AGE DISPARITIES IN CELLULAR SENESCENCE RESPONSES TO SIV INFECTION

AN ABSTRACT

SUBMITTED ON THE TWENTY-NINTH DAY OF NOVEMBER 2021

TO THE INTERDISCIPLINARY PH.D. PROGRAM IN AGING STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE SCHOOL OF MEDICINE

OF TULANE UNIVERSITY

FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY

BY

FEI WU

APPROVED:

Binhua Ling, Ph.D.
Committee Chair

Kate Baker, Ph.D.

Robert Blair, Ph.D.

Erik Flemington, Ph.D.
ABSTRACT

HIV infection plays a role in accelerating aging. A mechanism known as cellular senescence has been previously recognized as a hallmark of aging. However, how HIV/SIV influences brain aging or how it affects the progression of neurological disorders relevant to cellular senescence is not well understood. In this study, the SIV-infected rhesus macaque model was used to characterize the contribution of SIV in the aging process of the brain.

We performed in silico and in vitro experiments followed by in vivo studies. In addition, we developed a novel bioinformatic tool – MTD to facilitate the metatranscriptome analysis. Then, we further verify our findings by using MTD to perform data mining on the publicly shared raw data.

We found that SIV replication is inhibited in senescent cells. Moreover, our results revealed the age disparities in brain cellular senescence response to SIV infection. During SIV infection, young animals are more prone to have cellular senescence in the brain, whereas old animals are more resistant to be induced to cellular senescence in the brain. We further found that old animals have enhanced antiviral function in senescent brain cells than young ones. Given that senescent cells in the brain contribute to cognitive decline and neurodegeneration, our findings indicate that they also play an important role in the acceleration of brain aging in SIV-infected young hosts and possibly contribute toward the development of HIV-associated neurocognitive disorders even with long-term antiretroviral therapy.

Moreover, the association between abnormal behavior in rhesus macaques and SIV infection was studied by using the animal record database in TNPRC. Our study found that
the SIV infection and Chinese-origin were two factors positively associated with the expression of abnormal behaviors in rhesus macaques.
AGE DISPARITIES IN CELLULAR SENESCENCE RESPONSES TO SIV INFECTION

AN DISSERTATION

SUBMITTED ON THE TWENTY-NINTH DAY OF NOVEMBER 2021

TO THE INTERDISCIPLINARY PH.D. PROGRAM IN AGING STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE SCHOOL OF MEDICINE

OF TULANE UNIVERSITY

FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY

BY

FEI WU

APPROVED

Binhua Ling, Ph.D.
Committee Chair

Kate Baker, Ph.D.

Robert Blair, Ph.D.

Erik Flemington, Ph.D.
ACKNOWLEDGEMENTS

I want to express my appreciation for my mentor, Dr. Binhua Ling. I am so grateful for your support of the work I have been doing. I thank my co-mentor, Dr. Kate Baker, for your valuable advice on my behavior studies on the rhesus monkeys. Additionally, I would like to thank the other members of my dissertation committee, Dr. Robert Blair, who taught me the advanced knowledge of image analysis and immunohistochemistry methods, and Dr. Erik Flemington, who guided me to start the RNA-seq analysis.

I would like to thank the members of Dr. Ling's Lab past and present for their constant advice and coordination. Grey De La Torre for being the one who gave me worthy advice in academic writing in English. I thank Suvadip Mallick and Antonio Solis Leal, for your critical thinking and help in my work and living.

In addition, I am so thankful to Dr. Yaozhong Liu for your review of my bioinformatics pipeline for metatranscriptome analysis. Thanks to Dr. Xavier Alvarez and Cecily Midkiff for your guidance and support in my staining and microscopy works. Also, thankful to the histology lab and biochemistry lab in the TNPRC for your expert jobs.

I thank Sandra Smith and Brian Kopecki at the Department of Information Technology of Texas Biomedical Research Institute for maintaining HPC and installing the third-party programs asked. I am grateful to any contributors in our respectable open-source community and their virtue in sharing the questions and answers. Without those open access knowledge, it would be impossible for me to develop bioinformatic software.
I have gratitude to You Lu, May Chynces, Queen Chukwurah, Yan Du, and Katie Kwong, who are past and current students in the Tulane Interdisciplinary Ph.D. Program in Aging Studies. Thank you for your kindness and suggestions when I was new to this program. I am also thankful to the department coordinators and staff, Samayia Hodges, Traydell Beard, for your help in the process of my affairs.

I would like to thank the financial support from Tulane Center for Aging, Dr. Ling, the National Center for Research Resources, and all the other agencies, donors, and anonymous philanthropists behind those funding. Your generosity is essential to have kept me going.

I owe special gratitude to Dr. S. Michal Jazwinski and other committee members in the aging studies program for your faith in me and for giving me this precious opportunity to receive a doctoral education in interdisciplinary aging studies. This exceptional, foresighted, and challenging program at Tulane University, New Orleans, inlaid new lens on me to observe the world. I believe my study here is a start. The seed of interdisciplinary thinking in my heart will further grow into new pathways toward my career goal to increase the human healthy life span.

Thank you to my family, who has constantly supported me spiritually and financially to pursue my dream. This love, trust, and sacrifice passed down from generation to generation nourished me and guarded me to step to the current stage.

Finally, I want to thank my wife for being there with me every step of the way.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................................................. II

LIST OF TABLES ............................................................................................................................ IX

LIST OF FIGURES ........................................................................................................................... X

ABBREVIATION INDEX .................................................................................................................. XII

CHAPTER 1 – RESEARCH BACKGROUND .................................................................................. 1

1.1 HIV in the aging population ................................................................................................. 1

1.2 The Contribution of HIV and Cellular Senescence to Aging ............................................. 2

1.2.1 History of Cellular Senescence Research: A Short Introduction .................................. 2

1.2.2 Cellular Senescence and Viral Infection ......................................................................... 8

1.2.3 HIV Infection, Cellular Senescence and Age-associated Diseases ................................. 12

1.3 HIV and Behavior Issues .................................................................................................... 26

1.4 Aims of Study ....................................................................................................................... 28

List of References: ....................................................................................................................... 30

CHAPTER 2 - AGE DISPARITIES IN CELLULAR SENESCENCE RESPONSES TO SIV INFECTION .......................................................... 46

2.1 Introduction .......................................................................................................................... 46

2.2 Methods ................................................................................................................................ 49

2.2.1 Animals information ....................................................................................................... 49
2.2.2 SIV infection ........................................................................................................................50
2.2.3 Antiretroviral therapy (ART) .................................................................................................50
2.2.4 Serum sample collection ........................................................................................................50
2.2.5 Animal euthanasia and tissue collection ..................................................................................51
2.2.6 Cell culture ................................................................................................................................51
2.2.7 Lipofuscin measurement .........................................................................................................53
2.2.8 RT-qPCR ..................................................................................................................................53
2.2.9 Iron test ......................................................................................................................................54
2.2.10 Brain mononuclear cells isolation ........................................................................................55
2.2.11 Cell sorting ............................................................................................................................56
2.2.12 Bulk RNA-seq ......................................................................................................................56
2.2.13 Single cell RNA-seq analysis ................................................................................................57
2.2.14 Statistical analysis ................................................................................................................60

2.3 Results ........................................................................................................................................62

2.3.1 In silico simulation of the interaction between HIV-1 and cellular senescence ..................62
2.3.2 Reduced replication of SIV in senescent CEMx174 cells ......................................................65
2.3.3 SIV infection contributes to accelerating brain cellular senescence in young but not in old rhesus maeques ........................................................................................................68
2.3.4 Age difference in the change of the serum iron level after SIV infection and ART ............74
2.3.5 Age disparities in the anti-viral function of senescent cells by transcriptome analysis of BMC 76
2.3.6 Microbiome analysis of BMC ..................................................................................................89
2.3.7 Age difference in the anti-viral function of senescent brain cells .........................................96
2.3.8 Age difference in the brain cell types of viral infection .........................................................104
2.3.9 Cell-type correlation analysis ..............................................................................................106
2.4 Conclusion .................................................................................................................. 110

2.5 Discussion and future studies ..................................................................................... 112
  2.5.1 SIV and cellular senescence studies ...................................................................... 112
  2.5.2 Relationship between various viral species and cellular senescence ...................... 112
  2.5.3 Canonical cell type classification .......................................................................... 114
  2.5.4 Lipofuscin would be an ideal marker for cellular senescence measurement in FFPE tissue ... 114
  2.5.5 Association between cellular senescence and immune cells .................................. 115
  2.5.6 Improve cell senescence analysis technique in single-cell RNA-seq data ................ 117
  2.5.7 Assess tissue senescent cell burden by bulk RNA-seq data .................................... 118
  2.5.8 Refine the definition of immune cell senescence ................................................ 119
  2.5.9 Define the "true" cellular senescence ...................................................................... 119

List of References: ............................................................................................................. 123

CHAPTER 3 – DEVELOPMENT OF AN DUAL RNA-SEQ SOFTWARE FOR
METATRANSCRIPTOME ANALYSIS ................................................................................. 130

3.1 Introduction .................................................................................................................. 130

3.2 Methods ...................................................................................................................... 131
  3.2.1 Description of MTD development ....................................................................... 131
  3.2.2 Animal information and tissue collection ............................................................ 142

3.3 Results ......................................................................................................................... 144
  3.3.1 Application to bulk RNA-seq analysis: descending colon of rhesus macaques .......... 144
  3.3.2 Intersection of brain and gut microbiome ............................................................ 162
  3.3.3 Application to single-cell RNA-seq analysis ........................................................ 179
3.4 Conclusion............................................................................................................. 183

3.5 Discussion and future directions ................................................................. 184

List of References:.................................................................................................... 188

CHAPTER 4 - ASSOCIATION OF ABNORMAL BEHAVIOR AND SIV INFECTION
......................................................................................................................................... 198

4.1 Introduction ........................................................................................................ 198

4.2 Methods ............................................................................................................. 202

  4.2.1 Data collection................................................................................................. 202

  4.2.2 Data farming................................................................................................... 202

  4.2.3 Data cleaning.................................................................................................. 206

  4.2.4 Data analysis.................................................................................................. 207

4.3 Results ................................................................................................................ 209

4.4 Conclusion ......................................................................................................... 213

4.5 Discussion and future studies ......................................................................... 214

  4.5.1 SIV................................................................................................................ 214

  4.5.2 Origin........................................................................................................... 215

  4.5.3 Social enrichment.......................................................................................... 216

  4.5.4 Rearing type, sex, and age........................................................................... 216

  4.5.5 Association with cellular senescence ............................................................ 217

List of References:....................................................................................................... 218
APPENDIX .........................................................................................................................221

1. Reagents used in Chapter 2 ..........................................................................................221

2. Source code used in Chapter 2 single-cell RNA-seq analysis....................................222
LIST OF TABLES

Table 1. Animal information of cellular senescence test on FFPE brain tissues.............69
Table 2. Animal information of descending colon samples ........................................143
Table 3. Abnormal Behaviors ......................................................................................201
Table 4. Demographic Table of Abnormal Behavior Analysis ....................................207
Table 5. Logistic regression results ..............................................................................210
Table 6. Reagents used in Chapter 2 ...........................................................................221
LIST OF FIGURES

Figure 1. *In silico* simulation of the interaction between HIV-1 and cellular senescence.64

Figure 2. Induced senescence of CEMx174 cells restricts SIV infection *in vitro*.67

Figure 3. Frontal cortex lipofuscin quantification.70

Figure 4. Gene expression levels of cellular senescence markers in brain.73

Figure 5. Age disparities in the change of the serum iron level after SIV infection and ART.75

Figure 6. Different transcriptome profiles between groups.78

Figure 7. Pathway analysis results of BMC_H.82

Figure 8. Age difference in the functional changes of BMC_H.85

Figure 9. Cell type deconvolution analysis results.88

Figure 10. Heatmap of the microbiome species in BMC samples.90

Figure 11. The taxonomy graph of the microbiome in the BMC.91

Figure 12. Microbiome species analysis of BMC.94

Figure 13. Identification of canonical cell types and senescent cells.98

Figure 14. Virus-positive cells.100

Figure 15. Age difference in the anti-viral function of senescent brain cells.102

Figure 16. Changes in the proportions of cell populations from old and young mice brains.105

Figure 17. Cell-type correlation analysis.109

Figure 18. Overview of MTD.136
Figure 19 The automatic pipeline of MTD for dual-analyzing the bulk RNA-seq raw data.

........................................................................................................................................138

Figure 20. An example of the count matrix automatically generated by MTD for the single-cell microbiome analysis. ........................................................................................................141

Figure 21. The taxonomy graph of the microbiome in the descending colon. ...............145

Figure 22. Heatmap of the microbiome species in descending colon samples. ..........147

Figure 23. Microbiome species analysis of descending colon ........................................149

Figure 24. Differentially expressed microbiome species between groups .......................151

Figure 25 Microbial metabolic molecules in GO terms ....................................................152

Figure 26. Microbial metabolic molecules in KEGG terms ............................................154

Figure 27. Host gene expression analysis ........................................................................157

Figure 28. Association analyses of the microbiome and host genes or pathways in descending colon ..................................................................................................................................160

Figure 29. The intersections of brain and gut microbiome. ..............................................165

Figure 30. Detection of SIV in microglia cells from rhesus macaques .........................180

Figure 31. Co-expression microbiome and host gene .......................................................182

Figure 32. Data cleaning ....................................................................................................206

Figure 33. Multicollinearity and discrimination tests .......................................................212
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Alzheimer Disease</td>
</tr>
<tr>
<td>AGEs</td>
<td>Advanced Glycation End Products</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>AIDS</td>
<td>Acquired Immunodeficiency Syndrome</td>
</tr>
<tr>
<td>ANI</td>
<td>Asymptomatic Neurocognitive Impairment</td>
</tr>
<tr>
<td>ANOSIM</td>
<td>Analysis of Similarities</td>
</tr>
<tr>
<td>AO</td>
<td>Acridine Orange</td>
</tr>
<tr>
<td>ApoE4</td>
<td>Apolipoprotein E4</td>
</tr>
<tr>
<td>ARS</td>
<td>Animal Record System</td>
</tr>
<tr>
<td>ART</td>
<td>Antiretroviral Therapy</td>
</tr>
<tr>
<td>AUROC</td>
<td>Area Under the Receiver Operating Curve</td>
</tr>
<tr>
<td>BMC</td>
<td>Brain Mononuclear Cell</td>
</tr>
<tr>
<td>BMC_H</td>
<td>High Autofluorescent BMC</td>
</tr>
<tr>
<td>BMC_L</td>
<td>Low Autofluorescent BMC</td>
</tr>
<tr>
<td>cART</td>
<td>Combination Antiretroviral Therapy</td>
</tr>
<tr>
<td>CAV1</td>
<td>Caveolin-1</td>
</tr>
<tr>
<td>CDK</td>
<td>Cell Cycle Kinase</td>
</tr>
<tr>
<td>CDs</td>
<td>Cluster of Differentiations</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese Hamster Ovary</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>CMV</td>
<td>Cytomegalovirus</td>
</tr>
<tr>
<td>CNS</td>
<td>Central Nervous System</td>
</tr>
<tr>
<td>COPD</td>
<td>Chronic Obstructive Pulmonary Disease</td>
</tr>
<tr>
<td>CPE</td>
<td>CNS-Penetrating Effectiveness</td>
</tr>
<tr>
<td>CRP</td>
<td>C-Reactive Protein</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal Fluid</td>
</tr>
<tr>
<td>CVD</td>
<td>Cardiovascular Disease</td>
</tr>
<tr>
<td>DDR</td>
<td>DNA Damage Response</td>
</tr>
<tr>
<td>DEGs</td>
<td>Differentially Expressed Genes</td>
</tr>
<tr>
<td>DENV</td>
<td>Dengue Virus</td>
</tr>
<tr>
<td>Dox</td>
<td>Doxorubicin</td>
</tr>
<tr>
<td>DTG</td>
<td>Dolutegravir</td>
</tr>
<tr>
<td>EBV</td>
<td>Epstein–Barr Virus</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-Activated Cell Sorting</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal Bovine Serum</td>
</tr>
<tr>
<td>FFPE</td>
<td>Formalin-Fixed Paraffin-Embedded</td>
</tr>
<tr>
<td>FTC</td>
<td>Emtricitabine</td>
</tr>
<tr>
<td>GALT</td>
<td>Gut-Associated Lymphatic Tissue</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde-3-Phosphate Dehydrogenase</td>
</tr>
<tr>
<td>HAD</td>
<td>HIV-Associated Dementia</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>HANDs</td>
<td>HIV-Associated Neurocognitive Disorders</td>
</tr>
<tr>
<td>HBV</td>
<td>Hepatitis B Virus</td>
</tr>
<tr>
<td>HBV</td>
<td>hepatitis B virus</td>
</tr>
<tr>
<td>HCMV</td>
<td>Human Cytomegalovirus</td>
</tr>
<tr>
<td>HCV</td>
<td>Hepatitis C Virus</td>
</tr>
<tr>
<td>HDL</td>
<td>High-Density Lipoprotein</td>
</tr>
<tr>
<td>HERV</td>
<td>Human Endogenous Retroviruses</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>HIV-VL</td>
<td>HIV Viral Loads</td>
</tr>
<tr>
<td>HPC</td>
<td>High Performance Computing</td>
</tr>
<tr>
<td>HPV</td>
<td>Human Papillomavirus</td>
</tr>
<tr>
<td>HRSV</td>
<td>Human Respiratory Syncytial Virus</td>
</tr>
<tr>
<td>HSV</td>
<td>Herpes Simplex Virus</td>
</tr>
<tr>
<td>HTLV</td>
<td>Human T-cell Lymphotropic Virus</td>
</tr>
<tr>
<td>HUVEC</td>
<td>Human Umbilical Vein Endothelial Cells</td>
</tr>
<tr>
<td>IAV</td>
<td>Influenza A virus</td>
</tr>
<tr>
<td>IFN</td>
<td>Type I interferon</td>
</tr>
<tr>
<td>IL-6</td>
<td>Interleukin-6</td>
</tr>
<tr>
<td>IPF</td>
<td>Idiopathic Pulmonary Fibrosis</td>
</tr>
<tr>
<td>KSHV</td>
<td>Kaposi’s Sarcoma-Associated Herpesvirus</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>LDL</td>
<td>Low-Density Lipoprotein</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>MCP-1</td>
<td>Monocyte Chemoattractant Protein-1</td>
</tr>
<tr>
<td>MCPyV</td>
<td>Merkel Cell Polyomavirus</td>
</tr>
<tr>
<td>MHV</td>
<td>Mouse Hepatitis Virus</td>
</tr>
<tr>
<td>mNCD</td>
<td>Mild Neurocognitive Disorder</td>
</tr>
<tr>
<td>MoCA</td>
<td>Montreal Cognitive Assessment</td>
</tr>
<tr>
<td>MSCs</td>
<td>Mesenchymal Stem Cells</td>
</tr>
<tr>
<td>MTD</td>
<td>Meta-Transcriptome Detector</td>
</tr>
<tr>
<td>MV</td>
<td>Measles virus</td>
</tr>
<tr>
<td>NBF</td>
<td>Neutral Buffered Formalin</td>
</tr>
<tr>
<td>NES</td>
<td>Normalized Enrichment Scores</td>
</tr>
<tr>
<td>NHP</td>
<td>Nonhuman Primate</td>
</tr>
<tr>
<td>NK</td>
<td>Natural Killer</td>
</tr>
<tr>
<td>OAB</td>
<td>Other Abnormal Behavior</td>
</tr>
<tr>
<td>OIS</td>
<td>Oncogene-Induced Senescence</td>
</tr>
<tr>
<td>Old_H</td>
<td>BMC_H in Old Animals</td>
</tr>
<tr>
<td>Old_L</td>
<td>BMC_L in Old Animals</td>
</tr>
<tr>
<td>ONPRC</td>
<td>Oregon National Primate Research Center</td>
</tr>
<tr>
<td>OSAS</td>
<td>Obstructive Sleep Apnea Syndrome</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>p16</td>
<td>p16INK4a</td>
</tr>
<tr>
<td>p21</td>
<td>p21CIP1/WAF1</td>
</tr>
<tr>
<td>PAMPs</td>
<td>Pathogen-Associated Molecular Patterns</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PCoA</td>
<td>Principal Coordinate Analysis</td>
</tr>
<tr>
<td>PD</td>
<td>Population Doubling Number</td>
</tr>
<tr>
<td>PI</td>
<td>Propidium Iodide</td>
</tr>
<tr>
<td>POCT</td>
<td>Point-of-Care Testing</td>
</tr>
<tr>
<td>RAB</td>
<td>Repetitive Abnormal Behavior</td>
</tr>
<tr>
<td>RhCMV</td>
<td>Rhesus Cytomegalovirus</td>
</tr>
<tr>
<td>RIS</td>
<td>Radiation-Induced Senescence</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive Oxygen Species</td>
</tr>
<tr>
<td>RS</td>
<td>Replicative Senescence</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>Quantitative Reverse Transcription PCR</td>
</tr>
<tr>
<td>SA-β-gal</td>
<td>Senescence-Associated β-Galactosidase</td>
</tr>
<tr>
<td>SAHF</td>
<td>Senescence-Associated Heterochromatic Foci</td>
</tr>
<tr>
<td>SASP</td>
<td>Senescence-Associated Secretory Phenotype</td>
</tr>
<tr>
<td>Sc</td>
<td>Senescent Cell</td>
</tr>
<tr>
<td>SIB</td>
<td>Self-Injured Behavior</td>
</tr>
<tr>
<td>SIS</td>
<td>Stress-Induced Senescence</td>
</tr>
<tr>
<td>SNAREs</td>
<td>Serious Non-AIDS-Related Events</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>SOP</td>
<td>Standards of Operation</td>
</tr>
<tr>
<td>ssGSEA</td>
<td>single sample Gene Set Enrichment Analysis</td>
</tr>
<tr>
<td>TB</td>
<td>Mycobacterium Tuberculosis</td>
</tr>
<tr>
<td>TDF</td>
<td>Tenofovir Disoproxil Fumarate</td>
</tr>
<tr>
<td>TIS</td>
<td>Therapy-Induced Senescence</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-Like Receptor</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor Necrosis Factor</td>
</tr>
<tr>
<td>TNPRC</td>
<td>Tulane National Primate Research Center</td>
</tr>
<tr>
<td>VAT</td>
<td>Visceral Adipose Tissue</td>
</tr>
<tr>
<td>VBA</td>
<td>Visual Basic Application</td>
</tr>
<tr>
<td>VIF</td>
<td>Variance Inflation Factor</td>
</tr>
<tr>
<td>VSV</td>
<td>Vesicular Stomatitis Virus</td>
</tr>
<tr>
<td>VZV</td>
<td>Varicella Zoster Virus</td>
</tr>
<tr>
<td>Young_H</td>
<td>BMC_H in Young Animals</td>
</tr>
<tr>
<td>Young_L</td>
<td>BMC_L in Young Animals</td>
</tr>
</tbody>
</table>
CHAPTER 1 – RESEARCH BACKGROUND

1.1 HIV in the aging population

Acquired immunodeficiency syndrome (AIDS) is a malignant and fatal condition with progressive and severe immunodeficiency caused by an infection from the retrovirus, human immunodeficiency virus (HIV). It may take more than 10 years from the initial HIV infection to the development of AIDS. During the HIV latent phase, most people are asymptomatic and unaware of being infected unless specifically tested for HIV. Without receiving effective treatment, most patients generally die within 2 to 3 years after their first AIDS-defining illness [1].

The HIV/AIDS epidemic has continued to affect millions of people throughout the world for over 30 years since it was first discovered. In 2020, approximately 1.5 million people were newly infected with HIV-1, and around 0.7 million people died from AIDS-related diseases. Currently, there are about 37 million HIV carriers globally. However, nearly 30% of the infected persons do not have access to combination antiretroviral therapy (cART) due to economic and sociopolitical reasons [2].

Since the development and initiation of cART, life expectancy and quality of life of HIV patients have improved dramatically. Effective cART treatment has transformed AIDS into a manageable chronic disease so that few AIDS-related complications will develop among treated patients [3]. This has had a significant impact on the age distribution of the infected population. Today, about half of HIV-infected people in high-income countries are older than 50 years, and similar growth trajectories are expected to occur in low- and middle-income countries [4].
1.2 The Contribution of HIV and Cellular Senescence to Aging

HIV-infected patients suffer from many age-related comorbidities, which lead to the assumption that HIV patients experience premature aging phenotypes [5]. It is becoming increasingly evident that cellular senescence is a physiological phenomenon associated with aging and may contribute to the premature aging phenotypes associated with HIV infection [6]. Some senescent cells can secrete inflammatory cytokines, interleukins, growth factors, and proteases, which play an important role in inducing age-related chronic inflammation [7]. Senescent cells accumulate with aging in various tissues, including adipose tissue, skin, lung, bone marrow, gastrointestinal system, kidney, liver, and cardiovascular system [8]. Although HIV-infected patients and their comorbidities have been discussed in relation to premature aging in general, there have been limited studies of cellular senescence as a possible contributing factor of accelerating aging in HIV patients. Here, we introduce the current knowledge of the role of cellular senescence in the aging process of HIV-infected patients.

1.2.1 History of Cellular Senescence Research: A Short Introduction

Nowadays, the accumulation of senescent cells is widely recognized as a hallmark of aging. This acknowledgment is the remarkable outcome of the accumulation of discoveries in the past decades. Thus, it is worth introducing cellular senescence from a historical view and how our understanding of aging has evolved in the wake of scientific discoveries in the field.

Typically, cellular senescence is defined as a stable arrest of the cell cycle along with stereotypical changes in phenotypes. It has been 60 years since Leonard Hayflick first
observed the phenomenon of cell senescence in serial passage cultures of human diploid fibroblasts [9]. The accepted view at that time was that cultured cells could grow indefinitely under the right conditions. After a long debate against that preconceived idea, the validity of the new discovery was finally accepted and successfully repeated by other scientists. This finding created a paradigm shift in what was believed to be the cause of organismal aging from the exogenous, which is fully due to the damage from environments, to the endogenous, which admits mortality of cells and aging is an intrinsic attribute of the organism. Subsequent studies revealed a relationship between organismal and cellular aging. In 1970, Martin et al. found that the *ex vivo* cell replicative life span correlated with the age of the donor from which the cells were obtained [10]. In 1981, Röhme highlighted the correlation between the life span of the cell and the longevity of the donor species [11]. Nevertheless, since Cristofalo et al. failed to verify the relationship between cellular proliferative potential and donor age by using healthy donors of different ages in 1998 [12], this simple explanation of aging by cellular senescence *in vitro* raised criticism [13].

An important advance in the field of senescence was the discovery of cellular senescence biomarkers. In 1995, Goberdhan Dimri and Judith Campisi et al. found the senescence-associated β-galactosidase (SA-β-gal) [14]; and in 1997, Manuel Serrano et al. identified p16 [15]. Some of the markers have a long history of usage and have been increasingly validified in recent years, such as in the case of PAI-1 and lipofuscin [16, 17]. The list of markers is continuously increasing with the growing understanding of senescent cells in different scenarios [18]. These markers enable scientists to study the causal connection between cellular senescence and organismal aging *in vivo*. Senescent cell populations were later shown to increase with age [19, 20]. However, the burden of the
senescent cells remains relatively rare in vivo [21], which makes its effects in deteriorating aging questionable.

This query started to be answered gradually through the discovery of senescence-associated secretory phenotype (SASP) by Judith Campisi’s laboratory in 2008 [22]. The "bystander" effect of senescent cells may explain the influence of those minority cellular populations on the majority cells surrounding or even far from them [7, 23]. SASP involves the production of chemokines, cytokines, proteinases, and growth factors, all of which cause significant effects on neighboring cells, even further increasing the number of senescent cells. The most detrimental impact of SASP is that it increases chronic inflammation, which is a major cause of various age-related diseases [24-27].

In 2011, a breakthrough in the cellular senescence and aging research fields was achieved by a proof-of-concept experiment performed by Darren J. Baker, James L. Kirkland, and Jan M. van Deursen et al. in Mayo Clinic [28]. By using the progeroid transgenic mice to eliminate the senescent cells selectively, they directly demonstrated for the first time that the elimination of senescent cells could alleviate a group of age-related diseases and increase healthspan in vivo. A few years later, Darren J. Baker et al. further validated these observations in a naturally aged transgenic mice model [29]. In 2015, Yi Zhu et al. in James L. Kirkland's laboratory discovered specific drugs that can selectively ablate senescent cells in non-transgenic mice, alleviate symptoms of aging, and extend healthspan [30]. At the same time, the term "senolytics" was established to describe the agents that target the elimination of senescent cells. The research has proven the feasibility of the drug intervention for senescent cell clearance and concretized a novel aim for "anti-aging" drug development. Although it is still a new category of drug development,
senolytics has become an active research area that attracts both academic and industry entities [31-35].

Besides clarifying the role of cellular senescence in the aging process, another major target in the field consists of identifying the cause and the underlying molecular mechanisms of this phenotype. The senescence observed by Hayflick has come to be known as replicative cellular senescence. In 1990, Calvin B. Harley et al. proposed that telomere attrition is the cause of the replicate cellular senescence that Hayflick observed in fibroblasts [36]. This finding raised the notion that senescence may be a programmed process in which the telomeres function as a molecular clock. However, this idea was challenged by an earlier finding that replication senescence exhibits marked heterogeneity in cell division potential within populations and even within clonally-derived subpopulations [37]. These marked stochastic variations were further examined by the theoretical modeling of interactions between plausible mechanisms of molecular aging, including somatic mutations, mitochondrial dysfunction, and telomere erosion. The simulation results indicated that the observed heterogeneities in cell division potential could be the outcome of multiple mechanisms acting synergistically [38]. Therefore, evolutionary considerations argued against the idea that aging is programmed and that it is caused by a single molecular mechanism [39]. The subsequent experimental tests have validated that the random effects of diverse damaging conditions (radiation, oxidative stress, DNA damage, oncogene expression), which involved DNA damage, could induce cellular senescence [40, 41].

Because the pathways leading to the establishment of senescence were more complex than originally anticipated, combining traditional wet lab with bioinformatics and
system biology empowered functional analysis of gene regulation and offered important new perspectives. In recent years, we learned that cellular senescence is an alternative response to damage than cell death [42]. Other than being completely removed by a mechanism such as apoptosis, senescence allows cells to remain, but they lose the potential for division. This allows cellular senescence to act as an anti-cancer mechanism when cells become exposed to harmful stimuli (e.g., DNA damage, oncogene expression) that could induce abnormal proliferation. Under certain circumstances, it may be preferable to preserve the damaged cell while preventing further division [43]. Other than entirely depending on apoptosis and resultant regeneration by stem cells, cellular senescence may be a compromised outcome of natural selection that deal with the cellular damage. This option might have been preserved in evolution because, in nature, its side effects on health were negligible because most selection pressures only work until the species reach the reproductive age. In addition to the potential role of maintaining the damaged tissue, evidence in recent years demonstrated that cellular senescence is essential in development-related affairs such as embryonic development, tissue repair, and wound healing [44-46]. The cellular senescence response may be an example of the antagonistic pleiotropy theory [47], which states that a function (and genes underlying it) with beneficial effects in early life has detrimental effects causing aging later in life.

Since the cell damage response has been implicated as an anti-cancer mechanism, the cellular senescence research has been working tightly with and learning from the cancer research fields. The molecular hallmarks of cellular senescence are mainly related to DNA damage response (DDR), cell cycle kinase (CDK) inhibitors, secretory phenotype, and apoptosis resistance. For example, \(\gamma\)-H2AX foci and phosphorylated p53 in DDR;
CDKN2A encodes p16 and CDKN1A encodes p21 in CDK inhibitors; IL-6, IL-1a, IL-8, CCL2, MMP-1, and MMP3 in secretory phenotype; Bcl-2 and Bcl-xL in apoptosis resistance [48]. In addition, senescent cells also display metabolic changes and endoplasmic reticulum stress traits, but there is no consensus on the molecular markers [49]. Morphological alterations are the other hallmarks that can be observed directly via microscopy or through staining. For example, senescent cells are well-known for their enlarged cell size and irregular shape in cell culture, increased lysosomal content that can be measured by SA-β-gal or lipofuscins. They have dysfunctional mitochondria (fusion/fission) and produce high level of reactive oxygen species (ROS). Besides, they remarked by plasma membrane changes such as upregulating caveolin-1 (CAV1) and nuclear changes with increased senescence-associated heterochromatic foci (SAHF) or downregulated levels of LaminB1 [41, 49, 50]. However, none of the current markers can distinguish senescent cells unequivocally, so a combination of multiple markers may be the best way to identify senescence until further research can be conducted.

In addition to replicative senescence (RS), the cellular senescence caused by other sources were named accordingly. For example, oncogene-induced senescence (OIS), the stress-induced senescence (SIS), radiation-induced senescence (RIS), and therapy-induced senescence (TIS). It is noteworthy that the cellular senescence mentioned in the context of immunology, which contributes to immunosenescence, may describe different conditions than the classical cellular senescence (e.g., RS, OIS, SIS). This could be explained by historical disparities in research methods, subject development, and shared knowledge between fields such as immunology, infectious diseases, the biology of aging, and oncology. In classic senescence research, samples are often sourced from solid tissue,
whereas in the field of immunology, the primary target is mainly the circulatory system. Immune cellular senescence is usually defined solely by the cell surface markers or so-called "Cluster of Differentiations" (CDs). The findings of hallmarks of immunosenescence mentioned in the following section mainly relied on these CD markers. Whether other classical markers of cellular senescence (e.g., p16, p21, SA-β-gal, lipofuscin, telomere length, SASP) are associated with this phenotype remains unclear or without consensus [51]. Therefore, collaboration across various fields is required to illustrate cellular senescence under different circumstances.

1.2.2 Cellular Senescence and Viral Infection

Contrary to the knowledge of its role in cancer, the relationship between cellular senescence and infection is unclear. However, it is important to recognize that cellular senescence is an innate protective mechanism that prevents damage propagation and stimulates the immune system, both of which are involved in the response to pathogenic infections. On the other hand, pathogens like bacteria and viruses may also hijack the biological process of cell senescence to promote infectious diseases. Pathogenic microorganisms generate molecules that could induce senescence either directly or indirectly via pathogen-associated molecular patterns (PAMPs) such as bacterial flagella, lipopolysaccharide, and viral capsids [52]. Moreover, the actions and components of the infectious agent could interact with multiple aspects of senescence [53]:

1. Cellular senescent-related pathways regulating microbial invasion and replication.
2. Infection-induced interference of senescence could either inhibit the senescence causing cancer, or promote senescence by activating oncogenes or other stressors.

3. SASP modulates the tissue microenvironment, which may either build a favorable niche for infection or recruit immune cells to clear the invader.

4. Immunosenescence due to persistent infections.

The relationship between viral infection and cellular senescence is complex. Cellular senescence can serve as an antiviral defense mechanism that restricts viral propagation. However, some viruses are able to utilize the altered pathway in senescent cells to improve their replication. Additionally, cellular senescence can be either triggered or inhibited by different viruses [54].

1.2.2.1 Viral infection-induced cellular senescence

Many viruses induce cellular senescence through the DDR [55], such as the human respiratory syncytial virus (HRSV) [56], Merkel cell polyomavirus (MCPyV) [57], Epstein–Barr virus (EBV) [58], and Kaposi’s sarcoma-associated herpesvirus (KSHV) [59]. Another mechanism to trigger cellular senescence by some viruses is through activation of p53, p16, and p21 pathways, as observed in human cytomegalovirus (HCMV) [60], and the Hepatitis B virus (HBV) [61]. In addition, viruses have developed diverse mechanisms that could lead to cellular senescence. For example, the Measles virus (MV) through cell-to-cell fusion [62]; the Hepatitis C virus (HCV) by shortening the telomeres and then generating ROS in mitochondria [63]; and the H7N9 influenza A virus (IAV) which can increase nitric oxide synthase expression and nitric oxide release [64].
Some studies reported that HIV could induce cellular senescence, although most of the results came from *in vitro* experiments that use proteins or pseudo-virus [65-68]. This concept will be further examined in the following sections.

1.2.2.2 Cellular Senescence as an Antiviral Mechanism

In recent years, cellular senescence has been recognized as an antiviral response. One particular study showed that senescent cells could inhibit the replication of vesicular stomatitis virus (VSV) [69]. This inhibition seems independent of senescent induction stimuli and cell types. Importantly, they also demonstrated the senescent cell inhibition-effect on VSV replication *in vivo*.

In addition, a study reported that the Dengue virus (DENV) was inhibited in the senescent human umbilical vein endothelial cells (HUVEC) [70]. However, this inhibition may be cell type-dependent since senescent monocytes could facilitate DENV infection through increased expression of the receptor DC-SIGN [71].

Finally, MCPyV replication was shown to be inhibited by cellular senescence. It was mediated by KAP-1, phosphorylation of which was induced by infection and can further lead to senescence [57].

In general, it has been proposed that cellular senescence is a host defense mechanism that responds to viral infection [69, 72]. However, here are still a limited number of studies on the antiviral response of senescent cells and no published results about the antiviral effect of senescent cells on HIV/SIV.
1.2.2.3 Viral Inhibition of Cellular Senescence

Some viruses have developed mechanisms to subvert the antiviral response of cellular senescence in order to replicate in the host. Infection of EBV or KSHV initially activates DDR and leads to oncogene-induced senescence. However, EBV could express the latent proteins EBNA3C and LMP1 that attenuate the DDR and block the p16 pathway [73, 74]. Similarly, KSHV expresses the latent proteins vCyclin and v-FLIP that also inhibit cellular senescence by resistance to p16, p21, and degrade p27 [75]. In addition, it has been reported that HBV can express HBx protein to downregulate p16 and p21 to overcome senescence [76]. Moreover, human papillomavirus (HPV) expression of E6 and E7 proteins destabilize pRb and stimulate telomerase, resulting in senescence inhibition [77].

1.2.2.4 Promoted Viral Infection by Cellular Senescence

Some viruses can take advantage of the senescence program in various ways to increase their replication rates. Viral receptors, which may increase when cells enter senescence, can be utilized by these viruses. For example, senescent cells upregulate DC-SIGN, a receptor that DENV can attach to, to facilitate cell infection [71]. The increased DC-SIGN on monocytes might also increase infectivity by other viruses that could interact with this receptor, such as HCV, IAV, and SARS-Cov [78]. In addition, replicative senescent human bronchial epithelial cells or human dermal fibroblasts showed increased infection by IAV and Varicella Zoster Virus (VZV) [79]. This higher susceptibility was considered due to attenuated virus-induced type I interferon (IFN) in senescent rather than non-senescent cells.
1.2.3 HIV Infection, Cellular Senescence and Age-associated Diseases

For the HIV infection and cellular senescence, there are currently only a handful of studies that focus directly on this topic. Therefore, in addition to HIV and cellular senescence, we reviewed relevant literature about HIV-induced age-related morbidities and discussed the potential role of senescent cells in these aging processes.

1.2.3.1 HIV and Cellular Senescence in Chronic Inflammation

Even though antiretroviral therapy (cART) reduces the increased immune cell activation after HIV infection, several factors predispose to chronic inflammation. cART-treated patients with undetectable HIV viral loads (HIV-VL) seem to have ongoing low-level HIV replication [80], which stimulates chronic immune activation. Other contributors include thymic dysfunction leading to impaired T cell maturation [81] and co-infection with specific viruses, including cytomegalovirus (CMV), hepatitis B and C virus, and HPV [80]. Immune activation is also related to the translocation of intestinal microbial products into systemic circulation. Microbial translocation occurs when the gut-associated lymphatic tissue (GALT) is disrupted by HIV, which causes loss of CD4+ cells and epithelial injury and incomplete restoration by cART [82]. The biologically active products, including lipopolysaccharide (LPS), can pass into the bloodstream due to the incomplete restoration of GALT. They stimulate immune response by activating coagulation factors, macrophages and monocytes, as well as T and B cells [82]. Incomplete CD4+ recovery due to persistent immune activation is a risk factor for developing serious non-AIDS-related events (SNAREs) [83].
Chronic inflammation may play a core role in the human aging process. This "sterile" and low-grade inflammation has been coined "inflammaging" and has been reported to relate to major age-related diseases, including cardiovascular diseases, cancer, sarcopenia, bone diseases, type 2 diabetes, cartilage diseases, neurodegenerative diseases, and pulmonary diseases. In addition to these inflammation-driven chronic diseases in older adults, neuroendocrine activation acting via a chronically-stimulated hypothalamic-pituitary-adrenal axis can cause long-term system toxicities [84]. A variety of proinflammatory cytokines are elevated during aging, such as interleukin-6 (IL-6), IL-1β, C-reactive protein (CRP), and tumor necrosis factor (TNF)-α, which attract further immune system components [85].

In cART-treated HIV patients, chronic inflammation caused by ongoing low-grade HIV activity plays an important pathogenic role in the development of SNAREs. These include atherosclerosis, dementia, osteoporosis, and frailty. Also, the level of inflammatory markers such as IL-6, TNF, CRP, and the innate immune markers including CD14, CD163, and CXCL10 are elevated in the patients with cART. [86].

Some potential causes of age-related chronic inflammation include dysregulation of the immune system, immunosenescence, and chronic antigenic stimulation. Cellular senescence also plays an important role in this process through SASP [87]. The pro-inflammatory cytokines and proteases in the SASP exacerbate local microenvironment dysfunction and promote inflammaging.

Inflammation damages tissue by two mechanisms: 1) Infiltrated immune cells degrade tissues through the release of toxic molecules such as ROS. 2) Inflammatory cytokines induce changes in nearby cells that are independent of the immune system. For
example, IL-6 and IL-8 can dysregulate macrophages, initiate an innate immune response, induce migration and invasion of endothelial and epithelial cells, stimulate angiogenesis, and disrupt intercellular communication [88]. Moreover, the susceptibility to autoimmune diseases caused by chronic inflammation can lead to a limited capacity to initiate an immune response and to further increase inflammation in response to infection or other antigens (e.g., tumor, senescent cells, and immunization) [89].

The general decline in immune system function resulting from aging causes vulnerability to autoimmune disorders, cancers, and infections that correlate with increased mortality in older people [90]. Particular cell types undergo functional alterations during aging. For instance, the activation ability of T cells becomes damaged, and macrophage function is decreased. During aging, the anti-inflammatory pathway that keeps inflammation in check may also be impeded, leading to the low-grade, chronic sterile inflammatory state and tissue damage.

1.2.3.2 Immunosenescence

Chronic inflammation also contributes to the development of immunosenescence that is characterized by the quantitative and functional changes in immune parameters that occur in normal aging. Immunosenescence in older populations results in a decreased ability to control infectious diseases, loss of effectiveness in responding to vaccines, and an increased risk of age-related disorders [91].

Immunosenescence nearly affects the entire immune system, including monocytes, macrophages, neutrophils, natural killer (NK) cells, dendritic cells in the innate immunity, as well as T cells and B cells in the adaptive immunity. Despite successful antiretroviral
treatment, accelerated immunosenescence remains a detrimental consequence of infection [92].

Chronic stimulation of the immune system accompanied by the consistent telomere shortening and cellular senescence, changes the immune parameters in an age-related manner. Environmental and lifestyle factors also contribute to these changes. Lifelong exposure to antigens from the environment or latent viruses (e.g., CMV, EBV, herpesviruses, varicella) [93] can lead to chronic immune stimulation. Moreover, during human aging, increased gut microbial translocation and thymic atrophy attenuate the immune functions. The hallmarks of immunosenescence reflecting changes in the immune system due to senescence are:

1. Increased frequencies of the terminally differentiated CD28− and CD57+ T cells, especially within the CD8 subset with an altered expression of regulatory receptors and the release of the cytokines IL-6 and TNF-α, promote autoimmune diseases and chronic inflammation [94, 95].
2. The accumulation of memory T cells [96].
3. Reduced proportion of naïve T cells, which compromise the responses to novel antigens.
4. Inverted T helper to T suppressor cell ratio, which is a key feature to predict morbidity and mortality [97].
5. B cells have a reduced capacity to make high affinity antibodies.
6. T and B cell receptors have reduced diversities.
7. Reduced serum antibody (IgD and IgM) levels produced by naïve B lymphocytes.
The immunosenescence-related changes in older people and cART-treated HIV patients show similarities such as a reduction of the pool of naïve T cells, low proliferative potential, low CD4$^+$ to CD8$^+$ ratio, expansion of CD28$^-$ and CD8$^+$ T cells, elevated proinflammatory cytokines expression, reduced thymus function, and reduced vaccine responses. Further, decreases in telomerase activity, adenosine deaminase, and LRRN3 have been strongly associated with high CD38 in CD8$^+$ T cells [98]. This suggests CD8$^+$ T cells senescence may be a biomarker of the immune status in HIV infection. Accelerated aging due to HIV is corroborated by evidence of similarities between uninfected older adults and young HIV patients. Both groups show similar numbers of naïve T cells, accelerated rates of telomere shortening in CD28$^-$ and CD8$^+$ cells, and elevated level of p16, which is a cell senescence mediator and biomarker of aging [99].

1.2.3.3 Co-infections with HIV

Another similarity between HIV patients and older uninfected adults with other viral infections is the contribution of persistent viral infections to immune system dysfunction.

CMV coinfection is very common in HIV patients, ranging between 36-77% occurrence in developed countries, and nearly 100% in developing countries [100]. Long-term cART-treated patients have a high level of CMV-specific T cells, which is similar to that observed in the elderly but occurs at much younger ages. [101]. Additionally, CMV can increase cardiovascular risk. Woman that are coinfected with HIV and CMV are more likely to develop carotid artery stiffness [102]. CMV-specific T-cell responses are independently associated with increased carotid intima-media thickness in HIV-infected
people [103]. CMV contributes to the expansion of the CD8$^+$ and CD28$^-$ T cells [104] and to an inverted CD4$^+$ to CD8$^+$ ratio, which results in poor health outcomes. Latent herpes virus infections in older adults have been reported to account for up to 20% of the total memory T cell population. A strong anti-CMV response is positively correlated with the total and naïve CD4$^+$ T cell numbers and with immune recovery. In HIV patients with incomplete cART-mediated recovery, anti-CMV drug treatment can decrease CD8$^+$ T cell activation, rescue the CD4$^+$/CD8$^+$ ratios, and have fewer immunosenescence-related markers [105, 106]. The CMV-induced immune response may account for this accelerated atherosclerosis in HIV patients. CMV and HIV may act as co-drivers that increase the rate of immunological aging and the morbidity of chronic diseases among those receiving cART [107].

Other common co-infections with HIV include Mycobacterium tuberculosis (TB), hepatitis B virus (HBV), hepatitis C virus (HCV), Epstein-Barr virus (EBV), Herpes Simplex virus (HSV), Human T-cell lymphotropic virus (HTLV), Cryptococcus neoformans, and Plasmodium falciparum [108-110]. Studies have shown that chronic infection with these pathogens contributes to accelerated immunosenescence among the healthy elderly population, suggesting a role of these pathogens in aging. Interestingly, human endogenous retroviruses (HERV), which make up to 8% of the human genome, may still be active [111], and HERV-derived peptides can activate CD8$^+$ T-cells to destroy the cells co-presenting HIV and HERV antigens [112]. However, there is still not much information about the impact of these pathogens' latency on the host immunosenescence or their association with accelerated aging among HIV patients with successful cART. More research is needed to reveal the mechanisms by which latent pathogens induce
autoimmunity, chronic inflammation, and the aging process. It is unclear whether these latent pathogens share similar pathways with latent HIV in inducing immunosenescence and accelerating aging. Nonetheless, HIV increases the probability of patients being infected with other pathogens. Many of these pathogens, especially viruses, and their impact on human health are not well defined [113]. Unveiling the novel viruses that infect humans is promising not only for HIV patients but also for the benefit of public health. Moreover, the evolving sequencing technology and the expanding genome database could speed up the discovery of the virome in HIV carriers and in healthy aging population [114].

1.2.3.4 HIV and Cellular Senescence in Metabolic, Bone, and Lung Diseases

Regarding diabetes, senescent endothelial cells and preadipocytes accumulate in the adipose tissue of obese subjects. Telomere shortening has also been implicated in the morbidity of type 2 diabetes. Both in vitro and in vivo studies showed that some of the key SASP components, including monocyte chemoattractant protein-1 (MCP-1) and IL-6, contribute to insulin resistance. The chronic inflammation of adipose tissue in obesity is associated with type 2 diabetes [115].

In osteoporosis, cellular senescence is implicated in bone loss and osteoblast dysfunction. Senescent chondrocytes are also found in the joints of patients with osteoarthritis and are considered to be involved in its pathogenesis. Likewise, senolytic drugs increase vertebral bone mineral content and density in progeroid mice with osteoporosis [30]. Moreover, HIV proteins Tat and Nef can induce senescence in bone marrow mesenchymal stem cells (MSCs) and impair its osteoblastic potential [67]. This suggests that MSC senescence may be a contributor to osteoporosis in HIV patients.
In the lung, senescent cells accumulate in idiopathic pulmonary fibrosis (IPF), and in chronic obstructive pulmonary disease (COPD). In HIV-infected people, cellular senescence and chronic lung inflammation contribute to these age-related conditions. For instance, decreased telomere length in peripheral blood mononuclear cells, circulating levels of IL-6, a component of SASP, and CD4+ T-cell senescence were correlated with COPD [116]. Shorter telomere length in peripheral leukocytes in HIV patients correlates with decreased pulmonary function [117]. A recent study demonstrated that senolytic drugs improve pulmonary function and physical health of lungs with IPF [118].

The presence of senescent cells is related to the premature aging phenotype that accompanies HIV-infected people. The level of p16, a marker of senescence, is increased in the peripheral blood leukocyte and corneal endothelial cells in HIV patients [99, 119]. These cell populations display a decreased density and changed morphology as well as shortened telomere length in HIV patient, suggesting a correlation between senescence and HIV infection.

As mentioned above, cellular senescence and chronic inflammation are interrelated processes that reinforce each other and contribute to multiple age-related diseases. Drugs that target chronic inflammation and fundamental aging processes, such as increase in mTOR or cellular senescence, may play an important role in alleviating a range of age-related diseases. However, other than T cells, there are few studies that examine HIV-associated senescence in solid tissue with other types of immune cells, such as myeloid cells (e.g., macrophage, monocyte, and microglia) and non-immune cells of HIV patients. Especially in the cART era, the viral load can be suppressed, but the chronic presence of
viable viral genomes in the reservoir cells could still promote the generation of the senescent phenotype.

1.2.3.5 HIV and Cellular Senescence in Cardiovascular Disease

An epidemiologic study led by Triant et al. provided evidence that treated HIV patients have an increased risk of developing typical cardiovascular disease (CVD) which increases with age [120]. HIV interferes with the reverse cholesterol transport of macrophages, decreasing the high-density lipoprotein (HDL) cholesterol levels [121]. On the other hand, HIV viral component gp120 increases endothelial tissue inflammation factor levels, initiating proatherosclerotic signals [122].

Metabolic complications in older people contribute to increased CVD risk. A higher frequency of dyslipidemia has been observed in patients with HIV, which comprises of elevated triglyceride presence together with a low level of HDL cholesterol, and a high level of low-density lipoprotein (LDL) cholesterol [123]. Type 2 diabetes and insulin resistance are increased in HIV patients, and are influenced by factors such as lifestyle and treatment from ART drugs [124]. There is increased obesity in both cART treated and untreated patients [125], which likely reflects the prevalence of obesity in the general population. Just as in uninfected people, obesity is a major risk factor that contributes to it related complications in treated patients. Metabolic syndrome in older HIV patients is more prevalent than in young patients [126]. Body composition changes during aging, particularly increased visceral adipose tissue (VAT) may increase CVD risk in older patients [127, 128].
Atherosclerosis indicators such as calcium levels and thickness of artery walls were increased in the coronary artery, epicardial vessel, and carotid artery in older patients [129-131]. The chronic inflammatory state also has a proatherosclerotic effect on the cardiovascular system. Compared to those on continuous therapy, patients on episodic cART have an increased risk of SNAREs and CVD events [132]. The levels of the inflammatory markers IL-6 and D-dimer were elevated in patients on episodic cART [133]. The impact of chronic inflammation on atherosclerosis is further supported by the finding that HIV patients have increased cIMT and C-reactive protein levels after controlling for traditional CVD risks [134]. This study included a group of HIV patients who maintained normal immune profiles and undetectable viral loads and were untreated with cART. Studying the mechanisms behind the natural control of HIV may provide insight into the role of immune activation in disease progression [135].

There is a solid link between senescent cells and CVD. Senescent endothelial and smooth muscle cells accumulate in the plaques and are involved in atherosclerosis [136, 137]. Stressors that drive cellular senescence, such as oxidative stress and telomere attrition, are associated with major cardiovascular diseases including atherosclerosis, hypertension, and heart failure. Moreover, senolytic drugs which selectively eliminate senescent cells can reduce the age-related loss of cardiovascular function in old mice [29, 30]. Compared with the untreated group, both cardiac ejection fraction and carotid smooth muscle vascular reactivity were improved in the old group with senescent cell clearance.

Vascular endothelial cell senescence contributes to atherosclerosis in HIV patients. *In vitro* experiments showed that protease inhibitors could induce cellular senescence in arterial endothelial cells [138], and HIV proteins gp120 and Tat cause senescence in
vascular endothelial cells [65]. These proposed mechanisms could act by upregulating the expression of miR-34a, which targets SIRT1, leading to senescence [139].

1.2.3.6 HIV and Cellular Senescence in Cognitive Aging

HIV-associated neurocognitive disorders (HANDs) are a spectrum of cognitive impairments, including asymptomatic neurocognitive impairment (ANI), mild neurocognitive disorder (mNCD), and HIV-associated dementia (HAD). HANDs prevalence rate varies between 30% and 50% in HIV+ people [140], and between 20% and 30% in the early-treated or stably virally-suppressed individuals [141]. Although cART reduces the prevalence of HANDs significantly to 2-4%, a milder cognitive decline still exists [142]. ANI and mNCD occur in 20-25% and 25-30% of all patients, respectively. However, the prevalence is still controversial because a reliable assessment of morbidity remains challenging due to the difficulties with control selection, neuropsychologic testing, and determining functional status in younger subjects [141, 143].

Several factors contribute to the development of HANDs. Advanced age is a significant factor associated with neurocognitive decline and susceptibility to HANDs. Although there is debate as to whether HANDs represent an accelerated aging process, older people with HIV showed more rapid memory decline than younger people with HIV or uninfected controls [144]. Middle-aged HIV-positive patients performed worse in neurocognitive tests than HIV-negative controls, but similar to the older HIV-negative people [145]. These observations indicate that HIV is associated with accelerated cognitive aging in elders. In this case, premature brain aging is caused by direct damage from toxic viral products and indirectly through an increased risk of cardiovascular disease and
chronic ART drug use. HIV-related metabolic syndromes such as hypertension, abdominal obesity, dyslipidemia, and diabetes contribute to AD and vascular dementia. In treated HIV patients, hypertension, older age, abdominal obesity, proteinuria, and type 2 diabetes impair performance on standard cognitive function tests [146-148].

Although the HIV-VL in serum is undetectable in chronically ART-treated patients, HIV-VL and respective activation markers are still increased in the CSF [149]. Increasing evidence supports the association between cognitive decline and systemic inflammation in the population [150, 151]. Magnetic resonance brain imaging shows reduced cerebral blood flow, atrophy, and impaired neuronal networks due to HIV infection, and these brain dysfunctions are independent of aging [152]. Other factors such as a low serum CD4\(^+\) cell count and high HIV-VL could also predict cognitive dysfunction [142].

Whether HIV leads to neurodegeneration is still controversial. Some studies have suggested that neurodegeneration is a consequence of HANDs. Longitudinal studies show declined cognitive functions over time and a higher rate of HIV-infected people developing HANDs [153]. Brain pathophysiology also substantiates that HIV causes neurodegenerative changes. Alzheimer disease (AD) biomarkers in CSF are more prevalent in HIV patients than in controls. HIV patients with the apolipoprotein E4 (ApoE4) allele perform worse on cognitive tests than those without the allele [154]. Despite the common pathophysiological pathways, the big discrepancy between cognitive disturbance and brain dysfunction in HANDs and neurodegeneration makes HIV unlikely to cause neurodegeneration, per se [140]. AD has a higher A\(\beta\) accumulation rate than HIV or normal aging. The hippocampus is the first region to undergo neurodegeneration in AD, whereas in HANDs, it is first observed in the fronto-subcortical area.
Although HIV can cross the blood-brain barrier and is a neurotropic virus, it does not infect neurons directly. Immune activation of microglia, astrocytes, and CNS macrophages, followed by increasing proinflammatory cytokine levels in CNS and cerebrospinal fluid (CSF), could be a consequence of HIV brain infection [155]. Hence, neuronal function could be harmed by this detrimental inflammation.

Psychological and lifestyle-related disorders are prevalent in HIV patients, and their impact on neurobehavioral status and treatment outcomes should attract more attention. Co-infection with a persistent virus such as CMV or HCV may exacerbate the cognitive decline [156, 157].

Neurocognitive decline also relates to the presence of inflammatory markers. Recent research showed the association between markers of inflammation in the CSF and plasma. These markers could predict AD pathology and neuronal damage [158]. Bettcher et al. found that MIP-1β levels in CSF and plasma are strongly correlated. Levels of CSF and plasma MIP-1β, IL-8, and CSF IP-10 were positively correlated with the levels of p-Tau. Moreover, they identified positive correlations between CSF levels of IL-8 and Aβ42; plasma IL-8, MIP-1β and CSF sAβPPβ; CSF IL-8, IP-10, MIP-1β, and total levels of Tau in the CSF. As HIV contributes to chronic systemic inflammation as well as dementia, it would be interesting to investigate to what degree HIV can lead to AD via the inflammatory pathway.

However, there is a current lack of reliable methods to accurately diagnose ANI and mCND. Thus, developing a more advanced method for cognitive dysfunction screening is imperative. Current methods such as HIV Dementia Scale (HDS), standard
Montreal Cognitive Assessment (MoCA), and Mini-Mental Status Exam are ineffective in diagnosing HAND, particularly for mild cognitive impairment [159, 160].

Antiretroviral drugs demonstrate different CNS penetration capacities. Generally, they can reduce CNS inflammation and neurocognitive impairment. The variable CNS penetration ability can be examined by the HIV-VL in CSF. The CNS-penetrating effectiveness (CPE) score has been used to calculate the correlation between CNS penetration by drugs and neurocognitive symptoms [161]. A higher CPE scores generally precede better neuropsychological outcomes [162]. However, some antiretroviral drugs may be neurotoxic. For example, some patients may suffer neuropsychiatric toxicity when using efavirenz, an effective and commonly used anti-HIV drug [149]. Further, drug abuse accompanied by AIDS, particularly methamphetamine, can exacerbate HANDs and promote astrocyte senescence through the down-regulation of the Wnt/β-catenin pathway [163].

Over the past few years, accumulating research has shown the association between cellular senescence and Alzheimer's disease. Increased senescent cell makers have been found in astrocytes and in neurons with Tau-containing neurofibrillary tangles from Alzheimer patients [164, 165]. A Tau-dependent mouse model of AD showed that senescent astrocytes and microglia were aggregated in the brain and contributed to cognitive decline [166, 167]. Moreover, senolytic agents also delayed age-related neurodegenerative diseases in mice [30]. An in vitro experiment demonstrated that Env-pseudotyped HIV could induce senescence-like phenotype in human fetal microglia [68]. This is likely due to the Tat-mediated downregulation of SIRT3. In HIV patients, SIRT3 was downregulated in the prefrontal cortex along with the upregulation of senescence
markers [66]. Moreover, VSV-G-pseudotyped-HIV caused human fetal astrocytes senescence in culture. Also, HIV could induce astrocyte senescence in vivo in rodent models [168]. Although these results imply that HIV may contribute to the cellular senescence and cognitive decline in HIV patients, they were mainly in vitro experiments or on rodent models. No research was conducted on the non-human primate model, such as rhesus macaques with SIV infection.

1.3 HIV and Behavior Issues

Smoking, abuse of alcohol, and low physical activity levels are relatively popular in the population with HIV. These unhealthy behavior factors can affect the expression biomarker of aging, such as p16 and telomere length. Moreover, at the initiation of health care or academic study, these behaviors should be assessed. The strategies to correct the behavior problem should be offered to the patients, like nutritional advice and weight reduction programs. Also, the influence of these issue to the study results also have to be well considered. For instance, smoking has been proved to increase the risk of CVD, lung diseases, and cancer in HIV patients. It elevates the p16 in T-cells. [169]. The higher rate of obesity in HIV-positive individuals associate with the increased morbidity of diabetes and hypertension.

In addition to the addictive behavior factors affecting health outcomes in HIV patients. HIV could cause several behavior changes and possible detrimental consequences. People living with HIV have a higher chance of experiencing sleep disturbances. Up to 70% of them have insomnia and obstructive sleep apnea syndrome (OSAS) [170]. Neurobehavioral disturbance such as apathy and executive dysfunction also increased in
individuals with HIV [171]. Characterized by the damage to the frontostriatal system, HIV could induce cognitive and psychiatric syndromes. Alexithymia, a difficulty in affective and cognitive-emotional processing, is considered to associate with HIV [172]. Research has also shown that facial emotion recognition ability is impaired in the HIV patient [173].

Moreover, the presentation of behavioral pathologies such as anxiety, mania, and delirium are more common in HIV/AIDS-affected persons [174]. These abnormal behaviors can be caused by different factors, such as pre-existing mental disorders and substance misuse. It may also be caused by HIV infection itself, by opportunistic infections, or by medications being used to treat HIV. This topic will be further discussed and investigated in Chapter 4.
1.4 Aims of Study

The central hypothesis of this study is that there is an age disparity in response to SIV infection in the brain. Aged animals are less affected by the SIV infection than young animals due to the antiviral function of the senescent cells. Moreover, this antiviral function was more evident in the senescent cells from old than from young animals.

**Specific Aim 1:** Observe the age disparities in the brain after SIV infection

1. Compare senescent cell burden in the brain tissue from aged and young animals, with or without SIV infection.

**Specific Aim 2:** Examine the antiviral function of senescent cells during SIV infection

1. *In silico* simulation of the interaction between cellular senescence and HIV-1 infection
2. Test SIV replication in the senescent cells *in vitro*.
3. Examine the antiviral function of senescent brain cells *in vivo*.
4. Compare the antiviral function of senescent brain cells between old and young animals.

**Specific Aim 3:** Develop bioinformatics software to facilitate host-pathogen analysis.

1. Develop a user-friendly software for dual-analysis of the microbiome and host gene expression in both bulk and single-cell RNAseq data.
2. Use the software to study the relationship between cellular senescence and viral infection.

**Specific Aim 4:** Investigate the association between abnormal behavior and SIV infection in rhesus macaques

1. Use the data from the Animal Record System to test the association by logistic regression.
List of References:


156. Letendre, S., et al., *Higher Anti-Cytomegalovirus Immunoglobulin G Concentrations Are Associated With Worse Neurocognitive Performance During*


CHAPTER 2 - AGE DISPARITIES IN CELLULAR SENESCENCE RESPONSES TO SIV INFECTION

2.1 Introduction

HIV infection plays a role in accelerating aging. Cellular senescence has been recognized as a hallmark of aging. However, it is not well understood if HIV/SIV accelerates brain aging or neurological disorders pertaining to cellular senescence. Moreover, the role of cellular senescence in viral infections is still inconclusive. Thus, senescence could either inhibit or promote viral replication and cellular response to the viral infection. In addition, due to the natural increase of senescent cells in older adults, the age-related disparities between the virus and senescent cell interactions are unclear. However, partially due to the challenge of access to human brain tissues, it is hard to investigate the role of HIV infection in cellular senescence. Here, we used the SIV-infected rhesus macaque model to study whether SIV contributes to cell senescence and to evaluate the impact of cellular senescence response to viral infection in the brain.

In order to study the relationship between SIV/HIV infection and cellular senescence in vivo, a reliable marker for the detection and assessment of senescent cells in tissue is required. However, it remains challenging to detect and assess these cells in vivo with high levels of confidence [48]. Unlike in cell cultures, most cells in the brain tissue are terminally differentiated or quiescent, meaning that some of the markers used to identify senescent cells in vitro, which usually show a lack of proliferation and DNA synthesis, are not valid. Furthermore, the markers associated with the cell cycle or SASP can also be expressed in other cell types or during inflammation processes. In addition, morphological changes in vitro may also be different in vivo. Hence, to enrich and collect
the senescent cells population from tissues and assess their special characteristics in vivo, a proper cell sorting method is required. However, most markers are not suitable for the separation of senescent cells using fluorescence-activated cell sorting (FACS) due to the nonspecific concerns mentioned above and other technical difficulties. For example, the SA-β-gal staining can only be performed in fresh tissue, making it incompatible with other fixative sample preparation methods; and it is still a lack of good quality antibodies for p16 and p21 [175]. Substantial research is now being carried out to provide a more robust marker for identifying and assessing senescent cells in both fixed tissues and live animals.

Lipofuscin is an age pigment whose accumulation is considered a hallmark of cellular senescence [17, 41]. Autofluorescence is the most consistent feature of lipofuscin, and it has been used as a marker for the detection of cellular senescence [176-178]. Because it exempt the necessity of fixation for measuring the biomarker, it facilitates the collection of live senescent cells for downstream experiments. Therefore, the autofluorescence feature of lipofuscin turns it into the ideal marker to detect and sort the senescent cell in vivo.

By using this marker and through subsequent RT-qPCR analysis, we observed changes in cellular senescence in the SIV-infected and non-infected young and old rhesus monkeys. Next, we sorted the senescent cells from BMC of SIV-infected and ART-treated young and old rhesus macaques followed by RNA-seq and transcriptome analysis. The transcriptome of sorted cells showed features of senescence and enhancement of antiviral functions which were most significant in old animals. Hence, we wanted to verify this finding by interrogating the viral transcripts from the same RNA-seq data. However, there was lacking a proper method for this purpose. Thereby, we developed a dual-RNAseq
analysis software to facilitate this kind of metatranscriptome analysis. Possibly due to the presence of ART, viral reads were barely detectable in the BMC. To overcome this, we used public RNA-seq raw data that contained senescent cells from young and old animals. Finally, we further validated these findings by analyzing the cell-associated virus transcriptome on published single-cell RNA-seq data.
2.2 Methods

2.2.1 Animals information

Experimental rhesus macaques were housed at the Tulane National Primate Research Center (TNPRC), which is an AAALAC-accredited facility. IACUC-approved procedures were followed in accordance with the Guide for the Care and Use of Laboratory Animals. Animal housing rooms were maintained at temperatures between 64 and 84 °F (18–29 °C), with a light: dark cycle of 12:12 hours.

For the experiment to measure the senescent markers (lipofuscin, p21, CAV1, etc.) on the tissue sections by microscopy and RT-qPCR, we used archived formalin-fixed paraffin-embedded (FFPE) blocks. The tissues were cut into slides of 5 μm thickness for microscopy and 2×10 μm for RNA extraction. Frontal cortex tissues from 22 rhesus macaques (Table 1) were divided into four groups, which included SIVmac251-infected young (n=6, Mean 6.40 ± SD 0.79 years), SIVmac251-infected old-aged animals (n=6, Mean 17.83 ± SD 0.60 years), SIV-naïve young (n=6, Mean 6.91 ± SD 1.07 years) and old-aged (n=4, Mean 23.92 ± SD 3.94 years) animals for comparison.

For the iron test, which we discuss later, 7 SIV infected and ART-treated Chinese rhesus macaques were used, including four young (Mean 6.44 ± SD 0.48 years) and three old animals (Mean 17.45 ± SD 1.10 years) (Figure 4 A).

For the brain mononuclear cell (BMC) isolation, and RNA-seq for BMC, 6 SIV-infected and ART-treated Chinese rhesus macaques were used, which included three young (Mean 6.57 ± SD 0.49 years) and three old animals (Mean 17.45 ± SD 1.10 years) (Figure 6 A).
2.2.2 SIV infection

Rhesus macaques of Chinese origin were inoculated intravenously with 100 TCID50 of SIVmac239. The virus stock of SIVmac239 was collected via culture of CEMx174 cells and provided by the production core of TNPRC. Seven of the animals studied were intravenously inoculated with SIVmac239M, a barcoded virus with a 34 bp insertion between genomic regions of vpx and vpr in the SIVmac239 parent virus [179]. The SIVmac239M was generously provided by Dr. Brandon Keele from the Frederick National Laboratory for Cancer Research.

2.2.3 Antiretroviral therapy (ART)

A daily ART regimen of Tenofovir Disoproxil Fumarate (TDF) (5.1mg/kg), Emtricitabine (FTC) (30mg/kg), and Dolutegravir (DTG) (2.5mg/kg), was given subcutaneously once daily starting at 12 weeks post-infection and continuing for 8 months, followed by the first 21-day treatment interruption and resumed the ART for 4 months. Animals were euthanized 7 days after the second treatment interruption.

2.2.4 Serum sample collection

Individual 2 mL aliquots from whole blood samples collected by a trained veterinarian were stored in plastic tubes without anti-coagulants for biochemistry analysis. The blood samples were clotted at room temperature for 45 minutes. After centrifugation at 1500 g for 15 minutes, the serum was analyzed immediately for chemical levels.
2.2.5 Animal euthanasia and tissue collection

Euthanasia was performed in line with the recommendations from the Panel on Euthanasia of the American Veterinary Medical Association. Following Tulane IACUC standards of operation (SOP), SIV-infected and/or drug-treated macaques were euthanized firstly with telazol and buprenorphine, followed by an intravenous injection of sodium pentobarbital. The brain was perfused with PBS before necropsies. Fresh tissues from the brain were collected and soaked in 10% neutral buffered formalin (NBF) or stored in RNAlater™ Stabilization Solution during necropsies. The brain tissues were collected first in order to minimize the chances of contamination from the gut microbiome. After 16-32 hours following the necropsy, the fixed tissues in 10%NBF were used for paraffin-embedding.

2.2.6 Cell culture

CEMx174 cell line was maintained in RPMI 1640 Medium (ThermoFisher Cat# 11875093) supplemented with 10% heat-inactivated fetal bovine serum (FBS) (ThermoFisher Cat# 10438026) and antibiotics that were 100 units penicillin and 100 μg/mL streptomycin (Sigma-Aldrich Cat# P4333-100ML). The suspension culture was grown in culture flasks (VWR Cat# 10062-872 and Fisher Scientific Cat# 12-566-85) in a humidified incubator at 37°C with 5% CO2. The medium was renewed every 2 to 3 days. By adding or replacing fresh media, cultures can be maintained between $5 \times 10^5$ and $2 \times 10^6$ cells/mL until start the treatment.

DNA damage stress was used to induce cell senescence. Because of a lack of previous data on inducing cellular senescence on the CEMx174 cell line, we performed
preliminary tests and modified the working protocol [180]. The data from our preliminary tests showed that CEMx174 cells are prone to cell death after treatment. In order to acquire a proper number of cells for the following SIV infection test, we increased the initial number of cells for treatment.

CEMx174 cells in a concentration of $1 \times 10^6$ cells/mL (total $4.3 \times 10^6$ cells in 43 mL medium) were treated with 250 nM of doxorubicin (Sigma-Aldrich Cat# D1515-10MG) for 24 hours, followed by a wash with the medium. The cells were then incubated for 14 days to induce a senescent state. During this period, we monitored the cell viability, size, and concentration about every 2 days. Next, we centrifuged the cells at $130 \times g$ for 10 minutes, absorbed the medium carefully, then added a proper volume of fresh medium to maintain the cell concentration at $1 \times 10^6$ cells/mL.

On day 14, cells were divided into three wells in a new 24-well cell culture plate with $1 \times 10^6$ cells/mL in 0.5 mL medium for each well. 100 TCID$_{50}$ of SIVmac239 were added to each well. The cells were then harvested at time points of 48 hours, 72 hours, and 96 hours after adding the SIV. Lastly, the cells were centrifuged to a pellet at $250 \times g$ for 10 minutes then stored in RNAlater™ Stabilization Solution (ThermoFisher Cat# AM7021) for cell-associated SIV quantification. The virus stock of SIVmac239 was provided by the production core of TNPRC from CEMx174 cells [181].

Cell size was measured using an automated cell counter (Cellometer Auto 2000). The cell number count and viability analysis were performed by AO/PI Staining using ViaStain™ AOPI Staining Solution (Nexcelom Bioscience Cat# CS2-0106-5mL) on the Cellometer Auto 2000. Briefly, this method combines the use of nucleic acid binding dyes, acridine orange (AO) and propidium iodide (PI), to measure the viability and count the
numbers of cells. AO is cell-permeable, so it can stain all the cells to generate green fluorescence. PI can only enter cells with compromised membranes (e.g., dead cells) and generates red fluorescence. By quantifying the ratio of the two colors, cell viability can be calculated.

2.2.7 Lipofuscin measurement

The areas expressing lipofuscin were visualized by the overlapped autofluorescent signals in both fluorescent channels (423 nm and 555 nm excitation wavelength). The relative fold changes were calculated by the ratios of the lipofuscin area in the Regions of Interest (ROI) then normalized using the mean value of the young SIV- group. Image data quantification analysis was performed using the HALO software (Indica Labs).

2.2.8 RT-qPCR

Gene expression levels in the samples were measured by quantitative reverse transcription PCR (RT-qPCR), which was performed on the 7900HT Fast Real-Time PCR System. Pre-designed primers and probes were obtained from Thermofisher TaqMan Gene Expression Assays (Table 6), which has been successfully tested with the genome of rhesus macaque (Macaca mulatta) via NCBI BLAST [182]. The relative quantification of cellular senescence marker genes was expressed as $2^{-\Delta \Delta CT}$ by the comparative CT method and normalized to the expression of glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

For FFPE brain frontal cortex (Figure 3), RNA was extracted from FFPE brain frontal cortex using the Qiagen RNeasy FFPE Kit (QIAGEN Cat# 73504). cDNA was then
prepared with the Promega Reverse Transcription System (Promega Cat# A3500) and amplified using the TaqMan™ Fast Advanced Master Mix (ThermoFisher Cat# 4444963).

For doxorubicin-induced senescent CEMx174 cells (Figure 1B), about $3 \times 10^5$ cells were taken from the culture on day 14 to extract RNA using the RNeasy Mini Kit (QIAGEN Cat# 74104) followed by one-step RT-qPCR using TaqMan™ RNA-to-CT™ 1-Step Kit (ThermoFisher Cat# 4392938).

For cell-associated SIV quantification, the conserved gag region was used for designing the primers and probes that can detect SIVmac239, SIVmac251, and SHIV viruses. The primers and probe design, along with the SIV standards were provided by the Pathogen Detection and Quantification Core of Tulane National Primate Research Center [183]. RNase P gene was used to quantify genome copies and cell numbers in the sample. TaqMan™ RNase P Detection Reagents Kit (ThermoFisher Cat# 4316831), which includes the primers, probes, and standards, was used for RNase P gene quantification. The numbers of SIV copies were divided by the numbers of cells in the same sample.

2.2.9 Iron test

Serum iron quantification was performed using the iron measurement kit (OSR6286) on Beckman Coulter AU analyzers. Briefly, an acidic medium allows transferrin-bound iron to dissociate into apotransferrin and free ferric ions. Ferric ions are then reduced to a ferrous state by hydrochloric acid and sodium ascorbate. A blue-colored complex is formed when ferric ions react with chromogen TPTZ [2,4,6-Tri-(2-pyridyl)-5-triazine], which is detectable bichromatically at 600 and 800 nm. Transferrin-bound iron increases absorbance directly in proportion to its amount.
2.2.10 Brain mononuclear cells isolation

The brain mononuclear cells (BMC) isolated from the rhesus macaques on the day of the necropsy by using the protocol below:

1. Collect perfused brain tissues in the medium in 50 ml tubes.
2. Mince the tissue finely after removing the membrane and blood vessels. Then wash the tissue with 1xHBSS (with phenol red) once at 1500 RPM for 10 mins.
3. Discard the supernatant and collect the minced tissues in a flask.
4. Put approximately 15 ml of the enzymatic solution containing collagenase type 2 and DNase1 per 10 gm of minced tissue. Increase the volume of the enzymatic solution according to the minced tissue weight.
5. Put the flask in the 37°C in incubator for 45 mins and shake at 230 RPM.
6. Take out the flask and put the total solution in a 40 micro size steel mesh and plunge the tissue with the back of a syringe filter and collect it in 50 ml tubes.
7. Add 1xHBSS (with phenol red) to the filtrate 1:1 and spin the 50 ml tubes at 2000 RPM for 10 mins.
8. Discard the supernatant carefully as the fat layers tend to come out of the tube.
   Discard the fat layer as definitely as possible.
9. Wash the pellet once more with 1X HBSS at 1500 RPM for 10 mins.
10. Discard the supernatant.
11. Resuspend the pellet in 1.03 density Percoll. Use 25 ml of 1.03 density Percoll per 15 gm of tissue pellet (approx.).
12. Layer the suspension on top of 10 ml of 1.095 density Percoll.
13. Spin at 2500 RPM for 30 mins at 4 deg with zero ascending speed and zero brakes.

14. Discard the top fat layer very carefully and collect the cell layer in between the density gradients.

15. Wash the collected cell layer with three volumes of 1xHBSS (with phenol red) twice at 1500 RPM for 10 mins at 4 ℃.

16. Filter the cell pellet with a 40-micron cell filter.

17. Spin down the cell pellet and use ACK lysis buffer if RBC is present in the cell pellet.

18. Wash the cell pellet and count the cells.

2.2.11 Cell sorting

Isolated BMCs were stained with the LIVE/DEAD™ Fixable Aqua Stain then sorted based on their autofluorescence levels in the FL1 channel. Cells were kept on wet ice, and then sorted within one day from necropsy.

2.2.12 Bulk RNA-seq

RNA from sorted BMCs were prepared immediately using RNeasy Plus Mini Kit (QIAGEN Cat#74134). mRNA sequencing experiments were performed by Novogene. The PolyA-selected and non-strand specific library was prepared by NEBNext® UltraTM RNA Library Prep Kit for Illumina® (descending colon samples) or SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing by Takara (BMC samples). About 20 million
paired-end 150 bp reads per sample were generated from Illumina NovaSeq 6000 Sequencing System.

The RNA-seq raw reads stored in the FASTQ format were analyzed by using a homemade bioinformatic software - MTD, which can detect host and microbiome reads simultaneously and feature statical analysis and data visualization functions. A detailed description of MTD is outlined in the next chapter. The rhesus macaque genome (Mmul_10 assembly) and its annotation (Macaca_mulatta.Mmul_10.103.gtf) from the Ensembl genome database were used for reads alignment and counting. In addition, the step of differentially expressed genes (DEGs) analysis was modified for BMC samples. Here it can control the subject (animal) effect and the interaction between age and cellular autofluorescence. Significant DEG were defined by log2FoldChange > 0.5 and p-value < 0.05.

Enrichment analysis of DEGs was implemented independently by IPA (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis) [184], GSEA [185, 186], and the clusterProfiler R package [187].

2.2.13 Single cell RNA-seq analysis

Raw data of the singe-cell RNA-seq was downloaded from NCBI GEO with accession code GSE161340 (SRR13041553-13041560) using the SRA Toolkit. The detailed sample preparation method for the sequencing was described in the article [188]. Briefly, the single-cell suspension was prepared from hippocampi that were dissected from the brains of 4-month (young) and 24-month (old) C57BL/6 mice. This experiment utilized
a group of four young mice and another group of four old mice from which the hippocampi of two mice were pooled into one sample, creating a total of two samples per age group.

Reads type in SRA was first checked by the command like vdb-dump SRR13041553 -R1 -C READ_TYPE then converted to the raw FASTQ files by fasterq-dump function (e.g., fasterq-dump SRR13041553 --include-technical -S) in SRA Toolkit. Downstream processes were conducted on the corresponding biological and technical reads in FASTQ files.

The count matrix of the single-cell microbiome was generated through the MTD software. In R, the count matrixes of the microbiome and host from the same sample were first preprocessed and then merged by the rbind function. Downstream analysis was performed using the Seurat R package (Version 4.0.1) [189-192]. In Seurat, following the authors' method [188], the datasets were first normalized with the SCTransform function, then followed the integration workflow to joint analysis of multiple single-cell datasets.

SCTransform method is more effective than a traditional normalization workflow and better eliminates technical effects from the data. It allows higher PCs, which create a confounding factor that due to variation in sequencing depth, is substantially mitigated. Consequently, higher PCs represent more heterogeneity that is both subtle and biologically relevant. Thus, including them during downstream analysis can potentially improve results. Moreover, this method returns more variable features by default, which may represent more subtle biological changes [193].

Data under "SCT" assay of the Seurat object (e.g. project_name@assays[["SCT"]][@data]) was log-normalized using SCTTransform function. The top 3,000 highly variable genes/features among the datasets were used to find
"anchors" for integration. UMAP projections were calculated for the first 30 principal components of the log-normalized data. UMAP was chosen instead of the t-SNE because it could provide a better separation of identified cell types [194]. FindMarkers function with MAST methodology was used for computing the log-fold changes for each gene/feature between groups and their corresponding p- and adjusted p-values [195].

In addition, a homemade cell type classifier was used to identify senescent cells and other canonical cell types in the brain. This was achieved by performing the single sample Gene Set Enrichment Analysis (ssGSEA) [185, 196, 197] for each cell based on the gene expression profile of distinct cell types. We reference a list of senescence marker genes was from a recent single-cell analysis study on mice brains [198]. Marker gene lists of cell types were downloaded from the PanglaoDB database [199]. In the list, brain and vasculature were selected as organs containing cell types present in the brain.

A modified ssGSEA2.0 [197] program was used for the enrichment score calculation. High Performance Computing (HPC) cluster/server were used to accelerate the calculation. Because ssGSEA2.0 demands almost all (maximum threads -1) CPU processing capacity for its multi-threading computation, it makes its usage on the HPC cluster/server troublesome. This greedy occupation of the computational resources of the server can result in a failed run or create server issues for other users or administrators. Therefore, in order to make the program more compatible with its respective server, we have modified it to allow users to easily adjust the demands of threads to an appropriate value based on their actual computational environment.

Next, normalized enrichment scores (NES) were calculated for every cell within the data set based on the degree of absolute enrichment of a gene set. For each cell, the
cellular senescence score is equal to the NES of the senescence gene set, and the cell type is decided by the top-ranked NES of cell types. Senescent cells were defined as the NES > 3, based on the shape of the distribution curves.

The cell properties (e.g., cellular senescence group, senescent score, age group, viral infection status, cell types) were annotated and stored in the metadata of the Seurat object (e.g., project_name@meta.data). Viral abundance in the cell was indicated by the virus reads count ratio, which was calculated by dividing the total number of the reads that belong to the virus by the number of total reads in the cell. The correlations between cell types were examined by combining use of the cor (for correlation coefficient), cor.test (for p-value), and corrplot (for plot) functions in the Base R and corrplot R packages. The Spearman correlations between senescent cells and virus abundance were calculated and plotted by the ggscatter function in ggpubr R package. The source code of analysis is in the appendix.

2.2.14 Statistical analysis

An unpaired t-test was used to compare the difference between any given two groups. The Welch’s correlation was used when groups’ variances are unequal. A paired t-test was used to compare the difference between BMC groups of high or low autofluorescence. To assess correlations, Pearson's correlation tests were applied for parametric data, and Spearman's rank correlation tests were applied for non-parametric data. GraphPad Prism 9.0 statistical software (GraphPad Software, San Diego, California USA, www.graphpad.com) and R (version 4.0.3)[200] were used to analyze data. Statistical results were set two-tailed and \( P < 0.05 \) as significant unless otherwise specified. In
addition, data visualization was conducted through R (version 4.0.3), MTD software, GraphPad Prism 9.0, and Adobe Illustrator.
2.3 Results

2.3.1 *In silico* simulation of the interaction between HIV-1 and cellular senescence

To have an outlook of the relationship between HIV-1 and cellular senescence, we first performed an *in silico* simulation of their molecular interactions.

The data of proteins that interact with HIV-1 was downloaded from HIV-1 Human Interaction Database [201-203]. We further modified the dataset by categorizing the proteins that have clear interaction information as either up- or down-regulated by HIV-1 infection. The data of cell senescence-associated genes was downloaded from the CellAge database [18]. The dataset contained manually curated information about the senescence effect of the gene, such as inducing or inhibiting senescence. We coded the HIV-1 up-regulated proteins or senescence-inducing genes as a fold change equal to 2, otherwise as -2. Then, the *in silico* analysis was conducted on IPA.

Our analysis results predicted that HIV-1 infection could induce cellular senescence (Figure 1A-B). Furthermore, our results from gene functional enrichment showed that senescent genes could inhibit viral infection (Figure 1C). In the pathway analysis results of HIV-1 human interaction data, the neuroinflammation signaling pathway showed the highest \(-\log(p\text{-value})\) among all the enriched pathways, 125.484 for \(-\log(p\text{-value})\) and 6.965 for the z-score. The senescence pathway was enriched with both the high z-score (8.397) and \(-\log(p\text{-value})\) (66.346) (Figure 1B). Besides, the molecule activity prediction results also showed an increase in the cellular senescence after HIV-1 infection (Figure 1A). In the analysis results of cellular senescence genes, the functional annotation predicted the negative regulation of viral infections with a z-score of -1.671 for viral infections and -1.054 for replication of RNA virus (Figure 1C).
These analyses predict that, on the one hand, the HIV-1 infection induces cellular senescence, and on the other hand, the senescent cells may inhibit the HIV-1 infection. We further tested these hypotheses through \textit{in vitro} and \textit{in vivo} experiments by using the SIV and Rhesus Macaque animal models.
Figure 1. *In silico* simulation of the interaction between HIV-1 and cellular senescence.

A) The effect of HIV-1 on the cellular senescence pathway: The molecules in orange or blue colors indicate the predicted results of activation or inhibition. B) The representative pathway enrichment results of HIV-1 human interaction data. The numbers on each bar represent the z-score. C) The results of IPA functional annotation on the cellular senescent genes. The numbers on the heatmap indicated the z-score for the function.
2.3.2 Reduced replication of SIV in senescent CEMx174 cells

To test whether SIV infection could be inhibited by cellular senescence, we first performed an *in vitro* experiment using the CEMx174 cell line. CEMx174 is a cloned hybrid between the LCL 721 B lymphoblastic and CEM T lymphoblastic cell lines. It can be easily infected by SIV and has been widely used for SIV production and *in vitro* experiments. However, there was little knowledge about the senescent cell induction on this cell line. We treated CEMx174 with the chemotherapeutic DNA damaging agent doxorubicin (Dox) to induce it to senescence. We found that CEMx174 was more vulnerable to stimuli than the cell lines derived from the epithelium or fibroblasts. The live-cell population and cell viability decreased quickly in the first four days after treatment, then remained stable after day 6 (Figure 1A). This suggests that both the cell proliferation and apoptosis were reduced after day 6. On day 14, we evaluated the cells for the presence of cellular senescence markers. Doxorubicin-treated cells showed increased cell size (Figure 1B), and expression of a group of senescent markers (p16, p21, PAI-1, CCND1, CAV1, and IL6), and decreased expression of cell proliferation marker Ki67 (Figure 1C), which indicated senescence induction.

To test whether senescent cells were more resistant to SIV infection, untreated (non-senescence control) and induced senescent CEMx174 cells were inoculated with 100 TCID50 of SIVmac239. We quantified cell-associated SIV RNA by RT-qPCR at different time points after inoculation. As shown in Figure 1D, the virus copy numbers in senescent cells were lower than in non-senescent cells, indicating that senescent CEMx174 inhibits SIV replication. Although we observed high variance in the cell-associated SIV viral load within the group that increasing the p-values of the test (p = 0.077 at 72h and p = 0.099 at
96 h), the trend of the gaps of SIV copy numbers between senescent and non-senescent cells increased with time after inoculation. For example, all the triplicates of SIV copy numbers were much smaller in senescent cells than non-senescent at 96h. Based on these results, replication of SIV was inhibited in senescent CEMx174 compared to non-senescent cells.
Figure 2. Induced senescence of CEMx174 cells restricts SIV infection in vitro.

(A) CEMx174 cell number and viability after doxorubicin treatment. On day 14 after the treatment, CEMx174 cell size was measured (B), and gene expression levels of cellular senescence markers in treated CEMx174 cells compared to untreated control were analyzed by RT-qPCR (C). (D) Quantification of cell-associated SIV copy number by RT-qPCR on the time point 48, 72, and 96 hours after inoculation; the black dots are the senescent cell (Sc), and the circles are non-senescent cells (Ctl). Data were presented as mean value +/- SEM. *** p < 0.001 One-tailed Student's T-test.
2.3.3 SIV infection contributes to accelerating brain cellular senescence in young but not in old rhesus macaques.

2.3.3.1 Lipofuscin significantly increased in the brain of SIV-infected young, but not in old, rhesus macaques

In order to study the relationship between SIV and cellular senescence in the brain *in vivo*, we first analyzed archived FFPE frontal cortex tissue from rhesus macaques, which provided a sufficient sample size for groups in different SIV infection statuses and ages. The animal and grouping information are shown in Table 1. Lipofuscin detection by HALO software is exemplified in Figure 2A-D, and the quantification results are shown in Figure 2E.

As expected, in healthy SIV-naïve groups, a significantly higher amount of lipofuscin was observed in old animals than in young animals (Figure 2E). Interestingly, the significance of this age-dependent discrepancy disappeared between groups of young and old animals with SIV infection, although both groups had higher levels of lipofuscin than the young uninfected group. Moreover, the increase of lipofuscin was significantly higher in SIV-infected young animals than in those of age-matched animals without SIV infection. However, this increase was not observed between the older groups of animals with or without SIV infection.
Table 1. Animal information of cellular senescence test on FFPE brain tissues.

<table>
<thead>
<tr>
<th>Animal #</th>
<th>Sex</th>
<th>Age (Years)</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>H972</td>
<td>FEMALE</td>
<td>18.58</td>
<td>Old SIV+</td>
</tr>
<tr>
<td>I774</td>
<td>FEMALE</td>
<td>18.45</td>
<td>Old SIV+</td>
</tr>
<tr>
<td>FB04</td>
<td>MALE</td>
<td>17.86</td>
<td>Old SIV+</td>
</tr>
<tr>
<td>L618</td>
<td>MALE</td>
<td>17.6</td>
<td>Old SIV+</td>
</tr>
<tr>
<td>DT55</td>
<td>FEMALE</td>
<td>17.45</td>
<td>Old SIV+</td>
</tr>
<tr>
<td>P427</td>
<td>MALE</td>
<td>17.01</td>
<td>Old SIV+</td>
</tr>
<tr>
<td>N444</td>
<td>MALE</td>
<td>26.14</td>
<td>Old SIV-</td>
</tr>
<tr>
<td>I161</td>
<td>MALE</td>
<td>22.22</td>
<td>Old SIV-</td>
</tr>
<tr>
<td>FA98</td>
<td>MALE</td>
<td>19.26</td>
<td>Old SIV-</td>
</tr>
<tr>
<td>M083</td>
<td>FEMALE</td>
<td>28.06</td>
<td>Old SIV-</td>
</tr>
<tr>
<td>JH29</td>
<td>FEMALE</td>
<td>5.71</td>
<td>Young SIV+</td>
</tr>
<tr>
<td>IF44</td>
<td>FEMALE</td>
<td>6.42</td>
<td>Young SIV+</td>
</tr>
<tr>
<td>GD41</td>
<td>MALE</td>
<td>7.43</td>
<td>Young SIV+</td>
</tr>
<tr>
<td>IF29</td>
<td>FEMALE</td>
<td>6.94</td>
<td>Young SIV+</td>
</tr>
<tr>
<td>GB40</td>
<td>MALE</td>
<td>6.6</td>
<td>Young SIV+</td>
</tr>
<tr>
<td>GJ23</td>
<td>MALE</td>
<td>5.27</td>
<td>Young SIV+</td>
</tr>
<tr>
<td>JM95</td>
<td>FEMALE</td>
<td>7.35</td>
<td>Young SIV-</td>
</tr>
<tr>
<td>JR95</td>
<td>MALE</td>
<td>6.03</td>
<td>Young SIV-</td>
</tr>
<tr>
<td>KC32</td>
<td>MALE</td>
<td>5.16</td>
<td>Young SIV-</td>
</tr>
<tr>
<td>IL53</td>
<td>MALE</td>
<td>7.89</td>
<td>Young SIV-</td>
</tr>
<tr>
<td>IJ61</td>
<td>MALE</td>
<td>7.62</td>
<td>Young SIV-</td>
</tr>
<tr>
<td>IM10</td>
<td>FEMALE</td>
<td>7.41</td>
<td>Young SIV-</td>
</tr>
</tbody>
</table>

Table 1. Animal information of cellular senescence tests on FFPE brain tissues. Four groups of rhesus macaques were studied, which included SIVmac251-infected young (n=6, Mean 6.40 ± SD 0.79 years), SIVmac251-infected old animals (n=6, Mean 17.83 ± SD 0.60 years), SIV-naïve young (n=6, Mean 6.91 ± SD 1.07 years) and old aged (n=4, Mean 23.92 ± SD 3.94 years) animals for comparison.
Figure 3. Frontal cortex lipofuscin quantification.

Frontal cortex lipofuscin quantification. Frontal cortex of SIV infected (A&B) and naïve (C&D) young, rhesus macaques. SIV infected, young, rhesus macaques have higher numbers of cells that contain lipofuscin (A). In naïve, young, rhesus macaques, only few cells contain lipofuscin (C). Overlays of the HALO analysis results illustrate the autofluorescent area detected (yellow) in SIV-infected and naïve, young rhesus macaques (B&D respectively). The red arrow indicates the detected lipofuscin. The dotted box represents nonspecific autofluorescent areas which were excluded from our analysis. The relative fold change of lipofuscin in different groups (E). Plot: mean with SD, *: P < 0.05, **: P < 0.01. Student’s t-test.
2.3.3.2 SIV infection led to increased cellular senescence in the brain of SIV-infected young but not in old rhesus macaques

In addition to the lipofuscin, we examined gene expression levels of several senescent markers through RT-qPCR. We found CAV1, CCND1, and p21 genes differentially expressed between groups (Figure 3).

CAV1 gene expression was significantly increased in the SIV-infected young animals (Figure 3A). CAV1 is an important component of caveolae, cholesterol-enriched microdomains on the plasma membrane. It influences the morphology changes of senescent cells. Several studies have shown that CAV1 is elevated in diseases associated with organismal aging and that it regulates cellular senescence through activation of the p53 pathway and cooperates with the MAP kinase pathways [49]. Recently, it has been shown to be upregulated in senescent brain cells that have been exposed to supernatants from cells infected with HIV-1 [68].

CCND1 was significantly higher in uninfected older animals than in uninfected young animals, but SIV infection of young animals reduced this difference to insignificant (Figure 3B). CCND1 is a regulatory subunit of CDK4 or CDK6, whose activity is necessary for G1/S transition in the cell cycle [204]. In addition, it has been recently identified as a common senescent marker gene in several senescent cell types [205].

p21 showed a similar pattern of these changes, but it was not statistically significant (Figure 3C). In addition, p16 gene expression was undetectable in all brain frontal cortex samples. This was supported by previous findings that the p16 protein level was low and without obvious change in the brain microglia cells after HIV-1 infection [68], and its mRNA level was much lower than protein in the brain [206]. This may be due to the
relatively low proportion of senescent cells in the brain tissues (especially for p16), making the changes difficult to measure by the RT-qPCR method, which is further exacerbated by mRNA degradation in the archived FFPE tissue.

In general, in the young groups, SIV-infected animals had higher expression levels of CAV1, CCND1, and p21 than uninfected age-matched animals. However, these brain cellular senescence changes were not obvious in old animals.

We observed the age difference in the increasing of senescent cells abundance in the brain of SIV-infected animals. Given that senescent cells in the brain contribute to cognitive decline and neurodegeneration, our findings indicate that they play an important role in the acceleration of brain aging in young hosts and potentially contribute to the development of HIV-associated neurocognitive disorders.
Figure 4. Gene expression levels of cellular senescence markers in brain.

(A) CCND1 gene expression level (B) CAV1 gene expression level (C) p21 gene expression level. Gene expression levels were measured by RT-PCR. Plot: mean with SD, *: $P < 0.05$. Student’s t-test.
2.3.4 Age difference in the change of the serum iron level after SIV infection and ART

Serum iron levels have been associated with the severity of inflammation and infectious diseases [207, 208]. Higher serum iron levels strengthen the viral infection and is correlated with poor prognosis [209]. Previous studies showed that serum iron levels were increased in HIV-positive patients, and this increase did not change after ART [210].

We found age disparities in the change of the serum iron level after SIV infection and ART (Figure 5). In the first three months after SIV infection and ART, although the mean serum iron level was lower in old than in young animals, there was no significant difference (Figure 5B). However, after one year of SIV infection and ART, the difference became significant. Although iron levels in the old group did not have a noticeable change compared to one year ago, the younger rhesus macaques had a significantly higher serum iron level than the older cohort (Figure 5C).

These results indicated the age difference in serum iron levels in response to the SIV infection and ART. Old animals were more reluctant to changes in the iron levels than young animals. Thus, it may be inferred that young animals were more susceptible to iron metabolism dysregulation after SIV infection and ART. This iron derangement may contribute to increased oxidative stress and depletion of the CD4⁺ T cell population [211]. Moreover, the age disparities in iron metabolism and its health implication during SIV infection and ART warrants further investigation.
Figure 5. Age disparities in the change of the serum iron level after SIV infection and ART.

(A) Basic information of Chinese rhesus macaques used in this study. (B) Serum iron levels in the first three months after SIV infection and ART. (C) Serum iron levels in the three months after one year since SIV infection and ART. Plot: mean with SD, *: $P < 0.05$. Student’s t-test.
2.3.5 Age disparities in the anti-viral function of senescent cells by transcriptome analysis of BMC

To further study the effects of SIV infection on the brain, we focused on brain mononuclear cells (BMC), which are targets of SIV and play an important role in brain homeostasis. We performed mRNA-seq followed by transcriptome analysis to determine whether senescent BMCs have enriched pathways that inhibit the viral infection in vivo.

2.3.5.1 Distinctive transcriptome profiles between high and low autofluorescent BMC

BMC were isolated from SIV-infected Chinese rhesus macaques on ART for about 16 months. The brains were perfused during the necropsy to exclude the mononuclear cells from the blood vessel. The basic information of animals in this study can be found in Figure 6A. Animals were divided into two age groups, old (Mean 17.45 ± SD 1.10 years old) and young (Mean 6.57 ± SD 0.49 years old). For each animal, senescent cells were sorted based on their innate higher autofluorescent levels [176, 177]. Thus, we separated the BMC into two groups: high autofluorescent BMC (BMC_H) that enriched senescent cells, and low autofluorescent BMC (BMC_L) that enriched non-senescent cells.

The RNA-seq results showed that the transcriptional profiles were obviously different between BMC_H and BMC_L (Figure 6B-C). Their transcriptomes were clearly separated by Principal Component Analysis (PCA). The samples in the BMC_L group were on the left side, and samples in the BMC_H group were on the right side of the red dash line on the two-dimensional PCA plot (Figure 6C). This meant BMCs were separated into two groups of either low or high cell autofluorescent levels.

Moreover, the samples were further divided samples into four groups by their ages: BMC_H in old animals (Old_H), BMC_L in old animals (Old_L), BMC_H in young
animals (Young_H), and BMC_L in young animals (Young_L). On the PCA plot, we found that these four groups were also well separated from each other, which indicated the difference among their transcriptomes. In addition, the transcriptomes from the old animals were more likely to cluster together, compared to those from young animals. Such as the BMC samples in Old_H (dots in red color) or Old_L (dots in green color) were closer to each other than in Young_H (dots in blue color) or Young_L (dots in purple color). Thus, these results demonstrate the age disparities within BMC_H and BMC_L. BMCs from old animals had more homogeneous transcriptional profiles than BMCs from young animals. Therefore, age may change the function of BMCs. Next, we further examine if the senescent cells were enriched in BMC_H and the influence of age on these senescent BMCs.
Figure 6. Different transcriptome profiles between groups.

(A) Basic information of Chinese rhesus macaques in this study (B) The heatmap of differentially expressed genes (DEGs) between BMC_H and BMC_L; the higher or lower expressed genes were shown in red or blue, respectively. (C) PCA of the gene expression
patterns; each dot represents a sample, and each color represents a subgroup; four subgroups were generated by BMC_H or BMC_L in the old and young animals: BMC_H in old animals (Old_H), BMC_L in old animals (Old_L), BMC_H in young animals (Young_H), and BMC_L in young animals (Young_L); the dashed line separates the BMC_H and BMC_L groups. Data visualized on heatmap and PCA plots were normalized and transformed using the vst function in the Deseq2 R package.
2.3.5.2 Senescence-like functional changes in BMC_H

To check if the senescent cells were enriched in BMC_H, a pathways analysis was performed using the IPA software with the DEG of BMC_H that normalized to BMC_L. The enrichment results showed the distinctive changes in biological functions and pathways in BMC_H (Figure 7).

The functional annotation results (Figure 7A) showed that BMC_H cells increased tumor initiation, such as myeloid or lymphoid neoplasm and tumorigenesis. However, it decreased tumor promotion, including functions of cell movement, migration of cells, invasive tumor formation, invasion of cells, and proliferation of cells. In addition, it increased the functions of morbidity or mortality but reduced the death of immune cells, cell viability, and cellular homeostasis. Those functions of BMC_H were similar to the senescent cells that are known to be caused by tumorigenesis stimuli but respond with anti-tumor functions such as inhibiting cell proliferation and migration. Meanwhile, senescent cells have reduced cell homeostasis, viability, and cell death, leading to higher morbidity or mortality of the organism. Overall, these senescence-like characteristics suggested the enrichment of the senescent cells in BMC_H.

Moreover, the functions of viral infection (z-score= -2.367) and infection by RNA virus (z-score= -2.145) were decreased, which agreed with our previous in silico simulation. Interestingly, there was an increase in the function of infection of mammalia and the coronavirus pathogenesis pathway (Figure 7B). These results may suggest that the senescent BMC has a general function in inhibiting viral infection, but this inhibition effect is not absolute. It may not inhibit some other kinds of pathogens such as bacteria, parasites,
or some viruses. In coronavirus, it was even shown to promote the pathogenesis of the infection [212], which was in accordance with our results.
Figure 7. Pathway analysis results of BMC_H.

(A) Biological functional annotations of BMC_H; the length of the bar represents the number of molecules/DEG enriched in this function; results on the plot were all significant with p-value < 0.001 (B) Enriched pathways in BMC_H; the length of the bar represents -log(p-value). The z-score is shown beside each bar and represented by the color of the bar. The analysis was performed through IPA software. DEG of BMC_H was normalized to BMC_L. DEG was defined by log2FoldChange > 0.5 and p-value < 0.05.
2.3.5.3 Age difference in BMC_H

Next, we examined the age difference in the functional changes of BMC_H, which represented the senescent BMC, by comparing the functional and pathway enrichment results in old and young animals. In general, BMC_H from the old animals expressed enhanced functional characteristics of cellular senescence (Figure 8A). Furthermore, compared to the young, the old animals had more significant changes through either the z-scores or p-values of the functional annotations.

For the function of infections, we gathered three enrichment results, such as viral infection, infection by DNA virus, and infection of mammalia (Figure 8A). It is noteworthy that the viral infection was more significantly decreased in BMC_H from old (z-score = -3.833, -log(p-value) = 32.79) than from young (z-score = -1.47, -log(p-value) = 21.46) animals. In addition, the infection of mammalia increased less evidently in old (z-score = 1.838, -log(p-value) = 26.29) than young (z-score = 3.684, -log(p-value) = 21.17). However, only the young animal showed an enrichment result of infection by DNA virus with a z-score of 1.394.

In the pathway enrichment results (Figure 8B), cell death was more inhibited in BMC_H from old animals, including via ferroptosis or by induction of apoptosis by HIV-1. Furthermore, some signaling pathways showed opposing results between old and young. For example, the pathways that were increasing in old but decreasing in young animals were neuroinflammation, Toll-like receptor (TLR) signaling, the role of pattern recognition receptors in recognition of bacteria and viruses, inflammasome, mTOR, macropinocytosis, JAK/Stat signaling, and coronavirus pathogenesis pathways. Among them, some signaling pathways such as inflammasome, mTOR, macropinocytosis, and JAK/Stat signaling have
been well documented as characteristics of elevated expression in senescent cells [213-218]. It has been suggested that neuroinflammation could be the result of cellular senescence in the brain [219]. In addition, the elevated pathways of TLR signaling, and the role of pattern recognition receptors for bacteria and viruses may contribute to the additional virus inhibition function of senescent BMC in old animals. Moreover, the coronavirus pathogenesis pathway only increased in the senescent BMC from old animals, which might partially explain the age difference in response to the coronavirus infection.

Overall, we observed the age difference in the functions of senescent BMC, which demonstrated accentuated cellular senescence profiles in old animals. In addition, senescent BMC from old animals had a more remarkable ability to inhibit viral infection in general.
Figure 8. Age difference in the functional changes of BMC_H.

The figure shows the comparison of the biological function (A) and canonical pathways (B) of BMC_H from old and young animals. The color of the dot represents the z-score, and the size of the dot indicates the -log(p-value). Analysis was performed through IPA software. DEG of BMC_H was normalized to corresponding BMC_L. DEG was defined by log2FoldChange > 0.5 and p-value < 0.05.
2.3.5.4 Cell type deconvolution

Cell type deconvolution analyses were performed on the RNA-seq data to give us an idea of the cell type composite the BMC. We first used GSEA with the C8 on the MSigDB database that contains cell type signature gene sets to identify the most abundant cell type in the senescent BMC. The top three results showed that microglia is the primary cell type enriched in the senescent BMC (Figure 9A). Then we conducted a cell type deconvolution analysis using CIBERSORTx with the LM22, a validated leukocyte gene signature matrix [220, 221]. LM22 contains gene sets that identify 22 human hematopoietic cell phenotypes, including myeloid subsets, naïve and memory B cells, natural killer (NK) cells, plasma cells, and seven T-cell types. Although LM22 did not include microglia, it could help prospect cell-type classes in BMC. The results demonstrated that macrophage M2 was the most abundant cell type in the senescent BMC (Figure 9B). In contrast, it was almost undetectable in the non-senescent BMC. In non-senescent BMC, CD8⁺ T cells were the most abundant cell type, making up 31%~55% of the population (Figure 9B). In addition, the BMC from the old animals showed more evident enrichment of cell types in different senescence statuses: more macrophage M2 for senescence and more CD8⁺ T cells for non-senescence.

Studies have shown that macrophage M2 shows more similarity to senescent cells. The shared several phenotypes, including SA-β-gal and p16-positive [222], which could imply their transcriptional similarity. This may explain the results that more macrophage M2 were enriched in the senescent BMC.

Moreover, a new study showed the accumulation of CD8⁺ T cells in the aged brain [223]. This result was supported by our findings on the BMC of the old rhesus monkey.
The same researchers also demonstrated that these cytotoxic T cells drove axon degeneration in old mice and led to age-related cognitive and motor decline, implicating its detrimental role in the process of brain aging in rhesus monkey or human.

Overall, the cell type deconvolution results showed that myeloid cells were the most abundant cell type in the senescent BMC. It also supported our previous findings that the senescent BMC from old animals have a more distinct, relevant phenotype.
Figure 9. Cell type deconvolution analysis results.

(A) Top three results of GSEA with the C8 on the MSigDB database. (B) Result of CIBERSORTx by using the LM22 as the reference; the numbers in the table indicate the proportion of the cell type; the high proportion was in red color, and the low proportion is shown in white and blue.
2.3.6 Microbiome analysis of BMC

To further validate the findings that the anti-viral function of senescent BMC, directly interrogates the viral transcripts inside the same BMC RNA-seq data would be an ideal method. It allows us to verify whether senescent BMC harbored more abundance of viruses than non-senescent BMC. Therefore, we developed a novel analysis software - MTD, which readily fits this purpose.

Through our MTD software, the RNA of the microbiome (including virome) was examined using the same RNA-seq data as the host transcriptome analysis. Figure 10 is a heatmap showing all the microbiome species detected in the BMC samples. Figure 11 presents the taxonomic and phylogenetic trees of these microorganisms. Likely due to the rhesus macaques being under cART for more than one year, the brain tissues were less infected by the virus.
Figure 10. Heatmap of the microbiome species in BMC samples.

The higher abundance of microbiome reads is shown in the lighter shade of blue, and the lower abundance is shown in the darker shade. The data was normalized with Deseq2 and plotted by the phyloseq R packages, which were wrapped in MTD.
Figure 11. The taxonomy graph of the microbiome in the BMC.

The outer rings on the graph outline each group from our samples. The colored regions on the rings represent the species from the corresponding samples, and the darker-shaded colors highlight the differentially expressed species of each group, which was identified by DESeq2 R packages [224]. The blank regions across all the rings indicate that the p-values of the species were set to NA when performing the differential expression analysis. In our case, the possible reasons are either most of the samples have zero counts for that species
or it’s an extreme count outlier [225]. The graph was generated by a modified GraPhlAn program that was integrated into the MTD. Annotation levels were set to 2 (phylum), 3 (class), 4 (order), 5 (family), 6 (genus). Level 7 (species) was used as the external legend for the annotation. 50 most abundance clades were highlighted on the graph.
Figure 12A-B shows the differentially expressed microbiome species and the relative abundances of microbiomes at the phylum level. Only a few differentially expressed microbiomes were found between Young_H vs. Young_L, and none were found in the Old_H vs. Old_L. The clusters of old and young groups were separated by PCoA and indicated the general disparities in the microbiome diversity between the BMC from old and young animals (Figure 12C). In addition, the analysis of diversity showed a significant difference among the groups of Old_H, Old_L, Young_H, and Young_L (Figure 12D). We also examined reads that did not belong to the host, including all the microbiome information, classified and unclassified (Figure 12E). Only young_L showed a relatively higher non-host reads ratio than other groups, which imply a higher overall infection status in the non-senescent BMC from young animals. However, the alpha-diversity was lowest in the young_L (Figure 12F). The results may suggest that young_L infected with certain kinds of the microbiome have a higher abundance, whereas the other three groups had a higher within-group diversity but a lower abundance of those microorganisms in the infections.
Figure 12. Microbiome species analysis of BMC.

(A) The differentially expressed microbiome species. Data were normalized and transformed by vst function in the Deseq2 R package. (B) Relative abundance of the...
microbiome in the phylum level. (C) PCoA plot using Bray-Curtis distances (beta-diversity). (D) Analysis of diversity in groups using the anosim function in vegan R package [226]. (E) Difference of non-host reads ratio between groups examined by t-test. (F) Shannon alpha-diversity between groups examined by t-test.
2.3.7 Age difference in the anti-viral function of senescent brain cells

Because limited numbers of viruses were found in the BMC from ART animals, whether virus abundance was lower in the senescent cell than in non-senescent cells \textit{in vivo} remained unclear. Thus, we considered utilizing the publicly available omics data to answer this question. Here, we applied the single-cell RNA-seq data from an \textit{in vivo} study that included both young and aged animals, which ideally fit our purpose. Using the single-cell RNA-seq, senescent cells were identified from the tissues of young and old animals separately, then the viruses inside each cell were examined. Nonetheless, we still lacked a suitable tool for exploring virome in single-cell data. Therefore, we have developed a novel pipeline in MTD for the microbiome (including virome, bacteria, protozoa, fungi, plasmid, and vectors) analysis of the single-cell RNA-seq raw data. Hereby, we performed analysis by using the data recently published on GEO (GSE161340) that met our needs.

The raw data from single-cell RNA-seq that we analyzed was sourced from brain tissues of the healthy young and naturally aged mice. The clusters of different cell types in the brains were visualized on UMAP and split by animal age (Figure 13A). There was a total of 7,395 single-cell transcriptomes, including 2,926 from aged and 4,469 from young mice. Cell types were similar in both young and old brains.

Due to the inherent difficulty for single-cell sequencing to detect genes expressed at low levels, we calculated normalized the enrichment score (NES) for the whole set of senescence marker genes for each cell. The advantage of this method is that can assess both amplitude and number of senescence-related gene expressions in each cell. A density-smoothed distribution of senescence gene enrichment scores from young and aged brains is shown in Figure 13B. An overall distribution pattern shifted to higher senescence scores
was observed in the cells from old brains. This change varied across different cell types. Oligodendrocytes, neurons, microglia, and ependymal cells displayed the age-related shift in senescence scores more obviously. There were more microglia cells with high senescence gene enrichment scores in aged brains compared to young brains (Figure 13C). The senescent cells were visualized on UMAP as orange dots for young and old brains (Figure 13D). The old brain has a generally higher percentage of senescent cells than the young.

The microbiome in each cell was detected by using MTD software, and the cells contained virus reads were labeled. Figure 14A shows virus-positive cells in red on the UMAP. Old mice had a higher percentage of virus-positive cells in the brain compared to young. The most abundant viruses are listed in Figure 14B for the old and young brain.
Figure 13. Identification of canonical cell types and senescent cells.
(A) Plots of two-dimensional UMAP for a total of 7,395 cells based on the differentially expressed marker genes, colored according to cell types. There were 2,926 and 4,469 cells in aged and young brains, respectively. Cell type identity was assigned based on the homemade program as described in the “Methods.” Similar cell types were observed in samples derived from aged and young mouse brains. (B) A density-smoothed distribution of senescence gene enrichment scores in young and aged brains. (C) Distribution of senescence gene enrichment scores in main cell types. (D) Cells with high expression of senescence markers (senescence score > 3) are highlighted on UMAP plots, and the percentage of senescent cells is shown as pie charts for young and old brains, separately.
Figure 14. Virus-positive cells.

(A) Virus-positive cells are shown on UMAP in red color. The pie chart indicates the percentage of those cells in the total brain cells. (B) Top 10 abundant viruses in the brain of old and young mice.
Through the single-cell RNA-seq analysis, the viruses inside each cell were identified and the cells that harbored the virus were extracted. Then, the senescence status was classified for those cells. Finally, the difference of viral abundance could then be compared between senescent and non-senescent cells in aged and young animals separately.

We found that in old animals, viral abundance was significantly lesser in senescent than in non-senescent cells (Figure 15A). However, such a difference was not observed in young animals (Figure 15B). In addition to using the binomial definition of cellular senescence status, we also performed the Pearson correlation tests by using the cellular senescence score as a continuous variable. The correlation test showed a significant negative association between virus abundance and cell senescence score. This association is only apparent in the old animals (Figure 15C-D). Thus, those results consistently indicated that senescent brain cells have lesser viruses in old but not young animals.

Next, we conducted the IPA functional enrichment analysis on the transcriptomes of these brain senescent cells from old and young mice. As expected, the function of viral infection was decreased more evidently in senescent cells from the old than from the young brains, with both the degree (z-score: -4.197 vs. -0.791) and significance (-log(p-value): 12.313 vs. 4.312) of the inhibition (Figure 15E).

As a result, our findings highlighted the age difference in the anti-viral function of senescent cells in the brain. Senescent cells from old animals have a higher capability to manifest viral inhibition than those from young animals.
Figure 15. Age difference in the anti-viral function of senescent brain cells.
(A-B) Violin plot of virus abundance (virus reads count ratio) in senescent (orange) and non-senescent cells (green) from old (A) and young (B) mice. Each dot on the plot represents virus abundance in a cell. Comparison by t-test; ***: p-value < 0.001. (C-D) Spearman correlations between virus abundance and cellular senescence score in old (C) and young (D) mice brain. (E) IPA functional enrichment results of senescent cells in mice brain. Viral infection-related functional annotations were plotted. The dot in darker blue and larger size represents a smaller z-score and p-value. The p-value was -log10() transformed for visualization.
2.3.8 Age difference in the brain cell types of viral infection

The single-cell RNA-seq data offers researchers greater convenience than bulk RNA-seq to evaluate the cell type changes during the aging process and viral infection. Here, we found that the most prevalent population in the sample was microglia, whose distribution increased in the aged brain from 0.45 in young to 0.61 in old. In contrast, other cell types such as oligodendrocytes, endothelial cells, and astrocytes were decreased in the old brain (Figure 16A-B). The log2 fold changes of old versus young in these cell types are shown by the grey bars in Figure 16C.

Moreover, we found that the fractions of cell populations shifted in the virus-infected cells, and these changes varied with age difference. In the old brain, the proportion of microglia increased from 0.61 to 0.66, whereas in the young brain, there was not much change (0.45 to 0.44). The most apparent increase in cell types was the neurons, which increased from 0.08 to 0.12. In addition, astrocyte and endothelial cells were decreased in both old and young brains. Additionally, compared to the young, the old brain had more oligodendrocyte progenitor cells, and this proportion was largely increased in the virus-infected cells (Figure 16C).

Overall, microglia and oligodendrocyte progenitor cells in the old brain were more likely targeted by the viruses, whereas neurons were preferred in the young brain. In addition, the higher baseline proportion of microglia may explain a higher percentage of viral-infected cells in the old brain (Figure 14A). In young animals, the viruses were more prone to infect the neurons rather than microglia, which implicated neurological damage.
Figure 16. Changes in the proportions of cell populations from old and young mice brains. Bar graphs show changes in proportions of cell types in all cells and virus-infected cells in young (A) and old mice brain tissue (B). (C) shows log2 fold change of cell populations in old vs. young brains for all cells and viral-infected cells.
2.3.9 Cell-type correlation analysis

Canonical cell types were defined by various characteristics phenotypically and genotypically, such as tissue locations, morphologies, functions, and gene expression patterns. However, a defined "cell type" could have considerable heterogeneity and unstable status in the tissue [227]. This begins to blur the boundary among cell types and the definition of cell status, which triggers contemplation of the discrete versus continuous nature of cell types. Furthermore, cellular senescence adds an aging dimension to each canonical cell type. Therefore, it is compelling to investigate the internal associations of those cell types. Here, the single-cell RNA-seq data allowed us to examine their associations at the transcriptome level. Cell type scores based on gene expression patterns as the continuous variables were used in the correlation test.

We found that microglia and endothelial cells were the cell types that have the highest correlations with the senescent cells in old and young animals, respectively (Figure 17). Microglia have the highest correlation ($R = 0.29$), and endothelial cells have the second-highest correlation ($R = 0.28$) with senescent cells in old animals (Figure 17A). However, in the young animals, endothelial cells were the cell type most significantly correlated with senescent cells ($R = 0.34$), while the correlation of microglia decreased ($R = 0.2$) (Figure 17B). Nevertheless, the astrocytes that had been reported to play an important role in brain senescence did not show an obvious correlation with senescent cells in our results.

Considering that we found an enrichment of microglia in senescent BMC from old rhesus macaques (Figure 9A), our results indicate that microglia provide the greatest contribution to cellular senescence in the brain of old animals. In addition, the endothelial
cells also showed a distinct association with senescence, especially in the young brain, which emphasized its role in cellular senescence in the brain [198].
Cell-type correlation in old mice brain
Figure 17. Cell-type correlation analysis.

The plots show the Pearson correlations between cell types in the brain from old (A) and young (B) mice. Non-significant results are shown as blank; positive correlations are shown in red and negative correlations are shown in blue. Correlation coefficients can be found on each grid in the heatmap. Significance was determined by $P < 0.05$. 
2.4 Conclusion

Our *in silico* simulation predicted it would be extensive effects that, on the one hand, HIV-1 could induce cellular senescence, and on the other hand, senescent cells could inhibit viral infection. The subsequent *in vitro* experiment supported that senescent cells could inhibit SIV infection directly. Because cellular senescence is a hallmark of aging, the burden of which increased in old tissue such as the brain, we further examined the age difference in the brain after SIV infection by conducting *in vivo* experiments. We observed that the effect of SIV-induced brain cellular senescence diminished in old animals. Furthermore, the senescent brain mononuclear cells downregulated viral infection pathways, which was more significant in old animals. Then, we further validated the antiviral function in the senescent cell by performing a metatranscriptome analysis in single-cell RNA-seq data. Our results revealed that the total virus abundance was significantly decreased in the senescent brain cells from old animals exclusively, with a more evident downregulation of viral infection pathways.

This study investigated the relationship between cellular senescence and viral infection by combining the *in silico*, *in vitro*, and *in vivo* experiments, with a focus on the brain cellular senescence in SIV infection. Although it is well known that HIV-1 infection causes immune senescence *in vivo*, its role in cellular senescence remains less studied. Only a handful of studies in recent years directly investigate this question. However, those studies simply use HIV proteins (Tat, Nef, or gp120) or pseudo-virus to induce cellular senescence in a few cell types *in vitro*. Furthermore, only one study was conducted *in vivo* using the HIV-1 rodent model, without including the old animals.
To our knowledge, we have conducted the first study that demonstrated the age difference in the interaction between SIV and cellular senescence using the non-human primate model accompanied by data mining. Our main discoveries were:

1. Age difference in inducing cellular senescence in the brain.
2. Age difference in the antiviral function of cellular senescence in the brain.
3. Microglia could be the cell type that played the most important role in the age difference of senescent cells' antiviral functions.

Moreover, we first demonstrate the interaction between the virome and cellular senescence at the single-cell level by *in vivo* study. Identifying cellular senescence as a general antiviral mechanism improved our understanding of the organismal aging process. This study offered a method for systematically investigating the relationship between cellular senescence and the virome. Furthermore, the unique analytical methods we provided can be adapted to study the interaction between the microbiome and any cell type and forged a path in the study of host-pathogen interactions.

Given that senescent cells in the brain contribute to the cognitive decline and neurodegeneration, our findings suggest that they play an important role in the acceleration of brain aging in young hosts and possibly towards to the development of HIV-associated neurocognitive disorders.
2.5 Discussion and future studies

2.5.1 SIV and cellular senescence studies

Future *in vitro* experiments could use more cell types to test the inhibition of SIV/HIV in senescent cells. In this study, due to the ART, we were unable to detect the cell-associated SIV RNA in the BMC to further support the results from the pathway analysis, which demonstrated a general viral inhibition. Thus, SIV-infected rhesus macaques without ART could be used to facilitate measuring the SIV in the brain cells. Then, the change of SIV abundance and regulation of viral infection-related pathways could be associated directly in the cell. Besides, to *ex vivo* culture sorted BMC, comparing the difference in SIV infection between the senescent and non-senescent BMC is another way to be tested. Then, the antiviral function for SIV, in particular, would be further clarified. Moreover, it is interesting to observe if the viral inhibition effect in old animals due to cellular senescence would be diminished after removing those cells by senolytics.

2.5.2 Relationship between various viral species and cellular senescence

Our results revealed that compared to young, senescent cells in the old animal generally has a more remarkable increase in antiviral infection function. This effect warrants further study for specific viral species, and the results could vary for different kinds of viruses. For example, recent research showed β-coronavirus mouse hepatitis virus (MHV), which in the same family as SARS-CoV-1&2, caused high mortality exclusively in old mice. However, the coronavirus-related mortality in old mice was reduced dramatically by removing the senescent cells, suggesting that senescent cells in aged individuals may mediate the coronavirus pathogenesis [212]. Our results of senescent
BMCs further supported this finding by showing the increased coronavirus pathogenesis pathway in old animals only.

Furthermore, the elderly were reportedly more susceptible to some viral infections such as covid-19 and influenza. In addition to the viral-specific difference in inhibition ability of senescent cells, co-morbidities or the dysfunctional organ such as the aged lung can exacerbate the infection outcome. Overall, it appears to be a result of the reaction between viral inhibition at the cell level and co-morbidities at the global system level in the elderly. The degenerated organ systems in the elderly are more vulnerable to those stimuli even though the senescent cells provide inhibition to some extent.

One study reported that even when SIV pathogenesis was absent in aged Chinese rhesus macaques, they still developed marked viremia [228]. Our findings would partially explain this phenomenon that senescent cells in the elderly are not only reluctant to be affected by SIV, but also dampened their proliferation. However, the immunosenescence may offset this effect in the old animal. The defected virus clearance ability of the adaptive immune system in old may cause no difference in plasma viremia.

Moreover, the unclassified reads may contain unknown viruses, the distribution of which could be a confounding factor in the virome analysis. This influence will decrease gradually with the progress in the identification of new viruses and updates of the virus database in future.

Furthermore, the changes of molecules underlying this age difference in the antiviral function of senescent cells requires further investigation. The intervention on those targets might offer a new avenue for antiviral treatment.
2.5.3 Canonical cell type classification

In the future, other novel methods for canonical cell-type annotation could be further evaluated and included in our single-cell RNA-seq analysis, such as SingeIR (Aran et al., 2019), scCATCH (Shao et al., 2020), and SCSA (Cao, Wang, & Peng, 2020). In addition, the cell type deconvolution step of our BMC's bulk RNA-seq data will be improved when the single-cell RNA-seq data from BMCs of rhesus macaques becomes available.

2.5.4 Lipofuscin would be an ideal marker for cellular senescence measurement in FFPE tissue

Brain tissues from aged animals have a higher level of autofluorescence due to accumulated lipofuscin. It hinders the use of technologies that rely on fluorescent labeling such as immunofluorescence histochemistry and flow cytometry. Taking advantage of autofluorescence as a marker of senescence would be a game changer for detection in vivo since it is stable and can be easily observed. It also requires less equipment and reagents. Besides, it enables the utilization of an abundance of archived tissue samples in FFPE format for cellular senescence detection. This expansion of the utilizable “sample bank” would further promote cellular senescence research.

Moreover, autofluorescence in the tissue may include non-specific signals other than lipofuscin. We excluded those interferences manually for the ROI on each slide based on lipofuscin's appearance pattern. In the future, machine learning of the morphology characteristics of lipofuscin followed by auto-identification could make the quantification of the senescent cell on a large amount of sample size more efficient. Using the massive
archived FFPE tissues followed by an automatic slide scanner to images and analyze them using the auto-detector program, would accelerate our understanding of the association between cellular senescence and organ dysfunction or disease phenotypes.

2.5.5 Association between cellular senescence and immune cells

Non-immune cells develop macrophage-like features during cellular senescence [229] or under stimulus such as infection [230-232]. Similarly, we also found that some non-immune cells have gene expression profiles associated with microglia in the brain (Figure 17). To further explore this connection, it needs to identify potential macrophage-like functions of non-immune cells during senescence development. Thus, in addition to scoring each cell with cell types and correlation analyses in single-cell RNA-Seq data, future studies could do mediation analysis, which could test possible causal relationships. For instance, using linear regression to test if cellular senescence is a mediator between non-immune cells (independent variable) and a macrophage-like cell type (dependent variable). Then, ex-vivo experiments using histology methods could verify those morphological and functional changes.

Our results showed that among immune and non-immune cells in the brain, microglia and endothelial cells were the cell types most likely associated with cellular senescence. Furthermore, during aging, the association increased in microglia but decreased in endothelial cells. These changes in cellular feature associations were also accompanied by cell population changes with the increasing proportion of microglia but decreasing endothelial cells during brain aging. The reasons behind this shift are worth exploring. One of the possibilities is that endothelial cells may acquire some immune cell
characteristics (e.g., microglia) during senescence. However, the causality between cellular senescence, microglia, and endothelial cells remains unclear. It is unknown whether senescence in the endothelial cells leads to its cellular features shifting to microglia-like or if this cell type is just prone to be senescent without phenotypic changes. Hence, if it is solely due to higher susceptibility to senescence, why is the fraction corresponding to this cell type decreased in old brain tissue? The mediator analysis mentioned could help to clarify the first speculation about the cell type transformation. Thus, with the increase of available single-cell RNA-seq raw data, these questions could be better answered in the future by a meta-analysis that can include more samples.

Meanwhile, in addition to decreased viral abundance, we observed an increased viral infection rate in senescent cells (SC: 104/394=0.264 vs. NonSC: 531/2532=0.210 in old; SC: 92/573=0.161 vs. NonSC: 615/3896=0.158 in young), which indicate that the senescent cell is prone to infection, yet inhibits viral proliferation/function inside. Is this a cell-autonomous defense mechanism that engulfs more invaders and inhibits them similarly to innate immune cells? Or is it due to the chronologically accumulated opportunistic infection in the aged cells? It might be both. The changes in pathogen internalization in senescent non-immune cells associated with the acquisition of macrophage-like features can be quantified using the MTD software. Hence, the question can be further answered when other factors can be controlled, such as in vitro experiments or on germ-free animals that exclude the chance of accumulating opportunistic infection in aged cells. Because macrophage/microglia are key players in the innate immune system but also prone to be targeted and even become the reservoir of pathogens such as HIV/SIV, this proposed research would help uncover a new immune defense or infection mechanism.
2.5.6 Improve cellular senescence analysis technique in single-cell RNA-seq data

Our findings also raised the notion that senescent cells from different age groups may not function identically. Even though the cell has manifested in several senescent markers, the aged tissue environments might have additional effects on its senescent degree at the high dimensional omics level. A possible reason is that the methods currently used to define cellular senescence are incomplete to reflect the actual cell aging status.

During our single-cell RNA-seq analysis, a senescent cell was defined by the “senescent core gene set” [198], which combined the use of a group of markers (p16, p21, p53, etc.). In addition, we also collected the differentially expressed genes by using the p16 gene as a single marker to divide the senescent and non-senescent cells from young and old samples, which followed the definition used by the original author of this data [188]. IPA enrichment results showed differences in various biological functions between p16 positive brain cells from old and young mice. These age-associated functional changes of p16-positive cells included inhibiting tumor growth, increasing cellular movement, cell death, inflammatory response, encephalitis, etc. Notably, the viral infection pathway also slightly decreased in p16-positive cells from the aged animal with a z-score value of -0.618. Moreover, the senescence pathway (z-score: 2.1) was enriched in p16-positive cells from old animals but not from young animals. These results supported our findings of age difference in the function of senescent cells and further highlighted a demand to distinguish senescent cells from the high dimensional data.

In the single-cell RNA-Seq analysis, the definition of senescent cells largely relies on the gene set of cellular senescence markers and the scoring algorithm. However, there
is still a lack of either universal markers or the more pragmatical cell type-specific cellular senescence markers. Moreover, the senescent cell scoring method is based on the GSEA algorithm, and it cannot contain the down gene-regulation pattern in the gene set. For example, the proliferation maker MKI67 gene, which is downregulated in the senescent cells, will be omitted in the senescent scoring.

The senescent cell scoring method could be optimized. 1. The senescent gene set used in the scoring could be adjusted for different cell types in future. With the development of cell atlas and more single-cell RNA-Seq data including the healthy young and aged animals are published, the cell type-specific senescent markers can be defined. 2. The scoring algorithm could be improved by learning from other advanced annotation techniques for canonical cell types in the bioinformatics research fields of single-cell sequencing [233-235]; and by applying more sophisticated gene set enrichment analysis in the research fields of functional omics studies [236]. These may include overlap p-value and Activation Z-score methods in the IPA [184], which are able to count pattern matches between both up and down-regulated genes. In addition, topology-based methods could be suitable to score the cellular senescence which has relative distinguishable phenotype and interaction network with other pathways [237, 238].

2.5.7 Assess tissue senescent cell burden by bulk RNA-seq data

Because the RNA-seq data from bulk tissue are more easily accessible than the single-cell, senescent cell burden in tissue can be estimated by apply cell type deconvolution techniques. The method also will benefit from the optimized senescent gene set from advances in single-cell RNA-seq analysis. Non-senescent and senescent cell-
specific gene expression information can be used to illustrate the proportion of those cells in a tissue, which could be an indicator of the 'biological age' of the tissue. By utilizing vast amounts of bulk tissue RNA-seq data, this proposed approach will allow to further elucidate the senescent cell burden in various *in vivo* scenarios with different health conditions. Furthermore, it could synergize with the automatic lipofuscin analysis mentioned before, which could facilitate the retrospective study if the researcher restored the sample backup/archived tissue used in RNA-seq.

2.5.8 Refine the definition of immune cell senescence

Instead of using the umbrella term "senescence" to describe the age-related malfunction of immune cells by the "senescence score," which is like the biological age in the cell level, could more precisely illustrate the cell conditions. Furthermore, the disparity of understanding of cellular senescence between immunology and aging research communities may be unified by the novel definition methods that target the "truly" aging of the cell. Through this continuous variable/indicator, the status of the cell in the body could be monitored more accurately. Aging studies would benefit from a new way to understand aging at the cell level, thus offering opportunities for novel therapeutic strategies to treat age-related disorders.

2.5.9 Define the "true" cellular senescence

Studies have found senescent cells occupied 2% to 14% in different tissues in old mice [239], which are still a small minority and mystified its role in organismal aging. In addition to the explanation by SASP, another potential reason could be that current research
overlooked a proportion of senescent cells. Because of the classical definition of the senescent cells that only consider the late stage of the cellular senescence development, that results in defining the senescent as a binary categorical variable, which may neglect the cell level's aging process that could be a continuous variable.

In essence, rather than dichotomizing a cell to either senescent or non-senescent, cell aging should be considered one continuous process. Measuring the age of cells can hereby go beyond using the definition of "end-point" senescent stage. Training of machine learning could not only use the end-point cellular senescence stage as the standard but also include the mid-points as well as different stages of the cell’s lifespan. Learning from the studies of "biological age" in humans, in which we used the chronological age as a standard to train the model. At the cell level, we could combine the use of proper molecular and morphological changes as indicators such as the population doubling number (PD) and telomere length in the dividing cell types. We could also use the lipofuscin level and the degree of nuclei morphological changes in the postmitotic cells. This combination of multiple testing methods from cellular morphology/histology to the molecular level can help determine the cell age. Meanwhile, the gene set of cell aging could be reciprocally refined with the evolving model.

In addition to the supervised approach just mentioned, using unsupervised approaches could help obtain a more authentic age of cells. The supervised method will fully rely on the current knowledge of markers (e.g., PD, length of culture, SA-β-gal, p16) of cellular senescence by the researchers, which assumes we already know the "truth" of what senescence is. Then we use this "truth" to distinguish which cell is senescent in the sample and to train the machine to help us based on this standard. However, this assumption
may be problematic due to the complex nature of the senescence, which overlooks the latent factors that contribute to the biological age of cells. For example, the senescence trajectory may differ in adaptive immune cells from cells of sold tissue. As a result, aligning them with classic cellular senescence that relies on the currently used markers may be questionable. Thus, using the unsupervised approaches to identify previously unknown patterns from the high-dimensional multi-omics data is a promising approach to solve this problem. For example, people can easily define a fruit - apple correctly then label a large amount of data to train the machine to do so because our knowledge of the apple is sufficient to distinguish the true apple. Hence, people are the "supervisor" and can provide the standard reference for the machine to learn. However, there is still no consensus in the use of senescent markers, even the pattern of combination [41, 240]. The machine could not be well supervised by people to identify the "true" senescent cell. Nevertheless, researchers do know the basic phenotypic and functional characteristics of cellular senescence, which is the root of conceptualizing various types of senescent cells (e.g., RS, OIS, IRS). The gap seems to be just between the phenotypes & functions and those genomic & molecular traits. Therefore, future research to apply the unsupervised approach based on the primary functional pathways during cellular senescence would reveal the genuine aging status of cells. These senescent functional pathways may be relevant to anti-apoptosis, tumorigenesis, anti-proliferation, DNA damage, glycolysis, advanced glycation end products (AGE), and lysosome activity, etc. Noteworthy, a similar underlying concept successfully supported the development of the first generation of senolytic drug [30]. The researchers followed the functional characteristics of the senescent cells (e.g., anti-
apoptotic, similar to cancer but without division ability) to screen and target the senescence-associated genes.

In general, using powerful AI/machine learning methods such as the ensemble approach [241], new senescent patterns/markers can be condensed from high-dimensional data. Henceforth, the research paradigm would be shifted and evolved through the discovery of senescent markers. Nowadays, Cellular Senescence Network (SenNet) Program is a good start to reveal the characteristics of senescent cells at the single-cell level. Overall, defining the age of cells will be the fundamental commitment to solidify the edifice of senescence research.
**List of References:**


222. Hall, B.M., et al., *p16(Ink4a) and senescence-associated beta-galactosidase can be induced in macrophages as part of a reversible response to physiological stimuli*. Aging (Albany NY), 2017.


CHAPTER 3 – DEVELOPMENT OF AN DUAL RNA-SEQ SOFTWARE FOR METATRANSCRIPTOME ANALYSIS

3.1 Introduction

A variety of microorganism species have been recognized to contribute to the development of human diseases, including cancer, autoimmune diseases, psychological disorders. For example, Helicobacter pylori can cause stomach cancer [242], human papillomavirus (HPV) can lead to uterine cervix cancer [243], and HIV-1 can lead to AIDS [244] as well as HANDs [140]. In addition, the Epstein–Barr virus (EBV) may contribute to 1.5% of total human cancer of various types worldwide [245]. However, whether microorganisms are present in tissues and whether their state of activation contributes to disease development is yet to be determined. The pathogenic mechanism of a large proportion of opportunistic infections remains unexplored. Furthermore, different cell populations may contribute to the heterogeneous tropism in infections. Therefore, microbiome diversity and its abundance in the host cells are important to identify infected cells and their status.

RNA transcripts from tissue may contain RNAs from those microorganisms that have not been actively investigated. Several tools have been developed to detect the microbiomes in the RNA-seq data, such as Kraken2 [246], VIRTUS [247], and IDseq [248]. However, it is still challenging for a researcher without bioinformatics expertise to examine the state of the microbiome in host tissues and its relation to the endogenous expression of host genes, especially at the single-cell level. To address these issues, we developed a user-friendly pipeline for comprehensive microbiome analysis from bulk and single-cell RNA-seq data and named it MTD (Meta-Transcriptome Detector).
3.2 Methods

3.2.1 Description of MTD development

The MTD has two pipelines to detect and quantify microbiomes by analyzing bulk RNA-seq data and single-cell RNA-seq data, respectively (Figure 18). MTD is written in R (version 4.0.3) and Bash (version 4.2) languages and executed in GNU/Linux system. Users can easily install and run MTD using only one command and without requiring root privileges. The outputs (graphs, tables, count matrixes, etc.) are automatically generated and stored in the designated directory/folder defined by the user. The user manual and source code used in this chapter were disclosed on the webpage https://github.com/FEI38750/MTD. Here we describe the two sub-pipelines separately.

3.2.1.1 Dual analysis of bulk RNA-seq data

First, RNA-seq reads in the FASTQ file are trimmed and filtered by fastp (version 0.20.1) [249] with polyA/T trimming, and reads shorter than 40 bp (with the option --trim_poly_x --length_required 40) are discarded. Then, processed reads are classified based on the host genome by Kraken2 (version 2.1.1) [246] with default parameters. Finally, the host and non-host reads are collected separately in FASTQ format.

Host transcriptome analysis: The host species supported by MTD initially are Homo sapiens (Genome assembly: GRCh38), Mus musculus (Genome assembly: GRCm39), and Macaca mulatta (Genome assembly: Mmul_10). Additionally, users can add other host species by one command line. The host reads are aligned on the corresponding host genome by Hisat2 (version 2.2.1) [250] with default parameters and alignments are written as a SAM file. Quantification of reads for each host gene expression is done by featureCounts (version 2.0.1) [251]. Next, the count data is analyzed by DESeq2.
R package (version 1.32.0) [224] in Bioconductor to get the differentially expressed genes (DEGs). The gene annotation is done through the biomaRt R package (version 2.46.3) [252, 253] in Bioconductor. The data visualization and a comprehensive count matrix are automatically generated through R-written programs. The data visualization includes the heatmap, Venn Diagram, PCA, barplot, and volcano plot. The count matrix contains the Ensembl gene ID, gene symbol, chromosome name, gene position, functional descriptions, DEG results for each group comparison, raw read counts, normalized reads count, normalized and transformed reads counts. This comprehensive count matrix is saved in CSV format, and it facilitates the user to perform downstream analyses such as pathway enrichment and customized data visualization.

Microbiome transcriptome analysis: MTD supports a broad spectrum of microbiome species and vectors including viruses, bacteria, protozoa, fungi, plasmids, and vectors. At the time of writing, the viruses contain 16,275 species from Virus-Host DB [254] and SIV accession M33262. The rest of the microbiome are from the NCBI RefSeq database [255], which includes 63,237 species of bacteria, 13,970 fungi, 1,337 archaea, 573 protozoa, and 5,855 plasmids. In addition, vector contamination can be screened using the NCBI UniVec Database. Users can update the microbiome databases in MTD by one command line. The non-host reads are further classified by Kraken2 based on microbiome references with the default parameters, followed by a decontamination step that removes the microorganism under the genera reported as reagent and laboratory contaminants [256]. Users can easily customize the blacklist of contaminants based on their experimental environments. Then, the abundance of the microbiome in the species level is calculated by Bracken (version 2.6.0) [257] with the settings -r 75 -l S. Next, the count data is imported
into DESeq2 for analysis of the differentially expressed species. In DESeq2, we adjusted the abundance of microbiome species based on the transcriptome size of the sample. The rationale is that count from a microbial species should take into account the overall representation of the host transcriptome. This normalization step is conducted through the formula, design ~ group + transcriptome_size, where transcriptome_size is defined by the formula: log2 (of a transcriptome size) - mean (of all log2-transformed transcriptome sizes in a sample), and group is the group code (e.g., A or B) of a sample. As a result, the importance of a microbial species in a sample with a higher transcriptome representation is underscored. Meanwhile, the non-host reads (including both unclassified and classified reads by Kraken2 using the reference databases) are imported into the Humann3 (version 3.0) [258] for profiling microbial metabolic pathways and molecular functions. The ChocoPhlAn [259] and full UniRef90 [260] are used as reference databases for nucleotide and protein, respectively. Then the profiling results are translated to human-readable names to facilitate the downstream analyses. Next, the heatmaps of DEG, Venn Diagrams, PCA, bar plots, and volcano plots are generated for the microbial species, molecules, and metabolic pathways based on the Deseq2 results. Additionally, kraken-biom (version 1.0.1) [261] is used to format the data for diversity analyses and phylogenetic tree plotting. The phyloseq R package (version 1.34.0) [262] in Bioconductor and vegan R package (version 2.5-7) [226] are used to analyze the diversities, including alpha diversity (Shannon, Simpson) and beta diversity (Bray-Curtis). Then, the t-test comparison results of alpha diversities and abundance of unclassified reads between groups are shown as box plots. PCoA graphing and analysis of similarities (ANOSIM) are based on Bray-Curtis distance. The relative abundance of microbial in the total microbiome is shown as bar plots in the
Phylum level. The heatmap of total microbiome abundance is plotted using data normalized by Deseq2. Next, the phylogenetic trees are plotted through modified Graphlan [263], which is a tool for generating informative and integrative circular graph representing phylogenetic and taxonomic trees. However, the original program requires a specific data format as input and is not compatible with the output from Kraken2 and Bracken software despite conversion to BIOM-format tables using kraken-biom [261] and export2graphlan [264] software tools. Therefore, we wrote a converter program and integrated it into the pipeline to bridge the taxonomic classification software-Kraken2/Bracken and graph-making software GraPhlAn. It allows us to transform the output of Kraken2 and Bracken to match the data structure requirement by GraPhlAn. Furthermore, the default settings of colors have been optimized by modifying the source code of GraPhlAn. In addition to .biom format, data is also saved in .mpa and .krona formats to facilitate further downstream visualizations.

Finally, we examine the association between the microbiome and the host's characteristics, such as gene expression and pathways. Pathway enrichment for each sample is performed by a modified ssGSEA program mentioned in the previous chapter, with a MSigDB C2 database that contains 6,290 curated gene sets. Next, the covariance effects among groups are adjusted through the removeBatchEffect function in the limma R package (version 3.48.1) [265] in Bioconductor. Then, the association analyses are conducted through Halla (version 0.8.18) [266], which is set to compute hierarchical clustering of Spearman pairwise correlation. Figure 19 illustrates the MTD automatic pipeline for dual-analyzing the bulk RNA-seq raw data.
A

RNA-seq raw data

fastp

trimming & filtering

Customized DB (host):
Homo sapiens
Mus musculus
Rhesus macaques

Kraken2

taxonomic classification
separate host and Non-host reads

Host reads

Hisat2

Host's reads mapping

Humann3

ID translation

Kraeken2

unclassified reads collection

Microbiome reads taxonomic classification and counting

Decontamination

Bracken

Species level abundance estimation

FeatureCount

Host's transcript counting

Deseq2

Normalization
Differential expression analysis

Pathway analysis

ssGSEA

Annotation & Visualization & Converging

kraken-biom

Convert bracken report to biom format

HALLA

Association analysis

phyloseq

Diversity analysis

Convertor

Graphlan
Figure 18. Overview of MTD.

(A) Schematic views of MTD bulk mRNA-seq workflow. (B) Schematic views of MTD single-cell mRNA-seq workflow. In the workflow diagrams, the white boxes represent the reads in FASTQ format and the count matrix, the blue boxes show the bioinformatics
software used, and the green boxes are the additional tools for data processing. The white boxes with curved edges show the reference genome and databases. In the single-cell mRNA-seq workflow (B), the left side exemplifies the host reads process protocols, and the right side in yellow shadow shows the MTD automatic pipeline to calculate the count matrix for the microbiome reads and the correlation between microbiome and host genes.
Figure 19 The automatic pipeline of MTD for dual-analyzing the bulk RNA-seq raw data.
Analysis results are automatically saved into the folder assigned by the user. Examples of analysis outputs for the transcriptome of host and microbiome are demonstrated on the left side by the color red and on the right side by the color in blue. The shared procedures are in the purple boxes and include the input files and the association analyses. Besides the graphs, all the detailed information was included in the data sheets and stored in the corresponding output folder. To use the MTD, the user must simply put the FASTQ files in a folder with a sample sheet in CSV format that describes the sample names, groups, and comparisons, and then perform the analysis with one command line. For example, with the command line "bash MTD.sh -i ~/inputpath/samplesheet.csv -o ~/outputpath -h 9544 -t 20", the user enters the place of the sample sheet with the raw data after the flag "-i", and where he wants to place the results (after flag "-o"), the host taxonomic ID is after "-h", and the threads of CPU after "-t".
3.2.1.2 Dual analysis of single-cell RNA-seq data

The MTD supports automatic generation of the count matrix of the microbiome by using raw data in the FASTQ format and count matrix of host genes from two commonly used single-cell RNA-seq platforms, 10x Genomics and Drop-seq. First, the user can download the available count matrix of host genes (H5, Matrix, or .dge.txt format) or follow the corresponding workflow to process the raw data, such as through the 10x Cell Ranger software. The cell barcodes are identified from the count matrix of host genes, then the UMI and cell barcodes are extracted and add to read names by using the UMI-tools (version 1.1.1) [267]. Second, the reads are trimmed and filtered by fastp, then filtered out the host reads by Kraken2. The non-host reads are further classified by Kraken2 with the comprehensive microbiome databases and followed by a decontamination step. The steps from trimming to decontamination use the same settings as the description in the last section. Next, the taxonomic labels of the reads are extracted and aligned with corresponding cell barcodes through a step written by the AWK program language. Finally, UMI-tools is used to generate a count matrix through an R-written converter. Figure 20 exemplifies a count matrix automatically generated by MTD for the microbiome in each cell. Subsequently, MTD combines the count matrices of the host genes and microbiome to perform the correlation analysis automatically. An example of the correlation analysis result is showed in Figure 31 (A-B).

At the single-cell level, the Spearman correlations between microbial organisms and host genes are tested by using the top 3,000 most highly variable features, including the normalized data of both host and microbiome. Because the step of correlation analysis step is highly time-consuming for a large data matrix, parallelizing computing was applied
to speed up the computation by using the doParallel R package (Version 1.0.16). The other analysis methods through Seurat and homemade programs are described in Chapter 2. The diagram of the pipeline is demonstrated in Figure 18B.

Figure 20. An example of the count matrix automatically generated by MTD for the single-cell microbiome analysis.

For illustration, the figure shows part of the large count matrix. The name and taxonomy ID of the microbiome is in the first column and highlighted in the green box. The read counts are highlighted in the blue box, and the first row shows the cell barcodes.
3.2.2 Animal information and tissue collection

Here we demonstrated the application of MTD for dual analyzing the bulk RNA-seq data by using samples from the descending colon and BMC from rhesus monkeys. For bulk RNA-seq of descending colons, a total of 9 Chinese rhesus macaques were used, including 6 SIV-infected and ART-treated and 3 SIV-uninfected controls (Table 2). The information about BMC was described in Chapter 2.

Animal euthanasia was performed in line with the recommendations from the Panel on Euthanasia of the American Veterinary Medical Association. Following Tulane IACUC standards of operation (SOP). SIV-infected and/or drug-treated macaques were euthanized firstly with telazol and buprenorphine, followed by sodium pentobarbital intravenous injection. Fresh tissues from descending colon were collected and soaked in RNAlater™ Stabilization Solution during necropsies. The brain tissues were collected first in order to minimize the chances of contamination from the gut microbiome.

The RNA-seq protocol is the same as in the description in Chapter 2 Methods. Significant DEG were defined by log2FoldChange > 0.5 and p-value < 0.05.
Table 2. Animal information of descending colon samples

<table>
<thead>
<tr>
<th>Animal #</th>
<th>Sex</th>
<th>Age (Yrs)</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>DJ01</td>
<td>FEMALE</td>
<td>18.72</td>
<td>Old SIV+ART</td>
</tr>
<tr>
<td>EM77</td>
<td>FEMALE</td>
<td>16.87</td>
<td>Old SIV+ART</td>
</tr>
<tr>
<td>EP15</td>
<td>FEMALE</td>
<td>16.75</td>
<td>Old SIV+ART</td>
</tr>
<tr>
<td>KN41</td>
<td>FEMALE</td>
<td>6.92</td>
<td>Young SIV+ART</td>
</tr>
<tr>
<td>KT12</td>
<td>FEMALE</td>
<td>6.79</td>
<td>Young SIV+ART</td>
</tr>
<tr>
<td>LC31</td>
<td>FEMALE</td>
<td>6.01</td>
<td>Young SIV+ART</td>
</tr>
<tr>
<td>JJ57</td>
<td>FEMALE</td>
<td>8.15</td>
<td>Young NonSIV</td>
</tr>
<tr>
<td>JD56</td>
<td>FEMALE</td>
<td>9.07</td>
<td>Young NonSIV</td>
</tr>
<tr>
<td>JG77</td>
<td>FEMALE</td>
<td>8.95</td>
<td>Young NonSIV</td>
</tr>
</tbody>
</table>

- Old SIV+ART (Mean 17.45 ± SD 1.10 years)
- Young SIV+ART (Mean 6.57 ± SD 0.49 years)
- Young NonSIV (Mean 8.72 ± SD 0.50 years)
3.3 Results

3.3.1 Application to bulk RNA-seq analysis: descending colon of rhesus macaques

An example of application on BMC samples has been shown in Chapter 2. Here, we show the result of our analysis using the MTD on descending colon samples from rhesus macaques, which were assumed to have higher richness in microbiome species than BMCs. Then, the intersection of brain and gut microbiome were studied (section 3.3.2).

Through MTD, the transcripts of both the microbiome and the host were analyzed simultaneously using the same bulk RNA-seq raw data. Figure 21 presents the taxonomic and phylogenetic trees of microorganisms detected in the descending colon samples from rhesus macaques. Figure 22 is a heatmap showing the abundance of all the microbiome species in the samples from the descending colon. The results of the microbiome and host gene analyses of the descending colon are described in sections 3.3.1.1 and 3.3.1.2, respectively. The association analysis results of the microbiome to the host genes or pathways are in section 3.3.1.3.
Figure 21. The taxonomy graph of the microbiome in the descending colon.

The outer rings on the graph showcase each sample group. There are three groups: ART_Old (Old SIV+ART), ART_Young (Young SIV+ART), and Ctl_Young (Young NonSIV). The colored regions on rings represent the species of the corresponding samples, and the darker-colored parts outline the differentially expressed species of each group, which has been identified by DESeq2 R packages [224]. The blank regions across all the rings indicate that the p-values of the species were set to N/A when performing the
differential expression analysis. In our case, the possible reasons are that either most of the samples have zero counts for that species or that it is an extreme count outlier [225]. The graph was generated using a modified GraPhlAn program that was integrated into the MTD. Annotation levels were set to 2 (phylum), 3 (class), 4 (order), 5 (family), 6 (genus). Level 7 (species) was used as the external legend for the annotation. The 50 most abundant clades were highlighted on the graph.
Figure 22. Heatmap of the microbiome species in descending colon samples.

The higher abundance of microbiome reads is shown in a light blue color, and the lower abundance can be seen in a darker color. Data was normalized using the Deseq2 and plotted by the phyloseq R packages, which were wrapped in MTD.
3.3.1.1 Microbiome analysis of descending colon

An overview of differentially expressed microbiome species are shown in Figure 23A, and the relative abundances of microbiomes at the phylum level were plotted in Figure 23B. The heatmap with the detailed name of each differentially expressed species is stored in the same output folder defined by the user. The clusters groups were plotted on PCoA based on Bray-Curtis distance (Figure 23C). Three groups were not well separated in the PCoA, and the analysis of diversity did not find a significant difference among the groups (Figure 23D). A hundred and seven species were shared by these groups, and ART_Old animals had more unique species than other groups (Figure 23E). The reads that did not belong to the host were also examined, including all the microbiome information, classified and unclassified (Figure 23F). ART young animals showed a relatively lower non-host reads ratio than other groups, especially compared to the nonSIV control young animals, which implicated the effect of ART in reducing microbial abundance. However, the alpha-diversity was not significantly different among groups (Figure 23G). In general, the results indicated a lack of disparities in the microbiome diversity in the descending colons between the groups of ART_Old, ART_Young, and Ctl_Young. More detailed information about the significantly differentially expressed microbiome species between groups can be found in Figure 24.

The results of metabolic molecules analyses are shown in Figures 25-26. In addition, just a few microbial metabolic pathways were profiled in our samples. Among them, aerobic respiration I (cytochrome c) is the only significantly differently expressed pathway, which was higher in the Ctl_Young group than the other two groups. The table of pathway analysis results was stored in the corresponding output folder.
Figure 23. Microbiome species analysis of descending colon

(A) The significantly differentially expressed microbiome species. Data was normalized and transformed using the vst function in the Deseq2 R package. (B) Relatively abundance of the microbiome in the phylum level. (C) PCoA plot by using the Bray-Curtis distances
(beta-diversity). (D) Analysis of diversity in groups using the anosim function in vegan R package (Oksanen et al., 2020). (E) Venn diagram illustrated the number of shared microorganisms among groups. (F) Difference of non-host reads ratio between groups examined by t-test. (G) Shannon alpha-diversity between groups examined by t-test.
Figure 24. Differentially expressed microbiome species between groups

Bar plots (A, C) and volcano plots (B, D) represented the significantly differentially expressed microbial species between groups. Besides, the full lists of species, count matrixes, and DEG table of each comparison were stored in the corresponding output folder defined by the user.
Figure 25 Microbial metabolic molecules in GO terms

Only significantly differentially expressed microbial metabolic molecules were generated in the heatmap (A), bar (D, F), and volcano (E, G) plots. (B) PCA results. (C) Shared
number of molecules among groups. The heatmap and bar plot only show the thumbnail when there are a large number of results. The volcano plot labeled the top 20 biggest changes in the significant results. In addition, the full lists of molecules, count matrixes, DEG table of each comparison, and the heatmap or bar plot with species names were stored in the corresponding output folder defined by the user.
Figure 26. Microbial metabolic molecules in KEGG terms

Only significantly differentially expressed microbial metabolic molecules were generated in the heatmap (A), bar (D, F), and volcano (E, G) plots. (B) PCA results. (C) Shared
number of molecules among groups. The volcano plot labeled the top 20 biggest changes in the significant results. In addition, the full lists of molecules, count matrixes, DEG table of each comparison, and the heatmap with names were stored in the corresponding output folder defined by the user.
3.3.1.2 Host gene expression analysis of descending colon

ART_Young and Ctl_Young showed more dissimilarities in host gene expression patterns, whereas ART_Old cannot be separated from the other two groups (Figure 27 A-B). Most of the genes were shared among the three groups (Figure 27 C). The overview of significant DEG between groups are shown in the thumbnail bar plots (Figure 27 D, F) and the volcano plots (Figure 27 E, G). The top 20 genes with the most significant absolute fold changes are labeled in the volcano plots.

In addition, the comprehensive count matrix, the DEG table of each comparison, and the heatmap with gene names were stored in the corresponding output folder defined by the user. They can easily be used for downstream analyses, such as pathway enrichment, functional annotation, and molecular interaction network.
Figure 27. Host gene expression analysis

Only significantly differentially expressed host genes were generated in the heatmap (A), bar (D, F), and volcano (E, G) plots. (B) PCA results. (C) Shared number of genes among
groups. The heatmap and bar plot only shows the thumbnail when there are a large number of results. The volcano plot labeled the top 20 biggest changes in the significant results.
3.3.1.3 Association analyses of the microbiome to the host genes or pathways in descending colon

The correlations between the microbiome and host gene were illustrated in Figure 28A. For example, the *Debaryomyces hansenii* shows a significant positive correlation with host gene *UBE2I* and *IL27RA*. The expression of the host gene *C1QTNF8* positively correlated with a group of microbes, such as *Yarrowia lipolytica*, *Aspergillus chevalieri*, *Roseburia hominis*, and *Anaerostipes hadrus*.

The correlations between the microbiome and host pathways are shown in Figure 28B. For example, the pathway that controls the amplification of the 8q24 chromosome region (HEIDENBLAD_AMPLICON_8Q24_UP) was upregulated with the mRNA expression of *Saccharomyces eubayanus*. The complex I biogenesis signaling pathway (REACTOME_COMPLEX_I_BIOGENESIS) was negatively correlated with the expression of *Helicobacter cinaedi*. 
Figure 28. Association analyses of the microbiome and host genes or pathways in descending colon
The figure shows the correlation between the RNA expression level of microbiome species and host gene (A), or pathways (B). The x-axis is labeled with the names of the host genes or pathways, and the y-axis lists the names of microbiome species. Positive correlation coefficients are shown in red, and negative correlation coefficients are shown in blue. The significant results are marked with white dots and ranked by numbers. The results in the same cluster can be found in a box with the same number. The association was examined by pairwise Spearman correlation test based on hierarchical clustering. The tabulates of all comparison results and dot plots were saved in the corresponding output folder.
3.3.2 Intersection of brain and gut microbiome

All animals evolved in intimate interaction with microbiome - microbial communities comprising fungi, bacteria, archaea, protozoa, and viruses. These collections of microorganisms inhabiting inner organs and surface skin compose the microbiota in the body. Over the past decade, a flood of research in microbiology and neuroscience revealed the dynamic interactions between the microbiome and central nervous system (CNS) of their animal host. Nowadays, researchers are starting to investigate the impacts of the microbiome on the brain directly through infection, metabolic products, or the ability to induce inflammation response.

The 'gut-brain axis' refers to the pathways connecting the multiple biological systems that allow communication between the gut and brain. The gut microbiota representing the highest density and most abundance of microorganisms in the human body, plays a crucial role in maintaining the homeostasis of those functional pathways linked to the gut-brain axis such as the immunological, microbial, gastrointestinal, and central nervous systems of animals. Dysbiosis of gut microbiota has been studied in relation to neuropsychiatric disorders such as Alzheimer's disease (AD), Parkinson's disease, amyotrophic lateral sclerosis, and multiple sclerosis, etc. [268, 269].

In HIV-1 patients with antiretroviral therapy (ART), the role of the microbiome in the gut-brain axis is less understood. It is partially due to the practical challenges of acquiring tissue samples from both the CNS and the gut for human studies. Moreover, various lifestyle differences among individuals such as smoking, diet, and sleep patterns can influence research outcomes.
The nonhuman primate (NHP) such as rhesus macaques (*Macaca mulatta*) has been used as a common model in the research of SIV infection and ART. In addition, the NHP model has indispensable advantages in studying the microbiome in the animal:

1. Compared to rodents, which are the most widely used experiment models, NHP has a smaller disparity in the microbiota to humans [270].
2. In NHP studies, influence from the environment and host factors can be better controlled than in human studies. For example, medical histories such as pathogen infection and treatment can be fully described. As a result, confounding factors can be reduced or eliminated in NHP.
3. Unlike rodents, rhesus macaques have a higher similarity to humans in the pattern of behaviors (e.g., eating and sleeping), developing, and aging [271].
4. Rhesus macaques are more similar to humans than rodents in numerous biological aspects such as physiology, anatomy, endocrinology, immunology, and neurology [272].

Therefore, NHP can provide important information regarding characteristics of the microbiome during HIV/SIV infection and ART. In the last decade, the dynamics of the gut microbiome in healthy vs. SIV infection with or without ART have been demonstrated by cross-sectional or longitudinal studies [273-278]. However, the intersections of the microbiome in the gut and brain under ART, to our knowledge, has been rarely investigated. Moreover, the age and cell type differences in the microbiome shared by the gut and the brain remain unclear. Here, we investigated the intersected microbiome by the gut and brain under ART and the effect of age of the animals and brain cells on changes of the microbiome. Furthermore, we expand the breadth and depth of microbiome detection by
measuring a whole spectrum of the known microbiome, including the virus/virome, bacteria, protozoa, and fungi, to the species level.
Figure 29. The intersections of brain and gut microbiome.

The intersections of brain and gut microbiome. A: The samples taken from the brain (BMC) are shown in the grey-colored background, and from the gut (descending colon) in the pink
background. The dots linked by a line represent the intersection of microbiome species between samples. The intersections highlighted in orange indicate an overlapping of species between the tissue of the descending colon and the brain (either BMC_H or BMC_L or both) from the same animal. The vertical bar plots above each intersection show the number of shared species, the single dot without the lines means the absence of intersection, and the horizontal bar plots on the left side show the total microbiome species within each sample. B: The heatmap shows the shared species in the brain (BMC) and gut (descending colon) further illustrating the information of intersections highlighted in orange in figure A. The names of the shared species can be found on the right side of each row. The columns under the same animal ID indicate the intersection of species in the descending colon and BMC in each particular animal. The red sections represent the presence, while the white/blank represents the absence of overlap between the species in the sample. The groups of old and young animals are divided into the left and right sides, respectively.
Brain cells (BMCs) have an overall lower abundance of species than the gut (descending colon) tissues (Figure 29A). The species were not equally distributed in BMC_L and BMC_H cells. Although the species mainly did not overlap, some of them were present in both brain and gut. Among them, *Aspergillus oryzae* (*A. oryzae*), *Debaryomyces hansenii* (*D. hansenii*), and *Streptomyces fulvissimus* (*S. fulvissimus*) exist prevalently in both positions (Figure 29B). *A. oryzae* and *D. hansenii* showed a similar pattern of distribution between young and old rhesus macaques in the intersection of the gut-brain microbiome. *A. oryzae* was present in all the samples (12/12 intersections), and *D. hansenii* was identified in 10/12 intersections. *S. fulvissimus* displayed significant age-related differences in our animals. *S. fulvissimus* was seen exist in all the brain and gut in old animals in a total number of 9/12 intersections.

3.3.2.1 *A. oryzae* in the brain and gut

*A. oryzae* is a species under the genus of *Aspergillus* that includes hundreds of accepted fungal species including a few thousand strains and varieties [279, 280]. *Aspergillus* is a fungal genus which is widely distributed across the earth and has high economic importance in the fermentation and biotechnology industry. On the other hand, the members of *Aspergillus* can cause mycoses (e.g., pulmonary aspergillosis, keratitis, otomycosis), invasive mold infection, and foodborne contamination. Invasive aspergillosis is the most common invasive fungal infection in humans, particularly among immunocompromised individuals such as AIDS or chemotherapy patients. In addition to *Aspergillus fumigatus* which has been considered the primary source that accounts for
those infections, *A. oryzae* and other non-*fumigatus Aspergillus* species are emerging as cause of infection [281].

Central nervous system (CNS) infections due to *Aspergillus*, such as Aspergillosis, are relatively rare conditions but have more adverse effects on health and present additional treatment challenges. It usually occurs secondarily to hematogenous dissemination (fungemia) and is a consequence of local infection, such as the invasion from the lung due to inhalation of spores. Cases of CNS aspergillosis, including cerebral infarction, hemorrhage, and meningitis, have occasionally been reported in both the immunocompetent and immunocompromised patients [281-288]. Among them, there were a few cases involving *A. oryzae*.

Infection of *A. oryzae* in the human brain was first reported by M.D. Joseph Ziskind et al. in the Veterans Administration Hospital at New Orleans, Louisiana [286]. The specimen was taken from a cheek lesion of a 35-year-old male patient who was a farmer, and was used for culture. The cultures were sent to the United States Department of Agriculture in Illinois and the Department of Health, Communicable Disease Center in Georgia for species identification, which confirmed that the implicated organism was *A. oryzae*. A subsequent autopsy showed a significant change in the morphology of the head, including swelling of the face, gliosis, and irregular areas of softening region in the brain. In addition, brain inflammation also increased as seen by microscopical examination. However, no pathologic lesions were observed in the animal experiments, which consisted of inoculating the isolated organism into rats and guinea pigs. After this, a case of meningitis caused by *A. oryzae* was reported by Morris A. Gordon et al. in the Laboratories for Mycology, New York University School of Medicine [288]. CSF and serum specimens
were taken from a 34-year-old female patient who had a history of drug abuse. Then, *A. oryzae* species was identified through the combined use of culture and serologic test methods.

The pathogenic mechanisms of *Aspergillus* species invading the brain and then damaging the CNS are still elusive. *Aspergillus* can produce mycotoxins, including aflatoxins and gliotoxins, that inhibit phagocytosis of macrophages and loosen the blood-brain barrier, subsequently causing damage to neurons, microglia, and astrocytes [281]. For the *A. oryzae* species specifically, its toxins include maltoryzine, cyclopiazonic acid, and β-Nitropropionic acid, which can cause a variety of symptoms, including neurological damage.

Aspergillosis also occurs in a variety of animals, including reptiles, birds, dogs, cats, ruminants, horses, cetaceans, and non-human primates, especially under immune-compromising conditions [289]. In nonhuman primates, a case of invasive aspergillosis was reported in an 18-year-old female putty-nosed monkey [290]. The necropsy and histological examination found that the lung and brain tissue had extensive necrosis admixed with a severe infection of Aspergillus, and abundant fungal hyphae were surrounded by numerous inflammatory cells. The lungs had diffused pleuropneumonia, congestion, and multifocal abscesses, and the brain had diffused meningeal hyperemia, multiple necro-hemorrhagic foci, and multifocal suppurative meningitis. Further, a histopathological exam of a cynomolgus macaque showed infection at the heart and lung [291].

In addition, SIV could be a risk factor for the *Aspergillus* infection in rhesus macaques. A 6-year-old, indoor cage-housed, male rhesus macaque experimentally
infected with SIV developed dermatitis, and the skin lesion spread to the face and legs with pruritis in just two weeks [292]. The fungal culture identified Aspergillus. Followed by anti-fungal medication, the degree of dermatitis was significantly improved, and finally the results of follow-up fungal cultures came out negative for Aspergillus. This animal was infected with the SIVmac239 delta nef strain. It was an attenuated vaccine strain and did not typically lead to the typical disease without mutations. Interestingly, pathologic mutations had been reported in this animal's SIV strain, and a higher than typical viral load was observed.

Compared to the respiratory system and skin, which are the main entrances of the intruding path, the cases of invasive aspergillosis originating from the gut are extremely rare in the literature. Catia Dias et al. reported a case of primary intestinal aspergillosis in a 71-year-old HIV patient. [293]. She had experienced segmental enterectomy surgery but evolved with progressive clinical deterioration in the postoperative period. Postmortem histological examination found fungal hyphae of the Aspergillus accompanied by extensive lesions in the intestine. In addition, some other case reports also found aspergillosis in the gut of patients without pulmonary disease [294-296]. Hence, intestinal aspergillosis may be more common than estimated. The researchers suspected the infection might manifest as other abdominal diseases, such as enterocolitis and colonic ulcers [293].

Although patients with HIV infection are susceptible to Aspergillus infection, which increases morbidity and mortality, there are still not many reported cases of gut-originated Aspergillus invading the brain. Furthermore, its incidence and pathogenic route in animals such as rhesus macaques with SIV infection are largely unknown. Additionally, considering that the initial incidence of A. oryzae infection was reported in New Orleans,
which is the same geographic location as the Tulane National Primate Research Center (TNPRC) where our study animals were housed, more studies are needed to determine the source of the infection, and whether it was regional or more broadly-spread.

Moreover, as *A. oryzae* has been widely used in traditional fermentation industries, and its commercial strains have been considered non-toxic [297], it is easy to underestimate its pathogenic ability, especially in nature strains. More awareness of this species is required for future studies.

A prompt diagnosis is needed regarding the detrimental consequence and concern of the prevalence of infection by *Aspergillus* species. Applying the Point-of-Care Testing (POCT) with decentralized molecular diagnostic technologies would facilitate the screening work within the population, followed by the validation of traditional culture and histological examination. Thus, clinical deterioration by these infections could be prevented with the early detection and treatment.

3.3.2.2 *S. fulvissimus* in the brain and gut

*S. fulvissimus* is a spore-forming, mesophilic bacterium that builds an aerial mycelium. It is involved in the biosynthesis of an ionphore antibiotic known as valinomycin [298]. *S. fulvissimus* belongs to the *Streptomyces* genus, which contains over 800 validly named species of bacteria, and it is considered the largest genus under the Actinobacteria phylum [299].

*Streptomyces* are saprophytic organisms found ubiquitously in soil and decaying vegetation. Members of the genus *Streptomyces* have served as a source of antibiotics (antibacterial, antifungal, and antiparasitic drugs) for a long time. It is rarely known to
cause invasive infections by *Streptomyces* species. Other than mycetoma, the diseases caused by them are largely under-studied. However, it is noteworthy that two species of *Streptomyces*, *Streptomyces somaliensis* and *Streptomyces sudanensis*, have been reported to infect humans and cause actinomycetoma [300]. They result in a severe disease that could cause devastating outcomes including limb deformities, and could lead to poor social and economic prospects [301]. Actinomycetoma often occurs in the tropics due to the infection caused by thorns on plants carrying the *Streptomyces* species. The filamentous bacteria can grow into the skin, then the subcutaneous tissue, and in some cases invading the bone. In addition, the damage to the soft tissue can lead to secondary infections usually caused by staphylococci and streptococci. The health outcome is even worse in immunocompromised patients. For example, *Streptomyces* could spread into the bloodstream and lead to bacteremia [302, 303]. Although mycetoma is the most common symptom of Streptomyces infection, other kinds of invasive infections have also sporadically appeared. In 2003, pulmonary *Streptomyces* infection was identified in a 57-year-old man who had multiorgan sarcoidosis [304]. In 2014, *Streptomyces* was found in the culture of the pleural effusion taken from a 24-year-old man with Cushing syndrome [305]. Patients with immunocompromised conditions such as AIDS and cancer account for most of the cases in the past [305, 306].

The lung is the primary source of visceral infections with *Streptomyces*. However, recently, a few research demonstrated the presence of *Streptomyces* in the gut microbiome in both human and nonhuman primates [307, 308]. Interestingly, it showed that *Streptomyces* were more prevalent in nonhuman primates than humans. Some researchers
hypothesized that it is a protective mechanism for allergy, autoimmunity, and cancer, which can be justified by its antiproliferative and immunosuppressant functions [309].

Brain-associated infection was more rarely reported. A 41-year-old man with Still's disease was diagnosed with *Streptomyces* infection in the brain. A total of 92 species of *Streptomyces* were isolated from the brain abscess of the patient after the neurosurgery [310]. In addition, a previously healthy, 19-year-old man whose brain accidentally penetrated by a soil-contaminated object presented with a brain abscess caused by *Streptomyces* infection [311]. Nevertheless, literature that showing gut to brain invasion due to either *S. fulvissimus* particularly or *Streptomyces* is still limited.

The health implications of *Streptomyces* are still elusive. It can cause diseases in the skin or lung, but it is also hypothesized to play a role in anti-inflammation and anti-cancer functions in humans. We observed that *S. fulvissimus* was present in all the aged but not in the young animals. It is the only microorganism that showed significant age-related differences in the intersection of the gut-brain microbiome in our animals. Taking into account the many cases involving middle to older-aged patients, we propose that aging is another factor that determines the infection by this species. This could be due to the decline of immunity during aging in addition to the SIV infection. Furthermore, identifying specific species (e.g., *S. fulvissimus*) for those scenarios is required. Especially, the invasion into the brain from an organ like the gastrointestinal tract has substantial clinical value for investigation.
3.3.2.3 D. hansenii in the brain and gut

*D. hansenii*, also known as *Candida famata*, is a species of yeast that belongs to the family *Saccharomycetaceae* and genus *Debaryomyces*. It is a common species found in various types of dairy products and in the environment. Compared to *A. oryzae* and *S. fulvissimus*, *D. hansenii* has been more documented in human infections and accounts for up to 2% of invasive candidiasis cases [312]. For example, peritonitis, cholecystitis, mediastinitis, retinopathy, and bloodstream infections [313-316]. In most cases, *D. hansenii* infection occurred in immunocompromised hosts and the elderly. Nicholas D. Beyda et al. reviewed 10 reported cases showed that the overall mortality rate was low (10%) [312]. However, in a more recent review, Maria Karapetsa et al. demonstrated the mortality associated with *D. hansenii* is 5 in 9 total cases, which is a considerably high rate (56%) [317]. However, the review included patients with sepsis renal failure, which could also have contributed to the results. Hence, whether those symptoms are associated with more severe health outcomes and a higher mortality rate after *D. hansenii* infection warrants further investigation.

*D. hansenii* is also a member of the human gut microbiota. It has been found in both the healthy and unhealthy human gastrointestinal tract [318]. Studies have implicated the association between this gut microbe and CNS diseases. Specifically, *D. hansenii* was more prevalent (37.5%) in patients with multiple sclerosis than in controls (12.5%) [319]. Specifically, it was found in the serum as well as CSF specimens taken from multiple sclerosis patients. Further, the correlation of *D. hansenii* infection and Alzheimer's disease has also been reported [320-323]. It also plays a role in Candida meningitis in neurosurgical
patients. Moreover, Candida species, in general, can cause multiple brain abscesses [324, 325].

Although there is a considerable number of human infections related to *D. hansenii*, little is known about the virulence of this yeast species. A recent study examined the virulent potential of several isolated strains of *D. hansenii* from food as well as environmental and clinical situations [326]. The researchers used epithelial cell culture and Galleria mellonella as the model to test several physiological factors and parameters related to infection virulence in several yeast species that including *D. hansenii*. The results revealed that all the tested strains of *D. hansenii* were positive in at least a single virulent parameter, but no common pattern was observed. Specifically, no significant differences between clinical and non-clinical strains in phospholipase and pseudohyphal formation length were found. In addition, *D. hansenii* strains did not show clear virulence characteristics in the epithelial cell culture model. However, due to unclear virulence characteristics in the previous test, the researchers decided not to conduct Galleria mellonella experiments for *D. hansenii*. Because some of the tested strains were able to grow and produce pseudohyphae at 37°C and invade the agar plates showing characteristics of virulence similar to pathogenic yeast, the researchers suggested avoiding using those positive strains of *D. hansenii* in the food industry [326].

Nevertheless, the study of *D. hansenii* in rhesus macaques is extremely rare, particularly gut-to-brain invasion in these animals. Future studies could include more strains of *D. hansenii* from both non-human primates and human patients with different invasive infections, including strains from the gut and brain tissues. More research on the
phenotypic and genetic characteristics is required to unravel the virulence and the origin of those strains.

3.3.2.4 Discussion

Except for these three microbes introduced above, other species did not show a clear distribution pattern either across all samples or between groups (old vs. young animals or BMC_H vs. BMC_L). *A. oryzae* and *D. hansenii* had a similar distribution in all samples, whereas *S. fulvissimus* was more common in the old animals.

Two in three of these prevalent species shared by the gut and brain are fungus. Researchers have proposed that fungal toxin may play a role in the characteristic myelin degradation by destructing oligodendrocytes and astrocytes [319, 327]. Catherine B. Purzycki pointed out that certain pathogenic fungi such as *Aspergillus* and *Debaryomyces*, which we also detected, could mask from the immune system through their mannan coats, which are housed in non-neuronal tissue (e.g., gastrointestinal tract) and steadily releases toxins (e.g., gliotoxin) into the bloodstream. Once across the blood-brain barrier, these toxins target crucial cells in the brain such as oligodendrocytes, which provide nutritional maintenance for myelin, and astrocytes which are integral for maintaining the barrier. Without the proper support from glia cells, the blood-brain barrier weakens, and myelin degrades. Consequently, large-scale immune responses and brain inflammation could be triggered by that myelin debris. The symptoms of multiple sclerosis ensue with the further deconstruction of key components of the CNS. For example, ion imbalances, anoxia, conduction failure, mitochondrial depletion, redistribution of sodium channels, axon degeneration, and further demyelination [327].
Gliotoxin is a secondary metabolite produced by various species of *Aspergillus* and *Debaryomyces* [328]. It belongs to the class of immunosuppressive epipolythiodioxopiperazine toxins that have been associated with mycotoxicosis. It particularly destroys CNS astrocytes and oligodendrocytes and increases blood-brain barrier permeability, which leads to inflammation in CNS [327]. However, for *A. oryzae* and *D. hansenii* specifically, it is still unknown whether these two species produce gliotoxin, the types of mycotoxins, the virulence of those toxins, or the strain difference in the virulence.

In future studies, the symptoms, especially for those related to the gut and brain, need to be scrutinized among the animals carrying those species. Including control animals without infection is critical for investigating the impacts on the host health. In addition, we used isolated mononuclear cells that largely represent the innate immune cells in the brain. Future studies could use brain tissues from different regions, which may provide more microbiome information and allow further species detection. More samples from old and young animals could be tested to validate the age disparity in *S. fulvissimus* infection. Moreover, because these species are prevalent in food and the environment, the oral and skin microbiome could be tested from biopsies to help find the origin of the infections. However, the species were not examined by culture, and the strain information is yet to be clarified. Further research using the deep sequencing method could reveal the mutations and strain information as well as the expression of the gene related to toxic metabolites. These methods combined with culture and virulence test that to move beyond correlative studies could expand our understanding of these infections, therefore offering real potential to treat human disease.
Deciphering the relationship between diseases and microbes will meliorate the design of early screening methods for prominent species or strains and expand the current available strategies for preventing invasive infection and deterioration of health outcomes induced by those missing pathogenic factors.
3.3.3 Application to single-cell RNA-seq analysis

3.3.3.1 Microglia cells of SIV infected rhesus macaques

We next applied MTD to single-cell RNA-seq data from microglia cells isolated from SIV-infected rhesus macaques [329]. Because the analysis results from the authors identified SIV transcripts in the single cells, it is an ideal source of validation for the capacity of our MTD to process single-cell RNA-seq data.

First, count matrices of the microbiome were generated by MTD. Then, they were integrated with the host transcriptome and performed downstream analysis through the Seurat R package. The results showed that the major cell type in the sample was microglia, with a small portion of endothelial cells (Figure 30A). A cluster of microglia cells was highly infected with SIV (Figure 30B). For this cluster, the top 20 markers ranked by folder change are displayed in Figure 30C, such as PDE4A and SENP3. The cell subpopulation with these markers implicated a higher SIV tropism.

Overall, the results validated the capability of MTD for detecting specific microorganism species from single-cell RNA-seq data.
Figure 30. Detection of SIV in microglia cells from rhesus macaques

(A) Cell types in different colors on the UMAP plot. Cell type identity was assigned based on the homemade program described in "Methods" of Chapter 2. (B) SIV reads detected by MTD. The blue dots on the UMAP plot indicate the SIV-infected cells with the normalized reads quantity. (C) Markers of the cell cluster that harbor SIV. Analyses of the count matrix followed by visualization were performed through Seurat. Each dot in the UMAP plot represents every single cell. FindMarkers function with MAST methodology was used for computing the log2fold changes for each gene/feature between clusters and their corresponding adjusted p-values.
3.3.3.2 Brain cells of mice

We also applied MTD on the single-cell RNA-seq data of brain cells isolated from mice [188]. In addition to the analyses described in Chapter 2, we demonstrated the correlation analyses between the microbiome and host genes or pathways at the single-cell level.

We found *plasmodium vivax* (*P. vivax*) and host gene *GRIA2* had the highest correlation coefficient (Figure 31A). As shown in Figure 31B, *GRIA2* was tightly linked to the *P. vivax*-infected cells. We further collected all the host genes that highly correlated with *P. vivax* (Figure 31C), then performed pathway enrichment analysis. Results underscored the positive association of *P. vivax* with the function of the infected cell's plasma membrane region, such as cell junction and transmembrane transporter activity (Figure 31D). This result supports the cytoadherence phenomenon of *P. vivax* found in previous research [330-332]. Although *P. vivax* primary infects red blood cells, its cytoadherence on other cell types has been reported, such as in endothelial cells [330, 331]. Moreover, recent findings suggest that it has the ability to adhere to all Chinese hamster ovary (CHO) cells [331]. Our results bring insight into the interaction between *P. vivax* and host cells. It showed that *P. vivax* interacts with host cells that are incrementally expressing genes of the cellular membrane. Future research can study the causal effect of these molecules during infection, such as whether they contribute to pathogen adherence or if the infection leads to their increased expression.
Figure 31. Co-expression microbiome and host gene

(A) Visualization of the highest correlation on the UMAP plot. The cells expressed GRIA2 are represented by red dots, and the cells infected with *plasmodium vivax* are showed as green dots. The cells containing both are represented by the overlapping of the two colors (yellow). (B) A list of the top 20 correlations between host genes and microorganisms, and the highest pair is highlighted in red. (C) The results of pathway enrichment of the genes that were highly associated with *plasmodium vivax*, which were defined by $r > 0.2$ and $p < 0.05$. The top 3 results of each GO categories are highlighted. UMAP plots were drawn by the Seurat R package. The pathway enrichment was performed through g:GOSt in g:Profiler (version: e104_eg51_p15_3922dba, organism: mmusculus) [333].
3.4 Conclusion

MTD is a novel metatranscriptome detection and quantification pipeline, which can perform both bulk and single-cell RNA-seq data analysis. With this software, the activated microbiome (including the virome) can be detected in the sample, the cell type harboring them can be identified, and the differences in the hosts’ gene and microbiome expression can be calculated and associated. Thus, it would be a useful tool to improve our understanding of host-pathogen interactions, how the microbiome contributes to the host health, and what genes and pathways of the host correlate with a particular infection, then enlighten new ways of treatment or prevention of infectious diseases.
3.5 Discussion and future directions

RNA-seq data includes both host and non-host information. For the RNA-seq method, the microbiome must infect the cells and utilize the host’s cellular system to express its mRNA, which also means the activate microbiome in the host cells. These conditions differ from the microbiota in fecal samples, where the microbiome is not necessary to infect the host cells. Therefore, this transcriptome analysis of the active microorganism in the host cell provides further direct information about pathogen-host interactions.

MTD has several advantages compared to current tools in detecting metatranscriptome:

1. MTD enables simultaneous detection of the microbiome (including the virome) in the cells and the host gene expression in both bulk and single-cell RNA-seq data. To our knowledge, current tools have limited microbiome species coverage (partial bacteria or viruses) or are unable to identify the microbiome in single-cell RNA-seq data. This is partially due to the conventional alignment-based methods which have significantly high hardware requirements and slow processing speeds for detecting an extensive range of microorganisms. MTD uses an alignment-free detection approach, which is much faster and has relatively low hardware requirements.

2. Owing to the alignment-free detection approach, MTD is resistant to low sequence conservation due to mutation and recombination events [334] that often occur in microorganisms.
3. MTD has incomparable accessibility. Current analysis tools that depend on Docker, require administrator/root privilege to install, which restricts these tools' usage on most scientific computing resources such as HPC with multi-user systems. MTD is without the requirement of administrator/root privilege for users to install and use on HPC, which overcomes a hurdle for researchers in multi-user environments. Moreover, some tools can only be accessed through specific user accounts [248, 335], further limiting accessibility, especially for users not supported by NIH funding [335].

4. MTD warrants better data safety. MTD can be easily deployed on a local HPC of the institute/laboratory. Users do not need to apply particular accounts and upload their data to an outside cloud-based software.

5. MTD takes into account the host transcriptome size while modeling microbiome reads in differential abundance analysis, which better reflects the infectious severity of the exogenous species on the host.

6. MTD analyzes the association between the microbiome and the host gene expression or pathways. Thus, this information could offer researchers valuable insights into pathogen-host interaction.

7. MTD has an extensive virome detection ability. Comprehensive virome information from the Virus-Host database was included in the MTD. In addition to the NCBI RefSeq and GeneBank, it also contains references from UniProt, Viral Zone, and manually curated annotations.

8. MTD uses the two-step approach to gain better specificity [247]. First, host reads are extracted, then the microbiome is classified in the remaining non-
host reads. Thus, through the alignment-free approach, host reads could be identified and removed more explicitly, which further minimizes contamination from host reads during the microbiome analysis.

9. MTD is a user-friendly software that can be executed by one command line. It minimizes the programming knowledge requirements for its usage. Moreover, the classified reads in the intermediate steps are saved in the FASTQ format, which is beneficial for users who need additional analyses.

10. MTD has flexible host species choice. It initially supports three commonly used host species: human, mouse, and rhesus monkey. In addition, users can add other host species by just one command line.

Currently, it is common to perform the polyA tail enrichment during the library preparation for mRNA sequencing. Thus, MTD could avoid contamination from viruses in most cases because virus RNA only acquires polyA tail when it is transcribed in the host cell. Nevertheless, users need to be cautious about the exceptions of single-strand RNA viruses and other preparation methods that contain the polyA tail. Moreover, it is still challenging to avoid other contaminated microbes. There are tools to identify the potential contaminant by simply calculating the correlation of nucleic acid abundance between microbes and host [336]. However, these methods could lose their efficacy if the samples have similar abundance, heterogeneous contaminant patterns, or cross-contamination. MTD has a decontamination step that blacklists the common contaminant microbes in the laboratory environment [256]. Users can modify the list depending on the contaminant in their situations.
In the era of single-cell genomics consortia, the distribution of the microbiome in the cell population, tissues, and organ levels, and their associations with cell functions will be better analyzed and further understood. Researchers will be able to get insight into the pathogenesis of each microorganism identified. Furthermore, annotating the sample's geographic information with each microorganism would offer us a map of pathogens which could predict an epidemic. Thus, MTD could become a critical element for monitoring the spread of the microbiome and its pathogenesis in the future.
List of References:


CHAPTER 4 - ASSOCIATION OF ABNORMAL BEHAVIOR AND SIV INFECTION

4.1 Introduction

Behavioral disorders are common among HIV patients, and the presence of pre-existing psychopathology is the strongest predictor [337]. However, it remains challenging to distinguish the biological effect of HIV and the pre-morbid psychiatric illnesses contributing to the abnormal behaviors in humans. Moreover, captive nonhuman primates, usually the experimental animal used in HIV-associated research, also develop abnormal behaviors.

Abnormal behaviors occur in nonhuman primates due to both intrinsic and external factors. In rhesus macaques, the intrinsic risk factors could be an individual's sex, age, or genetic background. The external risk factors are rearing history, social experience, and housing situation [338, 339]. Impoverished early rearing experience was recognized as a strong predictor of abnormal behavior a long time ago [340, 341]. Harlow showed that the first six months of a monkey's life are critical to developing a pattern of abnormalities. William further suggested that the span between six and twelve months also have severe effects [342]. Although adolescent males were more aggressive than females [343], the sex effect did not occur in their first year [344]. Overall, behavioral issues declined with an individual's age increase [338, 343]. Moreover, genetic backgrounds including serotonin transporter gene polymorphism have contributed to behavioral differences [345]. In addition, past research has reported that mother-reared individuals were less likely to develop abnormal behaviors compared to their nursery-reared counterparts [346, 347]. Finally, single housing in adulthood has consistently been found to be a major predictor of
abnormal behaviors. The chances of developing either of these behaviors becomes higher over time [339, 348].

Because rhesus macaques are often used as an animal model in HIV/SIV research, their infection status and behavior are usually documented. Besides, their risk factors for abnormal behaviors such as the above mentioned can be better controlled in rhesus macaques than in humans. Thus, they could be good subjects to study the relationship between SIV infection and abnormal behaviors.

Furthermore, the geographic origin of rhesus macaques has been reported to contribute to their immunological and behavioral differences. For example, the major histocompatibility complex differs between rhesus monkeys of Indian and Chinese ancestry [349, 350]. In comparison with Indian rhesus macaques, SIV pathogenesis in Chinese rhesus macaques was closer to HIV-1 infections in untreated humans [351]. Indian and Chinese rhesus macaques also have different immune responses to the SIV antigen [352]. In addition, studies in Indian and Chinese-Indian hybrids demonstrated their difference in complete blood count and clinical biochemistry [353]. They also express disparities in temperament [354]. Compared with Indian origin, Chinese-Indian hybrid and Chinese origin exhibit more aggression and are more irritable toward other rhesus monkeys and humans [354]. Levels of CSF 5-HIAA, which has been inversely associated with impulsive and aggressive behaviors, are significantly lower in hybrid monkeys than in Indian origin [355]. A study from California National Primate Research Center found that more hybridized rhesus macaques are less fearful than purer Indian- and Chinese-origin macaques [356]. Moreover, Chinese-origin has been reported more likely to develop
repetitive behaviors than Indian-origin [357]. However, most of those studies have a small sample size and did not consider the influence of SIV infection on those behavior results.

TNPRC has a large colony of rhesus monkeys and an extensive archive of animal data in their animal record system, which became the ideal source for our retrospective analysis. It houses a total of about 27,000 rhesus macaques with about 10% being of Chinese origin. TNPRC has a long history of recording SIV inoculation, beginning in the year 1998. Furthermore, about 778 rhesus monkeys with abnormal behaviors have been recorded. The definitions of abnormal behaviors recorded during monitoring are shown in Table 3.

This study had two objectives. Our primary aim was to determine the relationship between the incidence of abnormal behavior and SIV infection in rhesus macaques. We further predicted that the Chinese geographic origin of the rhesus monkey would be associated with their abnormal behaviors.
Table 3. Abnormal Behaviors

<table>
<thead>
<tr>
<th>Abnormal Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-injured behavior (SIB)</strong></td>
</tr>
<tr>
<td>Non-injurious forceful self-directed behavior (nonbiting)</td>
</tr>
<tr>
<td>Sham-biting</td>
</tr>
<tr>
<td>Self-bite</td>
</tr>
<tr>
<td>Self-wounding</td>
</tr>
<tr>
<td><strong>Repetitive abnormal behavior (RAB)</strong></td>
</tr>
<tr>
<td>Bob</td>
</tr>
<tr>
<td>Bounce</td>
</tr>
<tr>
<td>Flip</td>
</tr>
<tr>
<td>Head toss</td>
</tr>
<tr>
<td>Pace</td>
</tr>
<tr>
<td>Repetitive licking</td>
</tr>
<tr>
<td>Rock</td>
</tr>
<tr>
<td>Spin</td>
</tr>
<tr>
<td>Swing</td>
</tr>
<tr>
<td>Other stereotypical locomotion</td>
</tr>
<tr>
<td><strong>Other abnormal behavior (OAB)</strong></td>
</tr>
<tr>
<td>Bizarre posture</td>
</tr>
<tr>
<td>Coprophagy</td>
</tr>
<tr>
<td>Feces paint</td>
</tr>
<tr>
<td>Self-clasp</td>
</tr>
<tr>
<td>Self-oral</td>
</tr>
<tr>
<td>Regurgitate</td>
</tr>
<tr>
<td>Urophagy</td>
</tr>
<tr>
<td>Eye poke/ salute (peri orbital contact)</td>
</tr>
<tr>
<td>Floating limb</td>
</tr>
</tbody>
</table>
4.2 Methods

4.2.1 Data collection

Raw data was downloaded from the LabKey animal record system (ARS) in TNPRC on August 06, 2020. The data was downloaded in batches in the .xlsx format which included demographics, inoculation history, necropsy, and housemate history. A total number of 33,715 animals, including 5,182 alive, 28,515 dead, and 18 transferred, were recorded in the LabKey ARS at the time of download.

4.2.2 Data farming

The raw data from LabKey ARS lacked records on rearing type (mother/nursery reared) and single housing, such as first single housing time and lifetime tenure in the single housing, which were necessary for our analysis. In order to get this essential information, we computed the relevant data in the housemate history and demographics. Two calculators were written through Visual Basic Application (VBA) and used in MS Excel Macros to compute single housing and rearing type information. Microsoft Office 365 or Office 2019 is required to run the calculators.

The number of days from the animal’s birth to when it was single-housed is denoted as its first single housing. We calculate it based on the first time point that is listed as empty under housemate information for each animal. Then, we categorize the data as less than a year (< 1 year), one year to less than two years (1- <2years), and equal to or more than two years (>=2 years).

Further, “Lifetime tenure in single housing” means the total days in the single housing in an animal's lifetime. We categorized this section as less than six months (<6
mons), six months to less than one year (6 mons - < 1 year), one year to less than two years (1- <2 years), and equal to or more than two years (>=2 years).

The source code of the calculator for single housing information is displayed below.

```vba
Sub HousingRoommatesCalculation()
    Dim Animalcount As Integer
    Dim i As Integer
    Dim ID As String
    Dim Birthday As Date
    Dim Independentday As Date
    Dim totalS As Integer
    Dim startS As Integer

    Worksheets("data").Cells(1, 17) = "First single housing (Days)"
    Worksheets("data").Cells(1, 18) = "Lifetime tenure in single housing (Days)"
    Worksheets("data").Cells(1, 19) = "First single housing: < 1 year"
    Worksheets("data").Cells(1, 20) = "1- <2years"
    Worksheets("data").Cells(1, 21) = ">=2years"
    Worksheets("data").Cells(1, 22) = "Life time tenure in single housing: <6 mons"
    Worksheets("data").Cells(1, 23) = "6 mons - <1 year"
    Worksheets("data").Cells(1, 24) = "1- <2 years"
    Worksheets("data").Cells(1, 25) = " >=2 years"

    Worksheets("data").Range("P1").Formula2 = "=UNIQUE(data!A:A)"

    Animalcount = Application.WorksheetFunction.CountIf(Worksheets("data").Range("P:P"), "*")

    For i = 2 To Animalcount
        ID = Worksheets("data").Cells(i, 16)
        Birthday = Application.WorksheetFunction.MinIfs(Worksheets("data").Range("E:E"),
            Worksheets("data").Range("A:A"), ID)
        Independentday = Application.WorksheetFunction.MinIfs(Worksheets("data").Range("E:E"),
            Worksheets("data").Range("A:A"), ID, Worksheets("data").Range("B:B"), "")
        If Independentday <> 0 Then
            startS = DateDiff("d", Birthday, Independentday)
        Else
            startS = 0
        End If
        Worksheets("data").Cells(i, 17) = startS
        totalS = Application.WorksheetFunction.SumIfs(Worksheets("data").Range("G:G"),
            Worksheets("data").Range("A:A"), ID, Worksheets("data").Range("B:B"), "")
        Worksheets("data").Cells(i, 18) = totalS
    Next i
End Sub
```
If startS > 0 And startS < 365 Then
    Worksheets("data").Cells(i, 19) = "Yes"
ElseIf startS >= 365 And startS < 365 * 2 Then
    Worksheets("data").Cells(i, 20) = "Yes"
ElseIf startS >= 365 * 2 Then
    Worksheets("data").Cells(i, 21) = "Yes"
End If
If totalS > 0 And totalS <= 365 * 0.5 Then
    Worksheets("data").Cells(i, 22) = "Yes"
ElseIf totalS >= 365 * 0.5 And totalS < 365 Then
    Worksheets("data").Cells(i, 23) = "Yes"
ElseIf totalS >= 365 And totalS < 365 * 2 Then
    Worksheets("data").Cells(i, 24) = "Yes"
ElseIf totalS >= 365 * 2 Then
    Worksheets("data").Cells(i, 25) = "Yes"
End If
Next i

End Sub

The rearing type was measured by whether the animal was reared either by the mother or by nursery. Mother reared was applicable if the monkey stayed within its dam for the vast majority of the first six months of life. Otherwise, the monkey was considered nursery reared. In a few cases, the baby or the mother may be separated for clinical treatment. So, if the monkeys were separated from their mothers for less than eight days within the first six months of life, it was still considered as mother reared.

The source code of the calculator for rearing type information is displayed here.

Sub Rearing()
    Dim birthdayC, birthdayR, bcend1, bcend2, damtime, totalM, Animalcount, i, m, damtimes As Integer
    Dim datevalue, cagemateend As Date
    Dim ID, dam As String

    Animalcount = Application.WorksheetFunction.CountIf(ActiveSheet.Range("L:L"), "*")
    For i = 2 To Animalcount
        ID = ActiveSheet.Cells(i, 12)
After the computation of all batches, the data was compiled into a single spreadsheet. Consequently, the data was grown and organized to facilitate the following analysis.
4.2.3 Data cleaning

The data cleaning process was conducted by removing the irrelevant records (Figure 32). Firstly, we removed the 18 transferred animals whose status was recorded as neither alive nor dead. Secondly, the 1,242 animals with undetermined or unknown sex were removed. Thirdly, 5,946 animals of species other than Rhesus macaques were removed. Fourthly, 12,697 animals that had a recorded death date earlier than the year 1998 were removed because there was no SIV inoculation history in ARS earlier than that year. Finally, 1,522 animals without records about their first single housing date were removed. As a result, a total of 12,290 animals remained after the data cleaning process (Table 4).

Figure 32. Data cleaning
Table 4. Demographic Table of Abnormal Behavior Analysis

<table>
<thead>
<tr>
<th>Individual-level variables</th>
<th>N</th>
<th>Percent</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>12290</td>
<td></td>
<td>5.4</td>
<td>5.0</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>6148</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>6142</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indian</td>
<td>11104</td>
<td>90.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>1178</td>
<td>9.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First single housing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;1 year</td>
<td>1728</td>
<td>14.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-&lt;2years</td>
<td>1363</td>
<td>11.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;=2years</td>
<td>9199</td>
<td>74.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime tenure in single housing (days)</td>
<td>12290</td>
<td></td>
<td>359.9</td>
<td>605.1</td>
</tr>
<tr>
<td>SIV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No inoculation</td>
<td>8964</td>
<td>72.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Innoculation</td>
<td>3326</td>
<td>27.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIV present length (days)</td>
<td>12290</td>
<td></td>
<td>116.8</td>
<td>321.4</td>
</tr>
<tr>
<td>Abnormal behaviors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>11825</td>
<td>96.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIB/RAB/OAB</td>
<td>465</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rearing types</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nursery reared</td>
<td>4147</td>
<td>33.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother reared</td>
<td>8143</td>
<td>66.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SIB: Self-injured behavior  
RAB: Repetitive abnormal behavior  
OAB: Other abnormal behavior

4.2.4 Data analysis

Next, the categorical variables were coded for statistical analysis. The nursery reared rhesus macaques were coded as 0, and mother reared were coded as 1. Since the birth of the animal, the first single housing time was coded as three scales: 1 for less than a year, 2 for one year to less than two years, and 3 for equal to or more than two years. An animal's lifetime tenure in the single housing was coded to four scales: 1 for less than six months, 2 for six months to less than one year, 3 for one year to less than two years, 4 for equal to or more than two years. Abnormal behavior was coded as 1, and normal behavior
was 0. The abnormal behavior contained self-injured behavior (SIB), repetitive abnormal behavior (RAB), and other abnormal behavior (OAB). The animal with a history of inoculation of any SIV strain was coded as 1, otherwise as 0. Rhesus macaques of Indian origin were code as 0 and Chinese origin as 1. The sex of female was coded as 0 and male as 1.

Logistic regression was used to analyze the relationship between the abnormal behavior and other variables, including rearing type, first single housing time, lifetime tenure in the single housing, SIV infection, SIV present days, sex, age, and geographic origin of the rhesus monkey. The backward stepwise (likelihood ratio) method was used to select the independent variables in the model. The quality of the model was examined by the pseudo R-square, multicollinearity, area under the receiver operating curve (AUROC), and concordance ratio.

Statistical analyses were performed using R (version 4.0.3) and SPSS software (IBM SPSS Statistics for Macintosh, Version 27.0.).
4.3 Results

The logistic regression results were shown in Table 3. Rhesus macaques with SIV inoculation history were about 1.8 times more likely to have abnormal behavior than those that had never been inoculated with SIV. Moreover, the Chinese origin rhesus macaques are nearly twice as likely to behave abnormally than those of Indian origin. In addition, the first single housing time was negatively associated with abnormal behavior, which means the longer time the animal lived with the housemate since birth, the less likely it behaved abnormally. Sex has an obvious impact on abnormal behaviors. Male rhesus macaques were 4.2 times more likely to have abnormal behaviors than females.

However, the lifetime tenure in the single housing and the age of the animal did not show an obvious influence on abnormal behaviors. Rearing type and SIV present days were not significant and were removed during the backward stepwise selection of the model.

Furthermore, the pseudo R-squared value of Nagelkerke's was 0.238 and McFadden’s was 0.210, which indicated the model was adequate to predict the outcome. To further acquaint the quality of the model, we tested the multicollinearity among the independent variables, sensitivity, specificity, AUROC, and concordance ratio.
Table 5. Logistic regression results.

Variables included in final model. Variables with positive beta (B) estimates positively predict abnormal behavior; variables with negative beta estimates negatively predict abnormal behavior.

Logistic regression analysis of abnormal behaviors of SIV-infected or uninfected rhesus macaques housed in the TNPRC

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>First single housing</td>
<td>-0.578</td>
<td>0.063</td>
<td>84.121</td>
<td>1</td>
<td>0.000</td>
<td>0.561</td>
</tr>
<tr>
<td>Lifetime tenure in single housing</td>
<td>0.001</td>
<td>0.000</td>
<td>210.429</td>
<td>1</td>
<td>0.000</td>
<td>1.001</td>
</tr>
<tr>
<td>SIV</td>
<td>0.570</td>
<td>0.108</td>
<td>27.938</td>
<td>1</td>
<td>0.000</td>
<td>1.769</td>
</tr>
<tr>
<td>Sex</td>
<td>1.440</td>
<td>0.127</td>
<td>128.778</td>
<td>1</td>
<td>0.000</td>
<td>4.222</td>
</tr>
<tr>
<td>Age</td>
<td>0.037</td>
<td>0.015</td>
<td>6.358</td>
<td>1</td>
<td>0.012</td>
<td>1.038</td>
</tr>
<tr>
<td>Origin</td>
<td>0.676</td>
<td>0.144</td>
<td>22.106</td>
<td>1</td>
<td>0.000</td>
<td>1.965</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.918</td>
<td>0.196</td>
<td>400.238</td>
<td>1</td>
<td>0.000</td>
<td>0.020</td>
</tr>
</tbody>
</table>

AUROC = 0.832
Pseudo R² = 0.238
n = 12,290
Multicollinearity diagnostics of the logistic regression model were conducted according to previously published guide [358]. We examined the tolerance, variance inflation factor (VIF), condition index, variance proportions, and correlations between predictors (Figure 23). No tolerance value was equal to or less than 0.1, and no VIF value was larger than 2.5 (Figure 23C). No condition index was larger than 15, and no pair of predictors was larger than 0.7 in the variance proportions (Figure 23B). Moreover, the correlation matrix presented in Figure 23A showed the correlation coefficients among the independent variables. No pair of variables was larger than 0.7 in the correlation coefficient. Overall, these results indicate that our model was without a noticeable issue of multicollinearity.

Next, the discrimination of the model was assessed using AUROC. The value of the AUROC is the probability that an animal who was abnormal behavior had a higher predicted probability than did an animal who was normal behavior. Using the plotROC function in the InformationValue R package to calculate the AUROC yielded a value of 0.832, indicating that the model discriminates well (Figure 23E). Additionally, the concordance ratio was 0.846, sensitivity was 0.106, and specificity was 0.997 using the corresponding functions in the same R package. The classification table from SPSS showed that our model was correctly classifying the outcome for 96.3% of the cases (Figure 23D). These results indicated that our model could confidently discriminate abnormal behavior from normal behavior.
Figure 33. Multicollinearity and discrimination tests
4.4 Conclusion

The association between SIV infection and abnormal behavior in non-human primates remains unclear. The animal record system in the non-human primate research center provided a pool of information to be further explored. Through data farming, we grew and harvested the data to get the necessary information for our study. Following data cleaning, we used logistic regression to analyze the 12,290 animal records from the LabKey ARS in TNPRC. Furthermore, we evaluated the model’s pseudo R-square, collinearity, and discrimination.

Therefore, we report two key findings:

1. Rhesus macaques with SIV inoculation history were about 1.8 times more likely to show abnormal behavior than one that has never been inoculated with SIV.

2. Chinese origin rhesus macaques are nearly twice as likely to behave abnormally than those of Indian origin.

In conclusion, SIV and geographic origin from China were risk factors of abnormal behavior in rhesus macaques. Because lifestyle variables were well controlled in the captive rhesus macaques, the impact of SIV on the host's behavior was illustrated more directly. Our findings could shed light on the HIV infection-induced behavioral disorders in human patients.
4.5 Discussion and future studies

We found that SIV infection, Chinese-origin, and sex are positively associated with abnormal behaviors in rhesus macaques, while social enrichment is negatively associated. Those findings corresponded to what we would have expected based on the literature.

4.5.1 SIV

So far in the published literature, only one case report has shown the SIV-related abnormal behavior [359]. A young rhesus macaque with SIV and Rhesus cytomegalovirus (RhCMV) coinfection displayed self-injurious behavior (SIB). The author suggested that immunosuppression due to SIV made RhCMV cause peripheral neuropathies, leading to SIB in the monkey. RhCMV is a common infection in macaques with nearly 100% prevalence in most colonies, including those from TNPRC. This suggests that RhCMV could be a "baseline" condition, and SIV infection is the key risk factor that triggers the change in behavior. Our study on a large population found that rhesus monkeys with SIV infection have about 1.8 times more likely to express abnormal behaviors. It controlled for rearing type, single housing, age, sex, and origin. Our study directly revealed the positive correlation between SIV infection and abnormal behavior in rhesus macaques.

In humans, the behavioral disorders caused by HIV may be mainly due to pre-existing psychiatric illnesses [337]. Hence, it would be better to count the pre-existing abnormal behaviors before the SIV inoculation in our study. Including this as a covariance in the future analysis could help differentiate the effects of pre-morbid behavior disorder from the biological effects of SIV. Nevertheless, this may not be a problem because experimentally, researchers tend to avoid enrolling any abnormal behavior animal in SIV-
related biomedical studies. Thus, abnormal behavior should be observed after the SIV inoculation.

Moreover, the effect of ART on mental health and behaviors in humans is still elusive. While it may reduce the damage of HIV on the CNS and devitalize the HIV-related risk behaviors, the toxicity of the drug could lead to neurobehavioral symptoms \cite{174, 360, 361}. In the future, the administration and total duration of ART could be included in the independent variables to test if the treatments help to reduce or increase the onset of abnormal behavior.

4.5.2 Origin

We found that rhesus macaques of Chinese origin were nearly twice as likely to display abnormal behavior than those of Indian origin. In 2015, researchers at the Oregon National Primate Research Center (ONPRC) had reported that Chinese rhesus monkeys were 49% more likely to display abnormal behavior than Indian origin \cite{357}. Our results further validated this claim with a much higher odds ratio.

Chinese and Indian origin rhesus have some known genetic and immunological differences \cite{349, 350, 362}. Although we controlled for differences in rearing types and social separation in our analysis, other factors may have contributed to the observed difference in abnormal behavior incidence between Chinese and Indian Rhesus monkeys. For example, Chinese origin may increase fearfulness in the presence of humans.

Since this origin-related behavior difference has been observed in two non-human Primate Research Centers - TNPRC and ONPRC, an extensive collaborative studies involving more centers housing both the Chinese and Indian rhesus monkeys would help
to validify these findings in the future. Moreover, the underlying cause of the discrepancy in abnormal behaviors requires further investigation.

4.5.3 Social enrichment

The results of several studies have indicated that macaques housed alone are more likely to show anxiety and display abnormal behaviors than macaques housed with others [338, 348, 363]. We further corroborate the findings that late single-housed of monkeys were significantly less likely to demonstrate abnormal behaviors, with only about 50% chance. In addition, it has been reported that lifetime tenure in single housing positively correlated with the expression of abnormal behavior [339, 347]. However, we did not observe the effect of this lifetime duration in single housing on the incidence of abnormal behaviors. Our results further suggest that early life stress such as social deprivation at an early age significantly increases the likelihood of developing abnormal behaviors.

4.5.4 Rearing type, sex, and age

In previous studies, differences in rearing type, age, and sex have been found to influence abnormal behavior in rhesus macaques [338, 339]. Thus, these factors were included as covariates in our model. We found that males tend to exhibit more abnormal behavior than females, which is consistent with previous research. While other groups have observed abnormal behavior to decrease with age, we did not see any apparent difference. In addition, we did not find significant disparities in abnormal behavior between mother-reared monkeys and nursery-reared monkeys, and so we removed this variable from the model during the backward stepwise selection.
4.5.5 Association with cellular senescence

Currently, there is not enough evidence to assert that rhesus macaques with abnormal behaviors have brain pathology. A recent study proved that senescent cells in some brain areas induce anxiety-related behaviors in rodents [364]. Therefore, it will be interesting to examine the cellular senescence in the brain of those abnormally behaved rhesus macaques. In the future, by taking advantage of the senescent cell measurement techniques mentioned before, such as quantifying lipofuscin and senescence burden in different brain regions, the association of cellular senescence and abnormal behavior in macaques could be understood. However, considering the difficulties of conducting a prospective study with a large population on those non-human primates, a retrospective study on ARS/database with archived brain tissues is more pragmatic. Hence, it would be an ideal scenario for automated lipofuscin measurement technology that avoids the shortcomings of archived tissue, thereby helping to associate an abnormal phenotype (behavior) with cellular senescence.
List of References:


APPENDIX

1. Reagents used in Chapter 2

Table 6. Reagents used in Chapter 2

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Vendor</th>
<th>Cat #</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNAlater™ Stabilization Solution</td>
<td>ThermoFisher</td>
<td>AM7021</td>
</tr>
<tr>
<td>RNeasy FFPE Kit</td>
<td>QIAGEN</td>
<td>73504</td>
</tr>
<tr>
<td>RNeasy Mini Kit (50)</td>
<td>QIAGEN</td>
<td>74104</td>
</tr>
<tr>
<td>RNeasy Plus Mini Kit (50)</td>
<td>QIAGEN</td>
<td>74134</td>
</tr>
<tr>
<td>QIAshredder (250)</td>
<td>QIAGEN</td>
<td>79656</td>
</tr>
<tr>
<td>Doxorubicin hydrochloride</td>
<td>Sigma-Aldrich</td>
<td>D1515-10MG</td>
</tr>
<tr>
<td>ACK Lysing Buffer</td>
<td>ThermoFisher</td>
<td>A1049201</td>
</tr>
<tr>
<td>RPMI 1640 Medium</td>
<td>ThermoFisher</td>
<td>11875119</td>
</tr>
<tr>
<td>PBS, pH 7.4</td>
<td>ThermoFisher</td>
<td>10010049</td>
</tr>
<tr>
<td>ViaStain™ AOPI Staining Solution</td>
<td>Nexcelom Bioscience</td>
<td>CS2-0106-5mL</td>
</tr>
<tr>
<td>Fetal Bovine Serum, heat inactivated</td>
<td>ThermoFisher</td>
<td>10438026</td>
</tr>
<tr>
<td>Penicillin-Streptomycin</td>
<td>Sigma-Aldrich</td>
<td>P4333-100ML</td>
</tr>
<tr>
<td>Hibernate™-A Medium</td>
<td>ThermoFisher</td>
<td>A1247501</td>
</tr>
<tr>
<td>HBSS, no calcium, no magnesium, no phenol red</td>
<td>ThermoFisher</td>
<td>14175095</td>
</tr>
<tr>
<td>HBSS, no calcium, no magnesium</td>
<td>ThermoFisher</td>
<td>14170112</td>
</tr>
<tr>
<td>Percoll</td>
<td>Sigma-Aldrich</td>
<td>GE17-0891-01</td>
</tr>
<tr>
<td>Reverse Transcription System</td>
<td>Promega</td>
<td>A3500</td>
</tr>
<tr>
<td>TaqMan™ Fast Advanced Master Mix</td>
<td>ThermoFisher</td>
<td>4444557</td>
</tr>
<tr>
<td>TaqMan™ RNA-to-CT™ 1-Step Kit</td>
<td>ThermoFisher</td>
<td>4392938</td>
</tr>
<tr>
<td>TaqMan™ RNase P Detection Reagents Kit</td>
<td>ThermoFisher</td>
<td>4316831</td>
</tr>
<tr>
<td>CDKN2A - TaqMan® Gene Expression Assays</td>
<td>ThermoFisher</td>
<td>Hs99999189_m1</td>
</tr>
<tr>
<td>CDKN1A - TaqMan® Gene Expression Assays</td>
<td>ThermoFisher</td>
<td>Rh02822707_m1</td>
</tr>
<tr>
<td>GAPDH - TaqMan® Gene Expression Assays</td>
<td>ThermoFisher</td>
<td>Rh02621745_g1</td>
</tr>
<tr>
<td>MKI67 - TaqMan® Gene Expression Assays</td>
<td>ThermoFisher</td>
<td>Rh01032443_m1</td>
</tr>
<tr>
<td>CCND1 - TaqMan® Gene Expression Assays</td>
<td>ThermoFisher</td>
<td>Rh04386126_m1</td>
</tr>
<tr>
<td>CAV1 - TaqMan® Gene Expression Assays</td>
<td>ThermoFisher</td>
<td>Rh02839690_m1</td>
</tr>
<tr>
<td>PAI-1 (SERPINE1) - TaqMan® Gene Expression Assays</td>
<td>Fisher Scientific</td>
<td>Rh02829204_m1</td>
</tr>
<tr>
<td>IL6 - TaqMan® Gene Expression Assays</td>
<td>ThermoFisher</td>
<td>Rh02621719_u1</td>
</tr>
<tr>
<td>Iron measurement kit</td>
<td>Beckman Coulter</td>
<td>OSR6286</td>
</tr>
<tr>
<td>LIVE/DEAD™ Fixable Aqua Dead Cell Stain Kit,</td>
<td>ThermoFisher</td>
<td>L34957</td>
</tr>
</tbody>
</table>
2. Source code used in Chapter 2 single-cell RNA-seq analysis

```r
### Setup the Seurat objects ###
library(Seurat)
library(patchwork)
library(ggplot2)
library(sctransform)
setwd("/master/fwu/R_Works")
# Load the 10X datasets (matrix; folder)
sc_OBrain1 <- Read10X(data.dir="/master/fwu/R_Works/GSE161340/GSE161340_RAW/sc_OBrain1_mm10/")
sc_OBrain3 <- Read10X(data.dir="/master/fwu/R_Works/GSE161340/GSE161340_RAW/sc_OBrain3_mm10/")
sc_YBrain1 <- Read10X(data.dir="/master/fwu/R_Works/GSE161340/GSE161340_RAW/sc_YBrain1_mm10/")
sc_YBrain3 <- Read10X(data.dir="/master/fwu/R_Works/GSE161340/GSE161340_RAW/sc_YBrain3_mm10/")
# transfer to dataframe for reprocessing
sc_OBrain1<-as.data.frame(sc_OBrain1)
sc_OBrain3<-as.data.frame(sc_OBrain3)
sc_YBrain1<-as.data.frame(sc_YBrain1)
sc_YBrain3<-as.data.frame(sc_YBrain3)
# clean the names of cell barcodes
colnames(sc_OBrain1)<-gsub("-1","",colnames(sc_OBrain1))
colnames(sc_OBrain3)<-gsub("-1","",colnames(sc_OBrain3))
colnames(sc_YBrain1)<-gsub("-1","",colnames(sc_YBrain1))
colnames(sc_YBrain3)<-gsub("-1","",colnames(sc_YBrain3))
## to uppercase the gene names
library(tibble)
library(dplyr)
Uppercase <- function(df){
  df<-add_column(df, NAME=toupper(row.names(df)), .before = 1)
  df<-df %>% group_by(NAME) %>% summarise_all(list(mean)) #average the duplicated rows
  df <- as.data.frame(df) # format tibble to dataframe
  row.names(df) <- df[,1] # rename the row name with uppercased gene names
  df <- df[,-1] # remove the "NAME" column
}
sc_OBrain1<-Uppercase(sc_OBrain1)
sc_OBrain3<-Uppercase(sc_OBrain3)
sc_YBrain1<-Uppercase(sc_YBrain1)
sc_YBrain3<-Uppercase(sc_YBrain3)
# combine host and microbiome Sc data
sc_OBrain1.micro<-read.delim("/Users/feiwu/Box/RNA-Seq/SingleCell/DualScseq/example/SC_Old_Mouse_1_and_2/Report_non-host_sc_OBrain1.kraken.c.tsv_Count.txt", row.names = 1)
```
sc_YBrain1.micro <- read.delim("/Users/feiwu/Box/RNA-Seq/SingleCell/Dual-Scseq/example/SC_Young_Mouse_1_and_2/Report_non-host_sc_YBrain1.kraken.c.tsv_Count.txt", row.names = 1)
sc_OBrain1 <- rbind(sc_OBrain1, sc_OBrain1.micro)
scc_OBrain3 <- rbind(sc_OBrain3, sc_OBrain3.micro)
scc_YBrain1 <- rbind(sc_YBrain1, sc_YBrain1.micro)
scc_YBrain3 <- rbind(sc_YBrain3, sc_YBrain3.micro)

# make dataframe to Seurat objects
sc_OBrain1 <- CreateSeuratObject(sc_OBrain1, project = "OBrain1")
scc_OBrain3 <- CreateSeuratObject(sc_OBrain3, project = "OBrain3")
scc_YBrain1 <- CreateSeuratObject(sc_YBrain1, project = "YBrain1")
scc_YBrain3 <- CreateSeuratObject(sc_YBrain3, project = "YBrain3")

# add Seurat objects into a list
Brain.list <- list(sc_OBrain1, sc_OBrain3, sc_YBrain1, sc_YBrain3)

# SCTransform for each dataset independently
Brain.list <- lapply(Brain.list, SCTransform)
features <- SelectIntegrationFeatures(object.list = Brain.list, nfeatures = 3000)
Brain.list <- PrepSCTIntegration(object.list = Brain.list, anchor.features = features)

### Perform integration ###
Brain.anchors <- FindIntegrationAnchors(object.list = Brain.list, normalization.method = "SCT", anchor.features = features)
Brain.combined.sct <- IntegrateData(anchorset = Brain.anchors, normalization.method = "SCT")

Brain.combined.sct <- RunPCA(Brain.combined.sct, verbose = FALSE)
Brain.combined.sct <- RunUMAP(Brain.combined.sct, reduction = "pca", dims = 1:30)

### find & label the clusters
Brain.combined.sct <- FindNeighbors(Brain.combined.sct, dims = 1:30)
Brain.combined.sct <- FindClusters(Brain.combined.sct, resolution = 0.5)

# find markers for every cluster compared to all remaining cells, report only the positive ones
Brain.markers <- FindAllMarkers(Brain.combined.sct, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25)
Brain.markers %>% group_by(cluster) %>% top_n(n = 2, wt = avg_log2FC)
Brain.markers <- Brain.markers[order(Brain.markers$avg_log2FC, decreasing = T),] #sort by avg_log2FC

# edit meta data
library(dplyr)
Brain.combined.sct$meta.data <- Brain.combined.sct$meta.data %>%
  mutate(age = if_else(Orig.ident %in% c("OBrain1", "OBrain3"), "24 months", "4 months"))
Brain.combined.sct$group <-
paste0(Brain.combined.sct$age, Brain.combined.sct$Senescent_Status)
Brain.combined.sct$group <- gsub("24 monthsNonSC", "Old_NonSC", Brain.combined.sct$group)
Brain.combined.sct$group <- gsub("24 monthsSC", "Old_SC", Brain.combined.sct$group)
Brain.combined.sct$group <- gsub("4 monthsNonSC", "Young_NonSC", Brain.combined.sct$group)
Brain.combined.sct$group <- gsub("4 monthsSC", "Young_SC", Brain.combined.sct$group)
rm(c(Brain.anchors, Brain.list))  # optional: to save memory
DimPlot(Brain.combined.sct, reduction = "umap", label = TRUE, repel = TRUE)
DimPlot(Brain.combined.sct, reduction = "umap", split.by = "age")

# Export the integrated count matrix
old <- as.data.frame(Brain.combined.sct[["RNA"]][, Brain.combined.sct$age == "24 months"])
young <- as.data.frame(Brain.combined.sct[["RNA"]][, Brain.combined.sct$age == "4 months"])
old <- old[which(rowSums(old) > 0),]  # remove rows with all 0
young <- young[which(rowSums(young) > 0),]

gct.making <- function(uppercased.df){
  gct1 <- add_column(uppercased.df, GENE=row.names(uppercased.df), Description=row.names(uppercased.df), .before=1)
  row1 <- "#1.2"[is.na(gct2[1:2,])] <- ""
  write.table(gct2, paste0(deparse(substitute(uppercased.df)), "_.gct"),
              col.names = F, row.names = F, sep="\t", quote = F)
}

gct.making(old)
gct.making(young)

# import cell type scored results
senescent.scored.old <- read.delim("/Users/feiwu/Box/RNA-Seq/SingleCell/GSE161340/SenescentCell-results-scores-old.gct",skip=2,row.names=1)
senescent.scored.old <- senescent.scored.old[rownames(old)]
senescent.scored.young <- read.delim("/Users/feiwu/Box/RNA-Seq/SingleCell/GSE161340/SenescentCell-results-scores-young.gct",skip=2,row.names=1)
senescent.scored.young <- senescent.scored.young[rownames(young)]
# add to meta data: senescent scores
Brain.combined.sct$Senescence_core_genes <-
c(t(senescent.scored.old),"Senescence_core_genes",t(senescent.scored.young),"Senescence_core_genes")

# divide to senescent and non-senescent groups
Brain.combined.sct@meta.data <- Brain.combined.sct@meta.data %>%
  mutate(Senescent_Status = if_else(Brain.combined.sct$Senescence_core_genes>3, "SC", "NonSC"))

### Microbiome Analysis ###
# VIRUS data extraction: senescent/non-senescent cell
old.SC.virus <-
old[grep("virus",row.names(old)),intersect(1:ncol(old),grep("^SC$",Brain.combined.sct$Senescent_Status))]
old.NSC.virus <-
old[grep("virus",row.names(old)),intersect(1:ncol(old),grep("^NonSC$",Brain.combined.sct$Senescent_Status))]
young.SC.virus <-
young[grep("virus",row.names(young)),intersect(1:ncol(young),grep("^SC$",Brain.combined.sct$Senescent_Status))]
young.NSC.virus <-
young[grep("virus",row.names(young)),intersect(1:ncol(young),grep("^NonSC$",Brain.combined.sct$Senescent_Status))]

# All Microbiome data extraction: senescent/non-senescent cell
old.SC.Mb <-
old[grep("taxid",row.names(old)),intersect(1:ncol(old),grep("^SC$",Brain.combined.sct$Senescent_Status))]
old.NSC.Mb <-
old[grep("taxid",row.names(old)),intersect(1:ncol(old),grep("^NonSC$",Brain.combined.sct$Senescent_Status))]
young.SC.Mb <-
young[grep("taxid",row.names(young)),intersect(1:ncol(young),grep("^SC$",Brain.combined.sct$Senescent_Status))]
young.NSC.Mb <-
young[grep("taxid",row.names(young)),intersect(1:ncol(young),grep("^NonSC$",Brain.combined.sct$Senescent_Status))]

## alpha diversity analysis ##
library(vegan)
## VIRUS ##
# test alpha diversity of viruses in total cells
old.SC.virus.diversity <-apply(old.SC.virus,2,diversity)
old.NSC.virus.diversity <-apply(old.NSC.virus,2,diversity)
t.test(old.SC.virus.diversity,old.NSC.virus.diversity)
t.test(old.SC.virus.diversity[old.SC.virus.diversity!=0],
      old.NSC.virus.diversity[old.NSC.virus.diversity!=0])
young.SC.virus.diversity <- apply(young.SC.virus, 2, diversity)
young.NSC.virus.diversity <- apply(young.NSC.virus, 2, diversity)
t.test(young.SC.virus.diversity, young.NSC.virus.diversity)

# viral infected cells
old.SC.virus.infected <- old.SC.virus[colSums(old.SC.virus) != 0]
old.NSC.virus.infected <- old.NSC.virus[colSums(old.NSC.virus) != 0]
old.SC.virus.infected.diversity <- apply(old.SC.virus.infected, 2, diversity)
old.NSC.virus.infected.diversity <- apply(old.NSC.virus.infected, 2, diversity)
t.test(old.SC.virus.infected.diversity, old.NSC.virus.infected.diversity)

young.SC.virus.infected <- young.SC.virus[colSums(young.SC.virus) != 0]
young.NSC.virus.infected <- young.NSC.virus[colSums(young.NSC.virus) != 0]
young.SC.virus.infected.diversity <- apply(young.SC.virus.infected, 2, diversity)
young.NSC.virus.infected.diversity <- apply(young.NSC.virus.infected, 2, diversity)
t.test(young.SC.virus.infected.diversity, young.NSC.virus.infected.diversity)

# add viral infectious status of each cells to metadata
Brain.combined.sct$Virus_infection <- "-
Brain.combined.sct$Virus_infection[names(old.SC.virus.infected)] <- "Old_SC_Viral_Pos"
Brain.combined.sct$Virus_infection[names(old.SC.virus[!colSums(old.SC.virus) == 0])] <- "Old_SC_Viral_Neg"

Brain.combined.sct$Virus_infection[names(old.NSC.virus.infected)] <- "Old_NSC_Viral_Pos"
Brain.combined.sct$Virus_infection[names(old.NSC.virus[!colSums(old.NSC.virus) == 0])] <- "Old_NSC_Viral_Neg"

Brain.combined.sct$Virus_infection[names(young.SC.virus.infected)] <- "young_SC_Viral_Pos"
Brain.combined.sct$Virus_infection[names(young.SC.virus[!colSums(young.SC.virus) == 0])] <- "young_SC_Viral_Neg"

Brain.combined.sct$Virus_infection[names(young.NSC.virus.infected)] <- "young_NSC_Viral_Pos"
Brain.combined.sct$Virus_infection[names(young.NSC.virus[!colSums(young.NSC.virus) == 0])] <- "young_NSC_Viral_Neg"

## All Microbiome ##
old.SC.Mb.diversity <- apply(old.SC.Mb, 2, diversity)
old.NSC.Mb.diversity <- apply(old.NSC.Mb, 2, diversity)
t.test(old.SC.Mb.diversity, old.NSC.Mb.diversity)

young.SC.Mb.diversity <- apply(young.SC.Mb, 2, diversity)
young.NSC.Mb.diversity <- apply(young.NSC.Mb, 2, diversity)
t.test(young.SC.Mb.diversity, young.NSC.Mb.diversity)

## All Microbiome (Exclude Unclassified reads) ##
old.SC.MbE.diversity <- apply(old.SC.MbE, 2, diversity)
old.NSC.MbE.diversity <- apply(old.NSC.MbE, 2, diversity)
t.test(old.SC.MbE.diversity,old.NSC.MbE.diversity)

young.SC.MbE.diversity <- apply(young.SC.MbE, 2, diversity)
young.NSC.MbE.diversity <- apply(young.NSC.MbE, 2, diversity)
t.test(young.SC.MbE.diversity, young.NSC.MbE.diversity)

# All Microbiome (Exclude Unclassified reads and VIRUS) #
old.SC.MbEV.diversity <- apply(old.SC.MbEV, 2, diversity)
old.NSC.MbEV.diversity <- apply(old.NSC.MbEV, 2, diversity)
t.test(old.SC.MbEV.diversity, old.NSC.MbEV.diversity)

young.SC.MbEV.diversity <- apply(young.SC.MbEV, 2, diversity)
young.NSC.MbEV.diversity <- apply(young.NSC.MbEV, 2, diversity)
t.test(young.SC.MbEV.diversity, young.NSC.MbEV.diversity)

## reads count RATIO analysis ##

## VIRUS ##

# viral infected cells
old.SC.virus.infected.ratio <-
colSums(old.SC.virus.infected)/colSums(old[, names(old.SC.virus.infected)])
old.NSC.virus.infected.ratio <-
colSums(old.NSC.virus.infected)/colSums(old[, names(old.NSC.virus.infected)])
t.test(old.SC.virus.infected.ratio, old.NSC.virus.infected.ratio)

young.SC.virus.infected.ratio <-
colSums(young.SC.virus.infected)/colSums(young[, names(young.SC.virus.infected)])
young.NSC.virus.infected.ratio <-
colSums(young.NSC.virus.infected)/colSums(young[, names(young.NSC.virus.infected)])
t.test(young.SC.virus.infected.ratio, young.NSC.virus.infected.ratio)

# comparison plot, t-test

violin_plot_comparison <- function(a,b,plotname){
  c1 <- c(rep(deparse(substitute(a)), length(a)),
         rep(deparse(substitute(b)), length(b)))
  c2 <- c(a,b)
  dat <- data.frame(group=c1, count.ratio=c2)
  p <- ggplot(dat, aes(group, count.ratio)) +
      geom_violin(aes(fill = group), trim = F) +
      geom_boxplot(aes(fill = group), width = 0.1, show.legend = F) +
      stat_boxplot(geom = 'errorbar', width = 0.1, color = 'black') +
      geom_jitter(alpha = 0.2) +
      theme_bw() + scale_fill_brewer(palette = 'Set2') +
      ggpubr::stat_compare_means(method = "t.test",
                               comparisons = list(c(deparse(substitute(a)),
                                                   deparse(substitute(b))))),
      label = 'p.signif', label.y = 0.00095)
  ggsave(paste0(deparse(substitute(plotname)), ".pdf"), width = 8, height = 8)
}
violin_plot_comparison(old.SC.virus.infected.ratio,old.NSC.virus.infected.ratio,old_count ratio_comparison)
violin_plot_comparison(young.SC.virus.infected.ratio,young.NSC.virus.infected.ratio,young _countratio_comparison)

## All Microbiome ##
old.SC.Mb.ratio = colSums(old.SC.Mb)/colSums(old[,names(old.SC.Mb)])
old.NSC.Mb.ratio = colSums(old.NSC.Mb)/colSums(old[,names(old.NSC.Mb)])
t.test(old.SC.Mb.ratio,old.NSC.Mb.ratio)

young.SC.Mb.ratio = colSums(young.SC.Mb)/colSums(young[,names(young.SC.Mb)])
young.NSC.Mb.ratio = colSums(young.NSC.Mb)/colSums(young[,names(young.NSC.Mb)])
t.test(young.SC.Mb.ratio,young.NSC.Mb.ratio)

## All Microbiome (Exclude Unclassified reads) ##
old.SC.MbE = old.SC.Mb[grep("(taxid 0)"),row.names(old.SC.Mb),invert=T],
old.NSC.MbE = old.NSC.Mb[grep("(taxid 0)"),row.names(old.NSC.Mb),invert=T],
old.SC.MbE.ratio = colSums(old.SC.MbE)/colSums(old[,names(old.SC.MbE)])
old.NSC.MbE.ratio = colSums(old.NSC.MbE)/colSums(old[,names(old.NSC.MbE)])
t.test(old.SC.MbE.ratio,old.NSC.MbE.ratio)

young.SC.MbE = young.SC.Mb[grep("(taxid 0)"),row.names(young.SC.Mb),invert=T],
young.NSC.MbE = young.NSC.Mb[grep("(taxid 0)"),row.names(young.NSC.Mb),invert=T],
young.SC.MbE.ratio = colSums(young.SC.MbE)/colSums(young[,names(young.NSC.MbE)])
young.NSC.MbE.ratio = colSums(young.NSC.MbE)/colSums(young[,names(young.NSC.MbE)])
t.test(young.SC.MbE.ratio,young.NSC.MbE.ratio)

# All Microbiome (Exclude Unclassified reads and VIRUS) #
old.SC.MbEV = old.SC.MbE[grep("virus"),row.names(old.SC.MbE),invert=T],
old.NSC.MbEV = old.NSC.MbE[grep("virus"),row.names(old.NSC.MbE),invert=T],
old.SC.MbEV.ratio = colSums(old.SC.MbEV)/colSums(old[,names(old.SC.MbEV)])
old.NSC.MbEV.ratio = colSums(old.NSC.MbEV)/colSums(old[,names(old.NSC.MbEV)])
t.test(old.SC.MbEV.ratio,old.NSC.MbEV.ratio)

young.SC.MbEV = young.SC.MbE[grep("virus"),row.names(young.SC.MbE),invert=T],
young.NSC.MbEV = young.NSC.MbE[grep("virus"),row.names(young.NSC.MbE),invert=T],
young.SC.MbEV.ratio = colSums(young.SC.MbEV)/colSums(young[,names(young.SC.MbEV)])
young.NSC.MbEV.ratio = colSums(young.NSC.MbEV)/colSums(young[,names(young.NSC.MbEV)])
t.test(young.SC.MbEV.ratio,young.NSC.MbEV.ratio)

### Identify differently expressed Microbime ###
## Extract Microbiome part of Seurat object
Brain.combined.sct.Mb = as.data.frame(Brain.combined.sct[["SCT"]][@data[grep("taxid",row.names(Brain.combined.sct[ ["SCT"])[@data]),]])
Brain.combined.sct.Mb = CreateSeuratObject(Brain.combined.sct.Mb)
Brain.combined.sct.Mb <-
AddMetaData(Brain.combined.sct.Mb,Brain.combined.sct@meta.data[,,-2:5])
Idents(Brain.combined.sct.Mb) <- "group"
# Identify differently expressed Microbime
# find markers for every cluster compared to all remaining cells, report only the positive ones
All.markers.Mb <- FindAllMarkers(Brain.combined.sct.Mb, test.use="MAST", min.pct = 0,
logfc.threshold = 0.1)
All.markers.Mb %>% group_by(cluster) %>% top_n(n = 2, wt = avg_log2FC)
write.table(All.markers.Mb,"All_markers_Mb.tsv",quote=F, sep="\t",col.names = NA)
# comparison between two groups
Old_SCvsNonSC.markers <- FindMarkers(Brain.combined.sct.Mb, test.use="MAST", ident.1 = "Old_SC", ident.2 = "Old_NonSC", min.pct=0, logfc.threshold=0.1)
Young_SCvsNonSC.markers <- FindMarkers(Brain.combined.sct.Mb, test.use="MAST", ident.1 = "Young_SC", ident.2 = "Young_NonSC", min.pct=0, logfc.threshold=0.1)
write.table(Old_SCvsNonSC.markers,"Old_SCvsNonSC_markers.tsv",quote=F, sep="\t",col.names = NA)
write.table(Young_SCvsNonSC.markers,"Young_SCvsNonSC_markers.tsv",quote=F, sep="\t",col.names = NA)
# Plot heatmap of All Markers
# re-arrange the dataframe
lt<-
for (i in unique(All.markers.Mb$cluster)){
  t1<-
    All.markers.Mb[All.markers.Mb$cluster==i & All.markers.Mb$p_val_adj<0.05,]
    t1<-
    t1[c("gene","avg_log2FC")]
    names(t1)[2] <- i
    lt[[i]] <- t1
}
lt.merge <- purrr::reduce(lt,full_join, by = "gene")
row.names(lt.merge)<-lt.merge$gene
lt.merge <- lt.merge[-1]
lt.merge[is.na(lt.merge)] <- 0
#heatmap require matrix
lt.merge.matrix <- as.matrix(lt.merge)
# draw heat map
library(pheatmap)
pheatmap(lt.merge.matrix, cluster_cols = T, scale="none")
ggsave("heatmap.pdf")
# adjust the rowname length
for (i in 1:length(row.names(lt.merge.matrix))){
  l<-nchar(row.names(lt.merge.matrix)[i])
  if (l > 35){
    row.names(lt.merge.matrix)[i]<-str_trunc(row.names(lt.merge.matrix)[i],35)
  }
}
# draw heat map
library(pheatmap)

hp_thumbnail <- pheatmap(transdata, cluster_cols = T, scale = "row",
annotation_col = Anno_col, annotation_colors = anno_colors,
show_rownames = F, cex = 1)

if (nrow(transdata) > 100) {
  hp <- pheatmap(transdata, cluster_cols = T, scale = "row",
                 fontsize_row = 5, annotation_col = Anno_col,
                 annotation_colors = anno_colors)
} else {
  hp <- pheatmap(transdata, cluster_cols = T, scale = "row",
                 annotation_col = Anno_col,
                 annotation_colors = anno_colors)
}

# Correlation test of microbiome and host gene among top 3,000 features
featured.microbiome <- subset(as.data.frame(Brain.combined.sct@assays[["SCT"]][@data]), row.names(Brain.combined.sct@assays[["SCT"]][@data] %in% grep("taxid", features, value = T)))
featured.host <- subset(as.data.frame(Brain.combined.sct@assays[["SCT"]][@data]), row.names(Brain.combined.sct@assays[["SCT"]][@data] %in% grep("taxid", features, value = T, invert = T))
featured.microbiomeXhost <- cor(t(featured.microbiome), t(featured.host))

write.table(featured.microbiome, "featured.microbiome.txt", sep = "\t", quote = F, col.names = NA)
write.table(featured.host, "featured.host.txt", sep = "\t", quote = F, col.names = NA)
write.table(featured.microbiomeXhost, "featured.microbiomeXhost.txt", sep = "\t", quote = F, col.names = NA)

library(corrplot)
color_adj <- rev(c("#67001F", "#B2182B", "#D6604D", "#F4A582", "#FDDBC7",
                     "#FFFFFF", "#D1E5F0", "#92C5DE", "#4393C3", "#2166AC", "#053061"))
corrplot(featured.microbiomeXhost,
          tl.col = "black", tl.srt = 45, tl.cex = 0.6,
          col = colorRampPalette(color_adj)(200))
corrplot(featured.microbiomeXhost)

### parallel correlation test ###
library(parallel)
library(doParallel)
cores <- 20
options('mc.cores' = cores)
registerDoParallel(cores)
p.corr <- function(testmatrix1, testmatrix2)
  x <- foreach (j = 1:ncol(testmatrix1),
    .combine = rbind,
    .multicombine = TRUE,
    .inorder = FALSE,
    .packages = c('data.table', 'doParallel')) %>%
foreach (i = 1:ncol(testmatrix2),
    .combine = cbind,
    .multicombine = TRUE,
    .inorder = FALSE,
    .packages = c('data.table', 'doParallel')) %dopar% {
    a <- cor.test(testmatrix1[,j], testmatrix2[,i], method = "spearman")
    a$p.value
} 

colnames(x)<-colnames(testmatrix2)
row.names(x)<-colnames(testmatrix1)
return(x)

# show the heatmap

p.mat<-p.corr(t(featured.host),t(featured.microbiome))

pdf("featured.microbiomeXhost.pdf",30,900)
corrplot(t(featured.microbiomeXhost), method="color",
    tl.col = "black", tl.srt = 45,tl.cex=0.6,
    col = colorRampPalette(color_adj)(200),
    addCoef.col = "black", number.cex=0.25,
    p.mat=p.mat,insig = "blank",diag=F)

dev.off()

# vertical

pdf("featured.microbiomeXhost.pdf",30,900)
corrplot(t(featured.microbiomeXhost), method="color",
    tl.col = "black", tl.srt = 45,tl.cex=0.6,
    col = colorRampPalette(color_adj)(200),
    addCoef.col = "black", number.cex=0.25,
    p.mat=p.mat,diag=F)

dev.off()

### acquire the significant correleated microbiome and host genes ###
sig.corr<-data.frame()
for (i in 1:nrow(p.mat)){
    for (j in 1:ncol(p.mat)){
        if (p.mat[i,j]<0.05 & & t(featured.microbiomeXhost)[i,j]>0.3){
            sig.corr <-
            rbind(sig.corr,data.frame(Microbiome=colnames(t(featured.microbiomeXhost))[j],
                Host_gene=rownames(t(featured.microbiomeXhost))[i],
                r=t(featured.microbiomeXhost)[i,j],
                p=p.mat[i,j]))
        }
    }
}
write.table(sig.corr, "MicrobiomeXhost_sigCorrelation.tsv", row.names = F, sep="\t",quote = F)

# show the most highest correlations
FeaturePlot(object = Brain.combined.sct, features = c("GRIA2", "Plasmodium vivax (taxid 5855)"), blend = TRUE)

### Host's gene differential expression analysis ###
## extract host's gene expression by age and senescent cells
Ids(Brain.combined.sct) <- "group"
Old_SCvsNonSC.MAST <- FindMarkers(Brain.combined.sct, ident.1 = "Old_SC", ident.2 = "Old_NonSC", test.use = "MAST")
Young_SCvsNonSC.MAST <- FindMarkers(Brain.combined.sct, ident.1 = "Young_SC", ident.2 = "Young_NonSC", test.use = "MAST")
write.table(Old_SCvsNonSC.MAST, "Old_SCvsNonSC.host.MAST.tsv", sep = "\t", quote = F, col.names = NA)
write.table(Young_SCvsNonSC.MAST, "Young_SCvsNonSC.host.MAST.tsv", sep = "\t", quote = F, col.names = NA)

## extract host's gene expression by age and senescent cells and viral_infection_status
Ids(Brain.combined.sct) <- "Virus_infection"
Old_Viral_Pos_SCvsNonSC.MAST <- FindMarkers(Brain.combined.sct, ident.1 = "Old_SC_Viral_Pos", ident.2 = "Old_NS_Viral_Pos", test.use = "MAST")
Young_Viral_Pos_SCvsNonSC.MAST <- FindMarkers(Brain.combined.sct, ident.1 = "young_SC_Viral_Pos", ident.2 = "young_NS_Viral_Pos", test.use = "MAST")
write.table(Old_Viral_Pos_SCvsNonSC.MAST, "Old_Viral_Pos_SCvsNonSC.MAST.tsv", sep = "\t", quote = F, col.names = NA)
write.table(Young_Viral_Pos_SCvsNonSC.MAST, "Young_Viral_Pos_SCvsNonSC.MAST.tsv", sep = "\t", quote = F, col.names = NA)

Old_SC_Viral_PosvsNeg.MAST <- FindMarkers(Brain.combined.sct, ident.1 = "Old_SC_Viral_Pos", ident.2 = "Old_SC_Viral_Neg", test.use = "MAST")
Old_NS_Viral_PosvsNeg.MAST <- FindMarkers(Brain.combined.sct, ident.1 = "Old_NS_Viral_Pos", ident.2 = "Old_NS_Viral_Neg", test.use = "MAST")
Young_SC_Viral_PosvsNeg.MAST <- FindMarkers(Brain.combined.sct, ident.1 = "young_SC_Viral_Pos", ident.2 = "young_SC_Viral_Neg", test.use = "MAST")
Young_NS_Viral_PosvsNeg.MAST <- FindMarkers(Brain.combined.sct, ident.1 = "young_NS_Viral_Pos", ident.2 = "young_NS_Viral_Neg", test.use = "MAST")
write.table(Old_SC_Viral_PosvsNeg.MAST, "Old_SC_Viral_PosvsNeg.MAST.tsv", sep = "\t", quote = F, col.names = NA)
write.table(Old_NS_Viral_PosvsNeg.MAST, "Old_NS_Viral_PosvsNeg.MAST.tsv", sep = "\t", quote = F, col.names = NA)
write.table(Young_SC_Viral_PosvsNeg.MAST, "Young_SC_Viral_PosvsNeg.MAST.tsv", sep = "\t", quote = F, col.names = NA)
write.table(Young_NS_Viral_PosvsNeg.MAST, "Young_NS_Viral_PosvsNeg.MAST.tsv", sep = "\t", quote = F, col.names = NA)

### Import cell type scored results ###
Cells.scored.old <- read.delim("/master/fwu/Dual-Scseq/test/ssGSEA/GSE161340_combined/Old/BrainCell-results-scores.gct", skip=2, row.names=1)
Cells.scored.old <- Cells.scored.old[names(old)]
Cells.scored.young <- read.delim("/master/fwu/Dual-Scseq/test/ssGSEA/GSE161340_combined/Young/BrainCell-results-scores.gct", skip=2, row.names=1)
Cells.scored.young <- Cells.scored.young[names(young)]

Cells.scored <- merge(Cells.scored.old, Cells.scored.young, by="row.names")
row.names(Cells.scored) <- Cells.scored[,1]
Cells.scored <- Cells.scored[,-1]
Brain.combined.sct <- AddMetaData(Brain.combined.sct, t(Cells.scored), row.names(Cells.scored))

### correlation test ###
# correlation of cell type and senescent cells

ggscatter(Brain.combined.sct@meta.data[,Brain.combined.sct$age=="24 months"], x = "Senescence_core_genes", y = "Microglia",
    add = "reg.line", conf.int = TRUE,
    cor.coef = TRUE, cor.method = "pearson",
    xlab = "Senescence_core_genes", ylab = "Microglia", title = "Old_mice_brain")

ggscatter(Brain.combined.sct@meta.data[,Brain.combined.sct$age=="4 months"], x = "Senescence_core_genes", y = "Microglia",
    add = "reg.line", conf.int = TRUE,
    cor.coef = TRUE, cor.method = "pearson",
    xlab = "Senescence_core_genes", ylab = "Microglia", title = "Young_mice_brain")

# all cell types correlations

Old.all_celltypes <-
data.frame(senescent_score=Brain.combined.sct$Senescence_core_genes[,Brain.combined.sct$age=="24 months"], t(Cells.scored.old))
Young.all_celltypes <-
data.frame(senescent_score=Brain.combined.sct$Senescence_core_genes[,Brain.combined.sct$age=="4 months"], t(Cells.scored.young))
cell.corr <- function(all_celltypes){
  res <- cor(all_celltypes)
  library(corrplot)
  color_adj <- rev(c("#67001F", "#821B2B", "#D6604D", "#F4A582", "#FDDBC7",
                     "#FFFFFF", "#D1E5F0", "#92C5DE", "#4393C3", "#2166AC", "#053061"))
  # corrplot(res, type = "upper", order = "hclust",
  # tl.col = "black", tl.srt = 45, tl.cex=0.6,
  # col = colorRampPalette(color_adj)(200))
  # to optimized the correlation plot:
  # mat : is a matrix of data
  # ... : further arguments to pass to the native R cor.test function
  cor.mtest <- function(mat, ...) {
    #
mat <- as.matrix(mat)
n <- ncol(mat)
p.mat <- matrix(NA, n, n)
diag(p.mat) <- 0
for (i in 1:(n - 1)) {
  for (j in (i + 1):n) {
    tmp <- cor.test(mat[, i], mat[, j], ...)
  }
}
colnames(p.mat) <- rownames(p.mat) <- colnames(mat)
p.mat

# matrix of the p-value of the correlation
p.matc <- cor.mtest(all_celltypes)
# LabelCol<-rep("black",ncol(res)) # try to highlight same labels
# LabelCol[c(1,11)]<"red"
pdf(paste0(deparse(substitute(all_celltypes)), ".pdf"), 8, 8)
corrplot(res, type = "upper", order = "hclust", method="color",
  tl.col = "black", tl.srt = 45,tl.cex=0.6,
  col = colorRampPalette(color_adj)(200),
  addCoef.col = "black", number.cex=0.25,
  p.mat=p.matc,insig = "blank",diag=F)
dev.off()
}
cell.corr(Old.all_celltypes)
cell.corr(Young.all_celltypes)

# extra test virus.ratio. X cell types
Brain.combined.sct$virus.ratio <-
c(young.SC.virus.infected.ratio,young.NSC.virus.infected.ratio,
old.SC.virus.infected.ratio,old.NSC.virus.infected.ratio)
Brain.combined.sct$virus.ratio[is.na(Brain.combined.sct$virus.ratio)] <- 0
Young.all_celltypes.virus<- data.frame(virus.ratio=Brain.combined.sct$virus.ratio[Brain.combined.sct$age="4 months"],Young.all_celltypes)
Old.all_celltypes.virus<- data.frame(virus.ratio=Brain.combined.sct$virus.ratio[Brain.combined.sct$age="24 months"],Old.all_celltypes)
cell.corr(Young.all_celltypes.virus)
cell.corr(Old.all_celltypes.virus)
pie(table(Brain.combined.sct@meta.data[Brain.combined.sct$virus.ratio!=0,"cell_type"]))
pie(table(Brain.combined.sct@meta.data,"cell_type"))

# frequency test for all cell types in both young and old animals
virus.celltype.Freq1 <- merge(data.frame(table(Brain.combined.sct@meta.data[Brain.combined.sct$virus.ratio!=0,"cell_type"])),
                      data.frame(table(Brain.combined.sct@meta.data["cell_type"])),
                      by = "Var1")

virus.celltype.Freq2 <- data.frame(Cell_type=virus.celltype.Freq1$Var1,
                        Viral_pos=virus.celltype.Freq1$Freq.x/sum(virus.celltype.Freq1$Freq.x),
                        All_cells=virus.celltype.Freq1$Freq.y/sum(virus.celltype.Freq1$Freq.y))

link_data <- virus.celltype.Freq2 %>%
  arrange(by=desc(Cell_type)) %>%
  mutate(Viral_pos=cumsum(Viral_pos),All_cells=cumsum(All_cells))

df.long <- virus.celltype.Freq2 %>%
  gather(group,abundance,Cell_type) %>%
  mutate(ratio=round(abundance,2))

df.long$ratio[df.long$ratio < 0.04]<-"

ggplot(df.long, aes(x=group, y=abundance, fill=Cell_type)) +
  geom_bar(stat = "identity", width=0.5, col="black") +
  geom_segment(data=link_data, aes(x=1.25, xend=1.75, y=All_cells, yend=Viral_pos)) +
  geom_text(aes(label= ratio), size = 3, position = position_stack(vjust = 0.5)) +
  theme_bw() +
  labs(x="Groups", y="Proportion", title="Young mice brain")

# frequency test for all cell types in young or old animals
virus.celltype.Freq1 <- merge(data.frame(table(Brain.combined.sct@meta.data[Brain.combined.sct$age=="4 months" &
Brain.combined.sct$virus.ratio!=0,"cell_type"])),
                      data.frame(table(Brain.combined.sct@meta.data[Brain.combined.sct$age=="4 months","cell_type"])),
                      by = "Var1")

virus.celltype.Freq2 <- data.frame(Cell_type=virus.celltype.Freq1$Var1,
                        Viral_pos=virus.celltype.Freq1$Freq.x/sum(virus.celltype.Freq1$Freq.x),
                        All_cells=virus.celltype.Freq1$Freq.y/sum(virus.celltype.Freq1$Freq.y))

link_data <- virus.celltype.Freq2 %>%
  arrange(by=desc(Cell_type)) %>%
  mutate(Viral_pos=cumsum(Viral_pos),All_cells=cumsum(All_cells))

df.long <- virus.celltype.Freq2 %>%
  gather(group,abundance,Cell_type) %>%
  mutate(ratio=round(abundance,2))

df.long$ratio[df.long$ratio < 0.04]<-"

ggplot(df.long, aes(x=group, y=abundance, fill=Cell_type)) +
  geom_bar(stat = "identity", width=0.5, col="black") +
  geom_segment(data=link_data, aes(x=1.25, xend=1.75, y=All_cells, yend=Viral_pos)) +
  geom_text(aes(label= ratio), size = 3, position = position_stack(vjust = 0.5)) +
  theme_bw() +
  labs(x="Groups", y="Proportion", title="Young mice brain")
ggsave("Celltype_stacked_barplot_young.pdf")
# for old animals:
virus.celltype.Freq1<-merge(data.frame(table(Brain.combined.sct@meta.data[Brain.combined.sct$age=="24 months"&Brain.combined.sct$virus.ratio!=0,"cell_type"])),
data.frame(table(Brain.combined.sct@meta.data[Brain.combined.sct$age=="24 months","cell_type"])), by = "Var1")
virus.celltype.Freq2<-data.frame(Cell_type=virus.celltype.Freq1$Var1,
Viral_pos=virus.celltype.Freq1$Freq.x/sum(virus.celltype.Freq1$Freq.x),
All_cells=virus.celltype.Freq1$Freq.y/sum(virus.celltype.Freq1$Freq.y))
link_data<-virus.celltype.Freq2 %>%
  arrange(by=desc(Cell_type)) %>%
  mutate(Viral_pos=cumsum(Viral_pos),All_cells=cumsum(All_cells))
df.long<-virus.celltype.Freq2 %>%
  gather(group,abundance,-Cell_type)
df.long$ratio<-round(df.long$abundance,2)
df.long$ratio[df.long$ratio < 0.04]<=""
ggplot(df.long, aes(x=group, y=abundance, fill=Cell_type)) +
geom_bar(stat = "identity", width=0.5, col="black") +
geom_segment(data=link_data, aes(x=1.25, xend=1.75, y=All_cells, yend=Viral_pos)) +
geom_text(aes(label=ratio), size = 3, position = position_stack(vjust = 0.5)) +
theme_bw() +
labs(x="Groups", y="Proportion", title="Old mice brain")
ggsave("Celltype_stacked_barplot_old.pdf")

# stock cell type ratios
Celltype_ratio.young<-df.long
Celltype_ratio.old<-df.long
# remove the unaligned line manually, then calculate cell type change:
Celltype_ratio.OldvsYoung<-data.frame(Cell_type=Celltype_ratio.old$Cell_type,
group=Celltype_ratio.old$group,
Fold_change=Celltype_ratio.old$abundance/Celltype_ratio.young$abundance)
Celltype_ratio.OldvsYoung$log2FC<-log2(Celltype_ratio.OldvsYoung$Fold_change)
ggplot(Celltype_ratio.OldvsYoung, aes(Cell_type,log2FC,fill=group)) +
geom_bar(position = "dodge",stat="identity") +
scale_fill_manual(values=c("#999999", "#E69F00")) +
ggtitle("Old vs Young") +
geom_text(aes(label=round(log2FC,2)),size=2.2,
  position=position_dodge(0.9),vjust=-0.5) +
theme_bw() +
theme(plot.title = element_text(hjust = 0.5),
  axis.text.x = element_text(angle = 45,hjust=1))
ggsave("Celltype_ratio.OldvsYoung_barplot.pdf")
### correlation test ###

# Senescent cell X virus #

# extract all virus counts in both senescent and non-senescent cells

old.virus <- old[grep("virus", row.names(old)),
young.virus <- young[grep("virus", row.names(young)),

#extract virus positive cells

old.viral.pos <- old.virus[colSums(old.virus)!=0]
young.viral.pos <- young.virus[colSums(young.virus)!=0]

# virus abundance in viral positive cells: ratio of raw counts

virus.ratio.cell <- function(viral.pos, df, Brain.combined.sct){
  virus.ratio.cell <- colSums(viral.pos)/colSums(df[,colnames(viral.pos)])
  Brain.combined.sct$Senescence_core_genes[names(virus.ratio.cell)]
  data.frame(Viral_count_ratio=virus.ratio.cell, Sc_score=virus.ratio.cell.Sc.score)
}

old.virus.ratio.cell <- virus.ratio.cell(old.viral.pos, old, Brain.combined.sct)
young.virus.ratio.cell <- virus.ratio.cell(young.viral.pos, young, Brain.combined.sct)

library("ggpubr")
ggscatter(old.virus.ratio.cell, x = "Viral_count_ratio", y = "Sc_score",
  add = "reg.line", conf.int = TRUE,
  cor.coef = TRUE, cor.method = "pearson",
  xlab = "Virus_count_ratio", ylab = "Senescent_score", title = "Old_mice_brain")
ggscatter(young.virus.ratio.cell, x = "Viral_count_ratio", y = "Sc_score",
  add = "reg.line", conf.int = TRUE,
  cor.coef = TRUE, cor.method = "pearson",
  xlab = "Virus_count_ratio", ylab = "Senescent_score", title = "Young_mice_brain")

### Mark cell types of clusters ###

cores <- 20
options(mc.cores = cores)
registerDoParallel(cores)
cells <- foreach (c = 1:ncol(Cells.scored)) %dopar% {
  c(cells, rownames(Cells.scored[Cells.scored[c]==max(Cells.scored[c]),]))
}
cells <- unlist(cells)
Brain.combined.sct$cell_type <- cells

Idents(Brain.combined.sct) <- "seurat_clusters"
ident.cell <- c()
for (l in levels(Brain.combined.sct)){
  table.cell <-
  table(Brain.combined.sct@meta.data[Brain.combined.sct$seurat_clusters==l,"cell_type"])
  ident.cell <- c(ident.cell, l, table.cell)
ident.cell<-c(ident.cell,names(table.cell[order(-table.cell)]))[1])
}

names(ident.cell) <- levels(Brain.combined.sct)

# Make Idents to the new cell labels
Brain.combined.sct$cell_type_clusters <- Brain.combined.sct$seurat_clusters
for (i in names(ident.cell)){
  Brain.combined.sct$cell_type_clusters <-
gsub(paste0("\",i,"\"",ident.cell[i],Brain.combined.sct$cell_type_clusters)
}
#Brain.combined.sct$cell_type_clusters <-
gsub(names(ident.cell),ident.cell,Brain.combined.sct$cell_type_clusters)

# Switch Idents to the new cell labels
#Idsents(Brain.combined.sct)<-Brain.combined.sct$cell_type_clusters
Idsents(Brain.combined.sct) <- "cell_type"

DimPlot(Brain.combined.sct) + DarkTheme() + NoAxes()
ggsave("umap_mice_brain_labeled_dark.pdf",width=12,height=10)

DimPlot(Brain.combined.sct, reduction = "umap")
ggsave("umap_mice_brain_split_labeled.pdf",width=10,height=5)

DimPlot(Brain.combined.sct, reduction = "umap", label = TRUE, repel = TRUE)
DimPlot(Brain.combined.sct, reduction = "umap", label = TRUE, pt.size=0.5) + NoLegend()
prop.table(table(Idsents(Brain.combined.sct)))

# Switch Idents to the senescent cell labels
Idsents(Brain.combined.sct)<-Brain.combined.sct$Senescent_Status

DimPlot(Brain.combined.sct, reduction = "umap", split.by = "age",cols=c("#999999","#d95f02"))
ggsave("SenescentCells_umap_mice_brain_split_labeled.pdf",width=10,height=5)
#DimPlot(Brain.combined.sct, reduction = "umap", split.by = "age",order="NonSC",cells.highlight=names(Brain.combined.sct$Senescent_Status[Brain.combined.sct$Senescent_Status=="SC"]))

# Ridge plot:
library(ggplot)
Celltype_RidgePlot<ggplot(Brain.combined.sct@meta.data, aes(x = Senescence_core_genes, y= cell_type_clusters, fill= age)) +
geom_density_ridges(alpha=0.25) +
geom_vline(aes(xintercept=3),
  color="black", linetype="dashed", size=0.5) +
ggtitle("Mice brain") +
xlab("Senescence Core Genes Enrichment Score") + ylab("") +
scale_fill_discrete(name="",
  breaks=c("24 months","4 months"),
  labels=c("aged", "young")) +
theme_bw() +
theme(plot.title = element_text(hjust = 0.5))
ggsave("Celltye_RidgePlot.pdf")

# Density plot:
SenescentCell_DensityPlot<- ggplot(Brain.combined.sct@meta.data, aes(x = Senescence_core_genes, group = age, fill = age)) +
  geom_density(alpha = 0.25) +
  geom_vline(aes(xintercept = 3),
              color = "black", linetype = "dashed", size = 0.5) +
  ggtitle("Mice brain") +
  xlab("Senescence Core Genes Enrichment Score") + ylab("Density") +
  scale_fill_discrete(name = "",
                      breaks = c("24 months", "4 months"),
                      labels = c("aged", "young")) +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5))
ggsave("SenescentCell_DensityPlot.pdf")

# Pie Chart of senescent cell percentages:
#install.packages("ggstatsplot")
df.SC.status<- read.delim("/Users/feiwu/Downloads/senescent_status.txt")
library(ggplot2)
ggstatsplot::ggpiestats(data = df.SC.status, x = Var1, counts = Freq,
                        bf.message = F, results.subtitle = F,
                        title = "Mice brain") +
  scale_fill_manual(breaks = c("SC", "NonSC"),
                   labels = c("Senescent cells", "Non-senescent cells"),
                   values = c("#d95f02", "#999999")) +
  theme(legend.title = element_blank()) +
  theme(plot.title = element_text(hjust = 0.5))

# just check cell types results (e.g. cluster 7)
Idents(Brain.combined.sct) <- Brain.combined.sct$seurat_clusters
View(Brain.combined.sct@meta.data[Brain.combined.sct$seurat_clusters == 7,])

FeaturePlot(object = Brain.combined.sct, features = c("CDKN2A"))
FeaturePlot(object = Brain.combined.sct, features =
            (Brain.combined.sct$Senescent_Status == "SC"))

# check cells in a cluster
which(Idents(Brain.combined.sct) == 8)
Idents(Brain.combined.sct)[Idents(Brain.combined.sct) == 8] # alternative to above

# check markers in a cluster (top10)
Brain.markers %>% group_by(cluster == 8) %>% top_n(n = 10, wt = avg_log2FC)
head(Brain.markers[Brain.markers$cluster == 8, 1:10])
BIOGRAPHY

Fei Wu was born in Jiangsu province, China, in 1988 and grew up in Suzhou metropolitan area. He has had a dream to be a scientist since early childhood. After studying the biology course in the first year of high school, he became interested in aging research and decided to pursue a career in this field. Therefore, he went to receive his degree in Bachelor of Science in Biotechnology from the Huaiyin Normal University. Then he started his aging study formally in the Institute for Ageing, Newcastle University, and received a degree in Master of Science in Medical Sciences. After that, he joined QIAGEN (Suzhou) as the first employee in the R&D department to help set up the laboratory and develop in vitro diagnostic technologies. He was promoted from assistant scientist to scientist during the three and half years of work. Meanwhile, he kept track of advances in aging research and maintained a wide range of scientific interests. In 2016, he was accepted by the Tulane Interdisciplinary Ph.D. Program in Aging Studies. Fei entered Dr. Ling’s Lab and conducted research on the relationships of SIV infection with cellular senescence and abnormal behaviors in rhesus macaques, and developed Point-of-Care Testing methods and a bioinformatics pipeline for metatranscriptome analysis. He was awarded second place in the Tulane Novel Tech Challenge. In addition, he received the Changemaker Catalyst Award, New Investigator Scholarship award of CROI 2020 conference, and did an oral presentation on 38th Annual Symposium on Nonhuman Primate Models for AIDS. He will continue to pursue his dream of being a scientist in aging studies. He hopes his work leads to a better understanding of the aging process and contributes to extending the healthy life span of human beings.