</tr>
<tr>
<td>PacBio</td>
<td>Pacific Biosciences</td>
</tr>
<tr>
<td>PBMC</td>
<td>peripheral blood mononuclear cell</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate-buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>PIC</td>
<td>pre-integration complex</td>
</tr>
<tr>
<td>PL</td>
<td>parietal lobe</td>
</tr>
<tr>
<td>PNGS</td>
<td>potential N-linked glycosylation sites</td>
</tr>
<tr>
<td>PR</td>
<td>protease</td>
</tr>
<tr>
<td>R5</td>
<td>CCR5</td>
</tr>
<tr>
<td>Rev</td>
<td>regulator of virion expression</td>
</tr>
<tr>
<td>RLU</td>
<td>relative light units</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RRE</td>
<td>Rev response element</td>
</tr>
<tr>
<td>RT</td>
<td>reverse transcriptase</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>sCD4</td>
<td>soluble CD4</td>
</tr>
<tr>
<td>SERINC</td>
<td>serine incorporator</td>
</tr>
<tr>
<td>SI</td>
<td>syncytia inducing</td>
</tr>
<tr>
<td>SIV</td>
<td>Simian Immunodeficiency Virus</td>
</tr>
<tr>
<td>SP1</td>
<td>spacer peptide 1</td>
</tr>
<tr>
<td>SP2</td>
<td>spacer peptide 2</td>
</tr>
<tr>
<td>STR</td>
<td>Short Tandem Repeat</td>
</tr>
<tr>
<td>TAR</td>
<td>trans-activation response region</td>
</tr>
<tr>
<td>Tat</td>
<td>trans-activator of transcription</td>
</tr>
<tr>
<td>T-Tropic</td>
<td>T cell tropic</td>
</tr>
<tr>
<td>UNAIDS</td>
<td>The Joint United Nations Programme on HIV/AIDS</td>
</tr>
</tbody>
</table>
Vif  viral infectivity factor
Vpr  viral protein r
Vpu  viral protein u
VSVG Vesicular Stomatitis Virus G protein
WHO World Health Organization
X4 CXCR4
CHAPTER I: INTRODUCTION

1.1 HIV and AIDS

1.1.1 Natural History and Discovery

On June 5, 1981, the Centers for Disease Control and Prevention (CDC) published a report documenting *Pneumocystis carinii* pneumonia in five healthy, young men who were homosexuals (CDC, 1981; Gottlieb et al, 1981). Doctors discovered that these patients also suffered from cytomegalovirus and mucosal candidiasis infections. This was surprising given that these types of opportunistic infections were usually limited to individuals with known immunocompromising diseases. Soon additional cases were described and doctors observed that these individuals had impaired T cell responses (Siegal et al, 1981). Initial reports concluded that this was an epidemic of the gay community, however similar patterns of opportunistic infections and acquired immunodeficiency emerged in other populations, including heterosexual men, women, and children (Masur et al, 1981; CDC, 1982a; CDC, 1982b; CDC, 1982c; CDC, 1982d; CDC, 1983). On September 24, 1982 the CDC called the disease Acquired Immune Deficiency Syndrome (AIDS) (CDC, 1982e). This was the start of a public health crisis and researchers began searching for the etiological agent responsible for the disease.

Luc Montagnier’s group from the Pasteur Institute was the first to publish a paper describing the Lymphadenopathy-Associated Virus (LAV) as the probable cause of AIDS (Barre-Sinoussi et al, 1983). This was followed by papers from
Robert Gallo and colleagues at the National Cancer Institute describing the isolation of Human T-Lymphotrophic Virus type III (HTLV-III) from patients with AIDS (Gallo et al, 1984; Gelmann et al, 1984; Popovic et al, 1984). Soon it was determined that they, as well as many other groups, were describing the same virus, which became known as the Human Immunodeficiency Virus (HIV) (Coffin et al, 1986a; Coffin et al, 1986b).

HIV is a lentivirus and a member of the Retroviridae family. Its origins can be traced back to West Central Africa where two viruses, HIV-1 and HIV-2, arose from separate zoonotic events. HIV-1 is found throughout the world and is responsible for the global HIV pandemic, while HIV-2 is endemic only to regions of West Africa. Using phylogenetic analyses and archived tissue and blood samples, the origins of HIV-1 have been mapped to Kinshasa, formerly Leopoldville, in the Democratic Republic of the Congo in the early part of the 20th century, no later than 1933 (Zhu et al, 1998; Korber et al, 2000; Worobey et al, 2008). From there HIV-1 strains spread across the globe. By the 1960s HIV-1 had arrived in the United States, though it would take almost twenty more years to be discovered (Gilbert et al, 2007).

HIV-1 is divided into four groups, which resulted from separate transmission events: M (Major or Main), N (New or non-M, non-O), O (Outlier), and P (Pending) (De Leys et al, 1990; Gurtler et al, 1994; Simon et al, 1998; Plantier et al, 2009; Vallari et al, 2010; Vallari et al, 2011). Genetic analyses have determined that HIV-1 group M and N viruses originated from Simian Immunodeficiency Virus (SIV) in
chimpanzees (Huet et al, 1990; Gao et al, 1999; Corbet et al, 2000; Keele et al, 2006). HIV-1 group P strains are most closely related to gorilla SIV, while the origin of group HIV-1 group O strains are unclear and may have come from chimpanzees or gorillas (Van Heuverswyn et al, 2006; Plantier et al, 2009; Takehisa et al, 2009; D'Arc et al, 2015). HIV-2 strains are genetically most similar to SIV from Sooty Mangabeys (Hirsch et al, 1989; Gao et al, 1992).

HIV-1 Group M viruses are responsible for the majority of HIV infections in the world and are grouped into nine subtypes, or clades: A, B, C, D, F, G, H, J, and K. Each clade is genetically distinct and dominant in specific geographic regions (Figure 1.1). In areas where more than one clade is present, viruses can recombine to create circulating recombinant forms (CRFs).

1.1.2 Transmission and Disease Progression

HIV-1 infection is acquired through three major modes of transmission: sexual contact, vertical transmission, or intravenous drug use. In the United States, approximately 63% of new infections, and 78% of new infections in men, are the result of men who have sex with men (MSM) (CDC, 2016). In contrast, 70% of new infections globally occur in sub-Saharan Africa, where the majority of infections arise from heterosexual sex (WHO, 2016). In about 80% of the transmissions, infection is established by a single viral variant. However, in the remaining cases two or more variants are transmitted (Keele et al, 2008). Once established, infection proceeds through acute, chronic, and late-stage phases (Figure 1.2).
During acute infection, the HIV-1 infected individual may experience flu-like symptoms. There is an initial spike in plasma HIV-1 RNA copies as the virus infects CD4+ T cells at infection sites and in immune tissues (Zhang et al, 1999). Virus then spreads to gut-associated lymphoid tissue (GALT) where large numbers of CD4+ T cells reside, become infected, and die (Veazey et al, 1998; Douek et al, 2003; Guadalupe et al, 2003; Brenchley et al, 2004; Li et al, 2005). Viral load drops to a stable setpoint during the chronic phase. At this time, the infected individual is asymptomatic and the virus remains latent. CD4+ T cell counts may briefly rebound, but eventually decline over a period of several years until the count is below 200 cells/mm$^3$. At this point the infected individual has entered the late-stage phase and is said to have Acquired Immune Deficiency Syndrome (AIDS). Plasma viremia will increase and the individual will become unable to mount an immune response to opportunistic infections and cancers, such as *Pneumocystis carinii* pneumonia, Kaposi’s Sarcoma, and mucosal candidiasis, which will ultimately lead to death.

### 1.1.3 Global Health

The World Health Organization (WHO) reported that in 2014 there were approximately 37 million people living with HIV globally, with 2 million new infections that year (WHO, 2016). The Foundation for AIDS Research (amfAR) estimates that an additional 5,753 people contract HIV each day, around 240 every hour, with those in developing countries and members of minority populations
disproportionally affected (Figure 1.3) (amfAR, 2016; WHO, 2016). The impact of the HIV epidemic has most severely affected sub-Saharan Africa and infections in this region account for the majority of new infections globally (UNAIDS, 2016). Despite this, there have been dramatic improvements in HIV rates with new infections declining 35% since 2000 and HIV-related deaths declining 42% since a peak in 2004 (UNAIDS, 2016).

In efforts to continue towards the goal of ending the HIV/AIDS epidemic, the Joint United Nations Programme on HIV/AIDS (UNAIDS) has proposed an ambitious 90-90-90 plan with three objectives to be accomplished by the year 2020: 1) 90% of infected people will know their HIV positive status, 2) 90% of infected people will receive antiretroviral treatment, and 3) 90% of those on treatment will have viral suppression (UNAIDS, 2016). Presently, it is estimated that 54% of those who are infected do not know they are infected and only 41% of those infected are receiving antiretroviral treatment (amfAR, 2016; WHO, 2016).

Antiretroviral treatments were established soon after the onset of the HIV crisis with monotherapies as the first treatments. The first drug approved by the Food and Drug Administration (FDA) was azidothymidine (AZT) on March 19, 1987 (FDA, 2016a). Others soon followed. However monotherapies did not perform well, as rapidly replicating HIV-1 was able to acquire resistance mutations and drug failure and viral rebound would follow. In the mid-1990s, new reverse transcriptase inhibitors and protease inhibitors were discovered and for the first time
Figure 1.1 Global distribution of HIV-1 subtypes and circulating recombinant forms (CRFs) While subtype B is predominant in developed countries, subtype C and CRFs are more prevalent in developing countries and are responsible for the largest number of HIV-1 infections. (Reproduced from Tyor et al, 2013 with permission)
Figure 1.2 Course of typical HIV-1 infection During the acute phase of HIV-1 infection, there is an initial spike in viral RNA copies in the plasma and a decline in CD4+ T cells, which may be accompanied by flu-like symptoms. This is followed by the chronic phase, where the infected individual is asymptomatic, viral load drops to a setpoint, and the CD4+ T cell count may be stable for a long period before it begins to decline. Late stage infection, or AIDS, is defined by a CD4+ T cell count that has dropped below 200 cells/mm$^3$. This is when opportunistic infections occur, resulting in a rapid decline in health and eventually death. (Reproduced from An and Winkler, 2010 with permission)
Figure 1.3 Number of HIV cases per region: Sub-Saharan Africa accounts for the majority of HIV infections, with approximately 70% of the total global infections. (Reproduced from AVERT, 2016 with no permission required)
administered in combination, which greatly improved treatment outcomes (Collier et al, 1996; D’Aquila et al, 1996; Staszewski et al, 1996). Mathematical modeling was used to determine that three drugs in combination would provide sufficient protection against drug resistance (Frost and McLean, 1994; Coffin 1995; Nowak et al, 1997; Stengel, 2008). Currently, the FDA has 39 medications approved for the treatment of HIV (FDA, 2016b; NIH, 2016). These include Nucleoside Reverse Transcriptase Inhibitors (NRTIs), Non-nucleoside Reverse Transcriptase Inhibitors (NNRTIs), Integrase Inhibitors, Protease Inhibitors, Entry Inhibitors, Fusion Inhibitors, Pharmacokinetic Enhancers, and Combination Drugs (FDA, 2016b; NIH, 2016).

Antiretroviral treatments have improved the health and extended the lifespans of HIV-1 infected individuals. However a vaccine is needed to eradicate the virus and thus vaccine development has become a priority of global HIV research. However, vaccine design has been challenging and traditional methods have not worked (Kwong et al, 2002; Wei et al, 2003; Burton et al, 2004; Lu, 2006). There have been six HIV-1 vaccine phase IIb/III clinical trials, with all but one having no efficacy (Schiffner et al, 2013). Only RV144, known as the Thai Trial, experienced a modest 31.2% efficacy using a two vaccine prime boost model using the ALVAC HIV and AIDSVAX B/E vaccines (Rerks-Ngarm et al, 2009). A successful vaccine will likely require both T cell responses and the generation of broadly neutralizing antibodies (bnAbs). Progress has been hampered by several factors, including the highly glycosylated structure of the HIV-1 envelope protein,
which shields it from antibodies, the high diversity of HIV-1, which contributes to the emergence of escape mutants, and the fact that effective HIV-1 immunogens have not been established.

1.2 Genome and Viral Proteins

1.2.1 Genome Organization

The HIV-1 genome is approximately 9.7 kb and consists of a positive sense single stranded RNA genome which is packaged as a dimer within the virion. (Figure 1.4). The integrated proviral DNA genome is flanked by 5’ and 3’ long terminal repeats (LTRs), which contain sites important for regulating viral transcription and the replication cycle.

1.2.2 Enzymatic Proteins

The polymerase gene (pol) encodes the three enzymatic proteins of HIV-1: protease (PR), reverse transcriptase (RT), and integrase (IN). Pol is synthesized along with Gag as Pr160\textsubscript{gag-pol} polyprotein, resulting from a ribosomal frameshift, which is then cleaved by protease (Jacks et al, 1988).

Protease (PR)

PR is responsible for the cleavage of HIV-1 polyproteins Gag-Pol (Pr160\textsubscript{gag-pol}), Gag (Pr55\textsubscript{Gag}), Pol, and Env (gp160) into their constituent proteins.
Figure 1.4 HIV-1 genome and virion structure Genome diagram depicts the organization of the HIV-1 genome with polyproteins and cleavage products noted. The locations of HIV-1 proteins are shown on the virion structure illustration. (Reproduced from Frankel and Young, 1998 with no permission required)
Reverse transcriptase (RT)

Retroviral reverse transcriptases function to convert the viral RNA template into DNA (Baltimore, 1970; Mizutani et al, 1970). In HIV-1 RT accomplishes this using RNA-dependent DNA polymerase activity, ribonuclease H, and DNA-dependent DNA polymerase activity. RT has a high error rate around $3 \times 10^{-5}$, which contributes to the diversity of HIV-1 (Preston et al, 1988; Mansky and Temin, 1995).

Integrase (IN)

Integrase is involved with mediating the incorporation of HIV-1 DNA into the host cell genome by facilitating the formation of the pre-integration complex (PIC) and insertion into the target chromosomal DNA (Bushman et al, 1990).

1.2.3 Structural Proteins

HIV-1 structural proteins include group-specific antigen (Gag) and envelope (Env) polyproteins and the proteins resulting from their cleavage by PR.

Group-specific antigen (Gag)

Gag is produced as polyprotein Pr55$^{\text{Gag}}$, which is cleaved into p17 matrix (MA), p24 capsid (CA), p7 nucleocapsid (NC), p6, and spacer peptides SP1 and SP2 by protease during maturation (Gottlinger et al, 1989). Gag is membrane associated in part via an N-terminal myristylation signal and recruits viral RNA and proteins to facilitate viral budding.
Matrix (MA)

MA is located at the N-terminal region of Pr55\textsuperscript{Gag} and plays a role in incorporation of Env into virions (Dorfman et al, 1994a; Freed and Martin, 1995b; Freed, 1998; Kiernan et al, 1998). Following cleavage by protease, it remains associated with the plasma membrane in the virus particle.

Capsid (CA)

CA proteins surround the viral core in the virion, protecting the nucleocapsid, and are involved with assembly during the replication cycle (Dorfman et al, 1994b; Franke et al, 1994; Thali et al, 1994; Ganser-Pornillos et al, 2007). CA is involved in Gag oligomerization and plays a role in the uncoating process (Gamble et al, 1996; Luban, 1996; Gamble et al, 1997).

Nucleocapsid (NC)

NC is associated with the genomic RNA dimer and is flanked by SP1 and SP2 in the Gag and Gag-Pol precursors, which allow proper processing. An important function of NC is facilitating the packaging of viral RNA into new virus particles by recognizing the packaging signal (Poznansky et al, 1991; Harrison and Lever, 1992). NC also mediates reverse transcription (Lapadat-Tapolsky et al, 1993; Levin et al, 2005).

p6

p6 facilitates incorporation of Vpr into the virion and release of virus particles from the plasma membrane through interactions with the host ESCRT pathway.

**Envelope (Env)**

Env is synthesized as polyprotein gp160, which is then cleaved by host furin and furin-like proteases into the gp120 surface glycoprotein and gp41 transmembrane domain. Env interacts with the CD4 receptor and a coreceptor, typically CCR5 or CXCR4, to initiate membrane fusion and host cell infection (Maddon et al, 1986; McDougal et al, 1986; Deng et al, 1996; Feng et al, 1996; Berger et al, 1998).

### 1.2.4 Regulatory Proteins

Two proteins play essential regulatory roles in HIV-1 replication, trans-activator protein (Tat) and regulator of virion expression (Rev).

**Trans-activator protein (Tat)**

Tat is an RNA binding protein that is required for replication (Feng and Holland, 1988; Ruben et al, 1989; Roy et al, 1990). It enhances HIV-1 transcription through binding the trans-activation response region (TAR) on RNA stem-loops at the start of HIV-1 transcripts and promotes elongation of full length transcripts (Dayton et al, 1986; Fisher et al, 1986; Kao et al, 1987; Feinberg et al, 1991).

**Regulator of virion expression (Rev)**

Rev regulates HIV-1 protein expression by binding the Rev response element (RRE) on viral mRNAs and exporting them out of the nucleus (Felber et al, 1989;

1.2.5 Accessory Proteins

HIV-1 has four accessory proteins: Vif, Vpu, Vpr, and Nef. These proteins were originally labeled as “accessory” because they are not required for infection in many cell lines. However, they have now been found to play important roles in HIV-1 infection in vivo.

Virus infectivity factor (Vif)


Viral protein U (Vpu)

Vpu is a membrane protein that enhances virion release from the host cell surface (Cohen et al, 1988; Strebel et al, 1988, Klimkait et al, 1990; Gottlinger et al, 1993). This is accomplished by degrading CD4 molecules, which trap Env proteins in the
ER, and by counteracting host cell tetherin/BST-2, which retains budding virions on the cell surface (Willey et al, 1992; Neil et al, 2008; Van Damme et al, 2008).

**Viral protein R (Vpr)**

Vpr is a virion-associated protein that aids in the nuclear import of HIV-1 preintegration complexes (Cohen et al, 1990; Heinzinger et al, 1994). It is also involved with cell cycle arrest and replication in non-dividing cells (Jowett et al, 1995; Rogel et al, 1995; Cohen et al, 1996).

**Negative factor (Nef)**

Nef is an HIV-1 gene that is expressed early and enhances viral infectivity (Kim et al, 1989; Kestler et al, 1991; Miller et al 1994). Nef decreases immune cell recognition of infected cells by downregulating MHC class I molecules along with CD4, which also enhances Env incorporation and viral budding (Garcia and Miller, 1992; Aiken et al, 1994; Schwartz et al, 1996). In addition, the Gottlinger and Pizzato labs recently discovered that Nef prevents serine incorporator 3 and 5 (serinc3 and serinc5) from being included in the HIV-1 virion and that this correlates with enhanced infectivity (Rosa et al, 2015; Usami et al, 2015).

### 1.3 HIV-1 Replication Cycle

#### 1.3.1 Replication

The HIV-1 replication cycle (Figure 1.5) can be divided into early and late phases. During the first steps of the early phase, HIV-1 enters host cells through receptor binding and membrane fusion, though there is some evidence that
endocytosis may also play a role (Miyauchi et al, 2009). The Env polyprotein gp160 is post translationally cleaved to surface protein gp120 and transmembrane protein gp41, which are non-covalently linked and are arranged in trimers on the viral surface (McCune et al, 1988; Lu et al, 1995). When a virus encounters a permissive cell, the Env trimers bind the host cell CD4 receptor through the CD4 binding site on gp120 (Maddon et al, 1986; McDougal et al, 1986; Kwong et al, 1998). Following engagement with the CD4 receptor, the Env undergoes a conformational change which allows interaction with a coreceptor (Kwong et al, 1998; Chen et al, 2005). C-C chemokine receptor type 5 (CCR5) and C-X-C chemokine receptor type 4 (CXCR4) are G protein–coupled receptors (GPCRs) and are the two main coreceptors for HIV-1 infection (Berger et al, 1998). They are expressed on HIV-1 target CD4+ T cells, macrophages, and dendritic cells. The Env then experiences another conformational change and a fusion pore begins to form (Markosyan et al, 2003). During this conformational change, N-terminal heptad repeat 1 (HR1) and C-terminal heptad repeat 2 (HR2) of gp41 interact forming a stable six helix bundle structure (Chan et al, 1997; Weissenhorn et al, 1997; Melikyan et al, 2000). Membrane fusion proceeds, allowing the viral core to enter the cytoplasm of the host cell.

Once inside the host cell, the viral core goes through an uncoating process to expose a reverse transcription complex (RTC) consisting of viral RNA, MA, CA, NC, RT, IN, and Vpr. HIV-1 genomic RNA is then bound by tRNA primer Lys3 at a primer binding site approximately 180 nucleotides from the 5’ end to initiate
HIV-1 binds to the host cell CD4 receptor and a coreceptor, usually CCR5 or CXCR4. Following membrane fusion, the viral core enters the cell and an uncoating process occurs. Reverse transcriptase synthesizes cDNA from viral RNA, which then becomes part of a pre-integration complex that is shuttled into the nucleus. Viral cDNA is integrated into the host cell DNA and then gives rise to HIV-1 proteins from unspliced genomic RNA, singly spliced mRNAs (ssRNA), and fully spliced mRNAs. These proteins, along with HIV genomes, are assembled and packaged into new virions, which bud off from the plasma membrane. Finally, polyproteins are cleaved by protease during the maturation process. (Reproduced from Peterlin and Trono, 2003 with permission)
DNA synthesis (Robert et al, 1990). An RNA-DNA hybrid molecule is formed, providing a substrate for RNase H, which then degrades the RNA strand. Finally, a second strand of DNA is synthesized via the RT DNA-dependent polymerase activity.

After a cDNA copy of the viral genome has been made, the pre-integration complex (PIC) is formed and is imported into the nucleus (Bowerman et al, 1989). IN facilitates the incorporation of viral DNA into the host’s genome through DNA cutting, joining, and repair mechanisms (Brown et al, 1987; Fujiwara and Mizuuchi, 1988; Brown et al, 1989). The integrated HIV-1 genome, called a provirus, is now a permanent part of the host cell’s genome and can be replicated using host cell machinery.

The late phase of replication begins after integration and includes the expression of viral genes, viral budding, and maturation. Viral RNAs are exported to the cytoplasm, proteins are translated and trafficked to the plasma membrane where they are then assembled, and virus particles bud off from the cell. This process is completed using the endosomal sorting complex required for transport (ESCRT) pathway of the host cell, which is recruited by the p6 domain of the Gag polyprotein. After budding occurs, the Gag polyprotein is cleaved by PR during the maturation process.
1.3.2 Evolution and Diversity

HIV-1 has extraordinary genetic diversity based on several aspects of its replication cycle. HIV-1 RT lacks proofreading mechanisms and has an error rate of $3 \times 10^{-5}$ (Preston et al, 1988; Mansky and Termin, 1995). This allows for mutations to arise in every replication cycle. During replication, recombination also occurs allowing for more genetic diversity (Jetzt et al, 2000; An and Telesnitsky, 2002; Rhodes et al, 2003; Charpentier et al, 2006; Chen et al, 2006; Mild et al, 2007; Zhang et al, 2010; Brown et al, 2011; Immonen et al, 2015; Sanborn et al, 2015). Considering that an infected cell can produce between $3.5 \times 10^3$ and $2.4 \times 10^5$ virus particles during the course of infection, it is clear how quickly genetic changes can accumulate (Ho et al, 1995; Wei et al, 1995; Althaus et al, 2014). Other factors shaping viral evolution include pressures from neutralizing antibodies, host restriction factors, antiretroviral drug treatment, and coinfections.

In a recent study examining full length genomic sequences, it was estimated that average nucleotide diversity was approximately 50% between HIV-1 and HIV-2, 37.5% between HIV-1 groups, 14.7% between HIV-1 subtypes, 8.2% within a single subtype, and less than 1% within an individual (Li et al, 2015). However, these numbers could be lower or higher in specific cases. For example, the env gene is particularly polymorphic and viruses within a single individual can change over 10% of their env sequence during an HIV-1 infection (Wolinsky et al, 1996; Shankarappa et al, 1999). The result of this genetic diversity is that a single HIV-1 positive individual is infected by populations of related but genetically distinct viral
variants called quasispecies (Goodenow et al, 1989). These quasispecies can have different properties, such as coreceptor use and cell tropism, discussed further below.

1.4 HIV-1 Envelope

The Env glycoprotein is synthesized on the rough endoplasmic reticulum (RER) from a bicistronic mRNA encoding both Env and Vpu (Hunter and Swanstrom, 1990; Freed and Martin, 1995a). Env gp160 polyprotein is targeted to the RER membrane by an N-terminal signal sequence, which is cotranslationally cleaved. Env then remains anchored to the membrane by a hydrophobic sequence in the gp41 domain (Berman et al, 1988; Haffar et al, 1988). Glycosylation occurs during translation, adding predominately N-linked oligosaccharides, but some O-linked oligosaccharides as well (Leonard et al, 1990; Bernstein et al, 1994). The gp160 polyproteins oligomerize into trimers and traffic to the Golgi where they are cleaved by furin and furin-like proteases into gp120 and gp41 proteins at a conserved K/R-X-K/R-R sequence (McCune et al, 1988; Freed et al, 1989; Hallenberger et al, 1992; Sen et al, 2007). The gp120 and gp41 proteins, associated by non-covalent interactions, are then trafficked to the plasma membrane where they are incorporated into virions and are essential for infectivity. Trimer conformation dynamics are not well understood, but trimers may adopt more open or closed states, or oscillate between these two, which could affect

The gp120 glycoprotein is composed of five regions of variable sequences (V1-V5) and five regions of comparatively constant sequences (C1-C5) (Starcich et al, 1986; Willey et al, 1986). Amino acids in different constant regions of gp120 are responsible for CD4 binding (Kowalski et al, 1987; Lasky et al, 1987; Olshevsky et al, 1990). Crystal structures have shown CD4 making contact with gp120 at 26 different residues (Kwong et al, 1998). The V3 loop region of gp120 confers coreceptor specificity (Hwang et al, 1991; Shioda et al, 1991; Cann et al, 1992).

The gp41 transmembrane glycoprotein plays an important role in mediating fusion of the viral and host cell membranes to allow the viral core to enter the host cell. It contains an extracellular domain with an N-terminal fusion peptide, two heptad repeat regions (HR1 and HR2), and a membrane proximal external region (MPER) (Bosch et al, 1989; Freed et al, 1990; Muñoz-Barroso et al, 1999; Salzwedel et al, 1999). In addition, gp41 has a transmembrane domain, which anchors Env in the membrane, and a long C-terminal cytoplasmic tail.

1.5 Tropism

Initial efforts to classify HIV-1 isolates by tropism were based on the ability of variants to induce syncytia in T cell lines, such as MT-2 cells. Syncytia inducing (SI) variants were defined as T cell tropic (T-tropic) and the non-syncytia inducing (NSI) variants were defined as macrophage tropic (M-tropic) (Gartner et al, 1986;
Tersmette et al, 1988). Studies aimed at identifying potential coreceptors on permissive cell types found that the chemokines RANTES, MIP-1 alpha, and MIP-1 beta could inhibit HIV-1 infection (Cocchi et al, 1995). This gave researchers clues that lead to the discovery of the coreceptors. The CXCR4 receptor, originally called LESTR or fusin, was the first to be identified and was defined as the predominant coreceptor for T-tropic HIV-1 variants (Feng et al, 1996). Soon after, it was discovered that the reason that RANTES, MIP-1 alpha, and MIP-1 beta inhibited infection was because they were binding to the CCR5 coreceptor and this was presumed to be the coreceptor for M-tropic HIV-1 variants (Deng et al, 1996).

This, however, turned out to be an incomplete story and variants described as NSI on cell lines were observed forming syncytia on primary CD4+ T cells and replicating in MT-2 cells (Forte et al, 1994; Todd et al, 1995; Berger et al, 1998; Goodenow and Collman, 2006). The syncytia induction tests being used to categorize viral variants were performed using the MT-2 cell line, which only expressed the CXCR4 coreceptor and not CCR5, though this was not known at the time (Koot et al, 1992). Thus this test had really been a test of coreceptor use and not cell tropism. A new definition of tropism was proposed based on coreceptor use with CCR5-using variants (R5), CXCR4-using variants (X4), and dual-tropic variants (R5X4) (Berger et al, 1998). In addition, these variants could be defined as T-tropic or M-tropic. Current tropism tests now include genotype analyses using prediction algorithms, such as WebPSSM and Geno2Pheno (Jensen et al, 2003; Sing et al, 2007), that predict whether an Env uses CCR5 or CXCR4, and
phenotype analyses, such as testing infection on primary cells or cell lines expressing one or both coreceptors.

T-tropic R5 variants are almost always the transmitted/founder viruses in a new infection (Keele et al, 2008; Salazar-Gonzalez et al, 2009; Wilen et al, 2011; Ochsenbauer et al, 2012). Tropism may then switch during disease progression and X4 variants emerge in approximately 50% of patients. There is some evidence that R5X4 strains represent an intermediate in the evolution of X4 variants (Yi et al, 1999; Huang et al, 2007; Irlbeck et al, 2008).

Coreceptor use has been mapped mainly to the V3 region of gp120 (O’Brien et al, 1990; Hwang et al, 1991; Shioda et al, 1991; Milich et al, 1997; Nelson et al, 1997; Nelson et al, 2000; Hoffman et al, 2002; Jensen et al, 2003). Accordingly, R5- and X4-using variants have V3 loops with different properties. For example, X4 variants have an overall higher positive charge compared to R5 variants (Fouchier et al, 1992). This may be in part due to changes to basic amino acids at positions 11 or 25 in the V3 region, which is associated with X4 tropism (Resch et al, 2001).

Once X4 variants are established in the HIV-1 population, disease prognosis is poor (Tersmette et al, 1989; Koot et al, 1993; Richman and Bozzette, 1994; Connor et al, 1997; Blaak et al, 2000). The reason for a rapid decline of CD4+ T cells and acceleration to AIDS and death is not known. One hypothesis is that as variants gain the ability to use the X4 receptor, they are able to rapidly deplete previously non-permissive T cell populations.
While the T cell population expresses both R5 and X4, there is a significant difference in the expression of R5 and X4 in T cell subsets: memory T cells express both R5 and X4 and naïve T cells only express X4 (Blaak et al, 2000). This means that in the overall T cell population, up to 90% of the T cells may be expressing X4, but only 15-35% may be expressing the R5 receptor (Bleul et al, 1997; Berkowitz et al, 1998; de Roda Husman et al, 1999). Given the abundance of permissive cells, it is unclear why X4 variants do not emerge sooner during the course of infection. It is possible that R5 variants out compete X4 variants by infecting CCR5+ CD4+ T cells recruited to the sites of infection or that immune pressures select against X4 variants. This is supported by the observation that X4 variants are more sensitive to CD4 binding site (CD4bs) antibodies and that they emerge later in disease progression when immune competence wanes (Bunnik et al, 2007). However, this explanation does not account for why X4 variants would not be transmitted and be present at earlier time points.

Cell tropism may also change as disease progression occurs, allowing viral variants within an individual to have different cell tropisms (Koyanagi et al, 1987). Though transmitted variants are usually T-tropic, M-tropic variants often arise and may predominate in some tissues. For example, R5 variants from the brain have higher frequency of M-tropism than R5 variants from blood and immune tissues do (Gorry et al, 2002; Peters et al, 2004; Peters et al, 2006; Peters et al, 2007). Macrophage tropism is associated with the ability to use low levels of CD4 and may become more prevalent in late stage infection (Li et al, 1999; Tuttle et al, 2002;
Gray et al, 2005; Thomas et al, 2007). Though they are often R5-using, macrophage tropic R5X4 and X4 variants have been isolated (Simmons et al, 1996; Simmons et al, 1998; Yi et al, 2003; Gray et al, 2009).

Several regions of gp120 have been implicated in the development of macrophage tropism, including regions flanking the CD4bs, V1V2, C2, and V4 (Koito et al, 1994; Imamichi et al, 2002; Walter et al, 2005; Dunfee et al, 2006; Dunfee et al, 2007; Duenas-Decamp et al, 2009; Musich et al, 2011). Specifically, an asparagine at position 283, loss of an N-linked glycosylation site at 386, and an E153G substitution have been identified as determinants of macrophage tropism (Dunfee et al, 2006; Dunfee et al, 2007; Musich et al, 2011). It has also been observed that macrophage tropic variants isolated from brain tissue have gp120s with lower charges in comparison with non-macrophage variants from immune tissues from the same patient (Gonzalez-Perez et al, 2012).

1.6 Compartmentalization

As discussed above, HIV-1 has extreme genetic diversity and this allows populations of viral quasispecies to arise with mutations that allow them to use different coreceptors, have different cell tropisms, develop drug resistance, and escape antibody neutralization, among other viral properties. Quasispecies may be particular to a specific tissue site, or compartment, and are said to be compartmentalized.
Compartmentalization may be the result of the availability of particular cell types in a specific tissue compartment, the presence or absence of antibodies (e.g., the brain is immune privileged), tissue specific coinfections (e.g., tuberculosis in the lungs), or the concentrations of antiretroviral drugs within in a tissue. Several studies have shown that HIV drugs do not penetrate tissues uniformly. This has been well documented in the brain, where drugs may be unable to cross the blood brain barrier (BBB), allowing the brain to potentially serve as a sanctuary site with continued replication despite controlled viremia in the plasma (Nowacek and Gendelman, 2009; Varatharajan and Thomas, 2009; Gomes et al, 2014). Recent studies have also shown that antiretroviral treatments can be lower in lymphatic tissues than blood, which similarly allows for continued replication even in patients on therapy (Fletcher et al, 2014; Lorenzo-Redondo et al, 2016). Yet another study reported that patients with undetectable viral loads in their plasma had virus detected in their semen and this correlated with lower semen drug concentrations (Lorello et al, 2009). This could have important implications for HIV-1 transmission.

The HIV-1 Env is particularly susceptible to selection pressures due to its position on the virion surface, and thus exposure to antibodies, tissue microenvironments, and host target cells. HIV-1 envelope compartmentalization has been observed in peripheral blood mononuclear cells (PBMCs), brain tissues/cerebrospinal fluid (CSF), immune tissues, male and female genital tracts, lung, and possibly other tissues (reviewed in Blackard, 2012).
1.7 Scope of Thesis

The emergence of diverse populations of intrahost viral quasispecies is a hallmark of HIV-1 infection. Understanding how these population behave and their ability to evade immune responses, infect different cell types, and establish reservoirs is essential to producing effective antiretroviral drugs, as well the development of curative treatments and a vaccine. To this end, the HIV-1 envelope remains a critically important target. Despite this, there have been no studies to date looking at full length envelopes from multiple tissues from a single HIV-1 infected individual with both genotype and phenotype characterization.

In this study, 75 full length envelopes were examined from the frontal lobe, occipital lobe, parietal lobe, colon, lung, and lymph node of an HIV-1 positive subject who died from end stage AIDS. Genetic studies were conducted where sequences were characterized based on the presence of features known to disrupt function (e.g., insertions, deletions, and premature stop codons) and examined for mutations known to change tropism (e.g., N283, loss of an N-linked glycosylation site at 386, E153G, changes to V3 loop, etc). In addition, length, charge, and number of potential N-linked glycosylation sites (PNGS) were calculated for specific env regions. Sequences were analyzed for hypermutation and recombination and assembled into phylogenetic trees to assess compartmentalization. Functional studies were used to characterize phenotypes of each Env, including functionality, coreceptor use, cell tropism, and susceptibility to inhibition by a panel of inhibitors and monoclonal antibodies. Finally, state of the
art Pacific Biosciences next generation sequencing was used to examine thousands of full length envs from each tissue, the first time this has been done with full length proviral HIV-1 env.

The results of this study provide new insights into HIV-1 populations in different tissue compartments, an area of HIV research that has not been adequately addressed. Analyzing these populations will provide a better understanding of infection in different tissues and how viral variants at these sites evolve differently. This could lead to more targeted drug regimens and aid in vaccine development.
CHAPTER II: MATERIALS AND METHODS

2.1 HIV-1 Positive Subject and Tissues

HIV-1 positive tissue was obtained from the National Disease Research Interchange (NDRI; Philadelphia, PA). Tissue was harvested postmortem and stored at -80°C until use. For this study, six tissues were examined from Subject 162, along with blood plasma and serum (Table 2.1). These included three brain lobes (frontal lobe (FL), occipital lobe (OL), and parietal lobe (PL)), colon (C), lung (L), and lymph node (LN). Subject 162 died of end stage AIDS.

2.2 Nucleic Acid Extraction

DNA was harvested from tissues and cells from plasma and serum using the QIAamp DNA Mini Kit (QIAGEN) as described by the manufacturer’s protocol. DNA was eluted in 300 μl nuclease free, PCR grade water and stored at -80°C.

RNA was isolated from plasma and serum using the High Pure Viral RNA Kit (Roche), eluted in 40 μl nuclease free, PCR grade water, and stored at -80°C.

2.3 PCR

Limiting Dilution Amplification

Nested PCR reactions were performed using limiting dilution amplification of HIV-1 proviral DNA. For the first PCR, outer primers ReenvA and EnvN were used (Table 2.2). Inner primers ReenvB_TOPO and Env-lo were used for the
<table>
<thead>
<tr>
<th>Subject</th>
<th>A/R/S</th>
<th>Tissues</th>
<th>Plasma Viral Load</th>
<th>CD4 Count</th>
<th>HIV Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>162</td>
<td>53/B/F</td>
<td>Brain - Frontal Lobe</td>
<td>undetectable (&lt;40 copies/ml)</td>
<td>733 cells/μl (5 months prior to death)</td>
<td>Emtricitabine, Tenofovir, Lopinavir, Ritonavir, Kaletra</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Brain - Occipital Lobe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Brain - Parietal Lobe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Colon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lung</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lymph Node</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plasma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Serum</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.1 Subject 162 data** Tissue was obtained from the National Disease Research Interchange (NDRI; Philadelphia, PA). A/R/S = age/race/sex. Viral load testing was performed at the Ann & Robert H. Lurie Children’s Hospital of Chicago.
<table>
<thead>
<tr>
<th>Primer</th>
<th>Use</th>
<th>Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>EnvN</td>
<td>LD/RT-PCR/PacBio outer</td>
<td>CTG CCA ATC AGG GAA GTA GCC TTG TGT</td>
<td>Gao et al, 1996</td>
</tr>
<tr>
<td>RevenvA</td>
<td>LD/PacBio outer</td>
<td>TAG AGC CCT GGA AGC ATC CAG GAA G</td>
<td>Salazar-Gonzalez et al, 2008</td>
</tr>
<tr>
<td>RevenvB_TOPO</td>
<td>LD inner</td>
<td>CAC CTA GGC ATC TCC TAT GGC AGG AAG</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>Env-lo</td>
<td>LD inner</td>
<td>GTC TCT TCC AGT CCC TTT TCT TTT AAA AAG</td>
<td>Revilla et al, 2011</td>
</tr>
<tr>
<td>SG3-up</td>
<td>RT-PCR</td>
<td>TAC AGT GCA GGG GAA AGA AT AATA GAC ATA ATAG</td>
<td>Revilla et al, 2011</td>
</tr>
<tr>
<td>SG3-lo</td>
<td>RT-PCR</td>
<td>AGA CCC AGT ACA GGC RAR AAG C</td>
<td>Revilla et al, 2011</td>
</tr>
<tr>
<td>Vif-2fw</td>
<td>RT-PCR</td>
<td>TAG GGA TTA TGG AAA ACA GAT GGC AG</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>TAT-1</td>
<td>RT-PCR</td>
<td>CCT AAA CTA GAG CCC TGG AAG CAT CC</td>
<td>Lee et al, 2000</td>
</tr>
<tr>
<td>T7*</td>
<td>sequencing/colony screen</td>
<td>TAA TAC GAC TCA CTA TAG GG</td>
<td>GENEWIZ universal primer</td>
</tr>
<tr>
<td>BGHR*</td>
<td>sequencing</td>
<td>TAG AAG GCA CAG TCG AGG</td>
<td>GENEWIZ universal primer</td>
</tr>
<tr>
<td>GP41F1</td>
<td>sequencing</td>
<td>GAG CAG CAG GAA GCA CTA T</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>GP41F2</td>
<td>sequencing</td>
<td>TGA ATA GAG TTA GGC AGG G</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>GP41R1</td>
<td>sequencing</td>
<td>ATA GTG CTT CTA GCT C</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>GP41R1Yang</td>
<td>sequencing</td>
<td>AAG GAC AAG GGT GAG TAT CCC TGC CTA A</td>
<td>Yang et al, 1999</td>
</tr>
<tr>
<td>M5</td>
<td>sequencing</td>
<td>CCA ATT CCC ATA CAT TAT TGT GCC CCA GC</td>
<td>Takehisa et al, 1998</td>
</tr>
<tr>
<td>MSR</td>
<td>sequencing/colony screen</td>
<td>CCA GCC GGG GCA CAA TAA TGT ATG GG</td>
<td>Takehisa et al, 1998</td>
</tr>
<tr>
<td>GP41R2</td>
<td>sequencing</td>
<td>CCC TGC CTA ACT CTA TTC A</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>V1/V2A1F</td>
<td>sequencing</td>
<td>CAG ATG CTA AAG CAT ATG</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>V1/V2A2F</td>
<td>sequencing</td>
<td>TCA AAG CCT AAA GCC ATG</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>V1/V2A3F</td>
<td>sequencing</td>
<td>CCC ATA CAT TAT TGT GCC</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>V1/V2A4F</td>
<td>sequencing</td>
<td>TCA ACT GAA CTA GGT TTA</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>FL3F</td>
<td>sequencing</td>
<td>GCT GTG GAA AGA TAC CTA</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>4VF</td>
<td>sequencing</td>
<td>GAC CCA GAA ATT GTA ATG C</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>470F</td>
<td>sequencing</td>
<td>CTT TCA ATG TCA CCA CAG GC</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>1200F</td>
<td>sequencing</td>
<td>AGT TTT AAT TGT GGA GGG G</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>2240F</td>
<td>sequencing</td>
<td>GAA GAA GAA GGT GGA GAG</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>2310F</td>
<td>sequencing</td>
<td>TCT AGA TCG ACC TGA AGA GC</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>2470F</td>
<td>sequencing</td>
<td>TCT AGA TCG ACC TGA AGA GC</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>2170R</td>
<td>sequencing</td>
<td>GGT GAT TCG CCC TGC CTA ACT C</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>2180R</td>
<td>sequencing</td>
<td>CGG GTG TGA AAC GAT AAT GG</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>2320R</td>
<td>sequencing</td>
<td>GGC TCT TCA GGT CGA TCT AG</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>2850R</td>
<td>sequencing</td>
<td>CTT CCT GTG CTT CTA GCC AGG C</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>PacBioF-long</td>
<td>PacBio inner</td>
<td>GAG CAG AAG ACA GTG GCA ATG AGA GTG A</td>
<td>Clapham Lab</td>
</tr>
<tr>
<td>PacBioR-long</td>
<td>PacBio inner</td>
<td>TTG ACC ACT TGC CAC CCA TCT TAT AGC A</td>
<td>Clapham Lab</td>
</tr>
</tbody>
</table>

**Table 2.2 HIV-1 env primers** Primers were used for limiting dilution PCR (LD), RT-PCR, to colony screen, for sequencing (Sanger), and PacBio sequencing. Primers marked with an asterisk were used for env sequencing, but were located in the pcDNA 3.1D/V5-His-TOPO plasmid.
second PCR (Table 2.2). The Phusion High-Fidelity DNA Polymerase Kit (New England BioLabs) was used according to the manufacturer’s protocol for 50 μl PCR reactions using “Phusion” thermocycler program with 40 cycles (Table 2.3). DNA was serially diluted to determine the concentration at which less than 30% of the reactions were positive. At this concentration, there is a >80% chance of PCR products being derived from single genomes and thus avoiding the generation of PCR generated recombinants during the amplification (Salazar-Gonzalez et al, 2008). Recent studies have demonstrated that there are minimal differences between products of limiting dilution PCR and bulk PCR (Etemad et al, 2015). However, limiting dilution PCR was used to be consistent with previously used methods and because recombination analyses were performed. PCR reactions were checked for positives by running 5 μl out on a 1.0% agarose gel with 0.5 μg/ml ethidium bromide in TAE buffer. Positive bands were approximately 3 kb, as determined using a 1 kb ladder for reference (New England BioLabs).

RT-PCR

RT-PCR was used to amplify env from viral RNA isolated from Subject 162 plasma and serum. SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Invitrogen) was used to synthesize cDNA and amplify env in a one-step PCR using primers SG3-up, SG3-lo, Vif2F, TAT1, and EnvN (Table 2.2) and the “RT-PCR” thermocycler program with 40 cycles (Table 2.3). RNA dilutions were used to achieve limiting dilution conditions, as described above. The second PCR was the same as described above, using primers ReenvB_TOPO and
<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>2</td>
<td>98°C</td>
<td>10 seconds</td>
</tr>
<tr>
<td>3</td>
<td>70°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>4</td>
<td>72°C</td>
<td>1.5 minutes</td>
</tr>
<tr>
<td>5</td>
<td>Repeat steps 2-4</td>
<td>40x</td>
</tr>
<tr>
<td>6</td>
<td>72°C</td>
<td>7 minutes</td>
</tr>
<tr>
<td>7</td>
<td>4°C</td>
<td>∞</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95°C</td>
<td>5 minutes</td>
</tr>
<tr>
<td>2</td>
<td>95°C</td>
<td>1 minute</td>
</tr>
<tr>
<td>3</td>
<td>55°C</td>
<td>1 minute</td>
</tr>
<tr>
<td>4</td>
<td>72°C</td>
<td>1 minute</td>
</tr>
<tr>
<td>5</td>
<td>Repeat steps 2-4</td>
<td>40x</td>
</tr>
<tr>
<td>6</td>
<td>72°C</td>
<td>7 minutes</td>
</tr>
<tr>
<td>7</td>
<td>4°C</td>
<td>∞</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95°C</td>
<td>1 minute</td>
</tr>
<tr>
<td>2</td>
<td>95°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>3</td>
<td>68°C</td>
<td>3 minutes</td>
</tr>
<tr>
<td>4</td>
<td>Repeat steps 2-3</td>
<td>35x</td>
</tr>
<tr>
<td>5</td>
<td>70°C</td>
<td>10 minutes</td>
</tr>
<tr>
<td>6</td>
<td>4°C</td>
<td>∞</td>
</tr>
</tbody>
</table>

**Table 2.3 Thermocycler programs used for PCR** Programs were used for amplifying *envs* from patient tissues (Phusion), for RT-PCR to amplify *envs* from plasma and serum (RT-PCR), for colony screening to detect *env*+ clones (General), and to produce amplicons for Pacific Biosciences sequencing (PacBio).
Env-lo (Table 2.2), along with the Phusion High-Fidelity DNA Polymerase Kit (New England BioLabs) and Phusion thermocycler program with 40 cycles (Table 2.3).

2.4 Cloning

When PCR reactions had fewer than 30% of the reactions positive, the positive reactions were run out on a 1.0% agarose gel with 0.8% crystal violet in TAE buffer. Bands were purified using the QIAquick Gel Extraction Kit (QIAGEN) as described by the manufacturer. PCR products were eluted in 30 μl nuclease free, PCR grade water and stored at -20°C until use.

Eluted PCR products were ligated into pcDNA 3.1D/V5-His-TOPO using TA Expression Kit (Invitrogen). Ligation reactions were performed as directed by the manufacturer using the following reaction mix:

- 1 μl PCR product
- 1 μl pcDNA 3.1D/V5-His-TOPO vector
- 1 μl salt solution
- 3 μl water

Ligation products were transformed into TOP10 chemically competent cells (Invitrogen) and then plated on LB plates with 100 μg/ml carbenicillin to select positive clones.

Clones were checked for positive inserts using Go Taq Green Master Mix (Promega) with primers T7 and M5R (Table 2.2) using the “General” thermocycler program with 40 cycles (Table 2.3). PCR reactions were run on a 1.0% agarose gel with 0.5 μg/ml ethidium bromide in TAE buffer and compared to a 100 kb ladder (New England BioLabs) for confirmation of an approximately 1 kb product.
2.5 Sanger Sequencing

Positive colonies were grown in LB with 100 μg/ml ampicillin cultures overnight, pelleted by centrifugation, and plasmid DNA was isolated using the QIAprep Miniprep Kit (QIAGEN). DNA was eluted in 50 μl nuclease free, PCR grade water and stored at -20°C until use.

Each clone was sequenced in both directions using a series of primers directed to conserved regions of the env sequence (Table 2.2). Sanger sequencing was performed by GENEWIZ (South Plainfield, NJ). The first eight sequencing primers in Table 2.2 were used for every clone and then additional primers were used as needed until the DNA was sequenced completely in both directions.

2.6 Sequence Analyses

Alignments were constructed for each envelope using SeqMan Pro (DNASTAR) and sequence traces were checked for quality using FinchTV (Geospiza). They were assembled into consensus sequences and checked against the BLAST databases at the National Center for Biotechnology Information (NCBI) and Los Alamos National Laboratory (LANL) to ensure that they were not generated from laboratory contamination.

Sequences were then examined for mutations known to affect tropism (e.g., N283, loss of an N-linked glycosylation site at 386, E153G, changes to V3 loop, etc), length and charge of env regions, and changes in potential N-linked glycosylation sites (PNGS). PNGS were counted using the LANL N-Glycosite
program (Zhang et al, 2004). The WebPSSM program (Jensen et al, 2003) was used to predict coreceptor use for each envelope based on the amino acid sequence of the V3 loop. Sequences were also checked for characteristics of non-functional envs, including insertions, deletions, and premature stop codons.

Sequences were analyzed for hypermutation and recombination by Thomas Leitner at LANL. Hypermutation was assessed using the Hypermut program (Rose and Korber, 2000). The Hypermut program detects G to A mutations that are likely the result of APOBEC3G or APOBEC3F activity of based on the context of the mutations in comparison to a reference sequence. Recombination was predicted using a hierarchical PHI test (Bruen et al, 2006; Immonen and Leitner, 2014). The PHI test measures the pairwise homoplasy index for sequences to predict recombination based on convergent mutations.

### 2.7 Phylogenetics

The env sequences from each tissue were aligned in GeneDoc (Pittsburgh Supercomputing Center) using manual adjustments with gaps excluded. Four HIV-1 group M subtype B reference sequences from LANL were included as well to be used as an outgroup:

Ref.B.FR.83.HXB2_LAI_IIIB_BRU.K03455  
Ref.B.US.98.1058_11.AY331295  
Ref.B.NL.00.671_00T36.AY423387  
Ref.B.TH.90.BK132.AY173951
Maximum likelihood phylogenetic trees were then constructed using Molecular Evolutionary Genetics Analysis (MEGA) software version 5 (Tamura et al, 2011). Trees were constructed for the following regions of env: gp160, gp120, gp41, V1-V2, V1-V5, V1, V2, V3, V4, V5, C1, C2, C3, C4, and C5. For full length env, the General Time Reversible Substitution Model with Gamma distribution was used and compartmentalization was confirmed using the Slatkin-Maddison test (Slatkin and Maddison, 1989). MEGA was used to select the best models for trees of individual env regions. In all cases, bootstrap analyses were used with 1,000 replicates and values ≥70% were noted on the trees.

2.8 Cell Culture

Three cell lines were used in this study: HEK 293T cells (American Type Culture Collection), HeLa TZM-bl cells (NIH AIDS Reagent Program), and HeLa HIJ cells (Platt et al, 1998). The HEK 293T cell line is a human embryonic kidney cell line containing the SV40 Large T-antigen and was used for the generation of pseudotyped viruses, or pseudoviruses (Graham et al, 1977; DuBridge et al, 1987; Pear et al, 1993). HeLa TZM-bl cells are indicator cells that express high levels of CD4, CCR5, and CXCR4 and have HIV-inducible luciferase and β-galactosidase genes (Platt et al, 1998; Derdeyn et al, 2000; Wei et al, 2002; Takeuchi et al, 2008; Platt et al, 2009). These were used to determine pseudovirus titers and for inhibition and neutralization assays. HeLa HIJ cells are a HeLa cell line that express CD4 and CXCR4, but not CCR5, and therefore can be used to evaluate
CXCR4 coreceptor tropism (Kabat et al, 1994; Platt et al, 1998). Cell lines were maintained in Dulbecco’s Modified Eagle Medium (DMEM) (Gibco) with 10% fetal bovine serum and 10 μg/ml gentamicin.

Macrophages were isolated from leuopaks (New York Biologics) using Ficoll separation. Following purification, cells were resuspended at 6.5 x 10^6 cells/ml in DMEM with 10% human plasma and 10 μg/ml gentamicin and plated on 150 mm non-tissue culture treated plates. Cells were incubated overnight and then washed with media the next day to remove non-adherent cells. Cells were allowed to incubate for an additional 6 days (7 days total) and then macrophages were replated for tropism experiments.

### 2.9 Pseudovirus Production

Pseudoviruses were constructed using env+ pcDNA 3.1D/V5-His-TOPO plasmids, carrying patient derived env genes, and an env- pNL4.3Δenv plasmid, an HIV-1 clone with a premature stop codon in the env gene (Bhattacharya et al, 2004). For each patient env, 1.25 μg of env+ plasmid and 1.25 μg of env- plasmid were cotransfected into 293T cells using the ProFection Mammalian Transfection System (Promega) as described by the manufacturer. Supernatants were harvested 48 hours post transfection and clarified by centrifugation. Cell free pseudovirus stocks were removed, aliquoted, frozen in liquid nitrogen, and stored at -152°C until use.
2.10 Viral Titer and Tropism

TZM-bl and HIJ cells were used to determine Env functionality and ability to use CXCR4 as a coreceptor, respectively. For each infectivity assay, cells were diluted to 4.0 x 10⁴ cells/ml and 500 µl/well were added to 48-well plates the day prior to infection. The day of infection, pseudovirus stocks were thawed at room temperature and then serially diluted in media to yield four concentrations: stock, 1/10, 1/100, and 1/1000. In addition, several laboratory strains were used as controls. For TZM-bl infection, B33, JRFL, and JRCSF were used as positive controls and L411 env- plasmid and L411 env- plasmid + pSVIII plasmid were used as negative controls. For HIJ infection, NL4.3 was used as a positive control and B33, JRFL, JRCSF, L411 env- plasmid, and L411 env- plasmid + pSVIII plasmid were used as negative controls.

Cells were infected by removing media and adding 100 µl of pseudovirus dilution. Plates were incubated for 3 hours at 37°C and then 400 µl of media were added to each well. Plates were incubated at 37°C for 48 hours. TZM-bl cells were fixed with 0.5% glutaraldehyde in phosphate-buffered saline (PBS). Cells were stained by rinsing twice in 1 x PBS and then adding 500 µl X-gal substrate (0.5 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, 3 mM potassium ferricyanide, 3 mM potassium ferrocyanide, 1 mM magnesium chloride). Plates were incubated at room temperature overnight and then β-galactosidase expressing cells were counted to determine the focus forming units (FFU)/ml of each pseudovirus stock. HIJ cells were fixed in cold 1:1 methanol:acetone,
washed, and immunostained for p24 using monoclonal antibodies 365 and 366 (UK Centre for AIDS Research), followed by an IgG-β-galactosidase conjugate (Southern Biotech), and then X-gal substrate. Plates were incubated overnight at room temperature and FFU were counted.

To determine macrophage tropism, 500 μl of macrophages at 2.5 x 10^5 cells/ml were plated in 48-well plates the day prior to infection. On the day of infection, media was removed from cells and 100 μl of 10 μg/ml DEAE dextran were added to each well and plates were incubated at 37°C for 30 minutes. Following incubation, 100 μl of serially diluted pseudovirus stocks, described above, were added to each well. Laboratory strains B33 and JRFL were used as positive controls and JRCSF, L411 env- plasmid and L411 env- plasmid + pSVIII plasmid were used as negative controls. Macrophages were spinoculated for 45 minutes at 1200 rpm in a benchtop centrifuge, incubated for 3 hours at 37°C, and then 300 μl media were added to each well. Plates were incubated for one week, adding additional media as needed. Macrophages were fixed, stained, and counted as described above for HIJ cells.

For viral titer and tropism experiments, a titer of 10^3 FFU/ml was used as a cutoff to determine whether a pseudovirus was functional on TZM-bl cells or could infect macrophages or HIJ cells. This titer was selected because it is sufficiently high to use the pseudovirus for inhibition and neutralization experiments. Below this value, it is difficult to observe changes in infection. However, it should be noted that while pseudoviruses below this value are categorized as “non-functional” for
the purpose of this study, they may still be infectious at a lower level. Likewise, they may be able to infect macrophages and HIJ cells at lower levels.

2.11 Inhibition and Neutralization Assays

Inhibition and neutralization assays were carried out as previously described to determine IC50 values, or the concentration of inhibitor or antibody at which infection is reduced by 50% (Peters et al, 2008). Two inhibitors and four monoclonal antibodies were tested: sCD4 (prepared in house), maraviroc (NIH AIDS Reagent Program), b12 (provided by Dennis Burton, The Scripps Research Institute), PG16 (Polymun Scientific), 17b (provided by George Lewis, Institute of Human Virology) and 447-52D (Polymun Scientific) (Table 2.4). Inhibitions were performed using HIV-1 inducible HeLa TZM-bl cells, which were diluted to 1.6 x 10^5 and plated at 50 μl/well in 96-well white-sided luminometer plates the day prior to the assay. The day of the assay, pseudovirus stocks were thawed at room temperature and diluted to 8000 FFU/ml. Several laboratory strains were used as controls for each experiment. Vesicular Stomatitis Virus G protein (VSVG) was used as a resistant control for all inhibitions and neutralizations. For sCD4, maraviroc, and b12, JRFL and JRCSF were used as sensitive controls. For PG16, JRFL was used as a resistant control and JRCSF was used as a sensitive control. In the 17b and 447-52D neutralization experiments, NL4.3 was used as a sensitive control and JRFL and JRSCF were used as resistant controls.
Eleven twofold serial dilutions were used for each inhibitor or antibody along with a media only, no inhibitor/antibody sample (Table 2.4). For each well of maraviroc inhibition, 25 μl serially diluted maraviroc were added to each well and incubated for 30 minutes at 37°C. Then 25 μl (200 FFU) pseudovirus were added to each well. For each well of inhibition or neutralization with sCD4, b12, PG16, 17b, and 447-52D, 25 μl (200 FFU) pseudovirus and 25 μl serially diluted inhibitor or antibody were combined, incubated for 1 hour at 37°C, and then added to TZM-bl cells. Plates were incubated at 37°C overnight and media was changed in the morning. The following day, media was removed via aspiration, 100 μl media without phenol red was added to each well and infection was detected with a luciferase readout using Beta-Glo (Promega). To each well, 100 μl Beta-Glo were added and plates were incubated at room temperature in the dark for 30 minutes. Plates were read using a BioTek Clarity luminometer (Winooski, VT) to obtain Relative Light Units (RLU).

2.12 Next Generation Sequencing

Full length envs were amplified for Pacific Biosciences (PacBio) next generation sequencing. DNA was isolated from tissues as described above and a nested PCR approach was used to amplify a 2.6 kb product. Outer primers ReenvA and EnvN and inner primers PacBioF-long and PacBioR-long were used (Table 2.2). The Advantage 2 PCR Kit (Clontech) was used as described in the manufacturer’s protocol for 50 μl reactions. The thermocycler program “PacBio”
was used with 35 cycles (Table 2.3). PCR products were cleaned up by running them out on a 2% E-Gel EX Agarose Gel (ThermoFisher Scientific) and purified with a QIAquick Gel Extraction Kit (QIAGEN). PCR products were eluted in 50 μl nuclease free, PCR grade water. Samples were sent to the Deep Sequencing Core Labs at the University of Massachusetts Medical School for sequencing using the PacBio RS II instrument. One SMRT Cell was analyzed for DNA from each tissue with 6-hour collection times.

Bioinformatics analyses were performed by Bioinfoexperts, LLC (Thibodaux, LA). The V3 loop nucleic acid sequences were isolated and aligned using Geneious software version 9 (Kearse et al, 2012) and Gene Cutter (Los Alamos National Laboratory), respectively. Any V3 amino acid sequence containing an ambiguous position was removed from the alignment. Tropism prediction was performed using WebPSSM (Jensen et al, 2003) and the x4r5 matrix. A python script was generated to randomly select 250 sequences from each tissue. Maximum likelihood phylogenies were generated with phyML (Guindon et al, 2010) using the HKY85 substitution model, the NNI tree improvement method, and the approximate likelihood ratio test (Anisimova and Gascuel, 2006). A midpoint rooted tree was graphed and colored according to tissue type using FigTree software (http://tree.bio.ed.ac.uk/software/figtree/).
<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Epitope</th>
<th>Starting Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>sCD4</td>
<td>CD4 binding site</td>
<td>50 µg/ml</td>
</tr>
<tr>
<td>Maraviroc</td>
<td>CCR5</td>
<td>100 ng/ml</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Epitope</th>
<th>Starting Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>b12</td>
<td>CD4 binding site</td>
<td>50 µg/ml</td>
</tr>
<tr>
<td>PG16</td>
<td>V2&amp;V3</td>
<td>10 µg/ml</td>
</tr>
<tr>
<td>447-52D</td>
<td>V3</td>
<td>50 µg/ml</td>
</tr>
<tr>
<td>17b</td>
<td>CD4i</td>
<td>50 µg/ml</td>
</tr>
</tbody>
</table>

Table 2.4 Inhibitors and antibodies A panel of HIV-1 inhibitors and neutralizing antibodies were tested for their ability to block infection of TZM-bl cells using pseudoviruses with patient derived Envs.
CHAPTER III: GENETIC CHARACTERIZATION OF HIV-1 ENVELOPES FROM TISSUES OF AN HIV-1 INFECTED SUBJECT

3.1 Introduction

A number of studies have addressed the diversity of HIV-1 env quasispecies. However one caveat of these studies is that they often look at only a particular region of the env (e.g., V1-V2 or V3), which may not be representative of the overall env sequence and may not be sufficient to infer phenotypes. Furthermore, these studies typically only examine a few tissue sites, such as brain, immune tissue, and plasma. In the first part of this study, envs were amplified from multiple tissues of an HIV-1 positive subject. They were then compared for compartmentalization, sequence variation, length, charge, possible N-linked glycosylation sites (PNGS), hypermutation, and recombination.

Previous studies have shown genetic compartmentalization in the brain compared to plasma and immune tissue, but there is limited data on compartmentalization in other tissues (Wong et al, 1997; van’t Wout et al, 1998; Wang et al, 2001; Caragounis et al, 2007; Schnell et al, 2010; Gonzalez-Perez et al, 2012; Sturdevant et al, 2015). Similarly, studies have been conducted to examine the overall charge of env regions, but they have been restricted to only a few tissue compartments. Nevertheless, it has been reported that CCR5-using envelopes carried an increased overall positive charge as disease progressed (Repits et al, 2008; Seclen et al, 2011; Gonzalez-Perez et al, 2012). In AIDS patients with neurological complications, envelopes in immune tissue carried a
higher positive charge in the V1-V5 region compared to those in brain tissue (Gonzalez-Perez et al, 2012).

Mutations in the HIV-1 genome can impact quasispecies diversity. These include mutations that result in amino acid substitutions as well as loss or addition of PNGS. Together with insertions and deletions, they play an important role in diversity and phenotypes of viral variants. Several groups have examined the relationship between sequence length and PNGS and disease progression. These studies have found that variants with shorter V1-V2 lengths and less PNGS may be selected for at transmission, but that length expanded and number of PNGS increased over the course of infection (Pollakis et al, 2001; Derdeyn et al, 2004; Chohan et al, 2005; Sagar et al, 2006; Liu et al, 2008; Curlin et al, 2010; Huang et al, 2012; Wang et al, 2013; Pollakis et al, 2015).

HIV-1 diversity can also be affected by human restriction factors, aimed at protecting host cells. Hypermutation resulting from the activity of APOBEC3G and F results in defective virus. Vif protects HIV-1 from APOBEC activity by inducing its degradation (Sheehy et al, 2002). However, in cells that express higher amounts of APOBEC or following a vif mutation, hypermutation can result (Lecossier et al, 2003). Previous studies have shown that APOBEC genes are differentially expressed in tissue compartments and that this can result in different levels of hypermutation (Koning et al, 2009; Refsland et al, 2010; Fourati et al, 2014).
Finally, recombination has been shown to increase genetic diversity in HIV-1 populations and increase the rate of evolution (Jetzt et al, 2000; An and Telesartsky, 2002; Rhodes et al, 2003; Charpentier et al, 2006; Chen et al, 2006; Mild et al, 2007; Zhang et al, 2010; Brown et al, 2011; Immonen et al, 2015; Sanborn et al, 2015).

To effectively treat HIV-1 and develop a vaccine, it is essential to understand its genetic variability and how this may be different in tissue compartments. This study represents the first time that full length env sequences have been amplified and analyzed from multiple tissue sites within an infected individual. In addition, all of the envelopes were functionally characterized, which is discussed in Chapter IV.

3.2 Results

3.2.1 Envelopes were amplified from all tissues except for plasma and serum

Limiting dilution PCR was used to amplify full length envelopes from proviral DNA at end point dilution from frontal lobe, occipital lobe, parietal lobe, colon, lung, and lymph node tissues, as well as from residual cellular material in plasma and serum of Subject 162 (Table 2.1). Envelopes were successfully amplified from all tissues, except plasma and serum, and used for further analyses (Table 3.1). Envelopes were named as follows: number of the PCR reaction – number of the PCR tube – number of the clone that was sequenced. A total of 8 envelopes were
<table>
<thead>
<tr>
<th>Tissue</th>
<th>Envelope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain - Frontal Lobe (FL)</td>
<td></td>
</tr>
<tr>
<td>8 envelopes</td>
<td>17-53-4</td>
</tr>
<tr>
<td></td>
<td>52-19-3</td>
</tr>
<tr>
<td></td>
<td>53-30-5</td>
</tr>
<tr>
<td></td>
<td>53-53-3</td>
</tr>
<tr>
<td></td>
<td>65-15-1</td>
</tr>
<tr>
<td></td>
<td>68-59-1</td>
</tr>
<tr>
<td></td>
<td>70-14-2</td>
</tr>
<tr>
<td></td>
<td>74-14-2</td>
</tr>
<tr>
<td>Brain - Occipital Lobe (OL)</td>
<td></td>
</tr>
<tr>
<td>10 envelopes</td>
<td>19-94-2</td>
</tr>
<tr>
<td></td>
<td>19-108-2</td>
</tr>
<tr>
<td></td>
<td>52-59-21</td>
</tr>
<tr>
<td></td>
<td>54-32-6</td>
</tr>
<tr>
<td></td>
<td>75-28-2</td>
</tr>
<tr>
<td></td>
<td>76-35-5</td>
</tr>
<tr>
<td></td>
<td>76-59-4</td>
</tr>
<tr>
<td></td>
<td>76-72-1</td>
</tr>
<tr>
<td></td>
<td>86-131-1</td>
</tr>
<tr>
<td></td>
<td>86-167-2</td>
</tr>
<tr>
<td>Brain - Parietal Lobe (PL)</td>
<td></td>
</tr>
<tr>
<td>5 envelopes</td>
<td>51-77-2</td>
</tr>
<tr>
<td></td>
<td>52-79-17</td>
</tr>
<tr>
<td></td>
<td>49-74-1</td>
</tr>
<tr>
<td></td>
<td>83-34-4</td>
</tr>
<tr>
<td></td>
<td>88-68-2</td>
</tr>
<tr>
<td>Colon (C)</td>
<td></td>
</tr>
<tr>
<td>11 envelopes</td>
<td>2-2-1</td>
</tr>
<tr>
<td></td>
<td>3-1-4</td>
</tr>
<tr>
<td></td>
<td>3-4-2</td>
</tr>
<tr>
<td></td>
<td>4-3-1</td>
</tr>
<tr>
<td></td>
<td>4-5-1</td>
</tr>
<tr>
<td></td>
<td>4-8-1</td>
</tr>
<tr>
<td></td>
<td>4-13-2</td>
</tr>
<tr>
<td></td>
<td>4-15-2</td>
</tr>
<tr>
<td></td>
<td>4-22-1</td>
</tr>
<tr>
<td></td>
<td>4-38-12</td>
</tr>
<tr>
<td></td>
<td>4-39-3</td>
</tr>
<tr>
<td>Lung (L)</td>
<td></td>
</tr>
<tr>
<td>29 envelopes</td>
<td>6-10-1</td>
</tr>
<tr>
<td></td>
<td>7-7-16</td>
</tr>
<tr>
<td></td>
<td>7-20-1</td>
</tr>
<tr>
<td></td>
<td>7-22-1</td>
</tr>
<tr>
<td></td>
<td>7-34-15</td>
</tr>
<tr>
<td></td>
<td>7-38-3</td>
</tr>
<tr>
<td></td>
<td>7-43-5</td>
</tr>
<tr>
<td></td>
<td>8-3-6</td>
</tr>
<tr>
<td></td>
<td>8-8-1</td>
</tr>
<tr>
<td></td>
<td>58-1-2</td>
</tr>
<tr>
<td></td>
<td>58-7-4</td>
</tr>
<tr>
<td></td>
<td>58-13-1</td>
</tr>
<tr>
<td></td>
<td>58-15-1</td>
</tr>
<tr>
<td></td>
<td>59-12-1</td>
</tr>
<tr>
<td></td>
<td>59-51-1</td>
</tr>
<tr>
<td></td>
<td>98-11-1</td>
</tr>
<tr>
<td></td>
<td>98-22-1</td>
</tr>
<tr>
<td></td>
<td>98-36-1</td>
</tr>
<tr>
<td></td>
<td>98-51-1</td>
</tr>
<tr>
<td></td>
<td>98-59-1</td>
</tr>
<tr>
<td></td>
<td>98-60-2</td>
</tr>
<tr>
<td></td>
<td>99-35-2</td>
</tr>
<tr>
<td></td>
<td>101-12-2</td>
</tr>
<tr>
<td></td>
<td>101-23-1</td>
</tr>
<tr>
<td></td>
<td>101-25-1</td>
</tr>
<tr>
<td></td>
<td>101-35-1</td>
</tr>
<tr>
<td></td>
<td>101-54-1</td>
</tr>
<tr>
<td></td>
<td>101-60-1</td>
</tr>
<tr>
<td></td>
<td>101-66-1</td>
</tr>
<tr>
<td>Lymph Node (LN)</td>
<td></td>
</tr>
<tr>
<td>12 envelopes</td>
<td>17-84-1</td>
</tr>
<tr>
<td></td>
<td>17-89-1</td>
</tr>
<tr>
<td></td>
<td>56-13-4</td>
</tr>
<tr>
<td></td>
<td>56-43-7</td>
</tr>
<tr>
<td></td>
<td>56-57-2</td>
</tr>
<tr>
<td></td>
<td>56-67-3</td>
</tr>
<tr>
<td></td>
<td>62-16-1</td>
</tr>
<tr>
<td></td>
<td>62-30-1</td>
</tr>
<tr>
<td></td>
<td>62-37-6</td>
</tr>
<tr>
<td></td>
<td>62-41-3</td>
</tr>
<tr>
<td></td>
<td>62-43-1</td>
</tr>
<tr>
<td></td>
<td>62-47-1</td>
</tr>
</tbody>
</table>

Table 3.1 **Subject 162 HIV-1 envelopes isolated from different tissues** HIV-1 *envs* were isolated from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node using nested PCR and DNA at limiting dilution.
isolated from frontal lobe, 10 from occipital lobe, 5 from parietal lobe, 11 from colon, 29 from lung, and 12 from lymph node for a total of 75 envelopes. Additional PCR reactions were performed to amplify smaller envelope regions (e.g., V1-V5 and V3) from plasma and serum proviral DNA, but attempts were not successful. However it was possible to amplify the human CCR5 gene (data not shown), suggesting that there was DNA in the samples, but proviral DNA was not present or had a very low copy number. Unfortunately, enriched white cells from blood were not available.

For plasma and serum, RT-PCR was used to amplify full length envelopes from viral RNA. However, no envelopes were isolated from either of these samples. As with the proviral DNA samples, subsequent attempts to PCR amplify smaller envelope regions from the plasma and serum were unsuccessful. Plasma and serum samples were analyzed for viral load by the Infectious Disease Laboratory at the Children’s Hospital of Chicago using Abbott RealTime HIV-1 RNA PCR. For both samples, viral load was undetectable with <40 copies/ml (Table 2.1). Subject 162 was on HAART (Table 2.1) and appears to have had well controlled viral replication at death.

3.2.2 No mutations were observed that have previously been associated with macrophage tropism or CXCR4 use

Sequences were examined for the presence of mutations associated with changes in coreceptor use or macrophage tropism. For example, the presence of either an arginine or lysine at residues 11 or 25 of the V3 loop is associated with
CXCR4 use, while the loss of an asparagine at N283, the glycan at N386 or an E153G substitution have been associated with changes in macrophage tropism (Peters et al, 2004; Dunfee et al, 2006; Dunfee et al, 2007; Duenas-Decamp et al, 2009; Musich et al, 2012). None of these were observed in any of the 75 sequences. There were some insertions/deletions observed, but they were not so large that they would clearly disrupt function. However many sequences, particularly in envs derived from the lung, had numerous premature stop codons. These may render the Env non-functional. All Envs were tested for functionality, discussed in Chapter IV.

There were many other polymorphisms that resulted in amino acid changes. None of these appeared to correlate exclusively with a particular tissue or phenotype, discussed in Chapter IV. For example, when neutralization by PG16 monoclonal antibody was tested, it was discovered that K130E was present in many envelopes that were resistant to neutralization. However there were also a number of envelopes with the K130E mutation that were sensitive to PG16.

3.2.3 Sequences were compartmentalized in brain tissue

Full length envelope gp160 sequences were isolated from frontal lobe, occipital lobe, parietal lobe, colon, lung, and lymph node (Table 3.1). The 53 functional sequences (assessed in Chapter IV) were used to construct a maximum likelihood phylogenetic tree (Figure 3.1) using Molecular Evolutionary Genetics
Figure 3.1 Phylogenetic analysis of Subject 162 gp160 sequences. Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 envs. Bootstrap values ≥70% are noted at branch points. Sequences from the brain regions are tightly compartmentalized. Compartmentalization was confirmed using the Slatkin-Maddison test with P < 0.001. Scale bar shows nucleotide substitutions per site.
Analysis (MEGA) version 5 software (Tamura et al, 2011). The tree was rooted using four unrelated clade B reference sequences. Genetic compartmentalization was observed between envelopes isolated from brain tissues and tissues from the body. Compartmentalization was confirmed using the Slatkin-Maddison test with P < 0.001.

Phylogenetic trees were also constructed for regions of the envelope: gp120, gp41, V1-V2, V1-V5, C1, C2, C3, C4, C5, V1, V2, V3, V4, and V5 (Appendix A). MEGA software was used to calculate mean genetic distances for each tree (Figure 3.2). The V1 and V5 regions had the greatest evolutionary divergence, while the C5 region had the least. The V3 loop was also highly conserved among tissues within the brain and in the periphery. This is consistent with previous observations, as it is known that all V regions except for the V3 loop have hypervariable sequences (Los Alamos National Laboratory, 2016).

3.2.4 Length of envelope regions did not differ significantly between different tissue sites

Functional envelopes were compared to determine if envelopes from different tissue compartments differed in the length of V1-V2 or V1-V5 (Figures 3.3 - 3.6). The length of the V1-V2 loops was more variable in peripheral tissues compared to those in brain. However, no significant differences were observed, when values were analyzed by unpaired, two-tailed t tests using GraphPad Prism software.
3.2.5 V1-V2 and V1-V5 net charges had greater variability in tissues from the body than in brain tissues

Net charges of V1-V2 and V1-V5 regions were calculated using the The European Molecular Biology Open Software Suite (EMBOSS) Pepstats program to count charged amino acids (Figures 3.7 - 3.10). Overall charges in V1-V2 and V1-V5 regions were more variable in envelopes derived from colon, lung, and lymph node compared to those in brain tissues. However, only a few envelopes from colon and lung carried a higher positive charge.

Using GraphPad Prism for unpaired, two-tailed t tests, it was determined that overall differences in V1-V2 and V1-V5 charges between envelopes present in brain tissues (frontal lobe, occipital lobe, and parietal lobe) and in tissues of the body (colon, lung, and lymph node) were not significant (Figures 3.8B and 3.10B). Nevertheless, significant differences were noted in V1-V2 charge between the following brain and body tissues, FL:OL, FL:C, FL:LN, OL:PL, PL:LN, as well as between C:L (Figure 3.7B). However, there was no consistent trend of higher charged envelopes in body tissues compared to brain tissues. Differences in the charges of the longer V1-V5 region were statistically significant between FL:OL, FL:LN, OL:PL, and PL:LN (Figure 3.9B). It was observed that there was much more variability in charges of envelope regions outside the brain than in the brain for both V1-V2 and V1-V5 regions, as shown by the means and standard deviations for each compartment (Figures 3.7C and 3.9C).
Figure 3.2 Evolutionary divergence of HIV-1 envs by region

MEGA version 5 software was used to calculate the mean distance (A) and evolutionary divergence (B) for each env region based on the number of base substitutions per site.
Figure 3.3 V1-V2 lengths of HIV-1 envs from different tissues

A) V1-V2 amino acid sequence lengths were calculated for each env and plotted using GraphPad Prism.

B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the env populations in each tissue compartment. An asterisk denotes that t test could not be performed because the samples in both tissues had the same values.

C) Mean and standard deviation are given to show the spread of the data range for the envs from each tissue compartment.

### Tissue Comparisons

<table>
<thead>
<tr>
<th>Tissue</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL:OL</td>
<td>0.4412</td>
</tr>
<tr>
<td>FL:PL</td>
<td>0.4730</td>
</tr>
<tr>
<td>FL:C</td>
<td>0.3998</td>
</tr>
<tr>
<td>FL:L</td>
<td>0.9003</td>
</tr>
<tr>
<td>OL:C</td>
<td>0.5283</td>
</tr>
<tr>
<td>OL:L</td>
<td>0.5935</td>
</tr>
<tr>
<td>C:L</td>
<td>0.9258</td>
</tr>
<tr>
<td>L:LN</td>
<td>0.4882</td>
</tr>
<tr>
<td>PL:LN</td>
<td>0.4899</td>
</tr>
</tbody>
</table>

### Mean and Standard Deviation

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>68.00</td>
<td>0</td>
</tr>
<tr>
<td>OL</td>
<td>68.00</td>
<td>0</td>
</tr>
<tr>
<td>PL</td>
<td>68.00</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>67.50</td>
<td>1.780</td>
</tr>
<tr>
<td>L</td>
<td>68.69</td>
<td>2.097</td>
</tr>
<tr>
<td>LN</td>
<td>68.09</td>
<td>1.868</td>
</tr>
</tbody>
</table>
Figure 3.4 V1-V2 lengths of HIV-1 *envs* from the brain and body

A) V1-V2 amino acid sequence lengths were calculated for each *env* in the brain and body and plotted using GraphPad Prism. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the *env* populations of the brain and body. C) Mean and standard deviation are given to show the spread of the data range for the *envs* from each compartment.
Figure 3.5 V1-V5 lengths of HIV-1 envs from different tissues A) V1-V5 amino acid sequence lengths were calculated for each env and plotted using GraphPad Prism. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the env populations in each tissue compartment. An asterisk denotes that t test could not be performed because the samples in both tissues had the same values. C) Mean and standard deviation are given to show the spread of the data range for the envs from each tissue compartment.
Figure 3.6 V1-V5 lengths of HIV-1* envs from the brain and body

A) V1-V5 amino acid sequence lengths were calculated for each *env* in the brain and body and plotted using GraphPad Prism.

B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the *env* populations of the brain and body.

C) Mean and standard deviation are given to show the spread of the data range for the *envs* from each compartment.

<table>
<thead>
<tr>
<th>Tissues</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain:Body</td>
<td>0.4034</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>339.0</td>
<td>0</td>
</tr>
<tr>
<td>Body</td>
<td>338.6</td>
<td>1.970</td>
</tr>
</tbody>
</table>
Figure 3.7 V1-V2 net charges of HIV-1 envs from different tissues  

A) V1-V2 amino acid net charges were calculated for each env and plotted using GraphPad Prism.  

B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the env populations in each tissue compartment. Statistically significant values are shown in red.  

C) Mean and standard deviation are given to show the spread of the data range for the envs from each tissue compartment.

<table>
<thead>
<tr>
<th>Tissues</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL:OL</td>
<td>0.0019</td>
</tr>
<tr>
<td>FL:PL</td>
<td>0.5059</td>
</tr>
<tr>
<td>FL:C</td>
<td>0.0290</td>
</tr>
<tr>
<td>FL:L</td>
<td>0.8296</td>
</tr>
<tr>
<td>FL:LN</td>
<td>0.0069</td>
</tr>
<tr>
<td>OL:PL</td>
<td>0.0016</td>
</tr>
<tr>
<td>OL:C</td>
<td>0.3124</td>
</tr>
<tr>
<td>OL:L</td>
<td>0.1796</td>
</tr>
<tr>
<td>OL:LN</td>
<td>0.4976</td>
</tr>
<tr>
<td>PL:C</td>
<td>0.0888</td>
</tr>
<tr>
<td>PL:L</td>
<td>&gt; 0.9999</td>
</tr>
<tr>
<td>PL:LN</td>
<td>0.0236</td>
</tr>
<tr>
<td>C:L</td>
<td>0.0349</td>
</tr>
<tr>
<td>C:LN</td>
<td>0.4809</td>
</tr>
<tr>
<td>L:LN</td>
<td>0.0519</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>-0.125</td>
<td>0.354</td>
</tr>
<tr>
<td>OL</td>
<td>-0.857</td>
<td>0.378</td>
</tr>
<tr>
<td>PL</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>-1.500</td>
<td>1.581</td>
</tr>
<tr>
<td>L</td>
<td>0</td>
<td>1.581</td>
</tr>
<tr>
<td>LN</td>
<td>-1.091</td>
<td>0.831</td>
</tr>
</tbody>
</table>
Figure 3.8 V1-V2 net charges of HIV-1 *envs* from the brain and body  

A) V1-V2 amino acid net charges were calculated for each *env* in the brain and body and plotted using GraphPad Prism.  

B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the *env* populations of the brain and body.  

C) Mean and standard deviation are given to show the spread of the data range for the *envs* from each compartment.
Figure 3.9 V1-V5 net charges of HIV-1 envs from different tissues A) V1-V5 amino acid net charges were calculated for each env and plotted using GraphPad Prism. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the env populations in each tissue compartment. Statistically significant values are shown in red. C) Mean and standard deviation are given to show the spread of the data range for the envs from each tissue compartment.
Figure 3.10 V1-V5 net charges of HIV-1 envs from the brain and body A) V1-V5 amino acid net charges were calculated for each env in the brain and body and plotted using GraphPad Prism. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the env populations of the brain and body. C) Mean and standard deviation are given to show the spread of the data range for the envs from each compartment.

<table>
<thead>
<tr>
<th>Tissues</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain:Body</td>
<td>0.2127</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>14.11</td>
<td>1.150</td>
</tr>
<tr>
<td>Body</td>
<td>13.44</td>
<td>2.120</td>
</tr>
</tbody>
</table>
3.2.6 Average number of predicted N-linked glycosylation sites was more variable in tissue compartments outside the brain

The low concentration of antibodies present inside the brain compartment may result in the evolution of envelopes with fewer glycans and an altered glycan shield. Potential N-linked glycosylation sites (PNGS) were predicted for functional envelope sequences using the Los Alamos National Laboratory N-Glycosite program for the V1-V5 region (Figures 3.11 and 3.12) and the full length envelope (Figures 3.13 and 3.14). Unpaired, two-tailed t tests were performed using GraphPad Prism. For the V1-V5 region, significant differences were found between FL:OL, OL:PL, and OL:LN (Figure 3.11B). For the envelope, significant differences were observed for FL:OL, FL:L, FL:LN, and OL:PL (Figure 3.13B). While regions outside of the brain had more variability (Figures 3.12C and 3.14C) in the number of PNGS than regions within the brain, these differences were not statistically significant (Figures 3.12B and 3.14B).

3.2.7 The lung compartment had a higher frequency of hypermutated envelopes than other tissues

All 75 envelope sequences were analyzed for hypermutation by Thomas Leitner at Los Alamos National Laboratory using the Hypermut program (Rose and Korber, 2000). A phylogenetic tree was constructed showing 19 envelope sequences in 5 lineages with hypermutation (Figure 3.15). Hypermutated sequences came from occipital lobe, colon, and lung. However the majority, 15 out of 19 envelopes, were from lung tissue. All of the envelopes with hypermutated
sequences were non-functional and unable to mediate infection of HeLa TZM-bl cells, as described in Chapter IV.

### 3.2.8 Recombination was extensive in all tissue compartments

Sequences were analyzed for potential recombination by Thomas Leitner at Los Alamos National Laboratory using a hierarchical PHI test (Bruen et al, 2006; Immonen and Leitner, 2014). Hypermutated sequences were not assessed for recombination and are removed from the counts below. Out of 56 remaining sequences, 39 were the product of recombination events (Figure 3.15). Of the non-hypermutated sequences, this includes 6 out of 8 frontal lobe sequences, 7 out of 7 occipital lobe sequences, 3 out of 5 parietal lobe sequences, 9 out of 10 colon sequences, 9 out of 14 lung sequences, and 5 out of 12 lymph node sequences. As discussed in Chapter IV, the majority of the recombinant sequences, 37 of the 39, were functional.

### 3.2.9 PacBio sequencing

A total of 32,152 sequences were generated from the parietal lobe, colon, lung, and lymph node of Subject 162 using PacBio sequencing (Table 3.2). Specifically, 13,791 were from the parietal lobe, 8,263 were from the colon, 66 were from the lung, and 10,032 were from the lymph node. PacBio sequencing was not performed for the frontal lobe or occipital lobe because it was not possible to amplify sufficient amounts of PCR product. For each sequence, the V3 loop was
analyzed to predict coreceptor use using WebPSSM (University of Washington).
All of the sequences were predicted to use CCR5 (Table 3.2).

To create phylogenetic trees, 250 random sequences were selected for each tissue. For lung, all 66 were used. A maximum likelihood tree was constructed, showing compartmentalization within the brain, as well as the colon (Figure 3.16). Sequences from the lung and lymph node were mixed.

3.3 Discussion

Subject 162 was an HIV-1 positive patient who died of complications from end stage AIDS. Despite this, it was difficult to amplify proviral DNA from patient tissue samples. Some tissues were more challenging than others, with plasma/serum and tissues of the brain being the most difficult. In the end, no envs were isolated from the plasma or serum. After many attempts to amplify viral RNA and proviral DNA, plasma and serum samples were sent out for viral load testing. It was discovered that neither sample had detectable viral loads. This would be expected in a subject that was on antiretroviral treatment with well controlled viremia. However this subject died from end stage AIDS, which is typically characterized by a spike in viral RNA copies in the plasma (Figure 1.2). It is unclear why a subject at this stage of HIV-1 infection would not have detectable viral loads. It is possible that the patient’s current drug regimen was effective at reducing the viral load in blood even though severe immune deficiencies led to their death.
Figure 3.11 V1-V5 potential N-linked glycosylation sites (PNGS) of HIV-1 *envs* from different tissues

A) V1-V5 PNGS were calculated for each *env* using N-Glycosite (LANL) and plotted using GraphPad Prism.

B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the *env* populations in each tissue compartment. Statistically significant values are shown in red.

C) Mean and standard deviation are given to show the spread of the data range for the *envs* from each tissue compartment.
Figure 3.12 V1-V5 potential N-linked glycosylation sites (PNGS) of HIV-1 envs from the brain and body 

A) V1-V5 PNGS were calculated for each env in the brain and body using N-Glycosite (LANL) and plotted using GraphPad Prism. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the env populations of the brain and body. C) Mean and standard deviation are given to show the spread of the data range for the envs from each compartment.
Figure 3.13 Env potential N-linked glycosylation sites of HIV-1 envs from different tissues  

A) Env PNGS were calculated for each env using N-Glycosite (LANL) and plotted using GraphPad Prism. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the env populations in each tissue compartment. Statistically significant values are shown in red. C) Mean and standard deviation are given to show the spread of the data range for the envs from each tissue compartment.
Figure 3.14 Env potential N-linked glycosylation sites of HIV-1 envs from the brain and body

A) Env PNGS were calculated for each env in the brain and body using N-Glycosite (LANL) and plotted using GraphPad Prism. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the env populations of the brain and body. C) Mean and standard deviation are given to show the spread of the data range for the envs from each compartment.
Figure 3.15 Hypermutation and recombination of gp160 sequences Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for all 75 HIV-1 envs. Bootstrap values ≥70% are noted at branch points. Hypermutation was assessed using the LANL Hypermut program and is shown by orange crosses (Rose and Korber, 2000). Sequences were analyzed for potential recombination using a hierarchical PHI test (Bruen et al, 2006; Immonen and Leitner, 2014). Recombinants are shown by purple diamonds. Scale bar shows nucleotide substitutions per site.
<table>
<thead>
<tr>
<th>Tissue</th>
<th>Total Sequences</th>
<th>CCR5-using</th>
<th>CXCR4-using</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
<td>Number</td>
</tr>
<tr>
<td>Parietal Lobe</td>
<td>13,791</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Colon</td>
<td>8,263</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Lung</td>
<td>66</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Lymph Node</td>
<td>10,032</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>32,152</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 3.2 Subject 162 PacBio sequencing results** The V3 loop nucleic acid sequences for all sequences in each data set were isolated using Geneious software version 9 (Kearse et al, 2012). Gene Cutter (Los Alamos National Laboratory) was used to codon align and translate sequences. The sequencing approach resulted in some gapped positions, so any V3 amino acid sequence containing an ambiguous position was removed from the alignment. Tropism prediction was performed using WebPSSM (Jensen et al, 2003) and the x4r5 matrix.
Figure 3.16 Phylogenetic analysis of Subject 162 PacBio sequences A python script was generated to randomly select 250 sequences from each tissue. Due to the small number of sequences generated, all lung sequences were used. Maximum likelihood phylogenies were generated with phyML (Guindon et al, 2010) using the HKY85 substitution model, the NNI tree improvement method, and the approximate likelihood ratio test (Anisimova and Gascuel, 2006). Tree was graphed and colored according to tissue type using FigTree software (http://tree.bio.ed.ac.uk/software/figtree/). Tree was midpoint rooted. Branch support is shown on the scale to the left, with red indicating branches with very high support.
Alternatively, plasma and serum samples may not have been handled optimally during tissue harvesting or shipment and RNA could have been degraded.

While it was also difficult to isolate envs from brain tissue, this was not surprising. It has previously been reported that viral loads are low in the brain tissues of neurologically normal subjects (Bell et al, 1993; Donaldson et al, 1994; McCrossan et al, 2006). Subject 162 did not have a diagnosis of any neurological disorders and therefore was considered neurologically normal. Regardless, 24 envs were isolated from brain tissues, along with 51 from other tissues (Table 3.1).

Tissues from the brain harbored envs that were compartmentalized from those in colon, lung, and lymph node. However, envs from these latter tissue sites were not. This compartmentalization was observed in full length env sequences, but not in all individual env regions (e.g., V3 loop). Evolutionary diversity was greatest in the V1 and V5 env regions (Figure 3.2), while other regions were more homogeneous. This is an important observation considering that many previous studies looked at only specific regions the of env sequences and in some cases concluded that there was no evidence of compartmentalization (Ince et al, 2009; Imamichi et al, 2011). For Subject 162, specific env regions including the V3 loop were nearly identical inside and outside the brain. This highlights the importance of examining full length env for studies of compartmentalization and viral diversity.

Sequence analysis revealed many mutations in env sequences, though none have been previously attributed to changes in macrophage tropism. An asparagine at position 283 and loss of an N-linked glycosylation site at 386 have
been shown to be associated with macrophage tropic brain variants in subjects with HIV-Associated Dementia (HAD) (Dunfee et al, 2006; Dunfee et al, 2007). An E153G mutation has also been shown to be a determinant of macrophage tropism in a pediatric subject with neurological complications (Musich et al, 2012). However none of the sequences had any of these mutations.

Few statistically significant differences were observed for the means of env length, charge, or number of PNGS, as determined using unpaired, two-tailed t tests. However, there was a clear difference in the variation of the range of values for env length, charge, and number of PNGS. In all cases, values for brain regions (frontal lobe, occipital lobe, and parietal lobe) were limited to a very narrow set of values. Interestingly, OL envs had a PNGS that was not present in FL or PL envs, as well as a lower net charge.

In contrast, env length, charge, and number of PNGS from colon, lung, and lymph node were much more varied. Several studies on V1-V2 length and PNGS have suggested that shorter V1-V2 regions and fewer PNGS are characteristic of transmitted variants and early infection (Pollakis et al, 2001; Derdeyn et al, 2004; Chohan et al, 2005; Sagar et al, 2006; Liu et al, 2008; Curlin et al, 2010; Huang et al, 2012; Wang et al, 2012; Pollakis et al, 2015). One hypothesis is that this could be due to fewer immune pressures early in infection, and thus less need for variation and protection by PNGS. However this does not explain how these variants would be present in an individual later during the course of infection such that they could be transmitted to others. In addition, brain envs did appear to be
slightly shorter and have fewer PNGS, though this was not statistically significant. This is consistent with previous findings, as it is likely brain *envs* would also not be subjected to as severe immune pressures as *envs* in tissues of the body would.

Recombination was widespread in all tissues types and did not affect the function of the *Envs*, as discussed in Chapter IV (Figure 3.15). On the other hand, predicted rates of hypermutation were varied between different tissue types (Figure 3.15). It was noted during sequence analysis that many of the sequences contained premature stop codons. Upon further inspection, it was discovered that these were almost entirely G to A mutations, a signature of APOBEC activity (reviewed in Refsland and Harris, 2013). When these sequences were analyzed using the Hypermut program, it was confirmed that these mutations were consistent with APOBEC activity. Three tissues had hypermutation: the occipital lobe, colon, and lung. The majority, 15 out of 19 hypermutated sequences, were from lung. Interestingly, studies have shown that the lung has high levels of APOBEC expression (Koning et al, 2009; Refsland et al, 2010; Fourati et al, 2014). However other sites, including immune tissues, have also been shown to express APOBEC but did not have a high level of hypermutation in this study.

PacBio sequencing data provided a first look at a large number of *env* sequences from multiple tissue sites. However, this data is preliminary and thus limited conclusions can be drawn from it. A major obstacle of this study was that no established protocol for data analysis existed prior to beginning this work.
Collaborators continue to develop a pipeline for analysis and this project remains in progress.

Based on preliminary data, the phylogenetic tree from PacBio sequences showed similar compartmentalization as was observed using PCR amplified clones (Figures 3.1 and 3.16). One difference was that colon sequences appeared much more compartmentalized on the tree of PacBio sequences. It is not clear why this would be, but may be the result of low proviral load in the tissues sampled. Thus only a small number of variants were sampled for PacBio sequencing, resulting in many of the same sequences. It is also possible that there were just too few PCR clones to see compartmentalization in the colon. Several small groups of colon sequences did cluster together on the PCR clone tree, but only 11 envs were examined, in comparison to 250 on the PacBio sequence tree. Compartmentalization may have become more clear if additional PCR clone sequences were added to the tree. It is not known why so few sequences were generated from the lung tissue. PacBio sequencing may be repeated for this tissue.

In summary, it is clear that envs outside of the brain are much more diverse than envs inside the brain. Though envs from the body were not genetically compartmentalized from one another, they did have differences in length, charge, number of PNGS, and rates of hypermutation. This may indicate that different tissue microenvironments result in unique viral variants, which could be explored further in future studies.
Envelopes from within the brain tissues were in an immune privileged site, with fewer permissive cell types. Accordingly, *envs* in the brain faced fewer and less diverse evolutionary pressures, resulting in more constrained sequences. In Chapter IV these Envs were tested for phenotypes to determine if they were similarly influenced by evolution in distinct tissue compartments.
CHAPTER IV: FUNCTIONAL CHARACTERIZATION OF HIV-1 ENVELOPES FROM TISSUES OF AN HIV-1 INFECTED SUBJECT

4.1 Introduction

In the first part of this study, discussed in Chapter III, 75 full length envs were amplified from six different tissue compartments and compared for genotypic variation. In this second part, these Envs were expressed on pseudoviruses and subjected to functional assays to compare phenotypic variation. Similar to previous genotypic studies, phenotypic studies have been limited to few tissue types and thus do not adequately address quasispecies variation within an infected individual. Here, Envs were characterized by their functionality, cell tropism, coreceptor use, and susceptibility to neutralization by inhibitors and monoclonal antibodies.

Coreceptor use and tropism have been previously studied in some tissues, such as the brain, immune tissues, and plasma, as reviewed in Chapter I. These studies have shown that T cell tropic R5-using variants are typically transmitted, but may evolve to use X4 or infect macrophages as disease progresses (Keele et al, 2008; Salazar-Gonzalez et al, 2009; Wilen et al, 2011; Ochsenbauer et al, 2012). Previous studies have also shown that Envs from the brain were macrophage tropic and those outside the brain were predominantly, but not always, non-macrophage tropic (Gorry et al, 2001; Peters et al, 2004; Gonzalez-Perez et al, 2012).
To further characterize envelope phenotypes and gain insight into possible structural differences between envelopes from different tissues, pseudoviruses were subjected to inhibition assays using sCD4 and maraviroc, and neutralization assays with monoclonal antibodies b12, PG16, 17b, and 447-52D. These reagents were selected to provide information on whether the trimer was closed or open, the V3 loop was exposed or occluded, and to further probe the extent that the CD4 binding site was exposed.

These functional studies provide new insight into phenotypic differences in quasispecies from various tissue compartments. The results of this study will aid in the understanding of how viral variants may evolve differently due to distinct tissue specific pressures.

4.2 Results

4.2.1 Most envelopes were able to infect HeLa TZM-bl cells

All envelopes isolated were tested for functionality based on the ability of Env+ pseudoviruses to infect HeLa TZM-bl cells (Figures 4.1 – 4.4). Infectivity was evaluated by titrating Env+ pseudoviruses and counting FFUs. Infectivity titers of $>10^3$ FFU/ml were considered sufficiently functional for further study, explained further in Chapter II. For brain, 100% of frontal lobe envelopes (8/8), 70% of occipital lobe envelopes (7/10), and 80% of parietal lobe envelopes (4/5) were functional. Outside brain, 91% of colon envelopes (10/11), 45% of lung envelopes (13/29), and 92% of lymph node envelopes (11/12) were also functional.
Figure 4.1 TZM-bl and macrophage titers of Envs from the brain

Envs derived from the brain tissues (frontal lobe, occipital lobe, and parietal lobe) of Subject 162 were made into pseudoviruses and tested for their ability to infect TZM-bl cells and macrophages. Laboratory strains B33, JRFL, and JRCSF were used as controls. B33 and JRFL were functional and macrophage tropic controls. JRCSF was a functional and non-macrophage tropic control. The L411 env- plasmid and L411 env- plasmid + pSVIII plasmid were used as negative controls. The red line shows the cut off titer value for an Env to be considered functional on TZM-bl cells or macrophage tropic.
Figure 4.2 TZM-bl and macrophage titers of Envs from the colon Envs derived from the colon tissue of Subject 162 were made into pseudoviruses and tested for their ability to infect TZM-bl cells and macrophages. Laboratory strains B33, JRFL, and JRCSF were used as controls. B33 and JRFL were functional and macrophage tropic controls. JRCSF was a functional and non-macrophage tropic control. The L411 env- plasmid and L411 env- plasmid + pSVIII plasmid were used as negative controls. The red line shows the cut off titer value for an Env to be considered functional on TZM-bl cells or macrophage tropic.
Figure 4.3 TZM-bl and macrophage titers of Envs from the lung Envs derived from the lung tissue of Subject 162 were made into pseudoviruses and tested for their ability to infect TZM-bl cells and macrophages. Laboratory strains B33, JRFL, and JRCSF were used as controls. B33 and JRFL were functional and macrophage tropic controls. JRCSF was a functional and non-macrophage tropic control. The L411 env- plasmid and L411 env- plasmid + pSVIII plasmid were used as negative controls. The red line shows the cut off titer value for an Env to be considered functional on TZM-bl cells or macrophage tropic.
Figure 4.4 TZM-bl and macrophage titers of Envs from the lymph node
Envs derived from the lymph node tissue of Subject 162 were made into pseudoviruses and tested for their ability to infect TZM-bl cells and macrophages. Laboratory strains B33, JRFL, and JRCSF were used as controls. B33 and JRFL were functional and macrophage tropic controls. JRCSF was a functional and non-macrophage tropic control. The L411 env-plasmid and L411 env- plasmid + pSVIII plasmid were used as negative controls. The red line shows the cut off titer value for an Env to be considered functional on TZM-bl cells or macrophage tropic.
4.2.2 All envelopes amplified used CCR5 and could not use CXCR4

Coreceptor use was predicted for all envelopes using the WebPSSM program (University of Washington). This program utilizes an algorithm that predicts coreceptor use from V3 loop sequences (Jensen et al, 2003). All envelopes were predicted to use CCR5 as a coreceptor. All shared similar charges and carried amino acids S and A at positions 11 and 25, respectively.

Functional pseudoviruses were able to infect HeLa TZM-bl cells, which express both CCR5 and CXCR4 coreceptors (Figures 4.1 – 4.4). However none of the functional pseudoviruses were able to infect HeLa HIJ cells which express CD4 and CXCR4 but not CCR5 (not shown). This is consistent with the WebPSSM prediction that all envelopes use CCR5 as a coreceptor and rules out the possibility that they could have been dual-tropic.

4.2.3 Brain derived envelopes were macrophage tropic while envelopes from tissues from the body were not

All pseudoviruses were tested for macrophage tropism on donor macrophages (Figures 4.1 – 4.4). All pseudoviruses constructed with envelopes derived from brain tissues were able to infect macrophages, with the exception of one occipital lobe envelope: 86-167-2. Conversely, only one pseudovirus with an envelope from tissue outside of the brain, 8-3-6 from the lung, was able to infect macrophages. All other envelopes from colon, lung, and lymph node were non-macrophage tropic.
4.2.4 Sensitivity to entry inhibitors and neutralizing monoclonal antibodies

sCD4

Consistent with previous studies, pseudoviruses prepared from envelopes from all three regions of the brain were inhibited at very low concentrations of sCD4 (Figures 4.5 and 4.6). In contrast, Envs from colon, lung, and lymph node varied in their sensitivity to sCD4, with colon Envs significantly more sensitive compared to those from lung and lymph node.

The difference in sensitivity to sCD4 between tissues from the brain (frontal lobe, occipital lobe, and parietal lobe) and the body (colon, lung, and lymph node) was also found to be highly significant ($P < 0.0001$) (Figure 4.6B). Highly significant differences were also noted for sCD4 sensitivity between FL:C, FL:L, FL:LN, OL:C, OL:L, OL:LN, PL:L, PL:LN, C:L, and C:LN (Figure 4.5B). Together these data are consistent with distinct selective pressures in brain and colon compared to lung and lymph node that influence the exposure of the CD4bs or the capacity of Envs to respond to CD4 binding.

Maraviroc

Maraviroc is an entry inhibitor that acts as a CCR5 antagonist and strongly inhibits entry of CCR5-using viral variants (Dorr et al, 2005; Wood and Armour, 2005). Maraviroc strongly inhibited most of the pseudoviruses (Figures 4.7 and 4.8). Envs from the lymph node were inhibited with IC50s ranging from 3.81-50.00 ng/ml. One colon Env, 4-38-12, was highly resistant with an IC50 >100 ng/ml. All other pseudoviruses were inhibited at low concentrations of maraviroc, with IC50s
between 2.35 and 21.88 ng/ml. These results do not indicate dramatic tissue dependent variation in CCR5 use. A significant difference was observed only between L:LN using an unpaired, two-tailed t test (Figure 4.7B).

**b12**

The b12 monoclonal antibody epitope maps to a region overlapping the CD4 binding site on gp120 (Burton et al, 1991; Barbas et al, 1992; Burton et al, 1994; Roben et al, 1994). Previous studies have shown that the majority of clade B macrophage tropic CCR5-using envelopes are sensitive to b12, while only a fraction of non-macrophage tropic envelopes are (Peters et al. 2008; Dunfee et al, 2009; O’Connell et al, 2013). For many of these latter envelopes, the b12 epitope is occluded within the trimer. Here, most Env+ pseudoviruses with envelopes from brain and other tissues were resistant or had IC50s in a narrow range of b12 concentrations of 35 - 50 µg/ml (Figures 4.9 and 4.10). However, several envelopes from lung and lymph node, and one from frontal lobe, were highly sensitive. Using an unpaired, two-tailed t test, a significant difference was observed between C:LN, but not with any other tissues or between the brain and body (Figures 4.9B and 4.10B). These data are not consistent with previous studies that mainly used Envs derived from subjects with dementia (Peters et al, 2008; Dunfee et al, 2009; Gonzalez-Perez et al, 2012).

**PG16**

The PG16 antibody recognizes a glycosylation and trimer dependent conformational epitope at the apex of the trimer association domain (Walker et al,
PG16 neutralized all pseudoviruses with brain envelopes at low concentrations (~0.01 μg/ml). Pseudoviruses with envelopes from other tissues were either strongly neutralized or strongly resistant (Figures 4.11 and 4.12). The resistance of these latter envelopes was not due to the absence of either of the critical N-linked glycosylation sites at N156 and N160 (McLellan et al, 2011; Julien et al, 2013; Wang et al, 2013).

A statistically significant difference was observed between IC50s of tissues from the brain (frontal lobe, occipital lobe, and parietal lobe) and tissues from the body (colon, lung, and lymph node) using an unpaired two-tailed t test in GraphPad Prism (Figure 4.12B). Specific tissues were also compared and it was found that only frontal lobe and colon had a statistically significant difference (Figure 4.11B). However, differences between other tissues had P values that approached significance (~0.05).

17b

The 17b antibody recognizes a CD4-induced epitope that overlaps the coreceptor binding site (Thali et al, 1993; Wyatt et al, 1995; Trkola et al, 1996; Kwong et al, 1998; Sullivan et al, 1998; Wyatt et al, 1998). 17b neutralizes several T cell line adapted laboratory isolates but does not usually neutralize primary HIV-1 strains. Almost all pseudoviruses were resistant to 17b neutralization with IC50s >50 μg/ml, the highest concentration used in this experiment (Figure 4.13 and 4.14). Only 58-13-1 from the lung and 62-30-1 from the lymph node showed some sensitivity, but with high IC50s at 33.17 μg/ml and 38.14 μg/ml, respectively.
The 447-52D antibody recognizes the KSIHIGPGRAF epitope of the V3 region of the HIV-1 envelope with the GPGR motif especially critical (Gorny et al, 1992; Gorny et al, 1993; Conley et al, 1994; Zolla-Pazner et al, 1995; Gorny et al, 1997; Binley et al, 2004; Stanfield et al, 2004; Krachmarov et al, 2006; Burke et al, 2009; Hioe et al, 2010). Sequence analysis showed that all envs had the 447-52D epitope. Despite this, only one pseudovirus was inhibited: 8-3-6 from the lung. All other pseudoviruses were strongly resistant to neutralization by 447-52D, even at the highest concentration of 50 μg/ml (Figures 4.15 and 4.16). There was no difference between envelopes from different tissue compartments. These data indicate that the V3 loop crown was occluded within the trimer association domain for all envelopes except 8-3-6 from lung.

4.3 Discussion

In this second part of the study, HIV-1 envs that were isolated and genotypically characterized in Chapter III were prepared as Env+ pseudoviruses to test their phenotypes. Once pseudoviruses were made, they were first tested on TZM-bl cells to determine whether or not they were functional. Due to the high rate of mutation in HIV-1, some envs acquire mutations that render the Env glycoprotein non-infectious. TZM-bl cells are a permissive cell line that expresses
Figure 4.5 sCD4 inhibition of HIV-1 Env from different tissues A) Each Env was tested for sensitivity to inhibition by sCD4. IC50s were calculated and plotted using GraphPad Prism. Laboratory strains JRFL and JRCSF were used as sCD4 sensitive controls. Vesicular Stomatitis Virus G protein (VSVG) was used as a sCD4 resistant control. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations in each tissue compartment. Statistically significant values are shown in red. C) Mean and standard deviation are given to show the spread of the data range for the Envs from each tissue compartment.
Figure 4.6 sCD4 inhibition of HIV-1 Envs from the brain and body

A) Each Env in the brain and body was tested for sensitivity to inhibition by sCD4. IC50s were calculated and plotted using GraphPad Prism. Laboratory strains JRFL and JRCSF were used as sCD4 sensitive controls. Vesicular Stomatitis Virus G protein (VSVG) was used as a sCD4 resistant control. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations of the brain and body. Statistically significant values are shown in red. C) Mean and standard deviation are given to show the spread of the data range for the Envs from each compartment.
Figure 4.7 Maraviroc inhibition of HIV-1 Env from different tissues

A) Each Env was tested for sensitivity to inhibition by maraviroc. IC50s were calculated and plotted using GraphPad Prism. Laboratory strains JRFL and JRCSF were used as maraviroc sensitive controls. Vesicular Stomatitis Virus G protein (VSVG) was used as a maraviroc resistant control. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations in each tissue compartment. Statistically significant values are shown in red. C) Mean and standard deviation are given to show the spread of the data range for the Envs from each tissue compartment.
Figure 4.8 Maraviroc inhibition of HIV-1 Envs from the brain and body

A) Each Env in the brain and body was tested for sensitivity to inhibition by maraviroc. IC50s were calculated and plotted using GraphPad Prism. Laboratory strains JRFL and JRCSF were used as maraviroc sensitive controls. Vesicular Stomatitis Virus G protein (VSVG) was used as a maraviroc resistant control. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations of the brain and body. C) Mean and standard deviation are given to show the spread of the data range for the Envs from each compartment.

<table>
<thead>
<tr>
<th>Tissues</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain:Body</td>
<td>0.0884</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>10.42</td>
<td>5.094</td>
</tr>
<tr>
<td>Body</td>
<td>17.96</td>
<td>18.460</td>
</tr>
</tbody>
</table>
Figure 4.9 b12 neutralization of HIV-1 Envs from different tissues  

**A)** Each Env was tested for sensitivity to neutralization by b12. IC50s were calculated and plotted using GraphPad Prism. Laboratory strains JRFL and JRCSF were used as b12 sensitive controls. Vesicular Stomatitis Virus G protein (VSVG) was used as a b12 resistant control. **B)** An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations in each tissue compartment. Statistically significant values are shown in red. **C)** Mean and standard deviation are given to show the spread of the data range for the Envs from each tissue compartment.

<table>
<thead>
<tr>
<th>Tissues</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL:OL</td>
<td>0.3865</td>
</tr>
<tr>
<td>FL:PL</td>
<td>0.6000</td>
</tr>
<tr>
<td>FL:C</td>
<td>0.2665</td>
</tr>
<tr>
<td>FL:L</td>
<td>0.7165</td>
</tr>
<tr>
<td>FL:LN</td>
<td>0.2295</td>
</tr>
<tr>
<td>OL:PL</td>
<td>0.7075</td>
</tr>
<tr>
<td>OL:C</td>
<td>0.7923</td>
</tr>
<tr>
<td>OL:L</td>
<td>0.2723</td>
</tr>
<tr>
<td>OL:LN</td>
<td>0.0593</td>
</tr>
<tr>
<td>PL:C</td>
<td>0.5895</td>
</tr>
<tr>
<td>PL:L</td>
<td>0.4659</td>
</tr>
<tr>
<td>PL:LN</td>
<td>0.1786</td>
</tr>
<tr>
<td>C:L</td>
<td>0.1684</td>
</tr>
<tr>
<td>C:LN</td>
<td>0.0220</td>
</tr>
<tr>
<td>L:LN</td>
<td>0.3400</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>42.15</td>
<td>17.210</td>
</tr>
<tr>
<td>OL</td>
<td>48.09</td>
<td>3.264</td>
</tr>
<tr>
<td>PL</td>
<td>47.04</td>
<td>5.915</td>
</tr>
<tr>
<td>C</td>
<td>48.62</td>
<td>4.370</td>
</tr>
<tr>
<td>L</td>
<td>38.87</td>
<td>21.150</td>
</tr>
<tr>
<td>LN</td>
<td>30.05</td>
<td>23.110</td>
</tr>
</tbody>
</table>
Figure 4.10 b12 neutralization of HIV-1 Envs from the brain and body

A) Each Env in the brain and body was tested for sensitivity to neutralization by b12. IC50s were calculated and plotted using GraphPad Prism. Laboratory strains JRFL and JRC5F were used as b12 sensitive controls. Vesicular Stomatitis Virus G protein (VSVG) was used as a b12 resistant control. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations of the brain and body. C) Mean and standard deviation are given to show the spread of the data range for the Envs from each compartment.
Figure 4.11 PG16 neutralization of HIV-1 Envs from different tissues

A) Each Env was tested for sensitivity to neutralization by PG16. IC50s were calculated and plotted using GraphPad Prism. Laboratory strain JRCSF was used as a PG16 sensitive control. Laboratory strain JRFL and Vesicular Stomatitis Virus G protein (VSVG) were used as PG16 resistant controls. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations in each tissue compartment. Statistically significant values are shown in red. An asterisk denotes that t test could not be performed because the samples in both tissues had the same values. C) Mean and standard deviation are given to show the spread of the data range for the Envs from each tissue compartment.
Figure 4.12 PG16 neutralization of HIV-1 Envs from the brain and body A) Each Env in the brain and body was tested for sensitivity to neutralization by PG16. IC50s were calculated and plotted using GraphPad Prism. Laboratory strain JRCSF was used as a PG16 sensitive control. Laboratory strain JRFL and Vesicular Stomatitis Virus G prote in (VSVG) were used as PG16 resistant controls. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations of the brain and body. Statistically significant values are shown in red. C) Mean and standard deviation are given to show the spread of the data range for the Envs from each compartment.
Figure 4.13 17b neutralization of HIV-1 Envs from different tissues A) Each Env was tested for sensitivity to neutralization by 17b. IC50s were calculated and plotted using GraphPad Prism. Laboratory strain NL4.3 was used as a 17b sensitive control. Laboratory strains JRFL and JRCSF and Vesicular Stomatitis Virus G protein (VSVG) were used as 17b resistant controls. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations in each tissue compartment. An asterisk denotes that t test could not be performed because the samples in both tissues had the same values. C) Mean and standard deviation are given to show the spread of the data range for the Envs from each tissue compartment.
Figure 4.14 17b neutralization of HIV-1 Envs from the brain and body

A) Each Env in the brain and body was tested for sensitivity to neutralization by 17b. IC50s were calculated and plotted using GraphPad Prism. Laboratory strain NL4.3 was used as a 17b sensitive control. Laboratory strains JRFL and JRCSF and Vesicular Stomatitis Virus G protein (VSVG) were used as 17b resistant controls.

B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations of the brain and body.

C) Mean and standard deviation are given to show the spread of the data range for the Envs from each compartment.
Figure 4.15 447-52D neutralization of HIV-1 Envs from different tissues

A) Each Env was tested for sensitivity to neutralization by 447-52D. IC50s were calculated and plotted using GraphPad Prism. Laboratory strain NL4.3 was used as a 447-52D sensitive control. Laboratory strains JRFL and JRCSF and Vesicular Stomatitis Virus G protein (VSVG) were used as 447-52D resistant controls. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations in each tissue compartment. An asterisk denotes that t test could not be performed because the samples in both tissues had the same values. C) Mean and standard deviation are given to show the spread of the data range for the Envs from each tissue compartment.
Figure 4.16 447-52D neutralization of HIV-1 Envs from the brain and body

A) Each Env in the brain and body was tested for sensitivity to neutralization by 447-52D. IC50s were calculated and plotted using GraphPad Prism. Laboratory strain NL4.3 was used as a 447-52D sensitive control. Laboratory strains JRFL and JRCSF and Vesicular Stomatitis Virus G protein (VSVG) were used as 447-52D resistant controls. B) An unpaired, two-tailed t test was used to test for statistical difference in the mean between the Env populations of the brain and body. C) Mean and standard deviation are given to show the spread of the data range for the Envs from each compartment.
the CD4 receptor, as well as both the CCR5 and CXCR4 coreceptors, so pseudoviruses with functional Envs should be able to infect them. Most of the Envs were functional, as assessed by their ability to infect TZM-bl cells (Figures 4.1 - 4.4). However 22 out of 75 were non-functional, with a disproportionate number coming from the lung. A total of 29 envelopes were amplified from the lung and 16, or 55%, were non-functional. Only 6 of the non-functional Envs were from other tissues. When lung env sequences were examined, as described in Chapter III, it was clear why they were non-functional. The envs derived from lung tissue had far higher rates of hypermutation than envs from other tissues. In fact, all but one of the non-functional envelopes from lung were the result of hypermutation, probably due to APOBEC activity. The remaining non-functional envelopes were non-functional due to hypermutation along with premature stop codons.

Of the functional Envs, all of them used CCR5 as a coreceptor. This was not surprising since this was predicted based on sequence using the WebPSSM program (Jensen et al, 2003). Envs from brain tissues were able to infect macrophages, consistent with previous findings (Gorry et al, 2001; Peters et al, 2004; Gonzalez-Perez et al, 2012). This was not due to the presence of known determinants, such as N283, loss of an N-linked glycosylation site at 386, or E153G. Only one Env from the colon, lung, or lymph node was macrophage tropic, 8-3-6 from the lung, although several others also conferred low level infection. Previous studies have described macrophage tropic HIV-1 variants in the lung and suggested that the lung may serve as a reservoir (Clarke et al, 1990; Jambo et al,
The 8-3-6 env did not have the same mutations observed in brain derived macrophage tropic envs and did not cluster with these sequences on the phylogenetic tree (Figure 3.1). It is therefore likely that macrophage tropism evolved independently in the brain and in the lung compartments.

The functional Envs were also tested for their sensitivity to sCD4 and maraviroc inhibitors, as well as a panel of monoclonal antibodies. It has been shown previously that sCD4 inhibition is correlated with macrophage tropism and suggests a trimer conformation with increased exposure of the CD4bs, although an enhanced interaction with CD4 is also possible (Peters et al, 2008; Gonzalez-Perez et al, 2012; O’Connell et al, 2013). Subject 162 macrophage tropic Envs were sensitive to sCD4 inhibition, consistent with these findings. However, there were also a number of non-macrophage tropic Envs from the colon, lung, and lymph node that were also sensitive to sCD4. In these tissues a range of IC50 values were observed, whereas with Envs from brain regions, inhibition occurred at a very low concentration of sCD4 for all Envs. An enhanced Env:CD4 interaction for Env trimers is presumably required for infection of resident macrophages of the brain, which express lower levels of CD4 compared to T cells (Lee et al, 1999; Bannert et al, 2000; Mori et al, 2000). In addition, Envs would also not need to be as tightly closed to protect from neutralizing antibodies as those outside brain, since the concentration of antibodies in the brain tissues is low (Triguero et al, 1989; Kuang et al, 2004). It’s not known why a range of resistant and sensitive phenotypes were present in the colon, lung, and lymph node tissues. Env trimers
in these tissues are more susceptible to immune pressures, so it is logical that they would remain in a more closed conformation to protect the CD4bs. It is possible that in this end stage AIDS subject, the immune system was failing and the resulting reduced immune pressure allowed for more open trimers to emerge and possibly the appearance of macrophage tropic variants. Macrophage tropic variants often do arise late in disease progression (Li et al, 1999; Tuttle et al, 2002; Gray et al, 2005; Thomas et al, 2007). Perhaps these more open, but not macrophage tropic, variants represent an intermediate phenotype in these tissue compartments, as recently suggested by work from the Swanstrom Lab (Arrildt et al, 2015).

Almost all Envs were inhibited by maraviroc, which is a strong inhibitor of R5-using HIV-1 variants. All of the Envs from this subject were shown to use the R5 coreceptor. Only one Env, 4-38-12 from the colon, was highly resistant. However, several from lymph node were modestly resistant, consistent with an enhanced Env:CCR5 interaction late in disease (Gorry et al, 2002; Reeves et al, 2002; Karlsson et al, 2004; Reeves et al, 2004; Gray et al, 2005; Repits et al, 2005; Repits et al, 2008; Etemad et al, 2009). Envs that can use the X4 coreceptor are resistant to maraviroc. However 4-38-12 was shown to use R5 both based on envelope sequence predictions and its inability to infect HeLa HIJ cells via CXCR4. It is possible that this Env can use an alternative coreceptor for entry that was not tested. Alternatively, some CCR5-using Envs are resistant to maraviroc. This may be due to enhanced affinity for CCR5 or the ability to use maraviroc-occupied

The b12 antibody overlaps the CD4 binding site and sensitivity has been associated with R5 macrophage tropism (Peters et al, 2008; Dunfee et al, 2009; O'Connell et al, 2013). In Subject 162, this was not the case. The majority of Envs were resistant to neutralization by b12 regardless of whether or not they were macrophage tropic. A few highly sensitive Envs were observed from the frontal lobe, lung, and lymph node. Neutralization by PG16 is dependent on Env glycosylation and conformation. PNGS at N156 and N160 have been shown to play a role in PG16 recognition (McLellan et al, 2011; Julien et al, 2013; Wang et al, 2013). All Envs had these PNGS and most were highly sensitive to PG16 neutralization. The frontal lobe, occipital lobe, and parietal lobe Envs were all neutralized at very low PG16 concentrations. Envs from the colon, lung, and lymph node compartments were either very resistant or very sensitive to neutralization, without any intermediate phenotypes. While no known mutations associated with PG16 resistance were observed, resistant envelopes had a K130E mutation. However several envelopes that were sensitive to PG16 also had this mutation, so this alone cannot be sufficient to confer resistance.

Using the final two antibodies, 17b and 447-52D, no differences were observed between different tissues compartments. All Envs were resistant to neutralization with 17b, which recognizes a CD4-induced epitope. Similarly, all Envs were resistant to neutralization by 447-52D, which recognizes a
KSIHIGPRAF epitope of the V3 loop. While all envelopes had this sequence, none of them were sensitive, indicating that the epitope was occluded within the structure. Only one Env, 8-3-6 from the lung, showed moderate sensitivity to neutralization by 447-52D. Interestingly, this Env is also macrophage tropic and sensitive to inhibition with sCD4. This suggests that this Env trimer has a more open conformation where the V3 loop and CD4bs are more exposed, thus helping to explain its ability to use low levels of CD4 to infect macrophages as well as its susceptibility to sCD4.

Chapter III investigated the genotypic differences in envs derived from different tissues of an HIV-1 positive individual. It was observed that envs from tissues within the brain were very similar in their sequences and properties, while envs from other regions of the body had much more varied characteristics. Here, these same Envs were examined for their phenotypes and a similar outcome was observed. Envs from regions of the brain behaved almost identically in all functional tests, while those from tissues of the body had a wide range of phenotypes. This again highlights the differences in these tissues and the evolutionary pressures that contribute to Env evolution. Envs in the brain must adapt to use low CD4 on the resident macrophages of the brain. However, they may be under less pressure from interaction with antibodies. Conversely, Envs in the body must evolve to avoid antibody neutralization but have access to several permissive cell types, allowing for more diversity in phenotypes. Studying these
different functional characteristics will aid in the understanding of viral quasispecies populations.
CHAPTER V: GENERAL DISCUSSION

5.1 Summary

In this study, full length HIV-1 envelopes were characterized in three different ways: by genotypic properties, through functional phenotypes, and using next generation sequencing. These envs were derived from the frontal lobe, occipital lobe, parietal lobe, colon, lung, and lymph node of an HIV-1 infected subject who died from end stage AIDS. This is the first study in which envelopes from an individual patient have been so extensively characterized. In addition, this is the first study to use Pacific Biosciences next generation sequencing to study full length proviral HIV-1 envs.

The first part of this study, discussed in Chapter III, used limiting dilution nested PCR to amplify full length HIV-1 envs from each tissue. These envs were examined for mutations known to affect tropism, length, charge, PNGS, hypermutation, and recombination. The env sequences were assembled into a phylogenetic tree and assessed for compartmentalization. The envs isolated from brain tissues (frontal lobe, occipital lobe, and parietal lobe) were very homogeneous in all areas. However, envs from tissues of the body (colon, lung, and lymph node) were very diverse in their genotypes.

The same outcome was observed in the second part of this study, discussed in Chapter IV. The Envs from each tissue were incorporated into pseudoviruses to examine functionality, coreceptor use, macrophage tropism, and
sensitivity to sCD4 and maraviroc inhibitors and neutralization by b12, PG16, 17b, and 447-52D monoclonal antibodies. Envs derived from tissues of the brain were similar in all phenotypes tested, while Envs from tissues of the body were more varied in their phenotypes.

5.2 Future Studies

The results of this study provide the first picture of how HIV-1 variants may evolve differently at various tissue sites within an infected individual. Additional studies could further elucidate the variation observed in different tissue compartments and help decipher the role this plays in disease progression.

Structural differences could explain changes in tropism or sensitivity to inhibitors and antibodies. Modeling programs could be used to predict structural changes of Envs based on sequences or structures could be examined using X-ray crystallography or Cryo-EM. Structural studies may reveal that some Envs are more open, such as those from the brain or 8-3-6 from the lung, allowing them to infect macrophages and be inhibited by sCD4. This could be supported by experiments to test envelope affinity for receptors, inhibitors, and antibodies.

None of the mutations known to change tropism (e.g., N283, loss of an N-linked glycosylation site at 386, and E153G) were observed. However there were other mutations throughout the env sequence, some of which were more frequent in a specific tissue compartment. In the case of the brain, it is possible that these mutations could be involved in macrophage tropism. To test this possibility, an
extensive mutagenesis study could be conducted in which all of the mutations that were specific to the brain tissues were introduced to envs from peripheral tissues with closely related sequences to see if they confer macrophage tropism. Likewise, it may be possible to determine what mutations are responsible for the differences observed in the inhibition and neutralization assays.

PacBio studies are ongoing and will be used to further examine env sequences in each tissue and will investigate whether tissue specific signature sequences and characteristics exist. In addition, time rooted phylogenetic trees will be constructed to determine which tissues may have been infected first and the migration patterns of viral variants throughout the body.

Finally, it would be useful to conduct further studies in additional subjects. A major limitation of this study is that it only examines one subject. It would be useful to determine if the observations are specific to Subject 162 or if they are common to all HIV-1 infected subjects. If possible, it would be helpful to work with a cohort where more information was available on the subject and their history. If a cohort could be established, additional questions could be asked as well. For example, how do these genotypes and phenotypes change over time? This is difficult to study in humans, as it is challenging or impossible to take longitudinal samples from some tissue sites. Thus previous longitudinal studies have been limited to few tissues compartments (Churchill et al, 2004; Schnell et al, 2010; Bull et al, 2013; Sturdevant et al, 2015; Vazquez-Santiago et al, 2015). It would be useful to know how quasispecies populations change over time within an
individual. The studies presented here could be expanded by modifying sampling methods to include biopsy where appropriate (e.g., colon and lymph node), bronchial lavage for lungs, and cerebrospinal fluid for brain.

5.3 Conclusions

This study added to the understanding of HIV-1 quasispecies populations by characterizing envelopes from multiple tissues within a single HIV-1 infected subject by both genotype and phenotype, which had not been done in any previous study. This is the most comprehensive study of HIV-1 envelope characteristics to date. It is clear in this subject that HIV-1 envelopes within the frontal lobe, occipital lobe, and parietal lobe of the brain have very low diversity. Envelopes from colon, lung, and lymph node peripheral tissues are much more diverse in both their genotypes and phenotypes. This diversity varied across the env sequence, highlighting the importance of studying full length env. Better understanding of viral quasispecies populations may aid in developing vaccines by defining new targets as well as promote discovery of curative treatments by describing populations in potential HIV-1 reservoir sites.
APPENDIX A: PHYLOGENETIC TREES FOR CHAPTER III

The following appendix contains phylogenetic trees for different env regions for Subject 162. Methods were followed as described in Chapter II to construct trees for gp120, gp41, V1-V2, V1-V5, C1, C2, C3, C4, C5, V1, V2, V3, V4, and V5. These phylogenetic trees were discussed in Chapter III and used to calculate evolutionary divergence (Figure 3.2).
Figure A.1 Phylogenetic analysis of Subject 162 gp120 sequences. Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env gp120 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.2 Phylogenetic analysis of Subject 162 gp41 sequences Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env gp41 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env V1-V2 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.4 Phylogenetic analysis of Subject 162 V1-V5 sequences

Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env V1-V5 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.5 Phylogenetic analysis of Subject 162 C1 sequences Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env C1 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.6 Phylogenetic analysis of Subject 162 C2 sequences. Full length HIV-1 *env* were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 *env* C2 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.7 Phylogenetic analysis of Subject 162 C3 sequences Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env C3 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.8 Phylogenetic analysis of Subject 162 C4 sequences Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env C4 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.9 Phylogenetic analysis of Subject 162 C5 sequences  Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env C5 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.10 Phylogenetic analysis of Subject 162 V1 sequences  Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env V1 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.11 Phylogenetic analysis of Subject 162 V2 sequences. Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env V2 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.12 Phylogenetic analysis of Subject 162 V3 sequences Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env V3 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Phylogenetic analysis of Subject 162 V4 sequences

Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env V4 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.14 Phylogenetic analysis of Subject 162 V5 sequences Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 env V5 regions. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
APPENDIX B: ANALYSIS OF AN HIV-1 INFECTED SUBJECT WITH TWO DISTINCT VIRAL POPULATIONS

Introduction

Previous chapters describe the characterization of HIV-1 envelopes derived from multiple tissues of Subject 162, an HIV-1 infected subject who died from end stage AIDS and was neurologically normal. Initial study plans included an additional subject, Subject 123, who also died of end stage AIDS, but had HIV-Associated Dementia (HAD). However preliminary sequence analyses revealed distinct, seemingly unrelated viral populations. These populations could not be directly compared for tissue specific genotypic and phenotypic changes because they may not have originated from the same viral variant. These populations were characterized to determine if they were the result of contamination or if Subject 123 was infected with two viral variants.

Methods

All methods were followed as described in Chapter II with additional methods described below.

Subject 123

HIV-1 positive tissue was obtained from the National Disease Research Interchange (NDRI; Philadelphia, PA). Tissue was harvested postmortem and stored at -80°C until use. Seven tissues were examined from Subject 123 (Table
A.1). These included three brain lobes (frontal lobe (FL), occipital lobe (OL), and parietal lobe (PL)), blood cells (BC), colon (C), lung (L), and lymph node (LN). Subject 123 died of end stage AIDS.

**Short Tandem Repeat Analysis**

Short Tandem Repeat (STR) Analysis (reviewed in Gettings et al, 2015) was performed by Jonathan Ball's laboratory at the University of Nottingham (Nottingham, UK). STR Analysis examines the number of repeats at specific loci in the human genome known to have varying numbers of repeats in different individuals. By examining several different STR loci, it can be determined with high probability whether or not different DNA samples came from the same person. The lab was sent ten coded DNA samples from the tissues of three different subjects. The number of subjects and tissues was not disclosed to the lab. The TH01, vWA, and D21S11 loci were examined using STR Analysis and then each DNA sample was assigned to a profile based on which DNA samples were believed to have come from the same subject.

**Next Generation Sequencing**

Methods for Pacific Biosciences (PacBio) sequencing were followed as described in Chapter II. The only change made for Subject 123 was that nested PCR was performed with a first PCR of 12 cycles, followed by a second PCR of 35 cycles.
Results

Envelopes were isolated from all tissues of Subject 123

Envelopes were isolated from Subject 123 frontal lobe, occipital lobe, parietal lobe, blood cells, colon, lung, and lymph node using limiting dilution PCR (Table A.2). In total, 31 envs were amplified: 14 from the frontal lobe, 1 from the occipital lobe, 1 from the parietal lobe, 1 from blood cells, 3 from colon, 9 from lung, and 2 from lymph node.

Subject 123 envs were compartmentalized in brain and body tissues

All 31 full length env sequences were used to construct a maximum likelihood phylogenetic tree using Molecular Evolutionary Genetics Analysis (MEGA) 5 software (Tamura et al, 2011). Compartmentalization was observed between tissues of the brain and tissues of the body (Figure A.15).

Sequences from brain and body envs were very different

Sequence alignments revealed that env sequences from the brain and body were substantially different and possibly not related. Alignment of the V3 loop region shows the extreme variability between the viral populations derived from tissues of the brain and tissues of the body (Figure A.16).
<table>
<thead>
<tr>
<th>Subject</th>
<th>A/R/S</th>
<th>Tissues</th>
<th>Plasma Viral Load</th>
<th>CD4 Count</th>
<th>HIV Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>43/C/M</td>
<td>Brain - Frontal Lobe, Brain - Occipital Lobe, Brain - Parietal Lobe, Colon, Lung, Lymph Node, Blood Cells</td>
<td>&lt;75 copies/ml (7 months prior to death)</td>
<td>not available</td>
<td>Viread, Norvir, Ziagen, Aptivus, Isentress, Intelence</td>
</tr>
</tbody>
</table>

**Table A.1 Subject 123 data** Tissue was obtained from the National Disease Research Interchange (NDRI; Philadelphia, PA). A/R/S = age/race/sex.
<table>
<thead>
<tr>
<th>Tissue</th>
<th>Envelope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain - Frontal Lobe (FL)</td>
<td></td>
</tr>
<tr>
<td>14 envelopes</td>
<td>33-2-2</td>
</tr>
<tr>
<td></td>
<td>33-10-5</td>
</tr>
<tr>
<td></td>
<td>33-25-3</td>
</tr>
<tr>
<td></td>
<td>35-11-3</td>
</tr>
<tr>
<td></td>
<td>35-15-4</td>
</tr>
<tr>
<td></td>
<td>35-20-1</td>
</tr>
<tr>
<td></td>
<td>35-21-2</td>
</tr>
<tr>
<td></td>
<td>35-22-1</td>
</tr>
<tr>
<td></td>
<td>35-28-4</td>
</tr>
<tr>
<td></td>
<td>35-31-2</td>
</tr>
<tr>
<td></td>
<td>35-35-1</td>
</tr>
<tr>
<td></td>
<td>35-47-1</td>
</tr>
<tr>
<td></td>
<td>35-49-1</td>
</tr>
<tr>
<td></td>
<td>35-79-1</td>
</tr>
<tr>
<td>Brain - Occipital Lobe (OL)</td>
<td></td>
</tr>
<tr>
<td>1 envelope</td>
<td>19-26</td>
</tr>
<tr>
<td>Brain - Parietal Lobe (PL)</td>
<td></td>
</tr>
<tr>
<td>1 envelope</td>
<td>20-32-1</td>
</tr>
<tr>
<td>Blood Cells (BC)</td>
<td></td>
</tr>
<tr>
<td>1 envelope</td>
<td>103-5-1</td>
</tr>
<tr>
<td>Colon (C)</td>
<td></td>
</tr>
<tr>
<td>3 envelopes</td>
<td>26-5-3</td>
</tr>
<tr>
<td></td>
<td>26-4-1</td>
</tr>
<tr>
<td></td>
<td>23-69-1</td>
</tr>
<tr>
<td>Lung (L)</td>
<td></td>
</tr>
<tr>
<td>9 envelopes</td>
<td>10-22-2</td>
</tr>
<tr>
<td></td>
<td>10-23-1</td>
</tr>
<tr>
<td></td>
<td>10-27-3</td>
</tr>
<tr>
<td></td>
<td>10-29-2</td>
</tr>
<tr>
<td></td>
<td>10-32-1</td>
</tr>
<tr>
<td></td>
<td>10-40-4</td>
</tr>
<tr>
<td></td>
<td>10-45-5</td>
</tr>
<tr>
<td></td>
<td>12-39-17</td>
</tr>
<tr>
<td></td>
<td>12-44-10</td>
</tr>
<tr>
<td>Lymph Node (LN)</td>
<td></td>
</tr>
<tr>
<td>2 envelopes</td>
<td>16-38-2</td>
</tr>
<tr>
<td></td>
<td>34-26-4</td>
</tr>
</tbody>
</table>

Table A.2 Subject 123 HIV-1 envs were isolated from different tissues. HIV-1 envs were isolated from three brain regions (frontal lobe, occipital lobe, and parietal lobe), blood cells, colon, lung, and lymph node using nested PCR and DNA at limiting dilution.
Figure A.15 Phylogenetic analysis of Subject 123 gp160 sequences Full length HIV-1 envs were amplified from three brain regions (frontal lobe, occipital lobe, and parietal lobe), blood cells, colon, lung, and lymph node. Four unrelated reference sequences were used as an outgroup. A maximum likelihood tree was constructed using MEGA version 5 for functional HIV-1 envs. Bootstrap values ≥70% are noted at branch points. Scale bar shows nucleotide substitutions per site.
Figure A.16 Alignment of Subject 123 V3 loop sequences V3 loop sequences were aligned using GeneDoc (Pittsburgh Supercomputing Center). Sequences from the brain (frontal lobe, occipital lobe, and parietal lobe) were compared to sequences from the body (blood cells, colon, lung, and lymph node).
Envs from the brain were predicted to use CCR5 and Envs from the body were predicted to use CXCR4

V3 loop sequences were used to predict coreceptor use using the WebPSSM program (University of Washington; Jensen et al, 2003) based on V3 net charge and amino acids at positions 11 and 25, which are known to play a role in coreceptor use (Table A.3). An increased net charge or an arginine or lysine at position 11 or 25 have been correlated with a switch to use the CXCR4 coreceptor. The WebPSSM program also compares envs to known and functional tested CCR5-using and CXCR4-using variants. All envelopes from brain tissues were predicted to use CCR5, while all envelopes from tissues of the body were predicted to use CXCR4 (Table A.3). Envelopes from the brain had a lower net charge around 4 and S and D at amino acid positions 11 and 25, respectively. However envelopes from the tissues of the body had a net charge around 6 with R at amino acid position 11 and E at amino acid position 25.

Different env populations in the brain and body tissues were from the same subject

To rule out the possibility of laboratory contamination or accidentally switching tissues with another subject, Subject 123 tissues were examined using STR Analysis to determine if they came from the same person. STR Analysis confirmed that tissues from Subject 123 had the same STR profile and therefore were from the same subject (Table A.4).
<table>
<thead>
<tr>
<th>Tissue</th>
<th>Genotype 11/25</th>
<th>Average Net Charge</th>
<th>Predicted Coreceptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal Lobe</td>
<td>SD</td>
<td>4</td>
<td>CCR5</td>
</tr>
<tr>
<td>Occipital Lobe</td>
<td>SD</td>
<td>4</td>
<td>CCR5</td>
</tr>
<tr>
<td>Parietal Lobe</td>
<td>SD</td>
<td>4</td>
<td>CCR5</td>
</tr>
<tr>
<td>Blood Cells</td>
<td>RE</td>
<td>6</td>
<td>CXCR4</td>
</tr>
<tr>
<td>Colon</td>
<td>RE</td>
<td>6</td>
<td>CXCR4</td>
</tr>
<tr>
<td>Lung</td>
<td>RE</td>
<td>6</td>
<td>CXCR4</td>
</tr>
<tr>
<td>Lymph Node</td>
<td>RE</td>
<td>6</td>
<td>CXCR4</td>
</tr>
</tbody>
</table>

Table A.3 Predicted coreceptor use for Subject 123 Envs WebPSSM (University of Washington) was used to determine amino acid genotype at positions 11 and 25 of the V3 loop, calculate the net charge, and predict coreceptor use.
Pacific Biosciences sequencing revealed low levels of mixing between CCR5- and CXCR4-using variants

Pacific Biosciences (PacBio) next generation sequencing was used to examine several thousand sequences from Subject 123 (Table A.5). In total, 44,630 full length env sequences were generated: 3,702 from the frontal lobe, 12,625 from the occipital lobe, 12,834 from the parietal lobe, 4,474 from the blood cells, 202 from the colon, 7,099 from the lung, and 3,694 from the lymph node.

Phylogenetic trees were assembled from 250 randomly selected sequences from each tissue (Figure A.17). Trees showed tight compartmentalization of envs from tissues of the brain (frontal lobe, occipital lobe, and parietal lobe) and from tissues of the body (blood cells, colon, lung, and lymph node). The majority of brain derived envelopes were predicted to use CCR5 as a coreceptor, while most of the body derived envelopes were predicted to use CXCR4 as a coreceptor (Table A.5). However interestingly, a small number of CXCR4-using variants were detected in the brain and CCR5-using variants were detected in each of the body tissues (Table A.5; Figure A.17).

Discussion

Subject 123 had two distinct viral populations that did not appear to be related. This was observed in the complete segregation of populations from the brain and body on the phylogenetic tree (Figure A.15), as well as by comparing sequence alignments of the V3 loop (Figure A.16). STR Analysis was used to
Table A.4 Short Tandem Repeat (STR) Analysis  STR Analysis was performed on ten DNA samples from three subjects for three different loci. It was determined that the DNA samples could be categorized into three profiles, A, B, and C, and that these corresponded to the DNA samples from the three different subjects.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Tissue</th>
<th>Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>Frontal Lobe</td>
<td>A</td>
</tr>
<tr>
<td>123</td>
<td>Occipital Lobe</td>
<td>A</td>
</tr>
<tr>
<td>123</td>
<td>Parietal Lobe</td>
<td>A</td>
</tr>
<tr>
<td>123</td>
<td>Lymph Node</td>
<td>A</td>
</tr>
<tr>
<td>123</td>
<td>Lung</td>
<td>A</td>
</tr>
<tr>
<td>162</td>
<td>Frontal Lobe</td>
<td>B</td>
</tr>
<tr>
<td>162</td>
<td>Lung</td>
<td>B</td>
</tr>
<tr>
<td>162</td>
<td>Colon</td>
<td>B</td>
</tr>
<tr>
<td>124</td>
<td>Frontal Lobe</td>
<td>C</td>
</tr>
<tr>
<td>124</td>
<td>Lung</td>
<td>C</td>
</tr>
</tbody>
</table>

Table A.4 Short Tandem Repeat (STR) Analysis  STR Analysis was performed on ten DNA samples from three subjects for three different loci. It was determined that the DNA samples could be categorized into three profiles, A, B, and C, and that these corresponded to the DNA samples from the three different subjects.
<table>
<thead>
<tr>
<th>Tissue</th>
<th>Total Sequences</th>
<th>CCR5-using</th>
<th>CXCR4-using</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
<td>Number</td>
</tr>
<tr>
<td>Frontal Lobe</td>
<td>3,702</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Occipital Lobe</td>
<td>12,625</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Parietal Lobe</td>
<td>12,834</td>
<td>99.98</td>
<td>3</td>
</tr>
<tr>
<td>Blood Cells</td>
<td>4,474</td>
<td>104</td>
<td>4,370</td>
</tr>
<tr>
<td>Colon</td>
<td>202</td>
<td>33.66</td>
<td>134</td>
</tr>
<tr>
<td>Lung</td>
<td>7,099</td>
<td>14</td>
<td>7,085</td>
</tr>
<tr>
<td>Lymph Node</td>
<td>3,694</td>
<td>101</td>
<td>3,593</td>
</tr>
<tr>
<td>Total</td>
<td>44,630</td>
<td>29,445</td>
<td>15,185</td>
</tr>
</tbody>
</table>

Table A.5 Subject 123 PacBio sequencing results The V3 loop nucleic acid sequences for all sequences in each data set were isolated using Geneious software version 9 (Kearse et al, 2012). Gene Cutter (Los Alamos National Laboratory) was used to codon align and translate sequences. The sequencing approach resulted in some gapped positions, so any V3 amino acid sequence containing an ambiguous position was removed from the alignment. Tropism prediction was performed using WebPSSM (Jensen et al, 2003) and the x4r5 matrix.
Figure A.17 Phylogenetic analysis of Subject 123 PacBio sequences A python script was generated to randomly select 250 sequences from each tissue. Due to the small number of sequences generated, all colon sequences were used. Maximum likelihood phylogenies were generated with phyML (Guindon et al, 2010) using the HKY85 substitution model, the NNI tree improvement method, and the approximate likelihood ratio test (Anisimova and Gascuel, 2006). Trees were graphed and colored according to tissue type (A) and predicted coreceptor usage (B) using FigTree software (http://tree.bio.ed.ac.uk/software/figtree/). Trees were midpoint rooted. Branch support is shown on the scale to the left, with red indicating branches with very high support.
confirm that all tissue samples did actually come from the same subject (Table A.4). It was therefore possible that Subject 123 had dual infection. This may arise as a result of coinfection or superinfection. Coinfection occurs when a subject is infected with more than one viral variant at the same time and both variants establish infection. Alternatively, Subject 123 may have been superinfected with a second variant after an initial infection event, with both variants establishing viral populations. It is also possible that the two populations originated from a single viral variant that acquired many mutations and became highly compartmentalized due to different selection pressures in the tissues of the brain and the body. Previous studies have shown that this level of diversity can be possible (Lin et al, 2012).

Though sequences from the tissues of the brain and the body were tightly compartmentalized, preliminary results from PacBio sequencing data showed low levels of CXCR4-using variants in the brain and CCR5-using variants in the body. PacBio studies are ongoing, as a pipeline for analysis had to be developed for this project. Future studies will examine whether recombination occurred between CXCR4-using and CCR5-using variants and will include additional phylogenetic analyses, such as constructing time rooted trees to investigate which tissues became infected first and how viral variants may have trafficked throughout the body. In addition, future studies will include phenotypic analyses to look at tropism (coreceptor use and macrophage tropism).
APPENDIX C: FUNDING SOURCES

Support for this research was provided by NIH grants AI089334, NS084910, and AI097265.
REFERENCES


compartmentalized to the female genital tract in cross-sectional analyses but genital lineages do not persist over time. J Infect Dis, 207(8), 1206-1215.


CDC. 1981. Pneumocystis pneumonia. MMWR. 30(21); 1-3.

CDC 1982a. Kaposi’s sarcoma and opportunistic infections in previously healthy persons. MMWR. 31(22); 294, 300-301.

CDC 1982c. Pneumocystis carinii pneumonia among persons with hemophilia A. MMWR. 31(27); 365-367.

CDC. 1982d. Unexplained immunodeficiency and opportunistic infections in infants. MMWR. 31(49); 665-667.


deleted HIV-1 in blood and cerebrospinal fluid of a long-term survivor who developed HIV-associated dementia. J Infect Dis, 190(12), 2181-2186.


the broadly reactive anti-V3 monoclonal antibody, 447-52D. J Virol, 68(11), 6994-7000.


orally bioavailable, and selective small-molecule inhibitor of chemokine receptor
CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity.
Antimicrob Agents Chemother, 49(11), 4721-4732.


DuBridge, R. B., Tang, P., Hsia, H. C., Leong, P. M., Miller, J. H., and Calos, M.
P. 1987. Analysis of mutation in human cells by using an Epstein-Barr virus

Determinants flanking the CD4 binding loop modulate macrophage tropism of

tropism of HIV in brain and lymphoid tissues is associated with sensitivity to the
broadly neutralizing CD4 binding site antibody b12. Retrovirology, 6, 69.

Dunfee, R. L., Thomas, E. R., Gorry, P. R., Wang, J., Taylor, J., Kunstman, K.,
Wolinsky, S. M., and Gabuzda, D. 2006. The HIV Env variant N283 enhances
macrophage tropism and is associated with brain infection and dementia. Proc
Natl Acad Sci U S A, 103(41), 15160-15165.

HIV envelope V4 region enhances macrophage tropism and is associated with

Etemad, B., Fellows, A., Kwambana, B., Kamat, A., Feng, Y., Lee, S., and Sagar,
M. 2009. Human immunodeficiency virus type 1 V1-to-V5 envelope variants from
the chronic phase of infection use CCR5 and fuse more efficiently than those

Single genome amplification and standard bulk PCR yield HIV-1 envelope
products with similar genotypic and phenotypic characteristics. J Virol Methods,
214, 46-53.

http://www.fda.gov/ForPatients/Illness/HIVAIDS/History/ucm151074.htm


(HTLV-III) from patients with AIDS and at risk for AIDS. Science, 224(4648), 500-503.


Gonzalez-Perez, M. P., O’Connell, O., Lin, R., Sullivan, W. M., Bell, J., Simmonds, P., and Clapham, P. R. 2012. Independent evolution of macrophage-tropism and increased charge between HIV-1 R5 envelopes present in brain and immune tissue. Retrovirology, 9, 20.


DNA may have important functions in the viral life cycle. Nucleic Acids Res, 21(4), 831-839.


Li, S., Juarez, J., Alali, M., Dwyer, D., Collman, R., Cunningham, A., and Naif, H. M. 1999. 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. J Virol, 73(12),
9741-9755.

Lin, N. H., Becerril, C., Giguel, F., Novitsky, V., Moyo, S., Makhema, J., Essex,
M., Lockman, S., Kuritzkes, D. R., Sagar, M. 2012. Env sequence determinants
in CXCR4-using human immunodeficiency virus type-1 subtype C. Virology,

The Vif protein of human and simian immunodeficiency viruses is packaged into

S., Maenza, J., Stevens, C. E., Stekler, J., Collier, A. C., Genowati, I., Deng, W.,
glycosylation following transmission of human immunodeficiency virus Type 1

Lorello, G., la Porte, C., Pilon, R., Zhang, G., Karnachow, T., and MacPherson,
P. 2009. Discordance in HIV-1 viral loads and antiretroviral drug concentrations
comparing semen and blood plasma. HIV Med, 10(9), 548-554.

Lorenzo-Redondo, R., Fryer, H. R., Bedford, T., Kim, E. Y., Archer, J.,
Kosakovsky Pond, S. L., Chung, Y. S., Penugonda, S., Chipman, J. G., Fletcher,
C. V., Schacker, T. W., Malim, M. H., Rambaut, A., Haase, A. T., McLean, A. R.,
and Wolinsky, S. M. 2016. Persistent HIV-1 replication maintains the tissue

from:
http://www.hiv.lanl.gov/content/sequence/VAR_REG_CHAR/variable_region_char
acterization_explanation.html.


Lu, S. 2006. Combination DNA plus protein HIV vaccines. Springer Semin
Immunopathol, 28(3), 255-265.

Luban, J. 1996. Absconding with the chaperone: essential cyclophilin-Gag
interaction in HIV-1 virions. Cell, 87(7), 1157-1159.

Lyumkis, D., Julien, J. P., de Val, N., Cupo, A., Potter, C. S., Klasse, P. J.,
Burton, D. R., Sanders, R. W., Moore, J. P., Carragher, B., Wilson, I. A., and


for viral replication in the CNS during presymptomatic HIV infection. Brain, 129(Pt 2), 503-516.


type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. J Virol, 67(7), 3978-3988.


